Application of Neural Networks for Stereo-Camera Calibration

Yongtae Do
School of Computer and Communication Engineering, Taegu University,
Kyungsan-City, Kyungpook, 712-714, Korea
ytdo@biho.taegu.ac.kr

Abstract

The position of a world point can be measured by the use of calibrated stereo cameras. Although simple linear methods for the calibration assuming an ideal projection model are available, their solutions are usually not accurate since most off-the-shelf lenses used in machine vision applications sustain considerable amount of nonlinear distortion. Recent research efforts on the problem have been thus concentrated on the modeling of lens distortion and its correction techniques. However, the types of lens distortion are various and the equations derived are more complicated if more precise model is employed for higher accuracy. In this paper, methods for calibrating stereo vision systems with neural networks are described. Different approaches are tested under various conditions and their results are compared.

1. Introduction

The calibration process is an important prerequisite for most computational vision tasks. Generally, a vision system is calibrated for largely two purposes; for computing the image coordinates from given world coordinates (projection) or for estimating the 3D position of a world point from its stereo image points (back-projection). Projection is primarily used in computer graphics, where an ideal camera model is often used. On the other hand, the purpose of back-projection is mainly to make position measurements for 3D applications including dimensional inspection and robotic manipulation. When cameras are used for back-projection applications, high accuracy in calibration is thus of importance.

Earlier camera calibration techniques usually employed a perfect pinhole camera model and the processes are simple and fast when applied [1]. However, due to the ignorance of nonlinearity that inherently exists in the imaging process, high accuracy is difficult to be obtained. A large portion of recent research work thus has been with the development of more precise camera models involving correction of lens distortion, which is identified as the major source of system nonlinearity. Tsai [2], for example, proposed an efficient technique which uses a series of equations to determine camera parameters in two calibration stages with a simplified model of radial lens distortion.

It should be noted, however, that techniques based on explicit models of physical vision system have a couple of practical disadvantages: (a) Optical features of cameras are different one another and a method that is proved to be effective for one system may be inefficient for others. It was reported that Tsai's method can be worse than even a simple linear method if lens distortion is relatively low [3]. (b) Practically no model is capable of describing a vision system perfectly. To get more accurate results, more elaborate modeling is required and this will bring more complicated mathematical equations. In Weng's model [4], for example, two additional types of lens distortion are considered besides radial distortion.

By the reasons stated above, techniques do not rely on explicit camera model may be useful in some applications. Stereoscopic 3D metrology is one of the examples, where physical camera parameters are not required to be identified if a 3D point observed by stereo cameras can be accurately determined by the use of some intermediate parameters. The two-plane method, originally proposed for back-projection purpose by Martins [5] and later developed further for even projection by Wei [6], may be referred to as the most widely known implicit technique. Wen [7], on the other hand, utilized a neural network to describe the part remained outside of the conventional explicit projection model. Considering a 3D coordinate can be represented in a straightforward manner from its stereo image coordinates, back-projection also is describable using a network [8]. Neural networks thus can be regarded as another effective way for implicitly calibrating a vision system. In this paper, methods for applying neural networks to stereoscopic metrology are studied and the results are comparatively analyzed.

2. Calibrating a Stereoscopic 3D Measurement System Using Neural Nets

Neural networks have several meaningful features for camera calibration. First, a network is made up of an interconnection of nonlinear neurons. Therefore, neural networks have the potential for learning the nonlinear imaging process. Second, the basic philosophies behind supervised neural net learning and camera calibration are the same. Both use a set of known data to find system parameters and apply the parameters later for the data unseen during the training stage. Third, using a neural network for a certain problem can be considered as a model-free solution for the problem, which is in common with the nature of implicit camera calibration techniques.

It is well known that multilayer neural networks can approximate any arbitrary continuous function to any desired degree of accuracy [9]. Thus, if we consider projection and back-projection as the mapping between world and image, the function identifying the mapping can be approximated without complicated mathematical modeling. In this section, different techniques for approximating projection and back-projection mapping functions using neural networks are described.

2.1. Direct mapping by neural networks

In stereoscopic back-projection, a 3D point is determined at the position where lines of sight ray from stereo image points intersect. A coordinate, *x,y,* or *z*, of 3D point can be described uniquely with two corresponding image points, i_1,j_1 and i_2,j_2, by a function *f*, which determines the back-projection of the stereo, *i.e.*,

$$p = f(i_1, j_1, i_2, j_2) \quad (1)$$

The function *f* cannot be represented in a closed-form as there is redundancy - only three of the four image coordinates are enough to compute the three coordinates of a world point [8]. In addition, the function is nonlinear due mainly to image distortion introduced by the lens used. Therefore, the calibration of a stereoscopic 3D measurement system includes correction of nonlinear lens distortion effect, determination of geometric and optical camera parameters and least squares of redundant measurements.

In this paper, the function of stereoscopic back-projection is approximated by neural network. The feasibility of the approach is based on the fact that multilayer feedforward networks are capable of approximating an arbitrary continuous nonlinear function [9] and solving least squares problem [10]. As a world coordinate *p* is obviously continuous, its images on the stereo are also continuous. So the mapping function between them is continuous too. Although, in reality, the position of an image on sensor plane is represented as a discrete value, it can be assumed to be continuous if precise measurement is made in a subpixel accuracy. For a two-layer network, an approximate realization of the function *f* can be represented by *g* like

$$g = a_2\left(\sum_{h=1}^{n} w_h a_1\left(\sum_{q=1}^{4} w_{hq} u_q - q_h\right) - q\right) \quad (2)$$

where a_1 and a_2 are activation functions, u_q is q^{th} input node value for a network having *n* number of hidden layer nodes. The goal of approximation is finding network parameters, *w* and *q*, satisfying the following for all input patterns and an arbitrary small positive value ε;

$$|f(i_1, j_1, i_2, j_2) - g(i_1, j_1, i_2, j_2)| \le e \quad (3)$$

Figure 1 shows the schematic diagram of the network designed to approximate the function of stereoscopic back-projection. Although stereo image positions are assigned to the input nodes in the figure, others such as higher orders of image coordinates also can be considered as additional input terms.

The concept of neural approximation of back-projection can be modified into the approximation problem of projection. In this case, a network should be constructed for each camera separately and input and output of the network are 3D world coordinates and their 2D image coordinates respectively.

There are two practical difficulties in neural approximation of projection or back-projection. First, the determination of proper network architecture for the problem is ambiguous like other neural net applications. It is difficult, for example, to find the optimal number of nodes in hidden layer. It is, however, proved that the number of hidden nodes for a two-layer network should equal *m-1* for perfectly learning *m* training data [11]. This can be upper bounds in determining the size of hidden layer; to avoid over-learning, the number of hidden nodes should be much less than the number of training data. Second, it generally takes too long time to achieve accurate approximation by training a neural network. The accuracy obtainable after reasonable training time cannot be comparable even with that of simple linear method. This problem can be tackled with the method described in the next subsection.

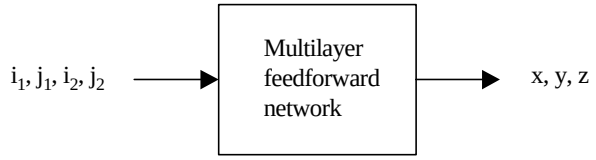

Figure 1. Approximation of stereoscopic back-projection using neural network

2.2. Learning the error of linear method

Certainly traditional calibration techniques based on ideal pinhole camera model have a practically important advantage; they are fast and simple as only linear equations in closed-form are needed to be solved. On the other hand, the disadvantage of the methods is also clear; they may not be accurate because the lens distortion effect is ignored.

From this observation, a linear method is used for 3D position estimation and its error is corrected by neural network. This approach is inspired by the work of Wen and Schweizer [7], where a neural network is employed to describe the unknown part remained after explicit calibration for projection. In this paper, however, the role assigned to neural network is the approximation of the error due mainly to lens distortion. It is expected that, with this approach, we can maintain the major advantage of linear methods and obtain improved accuracy without any complicated mathematical modeling process thank to nonlinear learning capability of neural network. As shown in figure 2, the network input consists of stereo image coordinates and the network is trained for *e*, 3D positional error of linear method. After training, the 3D position can be computed by summing output of linear method and the error correction term estimated by the network as $p_C=p_L+e_N$. The error *e* is due to the image error, say $\mathbf{e_U}$, for a set of image coordinates $\mathbf{u}=(i_1,j_1,i_2,j_2)$. On the other hand, $\mathbf{e_U}$ can be regarded as the sum of deterministic error $\mathbf{e_D}$ due mainly to lens distortion and random error $\mathbf{e_R}$ due to system uncertainty and electrical noise like

$$e_U = e_D + e_R \qquad (4)$$

where e_D for an image coordinate can be modeled as [4]

$$s_1(i'^2 + j'^2) + 3d_1 i'^2 + d_1 j'^2 + 2d_2 i' j' + k i' (i'^2 + j'^2)$$

for i coordinate, (5.a)

$$s_2(i'^2 + j'^2) + 2d_1 i' j' + d_2 i'^2 + 3d_2 j'^2 + k j' (i'^2 + j'^2)$$

for j coordinate. (5.b)

In the equations, (i',j') are ideal image coordinates measurable from optical image centre, s_1,s_2 are thin prism distortion parameters, d_1,d_2 are decentering distortion parameters, k is a radial distortion parameter. This error is deterministic and can be estimated by neural network while random error is difficult to be attacked with network.

This concept can be applied to the projection process either. In this case, a neural network can be employed to transform the distorted image to ideal image by training them to learn the error between real (distorted) image coordinates and ideal image coordinates. The corrected stereo images can be used for 3D position measurement as shown in Figure 3. Both input and output layers have only two nodes and the size of network can be smaller than that of the network built in Figure 2. But each camera of the stereo needs its own correction network

Neural networks used for back-projection and projection shown in figure 2 and 3 can be regarded as the postprocessor and preprocessor for the linear method respectively. The role of network is limited to a specific job while a network is used for approximating the entire back-projection (or projection) process in the direct mapping network scheme shown in figure 1.

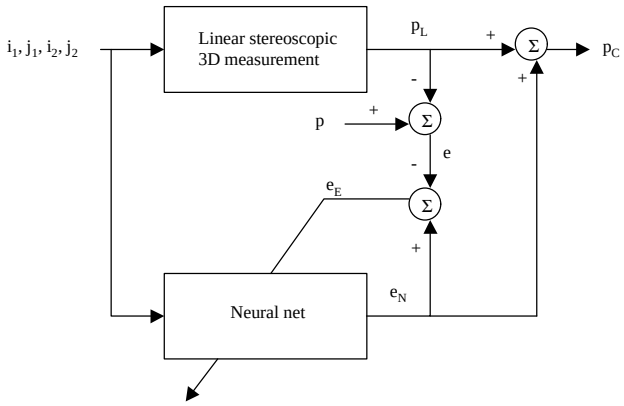

Figure 2. Linear method corrected by neural network

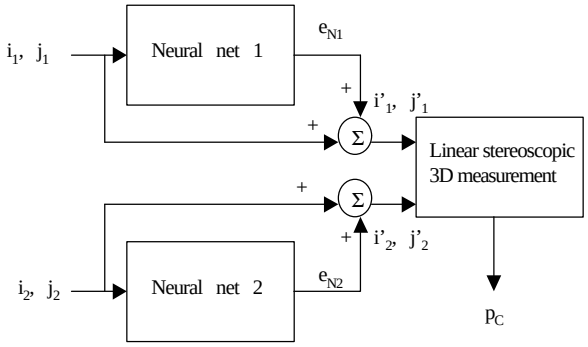

Figure 3. Neural nets used for correcting image distortion

3. Simulation Results

To evaluate the performance of techniques described in this paper, simulations with synthetic data are carried out. Two cameras are assumed to be positioned at (500, 260, 1000) and (550, 300, 1000) with z-y-x Euler angles of (180°, 0°, 5°) and (180°, 0°, -5°) respectively. All lengths are expressed in millimeters unless stated otherwise. Points at z=0 and 100 are used for calibration while points at z=50 are used for performance test. Optical parameters

assumed for both cameras are 15mm focal length, 11.25×15.00μm sized pixel, 512(H) × 480(V) pixels per image plane, and optical centre at (250,230). Three different data sets are assumed; data affected mainly by radial distortion, data affected by radial and considerable tangential distortion, and data corrupted by high level of noise in addition to radial distortion. The variance of Gaussian noise added is $\delta^2/12$ because the uniform quantization noise for pixel space δ equals $\delta^2/12$ in horizontal or vertical direction [4]. The noise level can be reduced if images are assumed to be measured in a subpixel accuracy. Specific parameters used for synthesizing the data sets are as the following:
Data type I: k=0.0003, $s_1=s_2=d_1=d_2=0$, additive noise by 1/5 subpixel accuracy measurement,
Data type II: k=0.0003, $s_1=0.0002$, $s_2=0.0005$, $d_1=0.0002$; $d_2=0.0001$, additive noise by 1/5 subpixel accuracy measurement,
Data type III: k=0.0003, $s_1=s_2=d_1=d_2=0$, additive noise by quantization without subpixel accuracy measurement.

For each data set, calibration methods described in the previous section are applied and the performance is observed. Table 1 shows the results, where method I is linear method, method II is direct back-projection neural network, method III is linear back-projection corrected by neural network, and method IV is linear back-projection after transformation from real image to ideal image by neural network. The average distance between real 3D points and computed points is used for the accuracy evaluation criterion. All the neural networks have 16 nodes at hidden layer and the results are obtained after 30000 training iterations by back-propagation algorithm. 60 data are used for calibration and 30 points are used for test.

From the table, it can be found that using a neural net with linear method is effective for the distorted data. The accuracy is improved by the factor of two approximately. However, by the reason described in section 2.2, the accuracy improvement is not significant for the data corrupted by significant random noise. Direct mapping by neural network, on the other hand, shows only rough estimation capability.

Table 1. Average 3D position measurement errors by various methods on different data sets

	Data type I	Data type II	Data type III
Method I	1.52186	1.81813	5.32947
Method II	7.08586	12.31885	15.15561
Method III	0.88862	0.90571	4.42610
Method IV	0.79186	0.95434	4.24230

4.Summary

Techniques for stereoscopic 3D position measurement using neural networks are studied. Several different approaches are studied and simulated with various data sets. Networks used for correcting errors of linear back-projection method and for transforming distorted image to ideal image are found to be useful for accurate measurement. Direct mapping by neural network provides the most straightforward and simple solution for the stereoscopic back-projection problem. However, high accuracy seems to be difficult to be obtained in reasonable training time. So the application of the direct mapping network may be limited to tasks where simplicity is more important than accuracy.

References

[1] 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.
[2] 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.
[3] S.-W.Shih, Y.-P.Hung, and W.-S.Lin, “When should consider lens distortion in camera calibration,” Pattern Recognition, Vol.28, No.3, pp.447-461, 1995.
[4] J.Weng, P.Cohen, and M.Herniou, “Camera calibration with distortion models and accuracy evaluation,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.14, No.10, pp.965-980, 1992.
[5] H.A.Martins, J.R.Birk, and R.B.Kelley, “Camera models based on data from two calibration planes,” Computer Graphics and Image Processing, Vol.17, pp.173-180, 1981.
[6] 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.
[7] J.Wen and G.Schweitzer, "Hybrid calibration of CCD cameras using artificial neural nets," Int. Joint Conf. Neural Networks, pp.337-342, 1991.
[8] Y.Do, S.-H.Yoo, and D.-S.Lee, “Direct calibration methodology for stereo cameras,” SPIE Conf. Vol. 3521: Machine vision systems for inspection and metrology VII pp.54-65, 1998.
[9] K.-I.Funahashi, "On the approximate realization of continuous mapping by neural networks," Neural Networks, Vol.2, pp.183-192, 1989.
[10] A.Cichocki and R.Unbehauen, “Simplified neural networks for solving linear least squares and total least squares problems in real time,” IEEE Trans. Neural Networks, Vol.5, No.6, pp.910-923, 1994.
[11] M.A.Sartori and P.J.Antsaklis, “A simple method to derive bounds on the size and to train multilayer neural networks,” IEEE Trans. Neural Networks, Vol.2, No.4, pp.467-471, 1991.