

The Comparison of Calibration Method of Binocular Stereo Vision System

Ke Zhang^{a*}, Zhao Gao^b

College of Engineering, Zhejiang Normal University, Jinhua, 321004, P.R.China

^ajhkeanu@zjnu.cn, ^bgz@zjnu.cn

Keywords: Stereo vision, Calibration method, Genetic algorithm, BP neural network

Abstract. Stereo vision is a non-contact method to realize 3D reconstruction of free-form surface, and has a wide range of applications in the field of reverse engineering and virtual reality. The technology of camera calibration has always been the research hotspot and difficulty. In this paper, firstly the experimental system of binocular stereo vision is constructed, and then the stereo vision system is calibrated respectively by using genetic algorithm and BP neural network. In the method of genetic algorithm, the encoding rule of adaptive adjustment of parameter search interval is proposed, which effectively realize the camera calibration of the high dimension and non-linear system of binocular stereo vision. In the calibration method of BP neural network, with the help of CNC precision mobile workstation to obtain the density calibration sample data, by means of BP neural network which simulating the mapping relation between the 3D space and 2D image plane of stereo vision system, the implicit calibration model of binocular stereo vision system is constructed, which avoids the system error because of imperfect mathematical models. The two methods have respective advantages, and are suitable for different application situation.

Introduction

The pioneering work of stereo vision started in the mid 60's of the last century. Roberts of America MIT extended the past simple 2D image analysis to the complex 3D scene by extracting 3D structure of simple and regular polyhedron, such as cube, wedges and prisms in digital images, and by describing the shapes of objects and spatial relationships, which marks the birth of stereo vision technology. With the development of research, the scope of study was enlarging from the feature extraction of edges and corners, and the analysis of geometrical elements such as lines, planes, surfaces, directly to the analysis of image shading, image texture, and movement and imaging geometry, and various data structure and rules of inference were established. Especially in the early 80 of the last century, Marr [1] generalized the research results in disciplines, such as image processing, mental physics, neuro-physiology, and clinical psychiatry, from the angle of information processing for the first time, and then created the theoretical framework of visual computing. The fundamental theory greatly promoted the development of stereo vision technology, so the complete system from image acquisition to visual surface reconstruction of 3D scene was formed, which made the stereo vision to be an important branch of the computer vision. After decades of development, the application of stereo vision in the fields of robot vision, aerial mapping, reverse engineering, military, medical imaging and industrial inspection become more widely [2,3].

Stereo vision simulates the mechanism of human vision using binocular cue to perceive the around scene, and realizes understanding and recognition of 3D information by sensing equipment. In the practical application, stereo vision always take images of the same object from different viewpoints using two cameras, and then reconstructs the 3D shape of space objects from the two images according to triangulation method.

In the system of binocular stereo vision, the geometry layouts between two cameras will directly affect the common vision of the two cameras and the search scope of image matching. Therefore, it is needed to perform analysis and selection of the system structure, and then the hardware system can be constructed.

The research of the problem of camera calibration of stereo vision has always been the focus and difficulty in the field of stereo vision. In the practical use, because of the existence of various distortion of camera lens in the system of stereo vision and various error of device installation, the

ideal linear model can not accurately describe the relationship of camera's imaging geometry. The existence of various non-linear factors makes the relationship between objects and images complex, and can't even use a precise mathematical model to express the relationship, the only way to do that is to adopt a simplified approximation of the actual model on the premise of satisfying the precision requirement.

There are two methods to solve the problem of camera calibration of stereo vision system, the one is based on non-linear mathematical model, which uses mathematical analysis equation to approximately express the non-linear geometry mapping relationship between the 3D geometry position of space object surface points and corresponding image points on the left and right image planes, and the camera calibration of stereo vision system is to solve all external and internal parameters which make up the mathematical model. The other one is based on artificial neural network, which has no need to establish the camera mathematical model of stereo vision system, and it constructs the mapping relationship between 2D image coordinates and 3D world coordinates by using non-linear fitting capability of artificial neural network.

Construction of Binocular Stereo Vision System

There are three optional structures of binocular stereo vision system, i.e. parallel optical axis structure, common optical axis structure and intersecting axis structure. In the stereo vision system of parallel optical axis structure [4], the focal length and other internal parameters of two cameras are all equal with the optical axis perpendicular to the camera's imaging plane, the horizontal axes of two cameras to be overlap, and the vertical axes to be parallel to each other, so the left camera can be coincident with the right camera by moving the left camera along the horizontal axis for a certain distance. The stereo vision system of this structure form has very convenient conditions when the matching relationship of images is being established, however it is very difficult to make the two axes keep parallel, because the axes can't be observed in the process of installing the cameras. In addition, in order to make two cameras observe of parallel optical axis at the same time the visible surface of the object to be measured, the two cameras must be installed very closely, but doing so will take the disparity between the projections on the left and right imaging planes to be smaller, which will affect the precision of measurement.

In the stereo vision system of common optical axis structure [5], the optical axes of two cameras of front and back layout are coincided. The epipolar plane composed of the space object point and two optical center of the camera intersect with the imaging planes of front and back layout, and the intersection lines are conjugate line, which go through their respective center of the imaging plane and have the same slope. Obviously, the conjugate lines through image center and with same slope provided convenient conditions for matching between the two images. However, as is the case with stereo vision system of parallel optical axis structure, it is also very hard to make the optical axis of the two cameras to be collinear. Meanwhile, in order to observe simultaneously the same part of the scene, the distance between the front and back cameras must be large enough, but, if the baseline distance is too large, then the occlusion on two images will appear, it is to say that the some surface area of the measured object will appear on the one image, and not on the other image, or will not appear on the two images. Even more so, when the curvature radius of the measured 3D surface is larger and the surface change is dramatic.

The intersecting optical axis structure is shown as Fig.1. In this structure form, the optical axes z_l , z_r of left and right cameras was arranged at an angle. The cameras of this structure form were installed easily, and the distance and slope direction of two cameras could be adjusted flexibly according to the characteristics of the measured object and system requirements, which could avoid from the structure the situations of small field of view and image occlusion existing in the above two forms. But, it can be observed from the figure that the distribution of conjugate epipolar lines p_l , p_r is not conducive to the matching of left and right images.

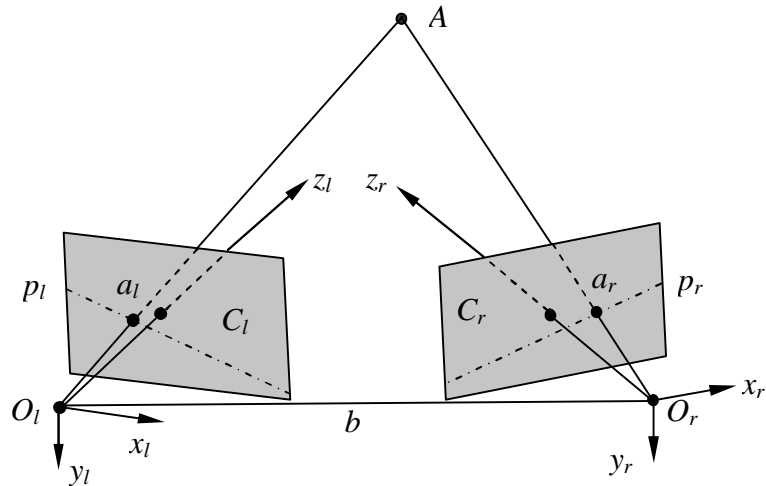


Fig.1 Intersecting optical axis structure

Considering comprehensively the advantages and disadvantages of three kinds of structures, the binocular stereo vision system set up in the intersecting optical axis structure both can be easily installed, adjusted, and can realize smoothly image matching under the existing conditions, which is a common structure form in the practical application.

System Calibration Based on Genetic Algorithm

Constructing the model of binocular stereo vision system, if considering the influence of radial distortion factor, the calibration of double camera is a nonlinear complex function optimization problem. The traditional optimization methods to solve such problems include Newton method, gradient descent method and conjugate gradient method, etc. But, these methods need to set the appropriate initial values, and if the initial value is not set properly, these methods can lead to poor convergence of optimization process or easy to fall into local extremum. Especially for the large scale nonlinear optimization problems, the difficulty of finding the global best solution will increase with the increase of local extremum.

Genetic algorithm [6] is an adaptive searching probabilistic method for global optimization, which is developed based on natural selection and genetic evolution mechanism. It uses group search technology to make the community evolve gradually to the state of including or nearing the optimal solution by imposing a series of operations on the current group, such as options, overlapping and mutation, and generating new group.

The application of genetic algorithm on the camera calibration has made great achievements, but it mainly focused on single camera parameter calibration. Qiang [7] adopted genetic algorithm to calibrate camera parameters, and got good performance in terms of convergence, accuracy and robustness. But, since the model of camera didn't include distortion factor of lens, the precision was difficult to ensure for 3D measurement. In addition, Roberts [8] pointed out that the standard genetic algorithm might not meet the precision requirements for the high-dimensional optimization space with number of parameters greater than 10. Meanwhile, his research results indicated that the performance of genetic algorithm would be degraded sharply with the increasing of the dimensions of optimization space.

The genetic algorithm applied in the camera calibration of stereo vision system should be improved to meet the various demands. The adaptive variable interval adjustment function was added to the standard genetic algorithm by improving the encoding scheme. The improved genetic algorithm can satisfy simultaneously the requirements of search space and encoding precision with the encoded length of the chromosome keeping unchanged, and eventually realize the parameters calibration of the high-dimension and nonlinear system of binocular vision with the improved genetic algorithm.

The adaptive adjustment method of variable search interval is achieved by moving through the

double interval. There are two intervals corresponding to the variable, the large interval $[A_i^{\min}, A_i^{\max}]$ represents the range of the possible variable value, and its boundary is fixed, the boundary value is determined by initial estimation. The small interval $[a_i^-, a_i^+]$ can move within the large intervals, and the interval size remains unchanged in the process of moving. When encoding variable, small interval is used so as to ensure the accuracy of coding. When searching the optimal solution, the search will cover the entire variable value range by making the small intervals moving within the large intervals.

The calibration plate with black and white pattern was taken images by left and right CCD camera. The total 126 test points were obtained. After image processing and actual measurement, corresponding 126 groups data were ready, and every group data included left and right image coordinates (u_l, v_l) and (u_r, v_r) , and 3D world coordinates (X_w, Y_w, Z_w) . The all test points data were divided into two groups of calibration and testing points, which 24 points were used as calibration calculation and the other 102 points used to test the 3D measurement precision after calibration of stereo vision system.

Taking the image coordinates (u_l, v_l) and (u_r, v_r) of testing points to be the input of the calibrated measurement system of binocular stereo vision, the 3D space coordinates (X'_w, Y'_w, Z'_w) can be obtained from 3D reconstruction. Then after comparison of the 3D space coordinates (X'_w, Y'_w, Z'_w) and the true 3D coordinates (X_w, Y_w, Z_w) of testing points, the average error of measurement of the stereo vision system was shown as in Table 1, which was compared with the measurement results of 24 calibration points.

Table 1 comparison of average measurement error

	X_w direction(mm)	Y_w direction(m m)	Z_w direction(mm)
Calibration point	0.37	0.56	1.32
Testing point	0.33	0.63	1.27

The experimental results indicates that although the solution space of the nonlinear complex and high dimension optimization problem of the binocular stereo vision system is very large, the improved genetic algorithm have good convergence in the near-optimal solution.

Calibration Based on BP Neural Network

For stereo vision system, the ultimate goal of camera calibration is to construct the mapping relationship between the 2D coordinates of object points and the 3D world coordinates, which make the 3D world coordinates of object points to be directly derived from the 2D coordinates on the left and right image planes. So there is not need to solve the every physical parameter of camera, and the physical parameters of the camera can be merged into some intermediate parameter to be calculated, which make the process of calculation simplified and improve the precision of camera calibration to some extent. This method is called implicit calibration. Corresponding to this method, the way to construct mathematical model of camera and accurately solve the internal and external physical parameters of the camera called explicit calibration of camera [9].

The internal mechanism of the Artificial Neural Network (ANN) is strikingly similar with the process of implicit calibration of a camera, and the both methods construct system model by some known data and then calculate unknown data. The method of camera calibration by means of ANN has no need to construct mathematical model of stereo vision system, which can reduce errors caused by imperfect mathematical model, and can avoid the defects of explicit calibration. Doing so can improve the measurement precision.

The ability of extension or generalization of ANN is an important index to measure the performance of neural network. A neural network of good generalization ability not only has higher matching effects for the training sample set, but also will simulate the approximate output of a new

sample input vector relative to destination vector. The standard BP algorithm is easy to get into trouble of the local minimum, slow convergence and excessive training, which not only affects training results and extends the convergence time, but also reduces ability to promote. The regularization method can improve the generalization ability by amendments to the training performance function of the standard BP neural network.

Generally, the performance function of BP algorithm adopts Mean Square Error function mse . Assuming that there are N training sample $[X_i, Y_i]$ ($i=1, 2, \dots, N$), and the network weights and initial value of thresholds is arbitrary, when there is input X_i , the output \bar{Y}_i is obtained, so the Mean Square Error of all sample is as follows.

$$mse = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y}_i)^2$$

(1)

In the method of regularization, the performance function is amended as follows.

$$E = g \cdot mse + (1-g) \cdot msw$$

(2)

Where, g represents scale factor and msw represents average sum of squares of all the network weights value.

$$msw = \frac{1}{M} \sum_{j=1}^M w_j^2$$

(3)

The method of regularization adaptively adjusts the value of g in the process of network training to make sure that the network has less weight in the case of network training error E as small as possible, which is equivalent to reducing the size of the network automatically to reduce the opportunity of overtraining, and obtain the objective to improve network generalization ability.

When acquiring images of calibration plate, positions of calibration are total 21 and calibration points of every calibration position are 216, so all calibration points are 4536. The true 3D coordinates of each calibration point are determined by the position of calibration plate and the position on the calibration plate. The image coordinates of every calibration point on the left and right image planes are obtained by means of corner extraction procedure, as shown in Fig.2.

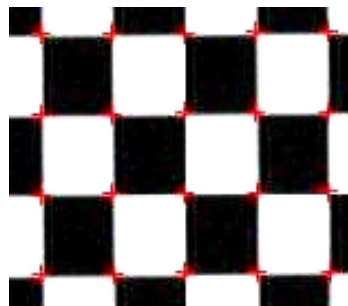


Fig.2 corner extraction

The data sample is composed of the left and right image coordinates (u_l, v_l, u_r, v_r) and the 3D world coordinates (X_w, Y_w, Z_w) corresponding to every calibration point. The total 4536 samples is divided into training samples 800 and testing samples 3736.

The training results of BP neural network are shown as in Table 2.

Table 2 training error of neural network

Average error(mm)			Max error(mm)		
$\square X$	ΔY	ΔZ	ΔX	ΔY	ΔZ
0.0548	0.0529	0.2744	0.2533	0.2467	0.9561

The testing errors are obtained by simulating the trained BP neural network using testing samples, shown as in Table 3.

Table 3 testing errors of neural network

Average error(mm)			Max error(mm)		
ΔX	ΔY	ΔZ	ΔX	ΔY	ΔZ
0.0579	0.0536	0.2972	0.3197	0.2739	1.3578

The experimental results indicate that the precision of training and testing of BP neural network are almost equal, which means the BP neural network has good generalization ability in the effective calibration 3D space.

Conclusion

The genetic algorithm used in the calibration of double camera can solve respectively the internal and external parameters of each camera. Because of the adoption of adaptive encoding method, the ability to solve the problem of nonlinear and complex function optimization got strengthened. The Artificial Neural Network can realize high approximation of the complex nonlinear mapping, which provides a effective method to construct camera model of stereo vision system. The method doesn't need to consider the effects of lens distortion and environmental factors, so it can reduce errors caused by imperfect of mathematical model and improve the measurement precision.

References

- [1] Marr D. Vision. W.H. Freeman and Company, San Francisco. Chinese version: computational theory of vision. G.Z. Yao , L.Liu and Y.J.Wang , Science Press. Beijing, China, 1988, pp. 6-37(In Chinese).
- [2] S.Y. You and G.Y. Xu: Journal of Image and Graphics. Vol. 2, No. 1, (1997), pp. 17~24(In Chinese).
- [3] Y.D. Jia: Machine Vision. Science Press. Beijing, China, 2000, pp. 1-25(In Chinese).
- [4] S.D. Ma and Z.Y. Zhang: Computer Vision-Fundamentals of Computational Theory and Algorithms. Science Press. Beijing, China, 1998, pp. 72-93(In Chinese).
- [5] M. Kong and S.M. Wang: Journal of Metrology. Vol. 25, No. 4, (2004), pp. 294~297(In Chinese).
- [6] M. Zhou and S.D Sun: Principle and Application of Genetic Algorithm. National Defense Industry Press. Beijing, China, 2002, pp. 4-64(In Chinese).
- [7] J. Qiang and Y.M. Zhang: IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans. Vol. 31, No. 2, (2001), pp. 120~130.
- [8] Roberts M and Naftel A J: Genetic Algorithms in Image Processing and Vision. IEE Colloquium on, 1994: 12/1~12/5
- [9] G.Q. Wei and S.D. Ma: IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 16, No. 5, (1994), pp. 469~480.