Journal of Achievements in Materials and Manufacturing Engineering
VOLUME 20
ISSUES 1-2
January-February 2007

An improved neural networks for stereo-camera calibration

Y.J. Xing*, J. Xing, J. Sun, L. Hu
Dalian University of Technology, Key Laboratory for Precision and Non-traditional Machining Technology of Ministry of Education, 116023, P. R. China
* Corresponding author: E-mail address: yjxing@ dlut.edu.cn

Received 02.11.2006; accepted in revised form 15.11.2006

Methodology of research

ABSTRACT

Purpose: Improve the generalization capability and speed of back-propagation neural network (BPNN).
Design/methodology/approach: In this paper, CCD cameras are calibrated implicitly using BP neural network by means of its ability to fit the complicated nonlinear mapping relation. Conventional BP algorithms easily fall into part-infinitesimal, slowing speed of convergence and exorbitance training that will influence the training result, delay convergence time and debase generalization capability. During our experiments, dense sample data are acquired by using high precisely numerical control platform, and the variances error (PVE) is adopted during training the neural network.
Findings: Experiments indicate that the neural network used PVE has great generalization. The error percentages obtained from our set-up are limitedly better than those obtained through Mean Square Error (MSE). The system is generalization enough for most machine-vision applications and the calibrated system can reach acceptable precision of 3D measurement standard.
Research limitations/implications: The value needs to be decided by experiments, and the reconstruction images will be distorted if the value is more than 6.
Originality/value: The variances error is be adopted in BPNN first.
Keywords: Image analysis; Stereo vision; BP neural network; Generalization

1. Introduction

Computer vision inspection systems are one of many important automated non-contact measurement and inspection devices in modern production lines. Camera calibration is considered as an important issue in computer vision. Accurate calibration of cameras is especially crucial for applications that involve quantitative measurements, depth from stereoscopy or motion from images [1]. Most current machine vision systems for automated industrial inspection are custom-designed, so they are suitable only for one specific application. The problem of camera calibration is to compute the camera extrinsic and intrinsic parameters. The extrinsic parameters of a camera indicate the position and the orientation of the camera with respect to the coordinate system, and the intrinsic parameters characterize the inherent properties of the camera optics, including the focal length, the image centre, the image scaling factor and the lens distortion coefficients. The number of parameters to be evaluated depends on the camera model being utilized. Their inherent inflexible and no versatile structures prevent them from being adapted to other applications, or even adapted to operate on same application but in a different operating environment. Neural networks are being applied in many scientific disciplines to solve a variety of problems in pattern recognition, prediction, and optimization associative memory and control [2]. None of the conventional approaches to these problems is flexible enough to perform well outside their domain. Neural networks provide exciting alternatives and many applications could benefit from them. This method need not construct the mathematic model of vision system, and can improve the precision of measure effectively [3]. In our problem, back-propagation neural network (BPNN)-a multilayer neural network -model is adopted because camera calibration problem is a nonlinear problem.

2. The theory of 3D measurement used two CCD cameras

The 3D measurement system composed by two CCD cameras is shown in figure1, $o_l x_l y_l z_l$ and $o_r x_r y_r z_r$ figure the reference frames of left and right cameras, o_l and o_r figure the optical centers of two cameras; $O_{il} X_l Y_l$ and $O_{ir} X_r Y_r$ figure the imaging plane reference frames of left and right cameras, O_{il} and O_{ir} are respectively the point of intersection that the optical axis of left and right cameras intersect with their own imaging plane, X_l, Y_l parallel x_l, y_l, X_r, Y_r parallel x_r, y_r. P is a random point on the surface of the object measured, it is presumed that the system has calibrated and it is sure that the point p_l on the image of left camera and the point p_r on the image of right camera are the image points of the same space point P. Because the point P is on both the line $o_l p_l$ and $o_r p_r$, the point P is the point of intersection of the line $o_l p_l$ and $o_r p_r$, namely the 3D position of point P is uniquely certain [4].

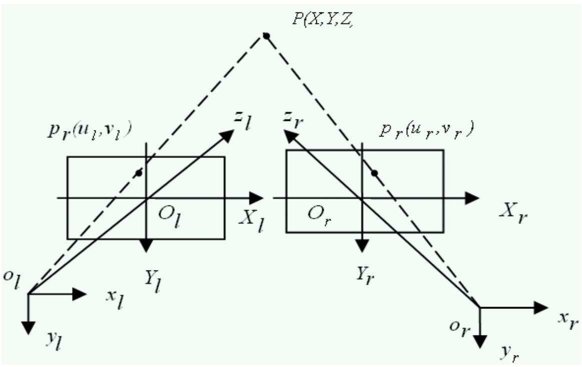


Fig. 1 The theory of 3D measurement

3. The structure and training of BPNN

3.1. The structure of BPNN

The BP neural network model in figure 2 is adopted for our simulations. It falls into the category of the feedforward class and the learning algorithm. The learning algorithm is based on a sliding mode approach, which guarantees very high speed of learning, and has been shown to be particularly effective in on-line implementations. Each output in a layer is connected to each input in the next layer [5, 6]. In this case, the output layer has simple linear neurons, while all the neurons in the two hidden layers have the same transfer function, with a sigmoidal nonlinearity. Also, because there is no feedback between layers, the effect of the feedforward neural net topology is to produce a nonlinear mapping between the input nodes and the output nodes.

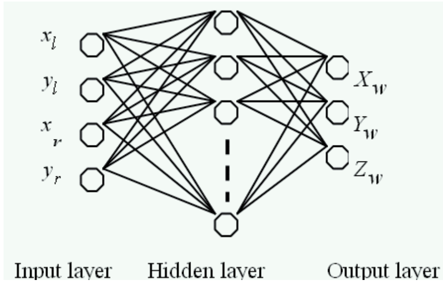


Fig. 2 Framework of BPNN

The model that we used consists of four input neurons, eight hidden neurons and three output neurons. The input neurons correspond to the image coordinates of matched points found on the stereo images (x_l, y_l) and (x_r, y_r). These points are generated by the same world point on both images and form the input of the neural network. The output neurons correspond to the actual world coordinates of the point (X, Y, Z) which are mapped as(x_l, y_l)and (x_r, y_r) on the two images[7]. The network is trained on a range of inputs and outputs, such that the network could, after training, give the world coordinates for any matched pair of points.

3.2. Improved capability of generalization

Generalization capability is an important guide line to scale the performance of neural networks. A neural network with good generalization capability not only has the better matching effect in training-swatch aggregate, but also will simulate a precise output relative to the new swatch input vector [8]. Conventional BP algorithms easily fall into part-infinitesimal, slowing speed of convergence and exorbitance training that will influence the training result, delay convergence time and debasing generalization capability [9]. Generally, the performance function of BP algorithm adopts Mean Square Error (MSE) function. If the number of training swatch [X_i, Y_i] is S, and net weights and thresholds is intercalated arbitrarily. When X_i is inputted, Y_i is outputted, so all MSE of the swatch is:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \overline{Y_i})^2 \tag{1}$$

In this paper, MSE function is modified to p variances error (PVE):

$$E = \mathrm{PVE} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{Y_i - \overline{Y_i}}{Y_i}\right)^{2p} \tag{2}$$

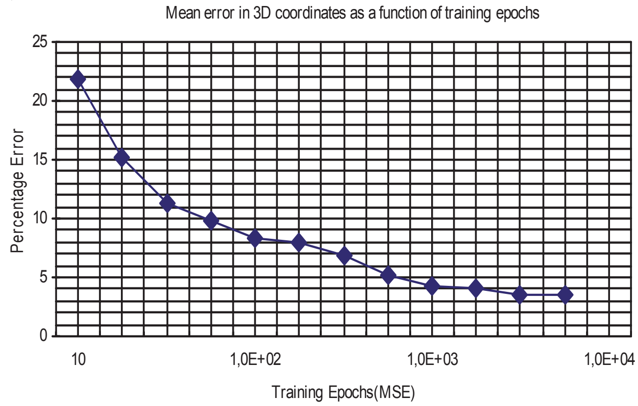
In formula (2), it is obvious that $0 < \left(\frac{Y_i - \overline{Y_i}}{Y_i}\right)^2 < 1$, so the value of $\left(\frac{Y_i - \overline{Y_i}}{Y_i}\right)^{2p}$ will decrease as p increase. That will make the net be more sensitive and insure the net has the lesser weights while the training error E is as soon as possible small. That means the scale of net reduces automatically, consequently, the chance of over-training will be reduced and the generalization capability of the net will be improved.

(a) p=2 (b) p=4 (c) p=7

Fig. 3 Reconstructed images with different p

Figure 3 shows that the relation between reconstructed images and the value p [10-12]. The value p is chose between 2 to 6 after experiment.

a)



b)

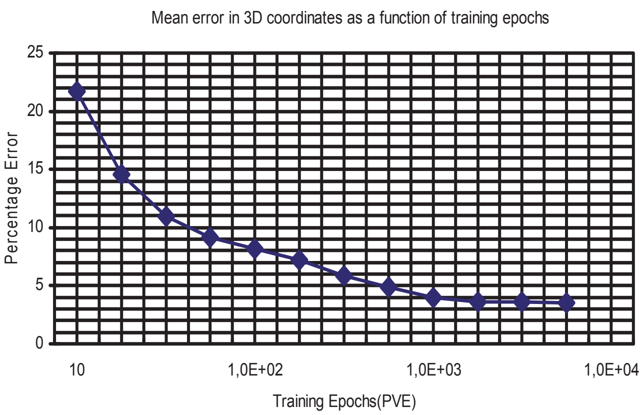


Fig. 4 Mean percentage error in computing 3D coordinates as a function of the number of epochs MSE, (b) PVE

As show in figure 4, the convergence speed of PVE is faster than the one of MSE. But the value of p can not be too big, that will result in the value E too small and the weight regulating too slow [13].

4. Experimental details and results

The distance between the two cameras is kept at approximately 70mm and did not align their optical axes precisely. Next a calibration chart consisting of a grid of lines 50mm apart is made. This chart is placed in front of the cameras at various distances from the world origin and its image captured from both cameras, without moving the cameras. The chart is placed at distances that were in the range within 200± 400mm in front of the cameras and captured images in this range at increments of 10mm [14].

After capturing the images of the calibration chart, these images are matched to obtain stereo pair points. For each stereo pair, we also knew the actual world coordinate, since the chart is placed at measured coordinate with respect to a world origin. There are 216 points on the calibration chart, so the 4320 stereo pairs and their actual three-dimensional (3D) world coordinate are got. The 1000 stereo pairs and their 3D world coordinate are made as the training set and the other as verifying set. Then the neural network is trained on this set of 1000 stereo pairs. The training is done by presenting the stereo pair points to the input of the network and presenting the 3D world coordinates at the output.

The net (p =4) converge after 4537 epochs of training. The results show as the table 1,

Table 1.
The training error of BPNN (absolute value) mm

Mean Error			Max Error		
ΔX	ΔY	ΔZ	ΔX	ΔY	ΔZ
0.0523	0.0391	0.2374	0.2012	0.1493	0.9504

The verifying set with 3320 stereo pairs is used to proof-test the BPNN trained, and the table2 shows the verifying error of the net.

Table 2.
The verifying error of BPNN (absolute value) mm

Mean Error			Max Error		
ΔX	ΔY	ΔZ	ΔX	ΔY	ΔZ
0.0502	0.0398	0.2847	0.2087	0.1982	1.1307

Comparing table1 and table2, it is evident that the training error and the verifying error correspond. That indicates that the BPNN have good generalization capability in 3D visual space [15].

5. Conclusions

In this paper, we have presented a unified approach of camera calibration and 3D world reconstruction for stereo-vision. We used a BPNN to train the system , when the system is presented with a matched pair of points, it automatically computes the world coordinates of the corresponding object point. We use P error variances to make the net has good generalization capability and increase the speed of convergence. The approach is simple in concept, independent of the camera model and obtains very good results.

The algorithm has good adaptability, in the future the aims should be focus on its wide applications.

Acknowledgements

This project is supported by the LiaoNing Provincial Natural Science Foundation in China (No. 2051057).

References

[1] L. Guan, H. Kong, A real-time machine vision system for visual information processing, Proceedings of International Conference on Systems, Man, and Cybernetics, 1994, 2:1375 – 1380.

[2] K. Niranjan Prasad, B. Ramamoorthy, Tool wear evaluation by stereo vision and prediction by artificial neural network ,Journal of Materials Processing Technology, 112(2001) 43-52.

[3] D. Yongtae. Application of neural networks for stereo-camera calibration. Proceedings of conference on Neural Networks, 1999, 2719 – 2722.

[4] K. Zhang, B. Xu, L.X.Tang, H.M.Shi, Camera Calibration of Binocular Vision System Based on BP Neural Network, Journal of Machinery & Electronics12 (2005) 12-14.

[5] B. Endelt, K.B. Nielsen, J. Danckert, New framework for on-line feedback control of a deep-drawing operation ,Journal of Materials Processing Technology, 177 (2006) 426-429.

[6] F. Cupertino, V. Giordano, E. Mininno, A neural visual servoing in uncalibrated environments for robotic manipulators. Proceedings of Conference on Systems, Man and Cybernetics, 6, 2004, 5362 – 5367.

[7] L.N. Smith, M.L. Smith, Automatic machine vision calibration using statistical and neural network methods Journal of Image and Vision Computing, 23(2005) 887-899.

[8] F. Dieterle, S. Busche, G. Gauglitz, Growing neural networks for a multivariate calibration and variable selection of time-resolved measurements, Journal of Analytica Chimica Acta, 490 (2003) 71-83.

[9] U. Zuperl, F. Cus, B. Mursec ,T. Ploj ,A generalized neural network model of ball-end milling force system,Journal of Materials Processing Technology, 175 (2006) 98-108.

[10] M. Lhuillier and L. Quan, Quasi-Dense Reconstruction from Image Sequence, Proc. Seventh European Conf. Computer Vision 2 (2002) 125-139.

[11] Q. Memony, S. Khan, Camera calibration and three-dimensional world reconstruction of stereo-vision using neural networks, Journal of Systems Science 32 (2001) 1155- 1159.

[12] N. Uchida, T. Shibahara, T. Aoki, H. Nakajima, K. Kobayashi, 3D Face Recognition Using Passive Stereo Vision, Proceedings of ICIP 2 (2005) 950-953.

[13] S.S. Panda, A.K. Singh, D. Chakraborty, S.K. Pal, Drill wear monitoring using back propagation neural network,Journal of Materials Processing Technology, 172(2006):283-290

[14] J.P. Wang, Y.C. Chueh**,** Neural network approach to recognize the grid patterns in experimental mechanics,Journal of Materials Processing Technology, 140 (2003) 90-94.

[15] A.W. Zhang, Z. Sui, S.X. Hu, J.J. Chen, M.Z.Li, Method for vision system calibration based on neural network, Journal of optical technique,27 (2001) 302-304.