Calibrating and Centering Quasi-Central Catadioptric Cameras

Miriam Schönbein[1], Tobias Strauß[1] and Andreas Geiger[2]

***Abstract*— Non-central catadioptric models are able to cope with irregular camera setups and inaccuracies in the manufacturing process but are computationally demanding and thus not suitable for robotic applications. On the other hand, calibrating a quasi-central (almost central) system with a central model introduces errors due to a wrong relationship between the viewing ray orientations and the pixels on the image sensor.**

In this paper, we propose a central approximation to quasi-central catadioptric camera systems that is both accurate and efficient. We observe that the distance to points in 3D is typically large compared to deviations from the single viewpoint. Thus, we first calibrate the system using a state-of-the-art non-central camera model. Next, we show that by remapping the observations we are able to match the orientation of the viewing rays of a much simpler single viewpoint model with the true ray orientations. While our approximation is general and applicable to all quasi-central camera systems, we focus on one of the most common cases in practice: hypercatadioptric cameras. We compare our model to a variety of baselines in synthetic and real localization and motion estimation experiments. We show that by using the proposed model we are able to achieve near non-central accuracy while obtaining speed-ups of more than three orders of magnitude compared to state-of-the-art non-central models.

I. Introduction

Recently, catadioptric cameras which combine the principles of refraction and reflection in one single optical systems have gained popularity in robotics. They are able to establish a 360° field of view with a very flexible geometry as the shape of the reflecting surface is a powerful design factor. Furthermore, efficient central projection models exist that allow for computing the point of reflection by intersecting a line with a quadric in case the single viewpoint (SVP) condition is fulfilled. Also, manufacturing costs are low as catadioptric cameras merely require combining a classical perspective camera with a mirror coated surface. Applications include robotic perception [1], [2] as well as driver assistance systems [3], [4], amongst others.

Catadioptric systems are usually designed to closely satisfy the SVP condition, i.e., all light rays are assumed to intersect at a single viewpoint. While leading to simple projection models [5], [6], the SVP assumption is often violated in practice [7], [8] (in the following we call such cameras 'quasi-central'). The reason for this is that off-the-shelve catadioptric cameras often use varifocal lenses where the viewpoint depends non-linearly on the focus and the focal length. Furthermore, precisely aligning the perspective camera center with the optical axis of the mirror is hard in practice. While a number of non-central projection models have been proposed [9], [10] they are slow as they either require non-linear optimization [9] or finding the solution of a complex root-finding problem [10]. This prevents them from being applied whenever real-time algorithms are required and computational resources are scarce.

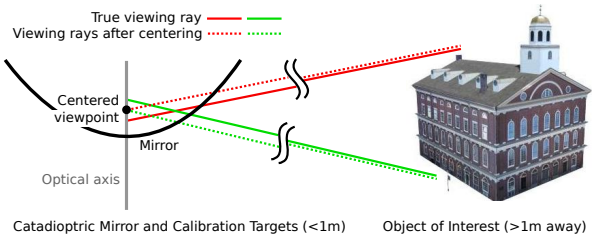


Fig. 1. **Illustration of Centered Camera Model.** This figure illustrates the true viewing rays emanating from the optical system (bold) as well as the rays of our centered approximation (dashed). Note that the distance to the objects of interest is usually very large compared to the deviation of the rays from the centered single viewpoint. Thus getting the ray orientation right is more important than considering translational errors.

In this paper, we take advantage of the fact that the distance to points in 3D is often large compared to the deviations from the SVP as illustrated in Fig. 1. This leads us to the conclusion that getting the orientation of the viewing rays right is more important than considering the translational deviation from the SVP correctly. Unfortunately, calibrating a quasi-central catadioptric camera using traditional SVP models [11], [12] introduces a bias in the orientation of the viewing rays due to the fact that the calibration patterns are presented in the vincinity of the camera.

Instead of using a central camera model for calibration, we propose to first calibrate the camera using a state-of-the-art non-central model [10] which we extend by a perspective camera and an appropriate distortion model in order to obtain the viewing ray orientations. Next, we remap the observations in a way such that the viewing ray orientations of a much simpler single viewpoint model coincide with those of the more accurate non-central model as illustrated in Fig. 1. This leads to a mapping where points at infinity are projected to the same pixels as in the non-central model and approximation accuracy gracefully degrades in the immediate vicinity of the camera center.

In our experiments we show that our model combines the efficiency of a single viewpoint model with the accuracy of a full (but slow) non-central model in tasks such as localization or motion estimation. Towards this goal, we construct a dataset using two catadioptric cameras and 17 landmarks. For

[1]Miriam Schönbein and Tobias Strauß are with the Institute of Measurement and Control Systems, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany. {strauss,miriam.schoenbein}@kit.edu

[2]Andreas Geiger is with Max Planck Institute for Intelligent Systems, Perceiving Systems Department, 72076 Tübingen, Germany. andreas.geiger@tue.mpg.de

performance evaluation we consider the 3D displacement in challenging monocular and stereo localization experiments using triplets of landmarks. We believe that these efforts are necessary for a fair comparison and to provide novel insights into the properties of the models and their parameters. Amongst others, we compare against the calibration toolboxes of Mei et al. [11] and Scaramuzza et al. [12] which are widely used in the robotics community. We also investigate the approximation accuracy when deviating from the single viewpoint in axial and radial direction as well as in terms of the distance to the points in 3D. Furthermore, we show that our *centered* model results in significant speed-ups which are important for real-time applications as in robotics or driver assistance systems. Our data and code in form of an easy-to-use calibration toolbox, including a fully automatic corner and checkerboard detector, is online available[1].

II. Related Work

Recently, a taxonomy for omnidirectional camera calibration methods has been proposed by Puig et al. [7]. They classify calibration methods into five catagories: Line-based calibration [13], [14], [15], 2D pattern calibration [12], [11], 3D pattern calibration [16], self-calibration [9] and polarization imaging [17]. As 2D patterns are easy to employ and constrain the problem sufficiently well, we focus on planar checkerboards as calibration targets in this paper.

The projection models that have been developed for omnidirectional cameras can be grouped into central and non-central models. The central case has been intensely researched over the last decades [5], [18], [14], [19], [20], [16], [15]. For instance, Geyer et al. [6] propose the sphere camera model which unifies all central catadioptric models. To account for the properties of real lenses, Mei et al. [11] add a perspective camera with distortions and provide their implementation[2]. Another widely used calibration toolbox[3] has been developed by Scaramuzza et al. [12] which describes the imaging function using a polynomial.

Unfortunatley, the SVP condition is hard to fulfill in practice [9] and even small deviations can impact performance when applying the model to accuracy sensitive tasks such as 3D reconstruction, localization or motion estimation. Furthermore, permitting more flexible arrangements allows to compromise image characteristics such as field of view with spatial resolution [8]. Consequently, a number of models for non-central catadioptric systems have been proposed. Early approaches [21], [9], [22], [23] make use of computationally expensive non-linear optimization and require an initial estimate of the pixel coordinates. Goncalves et al. [24] increase efficiency by reducing the optimization to a 1D search problem, but still require ~ 200 seconds for projecting 10.000 3D points to the image plane.

Recently, Agrawal et al. [10] derived an analytical forward projection (AFP) which requires solving an 8th degree polynomial equation. While the model has been applied in a bundle adjustment context with known intrinsics, no analysis in the context of calibrating the intrinsics and extrinsics jointly has been carried out yet. In this paper, we combine the AFP model with a perspective camera including lens distortions and derive an approximation that is more than three orders of magnitude faster (i.e., five orders of magnitude with respect to [24]), yet achieves similar accuracy in our experiments. Our results also reveal which parameters are important for modeling axial as well as lateral deviations from the single viewpoint, insights which have been missing in existing comparisons [7]. Furthermore, we investigate the impact of parameter initialization and suggest modifications to [12], resulting in improved convergence.

[1] http://www.mrt.kit.edu/software/
[2] http://www.robots.ox.ac.uk/cmei/Toolbox.html
[3] https://sites.google.com/site/scarabotix/ocamcalib-toolbox

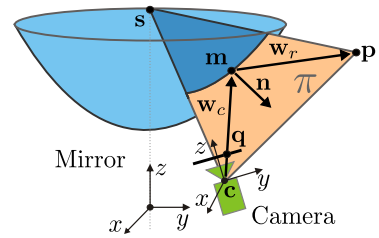


Fig. 2. **Geometric Projection Model with Parameters.** This figure shows the setup of the full non-central model which we use for obtaining the ray orientations in our centered approximation.

III. Projection Model

We start with a description of the geometric model [10] which we use as *base model* to obtain the ray orientations for the proposed centered model. While this specific choice of base model makes the assumption of a quadric mirror, our centered approximation (Section III-B) is more general and requires estimated light rays as the only input.

A. Geometric Model for Quadric Mirrors

Let us assume a quadric mirror with parameters A, B and C that can be described as

$$x^2 + y^2 + Az^2 + Bz - C = 0 \tag{1}$$

Further, let $\mathbf{p}$ denote the 3D world point, $\mathbf{m}$ a point on the mirror surface and $\mathbf{c}$ the perspective camera center (all $\in \mathbb{R}^3$) as illustrated in Fig. 2. Let the indices m, r and c denote the mirror, the rotated mirror and the camera coordinate systems, respectively. Using this notation, our geometric projection model maps a 3D world point $\mathbf{p}$ via the point of reflection on the mirror $\mathbf{m}$ to a pixel $(u,v)^\mathsf{T}$ in the image. Recently, Agrawal et al. [10] showed that the point of reflection $\mathbf{m}$ can be obtained analytically from the 3D world point $\mathbf{p}$ and the camera center $\mathbf{c}$. In the following, we briefly repeat their most important results and extend the model to include a perspective camera with distortions.

To reduce the order of the resulting polynomial forward projection equation, the world point $\mathbf{p}_m = (x_m, y_m, z_m)^\mathsf{T}$ and the camera location $\mathbf{c}_m = (r_m, s_m, t_m)^\mathsf{T}$ are rotated

$$\begin{aligned}\mathbf{p}_r &= \mathbf{R}_R\ \mathbf{p}_m \\ \mathbf{c}_r &= \mathbf{R}_R\ \mathbf{c}_m\end{aligned} \qquad \mathbf{R}_R = \begin{bmatrix} t\ (s_m+\varepsilon) & -t\ r_m & 0 \\ t\ r_m & t\ (s_m+\varepsilon) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

such that the rotated camera $\mathbf{c}_r = (0, s_r, t_r)^\mathsf{T}$ aligns with the y/z plane, where $t = 1/\sqrt{r_m^2 + (s_m + \varepsilon)^2}$. Note that in contrast to [10] we have introduced a small positive scalar ε which regularizes $\mathbf{R}_R$ against the identity matrix and prevents singularities in case r_m and s_m are both small.

Let π denote the plane of reflection on which $\mathbf{p}_r$, $\mathbf{c}_r$ and the (rotated) point of reflection $\mathbf{m}_r = (x_r, y_r, z_r)^\mathsf{T}$ are located. π can be represented by $\mathbf{p}_r$, $\mathbf{c}_r$ as well as the intersection of π with the z axis of the mirror coordinate system, which is given as $\mathbf{s} = (0, 0, z_r - Az_r - B/2)^\mathsf{T}$. As $(\mathbf{m}_r - \mathbf{s})$ is orthogonal to the normal of π, we have

$$(\mathbf{m}_r - \mathbf{s})^\mathsf{T} \cdot [(\mathbf{p}_r - \mathbf{c}_r) \times (\mathbf{s} - \mathbf{c}_r)] = 0 \tag{2}$$

The first constraint is obtained by solving (2) for z_r and plugging the result into the mirror equation (1). Similarly, a second constraint can be derived from the law of reflection. Combining both yields an 8th degree polynomial equation in z_r whose roots can be computed via eigenvalue decomposition of the companion matrix, yielding $\mathbf{m}_r$.

We now extend the model of Agrawal et al. [10] by a perspective camera that includes a lens distortion model. First, we rotate the point of reflection $\mathbf{m}_r$ back into mirror coordinates and map it into the camera coordinate system:

$$\mathbf{m}_c = (x_c, y_c, z_c)^\mathsf{T} = \mathbf{R}_C \mathbf{R}_R^\mathsf{T} \; \mathbf{m}_r + \mathbf{t}_C \tag{3}$$

Here, $\mathbf{R}_R$ is the pre-rotation matrix from above and $\mathbf{R}_c|\mathbf{t}_C$ denotes the 3D rigid transformation from mirror to camera coordinates. Let $\mathbf{q}_n = (x_n, y_n)^\mathsf{T} = (x_c/z_c, y_c/z_c)^\mathsf{T}$ be the normalized projection. The distorted point is given by

$$\begin{aligned}\mathbf{q}_d = &\left(1 + k_1 r_n^2 + k_2 r_n^4 + k_5 r_n^6\right) \mathbf{q}_n \\ &+ \begin{bmatrix} 2k_3 x_n y_n + k_4 (r_n^2 + 2x_n^2) \\ k_3 (r_n^2 + 2y_n^2) + 2k_4 x_n y_n \end{bmatrix}\end{aligned} \tag{4}$$

with $r_n = \sqrt{x_n^2 + y_n^2}$ and projected to the image via

$$\begin{bmatrix} \mathbf{q} \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & \alpha f_u & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{q}_d \\ 1 \end{bmatrix} \tag{5}$$

where $\mathbf{q} = (u, v)^\mathsf{T}$ denotes a pixel in the image and f, c, α, k are the intrinsic parameters of the camera which combined with A, B and C define the set of all calibration parameters.

B. Centered Model

While the presented non-central projection model is accurate, our experiments reveal that the complex analytic form of the AFP renders the projection slow compared to simple single viewpoint models, preventing real-time applications, e.g., on mobile devices.

However, we observe that in many situations of practical relevance the distance to the objects of interest is large compared to the deviation from the single viewpoint. Thus, given a calibrated non-central model we are able to specify a central approximation which maps observations according to the image residuals for points at infinity, i.e., in a way which lets the viewing ray orientations coincide with the true ones. Note that when applying the model this mapping can be pre-computed and efficiently applied to the whole input image or individual feature points, similar to undistortion or rectification maps for perspective cameras. Importantly, note that our representation is able to represent any SVP model accurately, i.e., without approximation, as long as the mapping between the observations is smooth.

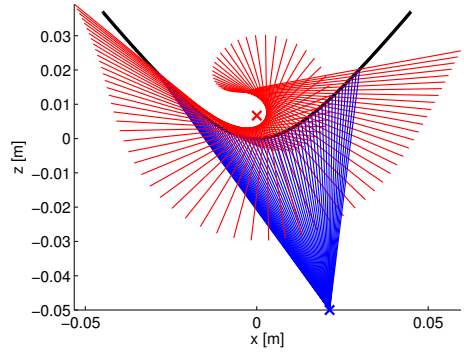


Fig. 3. **Centered Viewpoint Estimation.** The viewpoint $\mathbf{v}$ (red cross) of the centered model is obtained as the one that minimizes the distance to all reflected rays $\mathbf{w}_r$ (red). The camera center $\mathbf{c}$ and the camera rays $\mathbf{w}_c$ are shown in blue.

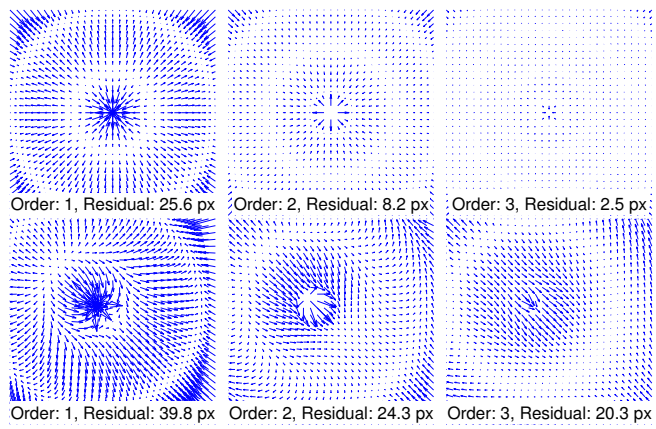


Fig. 4. **Centered Residual Field** for re-mapping the observations of the centered model for polynomials of different order. Top row: SVP scenario. Bottom: 5 mm deviation from the SVP.

As detailed in the following sections, we first compute the *optimal* single viewpoint, i.e., the point that comes closest to all viewing rays as illustrated in Fig. 3. In a second step, we derive an efficient centered camera model which we use to remap the image observations (see Fig. 4). Importantly, the quality of the centered model itself *does not* depend on the actual choice of the projection function as all centered models establish the same relationship between (remapped) pixel coordinates and the corresponding viewing ray orientations as detailed in the appendix. While our exposition is based on the geometric model from the previous section, the proposed centered model can be utilized as surrogate for any other projection function as well.

Viewpoint: Let $\mathbf{c}$ be the estimate of the true camera position, obtained by calibration using the non-central model presented in Section III-A. Here, we drop the mirror coordinate index m for clarity. Denote $\Omega = \{\mathbf{m}\}$ the set of all points on the mirror surface, defined by Eq. 1. For any $\mathbf{m} \in \Omega$, the

reflected ray $\mathbf{w}_r(\mathbf{m})$ can be computed as

$$\mathbf{w}_r = \mathbf{w}_c - 2\mathbf{n}\frac{\mathbf{w}_c^\mathsf{T}\mathbf{n}}{\mathbf{n}^\mathsf{T}\mathbf{n}} \tag{6}$$

with camera ray $\mathbf{w}_c$ and normal $\mathbf{n}$ given by

$$\mathbf{w}_c = \frac{\mathbf{m}-\mathbf{c}}{\|\mathbf{m}-\mathbf{c}\|_2} \qquad \mathbf{n} = (x, y, Az + B/2)^\mathsf{T} \tag{7}$$

Our goal is to find the viewpoint $\mathbf{v}$ that is optimal in the least-squares sense, i.e., that minimizes the squared distance to the set of all reflected rays $\{\mathbf{m} + \lambda\mathbf{w}_r(\mathbf{m}) \mid \mathbf{m} \in \Omega\}$. This requirement can be formalized as

$$\mathbf{v} = \underset{\tilde{\mathbf{v}}}{\operatorname{argmin}} \int_\Omega \|\tilde{\mathbf{v}} - \mathbf{m}\|_2^2 - ([\tilde{\mathbf{v}} - \mathbf{m}]^\mathsf{T}\mathbf{w}_r(\mathbf{m}))^2 \; d\mathbf{m} \tag{8}$$

The solution to Eq. 8 is given by the integral equation

$$\int_\Omega \mathbf{v} - \mathbf{m} - \mathbf{w}_r(\mathbf{m})\,(\mathbf{v}-\mathbf{m})^\mathsf{T}\,\mathbf{w}_r(\mathbf{m}) \; d\mathbf{m} = \mathbf{0} \tag{9}$$

which is linear in $\mathbf{v}$. This integral can be approximated to arbitrary precision by a summation over a discretized set of surface points $\{\mathbf{m}\}$. In practice, we found $2,500$ equidistantly sampled rays to yield a sufficiently good approximation. The optimal viewpoint $\mathbf{v}$ is illustrated in Fig. 3 for a 2D example. Note that even in the case of deviations from the SVP model, distant objects are mapped accurately due to the small displacement of $\mathbf{v}$ from the viewing ray. Importantly, the orientation of each viewing ray is left unaltered by means of a residual field applied to the observations which we describe in the following section.

Centered Camera Model: Given the optimal viewpoint $\mathbf{v}$, we seek for a simple and efficient camera model that maps a world point $\mathbf{p} = (x, y, z)^\mathsf{T}$ into image coordinates $\mathbf{q} = (u, v)^\mathsf{T}$. Note that here $\mathbf{q}$ refers to the remapped image observations as described in the following.

We propose an angle-based representation

$$\mathbf{q}(\varphi, \theta; c_u, c_v, \gamma) = \begin{bmatrix} c_u \\ c_v \end{bmatrix} + \sum_{i=0}^{k} \gamma_i \varphi^i \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \tag{10}$$

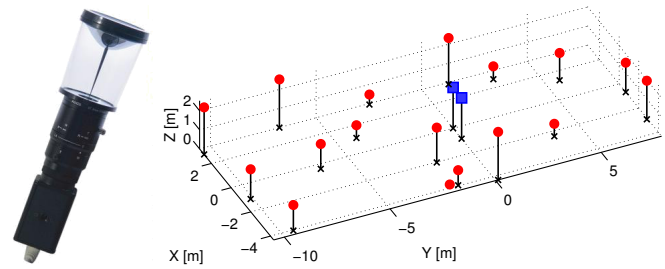
with polynomial order k, pitch φ and yaw angle θ of the light ray defined by

$$\varphi(\mathbf{p}) = \operatorname{atan}\left(\frac{z}{\sqrt{x^2+y^2}}\right) \qquad \theta(\mathbf{p}) = \operatorname{atan}\left(\frac{y}{x}\right) \tag{11}$$

Here, $(c_u, c_v)^\mathsf{T}$ denotes the location of the image center and $\gamma = (\gamma_0, \ldots, \gamma_k)^\mathsf{T}$ are the polynomial coefficients that describe the relationship between the pitch angle and the distance from the image center. The parameters of the model (c_u, c_v, γ) are obtained via non-linear least-squares: Using Eq. 6, we calculate the viewing ray orientations φ and θ for each pixel $\mathbf{q}'_{\varphi,\theta}$ in the calibrated geometric camera model and minimize

$$c_u, c_v, \gamma = \underset{\tilde{c}_u, \tilde{c}_v, \tilde{\gamma}}{\operatorname{argmin}} \sum_{\varphi,\theta} \|\mathbf{q}'_{\varphi,\theta} - \mathbf{q}(\varphi, \theta; \tilde{c}_u, \tilde{c}_v, \tilde{\gamma})\|_2^2 \tag{12}$$

to obtain the parameters of the centered model c_u, c_v, γ. Here, $\mathbf{q}'_{\varphi,\theta}$ denotes the pixel associated with φ and θ. The image residuals $\mathbf{q}' - \mathbf{q}$ after optimization define the residual displacement field which is applied to the image observations before using the centered model projection. Note that the degree of the polynomial in Eq. 10 does not impact the quality of the approximation, but only affects the smoothness of the residual displacement field, as illustrated for polynomials of differing order in Fig. 4.



(a) Camera (b) Landmarks for Localization Experiments

Fig. 5. **Experimental Setup.** Fig. (a) shows the omnidirectional camera we use (VS-C450U). Fig. (b) depicts the landmarks used in our localization experiments in red, their footpoints ($z = 0$) in black and the true camera locations of our stereo setup in blue.

Experimentally, we found a third order polynomial ($k = 3$) sufficient for providing smooth displacement fields. By initializing c_u, c_v to half the image size and γ to zero the optimization of Eq. 12 always converged to the true parameters. For a given camera calibration, the displacements can be densely precomputed. When applying the centered model, all observations are mapped through this field such that the centered model in Eq. 10 applies. Note that for points at infinity (or all central models) the centered model is exact, i.e., it is equivalent to the model it is derived from. For details we refer the reader to the appendix.

IV. Experimental Evaluation

We evaluate the proposed centered model aginst a variety of popular calibration models [12], [11], [6] on real experiments and in simulation. Towards this goal, all projection models have been integrated into a multi-camera bundle-adjustment framework.

A. Localization

Our setup consists of two hypercatadioptric cameras (see Fig. 5(a)) as it could be used in robotics or driver assistance applications. In our setup, the camera violates the single viewpoint assumption by approximately 20 mm in axial and 1 mm in lateral direction. The location of the cameras is marked in blue in Fig. 5(b). Furthermore, we installed 17 landmarks (red circles in Fig. 5(b)) at various heights ($0-2.5$ m) and distances ($2.5-10$ m), spanning an area of 10×20 meters.

For accurate 3D ground truth we made use of a high precision laser range finder to measure the distances between all pairwise combinations of cameras and landmarks, as well as their heights above ground level (black lines in Fig. 5(b)). We initialized the 3D camera and landmark locations manually and refined them by minimizing all distance and height errors using non-linear least-squares.

To obtain the calibration parameters for all camera models we presented 67 checkerboard calibration patterns to the cameras and automatically extracted the corners at sub-pixel accuracy using [25], which we extended by a tracking stage and modified to better handle catadioptric image distortions. In particular, we apply the detector of [25] on two image scales and predict corners non-linearly in the association stage. Note that for calibration we only make use of the checkerboard calibration patterns for which the pose is unknown (allowing for a simple standard calibration procedure). The 3D landmark locations are not used for calibration but only in our localization experiment. Using the automatically detected checkerboard corners, we jointly optimize the intrinsic and extrinsic calibration parameters of the catadioptric stereo camera rig described in the previous section using the geometric model, the centered model, the unified central model [6] and the models of Mei et al. [11] (unified+distortions) and Scaramuzza et al. [12]. We initialize the mirror parameters of the geometric model to the manufacturer settings which is a reasonable assumption for commercial systems where these values are usually available. All other parameters are initialized automatically assuming a SVP model. The parameters of the competing methods have been initialized as proposed by the authors.

We evaluate camera localization performance by minimizing the reprojection errors of the landmarks in the images using a monocular and a stereo setup. We selected 29 non-collinear landmark triplets as minimum sets for localization. Our results are shown in Table I. For the geometric model, we always calibrate the focal length and the principal point parameters. Distortions (k), perspective camera pose (C) and mirror parameters (M) are optimized when indicated. Note that single viewpoint models without distortions (first row and last row) fail completely in capturing the system geometry. Including distortion parameters (e.g., second row, Mei et al.) improves calibration results. As expected, using a full non-central model yields the best results, e.g., reduces stereo localization errors by 23% with respect to Mei et al. [11] and by 29% with respect to Scaramuzza et al. [12] Importantly, note that our centered model achieves almost the same performance as the geometric model from which it has been derived (in this case we used Geometric+C, third row) while being significantly faster as shown in Section IV-E. In the absence of suitable regularizers, including all parameters of the geometric model degrades calibration quality due to the complex interplay between the mirror, camera center and distortion parameters. This overfitting behavior also indicates that reprojection errors – which are often employed to judge calibration performance – are insufficient to assess the true accuracy of a calibration model.

Furthermore, we observed that the method of Scaramuzza et al. [12] is sensitive with respect to the initialization when deviating from the single viewpoint. We found the reason for this to be mainly numerical instabilities which can be mitigated by normalizing the polynomial coefficients appropriately. Using this modification and optimizing all cameras and checkerboards jointly as opposed to the two-step procedure suggested in [12] improved calibration results significantly ('Scaramuzza Improved' vs. 'Orig' in Table I) compared to the original work.

Method	Parameter	Reproj. Error [Pixels]	Local. Error Mono [mm]	Local. Error Stereo [mm]
Geometric SVP Model		1.5944	207.23	166.83
	k	0.6241	50.89	45.56
Geometric Non SVP Model	C	0.5989	42.11	36.21
	C+k	0.5864	40.04	34.64
	C+M	0.5977	43.39	38.15
	C+M+k	0.5850	89.31	86.43
Centered Model		-	42.14	36.26
Scaramuzza et al. [12]	Improved	0.6241	49.51	48.08
	Orig.	3.4143	771.93	687.86
Mei et al. [11]		0.6229	50.48	44.78
Geyer et al. [6]		0.6421	127.45	122.17

TABLE I

Calibration and Localization Experiments. THIS TABLE SHOWS OUR EXPERIMENTS ON REAL DATA IN TERMS OF THE REPROJECTION ERRORS OF THE CHECKERBOARD CORNERS AFTER CALIBRATION AND THE LOCALIZATION ERRORS, AVERAGED OVER ALL TRIPLETS.

B. Motion Estimation

In a second experiment, we mounted the hypercatadioptric stereo setup from Section IV-A onto our driving platform AnnieWAY [26] which is equipped with a high-precision GPS/IMU that delivers the ground truth motion [27]. We calibrate both cameras jointly using 70 images of checkerboards at various locations and orientations. Next, we recorded four challenging inner-city and rural sequences.

Feature detection and matching between two consecutive stereo pairs is performed using BRIEF [28] features in combination with the FAST corner detector [29] which gave the best matching results in our experiments. Using RANSAC for robustness against outliers, we estimate the motion between two consecutive frames by triangulating the 3D points in the previous frame and minimizing the reprojection errors with respect to the observations in the current frame. Note that while more sophisticated structure-from-motion pipelines [30], [31], [32], [33] could be used, we focus on the two-frame motion estimation problem here to keep things simple and avoid side-effects from the respective method. We expect that improvements in this task directly translate into improvements when using more advanced algorithms.

For comparing different calibration approaches, we show endpoint errors after 200 frames (corresponding to a driven path of up to 300 m depending on the driving speed) for all frames of each sequence. The results are shown in Fig. 6 (d-g) , with averages given in the legend of the error plot. Fig. 6 (b+c) shows the trajectories for sequence 1 in bird's eye view and from the side. Due to the large axial deviation from the single viewpoint (20 mm), the centered model is able to significantly reduce drift in z-coordinate direction (corresponding to altitude), while projected on the ground plane, all methods perform similarly. Overall, the proposed centered calibration model is able to reduce 3D translation errors by more than a factor of two compared to the baselines

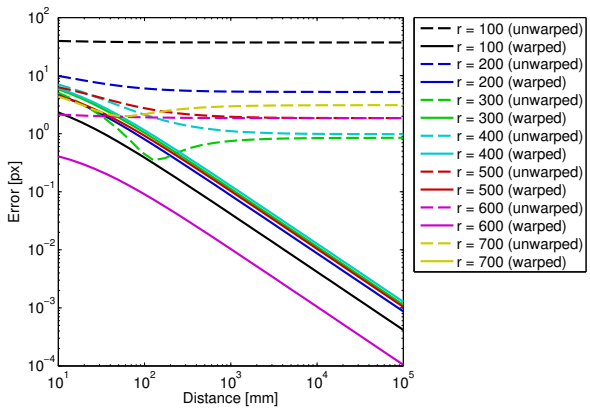


Fig. 7. **Approximation Errors of the Centered Model.** This figure shows the average approximation error of the projection in pixels over the distance of the 3D point from the camera center for simulations of the camera setup in Section IV-A. Here, r denotes the distance from the image center. The errors of the projection without warping the observations are given for reference.

which suffer from inaccuracies in the orientations of the viewing rays.

C. Single Viewpoint Violation

To further assess the sensitivity of the methods with respect to deviations from the SVP, we simulate a set of scenarios by assuming a perfect hypercatadioptric system as described by the geometric model in Section III-A using the intrinsic and extrinsic parameters from our real camera. To emphasize the differences we consider the noise-free case here. We vary the location of the perspective camera laterally (x-direction) and along the mirror axis (z-direction). For each setting, we project the checkerboards and the landmarks onto the image plane using the geometric model and calibrate the cameras using the previously described models. Next, we localize the camera as described in the previous section.

Fig. 8 shows the reprojection errors of the checkerboards after calibration (left) and the localization errors (right) for axial (top) and lateral (bottom) displacements. Again we observe that the localization errors for axial displacements are small when using a model that includes distortions. On contrast, lateral displacements impact performance much stronger. However, note that the centered model can handle both cases much better than the baseline methods. The fact that the reprojection errors of the centered model are larger can be attributed to the proximity of the calibration patterns to the camera ($0.1 - 1$m), compared to the landmarks ($>$ 2.5m). Again, small checkerboard reprojection errors are not an indicator for a well calibrated camera with respect to some target application, e.g., localization. In fact, this confirms our assumption that getting the viewing ray orientation right is much more important as small orientation errors propagate to large translation errors at distance.

D. Approximation Analysis

In this section, we analyse the approximation properties of the centered model experimentally. Towards this goal, we make use of the geometric model parameters from our real catadioptric camera system described in our experiments in Section IV-A. Fig. 7 depicts the reprojection errors with respect to the distance of the 3D point for various radii, i.e., distances from the image center. For reference we also plot the reprojection errors with respect to the original unwarped observations. This corresponds to applying an SVP model without distortions to the non-central calibration problem. Fig. 7 shows that the quality of our calibration model degrades gracefully: All errors fall below 0.1 pixel at 1 meter distance and below 0.01 pixel at 10 meters distance, even though the SVP has been violated by 20 millimeters.

	Running Time
Numeric Non-Central [24]	$\sim$ 185,000.00 ms
Geometric Model*	2,919.98 ms
Scaramuzza et al. [12]	913.93 ms
Mei et al. [11]	6.58 ms
Centered Model	3.42 ms
Centered Model*	1.79 ms

TABLE II

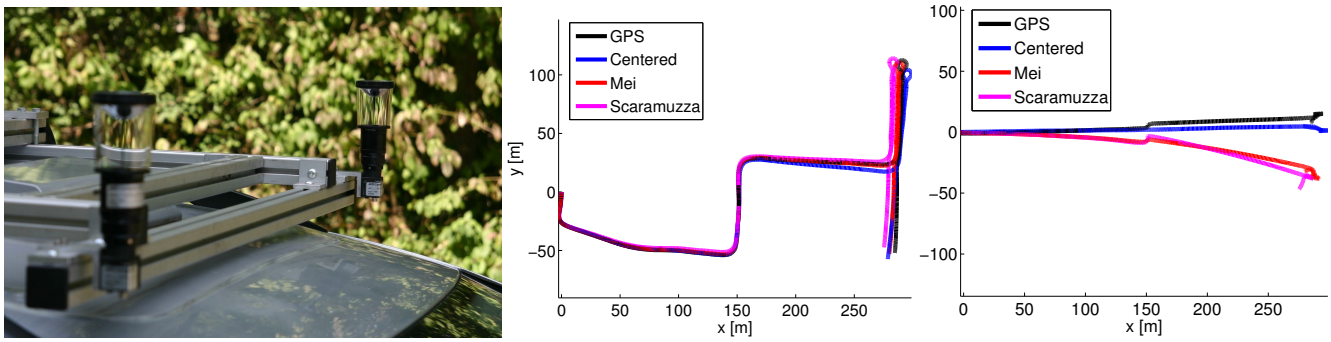
RUNNING TIMES FOR PROJECTING $10,000$ POINTS IN MATLAB. METHODS MARKED WITH AN ASTERISK (*) ARE WRAPPED IN C++.

E. Running Times

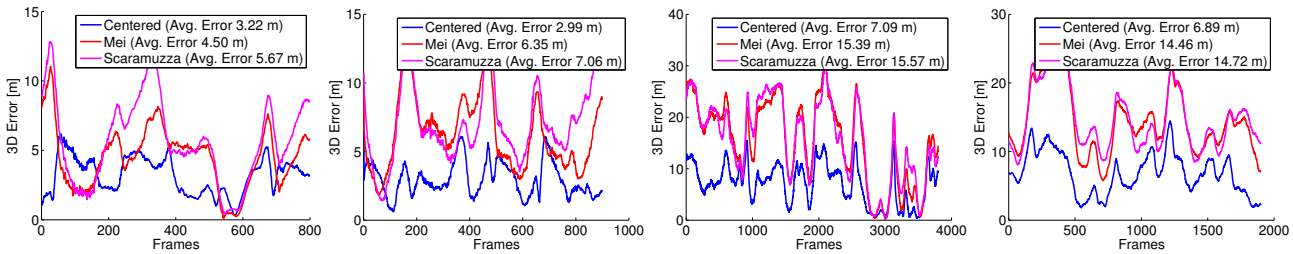
As applications (e.g., in robotics) often require real-time performance, we analyze the time required for projecting a 3D world point onto the image plane. Whenever the reprojection error needs to be computed, e.g., during localization, motion estimation or 3D reconstruction, this function is called (frequently) by the optimization routine. Table II shows the time required for projecting $10,000$ random 3D points to the image plane. The running time for the numeric non-central projection has been taken from [24] and is considered approximate. However our results agree with the observations in [10]. All the remaining running times have been computed on an Intel i7 2.67 Ghz machine using a single CPU core. As expected, the numeric optimization in the non-central model is the slowest, requiring $\sim$ 185 seconds for projecting the points. The run-time of the geometric model is heavily dominated by the computation of the analytical forward projection [10] (provided to us by the authors) which involves the computation of the polynomial coefficients and the evaluation of MATLAB's `roots` function for finding the polynomial roots. The latter is also responsible for the relatively slow evaluation of Scaramuzza's projection function [12]. On contrary, the proposed centered model and the relatively simple projection model of Mei et al. [11] are very fast and able to project $10,000$ points in a couple of milliseconds only. Compared to the geometric model, our centered approach yields a speed-up of more than three orders of magnitude.

V. CONCLUSION

In this paper we have proposed to center quasi-central catadioptric cameras while ensuring that the viewing ray orientation matches the true one. We showed that the proposed centered camera model approximates non-central catadioptric systems sufficiently well as long as the distance to the observed points is relatively large compared to the



(a) Cameras on top of our recording Platform AnnieWAY (b) Sequence 1: Trajectory (bird's eye view) (c) Sequence 1: Trajectory (side view)



(d) Sequence 1: Translation Error (e) Sequence 2: Translation Error (f) Sequence 3: Translation Error (g) Sequence 4: Translation Error

Fig. 6. **Motion Estimation Experiment.** This figure shows the hypercatadioptric cameras on top of our recording platform which we used in our experiments (a), the bird's eye and side view of the estimated and ground truth trajectories of the first (and shortest) sequence (b+c) and the end-point errors after estimating the vehicle's motion over 200 frames (d-g) corresponding to $0-300$ meters for sequences of various length.

deviation from the single viewpoint, which is a reasonable assumption in practice, e.g., for mobile robots. Unlike exact forward projections for the non-central case, its efficiency makes the proposed projection model suitable for real-time applications. We provide our dataset and code in form of a easy-to-use calibration toolbox.We believe this to be a step forward towards the practical use of catadioptric cameras in settings where both, calibration accuracy and runtime matters.

Appendix

Claim 1: Every central base model can be represented exactly by the centered model presented in Section III-B.

Proof: Assuming an undistorted perspective camera, every central projection can be represented in the form

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} c_u \\ c_v \end{bmatrix} + f(\varphi) \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \tag{13}$$

where (c_u, c_v) denotes the principal point, φ and θ are the pitch and yaw angle of the viewing ray and f is an arbitrary monotonic and smooth function. By rearranging the terms we see that

$$\varphi = f^{-1}\left(\sqrt{(u-c_u)^2 + (v-c_v)^2}\right) \tag{14}$$

Substituting Eq. 14 into Eq. 10 yields a displacement field which maps the image observations into the coordinates of the centered model. ■

Claim 2: The centered model maps points at infinity exactly for any non-central base model.

Proof: Let $\mathbf{p} = \lambda(x,y,z)^\top$ denote a 3D world point and $\mathbf{t} = (t_x, t_y, t_z)^\top$ an arbitrary finite translation of the viewing ray with pitch angle φ and yaw angle θ, i.e.,

$$\varphi = \operatorname{atan}\left(\frac{\lambda z + t_z}{\sqrt{(\lambda x + t_x)^2 + (\lambda y + t_y)^2}}\right) \quad \theta = \operatorname{atan}\left(\frac{\lambda y + t_y}{\lambda x + t_x}\right)$$

For $\lambda \to \infty$ we obtain

$$\varphi \to \operatorname{atan}\left(\frac{z}{\sqrt{x^2+y^2}}\right) \qquad \theta \to \operatorname{atan}\left(\frac{y}{x}\right)$$

Thus we can represent the viewing ray orientations exactly using a centered model. ■

References

[1] A. Voigtländer, S. Lange, M. Lauer, and M. A. Riedmiller, "Real-time 3d ball recognition using perspective and catadioptric cameras," in *ECMR*, 2007.

[2] O. Tahri and H. Araujo, "Non-central catadioptric cameras visual servoing for mobile robots using a radial camera model," in *IROS*, 2012.

[3] T. Ehlgen, T. Pajdla, and D. Ammon, "Eliminating blind spots for assisted driving," vol. 9, no. 4, pp. 657–665, 2008.

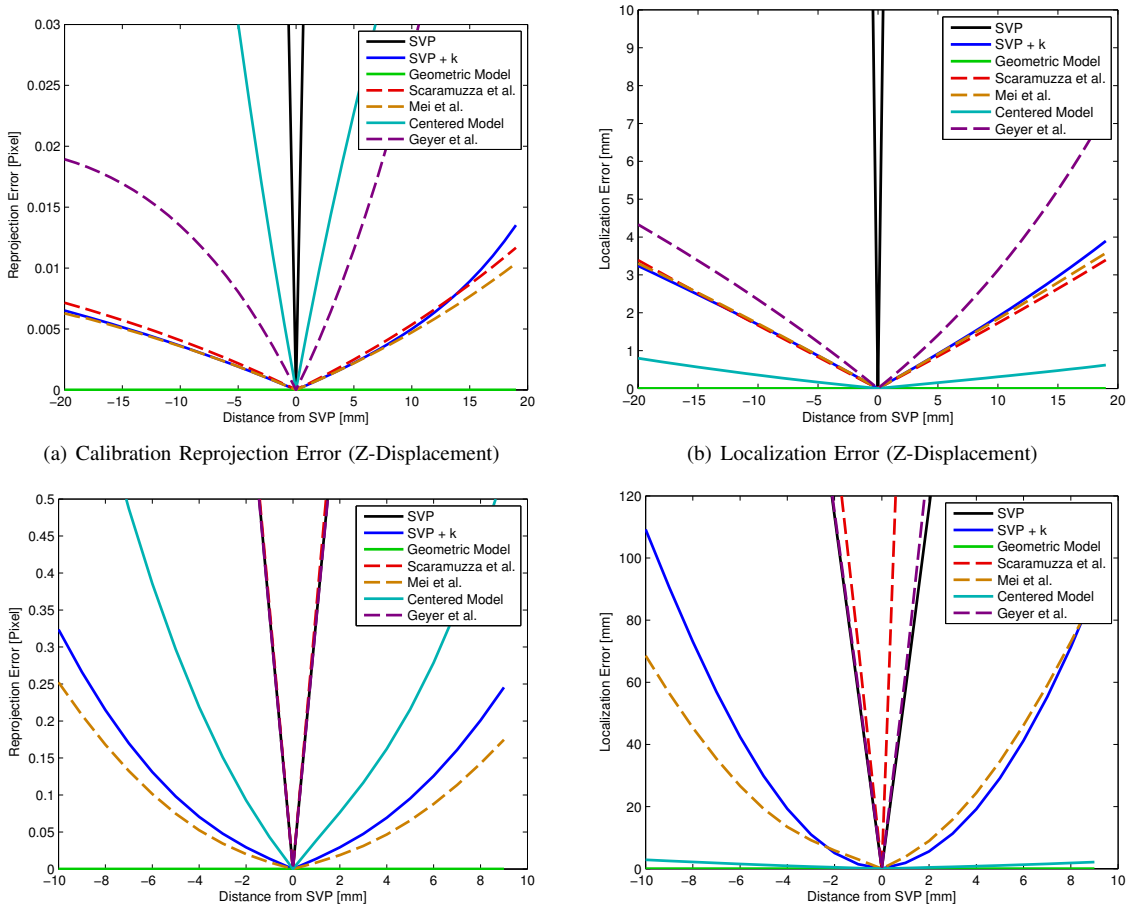
[4] D. Scaramuzza and R. Siegwart, "Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles," vol. 24, no. 5, pp. 1015–1026, 2008.

[5] S. Baker and S. K. Nayar, "A theory of single-viewpoint catadioptric image formation," *IJCV*, vol. 35, no. 2, pp. 1 – 22, 1999.

[6] C. Geyer and K. Daniilidis, "A unifying theory for central panoramic systems and practical implications," in *ECCV*, 2000.

[7] L. Puig, J. Bermúdez, P. Sturm, and J. J. Guerrero, "Calibration of omnidirectional cameras in practice: A comparison of methods," *CVIU*, vol. 116, no. 1, pp. 120–137, 2012.

[8] R. Swaminathan, M. D. Grossberg, and S. K. Nayar, "Non-single viewpoint catadioptric cameras: Geometry and analysis," *IJCV*, vol. 66, no. 3, pp. 211–229, 2006.



(a) Calibration Reprojection Error (Z-Displacement)

(b) Localization Error (Z-Displacement)

(c) Calibration Reprojection Error (X-Displacement)

(d) Localization Error (X-Displacement)

Fig. 8. **Simulated Displacements from the Single Viewpoint.** This figure shows the reprojection errors of the checkerboard corners after calibration (left) and the localization errors from the localization experiment (right) using triplets of points when displacing the camera center axially (top row) and laterally (bottom row). The geometric model (green) which has been used for generating the data produces zero error and is given for reference only.

[9] B. Micusík and T. Pajdla, "Autocalibration & 3d reconstruction with non-central catadioptric cameras," in *CVPR*, 2004.

[10] A. Agrawal, Y. Taguchi, and S. Ramalingam, "Beyond alhazen's problem: Analytical projection model for non-central catadioptric cameras with quadric mirrors," in *CVPR*, 2011.

[11] C. Mei and P. Rives, "Single view point omnidirectional camera calibration from planar grids," in *ICRA*, 2007.

[12] D. Scaramuzza and A. Martinelli, "A toolbox for easily calibrating omnidirectional cameras," in *IROS*, 2006.

[13] C. Geyer and K. Daniilidis, "Paracatadioptric camera calibration," *PAMI*, vol. 24, no. 5, pp. 687–695, 2002.

[14] J. P. Barreto and H. Araujo, "Geometric properties of central catadioptric line images and their application in calibration," *PAMI*, vol. 27, no. 8, pp. 1327–1333, 2005.

[15] X. Ying and Z. Hu, "Catadioptric camera calibration using geometric invariants," *PAMI*, vol. 26, pp. 1260–1271, 2004.

[16] L. Puig, Y. Bastanlar, P. Sturm, J. Guerrero, and J. Barreto, "Calibration of central catadioptric cameras using a dlt-like approach," *IJCV*, vol. 93, pp. 101–114, 2011.

[17] O. Morel, R. Seulin, and D. Fofi, "Catadioptric camera calibration by polarization imaging," in *IAPR*, 2007.

[18] S. Baker and S. K. Nayar, "Single viewpoint catadioptric cameras," in *Panoramic Vision*. Springer, 2001.

[19] J. Barreto, "A unifying geometric representation for central projection systems," *CVIU*, vol. 103, no. 3, pp. 208–217, 2006.

[20] S. Gasparini, P. Sturm, and J. Barreto, "Plane-based calibration of central catadioptric cameras," in *ICCV*, 2009.

[21] V. Caglioti, P. Taddei, G. Boracchi, S. Gasparini, and A. Giusti, "Single-image calibration of off-axis catadioptric cameras using lines," in *Omnivis*, 2007.

[22] D. Strelow, J. S. Mishler, D. Koes, and S. Singh, "Precise omnidirectional camera calibration," in *CVPR*, 2001.

[23] M. Lhuillier, "Automatic scene structure and camera motion using a catadioptric system," *CVIU*, vol. 109, no. 2, pp. 186–203, 2008.

[24] N. Goncalves and A. Nogueira, "Projection through quadric mirrors made faster," in *Omnivis*, 2009.

[25] A. Geiger, F. Moosmann, O. Car, and B. Schuster, "Automatic calibration of range and camera sensors using a single shot," in *ICRA*, 2012.

[26] A. Geiger, M. Lauer, F. Moosmann, B. Ranft, H. Rapp, C. Stiller, and J. Ziegler, "Team annieway's entry to the grand cooperative driving challenge 2011," *TITS*, 2012.

[27] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *IJRR*, vol. 32, pp. 1229 – 1235, 2013.

[28] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua, "Brief: Computing a local binary descriptor very fast," *PAMI*, vol. 34, no. 7, pp. 1281–1298, 2012.

[29] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *ECCV*. Springer Berlin Heidelberg, 2006, pp. 430–443.

[30] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *IJRR*, vol. 31, pp. 217–236, 2012.

[31] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski, "Building rome in a day," in *ICCV*, 2009.

[32] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *PAMI*, vol. 29, no. 6, pp. 1052–1067, 2007.

[33] P. Alcantarilla, L. Bergasa, and F. Dellaert, "Visual odometry priors for robust EKF-SLAM," in *ICRA*, 2010.