Czech Pattern Recognition Workshop 2000, Tomáš Svoboda (Ed.)
Peršlák, Czech Republic, February 2–4, 2000
Czech Pattern Recognition Society

Camera Calibration with a Simulated Three Dimensional Calibration Object

Hynek Bakstein [1] and Radim Halíř [2]

[1] Czech Technical University
Center for Machine Perception
Karlovo nám. 13, 121 35 Prague, Czech Republic
bakstein@cmp.felk.cvut.cz
[2] Charles University
Department of Software Engineering
Malostranské nám. 2/25, 118 00 Prague, Czech Republic
halir@ms.mff.cuni.cz

Abstract *A new camera calibration method based on DLT model is presented in this paper. Full set of camera parameters can be obtained from multiple views of coplanar calibration object with coordinates of control points measured in 2D. The method consists of four steps which are iterated until convergence. The proposed approach is numerically stable and robust in comparison with calibration techniques based on nonlinear optimization of all camera parameters. Practical implementation of the proposed method has been evaluated on both synthetic and real data. A MATLAB toolbox has been developed and is available on WWW.*

1 Introduction

Camera calibration is one of the basic tasks in computer vision [4]. It is used to determine the optical characteristics of the camera (so called intrinsic parameters) and position of the camera in the scene (extrinsic parameters). One or more images of a priori known calibration object are acquired by a camera and several well-defined features (so called *control points*) are detected in the images. The coordinates of the features in the scene can be unknown (so called *self-calibration* [10]) or they can be measured in advance. In this paper, a known calibration object with coordinates of the features measured in 2D is assumed.

Various camera calibration methods were introduced in a literature. The classical approach [9], based on methods used in photogrammetry, gives precise results but it is computationally extensive. Several simplification were made (such as [11] and [13]), but the nonlinear search used there may lead into computational instability. There are also methods which combine both minimization and closed-form solution [4, 5, 7, 15]. All these methods are based on physical camera parameters. Implicit camera calibration methods [12], on the other hand, use an interpolation between coordinates of the points on multiple planes or images and they do not explicitly compute camera parameters.

The calibration object can be three dimensional [1], two dimensional (also called coplanar) [7] or the method can use multiple views of a coplanar object to simulate the three dimensional one. The three dimensional object gives better results and the complete set of camera parameters can be acquired, but such object is hard to manufacture. In addition, the coordinates of the detected features have to be measured in three dimensional space. In case of multiple views of a coplanar object, it can be moved freely between the views [5, 15] or the movement should be a priori known [11]. The known movement of a coplanar object requires a robot which may not be available.

Above mentioned facts lead to a conclusion that the most effective approach is to simulate a three dimensional calibration object by multiple views of coplanar one. The movement of the coplanar target between image acquisitions can be unconstrained.

2 Simulation of the 3D calibration target using multiple unconstrained views of a coplanar object

The proposed calibration method uses multiple views of a 2D calibration target (so called *calibration plate*) to simulate a 3D object, which is needed for precise and reliable calibration results. The calibration plate can be moved freely between the image acquisitions. An important assumption of the method is that the intrinsic parameters of the camera are constant for all views. This constraint allows the simulation of the 3D object to be based on relative positions of the plate between particular acquisitions.

The method consists of the following four steps (see Figure 1):

1. an initial estimation of the intrinsic parameters of a camera,

2. an estimation of the extrinsic camera parameters for each view,

3. a construction of a virtual 3D calibration object from multiple 2D planes, and

4. a complete camera calibration based on the virtual 3D object.

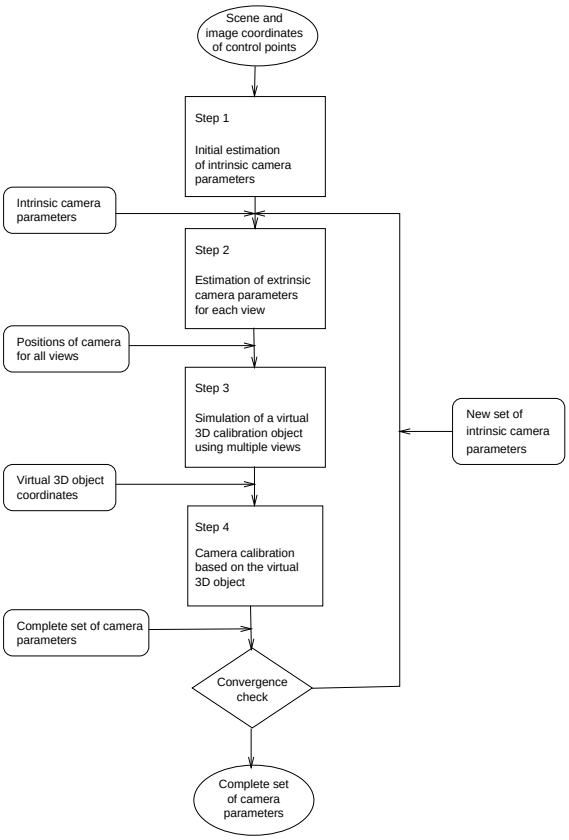


Figure 1: A scheme of the proposed calibration method

Due to the employment of a coplanar calibration target, an initial guess of the intrinsic camera parameters is required, because the coordinates of the control points are measured only in 2D. These parameters are supplied by the user in the first step of the proposed method.

In the second step, the extrinsic camera parameters are estimated for each view with respect to the provided intrinsic parameters. Any explicit coplanar camera calibration method can be applied here.

In the third step, the virtual 3D calibration object is constructed. The construction exploits the knowledge on the camera positions for all views provided by the previous step.

The pairs of scene and image coordinates of the control points of the simulated calibration object are finally passed to the fourth step of the proposed method. A complete set of the camera parameters is estimated there. Any explicit 3D based camera calibration method can be used for this purpose.

As a result of the calibration, both intrinsic and extrinsic camera parameters are provided. Regarding the fact that the proposed calibration method needs an initial guess of the internal characteristics of the camera, its accuracy can be improved as follows: The obtained intrinsic parameters are put back to the second step of the method and a new set of the camera parameters is estimated (as depicted in Figure 1. The whole process is iterated until convergence.

3 Implementation of the method using DLT camera model

The outline of the proposed calibration method presented in the previous section illustrates only main ideas of the novel approach. All the estimation steps are relatively independent on concrete calibration method. Practical realization, however, has to be based on particular models of a camera.

A coplanar variant of the DLT [1] camera model (so called CDLT) has been chosen for implementation of the second step of the method. The model can be expressed in form of a linear matrix equation which allows direct extraction of camera parameters. Although it does not compensate nonlinear distortions, it still gives a very good approximation of a camera. Due to the fact that all camera parameters cannot be extracted from the CDLT matrix [7], an initial estimate of at least three of them has to be provided to the first step of the method.

CDLT model allows the image formation to be expressed in homogeneous coordinates by the following matrix equation:

$$q_i = \mathbf{A}\, p_i \ , \tag{1}$$

where $\mathbf{A}$ is the coplanar DLT (CDLT) matrix of the size 3×3, $p_i = [x_i, y_i, 1]^T$ are homogeneous scene coordinates of the i-th point and $q_i = [w_i u_i, w_i v_i, w_i]^T$ are the appropriate coordinates in the image plane. The CDLT matrix $\mathbf{A}$ of the size 3×3 can be written as

$$\mathbf{A} = \lambda \mathbf{V}^{-1}\mathbf{B}^{-1}\mathbf{F}\mathbf{M}\mathbf{T} \ , \tag{2}$$

where matrices $\mathbf{V}$ compensates the shift of the image origin, $\mathbf{B}$ represents the difference in scale and lack of orthogonality between image axes, $\mathbf{F}$ stands for focal length, $\mathbf{M}$ for rotation and $\mathbf{T}$ represents translation, respectively and λ is a non zero scalar.

An estimate of the extrinsic camera parameters is obtained by a decomposition of the CDLT matrix $\mathbf{A}$ in the second step. Values of these parameters have to be refined, especially the matrix of rotation $\mathbf{M}$ need not to be orthonormal as required. Geometrical approach to transform the matrix $\mathbf{M}$ to a proper matrix of rotation was chosen in this particular realization of the method (see Figure 2):

First two columns of $\mathbf{M}$ determine a 2D plane p. Changing the angle α between these vectors $\mathbf{m}_1$ and $\mathbf{m}_2$ to the value $\frac{\pi}{2}$ and computing the third column of $\mathbf{M}$ as vector product of $\mathbf{m}_1$ and $\mathbf{m}_2$ gives a proper matrix of rotation similar to $\mathbf{M}$. The translation is then computed directly from Equation 2.

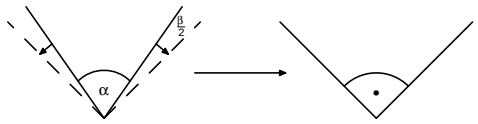


Figure 2: Orthogonalization of two vectors in a plane: t_1 and t_2 are the original vectors, o_1 and o_2 the orthogonalized ones

The above closed-form computation of extrinsic parameters gives only a rough approximation of camera parameters. Therefore a nonlinear minimization is exploited to refine the values of the parameters. The Euclidean distance between observed q_j and reprojected $\widetilde{q}_j$ image coordinates of control points is used as a criterion for this minimization. Reprojection is performed by applying the CDLT matrix $\mathbf{A}$ composed from estimated camera parameters on scene coordinates of control points p_j:

$$\widetilde{q}_j = \mathbf{A}_i \, p_j . \tag{3}$$

In [3], a reduction of the search space to four is introduced. This is achieved by representing rotation with three parameters and translation only by one parameter. Remaining two translation values are computed from the coordinate pairs of control points and their images. More details can be found in [3]. In practical realization of the proposed method, SolvOpt optimization toolkit [6] for MATLAB was used.

The optimization can be described by equation:

$$\min_{z_0,\sigma,\rho,\phi} \sum_{j=1}^{n} \left\| q_j - \widetilde{q}_j \right\|^2 \quad , \tag{4}$$

where n is a number of control points, z_0 is a parameter of translation and σ, ρ, ϕ represent rotation. These variables are used is computation of reprojected coordinates $\widetilde{q}_j$.

Construction of the virtual 3D calibration object is performed in the third step of the method. One of the views is selected as a *reference*. Let it be the first one denoted by subscript 1. A virtual camera is created and placed into the position respective to the reference view. Then, the images of control points in all views are merged into that reference one. Scene coordinates of control points are translated and rotated into a new position. This position is chosen so that if the points are observed by the virtual camera, they are projected into images acquired by camera in position respective to particular views (see Figure 3).

The construction of virtual 3D object can be expressed as:

$$q_i = \mathbf{A}_1 \mathbf{R}_i p_i \quad , \quad i = 1 \ldots N \quad , \tag{5}$$

where p_i are scene coordinates of control points, q_i are their images taken by a camera in reference position represented by matrix $\mathbf{A}_1$ and N is the number of views. Matrices $\mathbf{R}_i$ express the transformation of scene coordinates of control points described above. They can be computed as:

$$\mathbf{R}_i = \mathbf{T}_i \mathbf{M}_i \mathbf{M}_1^{-1} \mathbf{T}_1^{-1} \quad , \tag{6}$$

where $\mathbf{T}_i$ and $\mathbf{M}_i$ are translation and rotation matrices corresponding to the i-th view. Value $i = 1$ denotes the reference

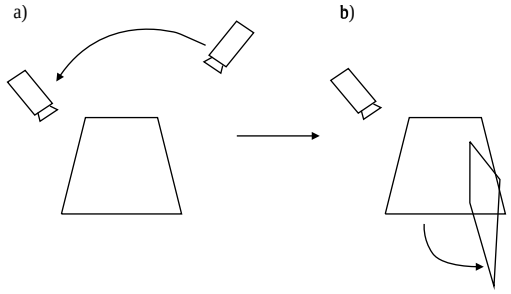


Figure 3: Using more views of 2D plane (a) to simulate a 3D object (b)

view. Note similarity of Equation 6 with derivation of Epipolar constraint [8].

4 Experimental results

For the following experiments, a MATLAB toolkit implementing the proposed method, available on WWW [2], was used. The tests were performed both on synthetic and real data. The synthetic experiments have proved that the method gives appropriate results. Image size was set to 704 by 573 pixels, focal length to 1136 pixels and coordinates of principal point to 363 and 280 pixels. Calibration target was a generated grid of 30 control points in 5 rows and 6 columns. 6 images from various angles were computed. The movement of the simulated camera was regularly distributed around the grid.

Speed of convergence of the method depends on the precision of the initial guess. Higher error in the estimate of the initial values of parameters is compensated by higher number of iteration. After a small number of steps, the improvement of the accuracy was negligible. It was also observed, that the best convergence was obtained when the movement of the calibration object performed in all axes. This means that the object was not moved only in one direction.

The mean Euclidean distance between the observed coordinates of control points and their values predicted by the model is referred to as a *error of reprojection* of estimation of the camera parameters. Predicted coordinates were computed by a reprojection of the control points by the model of a camera composed from the estimated parameters. 10 cycles were performed for each level of additive noise.

Table 1 presents the error of reprojection, focal length f and coordinates of principal point u_0 and v_0 in case of ideal noise-free data. All values are in pixels. Each row of the table, except the first and the last one, represents different number of views. The first row corresponds to an initial guess. The last one contains ground truth for comparison.

As can be seen, the error value decreases with larger number of views with exception of the case of two views. This ambiguity is caused by numerical computation of the error of reprojection. Two views do not suffice to ensure stability of

the method which can be demonstrated on incorrect estimation of camera parameters (see Table 1).

Table 1: Results of the calibration on synthetic data: ideal data, no noise. See text for a description of the table

No. of views	Error	f	u_0	v_0
Initial guess	5.5349	1300	353	286
2	0.2683	1095	340	304
3	0.2806	1110	363	276
4	0.2295	1122	357	285
5	0.2208	1132	355	284
6	0.2119	1112	356	278
Ground truth	0.0	1136	363	280

Table 2 contains results for experiment with synthetic data blurred by Gaussian noise with standard deviation of 2 pixels. Structure of the table is the same as in the previous case. Again, note that the error value decreases with larger number of views.

Table 2: Results of the calibration on synthetic data: the data were blurred by Gaussian noise with standard deviation of 2 pixels.

No. of views	Error	f	u_0	v_0
Initial guess	5.3482	1300	353	286
2	2.6889	1120	326	295
3	2.6025	1169	366	277
4	2.5637	1183	364	264
5	2.4727	1167	356	277
6	2.4720	1124	355	268
Ground truth	$\approx$2.0	1136	363	280

Figure 4 presents results of experiments on synthetic data. Figure 4(a) depicts the calibration error in the image plane. Data were blurred by Gaussian noise with standard deviation 0.5 pixels. Note that the arrows are scaled, mean of the error is 0.63 pixels. As can be seen, there is no correlation between error vectors (this means no systematic linear distortion) which means that the proposed method did estimate the parameters of the DLT model correctly.

Figure 4(b) shows dependency of the calibration error on noise. Observe that the dependency is linear and the values of noise and error are of the same magnitude. This indicates robustness of the approach.

Experiments on real data were performed with a commonly available equipment. The calibration target was a regular grid printed on a laser printer. The lines forming the grid were 4 cm in distance. The total number of 12 images sampled to 786 x 576 pixels were taken from various angles (see Figure 5 for an example). All images were taken under the same camera settings.

Lines crossings were chosen as the control points (depicted in Figure 6). From each image about 20 of them were manually detected. Precision of the detection was about 0.5 pixels. A set of experiments was performed using combinations of these images.

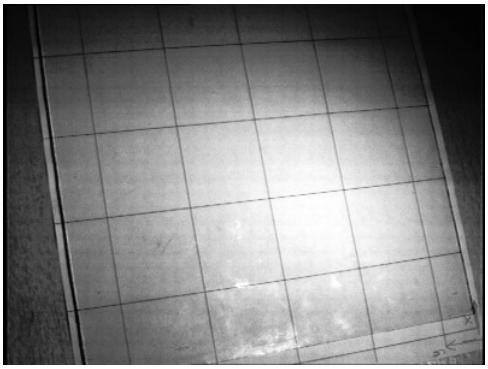


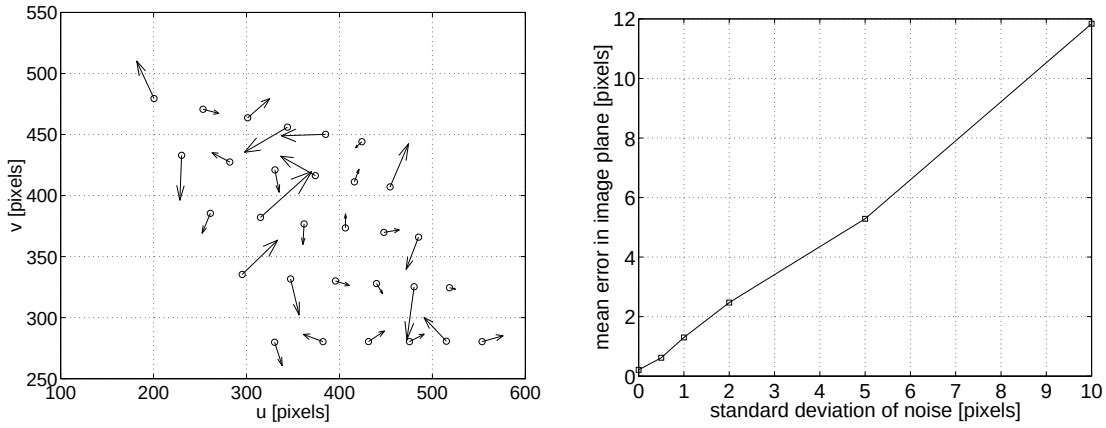
Figure 5: An image of calibration grid

Figure 6: Detail of the calibration grid with manually detected control point (marked with 'x')

Figure 7 illustrates the results of the experiment on real data. Figure 7(a) shows the error of reprojection in image plane. Again, the arrows are scaled, mean error value is 0.89 pixels. Again, it can be seen that there is no systematic linear distortion. But, unlike in the synthetic case, one of the coordinates is detected incorrectly. This can be observed as a big arrow on the right side of the graph. Moreover there is nonlinear distortion present in real cameras. Barrel radial distortion can be observed in the figure.

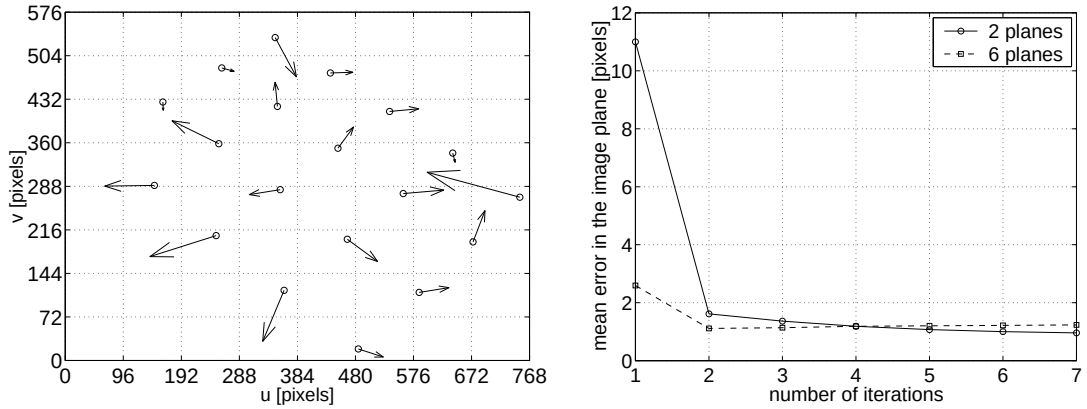
Figure 7(b) depicts the development of the error of reprojection for increasing number of iterations for 2 and 6 views. Note that after a small number of iterations, the change in estimates is negligible which means that the method converges rapidly. Gross error in the initial estimate of the intrinsic parameters causes larger number of iterations. It should be mentioned that even if the value of the error does improve slowly after several steps, the values of the camera parameters do change.

Table 3 contains the estimated values of intrinsic camera parameters and error for various number of views. The format



(a) (b)

Figure 4: Test with synthetic data: (a) the calibration error (note that the arrows are scaled, mean value is 0.63 pixels), (b) dependence of the calibration error on noise



(a) (b)

Figure 7: Experiments on real data: (a) calibration error for the reference view (note that the arrows are scaled, mean value is 0.89 pixels), (b) calibration error for different number of iterations (2 and 6 planes)

Table 3: Calibration error and intrinsic parameters for different number of views, real data. See text for a description

No. of views	Error	f	u_0	v_0
Initial guess	3.8014	1500	353	286
3	1.0019	1483	400	297
4	0.9718	1479	404	279
5	0.9642	1483	397	275
6	0.9171	1484	397	277

of the table is the same as for tables 1 and 2. Similar development of the error of reprojection in relation to number of views as in the case of synthetic data can be observed. The error is smaller with increasing number of views.

5 Comparison with other methods

The proposed method was compared with other freely available camera calibration approaches. The most popular one is the technique of Tsai [14] although it does not provide reliable results in the case when a coplanar calibration object is used. Another method distributed freely was developed by Heikkilä and Silvén [5] which is based on nonlinear optimization of all camera parameters.

The result of the comparison can be seen in Table 4. The column marked 'Tsai (basic)' corresponds to the method of Tsai with no optimization, while the column 'Tsai (full optim)' contains the values obtained by the Tsai's approach with full optimization which also includes an estimation of nonlinear distortions. The last row contains time of computation. Each other row corresponds to one camera parameter. Approximations of unknown camera parameters are listed in the first column for comparison. Other columns contain estimations of the parameters for each calibration method. Because the proposed approach does not compensate nonlinear distortion, the error of reprojection is bigger than in case of techniques which use nonlinear optimization of all parameters. But as can be seen on the results of the method of Tsai, numerical instability may lead to incorrectly estimated values.

Exact values of camera parameters were not available for the tests. Only basic camera parameters like the nominal focal length and the camera resolution were obtained from data sheet of the camera. The size of the CCD chip in millimeters needed for the conversion of the focal length from pixels to millimeters was unknown. Therefore the alternative values of the focal length f in Table 4 could not be computed.

Figure 8 compares the calibration approaches with respect to their sensitivity to noise among data. This test was performed on synthetic data blurred with Gaussian noise with standard deviation set to 0.25, 0.5, 1, 2, 3, 4 and 5 pixels. It can be seen that the proposed method gives the precision of the same order as the method of Heikkilä and Silvén. This means that the proposed approach gives results of the same quality without nonlinear optimization of all camera parameters.

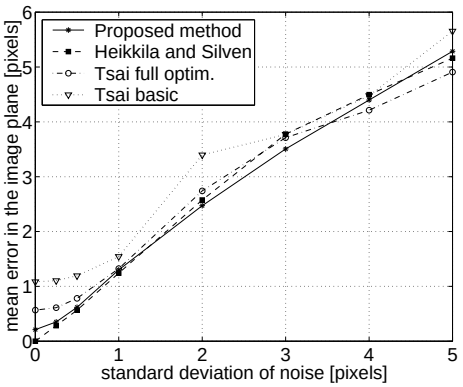


Figure 8: A comparison of various calibration methods with respect to noise among data

Next two comparisons were also made on synthetic data. They demonstrate, how the estimation of intrinsic parameters is sensitive to noise. Figure 9(a) depicts the precision of estimation of focal length in comparison with ideal value. Note that the proposed method shows smallest variations of the computed values from the selected methods. This indicates better stability than approaches based on nonlinear optimization of all parameters. Also the estimated focal length is close to the ideal value even for high level of noise.

Precision of estimation of the coordinates of principal point can be seen in Figure 9(b). Resulting values are computed as distances of predicted principal point from its correct position. Again, the proposed approach performs more robustly than methods based on nonlinear optimization of all parameters. The estimated parameter is close to the correct position even when large noise is present in data. As can be noticed, the method of Tsai with full optimization is not suitable for use with coplanar calibration object although it works very well with 3D objects. When the optimization is disabled, their results depend on the initial guess of the parameters.

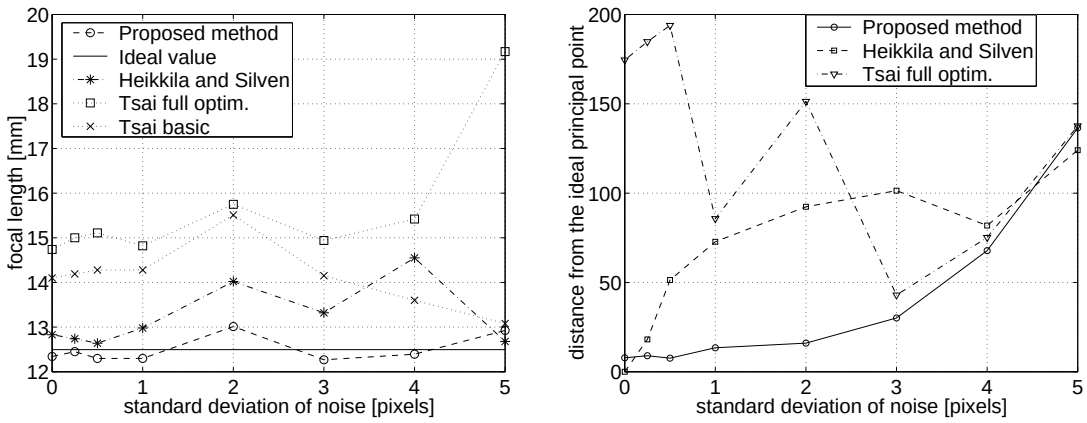
6 Conclusion

A new camera calibration was presented in this paper. The approach is capable to extract full set of camera parameters from multiple views of a coplanar target. The target can be moved without any constraint between the image acquisitions. Tests both with synthetic and real data were performed. They evaluated that the method is robust and stable and gives reliable results. Method was compared with other approaches. The proposed method exhibits significant improvements in comparison to the technique of Tsai with coplanar object and better numerical stability than the methods based on nonlinear optimization of all camera parameters.

The method models only linear distortion but it can be extended as in it is shown in [7] or [5]. The main disadvantage

Table 4: Comparison of methods, real data. See text for a description

Parameter name	Approx. value	Proposed method	Heikkilä and Silvén	Tsai (basic)	Tsai (full optim.)
f[pixels]	?	1513	?	?	?
f[mm]	≈16.20	?	16.64	20.32	19.55
u_0[pixels]	≈384	400	428	353	197
v_0[pixels]	≈288	292	314	286	330
error [pixels]	≈0.5	0.92	0.36	1.59	0.66
time [s]	—	245	11	0.06	1.75



(a) (b)

Figure 9: A comparison of the calibration methods. (a) estimation of the focal length, (b) the distance of the estimated principal point from its ideal position in pixels

is the time of the computation which is due to use of general optimization routine in the second step of the method. Providing gradients should speed up the execution. Future work should be targeted mainly to decrease the time needed for estimation of extrinsic camera parameters performed in the second step of the method.

References

[1] Y. I. Abdel-Aziz and H. M. Karara. Direct linear transformation into object space coordinates in close-range photogrammetry. In *Proc. of the Symposium on Close-Range Photogrammetry, Urbana, Illinois*, pages 1–18, 1971.

[2] H. Bakstein. Camera calibration toolbox, March 1999. Available at, http://terezka.ufa.cas.cz/hynek/toolbox.html.

[3] H. Bakstein. A complete DLT-based camera calibration with a virtual 3D calibration object. Master's thesis, Charles University, Prague, 1999.

[4] O. Faugeras. *Three-dimensional computer vision — A geometric viewpoint*. MIT Press, 1993.

[5] J. Heikkilä and O. Silvén. A four-step camera calibration procedure with implicit image correction. In *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97), San Juan, Puerto Rico*, 1997.

[6] A. Kuntsevitch and F. Kappel. Solvopt — the solver for local nonlinear optimization problems, 1997. http://bedvgm.kfunigraz.ac.at:8001/alex/solvopt/index.html.

[7] T. Melen. *Geometrical modelling and calibration of video cameras for underwater navigation*. PhD thesis, Institut for teknisk kybernetikk, Norges tekniske høgskole, Trondheim, November 1994.

[8] T. Moons. A guided tour through multiview relations. In R. Koch and L. Van Gool, editors, *European Workshop, SMILE'98, Freiburg, Germany*, June 1999.

[9] C. C. Slama. *Manual of Photogrammetry*. American Society of Photogrammetry, Falls Church, Virginia, 4th edition, 1980.

[10] B. Triggs. Autocalibration from Planar Scenes. In H. Burkhardt and B. Neumann, editors, *Proc of the 5th European Conference on Computer Vision (ECCV'98), Freiburg, Germany*, volume 1, pages 158–174. Springer-Verlag, June 1998.

[11] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine metrology using off-the-shelf cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, 1987.

[12] G. Q. Wei and S. D. Ma. A complete two-plane camera calibration method and experimental comparsion. In *Proc. of the 4th International Conference on Computer Vision (ICCV'93), Berlin, Germany*, pages 439–446, 1993.

[13] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accuracy evaluation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(10):965–980, October 1992.

[14] R. Willson. Tsai camera calibration software, 1995. http://www.cs.cmu.edu/ rgw/TsaiCode.html.

[15] Z. Zhang. A flexible new technique for camera calibration. Technical Report MSR-TR-98-71, Microsoft Research, 1998.