On the Issue of Camera Calibration with Narrow Angular Field of View

Klaus H. Strobl, Wolfgang Sepp, and Gerd Hirzinger

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

***Abstract*— This paper considers the issue of calibrating a camera with narrow angular field of view using standard, perspective methods in computer vision. In doing so, the significance of perspective distortion both for camera calibration and for pose estimation is revealed. Since narrow angular field of view cameras make it difficult to obtain rich images in terms of perspectivity, the accuracy of the calibration results is expectedly low. From this, we propose an alternative method that compensates for this loss by utilizing the pose readings of a robotic manipulator. It facilitates accurate pose estimation by nonlinear optimization, minimizing reprojection errors *and* errors in the manipulator transformations at the same time. Accurate pose estimation in turn enables accurate parametrization of a perspective camera.**

I. Introduction

The foundation pillars for standard, perspective camera calibration in computer vision applications are the following: appropriate definition of the camera model (a), successful initial estimation of its parameters (b), availability of enough evidence on perspective distortion (c), and finally proper estimation of scene structure (d). Whenever one of these pillars is shaking, the accuracy of standard camera calibration is compromised.

It is remarkable that the geometric model of ancient *pinhole* cameras still holds for accurately describing the main functioning principle of a number of modern cameras (a). An approach for accurate, simple parameter initialization within this model (b) was proposed in 1999 by Zhang and Sturm and Maybank [1], [2]; this approach proved extremely useful, thus most successful. In 2008 Strobl and Hirzinger noted a predominant error source for correct scene structure estimation (d) and brought forward an alternative formulation [3]. In the present work we focus on a critical aspect concerning the remaining pillar: the requirement for satisfactory evidence on perspective distortion, in particular in relation to the limited angular field of view (AOV) of some cameras.

Perspective distortion is a direct consequence of the use of pinhole model-like cameras. It describes the mapping of a 3-D scene onto its 2-D image and can be roughly summed up by these two circumstances: close objects project bigger, and differently distant objects may project onto the same region—i.e. range gets lost. These circumstances are regularly used for camera calibration since they help to discriminate between the Euclidean structure of both scene and camera, and the camera magnifying characteristics themselves. The images in Fig. 1 show different perspective distortion effects on images

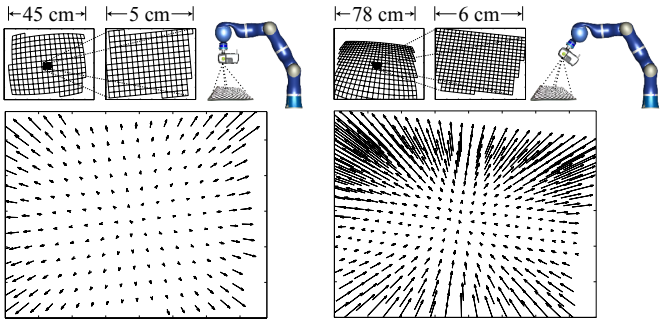


Fig. 1. Camera projection of chessboard calibration patterns distant $32\,cm$ and perpendicular to the principal axis of the camera (a), and distant $41\,cm$ and tilted $37°$ (b), for two different scaling parameters $\alpha=\beta=482$ ($86°\times 65°$) and $\alpha=\beta=4820$ ($9.5°\times 7°$), radial lens distortion $k_1=0.155$, and image size 780×580. The object pattern sizes differ for each projection. Corresponding points within each image are linked together in order to show the very significant evidence on perspective distortion from oblique views (b). The residuals in (a) are solely owing to static, radial lens distortion. At the top the projections are separately depicted; two illustrations show the mentioned vantage points with cameras mounted at the DLR Light-Weight Robot 3.

of a planar pattern in relation to both the external orientation of the camera—(a) against (b)—and different magnifying characteristics of its perspective camera model—within each figure.

In this respect, the issue of calibrating a camera with very limited AOV is addressed in this work. It is difficult to gather enough evidence on perspective distortion with these cameras, thus calibration accuracy gets compromised. Even though the particularities of wide AOV have been often addressed [4], to the best of our knowledge the issue in this paper has been left largely untreated in computer vision—apart from changing into affine camera approximations.[1] According to our experience on maintaining the camera calibration toolbox DLR CalDe and DLR CalLab [8], there is less concern among the users about applying regular camera calibration methods for camera systems with extremely *narrow* AOV.[2] Of course, there exist photogrammetric approaches to deal with this problem

[1] Affine camera models are tolerable approximations of perspective projection cameras when the AOV and the relative variation of depth are small. Their models are linear (instead of merely linear projective in the case of perspective models), thus allow for linear algebra solutions (instead of nonlinear solutions) to a number of algorithms. In addition, affine camera calibration is better conditioned for narrow AOV. However, affine camera models are still approximated and very limiting in several aspects [5]–[7].

[2] Bad relative positioning choices between camera and calibration plate is reportedly the other most common reason for erroneous camera calibration [8].

since it occurs very frequently in that field; however, these methods and the required equipments (e.g. optical collimators) are rarely available outside photogrammetric labs.

What makes it all the worse are the *numerous applications* of narrow AOV cameras, and what is more that these applications are mostly justified precisely by the high accuracy that they are supposed to provide. A collection of examples: long focal length cameras for feedback control of robotic manipulators in industry (e.g. laser beam welding), high-accuracy positioning by gazing at landmarks in structured environments, foveal vision e.g. for anthropomorphic research, manufacturing inspection in intricate cavities, etc. In practice, it is mostly only possible with *this* type of cameras to further increase the already high accuracy of current robotic manipulators.

The remainder of this article is as follows: Section II further introduces the fundamentals of the pinhole camera model. These serve as a basis for Section III, where the potential accuracy in camera calibration and utilization is considered in relation to the AOV of the camera. In Section IV we propose a calibration method to amend the deficiencies shown in the latter section; it uses positioning information from a robotic manipulator. The work is recapitulated in Section V and a statement of future work is delivered.

II. The Pinhole Camera Model and the Role of the Focal Length

The pinhole camera model is the main part of the projection model of most cameras in computer vision applications. It represents the perspective projection taking place when mapping the 3-D world scene onto the 2-D imaging plane by rays of light passing through a (conceptual) point, the camera focal point or focus. In reality, the imaging plane is usually instantiated by an electronic imaging sensor like charge-coupled devices (CCD) or CMOS chips, and the focus of the camera is located at the aperture center of the frontal lens. Further potential parts of the camera model are the digitization process, the lens distortion model, or the extrinsic rigid-body transformation from the camera to an external point.

The geometrical mapping of 3-D points ${}_O\boldsymbol{x}$ in the world/**o**bject reference frame S_O onto their projections $\boldsymbol{m}$ in the imaging plane can be mathematically represented as follows:

$$s\,\boldsymbol{m} = \underbrace{\boldsymbol{A}\,{}_C\boldsymbol{T}^O}_{\boldsymbol{P}_{3\times 4}}\,{}_O\boldsymbol{x}, \quad \boldsymbol{A} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

where s is an arbitrary scale factor and $\boldsymbol{P}$ is the perspective projection matrix, which consists of the camera intrinsic matrix $\boldsymbol{A}$ and the rigid body transformation ${}_C\boldsymbol{T}^O$ from the **c**amera frame S_C to the object/world frame S_O at a particular imaging instant (time and point-related indexes have been omitted for the sake of clarity). The matrix $\boldsymbol{A}$ is in turn composed of the scaling parameters α and β, which are directly proportional to the focal length, the skew parameter γ, which represents slight skewness in the image plane coordinates, as well as the 2-D coordinates u_0 and v_0, which locate the principal point in the image frame. The principal point is supposed to be its closest point to the focal point—usually it is *not* [9].

This is the generally established formulation but other formulations exist as well. In the past, the model was much more related to actual camera parameters like the sizes of the picture elements in different directions, or to the focal length, cf. Refs. [10], [11]. However, this does not pay off for both calibration and utilization of regular cameras, and all-encompassing parameters like α or β are currently preferred [8], [12]. Nonetheless, the user should bear in mind both their origin and nature.

The focal length is one of the main camera parameters that have to be taken into account *either in reconstruction*, in order to extract information from the image projection, *or in acquisition*, to determine both scene and camera position and orientation (pose) so that the user eventually obtains the desired image projection. In reality, it modifies the perpendicular distance between the focal point and the image frame; for instance, the angular area of the projected scene *reduces* when the focal length increases (paradoxically this is what we get to call image *amplification*), which is due to the limited size of the sensor chip.

But strictly speaking, in the perspective distortion issue it is all about the pose of the camera w.r.t. the scene, since it primarily defines the *potential* perspective distortion that we can expect from the *whole* scene, whereas the focal length relates to the AOV by narrowing or broadening it (thus determining the absolute scale of the projection) but without modifying its appearance. When one speaks of decreasing the perspective distortion by increasing the focal length what actually occurs is *either* that the projected scene reduced to a small section of the original one, without moving the camera nor the potential perspective distortion, *or* that the camera departed from the scene and the focal length had to be increased in order the same part of the original scene to remain on camera, losing some perspective distortion all this way. In this second case, increasing the focal length is just a by-product of the action of moving the camera since else the imaging chip would get a huge viewing area, wasting most of its valuable pixels for void space.[3]

In the next sections performance both in final camera utilization as well as in calibration will be discussed in relation to the AOV. All through the work a number of simulation results will be featured; they are strongly based on real data—therefore the seemingly arbitrary choice of coordinates. In the simulations, the poses of the camera with respect to (w.r.t.) the scene remain constant for every set of images; this allows for fair comparisons regarding precision in pose estimation. The focal length (and the size of the pattern) are of course changed. It is useful to first clarify the relationship between AOV and the focal length, which is a nonlinear one, see Fig. 2 (a).

In general, of course, the longer or shorter the focal length, the smaller or bigger the AOV, respectively. However, the reduction of AOV in a couple of degrees when it is already small takes much bigger an increase in focal length than it

[3] Since A. Hitchcock's *Vertigo* this effect is also used by filmmakers to provoke a disquieting sensation, or the character's reassessment of a situation.

would take if the AOV were bigger. This is inconvenient e.g. if it is required to represent simulation data w.r.t. the AOV (as we do in this work for more natural and general reading), since uniformly distributed sampling in focal length implies highly non-uniform distributions in AOV. This issue is easily handled by uniformly distributing *on the inverse* of the focal length, which almost linearly corresponds to the AOV, see Fig. 2 (b). In this work all simulations are going to be performed on this distribution—yet represented in AOV.

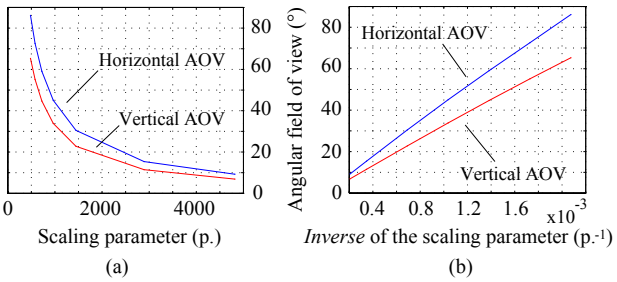


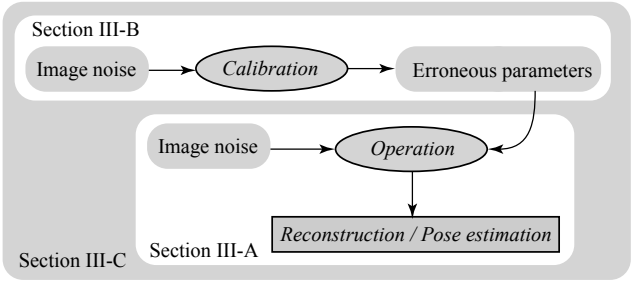
Fig. 2. Relation of the AOV with the scaling parameter $\alpha = \beta$ (a) and with its inverse (b). Ordinary radial distortion has been also taken into account.

For the rest, customary camera parameters are used. It is worth mentioning that the resolution is invariably set to moderate $780 \times 580\, p.$ — this value is relevant only in direct conjunction with the scaling parameters and the image processing noise, which follows an homogeneous 2-D zero-mean Gaussian distribution with standard deviation $\sigma_{\{x,y\}} = 0.4\, p.$ This can potentially map to diverse actual camera systems.

III. Erroneous Camera Model Utilization and Parametrization

One of the main points of this work is to assess the accuracy that we can *finally* expect from vision-based algorithms in general. On the way to the final application, the accuracy becomes compromised in several steps. It is indicated to separately study these error sources in this section.

The following sequence of events is responsible for inaccurate vision-based estimation in most computer vision applications that require calibrated cameras. Starting out from the calibration process: In the beginning was only image processing noise when detecting features in images for calibration. Through calibration we may get a parameterized camera model, but its values are erroneous to some extent, and what is more, even the model is only approximated. Eventually, in final camera operation, both the erroneously estimated camera model and *further* image processing noise jointly affect the accuracy of the final estimation adversely.



In Section III-A operation from noisy image processing will be treated. Then we also consider erroneous calibration from noisy image processing in Section III-B. On the basis of these latter results, we extend the former initial results on noisy operation taking also camera model parametrization errors into account; this is in Section III-C.

A. Image-based estimation from noisy image processing

In this section we present ordinary results on camera *pose estimation* from known scenery on the pretentious assumption that both the camera parametrization and its model are known and *flawless*. The scenery corresponds to a perfectly known planar chessboard pattern as used for camera calibration. The projections of the pattern are affected by homogeneous Gaussian noise as above mentioned. The pose estimation algorithm is an optimal nonlinear optimization process that minimizes the sum of squared prediction errors of object reprojections—optimal provided that the estimation is initialized on the convex area of the absolute minimum. This frugal example is in preparation for more complex ones in the following.

In Fig. 3 the pose estimation precision for different AOVs is shown; the camera is at a fixed distance and perpendicular view to the plate. The curves result from 7 AOV points, uniformly sampled on the inverse scaling parameter space (or inverse focal length space). The data stem from a Monte Carlo simulation consisting of 1000 pose estimation optimizations repeated with independent image noise, for each AOV. The images in Fig. 1 (a) correspond to the horizontal extremes in Fig. 3, i.e. with the widest and the narrowest AOVs.

The figure shows a considerable worsening of both positioning and orientation estimation precision for small AOVs—even though the camera model holds perfectly. It was mentioned in Section II that it is the camera pose that is responsible for perspective distortion in the images. Since planar structure points from perpendicular images present similar distances, these images provide less *variation* in perspective distortion w.r.t. the camera pose (cf. Fig. 1 (a)), which comes near by affine projection and ambiguities like the Necker reversal. Therefore, pose estimation becomes bad conditioned. It is only due to both the known structure and the known camera scaling (focal length) that at least the range estimation is good conditioned (see z in Fig. 3).

It is interesting to compare this simulation with the results when the camera *is* tilted w.r.t. the calibration plate (Figs. 5 and 1 (b)). The perspective distortion due to an inclination of 37° is more pronounced because different ranges appear and differently distant parts project in different sizes, which makes the *relative* pose estimation better conditioned. This is so pronounced that the accuracy becomes virtually independent of the actual focal length. Furthermore, the known scaling parameter of the camera along with the known structure allow for accurate range estimation, thus *absolute* pose estimation.

It is alarming news that perpendicular views to planar objects are most common both in final applications as well as in camera calibration.

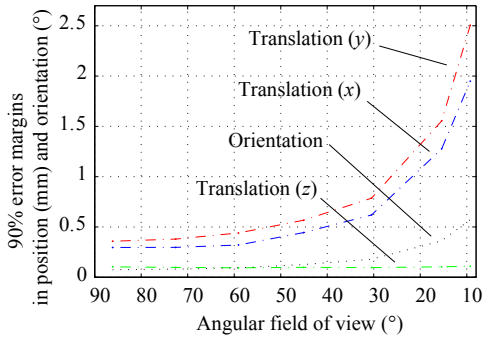


Fig. 3. Positioning and orientation accuracy (90% error margin) w.r.t. the AOV, with camera range $32\,cm$ and perpendicular view to the planar object. The estimation biases are insignificant.

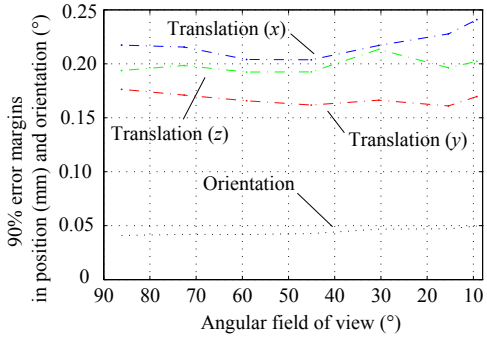


Fig. 5. Positioning and orientation accuracy (90% error margin) w.r.t. the AOV, with camera range $41\,cm$ and tilted $37°$ w.r.t. the perpendicular of the planar object. The estimation biases are insignificant.

These results still represent the minimum errors that the user should expect. Image processing errors rarely spread homogeneously in the image nor are clear of outliers, and neither the pinhole camera model nor its parametrization completely hold but in simulations. It is also worth mentioning that, on occasions, camera parametrization inaccuracies are implicitly assumed *within* the imaging noise in normal operation—the validity of this assumption has to be determined elsewhere.

B. Erroneous camera model parametrization

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 section applies the most common algorithms in computer vision for camera calibration (e.g. [8], [12]) for different AOVs. Noteworthy details are the following: The used camera parametrization corresponds to Ref. [1], and the parameters initialization is also performed by the algorithms of Refs. [1], [2]. The algorithm requires a perfectly known calibration plate [3] which confines the user to close- to mid-range imaging. In a nutshell: The camera calibration process consists in optimally estimating the pinhole camera parameters (mainly the focal length) by numerically minimizing image reprojection errors for several object views. In intrinsic camera calibration, *several views* are mainly required for the parameters initialization stage, cf. [1], [2]. In extrinsic camera calibration, at least three views (specifically two movements with nonparallel rotation axes) are required [13]. In addition, the central limit theorem requests a sufficiently large amount of data, and therefore so does statistical optimality.

Both the principal point location and the distortion parameters are set to fixed, realistic values and are not being estimated. This is because potential variation of these parameters directly implies a motion of the camera frame;[4] in turn, a motion of the camera frame implies drifts in the remaining intrinsic parameters—including the focal length. This interplay would cover up the intrinsic weakness that we want to show up in this section concerning the interdependence of the focal length estimation and the estimation of the camera poses in the presence of noisy image data and a limited AOV. This adoption fixing distortion parameters is realistic since they can—and on occasions even should—be estimated *in advance* of pinhole camera model calibration by e.g. the approach in Ref. [15]. Furthermore, lens distortion is scarcely noticeable in narrow AOV camera systems, cf. Fig. 1 (a). In addition to this, the ground-truth camera model used in this study also lacks of skewness, and the relative projection scaling α/β is enforced to unity (i.e. $\alpha \triangleq \beta$). In this way, the only remaining camera parameter is the focal length, which is the central parameter of the pinhole camera model after all. These measures support the results on calibration accuracy presented here since they make this study a best case scenario for camera calibration, where fundamental weaknesses for general models are to be clearly identified.

[4] The translation of the principal point of a pinhole camera model implies primarily shifting the origin of the lens distortion effects [9], [14], secondarily a rotation of the camera frame, and thirdly a slight displacement of it [10].

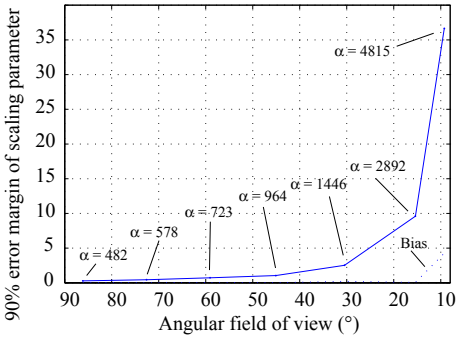


Fig. 4. Scaling parameter estimation error (90% error margin) w.r.t. the actual scaling parameter α after 150 standard camera calibrations for each AOV.

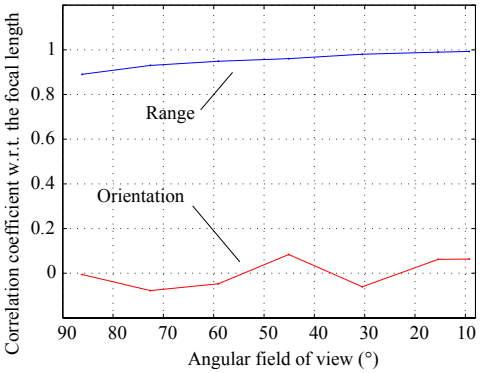


Fig. 6. Correlation coefficient relating the focal length estimation error with the range and orientation estimation errors for a typical calibration image.

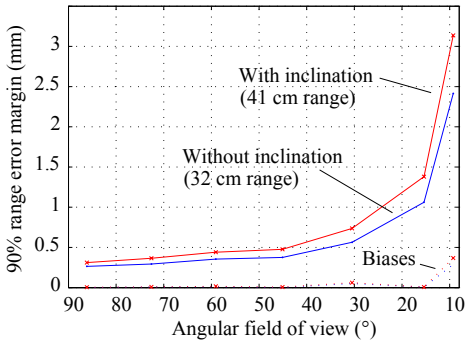


Fig. 7. Range estimation error (90% error margin) after 150 standard camera calibrations for each AOV (for the images with and without inclination).

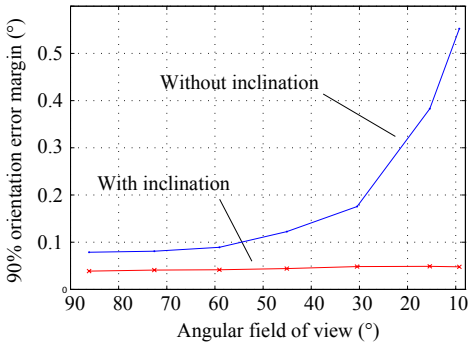


Fig. 8. Orientation estimation error (90% error margin) after 150 standard camera calibrations for each AOV (for the images with and without inclin.).

Next the statistical results from 150 intrinsic calibrations for each of the 7 different focal lengths/AOVs are presented. For each calibration, image noise is independently generated for 12 convenient, different calibration images. In Fig. 4 the focal length estimation results (in the form of the scaling parameter) are compared to ground-truth. It can be seen that the estimation accuracy of the focal length strongly depends on the AOV of the camera; it worsens for narrow AOVs.

Similar to the last section, the camera poses are also unknown and have to be estimated. In Figs. 7 and 8 the accuracies of these extrinsic estimations are depicted w.r.t. the ground-truth for the two same images treated in the last section, perpendicular and tilted (which are included in the 12 images used for calibration). The positioning accuracy (in this case its range, i.e. $\sqrt{{}_C x^2 + {}_C y^2 + {}_C z^2}$) worsens for narrower AOVs, similar to the results in Fig. 3. However, in the case of tilted views, the results are very different since now they also suffer from positioning inaccuracy, cf. Figs. 5 and 7. This was expected since the (erroneous) focal length is responsible for the absolute scaling of images, refer to Section II.

In general, two reasons account for the inaccurate estimation of the scaling parameter: On the one hand, a reduction of the AOV (without relocating the camera) implies that a smaller area of the scene will be seen, and therefore that there will be less evidence for accurate estimation—notwithstanding some more precision in the measurements. As we mentioned in Section II, this is not because of any change in the potential perspectivity of the scene, but because of the limited size of the imaging chip. However, the comparison in Section III-B made clear that individual tilted images still contain perspectivity evidence for very accurate camera pose estimation. Exactly the same in camera calibration, it is the perspective distortion that differentiates camera range from focal length, and therefore one would expect that camera calibration does a better job in the estimation of the camera pose of tilted images, cf. Fig. 7. One the other hand, during the camera calibration process the intrinsic camera parameters are continously being shared between all calibration images. Erroneous pose estimation by certain images (e.g. the perpendicular ones, see Fig. 3) will spread to images with sufficient perspectivity information simply because they share the focal length parameter. This point intensifies our conflict with perpendicular images mentioned in Section III-A, even though perpendicular images may be useful for reliable lens distortion estimation.

Fig. 8 shows that the accuracy of the camera orientation estimation is not affected by the concurrent estimation of the focal length (cf. with Figs. 3 and 5), and Fig. 6 shows extreme correlation between range and focal length estimations and no correlation between orientation and focal length estimations. This is because the projective effects of camera rotations and focal length adaption are clearly differentiated.

The results in Figs. 7 and 8 could also help to define a threshold for the proper definition of a potential, subsequent hand-eye calibration [13].

But for all that, it is often not sensible to validate results in relation to the estimation accuracy of particular parameters—for instance, one may expect less accuracy in long-range pose estimation than in short-range, which is perfectly normal. In the next section the consequences of this issue in *final* camera operation will be shown.

C. *Image-based estimation from noisy image processing* and *erroneous models*

After each calibration process it is convenient to be able to properly assess the calibration results. The most common practice is to mention the root mean square error (RMS) in reprojection after intrinsic calibration. Whereas this is acceptable for regular cameras with reasonable AOV, proper camera and object models, and valid image processing and optimization processes, this practice is intrinsically wrong. This is because during optimization it is explicitly rewarded to minimize precisely that RMS error at expenses of the model parametrization. Two evils come on scene: erroneous camera parameters and wrongful reprojection residuals, thus wrongful assessment.

In order to assess the model parametrization of the last section *in final operation*, the following simulation was performed: For *each* of the calibration results from the last section (i.e. 150 intrinsic calibrations for each of the 7 different AOVs), *again* 250 sets of simulated noisy points were generated for all images, on the ground-truth projections *at the ground-truth camera poses*. Only in this way the real parametrization errors emerge—as opposed to the residuals after calibration—since the estimated camera poses are not a valid outcome of the calibration process. For perfect model parametrization, this RMS reprojection error should average the Gaussian image noise $\sqrt{\sigma_x^2 + \sigma_y^2} = 0.56\,p.$, but Fig. 9 shows that this only happens for wide AOVs; for narrow AOV the residuals fairly surpass that level.

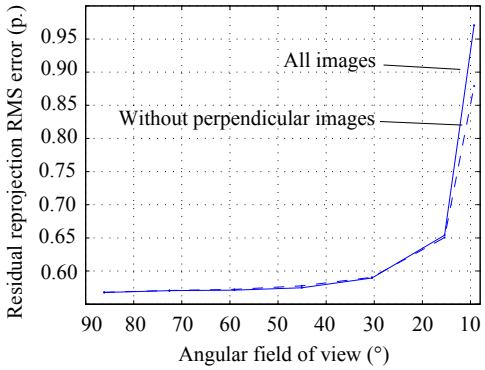


Fig. 9. Residual reprojection RMS error after standard calibration *and* further erroneous reprojection, for each AOV.

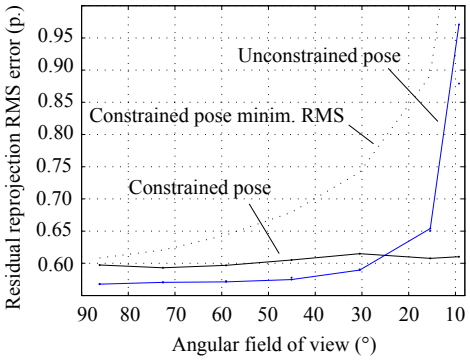


Fig. 10. Residual reprojection RMS error after calibration *and* further erroneous reprojection for intrinsic calibration supported by the robotic manipulator. The standard intrinsic calibration results of Fig. 9 are overlaid.

Fig. 9 also shows the accuracy results when the 5 images where the principal axis of the camera and the perpendicular to the calibration plate comprise less than $15°$ are omitted from all calibrations. The remaining 7 images do yield slightly more accurate results. However, it is inconvenient to further omit tilted images since image processing performs worse with strong perspective distortion. In general, it is very difficult to further improve in this way.

A comparison between these results and the ones in Section III-A could bring light into the question of whether and when is it really appropriate to consider the calibration errors as an extra level of imaging noise during final camera operation.

IV. Proposed Calibration Method: Joint Intrinsic and Hand-Eye Estimations

Inspired by photogrammetric procedures but recasting them in the typical computer vision scenario of robotics, we propose the use of pose readings of a robotic manipulator in order to support the intrinsic calibration of a camera, being the camera mounted on its end-effector. The method should be used in the case of images showing lack of perspectivity, e.g. with narrow AOV cameras, constrained camera placements (e.g. few tilted images), or with visually impaired cameras like some endoscopes. Furthermore, in this way the processes of intrinsic and extrinsic camera calibration merge and former intrinsic inaccuracies do not harm the latter, potential extrinsic (hand-eye) calibration anymore.

The pose readings of the kinematic chain of the manipulator represent the rigid-body motions ${}_B\tilde{\boldsymbol{T}}_i^T$ between the **b**ase of the manipulator S_B and its end-effector (**T**CP) S_T in different instants i. Along with the fixed (yet unknown) object-to-base ${}_O\boldsymbol{T}^B$ and hand-eye ${}_T\boldsymbol{T}^C$ transformations, they define the pose of the camera S_C in S_O: ${}_O\hat{\boldsymbol{T}}_i^C = {}_O\hat{\boldsymbol{T}}^B\ {}_B\tilde{\boldsymbol{T}}_i^T\ {}_T\hat{\boldsymbol{T}}^C$. The idea suggests itself to directly include these extrinsic transformations in place of the unknown poses of the camera, performing a common minimization of reprojection errors for estimation of the camera parameters. Even though this may result in lower RMS error after calibration, simulations like in Section III-C show that this approach *worsens* the accuracy in the estimations, see Fig. 10. Similar to the motivation of our earlier work in Ref. [13], we understand that optimal stochastic estimation by minimization can only be achieved if all significant error sources are minimized (according to their statistical distributions). By the inclusion of manipulator readings that are naturally noisy, substantial deviations appear, and these deviations are of similar effect than image noise.

An approach for optimal hand-eye calibration on noisy manipulator readings was previously presented in Ref. [13]; it consists in a minimization of transformation errors of a robotic manipulator. Translational and rotational errors ($\mathcal{O}^{tra}$ and $\mathcal{O}^{rot}$) are considered separately, but are minimized at the same time in relation to the precision characteristics of the system. Here we extend this formulation for simultaneous intrinsic and extrinsic camera calibration by including reprojection errors in the minimization; furthermore, the algorithm is able to automatically adapt its weighting factors ${}^\star\sigma_{x|y}$, ${}^\star\sigma_{rot}$, and ${}^\star\sigma_{tra}$ according to the precision characteristics of the system iteratively, refer to [13]. The extended optimization problem now reads:

$$\{{}_T\boldsymbol{T}^C, {}_O\boldsymbol{T}^B, \alpha\}^\star = \arg\min_{{}_T\boldsymbol{T}^C,\, {}_O\boldsymbol{T}^B,\, \alpha} \left(\sum_{i=1}^{n} \frac{(\mathcal{O}_i^{im})^2}{{}^\star\sigma_{x|y}^2} + \frac{(\mathcal{O}_i^{rot})^2}{{}^\star\sigma_{rot}^2} + \frac{(\mathcal{O}_i^{tra})^2}{{}^\star\sigma_{tra}^2} \right)$$

where $\mathcal{O}_i^{im} = \sum_{p=1}^{n_i} ({}_p^i\Delta_x^2 + {}_p^i\Delta_y^2)$ accumulates the n_i square reprojection residuals $\Delta_x^2 + \Delta_y^2$ in image i.

Next the analogous simulation to Section III-C is performed. In addition, noisy manipulator poses were generated over ground-truth manipulator poses inspired by real calibration scenarios. The error was added to the pose of the end-effector[5] as follows: The angles θ of the angle-axis representation $\{\theta, \boldsymbol{p}\}$ of the added rotational errors follow a zero-mean Gaussian distribution with $\sigma_\theta = 0.1°$ and their axes $\boldsymbol{p}$ are uniformly distributed, i.e. their azimuth 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 $pdf(\psi) \propto \sin(\psi/90)\,[°]^{-1}$. The translational errors $\boldsymbol{t}$ also follow a zero-mean Gaussian distribution in range with $\sigma_t = 0.5\,mm$ and the directions are again uniformly distributed. These relative precision levels are conservative and are readily surpassed by most commercial robotic manipulators.

[5] Translational residual errors at the end-effector S_T are more realistic than at the base S_B because of the rigid base of manipulators; note that, for generality, the minimization algorithm assumes translational errors both at S_B and at S_T. Orientational errors in S_B and in S_T are of course equivalent.

In Fig. 10 the results of the intrinsic calibration aided by the robotic manipulator (*constrained pose*) are superimposed on the former results of the standard intrinsic approach in Fig. 9 (*unconstrained pose*). The constrained approach is insensitive to the narrowness of the AOV and reaches slightly worse intrinsic accuracy than optimal due to the noise in the manipulator readings. This very low error level seems preferable to the dangers of using affine camera models. The figure also shows the level of narrowness at which this method should be preferred to standard perspective camera calibration ($\sim 25°$). For bigger AOV this positioning aid should not be used. The figure also shows the accuracy reached by minimizing only RMS reprojection errors, i.e. not considering errors in the manipulator readings; the algorithm performs even worse than standard intrinsic camera calibration at these positioning accuracy levels.

V. Conclusion and Future Work

This work considers the issue of camera calibration for computer vision applications with the particularity of narrow angular fields of view. We reveal deficiencies not in the validity of the pinhole camera model, but in the ability of standard camera calibration algorithms to accurately parameterize it. Narrow angular fields of view make it difficult to obtain the required evidence on perspectivity in images; this compromises algorithms that rely on this evidence and furthermore consider several images at the same time, like camera calibration.

The work starts out with an overview on camera calibration for computer vision applications. This justifies clearly why there is need to address this problem. Crucially, a significant number of major application areas are listed. Next we descriptively explain the roles of focal length and camera pose for the achievement of perspectivity richness in images. We also demonstrate the consequences of critical evidence on perspectivity for exemplary computer vision applications as well as for standard camera calibration, and lay emphasis on the detrimental effects for image-based estimation of images taken perpendicular to planar objects.

Since perspective distortion is primarily defined by the pose of the camera, it will be difficult for any algorithm to accurately discern pose on the basis of insufficient evidence on perspectivity; the same holds for pose *and* focal length estimation, i.e. standard camera calibration. For this reason we propose an alternative method that uses positioning information from a robotic manipulator in order to support intrinsic camera calibration. Experiments show that the direct insertion of this extrinsic information in the optimization problem, still by minimizing reprojection residuals only, does not support intrinsic camera calibration but compromises it. This is due to the naturally noisy readings of the robotic manipulator. We introduce a novel method that optimizes the intrinsic and extrinsic parameters by the minimization of a *hybrid residual term*; it consists of translational and rotational errors in the kinematic transformation of the robot as well as image reprojection errors. This method extends former work by the authors on accurate hand-eye calibration [13].

Concluding, accuracy assessments compare this formulation with current intrinsic camera calibration approaches, and prove its better performance for narrow angular fields of view.

In the future extensive research has to be carried out in order to develop the optimal camera calibration algorithm capable of *automatically* pointing at the most convenient method for calibrating any given camera system. Finally, these extensions will be included in the camera calibration toolbox DLR CalDe and DLR CalLab [8], and offered to the research community.

References

[1] 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.

[2] P. F. Sturm and S. J. Maybank, "On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR*, Fort Collins, USA, June 1999, pp. 432–437.

[3] K. H. Strobl and G. Hirzinger, "More Accurate Camera and Hand-Eye Calibrations with Unknown Grid Pattern Dimensions," in *Proceedings of the IEEE International Conference on Robotics and Automation ICRA*, Pasadena, CA, USA, May 2008, pp. 1398–1405.

[4] S. S. Brandt and J. Kannala, "A Generic Camera Model and Calibration Method for Conventional, Wide-Angle, and Fish-Eye Lenses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1335–1340, August 2006.

[5] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[6] O. Faugeras, Q.-T. Luong, and T. Papadopoulo, *The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications*. The MIT Press, ISBN: 0262062208, 2001.

[7] S. Christy and R. Horaud, "Euclidean Shape and Motion from Multiple Perspective Views by Affine Iterations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 11, pp. 1098–1104, 1996.

[8] K. H. Strobl, W. Sepp, S. Fuchs, C. Paredes, and K. Arbter. DLR CalDe and DLR CalLab. Institute of Robotics and Mechatronics, German Aerospace Center (DLR). Oberpfaffenhofen, Germany. [Online]. Available: http://www.robotic.dlr.de/callab/

[9] R. G. Willson and S. A. Shafer, "What is the Center of the Image?" *Journal of the Optical Society of America A*, vol. 11, no. 11, pp. 16–29, November 1994.

[10] 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.

[11] O. D. Faugeras and G. Toscani, "Camera Calibration for 3D Computer Vision," in *Proceedings of the International Workshop on Machine Vision and Machine Intelligence*, Tokyo, Japan, February 1987, pp. 240–247.

[12] J.-Y. Bouguet. Camera Calibration Toolbox for Matlab. Computer Vision Research Group, California Institute of Technology (Caltech). Pasadena, CA, USA. [Online].
Available: http://www.vision.caltech.edu/bouguetj/calib_doc/index.html

[13] K. H. Strobl and G. Hirzinger, "Optimal Hand-Eye Calibration," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems IROS*, Beijing, China, October 2006, pp. 4647–4653.

[14] 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.

[15] F. Devernay and O. Faugeras, "Straight Lines Have to be Straight: Automatic Calibration and Removal of Distortion from Scenes of Structured Enviroments," *Machine Vision and Applications*, vol. 13, no. 1, pp. 14–24, August 2001.