UNIVERSAL CAMERA CALIBRATION WITH AUTOMATIC DISTORTION MODEL SELECTION

Vitaliy Orekhov, Besma Abidi, Chris Broaddus, and Mongi Abidi

Imaging, Robotics and Intelligent Systems Laboratory
The University of Tennessee, Knoxville, TN 37996

ABSTRACT

We present an optimized full scale automatic camera calibration approach which is both accurate and simple to implement. The method can be applied to a wide range of cameras equipped with normal, wide-angle, fish-eye, and telephoto lenses. The procedure does not require prior knowledge of any parameters. The method uses a simple planar calibration pattern which is observed from different positions. Closed-form estimates for the intrinsic and extrinsic parameters are computed followed by nonlinear optimization. A polynomial function describes the lens projection instead of the commonly used radial model. Statistical information criteria are used to automatically determine the complexity of the lens distortion model. The proposed complete calibration method is shown to obtain lower reprojection error than some commonly used methods while maintaining a low complexity distortion model.

***Index Terms*—** Calibration, Distortion, Lenses

1. INTRODUCTION

A large number of camera models and procedures are available today to calibrate a camera but the challenge is in choosing both a model which provides accurate results and a procedure which is simple. The assumptions and constraints in many of these methods place limitations on having a complete and universal calibration method. The most popular calibration methods take several images of a known calibration object from different camera positions. The projection of calibration object's features onto the image sensor is approximated with the pinhole camera model. The deviation of features from the pinhole camera is modeled with radial and tangential distortions [1,2]. Most such methods either require prior knowledge, namely the focal length, or are restricted to normal cameras with perspective projections. As a result these methods can not be used on cameras where prior knowledge of some parameters is not known or on camera systems which are equipped with wide-angle or fish-eye lenses. Such camera lenses exhibit significant amount of lens distortion which must be given special consideration.

Much research has been made in the area of distortion calibration on wide-angle and fish-eye lenses [3,4]. Many of these methods use calibration patterns to fix the distortion [4,5] while other nonmetric methods depend on the presence of certain features in the scene [3]. The limitation with these methods is that they are concerned with only correcting the distortion while leaving the rest of the camera parameters to be estimated with other methods.

Even with all of the proposed distortion calibration methods, few address the issues of a generic distortion model [4] which can work on a wide range of cameras or consider distortion model complexity selection [3,7]. All together the methods which focus on complete generic calibration for wide range of cameras are few.

In this paper, we present a unified framework for a full scale camera calibration technique which addresses a number of shortcomings of previous methods. The result is a complete generic calibration procedure with automatic distortion model selection which can be applied to normal, telephoto, wide-angle, and fish-eye lenses. We propose to use a closed form solution to estimate the intrinsic and extrinsic parameters and refine the values using an optimization step with bundle adjustment. To be able to model distortion on a wide range of lenses we use the lens projection polynomial model with statistical model complexity selection. We apply our method on both synthetic and real data captured using a wide range of camera lenses.

2. INTRINSIC AND EXTRINSIC CALIBRATION

In a pinhole camera, a point in space $\mathbf{M} = (x, y, z)^{\mathrm{T}}$ is projected onto the *image plane* to image point $\mathbf{m} = (u, v)^{\mathrm{T}}$ so the ray from $\mathbf{M}$ to $\mathbf{m}$ passes through the *camera center* $\mathbf{C}$. Points $\mathbf{M}$ and $\mathbf{m}$ are related by the projection $\mathbf{P} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]$, where $\mathbf{R}$ is a 3×3 rotation matrix, $\mathbf{t}$ a 3×1 translation vector, and $\mathbf{K}$ the intrinsic calibration matrix. The *intrinsic* parameters include the focal length f, aspect ratio α/β, skew s, and principal point (u_0, v_0). We will denote the homogenous representation of $\mathbf{M}$ and $\mathbf{m}$ as $\hat{\mathbf{M}}$ and $\hat{\mathbf{m}}$ respectively.

2.1. Homography Estimation

By using a planar calibration grid, the projection matrix reduces to a 2D to 2D mapping. Several techniques exist to estimate the homography $\mathbf{H}$ expressed as:

$$\hat{\mathbf{m}} \sim \mathbf{H}\hat{\mathbf{M}} \text{ with } \mathbf{H} = \mathbf{K}(\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}). \tag{1}$$

Normalized direct linear transformation (NDLT), followed by nonlinear optimization is used to compute $\mathbf{H}$ [2,8].

2.2. The Closed-form Solution

The image of the absolute conic $\omega = \mathbf{K}^{-\mathrm{T}}\mathbf{K}^{-1}$ and the homography $\mathbf{H}$ relating a model plane in the world coordinate system to its image places two constraints on the intrinsic parameters [2]. Since $\mathbf{r}_1$ and $\mathbf{r}_2$ are orthonormal, using $\mathbf{H} = \mathbf{K}(\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t})$ we obtain our two constraints $\mathbf{h}_1^{\mathrm{T}}\omega\mathbf{h}_1 = \mathbf{h}_2^{\mathrm{T}}\omega\mathbf{h}_2 = 1$ and $\mathbf{h}_1^{\mathrm{T}}\omega\mathbf{h}_2 = 0$. Given homography $\mathbf{H}$ we may write $\mathbf{H}^{\mathrm{T}}\omega\mathbf{H}$. Writing ω in terms of $(\alpha, \beta, s, u_0, v_0)$ gives a symmetric matrix that may be defined by a 6D vector $\hat{\omega} = (\omega_{11}, \omega_{12}, \omega_{22}, \omega_{13}, \omega_{23}, \omega_{33})^{\mathrm{T}}$. Writing the i^{th} column of $\mathbf{H}$ as $\mathbf{h}_i = (h_{1i}, h_{2i}, h_{3i})$ we obtain

$$\mathbf{h}_i^{\mathrm{T}}\hat{\omega}\mathbf{h}_j = v_{ij}^{\mathrm{T}}\hat{\omega}. \tag{2}$$

Combining the constraints as a homogenous system gives

$$\mathbf{V}\hat{\omega} = \begin{bmatrix} v_{12}^{\mathrm{T}} \\ (v_{11} - v_{22})^{\mathrm{T}} \end{bmatrix}\hat{\omega} = \mathbf{0}. \tag{3}$$

If we have n images of the model plane, then stacking equation (3) makes $\mathbf{V}$ a $2n \times 6$ matrix with a unique solution when $n \geq 3$. Once we have ω we can make various substitutions to solve for $(\alpha, \beta, s, u_0, v_0, \lambda)$ with λ being the scale factor.

Once the intrinsic parameters have been solved for, the extrinsic parameters are computed as:

$$\begin{gathered} \mathbf{r}_1 = \lambda\mathbf{K}^{-1}\mathbf{h}_1,\ \mathbf{r}_2 = \lambda\mathbf{K}^{-1}\mathbf{h}_2,\ \mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 \\ \mathbf{t} = \lambda\mathbf{K}^{-1}\mathbf{h}_3 \end{gathered} \tag{4}$$

with $\lambda = 1/\|\mathbf{K}^{-1}\mathbf{h}_1\| = 1/\|\mathbf{K}^{-1}\mathbf{h}_2\|$. Because of noise in the data, the rotation matrix will not necessarily satisfy all the properties of a rotation matrix. The best rotation matrix $\mathbf{R}$ approximating a given matrix $\mathbf{Q}$ under the Frobenius norm is the one that minimizes $\min_{\mathbf{R}}\|\mathbf{R} - \mathbf{Q}\|_{\mathrm{F}}$. The solution is then $\mathbf{R} = \mathbf{U}^{\mathrm{T}}\mathbf{V}$ with $\mathbf{Q} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}}$ being the singular value decomposition.

3. SOLVING FOR DISTORTION

To be able to apply the calibration method to an entire spectrum of lenses from telephoto to fisheye we use lens projection model rather than the commonly used radial distortion model. This results in a decreased error and complexity of the overall approach as compared to modeling radial distortion. Standard cameras are built to follow a perspective projection. However, perspective projection has an asymptote at 180° FOV which makes it extremely difficult to build a rectilinear lens above 100° FOV. Other types of projections have been proposed [9] to overcome the limitation and are listed in Table 1.

Table 1: Types of lens projections

	Name	Formula
1	Perspective	$r = f \tan\phi$
2	Stereographic	$r = 2f \tan(\phi/2)$
3	Equisolid	$r = 2f \sin(\phi/2)$
4	Orthogonal	$r = f \sin\phi$

In practice, real cameras do not exactly follow the projections in Table 1. A polynomial is used to approximate the real lens projection in the form:

$$r(\phi) = f\sum_{i=1}^{p}\kappa_i\phi^{2i-1} = f(\kappa_1\phi + \kappa_2\phi^3 + \ldots). \tag{5}$$

Once a solution has been computed for the calibration matrix, rotation matrix and translation vector, a least-squares solution to p lens projection coefficients $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \ldots \kappa_p)^{\mathrm{T}}$ is computed.

In optical systems the centers of lens elements are not strictly collinear and are subject to various amounts of decentering distortion [1]. This distortion has both radial and tangential components and can be modeled as:

$$\Delta\mathrm{d} = \begin{pmatrix} (2\rho_1 uv + \rho_2(r^2 + u^2))(1 + \rho_3 r + \rho_4 r^3 + \ldots) \\ (\rho_1(r^2 + v^2) + 2\rho_2 uv)(1 + \rho_3 r + \rho_4 r^3 + \ldots) \end{pmatrix} \tag{6}$$

with q decentering coefficients $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots \rho_q)$. Since decentering distortion is usually small, initial estimates are set to zero and then later optimized with other parameters.

4. FINAL PARAMETER OPTIMIZATION

Once the close-from solutions to the camera parameters are computed, including the distortion coefficients, the results are refined using maximum likelihood estimation (MLE). From our experiments, as the lens projection deviates from the perspective projection, alternating between refining $(\mathbf{K}, \mathbf{R}, \mathbf{t})$ and $(\boldsymbol{\kappa}, \boldsymbol{\rho})$ produces significantly better results. Levenberg-Marquardt algorithm is used to perform MLE.

5. DISTORTION MODEL SELECTION

Distortion model selection is the task of choosing the best model for a given system when several competing models can represent the distortion. Even though any model can be incorporated into the algorithm introduced above, the use of the most fitting and concise model will provide both better accuracy and reduced computational complexity. In most cases the model with more degrees of freedom will fit the data closer than other less complex models but higher order terms in the polynomial radial distortion model may cause numerical instability [1,2]. To help with stability and since higher order terms are comparatively insignificant in some systems, the number of distortion coefficients are kept low when modeling standard cameras [1,8] but higher order terms may be necessary when modeling wide-angle lenses.

The goal is to develop an automatic model selection criterion based on a quantitative measure which will select the model with a reduced number of coefficients without sacrificing accuracy. Akaike laid the foundation for statistical model selection by introducing the information theoretic criterion (AIC) [12]. In AIC, the model selected is the one that minimizes the error of a new observation. It has the form

$$\mathrm{AIC} = -2\log L(\boldsymbol{\theta};\mathbf{m}_i) + 2k\ , \tag{7}$$

where $k = p + q$ is the number of parameters in the model and $L(\boldsymbol{\theta};\mathbf{m}_i)$ is the likelihood of the model parameters $\boldsymbol{\theta} = (\mathbf{K},\mathbf{R},\mathbf{t},\boldsymbol{\kappa},\boldsymbol{\rho})$ given observations $\mathbf{m}_i$. The model with the lowest AIC score is selected. The first term in equation (7) is a measure of the goodness of fit of the model, and the second term penalizes higher complex models.

The sum-square-error (SSE) is computed as $\mathrm{SSE} = \sum_i r_i^2$ with $r_i = \|\mathbf{m}_i - \breve{\mathbf{m}}_i\|$ the difference between the measured and estimated image points. Assuming the noise in the data is Gaussian distributed, the probability of $\mathbf{m}_i$ given the model $\boldsymbol{\theta}$ is the product of the individual probability density functions (PDFs) of each point, assuming the errors on all points are independent. The PDF of the noise perturbed data is given by

$$\Pr(\mathbf{m}_i \mid \boldsymbol{\theta}) = \prod_i \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) e^{-r_i^2/(2\sigma^2)}, \tag{8}$$

where σ^2 is the variance of noise. The *log-likelihood* of the model parameters $\boldsymbol{\theta}$ given the observations $\mathbf{m}_i$ is then:

$$\begin{aligned}\log L(\boldsymbol{\theta};\mathbf{m}_i) &= \arg_{\boldsymbol{\theta}}\left[\log \Pr(\mathbf{m}_i \mid \boldsymbol{\theta})\right] \\ &= -\frac{1}{2\sigma^2}\sum_i r_i^2 + \text{constant.}\end{aligned} \tag{9}$$

The *maximum log-likelihood estimate (MLE)* is the set of parameters $\boldsymbol{\theta}$ that maximizes $\log L(\boldsymbol{\theta};\mathbf{m}_i)$. We observe that minimizing the SSE is equivalent to maximizing the log-likelihood, which is in-turn equivalent to maximizing the likelihood of the model parameters $\boldsymbol{\theta}$. Therefore, by substituting equation (9) into equation (7) and simplifying, we can write AIC in the following form:

$$\mathrm{AIC} = \frac{1}{\sigma^2}\sum_i r_i^2 + 2k\ . \tag{10}$$

We use the formulation in [10] to calculate the variance σ^2 of unknown Gaussian noise:

$$\sigma^2 = \sum_i r_i^2 \Big/ (N - \hat{k}), \tag{11}$$

where N is the number of samples and $\hat{k}$ is the number of coefficients of the most complex competing model. In summary, using the proposed distortion model, a variety of pq distortion models are fit to the data. Then one of the criterions listed in Table 2 is used to select the distortion model. The selection of which criterion to used should be determined by the application, data size, noise, and model library [10].

Table 2: Model selection criterions

Name	Formula
AIC [11]	$-2\log L(\boldsymbol{\theta};\mathbf{m}_i) + 2k$
MDL [12]	$-2\log L(\boldsymbol{\theta};\mathbf{m}_i) + 1/2\,k\log N$
BIC [13]	$-2\log L(\boldsymbol{\theta};\mathbf{m}_i) + 2k\log N$
SSD [12]	$-2\log L(\boldsymbol{\theta};\mathbf{m}_i) + k\log[(N+2)/24] + 2\log(k+1)$
CAIC [14]	$-2\log L(\boldsymbol{\theta};\mathbf{m}_i) + k(\log N + 1)$

6. EXPERIMENTAL RESULTS

6.1. Distortion model selection

To test the performance of the proposed model selection we generated synthetic data consisting of 8 images. Each 800x600 image contained sixty four control points and was transformed with varying rotations and translations. Distortion was applied to the data with a fixed number of radial and tangential coefficients. Model selection was performed between 12 competing distortion model complexities. The distortion models range from simplest model with no distortion to the most complex model with 5 coefficients to model radial distortion and 2 coefficients to model tangential distortion. Results show that the model selection can successfully select the same model complexity as the model complexity used to generate the synthetic data.

To test the robustness of model complexity selection we generated synthetic data with varying noise levels of Gaussian distribution with zero mean and standard deviation, σ, ranging from zero to 1.2 pixels. The simulation was repeated 250 times at various noise levels. Since we did not see a significant dependence on noise we summarized the results for the tested criterions in Table 3. Results with synthetic data, which were generated to simulate real data, verify the correct model complexity selection with the tested criterions.

Table 3: Model Complexity Selection Accuracy

Model criterions	AIC	MDL	BIC	SSD	CAIC
% Accuracy	94.8%	98.0%	99.6%	99.2%	99.2%

6.2. Complete calibration with real data

We applied our calibration algorithm with distortion model selection to five cameras: (1) 210mm Nikon Vari-focal Lens, (2) PULNiX CCD camera with 6mm lens [2], (3) IQEye3 with a FUJINON 1.4-3.1 mm lens set to wide angle, and a Nikon fisheye FC-E8 lens set to two different zoom settings to produce a (4) full frame fisheye (FOV of 180° across the diagonal) and (5) circular fisheye (FOV of 180° in all directions). Table 4 and figure 1 show the results with the mean-square-error (MSE) for the different cameras comparing the approach using the proposed *lens projection* with *decentering distortion (LPDD)* model and the traditional *radial distortion* with *decentering distortion (RDDD)* model. Figure 2 shows the complexities as selected by the MDL criterion. MDL was chosen over other criterions because it always selected a complexity less than or equal to other criterions without significantly sacrificing the error. The algorithm was modified for the verifocal telephoto lens data to use the zoom of expansion as the initialization value for the principal point.

Table 4: MSE error with RDDD and LPDD models.

	(1) Telephoto	(2) Normal	(3) Wide angle	(4) Full frame	(5) Fisheye
RDDD	0.4446	0.0287	1.3790	2.3857	8.1422
LPDD	0.4597	0.0298	0.9520	0.6639	0.9405

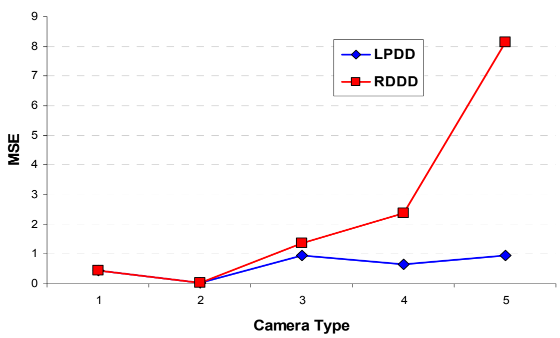

Fig. 1. Graph of the mean-square-error (MSE) from Table 4 for different cameras with LPDD and RDDD distortion models.

Zhang achieved a root-mean-square (RMS) error on his publicly available dataset of 0.335, where he only modeled radial distortion with two coefficients [2]. This corresponds to an MSE of approximately 0.1122. Our LPDD method achieved an MSE of 0.0298 using the same number of coefficients. The RDDD method also achieved a lower MSE than Zhang at the expense of two extra coefficients for decentering distortion, as selected by MDL.

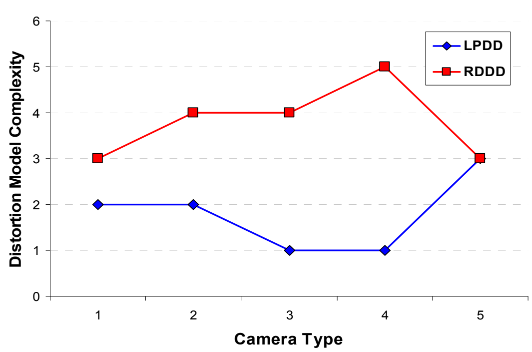

Fig. 2. Graph of the model complexity for the different cameras with LPDD and RDDD distortion models.

In all the experiments, the LPDD method outperformed the RDDD method except for normal and tele-photo lens, but were lower only by a small margin. We can clearly see in Figure 1 the exponential increase of the MSE for RDDD as the camera approaches a circular fisheye whereas the MSE for LPDD is small and stable for all cameras. Also, the complexity of the model is always less than or equal to that of RDDD shown in Figure 2.

6. CONCLUSIONS

We have presented a complete automatic camera calibration technique for use with a large spectrum of cameras. To model distortion, lens projection polynomial and decentering distortion are incorporated in the overall unified framework. A number of information criteria were shown to be successful in automatically selecting the complexity of the camera distortion model. Results show the advantage of using the LPDD model and the superior performance of our method on a wide range of cameras, varying from telephoto to fisheye.

7. REFERENCES

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

[2] Z. Zhang, "A flexible new technique for camera calibration," *In IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), pp. 1330-1334, Nov. 2000.

[3] M. T. El-Melegy and A. A. Farag, "Nonmetric lens distortion calibration: closed-form solutions, robust estimation and model selection," in *Proceedings Ninth Int. Conf. on Computer Vision*, vol. 1, pp. 554-559, 2005.

[4] R. Hartley and S. B. Kang, "Parameter-free radial distortion correction with center of distortion estimation," in *Proc. ICCV-2005*, Oct. 2005.

[5] S. Graf, T. Hanning, "Analytically solving radial distortion parameters", in CVPR *2005 IEEE computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 20-26, June 2005.

[6] S. B. Kang, "Radial distortion snakes," *IEICE Trans. Inf. & Syst.*, vol E84-D, No. 12, pp.1603-0611, Dec. 2001.

[7] K. Kinoshita and L. Lindenbaum, "Camera model selection based on geometric AIC," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. 514-519, 2000.

[8] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[9] J. Kannala and S. Brandt, "A Generic Camera Calibration Method for Fish-eye Lenses," in *Proc. of the IEEE Int. Conf. on Pattern Recognition*, pp. 10-13, 2004.

[10] N. Gheissari and A. Bab-Hadiashar, "Model Selection Criteria in Computer Vision: Are They Different?" *In Proc. of VII Digital Image Computing: Techniques and Applications*, pp.185-194, Dec. 2003.

[11] H. Akaike, "A New Look at the Statistical Model Identification," *Automatic Control, IEEE Trans.on*, vol. 19, no. 6, pp. 716-723, Dec. 1974.

[12] J. Rissanen, "Modeling by Shortest Data Description," *Automatica*, vol. 14, pp. 465-471, 1978.

[13] G. Schwarz, "Estimating the Dimension of a Model," *Annals of Statistics*, vol. 6, pp. 461-464, 1978.

[14] H. Bozdogan, "Model Selection and Akaike's Information Criterion," *Psychmetrika*, 53(3), pp. 345-370, 1987.