2008 IEEE International Conference on
Robotics and Automation
Pasadena, CA, USA, May 19-23, 2008

More Accurate Camera and Hand-Eye Calibrations with Unknown Grid Pattern Dimensions

Klaus H. Strobl and Gerd Hirzinger
Institute of Robotics and Mechatronics
German Aerospace Center (DLR)
D-82230 Wessling, Germany
Klaus.Strobl@dlr.de

Abstract— **This paper presents two novel approaches for accurate intrinsic and extrinsic camera calibration. The rationale behind them is the widespread violation of the traditional assumption that the metric structure of the calibration object is perfectly known. A novel formulation parameterizes a checkerboard calibration pattern in such a way that the calibration performs optimally irrespective of its actual dimensions. Simulations and experiments show that it is very rare for traditional calibration methods to come by the accuracy readily attained by this approach.**

I. Introduction

Camera calibration is the process of estimating the parameters of a camera model that is capable of adequately reflecting the operation of the actual camera at hand. This is usually accomplished by *comparing its expected model-based operation with the actual one*, followed by a sensible minimization of the resulting discrepancies. The parameterized model will enable the user to infer in 3D Euclidean space from the evidence of the 2D information in the image projections.

In this work we proceed on the assumption that the camera cannot intrinsically change during operation, but only extrinsically in its pose (i.e. position and orientation) with respect to (w.r.t.) the scene. It is therefore possible for the user to *estimate the parameters in advance* of regular operation. Moreover, since cameras only passively record the environment, it makes sense to set conditions on the scene during calibration in order to support both robustness and accuracy. Frequently, these conditions concern some *a priori knowledge of the metric structure of the scene* and are aimed at maximizing both the amount of data for calibration as well as the quality of the predicted operation during calibration:

- The more *diverse* the scene is (e.g. a general 3D scene), the more independent evidence for the calibration will be.
- The more *accurate* knowledge of the scene exists, the more accurate predictions of the camera operation will be.

Unfortunately, these points imply a trade-off since advantageous conditions (e.g. general precisely known scenes) suggest elaborated and expensive calibration setups, which are cumbersome for general computer vision applications. On the other hand, less advantageous but convenient conditions, such as planarity or point correspondences, are not sufficient for accurate camera calibration; these approaches are called self-calibration and are usually less reliable, e.g. [1].

But for all that, the authors consider that research on camera calibration for computer vision applications has arrived at a point where *most* of its components have become standard

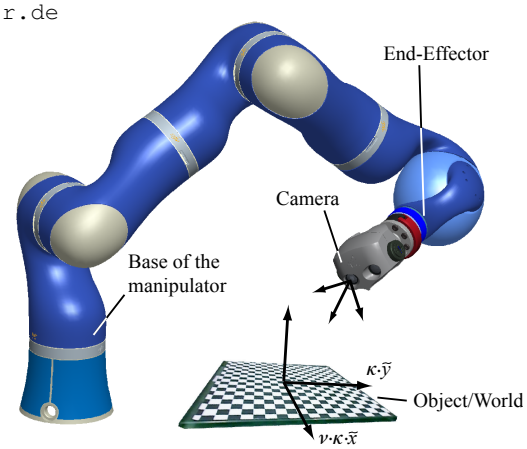


Fig. 1. Stereo camera mounted at the top of the DLR Light-Weight Robot 3.

practice: The perspective projection model (pinhole camera model) [2], the radial and tangential lens distortion models [3], the imaging noise assumption [4], the feature detection algorithms [5], the planar calibration object [6], and even the estimation algorithms [7], [8] meet the demands of the computer vision community. Next we summarize the most significant contributions that led us to this point.

Until the mid eigthies there only was photogrammetric work, which mainly relied on full-scale nonlinear optimizations for very elaborate models of imaging and calibration objects – refer to the work of Brown and Faig. This was not suitable for computer vision applications since firstly, their hardware requirements (computational among others) were too high and secondly, the complexity of their camera models exceeded the required for solid-state imaging devices. The work of Abdel-Aziz and Karara paved the way for computer vision applications; their *direct linear transformation* simply consists in finding solutions to linear equations based on the basic camera model of collinearity. However, since ignoring lens distortion is mostly unacceptable, Tsai in the mid eigthies used a more complete camera model [6]. He was able to simplify the formulation by using the *radial alignment constraint*, which reduces the dimensionality of the problem and allows its decomposition in two independent stages – at the risk of losing valuable information in the radial component when the distortion is small. There still apply severe scene restrictions: the procedure requires either 3D calibration objects or accurately moving a planar calibration plate in its perpendicular direction (2.5D). Both the works of Weng et al. [3] and Faugeras and Toscani differ in algorithms, but are subject to similar requirements. The former used a

very extensive lens distortion model, which was the reason for them to strongly rely on an iterative coupling of local nonlinear optimizations.

A major contribution towards simplicity in camera calibration was simultaneously made in the late nineties by Zhang [7] and Sturm and Maybank [8]. The suitability of their algorithm in computer vision applications made both their algorithmics and the used models standard practice. They presented a closed-form solution by linear least-squares techniques for the initialization of a nonlinear optimization. Most importantly, they relax conditions on the scene allowing for *freely* moving a *precisely known* planar calibration pattern for collecting data. In short, it recovers the intrinsic camera parameters from the readily obtained object-to-camera homographies using the camera model as well as rigid body motion constraints. The approach represented a step towards self-calibration since no purely implicit 3D information was required anymore, but only 2D. Malm and Heyden [9] extended this formulation for the case of stereo camera systems by the addition of further rigid body motion constraints.

Moreover, it is frequently the case that the camera system is rigidly attached to the end-effector of a robot manipulator (eye-in-hand system), and its pose w.r.t. the latter (hand-eye transformation) is required. For literature on this kind of extrinsic calibration (hand-eye calibration) see Ref. [10].

As regards the intrinsic calibration, historically there was a steady reduction in complexity of the calibration objects: Faugeras (3D calibration object) $\rightsquigarrow$ Tsai (2.5D) $\rightsquigarrow$ Zhang, Sturm, and Maybank (2D). It was mostly motivated by the fact that *this deskills the calibration procedure*, which of course is a valuable asset for computer scientists most impatient to concentrate all efforts on the end applications. In this work the authors agree on this trend, but do not completely agree on its motivation. We instead argue that, since the principal purpose of calibration is achieving highest accuracy in the estimation of the parameters, the best substantiation of significance of a calibration approach is necessarily the *improvement of accuracy* – if it ever exists. That is because eventual erroneous operation of the end applications originates in part in erroneous calibration (along with noisy operation or model simplifications), but not in any way in the simplicity or the speed of the calibration process. Erroneous calibration in turn mainly stems from deviations of the models from the actual systems (the topic of this paper), and secondarily from non-optimal definition of the residuals (the topic of Ref. [10]).

In this spirit we still advocate procedures concerning a relaxation of diversity conditions on the scene, since this also tends to increase estimation accuracy. This is *not* because extensive knowledge compromises accuracy but because lighter conditions prevent damage to the calibration process from human inaccuracies and mistakes (e.g. while measuring the calibration object), being a by-product of lighter conditions that they also deskill the procedure. Several works (e.g. recently Ref. [4]) give experimental evidence on this point. On top of that the authors consider it necessary to take the matter further, relaxing requirements on the knowledge of the scene and thus narrowing the gap on the self-calibration approaches while still performing accurate ground truth based camera calibration.

Planar calibration objects do indeed provide convenient ground truth for camera calibration for different reasons: Firstly, they are easy to manufacture, use, and store; secondly, they are naturally well adapted to the calibration of lens distortion since they can easily occupy whole images; thirdly and most importantly, high geometrical accuracy can be (cheaply) achieved. However, that is unfortunately not the case for the 2D pattern imprinted on it. Regular printers dramatically lack of accuracy and it is therefore standard practice to gauge it by hand, which is in turn prone to errors because of the use of inaccurate or inappropriate rulers, or even the indolent commitment of the user.

In this work we aim at a *calibration method that relaxes restrictions on accurate knowledge of the imprinted pattern dimensions*, while still taking advantage of a priori knowledge of both its planarity and the regularity of the pattern. We furthermore claim that in most common systems highest accuracy camera calibration is still possible by this means. This would be a significant contribution in order to avoid commonplace mistakes and therefore increase calibration accuracy.[1]

The remainder of this article is as follows: Section II presents the state of the art of camera calibration and expounds its problems, which are best alleviated with the novel methods of Section III. Sections IV and V illustrate it by means of simulations and experiments, and conclude with a discussion.

II. Problem Description: Camera Calibration

In this section the conventional approach to camera calibration – along with its shortcomings – is presented.

Firstly, the general procedure has to be appointed. We opt for off-line (stereo) camera calibration with a priori ground truth knowledge of the scene: a known calibration object is imaged by the camera from N different vantage points.[2] The discrepancies between the expected and the actually detected projections are to be minimized to refine the intrinsic parameters. If the camera is a constituent part of an eye-in-hand system, the extrinsic hand-eye transformation can be collected at the same time if the N poses of the end-effector at the different imaging moments are also recorded.

Secondly, adequate models are assigned. The calibration object is a *planar checkerboard pattern*[3] and the 2D relative

[1] The authors of Ref. [11] bring forward a similar argument. However, they deal with 3D calibration objects and assume independent randomly distributed errors in the 3D coordinates of the control points. Thus, the estimation process lacks of any knowledge of the structure of the scene (but the required accurate initial estimation) and therefore corresponds to the photogrammetic bundle adjustment problem, which is provably worse conditioned for camera calibration than the work that will be presented here.

[2] Multiple vantage points are not always necessary for camera calibration, but they facilitate intrinsic initialization and allow for hand-eye calibration. In addition to that, the central limit theorem states that when the amount of independent and identically distributed (i.i.d.) data grows, the error distribution tends to Gaussianity, which eventually facilitates optimal estimation (especially the hand-eye [10]). In fact, at least 10 vantage points are repeatedly recommended in the literature, cf. Refs. [1], [6]–[8], [10].

[3] The authors opt for a checkerboard pattern since it was recently appointed as the most convenient one in terms of accuracy in points detection [5], [12].

positions of its control points are perfectly known – this condition is about to be lifted in part in Section III. The control points are then perspectively projected by the camera to 2D images that stack up into a discrete memory array consistently with the discretized pinhole camera model. The subsequent detection of the points is performed with sub-pixel accuracy and with errors according to a 2D i.i.d. zero-mean Gaussian distribution.[4] Other than that, in the case of an eye-in-hand system it is quite natural that the absolute pose of the end-effector is also erroneous to some extent, see Ref. [10].

Thirdly, the estimation algorithms are specified. Intrinsic and extrinsic calibrations are separatedly performed.[5]

A. Intrinsic camera calibration

Intrinsic camera calibration aims at the estimation of the intrinsic parameters of the camera model and, as a by-product, of the absolute poses of the camera in the world/object frame (here called absolute extrinsics). Next we outline the well-known approach of Refs. [7], [8].

It is critical for accurate camera calibration to start out properly on the detection and identification of the control points ${}_0\boldsymbol{x}_i = [x_i\ y_i\ z_i\ 1]^T$ perspectively projected onto the image frame, and erroneously detected in image $n \in \{1, \ldots, N\}$ as ${}_n\tilde{\boldsymbol{m}}_i$, see Ref. [12]. These are to be compared with the expected ones ${}_n\hat{\boldsymbol{m}}_i = [{}_nu_i\ {}_nv_i\ 1]^T$ which are estimated using an Euclidean decomposition of the perspective projection matrix $\boldsymbol{P}$ as follows:

$$s\,\hat{\boldsymbol{m}} = \underbrace{\boldsymbol{A}\ {}_C\boldsymbol{T}^0}_{\boldsymbol{P}_{3\times4}}\ {}_0\boldsymbol{x} = \underbrace{\begin{bmatrix}\alpha & \gamma & u_0\\ 0 & \beta & v_0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\boldsymbol{r}_1\ \boldsymbol{r}_2\ \boldsymbol{t}\end{bmatrix}}_{\boldsymbol{H}_{3\times3}}\begin{bmatrix}x\\ y\\ 1\end{bmatrix} \tag{1}$$

where s is an arbitrary scale factor, $\boldsymbol{A}$ the camera intrinsic matrix [7], and ${}_C\boldsymbol{T}^0$ the rigid body transformation from the camera frame to the object/world frame in the image n (indexes i and n have been omitted for the sake of clarity). Since $z = 0$ the formulation can be simplified to the linear projective transformation $\boldsymbol{H} = [\boldsymbol{h}_1\,\boldsymbol{h}_2\,\boldsymbol{h}_3]$, which equals the homography between the calibration plane and the image. The N homographies $\widehat{\boldsymbol{H}}_n$ can be readily estimated up to scale on the basis of both ${}_0\boldsymbol{x}_i$ and ${}_n\tilde{\boldsymbol{m}}_i$. After imposing $\widehat{\boldsymbol{H}} \propto \boldsymbol{A}\,[\boldsymbol{r}_1\,\boldsymbol{r}_2\,\boldsymbol{t}]$ and the orthonormality restrictions[6] $\boldsymbol{r}_1 \cdot \boldsymbol{r}_2 = 0$, $\boldsymbol{r}_1 \cdot \boldsymbol{r}_1 = 1$, and $\boldsymbol{r}_2 \cdot \boldsymbol{r}_2 = 1$, i.e. ${}_C\boldsymbol{R}^0 \in SO(3)$, and sorting the scale out:

$$\left.\begin{array}{l}(A^{-1}\boldsymbol{h}_1)^T \cdot (A^{-1}\boldsymbol{h}_2) = 0\\ (A^{-1}\boldsymbol{h}_1)^T \cdot (A^{-1}\boldsymbol{h}_1)\\ -(A^{-1}\boldsymbol{h}_2)^T \cdot (A^{-1}\boldsymbol{h}_2) = 0\end{array}\right\} \Leftrightarrow \left.\begin{array}{l}\boldsymbol{h}_1^T\,\boldsymbol{\omega}_\infty\,\boldsymbol{h}_2 = 0\\ \boldsymbol{h}_1^T\,\boldsymbol{\omega}_\infty\,\boldsymbol{h}_1 = \boldsymbol{h}_2^T\,\boldsymbol{\omega}_\infty\,\boldsymbol{h}_2\end{array}\right\} \tag{2}$$

with the so-called absolute conic $\boldsymbol{\omega}_\infty = \boldsymbol{A}^{-T}\boldsymbol{A}^{-1}$. These two equations hold for every N images, leading to $2N$ constraints for 5 intrinsic unknowns, which in this particular case can be solved for with a linear least-squares criterion if at least three different views are available. It is worth noting that they only depend on the orientation of the plane and *not* on its distance or scale, i.e. the formulation works both on Euclidean and similarity geometries [2].

If a stereo configuration exists (i.e. N_c additional cameras rigidly attached to the main one), it is possible to unify their absolute extrinsics, see Ref. [9]. The rigid body constraint ${}_{C_c}\boldsymbol{T}^0 = {}_{C_c}\boldsymbol{T}^{C}\,{}_{C}\boldsymbol{T}^0$ for every additional camera C_c, $c \in \{1, 2, \ldots, N_c\}$, can be easily transcribed for homographies: $\boldsymbol{h}_1 = {}^{s_c}\!/_{s}\ {}_c\boldsymbol{H}_\infty\ {}_c\boldsymbol{h}_1$ and $\boldsymbol{h}_2 = {}^{s_c}\!/_{s}\ {}_c\boldsymbol{H}_\infty\ {}_c\boldsymbol{h}_2$ for every N images and N_c additional cameras, from where the infinite homographies ${}_c\boldsymbol{H}_\infty = \boldsymbol{A}\ {}_C\boldsymbol{R}^{C_c}\,{}_c\boldsymbol{A}^{-1}$ can be estimated. These result in constraints for the intrinsic matrices as follows: $\boldsymbol{\omega}_\infty = {}_c\boldsymbol{H}_\infty^T\ {}_c\boldsymbol{\omega}_\infty\ {}_c\boldsymbol{H}_\infty$. These six linear constraints may stack up in the system of $2N$ linear equations (2) for the N_c cameras.

After determining the intrinsic matrix, the absolute extrinsics for image n are readily computed from the homographies as follows: $\boldsymbol{r}_1 = {}^1\!/_{s}\,\boldsymbol{A}^{-1}\boldsymbol{h}_1$, $\boldsymbol{r}_2 = {}^1\!/_{s}\,\boldsymbol{A}^{-1}\boldsymbol{h}_2$, $\boldsymbol{r}_3 = \boldsymbol{r}_1 \times \boldsymbol{r}_2$, $\boldsymbol{t} = {}^1\!/_{s}\,\boldsymbol{A}^{-1}\boldsymbol{h}_3$, and $s = ||\boldsymbol{A}^{-1}\boldsymbol{h}_1|| = ||\boldsymbol{A}^{-1}\boldsymbol{h}_2||$.

Unfortunately, lens distortion spoils the neat linear projective formulation of any camera.[7] Although it is possible to estimate exclusively the distortion parameters with a linear least squares criterion, this only works – to some extent – in alternation with the former intrinsic estimation [7], [14]. For accurate estimation it is still necessary to perform a subsequent nonlinear optimization process, which can be initialized with the above described estimations. This nonlinear optimization will necessarily be exclusively responsible for the final precision of the calibration if the initialization ends up within a *fairly broad convergence region* in the parameter space.

Using the maximum likelihood criterion, and since the only data considered as *erroneous* is the detected projections ${}_n\tilde{\boldsymbol{m}}_i$ of the control points in the images (viz. with i.i.d. zero-mean Gaussian error distribution), it follows the optimal parameters estimation that minimizes the sum of squared prediction errors in the projections of the control points:

$$\widehat{\boldsymbol{\Omega}}_\star = \arg\min_{\widehat{\boldsymbol{\Omega}}} \sum_{n=1}^{N}\sum_{i} ||{}_n\tilde{\boldsymbol{m}}_i - {}^d_n\hat{\boldsymbol{m}}_i(\widehat{\boldsymbol{\Omega}},\ \boldsymbol{\Upsilon}({}_0\boldsymbol{x}_i))||^2 \tag{3}$$

where ${}^d_n\hat{\boldsymbol{m}}_i$ are the distorted[8] projections of the control points ${}_0\boldsymbol{x}_i$ expected in the image frame. They depend both on the calibration parameters $\boldsymbol{\Omega}$ to be estimated (intrinsic, distortion, and absolute extrinsic parameters) and on the system models $\boldsymbol{\Upsilon}$, which include the camera and lens distortion models as well as the calibration object model (e.g. ${}_0\boldsymbol{x}_i$).

[4] It is a tacit assumption that these errors actually encompass both the pinhole camera model simplification error and its discretization error as well.

[5] This is the commonly used method. However, on occasions (e.g. with very low resolution images, high noise, or narrow field of view) it may be meaningful to estimate the intrinsic parameters and the hand-eye transformation *at the same time* in order to make full use of the extrinsic positioning accuracy.

[6] The geometrically inclined reader may prefer the interpretation concerning the constraints on the critical points of each calibration plane by the intersection of the lines at infinity of the respective planes with the absolute conic (at the plane at infinity). This projective interpretation belongs together with a complexification of the homogeneous Euclidean vector space, cf. [2], [7], [8].

[7] For centuries lenses take the place of the notional pinholes mainly in order to increase light gathering while widely maintaining sharpness. The complex path of light in a lens accounts for deviations from the straight line projection assumption. This problem cannot be completely avoided since there are multiple factors to be concurrently optimized in lens design [13].

[8] Distorted projections are mostly calculated following an undistorted-to-distorted (*U-D*) model like in Ref. [7] – optionally including decentering and thin prism distortions [3], [12]. Although in the past the distortion direction was differently applied [15], the *U-D* model is usually closer to the radial distortion reality than the opposite *D-U* one. Admittedly this statement – like many others in this area – is still subject to the equipment at hand.

As mentioned in Section I, most of these models are widely accepted in the computer vision community and it does not seem necessary to question their suitability anymore. However, it is apparent that there still exists one potential error source that has not yet been addressed, namely the allegedly known positions of the control points in the calibration object, that on the one hand support calibration accuracy, but on the other very easily feed incorrect data into the estimation. In fact, *the pattern on the calibration plate is usually inaccurately imprinted*. Off-the-shelf printers especially fail in scaling the pattern, which independently (and regularly) occurs in the two perpendicular directions – skew patterns rarely occur. In order to cope with this problem it is standard practice to carry out subsequent 2D measurements of the positions of the control points, which is difficult to perform by hand with high accuracy[9] – Figs. 2 (a), 3, and 4 in Section IV-A.1 explicitly show the negative effects of this limitation.

However, it is still possible to rescale the pattern back to the truth with only two parameters. For this purpose, in Section III *the scaling factor κ and the aspect ratio ν are introduced for the parameterization of the calibration pattern*.

B. Extrinsic camera calibration

In the context of eye-in-hand systems we define extrinsic camera calibration (or hand-eye calibration) as the estimation of the rigid body transformation ${}_t\boldsymbol{T}^C$ relating the end-effector frame of the robot manipulator (hand) to the camera frame (eye) – full details in Ref. [10]. In short, this is implemented minimizing the discrepancies ($\mathcal{O}_n$) between expected and measured transformations. Expected values stem in part from the intrinsic calibration by the absolute extrinsics ${}_C\widehat{\boldsymbol{T}}^0_\star$. Measured values correspond to the (erroneous) motion readings of the manipulator ${}_b\widetilde{\boldsymbol{T}}^t$. Following the notation of Ref. [10]:

$$\{{}_t\widehat{\boldsymbol{T}}^C_\star,\, {}_b\widehat{\boldsymbol{T}}^0_\star\} = \arg\min_{{}_t\widehat{T}^C,\, {}_b\widehat{T}^0} \sum_{n=1}^{N} \mathcal{O}_n({}_t\widehat{\boldsymbol{T}}^C,\, {}_b\widehat{\boldsymbol{T}}^0,\, {}_C\widehat{\boldsymbol{T}}^0_\star,\, {}_b\widetilde{\boldsymbol{T}}^t).$$

From this it follows that wrong absolute extrinsics inevitably lead to wrong extrinsic camera calibration. In Section III it will be shown that a correct intrinsic camera calibration can still provide erroneous absolute extrinsics depending on the scaling factor κ used. Again, Fig. 2 (b) in Section IV-A.1 explicitly shows these negative effects.

III. Two Novel Methods for the Simultaneous Determination of the Grid Dimensions

The last section suggested that slightly incorrect data on the pattern dimensions leads to biased calibration results. We proposed a parameterization for the grid pattern of the planar calibration object by two parameters only: the scaling factor κ and the aspect ratio ν. This is a convenient parameterization not only because it very closely corresponds to the actual limitations of conventional printing equipment, but also because the effects of these parameters on the calibration process can be clearly differentiated: whereas an erroneous aspect ratio $\tilde{\nu}$ does affect the estimation of the intrinsic parameters, an erroneous scaling factor $\tilde{\kappa}$ allows optimal intrinsic calibration; it only affects (in range) the absolute extrinsics of each camera. Furthermore, they are a tight object model representation that makes it still possible for the calibration process to take advantage of accurate knowledge of the scene in the form of the high planarity and regularity of the imprinted pattern.

In this section, we present two different methods for camera calibration that do the job properly irrespective of these unknown parameters; this mostly goes with the simultaneous estimation of these parameters. We furthermore claim that very often the accuracy of these estimations surpasses the accuracy that the common user is able to give gauging it by hand.

The authors build upon the planar approach of Refs. [7], [8] (intrinsic calibration) and the hand-eye calibration of Ref. [10] (extrinsic calibration). Firstly, the intrinsic parameters are roughly estimated with linear least-squares techniques. Secondly, the complete set of parameters of the camera model is refined by nonlinear optimization. Thirdly, the extrinsic parameters are also roughly estimated. Lastly, the extrinsics are refined by nonlinear optimization.

A. Initial intrinsic closed-form solution

Since the actual Euclidean coordinates of the control points ${}_0\boldsymbol{x}_i$ are no longer known, but only the erroneously scaled ones ${}_0\tilde{\boldsymbol{x}}_i$, the solution of the system of Eqs. (2) may lead strongly biased results. That is because the decomposition of the calculated homographies $\widehat{\boldsymbol{H}}$ (so that $s\,\hat{\boldsymbol{m}} = \widehat{\boldsymbol{H}}\,[\tilde{x}\;\tilde{y}\;1]^T$) has changed – cf. Eq. (1). Now:

$$\widehat{\boldsymbol{H}} \propto \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{r}_1 & \boldsymbol{r}_2 & \boldsymbol{t} \end{bmatrix} \begin{bmatrix} \kappa\nu & 0 & 0 \\ 0 & \kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{such that} \quad \left.\begin{array}{r} \boldsymbol{r}_1 \cdot \boldsymbol{r}_2 = 0 \\ \boldsymbol{r}_1 \cdot \boldsymbol{r}_1 = 1 \\ \boldsymbol{r}_2 \cdot \boldsymbol{r}_2 = 1 \end{array}\right\} \quad (4)$$

$\forall\ \nu, \kappa \in \mathbb{R} \,/\, \nu, \kappa \neq 0$, and it follows:

$$\left.\begin{array}{r} (A^{-1}\boldsymbol{h}_1)^T \cdot (A^{-1}\boldsymbol{h}_2) = 0 \\ 1/\nu^2 \cdot (A^{-1}\boldsymbol{h}_1)^T \cdot (A^{-1}\boldsymbol{h}_1) \\ -(A^{-1}\boldsymbol{h}_2)^T \cdot (A^{-1}\boldsymbol{h}_2) = 0 \end{array}\right\} \Leftrightarrow \left.\begin{array}{l} \boldsymbol{h}_1^T\, \boldsymbol{\omega}_\infty\, \boldsymbol{h}_2 = 0 \\ \boldsymbol{h}_1^T\, \boldsymbol{\omega}_\infty\, \boldsymbol{h}_1 = \nu^2 \cdot \boldsymbol{h}_2^T\, \boldsymbol{\omega}_\infty\, \boldsymbol{h}_2 \end{array}\right\}.$$

The introduction of one further unknown (the aspect ratio ν) does away with the former linear formulation in part. Although ν could be otherwise calculated (it is unique for the N images), there is no point bothering about it since intrinsic calibration is not about estimating the value of ν, but of the intrinsic parameters only. It is possible to only orthogonalize both $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ (i.e. $\boldsymbol{h}_1^T\, \boldsymbol{\omega}_\infty\, \boldsymbol{h}_2 = 0$) without normalizing them (i.e. without forcing $\boldsymbol{h}_1^T\, \boldsymbol{\omega}_\infty\, \boldsymbol{h}_1 = \nu^2 \cdot \boldsymbol{h}_2^T\, \boldsymbol{\omega}_\infty\, \boldsymbol{h}_2$) thereby resulting in N constraints for 5 intrinsic unknowns, and there still are more equations than unknowns after all. In addition, the stereo constraints in Ref. [9] still hold (unmodified) in this case. This formulation is now closer to the actual intrinsic and absolute extrinsic values than the traditional one in the ubiquitous case of erroneous knowledge of the aspect ratio ν.

However, the authors actually recommend to decrease the number of unknowns in this first estimation. It is a pointless effort to aim at success in accurately estimating very sensitive parameters, such as the skew γ or the principal point[10]

[9] Conventional rulers are accurate to say $1\,mm$ markers [4]. The reader is invited to check different rulers against each other.

[10] This is because the principal point suffers from a severe identity disorder, its key role being largely to locate the origin of the radial lens distortion [6]. In addition to that, partial imaging of the calibration object can also lead to problems in the estimation of the principal point [9].

$[u_0,\, v_0]$, prior to the estimation of the lens distortion. It is more advisable to include some prior knowledge of the parameters in the following form: $\gamma = 0$ and $[u_0,\, v_0]$ be located at the image center, solely remaining the scale factors α and β to be estimated. It is more likely that the estimations resulting from *this* method fall in the convergence region required for successful nonlinear optimization, rather than the numerous former parameters afflicted with biases. Alternatively, the iterative method in Ref. [14] could be used with the omission of the normalization constraint. Nevertheless, it is fair to say that the traditional approach does also mostly fall in that region – simulations and experiments show that most parameters can be safely initialized with an inaccurate prior without affecting the eventual accuracy after the nonlinear optimization.

After determining the intrinsic matrix, the absolute extrinsics for every image n are computed from Eq. (4) as follows: $\boldsymbol{r}_1 = {}^{\nu}/_{s\kappa}\, \boldsymbol{A}^{-1}\boldsymbol{h}_1$, $\boldsymbol{r}_2 = {}^{1}/_{s\kappa}\, \boldsymbol{A}^{-1}\boldsymbol{h}_2$, $\boldsymbol{r}_3 = \boldsymbol{r}_1 \times \boldsymbol{r}_2$, $\boldsymbol{t} = {}^{1}/_{s\kappa}\, \boldsymbol{A}^{-1}\boldsymbol{h}_3$, $s\kappa = ||\boldsymbol{A}^{-1}\boldsymbol{h}_1||$, and even $\nu = ||\boldsymbol{A}^{-1}\boldsymbol{h}_1|| / ||\boldsymbol{A}^{-1}\boldsymbol{h}_2||$! Again, these object-to-camera absolute extrinsics (and the camera-to-camera translations in the case of stereo) may be incorrectly scaled (in range), both after this initial estimation and after the following nonlinear optimization, *if* the assumed scaling factor κ is far from reality; the potential extrinsic calibration will be able to correct them, though. The intrinsic parameters remain unaffected, since they can be estimated irrespective of the latter – recall Eqs. (4).

These values finally do pave the way for an optimization of the intrinsic parameters similar to Eq. (3), with subsequent extrinsic calibration, the only difference being the additional unknown parameters aspect ratio ν and scaling factor κ.

B. Method #1: aspect ratio ν from projection errors minimization; scaling factor κ from extrinsic transformation errors minimization

Since an erroneous aspect ratio $\tilde{\nu}$ implies systematic errors between the estimated $\hat{\boldsymbol{m}}$ and the actually projected $\tilde{\boldsymbol{m}}$ control points in the image, *the Gaussian distribution of the error metric is violated which prevents (unbiased) optimal estimation of the intrinsic parameters*. In addition to that, it can be observed in the simulations in Section IV-B that *the released intrinsic parameters cannot completely compensate for these systematic errors if multiple images from different vantage points are taken*. From these observations it follows that *firstly*, only the correct value for ν truly minimizes the projection errors after nonlinear optimization (see Fig. 4), and *secondly*, the aspect ratio can be estimated *at the same time* along with the other parameters. In this way, the erroneous data is again the projections ${}_n\tilde{\boldsymbol{m}}_i$ of the control points in the images only. The following minimization provides now the optimal internal parameters:

$$\widehat{\boldsymbol{\Omega}}^s_\star = \arg\min_{\widehat{\boldsymbol{\Omega}}^s} \sum_{n=1}^{N} \sum_i ||{}_n\tilde{\boldsymbol{m}}_i - {}^d_n\hat{\boldsymbol{m}}_i(\widehat{\boldsymbol{\Omega}}^s(\hat{\nu}),\ \boldsymbol{\Upsilon}({}_0\tilde{\boldsymbol{x}}_i))||^2. \quad (5)$$

In contrast to Eq. (3) here the optimization vector of calibration parameters $\boldsymbol{\Omega}^s$ includes the aspect ratio ν that in turn, together with the erroneous object model in the vector of system models $\boldsymbol{\Upsilon}$, eventually generates ${}_0\hat{\boldsymbol{x}}_i = [\kappa\cdot\nu\cdot\tilde{x}_i \;\; \kappa\cdot\tilde{y}_i \;\; 0 \;\; 1]^T$.

If a subsequent hand-eye calibration is required, it has to be considered the above mentioned fact that the estimated absolute extrinsics ${}_C\widehat{\boldsymbol{T}}^0_\star$ may strongly differ from the actual ones. The transformation errors in the rotational and translational metrics within $SE(3)$ do not present the required unbiased Gaussian distributions anymore, and the optimal estimation process becomes strongly corrupted [10]. Therefore, the hand-eye calibration algorithm has to be modified in order to estimate the scaling factor κ in which the intrinsic calibration was actually performed. In doing so, the absolute extrinsics (and, if stereo, the camera-to-camera transformations[11] as well) have to be scaled accordingly. Since the released extrinsic parameters cannot compensate for erroneous scales/ranges in all the absolute extrinsics at the same time, the *simultaneous estimation of the hand-eye transformation and the scaling factor κ for multiple images* tends to restore the error distribution to its reputed unbiased Gaussian nature, and the calibration along with it to optimal (unbiased) operation:

$$\{{}_t\widehat{\boldsymbol{T}}^C_\star, {}_b\widehat{\boldsymbol{T}}^0_\star, \hat{\kappa}_\star\} = \arg\min_{{}_t\widehat{T}^C,\, {}_b\widehat{T}^0,\, \hat{\kappa}} \sum_{n=1}^{N} \mathcal{O}_n(\Phi({}_C\widehat{\boldsymbol{T}}^0_\star,\, \hat{\kappa}),\ {}_b\widetilde{\boldsymbol{T}}^t,\ \ldots) \quad (6)$$

where the function Φ scales the estimated absolute extrinsics ${}_C\widehat{\boldsymbol{T}}^0_\star$ in range according to the estimated scaling factor $\hat{\kappa}$.

C. Method #2: both aspect ratio ν and scaling factor κ from extrinsic transformation errors minimization

Alternatively, and only if a subsequent hand-eye calibration has to be performed, it is also possible to estimate the aspect ratio ν – and again the scaling factor κ – by minimizing the extrinsic transformation errors. As mentioned above, any intrinsic calibration with incorrect aspect ratio ν will yield erroneous parameters and thus erroneous absolute extrinsics along with them – not only in range. This will necessarily compromise the hand-eye calibration (even if it also estimates κ) and therefore once again only the correct value for the aspect ratio ν will make it to truly minimize the extrinsic residuals:

$$\{{}_t\widehat{\boldsymbol{T}}^C_\star, {}_b\widehat{\boldsymbol{T}}^0_\star, \hat{\kappa}_\star, \hat{\nu}_\star\} = \arg\min_{{}_t\widehat{T}^C,\, {}_b\widehat{T}^0,\, \hat{\kappa},\, \hat{\nu}} \sum_{n=1}^{N} \mathcal{O}_n(\Phi({}_C\widehat{\boldsymbol{T}}^0_\star,\, \hat{\kappa}),\ {}_b\widetilde{\boldsymbol{T}}^t, \ldots)$$

with ${}_C\widehat{\boldsymbol{T}}^0_\star \in \arg\min_{\widehat{\boldsymbol{\Omega}}} \sum_{n=1}^{N} \sum_i ||{}_n\tilde{\boldsymbol{m}}_i - {}^d_n\hat{\boldsymbol{m}}_i(\widehat{\boldsymbol{\Omega}},\ \boldsymbol{\Upsilon}^s(\hat{\nu}, {}_0\tilde{\boldsymbol{x}}_i))||^2$

where ν is not included in the optimization vector of the intrinsic estimation $\boldsymbol{\Omega}$ anymore, but in the new vector of system models $\boldsymbol{\Upsilon}^s$. This method is computationally more expensive since a complete optimization for calculating ${}_C\widehat{\boldsymbol{T}}^0_\star$ is taking place for every single extrinsic iteration. Its main motivation are systems where the positioning accuracy is very high, the errors in the chosen metrics in $SE(3)$ close to Gaussian, and the imaging errors are neither small nor Gaussian (e.g. with very low resolution or oddly distorted images). In this case, a feasible solution can be obtained as follows: A first solution by *Method #1* can serve as a good initialization for *Method #2*, which in this way could only consist of a very restricted local search on ν over both the traditional intrinsic optimization in Eq. (3) and the subsequent extrinsic optimization in Eq. (6).

[11] The transformation ${}_{C_c}\boldsymbol{T}^C$ can actually be considered as an intrinsic parameter of a more abstract camera system concerning stereo cameras.

IV. Simulation Results

Simulations were conducted in order to illustrate the fundamental weaknesses of the traditional calibration methods and to put the novel methods presented in the last section to the proof. Ground truth data was adopted from the intrinsic and extrinsic results of the real (monocular[12]) camera calibration of Section V (left-hand side of Fig. 6) and the assumed pattern dimensions. In doing so, the ideal image projections and robot motions were calculated, and noisy image and positioning data generated on them. Next, the effects of errors in the assumed pattern dimensions and noise levels are studied.

A. Effects of erroneous pattern dimensions

In this section the errors in the estimation of the parameters of the camera model after traditional calibrations with inaccurate knowledge of both pattern parameters ν and κ from 0.99 to 1.01 is shown – ground truth actually amounts to $\nu=\kappa=1$. Afterwards, the camera performance is assessed.

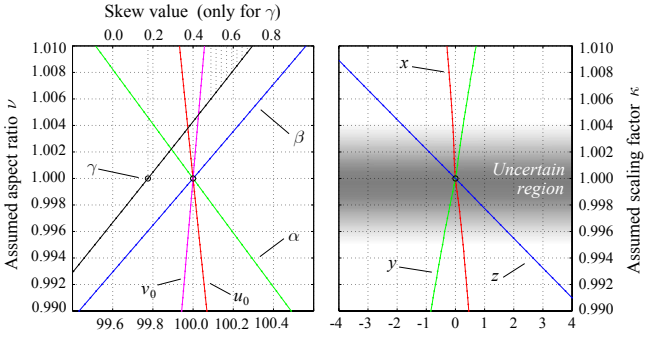
Noisy visual data was generated over the ideal image projections with $\sigma_u=\sigma_v=0.15\,p$. For the extrinsic calibration, random noisy transformations ${}_0\widetilde{\boldsymbol{T}}^t$ were generated from the N ideal absolute extrinsics and the ground truth hand-eye transformation. The noise was added to the ideal pose of the end-effector of the robot as follows: The angles θ of the angle-axis representation $\{\theta,\boldsymbol{p}\}$ of the added noise follow a zero-mean Gaussian distribution with $\sigma_\theta=0.05°$ and their axes $\boldsymbol{p}$ are uniformly distributed, i.e. their azimut and elevation angles ϕ and ψ are $\phi\in[-90°,90°)$ according to the **p**robability **d**ensity **f**unction $pdf(\phi)=180^{-1}\,[°]^{-1}$ and $\psi\in[-90°,90°)$ with $pdf(\psi)\propto\arcsin(\psi/90)\,[°]^{-1}$. The translation errors $\boldsymbol{t}$ also follow a zero-mean Gaussian distribution in range with σ_t=0.25 mm and the directions are again uniformly distributed.

1) Erroneous estimation of the parameters: Figs. 2 and 3 represent the erroneously estimated camera parameters in relation to the assumed parameters ν and κ. Fig. 2 (a) shows notable drifts in the intrinsic parameters of the camera model in relation to the assumed aspect ratio (irrespective of κ). Fig. 3 does the same for the resulting absolute extrinsics. The results with $\nu\neq1$ will of course also imply erroneous extrinsic calibration. Even if $\nu=1$ there still exists the possibility that an erroneous scaling factor $\kappa\neq1$ yielded badly scaled absolute extrinsics (in range), even though the intrinsic parameters were optimally estimated. Fig. 2 (b) shows the error in the estimation of the hand-eye transformation in this last case.

2) Performance after erroneous calibration: Erroneous calibration sharply deteriorates performance. In Fig. 4 the *RMS*[13] of the intrinsic projection errors between the expected and the actually imaged projections of the *real* calibration object for all images are shown. The expected projections are obtained with the intrinsic parameters from traditional calibrations with $\tilde{\nu}\in[0.99,\,1.01]$ but with ground truth absolute extrinsics. For optimal calibration ($\nu=1$) the projection error is expectedly

[12] *Monocular* calibration is used since it is worse conditioned than *stereo*.

[13] *RMS*: **R**oot **M**ean **S**quare error, mostly in projected pixel distances. In this section only the *RMS* error representation is used for performance assessment since other meaningful evaluation metrics show similar trends [4].



(a) Estimated vs actual intrinsics [percent] (b) Errors in the hand-eye transformation [*mm*]

Fig. 2. Percent of error in the intrinsic parameters (a) and translation error in the hand-eye transformation (b) in relation to the pattern scaling parameters assumed for traditional calibration. The actual parameters are $\nu=1$ and $\kappa=1$.

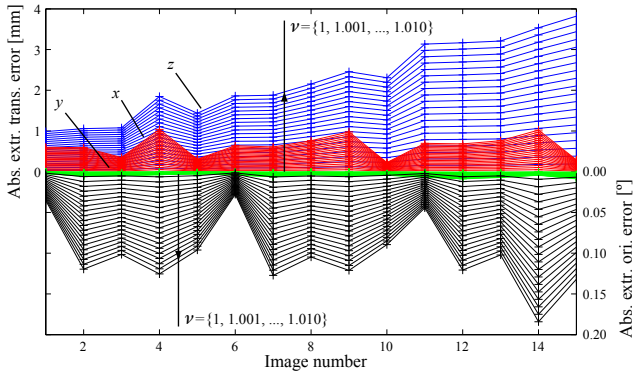


Fig. 3. Error in translation and orientation of the absolute extrinsics in relation to the aspect ratio assumed for traditional calibration. In reality $\nu=1$.

minimal and identical to the "*virtual*" residual after calibration ($0.21\,p$. *RMS*). On the contrary, for traditional calibrations with $\nu\neq1$ the error scales up (approx. linearly) to e.g. $0.4\,p$. *RMS* for $\nu=0.9975$ (0.25 % aspect ratio error, i.e. only 0.75 mm discrepancy between the x and y lengths when measuring a $30\times30\,cm$ section of the pattern as a whole). In addition, the "*virtual*" residuals after calibration with $\tilde{\nu}\in[0.99,\,1.01]$ are depicted; these reflect the operation of Eq. (5) relative to the assumed aspect ratio ν, where erroneous absolute extrinsics try their hardest to compensate for erroneous intrinsic parameters.

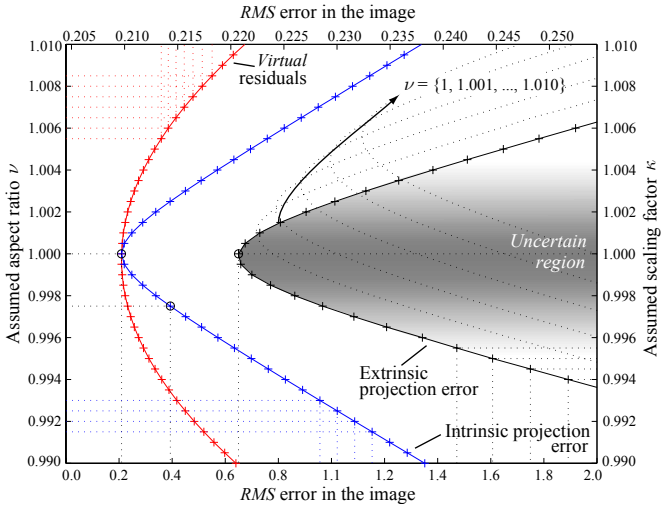


Fig. 4. Image projection errors in relation to the scaling parameters for traditional calibration. The actual parameters: $\nu=1$, $\kappa=1$. Note the different axes.

If the intrinsic calibration was followed by the extrinsic one, Section III showed that the extrinsic estimation may also become inaccurate, and naturally the eventual performance will get worse as well. The set of curves on the right-hand side in Fig. 4 show the projection errors where the actual noisy readings of the manipulator ${}_0\widetilde{\boldsymbol{T}}^t$, along with the traditionally estimated hand-eye transformations ${}_t\widehat{\boldsymbol{T}}^C$ with $\tilde{\nu} \in [1.00,\ 1.01]$ and $\tilde{\kappa} \in [0.99,\ 1.01]$, take the place of the former absolute extrinsics. For the ground truth parameters ($\nu = \kappa = 1$) the error scales up to $0.65\,p.$ *RMS*. The (small) noise in the manipulator readings accounts for this increment. Slightly erroneous pattern parameters skyrocket the error.

It is worth noting that in this last case and for traditional hand-eye calibration it clearly exists a *fundamental uncertainty region* where it is not possible for the user to assess the calibration accuracy, since it is subject to the absolute accuracy of the ruler at hand – refer to Footnote 9. For incorrect aspect ratio $\tilde{\nu}$ this is not clearly defined, since relative dimensions can be determined with high precision using inaccurate rulers.

B. Convergence of the novel estimation methods under noise

Simulations were conducted with variable noise levels in the positions of the control points detected in the images, and in the positioning accuracy of the robot manipulator. Fig. 5 shows the "*virtual*" residuals after traditional intrinsic calibrations with different image noise levels $\sigma_{\{u,v\}} \in [0.1,\ 1.0]\,p.$ and assumed aspect ratios $\tilde{\nu} \in [0.99,\ 1.01]$, as percentage w.r.t. the optimal results when $\nu = 1$. The residuals reflect the operation of Eq. (5). The minimum residual is unequivocal for the optimal solution $\nu = 1$ and shows that in this context *the erroneous intrinsic and absolute extrinsic parameters cannot completely compensate for erroneous knowledge of the aspect ratio of the calibration pattern* (refer to Section III-B). This result is basis for the intrinsic optimization in Eq. (5) of *Method #1*, since it clearly shows the existence of an unique unbiased minimum for the optimization. Similar results are achieved for the extrinsic calibration with scaling factor codetermination in Eq. (6), as well as for the optimizations of *Method #2*. In general, the methods do not only converge for the initial parameters shown in these simulations, but for significantly worse ones; aspect ratio and scaling factor errors of up to only $\pm 1\,\%$ were used in this section in order to visualize the absence of biases in the final estimations.

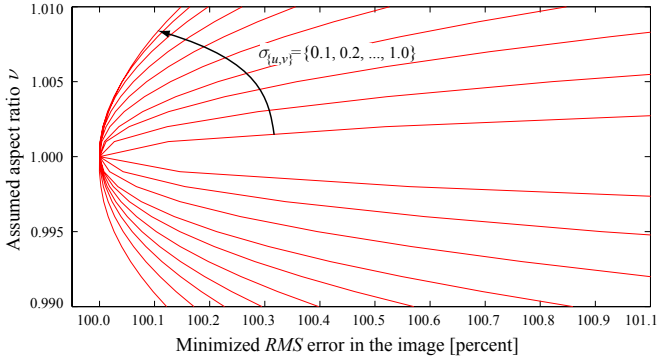


Fig. 5. Minimized *virtual* image projection errors in relation to the aspect ratio and different image noises $\sigma_{\{u,v\}} \in \{0.1,\ 0.2,\ \ldots,\ 1.0\}\,p.$

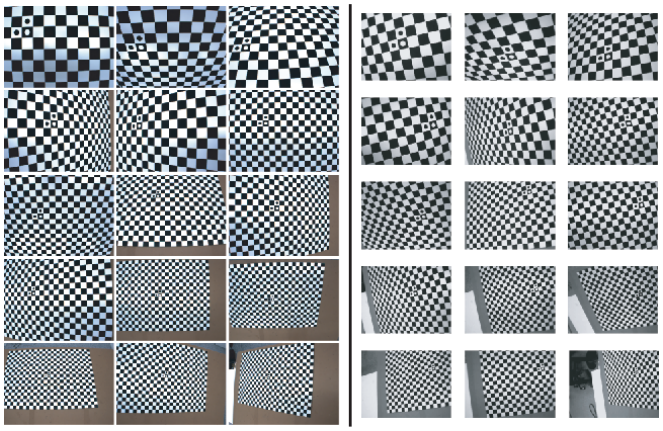


Fig. 6. The fifteen images used for calibration with the AVT Marlin camera (left, $780{\times}580\,p.$) and the Typhoon™ EasyCam camera (right, $640{\times}480\,p.$). The extrinsic poses were taken by a KUKA KR16. The pattern is size A2.

V. Experimental Results

In this section the performance of the algorithms in real systems is studied in order to validate both models and algorithms. Since the validity of the traditional camera calibration methods is out of the question,[14] here only the novelty concerning the codetermination of the pattern scaling parameters ν and κ is evaluated. Fortunately, and in contrast to the intrinsic parameters estimation, in doing so it is possible to directly assess the accuracy in the determination of these pattern parameters, since they can also be directly measured. If the determination is accurate, the estimation of the further parameters is necessarily the equivalent of the well-established traditional calibration methods, which consequently validates the novel methods in this work for the systems in test.

With the idea of validating in a wide range of systems, two different (monocular) cameras were used: On the one hand, an accurate progressive scan AVT Marlin camera with a SVGA $1/2$" Sony CCD chip and a Sony VCL-06S12XM $6\,mm$ objective worth $\$1{,}500$ altogether; on the other an off-the-shelf VGA $1/4$" CMOS $4\text{-}6\,mm$ Typhoon™ EasyCam webcam worth $\$15$. Both cameras are rigidly attached to the end-effector of a precise robot manipulator KUKA KR16 and take pictures of a precisely imprinted A2-sized checkerboard calibration plate, see Fig. 6.

Starting out on the experiment, ground truth data was obtained by visually measuring an extended patch of the checkerboard pattern with a metallic precision ruler – it was assumed that the checkerboard pattern regularly spreads in x and y directions. Specifically, the lengths d_x and d_y of the segments defined by 28 and 19 squares the size of $u_x \times u_y$ ($u_x \approx u_y \approx 2\,cm$) were measured, and the optimal parameters $\nu_\star = \kappa_\star = 1$ were assigned to them. After that, *Methods #1* and *#2* were used to estimate the correcting parameters $\hat{\nu}$ and $\hat{\kappa}$ against potentially erroneous pattern data, which in turn lead to the estimated lengths $\hat{d}_x$ and $\hat{d}_y$. The results in Tab. I show a formidable consistency of the estimated and measured

[14] This statement refers to the common set-up concerning cameras with relatively high resolution, reasonable field of view, and known pattern dimensions.

dimensions, even though the algorithms were initialized with dramatically wrong dimensions like $u_x = 3\,cm$ and $u_y = 1\,cm$.

TABLE I

		κ	ν	d_x	d_y	RMS_{int}
Precision ruler		1.00000	1.00000	559.6	379.0	—
AVT Marlin	*M. #1*	0.99906	1.00096	559.59	378.63	0.1735
	M. #2	0.99936	0.99910	559.34	379.55	0.1737
Typhoon™ EasyCam	*M. #1*	0.99967	1.00056	559.47	378.78	0.6452
	M. #2	1.00081	0.99780	558.82	379.31	0.6453
				$[mm]$	$[mm]$	$[pixel]$

These values are more accurate than the ones any user is able to obtain with the sole aid of a regular ruler over a smaller patch of the pattern.[15] As regards the computational costs – which is a quite immaterial issue for general calibration to begin with, *Method #1* hardly affects them, especially if the parameters are reasonably close to the true values. *Method #1* increases the lengths of the optimization vectors in only one parameter each, being them usually the size of $5+2+(6\times N)$ parameters intrinsically, and 6×2 extrinsically. In contrast, *Method #2* does significantly increase costs, since it implies an iterative process of numerical optimizations. In this work the *Levenberg-Marquardt* optimization method was used both for the intrinsic and the extrinsic calibrations.

These results call for the following **discussion**: In general systems, where the Gaussian image noise assumption largely holds, optimal intrinsic camera calibration is only attained if the aspect ratio of the pattern is perfectly known. Since this never holds outside of simulation scenarios, the user should opt for one of the methods in this work if he or she is not able to determine the aspect ratio u_x/u_y with an accuracy of say one part in a thousand (i.e. $0.3\,mm$ in a $30\times30\,cm$ patch). As to which method to use, the fact that the projection residuals in the images are mostly numerous and small, and conversely the camera vantage points fewer (typically 10 to 15) and their positioning errors of arbitrary size (depending on the system), suggests that the former errors distributions present much more a Gaussian nature than the latter ones. Therefore, *Method #1* should usually perform more accurately than *Method #2* for the codetermination of the aspect ratio ν. As regards the extrinsic calibration, the user should also opt for one of these methods if he or she is not able to determine the size of the plate with an accuracy of say one part in a thousand (i.e. $0.3\,mm$ accuracy in $30\,cm$), which actually is *very* often the case.

Acknowledgment: The authors would like to thank Dr. H. Hirschmüller, Dr. U. Hillenbrand, Prof. D. Burschka, Dr. J. F. Seara, Mr. W. Sepp, Mr. J. Dunn, and the anonymous reviewers for helpful suggestions. This work has been partially funded by the European Commission's Sixth Framework Programme under grant no. 011838 as part of the integrated project *SMErobot*.

[15] To be precise, the upper and lower lengths d_x of the measured patch actually differ as much as $0.3\,mm$ – here the mean value of $559.6\,mm$ was used. In addition to that, the measurements were in general the result of an interpolation between two different metallic precision rulers which differ in length about $0.3\,mm$ over their whole length of $1\,m$. Thus, these results come by the highest measurable accuracy that the authors were able to achieve.

VI. Conclusion and Future Work

The two novel approaches presented in this paper fix a shortcoming that frequently arises out of the traditional camera calibration methods. We note that highly accurate knowledge of the dimensions of the calibration pattern rarely exists, and furthermore that this violation has very negative effects on the proper estimation of the camera parameters.

This work starts out with an overview of the literature on camera calibration. We bring into focus that there is a tendency to decrease the complexity of the calibration object, motivated by the fact that this deskills the calibration procedure. We elaborate on this motivation and suggest that in fact this trend is appropriate, as less complex objects prevent damage to the calibration due to metric inaccuracies. Yet *we take the matter further, easing requirements of knowledge of the metric dimensions of the calibration pattern*.

It turns out that there is a simple parameterization of the *checkerboard pattern* that on the one hand corresponds to the actual inaccuracies resulting from regular printing equipment, and on the other allows for *optimal intrinsic and extrinsic calibration* irrespective of the actual values of the parameters.

We substantiate our allegation on accuracy improvement with simulations and impressive experimental results. The camera calibration toolbox DLR CalDe and DLR CalLab [12] will offer these algorithms in the near future.

References

[1] B. Triggs, "Autocalibration from Planar Scenes," in *Proc. of the ECCV*, pp. 89–105, Freiburg, Germany, 1998.
[2] O. Faugeras and Q.-T. Luong, *The Geometry of Multiple Images*. The MIT Press, Cambridge, MA, USA, 2004.
[3] J. Weng, P. Cohen, and M. Herniou, "Camera Calibration with Distortion Models and Accuracy Evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 10, pp. 965–980, 1992.
[4] W. Sun and J. R. Cooperstock, "An Empirical Evaluation of Factors Influencing Camera Calibration Accuracy Using Three Publicly Available Techniques," *Mach. Vision and App.*, vol. 17, pp. 51–67, March 2006.
[5] J. Mallon and P. F. Whelan, "Which Pattern? Biasing Aspects of Planar Calibration Patterns and Detection Methods," *Pattern Recog. Letters*, vol. 28, no. 8, pp. 921–930, June 2007.
[6] 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*, vol. 3, no. 4, pp. 323–344, August 1987.
[7] Z. Zhang, "A Flexible new Technique for Camera Calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, November 2000.
[8] P. F. Sturm and S. J. Maybank, "On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications," in *Proc. of the IEEE Conf. on Comp. Vision and Pattern Recog. CVPR*, pp. 432–437, Fort Collins, USA, June 1999.
[9] H. Malm and A. Heyden, "Stereo Head Calibration from a Planar Object," in *Proc. of the IEEE Conf. on Comp. Vision and Pattern Recog. CVPR*, vol. 2, pp. 657–662, Kauai, Hawaii, USA, December 2001.
[10] K. H. Strobl and G. Hirzinger, "Optimal Hand-Eye Calibration," in *Proc. of the IEEE/RSJ International Conf. on Intelligent Robots and Systems IROS*, pp. 4647–4653, Beijing, China, October 2006.
[11] 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*, pp. 158–174, Freiburg, Germany, 1998.
[12] DLR CalDe and DLR CalLab. Institute of Robotics and Mechatronics, German Aerospace Center [Online]. http://www.robotic.dlr.de/callab/
[13] R. Kingslake, *Optics in Photography*. SPIE, Bellingham, USA, 1992.
[14] J. A. Sánchez et al., "Plane-based Camera Calibration without Direct Optimization Algorithms," in *IV Jor. Arg. de Robótica*, Argentina, 2006.
[15] T. Tamaki, "Unified Approach to Image Distortion: D-U and U-D Models," *IEICE Transactions on Inf. & Systems*, pp. 1086–1090, 2005.