Neural-Network-Based Model-Free Calibration Method for Stereo Fisheye Camera

Yuwei Cao[1], Hui Wang[1,2], Han Zhao[3]* and Xu Yang[1]

[1]School of Automation, Wuhan University of Technology, Wuhan, China, [2]Key Laboratory of Icing and Anti/De-icing, China Aerodynamics Research and Development Center, Mianyang, China, [3]School of Integrated Chinese and Western Medicine, Anhui University of Chinese Medicine, Hefei, China

The fisheye camera has a field of view (FOV) of over 180˚, which has advantages in the fields of medicine and precision measurement. Ordinary pinhole models have difficulty in fitting the severe barrel distortion of the fisheye camera. Therefore, it is necessary to apply a nonlinear geometric model to model this distortion in measurement applications, while the process is computationally complex. To solve the problem, this paper proposes a model-free stereo calibration method for binocular fisheye camera based on neural-network. The neural-network can implicitly describe the nonlinear mapping relationship between image and spatial coordinates in the scene. We use a feature extraction method based on three-step phase-shift method. Compared with the conventional stereo calibration of fisheye cameras, our method does not require image correction and matching. The spatial coordinates of the points in the common field of view of binocular fisheye camera can all be calculated by the generalized fitting capability of the neural-network. Our method preserves the advantage of the broad field of view of the fisheye camera. The experimental results show that our method is more suitable for fisheye cameras with significant distortion.

Keywords: fisheye camera, stereo calibration, phase unwrapping, neural-network, large field of view

1 INTRODUCTION

The ordinary camera has a limited FOV. It can no longer meet the needs of some research projects without adding other auxiliary facilities. The appearance of the fisheye camera overcomes this shortcoming. The fisheye camera has a small focal length, and the field of view can generally reach more than 180˚ (Arfaoui and Thibault, 2013). Due to the large field of view of the fisheye camera, one single fisheye image can show a large portion of the surrounding environment without image splicing (Hou et al., 2012).

The research on stereo calibration technology of fisheye cameras is more meaningful. Compared with ordinary camera, the structure of fisheye camera is more complicated. Fisheye cameras introduce severe distortion, especially barrel distortion, during imaging (Kanatani, 2013). This strong optical distortion results in high image center separation and low resolution at the edges of fisheye images (Hughes et al., 2010). Consequently, the stereo calibration accuracy of the fisheye camera is also limited to some extent.

The traditional stereo calibration method for fisheye cameras requires an imaging model in a specific mathematical format. Before this paper, there were some studies on stereo calibration techniques for fisheye cameras. For example, first combined fisheye camera calibration and epipolar

rectification applied in a stereo fisheye camera system. They accomplished 3D reconstruction of specific points from authentic fisheye images Abraham and Forstner (2005). Designed a novel measurement system based on a binocular fisheye camera. The measurement system uses the dynamic angle compensation method, which can achieve high-precision 3D positioning in a dynamic environment Cai et al. (2019). Proposed a new strategy for computing parallax maps from hemispherical stereo images taken by fisheye camera. They considered only matches of the same class by segmenting the textures in the scene Herrera et al. (2011). Presented a method to calibrate multiple fisheye cameras with a wand that can move freely. The internal and external parameters and the 3D coordinates of the fisheye camera could be obtained Fu et al. (2015). Proposed a panoramic stereoscopic imaging system, which could provide stereoscopic vision of 360° horizontal field Li and Li (2011). Analyzed existing dense stereo systems. They combined the epipolar rectification model of the binocular fisheye camera with the dense method, able to provide dense 3D point clouds at 6–7 Hz Schneider et al. (2016). These stereo calibration methods usually require correcting fisheye images with significant distortion to perspective projection images. However, this distortion removal process leads to the loss of information at the image edges, losing the advantage of the large field of view of the fisheye camera. So the results of performing stereo matching on fisheye images are unsatisfactory. In addition, stereo matching also has strict restrictions on the scene. Some factors such as too much scene noise pollution and too much repetitive texture may impact the matching accuracy.

With the development of artificial intelligence, Deep Learning (DL) is increasingly used in the field of computer vision (Huang et al., 2021). There have been many studies applying DL to the distortion correction of fisheye images. Proposed a Distortion Rectification Generative Adversarial Network (DR-GAN) for the severe barrel distortion of wide-angle camera images. DR-GAN is the first end-to-end trainable adversarial framework for radial distortion correction Liao et al. (2020). Considered the characteristics of fisheye images and proposed an unsupervised fisheye camera distortion correction network. The network can predict distortion parameters and implement direct mapping from fisheye images to rectified images Yang et al. (2020). DL-based methods are computationally fast. However, they trained the network with a large number of fisheye images, which consumes a lot of resources. In addition, this method is very sensitive to the scene.

To overcome these shortcomings, we propose the application of neural-network to the stereo calibration of binocular fisheye camera. Take the image coordinates of the left and right fisheye cameras as the input training set. The spatial coordinates corresponding to the image coordinates in the scene are used as the output training set. The trained neural-network can implicitly describe the mapping relationship from the 2D image plane to the 3D space. With the nonlinear fitting ability of the neural-network, it can directly predict the spatial coordinates of the target point based on the trained network. Compared with traditional stereo calibration, the proposed method is model-free. There is no need to establish an accurate mathematical imaging model, nor does it need to know the intrinsic and extrinsic parameters of the fisheye camera. Experiments have been conducted, and their results verify the performance of the proposed method.

To obtain the training set of the neural-network, a large number of feature points with known image coordinates and spatial coordinates are required. Some 2D targets such as chessboard are the most commonly used. The chessboard-based calibration method has good calibration accuracy for ordinary cameras (Zhang, 2000). However, the chessboard images taken by the fisheye camera have severe barrel distortion, which leads to low feature detection accuracy or failure to detect feature points located at the edge of the images. To overcome this shortage, active targets are used (Schmalz et al., 2011). Active phase targets are widely used in optical measurement due to their high accuracy and high speed (Wang et al., 2011; Xu et al., 2017; Wang et al., 2020; Wang et al., 2022). This paper uses a feature extraction method based on three-step phase-shift method and a multi-frequency method (Wang et al., 2019). This feature extraction method has high precision and strong robustness (Schmalz et al., 2011). Therefore, it is more suitable for fisheye cameras with severe distortion.

The remained parts of the paper are as follows. **Section 2** describes the fisheye camera model and the stereo calibration model of the binocular fisheye camera. **Section 3** presents the training process of the neural-network, the setting of the neural-network parameters, and the acquisition of the training sets. **Section 4** describes the experiment. Finally, **Section 5** concludes as well as some prospects for the future.

2 PRINCIPLE

2.1 Single Fisheye Camera Model

Fisheye cameras take non-similar imaging and introduce large barrel distortion in the imaging process. Compressing the diameter space breaks the limitation of the imaging field of view to achieve wide-angle imaging (Wei et al., 2012). The projection refraction angle and the incident angle of fisheye cameras are not equal and will deviate from the direction of the optical axis for refraction. There are four basic imaging models of fisheye cameras: equidistant projection model, equisolid-angle projection model, orthographic projection model, and stereographic projection model (Schneider et al., 2009).

The projection equation for equidistant projection is shown as: **Eq. 1**

$$r_d = f\theta \tag{1}$$

where r_d is the radial distance; f is the focal length of the fisheye camera; θ is the angle of incidence of the light.

The projection equation for equisolid-angle projection is shown as: **Eq. 2**
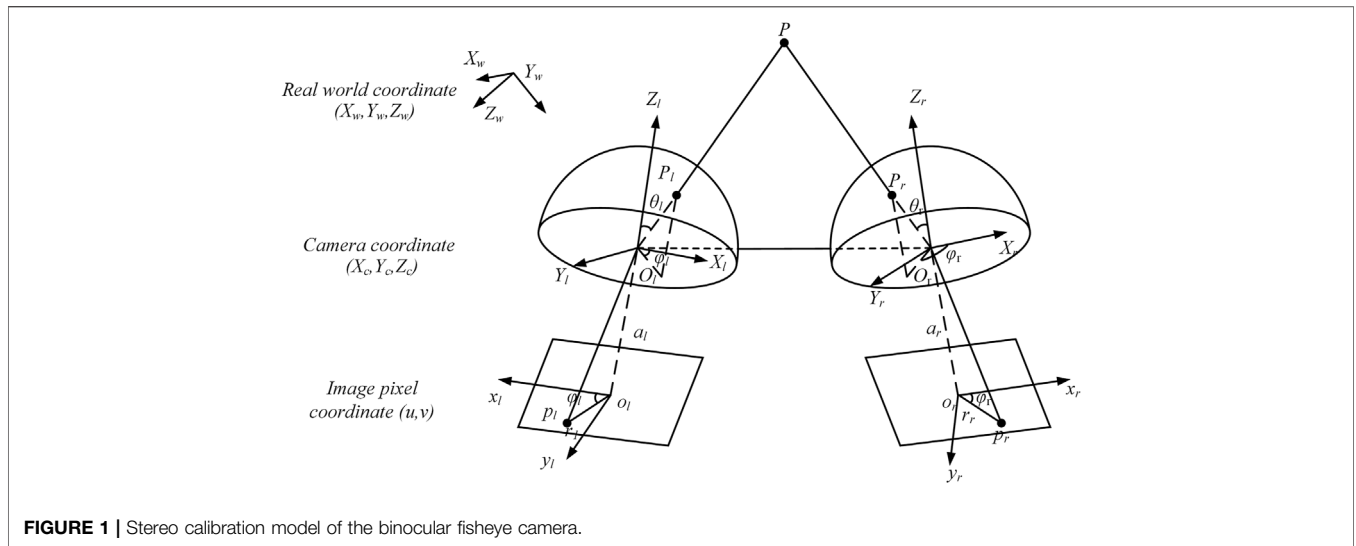
$$r_d = 2f \sin\left(\frac{\theta}{2}\right) \tag{2}$$

FIGURE 1 | Stereo calibration model of the binocular fisheye camera.

The projection equation for orthographic projection is shown as: **Eq. 3**

$$r_d = f\sin\theta \qquad (3)$$

The projection equation for stereographic projection is shown as: **Eq. 4**

$$r_d = 2f\tan\left(\frac{\theta}{2}\right) \qquad (4)$$

Traditional distortion models cannot guarantee the accuracy of fisheye camera. Several models have been developed to represent the distortion of the fisheye camera, including the polynomial model, the field-of-view (FOV) model (Devernay and Faugeras, 2001), and the fisheye transform (Base, 1995).

2.2 Stereo Calibration Model of the Binocular Fisheye Camera

Figure 1 shows the stereo calibration model of the binocular fisheye camera. The fisheye camera coordinate system on the left is denoted by $O_l - X_lY_lZ_l$, and the fisheye camera coordinate system on the right is denoted by $O_r - X_rY_rZ_r$. Since the fisheye camera imaging is nonlinear, the camera coordinate system is denoted by unit spherical coordinates. The world coordinate system is denoted by $O_w - X_wY_wZ_w$. The relative positions of the left and right cameras are fixed, and their relationship can be expressed as:

$$P_l = R_lP_w + T_l \qquad (5)$$
$$P_r = R_rP_w + T_r \qquad (6)$$

where R_l and R_r represent the rotation vectors corresponding to the world coordinate system and the left and right fisheye camera coordinate systems, respectively; T_l and T_r represent the translation vectors; P_w represents the world coordinate of any point P ; P_l and P_r respectively represent the coordinate of the
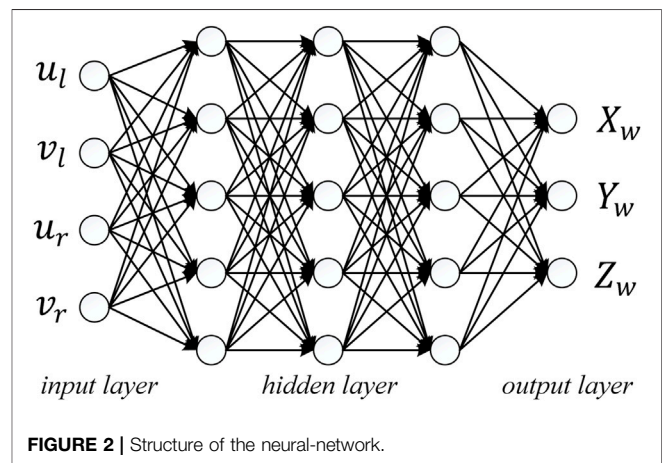

FIGURE 2 | Structure of the neural-network.

point P in the left and right camera coordinates. Combining **Eqs 5, 6**, we can obtain the spatial position conversion relationship between the left and right camera coordinate systems: **Eq. 7**

$$P_r = R_rR_l^{-1}P_l + T_r - R_rR_l^{-1}T_l \qquad (7)$$

Therefore, the rotation vector of the left fisheye camera to the right fisheye camera is $R = R_rR_l^{-1}$, and the translation vector is $T = T_r - R_rR_l^{-1}T_l$.

The solution R of and T is essentially the process of stereo calibration (Li, 2008; Beekmans et al., 2016).

3 NEURAL-NETWORK MODEL

3.1 Training Process of Neural-Network
In recent years, the emergence of some bio-inspired algorithms that simulate natural ecosystems provide new ideas for solving complex optimization problems. These bio-inspired algorithms include genetic algorithms (Liu X et al., 2022), particle swarm algorithms

(Liu Y et al., 2022; Wu et al., 2022; Zhao et al., 2022), predictive modeling algorithms (Chen et al., 2021a; Chen et al., 2021b; Chen et al., 2022), convolution neural network algorithm (Huang et al., 2022; Tao et al., 2022; Yun et al., 2022), artificial bee colony algorithm (Sun et al., 2022), etc. A neural-network is a multi-layer feed forward network that follows an error back propagation algorithm, as shown in **Figure 2**. The basic component units of a neural-network are neurons, also called network nodes. The essence of each neuron is a nonlinear transformation of the input data. Theoretically, a neural-network can accomplish any form of nonlinear mapping (Parma et al., 1999). A neural-network can provide a nonlinear model $h_{w,b}(x)$ to fit the output y. The network parameters are the weights w and the bias b. Training a neural-network is to continuously update these two parameters under the stimulus of external inputs so that the output keeps approaching the desired output. The training process consists of a forward propagation process of the input information and a backward propagation process of the error information.

The forward propagation process can be understood as follows: the output of the previous layer is used as the input of the next layer, and the output of the next layer is calculated until the operation reaches the output layer. Let the activation value of the ith node in the lth layer of the neural-network be a_i^l. The jth node in the $(l-1)^{th}$ layer to the ith node in layer l^{th} node is connected by the weight w_{ij}^l, b_i^l is the bias of the ith node in layer l^{th}. It is not difficult to see from the structure of the neural-network that a_i^l depends on the activation of the previous layer.

$$a_i^l = f\left(\sum w_{ij}^l a_i^{l-1} + b_i^l \right) \tag{8}$$

where $f()$ is the activation function. In this paper, the activation function uses the tanh function. Expressing **Eq. 8** in matrix form:

$$\begin{cases} z^l = W^l a^{l-1} + b^l \\ a_i^l = f(z^l) \end{cases} \tag{9}$$

where z^l is the input of each layer. Use **Eq. 9** to calculate the activation value of the network layer by layer. Finally, the output of the network can be obtained.

Before explaining the back propagation algorithm, it is first necessary to define the loss function. The loss function can measure the loss between the output computed by the training samples and the actual output.

The purpose of the back propagation process is to adjust the network parameters. Its essence is to find the optimal weights and biases by minimizing the loss function. So it is necessary to calculate the partial derivatives of the loss function to the weights and biases. The gradient of the variables in each layer of the neural-network can be obtained by finding the partial derivatives. The stochastic gradient descent algorithm (SGD) is commonly used to update the network parameters. The SGD algorithm can be summarized as: **Eq. 10**

$$\begin{cases} w_{ij}^l = w_{ij}^l - \eta \dfrac{\partial}{\partial w_{ij}^l} J() \\[2em] b_i^l = b_i^l - \eta \dfrac{\partial}{\partial b_i^l} J() \end{cases} \tag{10}$$

where η is the learning rate; $J()$ is the loss function. After the network parameters are updated, we determine whether the current model meets the requirements. If the requirements are not met, the following forward and backward propagation is performed. The network parameters continue to be updated. Until the current model meets the requirements, the neural-network training is completed.

3.2 Setting of Neural-Network Parameters
This paper uses a neural-network to implicitly describe the nonlinear mapping relationship between image coordinates and their corresponding spatial coordinates in the scene. The settings of neural-network parameters include the structure of the neural-network, loss function, activation function, and optimizer. With the training sets keep constant, different network parameters can significantly impact the convergence speed and prediction accuracy of the network.
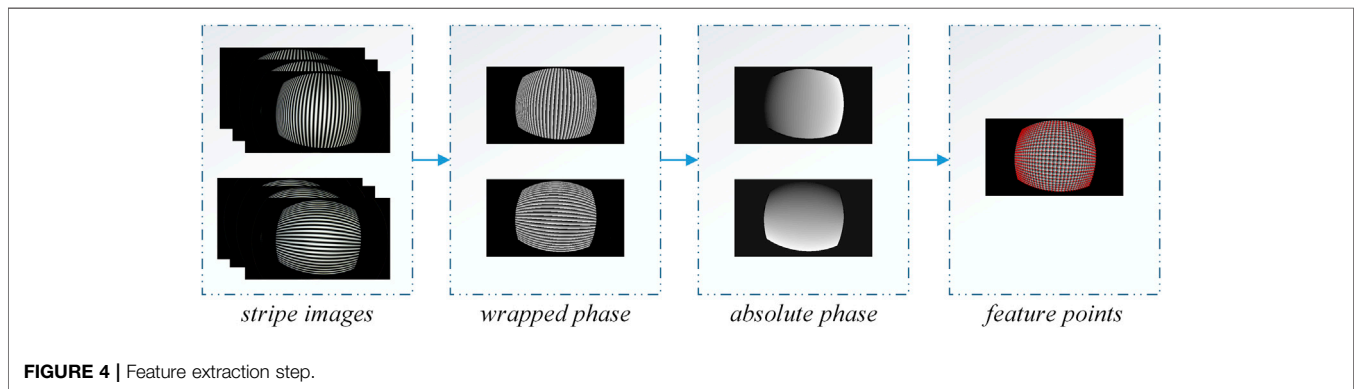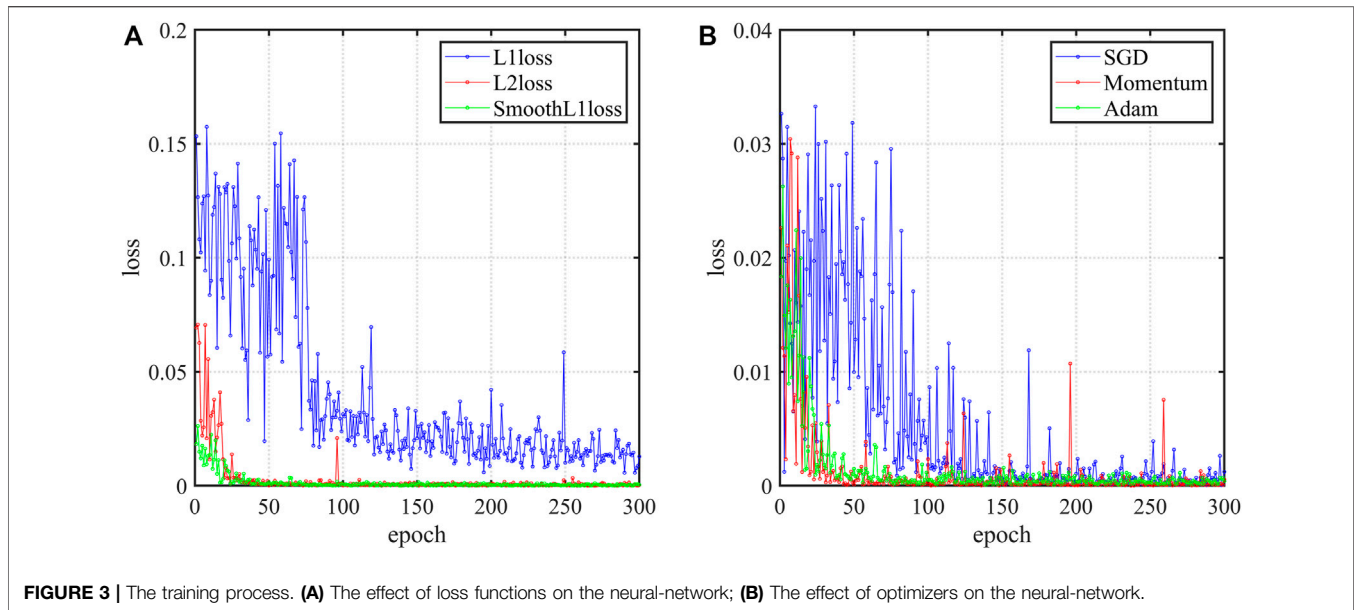
3.2.1 Structural Parameters
The neural-network structure proposed in this paper contains five layers, as shown in **Figure 2**. There is one input layer, three hidden layers, and one output layer. The input layer has four nodes. (u_l, v_l) represent the left image coordinates of feature points. (u_r, v_r) represent the right image coordinates of feature points. The output layer has three nodes. (X_w, Y_w, Z_w) represents spatial coordinates of feature points. Each hidden layer contains five nodes.

3.2.2 Loss Function
This study is essentially a regression problem. The most commonly used loss functions for regression problems are the mean square error (L2loss) and the mean absolute error (L1loss). L2loss function curve is smooth and can converge quickly to a minimum even at meager learning rates. However, when outliers exist in the training set, L2loss gives higher weight to the outliers, affecting the overall performance (Natekin and Knoll, 2013). L1loss performs sluggishly for outliers but converges slowly. So it is natural to think of the SmoothL1loss function. The SmoothL1loss function converges faster than L1loss. Compared to L2loss, it is insensitive to outliers. To further verify the effect of loss function on the neural-network, **Figure 3A** shows the training process with three different loss function settings. L1loss has the slowest convergence speed and relatively low training accuracy. In contrast, SmoothL1loss has the fastest convergence speed and the best training accuracy. Therefore, SmoothL1loss is finally chosen as the loss function.

3.2.3 Optimizers
Optimizers can optimize neural-network to improve training accuracy and save training time. The most basic optimizer is the SGD algorithm, initially introduced in the previous subsection. The SGD algorithm is computationally efficient and only requires solving the first-order derivatives of the loss function. However, The SGD algorithm makes the results fall into saddle points and find local optimal solutions because of the direction. Consequently, this paper uses an adaptive optimization algorithm that can update the learning rate

FIGURE 3 | The training process. **(A)** The effect of loss functions on the neural-network; **(B)** The effect of optimizers on the neural-network.

FIGURE 4 | Feature extraction step.

automatically. To further verify the effect of the optimizer on the neural-network, **Figure 3B** shows the training process under the three optimizer settings of SGD, SGD with momentum, and Adam. SGD has the worst optimization effect and the slowest speed. As a modified version of SGD, Momentum is much better. Adam is the best and the fastest convergence speed. So Adam optimizer is chosen.

3.3 Generation of the Training Set

The input set is (u_l, v_l), (u_r, v_r) and the output set is (X_w, Y_w, Z_w). A common practice is to use the corner points of the ordinary chessboard as feature points. This method is simple and easy to operate. However, the fisheye camera distortion is severe. The chessboard will be severely distorted at the location closer to the camera, with low or even undetectable corner point detection accuracy at the edge.

To solve the above problems, the active phase target is used. Firstly, the wrapped phase of the sinusoidal periodic stripe image is solved using the three-step phase shift equation. According to the multi-frequency method, the

phase is unwrapped to obtain the absolute phase. Finally, we select the eligible pixel points as feature points according to the absolute phase. A set of exactly matched image coordinates and spatial coordinates will be obtained if the unwrapping is successful. The feature points extracted using our method have the advantage of quantity and are minimally affected by the fisheye camera distortion. **Figure 4** shows the specific implementation flow chart, summarized as follows:

1) Generate three-frequency three-step stripe images with equal-step phase shift increments of $2\pi/3$. Their intensities can be expressed as: **Eq. 11**.

$$\begin{cases} I_1 = I'(x, y) + I''(x, y)\cos\left[\phi(x, y) - \dfrac{2\pi}{3}\right] \\ I_2 = I'(x, y) + I''(x, y)\cos\left[\phi(x, y)\right] \quad\quad (11) \\ I_3 = I'(x, y) + I''(x, y)\cos\left[\phi(x, y) + \dfrac{2\pi}{3}\right] \end{cases}$$

where I_1, I_2, and I_3 are the grayscale values of the phase diagram; $I'(x, y)$ is the background light intensity; $I''(x, y)$ is the intensity modulation parameter; $\phi(x, y)$ is the wrapped phase to be solved. The horizontal and vertical phase shift stripes are displayed sequentially on the LCD.

2) Two fisheye cameras are fixed on the overhead camera mount, and the LCD monitor is fixed on the high-precision horizontal elevator. The fisheye cameras can shoot the LCD overhead. The high-precision horizontal elevator controls the LCD to move in the Z_w direction in steps (the displacement error is 0.05 mm). Two fisheye cameras are controlled in each plane to acquire stripe images simultaneously.

3) According to **Eq. 12**, a three-step phase shift algorithm is used to calculate the two wrapped phases ϕ_u and ϕ_v of the streak image. The value domain of the Arctangent function is within $[-\pi, \pi]$. So if the streak image with more than one period is used for decoding, the calculated wrapped phase is discontinuous. Therefore, the wrapped phases are unwrapped using the multi-frequency method to obtain the continuous absolute phases Φ_u and Φ_v.

$$\phi(u, v) = \tan^{-1}\left(\sqrt{3} \ \frac{I_1 - I_3}{2I_2 - I_1 - I_3}\right) \qquad (12)$$

4) Any point on the stripe image, calculate its absolute phase Φ_u and Φ_v. Some alternative feature points can be extracted if they satisfy the following relationships: **Eq. 13**.

$$\begin{cases} |\Phi_u - 2\pi n| < \tau \\ |\Phi_v - 2\pi m| < \tau \end{cases} \qquad (13)$$

where τ is an artificially set threshold; m and n are integers. Then, among these alternative feature points, the coordinates such that $|\Phi_u - 2\pi n| + |\Phi_v - 2\pi m|$ obtains the minimum value are searched for as the desired feature points. Finally, least-squares linear interpolation is used to optimize the feature points to the sub-pixel level.

5) The absolute phase is converted to spatial coordinates for each feature point with the following equation: **Eq. 14**.

$$\begin{bmatrix} X_w \\ Y_w \end{bmatrix} = \frac{qP}{2\pi} \begin{bmatrix} \Phi_u \\ \Phi_v \end{bmatrix} \qquad (14)$$

where P represents the number of pixels in the stripe cycle; q represents the pixel size of the LCD. Use the reading of the high-precision horizontal elevator as the Z_w coordinate of the feature point. In this paper, the fisheye camera has a large field of view. The field of view can cover the whole LCD screen even at a position very close to the camera. So the spatial coordinates of the feature points determined by the left and right cameras are the same.

After determining the input and output data sets, the image and spatial coordinates have different value ranges. So it is necessary to normalize the data. Normalization can improve the convergence speed of the neural-network and the model's accuracy. We use the polar difference transformation method.
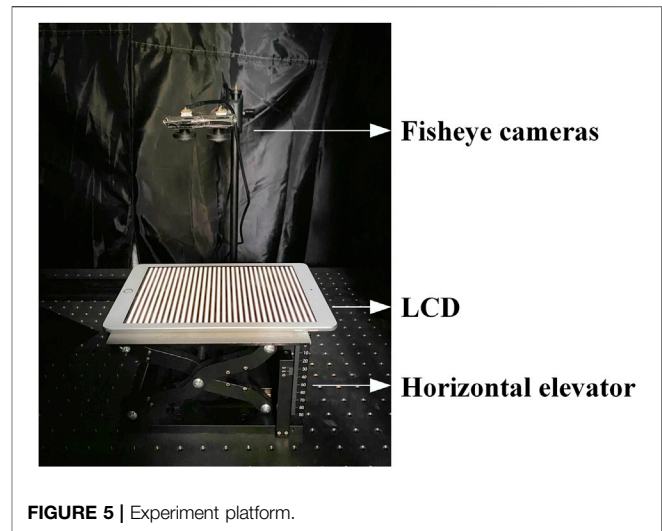

FIGURE 5 | Experiment platform.
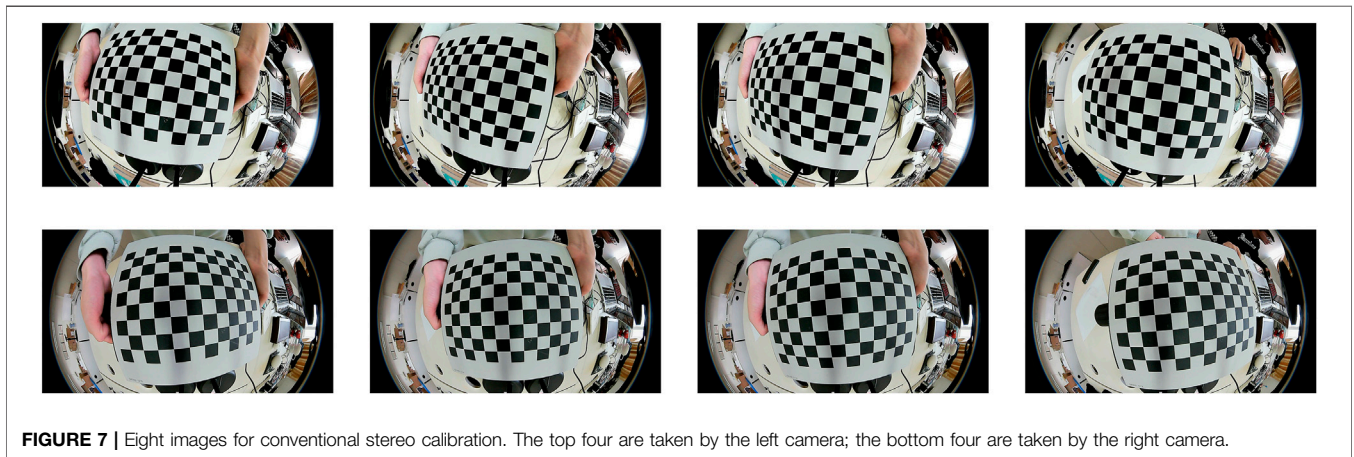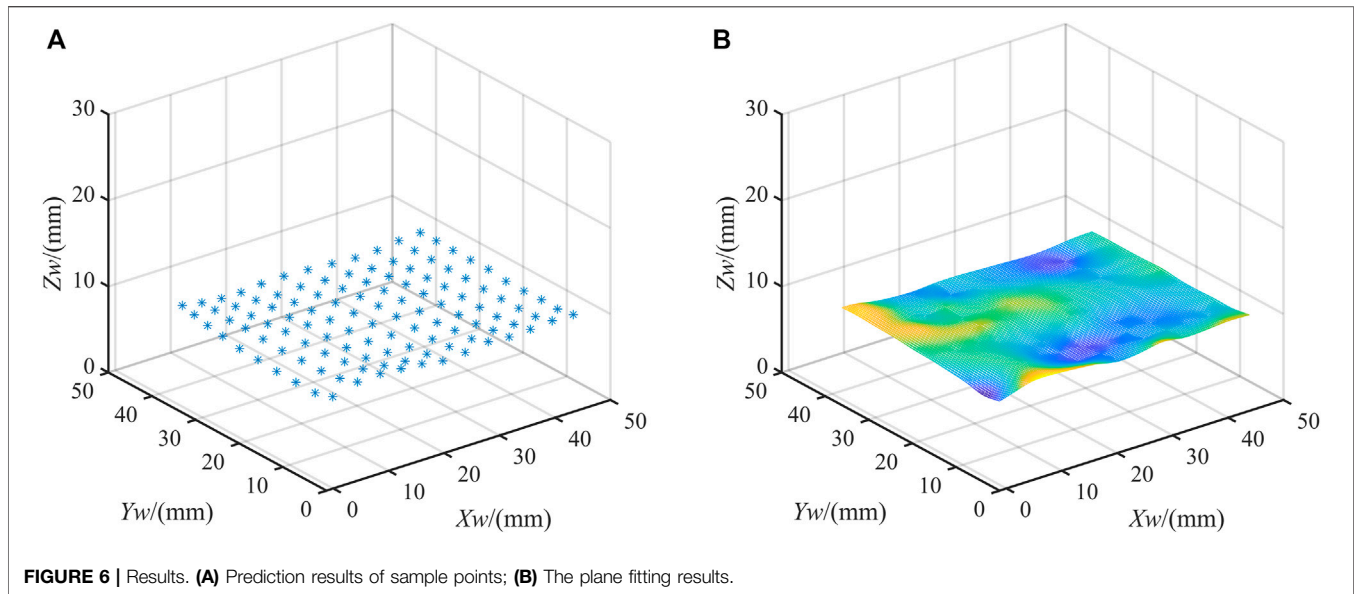
4 EXPERIMENTS

To verify the accuracy of the proposed method, an experimental platform was built. **Figure 5** shows the experiment platform. The experimental platform includes two identical cameras (AR0230AT), a high-precision horizontal elevator (HTZ210), an LCD (iPad A1893), and a chessboard calibration plate. The fisheye lens (LRCP12014_27 1/2) mounted on the camera has a focal length of 1.4 mm and a field of view of 220°. Two comparison experiments were conducted in different configurations. Finally, the trained neural-network is used to reconstruct the corner points of the chessboard and part of the surface of the sphere.

4.1 Neural-Network Method vs. Traditional Fisheye Camera Model Method

The first experiment compares the neural-network-based fisheye camera stereo calibration (the proposed method) with traditional fisheye camera stereo calibration. As shown in **Figure 5**. Two fisheye cameras are mounted on the overhead camera mount. The LCD is fixed on a high-precision horizontal elevator. The LCD is used to display the three-frequency, three-step stripe images. The high-frequency stripe period is 64, and the high, medium, and low frequencies multiplier is 6. The LCD resolution is 2048 × 1536 pixels, and the pixel size is 0.096 mm. The high-precision horizontal elevator controls the gradual movement of the LCD in the Z_w direction.

The training set is obtained by following the steps described in **Section 3.3**. The neural-network is configured according to **Section 3.2**.

Based on the trained network, the sample data can be predicted. **Figure 6A** shows the prediction results of 120 sample points. The actual values of the spatial coordinates are known. So we can quantitatively analyze the deviations in three directions. The mean error of X_w is 0.416 mm, Y_w is 0.253 mm, and Z_w is 0.271 mm. To visually show the prediction results, the predicted results of the spatial coordinates of the sample points are linearly interpolated. **Figure 6B** shows the fitted plane.

FIGURE 6 | Results. **(A)** Prediction results of sample points; **(B)** The plane fitting results.

FIGURE 7 | Eight images for conventional stereo calibration. The top four are taken by the left camera; the bottom four are taken by the right camera.

We perform the traditional fisheye camera stereo calibration using the fisheye camera calibration method in opencv3.0. The specific principle can be referred to (Kannala and Brandt, 2006) and is not described in detail here. This method requires two fisheye cameras to take pictures of the target in different directions. A total of 25 images were taken. **Figure 7** shows some of the 25 images.

Table 1 compares the reconstruction accuracy of the neural-network model-based method with the traditional fisheye camera model method. The experimental results show that the neural-network-based method proposed in this paper has higher accuracy and is more suitable for fisheye camera with larger distortion.

4.2 Active Phase Target vs. Chessboard

The second experiment compares two different methods of obtaining the training set for the neural-network. One is to use active phase targets as proposed in this paper, and the

TABLE 1 | Mean errors in Z_w.

Stereo calibration methods	Mean error in (mm) Z_w
Neural-network model	0.271
Fisheye camera model	3.967

other is to use a chessboard as the target. The experimental procedure using the active phase target has been described in **Section 4.1**.

The experimental chessboard contains 88 corner points with a spacing of 15 mm and a manufacturing error of 0.01 mm. The Harris corner point detection algorithm can obtain the sub-pixel image coordinates of the chessboard corner points. The corner points of the chessboard are used as feature points. To ensure the consistency of the experimental conditions, the positions of the fisheye
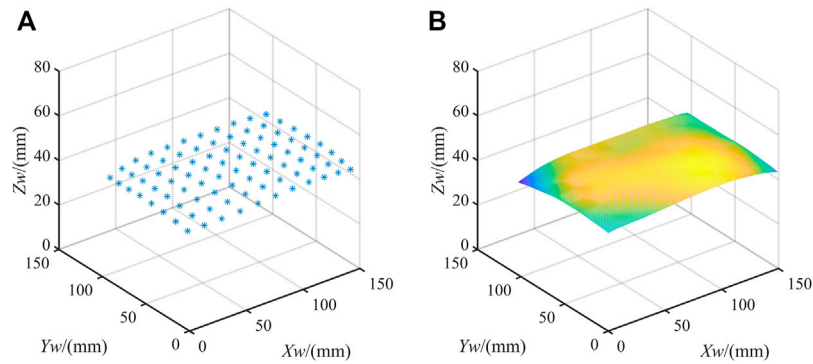
FIGURE 8 | Results. **(A)** Prediction results of sample points; **(B)** The plane fitting results.

TABLE 2 | Mean errors in X_w, Y_w, and Z_w.

Stereo calibration methods	Mean error in (mm) X_w	Mean error in (mm) Y_w	Mean error in (mm) Z_w
Active phase targets	0.416	0.253	0.271
Chessboard	1.105	0.894	1.177
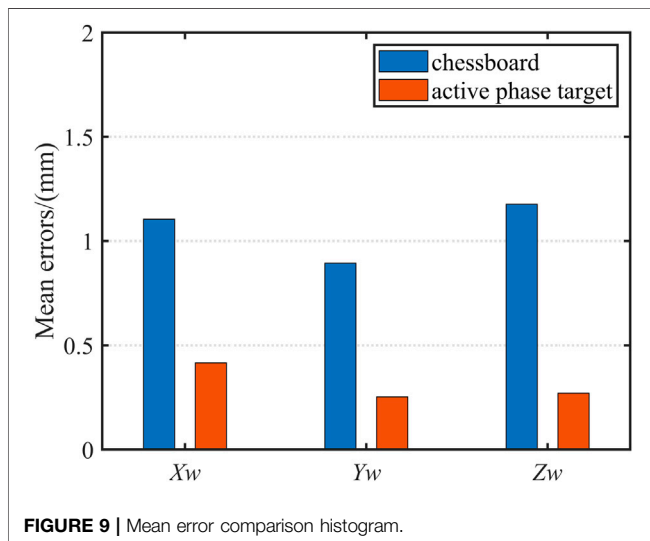
FIGURE 9 | Mean error comparison histogram.

cameras are not changed. The chessboard is fixed on the high-precision horizontal elevator. The high-precision horizontal elevator controls the chessboard to move in the Z_w direction in steps. The readings of the high-precision horizontal elevator are used as the Z_w coordinates of the feature points. The neural-network parameter settings are not changed.

The sample data are then predicted based on the trained network model. **Figure 8A** shows the results. The actual values of the spatial coordinates of these points are known. So we can quantitatively analyze the deviation in three directions. The mean error of X_w is 1.105 mm, Y_w is 0.894 mm, and Z_w is 1.177 mm. To demonstrate the experimental results more

intuitively, linear interpolation is performed on the predicted results. **Figure 8B** shows the fitted plane.

Table 2 compares the reconstruction accuracy comparison of the training set obtained using the active phase target and the chessboard. **Figure 9** shows the mean error comparison graph. It is clear that the method using the active phase target to extract feature points as the training set is more accurate, especially in the Z_w direction. The experimental results prove that the active phase target has the advantage of the number of feature points and is more suitable for the calibration of the fisheye camera.

4.3 3D Reconstruction

To further verify the practicability of the proposed method, 3D reconstructions of the chessboard corners and a partial plane of the sphere are performed.

The experimental chessboard contains 88 corner points with a spacing of 4.9 mm. The binocular fisheye camera takes pictures of the chessboard in different poses at the same time. The subpixel image coordinates of the chessboard corners are obtained using the Harris corner detection algorithm. The spatial coordinates of these corners are then reconstructed using the trained neural-network model. **Figure 10** shows the reconstruction results. We calculate the square size of the chessboard based on the spatial coordinates and compare it with the true value. Among them, the reconstruction error of the corners located at the edge of the chessrboard is larger, and the reconstruction errors of the middle corners is smaller. This is due to the characteristics of the fisheye image itself. The edges of the fisheye image are stretched due to the severe distortion of the camera. **Table 3** shows the mean error for the chessboard square size.

Similarly, we reconstructed a partial plane of the sphere. We recover the absolute phase of the sphere by projecting a fringe

FIGURE 10 | Chessboard cornors reconstruction results.
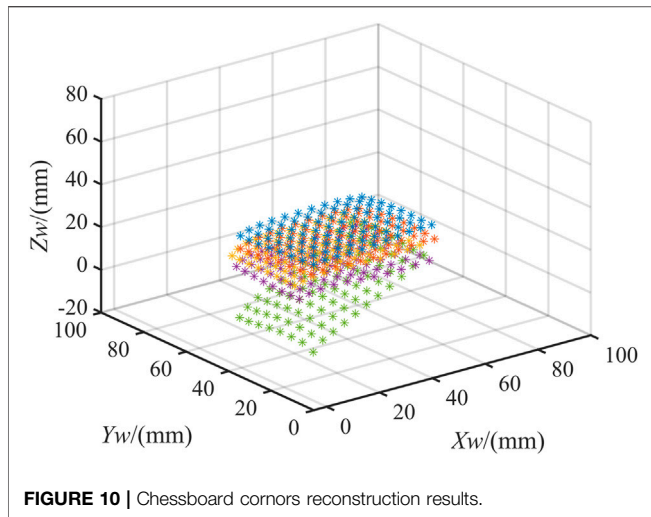
TABLE 3 | Mean errors in square size.

Pose	1 (mm)	2 (mm)	3 (mm)	4 (mm)	5 (mm)
Mean errors	0.174	0.066	0.159	0.142	0.345

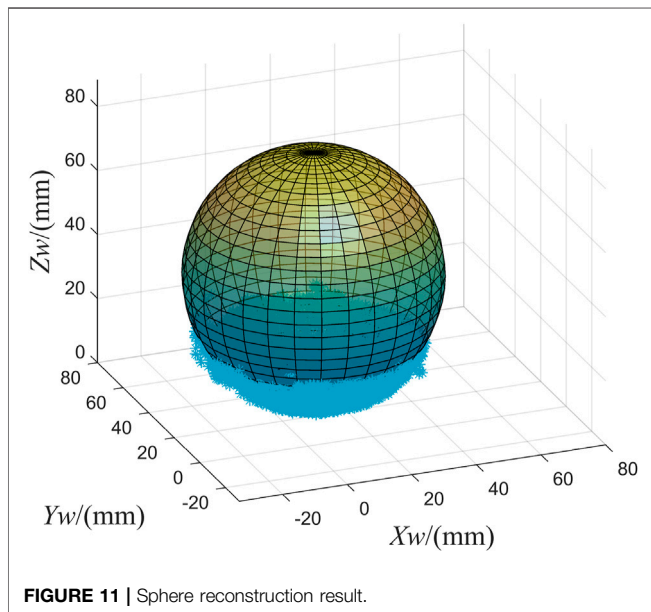FIGURE 11 | Sphere reconstruction result.

image on the upper surface of the sphere. According to the matching relationship between the absolute phases of the sphere in the left and right cameras, the pixel points are matched (Chen X. et al., 2021). Then we use the trained neural-network to predict the spatial coordinates of these

points. **Figure 11** shows the reconstruction results. We performed a least squares fit to the results for the sphere. The real diameter of the sphere is 71 mm. The fitted diameter is 73.7288 mm. So the reconstruction error is 2.7288 mm. Experiments show that the neural-network-based method proposed in this paper has high measurement accuracy.

5 CONCLUSION

This paper applies a neural-network to the fisheye camera stereo calibration technique. There is no need to pre-build the fisheye camera model. The proposed method is model-free. A nonlinear mapping relationship between image coordinates and spatial coordinates is established using neural-network. The use of the active phase target enables the extraction of feature points with a larger number and higher precision, which is more suitable for the calibration of fisheye cameras. Due to the flexible structure of the neural-network, the neural-network model can be easily extended to the joint calibration of multiple fisheye cameras and the joint calibration of asymmetric fisheye camera layouts. These are expected to be further investigated and implemented in the future.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

YC and HZ conceived the overall research goals as well as the experimental protocol. YC completed the validation of the experimental design, the development of related software, and the writing of the first draft. HZ participated in and guided the entire experimental process, and provided constructive comments on the research plan. XY and HW provided the experimental equipment and organized the experimental data. All authors contributed to the revision of the manuscript.

FUNDING

REFERENCES

Abraham, S., and Förstner, W. (2005). Fish-eye-stereo Calibration and Epipolar Rectification. *Isprs J. Photogrammetry Remote Sens.* 59 (5), 278–288. doi:10.1016/j.isprsjprs.2005.03.001

Arfaoui, A., and Thibault, S. (2013). Fisheye Lens Calibration Using Virtual Grid. *Appl. Opt.* 52 (12), 2577–2583. doi:10.1364/ao.52.002577

Base, A. J. P. R. L. (1995). Alternative Models for Fish-Eye Lenses. *Pattern Recognit. Lett.* 16 (4), 433–441.

Beekmans, C., Schneider, J., Läbe, T., Lennefer, M., Stachniss, C., and Simmer, C. (2016). Cloud Photogrammetry with Dense Stereo for Fisheye Cameras. *Atmos. Chem. Phys.* 16 (22), 14231–14248. doi:10.5194/acp-16-14231-2016

Cai, C., Qiao, R., Meng, H., and Wang, F. (2019). A Novel Measurement System Based on Binocular Fisheye Vision and its Application in Dynamic Environment. *Ieee Access* 7, 156443–156451. doi:10.1109/access.2019.2949172

Chen, T., Peng, L., Yang, J., Cong, G., and Li, G. (2021a). Evolutionary Game of Multi-Subjects in Live Streaming and Governance Strategies Based on Social Preference Theory during the COVID-19 Pandemic. *Mathematics* 9 (21), 2743. doi:10.3390/math9212743

Chen, T., Rong, J., Yang, J., and Cong, G. (2022). Modeling Rumor Diffusion Process with the Consideration of Individual Heterogeneity: Take the Imported Food Safety Issue as an Example During the COVID-19 Pandemic. *Front. Public Health* 10, 781691. doi:10.3389/fpubh.2022.781691

Chen, T., Yin, X., Yang, J., Cong, G., and Li, G. (2021b). Modeling Multi-Dimensional Public Opinion Process Based on Complex Network Dynamics Model in the Context of Derived Topics. *Axioms* 10 (4), 270. doi:10.3390/axioms10040270

Chen, X., Chen, Y., Song, X., Liang, W., and Wang, Y. (2021c). Calibration of Stereo Cameras with a Marked-Crossed Fringe Pattern. *Opt. Lasers Eng.* 147, 106733. doi:10.1016/j.optlaseng.2021.106733

Devernay, F., and Faugeras, O. (2001). Straight Lines Have to Be Straight. *Mach. Vis. Appl.* 13 (1), 14–24. doi:10.1007/pl00013269

Forster, F., Forster, F., and Angelopoulou, E. (2011). Camera Calibration: Active versus Passive Targets. *Opt. Eng.* 50 (11), 113601. doi:10.1117/1.3643726

Fu, Q., Quan, Q., and Cai, K. Y. (2015). Calibration of Multiple Fish-eye Cameras Using a Wand. *Iet Comput. Vis.* 9 (3), 378–389. doi:10.1049/iet-cvi.2014.0181

Herrera, P. J., Pajares, G., Guijarro, M., Ruz, J. J., and Cruz, J. M. (2011). A Stereovision Matching Strategy for Images Captured with Fish-Eye Lenses in Forest Environments. *Sensors* 11 (2), 1756–1783. doi:10.3390/s110201756

Hou, W., Ding, M., Qin, N., and Lai, X. (2012). Digital Deformation Model for Fisheye Image Rectification. *Opt. Express* 20 (20), 22252–22261. doi:10.1364/oe.20.022252

Huang, L., Chen, C., Yun, J., Sun, Y., Tian, J., Hao, Z., et al. (2022). Multi-Scale Feature Fusion Convolutional Neural Network for Indoor Small Target Detection. *Front. Neurorobot.* 16, 881021. doi:10.3389/fnbot.2022.881021

Huang, L., Fu, Q., He, M., Jiang, D., and Hao, Z. (2021). Detection Algorithm of Safety Helmet Wearing Based on Deep Learning. *Concurr. Comput. Pract. Exper.* 33 (13). doi:10.1002/cpe.6234

Hughes, C., Denny, P., Jones, E., and Glavin, M. (2010). Accuracy of Fish-Eye Lens Models. *Appl. Opt.* 49 (17), 3338–3347. doi:10.1364/ao.49.003338

Kanatani, K. (2013). Calibration of Ultrawide Fisheye Lens Cameras by Eigenvalue Minimization. *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (4), 813–822. doi:10.1109/tpami.2012.146

Kannala, J., and Brandt, S. S. (2006). A Generic Camera Model and Calibration Method for Conventional, Wide-Angle, and Fish-Eye Lenses. *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (8), 1335–1340. doi:10.1109/tpami.2006.153

Li, W., and Li, Y. F. (2011). Single-camera Panoramic Stereo Imaging System with a Fisheye Lens and a Convex Mirror. *Opt. Express* 19 (7), 5855–5867. doi:10.1364/oe.19.005855

Liao, K., Lin, C., Zhao, Y., and Gabbouj, M. (2020). DR-GAN: Automatic Radial Distortion Rectification Using Conditional GAN in Real-Time. *IEEE Trans. Circuits Syst. Video Technol.* 30 (3), 725–733. doi:10.1109/tcsvt.2019.2897984

Liu, X., Jiang, D., Tao, B., Jiang, G., Sun, Y., Kong, J., et al. (2022). Genetic Algorithm-Based Trajectory Optimization for Digital Twin Robots. *Front. Bioeng. Biotechnol.* 9, 793782. doi:10.3389/fbioe.2021.793782

Liu, Y., Jiang, D., Yun, J., Sun, Y., Li, C., Jiang, G., et al. (2022). Self-Tuning Control of Manipulator Positioning Based on Fuzzy PID and PSO Algorithm. *Front. Bioeng. Biotechnol.* 9, 817723. doi:10.3389/fbioe.2021.817723

Natekin, A., and Knoll, A. (2013). Gradient Boosting Machines, a Tutorial. *Front. Neurorobot.* 7, 00021. doi:10.3389/fnbot.2013.00021

Parma, G. G., Menezes, B. R. D., and Braga, A. P. (1999). Neural Networks Learning with Sliding Mode Control: the Sliding Mode Backpropagation Algorithm. *Int. J. Neur. Syst.* 09 (3), 187–193. doi:10.1142/s0129065799000174

Schneider, D., Schwalbe, E., and Maas, H.-G. (2009). Validation of Geometric Models for Fisheye Lenses. *Isprs J. Photogrammetry Remote Sens.* 64 (3), 259–266. doi:10.1016/j.isprsjprs.2009.01.001

Schneider, J., Stachniss, C., Forstner, W., and Letters, A. (2016). On the Accuracy of Dense Fisheye Stereo. *IEEE Robot. Autom. Lett.* 1 (1), 227–234. doi:10.1109/lra.2016.2516509

Shigang Li, S. G. (2008). Binocular Spherical Stereo. *IEEE Trans. Intell. Transp. Syst.* 9 (4), 589–600. doi:10.1109/tits.2008.2006736

Sun, Y., Zhao, Z., Jiang, D., Tong, X., Tao, B., Jiang, G., et al. (2022). Low-Illumination Image Enhancement Algorithm Based on Improved Multi-Scale Retinex and ABC Algorithm Optimization. *Front. Bioeng. Biotechnol.* 10, 865820. doi:10.3389/fbioe.2022.865820

Tao, B., Wang, Y., Qian, X., Tong, X., He, F., Yao, W., et al. (2022). Photoelastic Stress Field Recovery Using Deep Convolutional Neural Network. *Front. Bioeng. Biotechnol.* 10, 818112. doi:10.3389/fbioe.2022.818112

Wang, Y., Cai, J., Zhang, D., Chen, X., and Wang, Y. (2022). Nonlinear Correction for Fringe Projection Profilometry with Shifted-Phase Histogram Equalization. *IEEE Trans. Instrum. Meas.* 71, 1–9. doi:10.1109/tim.2022.3145561

Wang, Y., Liu, L., Cai, B., Wang, K., Chen, X., Wang, Y., et al. (2019). Stereo Calibration with Absolute Phase Target. *Opt. Express* 27 (16), 22254–22267. doi:10.1364/oe.27.022254

Wang, Y., Liu, L., Wu, J., Song, X., Chen, X., and Wang, Y. (2020). Dynamic Three-Dimensional Shape Measurement with a Complementary Phase-Coding Method. *Opt. Lasers Eng.* 127, 105982. doi:10.1016/j.optlaseng.2019.105982

Wang, Y., Zhang, S., and Oliver, J. H. (2011). 3D Shape Measurement Technique for Multiple Rapidly Moving Objects. *Opt. Express* 19 (9), 8539–8545. doi:10.1364/oe.19.008539

Wei, J., Li, C.-F., Hu, S.-M., Martin, R. R., and Tai, C.-L. (2012). Fisheye Video Correction. *IEEE Trans. Vis. Comput. Graph.* 18 (10), 1771–1783. doi:10.1109/tvcg.2011.130

Wu, X., Jiang, D., Yun, J., Liu, X., Sun, Y., Tao, B., et al. (2022). Attitude Stabilization Control of Autonomous Underwater Vehicle Based on Decoupling Algorithm and PSO-ADRC. *Front. Bioeng. Biotechnol.* 10, 843020. doi:10.3389/fbioe.2022.843020

Xu, Y., Gao, F., Ren, H., Zhang, Z., and Jiang, X. (2017). An Iterative Distortion Compensation Algorithm for Camera Calibration Based on Phase Target. *Sensors* 17 (6), 1188. doi:10.3390/s17061188

Yang, S., Lin, C., Liao, K., Zhao, Y., and Liu, M. (2020). Unsupervised Fisheye Image Correction through Bidirectional Loss with Geometric Prior. *J. Vis. Commun. Image Represent.* 66, 102692. doi:10.1016/j.jvcir.2019.102692

Yun, J. T., Jiang, D., Liu, Y., Sun, Y., Tao, B., Kong, J. Y., et al. (2022). Real-time Target Detection Method Based on Lightweight Convolutional Neural Network. *Front. Bioeng. Biotechnol.* 10, 861286. doi:10.3389/fbioe.2022.861286

Zhang, Z. (2000). A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (11), 1330–1334. doi:10.1109/34.888718

Zhao, G., Jiang, D., Liu, X., Tong, X., Sun, Y., Tao, B., et al. (2022). A Tandem Robotic Arm Inverse Kinematic Solution Based on an Improved Particle Swarm Algorithm. *Front. Bioeng. Biotechnol.* 10, 832829. doi:10.3389/fbioe.2022.832829