논문 98-7-6-06

3차원 위치측정을 위한 스테레오 카메라 시스템의 인공 신경망을 이용한 보정

都勇兌*, 李大植*, 柳奭桓*

Calibrating Stereoscopic 3D Position Measurement Systems Using Artificial Neural Nets

Yongtae Do*, Dae-Sik Lee*, Seog-Hwan Yoo*

요 약

로봇을 비롯한 자동화 기계의 3차원 작업에서 스테레오 카메라는 가장 널리 사용되는 센서 장치이다. 스테레오 카메라를 사용함으로써 3차원 실세계 공간내 임의 목표점의 위치를 측정할 수 있으며, 카메라의 보정은 이를 위한 중요한 선행작업이다. 기존의 카메라 보정법은 크게 선형과 비선형의 기법으로 나눌 수 있는데, 선형의 기법은 간단하나 정확도의 면에서 문제점을 지니고, 비선형 기법은 렌즈의 왜곡을 보상하기 위한 모델링 과정과 이의 비선형 해를 구하는 비교적 복잡한 과정을 필요로 한다는 문제가 있다. 본 논문에서는 이러한 문제의 한 해결방안으로 인공신경망을 적용하는 방법을 연구하고 그 결과를 제시한다. 특히 역전파 알고리즘에 의해 학습된 다층 신경망의 함수 근사화 능력을 활용하여 선형기법의 오차 패턴을 학습함으로써 간단하고 효과적으로 계측의 정확도를 향상시킬 수 있음을 실험을 통하여 보인다.

ABSTRACT

Stereo cameras are the most widely used sensing systems for automated machines including robots to interact with their three-dimensional(3D) working environments. The position of a target point in the 3D world coordinates can be measured by the use of stereo cameras and the camera calibration is an important preliminary step for the task. Existing camera calibration techniques can be classified into two large categories - linear and nonlinear techniques. While linear techniques are simple but somewhat inaccurate, the nonlinear ones require a modeling process to compensate for the lens distortion and a rather complicated procedure to solve the nonlinear equations. In this paper, a method employing a neural network for the calibration problem is described for tackling the problems arisen when existing techniques are applied and the results are reported. Particularly, it is shown experimentally that by utilizing the function approximation capability of multi-layer neural networks trained by the back-propagation(BP) algorithm to learn the error pattern of a linear technique, the measurement accuracy can be simply and efficiently increased.

1. INTRODUCTION

The use of sensors allows robots and other automatic machines to interact with their target objects and surroundings. This interaction can be performed by contacting or noncontacting ways

* 대구대학교 정보통신공학부 (School of Computer & Communication Engineering, Taegu University)
<접수일자 : 1998년 9월 20일>

endowing robots with flexibility. A camera is the most widely used noncontacting sensing device for robots thanks to its versatility. Visual perception using cameras has been studied since the early stage of robotics research[1] and its applications has been expanded from simple 2D tasks to sophisticated spatial manipulations in unstructured environments. Sensing the 3D world reliably in real time is thus of growing importance.

3D information can be acquired by employing stereo cameras. The real world and corresponding stereo images can be related if the optical and geometrical parameters of the two cameras are known. Camera calibration is the process of determining the intrinsic parameters (such as focal length and optical image centre) and extrinsic parameters (such as geometric position and orientation) of a camera implicitly or explicitly for establishing the projection or back-projection relation between the 3D world and 2D image[2].

Most existing calibration techniques may be classified into two large categories; linear and nonlinear techniques[3]. In the techniques belong to the former[4,5], parameters are usually determined analytically based on the ideal pin-hole camera geometry. Simple and fast processing is the major merit of the approach. However, since the imaging process by most off-the-shelf camera systems is somewhat nonlinear, high accuracy may not be expected. To increase the accuracy, the nonlinearity due mainly to the lens distortion has been modeled and corrected[6-8]. When using the techniques belong to the latter[9,10], the unknown parameters of a camera are determined by iterative optimization. The solutions are relatively accurate but dependent upon initial guess for the iterative search and the system equations used are rather complicated.

We have tried to find an easy and accurate calibration method overcoming the practical difficulties in utilizing existing techniques. As a result, neural networks are employed for the stereo calibration. There are several encouraging facts in using neural networks for the problem including the nets' capabilities of nonlinear mapping, model-free learning, and function approximation[11,12]. When stereo cameras are used for 3D positional measurement, implicit calibration[2] is enough and this can be an additional advocating fact of using a neural net for the problem. We start from the previous research where neural networks are trained to learn the unknown part of camera model[13] and to localize a camera[14]. However, in this paper, neural networks are employed for the stereoscopic 3D position measurement rather than modeling or calibrating a single camera. Two ways of neural solution are studied and compared; a neural network for direct stereo-to-world mapping and neural learning the error pattern of a linear stereo model.

This paper organized as follows. In Section 2, a method for stereoscopic 3D position measurement is briefly described. Then, the techniques of employing feedforward neural networks are presented in Section 3. The experimental results are presented in Section 4 and conclusions are followed in Section 5.

2. STEREOSCOPIC 3D POSITION MEASUREMENT

Most camera calibration algorithms are based on the simple pin-hole camera model[15]. As shown in Figure 1, an image point at (u,v) can be related to a ray through a 3D world point $P(X,Y,Z)$. The projection model can be described by the following two equations:

$$u = (P - C, H)/(P - C, A) \quad (1.a)$$

$$v = (P - C, V)/(P - C, A) \quad (1.b)$$

where $C(C_X,C_Y,C_Z)$, $H(H_X,H_Y,H_Z)$, $V(V_X,V_Y,V_Z)$, $A(A_X,A_Y,A_Z)$ are the positional, horizontal, vertical, aiming vectors of the camera respectively and (M,N) is the scalar product of the vectors M and N. The derivation of the equations and the detail

meanings of the parameters are described in [4,15,16].

Using n number of control points and their image coordinates, eq.(1) can be rewritten as

$$X_m u_m A_X + Y_m u_m A_Y + Z_m u_m A_Z - X_m H_X - Y_m H_Y - Z_m H_Z - u_m C_A + C_H = 0 \quad (2.a)$$

$$X_m v_m A_X + Y_m v_m A_Y + Z_m v_m A_Z - X_m V_X - Y_m V_Y - Z_m V_Z - v_m C_A + C_V = 0 \quad (2.b)$$

where $m=1,\ldots,n$, $C_H = (C, H)$, $C_V = (C, V)$, $C_A = (C, A)$. In the case of $n>6$, the twelve parameters are over-determined by $2n$ equations and can be found by standard least-squares techniques.

Once one of the stereo cameras is calibrated, rearranging eq.(2) for a certain 3D point P yields the followings:

$$(uA_X - H_X)X + (uA_Y - H_Y)Y + (uA_Z - H_Z)Z = uC_A - C_H \quad (3.a)$$

$$(vA_X - V_X)X + (vA_Y - V_Y)Y + (vA_Z - V_Z)Z = vC_A - C_V \quad (3.b)$$

As two additional equations can be obtained using the other camera calibrated by the same way, four equations for three unknown positional coordinates are available. The unknowns, $X,Y,$ and $Z,$ can then be computed using the least squares.

Only linear equations are to be solved if the technique described so far is used. However, due to the nonlinearity practically existing in imaging process of a camera, the technique may not be accurate enough for some tasks especially in precise inspection and measurement. Assuming the nonlinearity comes mainly from the distortion by the lens, the distortion is tried to be approximated as the followings[6-8]:

$$u = u_D + u_D(k_1 r^2) + t_1(r^2 + 2u_D^2) + 2t_2 u_D v_D \quad (4.a)$$

$$v = v_D + v_D(k_1 r^2) + 2t_1 u_D v_D + t_2(r^2 + 2v_D^2) \quad (4.b)$$

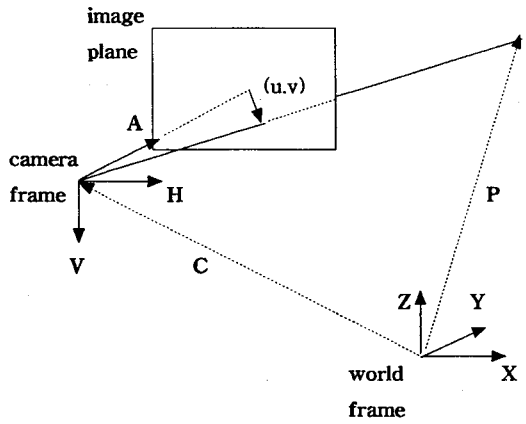


Figure 1. Imaging geometry

where u_D, v_D are distorted image coordinates, $r = (u_D^2 + v_D^2)^{1/2}$, k_1 is one term radial distortion coefficient, t_1 and t_2 are two-term tangential distortion coefficients of a lens. One problem of this nonlinearity compensation method is that there are still some parts not included in the model. So, Wen and Schweizer[13] tried to approximate the part remained outside of the model using a neural network. In this paper, however, we employed neural networks in different ways; for the direct stereo-to-world mapping and learning the error patterns of linear stereo vision model as described in the next section.

3. THE USE OF NEURAL NETWORKS FOR THE STEREO CALIBRATION

Multilayer feedforward neural networks are known to be capable of approximating any arbitrary continuous function[11,12]; if a neural net has n input units, m output units and at least one hidden layer whose nodes have sigmoid activation function,

the net can approximate a continuous mapping from n-dimensional Euclidean space to m-dimensional Euclidean space.

An efficient and probably the most widely used training algorithm for multilayer feedforward neural nets is the BP algorithm. Neural nets trained by the BP algorithm have many tempting features to be useful for the problem stated in this paper. First, since the BP is a supervised learning algorithm, it has a fundamental similarity with traditional calibration techniques, where the mapping error between control points and their stereo images are tried to be minimized. Second, an interconnection of nonlinear neurons can be helpful to overcome the limit of linear techniques. Particularly, increased accuracy and noise insensitivity are expected. Third, the massively parallel nature of a neural net makes it potentially fast for any computation task. Thus, the major advantage of linear system may not be hurt by employing a neural net. Finally, using a neural network is basically a model-free approach. Most existing techniques have tried to improve the accuracy by reflecting the real physical nature more precisely and usually required more complex model. For example, Tsai[7] tried to correct symmetric lens distortion and Faugueras and Toscani[6] later corrected even asymmetric lens distortion for higher accuracy. Neural nets, in comparison, have much larger adaptability and generality.

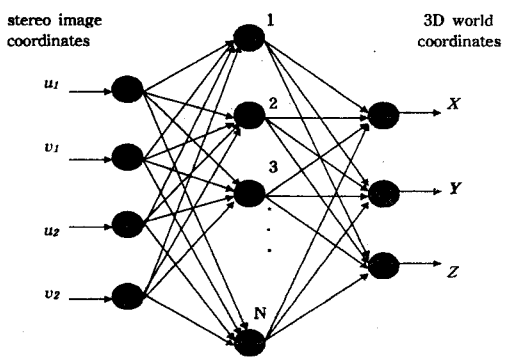


Figure 2. A neural net for stereo-to-world mapping

An intuitive neural net implementation for the problem given is in the architecture of Figure 2. The elements of input vector are four stereo image coordinates and those of output vector are three world coordinates. Nonlinear activation function such as sigmoid function employed for the neurons in hidden and/or output layers may be helpful for approximating the nonlinear mapping. The network can be trained using a number of control points and their stereo observations. For the generalization performance test after learning the mapping, some world points and their corresponding stereo images which are not used for the training are also needed.

Practically, however, a neural network in the architecture of Figure 2 may have several problems as we have found in our experimental study. One of them is that the performance of a network depends severely on its training data. Both the number and the properties of the data affect the performance. In existing camera calibration techniques, around twenty or more number of noncoplanar control points are enough to get reasonably accurate results. However, when the neural network are trained with this number of points, the accuracy is far worse than those obtainable by even the simplest linear method (as shown later in Section 4). Since collecting accurate control points and their stereo observations is not easy job, the requirement of a large number of data for the learning certainly is a limiting factor of using the network. Another practical difficulty is that the network may have very slow convergence rate. Although it is known that a neural networks with sufficient number of hidden neurons driven by the BP algorithm can approximate any continuous multivariate function to any desired degree of accuracy, it is not practically useful if the desired accuracy can not be obtained within reasonable time. One cause of the slow convergence, when applied into the back-projection applications, is that the error values we wish to meet are much less than the absolute output values. The convergence of network output is also dependent on network parameters such as learning rate, momentum, and

size of hidden layer. It is, however, difficult to decide them optimally and the optimality varies depending on data set even for the same problem domain.

By the reasons stated till now another way of neural net implementation is searched. We start with several observations; (a) linear techniques suffer from accuracy problem but they are fast and simple practically, (b) neural networks can learn any continuous nonlinear functions and camera systems are somewhat nonlinear, (c) neural networks converge very slowly when used for the stereoscopic measurements and a large number of data are required for the generalization.

A paradigm devised from these observations is shown in Figure 3, where $\hat{E}$ is the estimate of error between real 3D coordinates and those computed by a linear technique, and $\hat{P}$ is the estimated position after being corrected by the neural network. The network is used here to learn the error pattern of a linear technique using network's nonlinear function learning capability.

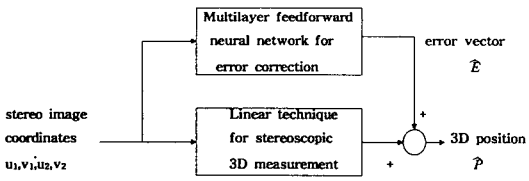


Figure 3. A neural net for correcting linear stereoscopic vision model

4. EXPERIMENTAL RESULTS

The performance of a feedforward neural network having the structure of Figure 2 is tested by an experiment. Real data of three calibration planes are used for the experiment: Points of two planes at near and far distances from the camera system and their stereo images are used for training and those of the middle plane are used for generalization performance test. The distance between the middle plane and cameras is about 2300([mm]; hereafter all lengths are expressed in millimeters) and the planes are placed at intervals of 200.

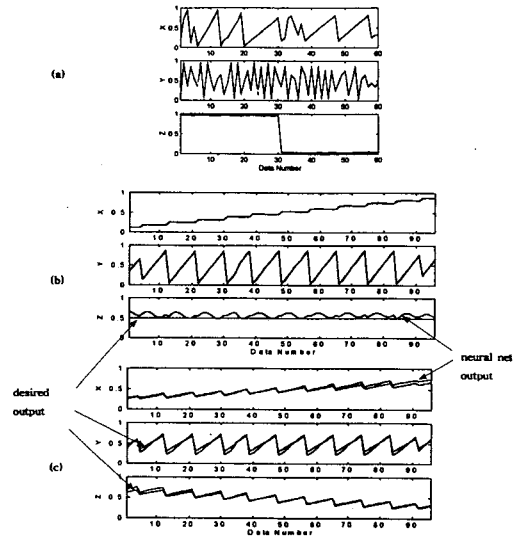


Figure 4. Approximation by a stereo-to-world mapping network; data are scaled between 0.05 and 0.95, 30000 iterations, number of hidden nodes=16
(a) learning with 60 training data,
(b) testing for generalization performance with 96 data
(c) testing for generalization performance with data defined in a rotated world frame

Figure 4 shows a test result. As shown in the Figure 4(a) the network can be trained to learn the mapping between the stereo and the world coordinates of given points. However, the approximation error of the Z coordinate is much larger than those of X and Y coordinates in a generalization performance test as shown in Figure 4(b). One probable reason is that the data distribution along Z axis is not diverse - points only two Z coordinates are used for the training. Although this difficulty may be avoided if we use the points well distributed along the all three axes, it is also practically common to use only limited number of calibration planes because of difficulties in precisely measuring the control points. We thus

rotate the world frame so that the calibration data can be represented evenly along the three axes without losing generality. By this simple process the accuracy can be improved from (average error, maximum error) = (41.43, 88.05) to (37.20, 57.09). Table 1 presents the summarized results of the experiment. These results are obtained after 20000 iterations with feedforward neural nets having different number of training points and nodes in a hidden layer. Sigmoid function is used for both hidden and output nodes. It is shown that the generalized mapping accuracy obtained is far worse than that of solving the linear equations described in Section 2. The error is defined as the average distance between the computed position and real position of 3D points as

$$1/n \sum_{m=1}^{n} \sqrt{(X_m - \hat{X_m})^2 + (Y_m - \hat{Y_m})^2 + (Z_m - \hat{Z_m})^2} \quad (5)$$

where n is the number of testing points, $(\hat{X}, \hat{Y}, \hat{Z})$ is the computed value for the real position (X,Y,Z).

number of training points	20			60		
method	linear solution	neural net		linear solution	neural net	
number of hidden nodes		8	24		8	24
average error	5.54	58.28	55.02	4.75	68.33	35.71
maximum error	10.75	75.05	71.87	10.11	84.33	52.89

Table 1. The error of 3D position computation by neural network compared with the results by solving linear equations

The learning of neural network for the error pattern of the linear system is also experimented. The desired output of the network is the error vector between the real coordinates of the training data and the actual solution of the linear system. After the learning, the network computes the error pattern for the generalization performance testing data and the pattern is added to the output of the linear system to estimate the real coordinates of the test data. Figure 5 shows an example of the net's error learning. Experimental results with different number of hidden nodes and training data are presented in Table 2. Significant reduction of error is noticed compared with the results of Table 1.

number of training points	20		60	
number of hidden nodes	8	24	8	24
average error	1.43	1.73	0.76	0.83
maximum error	6.88	6.86	2.38	2.60

Table 2. The error of 3D position computation by linear solution and neural correction

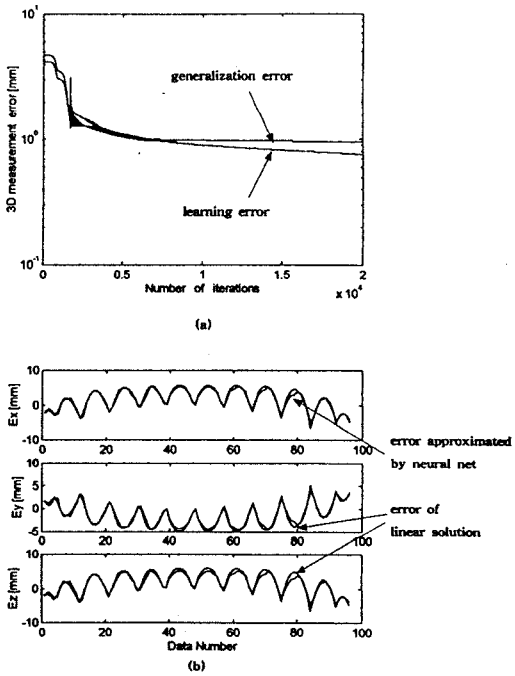


Figure 5. Neural correction for the error of linear 3D position measurement
(a) error plot: 60 control points for training data, 8 hidden nodes
(b) error of linear system solution and approximation by neural net for 96 generalization performance checking data

5. CONCLUSIONS

The problem of stereoscopic 3D position measurement has been dealt with. We first discussed the practical problems of existing camera calibration techniques. A neural network is then introduced to the problem as a method to tackle the problems. Training a neural network for direct stereo-to-world mapping is simple in concept but difficult to obtain high accuracy within reasonable training time. So, the neural stereo-to-world mapper can only be used for limited applications where simplicity is more important than the accuracy; for example, the world modeling of a autonomous navigator. Learning the error pattern of linear stereo model is another approach studied. Using the neural network's capability of nonlinear function approximation, the drawbacks of linear techniques can be largely overcome without introducing complex physical or mathematical modeling process. Since a neural network is in parallel computation architecture, the major advantages of using linear techniques, the simple and fast processing, are reduced only little when corrected by the network. As the demand for real-time stereo machine is very high, the paradigm proposed in this paper can be practically useful.

REFERENCES

[1] A.Pugh (ed.), *Robot Vision*, IFS, Bedford, 1983.

[2] G-Q.Wei and S.D.Ma, "Implicit and explicit camera calibration: theory and experiments," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.16, No.5, pp.469-480, 1994.

[3] M.Ito, "Robot vision modelling - camera modelling and camera calibration," *Advanced Robotics*, Vol.5(3), pp.321-337, 1991.

[4] Y.Yakimovsky and R.Cunningham, "A system for extracting three-dimensional measurements from a stereo pair of TV cameras," *Computer Graphics and Image Processing*, Vol.7, pp.195-210, 1978.

[5] S.Ganapathy, "Decomposition of transformation matrices for robot vision," in *Proc. IEEE Int. Conf. Robotics and Automation*, pp.130-139, 1984.

[6] O.D.Faugueras and G.Toscani, "The calibration problem for stereoscopic vision," in *Sensor Devices and Systems for Robotics* (A.Casals, ed.), NATO ASI Series, Vol.F52, Springer-Verlag, Berlin, pp.195-213, 1989.

[7] R.Y.Tsai, "A versatile camera calibration technique for high accuracy 3d machine vision metrology using off-the-shelf TV cameras and lenses," IEEE J.Robotics & Automation, Vol. RA-3, No.4, pp.323-344, 1987.

[8] H.Bacakoglu and M.S.Kamel, "A three-step camera calibration method," *IEEE Trans. Instrumentation and Measurement*, Vol.46, No.5, pp.1165-1172, 1997.

[9] I.W.Faig, "Calibration of close-range photogrammetric system: mathematical formulation," *Photogrammetric Engineering and Remote Sensing*, Vol.41, No.12, pp.1479-1486, 1975.

[10] K.W.Wong, "Mathematical formulation and digital analysis in close-range photogrammetry" *Photogrammetric Engineering and Remote Sensing*, Vol.41, No.11, pp.1355-1373, 1975.

[11] K-I.Funahashi, "On the approximate realization of continuous mapping by neural networks," *Neural Networks*, Vol.2, pp.183-192, 1989.

[12] K.M.Hornik *et al.*, "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks," *Neural Networks*, Vol.3, pp.551-560, 1990.

[13] J.Wen and G.Schweitzer, "Hybrid calibration of CCD cameras using artificial neural nets," *Int. Joint Conf. Neural Networks*, pp.337-342, 1991.

[14] D-H.Choi and S-Y.Oh, "Real-time neural network based camera localization and its extension to mobile robot control," *Int. J. Neural Systems*, Vol.8, No.3, pp.279-293, 1997.

[15] A.M.Thompson, "Camera geometry," in *Robotics*

Age; In the Beginning (C.T.Helmers, ed.), Hayden, Hasbrouck Heights, pp.102-109, 1983.

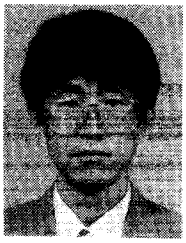
[16] L.A.Gerhardt and W.I.Kwak, "An imporved adaptive stereo ranging method for three-dimensional measurements," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp.21-26, 1986.

이 논문은 1998학년도 대구대학교 학술연구비 지원에 의한 논문임

著 者 紹 介

都勇兌 (Yongtae Do)

『센서학회지 제4권 4호』 논문 95-4-4-07, p54 참조, 현재 대구대학교 정보통신공학부 부교수



李大植 (Dae-Sik Lee)

1960년 5월 5일 생, 1982년 경북대 전자공학과 (공학사), 1984년 KAIST 전기 및 전자공학과 (공학석사), 1991년 KAIST 전기 및 전자공학과 (공학박사), 현재 대구대학교 정보통신공학부 부교수, 관심분야: 지능제어, 자동화 및 로봇공학

柳奭桓 (Seog-Hwan Yoo)

1956년 1월 3일생, 1975년 서울대 전기공학과 (공학사), 1979년 서울대 전기공학과 (공학석사), 1989년 University of Florida 전기공학과 (Ph.D), 현재 대구대학교 정보통신공학부 부교수, 관심분야: 견실제어, 신호처리 및 로봇공학