A Data-Driven Algorithm for Large-Scale Multi-camera Calibration

Zijun Wang, Yuhui Wan, Kunlin Zhong, Yixin Zhang(✉), and Jian Wang(✉)

School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China
{zyixin,wangjnju}@nju.edu.cn

Abstract. Multi-camera calibration, which is the establishment of a mapping relationship between the 2D coordinates of individual cameras and the 3D world coordinates, has been a major challenge in computer vision technology. The model-driven multi-camera calibration, which starts from the imaging model of the camera, is computationally complex and difficult to consider the imaging distortion comprehensively, while the data-driven multi-camera calibration is often difficult to meet the needs of large-scale scenes in terms of the calibration range. To solve the above two common problems, this paper designs a multi-camera image acquisition system, which collects millions of point cloud coordinate samples in a large-scale space and uses neural networks to infinitely approximate the transformation model of 2D images and 3D world. After the calibration experiments, the scheme is simple and effective, and it can accomplish the requirement of high precision calibration for large-scale space in both spatial positioning and reprojection.

Keywords: Multi-camera calibration · Large-scale · Data-driven · Neural networks

1 Introduction

With the continuous development of computer vision technology, multi-camera has gained widespread attention in the fields of 3D reconstruction, part measurement, defect detection, and autopilot due to their advantages such as extensive measurement range, multiple perspectives, and rich 3D information [1]. The accuracy of the calibration results and the stability of the algorithm directly affect the subsequent 3D reconstruction work. The size of the calibration space also directly limits the use of multi-camera, so how to perform high-precision calibration in large-scale space is the focus of this work.

The current multi-camera calibration work can be broadly divided into two categories. One is defined as model-driven from the imaging model of the camera. And another is defined as data-driven camera calibration from the coordinate data of pairs of massive 2D pixel points and corresponding 3D world

X. Jiang (Ed.): MLICOM 2022, LNICST 481, pp. 75–85, 2023.
https://doi.org/10.1007/978-3-031-30237-4_7

points. Based on the model-driven calibration method, the geometric camera imaging model must be established to consider various possible aberrations in the imaging process. However, the more cameras there are, the more comprehensive the aberrations are considered, and the more accurate the imaging model is. Still, the more computationally tricky it is to perform the calibration work. Unlike model-driven calibration methods, data-driven multi-camera calibration does not require complex model computation and analysis. The mapping relationship between two-dimensional images and the three-dimensional world can be accurately solved using the powerful fitting ability of neural networks. Based on the above reasons, this paper establishes a mapping model with the help of neural networks using a data-driven calibration method. It verifies the feasibility of performing high-precision multi-camera calibration work even in a large-scale range.

2 Related Works

A new algorithm is proposed in [2] for calibrating cameras using occlusion contours of spheres, requiring the camera to observe the sphere at three or more locations. By specifying the problem in dyadic space, [2] recovers the camera parameters optimally using semidefinite planning. The solution is simple, flexible, enabling simultaneous calibration of multiple cameras. However, this approach can be susceptible to degradation or camera aberrations. Lu and Li [3] have designed a theodolite coordinate measurement system (TCMS), a global calibration scheme. The system determines the 3D coordinates of the feature points. Then, the global calibration is performed by direct and indirect transformation methods according to the calibration target position of the camera. But the calibration accuracy of the system depends on the theodolite coordinate measurement system, and its establishment process is slow. Other calibrations of multi-camera use auxiliary tools such as mirrors [4] [5]. Shen and Hornsey [6] present a novel non-planar target for fast calibration of inward-looking visual sensor networks (VSNs). Two spheres built on a supporting rod are used as calibration targets. However, the non-planar targets are applied separately rather than to all cameras. All of the above work is model-driven, but model-driven starts from a complex imaging model, which is difficult to derive, challenging to estimate distortions, and computationally complex and time-consuming.

As artificial neural network technology [7] continues to develop and advance, more and more problems in computer vision can be solved using neural network methods [8]. The same is true for camera calibration. Ahmed et al. [9] propose a method for camera calibration using neural networks, whose network solves the perspective projection matrix between the world's 3D points and the associated 2D image pixels, solving four calibration problems: (i) Estimating all camera parameters simultaneously. (ii) Estimating other parameters given the image center. (iii) Estimating extrinsic parameters given intrinsic parameters. (iv) Estimating intrinsic parameters given extrinsic parameters. However, this algorithm is only suitable for single-camera calibration work. Chen [10] proposes

an improved genetic simulated annealing algorithm to optimize Back Propagation neural networks for binocular camera calibration, which has high calibration accuracy and significantly improves the calibration speed but has a limited calibration space size.

3 Method

To solve the problems of complex calculation and limited calibration range, this paper designs the following system for data acquisition and uses a neural network algorithm for high-precision multi-camera calibration.

3.1 Image Acquisition System

In this paper, we design an auto-system using the multi-camera to acquire images of calibrators at different heights. Based on the system, a large and highly dense calibration point cloud [11] is generated, which provides data support for data-driven multi-camera calibration based on 8 cameras, arranged in the order shown in Fig. 1, with the image acquisition logic shown in Fig. 2.

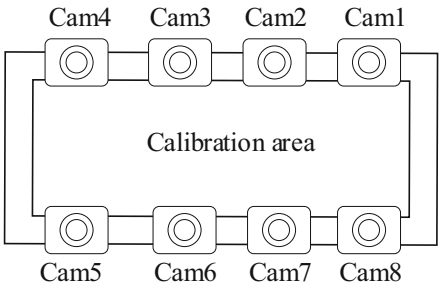

Fig. 1. Camera Array.

3.2 Multi-camera Calibration Process

The camera calibration aims to solve the mapping relationship between the world and pixel coordinate systems, involving the coordinate transformation relationships shown in Fig. 3.

The model-driven calibration of a multi-camera relies on an accurate mathematical imaging model. However, it is challenging to consider the full range of imaging aberrations and accurately describe the non-linear aberrations between the ideal image plane and the actual image plane. All non-linear factors are included in the neural network. Only the coordinates of many point clouds in the camera imaging space are needed to infinitely approximate the mapping between 3D world coordinates and pixel coordinates. In summary, the data-driven calibration method solves the pain points of the model-driven method with the amount of data and skips the intermediate transformation process. The mapping relationship between the two is shown in Fig. 3.

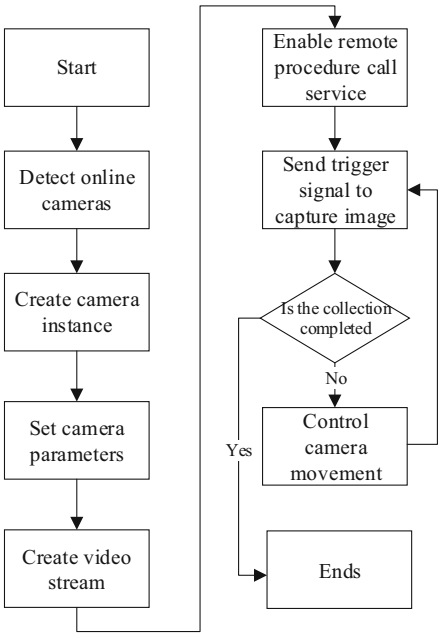

Fig. 2. Image acquisition logic.

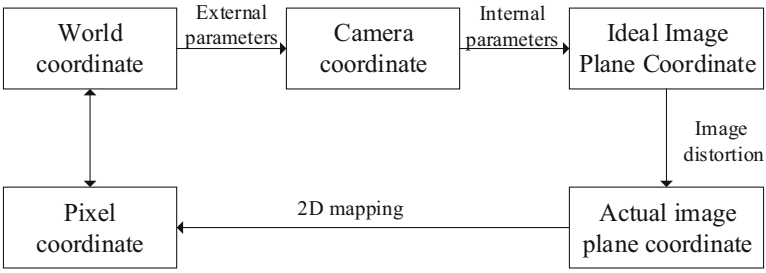

Fig. 3. Coordinate conversion relationships.

Neural Network Structure and Principles. With the development of artificial intelligence, it has been shown that neural networks are capable of approximating arbitrary continuous functions with infinitely minor errors, incorporating all non-linearities. Based on this property, using neural networks makes it possible to perform multi-camera calibrations.

A neural network consists of three parts, an input layer, hidden layers and an output layer. The neural network structure used in this paper is shown in Fig. 4.

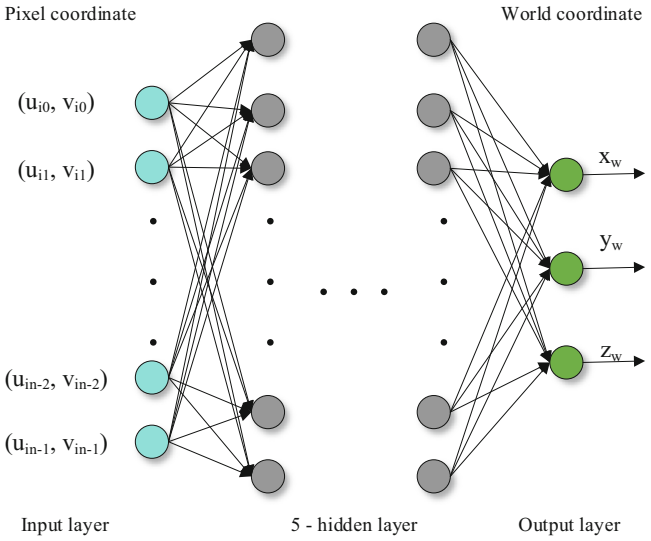

Fig. 4. Neural network structure.

Input Layer. Receives the pixel coordinates from the multi-camera and passes the pixel coordinate information to the hidden layers.

Hidden Layers. The hidden layers used in this paper are all fully connected, the number of neurons in each layer is 32, 64, 128, 64, and 32, respectively, and the principle of action is as in Eq. 1.

$$output = input \times kernel + b \tag{1}$$

Input is a fully connected network input, the output of the fully connected layer, kernel is the internal weight matrix, and b is the bias. The function of the fully connected layer is to perform an affine transformation on the feature data. And the multiple superimposed affine transformations are still essentially affine transformations. The appropriate activation function can introduce non-linear parameters that perfectly fit the mapping relationship between the pixel and 3D world space.

Output layer. The network predicts three neurons for the 3D world coordinates.

Evaluation Indicators. The calibration network uses a multi-layer fully-connected neural network for multi-camera calibration. The core operation of a fully connected neural network is matrix multiplication, which is essentially a feature space transformation. The neural networks-based multi-camera calibration method is driven by data, which first requires extracting the pixel coordinates of calibration points in the chessboard image and establishing a pixel

coordinate point cloud of calibration points with a 3D world coordinate point cloud. The error between the predicted and actual values is calculated based on the loss function, and the network parameters are updated at the end of each epoch (Fig. 5).

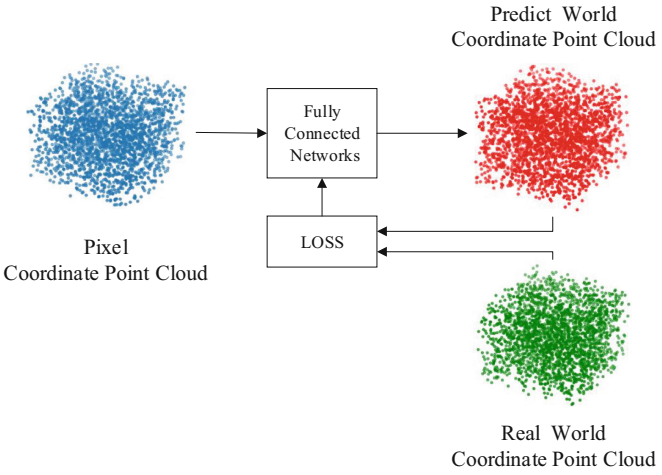

Fig. 5. Schematic of a fully connected calibration network

The neural network loss function is set to MAE(Mean Absolute Error), as shown in the following:

$$MAE = \frac{\sum_{i=1}^{n} [|x_w - x'_w| + |y_w - y'_w| + |z_w - z'_w|]}{n} \tag{2}$$

where x'_w,y'_w,z'_w are the corresponding predicted values, respectively. Through continuous iterative training, different weights and biases in the three-dimensional error space will cause the mean absolute error between the true and predicted values to change, so the magnitude of the MAE value is also an indicator of intuitively understanding the performance of the calibration network.

4 Experiments and Results

4.1 Calibration Data Set Processing

Experiments will be conducted to verify whether the fully connected network can calibrate the multi-camera with high accuracy in large-scale space. As the chessboard [12] calibration board has the characteristics of perspective invariance and high accuracy in large-scale calibration space, it can present apparent edge features at any position and accurately identify saddle points, which helps

to improve spatial positioning accuracy. Therefore, this paper adopts the chessboard calibration board as a calibration tool to accurately and conveniently prepare 3D world coordinates and pixel coordinates datasets to provide tremendous data support for the neural network.

Figure 6 shows the use of a chessboard specification with a grid size of 3.81 mm×3.81 mm, and the calibration points are arranged as 200 × 299. The chessboard has 59800 calibration points, covering 1143×765.81 mm^2 of space.

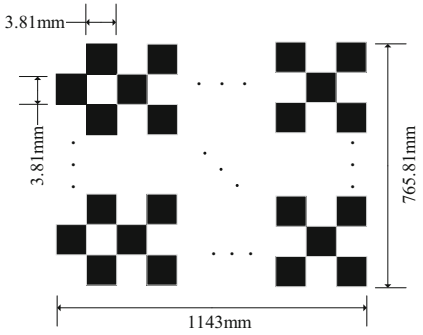

Fig. 6. Chessboard specifications

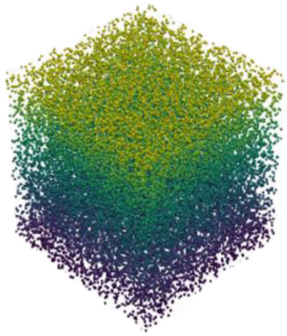

Fig. 7. Neural network train point cloud

The multi-camera calibration system designed in this paper has 8-cameras with an image resolution of 4608 × 3456. There are 76 images on the z-axis of the calibration board, which is uniformly at 4 mm intervals.

Using the OpenCV tool library to identify chessboard corner points with sub-pixel accuracy, a total of 4.5 million samples were obtained, forming a calibrated coordinate point cloud in 3D space, as shown in Fig. 7.

4.2 Calibration Results and Accuracy Analysis

The neural network was trained on a GPU (Graphics Processing Unit), with the number of iterations set to 1000, the learning rate set to 0.0001, and the MAE defined by equation (1) as the loss function.

Spatial Positioning Results. Divide the collected samples according to the ratio of the train set, validation set, and test set 8:1:1. The calibration results from pixel space to 3D world space are shown in Table 1.

Table 1. Spatial positioning accuracy (mm)

MAE for each coordinate axis	Value
x-axis	0.14
y-axis	0.12
z-axis	0.15
Average	0.14

The average spatial positioning accuracy of all three coordinate axis directions in the large-scale calibration space can reach 0.15 mm. To evaluate the prediction effect more intuitively, we fit the predicted coordinate values of the 3D world by the network with the actual values. As shown in Fig. 8, the fully connected network-based multi-camera is calibrated at x, y, z coordinate axes have a high degree of agreement, all of which can perform the task of multi-camera calibration at large scales well. The accuracy is shown in Fig. 9, where the MAE of the test sample is fitted on the 3D surface. In the large-scale calibration space, the local positioning accuracy of the 3D spatial points is more concentrated, and calibration errors in fully connected networks are mainly within 0.2 mm. From the error surface, the calibration accuracy is poorer at the edges, and extreme points with significant errors can occur, probably due to less information around the edge calibration points.

Reprojection Results. Reprojection is an inverse operation with spatial localization to map the 3D world coordinates to the pixel coordinates of the eight cameras. The projected pixel errors for the eight cameras are averaged, and the reprojection accuracy is shown in Table 2.

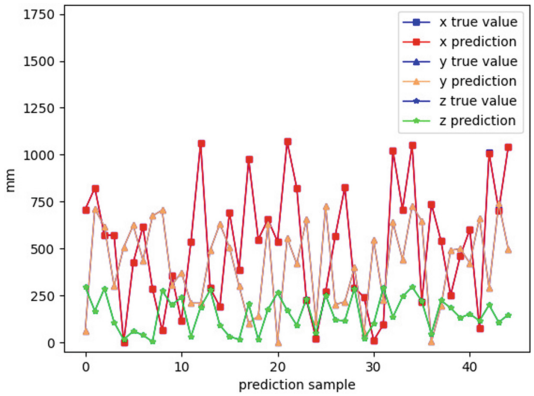

Fig. 8. Results of fitting each axis of the fully connected network.

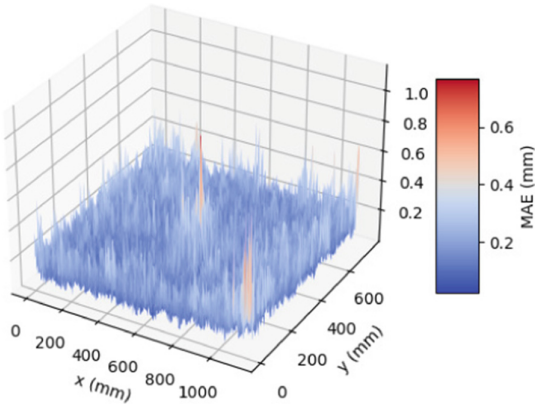

Fig. 9. Fully connected network error surfaces.

Table 2. Reprojection accuracy (in pixels)

MAE for each coordinate axis	Value
x-axis	0.29
y-axis	0.31
Average	0.30

Reprojecting from the 3D world space to the pixel plane, the difference between the results in the x and y axes is small, with an average value of 0.30 pixels. The reprojection error of 0.3 pixels satisfies the need for high precision calibration in large-scale calibration. The reprojection results for each camera are similar, so we selected the results of one of them. The results are as shown in Fig. 10. Figure 11 shows its reprojection error surface.

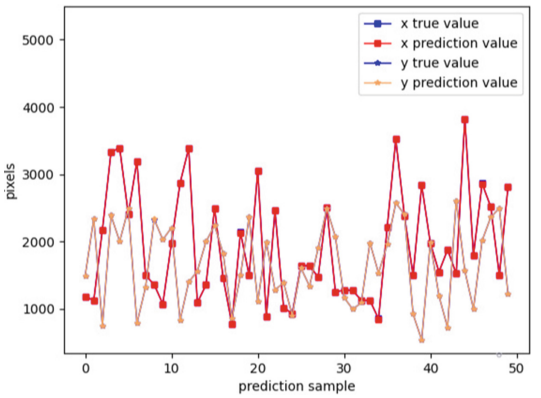

Fig. 10. Results of reprojection fitting for fully connected networks

From the fitting results, the x-axis and y-axis fitting accuracies are high, fully satisfying the needs of large-scale calibration. But the error surface shows an obvious edge error. The reason is that the reprojection model input is 3D spatial coordinates, and the output is the pixel coordinates of 8 cameras. The output dimension is much larger than the input dimension, which makes the model difficult to converge, and the calibration accuracy is relatively poor.

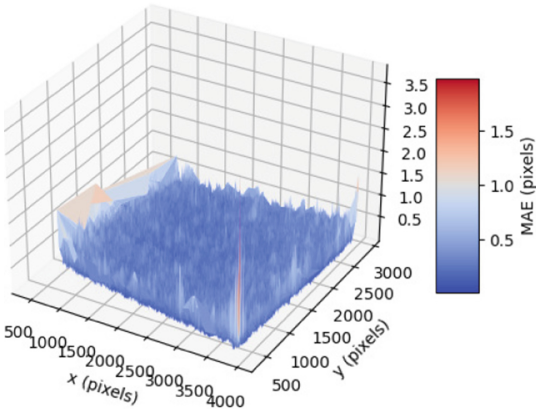

Fig. 11. Fully connected network error surfaces.

5 Conclusion

This paper designs and implements an 8-camera calibration image acquisition system, and the system size is $1143 \times 765.81\,\text{mm}^2$. A fully connected network calibration was carried out using 4.5 million data in a large-scale scene. Based on the experimental results, the network achieves an accuracy of 0.14 mm from pixel coordinates to 3D world coordinates and a reprojection accuracy of up to 0.3 pixels. The result demonstrates the effectiveness of the network in calibrating multi-camera in large-scale space and provides a sound basis for subsequent 3D reconstruction work with high accuracy. To solve the edge accuracy degradation problem, in the subsequent work, we can try to change the network structure or enrich the position information of the edge calibration points to improve the multi-camera calibration performance further.

Acknowledgements. This work was supported by Jiangsu Key R&D Plan (Industry Foresight and Common Key Technology) (BE2018114) and Postgraduate Research & Practice Innovation Program of Jiangsu Province (SJCX22_0014).

References

1. Olagoke, A.S., Ibrahim, H., Teoh, S.S.: Literature survey on multi-camera system and its application. IEEE Access **8**, 172892–172922 (2020). https://doi.org/10.1109/ACCESS.2020.3024568

2. Agrawal, M., Davis, L.S.: camera calibration using spheres: a semi-definite programming approach. In: Proceedings of the Ninth IEEE International Conference on Computer Vision, ICCV 2003, vol. 2, p. 782. IEEE Computer Society, USA (2003)
3. Lu, R., Li, Y.: A global calibration method for large-scale multi-sensor visual measurement systems. Sens. Actu. A Phys. **116**(3), 384–393 (2004). https://doi.org/10.1016/j.sna.2004.05.019, https://www.sciencedirect.com/science/article/pii/S0924424704003279
4. Lébraly, P., Deymier, C., Ait-Aider, O., Royer, E., Dhome, M.: Flexible extrinsic calibration of non-overlapping cameras using a planar mirror: application to vision-based robotics. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 5640–5647 (2010). https://doi.org/10.1109/IROS.2010.5651552
5. Xu, Z., Wang, Y., Yang, C.: Multi-camera global calibration for large-scale measurement based on plane mirror. Optik **126**(23), 4149–4154 (2015). https://doi.org/10.1016/j.ijleo.2015.08.015, https://www.sciencedirect.com/science/article/pii/S0030402615007743
6. Shen, E., Hornsey, R.: Multi-camera network calibration with a non-planar target. IEEE Sens. J. **11**(10), 2356–2364 (2011). https://doi.org/10.1109/JSEN.2011.2123884
7. Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A., Arshad, H.: State-of-the-art in artificial neural network applications: a survey. Heliyon **4**(11), e00938 (2018). https://doi.org/10.1016/j.heliyon.2018.e00938, https://www.sciencedirect.com/science/article/pii/S2405844018332067
8. Zhou, Y.T., Chellappa, R.: Computational Neural Networks, pp. 6–14. Springer, New York (1992). https://doi.org/10.1007/978-1-4612-2834-9_2
9. Ahmed, M., Hemayed, E., Farag, A.: Neurocalibration: a neural network that can tell camera calibration parameters. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 1, pp. 463–468 (1999). https://doi.org/10.1109/ICCV.1999.791257
10. Chen, L., Zhang, F., Sun, L.: Research on the calibration of binocular camera based on bp neural network optimized by improved genetic simulated annealing algorithm. IEEE Access **8**, 103815–103832 (2020). https://doi.org/10.1109/ACCESS.2020.2992652
11. Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M.: Deep learning for 3D point clouds: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **43**(12), 4338–4364 (2021). https://doi.org/10.1109/TPAMI.2020.3005434
12. De la Escalera, A., Armingol, J.M.: Automatic chessboard detection for intrinsic and extrinsic camera parameter calibration. Sensors **10**(3), 2027–2044 (2010). https://doi.org/10.3390/s100302027, https://www.mdpi.com/1424-8220/10/3/2027