Forget the checkerboard: practical self-calibration using a planar scene

Daniel Herrera C. Juho Kannala Janne Heikkilä
University of Oulu
Finland

Abstract

We introduce a camera self-calibration method using a planar scene of unknown texture. Planar surfaces are everywhere but checkerboards are not, thus the method can be more easily applied outside the lab. We demonstrate that the accuracy is equivalent to a checkerboard-based calibration, so there is no need for printing checkerboards any more. Moreover, the use of a planar scene provides improved robustness and stronger constraints than a self-calibration with an arbitrary scene. We utilize a closed-form initialization of the focal length with minimal and practical assumptions. The method recovers the intrinsic and extrinsic parameters of the camera and the metric structure of the planar scene. The method is implemented in a real-time application for non-expert users that provides an easy and practical process to obtain high accuracy calibrations.

1. Introduction

Calibrating a camera's intrinsics (focal length, principal point, and distortion coefficients) is a fundamental problem in computer vision. A calibrated camera is needed to perform a metric reconstruction of a scene, otherwise only a projective reconstruction is possible [7]. Some of the most interesting applications of computer vision, like simultaneous localization and mapping, augmented reality, and 3D reconstruction, require a metric reconstruction of the scene. Nowadays, cameras are most often calibrated offline using a calibration target. A planar target with a checkerboard or circular pattern of known structure is a well established and popular method for camera calibration [13, 15, 4, 8].

Planar scenes are a very convenient calibration target because they are easy to detect, match, and the observed motion can be completely described by a homography. A planar target is much easier to manufacture than a 3D target of known structure, for example a paper checkerboard can be produced by a standard printer and attached to a table. Homography-based known-target calibration methods like [13, 15] extract the intrinsic parameters of a camera from a set of homographies between the known points in metric space and the matched points in image space. This produces a very accurate calibration because the homographies are very robust to noise and outliers.

Although this is an accurate and practical method, it is not as practical as it can be. Checkerboard targets are often inconvenient and not always available, especially outside the lab. There are a myriad of well-textured planar targets in the wild (books, paintings, billboards) but the metric structure of their texture is not known a-priori and are thus not suited for a method like [15]. We explore the problem of *planar self-calibration* which attempts to simultaneously calibrate the camera and recover the metric structure of the scene under the assumption that the scene is planar.

Planar self-calibration is attractive for several reasons. It is much more practical than a known-target calibration because we can use any planar structure for calibration. On the other hand, when compared to a generic self-calibration approach, the planar-scene constraint significantly reduces the degrees of freedom of the problem and increases the robustness and accuracy of the calibration. Moreover, homography estimation is a much simpler and robust process than feature matching of arbitrary 3D points.

An internal and necessary component of planar self-calibration is *homography-based self-calibration* which takes a set of homographies and estimates the intrinsic parameters of the camera. So far, homography-based self-calibration has been proven possible [14] but it has not led to a practical implementation that can replace known-target calibration methods due to its lack of a closed-form solution for initialization. Checkerboard-based calibration is still the standard calibration technique due to several reasons. Homography-based self-calibration by itself cannot reach the accuracy of a known-target calibration because the camera poses are implicitly fixed within the homographies. There hasn't been a proper comparison of the performance of both calibration types that would prove that it is safe to use planar self-calibration without sacrificing accuracy. Moreover, there is no publicly available implementation as there is for known-target calibration, *e.g.* [4].

In this paper we address these issues and describe a complete planar self-calibration system that rivals the performance of known-target calibration methods. We compare the limits of both calibration types and show that in almost any practical situation, planar self-calibration can be used to obtain calibrations with the same accuracy as that of known-target calibration methods.

Previous work

Hartley and Zisserman [7] present a comprehensive analysis of camera calibration, including known-target and self-calibration methods. In the area of known-target calibration the landmark paper of Zhang [15] is nowadays the defacto standard for camera calibration and has been implemented for many platforms [4, 5]. It is interesting to notice that the calibration constraints used there have the same nature as the ones presented here. However, because the metric structure of the world is known the equations simplify considerably, there are less degrees of freedom, and the optimization is performed directly in metric space.

There has been considerable progress in the area of self-calibration when the observed scene has a 3D structure. Closed-form and linear solutions have been obtained to recover the focal length of a moving camera [3]. However, these methods fail when the scene is planar. The planar self-calibration constraints were first introduced by Triggs [14]. Triggs encoded the plane structure with two circular points using 4 degrees of freedom (DoF). Bocquillon *et al.* [2] later reformulated these constraints to encode the plane structure using only the plane normal (2 DoF) and solved the calibration problem using interval analysis which results in an exhaustive search through parameter space.

Gurdjos and Sturm [6] took a different approach using the *centre line constraint*. Their formulation keeps the DoF constant even with a varying focal length. However, it is less constrained than the formulation of [2] and requires more images to reach the same accuracy. Gurdjos and Sturm also provide a closed-form solution under a minor assumption, namely that the reference image is close to fronto-parallel with the scene plane.

Our system combines the ideas from [2] and [6] and extends them into a complete planar self-calibration system. We take a novel approach to derive the homography-based self-calibration constraints which highlights a weakness of the existing formulation. We solve this by providing a new set of normalized constraints. We use the same assumption as [6] to obtain a closed-form solution and bootstrap the calibration. Then, we obtain a significant improvement in the final accuracy by normalizing the constraints and adding a final bundle adjustment in metric space. Our contributions can be summarized as follows

- A new set of normalized planar self-calibration constraints that are more robust to noise.

- A novel derivation of the planar self-calibration constraints that is more intuitive and provides insight into the nature of the constraints.

- A complete planar self-calibration pipeline that robustly obtains a final metric reconstruction and calibration with an accuracy significantly higher than that of a homography-based self calibration.

- A direct comparison of the accuracy of planar known-target and self-calibration methods.

- An open source implementation of the planar self-calibration system.

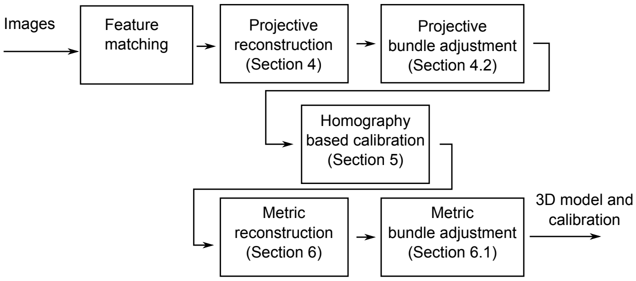

Images
Feature matching
Projective reconstruction (Section 4)
Projective bundle adjustment (Section 4.2)
Homography based calibration (Section 5)
Metric reconstruction (Section 6)
Metric bundle adjustment (Section 6.1)
3D model and calibration

Figure 1: The outline of our planar self-calibration algorithm. It first performs a projective reconstruction, then recovers the calibration matrix from the obtained homographies, and then upgrades it to a metric reconstruction.

2. The planar self-calibration system

The homography-based self-calibration stage as described by [2] and [6] is a key component for planar self-calibration but it is not enough to produce the best calibration results. A projective reconstruction stage is needed before it to prepare the homographies and a metric recronstruction stage is needed after it to perform a final bundle adjustment. The structure of the final system is shown in Fig. 1. We first describe our camera model in Section 3.

The projective reconstruction stage receives the input images and extracts the homographies between them. It also gives an initial estimate of the distortion. It initially estimates each homography individually and then performs a global bundle adjustment of all homographies, the distortion coefficients, and the observed points in projective space. Section 4 describes this stage in detail.

Once the optimal set of homographies is found, they are fed into a homography-based self-calibration stage that simultaneously recovers the camera calibration and the pose of the reference camera. We derive the planar self-calibration constraints and propose a novel set of normalized constraints in Section 5.

The calibration obtained allows a metric reconstruction of the scene. However, this calibration is not optimal because the camera poses were not optimized together with the intrinsic parameters. Thus, a metric reconstruction is performed and a final bundle adjustment is done to obtain the optimal calibration. Details of this reconstruction and final optimization are described in Section 6.

Notation: We denote vector quantities as bold lowercase letters (*e.g.* $\mathbf{x}$,$\mathbf{p}$,$\mathbf{t}$), matrix quantities as bold uppercase letters (*e.g.* $\mathbf{R}$,$\mathbf{K}$), and scalars as lowercase italic letters (*e.g.* f_x,u_0). We denote the homogeneous representation of a vector $\mathbf{x}$ with $\hat{\mathbf{x}}$. The transformation back from homogeneous coordinates, denoted by $\check{\nu}(\hat{\mathbf{x}})$, is performed by dividing a vector by its last component and discarding it.

3. Camera model

We model our camera using the well-known pinhole model with radial distortion. The projection function $\mathbf{p} = \mathcal{P}(\mathbf{x})$ transforms a point in 3D world space $\mathbf{x} = [x, y, z]^\top$ to a 2D pixel position $\mathbf{p} = [u, v]^\top$. The projection function is the composition of three functions: the extrinsic transform $\mathcal{T}$, the intrinsic transform $\mathcal{K}$, and the distortion function $\mathcal{D}$, *i.e.* $\mathcal{P}(\mathbf{x}) = \mathcal{D} \circ \mathcal{K} \circ \mathcal{T}(\mathbf{x})$.

The extrinsic transform is a rigid 3D transform that aligns the point with the camera reference frame

$$\mathbf{x}_c = \mathcal{T}(\mathbf{x}) = \mathbf{R}\mathbf{x} + \mathbf{t} \tag{1}$$

where the rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$ are the extrinsic parameters of the camera, *i.e.* its rigid pose. The intrinsic function converts the point from metric to pixel units

$$\mathbf{p}_n = \mathcal{K}(\mathbf{x}_c) = \check{\nu}(\mathbf{K}\mathbf{x}_c) \tag{2}$$

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

.

Finally, the distorted point is obtained by applying

$$\mathbf{p} = \mathcal{D}(\mathbf{p}_n) = (\mathbf{p}_n - \mathbf{p}_0)(1 + r^2 d_0 + r^4 d_1) + \mathbf{p}_0 \tag{4}$$

where $\mathbf{p}_0 = [u_0, v_0]^\top$ is the principal point and $r = \|\mathbf{p}_n - \mathbf{p}_0\|$. The vector $\mathbf{d} = [d_0, d_1]^\top$ contains the distortion coefficients.

The camera model contains 6 degrees of freedom (DoF) for the extrinsic parameters (three for rotation and three for translation) and 6 DoF for the intrinsic parameters (two focal lengths, two for the principal point, and two distortion coefficients).

4. Projective reconstruction

Planar geometry is well understood and extensively documented. Here we provide a brief review of what is relevant to this paper. Further details can be found in [7]. We first approach planar geometry using a pinhole camera with no distortion. Section 4.1 addresses the effects of distortion.

A planar scene induces a homography between two pinhole cameras. That is, the measurements of image i are related to those of image j by

$$\hat{\mathbf{p}}_j \propto \mathbf{H}_{ij}\hat{\mathbf{p}}_i \tag{5}$$

where $\mathbf{H}_{ij}$ is a 3×3 full-rank matrix. The equation is only up to scale because all elements are in homogeneous coordinates and thus $\mathbf{H}_{ij}$ has only 8 DoF.

The first step in our algorithm is to obtain a projective reconstruction of the scene. The reconstruction includes the position of the observed points in world coordinates and the poses of the cameras. The points can be represented with a 2D position vector $\mathbf{y} = [x, y]^\top$ because the scene is planar. The pose of a camera i can be represented by a homography $\mathbf{H}_i$ that translates the points from world to image space.

A projective reconstruction is defined up to a homography. Thus, without loss of generality we select one of the image frames as the reference frame, say image 0, which has the same coordinate frame as the world, *i.e.* $\mathbf{H}_0 = \mathbf{I}_{3\times 3}$. This also implicitly fixes the position of the world points $\mathbf{y} = \mathbf{p}_0$. The poses of the other frames can be determined independently by computing the homography between them and the reference frame.

4.1. Distortion

The relation of Eq. (5) only holds for a pinhole camera without distortion. In the case of distortion, the undistorted measurements will still follow Eq. (5). We can thus estimate the distortion model by finding the coefficients that allow the measurements to be modelled by a homography. We use the distortion model from Eq. (4) to remove the distortion in image space.

Although there are methods of estimating the distortion parameters from a set of uncalibrated images, for simplicity we rely here on robust homography estimation methods [7] to obtain an initial projective reconstruction assuming no distortion. The distortion coefficients are then estimated during a projective bundle adjustment step. This has proven to work well with moderate distortion levels.

4.2. Projective bundle adjustment

The projective reconstruction obtained so far is not optimal for two reasons. First, the homographies have all been computed in a pair-wise manner between the reference image and the all others. This ignores the fact that the other images might constrain each other's projective pose as well. In fact, as the camera moves farther away from the reference pose, matching becomes harder and nearby images might constrain the pose better. Second, distortion has not been estimated yet. Even under mild distortion, this can add a considerable noise to homographies that are estimated from distorted pixel correspondences.

Although the final metric bundle adjustment might recover from these intermediate errors, there is no guarantee. Thus, we perform a robust non-linear minimization [1] over all parameters to obtain the optimal projective reconstruction. The formulation of the minimization problem is as follows,

$$\operatorname*{arg\,min}_{\mathbf{d}',\mathbf{p}_0,\{\mathbf{H}_i\},\{\mathbf{y}_k\}} \sum_i \sum_k \rho(\|\mathbf{p}_{ik} - \mathcal{D}(\mathbf{H}_i\mathbf{y}_k)\|)^2 + \lambda\|\mathbf{p}_0 - \bar{\mathbf{p}}_0\|^2 \tag{6}$$

where $\rho(\cdot)$ is a robust function to reduce the influence of outliers (*e.g.* the Cauchy loss function [1]). The first term transforms the projective positions of a point $\mathbf{y}_k$ to the frame of camera i through its homography $\mathbf{H}_i$, distorts it, and compares it with the measured position $\mathbf{p}_{ik}$. The homography of the reference camera is kept fixed to remove the projective ambiguity. However, all point coordinates $\mathbf{y}_j$ are optimized to remove any bias towards the reference camera. The final term regularizes the center of distortion, biasing it towards the center of the image $\bar{\mathbf{p}}_0$ in case of no distortion. The weighting term λ is set to the inverse of the expected variance of the principal point. However, this regularization is only a safety measure and is negligible with even minor distortions.

5. Homography-based self-calibration

We first derive the constraints for homography-based self-calibration. We take a different route than [2] for clarity and to gain insight into the nature of the constraints. Yet, the constraints obtained in this section are equivalent to those of [2], as is shown in Section 5.1. In Section 5.3 we show that these constraints are biased and propose a new set of normalized constraints. We then derive a closed-form solution to obtain an initial guess for the focal length in Section 5.2.

The input to a homography-based self-calibration stage are a series of homographies $\{\mathbf{H}_i\}$ that relate all images to a reference image. The calibration recovers the intrinsic parameters f_x,f_y, and $\mathbf{p}_0$. It assumes no distortion since the distortion has been corrected beforehand during the projective reconstruction stage.

The calibration constraints come from our knowledge of the metric structure of the world applied to the obtained projective reconstruction. To derive the constraints we arbitrarily choose our metric world coordinate frame to coincide with the camera frame of the reference image. We describe the scene plane in this reference frame by its 3D normal vector $\mathbf{n}_0$ (with unit norm) and the distance to the origin.

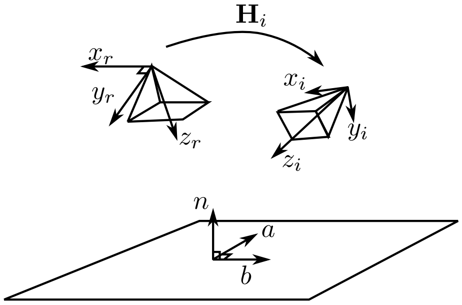


Figure 2: Planar self-calibration geometry.

We encode the Euclidean structure of the plane with two orthogonal basis vectors with equal norm that span the plane: $\mathbf{a}_0$ and $\mathbf{b}_0$. Like the normal vector, these vectors represent directions in 3D space, see Fig. 2. Note that these vectors are not unique since we can rotate them around $\mathbf{n}_0$ and they are still orthonormal basis vectors. With this in mind, the following equations encode the orthogonality and equal norm constraints respectively

$$\mathbf{a}_i^\top \mathbf{b}_i = 0, \tag{7}$$
$$\mathbf{a}_i^\top \mathbf{a}_i - \mathbf{b}_i^\top \mathbf{b}_i = 0, \tag{8}$$

where the index i has been kept variable because these constraints apply to all cameras.

By definition, we can derive these vectors for the reference camera. To ensure orthogonality, we select $\mathbf{a}_0$ as the cross product of $\mathbf{n}_0$ with an auxiliary fixed vector $\mathbf{e}$

$$\mathbf{a}_0 = \mathbf{n}_0 \times \mathbf{e} = [\mathbf{n}_0]_\times \mathbf{e} \tag{9}$$
$$\mathbf{b}_0 = \mathbf{n}_0 \times \mathbf{a}_0 = [\mathbf{n}_0]_\times^2 \mathbf{e} \tag{10}$$

Note that $\mathbf{e}$ can be any vector as long as it is not parallel to $\mathbf{n}_0$. A different $\mathbf{e}$ will produce a different orientation for the basis but Eqs. (7) and (8) will remain unchanged.

Given the basis vectors in the reference camera they can be transformed to the frame of camera i through its homography

$$\mathbf{a}_i = \mathbf{K}^{-1}\mathbf{H}_i\mathbf{K}\mathbf{a}_0, \tag{11}$$
$$\mathbf{b}_i = \mathbf{K}^{-1}\mathbf{H}_i\mathbf{K}\mathbf{b}_0. \tag{12}$$

Note that Eqs. (7) and (8) for the reference camera are satisfied by definition but the other cameras constrain $\mathbf{K}$ using the reconstructed homographies. These are the self-calibration constraints that will allow us to recover $\mathbf{K}$.

5.1. Connection with previous work

Although our derivation takes a different route, the constraints obtained so far are identical to those proposed in [2]. The components of the *circular points* in [2] $\mathbf{x}_1$ and $\mathbf{x}_2$ are simply the images of the basis vectors (*i.e.* $\mathbf{x}_1 = \mathbf{K}\mathbf{a}_0$ and $\mathbf{x}_2 = \mathbf{K}\mathbf{b}_0$). By using the notation $\omega = \mathbf{K}^{-\top}\mathbf{K}^{-1}$, Eqs. (7) and (8) can be expanded to be expressed as in [2]

$$\begin{aligned} \mathbf{a}_i^\top \mathbf{b}_i &= (\mathbf{K}^{-1}\mathbf{H}_i\mathbf{K}\mathbf{a}_0)^\top (\mathbf{K}^{-1}\mathbf{H}_i\mathbf{K}\mathbf{b}_0) \\ &= \mathbf{x}_1^\top \mathbf{H}_i^\top \boldsymbol{\omega} \mathbf{H}_i \mathbf{x}_2 = 0 \end{aligned} \tag{13}$$

$$\begin{aligned} \mathbf{a}_i^\top \mathbf{a}_i - \mathbf{b}_i^\top \mathbf{b}_i &= (\mathbf{K}^{-1}\mathbf{H}_i\mathbf{K}\mathbf{a}_0)^\top (\mathbf{K}^{-1}\mathbf{H}_i\mathbf{K}\mathbf{a}_0) \\ &\quad - (\mathbf{K}^{-1}\mathbf{H}_i\mathbf{K}\mathbf{b}_0)^\top (\mathbf{K}^{-1}\mathbf{H}_i\mathbf{K}\mathbf{b}_0) \\ &= \mathbf{x}_1^\top \mathbf{H}_i^\top \boldsymbol{\omega} \mathbf{H}_i \mathbf{x}_1 - \mathbf{x}_2^\top \mathbf{H}_i^\top \boldsymbol{\omega} \mathbf{H}_i \mathbf{x}_2 = 0 \end{aligned} \tag{14}$$

which are the same constraints and with the same parametrization as in Eq. (4) of [2]. The resulting unknowns are $\mathbf{K}$ and $\mathbf{n}_0$. The normal $\mathbf{n}_0$ has 3 variables with 2 DoF. The intrinsic matrix $\mathbf{K}$ varies between 5 DoF and 1 DoF depending on the camera model. Vector $\mathbf{e}$ is not an unknown of the problem. It can be chosen arbitrarily and fixed as long as it is not parallel to the normal vector.

5.2. Closed-form with a known plane normal

The normalized constraints can be readily used in an iterative non-linear minimization procedure to find the optimum values for the intrinsic parameters and the normal vector. However, we need an initial guess that we'd like to obtain in closed-form. A closed-form solution can be obtained if we assume a known plane normal and a simplified camera model. Gurdjos and Sturm [6] derived a closed-form solution using their formulation. Here we show that our formulation so far is equivalent and reaches the same closed-form solution.

We reduce the intrinsic matrix to have 1 DoF, *i.e.* $\mathbf{K} = \mathrm{diag}(f, f, 1)$, and assume a reference image fronto-parallel to the plane, *i.e.* $\mathbf{n}_0 = [0, 0, 1]^\top$. This seems like a strong assumption, however it is only used to obtain an initial guess of the focal length. The constraint is dropped during the optimization step and the real normal is found. Moreover, different initial guesses for the focal length can be obtained very cheaply by repeating this procedure with a different input image selected as the reference frame. The best initial guess from these can then be used for further optimization. The experiments presented in Section 7 demonstrate the robustness of the algorithm to violations of this assumption.

Under these assumptions the constraints (7) and (8) reduce to a pair of quadratic polynomials in one variable

$$h_{31}\, h_{32}\, f^2 + h_{11}\, h_{12} + h_{21}\, h_{22} = 0 \tag{15}$$

$$f^2\, {h_{31}}^2 - f^2\, {h_{32}}^2 + {h_{11}}^2 + {h_{21}}^2 - {h_{12}}^2 - {h_{22}}^2 = 0 \tag{16}$$

Note that since $\mathbf{n}_0$ is known the choice of $\mathbf{e}$ vanishes. Choosing either $\mathbf{e} = [1, 0, 0]^\top$ or $\mathbf{e} = [0, 1, 0]^\top$ result in the same equations. However, choosing $\mathbf{e} = [0, 0, 1]^\top$ (*i.e.* $\mathbf{e} = \mathbf{n}_0$) results in the trivial constraint $0 = 0$. Each homography imposes two constraints on the focal length. Since the focal length only appears in quadratic form Eqs. (15) and (16) simplify to linear constraints that can the be solved directly as an overdetermined linear system to obtain f. In case of outliers, a RANSAC approach can be used here to discard homographies that were poorly estimated.

5.3. Scale and normalization

In the absence of noise, the self-calibration constraints are homogeneous and their scale does not matter. However, with real data the constraints will have non-zero residuals and their scale affects the solution obtained. For example, Eq. (7) can also be expressed as

$$\mathbf{a}_i^\top \mathbf{b}_i = \|\mathbf{a}_i\| \|\mathbf{b}_i\| \cos\theta_{a_i b_i} \tag{17}$$

where $\theta_{a_i b_i}$ is the angle between the vectors. We see that the scale of the residuals is directly proportional to the norm of the vectors. However, this constraint should only depend on the angle because it is an orthogonality constraint. Similarly, Eq. (8) should only constrain the norms of the vectors to be equal regardless of their scale. The constraints from both [2] and [6] suffer from this problem.

There are two factors that affect the norm of the basis vectors: the homography's scale and the angle between $\mathbf{n}_0$ and $\mathbf{e}$, *i.e.*

$$\|\mathbf{a}_i\| \propto \|\mathbf{H}_i\| \sin\theta_{\mathbf{n}_0 \mathbf{e}} \tag{18}$$

Both of these bias the calibration process if not accounted for. A homography is a homogenous quantity and can have any arbitrary scale. For example, if it is obtained through linear means, it is usual to set $h_{33} = 1$ and thus fix the scale. However, if it is obtained through SVD it is common to set $||\mathbf{H}||_F = 1$. Even though they are equivalent, these two normalizations produce very different results when using constraints (7) and (8), because they give a different and unequal weight to the residuals.

In a similar fashion, the angle between $\mathbf{n}_0$ and $\mathbf{e}$ will also bias the solution. Even though $\mathbf{e}$ can be fixed, the normal of the plane will vary during the optimization. A larger angle will lead to a smaller norm for the basis vectors, thus reducing the cost of the solution and biasing the optimization towards larger angles between $\mathbf{n}_0$ and $\mathbf{e}$. In the extreme, when the vectors are perpendicular, all residuals vanish. Even if $\mathbf{e}$ is adjusted after each iteration to avoid vanishing residuals, it is not possible to normalize the homographies to give an equal norm to all basis vectors. To remove this bias and improve the obtained calibration we propose a pair of normalized constraints that are scale independent

$$\cos\theta_{a_i b_i} = \frac{\mathbf{a}_i \cdot \mathbf{b}_i}{\|\mathbf{a}_i\| \|\mathbf{b}_i\|} = 0 \tag{19}$$

$$1 - \frac{\|\mathbf{b}_i\|^2}{\|\mathbf{a}_i\|^2} = 0 \tag{20}$$

These constraints are shown in Section 7 to produce a more stable and accurate calibration than the original constraints of [2] and [6].

Normalizing the homographies with their Frobenious norm is recommended to avoid possible singularities [7]. We still tried both normalizations for the homographies and our normalized constraints consistently produced a better calibration in both cases.

5.4. Non-linear optimization

The closed-form solution provides an initial guess under the fixed-normal assumption. Although rough, this is a suitable starting point for a non-linear optimization using the constraints (19) and (20). The formulation of the minimization problem is as follows,

$$\arg\min_{\mathbf{K},\mathbf{n}_0} \sum_i \frac{\mathbf{a}_i \cdot \mathbf{b}_i}{\|\mathbf{a}_i\|\|\mathbf{b}_i\|} + \sum_i 1 - \frac{\|\mathbf{b}_i\|^2}{\|\mathbf{a}_i\|^2} \tag{21}$$

where the normal vector is constrained to have unit norm and $\mathbf{K}$ is allowed to have 4 DoF as in (3). Although this non-linear minimization is necessary to improve the initial guess, it is still not optimal because the camera poses are encoded into the homographies and cannot be optimized. The optimal solution will be obtained by optimizing in metric space.

6. Metric reconstruction

Once the intrinsic parameters have been recovered, upgrading the projective reconstruction to a metric reconstruction is straightforward. We define a new world reference frame so that the scene's plane lies at $z = 0$. Under this reference frame, the extrinsic parameters of the reference camera are initialized to $\mathbf{R} = [\mathbf{a}, \mathbf{b}, \mathbf{n}]^\top$ and $\mathbf{t} = -\mathbf{R}\mathbf{n}$, which positions the camera exactly one unit away from the plane center and aligns it with the recovered normal.

The position $\mathbf{x} = [x, y, 0]$ of the observed points can be obtained by intersecting the optical ray of the reference camera with the scene plane at $z = 0$. The extrinsic parameters of the other cameras are implicitly contained in the homographies and could be directly recovered from them using non-linear means. Alternatively, since the observed points are already triangulated and the camera is calibrated we can use well-known perspective-n-points techniques to estimate the extrinsics [9].

This produces a complete and calibrated 3D reconstruction of the scene and the cameras, including the intrinsic parameters, $\mathbf{K}$ and $\mathbf{d}$, the extrinsic parameters, $\{\mathbf{R}_i\}$ and $\{\mathbf{t}_i\}$, and the point 3D positions, $\{\mathbf{x}_k\}$.

6.1. Metric bundle adjustment

The final solution is obtained by a non-linear optimization which minimizes the reprojection errors of 3D points on image space. That is, the point coordinates are now in metric space and the quantity minimized is in pixel units. This implicitly enforces all known constraints about the scene over all measurements. The formulation of the minimization problem is as follows,

$$\arg\min_{\mathbf{K},\mathbf{d},\{\mathbf{R}_i\},\{\mathbf{t}_i\},\{\mathbf{x}_k\}} \sum_i \sum_k \rho(\|\mathbf{p}_{ik} - \mathcal{P}_i(\mathbf{x}_k))\|)^2 \tag{22}$$

where the position of the points is constrained so that $z = 0$.

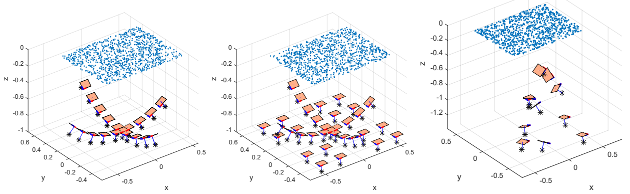
7. Experimental results

We present a series of experiments that highlight the accuracy, robustness, and practicality of our calibration approach. Most of the experiments are done with synthetically generated datasets in order to have absolute ground truth for comparison. The final experiments use real cameras with both a checkerboard and an arbitrary planar surface to validate our method against a popular calibration toolbox.

7.1. Synthetic experiments

All synthetic tests were produced using a virtual camera with images of size 640 × 480 and intrinsic parameters $f_x = f_y = 600$, $\mathbf{p}_0 = [320, 240]^\top$, and $\mathbf{d} = [0.1, -0.01]^\top$. We use three different position set-ups for the virtual cameras, shown in Fig. 3. Scene `Fixed-A` represents a common pattern by users capturing a plane from different angles. Although common, this motion is close to degenerate as described in [12] and our experiments show that it should be avoided. Scene `Fixed-B` augments this camera arrangement with a series of cameras that are translated on the XY plane, thus better constraining the problem. These scenes have a fixed number of 35 cameras. Half of the cameras for each scene were rotated by 90° to stabilize and constrain the calibration [10]. To test the influence of the number of images used for calibration we use scene `Random` which samples cameras uniformly in space and points them at a random point in the plane. In all scenes the reference camera was positioned at a fixed central location and the angle between the optical axis and the plane normal was chosen according to the experiment. For all scenes, a uniformly distributed set of 1000 features was generated on the plane and projected onto the images, keeping only those measurements within the image bounds. Gaussian noise was added to the image point measurements with a σ depending on the experiment. Each experiment was repeated 300 times to obtain valid statistical results. All errors are calculated as the root-mean-squared value from these iterations and they are reported as a percentage of the ground truth value.

Figure 4 shows the robustness of the algorithm to violations of the assumption used to derive the closed-form solution. The reference camera was randomly sampled to have an increasing angle with the scene's plane. We observe that



(a) Scene Fixed-A (b) Scene Fixed-B (c) Scene Random

Figure 3: Synthetically generated scenes used for testing. For all scenes $f_x = f_y = 600$, $u_0 = 320$, $v_0 = 240$, $d_0 = 0.1$, and $d_1 = -0.01$.

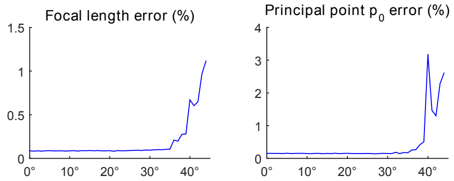


Figure 4: Robustness of the calibration system to non-perpendicular reference cameras. The calibration is robust to angles of up to 35°.

the closed-form focal length estimation was able to produce a good enough initial guess for the system to find the correct calibration with angles of up to 35°. Scene `Fixed-B` was used with noise of $\sigma = 1$px.

Figure 5 presents a comparison of our normalized constraints with the formulation from [2]. Scene `Fixed-B` with noise of variable σ was used. To show a direct comparison the metric reconstruction stage was disabled, Fig. 5 shows the results after the non-linear minimization of Eq. (21). The angle between the reference camera and the plane was set to 15°. The results show that the normalized equations are more accurate and more robust to noise. These results correspond to those obtained by Gurdjos and Sturm [6] with roughly 5% error in the focal length. However, they only tested with small amounts of noise.

Figure 6 evaluates the accuracy of the calibration in the presence of noise. We generate a synthetic scene with noise of increasing σ. We calibrate using our system (blue solid line for `Fixed-B` and dotted line for `Fixed-A`). As a comparison, we also show two cases for scene `Fixed-B` where the final metic BA is initialized with the ground truth. In the first (green) the points are allowed to vary along the plane, in the second (black) the points are kept fixed at the ground truth to simulate an ideal known-target calibration.

Figure 6 shows the clear difference in accuracy between scenes `Fixed-A` and `Fixed-B`. The translated images help constrain the distortion coefficients and result in a more accurate calibration. We observe that the green line is very close to our results, which means that our calibration is very close the theoretical optimum using planar self-calibration constraints.

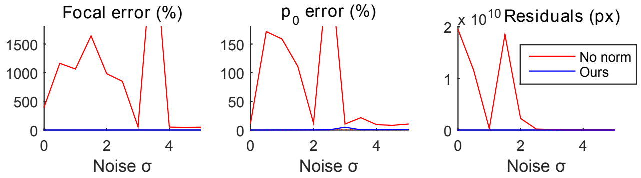


Figure 5: Comparison of our normalized constraints (blue, coincides with horizontal axis) and those of Bocquillon [2] (red) using $||\mathbf{H}||_F = 1$. The calibration fails without the normalized constraints, resulting in errors several of magnitude larger. The results shown are after the metric BA, before the BA the results were even worse for [2]. Note that this is the same experiment as in Fig. 6 but with a different scale.

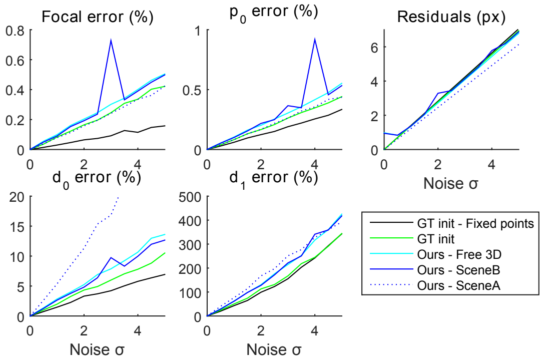


Figure 6: Comparison of our calibration accuracy with different scenes and increasing noise (blue and red). Results for a final BA initialized from ground truth using fixed (black) and variable (green) scene structure also shown. Our result is only marginally off from the theoretical optimum for self-calibration. Note that this is the same experiment as in Fig. 5 but with a different scale.

Self-calibration is more susceptible to noise than a known-target calibration. This is to be expected because it has more degrees of freedom. However, we can correct this by using more images. Fig. 7 shows the behaviour of the algorithm with a varying number of images used for calibration. We see that the accuracy of the self-calibration converges to that of the known-target calibration when more images are used. For modern cameras it is trivial to capture more images for calibration. In fact, we routinely use a video of the camera moving around looking at a book cover for calibration.

7.2. Real cameras

We show the results of calibrating real cameras using the popular Bouguet toolbox [4] and our method. Bouguet's calibration uses planar checkerboards of known structure

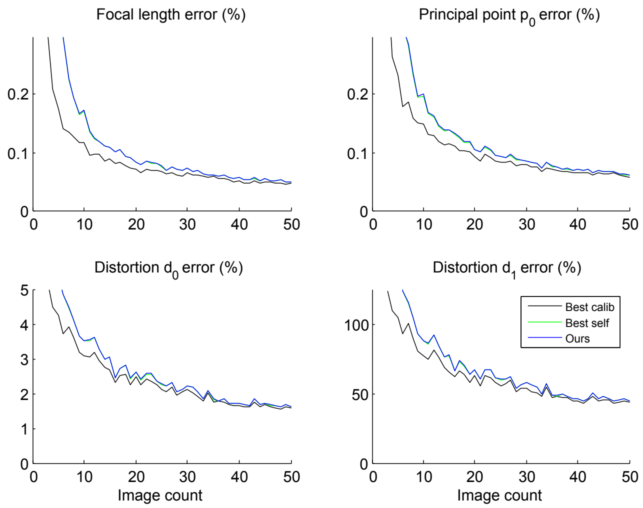


Figure 7: Calibration accuracy with increasing number of images (scene `Random`, $\sigma = 1$). The self-calibration converges with the known-target calibration.

to perform calibration. We use 30 images for calibration and 15 for validation. We use the same hand-picked corner matches for our method without using the known structure to perform a direct comparison. This is shown in Table 1 as *user match*. Additionally, we perform calibration using a video of a flat Lego box with unknown texture structure. ORB features [11] were automatically matched between frames to provide the necessary correspondences. The results of this video calibration are called *automatic match*. The reprojection error of the calibration images is shown as e_{train}. The pose of the validation images is optimized and their reprojection error is listed as e_{val}.

Our method provides an equivalent calibration to that of [4]. Using the same matches the calibrations are practically identical and the reprojection errors are comparable. Our training reprojection error is marginally lower due to the extra degrees of freedom and the validation error is marginally higher. The obtained accuracy is in line with the synthetic results, showing a 0.15% focal length error. Moreover, although it is a good reference, the calibration from [4] is not a true ground truth and is also noisy.

For the video sequence, the distortion parameters are considerably underestimated which results in a higher reprojection error. This is due to inaccurate matching and outliers, as indicated by the very high calibration reprojection error. Yet, the reprojection error is still below 1px and the obtained accuracy is better than that reported by [6] which demonstrates the robustness of the calibration approach. The feature matching stage can be improved but it is out of the scope of this paper.

Finally, to showcase the flexibility and applicability of our method we present a simple augmented reality application. A virtual cube is overlaid on top of the scene plane and rigidly attached to it. This is best viewed as a video sequence in the supplemental material . Plane-based augmented reality is a popular application of computer vision due to its robustness and accuracy. In our case, no knowledge about the plane is needed. The camera can be self-calibrated on the fly and the plane can be augmented with any virtual scene.

Table 1: Comparison of obtained calibrations using real cameras. Bouguet's values are given as a reference and our calibration is expressed as a relative deviation percentage from it. We achieve practically the same calibration and reprojection errors below 1px on validation.

	Bouguet [4]	Our method (%)	
		user match	automatic match
f_x	1384.77	-0.05	0.69
f_x	1384.92	-0.15	0.93
u_0	953.78	-0.27	2.00
v_0	528.42	-0.41	-0.37
d_0	0.094	4.70	-37.63
d_1	-0.158	2.10	-14.45
e_{train}	0.25px	0.23px	1.94px
e_{val}	0.26px	0.28px	0.84px

8. Conclusions

We have presented a planar self-calibration system that rivals the state-of-the-art calibration algorithms in accuracy and is considerably more practical to use. Our novel derivation of the planar self-calibration constraint shows that the previous formulations are biased. We proposed a set of normalized self-calibration constraints that eliminates this bias and is more robust to noise due to proper normalization. We demonstrated that the assumption used to obtain the closed-form solution to estimate the focal length is not very strict and for most practical purposes poses no limitations on the system. The system proved to be very robust to violations of this assumption, obtaining correct calibrations with angles of up to 35° between the reference camera and the scene plane. Our calibration system improves the accuracy of calibration over a purely homography-based self-calibration by over an order of magnitude.

We showed that our system has such a high accuracy that with enough input images it reaches the same accuracy as a known-target calibration, thus eliminating the practical need of printing checkerboards for camera calibration. As a possible application, our system enables plane-based augmented reality without the need for any prior knowledge about the plane or camera. Finally, we release the code of our calibration system so that the computer vision community can easily adopt this calibration method.

References

[1] S. Agarwal, K. Mierle, and Others. Ceres solver. http://ceres-solver.org. 4

[2] B. Bocquillon, P. Gurdjos, and A. Crouzil. Towards a guaranteed solution to plane-based self-calibration. In *ACCV*, volume 3851 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2006. 2, 4, 5, 6, 7

[3] S. Bougnoux. From projective to euclidean space under any practical situation, a criticism of self-calibration. In *Computer Vision, 1998. Sixth International Conference on*, pages 790–796. IEEE, 1998. 2

[4] J.-Y. Bouguet. Camera calibration toolbox for Matlab. http://www.vision.caltech.edu/bouguetj/calib_doc. 1, 2, 7, 8

[5] G. Bradski. The OpenCV library. *Dr. Dobb's Journal of Software Tools*, 2000. 2

[6] P. Gurdjos and P. Sturm. Methods and geometry for plane-based self-calibration. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I–491. IEEE, 2003. 2, 5, 6, 7, 8

[7] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2003. 1, 2, 3, 6

[8] J. Kannala, J. Heikkil, and S. Brandt. Geometric camera calibration. *Wiley Encyclopedia of Computer Science and Engineering*, 2008. 1

[9] L. Quan and Z. Lan. Linear n-point camera pose determination. *PAMI*, 21:774–780, 1999. 6

[10] F. Remondino and C. Fraser. Digital camera calibration methods: considerations and comparisons. In *ISPRS*, volume 36, pages 266–272, 2006. 6

[11] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. ORB: an efcient alternative to SIFT or SURF. In *ICCV*, 2011. 8

[12] P. Sturm. Critical motion sequences for monocular self-calibration and uncalibrated Euclidean reconstruction. In *CVPR*, pages 1100–1105, 1997. 6

[13] P. Sturm and S. Maybank. On plane-based camera calibration: A general algorithm, singularities, applications. In *CVPR*, 1999. 1

[14] B. Triggs. Autocalibration from planar scenes. In *Computer VisionECCV'98*, pages 89–105. Springer, 1998. 1, 2

[15] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *IEEE International Computer Vision Conference (ICCV)*, pages 666–673, 1999. 1, 2