More Accurate Pinhole Camera Calibration with Imperfect Planar Target

Klaus H. Strobl and Gerd Hirzinger
Institute of Robotics and Mechatronics
German Aerospace Center (DLR)
Klaus.Strobl@dlr.de

Abstract

This paper presents a novel approach to camera calibration that improves final accuracy with respect to standard methods using precision planar targets, even if now inaccurate, unmeasured, roughly planar targets can be used. The work builds on a recent trend in camera calibration, namely concurrent optimization of scene structure together with the intrinsic camera parameters [4, 8, 1]. A novel formulation is presented that allows maximum likelihood estimation in the case of inaccurate targets, as it extends the camera extrinsic parameters into a tight parametrization of the whole scene structure. It furthermore observes the special characteristics of multi-view perspective projection of planar targets. Its natural extensions to stereo camera calibration and hand-eye calibration are also presented. Experiments demonstrate improvements in the parametrization of the camera model as well as in eventual stereo reconstruction.

1. Introduction

Camera calibration is the process of estimating the parameters of a camera model that is capable of adequately reflecting the operating principle of the actual camera. Accurately calibrated cameras are prerequisite to most vision-based algorithms. However, researchers still find it challenging to achieve the required accuracy in particular areas like stereo vision or SLAM. Ever since the advent of high-resolution, stereo vision algorithms are demanding higher accuracy (i.e., more accurate epipolar geometry) to keep computational costs in practical terms; SLAM puts similar requirements on calibration accuracy, mainly to reduce dead reckoning drift, thus improving overall performance.

In this paper we are reworking the standard method for camera calibration. On the basis of an adequate camera model, the standard method optimally estimates camera parameters by sensibly minimizing discrepancies between actually collected and expected, model-based data. Of course, the quality of gathered data plays an important role. Data quality, quantity and diversity can be enhanced by controlled scene structure, i.e., using calibration targets. In addition, accurate knowledge of the target's geometry can be provided. It is this last consideration that recently became matter in question [4, 8, 1] and motivates our contribution.

In Section 2 the dangers arising from the requirement of accurate knowledge of the scene structure are discussed. If the latter requirements are lifted, optimal estimation by sensible minimization of reprojection discrepancies must be re-engineered. In Section 3 the standard formulation for camera calibration will be adapted to this new paradigm.

2. Calibration by scene structure estimation

In the early years of computer vision, camera calibration was a cumbersome process. 3-D knowledge of the scene structure was a hard requirement [3, 2, 12] and high quality targets were difficult to achieve. In contrast to this, the possibility to *estimate* the scene structure also exists, as a by-product, along with regular camera calibration. In fact, this was the most important trend in camera calibration since the inclusion of optical distortion models. Tsai strikes this new path by two calibration methods, where the pose of the target (either 3-D or a planar, accurately shifted target) with respect to (w.r.t.) the camera is being estimated [12, 13]. The most significant contribution, however, was simultaneously presented in the late nineties by Zhang, Sturm and Maybank [14, 10]. Their approach allows free motion of a *precisely known* planar calibration target. Their formulation obtains an approximate solution for both the target pose and the camera parameters from the readily obtained object-to-camera homographies, by means of rigid body motion constraints. The approach is flexible and accurate enough to become standard practice to this day. This is *not* because extensive 3-D knowledge of the scene directly compromises calibration accuracy—the contrary is true, but because its flexibility and simplicity prevent damage to calibration owing to human inaccuracies and mistakes [11, 8].

It is pertinent to address this trend towards camera-to-target pose estimation in the context of scene structure estimation, even though they do not yet include the geometry of the calibration target into the optimization. From the camera's point of view, the scene structure is equally determined by both the target's geometry and its relative pose w.r.t. the camera. In other words, the 6 **D**egrees **o**f **F**reedom (DoF) of the target's pose combine with local target geometry to form the actual scene structure that eventually projects unto the camera. This trend therefore provides a clear indication to *additionally estimate the target's geometry*.

2011 IEEE International Conference on Computer Vision Workshops
978-1-4673-0063-6/11/$26.00

In fact, a few authors already made an attempt at this. Since manufacturing accurate 3-D calibration targets is more laborious than manufacturing planar ones, the approach was first taken in 3-D by Lavest *et al.* [4]. Even though their results seem convincing, the method did not become popular, probably because it is formulated in 3-D and, from 1999 on, researchers largely opted for planar calibration targets. Strobl and Hirzinger in Ref. [8] were the first to deal with structure estimation in 2-D. They noticed that off-the-shelf printers systematically cause errors, both in global scale and in aspect ratio of the printed pattern. The pattern is then minimally modeled by these two parameters, which can be simultaneously estimated during intrinsic and hand-eye calibrations, respectively. Albarelli *et al.* [1] go one step further; they observe anisotropic error distribution in reprojected object coordinates after calibration, which leads them to believe that significant, systematic pattern errors are actually pervasive—however small. In addition, the considerable reduction *in residual reprojection errors* reached after full camera and scene structure optimization further strengthens their position, that slightest pattern corrections really imply a significant accuracy improvement in camera calibration. Even though their initial rationale is wrong (anisotropic error distributions are actually expected after nonlinear reprojection of isotropic image noise), their approach is undoubtedly convenient, at least when major inaccuracies in the calibration target occur.

Although the algorithms in Refs. [4] and [1] are indeed very similar (incidentally, the latter fail to cite the former), their conveyed messages differ. While Lavest *et al.* claim that, by using their method, small inaccuracies will not get to harm camera calibration, Albarelli *et al.* on the other hand affirm that, by target geometry optimization, the user will even be able to come by accuracy levels that are otherwise virtually impossible to achieve at moderate cost. Whatever message they convey, the two main differences in their methods are: *Firstly*, Albarelli *et al.* assume planar patterns and have the opportunity to make use of the convenient algorithms in Refs. [14, 10, 8], whereas Lavest *et al.* require 3-D calibration targets and a laborious initialization step. *Secondly*, Lavest *et al.* seem to directly include *all* 3-D geometry of the target into the optimization, for several images, without further ado; Albarelli *et al.* on the contrary are forced to construct an iterative algorithm that decouples geometric estimation of the target from intrinsic parameters estimation. It is mainly this last detail that we rework here.

As stated above, when considering together the target's geometry and its relative pose w.r.t. the camera, they form together the scene structure. A parametrization using both, a rigid body transformation (6 DoF) and $3 \times M$ Euclidean coordinates (in 3-D) for all M feature points, is clearly overparameterized, cannot be estimated unambiguously, and will not converge during nonlinear optimization. A tight parametrization is achieved, e.g., by merely releasing $3 \times M$ Euclidean coordinates. However, it is sensible to take advantage of the relative transformation between the calibration target and the camera ${}_c\boldsymbol{T}^o$, because the local geometry model will then still hold, unmodified, from a different vantage point—this is convenient for multi-view optimization. In the latter case, however, the local geometry model of the calibration target has to be restricted to $3 \times M - 6$ parameters. The authors in Ref. [4] do not mention this issue, which may have been another reason for the limited popularity of their approach. In Ref. [1] the authors encounter this problem; they deal with it by strictly decoupling target geometry and camera parameters estimation in an iterative way. While the latter approach should work, it is not necessary to detach scene structure estimation from intrinsic optimization if a tight parametrization is used. The perspective distortion captured by the images ought to be sufficient to distinguish between camera magnification (i.e. focal length) and scene structure (target geometry and poses ${}_c\boldsymbol{T}^o$)—*up to scale*, provided a multi-view calibration approach was chosen. Furthermore, their rescaling step back to original absolute scale is superfluous, as correct monocular intrinsic calibration is possible irrespective of absolute scale [8].

3. Proposed method: intrinsic camera calibration *and* full scene structure estimation

In this section we are presenting a calibration method that completely releases target geometry and performs jointly with intrinsic parameters estimation. The approach is similar to the standard calibration methods in Refs. [14, 10, 8]. Expected, model-based operation is compared with actual projections; after that, the resulting discrepancies are minimized by tuning parameters in the projection model. In this work the modification w.r.t. standard methods will be the *tight* release of the target's geometry during the final nonlinear optimization. Critically, requirements on the calibration target are now drastically lifted so that unmeasured patterns (e.g., a checkerboard printed on paper using off-the-shelf printers) can be used, even on an uneven surface. The only requirement now is that the pattern remains static during calibration—unless it is rigid material. If stereo camera calibration is intended, a sole scale parameter (e.g., absolute distance between two arbitrary corner features) is required. A potential hand-eye calibration in turn waives this last requirement.

3.1. Feature detection

It is of paramount importance for accurate camera calibration to precisely and robustly detect calibration target features on the images. In fact, Lavest *et al.* argue that, by following this paradigm of concurrent target geometry estimation, the calibration results will no longer depend on the

(lack of) accuracy of the pattern, but mainly on the accuracy of feature detection [4]. Planar checkerboard patterns are certainly convenient in terms of (sub-pixel) localization accuracy of their corners [5, 9], thus our method is conceived for (not restricted to) this type of data.

3.2. Initial closed-form solution

Like most optimization processes that are formulated as residual minimization problems, camera calibration is vulnerable to local solutions. The current standard for its initialization stems from Refs. [14, 10].

In projective geometry, homogeneous plane coordinates are transformed following a *linear* projective transformation called homography. Since the planar sensor of the camera and the planar target are approx. related by a projective transformation, homographies ${}_n\boldsymbol{H}$ between image projections ${}_n\boldsymbol{m}_i = [{}_nu_i\ {}_nv_i\ 1]^T$ and pattern features $\boldsymbol{x}_i = [x_i\ y_i\ 1]^T$ can be estimated ($\widehat{\ }$) from at least four (three out of four non-collinear) correspondences i, for every image n, so that ${}_n\boldsymbol{m}_i \approx {}_n\widehat{\boldsymbol{H}}\boldsymbol{x}_i,\ \ \forall\, n \in \{1,\ldots,N\},\ i \in \{1,\ldots,M\}$.

We aim at the pinhole camera model represented by its intrinsic matrix $\boldsymbol{A}$, which together with the rigid body transformation between the camera frame and the object frame ${}_{c_n}\boldsymbol{T}^o = [{}_n\boldsymbol{r}_1\ {}_n\boldsymbol{r}_2\ {}_n\boldsymbol{r}_3\ {}_n\boldsymbol{t}]$, project 3-D coordinates as follows:

$$ {}_n\boldsymbol{m}_i \approx \boldsymbol{A}\ {}_{c_n}\boldsymbol{T}^o\,[x_i\ y_i\ z_i\ 1]^T. \tag{1} $$

For planar targets ($z_i \triangleq 0$) we have ${}_n\widehat{\boldsymbol{H}} \propto \boldsymbol{A}[{}_n\boldsymbol{r}_1\ {}_n\boldsymbol{r}_2\ {}_n\boldsymbol{t}]$. Since $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are orthonormal: $\boldsymbol{r}_1 \cdot \boldsymbol{r}_2 = 0$, $\boldsymbol{r}_1 \cdot \boldsymbol{r}_1 = 1$, and $\boldsymbol{r}_2 \cdot \boldsymbol{r}_2 = 1$, i.e., ${}_c\boldsymbol{R}^o \in SO(3)$. Sorting the scale out we obtain $\boldsymbol{h}_1^T\,\boldsymbol{\omega}_\infty\,\boldsymbol{h}_2 = 0$ and $\boldsymbol{h}_1^T\,\boldsymbol{\omega}_\infty\,\boldsymbol{h}_1 = \boldsymbol{h}_2^T\,\boldsymbol{\omega}_\infty\,\boldsymbol{h}_2$, where $\boldsymbol{\omega}_\infty = \boldsymbol{A}^{-T}\boldsymbol{A}^{-1}$ is the so-called absolute conic. These two equations hold for N images, leading to $2N$ constraints for e.g. 5 intrinsic unknowns, which can be solved for using a linear least-squares criterion, if $N \geq 3$. The system of equations can be extended if stereo is used [6]. In Ref. [8] Strobl and Hirzinger noted that, if the pattern coordinates are only known up to aspect ratio, normalization of the rotation matrices ($\boldsymbol{h}_1^T\,\boldsymbol{\omega}_\infty\,\boldsymbol{h}_1 = \boldsymbol{h}_2^T\,\boldsymbol{\omega}_\infty\,\boldsymbol{h}_2$) should not be performed. The solution produced hereby is irrespective both, of the absolute scale and of the aspect ratio of the planar pattern, and it suffices to bootstrap nonlinear optimization.

3.3. Optional: nonlinear intrinsic optimization

Since optical distortion has not yet been compensated for during initialization, the user may insert a *standard*, nonlinear optimization in order to support eventual convergence. At this point, the user may choose between the traditional approach and the approach in Ref. [8], where the pattern aspect ratio is being estimated. However, if the expected (prior to printing) pattern dimensions are provided and off-the-shelf printers are used, experiments show that this step can be readily skipped.

Regular nonlinear optimization is maximum likelihood estimation only if perfect knowledge of the target geometry is assumed. It optimizes parameters by minimizing residuals as follows:

$$ \widehat{\boldsymbol{\Omega}}_\star = \arg\min_{\widehat{\boldsymbol{\Omega}}} \sum_{n=1}^{N}\sum_{i=1}^{M} ||{}_n\tilde{\boldsymbol{m}}_i - {}_n\hat{\boldsymbol{m}}_i(\widehat{\boldsymbol{\Omega}},\, \boldsymbol{x}_i)||^2\,, \tag{2} $$

where the optimized ($_\star$) parameters $\boldsymbol{\Omega}_\star$ include the intrinsic matrix $\boldsymbol{A}$, the distortion parameters $\boldsymbol{k} = \{k_1,\, k_2,\, \ldots\}$, and the camera poses ${}_{c_n}\boldsymbol{T}^o$. ${}_n\tilde{\boldsymbol{m}}_i$ are actually measured image projections and ${}_n\hat{\boldsymbol{m}}_i$ are expected, distorted projections of the pattern features $\boldsymbol{x}_i$. The optimization can be extended with the intrinsics of further cameras (stereo camera) along with their rigid transformations w.r.t. the reference camera.

3.4. Simultaneous intrinsic camera calibration and full scene structure estimation

As stated in Section 2, it is sensible to fully release scene structure, extending optimization parameters to the target's geometry. Three recent approaches were reviewed that are either erroneous, incomplete, or needlessly complex. Here we bring forward a novel target parametrization that is perfect complement to the N relative transformations ${}_{c_n}\boldsymbol{T}^o$, to jointly model full scene structure.

Target geometry is a parameter to reprojection in Eq. (2), but it is not part of the optimization parameters $\boldsymbol{\Omega}$. The blunt inclusion of M 3-D target points is suggested in Ref. [4]. This leads to overparametrization when coupled with the N unknown transformations ${}_{c_n}\widehat{\boldsymbol{T}}^o$ ($3\times M+6$ DoF at every station n) and estimations change uncontrollably during optimization, which precludes absolute convergence. To obtain a tight representation, 6 DoF have to be subtracted from the geometric model of the target (now $3\times M-6$ DoF) to overall $3\times M$ DoF at every station n—and scene structure is uniquely defined. However, since intrinsic camera calibration is possible irrespective of the absolute scale of the scene, a further DoF has to be subtracted. In Fig. 1 the 7 DoF that are excluded from optimization are depicted; they involve three corner features—their choice is arbitrary as long as they are non-collinear. Feature $\boldsymbol{x}_1^{\mathrm{3D}} = [0\,0\,0]^T$ is fixed to be pattern origin since else it couples with ${}_n\boldsymbol{t}$. Two other fixed points are $\boldsymbol{x}_2^{\mathrm{3D}} = [d\,0\,0]^T$ and $\boldsymbol{x}_3^{\mathrm{3D}} = [x_3\ y_3\ 0]^T$. $y_2 \triangleq 0$, $z_2 \triangleq 0$, and $z_3 \triangleq 0$ fix the target orientation so that it will not get coupled with the orientation in ${}_{c_n}\widehat{\boldsymbol{T}}^o$. $x_2 \triangleq d$ fixes the absolute pattern scale to an *arbitrary* value. In spite of these constraints, the target geometry is still released up to its absolute scale—which cannot be estimated during calibration after all.

The new optimization parameters $\boldsymbol{\Omega}^+$ include x_3, y_3, and $\boldsymbol{x}_i^{\mathrm{3D}}\ \forall i \in \{4,\ldots,M\}$, i.e. $3\times(M-3)+2$ variables:

$$ \widehat{\boldsymbol{\Omega}}_\star^+ = \arg\min_{\widehat{\boldsymbol{\Omega}}^+} \sum_{n=1}^{N}\sum_{i=1}^{M} ||{}_n\tilde{\boldsymbol{m}}_i - {}_n\hat{\boldsymbol{m}}_i(\widehat{\boldsymbol{\Omega}}^+,\, d)||^2. \tag{3} $$

In doing so, we cast the former Eq. (2) into a much harder optimization task as the parameters vector length skyrockets from e.g. $5+2+6\times N$ to $5+2+6\times N+3\times(M-3)+2$, where $M \gg N$. Being the residuals vector already long (up to $2\times M\times N$), the required Jacobian matrix increases exponentially in size. Even though computing efficiency is uncritical in camera calibration, we recommend providing Jacobian sparsity patterns to this optimization.[1]

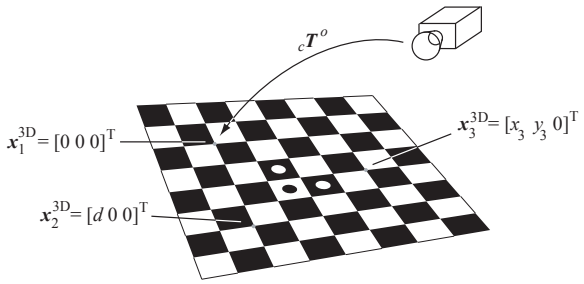

Figure 1. Pattern features $\boldsymbol{x}_1^{3D}$, $\boldsymbol{x}_2^{3D}$ and $\boldsymbol{x}_3^{3D}$ that will be (in part) fixed during joint intrinsic and full scene structure optimization.

3.5. Extension #1: stereo camera calibration

A natural extension of this work is in the case of stereo:

$$\widehat{\boldsymbol{\Omega}}_\star^{\oplus} = \arg\min_{\widehat{\boldsymbol{\Omega}}^{\oplus}} \sum_{c=1}^{C}\sum_{n=1}^{N}\sum_{i=1}^{M} ||{}_n^c\tilde{\boldsymbol{m}}_i - {}_n^c\hat{\boldsymbol{m}}_i(\widehat{\boldsymbol{\Omega}}^{\oplus},\, d)||^2. \quad (4)$$

Compared to $\boldsymbol{\Omega}^+$ the optimization parameters $\boldsymbol{\Omega}^{\oplus}$ additionally include the intrinsics of further cameras ($\boldsymbol{A}_c$, $\boldsymbol{k}_c$) and their rigid, relative transformations ${}_{c1}\boldsymbol{T}^{cc}$ w.r.t. the reference camera $c1$. If two cameras are used, the residuals vector length amounts to up to $2\times 2\times M\times N$, and the parameters vector to, e.g., $(5{+}2){\times}2{+}6{+}6{\times}N{+}3{\times}(M{-}3){+}2$. It is worth noting that the value of d is *not* arbitrary anymore, if stereo ought to be calibrated to correct metric scale; the user has to provide a valid distance d between two (arbitrary) features.

3.6. Extension #2: hand-eye calibration

Hand-eye calibration is the estimation of the rigid body transformation ${}_t\boldsymbol{T}^{c1}$ relating the end-effector frame of, e.g., a robot manipulator (hand, t) to the reference camera frame (eye, $c1$) mounted on it [7]. Similar to stereo calibration, *regular* hand-eye calibration *requires* correct metric scale. Since more often than not hand-eye calibration is decoupled from intrinsic camera calibration, the hand-eye calibration method presented by Strobl and Hirzinger in Ref. [8], *Method #1*, still holds. In a nutshell: The discrepancies ($\mathcal{O}_n$) between *expected* and *measured* transformations are minimized. Expected transformations ${}_{c1}\widehat{\boldsymbol{T}}_\star^o$ stem from intrinsic calibration; measured transformations ${}_b\widetilde{\boldsymbol{T}}^t$ from the (erroneous) motion readings of the manipulator. Note that here the absolute scale d can be simultaneously estimated during optimization. Following the notation in Refs. [7, 8]:

$$\{{}_t\widehat{\boldsymbol{T}}_\star^c, {}_b\widehat{\boldsymbol{T}}_\star^o, \hat{d}_\star\} = \arg\min_{{}_t\widehat{T}^c,\, {}_b\widehat{T}^o,\, d} \sum_{n=1}^{N} \mathcal{O}_n(\Phi({}_c\widehat{\boldsymbol{T}}_\star^o,\, \hat{d}),\, {}_b\widetilde{\boldsymbol{T}}^t,\, ...)$$

where the function Φ scales the transformations ${}_c\widehat{\boldsymbol{T}}_\star^o$ in range—according to the scaling factor $\hat{d}$ being estimated. If this method is used, the user does not even need to provide a valid distance d for stereo calibration; he or she only needs to rescale ${}_{c1}\widehat{\boldsymbol{T}}_\star^{cc}$ back to correct metric scale using $\hat{d}_\star$.

4. Experiments

In this section we are analyzing the results of last section's method, both on calibration data and in independent validation experiments. After that we shall discuss on the utility of the presented approach.

A stereo camera was used consisting of two progressive scan AVT Marlin cameras with SVGA 1/2" Sony CCD chips and Sony VCL-06S12XM 6 mm objectives; experiments show that a radial distortion model using only two parameters (3^{rd} and 5^{th} degree) suffice to model the optics' geometric distortion. Stereo base distance is approx. 5 cm.

Two calibration targets are used: On the one hand a precision pattern size A3 printed on a metallic plate; on the other hand a printed A3 sheet of paper with the same checkerboard pattern of $14\times 20 = 280$ corner features. The distance between features is approx. 2 cm. The paper pattern was folded previous to calibration to affect its planarity, thus represents a worst-case scenario, see Fig. 2. In both cases the calibration consists of 12 *tilted* images at three different heights w.r.t. the pattern: 20, 40 and 80 cm. Of course, not all corner points are seen in every image.

At this point it is worth mentioning the reason for taking additional images at different heights, since usually 4 or 8 images suffice: It is critical to optical distortion estimation to fill in images with features, so that distortion can be correctly estimated in the image corners [9]. Naturally, some features in the image corners might be imaged only once. Using our novel method, those lone features are now totally released in 3-D to match their actual image projections, thus will not enforce correct distortion model parametrization. To avoid lone features, we additionally take distant images.

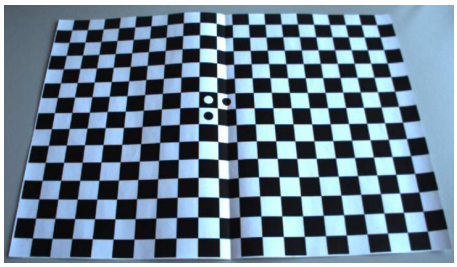

Figure 2. Wrinkled paper calibration target size A3.

[1] The use of analytical Jacobians is here, however, discouraged as residuals are in distorted image space, Jacobians are hard to get, and it is too costly to perform variable substitution on them in the first place.

Table 1. Intrinsic parameters after standard (Std.) and simultaneous scene structure and stereo calibration (Full), using a precision target.

	Left camera							Right camera							*(both)*
	${}_L\alpha$	${}_L\beta$	${}_Lu_0$	${}_Lv_0$	${}_Lk_1$	${}_Lk_2$	*RMS*	${}_R\alpha$	${}_R\beta$	${}_Ru_0$	${}_Rv_0$	${}_Rk_1$	${}_Rk_2$	*RMS*	*RMS*
Std.	724.79	724.12	372.42	272.34	−0.1963	0.0995	0.155	728.31	728.01	391.73	270.35	−0.1962	0.1008	0.173	0.165
Full	724.32	724.35	372.20	271.22	−0.1973	0.0993	0.078	727.85	728.40	391.55	269.23	−0.1982	0.1033	0.077	0.077

Calibration starts out by accurately detecting and locating corners in the images using DLR CalDe [9], with subpixel accuracy. Since calibration will also estimate the target's geometry, it is not necessary to provide accurate pattern dimensions—experiments in Ref. [8] showed strong convergence in a similar scenario. However, the metallic plate was initially meant to deliver ground-truth geometry, or rather to show the potential precision in target geometry estimation, thus we do adopt accurate pattern dimensions in that case (actual square size is 19.985×19.950 mm).

The optimization method in Sections 3.4 and 3.5 was implemented in MATLAB®; the used Levenberg-Marquardt optimization function is lsqnonlin in its large-scale variant. We choose to provide Jacobian patterns for its sparse numerical implementation to keep computational costs low.

4.1. Joint optimization of camera and scene

Next we are showing the resulting camera parameters and scene structure as well as the residuals after calibration both in image and in 3-D target coordinates.

4.1.1 Using an accurate, planar metallic target

Planar calibration targets imprinted on metallic plates provide both structural stability and high planarity. This is a best-case scenario to camera calibration, thus less profit is expected from concurrent scene structure estimation.

In fact, both monocular (Section 3.4) and stereo (Section 3.5) joint intrinsic and full scene structure estimations deliver almost identical camera parameters w.r.t. the standard approaches, cf. Tabs. 1 and 2. The reason is the very slight optimization of the pattern structure achieved, see Fig. 3, in the region of a tenth of a millimeter. The target optimization is mainly in its 2-D imprinted pattern because, apparently, it is still subject to inaccurate printing errors similar to off-the-shelf paper printers, see Fig. 6 (a). Fig. 3 (a) shows a planarity correction in the order of a tenth over 200 mm—a very slight bending of the plate.

A remarkable result is, however, the significant reduction both in image and object reprojection residuals, see Figs. 5 and 6. Image reprojection residual errors are measured by their **R**oot **M**ean **S**quare error (RMS). Nevertheless, these reductions result from calibration-related minimizations and their potential effects in final accuracy still have to be experimentally verified, see Section 4.2.

Table 2. Intrinsics after standard (Std.) and simultaneous scene structure and monocular calibration (Full), using a precision target.

	${}_L\alpha$	${}_L\beta$	${}_Lu_0$	${}_Lv_0$	${}_Lk_1$	${}_Lk_2$	*RMS*
Std.	724.58	723.93	372.44	272.17	−0.1960	0.0994	0.151
Full	724.50	723.69	371.92	271.08	−0.1955	0.0975	0.063

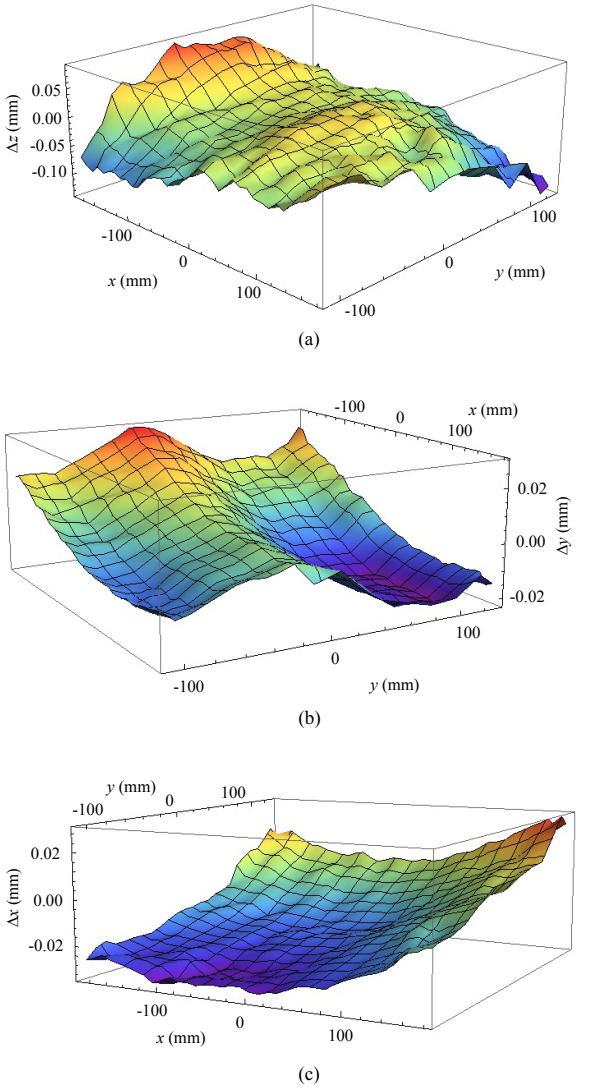

Figure 3. Corrected feature positions Δz (height), Δy and Δx (in 2-D) after joint intrinsic and full scene structure estimation on the precision target. Corrections are consistent after monocular and stereo approaches.

Table 3. Intrinsics after standard (Std.) and simultaneous scene structure and stereo calib. (Full), using an unknown, wrinkled paper target.

	Left camera							Right camera							*(both)*
	${}_L\alpha$	${}_L\beta$	${}_Lu_0$	${}_Lv_0$	${}_Lk_1$	${}_Lk_2$	*RMS*	${}_R\alpha$	${}_R\beta$	${}_Ru_0$	${}_Rv_0$	${}_Rk_1$	${}_Rk_2$	*RMS*	*RMS*
Std.	718.99	724.10	362.60	268.97	−0.2534	0.1706	2.180	719.64	724.71	393.26	271.03	−0.2050	0.0840	2.084	2.133
Full	724.71	724.07	372.53	270.87	−0.1968	0.0971	0.111	728.18	728.08	391.74	268.89	−0.1981	0.1013	0.115	0.113

4.1.2 Using an unknown, wrinkled paper target

Checkerboard patterns on paper using off-the-shelf printers are the most convenient calibration targets that still guarantee accuracy and repeatability in detection and location of their corner projections, through several images.[2] Indeed, printed patterns are the most used calibration targets worldwide [5, 9]. Researchers struggle to stick them on planar surfaces and, more often than not, to measure up their dimensions. Eventually they get humid and bumpy and need to be replaced.

For reasons of space we are addressing a worst-case scenario where the pattern is *not* measured after printing. We assume corner distances of 2 cm as in the original PostScript® file. On top of that, the paper target has a folding mark in the middle so that it is clearly not planar, see Fig. 2.

Standard camera calibration cannot deliver accurate results over this pattern, see Tabs. 3 and 4. The image reprojection residuals after calibration in Fig. 7 (a) are very high, owing to strong systematic errors in the object model, see Fig. 8 (a). The presented methods in Sections 3.4 and 3.5 *do* compensate for these object model errors, see Figs. 7 (b) and 8 (b), so that the intrinsic camera parameters virtually match former results in Tabs. 1 and 2.[3] The object model optimization performed during calibration is depicted in Fig. 4. The results correspond with the expected deformation showing unevenness of approx. 6 mm.

A further drawback of using the standard method with this type of patterns is that measuring its dimensions is difficult, as the pattern is delicate and easy to deform. By using our method this step is rendered superfluous. In the case of stereo calibration (Section 3.5), the input of a single absolute distance between two arbitrary pattern corners suffices. If hand-eye calibration is additionally performed, the user can spare this last measurement.

[2] The only more convenient calibration target is unstructured scenery (self-calibration), which does not, however, guarantee accurate and robust feature detection and localization.

[3] In the case of stereo calibration, residuals do not quite reach the levels of the metallic pattern, cf. Tabs. 1 and 3. If the paper was not folded but directly put on a table after printing, results do match exactly, irrespective of natural paper bending. The difference can be explained either by noisy detection of pattern features due to local shadows, or by stagnant convergence of the nonlinear optimization. Either way, the validity of the parametrization is not stated by the calibration RMS but by independent validation experiments, see Section 4.2.

Table 4. Intrinsics after standard (Std.) and simult. scene structure and monoc. calibration (Full), using an unknown, wrinkled paper.

	${}_L\alpha$	${}_L\beta$	${}_Lu_0$	${}_Lv_0$	${}_Lk_1$	${}_Lk_2$	*RMS*
Std.	718.65	723.05	370.78	268.53	−0.2518	0.1721	2.105
Full	724.35	723.58	372.18	270.90	−0.1943	0.0946	0.069

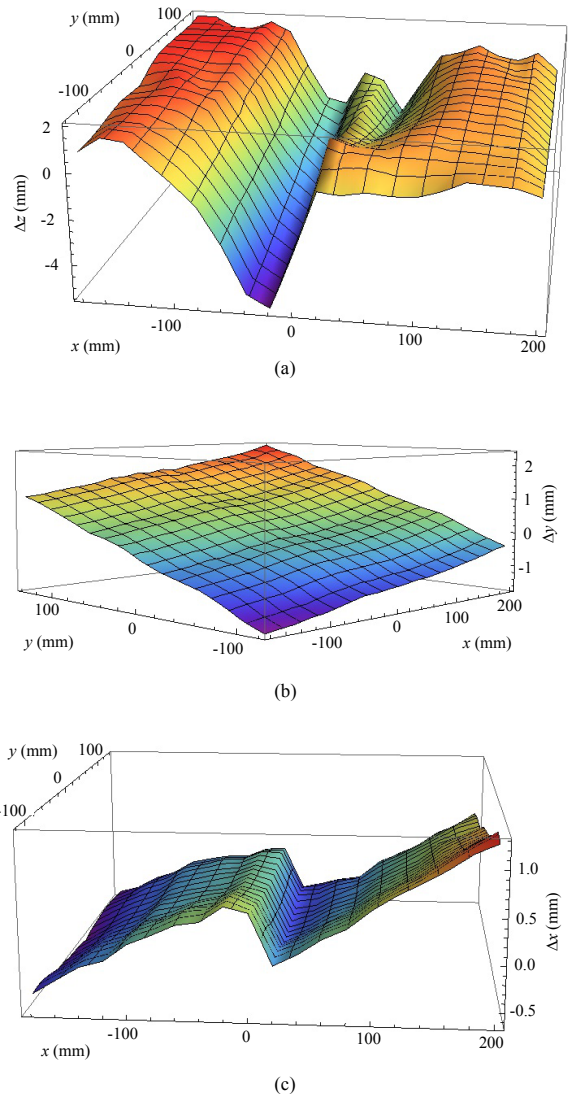

Figure 4. Corrected feature positions Δz (height), Δy and Δx (in 2-D) after joint intrinsic and full scene structure estimation on the paper target. Corrections are consistent after monocular and stereo approaches.

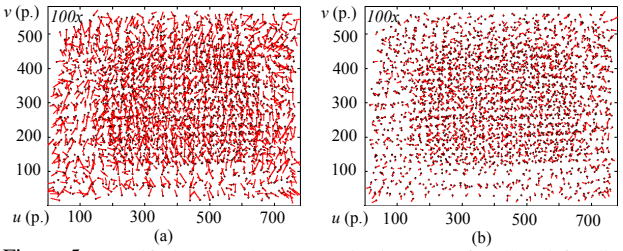

Figure 5. Magnified (100×) image reprojection errors for all 12 left calibration images after std. camera calibration (a) and after full estimation (b), using a precision pattern. RMS error reduces from 0.151 to 0.063 p.

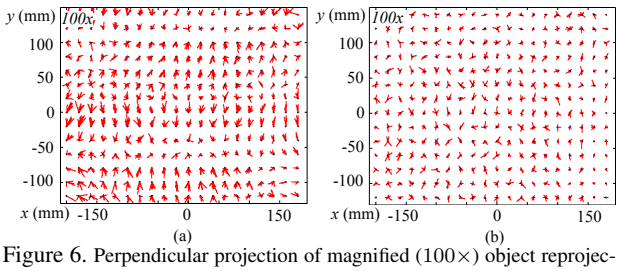

Figure 6. Perpendicular projection of magnified (100×) object reprojection errors for all 12 left calibration images after standard camera calibration (a) and after full estimation (b), using a metallic precision pattern.

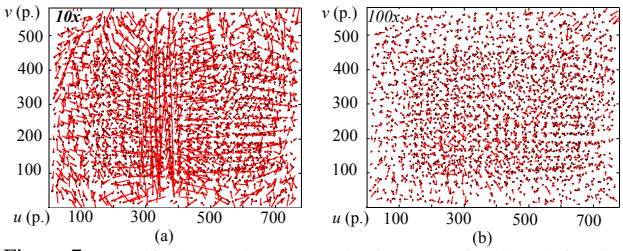

Figure 7. Magnified (10×) image reprojection errors for all 12 left calibration images after std. camera calibration (a) and after full estimation (b), using a wrinkled paper pattern. RMS error reduces from 2.105 to 0.069 p.

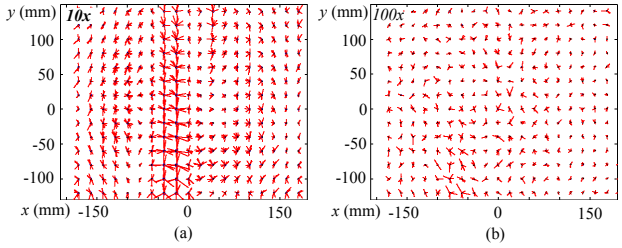

Figure 8. Perpendicular projection of magnified (10×) object reprojection errors for all 12 left calibration images after standard camera calibration (a) and after full estimation (b), using a wrinkled paper pattern.

4.2. Accuracy evaluation

Next we are showing stereo triangulation results on data independent from calibration that will be used to check calibration methods against each other. We replicate the validation experiment in Ref. [1], which measures the distance d *between* two rigid points in 3-D space. The camera continuously moves in the direction of its optical axis. In order to reach optimal feature localization accuracy, we take two particular corner features in a checkerboard pattern that is standing perpendicular to the camera motion. The features are approx. 22 cm apart from each other.

The measured distance d is irrelevant to our analysis as it ultimately depends on the accuracy when measuring the pattern scale by hand during calibration, which is naturally limited (refer to Section 3.5). A valid hint for calibration accuracy is, however, the *consistency* of the distance estimation at different triangulation ranges [1]. Fig. 9 (a) shows that, both with and without full scene structure estimation, the metallic plate-based stereo camera calibration delivers near-constant estimations, drifting half a millimeter (out of 220 mm) from 0.3 to 1 m range. Paper target-based calibration causes a major drift of 2 mm unless full structure estimation is performed—then results match the former.

Flawless stereo triangulation is of course impeded by inaccurate feature detection and imperfect camera calibration—i.e., estimated ray directions will not intersect. As a triangulation result for a particular feature, we choose the 3-D point $\boldsymbol{i}$ in the middle of the segment of minimum distance between the left (camera $c1$) and the right (camera $c2$) stereo rays ${}_{c1}\boldsymbol{l}$ and ${}_{c2}\boldsymbol{r}$. It can be represented as follows:

$$ {}_{c1}\boldsymbol{i} \;=\; L\,{}_{c1}\boldsymbol{l} + \frac{N}{2}\,{}_{c1}\boldsymbol{n} \;=\; R\left({}_{c1}\widehat{\boldsymbol{R}}_{\star}^{c2}\,{}_{c2}\boldsymbol{r}\right) + {}_{c1}\hat{\boldsymbol{t}}^{c2} - \frac{N}{2}\,{}_{c1}\boldsymbol{n} \;\Big/ $$
$$ {}_{c1}\boldsymbol{n} \;=\; {}_{c1}\boldsymbol{l} \times {}_{c1}\widehat{\boldsymbol{R}}_{\star}^{c2}\,{}_{c2}\boldsymbol{r}, \quad L \in \mathbb{R}, \quad N \in \mathbb{R}, \quad R \in \mathbb{R}\,. \qquad (5) $$

Eq. (5) forms a linear system of 3 equations and 3 unknowns L, N and R that is solved using LU factorization. Similar to consistency in distance estimation in Fig. 9 (a), the minimum distance N between stereo reprojection rays also indicates calibration accuracy. Fig. 9 (b) shows its mean value for both corner points w.r.t. camera range. For the metallic plate-based calibration, stereo triangulation is performed with half a tenth of a millimeter triangulation error at any distance tested. Scene structure estimation does slightly improve consistency (9.9% error decrease). Results are clearer for the paper target-based calibration, where triangulation errors increase to four tenths of a millimeter at far range, if the standard calibration method is used. If scene structure estimation was performed, error levels shrink anew to half a tenth of a millimeter (**72**% error decrease), exactly as when using the metallic plate.[4]

It is worth noting that the extrinsic rigid transformation between cameras ${}_{c1}\widehat{\boldsymbol{T}}^{c2}$ is mainly responsible for the results presented here. Contrary to the experiment in Ref. [1], in this work the stereo transformation *fully* results from the full structure estimation paradigm in question, see Section 3.5. Furthermore the examined range extends to 1 m.

[4] More specifically 7.8% worse than after full scene structure estimation using the metallic plate, but then 3.4% *better* than standard calibration using the precision metallic pattern.

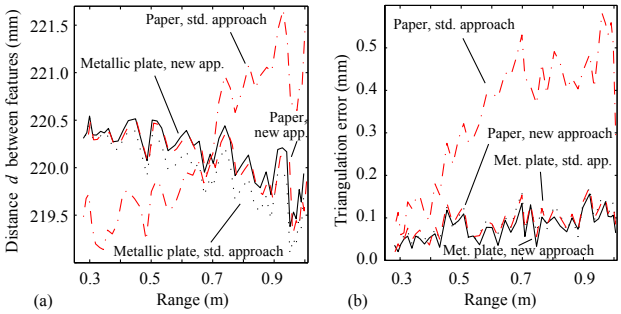

Figure 9. Validation by stereo: (a) Distance d between rigid points and (b) mean value of the triangulation error, w.r.t. camera range.

4.3. Discussion

At first the above results make for somewhat of a disappointment. If camera calibration is dutifully performed, less extra accuracy is attained by simultaneous estimation of full scene structure.[5] All things considered, however, it is very difficult for most users to produce an exact calibration target and, on top of that, it comes at no cost to calibration accuracy to perform simultaneous intrinsic and full scene structure estimation as long as two slight limitations are observed: Firstly, to avoid gathering features in image corners with exclusive support; additional images are encouraged where the pattern is fully captured.[6] Secondly, the calibration target has to remain static unless it is rigid material; it is the camera and not the calibration target that should be shifted for grabbing images.

The preliminary results in this work show that simultaneous intrinsic and full scene structure estimation should be performed in any situation where the calibration target is expected to be nearly planar. Apart from delivering results at least as accurate as from a flawless standard implementation, the method deskills the calibration procedure, thus prevents damage from pattern inaccuracies and human mistakes. This is especially true in the case of printed paper patterns or bigger targets (e.g. patterns projected by an overhead projector), which are difficult to measure accurately. In view of the blatant similarity to bundle adjustment—gold standard for structure from motion approaches, the current methods have the potential to be considered *gold standard for pinhole camera calibration using planar targets*.

[5] At Ref. [1] Albarelli *et al.* observe that, using their method on an accurate planar target, scene structure is optimized prior to camera parameters since this minimizes residuals faster—they cannot provide an explanation for that. Our read of this phenomenon is that, since it is only scene structure optimization that minimizes residual errors, camera parameters do not *significantly* change. Standard, least-squares optimization with abundant, redundant data *already compensated* for the former structure inaccuracies, thereby delivering optimal, accurate intrinsic parameters in the first place.

[6] For that matter, it is widespread to take *only* this type of images during camera calibration, which is harmful to accurate estimations.

5. Conclusion

The novel approach presented in this paper enables researchers to perform camera calibration using, e.g., freshly printed patterns, outperforming conventional methods that *require* precision patterns. The pattern does not even need to be measured for monocular camera calibration. If stereo camera calibration with correct metric scale is intended, a single distance measurement has to be provided unless hand-eye calibration following Ref. [8] is performed—then again no pattern measurements are required.

Experiments on real calibration data and an evaluation of stereo reconstruction accuracy validate the approach. The algorithm will be soon included in the camera calibration toolbox DLR CalDe and DLR CalLab [9].

References

[1] A. Albarelli, E. Rodolá, and A. Torsello. Robust Camera Calibration using Inaccurate Targets. In *Proc. of the BMVC*, Aberystwyth, UK, Sep. 2010.

[2] O. D. Faugeras. *Three-Dimensional Computer Vision*. Artificial Intelligence. MIT Press, Cambridge, MA, USA, 1993.

[3] O. D. Faugeras and G. Toscani. Camera Calibration for 3D Computer Vision. In *Proc. of the Int. Workshop on MVMI*, p. 240, Tokyo, Japan, Feb. 1987.

[4] J.-M. Lavest, M. Viala, and M. Dhome. Do We Really Need an Accurate Calibration Pattern to Achieve a Reliable Camera Calibration? In *Proc. of the ECCV*, volume 1, p. 158, Freiburg, Germany, 1998.

[5] J. Mallon and P. F. Whelan. Which Pattern? Biasing Aspects of Planar Calibration Patterns and Detection Methods. *Pattern Recognition Letters*, 28(8):921–930, June 2007.

[6] H. Malm and A. Heyden. Stereo Head Calibration from a Planar Object. In *Proc. of the CVPR*, p. 657, USA, Dec. 2001.

[7] K. H. Strobl and G. Hirzinger. Optimal Hand-Eye Calibration. In *Proc. of the IROS*, p. 4647, Beijing, China, Oct. 2006.

[8] K. H. Strobl and G. Hirzinger. More Accurate Camera and Hand-Eye Calibrations with Unknown Grid Pattern Dimensions. In *Proc. of the ICRA*, p. 1398, Pasad., USA, May 2008.

[9] K. H. Strobl *et al.* DLR CalDe and DLR CalLab [Online].

[10] P. F. Sturm and S. J. Maybank. On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications. In *Proc. of the CVPR*, p. 432, Fort Collins, USA, June 1999.

[11] W. Sun and J. R. Cooperstock. An Empirical Evaluation of Factors Influencing Camera Calibration Accuracy Using Three Publicly Available Techniques. *Machine Vision and Applications*, 17(1):51–67, Mar. 2006.

[12] R. Y. Tsai. An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. In *Proc. of the CVPR*, p. 364, Miami, Florida, USA, 1986.

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

[14] Z. Zhang. A Flexible new Technique for Camera Calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, Nov. 2000.