Zoom-lens Camera Calibration From Noisy Data With Outliers

Moumen T. Ahmed and Aly A. Farag
Computer Vision and Image Processing Laboratory
University of Louisville
Louisville, KY 40292, USA
moumen,farag@cvip.uofl.edu

Abstract

Camera systems with zoom lenses are inherently more useful than those with passive lenses due to their flexibility and controllability. However, their calibration raises several challenges. In this paper, we present a neural framework for zoom-lens camera calibration that can capture complex variations in the camera model parameters across continuous ranges in the lens control space, while minimizing the calibration error over all the calibration data. To automate the tedious process of collecting calibration data, the calibration approach should be prepared to handle possible outliers in the data. We demonstrate how the calibration approach can be robust and less sensitive to outliers. The validity and performance of our approach are tested using both synthetic data with outliers, and with real experiments to calibrate two Hitachi CCD cameras with H10x11E Fujinon active lenses.

1 Introduction

The goal of machine vision is to understand the visible world by inferring 3D properties from 2D images. Making such an inference requires modeling of the relationship between the 2D images and the 3D world. Camera calibration is a process which models this relationship. In this work, we report our approach to calibrate zoom-lens cameras that are part of the CardEye, an active vision system developed at our lab. The system has a trinocular head with basic mechanical properties such as pan, tilt, roll, focus, zoom, aperture, vergence and baseline. The system uses an active lighting device (a laser pattern generator with different diffractor filters), located at the center of the head, to assist in surface reconstruction process and other tasks. The system makes use of the flexibility and controllability of the zoom-lens cameras to improve the performance of the system in its tasks. To achieve this, the cameras should be calibrated across continuous ranges of zoom and focus that cover the working space of the system. However, calibrating cameras with zoom-lenses is rather difficult and raises several challenges. First, the dimensionality of data collected for calibration is large. A second challenge is the potential difficulty in taking measurements across a wide range of imaging conditions (e.g., defocus and magnification changes) that can occur over

the range of zoom and focus control parameters.

The calibration problem of these cameras relies on formulating functions that describe the relationships between the camera model parameters and the lens settings. This is usually achieved by calibrating a conventional static camera model (commonly the pinhole model) at a number of lens settings which span the lens control space using traditional calibration techniques. The calibrated model parameters at each lens setting are then stored in lookup tables [2],[8], or polynomials (or perhaps other functions) are formulated to model the parameters [9],[5],[4]. The following remarks can be drawn on the previous approaches:

- using interpolation to obtain each model parameter at intermediate lens settings in tables or fitting a function to each parameter independently emphasizes only the fitting error for this particular parameter at the expense of of the overall calibration accuracy, which is completely ignored in this step. This step, subsequently, does not consider the interaction between all the model parameters to represent the underlying camera model.

- Polynomials, in many cases, fail to follow the complex variations in some model parameters. Although other alternatives such as exponential functions, Chebyshev polynomials and Legendre polynomials can be exploited, the question about the optimal (or best) function type often remains difficult to answer. Lookup tables have been previously used to get around this problem.

To remedy the first point, an additional step of global optimization over all the calibration data is carried out in [10], [9],[4] to refine the coefficients of fitted polynomials. This step involves optimizing for each polynomial in turn until one (or more) cycle is completed for all the parameters. However, as noted in [4], the sequence of fitting the polynomials to the data affects the final calibration error. Therefore, several experiments may be needed to reach a good sequence.

To calibrate the CardEye's zoom-lens cameras, we resort to proven power of multi-layered feedforward neural networks (MLFNs) as universal approximators [6] to provide suitable parameter formulation/fitting. This can take care of the second point, but one still needs to consider the interaction between these MLFNs in such way that minimizes the calibration error over all data. We have recently [11] devised a solution by introducing the neurocalibration approach, which casts the camera calibration problem into a learning problem of a MLFN. Therefore, we propose to add the neurocalibration network to a number of MLFNs, thus building an all-neural framework in which all the MLFNs learn concurrently, independently and cooperatively, to capture the variations of model parameters across continuous ranges of lens settings. This framework can provide handy solutions to the previous two points.

In an addition to a brief description of our approach to zoom-lens calibration, in this paper, we address another related problem. Collecting calibration data that covers the working space of the system is rather tedious. So we are planning to have the system collect the data by itself. This is can be accomplished by using special markers embedded in the system working space. Moreover, we are currently investigating the integration of the active lighting device with a range finder such that the two use the same laser beam. The integrated active device

will have two modes: as pattern generator and as range finder. Such a device is useful in both surface reconstruction and getting data for calibration. Automating the data collection phase makes the system easy to re-calibrate whenever necessary. However, it may introduce outliers in the data due to poor localization of the markers or the laser pattern in the captured images, or due to incorrect correspondence between 3D points and their corresponding image pixels. It is worth mentioning here that usually off-line calibration is done under human supervision and outliers are less likely to occur. That is why, to the best of our knowledge, we are not aware of any previous calibration work that has dealt with outliers in the data.

To cope with the new possibility of outlying calibration data, a robust approach to calibration should be pursued; One bad outlier would skew the results of any approach based on the widely-used least squares estimates. We therefore propose to use the maximum likelihood estimator *M-estimator* [13] as a more robust estimator.

The rest of the paper is organized as follows. We describe the camera model and state the calibration problem in the next section. In Section 3, we briefly describe the neurocalibration network, followed by a more robust approach in Section 4. A brief overview of the framework for zoom-lens calibration is given in Section 5. Sections 6 and 7 describe our experiments and our concluding remarks.

2   Camera Calibration Problem

The result of camera calibration is an explicit transformation that maps a 3D world point $M = (X, Y, Z, 1)^\top$ into a 2D pixel $m = (u, v, 1)^\top$. This mapping can be represented by a 3×4 projection matrix, $\mathbf{P}$, that encompasses 11 physical parameters: rotation angles R_x, R_y and R_z, translations t_x, t_y and t_z, the coordinates of the principal point (u_0, v_0), two scale factors α_u and α_v, and the skewness c between the image axes. This camera model thus ignores lens distortion which is often accounted for in the camera model by adding some distortion parameters [1]. However, these parameters can be estimated in the captured images by a pre-calibration process [3],[7]. Then the images (or image features) can be undistorted before calibration proceeds. The decoupling between distortion parameters from the others will allow us to maintain the simple relation of the distortion-free model thus making following vision tasks (e.g., stereo reconstruction) easier. Moreover, the decoupling would reduce the effect of the correlation between lens distortion coefficients and other camera model parameters [10] on parameter estimation. The analysis and effects of this pre-calibration process are deferred to a follow-up work.

Given a sufficient number, N, of reference world points, $\mathbf{M}_i$, as well as their corresponding pixel positions, $\mathbf{m}_i$, the camera calibration problem is to estimate the 11 camera parameters or the projection matrix $\mathbf{P}$, that minimize

$$E = \sum_{i=1}^{N} \parallel \mathbf{P} \ \mathbf{M}_i - \mathbf{m}_i \parallel^2 . \tag{1}$$

3   Neurocalibration

In [11] we proposed a MLFN (see the central net in Fig. 1) that not only learns perspective projection mapping of a camera, but also can specify explicitly the calibration parameters. Here, we give a brief summary of our approach. The interested reader can refer to [11],[15] for more details. The network tries to minimize the calibration error in (1), which after some manipulation can be put in the familiar form

$$E = \sum_{i=1}^{N} \sum_{j=1}^{3} (d_{ij} - O_{ij})^2, \tag{2}$$

where $\mathbf{d}_i = \mathbf{m}_i$ and $\mathbf{O}_i$ denotes network output vector. After training, the projection matrix can be computed from the network weights [11]. Moreover, the net can explicitly solve for the camera model parameters after mapping each network weight to one camera parameter by enforcing the orthogonality constraints on the rotational part of the model during network learning [11].

Our extensive simulations and tests on practical images [11] yielded very low calibration error and have shown that this neurocalibration approach has the following features:

- it relaxes the requirement of a good initial starting point, which is common to other non-linear optimization techniques (e.g., Levenberg-Marquardt algorithm). These techniques often fail without this condition. In all the experiments conducted, the network has converged starting from random initial weights without sacrificing the calibration accuracy.

- experiments have shown very small sensitivity of the network learning to network parameters, e.g., learning constants.

- the optimization procedure takes account of the structure of the orthonormal rotation matrix without extra optimization steps (e.g., [1]).

- the technique is completely parallel thanks to its neural basis. Speedup can be greatly achieved if implemented on parallel computers (or even onto VLSI chips).

4   Robust Neurocalibration

Various machine vision algorithms found in literature optimize a least squares criterion, which is optimum and reliable when the underlying noise in the data is Gaussian. However, when outliers are present in the data, the Gaussian assumption is violated and the least squares result is skewed. During the last three decades, many robust techniques have been proposed [12],[13],[14] to handle outliers and these techniques have gained popularity in computer vision. Robust estimates include M-estimates (Maximum likelihood estimates), L-estimates (linear combination of order statistics), R-estimates (estimates based on rank transformations) and LMedS estimates (Least Median Square). If r_i denotes the residual error of the i-th data item, M-estimators try to reduce the effect of outliers by

minimizing another function of the residuals, $\sum_i \rho(r_i)$, where ρ is a symmetric, positive-definite function with a unique minimum at zero, and is chosen to be less increasing than square. Many of such ρ function have been suggested [13],[14], which yield breakdown points of about $1/p$, where p is the number of unknowns ($p = 11$ in case of camera calibration). M-estimators have high efficiencies, typically 0.9 [14].

To robustify our calibration approach, instead of minimizing the error in (2), the network will minimize

$$E = \sum_{i=1}^{N} \sum_{j=1}^{3} \rho(d_{ij} - O_{ij}). \tag{3}$$

We have selected for the function ρ the redescending function suggested by Tukey [14], which has provided better results as compared to other function such as Huber's function [13]. The new updating rules for networks weights are re-derived according to the new error in (3).

It is important to note that M-estimate methods tend to be extremely susceptible to the initial solution to the non-linear optimization method. Most calibration approaches found in the literature initialize a non-linear optimization algorithm with the closed-form solution of a least squares linear approach. However, due to outliers, the linear approach would not provide a reliable initial solution to the non-linear optimization algorithm, which subsequently fails to yield any improvement. Our neural approach does not suffer from this problem since it starts with random initial weights.

After network training, one can make a good, *robust* estimate of the standard deviation of the errors of good data (inliers). This estimate is related to the median of the absolute values of the residuals, $\hat{\sigma}$ [12]. Any data item whose whose error is larger than a certain number (e.g. $2.5 - 3.0$) of $\hat{\sigma}$ can be considered as an outlier and removed. We are currently investigating another approach based on LMedS estimates [12], which theoretically has the largest possible breakdown point (0.5) but lower efficiency and higher computational complexity.

5 Zoom-Lens Camera Calibration

In this section, we outline our framework for zoom lens calibration in the following three steps.

5.1 Passive Camera Calibration

The calibration process starts with collecting the calibration data at a number of different lens optical settings (zoom, focus and/or aperture) covering the operating space of the system. At each fixed lens setting, the fixed camera model parameters are estimated using the neurocalibration technique (any other calibration technique may be used for this step). If data is likely to have outliers, the robust neurocalibration technique is used, then outliers are identified and thrown away.

5.2 Initial Parameter Formulation

Having obtained the parameter values at the different positions, we are ready to fit functions to these values. The skew, c, is fixed to zero. For ease of use of the camera model, excluding t_z, the position and the orientation of the camera coordinate frame relative to the world coordinates are kept unchanged as the lens parameters are varied [4],[10]. Therefore, R_x, R_y, R_z, t_x and t_y are modeled with constants (zero-order terms). Constraining these parameters to be independent of optical settings makes use of the extra-degrees of freedom in the calibration (i.e., the dependency and correlation between some camera parameters for small variations [10]). The initial values for the former five terms are set to their average values throughout the whole data. Then, for each of the remaining parameters, $\alpha_u, \alpha_v, u_0, v_0$, and t_z, a function is fitted across the optical settings using a MLFN. The topology of each (from now on, *parameter*) MLFN is determined experimentally independently from the others.

5.3 Global Optimization

The neurocalibration net has a central role in the global optimization step. The five zero-order terms are represented by five network weights, while the five parameter MLFNs serve to provide the central network with the rest of the weights representing the parameters that vary with the lens optical setting. Fig. 1 illustrates the central neurocalibration network and its associated MLFNs. The novel
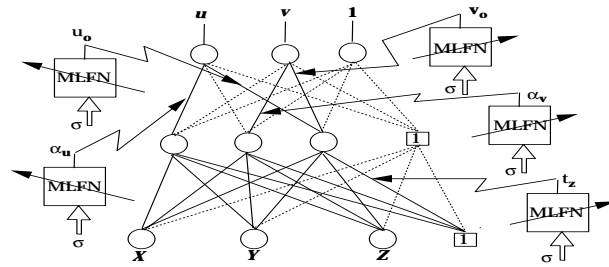

Figure 1: Global optimization step: five parameter MLFNs cooperate with the central neurocalibration net during the optimization, σ denotes the lens controllable optical parameters and dashed links in the central network designate link weights fixed during learning at 0 or 1.

network of neural networks, shown in Fig. 1, presents a non-typical structure of feedforward neural networks. We have developed an extended variant of the known Backpropagation algorithm to train these networks minimizing the overall calibration error for all the collected calibration data. At each lens setting, the outputs of the five parameter MLFNs provide the corresponding five model parameter values to the central network, which uses these values along with its weights representing the zero-order term parameters to project all the input vectors, to compute the calibration error at this lens setting, and to update all its weights. Five of the updated weights are propagated back to the parameter MLFNs to update their

functional mapping between the parameters and the lens settings. Therefore, the formulations of the parameters are updated in this step in a way that minimizes the overall calibration error (this functions as an automated change and refinement of the kernels or function types used for fitting). Note that each parameter MLFN minimizes its own fitting error, which is different from the calibration error computed by the central neurocalibration network. However, the fitting error of each parameter MLFN affects the calibration error. While it is rather difficult to prove it mathematically, the training process of all networks has indeed converged to a small calibration error in all of our experiments. Each parameter MLFN, then, will have the final functional relationship of that particular parameter versus lens settings while the central network will have the final values of the zero-order term parameters. More details about the training algorithm and the calibration approach can be found in [15].

6   Experimental Results

In this section, the performance of the robust calibration technique is tested with synthetic data with outliers. Then a real experiment to calibrate the CardEye's cameras is briefly described.

6.1   Calibration from Synthetic Data with Outliers

Using a set of specific external and internal camera parameters, a set of 440 3D points are projected to form a set of 2D points in an image of size 242×320. Then, Gaussian noise with standard deviation 0.2 pixels is added to the 2D image coordinates to represent the uncertainty in detecting these 2D points. A fraction, η, of the 2D points is selected and replaced with points generated randomly from uniform distribution that spans the whole image. Using the set of 2D points with the introduced outliers and the set of 3D points, we have performed a series of calibration experiments using different η. Figure 2(a) shows the set of 2D points at $\eta = 5$ and the obtained residual plot of the points after minimization of the error in (3) using our network is depicted in Fig. 2(b). The points with gross error correspond to the actual outliers in the data, and thus they can be identified and removed. For comparison sake, two other calibration techniques have been tested: a linear calibration approach and a nonlinear technique that starts with the solution of the linear approach minimizing the error in (1) using the well-known Levenberg-Marquardt algorithm. For each approach, the obtained projection matrix is used to project the 3D points. Then the *rms* error of the deviations between the point projections and the correct ones is computed and plotted against η in Fig. 2(c). Both the linear and Levenberg-Marquardt algorithm minimize a least squares criterion and thus they are largely affected by outliers in the data. The robust neurocalibration approach is less sensitive to outliers and produces error within 2 pixels till about 12% outlier percentage, which is around the theoretical upper bound of the breakpoint (9%). Fig. 2(d) shows the performance of our approach in terms of the percentage of correctly identified outliers versus η.
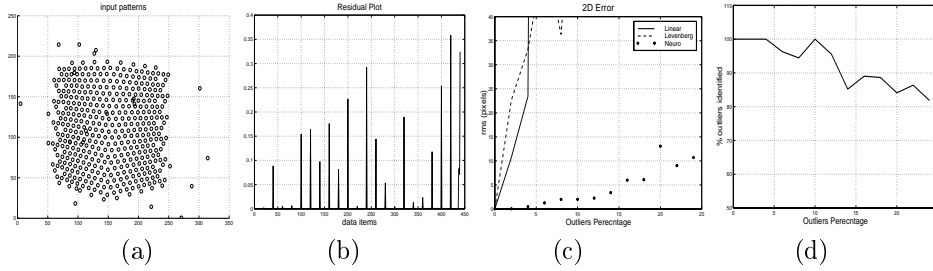
Figure 2: Robust calibration: (a) 2D points with 5% outliers (b) Residual plot of error at $\eta = 5$, (c) rms of point errors in pixels for three approaches and (d) Outlier identification rate versus η.

6.2 Zoom-lens Calibration

In this section, due to space limitations, we briefly describe the experiments conducted to calibrate two Hitachi HP-M1 CCD cameras with H10x11E Fujinon active lenses that are part of the CardEye system. More experiments and analysis can be found in [15]. For the two cameras, calibration data collection was human-supervised to easily quantify the performance of our approach, therefore no outliers were present. For the operating range, we have chosen a normalized focus range of $0.5 \leq m_f < 1.0$ and zoom range of $0.5 \leq m_z < 1.0$. We used a regular 7×7 sampling of the selected zoom and focus ranges. At each lens setting, an image of a checkerboard calibration pattern was captured by the camera, thus a total of 49×440 calibration points were obtained. The calibration approach, explained in Section 5, was applied to the collected calibration data of each camera independently.

For comparison sake, we applied Willson's approach [4] to the same collected calibration data (but using the camera model in Section 2 instead of Tsai's). In the global optimization step using the Levenberg-Marquardt algorithm in Willson's approach [4], the sequence of fitting the parameter polynomials to the data affects the final calibration error. So a greedy algorithm was used to find the best sequence. The rms of calibration error in pixels is shown in Table 1 for the two cameras, computed over all data points, before and after the global optimization step for the two methods.

Since we have not imposed on the calibration procedure the fact that the aspect

	Camera 1		Camera 2	
Approach	Init	Fin	Init	Fin
Willson's	5.41	1.82	6.21	1.30
Ours	2.53	0.12	1.88	0.14

Table 1: Comparison of the rms of calibration error in pixels before and after global optimization between our proposed approach and Willson's.

ratio, α_v/α_u, should be nearly constant (from our earlier experience with these

cameras, it is equal to 1) across the different lens settings, we used this to assess the results of our calibration. Our results, for both cameras, have shown an aspect ratio of 1 ± 0.05 across zoom and focus calibrated ranges.

Moreover, since we often captured images of the same calibration patterns by both cameras, one can compute for these images the rms of the error in 3D reconstruction of the calibration pattern. This will serve as a quantitative measure of calibration accuracy. Table 2 shows this measure computed, using 20 images, for the two approaches using the calibrated parameters before and after the global optimization step. In fact, we have used this measure to validate the different parameter models and to circumvent over/under-fitting given the the size of the available data. The third column shows the average run-time of the two approaches on an SGI-Indigo2 machine. Although the neural approach takes longer time, time can be greatly reduced if the approach is implemented in parallel. Moreover, this type of calibration is done off-line so accuracy not speed is the major issue. The calibration accuracy can be further improved if the images are corrected for lens distortion before calibration. Moreover, we can make use of more sampling positions (and thus more collected data size) to improve the accuracy. One can

Approach	Initial	Final	Time (min.)
Willson's	2.65	0.91	6
Ours	1.41	0.32	16

Table 2: Rms of 3D reconstruction error in cm before and after global optimization.

conclude from Tables 1 and 2 the importance of the global optimization step to decrease the calibration error and that our approach is more able to capture the variations of the parameters across the lens settings.

7   Conclusions

We have presented a neural framework for zoom-lens camera calibration, which makes use of our recently-introduced neurocalibration approach [11]. To automate the tedious process of collecting calibration data, the calibration approach should be able handle possible outliers in the data. We have demonstrated how our approach can be robust and less sensitive to outliers. Our experimental results have demonstrated better performance of our approach compared to Willson's approach, which is a reference of this domain. To improve the accuracy of our approach, more sampling positions during data collection are needed, along with removal of lens distortion [3],[7] from images before calibration. These two goals define our future directions in this work. We believe that this approach has the following key features, as opposed to other techniques (e.g., [9],[5],[4],[10]):

1. it is general; it can consider, in a straightforward manner, any number/combination of lens control parameters, e.g., zoom, focus and/or aperture.

2. Since no *a priori* knowledge about how lens settings affect the model parameters can be assumed available, our framework is flexible enough to capture

complex variations in the model parameters across continuous ranges of control space.

3. it integrates parameter formulation with the minimization of the overall calibration error; the formulations for all model parameters are refined at the same time, while in other approaches [9],[5],[4],[10], one parameter is fitted at a time and the final level of error generally depends on the sequence in which the models are fit to the data.

References

[1] J. Weng, Paul Cohen and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *PAMI*, Vol. 14, No. 10, Oct 1992.

[2] Mengxiang Li and Jean-Marc Lavest, "Some aspects of zoom lens camera calibration," *PAMI*, Vol. 18, Nov. 1996.

[3] R. Swaminathan and S. Nayar, "Non-metric calibration of wide-angle lenses and polycameras," *Proc. CVPR99, Fort Collins, Colorado*, June 1999.

[4] R. G. Willson, "Modeling and calibration of automated zoom lenses", *PhD dissertation, Dept. Elect. Comp. Eng., Carnegie Mellon Univ.*, 1994.

[5] A. Wiley and K. Wong, "Geometric calibration of zoom lenses for computer vision metrology," *Photogrammetric Eng. Remote Sensing*, Vol. 61, No. 1, jan. 1995.

[6] F. Hornik, "Multilayer feedforward networks are universal approximators," *Neural Networks*, Vol. 2, 1989.

[7] B. Prescott and G. McLean, "Line-based correction of radial lens distortion," *Graph. Models and Img. Process.*, Vol. 59, No. 1, Jan. 1997.

[8] K. Tarabanis, R. Tsai and D. Goodman, "Calibration of a computer controlled robotic vision sensor with a zoom lens," *CVGIP*, Vol. 59, No. 2, Jan. 1994.

[9] W. Seales and D. Eggert, "Active-camera calibration using iterative image feature localization," *Proc. Conf. Analysis of Images and Patterns, Prague*, Sept. 1995.

[10] S. Shih, Y. Hung and W. Lin, "Calibration of an active binocular head," *IEEE Trans. Man, Sys. and Cybernetics*, Vol. 28, No. 4, July 1998.

[11] M. Ahmed, E. Hemayed and A. Farag, "Neurocalibration: a neural network that can tell camera calibration parameters," *Proc. ICCV, Korfu, Greece*, Sept. 1999.

[12] P. Rousseeuw and A. Roy, "Robust regression and outlier detection," *John Wiley and Sons*, 1987.

[13] P. Huber, "Robust statistics," *John Wiley and Sons*, 1981.

[14] F. Hampel, E. Ronchetti, P. Rousseeuw and W. Stahel, "Robust statistics: the approach based on influence functions," *John Wiley and Sons*, 1986.

[15] M. Ahmed and A. Farag, "A neural optimization framework for zoom-lens camera calibration," *Technical Report, CVIP Lab, University of Louisville, http://www.cvip.uofl.edu/*, May 2000.