3D Modeling of Objects by Using Resilient Neural Network

Erkan Beşdok

Erciyes University, Engineering Faculty, Photogrammetry Division, Soft-Computing and Photogrammetric-Vision Lab., 38039, Kayseri, Turkey, ebesdok@erciyes.edu.tr

Abstract. Camera Calibration (CC) is a fundamental issue for Shape-Capture, Robotic-Vision and 3D Reconstruction in Photogrammetry and Computer Vision. The purpose of CC is the determination of the intrinsic parameters of cameras for metric evaluation of the images. Classical CC methods comprise of taking images of objects with known geometry, extracting the features of the objects from the images, and minimizing their 3D backprojection errors. In this paper, a novel implicit-CC model (CC-RN) based on Resilient Neural Networks has been introduced. The CC-RN is particularly useful for 3D reconstruction of the applications that do not require explicitly computation of physical camera parameters in addition to the expert knowledge. The CC-RN supports intelligent-photogrammetry, photogrammetron. In order to evaluate the success of the proposed implicit-CC model, the 3D reconstruction performance of the CC-RN has been compared with two different well-known implementations of the *Direct Linear Transformation* (DLT). Extensive simulation results show that the CC-RN achieves a better performance than the well-known DLTs in the 3D backprojection of scene.

1 Introduction

Camera calibration (CC) is an obligatory process for metric Shape-Capture, Robotic-Perception and 3D Scene Reconstruction in Photogrammetry, Photogrammetron and Computer Vision [1–9]. The purpose of CC is the computation of the physical imaging parameters of cameras for metric evaluation of the images. Classical CC methods [1–3] comprise of taking images of objects with known view-geometry of camera, extracting 2D positions of the objects from the images, and minimizing their 3D backprojection errors. CC has been studied extensively [1–9] in 3D Vision technologies (e.g. Photogrammetry, Photogrammetron and Computer Vision) and the proposed techniques in the literature include the techniques that use 3D calibration patterns, 2D calibration images and self-calibration. *Heuristic tools* (e.g. artificial neural networks, genetics, fuzzy systems) have also been used to solve some common Computer Vision problems such as Stereo Vision, Motion Tracking, Image Classification and Image Restoration.

The goal of implicit CC is calibrating view-geometry of the cameras without computing their physical parameters. Most of the classical CC methods [2, 3] use a fixed number of parameters which can limit the measurement-accuracy of

Proceedings of the ICVS Workshop on Camera Calibration Methods for Computer Vision Systems - CCMVS2007
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany

the CC and most of them require using complex optimization methods [1–3, 5] for the solution of nonlinear equations. This paper proposes a novel implicit CC method [4, 6–9] which describes a mathematical mapping between a point on an image plane and its corresponding 3D world coordinates. The proposed CC method benefits from the parameters of a Resilient Neural Network (RN) [10–12] to realize this mathematical mapping.

The proposed CC-RN can be used at calibrating a single camera to make 2D reconstruction and also can be used at the 3D reconstruction by using multiview images, which has been investigated in this paper. The proposed CC-RN describes the camera calibration parameters as the weights, bias values and transfer functions of the RN [10–12].

Full automation of the Photogrammetry which is referred as *Intelligent Photogrammetry* (Photogrammetron) [7] consists of ideas, methods and applications from Digital Photogrammetry, Intelligent Agents and Active Vision. Photogrammetron can only be possible with an autonomous and intelligent agent system such as Artificial Neural Networks [8, 9], Fuzzy Systems, Genetic Algorithms [4, 6], and other intelligent computing techniques. Photogrammetron is basically a Photogrammetric system that has a full functionality of Photogrammetry in addition to an intelligent agent system and a physical structure of active vision. It may have different forms as coherent stereo photogrammetron, separated stereo photogrammetron and multi-camera network photogrammetron. Photogrammetron can be used in various applications including photogrammetry-enabled robotic guidance, intelligent close-range photogrammetry, intelligent 3D multimedia-video indexing and real-time digital Videogrammetry.

A number of CC methods that implement an intelligent agent have been introduced in the literature [4, 6, 8, 9]. Most of these methods implement the structure of an intelligent agent either to learn the mapping from 2D image to 3D world coordinates or to improve the performance of other existing methods. Knowing that the CC parameters are important in various computer vision applications such as stereo-reconstruction, the CC-RN goes beyond the existing ones by providing 3D reconstruction from multi-view images besides stereo-images [1–9].

The proposed CC-RN adaptively models the imaging-sensor for each 3D backprojection points by using resilient-backpropagation learning method in order to achieve 2D to 3D mapping more accurately. The CC-RN can be used with automated active lenses and does not require a good initial guess of classical CC parameters.

The rest of the paper is organized as follows: *Resilient Neural Networks* are explained in Section 2. *Proposed Method* and *Experiments* are given in Section 3 and Section 4, respectively. Finally, *Conclusions* are given in Section 5.

2 Resilient Neural Networks

Artificial Neural Network (ANN) [8–12] is an advanced learning and decision-making technology that mimics the working process of a human brain. Various kinds of ANN structures and learning algorithms have been introduced in the literature [8–12]. In this paper, RNs have been used for the CC.

In contrast to other gradient algorithms, this algorithm does not use the magnitude of the gradient. It is a direct adaptation of the weight step based on local gradient sign. The RN generally provides faster convergence than most other algorithms [10–12]. The role of the RN is to avoid the bad influence of the size of the partial derivative on the weight update. The size of the weight change is achieved by each weight's update value, $A_{ji}(k)$, on the error function $E(k)$, which is used to calculate the delta weight as in Equation 1.

$$\Delta w_{ji}\left(k\right)=\begin{cases}-A_{ji}\left(k\right) & if\;\,B\left(k\right)>0\\ +A_{ji}\left(k\right) & if\;\,B\left(k\right)<0\\ 0 & else\end{cases} \tag{1}$$

where $B\left(k\right)$ is $\frac{\partial E}{\partial w_{ji}}\left(k\right)$ and

$$A_{ji}=\begin{cases}\eta A_{ji}\left(k-1\right), & B\left(k-1\right)B\left(k\right)>0\\ \mu A_{ji}\left(k-1\right), & B\left(k-1\right)B\left(k\right)<0\\ A_{ji}\left(k-1\right), & else\end{cases} \tag{2}$$

where $B\left(k-1\right)$ is $\frac{\partial E}{\partial w_{ji}}\left(k-1\right)$, η and μ are the increase and decrease factors, respectively where $0<\mu<1<\eta$.

More details about the algorithm can be found in [10–12].

A number of calibration methods implementing intelligent agents have been introduced recently [4, 6, 8, 9]. The intelligent agents based camera calibration methods introduced in the literature generally require a set of image points with their 3D world coordinates of the control points and the corresponding 2D image coordinates for the learning stage. The CC-RN can merge multisource camera images, for the reconstruction of a 3D scene, due to its high flexible properties of learning from examples.

3 Proposed Method

In this paper, the algorithm of the CC-RN and the preparation steps of the Learning Data and Validation Data used in the training RN structure of the CC-RN are explained below:

1. Determine the Image Coordinates of the Control Points (u,v), which are over the Calibration Patterns as seen in Fig.1-a and obtain 3D world-coordinates, (X,Y,Z), of the related (u,v). Each Calibration Pattern plane involves 11x7=77 points, hence, totally 539 calibration points over the seven Calibration Patterns planes have been used in this study.
2. Select 300 Control Points randomly for training and use the remaining 239 Control Points for validation of training in order to avoid from over-training disaster of RN structures.

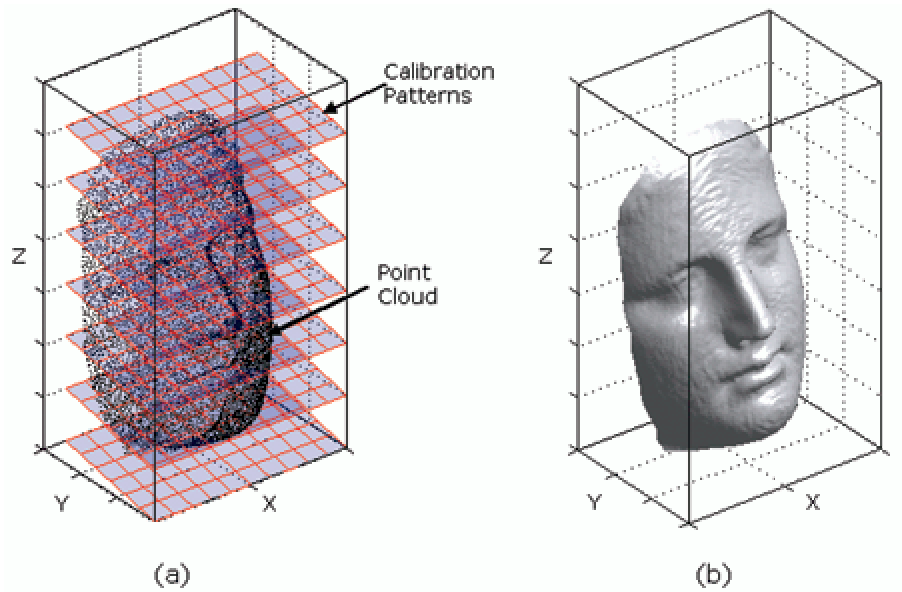

Fig. 1. (a) The seven Calibration Patterns and the Point Cloud obtained by using the proposed method for the Test Pattern. (b) Solid model of the Point Cloud which was obtained by using the proposed method for Test Pattern.

3. Train the RN-based neural structure that achieves a mapping from 2D (u,v) image space to 3D world-space (X,Y,Z) (Fig.1-b). The input data of the neural structure are the image coordinates (u,v) of the Control Points and the output data are the corresponding (X,Y,Z) world-coordinates of the input data (u,v).
4. Apply the image coordinates, (u,v), to the neural structure in order to compute the world coordinates, (X,Y,Z), of the related image points.

Extensive simulations exposed that accuracy of 3D reconstruction of the neural network systems increases with the increasing number of the cameras used in the imaging system. Therefore, the imaging system used in this paper comprises four virtual Fire-i IEEE 1394 cameras. Consequently, neural structures have eight inputs $(u_1, v_1, u_2, v_2, u_3, v_3, u_4, v_4)_p$ and three outputs $(X, Y, Z)_p$ for each point p.

The maximum epoch of the training phase was predefined as 250.000 for the neural structure but in order to control the computational burden of the proposed method, the training phase ended when the error goal of the RN was smaller than 0.001. Extensive simulations have been realized by using different initial conditions for neural structure and the best solution of problem of RN has been used in this paper. In the proposed method, the *Purelin Training Function* of the Matlab® [10] has been used for the neural structures.

Table 1. Mean-Squared-Error (MSE) values of the proposed CC-RN, DLT [1], and M-DLT [5] methods

	MSE of X	**MSE of Y**	**MSE of Z**
DLT (Train)	0.0025	0.1372	0.0008
DLT (Validation)	0.0027	0.1537	0.0010
M-DLT (Train)	0.0025	0.1374	0.0008
M-DLT (Validation)	0.0027	0.1543	0.0010
CC-RN (Train)	0.0022	0.1268	0.0009
CC-RN (Validation)	0.0024	0.1479	0.0010

4 Experiments

In this paper, a set of real images have been employed in the experiments of the CC-RN. The obtained results have then been compared with both DLT [1] and M-DLT [5] methods. Four virtual Fire-i IEEE 1394 cameras have been used in the experiments. Distortion models of the virtual cameras have been obtained by using the well known camera calibration toolbox given in [13]. The neural structure of the proposed method has $4_{Camera} * 2_{Image\ Corrdinates} = 8$ inputs. Each of the Calibration Pattern Planes have 11*7=77 Control Points. Image Coordinates (u,v) of the Control Points were acquired by using the well-known *normalized cross correlation matching algorithm* with subpixel accuracy for [13]. 10.000 measurements have been acquired over the Test Object. The seven Calibration Patterns and The Point Cloud obtained by using the proposed method for the Test Pattern was illustrated in Fig.1-a. The Calibration Patterns have been used for setting Calibration Data sets of the proposed method, DLT and M-DLT. Planimetric and depth reconstruction accuracies were evaluated in Mean-Squared-Error (MSE) as seen in the Table 1. Extensive simulations show that CC-RN supplies statistically acceptable and more accurate results as seen in Table 1. The method proposed in this paper was implemented by using the Image Processing, Image Acquisition and Neural Network toolboxes of Matlab® v7.3.0.

For the analysis of the CC-RN, the images obtained from the camera have been employed directly. The image coordinates of the control points have been extracted by employing the *normalized cross correlation matching* algorithm with *subpixel accuracy* and no deformation corrections have been applied to the images.

Before using the DLT methods (DLT and M-DLT), the required distortion corrections have been applied to the coordinates obtained as a result of the image matching. On the other hand, no distortion corrections have been applied to the corresponding coordinates in the CC-RN method.

Solid model of the Point Cloud which was obtained by using the proposed method for the Test Pattern was illustrated in Fig.1-b. Solid models obtained from 3D measurement systems are often noisy. Therefore, mesh smoothing is required for realistic mesh models. In this paper a variation of the method proposed in [14] has been used as a smoothing operator for illustration of Fig1-b.

4.1 Statistical Analysis

In the Manova [15, 16], mean vectors of a number of multidimensional groups are compared. Therefore, the Manova is employed to find out whether the differences in the results of CC-RN, DLT and M-DLT are statistically significant or not.

The tested null hypothesis is;

H_0: $\mu_1 = \mu_2 = \ldots = \mu$

H_1: at least two μ's are unequal

where $\mu_1$is the mean vector of the group # 1, μ_2 is the mean vector of the group # 2 and μ is the population mean vector.

As a result of the implemented Manova, no statistically significant difference has been found between the results of CC-RN, DLT and M-DLT. That means the null hypothesis cannot be rejected. This hypothesis test has been made using Wilk's Λ and χ^2 tests. The details of both of the tests and Manova can be found in [15].

For the Wilk's Λ test, the test and critical values are computed as 0.9953 and 0.7244, respectively, given that α significance level is 0.05 and degrees of freedom for the sum of squares and cross-products are 29997 and 29999, respectively. Due to the condition of $\Lambda_{test} > \Lambda_{critic}$, the null hypothesis cannot be rejected.

For the χ^2 test [15], the test and the critical values are computed as 2.00 and 6.00, respectively, given that α significance level is 0.05 and degrees of freedom is 3. Due to the condition of $\chi^2_{test} < \chi^2_{Critic}$, the null hypothesis cannot be rejected.

That is to say that there is no statistically significant difference between the results of CC-RN, DLT and M-DLT.

This outcome statistically verifies the advantages of the CC-RN method in various perspectives.

5 Conclusions

A Resilient Neural Network based camera calibration method for 3D information recovery from images is proposed in this paper. The obtained results have been compared with the traditional DLT [1] and M-DLT [5].

The main advantages of the CC-RN are as follows: It does not require the knowledge of complex mathematical models of view-geometry and an initial estimation of camera calibration, it can be used with various cameras by producing correct outputs, and it can be used in dynamical systems to recognize the position of the camera after training the ANN structure. Therefore, the CC-RN is more flexible and straightforward than the methods introduced in the literature.

The advantages of the CC-RN may be summarized as follows:

- Offers high accuracy both in planimetric (X,Y) and in depth (Z).
- Simple to apply and fast after training.

- Suitable for robotic 3D-vision.
- The CC-RN can use same or different type cameras together or separately.
- Optimization algorithms are not employed during 3D reconstruction in contrary to the same well-known 3D acquisition methods.
- Does not use physical parameters of cameras.
- Users do not need to know complex mathematical models of the view-geometry, therefore, proposed CC-RN does not need to have detailed technical knowledge on CC.
- Does not have any camera type restriction.
- An approximated solution for initial step of CC is not employed.
- No image distortion model is required.
- The CC-RN is flexible and significantly easier to implement than the CC methods presented in the literature.

Acknowledgments

This study has been supported by the Research Foundation of Erciyes University under the project entitled "3D Laser-Scanner Design by Using Soft-Computing", 2007.

Bibliography

[1] H., Hatze, High-Precision Three-Dimensional Photogrammetric Calibration and Object Space Reconstruction Using a Modified DLT Approach., J. Biomechanics, 21, 533–538, (1988).

[2] R., Tsai, A Versatile Camera Calibration Technique for High Accuracy 3d Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses, IEEE Journal of Robotics and Automation, 3,4, 323-344, (1987).

[3] Z., Zhang, A Flexible New Technique for Camera Calibration, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 11, 1330-1334, (2000).

[4] S., Hati, S., Sengupta, Robust Camera Parameter Estimation Using Genetic Algorithm, Pattern Recognition Letters, 22, 289–298, (2001).

[5] H.J., Woltring, R., Huiskes, Stereophotogrammetry, in Biomechanics of Human Movement: Application in Rehabilitation, Sports and Ergonomics, Bertec Corp., 108-127, (1990).

[6] J., Qiang, Z., Yongmian, Camera Calibration with Genetic Algorithms, IEEE Transactions on Systems, Man, and Cybernetics- Part A: Systems and Human, 31, 2, 120–130, (2001).

[7] H.P., Pan, C.S., Zhang, System Calibration of Intelligent Photogrammetron, International Archives of Photogrammetry and Remote Sensing, 34, 2, (2002).

[8] M.T., Ahmed, E., Hemayed, A., Farag, Neurocalibration: A Neural Network that can Tell Camera Calibration Parameters, Proc. of the International Conference on Computer Vision, Korfu, Greece, 1, 463–468, (1999).

[9] J., Jun, K., Choongwon, Robust Camera Calibration Using Neural Network, IEEE Region 10 Conference TENCON 99, 1, 1694–697, (1999).

[10] Mathworks Inc., Matlab, Neural Networks Toolbox, Image Poccessing Toolbox, Image Acquisition Toolbox, Mathworks, (2005).

[11] M., Reidmiller, Rprop- Description and Implementation Details, Technical Report, University of Karlsruhe, Germany, (1994).

[12] M., Reidmiller, H., Braun, A Direct Adaptive Method for Faster Backpropogation Learning: The Rprop Algorithm, Proceedings of the IEEE Int. Conf. On Neural Networks, San Francisco, CA, 586–591, (1993).

[13] J.Y., Bouguet, http://www.vision.caltech.edu/bouguetj/calib_doc/, (2004).

[14] P., Civicioglu, Using Uncorrupted Neighborhoods of the Pixels for Impulsive Noise Suppression with ANFIS, IEEE Transactions on Image Processing, 16, 3, 759-773, (2007).

[15] A.C., Rencher, Methods for Multivariate Analysis, Wiley-Interscience, John Wiley-Sons , Inc., (2002).

[16] P., Civicioglu, E., Besdok, Implicit Camera Calibration by Using Resilient Neural Networks, Lecture Notes in Computer Science, 4233, 632-640, (2006).

Proceedings of the ICVS Workshop on Camera Calibration Methods for Computer Vision Systems - CCMVS2007
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany