Comparison of Methods for Geometric Camera Calibration using Planar Calibration Targets

H. Zollner and R. Sablatnig
Pattern and Image Processing Group (PRIP)
Vienna University of Technology
zollnerh@prip.tuwien.ac.at, http://www.prip.tuwien.ac.at

Abstract:
The quest for the best solution to the geometrical camera calibration problem is ongoing up to the present. For simplicity and good adaptivity to the pinhole model many methods use sets of co-planar fiducial points. In this paper we do not invent a new technique, but investigate and describe major and minor differences of three most widely used plane-based calibration algorithms. Formulation and computation differ in their way of implementation, but also in the mathematical substance. As we could prove, each of the presented method is capable of delivering reliable results on the intrinsic camera parameters of our specific setup in the lab. We can also make a statement about which calibration method is favorable for specific configurations and which is not. A representative table of test results is given, which substantiates our argumentation.

1 Introduction

We compare three fundamentally different approaches [1, 3, 4] to the problem of getting a camera's projective features by taking a couple of images of a calibration object consisting of a planar pattern. This setup is as common in computer vision as are the solutions presented. The goal of this work is a quantitative accuracy estimation of the sample algorithms allowing a practical evaluation of odds and benefits of each particular technique. Generally a model of the physical projection of the camera must be provided for enabling exact measurement of 3D objects with a digital imaging sensor. A specific set of model parameters for a certain setup is called a *camera model*. It contains information about the *outer orientation*[1)], and internal projective features, described by the *intrinsic parameters* defining the so-called *inner orientation*. We are primarily interested in the inner orientation, since the intrinsic model serves as set of input parameters for 3D modeling procedures.

[1)] camera pose toward a pre-defined world coordinate system

2 Projection Models for Cameras

The imaging process produced by a projective camera can be interpreted as a sequence of three projective transformations. The overall mapping from image to world coordinates can be written as set of matrices for the i-th view:

$$\mathbf{P_i} = \mathbf{K}\mathbf{H_i}\mathbf{H_{ii}} = \mathbf{K}(\mathbf{R_i}\mathbf{t_i}), \mathbf{K} = \begin{pmatrix} \alpha & s & u_0 \\ 0 & k\alpha & v_0 \\ 0 & 0 & 1 \end{pmatrix}, \mathbf{H_i} = \begin{pmatrix} \mathbf{R_i} & \mathbf{t_i} \\ \mathbf{0}^T & 1 \end{pmatrix}, \mathbf{H_{ii}} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \end{pmatrix} \quad (1)$$

The points are given in homogeneous coordinates and therefore this is a linear projection $\mathcal{P}^3 \rightarrow \mathcal{P}^2$. $\mathbf{R_i}$ is a rotation matrix[2)], $\mathbf{t_i}$ is a translation vector, $\mathbf{I}$ is the 3x3 identity matrix, $\mathbf{0}^T$ a transposed null vector in $\mathcal{P}^3$, $\mathbf{K}$ a 3x3 matrix describing the specific projection of the camera optics, where α is the *focal length*, (u_0, v_0) are the pixel coordinates of the *principal point*, k the aspect ratio and s the *skewness*. The matrix $\mathbf{P_i}$ can alternatively be decomposed in the following way:

$$\mathbf{P_i} = \lambda \mathbf{V}^{-1}\mathbf{B}^{-1}\mathbf{F}[\mathbf{R_i}\mathbf{t_i}], \mathbf{V} = \begin{pmatrix} 1 & 0 & -u_0 \\ 0 & 1 & -v_0 \\ 0 & 0 & 1 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 1+b_1 & b_2 & 0 \\ b_2 & 1-b_1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (2)$$

where λ is an overall scaling factor, $F = \alpha\Pi$. $\mathbf{V}$ is clearly invertible, whereas $\mathbf{B^{-1}}$ exists if ${b_1}^2 + {b_2}^2 < 1$. This decomposition was used in [2] in order to compute some intrinsic parameters from $\mathbf{P_i}$. Radial distortion can be narrowed down by a Taylor polynomial of order 2 adding to additional calibration parameters κ_1 and κ_2.

3 Outline of 3 Calibration Algorithms

The description of the algorithms is only a very short outline, without any details about further concepts in projective geometry, parameter optimization or image preprocessing. This section will only point out fundamental algorithm differences. These differences are the reason why we choose to compare exactly these three algorithms. Indeed, most other calibration algorithms dealing with plane-based calibration have inherited their motivation from one of these approaches.

Direct Linear Transform (DLT)

This calibration procedure consists of two steps. In the first step the linear transformation from the object coordinates (x_j, y_j, z_j) to image coordinates (u_{ij}, v_{ij}) is solved. This matrix is

[2)] The rotation matrix is generally an orthonormal matrix. It is an alternative representation of a rotation around all three coordinate axes. Nevertheless rotation can also express mirroring and its interpretation is ambiguous. It is generally not task of calibration routine to uniquely define the pose of calibration target.

represented by a 3x4 matrix P_i for the i-th projection and N fiducial points. The parameters $p_{11} \dots p_{34}$ of the DLT matrix can be solved by a homogeneous matrix equation:

$$\mathbf{L}\mathbf{p_i} = \mathbf{0}, \tag{3}$$

where $\mathbf{L}$ is a Nx12 matrix (see [2, 1] for details), constituted by corresponding world and image coordinates and $\mathbf{p_i} = (p_{11}, \dots p_{34})$. Using equation 2 a five step algorithm for solving the parameters is given in [2]. In case of a coplanar control point structure the DLT matrix $\mathbf{P_i}$ becomes singular and a 3x3 submatrix $p_{11} \cdots p_{33}$ has to be used. In this case the decomposition of the submatrix can only deliver a subset of estimates of the camera parameters (see [2]). After solving the system for these parameters a subset of them can be used as start values for a classical bundle block calibration [1].

R. Y. Tsai's Method

The method proposed by R.Y. Tsai in [3] linearizes a huge part of the computation by restricting the lens distortion effect to radial distortion. The model generally does not recognize skewness or lack of orthogonality of the projection. By simplifying the camera model in this way, the observation illustrated by Figure 1 is made. Tsai's two stage calibration technique is fast and no initial guess of the calibration parameters is needed. In the first stage all extrinsic parameters except for t_z are computed by using the *Parallelism constraint*. In the second stage all missing parameters are evaluated by non-linear optimization. The optimization does not use the full camera model in order to speed up performance. Thus the computed residual is quite irrelevant for error measurement, which is done in this case separately by building a full camera model by collecting all resulting parameters of the previous steps. The final error is given here by the difference between back-projected fiducial 3D world points and their corresponding image points. The solution, generally designed for mono-view calibration, was also applied for multiple viewing position calibration. In this case a planar pattern is moved to different levels by a z stage for multiple calibration images [3].

Z. Zhangs Method

In contrast to Tsai's technique, we need at least 3 different projections of a planar calibration target. From the image coordinates $\tilde{\mathbf{p}} = (u, v, 1)$ and its corresponding known set of world coordinates $\tilde{\mathbf{P}} = (X, Y, Z = 0, 1)$, a *homography* $\mathbf{A} = (\mathbf{a_1 a_2 a_3})$ [4] can be derived:

$$s\tilde{\mathbf{p}} = \mathbf{A}\tilde{\mathbf{P}}, \mathbf{A} = \lambda\mathbf{K}(\mathbf{r_1 r_2 t}), \tag{4}$$

where $\mathbf{r_1}$ and $\mathbf{r_2}$ are the first two column vectors of rotation matrix $\mathbf{R}$ Using the constraint that these two column vectors are orthonormal the following two identities are derived:

$$\mathbf{a_1}^\mathbf{T}\mathbf{K}^{-\mathbf{T}}\mathbf{K}^{-\mathbf{1}}\mathbf{a_2} = 0, \quad \mathbf{a_1}^\mathbf{T}\mathbf{K}^{-\mathbf{T}}\mathbf{K}^{-\mathbf{1}}\mathbf{a_1} = \mathbf{a_2}^\mathbf{T}\mathbf{K}^{-\mathbf{T}}\mathbf{K}^{-\mathbf{1}}\mathbf{h_2} \tag{5}$$

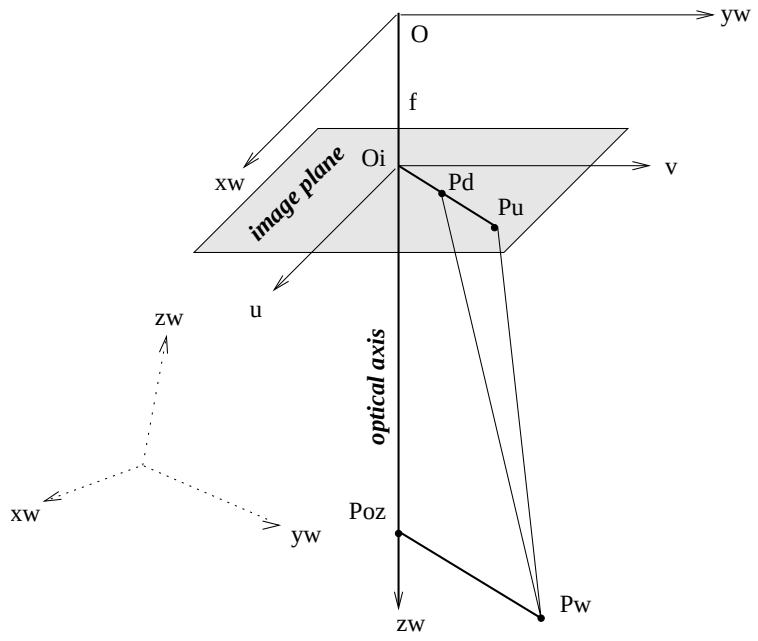


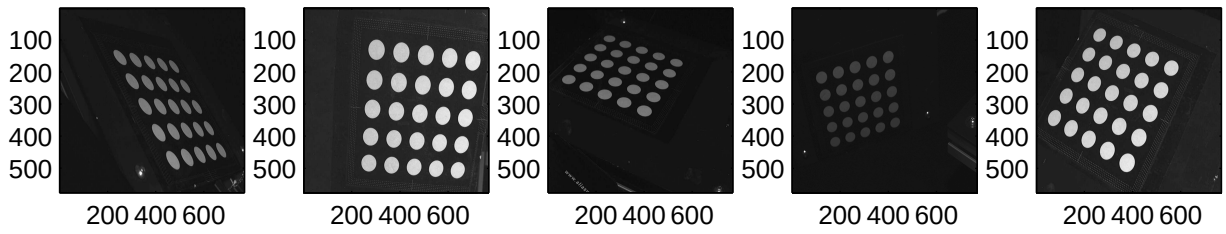
Figure 1: No matter how large the distortion factor is, as soon as it is only a radial effect, the direction of the vector $\overline{O_i P_d}$ extending from the principal point O_i in the image plane to the image point P_d remains unchanged, and is parallel with the vector $\overline{P_{oz} P_w}$ extending from the point P_{oz}, that lies on the *optical axis* and whose z coordinate (here, we refer to the camera coordinate system) is the same as that for the object point P_w, to the real world object point P_w. This is the so-called *Parallelism Constraint.* The constraint that $\overline{O_i P_d} \parallel \overline{P_{oz} P_w}$ for every point P_w, being shown to be independent of radial distortion parameters, the effective focal length, and the z component of 3D translation, is actually sufficient to determine the 3D rotation and the translation components in x and y of the transformation matrix from the world coordinate system (x_w, y_w, z_w) to the camera coordinate system (x_c, y_c, z_c).

The image of the absolute conic $\mathbf{B} = \mathbf{K^{-T} K^{-1}}$ is generally symmetric and can be described as a 6D vector $\mathbf{b} = (b_{11}, \ldots b_{33})^T$. So for each of the N projections of the target we get a set of 2 equations, to solve for $\mathbf{b}$, which implies that we need at least 3 views of the calibration target to solve the system for all unknowns.

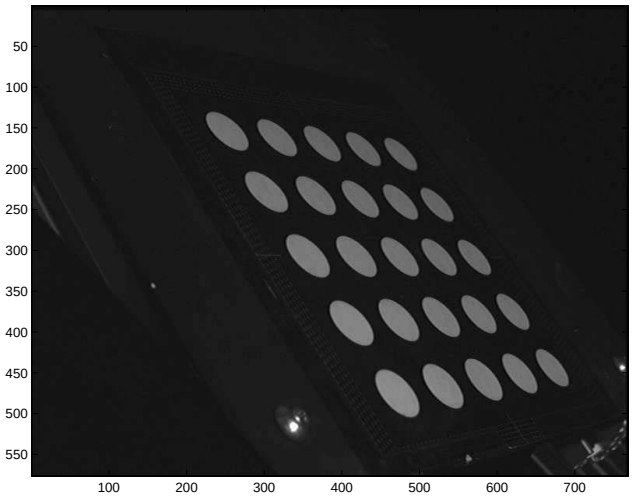
Implementation

Chronologically Tsai's algoritm is oldest one and is still widely used in computer vision, since there are numerous proper implementations in C/C++ or other high-level computer languages. The DLT-based model presented by Janne Heikkilla uses older concepts of Melen [2] and techniques used in photogrammetry. The original implementation is available for matlab. Z. Zhangs method is the newest one and makes use of advanced concepts in projective geometry. Unfortunately only binaries are available from the author.
For comparability of performance we implemented all three algorithms independently in a matlab toolbox. For non-linear estimation procedures we use the original optimization toolbox of matlab version 6.5.

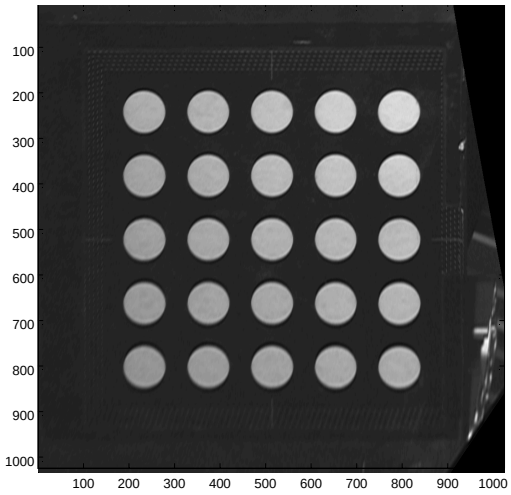
4 Comparison of Real World Test Results



(a)



(b)



(c)

Figure 2: Illustration of the back-projection: (a) All sample images, (b) First sample image, (c) Surface of the calibration target in (b) back-projected to the system plane; The camera model was take from Z.Zhangs method for multiple images in Table 1.

We choose a 5x5 square grid of measured size as calibration pattern giving 25 fiducial points for each view of the target. Concerning illumination, camera/sensor features and pose the setup for the experiments was the same as described in [5],.in the multi-view case we used 5 views of the calibration target.

The fiducial points themeselves are represented by circle centroids. Geometrically this is problematic, since the centroids of projected circles are not invariant. Indeed, in [1] a method to correct for the projective asymmetry was introduced. The correction term is dependent on focal length, circular radius and the outer orientation parameters and is exactly derived by geometrical interpretation. There are basically two approaches to include the correction term in the optimization procedure. One is to directly include the correction term in the camera model, which, as experiments show, leads to poor performance and convergence. Secondly, the problem can be solved iteratively doing the calibration procedures without regarding the asymmetry and computing the correction terms out of the resulting calibration parameters.

Finally the calibration procedure is repeated with corrected center points. The later method does not guarantee an optimal result in a least square sense. Nevertheless, the magnitude of the correction lies in between 0.5-1 [px][3)], and is therefore a considerably small error factor.
After each calibration run we measure the mean squared error by back-projection of the ellipse center points to the system plane[4)] , since this error measurement can be applied to all introduced calibration techniques. Also the whole image information can be back-projected as illustrated in Figure 2 (a) and (b).
Tsai's method is not appropriate for comparison in the multi-view case, since this would require a special recording setup [3]. Z.Zhang's method is generally not applicable to calibrate for a complete model of the inner orientation in the mono-view case. Table 1 contains the estimated values of all intrinsic parameters introduced in Section 1 for all possible four configurations.
Comparing the two runs with a single input image, the final overall error of the DLT-based estimation is significantly smaller than Tsai's method, but yielding a much slower performance with respect to convergence (iterations) and raw computation time. Though this is only one sample experiment the figures generally reflect the overall behavior of both algorithms. As we used only 25 fiducial points building the nonlinear equation system to solve for 12 independent variables in the DLT-based solution it is not that much overdetermined. Tsai's algorithm does not estimate all its parameters simultaneously and defines only 9 free parameters at the final stage of the optimization. In consequence Tsai's partly linearized algorithm has bigger average point errors, even though point errors are never significantly larger than 1 [px] in our experiments.
In the multi-view case we compare the DLT-based method with final bundle-block adjustment [1] with Z.Zhang's method [4]. The later method requires a lot of computation effort at the initial step, which is the iterative non-linear estimation of the homographies for each view, whereas the solution of equations L is a linear least square problem, which gives the DLT-based method an edge in terms of performance. The clear disadvantage of the DLT-based method is that it uses two different camera models, one for the DLT matrix evaluation and one for the final bundle block adjustment. In addition the two parameters b_1 and b_2 in Equation 2 cannot be easily translated to a skewness parameter in the standard camera model for calibration matrix K. Thus, these parameters are simply dropped[5)] for the second calibration stage.We also did not include the skewness parameter in the bundle block adjustment since this turns out to produce even worse estimation results. Hence the DLT-based method generates slightly larger errors than Zhang's (see Table 1).
The values estimated for the intrinsic parameters were in plausible ranges for all configurations.

[3)]We use the unit [px] as length measure for point-to-point distances measured in units equal to the side length of an image pixel

[4)]The system plane is orthogonal to the optical axis with distance f to the optical center representing the (x,y) world coordinates of the calibration target on that plane.

[5)]Parameters become constants of value 1.

By a given nominal focal length of 16mm the estimated values of the calibration parameter vary between 16-20mm, which is a feasible result, since the nominal focal length in the data sheet description of the camera corresponds to true lens projection describing physical characteristics of the optics rather than the ideal perspective projection. The aspect ratio k is only slightly adopted. It is doubtful that this value should be used as free parameter at all, since most cameras of today are pixelsynchronious. As we used 16mm for an 8.8x6.6mm CCD sensor camera with standard mounting, a considerable radial distortion effect can be expected. This is also mirrored by results by parameters κ_1 and κ_2. The estimated coordinates of the principal point (u_0, v_0) hardly coincide with the image centers. Indeed we see a lot of variation in these parameters. We presume that this could be caused by the assumption that the distortion minimum coincides with the principal point. In order to maintain the original concepts of the presented calibration methods we decided not to use more complex camera models which can model also a de-centered radial distortion effect. The skewness parameter is only computed by Zhang since it was not recognized by the other two authors. Surely this parameter can be added to camera model for the other two methods. As we discuss only the original approaches we abstained from experimenting with these kinds of configurations.

Method	**planar DLT**		**R.Y.Tsai**	**Z. Zhang**
Views	single	multiple	single	multiple
k	1.1801	1.0440	=1.05	1.0658
f[mm]	16.3227	17.3652	19.9447	16.4898
u_0[px]	264.1367	378.3158	=384	500.8886
v_0[px]	136.4254	294.5510	=288	355.8066
κ_1	0.0018	0.0011	0.0016	0.0021
κ_2	8.2147e-004	-3.3064e-006	4.0762e-06	2.3456e-06
s(skewness)[deg]	—	—	—	91.23
Error/point[px]	0.098517	0.3874	0.897405	0.1155
Iterations	288	26	74	104
ClkTicks[sec]	3.3430	1.2030	0.6321	8.4690

Table 1: Experimental results in mono- and multiview case. The leading equation signs designate constants,others are the final optimization parameters. The input images are depicted in Figure 2 (a)

5 Conclusion

We compared three different approaches for the solution of the single camera calibration problem by a planar calibration target in the mono- and multi-view case. Results show that the investigated algorithms produce feasible values for determination of a complete camera model for close range photogrammetry.

Though accuracy is comparable the differences in the camera models determine applicability. In the mono-view case with strong radial distortion present R.Y. Tsai's method performs best, since it directly models this feature. Z. Zhang's method has probably the best convergence features in the multi-view case, but is quite low compared to the DLT-based method since more iterative non-linear estimation is required. Setups producing singularities are very rare and not easy to reproduce with real data. For detailed analysis of singularities and error bounds we refer to [2], [3] and [4].
As mentioned we will try to apply more complex camera models working not only with radial, but also with tangential and prism distortion to presented calibration algorithms in future work.

References

[1] J. Heikkila and O. Silven. A four-step camera calibration procedure with implicit image correction. In *IEEE Computer Society Conference on Computer Vision and Patern Recognition (CVPR'97)*, San Juan, Puerto Rico, 1997. IEEE.

[2] T. Melen. *Geometrical modelling and calibration of video cameras for underwater navigation.* PhD thesis, Institutt for teknisk kybernetikk, Norges tekniske hogskole, 1994.

[3] R. Y. Tsai. An efficient and accurate camera calibration technique for 3d-machine vision. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 364–374, Miami Beach, Florida, 1986. IEEE.

[4] Z. Zhang. A flexible new technique for camera calibration. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages pp. 1330–1334, 2000.

[5] H. Zollner. Camera calibration using pose estimation. Technical Report 85, Pattern and Image Processing Group (PRIP), Institute of Computer Aided Automation, Computer Science Dept., Vienna University of Technology, Vienna, Austria, August 2003.