

Why Having 10,000 Parameters in Your Camera Model is Better Than Twelve

Thomas Schöps¹

Viktor Larsson¹

Marc Pollefeys^{1,2}

Torsten Sattler³

¹Department of Computer Science, ETH Zürich ³Chalmers University of Technology

²Microsoft Mixed Reality & AI Zurich Lab

Abstract

Camera calibration is an essential first step in setting up 3D Computer Vision systems. Commonly used parametric camera models are limited to a few degrees of freedom and thus often do not optimally fit to complex real lens distortion. In contrast, generic camera models allow for very accurate calibration due to their flexibility. Despite this, they have seen little use in practice. In this paper, we argue that this should change. We propose a calibration pipeline for generic models that is fully automated, easy to use, and can act as a drop-in replacement for parametric calibration, with a focus on accuracy. We compare our results to parametric calibrations. Considering stereo depth estimation and camera pose estimation as examples, we show that the calibration error acts as a bias on the results. We thus argue that in contrast to current common practice, generic models should be preferred over parametric ones whenever possible. To facilitate this, we released our calibration pipeline at https://github.com/puzzlepai/nt/camera_calibration, making both easy-to-use and accurate camera calibration available to everyone.

1. Introduction

Geometric camera calibration is the process of determining where the light recorded by each pixel of a camera comes from. It is an essential prerequisite for 3D Computer Vision systems. Common parametric camera models allow for only a few degrees of freedom and are thus unlikely to optimally fit to complex real-world lens distortion (cf. Fig. 1). This can for example be aggravated by placing cameras behind windshields for autonomous driving [1]. However, accurate calibration is very important, since calibration errors affect all further computations. Even though the noise introduced by, for example, feature extraction in the final application is likely much larger than the error in the camera calibration, the latter can still be highly relevant since it may act as a bias that cannot be averaged out.

Generic camera models [15] relate pixels and their 3D observation lines resp. rays outside of the camera optics in a purely mathematical way, without offering a physical inter-

Figure 1. Residual distortion patterns of fitting two parametric camera models (left, center) and a generic model (right) to a mobile phone camera. While the generic model shows mostly random noise, parametric models show strong systematic modeling errors.

Figure 2. 2D sketch of the two generic camera models considered in this paper. (a) In image space (black rectangle), a grid of control points is defined that is aligned to the calibrated area (dashed pink rectangle) and extends beyond it by one cell. A point (red) is unprojected by B-Spline surface interpolation of the values stored for its surrounding 4x4 points (blue). Interpolation happens among directions (gray and blue arrows) starting from a projection center (black dot) for the central model (b), and among arbitrary lines (gray and blue arrows) for the non-central model (c).

pretation of the camera geometry. They densely associate pixels with observation lines or rays; in the extreme case, a separate line is stored for each pixel in the camera image. Due to their many degrees of freedom, they may fit all kinds of cameras, allowing to obtain accurate, bias-free calibrations. Fig. 2 shows the models considered in this paper.

Previous generic calibration approaches (cf. Sec. 2) have seen limited use in practice. On the one hand, this might be since there is no readily usable implementation for any existing approach. On the other hand, the community at large does not seem to be aware of the practical advantages of generic calibration approaches over parametric models.

Our contributions are thus: 1) We propose improvements to camera calibration, in particular to the calibration pattern and feature extraction, to increase accuracy. 2) We show the benefits of accurate generic calibration over parametric models, in particular on the examples of stereo depth and camera pose estimation. 3) We publish our easy-to-use calibration pipeline and generic camera models as open source.

2. Related Work

In this section, we present related work on calibration with generic camera models. We do not review calibration initialization, since we adopt [26] which works well for this.

Pattern design and feature detection. Traditionally, checkerboard [5, 6] and dot [19] patterns have been used for camera calibration. Feature detection in dot patterns is however susceptible to perspective and lens distortion [21]. Recent research includes more robust detectors for checkerboard patterns [11, 24], the use of ridge lines for higher robustness against defocus [10], and calibration with low-rank textures [48]. Ha et al. [16] propose the use of triangular patterns, which provide more gradient information for corner refinement than checkerboard patterns. Our proposed calibration pattern similarly increases the available gradients, while however allowing to vary the black/white segment count, enabling us to design better features than [16].

Non-central generic models. Grossberg and Nayar [15] first introduced a generic camera model that associates each pixel with a 3D observation line, defined by a line direction and a point on the line. This allows to model central cameras, i.e., cameras with a single unique center of projection, as well as non-central cameras. [26, 39] proposed a geometric calibration approach for the generic model from [15] that does not require known relative poses between images. [26] focus on initialization rather than a full calibration pipeline. Our approach extends [26] with an improved calibration pattern / detector and adds full bundle adjustment.

Central generic models. Non-central cameras may complicate applications, e.g., undistortion to a pinhole image is not possible without knowing the pixel depths [41]. Thus, models which constrain the calibration to be central were also proposed [1, 3, 13, 23]. For a central camera, all observation lines intersect in the center of projection, simplifying observation lines to observation rays / directions.

Per-pixel models vs. interpolation. Using a observation line / ray for each pixel [13, 15] provides maximum flexibility. However, this introduces an extreme number of parameters, making calibration harder. In particular, classical sparse calibration patterns do not provide enough measurements. Works using these models thus obtain dense matches using displays that can encode their pixel positions [2, 3, 13, 15], or interpolate between sparse features [26].

Since using printed patterns can sometimes be more practical than displays, and interpolating features causes inaccuracy [13], models with lower calibration data requirements have been proposed. These interpolate between sparsely stored observation lines resp. rays. E.g., [22] propose to interpolate arbitrarily placed control points with radial basis functions. Other works use regular grids for more convenient interpolation. [1] map from pixels to observation directions with a B-Spline surface. [32, 33] use two spline

surfaces to similarly also define a non-central model. In this work, we follow these approaches.

The above works are the most similar ones to ours regarding the camera models and calibration. Apart from our evaluation in real-world application contexts, we aim to achieve even more accurate results. Thus, our calibration process differs as follows: 1) We specifically design our calibration pattern and feature detection for accuracy (cf. Sec. 3.1, 3.2). 2) [1, 33] approximate the reprojection error in bundle adjustment. We avoid this approximation since, given Gaussian noise on the features, this will lead to better solutions. 3) [1, 33] assume planar calibration patterns which will be problematic for imperfect patterns. We optimize for the pattern geometries in bundle adjustment, accounting for real-world imperfections [37]. 4) We use denser control point grids than [1, 33], allowing us to observe and model interesting fine details (cf. Fig. 7).

Photogrammetry. The rational polynomial coefficient (RPC) model [20] maps 3D points to pixels via ratios of polynomials of their 3D coordinates. With 80 parameters, it is commonly used for generic camera modeling in aerial photogrammetry. In contrast to the above models, its parameters globally affect the calibration, making it harder to use more parameters. Further, this model works best only if all observed 3D points are in a known bounded region.

Evaluation and comparison. Dunne et al. [12] compare an early variant [38] of Ramalingam and Sturm’s series of works [25–28, 38, 39] with classical parametric calibration [40, 47]. They conclude that the generic approach works better for high-distortion cameras, but worse for low-to medium-distortion cameras. In contrast, Bergamasco et al. [2] conclude for their approach that even quasi-pinhole cameras benefit from non-central generic models. Our results also show that generic models generally perform better than typical parametric ones, and we in addition evaluate how this leads to practical advantages in applications.

3. Accurate Generic Camera Calibration

The first step in our pipeline is to record many photos or a video of one or multiple calibration patterns to obtain enough data for dense calibration (Sec. 3.1). We propose a pattern that enables very accurate feature detection. The next step is to detect the features in the images (Sec. 3.2). After deciding for the central or non-central camera model (Sec. 3.3), the camera is calibrated: First, a dense per-pixel initialization is obtained using [26]. Then, the final model is fitted to this and refined with bundle adjustment (Sec. 3.4).

All components build relatively closely on previous work, as indicated below; our contributions are the focus on accuracy in the whole pipeline, and using it to show the limitations of parametric models in detailed experiments. Note that our approach assumes that observation rays / lines vary smoothly among neighbor pixels, without discontinuities.

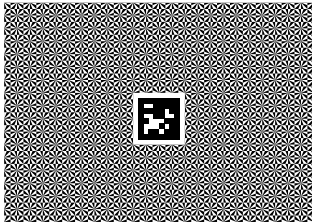


Figure 3. **Left:** Our (downsized) calibration pattern, allowing for unique localization using the AprilTag, and for very accurate feature detection using star-shaped feature points. Note that one should ideally adapt the density of the star squares to the resolution of the camera to be calibrated. **Right:** Repeating pattern elements for different star segment counts. Top to bottom and left to right: 4 (checkerboard), 8, 12, 16, 24, 32 segments.

3.1. Calibration Pattern & Data Collection

For data collection, we record images of a known calibration pattern. This allows for virtually outlier-free and accurate localization of feature points on the pattern. Thus, compared to using natural features, less data is required to average out errors. Furthermore, the known (near-planar) geometry of the pattern is helpful for initialization.

As mentioned in Sec. 2, dot patterns make it difficult for feature detection to be robust against distortion [21]. We thus use patterns based on intersecting lines, such as checkerboards. Checkerboards have several disadvantages however. First, there is little image information around each corner to locate it: Only the gradients of the two lines that intersect at the feature provide information. As shown in [16], using 3 instead of 2 lines improves accuracy. This raises the question whether the number of lines should be increased further. Second, checkerboard corners change their appearance strongly when viewed under different rotations. This may make feature detectors susceptible to yield differently biased results depending on the orientation of a feature, e.g., in the presence of chromatic aberration.

To address these shortcomings, we propose to use star-based patterns (cf. Siemens stars, e.g., [31]) as a generalization of checkerboards. Each feature in this type of pattern is the center of a star with a given number s of alternating black and white segments. For $s = 4$, the pattern corresponds to a checkerboard. For $s = 6$, the features resemble those of the deltille pattern [16] (while the feature arrangement differs from [16], however). We constrain the area of each star to a square and align these squares next to each other in a repeating pattern. Additional corner features arise at the boundaries of these squares, which we however ignore, since their segment counts are in general lower than that of the feature in the center. We also include an AprilTag [45] in the center of the pattern to facilitate its unambiguous localization (cf. [1, 16]). See Fig. 3 for an image of the pattern, and squares with different numbers of segments. The number of segments needs to balance the amount of gradient information provided and the ability for

the pattern to be resolved by the display or printing device and the camera; as justified in Sec. 4.1, we use 16 segments.

The pattern can be simply printed onto a sheet of paper or displayed on a computer monitor. If desired, multiple patterns can be used simultaneously, making it very easy to produce larger calibration geometries. Strict planarity is not required, since we later perform full bundle adjustment including the calibration patterns' geometries. However, we assume approximate planarity for initialization, and rigidity.

During data collection, we detect the features in real-time (cf. Sec. 3.2) and visualize the pixels at which features have been detected. This helps to provide detections in the whole image area. Image areas without detections either require regularization to fill in an estimated calibration, or need to be excluded from use. For global-shutter cameras, we record videos instead of images for faster recording.

3.2. Feature Extraction

Given an image of our 'star' calibration pattern (Fig. 3), we must accurately localize the star center features in the image. We first detect them approximately and then refine the results. For detection, we establish approximate local homographies between the image and the pattern, starting from the detected AprilTag corners. Detected features add additional matched points and thus allow to expand the detection area. For details, see the supplemental material. In the following, we only detail the refinement, which determines the final accuracy, as this is the focus of this paper.

The refinement step receives an approximate feature location as input and needs to determine the feature's exact subpixel location. To do so, we define a cost function based on symmetry (similar to the supplemental material of [35]), cf. Fig. 4: In pattern space, mirroring any point at a feature point must yield the same image intensity as the original point. This is generally applicable to symmetrical patterns.

We define a local window for feature refinement which must include sufficient gradients, but should not include too much lens distortion. The optimum size depends on factors such as the blur from out-of-focus imaging, internal image processing in the camera, and clarity of the calibration pattern. It should thus be suitably chosen for each situation; in this paper, we usually use 21×21 pixels. It is not an issue if the window covers multiple 'stars' since the pattern is symmetric beyond a single star. Within this window, we sample eight times as many random points as there are pixels in the window, in order to keep the variance due to random sampling low. The initial feature detection (cf. supp. PDF) provides a homography that locally maps between the pattern and image. With this, we transform all n random samples into pattern space, assuming the local window to be centered on the feature location. A cost function C_{sym} is then defined to compare points that are mirrored in pattern space:

Figure 4. Symmetry-based feature refinement: Both a sample and its mirrored position (orange circles) are transformed from pattern to image space with homography H . H is optimized to minimize the differences between sampling both resulting positions.

$$C_{\text{sym}}(H) = \sum_{i=1}^n \|I(H(s_i)) - I(H(-s_i))\|^2. \quad (1)$$

Here, H is the local homography estimate that brings pattern-space points into image space by homogeneous multiplication. For each feature, we define it such that the origin in pattern space corresponds to the feature location. s_i denotes the pattern-space location of sample i , and with the above origin definition, $-s_i$ mirrors the sample at the feature. I is the image, accessed with bilinear interpolation.

We optimize H with the Levenberg-Marquardt method to minimize C_{sym} . We fix the coefficient $H_{2,2}$ to 1 to obtain 8 remaining parameters to optimize, corresponding to the 8 degrees of freedom of the homography. After convergence, we obtain the estimated feature location as $(H_{0,2}, H_{1,2})^T$.

The sample randomization reduces issues with bilinear interpolation: For this type of interpolation, extrema of the interpolated values almost always appear at integer pixel locations. This also makes cost functions defined on a regular grid of bilinearly-interpolated pixel values likely to have extrema there, which would introduce an unjustified prior on the probable subpixel feature locations. Further, note that bilinear interpolation does not account for possible nonlinearities in the camera’s response function; however, these would be expected to only cause noise, not bias.

3.3. Camera Model

Accurate camera calibration requires a flexible model that avoids restricting the representable distortions. Storing a separate observation ray for each pixel, indicating where the observed light comes from, would be the most general model (assuming that a ray sufficiently approximates the origin directions). Such a model requires multiple feature observations for each pixel, or regularization, to be sufficiently constrained. Obtaining fully dense observations is very tedious with point features. It would be more feasible with dense approaches [17, 29, 30], which we consider out of scope of this paper, and it would be possible with displayed patterns [2, 3, 13, 15], which we do not want to require. We thus reduce the parameter count by storing observation rays in a regular grid in image space and interpolating between them (like [1, 32, 33]). This is appropriate for all cameras with smoothly varying observation directions.

A non-central model, while potentially more accurate, may complicate the final application; images in general cannot be undistorted to the pinhole model, and algorithms de-

signed for central cameras may need adaptation. We consider both a central and a non-central model (cf. Fig. 2).

Central camera model. For the central model, we store a unit-length observation direction at each grid point. For un-projecting a given image pixel, these directions are interpolated as 3D points using a cubic B-Spline [9] surface. The interpolated point is then re-normalized to obtain the observation direction. We also considered bicubic interpolation using Catmull-Rom splines [8], however, the resulting surfaces tend to contain small wiggles as artifacts.

Non-central camera model. When using the non-central model, each grid point stores both a unit-length direction and a 3D point p on the observation line. In un-projection, both points are interpolated with a cubic B-Spline surface, and the direction is re-normalized afterwards. The result is a line passing through the interpolated 3D point with the computed direction. Note that the interpolated lines may change if p is moved along the line. Since, in contrast to the directions, there is no obvious normalization possibility for points p , we keep this additional degree of freedom.

Projection. The presented camera models define how to un-project pixels from the image to directions respectively lines in closed form. For many applications and the later bundle adjustment step (cf. Sec. 3.4), the inverse is also required, i.e., projecting 3D points to pixels, which we find using an optimization process. Note that this is different from many parametric models, which instead define projection in closed form and may require an optimization process for un-projection if they are not directly invertible.

To project a 3D point, we initialize the projected position in the center of the calibrated image area. Then, similar to [33], we optimize it using the Levenberg-Marquardt method such that its un-projection matches the input point as closely as possible. Pixel positions are constrained to the calibrated area, and we accept the converged result only if the final cost is below a very small threshold. For speedup, if the same point was projected before with similar pose and intrinsics, the previous result can be used for initialization.

This approach worked for all tested cameras, as long as enough calibration data was recorded to constrain all grid parameters. For cameras where the procedure might run into local minima, one could search over all observation directions / lines of the camera for those which match the input point best [33]. This also helps if one needs to know all projections in cases where points may project to multiple pixels, which is possible with both of our camera models.

Performance. Tasks such as point (un)projection are low-level operations that may be performed very often in applications, thus their performance may be critical. We thus shortly discuss the performance of our camera models.

For central cameras, images may be ‘undistorted’ to a different camera model, usually the pinhole model. This

transformation can be cached for calibrated cameras; once a lookup table for performing it is computed, the choice of the original camera model has no influence on the run-time anymore. For high-field-of-view cameras, e.g., fisheye cameras, where undistortion to a pinhole model is impractical, one may use lookup tables from pixels to directions and vice versa. Thus, with an optimized implementation, there should be either zero or very little performance overhead when using generic models for central cameras.

For non-central cameras, image undistortion is not possible in general. Un-projection can be computed directly (and cached in a lookup table for the whole image). It should thus not be a performance concern. However, projection may be slow; ideally, one would first use a fast approximate method (such as an approximate parametric model or lookup table) and then perform a few iterations of optimization to get an accurate result. The performance of this may highly depend on the ability to quickly obtain good initial projection estimates for the concrete camera.

We think that given appropriate choice of grid resolution, the initial calibration should not take longer than 30 minutes up to sufficient accuracy on current consumer hardware.

Parameter choice. The grid resolution is the only parameter that must be set by the user. The smallest interesting cell size is similar to the size of the feature refinement window (cf. Sec. 3.2), since this window will generally ‘blur’ details with a kernel of this size. Since we use 21×21 px or larger windows for feature extraction, we use grid resolutions down to 10 px/cell, which we expect to leave almost no grid-based modeling error. If there is not enough data, the resolution should be limited to avoid overfitting.

3.4. Calibration

Given images with extracted features, and the chosen central or non-central camera model, the model must be calibrated. Our approach is to first initialize a per-pixel model on interpolated pattern matches using [26]. Then we fit the final model to this, discard the interpolated matches, and obtain the final result with bundle adjustment. See [26] and the supp. material for details. In the following, we focus on the refinement step that is responsible for the final accuracy.

Bundle Adjustment. Bundle adjustment jointly refines the camera model parameters, image poses (potentially within a fixed multi-camera rig), and the 3D locations of the pattern features. We optimize for the reprojection error, which is the standard cost function in bundle adjustment [43]:

$$C(\mathcal{C}, \mathcal{M}, \mathcal{T}, \mathcal{p}) = \sum_{c \in \mathcal{C}} \sum_{i \in \mathcal{I}_c} \sum_{o \in \mathcal{O}_i} (r_{c,i,o}^T r_{c,i,o}) \quad (2)$$

$$r_{c,i,o} = \mathcal{C}_c(\mathcal{M}_c \mathcal{T}_i \mathcal{p}_o) - \mathcal{d}_{i,o}$$

Here, \mathcal{C} denotes the set of all cameras, \mathcal{I}_c the set of all images taken by camera c , and \mathcal{O}_i the feature observations in image i . \mathcal{p}_o is the 3D pattern point corresponding to obser-

vation o , and $\mathcal{d}_{i,o}$ the 2D detection of this point in image i . \mathcal{T}_i is the pose of image i which transforms global 3D points into the local rig frame, and \mathcal{M}_c transforms points within the local rig frame into camera c ’s frame. \mathcal{C}_c then projects the local point to camera c ’s image using the current estimate of its calibration. $\|r_{c,i,o}\|^2$ is a loss function on the squared residual; we use the robust Huber loss with parameter 1.

As is common, we optimize cost C with the Levenberg-Marquardt method, and use local updates for orientations. We also use local updates (x_1, x_2) for the directions within the camera model grids: We compute two arbitrary, perpendicular tangents t_1, t_2 to each direction g and update it by adding a multiple of each tangent vector and then re-normalizing: $\frac{g + x_1 t_1 + x_2 t_2}{\|g + x_1 t_1 + x_2 t_2\|}$. Each grid point thus has a 2D update in the central model and a 5D update in the non-central one (two for the direction, and three for a 3D point on the line, cf. Sec. 3.3). The projection involves an optimization process and the Inverse Function Theorem is not directly applicable. Thus, we use finite differences to compute the corresponding parts of the Jacobian.

The optimization process in our setting has more dimensions of Gauge freedom than for typical Bundle Adjustment problems, which we discuss in the supplemental material. We experimented with Gauge fixing, but did not notice an advantage to explicitly fixing the Gauge directions; the addition of the Levenberg-Marquardt diagonal should already make the Hessian approximation invertible.

The Levenberg-Marquardt method compares the costs of different states to judge whether it makes progress. However, we cannot always compute all residuals for each state. During optimization, the projections of 3D points may enter and leave the calibrated image area, and since the camera model is only defined within this area, residuals cannot be computed for points that do not project into it. If a residual is defined in one state but not in another, how should the states be compared in a fair way? A naive solution would be to assign a constant value (e.g., zero) to a residual if it is invalid. This causes state updates that make residuals turn invalid to be overrated, while updates that make residuals turn valid will be underrated. As a result, the optimization could stall. Instead, we propose to compare states by summing the costs only for residuals which are valid in both states. This way, cost comparisons are always fair; however, some residuals may not be taken into account. Theoretically, this may lead to oscillation. We however did not observe this in practice, and we believe that if it happens it will most likely be very close to the optimum, since otherwise the remaining residuals likely outweigh the few which change validity. In such a case, it then seems safe to stop the optimization.

4. Evaluation

Tab. 1 lists the cameras used for evaluation. The camera labels from this table will be used throughout the evaluation.

Label	Resolution	Field-of-view (FOV)	Description
D435-C	1920 × 1080	ca. 70 × 42	Color camera of an Intel D435
D435-I	1280 × 800	ca. 90 × 64	Infrared camera of an Intel D435
SC-C	640 × 480	ca. 71 × 56	Color camera of a Structure Core
SC-I	1216 × 928	ca. 57 × 45	Infrared camera of a Structure Core

Table 1. Specifications of the cameras used in the evaluation (more cameras are evaluated in the supplemental material). FOV is measured horizontally and vertically at the center of the image.

Figure 5. Median reprojection errors (y-axis) for calibrating cameras D435-C and D435-I with patterns having different numbers of star segments (x-axis). The feature refinement window was 31×31 pixels for D435-C and 21×21 pixels for D435-I. The Deltille results were obtained with the feature refinement from [16].

We evaluate the generic models against two parametric ones, both having 12 parameters, which is a high number for parametric models. The first is the model implemented in OpenCV [6] using all distortion terms. The second is the Thin-Prism Fisheye model [46] with 3 radial distortion terms, which was used in a stereo benchmark [36]. We also consider a “Central Radial” model (similar to [7, 18, 42]) based on the OpenCV model, adding the two thin-prism parameters from the Thin-Prism Fisheye model and replacing the radial term with a spline with many control points. With this, we evaluate how much improvement is obtained by better radial distortion modeling only. Note that unfortunately, no implementations of complete generic calibration pipelines seemed to be available at the time of writing. This makes it hard to compare to other generic calibration pipelines; we released our approach as open source to change this. In addition, since we aim to obtain the most accurate results possible, and since we avoid synthetic experiments as they often do not reflect realistic conditions, there is no ground truth to evaluate against. However, our main interest is in comparing to the commonly used parametric models to show why they should (if possible) be avoided.

4.1. Calibration Pattern Evaluation

We validate our choice of pattern (cf. Sec. 3.1) by varying the number of star segments from 4 to 32 (cf. Fig. 3). For the 4-segment checkerboard and 6-segment ‘deltille’ [16] patterns, we also compare against the feature refinement from [16]. For each pattern variant, we record calibration images with the same camera from the same poses. We do this by putting the camera on a tripod and showing the pattern on a monitor. For each tripod pose that we use, we cycle through all evaluated patterns on the monitor to record a set of images with equal pose. Since not all features are detected in all patterns, for fairness we only keep those feature detections which succeed for all pattern variants.

Since there is no ground truth for feature detections, we compare different patterns via the achieved reprojection er-

Figure 6. Median reproj. error (y-axis) for calibrating the cameras on the x-axis with different feature refinement schemes (colors). For SC-C, cornerSubPix() results were too inconsistent.

rors. We calibrate each set of images of a single pattern separately and compute its median reprojection error. These results are plotted in Fig. 5. Increasing the number of segments starting from 4, the accuracy is expected to improve first (since more gradients become available for feature refinement) and then worsen (since the monitor and camera cannot resolve the pattern anymore). Both plots follow this expectation, with the best number of segments being 12 resp. 20. The experiment shows that neither the commonly used checkerboard pattern nor the ‘deltille’ pattern [16] is optimal for either camera (given our feature refinement). For this paper, we thus default to 16 segments as a good mean value. The results of [16] have higher error than ours for both the checkerboard and ‘deltille’ pattern.

4.2. Feature Refinement Evaluation

We compare several variants of our feature refinement (cf. Sec. 3.2): i) The original version of Eq. (1), and versions where we replace the raw intensity values by ii) gradient magnitudes, or iii) gradients (2-vectors). In addition, we evaluate OpenCV’s [6] cornerSubPix() function, which implements [14]. In all cases, the initial feature positions for refinement are given by our feature detection scheme. For each camera, we take one calibration dataset, apply every feature refinement scheme on it, and compare the achieved median reprojection errors. Similarly to the previous experiment, we only use features that are found by all methods. The results are plotted in Fig. 6. Intensities and X/Y gradients give the best results, with X/Y gradients performing slightly better for the monochrome cameras and intensities performing slightly better for the color cameras.

4.3. Validation of the Generic Model

We validate that the generic models we use (cf. Sec. 3.3) can calibrate cameras very accurately by verifying that they achieve bias-free calibrations: The directions of the final reprojection errors should be random rather than having the same direction in parts of the image, which would indicate an inability of the model to fit the actual distortion in these areas. Fig. 7 shows these directions for different cameras, calibrated with each tested model. We also list the median reprojection errors, both on the calibration data and on a test set that was not used for calibration. The latter is used to confirm that the models do not overfit. As a metric of biasedness, we compute the KL-Divergence between the 2D

	OpenCV (12 parameters)	Thin-Prism Fisheye (12 parameters)	Central Radial (258 parameters)	Central Generic ca. 30 px/cell	Central Generic ca. 20 px/cell	Central Generic ca. 10 px/cell	Noncentral Generic ca. 20 px/cell
D435-C (968 images)							
Errors ¹	0.092 / 0.091 / 0.748	0.163 / 0.161 / 1.379	0.068 / 0.070 / 0.968	0.030 / 0.039 / 0.264	0.030 / 0.039 / 0.265	0.029 / 0.040 / 0.252	0.024 / 0.032 / 0.184
D435-I (1347 images)							
Errors ¹	0.042 / 0.036 / 0.488	0.032 / 0.026 / 0.365	0.042 / 0.037 / 0.490	0.023 / 0.018 / 0.199	0.023 / 0.018 / 0.198	0.023 / 0.018 / 0.189	0.022 / 0.017 / 0.179
SC-C (1849 images)							
Errors ¹	0.083 / 0.085 / 0.217	0.083 / 0.084 / 0.215	0.082 / 0.084 / 0.200	0.069 / 0.072 / 0.055	0.069 / 0.072 / 0.054	0.068 / 0.072 / 0.053	0.065 / 0.069 / 0.040
SC-I (2434 images)							
Errors ¹	0.069 / 0.064 / 0.589	0.053 / 0.046 / 0.440	0.069 / 0.064 / 0.585	0.035 / 0.030 / 0.133	0.035 / 0.030 / 0.139	0.034 / 0.030 / 0.137	0.030 / 0.026 / 0.120

Figure 7. Directions (see legend on the left) of all reprojection errors for calibrating the camera given by the row with the model given by the column. Each pixel shows the direction of the closest reprojection error from all images. Ideally, the result is free from any systematic pattern. Patterns indicate biased results arising from inability to model the true camera geometry. Parameter counts for generic models are given in the images. ¹Median training error [px] / test error [px] / biasedness.

normal distribution, and the empirical distribution of reprojection error vectors (scaled to have the same mean error norm), in each cell of a regular 50×50 grid placed on the image. We report the median value over these cells in Fig. 7.

The generic models achieve lower errors than the parametric ones throughout, while showing hardly any signs of overfitting. This is expected, since – given enough calibration images – the whole image domain can be covered with training data, thus there will be no ‘unknown’ samples during test time. Interestingly, the non-central model consistently performs best for all cameras in every metric, despite all of the cameras being standard near-pinhole cameras.

All parametric models show strong bias patterns in the error directions. For some cameras, the generic models also show high-frequency patterns with lower grid resolutions that disappear with higher resolution. These would be very hard to fit with any parametric model. The central radial model only improves over the two parametric models for one camera, showing that improved radial distortion modeling alone is often not sufficient for significant improvement.

4.4. Comparison of Different Models

We now take a closer look at the differences between accurate calibrations and calibrations obtained with typical parametric models. We fit the Thin-Prism Fisheye model to our calibrations, optimizing the model parameters to minimize the two models’ deviations in the observation directions per-pixel. At the same time, we optimize for a 3D rotation applied to the observation directions of one model, since consistent rotation of all image poses for a camera can

D435-C D435-I SC-C SC-I

Figure 8. Differences between calibrations with the central-generic model and fitted Thin-Prism Fisheye calibrations, measured as reprojection errors. **Top:** Medium gray corresponds to zero error, while saturated colors as in Fig. 7 correspond to 0.2 pixels difference. **Bottom:** Alternative visualization showing the error magnitude only, with black for zero error and white for 0.2 pixels error.

be viewed as part of the intrinsic calibration. After convergence, we visualize the remaining differences in the observation directions. While these visualizations will naturally be similar to those of Fig. 7 given our model is very accurate, we can avoid showing the feature detection noise here. Here, we visualize both direction and magnitude of the differences, while Fig. 7 only visualizes directions. The results are shown in Fig. 8, and confirm that the models differ in ways that would be difficult to model with standard parametric models. Depending on the camera and image area, the reprojection differences are commonly up to 0.2 pixels, or even higher for the high-resolution camera D435-C.

4.5. Example Application: Stereo Depth Estimation

So far, we showed that generic models yield better calibrations than common parametric ones. However, the differences might appear small, and it might thus be unclear

Figure 9. Distances (black: 0cm, white: 1cm) between corresponding points estimated by dense stereo with a generic and a parametric calibration, at roughly 2 meters depth.

how valuable they are in practice. Thus, we now look at the role of small calibration errors in example applications.

Concretely, we consider dense depth estimation for the Intel D435 and Occipital Structure Core active stereo cameras. These devices contain infrared camera pairs with a relatively small baseline, as well as an infrared projector that provides texture for stereo matching. The projector behaves like an external light and thus does not need to be calibrated; only the calibration of the stereo cameras is relevant.

Based on the previous experiments, we make the conservative assumption that the calibration error for parametric models will be at least 0.05 pixels in many parts of the image. Errors in both stereo images may add up or cancel each other out depending on their directions. A reasonable assumption is that the calibration error will lead to a disparity error of similar magnitude. Note that for typical stereo systems, the stereo matching error for easy-to-match surfaces may be assumed to be as low as 0.1 pixels [35]; in this case, the calibration error may even come close to the level of noise. The well-known relation between disparity x and depth d is: $d = \frac{bf}{x}$, with baseline b and focal length f . Let us consider $b = 5\text{cm}$ and $f = 650\text{px}$ (roughly matching the D435). For $d = 2\text{m}$ for example, a disparity error of $\pm 0.05\text{px}$ results in a depth error of about 0.6cm. This error grows quadratically with depth, and since it stays constant over time, it acts as a bias that will not easily average out.

For empirical validation, we calibrate the stereo pairs of a D435 and a Structure Core device with both the central-generic and the Thin-Prism Fisheye model (which fits the D435-I and SC-I cameras better than the OpenCV model, see Fig. 7). With each device, we recorded a stereo image of a roughly planar wall in approx. 2m distance and estimated a depth image for the left camera with both calibrations. Standard PatchMatch Stereo [4] with Zero-Mean Normalized Cross Correlation costs works well given the actively projected texture. The resulting point clouds were aligned with a similarity transform with the Umeyama method [44], since the different calibrations may introduce scale and orientation differences. Fig. 9 shows the distances of corresponding points in the aligned clouds. Depending on the image area, the error is often about half a centimeter, and goes up to more than 1 cm for both cameras. This matches the theoretical result from above well and shows that one

should avoid such a bias for accurate results.

4.6 Example Application: Camera Pose Estimation

To provide a broader picture, we also consider camera pose estimation as an application. For this experiment, we treat the central-generic calibration as ground truth and sample 15 random pixel locations in the image. We unproject each pixel to a random distance to the camera from 1.5 to 2.5 meters. Then we change to the Thin-Prism Fisheye model and localize the camera with the 2D-3D correspondences defined above. The median error in the estimated camera centers is 2.15 mm for D435-C, 0.25 mm for D435-I, 1.80 mm for SC-C, and 0.76 mm for SC-I.

Such errors may accumulate during visual odometry or SLAM. To test this, we use Colmap [34] on several videos and bundle-adjust the resulting sparse reconstructions both with our Thin-Prism-Fisheye and non-central generic calibrations. For each reconstruction pair, we align the scale and the initial camera poses of the video, and compute the resulting relative translation error at the final image compared to the trajectory length. For camera D435-I, we obtain $1.3\% \pm 0.3\%$ error, while for SC-C, we get $5.6\% \pm 2.2\%$. These errors strongly depend on the camera, reconstruction system, scene, and trajectory, so our results only represent examples. However, they clearly show that even small calibration improvements can be significant in practice.

5. Conclusion

We proposed a generic camera calibration pipeline which focuses on accuracy while being easy to use. It achieves virtually bias-free results in contrast to using parametric models; for all tested cameras, the non-central generic model performs best. We also showed that even small calibration improvements can be valuable in practice, since they avoid biases that may be hard to average out.

Thus, we believe that generic models should replace parametric ones as the default solution for camera calibration. If a central model is used, this might not even introduce a performance penalty, since the runtime performance of image undistortion via lookup does not depend on the original model. We facilitate the use of generic models by releasing our calibration pipeline as open source. However, generic models might not be suitable for all use cases, in particular if the performance of projection to distorted images is crucial, if self-calibration is required, or if not enough data for dense calibration is available.

Acknowledgements. Thomas Schöps was partially supported by a Google PhD Fellowship. Viktor Larsson was supported by the ETH Zurich Postdoctoral Fellowship program. This work was supported by the Swedish Foundation for Strategic Research (Semantic Mapping and Visual Navigation for Smart Robots).

References

- [1] Johannes Beck and Christoph Stiller. Generalized B-spline camera model. In *Intelligent Vehicles Symposium*, 2018. 1, 2, 3, 4
- [2] Filippo Bergamasco, Andrea Albarelli, Emanuele Rodolá, and Andrea Torsello. Can a fully unconstrained imaging model be applied effectively to central cameras? In *CVPR*, 2013. 2, 4
- [3] Filippo Bergamasco, Luca Cosmo, Andrea Gasparetto, Andrea Albarelli, and Andrea Torsello. Parameter-free lens distortion calibration of central cameras. In *ICCV*, 2017. 2, 4
- [4] Michael Bleyer, Christoph Rhemann, and Carsten Rother. PatchMatch stereo - stereo matching with slanted support windows. In *BMVC*, 2011. 8
- [5] Jean-Yves Bouguet. Camera calibration toolbox for Matlab, 2004. 2
- [6] Gary Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000. 2, 6
- [7] Federico Camposeco, Torsten Sattler, and Marc Pollefeys. Non-parametric structure-based calibration of radially symmetric cameras. In *ICCV*, 2015. 6
- [8] Edwin Catmull and Raphael Rom. A class of local interpolating splines. In *Computer aided geometric design*, pages 317–326. Elsevier, 1974. 4
- [9] Carl De Boor. A practical guide to splines, volume 27. Springer-Verlag, 1978. 4
- [10] Wendong Ding, Xilong Liu, De Xu, Dapeng Zhang, and Zhengtao Zhang. A robust detection method of control points for calibration and measurement with defocused images. *IEEE Transactions on Instrumentation and Measurement*, 66(10):2725–2735, 2017. 2
- [11] Alexander Duda and Udo Frese. Accurate detection and localization of checkerboard corners for calibration. In *BMVC*, 2018. 2
- [12] Aubrey K. Dunne, John Mallon, and Paul F. Whelan. A comparison of new generic camera calibration with the standard parametric approach. In *IAPR Conference on Machine Vision Applications*, 2007. 2
- [13] Aubrey K. Dunne, John Mallon, and Paul F. Whelan. Efficient generic calibration method for general cameras with single centre of projection. *CVIU*, 114(2):220–233, 2010. 2, 4
- [14] Wolfgang Förstner and Eberhard Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *ISPRS intercommission conference on fast processing of photogrammetric data*, 1987. 6
- [15] Michael D. Grossberg and Shree K. Nayar. A general imaging model and a method for finding its parameters. In *ICCV*, 2001. 1, 2, 4
- [16] Hyowon Ha, Michal Perdoch, Hatem Alismail, In So Kweon, and Yaser Sheikh. Deltile grids for geometric camera calibration. In *ICCV*, 2017. 2, 3, 6
- [17] Morten Hannemose, Jakob Wilm, and Jeppe Revall Frisvad. Superaccurate camera calibration via inverse rendering. In *Modeling Aspects in Optical Metrology VII*, volume 11057, page 1105717. International Society for Optics and Photonics, 2019. 4
- [18] Richard Hartley and Sing Bing Kang. Parameter-free radial distortion correction with center of distortion estimation. *PAMI*, 29(8):1309–1321, 2007. 6
- [19] Janne Heikkila. Geometric camera calibration using circular control points. *PAMI*, 22(10):1066–1077, 2000. 2
- [20] Yong Hu, Vincent Tao, and Arie Croitoru. Understanding the rational function model: methods and applications. In *International archives of photogrammetry and remote sensing*, 20(6), 2004. 2
- [21] John Mallon and Paul F. Whelan. Which pattern? Biasing aspects of planar calibration patterns and detection methods. *Pattern recognition letters*, 28(8):921–930, 2007. 2, 3
- [22] Pedro Miraldo and Helder Araujo. Calibration of smooth camera models. *PAMI*, 35(9):2091–2103, 2012. 2
- [23] David Nister, Henrik Stewenius, and Etienne Grossmann. Non-parametric self-calibration. In *ICCV*, 2005. 2
- [24] Simon Placht, Peter Fürsattel, Etienne Assoumou Mengue, Hannes Hofmann, Christian Schaller, Michael Balda, and Elli Angelopoulou. Rochade: Robust checkerboard advanced detection for camera calibration. In *ECCV*, 2014. 2
- [25] Srikumar Ramalingam and Peter Sturm. Minimal solutions for generic imaging models. In *CVPR*, 2008. 2
- [26] Srikumar Ramalingam and Peter Sturm. A unifying model for camera calibration. *PAMI*, 39(7):1309–1319, 2016. 2, 5
- [27] Srikumar Ramalingam, Peter Sturm, and Suresh K. Lodha. Theory and experiments towards complete generic calibration. In *Rapport de Recherche 5562*, 2005. 2
- [28] Srikumar Ramalingam, Peter Sturm, and Suresh K. Lodha. Towards complete generic camera calibration. In *CVPR*, 2005. 2
- [29] Joern Rehder, Janosch Nikolic, Thomas Schneider, and Roland Siegwart. A direct formulation for camera calibration. In *ICRA*, 2017. 4
- [30] Joern Rehder and Roland Siegwart. Camera/IMU calibration revisited. *IEEE Sensors Journal*, 17(11):3257–3268, 2017. 4
- [31] Ralf Reulke, Susanne Becker, Norbert Haala, and Udo Tempelmann. Determination and improvement of spatial resolution of the CCD-line-scanner system ADS40. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(2):81 – 90, 2006. 3
- [32] Dennis Rosebrock and Friedrich M Wahl. Complete generic camera calibration and modeling using spline surfaces. In *ACCV*, 2012. 2, 4
- [33] Dennis Rosebrock and Friedrich M. Wahl. Generic camera calibration and modeling using spline surfaces. In *Intelligent Vehicles Symposium*, 2012. 2, 4
- [34] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *CVPR*, 2016. 8
- [35] Thomas Schöps, Torsten Sattler, and Marc Pollefeys. BAD SLAM: Bundle adjusted direct RGB-D SLAM. In *CVPR*, 2019. 3, 8

- [36] Thomas Schöps, Johannes L. Schönberger, Silvano Galliani, Torsten Sattler, Konrad Schindler, Marc Pollefeys, and Andreas Geiger. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 6
- [37] K. H. Strobl and G. Hirzinger. More Accurate Pinhole Camera Calibration with Imperfect Planar Target. In *ICCV Workshops*, 2011. 2
- [38] Peter Sturm and Srikumar Ramalingam. A generic calibration concept: Theory and algorithms. In *Rapport de Recherche 5058*, 2003. 2
- [39] Peter Sturm and Srikumar Ramalingam. A generic concept for camera calibration. In *ECCV*, 2004. 2
- [40] Peter F. Sturm and Stephen J. Maybank. On plane-based camera calibration: A general algorithm, singularities, applications. In *CVPR*, 1999. 2
- [41] Rahul Swaminathan, Michael D. Grossberg, and Shree K. Nayar. A Perspective on Distortions. In *CVPR*, 2003. 2
- [42] Jean-Philippe Tardif, Peter Sturm, Martin Trudeau, and Sebastien Roy. Calibration of cameras with radially symmetric distortion. *PAMI*, 31(9):1552–1566, 2008. 6
- [43] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon. Bundle adjustment – a modern synthesis. In *International workshop on vision algorithms*, pages 298–372. Springer, 1999. 5
- [44] Shinji Umeyama. Least-squares estimation of transformation parameters between two point patterns. *PAMI*, 13(4):376–380, 1991. 8
- [45] John Wang and Edwin Olson. AprilTag 2: Efficient and robust fiducial detection. In *IROS*, October 2016. 3
- [46] Juyang Weng, Paul Cohen, and Marc Herniou. Camera calibration with distortion models and accuracy evaluation. *PAMI*, 14(10):965–980, 1992. 6
- [47] Zhengyou Zhang. A flexible new technique for camera calibration. *PAMI*, 22, 2000. 2
- [48] Zhengdong Zhang, Yasuyuki Matsushita, and Yi Ma. Camera calibration with lens distortion from low-rank textures. In *CVPR*, 2011. 2