Universidad Zaragoza
1542

Master's Thesis

Simplifying geometric camera calibration with minimal requirements on calibration objects or angular directions

Author

María Sánchez Sarsa

Supervisors

Javier Civera Sancho

Klaus H. Strobl

ESCUELA DE INGENIERÍA Y ARQUITECTURA
2025-2026

Acknowledgments

This work has been possible thanks to many people, whom I would like to thank.

First, I would like to sincerely thank *Javier* and *Klaus*, for giving me the opportunity to carry out this work at the Deutsches Zentrum für Luft- und Raumfahrt (German Aerospace Center) and for their continuous guidance and support throughout the project. Their expertise, insightful feedback, and supervision have been fundamental to the development of this thesis.

I would also like to thank my friends, who are much more than just office mates, from *Room 2110* and *2102*: María, Matthias, Leon, Kristina, Kristian, Jakob and Yasin. Thank you for the very fun working environment, the even better times shared outside the office, and all the memories I will cherish from these months.

Finally, I would like to thank my parents, who are the reason why I was able to take part in such a wonderful and enriching path, and who have always supported me throughout it.

Abstract

Geometric camera calibration is a fundamental step in computer vision, as it provides the intrinsic parameters that define the camera projection model. Classical calibration methods, such as Zhang's technique, achieve high accuracy but require multiple images of a known calibration pattern and careful data acquisition, which limits their usability in constrained or low-effort scenarios.

This thesis explores camera calibration under minimal data requirements, with the aim of simplifying the calibration process while preserving geometric reliability. Two approaches are investigated. First, a one-shot calibration method based on Zhang's formulation is studied, reducing calibration to a single image of a planar object with the minimum number of features. The influence of camera–pattern pose on parameter observability is analyzed, leading to practical guidelines for selecting configurations that provide sufficient perspective and radial distortion evidence.

Second, a zero-shot calibration method is introduced, which eliminates the need for calibration images and patterns. By modeling the camera as an angular measurement device, intrinsic parameters are estimated from angular constraints between viewing rays, including explicit radial distortion modeling. Theoretical and numerical analyses are used to characterize parameter coupling and identify well-conditioned measurement configurations.

The results show that meaningful camera calibration is possible with significantly reduced data, provided that sufficient geometric information is available. This work contributes to a better understanding of intrinsic parameter observability and supports the development of simpler and more accessible calibration procedures.

Index

Chapter 1

Introduction and objectives

Geometric camera calibration is the process of estimating the parameters that define the mathematical model of image formation, including focal length, principal point, lens distortion, and, when required, the camera pose. This model describes how three-dimensional scene geometry is mapped onto the image and establishes the geometric relationship needed to recover spatial information from image measurements. Accurate calibration is therefore critical in applications such as 3D reconstruction, robotics, and augmented reality, where small parameter errors can propagate into significant geometric inaccuracies.

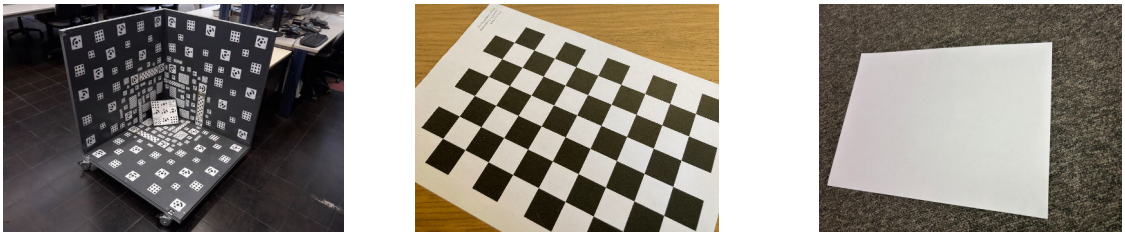
For this reason, the search for precise and practical calibration methods has been ongoing for decades. Early calibration approaches emerged from photogrammetry, where cameras were calibrated by observing a very precise 3D calibration object [1]. Classical works, such as Brown's close-range calibration and distortion formulation, exemplify the compromise that these methods presented: they achieved high accuracy but required precise 3D calibration objects and extensive knowledge of the method. As a result, camera calibration was not broadly accessible and was typically confined to specialized, well-equipped environments.

A major step was taken with the introduction of more flexible calibration techniques in computer vision. Tsai proposed a 2.5D calibration method, which reduced the dependence on complex 3D objects, although it still required constrained camera–target motion and assumptions about the structure of the calibration object [2]. The final improvement came with Zhang's method [3], which demonstrated that full camera calibration could be performed using only a planar pattern observed from multiple viewpoints at unknown poses. By exploiting homographies between the image plane and the calibration target, this technique eliminated the need for 3D objects. Unlike Tsai's method, it did not require constrained camera-to-target motion or assumptions about the structure of the calibration object, allowing for simple, low-cost calibration setups while maintaining high accuracy.

Despite these advances, there is still a need for further simplification of camera calibration processes. Designing more straightforward methods is crucial for wider adoption of the geometric camera projection model both in industry and in daily life applications, as simpler technical expertise requirements would make computer vision systems more accessible and user-friendly.

In this context, recent research has begun to explore calibration strategies that operate under increasingly minimal data conditions. In particular, *one-shot* calibration methods aim to estimate camera parameters from a single image of a calibration target, while *zero-shot* approaches seek to eliminate the need for explicit calibration objects altogether, relying instead on geometric or angular constraints inherent to the imaging process. These methods provide a useful framework for studying the fundamental requirements of camera calibration and motivate the objectives of the present thesis.

Figure 1.1 illustrates this evolution, from 3D calibration objects to planar and minimal calibration setups.



(a) 3D calibration object [4]

(b) Planar checkerboard calibration

(c) Proposed minimal calibration pattern

Figure 1.1: Evolution of camera calibration setups, from complex 3D objects to minimal approaches.

1.1 Objectives and scope

The objective of this thesis is to explore geometric camera calibration with minimal reliance on calibration objects or measurements, with the goal of reducing the complexity of the process while maintaining precision. The temporal distribution of the work is shown in Table 1.1.

The main contributions of this thesis are:

1. We reduce the data requirements of classical camera calibration by adopting Zhang's method [3] as a baseline and decreasing both the number of required calibration views and the complexity of the calibration pattern.

2. For that minimal Zhang method, we conduct an analysis of the camera pose and pattern configurations to identify the conditions that best constrain the

calibration parameters. Based on this analysis, we propose practical guidelines for selecting optimal camera–to-pattern poses.

3. We introduce a zero-shot camera calibration method based on angular measurements, which incorporates radial distortion modeling (including the estimation of its center of distortion), without requiring camera pose estimation.

4. For our zero-shot calibration method based on angular measurements, we analyze how different angular measurement configurations affect parameter observability and provide guidelines for their optimal selection.

Task	Aug	Sep	Oct	Nov	Dec	Jan
Research						
Minimal Zhang method						
Zero-Shot method						
Studying simulations						
Writing & results						

Table 1.1: Gantt diagram representing the distribution of the work along the months.

1.2 Structure of the thesis

The work is structured as follows:

1. **Chapter 2** reviews state-of-the art of camera calibration methods, including classical approaches and recent studies on pattern-free calibration.

2. **Chapter 3** presents the camera model used throughout the thesis, namely the pinhole camera model with radial distortion.

3. **Chapter 4** analyzes the classical Zhang calibration method, and proposes a simplified formulation in which camera parameters are estimated from a single image of a planar everyday object (a sheet of paper). The influence of the camera pose with respect to (w.r.t.) the calibration object and its relevance are studied through simulations.

4. **Chapter 5** explores an approach that relies neither on calibration targets nor on image processing, using external angular measurements of the camera's field of view instead.

5. **Chapter 6** summarizes the results of the simulations, and discusses the robustness and limitations of each approach.

All the code developed for this thesis is accessible upon request

Use of Artificial Intelligence tools

Artificial intelligence tools were used during the development of this thesis to support code for data visualization, the creation of plots and the automation of results from simulation outputs. These tools were also used to support code debugging, as well as to check for writing mistakes and ease the correct use of LaTeX commands, such as table and figure formatting.

Chapter 2

State-of-the-art

Geometric camera calibration has evolved significantly since the first widely adopted methods [1, 2]. Over time, the field has moved towards more flexible and user-friendly approaches that reduce the data-collection requirements during calibration. This trend is motivated not only by the desire for more ease of use but also by the aim of avoiding high-precision experimental setups, which may introduce errors when not carefully designed or measured. The present work builds primarily on Zhang's method, which has become the *de facto* standard for geometric camera calibration. In addition, object-free calibration approaches are also explored [5].

2.1 The Zhang method

Zhang, concurrently with Sturm and Maybank, introduced a flexible calibration technique that changed the computer vision field by lowering the setup requirements [3, 6]. Unlike previous methods that required known 3D calibration objects, this approach only needs a planar checkerboard to be observed in at least three different orientations. Either the camera or the pattern can be moved freely, and this motion does not need to be known in advance. The key feature of this calibration method is reducing the dependence on a 3D calibration object or on the precise motion of a 2D pattern to a 2D pattern of unmeasured pose. Mathematically, this dimensionality reduction simplifies the problem: estimating the fundamental matrix is replaced by estimating a homography for each pose. Methodologically, it is also simpler for the user, since a planar printed pattern is sufficient.

More than 25 years after its introduction, Zhang's technique remains the standard and most widely used method for geometric camera calibration due to its simplicity. The user is only required to print the calibration pattern, take some photos from different orientations and upload them to a computer vision toolkit, such as OpenCV, the MATLAB Calibration Toolbox, or DLR CalLab [7, 8, 9].

2.1.1 Mathematical formulation

The mathematical foundation of Zhang's method relies on the geometric relationship between 3D world points and their 2D image projections.

The projection of a 3D point $\mathbf{X}$, expressed in the coordinate frame of the calibration target, onto the image plane results in a 2D point $\mathbf{x}$ is

$$\mathbf{x} \sim \mathbf{K}\,[\mathbf{R} \mid \mathbf{t}]\,\mathbf{X}^\top \tag{2.1}$$

where $\mathbf{R}$ and $\mathbf{t}$ describe the transformation that maps the points from the 3D to 2D world. In particular, $\mathbf{R}$ represents the orientation of the target w.r.t. the camera, and $\mathbf{t}$ its position. Lastly, $\mathbf{K}$ is the intrinsic calibration matrix describing the internal geometry of the camera (Chapter 3 for a detailed description).

For a planar target, all points lie on a single image plane, $Z = 0$, the mapping reduces to a homography:

$$\mathbf{x} \sim \mathbf{H}\,\mathbf{X}^\top, \quad \text{being} \quad \mathbf{H} \quad \text{the homography} \quad \mathbf{H} = \mathbf{K}\,[\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{t}]\,, \tag{2.2}$$

where $\mathbf{X}^\top = [X, Y, 1]^\top$ denotes the planar point coordinates in the target reference frame, and $\mathbf{r}_1, \mathbf{r}_2$ are the first two columns of the rotation matrix, $\mathbf{R}$, as shown in Figure 2.1.

The key observation exploited by Zhang, Sturm, and Maybank is that $\mathbf{R}$ is a rotation matrix and therefore it is orthogonal. This implies the following constraints

$$\mathbf{r}_1^\top \mathbf{r}_2 = 0, \qquad \mathbf{r}_1^\top \mathbf{r}_1 = \mathbf{r}_2^\top \mathbf{r}_2. \tag{2.3}$$

Let $\mathbf{H} = [\mathbf{h}_1\ \mathbf{h}_2\ \mathbf{h}_3]$ denote the homography matrix, where $\mathbf{h}_i$ are its column vectors. From $\mathbf{H} = \mathbf{K}[\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{t}]$, it follows that $\mathbf{r}_i = \mathbf{K}^{-1}\mathbf{h}_i$. Using this relation, equations (2.3) become

$$\mathbf{h}_1^\top \mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_2 = 0, \qquad \mathbf{h}_1^\top \mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_1 = \mathbf{h}_2^\top \mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_2. \tag{2.4}$$

Stacking these constraints across multiple views yields a system of equations that provides a closed-form estimate of $\mathbf{K}$, which is then refined via nonlinear optimization.

2.1.2 Practical workflow

The standard calibration workflow consists of the following steps: a user captures multiple images of a checkerboard, the software processes the images, and then it computes the intrinsic matrix (focal lengths f_x, f_y, principal point c_x, c_y) and distortion coefficients (typically k_1, k_2 for radial and p_1, p_2 for tangential distortion), together with the camera poses w.r.t. the calibration plate.

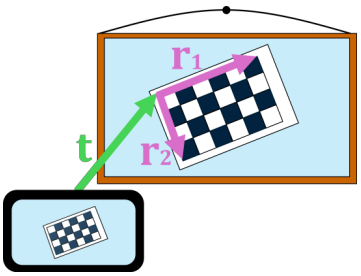


Figure 2.1: Camera–pattern setup with a planar target observed from a fixed viewpoint.

Figure 1.1b shows a typical calibration image where the checkerboard pattern is positioned at a certain orientation relative to the camera. The user must capture multiple images with the pattern at different angles and distances to provide sufficient geometric variation. When enough images are collected, typically 10 to 15, the software processes the images and runs Zhang's algorithm to compute the camera parameters.

2.2 Constraints and limitations

Despite the popularity and flexibility of the method, there are still significant constraints and limitations. While the theoretical requirements are low (historically three views), the practical requirements for high-accuracy calibration are more demanding. Consequently, Zhang's method still presents a significant gap between its theoretical elegance and the practical usability.

2.2.1 The multi-view requirement

The most immediate practical constraint is the number of images required. Although the original mathematical formulation suggests three views, significantly more photos are used, usually from 10 to 15 for reliable results [3].

This requirement arises from a geometric degeneracy of the calibration equations: when the captured images do not provide sufficient perspective variation, the focal length and the distance to the calibration pattern become strongly attached and cannot be reliably uncoupled [10]. For instance, a camera with twice the focal length observing a pattern from twice the distance produces exactly the same image. As more views are taken from different angles and distances, the ambiguity decreases, and the system of equations becomes better conditioned. In addition, a homogeneous parametrization of the lens distortion model across the entire image field is desired, and consequently an evenly distributed projection of the pattern on the images is sought. This turns calibration from a simple task into a systematic data collection procedure requiring the user to plan for the optimal camera motions.

2.2.2 Calibration pattern quality

Another significant limitation lies in the often overlooked assumption that the calibration pattern is perfectly planar and with precisely known geometry. The influence of calibration pattern accuracy has been questioned in the literature. Early works suggested that calibration errors were dominated by image measurement noise rather than pattern inaccuracies [11]. However, more recent studies by Strobl and Hirzinger [12, 13] demonstrated that even minor geometric deviations in the physical pattern can significantly bias both intrinsic and extrinsic parameter estimates and that these deviations can be better estimated during calibration.

In most daily use cases, the pattern is printed with a standard office printer on a simple paper sheet. This introduces multiple sources of error, such as deformation, printing inaccuracy in both dimensions or checkerboard edge quality, which deteriorate the quality of the calibration.

2.2.3 Environmental and user factors

Finally, the calibration process is highly sensitive to the setup environment and human operation. Light changes and high contrast within the checkerboard pattern play an important role to achieve unbiased image processing results. Uneven lighting can introduce systematic localization errors in corner detection, deviating from the assumption of zero-mean Gaussian noise.

Furthermore, the method assumes that the users understand how to capture "good" calibration data. However, Zhang's method provides no guidance on optimal capture strategies. The users must verify the calibration result by checking reprojection error metrics or testing 3D reconstructions when the process is done.

2.2.4 Toward minimal calibration constraints

The above-mentioned constraints and limitations reveal a common issue: Zhang's method, despite only requiring a planar pattern as a calibration object, still demands significant effort and expertise from the user. The user must still capture numerous planned images, ensure pattern quality, control the conditions, and have sufficient expertise to evaluate the calibration results.

Some methods support increasingly complex camera models with a large number of parameters to capture subtle optical effects, like in [14]. Other works, which align with this thesis' objective, pursue simplification on requirements on equipment, environment, and user knowledge. These approaches range from exploiting scene geometry with straight lines [15] and circles [16] to self-calibration techniques that

eliminate explicit calibration objects [17, 18].

More recently, learning-based approaches have also investigated camera calibration under minimal assumptions. In particular, AnyCalib proposes a model-agnostic, single-view calibration framework based on on-manifold learning, capable of estimating camera parameters without explicit calibration objects [19]. While such methods rely on strong data-driven priors learned from large datasets, they demonstrate that calibration without patterns is feasible beyond purely geometric formulations.

The present work explores camera calibration from minimal geometric constraints, searching to establish the baseline requirements that make camera calibration feasible.

Chapter 3

Camera model

3.1 The pinhole camera model

The pinhole camera model, shown in Figure 3.1a, represents the most fundamental geometric description of image formation in computer vision [20]. It serves as the baseline upon which more complex optical effects, such as lens distortion aberration or thick lenses, are added. This model describes the ideal projection of a 3D point $\mathbf{X} = [x, y, z]^\top$, expressed in the camera coordinate system, onto the image plane, assuming that no lens distortion is present. The resulting projection, in pixel coordinates is given by

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} x/z \\ y/z \\ 1 \end{bmatrix}, \tag{3.1}$$

where $\mathbf{K}$ is the intrinsic calibration matrix

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.2}$$

Here, f_x and f_y are the focal lengths, in pixels, and (c_x, c_y) represents the principal point of the camera, which is the intersection of the optical axis with the image plane, also in pixels. In the following, we assume square pixels and denote the focal length by $f = f_x = f_y$. This represents a common simplifying assumption that reduces the number of intrinsic parameters without altering the fundamental pinhole projection model.

3.1.1 Angular field of view

The angular field of view (AoV) of a camera describes the angular scope of the observable scene that is projected onto the image plane. In the pinhole camera model,

shown in Figure 3.1b, the AoV is directly related to the focal length and the image sensor dimensions [20].

For a pinhole camera with focal length f and image size (W, H) (in pixels), the horizontal and vertical field of view θ_{h} and θ_{v} are defined as

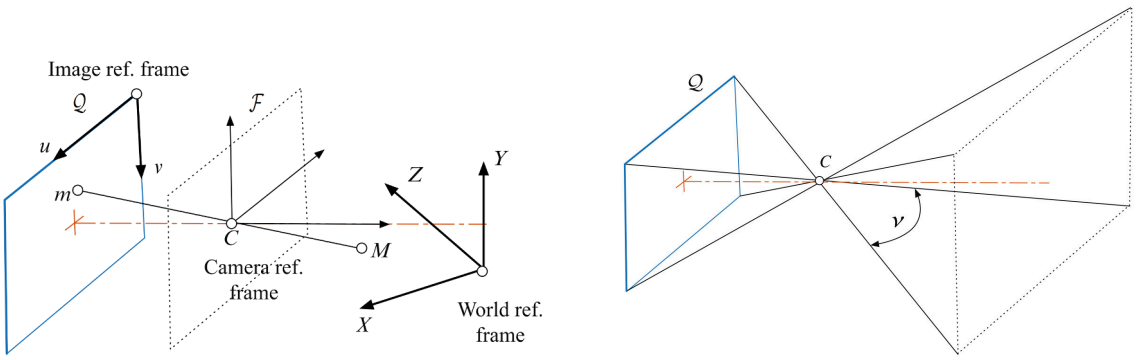
$$\theta_h = 2\arctan\left(\frac{W}{2f}\right) \qquad \theta_v = 2\arctan\left(\frac{H}{2f}\right). \tag{3.3}$$

For a fixed sensor size, shorter focal lengths produce wider angular field of view, while longer focal lengths narrow the angular range.

When the field of view and the image dimensions are known, the focal lengths can be recovered with

$$f = \frac{W}{2\tan(\theta_{\mathrm{h}}/2)} = \frac{H}{2\tan(\theta_{\mathrm{v}}/2)}. \tag{3.4}$$

Expressing focal length in terms of the AoV provides a more intuitive parameterization and highlights the direct link between angular information and intrinsic camera calibration.



(a) Pinhole camera model geometry (b) Definition of the angular field of view

Figure 3.1: The basis of the camera model used throughout this thesis [20].

3.1.2 Perspective distortion

The pinhole camera model introduces a geometric deformation inherent to all perspective projections, known as perspective distortion. This effect is not an optical aberration but a consequence of projecting 3D space onto a 2D plane through a theoretical pinhole. Objects closer to the camera appear larger, while distant objects appear smaller.

Under the pinhole camera model, the radial distance of an image point from the principal point is known as the perspective projection radius

$$r_{\mathrm{pers}} = f\tan(\theta), \tag{3.5}$$

where θ is the angle between an incoming ray and the optical axis. This relationship is a direct consequence of the projective geometry of the pinhole model and exists independently of any lens imperfections or optical distortion.

3.2 The stereo camera model

The fundamental limitation of monocular vision is that depth cannot be directly observed. This can be overcome by introducing a second viewpoint. A stereo camera system uses two cameras with known relative pose to observe the same scene, enabling depth recovery through correspondence search and subsequent triangulation.

In the common stereo configuration, shown in Figure 3.2, the image planes are parallel and the relative translation is aligned with the horizontal axis. A 3D point is projected onto the left and right image planes according to their respective camera models. Corresponding points appear on the same image row, and their horizontal displacement, known as disparity, is defined as

$$D = u_{\mathrm{L}} - u_{\mathrm{R}}, \tag{3.6}$$

where u_{L} and u_{R} denote the horizontal pixel coordinates of the point in the left and right images, respectively.

Since disparity is inversely proportional to depth, it provides a direct depth measurement:

$$Z = \frac{fb}{D}, \tag{3.7}$$

where b is the baseline distance between the two camera centers.

This illustrates how by introducing a second viewpoint, depth is directly observable, in contrast to the monocular case, where depth must be inferred from additional geometric or calibration constraints or prior information.

3.3 The Brown-Conrady distortion model

The pinhole camera model provides an ideal geometric description of image formation but does not account for optical imperfections present in real lenses. To model these effects, additional distortion terms are commonly introduced on top of the pinhole projection.

Between the different aberration effects, the most relevant in geometric computer vision is the radial distortion. A model that describes this is the Brown–Conrady model [1], which expresses radial distortion in 2D image coordinates as

$$\begin{aligned} x_{\mathrm{d}} &= x_{\mathrm{u}} \left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6 ...\right), \\ y_{\mathrm{d}} &= y_{\mathrm{u}} \left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6 ...\right), \end{aligned} \tag{3.8}$$

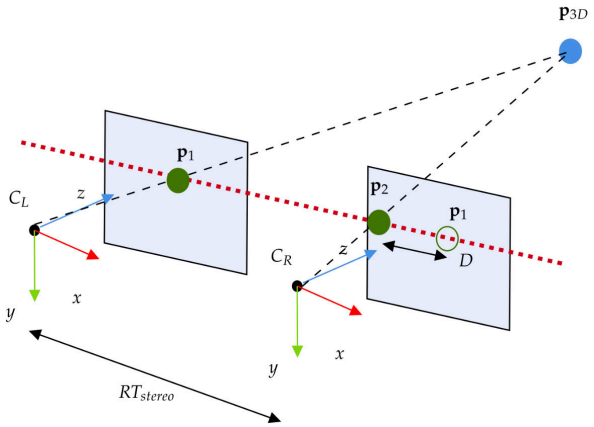


Figure 3.2: Stereo vision geometry with two cameras separated by a baseline and observing a common scene [21].

where $(x_\mathrm{d}, y_\mathrm{d})$ are the distorted coordinates, $(x_\mathrm{u}, y_\mathrm{u})$ the undistorted, and $k_1, k_2, k_3, ...$ are the radial distortion coefficients.

The impact of each coefficient varies significantly. The first order coefficient k_1 dominates in most lenses and captures the primary distortion behavior. In fact, Tsai in [2] states that *"any more elaborate modeling not only would not help but also would cause numerical instability."* Here, positive coefficients correspond to barrel distortion, typically seen in wide-angle lenses, while negative coefficients generate pincushion distortion. The second-order coefficient k_2 becomes relevant for very wide-angle lenses, and third-order coefficient k_3 is rarely used except for fisheye or very distorted optical systems.

Chapter 4

Minimal Zhang method

4.1 Motivation

Camera calibration remains a fundamental step for most computer vision tasks. Traditional methods [3], while robust, typically require a tedious data acquisition process involving multiple images of a checkerboard pattern taken from various angles. This requirement often creates a barrier for the user.

This chapter explores a simplified calibration approach. The presented approach is an instance of the Zhang calibration method relying on a minimal amount of input data. The objective is to study whether camera calibration can be performed using a reduced calibration setup, both in terms of the complexity of the calibration pattern and the number of required images. In particular, a **DIN A4 paper sheet** is proposed as the calibration object and the use of **one single calibration image** for the calibration process, hence a one-shot method.

This research builds upon the theoretical framework developed in 2016 by Strobl and Lingenauber in their work on *"Stepwise calibration of focused plenoptic cameras"* [22], specifically leveraging their techniques for decoupling intrinsic parameters through principal point centralization.

4.2 Problem formulation

4.2.1 Minimal data and problem dimensionality

The proposed one-shot calibration approach aims to estimate a set of camera parameters. In particular, the intrinsic parameters considered are the focal length f, the principal point (c_x, c_y), and the dominant radial distortion coefficient k_1 (see Chapter 3). In addition, the price for using the convenient Zhang method is the necessity to estimate the 6-DoF pose of the calibration object w.r.t. the camera for each calibration image (just one in this work). Altogether, this results in a total of ten

unknown parameters

$$\mathbf{x} = (f, k_1, c_x, c_y, r_1, r_2, r_3, t_1, t_2, t_3). \tag{4.1}$$

In order to obtain a well-defined calibration problem, a sufficient number of independent measurements is required. For this approach, 4 planar (2D) projections of 3D points on the image (the corners of the A4 paper sheet) are chosen. On the one hand, that is the minimum amount of data that allows the computation of a homography between planar projections in 3D (see Appendix A), which will be used to initialize the camera parameters following the Zhang method. On the other hand, because it amounts to eight measurements that, together with the measured size of the calibration object (width and height, or aspect ratio and absolute scale[1]), add up to the ten measurements that can allow a well-defined parametrization of the above camera model if the data is varied enough such that the influence of each parameter on the measurements can be clearly separated.

Therefore, a DIN A4 paper sheet is a well-suited, standardized object for this task.

4.2.2 The challenge of minimal calibration

The use of minimal data fundamentally changes the nature of the calibration problem. While traditional multi-image Zhang calibration benefits from redundant measurements that can statistically compensate for noise, the one-shot approach with only four corner correspondences creates an exactly determined system which introduces vulnerability.

With minimal data, however, each of the four corners contributes critically to the solution. The eight measurements (four 2D points) precisely match the degrees of freedom in the homography matrix. Any noise in corner detection propagates directly into the estimated parameters. When optimization achieves zero reprojection error—which is expected in an exactly determined system—this exact fit is deceptive. The measurement errors have not been canceled; they have been absorbed into the estimated focal length, distortion, principal point, and pose.

This fundamental limitation reveals that in minimal calibration scenarios, the geometric configuration of the calibration setup becomes very relevant. The quality of the evidence provided by the camera-to-object pose—specifically, the amount of perspective distortion and the distribution of radial distances—determines whether the parameters can be estimated in a stable, uncoupled manner. Without careful geometric

[1]Note that camera calibration is still possible without knowledge of the absolute pattern scale (i.e., with nine measurements) if the pattern's absolute scale is fixed to an aleatory value and the camera-to-pattern range, which is a by-product of calibration, is disregarded [12].

conditioning, certain parameters become interchangeable (coupled); see Appendix B, and the calibration is unreliable regardless of the precision of the measurement.

The following section examines the relevance of geometric configurations to provide sufficient evidence for robust single-image calibration.

4.3 Geometric conditioning

Geometric conditioning plays a central role in determining the feasibility and stability of camera calibration from minimal requirements. In this section, we analyze how camera pose parameters influence calibration performance, with evidence drawn from the experimental results presented in Chapter 6.

In the following, the orientation of the calibration pattern w.r.t. the camera is described using the standard Euler angle convention: pitch, yaw, and roll. Pitch denotes a rotation around the horizontal image axis, yaw a rotation around the vertical image axis, and roll a rotation around the optical axis of the camera.

4.3.1 The role of perspective evidence

The coupling between geometric projection and lens distortion under limited viewpoints has previously been highlighted, particularly in minimal calibration scenarios [23].

Interestingly, in this extreme one-shot camera calibration scenario, the quality of the evidence on perspective projection (i.e., the perspective distortion), provided by the relative pose between calibration object and camera, becomes more relevant than precise corner localization. For example, even if corner localization was twice as noisy, the estimated parameters would just slightly worsen. However, a single calibration image without sufficient perspective distortion will not allow the valid and uncorrelated estimation of parameters that can become coupled, such as focal length and radial distortion.

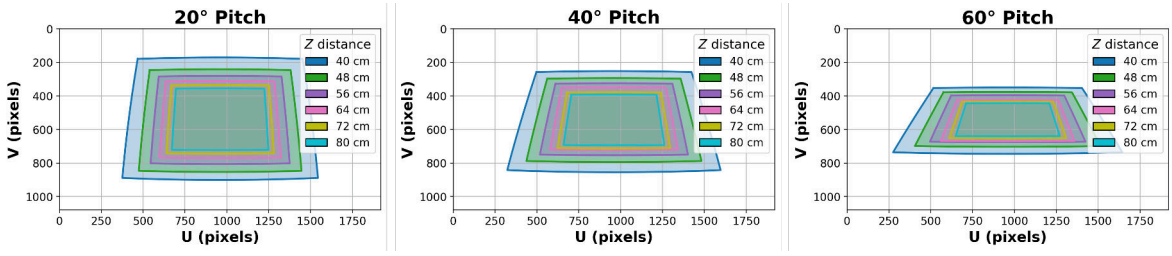


Figure 4.1: Effect of camera pitch and depth on calibration pattern projection.

The importance of perspective evidence can be directly observed in Figure 4.1, where the calibration pattern is shown at varying Z distances from the camera, under

different pitch angles (20°, 40° and 60°). At small pitch angles (20°), the projected objects corresponding to different depths remain almost uniform in shape across Z. This configuration provides weak perspective cues. As the pitch increases to 40°, perspective deformation becomes more apparent, and at 60° pitch, perspective effects dominate: the shape of the object stretches unevenly across the image.

4.3.2 The role of radial distortion evidence

Another critical aspect for this method is obtaining sufficient evidence to parametrize the radial lens distortion model (Brown-Conrady model) from a single-image containing the four 2D locations of the calibration object. Since radial distortion grows nonlinearly with the distance from the principal point, placing the calibration object near the image center provides little information at large radii, where distortion effects are most pronounced. As a result, the optimizer struggles to accurately estimate the distortion parameters. In the absence of sufficient radial distortion evidence, the optimization process tends to compensate by adjusting other parameters, such as the focal length, principal point, or camera pose, leading to biased estimates.

To address this, it is crucial to choose a camera-to-object pose that meets two criteria: first, fill the image with the calibration object as much as possible, and second, provide a well-distributed set of radial distance measurements. The latter ensures that perspective effects and radial distortion can be decoupled. A varied-distribution can be achieved by creating asymmetry in the projected pattern corners.

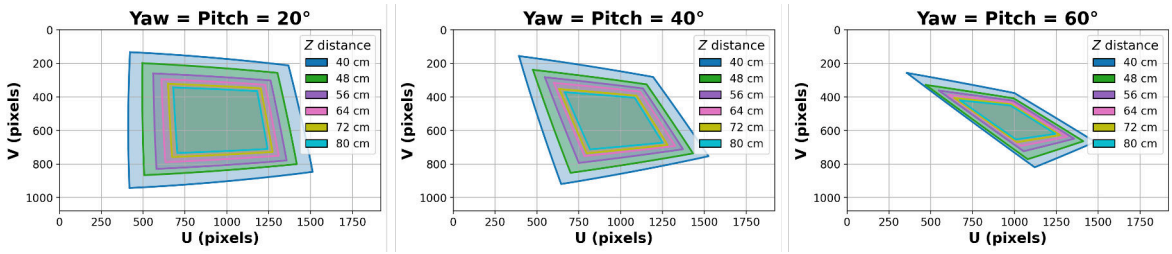


Figure 4.2: Effect of combined camera pitch, yaw, and depth on calibration pattern projection.

The importance of radial distance distribution is shown in Figures 4.2 and 4.3. Radial distortion evidence can be observed in the figures by how the projected corners move across the image as yaw and roll increase. With small pitch/yaw angles, most corners remain close to the image center, providing limited information about radial distortion. As pitch/yaw increases, the pattern shifts toward the image periphery, and the corners sample larger radial distances where distortion effects are stronger. Introducing a roll angle further breaks symmetry, spreading the corners unevenly across

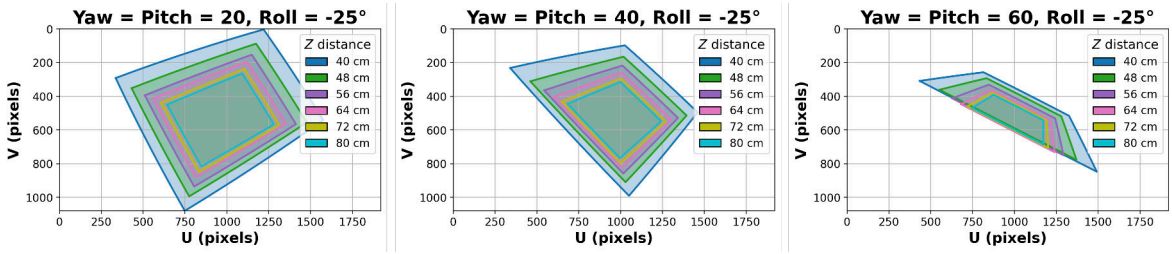


Figure 4.3: Effect of combined camera pitch, yaw, roll and depth on calibration pattern projection.

the image and improving the perception of radial distortion. Together, larger pitch/yaw angles and asymmetric poses provide richer radial evidence, improving the robustness of distortion parameter estimation in a single-image calibration approach.

4.3.3 Optimal pose selection

Considering all of the above, the following question arises: *Is there an optimal camera-to-object relative pose for camera calibration?* This is especially relevant in the case of single-image calibration.

The conditioning of the optimization equations to estimate the parameters is complex, since their effects are often coupled. These include the three translational parameters (t_x, t_y, t_z) representing lateral offset, vertical offset, and distance from the camera, and three rotational parameters (*pitch*, *yaw*, *roll*) defining the orientation of the pattern. Each of these parameters influences the quality of the geometric evidence available for parameter estimation.

Geometric conditioning

The main features required for good geometric conditioning are, first, *frame visibility*, as all four corners must remain visible within the image boundaries; second, *image fill*, since the calibration pattern should occupy as much portion of the image as possible, providing stronger geometric constraints that help decouple the parameters; and finally, *radial distance distribution*, meaning that the corners should span a sufficient and evenly-distributed range of radial distances from the principal point to enable reliable decoupling of distortion from focal length estimation and camera-to-pattern range.

Breaking radial symmetry: The most difficult task relies on obtaining a good radial distance distribution across the corners of the A4 paper sheet. There are different ways of breaking the radial symmetry among the points, but not all of them do it with the same outcome.

Translational asymmetry can be used to achieve radial asymmetry by, for example, laterally shifting the paper along the x-axis. However, it introduces bias, since the system becomes better conditioned for estimating one coordinate of the principal point (c_x), but not the other (c_y). Since the paper should remain fairly centered in the image to achieve optimal focal length conditioning, a displacement along both axes is not ideal.

A more effective approach is to introduce *rotational asymmetry* through coupled pitch and yaw rotations. By rotating the calibration sheet simultaneously around both the x- and y-axes with similar angles, the sheet moves away from the optical axis in a balanced manner. Neither c_x nor c_y becomes preferentially conditioned, and radial symmetry is broken. Additionally, the roll (around the z-axis, perpendicular to the image plane) further spreads the radial distances of the corners relative to the principal point. This enhances the sensitivity of the estimation to both focal length and distortion parameters, especially when the calibration pattern has an asymmetric aspect ratio, such as A4 paper.

An important practical consideration is the dependence between rotation angles and the positioning of the pattern within the image. When the calibration sheet is tilted (pitch rotation), perspective projection causes the sheet to appear shifted vertically in the image, tilting away from the camera moves the projection toward the top of the frame. Similarly, yaw rotation introduces horizontal displacement. To maintain a good radial distance distribution while maintaining all corners within the image, the sheet's position must be adjusted through a small lateral translation (t_x, t_y) to compensate for this perspective shift. The simulation tool presented with this work accounts for these dependencies by optimizing (t_x, t_y) displacements to ensure centering of the calibration object in the image.

Recommended angular configuration

Through systematic simulation across the parameter space, it has been identified that oblique angles in the range of 30° to 50° for pitch and yaw (-25° for roll) typically provide robust calibration performance. These angles offer several benefits.

Firstly, sufficient perspective evidence is generated, as the oblique projection produces enough perspective distortion to decouple focal length from distance and object size. Secondly, adequate coverage is achieved, since the projection fills a

considerable portion of the image when the range (t_z) is appropriately chosen. Finally, radial asymmetry is ensured, as the corners distribute across a wide range of radial distances, providing strong constraints on distortion parameters.

More extreme angles (>60°) can cause corners to exit the frame or create excessive perspective distortion that reduces image coverage. On the other hand, small angles (<20°) fail to introduce enough perspective evidence, leading to parameter coupling and poor conditioning.

Camera-dependent optimization

Optimal pose configuration depends critically on the camera specifications as different cameras exhibit different trade-offs.

On the one hand, *wide-angle lenses* naturally capture a larger field of view and depth of the scene, but they exhibit strong radial distortion; consequently, they require less extreme rotation angles to generate sufficient perspective evidence. On the other hand, *narrow-angle lenses* have minimal inherent distortion and weaker perspective effects, and therefore benefit from more aggressive oblique poses to generate the necessary geometric constraints [10]. Finally, *the sensor's aspect ratio* affects the required balance between horizontal and vertical angles, as cameras with wide aspect ratios may require different pitch-to-yaw ratios compared to square sensors.

This camera-specific variability motivates the need for a tool that allows users to explore the parameter space for their particular cameras.

Several key observations emerge from these visualizations. First, the optimal rotation angles differ across camera types: about 45° for standard lenses (Fig. 4.4), 50° for narrow-angle lenses (Fig. C.1), and 35° for wide-angle lenses (Fig. C.2), see Appendix C. This shows that, the pose should balance perspective evidence with image coverage: wide-angle lenses need only moderate tilts, whereas narrow-angle cameras benefit from more oblique views.

Second, the optimal distance t_z depends on both angle and camera type. As rotation increases, foreshortening reduces the projected size of the pattern, so the optimal distance shifts closer to preserve image fill. Wide-angle cameras generally require larger distances to keep all corners within the frame, while narrow-angle cameras operate closer to achieve comparable coverage.

It was found that, typically, an optimal pose chooses a distance that fills the image with the pattern and includes yaw, tilt, and roll angles for oblique projection of approximately 30° to 50°, as was already noticed in [10] with the motivation of maximizing perspective distortion in calibration images.

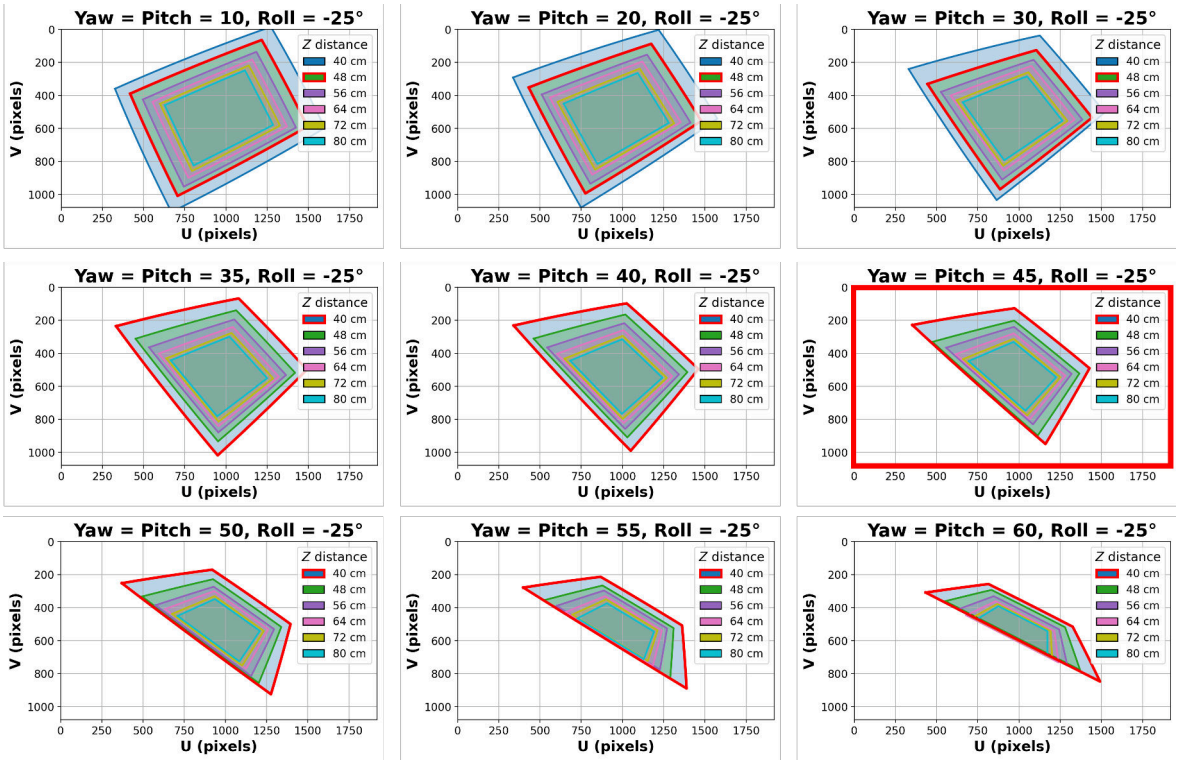


Figure 4.4: Calibration pattern image projection for increasing camera pitch/yaw angles (yaw = pitch from 10° to 60°) at different depths (40 to 80 cm). For each pitch/yaw angle, the red contour indicates the depth that yields the best calibration result. Among all configurations, the pose whose subplot exhibits the best calibration is chosen as the optimal camera set-up (external red contour).

4.4 Methodology

4.4.1 Calibration pipeline

The pipeline implemented in this study follows a three-stage process:

1. **Pose simulation (research step):** Before computing any homography, a synthetic pose grid is generated to identify configurations that provide strong perspective and radial-distortion evidence. The analysis explores ranges of tilt angles and camera-to-pattern distances and, for each candidate configuration, evaluates small lateral translations to ensure full visibility of the calibration sheet while avoiding symmetric projections. This step yields a set of poses that are well-conditioned for a one-shot initialization.

2. **Homography extraction and parameter initialization:** From the four corner points of the calibration sheet, whose real-world dimensions are known, and their detected 2D image projections, a planar Direct Linear Transformation (DLT) system is built to estimate via Singular Value Decomposition (SVD) the homography $\mathbf{H}$ relating the calibration plane to the image. Under the pinhole

camera model, $\mathbf{H}$ provides algebraic estimates of the focal length f and the camera-to-pattern pose parameters $(\mathbf{R}, \mathbf{t})$. These estimates are not optimal but serve as a consistent initialization for the following nonlinear refinement.

3. **Non-linear refinement:** A Levenberg-Marquardt least-squares optimization minimizes the reprojection error, defined as the Euclidean distance between observed image points and those projected by the camera model.

4.4.2 Novel focal length initialization

As part of the homography-based parameter initialization described in the previous section, a dedicated strategy is adopted to obtain a robust initial estimate of the focal length when operating under minimal data conditions.

The proposed calibration approach follows Zhang's method but operates on minimal data. Following Strobl and Lingenauber [22], main point centralization is employed to decouple the intrinsic matrix from the main point location. This simplification allows to derive a direct closed-form solution for the focal length from the homography matrix.

As it has been seen in Chapter 2, the relationship between the 3D world points on a plane and their 2D image points is defined by a homography matrix $\mathbf{H}$. The homography matrix $\mathbf{H}$ can be decomposed to retrieve the camera's intrinsic and extrinsic parameters. The calibration matrix is simplified through main point centralization from eq. (3.2) to

$$\mathbf{K} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.2}$$

Let the homography matrix be written as

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix},$$

where h_{ij} denote the individual entries of the homography matrix, defined up to an arbitrary scale factor. In the following, only the first two columns of $\mathbf{H}$ are involved, as they are directly related to the first two rotation vectors of the camera pose.

With this simplified calibration matrix, the two constraints from the orthonormality of the rotation matrix, eq. (2.3), can be written explicitly in terms of the homography entries as

$$1. \quad h_{11}h_{12} + h_{21}h_{22} + h_{31}f^2h_{32} = 0$$

$$2. \quad h_{11}^2 + h_{21}^2 + h_{31}^2f^2 = h_{12}^2 + h_{22}^2 + h_{32}^2f^2$$

Now, solving for the focal length for each the equations yields:

$$\boxed{f_1 = \pm\sqrt{\frac{-(h_{11}h_{12} + h_{21}h_{22})}{h_{31}h_{32}}}} \qquad \boxed{f_2 = \pm\sqrt{\frac{h_{12}^2 + h_{22}^2 - h_{11}^2 - h_{21}^2}{h_{31}^2 - h_{32}^2}}}$$

Although both f_1 and f_2 satisfy the orthonormality constraints algebraically, we keep the solution f_2. This choice is motivated by its better numerical conditioning w.r.t. perspective distortion, as also reported by Strobl and Lingenauber [22], who show that f_2 remains more stable.

With the selected focal length, the calibration matrix is known, eq (4.2), the homography matrix can be decomposed to recover the camera pose up to scale. Let the homography be expressed in column form as $\mathbf{H} = [\mathbf{h}_1\ \mathbf{h}_2\ \mathbf{h}_3]$. The first two columns are proportional to the first two rotation axes of the camera, while the third column encodes the translation vector.

Since the homography is defined up to an arbitrary scale factor, the rotation vectors r_1, r_2, and the translation vector t are recovered by normalizing the transformed columns, and the third rotation axis is obtained as the vector product of the first two to enforce orthonormality

$$\mathbf{r}_1 = \frac{\mathbf{K}^{-1}\mathbf{h}_1}{\|\mathbf{K}^{-1}\mathbf{h}_1\|}, \quad \mathbf{r}_2 = \frac{\mathbf{K}^{-1}\mathbf{h}_2}{\|\mathbf{K}^{-1}\mathbf{h}_1\|}, \quad \mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2, \quad \mathbf{t} = \frac{\mathbf{K}^{-1}\mathbf{h}_3}{\|\mathbf{K}^{-1}\mathbf{h}_1\|}.$$

This estimate is then used as initialization for the subsequent nonlinear refinement.

In summary, with the detection of the four corners of the A4 paper, the homography $\mathbf{H}$ can be estimated and from it initialize the focal length and camera pose parameters.

4.4.3 Extension to stereo camera calibration

The minimal Zhang method presented in this section can be naturally extended to stereo camera calibration. By acquiring a single stereo view of the calibration pattern, the intrinsic parameters and poses of both cameras w.r.t. the pattern can be initialized independently using the same procedure described above. These initial estimates are then jointly refined through nonlinear optimization.

Once the poses of both cameras relative to the calibration pattern are known, the rigid camera-to-camera transformation (the stereo baseline) can be directly obtained by composition of the two poses. In contrast to traditional stereo calibration, where the relative pose between the cameras must be explicitly parameterized and constrained during optimization, the proposed approach estimates one 6-DoF pose per camera without requiring additional modifications to the optimization formulation.

Chapter 5

Zero-shot calibration from angular constraints

In this work, the term *zero-shot* refers to the absence of calibration images and calibration objects, rather than the absence of measurements. Unlike classical and one-shot calibration approaches, the proposed method does not rely on captured images of a scene or on feature detection. Instead, the camera is used as an angular measurement device, providing angular field-of-view and ray-angle information without explicitly taking images.

5.1 Motivation and relation to the minimal Zhang method

Camera calibration is typically performed using multiple views of a known pattern (like a checkerboard). This yields accurate results, but it requires several images, an accurate calibration pattern, and image processing (corner detection). The proposed one-shot calibration method simplifies these requirements. However, it still relies on images of a known calibration pattern to establish the needed metric correspondences.

Ideas on calibrating without taking images of a known calibration pattern have appeared in the context of implicit calibration, where camera parameters are not explicitly recovered but estimated through projective constraints [24]. There are also calibration methods based on geometric scene regularities, such as straight lines, that further illustrate that explicit calibration patterns are not strictly necessary [15]. In this chapter, we explore whether camera calibration can be further simplified by eliminating the need of using a calibration pattern and instead exploiting angular camera constraints [5], such as the angular field of view.

An early and closely related approach was proposed by Kolesnik [25], who demonstrated that the internal camera parameters can be recovered directly from

known angular constraints using a single-image. While similar to the zero-shot formulation adopted in this work, that method keeps the observed laser cross within the interior of the image, avoiding measurements near the image borders and limiting the range of radial distances considered. At the other end, the recent approach by Lee in [5] only uses the sensor angular field of view (i.e., measurements at the borders of the images, or rather its specifications) yet it does not include a radial distortion model that is suitable for computer vision.

The approach proposed in this chapter exploits large incident angles and measurements near the image edges, where radial distortion effects are more dominant. Our method can be interpreted as a fusion of the angle-based formulation of [25] with later approaches that emphasize the use of image borders, such as the method of Lee [5], while additionally incorporating explicit radial distortion modeling.

The proposed geometric approach is as follows: by considering angles between rays that pass through known image points, it is possible to constrain the camera intrinsic parameters by simply using angular measurements. This is opposed to the traditional approach, which also relies on the pose (6 DoF) of a known calibration object in front of the camera, for N calibration images. In this way, $6 \times N$ of the traditional $4+6\times N$ optimization parameters are no longer required, which significantly simplifies the optimization problem and avoids the coupling between parameters with broadly similar effects in projections like the focal length, the object range, and the radial distortion parameters. These angular measurements, each consisting of two 2D image locations and the measured outwards angle relating these projections, are then used as the basis for a non-linear optimization that estimates the intrinsic camera parameters.

5.2 Information without a calibration pattern

We have seen that, when a calibration pattern is used, the required information is obtained with metric correspondences between known 3D points (or 2D points when using a planar object) and their 2D image projections. These correspondences allow the estimation of homographies and camera poses, which constrain the intrinsic parameters.

In the absence of a calibration pattern, the metric correspondences are no longer available. Therefore, the main question is: what information remains available in a single-image that can still constrain the camera intrinsics?

A key insight from the zero-shot calibration approach is that the focal length and radial distortion are directly related to the angular field of view of the camera (see Section 3.1.1). Specifically, the radial distance is

$$r_{\text{radial}} = r_{pers} + r_{dist} = f\tan(\theta) + fk_1\tan^3(\theta) = f\tan(\theta)\left(1 + k_1\tan^2(\theta)\right). \qquad (5.1)$$

This relationship implies that any angular measurement directly constrains the intrinsic parameters, without requiring knowledge of scene geometry or camera pose. Therefore, if angular information is available, the estimation of the intrinsic parameters of the camera from scratch is possible even without a calibration object.

5.3 Methodology

The proposed zero-shot calibration method estimates the intrinsic parameters of a camera using only angular constraints, without relying on a calibration pattern, scene geometry, or camera pose estimation.

5.3.1 Problem setup

Given a set of image points $(u_i, v_i)_{i=1}^N$ in a single-image, each point corresponds to a ray from the camera center, with its direction defined by an incidence angle θ_i relative to the optical axis.

The goal of the zero-shot calibration is to estimate the four-dimensional intrinsic parameter vector

$$\mathbf{x} = (f, c_x, c_y, k_1).$$

5.3.2 Angular measurement model and cost function

The input for this angular calibration approach are angular measurements. An angular measurement refers to a pair of image points (u_i, v_i) and (u_j, v_j) together with the measured angle α_{ij} between their corresponding viewing rays.

Under the pinhole camera model, each image point (u, v) defines a unique viewing ray originating at the camera center. This ray can be recovered by undistorting the pixel coordinates, mapping them to normalized camera coordinates, and normalizing the resulting vector to unit length.

The angular separation between two viewing rays is then computed as

$$\alpha_{ij} = \arccos(\mathbf{d}_i \cdot \mathbf{d}_j) \tag{5.2}$$

where $\mathbf{d}_i$ and $\mathbf{d}_j$ are the unit direction vectors associated with image points (u_i, v_i) and (u_j, v_j), respectively.

We formulate the calibration problem as a nonlinear least-squares optimization. Given N angular measurements $\{\alpha_{ij}^{\text{meas}}\}$, we estimate the parameters by minimizing:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \sum_{(i,j)\in\mathcal{P}} \left(\alpha_{ij}^{\text{pred}}(\mathbf{x}) - \alpha_{ij}^{\text{meas}}\right)^2, \tag{5.3}$$

where $\mathbf{x} = [f, c_x, c_y, k_1]^\top$, $\alpha_{ij}^{\text{pred}}(\mathbf{x})$ is the predicted angle given the current parameter estimates and $\mathcal{P}$ denotes the set of selected point pairs used to form the angular measurements.

5.3.3 Direct focal length initialization from the angular field of view

Before introducing radial distortion, an initial estimate of the focal length can be obtained directly from angular field-of-view (AoV) measurements under the pinhole camera model. Assuming square pixels and neglecting lens distortion, the focal length is uniquely related to the horizontal or vertical field of view and the image dimensions. In particular, for a known image width W and horizontal field of view θ_h, the focal length can be recovered with eq. (3.4).

This closed-form initialization provides a physically meaningful starting point for the subsequent angular optimization that incorporates the radial distortion parameter and main point coordinates.

5.3.4 Decoupling focal length and radial distortion

A fundamental challenge in zero-shot calibration is the coupling between focal length f and radial distortion k_1. This coupling arises from the mathematical form of the radial projection model, see eq. (5.1). The equation shows that the radial distance consists of two terms:

- A **linear term** $f\tan(\theta)$ representing the perspective projection.
- A **cubic term** $fk_1\tan^3(\theta)$ representing the radial distortion by the lens.

When the selected angle measurements span a narrow range, the variation of $\tan(\theta)$ is limited, making it difficult to distinguish between the two terms.

Figure B.2a (Appendix B) illustrates this problem: a single angular measurement at 45° can be explained by a continuous family of (f, k_1) solutions. All curves pass through the same measured point, but predict different intrinsics at other angles.

The key to decoupling f and k_1 is **radial angular diversity**. Different angle ranges provide sensitivity to different parameters (see Appendix B):

- **Small angles** ($\theta < 20$): At small angles, $\tan(\theta)$ is small but $\tan^3(\theta)$ is negligible. The radial distance is dominated by the linear term $f\tan(\theta)$, conditioning f.
- **Large angles** ($\theta > 40$): At large angles, $\tan^3(\theta)$ grows rapidly. The cubic term $fk_1\tan^3(\theta)$ becomes significant, providing conditioning to k_1.

This is why the best-conditioned solutions require combining measurements from diverse radial angles. Without large angles, there is insufficient evidence of radial distortion. Similarly, without moderate angles, k_1 cannot be distinguished from f.

5.3.5 Minimal number of pair-angle measurements

In the proposed pair-angle formulation, each measurement is defined by selecting a pair of image points (i, j) and evaluating the angular separation between their corresponding viewing rays.

The camera parameters are estimated by minimizing the discrepancy between predicted and measured angles using a nonlinear least-squares formulation. Collecting all residuals, calculated with the equation 5.4, the parameter estimate is obtained as shown in eq. (5.3).

$$\mathbf{r}(\mathbf{x}) = \begin{bmatrix} \alpha_1^{\text{pred}}(\mathbf{x}) - \alpha_1^{\text{meas}} \\ \vdots \\ \alpha_N^{\text{pred}}(\mathbf{x}) - \alpha_N^{\text{meas}} \end{bmatrix} \in \mathbb{R}^N, \tag{5.4}$$

The question of how many pair-angle measurements are required to constrain the parameter vector can be addressed by examining whether the available measurements provide sufficient independent constraints. In the proposed formulation, the camera model is described by four intrinsic parameters,

$$\mathbf{x} = (f, c_x, c_y, k_1),$$

which must be simultaneously estimated from the angular measurements.

Let $\mathbf{J}(\mathbf{x}) = \partial \mathbf{r} / \partial \mathbf{x} \in \mathbb{R}^{N \times 4}$ denote the Jacobian of the residual vector w.r.t. these four parameters. A necessary condition for the problem to be locally solvable is that the Jacobian has full column rank,

$$\operatorname{rank} \mathbf{J}(\hat{\mathbf{x}}) = 4. \tag{5.5}$$

This implies that fewer than four independent pair-angle measurements cannot determine the four unknown parameters in general, since the rank of the Jacobian is bounded by the number of residuals, and each angular measurement yields one residual equation.

Degrees of freedom and conditioning

It is important to distinguish between the theoretical number of degrees of freedom and the practical conditioning of the estimation problem. With the degrees of freedom of our system in mind, four independent angular measurements are needed to estimate the four unknown parameters (f, k_1, c_x, c_y). However, this minimal count can only

guarantee that the system is algebraically determined, not that the solution is stable or robust to noise. In practice, the conditioning of the problem depends strongly on the geometric configuration of the selected measurements. Satisfying this minimal count is not sufficient on its own. Even when $N = 4$, the Jacobian may become rank-deficient or poorly conditioned depending on the geometric configuration of the selected point pairs.

Achieving a full-rank and well-conditioned Jacobian requires selecting pair-angle measurements that provide complementary geometric information. In particular, the selected pairs should span different incidence angles to decouple focal length from radial distortion, and cover multiple image directions so that variations in the principal point cannot be compensated by changes in the remaining parameters.

5.4 Angle measurement methods

A wide range of methods can be employed to measure angles, ranging from low-cost manual techniques to hardware-assisted approaches. Each method has its own trade-off between feasibility and measurement accuracy.

5.4.1 Low-precision measurement with minimal hardware

For some applications, approximate angular measurements are sufficient. In these cases, simple and easily available tools or methods can be used.

Mechanical ruler or protractor. The angle is estimated manually using a ruler, protractor, or printed angular markings on the experimental setup. While direct, this method is affected by alignment errors, parallax, and limited resolution.

Smartphone sensors. Modern smartphones provide inertial sensors capable of estimating orientation angles. These sensors enable fast and convenient measurements, but they are prone to sensor noise, bias drift, and limited accuracy. As a result, their use is limited to unrefined angle estimation or experimental setup verification.

Camera specifications sheet. Angular fields of view (AoVs) can be provided by manufacturers in data sheets. In fact, Lee in [5] used these values to parameterize a camera model; however, the used distortion model was unsuitable for computer vision applications, both due to its limited adoption and because the distortion center was not estimated but fixed at the image center. Manufacturers often only list generic horizontal and vertical AoVs. Because cameras and lenses are typically

acquired separately, relying on these datasheet specifications is rarely useful for precise calibration.

In these methods, angular measurements in the image are commonly derived from points located at the natural image boundaries, in particular the image corners. This is due to the simplicity and ease of identification, especially when the measurement relies on visual or manual precision instead of image processing.

5.4.2 High-precision calibration with additional hardware

Hardware-based approaches typically yield higher precision angular measurements.

LED-Based geometric angle estimation. A commonly used laboratory technique consists of mounting multiple LEDs on a rigid structure attached to either the calibration target (e.g. a wall) or the camera mount, refer to [26] for a setup with two LEDs. The relative 3D positions of the LEDs w.r.t. the camera are known with high accuracy.

After one or several image acquisitions, LEDs are detected robustly and precisely as high-contrast point features. Then, their 2D image coordinates are matched to the known 3D sources. Ray directions are estimated from the 2D image coordinates, based on an initial, rough camera model. Finally, the camera model is optimized for the estimated angles between directions to match the measured angles.

This method provides high angular accuracy and repeatability. Its main limitations are the experimental complexity and the initial geometric calibration of the LED rig.

In contrast to manual or visually guided methods, this approach removes the restriction of relying exclusively on points located at the image boundaries. Angular measurements can be obtained from point sources distributed across the entire image, enabling more uniform angular coverage.

Chapter 6

Results

In this chapter, we showcase our implementations of the methods contributed in Chapters 4 and 5.

6.1 Results of minimal Zhang calibration method

This section presents the experimental evaluation of the proposed one-shot calibration method in Chapter 4, analyzing parameter dependencies and robustness under different geometric conditions.

6.1.1 Standard approach

First, the convergence behavior of the proposed one-shot calibration method is analyzed over 40 independent trials under baseline conditions, which are yaw=pitch (for all angles), roll=-25°, typical image processing noise (0.5 px standard deviation), and the standard initialization of parameters (f_{est}, r_{est}, t_{est}, $k_1 = 0$, $c_x = W/2$, $c_y = H/2$) explained in Section 4.4.2. The ground-truth parameters used in the simulations are chosen to represent a typical consumer camera and are set to $f = 1500$ px, $k_1 = -0.2$, and a full-HD image resolution of $W = 1920$ px and $H = 1080$ px.

With the aim of finding the optimal camera-to-pattern calibration poses for this method, we will simulate an extensive set of orientations and ranges. Since the method's outcome naturally depends on the relative pose, convergence is not expected for every arbitrary pose. The goal of this analysis is therefore to identify pose configurations that provide sufficient geometric constraints.

Consistent convergence is observed for intermediate pitch/yaw angles, approximately between 35° and 55°, see Figure 6.1 and Table 6.1. In this range, the focal length converges to the ground-truth for different depth values, particularly at smaller Z distances.

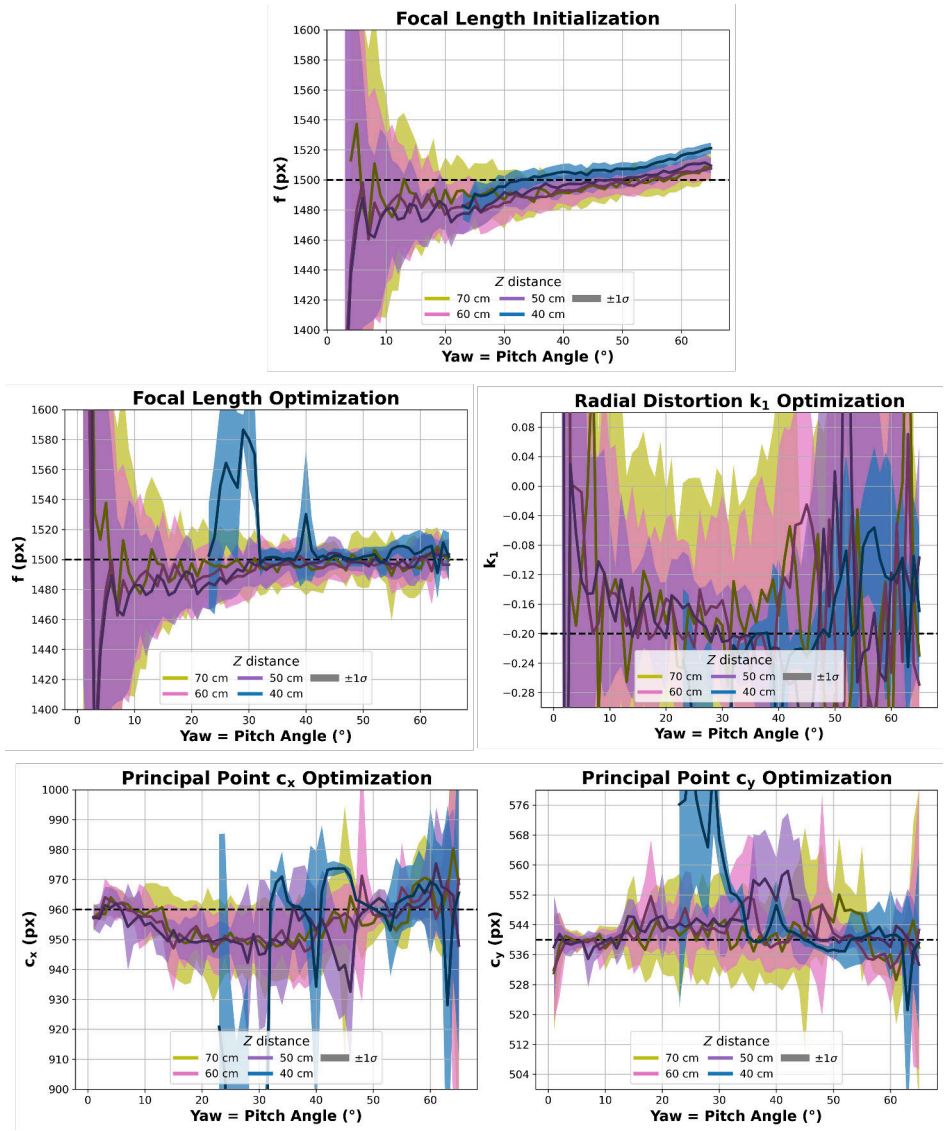


Figure 6.1: Standard minimal Zhang approach calibration results across pose angle and depth. The estimated focal length is shown for the initialization of the optimization. Curves are reported for several target depths (Z), as a function of the tilt angle (yaw=pitch) with roll -25°.

Pitch/Yaw (°)	**Z (m)**	**f (px)**	**k_1**	**c_x (px)**	**c_y (px)**
35	0.40	1501.33 ± 5.54	-0.21 ± 0.02	962.89 ± 3.49	540.93 ± 2.09
35	0.70	1491.59 ± 15.23	-0.16 ± 0.17	950.30 ± 13.22	541.08 ± 12.97
45	0.40	1502.60 ± 4.95	-0.23 ± 0.03	973.62 ± 1.25	539.85 ± 1.73
45	0.70	1501.42 ± 11.29	-0.23 ± 0.13	952.49 ± 41.96	540.10 ± 29.85
55	0.40	1506.87 ± 5.37	-0.08 ± 0.07	963.16 ± 10.56	540.51 ± 7.00
55	0.70	1489.23 ± 30.29	-0.19 ± 0.73	955.39 ± 8.71	546.75 ± 18.16

Table 6.1: Estimated calibration parameters obtained using the standard minimal Zhang approach.

Principal point estimation exhibits higher sensitivity. While c_x and c_y remain close to the ground-truth at small angles, increased perspective distortion at larger pitch/yaw angles leads to higher estimation variance and stronger depth dependence. This behavior can be explained by the spatial distribution of the projected points. Strong perspective distortion tends to produce compact and uneven point configurations, which provide limited radial coverage of the image. Since the principal point defines the center of the radial distortion model, insufficient radial distribution makes its estimation more sensitive and increases variance.

The radial distortion coefficient k_1 remains the most weakly constrained parameter. While projections near the image borders (large radial angles) are strictly necessary to robustly estimate higher-order distortion terms such as k_2, even the first-order coefficient k_1 benefits from increased radial coverage. In minimal single-image settings, limited radial diversity leads to stronger coupling between k_1, focal length, and depth.

Overall, these results show that the proposed method converges consistently under specific pose configurations that provide sufficient perspective and radial evidence, while exposing the geometric limits of parameter observability in minimal calibration scenarios.

6.1.2 Camera type

To evaluate the generalization capability of the method, two additional camera models are evaluated: a narrow-angle lens ($f = 2200$ px, $k_1 = -0.05$) and a wide-angle lens ($f = 800$ px, $k_1 = -0.35$).

For the narrow-angle configuration, convergence behavior remains comparable to the baseline model (see Figure C.3 in Appendix C). Stable f estimates are obtained in a similar pitch/yaw angle range, although the effective convergence region is slightly narrower and more sensitive to depth, due to the reduced angular field of view. Radial distortion estimation remains weak, as expected for lenses with mild distortion.

In contrast, the wide-angle configuration shows strong non-linear projection effects increasing the sensitivity of the optimization to depth and pose estimation, see Figure C.4 (Appendix C). In particular, principal point estimation becomes less stable across several angle/distance combinations, and focal length convergence is more pose-dependent than in the narrow-angle case. This indicates that, in minimal single-image settings, increased field of view can degrade conditioning of the convergence region.

6.1.3 Main point shift

The sensitivity to the displacement of the principal point is also studied. Two levels are considered: a moderate shift of 10 to 15 px and a larger shift of 20 to 25 px.

For moderate shifts, focal length estimation remains stable in pose configurations with sufficient perspective and radial distortion, see Figure C.5 (Appendix C). Although c_x and c_y show increased variability, values close to the ground-truth are recovered for most poses, indicating tolerance to small deviations from central projection.

Larger shifts, however, lead to a clear degradation across all parameters, see Figure C.6 (Appendix C). The increased dispersion arises from the loss of radial symmetry, which directly affects the assumed geometric center of the projection.

6.1.4 Image processing noise

Robustness to image processing noise is evaluated by increasing the Gaussian-distributed corner localization error to standard deviations of 1 px and 2 px.

As the noise level increases, the variance of all estimated parameters grows, with the radial distortion coefficient k_1 being the most affected. This behavior arises because, when the corner detection error becomes comparable to or larger than the displacement induced by radial distortion, the geometric evidence required to constrain k_1 is effectively masked by the noise.

Despite this degradation, focal length estimation remains stable within a consistent intermediate range of camera poses across all noise levels. In contrast, principal point estimation becomes increasingly sensitive to noise, particularly at larger tilt angles. This is expected, as the main role of the principal point is to fix the origin of the radial lens distortion [27]. This can be observed in Figures C.7 and C.8 (Appendix C).

Overall, these results indicate that the proposed method tolerates moderate image processing noise while preserving reliable focal length estimation, but that accurate radial distortion and principal point recovery requires noise levels sufficiently low to maintain observable radial displacement. Alternatively, more images have to be taken to be able to statistically compensate for these errors.

6.1.5 Camera poses

Finally, the influence of camera orientation on the stability of the one-shot calibration is analyzed by comparing pitch-only configurations with joint pitch/yaw configurations.

In the pitch-only case (Figure C.9, Appendix C), the calibration pattern is tilted exclusively around a single axis while remaining symmetric w.r.t. the image

center. Although increasing pitch introduces perspective distortion, this configuration preserves a high degree of geometric symmetry. As a result, the focal length can often be recovered with reasonable accuracy for moderate tilt angles, but the estimation of radial distortion and principal point parameters remains unstable and exhibits large variance. This behavior reflects the limited diversity of radial distances and ray directions induced by pure pitch motion.

When pitch and yaw rotations are combined (Figure C.10), this symmetry is partially broken, leading to a more heterogeneous distribution of projected points across the image. While focal length convergence remains comparable to the pitch-only case, the additional angular diversity provides improved geometric evidence for radial distortion estimation in some configurations. Overall, joint pitch/yaw rotations tend to improve, but not fully resolve, the conditioning limitations inherent to minimal single-image calibration.

These results highlight that camera orientation plays a critical role in minimal calibration scenarios. While pitch-only motions can be sufficient for focal length initialization, reliable estimation of the full intrinsic parameter set requires asymmetric poses that jointly excite multiple degrees of freedom. This observation motivates the use of coupled pitch/yaw and roll rotations as a practical guideline for one-shot calibration.

6.1.6 Stereo camera calibration

The proposed minimal Zhang formulation is extended to the stereo case by jointly calibrating two cameras from a single stereo observation of the planar calibration pattern. Both cameras are initialized independently using the one-shot homography-based approach and subsequently refined through a joint nonlinear optimization, while enforcing the rigid stereo geometry.

Figure 6.2a shows the calibration results for a stereo system with a baseline of 5 cm, while Figure 6.2b corresponds to a wider baseline of 10 cm. For both configurations, the focal length estimates of the left and right cameras converge reliably toward the ground-truth value as the pitch/yaw angle increases, confirming that sufficient perspective distortion is essential for stable initialization and optimization.

In the narrow-baseline configuration, the intrinsic parameters exhibit slightly higher variability at low tilt angles, particularly during the initialization stage. This behavior is expected, as smaller baselines reduce the geometric diversity between views and therefore provide weaker constraints on depth-related parameters. Nevertheless, after nonlinear refinement, the focal length and principal point estimates stabilize across a wide angular range.

The larger baseline configuration provides improved conditioning of the problem. As observed in Fig. 6.2b, the optimization stage yields faster convergence and reduced variance for both intrinsic and distortion parameters. The radial distortion coefficient k_1 is also more consistently recovered when larger incident angles are available, especially near the image borders.

Overall, these results demonstrate that the minimal Zhang method can be directly applied to stereo camera calibration without modifying the optimization structure. A single stereo view is sufficient to initialize both cameras and recover a consistent stereo geometry, provided that the calibration pattern is observed under sufficiently oblique viewing angles. This significantly reduces the data acquisition effort compared to traditional multi-view stereo calibration pipelines.

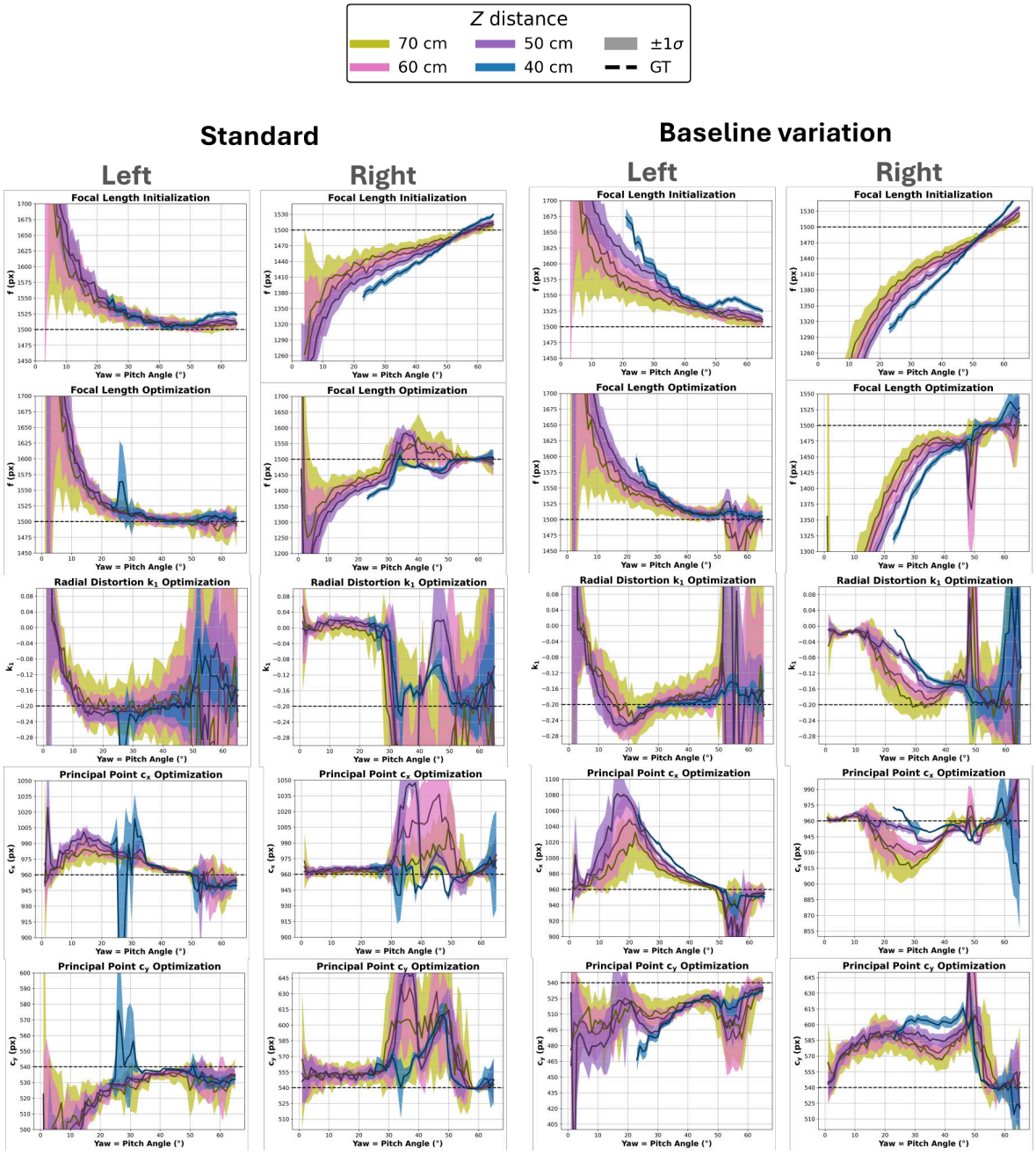
6.2 Zero-shot camera calibration results

This section analyzes the convergence behavior of the zero-shot calibration method proposed in Chapter 5, under different settings. These representative configurations are evaluated using the same experimental conditions: 20 repetitions for statistical measures, angular measurement noise of 0.2°, a set of angular measurements formed by point pairs selected from Figure 6.3, and the same set of released intrinsic camera parameters (f, c_x, c_y, k_1). With the variable settings, the importance of geometric conditioning for the zero-shot formulation will be revealed.

6.2.1 Well-conditioned optimization

The corresponding radial–angular representation $r(\theta)$ in Figures 6.4 and 6.5 shows that the selected measurements span a wide range of small-to-large incidence angles, while preserving a coherent and dispersed radial structure. Radially sampling camera directions provides strong geometric constraints, allowing the detached estimation of focal length, principal point, and radial distortion parameters at the same time. In particular, measurements at moderate and large angles contribute decisively to stabilizing the estimation of the distortion parameter k_1 by making the curvature in the $r(\theta)$ relationship observable.

Figure 6.4, four angle measurements, and Figure 6.5, six angle measurements, show that all parameters lie close to the ground-truth, indicating stable convergence to the ground-truth, although the six-measurement configuration yields more precise estimates.



(a) Stereo calibration with baseline of 5 cm (b) Stereo calibration with baseline of 10 cm

Figure 6.2: Comparison of minimal Zhang stereo calibration results for different baseline lengths.

6.2.2 Poorly-conditioned optimization

In contrast, this experiment analyzes failure cases of the zero-shot calibration method under geometrically weak configurations. All experiments use identical noise levels, number of repetitions, and optimization settings, differing only in the radial configuration of the selected measurements.

In the first configuration (Figure 6.6) only three angular pairs are used, which are

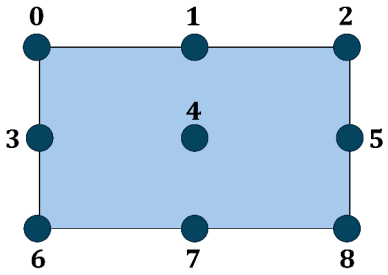


Figure 6.3: Image pixel points considered for the formation of point pairs and angular measurements.

insufficient to condition the system of equations. In particular, the vertical principal point coordinate c_y deviates significantly from the ground-truth, indicating strong parameter coupling due to under-minimal constraints. This parameter is the most weakly constrained due to the sensor geometry: since the image height is smaller than the width, the angular range in the vertical direction is reduced.

Increasing the number of angular pairs to four (Figures 6.7 and 6.8) does not resolve this issue when the additional measurements remain in a similar radial region, providing insufficient geometric diversity. As a result, focal length, radial distortion, and principal point estimates remain unstable and exhibit increased dispersion.

These results indicate that zero-shot calibration performance depends primarily on the radial diversity of the selected measurements rather than on their number alone. When the angular configuration fails to describe the curvature in the radial–angular relationship, the optimization becomes ill-conditioned since multiple parameter combinations can explain the observations similarly well.

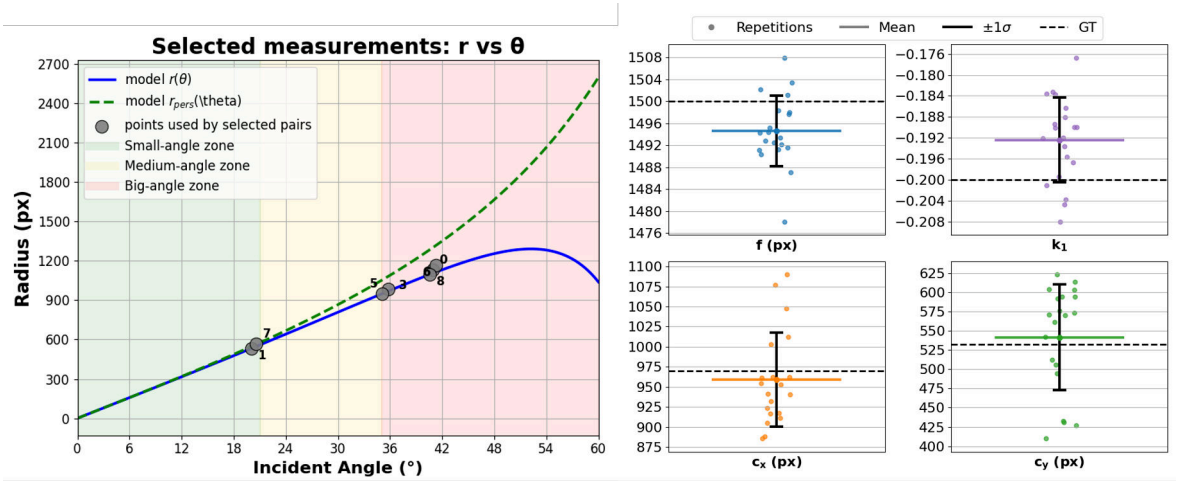


Figure 6.4: Zero-shot calibration under well-conditioned geometry using four angular pairs (the minimum number of measurements). The selected measurements span small-to-large incidence angles.

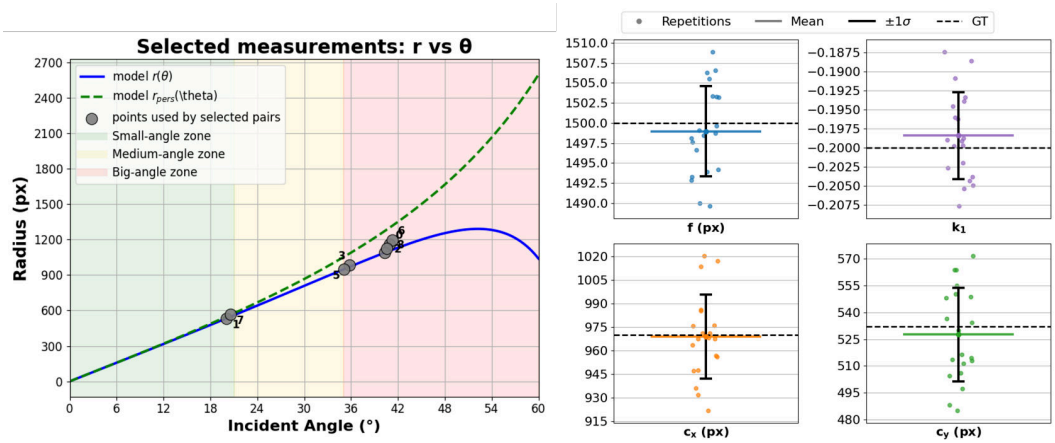


Figure 6.5: Zero-shot calibration under well-conditioned geometry using six angular pairs. The selected measurements span small-to-large incidence angles.

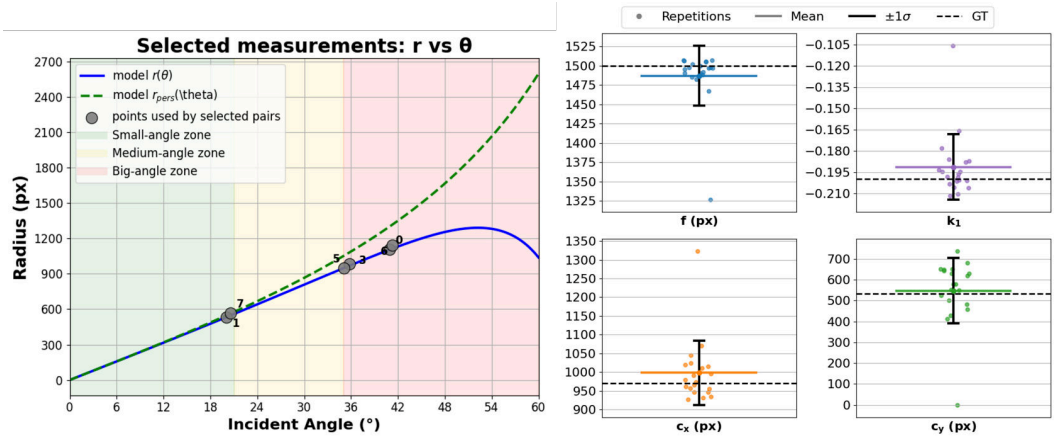


Figure 6.6: Zero-shot calibration under poorly conditioned geometry using only three angular pairs (an under determined system). The selected measurements span small-to-large incidence angles.

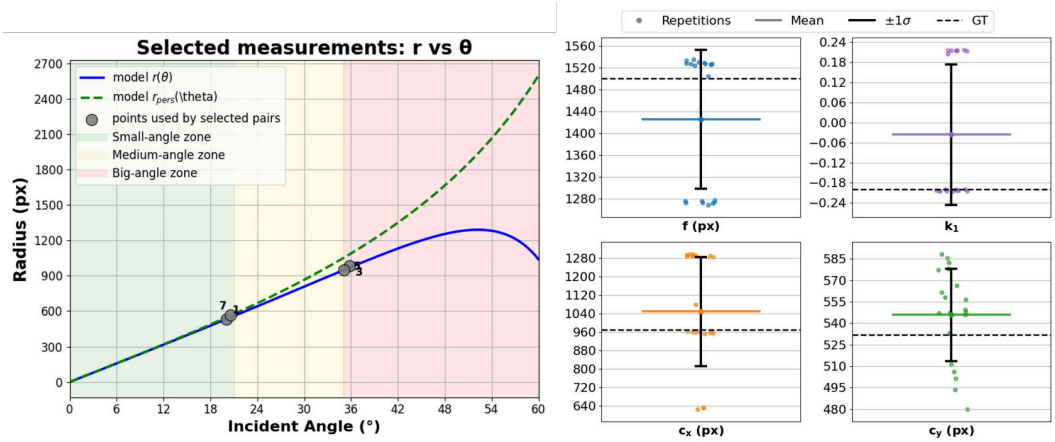


Figure 6.7: Zero-shot calibration under poor-conditioned geometry using four angular pairs (the minimum number of measurements). The selected measurements span a limited radial diversity and f and k_1 cannot be detached.

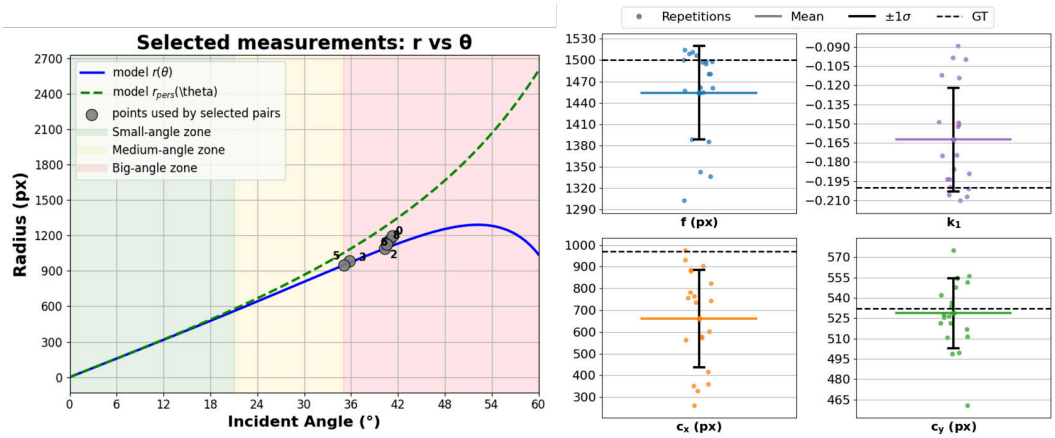


Figure 6.8: Zero-shot calibration under poor-conditioned geometry using four angular pairs (the minimum number of measurements). The selected measurements only span large incidence angles and f and k_1 cannot be uncoupled.

Chapter 7

Conclusions

This thesis has explored geometric camera calibration under minimal requirements, with the objective of reducing the complexity of traditional calibration procedures while preserving their practical usefulness and accuracy. Two approaches have been investigated: a one-shot calibration method based on a single image of a planar everyday object, and a zero-shot calibration method relying solely on angular constraints without the need for calibration images or patterns. Together, these approaches contribute to a deeper understanding of the fundamental information required for camera calibration and provide practical guidelines for operating under strongly reduced data conditions.

Beyond the specific methods proposed, this thesis highlights several broader insights into camera calibration:

1. Calibration is not only a problem of data quantity, but fundamentally a problem of information content. Poorly conditioned configurations cannot be compensated for by noise reduction alone.

2. Perspective distortion, often regarded as an inconvenience in vision applications, becomes a valuable source of information in minimal calibration scenarios.

3. Radial distortion estimation critically depends on sampling large incidence angles; measurements near the image center alone are insufficient.

4. Reducing calibration requirements shifts responsibility from redundancy to geometric design, making pose selection and measurement configuration crucial.

These insights are relevant not only for minimal calibration methods, but also for understanding failure modes and design choices in standard calibration pipelines.

7.1 Challenges and limitations

Despite its contributions, this work encountered several limitations. First, both proposed methods have been primarily validated through simulation. While this allows systematic exploration of geometric effects, real-world experiments would be necessary to fully assess robustness against unmodeled effects such as imperfect pattern planarity, sensor noise, or mechanical misalignment in angular measurements.

Second, the distortion model is limited to a single radial coefficient. While this choice is justified for many lenses and aligns with the goal of minimal parameterization, more complex optical systems may require higher-order models that are more difficult to constrain under minimal data conditions.

Finally, the zero-shot approach assumes access to angular measurements, which may require additional hardware or careful experimental setup. While this is acceptable in some contexts, it limits immediate applicability in purely vision-based consumer scenarios.

7.2 Future work

Several directions for future research naturally follow from this thesis. An immediate extension would be the experimental validation of both methods on real camera systems, including a quantitative comparison with standard calibration techniques.

Further work could also investigate adaptive strategies for automatic measurement selection, guiding users toward geometrically informative configurations in both one-shot and zero-shot settings. Finally, combining the geometric insights developed here with learning-based approaches could lead to hybrid methods that keep interpretability while benefiting from data-driven priors.

Bibliography

[1] Duane Brown. Close-range camera calibration. *Photogramm. Eng.*, 37, 12 1971.

[2] Roger Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, August 1987.

[3] Zhengyou Zhang. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22:1330 – 1334, 12 2000.

[4] Niclas Zeller, Franz Quint, and Uwe Stilla. A synchronized stereo and plenoptic visual odometry dataset, 07 2018.

[5] Jae-Yeong Lee. Zero-shot estimation of radial distortion of wide-angle camera. *2024 IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–8, 2024.

[6] Peter F. Sturm and Stephen J. Maybank. On plane-based camera calibration: A general algorithm, singularities, applications. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 432–437, Fort Collins, CO, USA, June 1999.

[7] OpenCV. Camera calibration and 3d reconstruction. `https://docs.opencv.org/4.x/dc/dbb/tutorial_py_calibration.html`. Accessed: January 2026.

[8] Azra Fetić, Davor Jurić, and Dinko Osmanković. The procedure of a camera calibration using camera calibration toolbox for matlab. *2012 Proceedings of the 35th International Convention MIPRO*, pages 1752–1757, 2012.

[9] K. H. Strobl, W. Sepp, S. Fuchs, C. Paredes, M. Smisek, and K. Arbter. DLR CalDe and DLR CalLab. `https://rm.dlr.de/callab`.

[10] Klaus H. Strobl, Wolfgang Sepp, and Gerd Hirzinger. On the issue of camera calibration with narrow angular field of view. *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 309–315, 2009.

[11] Jean Marc Lavest, Marc Viala, and M. Dhome. Do we really need an accurate calibration pattern to achieve a reliable camera calibration? In Hans Burkhardt

and Bernd Neumann, editors, *Computer Vision — ECCV'98*, pages 158–174, Berlin, Heidelberg, 1998. Springer Berlin Heidelberg.

[12] Klaus H. Strobl and Gerd Hirzinger. More accurate camera and hand-eye calibrations with unknown grid pattern dimensions. *2008 IEEE International Conference on Robotics and Automation*, pages 1398–1405, 2008.

[13] Klaus H. Strobl and Gerd Hirzinger. More accurate pinhole camera calibration with imperfect planar target. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV), 1st IEEE Workshop on Challenges and Opportunities in Robot Perception*, pages 1068–1075, Barcelona, Spain, November 2011.

[14] Thomas Schöps, Viktor Larsson, Marc Pollefeys, and Torsten Sattler. Why having 10,000 parameters in your camera model is better than twelve. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2532–2541, United States, 2020. IEEE - Institute of Electrical and Electronics Engineers Inc.

[15] Frédéric Devernay and Olivier Faugeras. Straight lines have to be straight automatic calibration and removal of distortion from scenes of structured environments. *Mach Vis Appl*, 13, 08 2001.

[16] Janne Heikkilä. Geometric camera calibration using circular control points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(10):1066–1077, 1997.

[17] Stephen J. Maybank and Olivier D. Faugeras. A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–151, 1992.

[18] Bill Triggs. Autocalibration from planar scenes. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Lecture Notes in Computer Science, pages 89–105. Springer, 1998.

[19] Javier Tirado-Garín and Javier Civera. Anycalib: On-manifold learning for model-agnostic single-view camera calibration, 2025.

[20] Andrea Fusiello. The pinhole camera model. In *Computer Vision: Three-dimensional Reconstruction Techniques*, pages 15–33, Cham, 2024. Springer International Publishing.

[21] D. Vivet, Jordi Vilà-Valls, Gael Pages, and E. Chaumette. Robust filter-based visual navigation solution with miscalibrated bi-monocular or stereo cameras. *Remote Sensing*, 14:1470, 03 2022.

[22] Klaus H. Strobl and Martin Lingenauber. Stepwise calibration of focused plenoptic cameras. *Computer Vision and Image Understanding*, 145:140–147, 2016. Light Field for Computer Vision.

[23] A.W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 1:I–I, 2001.

[24] Guo-Qing Wei and Song De Ma. Implicit and explicit camera calibration: theory and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5):469–480, 1994.

[25] Marina Kolesnik. Direct recovery of the camera internal parameters using known angles. In *Proceedings of the Asian Conference on Computer Vision (ACCV) 2002, Volume II*, pages 430–435. ACCV, 2002.

[26] Michal Smisek. Omnidirectional camera calibration report. Technical report, German Aerospace Center (DLR), Institute of Robotics and Mechatronics, 08 2011.

[27] Reg G. Willson and Steven A. Shafer. What is the center of the image? *Journal of the Optical Society of America A*, 11(11):16–29, November 1994.

Appendix A

Minimal point requirements for homography estimation

A planar homography describes a projective transformation between two planes. In the context of camera calibration, it relates points on a planar calibration object to their corresponding projections in the image.

A homography is represented by a 3×3 matrix $\mathbf{H}$, defined up to an arbitrary scale factor, such that for a point $\mathbf{X} = [X, Y, 1]^\top$ on a plane and its corresponding image point $\mathbf{x} = [u, v, 1]^\top$, the following relation holds:

$$\mathbf{x} = \mathbf{H}\,\mathbf{X}. \tag{A.1}$$

Since $\mathbf{H}$ has nine elements but is defined up to scale, effectively it has eight independent degrees of freedom.

Each point correspondence between the plane and the image provides two independent linear constraints on the entries of $\mathbf{H}$. Consequently, at least four non-collinear point correspondences are required to fully constrain the homography, since $4 \times 2 = 8$ constraints.

The system is underdetermined with fewer than four point correspondences, admitting infinitely many solutions. Conversely, four point correspondences yield a determined linear system whose solution defines the homography uniquely (up to scale), provided that the points are non-collinear.

In practice, homographies are estimated using more than four correspondences. However, the four-point case represents the minimal configuration necessary for homography estimation and is therefore of particular interest in this work on minimal calibration scenarios. This motivates the use of planar calibration patterns with four well-defined corner points in Chapter 4.

Appendix B

Disentangling focal length from radial lens distortion

The camera model used in this thesis (and largely in computer vision) is subject to two types of distortions: perspective and radial. The perspective distortion is defined by the relative camera-to-object pose and the focal length for scaling, while the radial distortion is parameterized by the radial distortion parameter, in our case only k_1. It is the combination of the two effects that we observe when taking a picture. Now, how can we distinguish one from another, disentangling the focal length from the lens distortion parameter?

This question is of great importance in our context, and even more important is the question: *What are the minimum requirements for decoupling the f and the k_1?*

B.1 Total radial image projection

To answer these questions, we begin by understanding how the total distortion is described in terms of the incidence angle with respect to the principal axis of the camera. The equation for the total radial image distortion from perspective projection and lens distortion is the following:

$$r = r_{\text{pers}} + r_{\text{radial}} = f \cdot \tan(\theta) + f \cdot k_1 \cdot \tan^3(\theta) = f \cdot \tan(\theta) \cdot (1 + k_1 \tan^2(\theta)) \quad \text{(B.1)}$$

where:

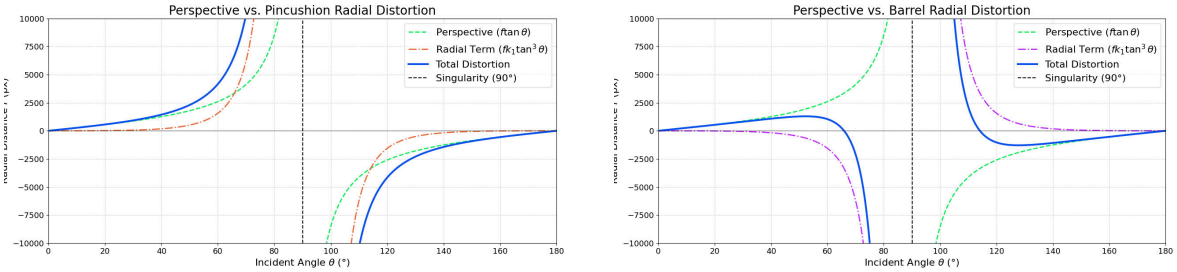
r: Radial distance from the principal point (pixels).

θ: Incident angle of the light ray.

f: Focal length (pixels).

k_1: First radial distortion coefficient.

The objective is to solve for the unknowns (f, k_1) with the least number of measurements of (r_i, θ_i).



(a) Pincushion radial distortion $k_1 > 0$ (b) Barrel radial distortion $k_1 < 0$

Figure B.1: Decomposition of the radial distance in perspective, radial distortion terms and total distortion (Eq. B.1) .

Figure B.1 illustrates the decomposition of the radial distance into its perspective and radial distortion components as a function of the incident angle θ, according to eq. B.1. In both cases, the perspective term $f \tan\theta$ dominates at small angles and grows rapidly as θ approaches 90°. The radial term $fk_1 \tan^3\theta$ becomes significant only at larger angles, where its cubic dependence amplifies deviations from the pure perspective model. For pincushion distortion ($k_1 > 0$, Fig. B.1a), the radial term reinforces the perspective expansion, leading to an increased total radial distance. Conversely, for barrel distortion ($k_1 < 0$, Fig. B.1b), the radial term compensates the perspective distortion. These plots highlight that decoupling of the focal length f and the distortion parameter k_1 requires measurements at sufficiently large incident angles, where both contributions exhibit distinct and observable behaviors.

B.2 One measurement

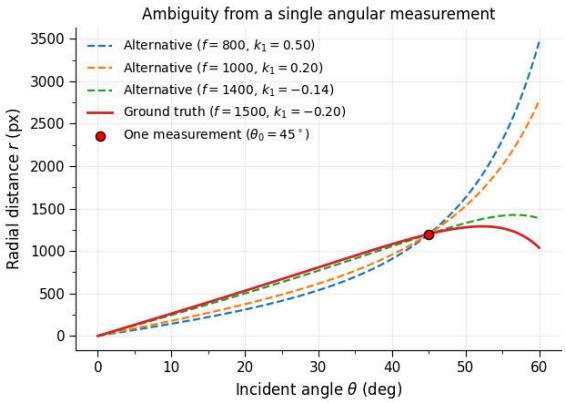
Ideally, we would like to use only one measurement to solve the system, but with one single measurement, the number of results that satisfy the system is infinite. Given a single observation (r_1, θ_1), substituting in eq. B.1:

$$r_1 = f \tan(\theta_1) + fk_1 \tan^3(\theta_1). \tag{B.2}$$

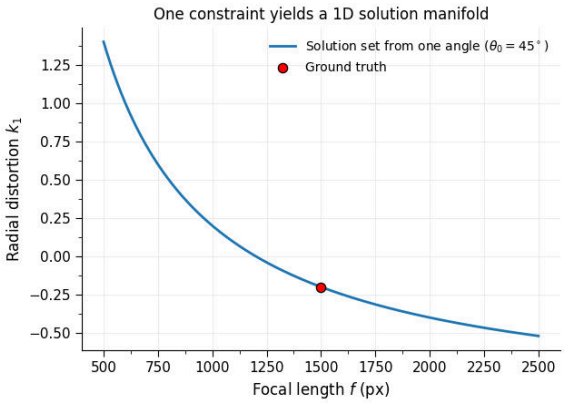
If we try to solve for k_1 in terms of f we get:

$$k_1(f) = \frac{\frac{r_1}{f \tan(\theta_1)} - 1}{\tan^2(\theta_1)} \tag{B.3}$$

such that for any arbitrary choice of focal length f (where $f \neq 0$), there exists a corresponding k_1 that perfectly satisfies the equation. This defines a family of curves that satisfy the system. Therefore f and k_1 are fully coupled and unobservable with one single measurement.



(a) Single angular measurement providing insufficient constraint on f and k_1



(b) Solution set of (f, k_1) satisfying the same angular constraint

Figure B.2: Non-identifiability of (f, k_1) from a single angular constraint.

B.3 Two measurements

Given two independent measurements (r_1, θ_1) and (r_2, θ_2):

$$\begin{aligned} r_1 &= f \tan(\theta_1) \left(1 + k_1 \tan^2(\theta_1)\right), \\ r_2 &= f \tan(\theta_2) \left(1 + k_1 \tan^2(\theta_2)\right). \end{aligned} \tag{B.4}$$

Dividing each equation by $\tan(\theta_i)$ yields

$$\begin{aligned} \frac{r_1}{\tan(\theta_1)} &= f \left(1 + k_1 \tan^2(\theta_1)\right), \\ \frac{r_2}{\tan(\theta_2)} &= f \left(1 + k_1 \tan^2(\theta_2)\right). \end{aligned} \tag{B.5}$$

Subtracting the two equations eliminates the focal length f and leads to

$$\frac{r_1}{\tan(\theta_1)} - \frac{r_2}{\tan(\theta_2)} = f k_1 \left(\tan^2(\theta_1) - \tan^2(\theta_2)\right). \tag{B.6}$$

Provided that $\theta_1 \neq \theta_2$, the system admits a unique solution for k_1:

$$k_1 = \frac{\frac{r_1}{\tan(\theta_1)} - \frac{r_2}{\tan(\theta_2)}}{f \left(\tan^2(\theta_1) - \tan^2(\theta_2)\right)}. \tag{B.7}$$

Substituting this result back into either equation allows the focal length f to be computed uniquely. Thus, two measurements at different incidence angles are enough to decouple the focal length from the radial distortion parameter.

It may seem that any two non-equal angles are enough for determining the focal length and distortion parameter. However, if we go back to the distortion graph, we see that not all angles pairs provide the same amount of information. In order to distinguish the contribution of perspective projection from the radial distortion, the angle measurements must be in different regions of the distortion curve.

For small incidence angles, the distortion term is negligible compared to the perspective term. On the other hand, measurements of larger incidence angles are more affected by radial distortion rather than by perspective distortion. Therefore, effective decoupling of focal length and radial distortion requires not only two distinct angles, but sufficient angular diversity such that both distortion terms are observable, see Figure B.3.

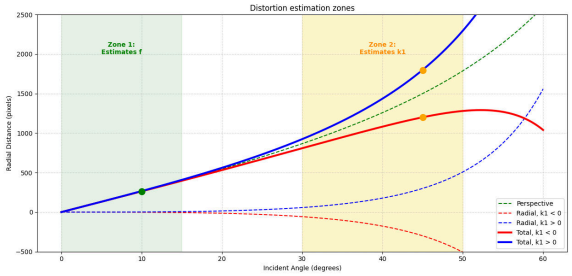


Figure B.3: Incident angle regions that condition best f and k_1.

B.4 Role of the principal point in decoupling

The derivation of eq. B.1 assumes that the radial distance r is measured with respect to the true principal point (c_x, c_y). Under this assumption, the distortion model depends only on the incident angle θ and is radially symmetric with respect to the main point.

If the principal point is incorrectly specified or unknown, this symmetry is broken and the measured radius becomes a function of both θ and the azimuth direction ϕ in the image. As a consequence, the relation $r = r(\theta)$ is no longer valid and additional directional information is required to describe the observed distortions.

Figure B.4 illustrates this effect by showing the measured radial distance for different principal point offsets. The deviation from the ground-truth curve increases with the angle, highlighting that accurate knowledge of (c_x, c_y) is a necessary condition for decoupling the focal length f and the radial distortion parameter k_1 from two angular measurements alone.

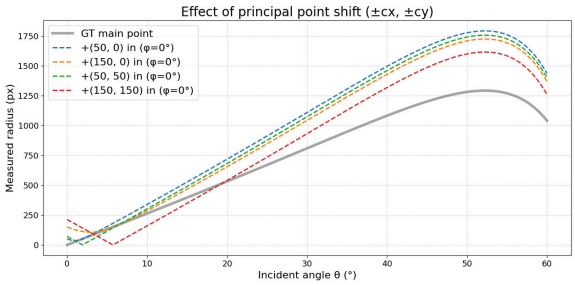


Figure B.4: Effect of the shift in the camera main point with respect to the image center.

Appendix C

Minimal Zhang method results

This chapter presents the experimental results obtained using the one-shot calibration approach under different camera configurations.

C.1 Best poses for different camera types

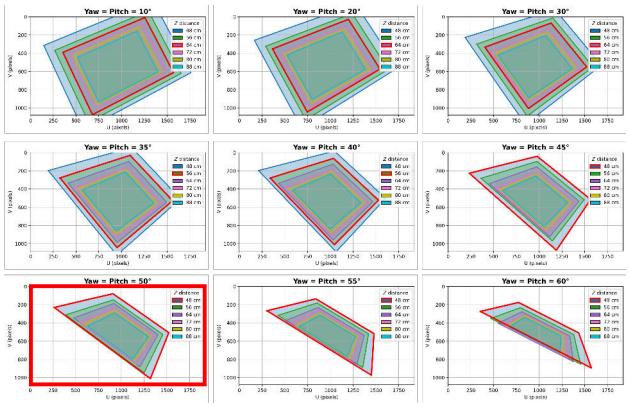


Figure C.1: Calibration pattern image projection with a narrow-angle camera.

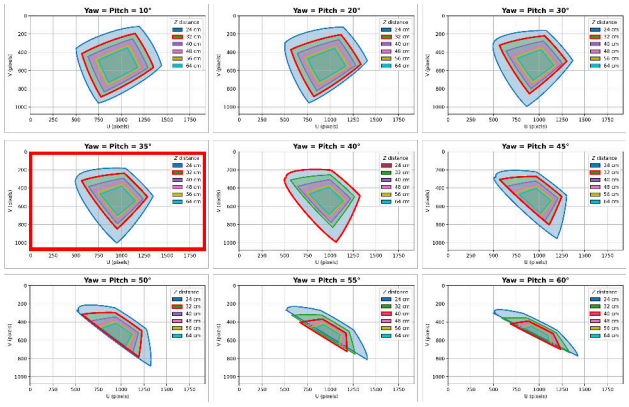


Figure C.2: Calibration pattern image projection with a wide-angle camera.

C.2 Camera type

Narrow-angle camera

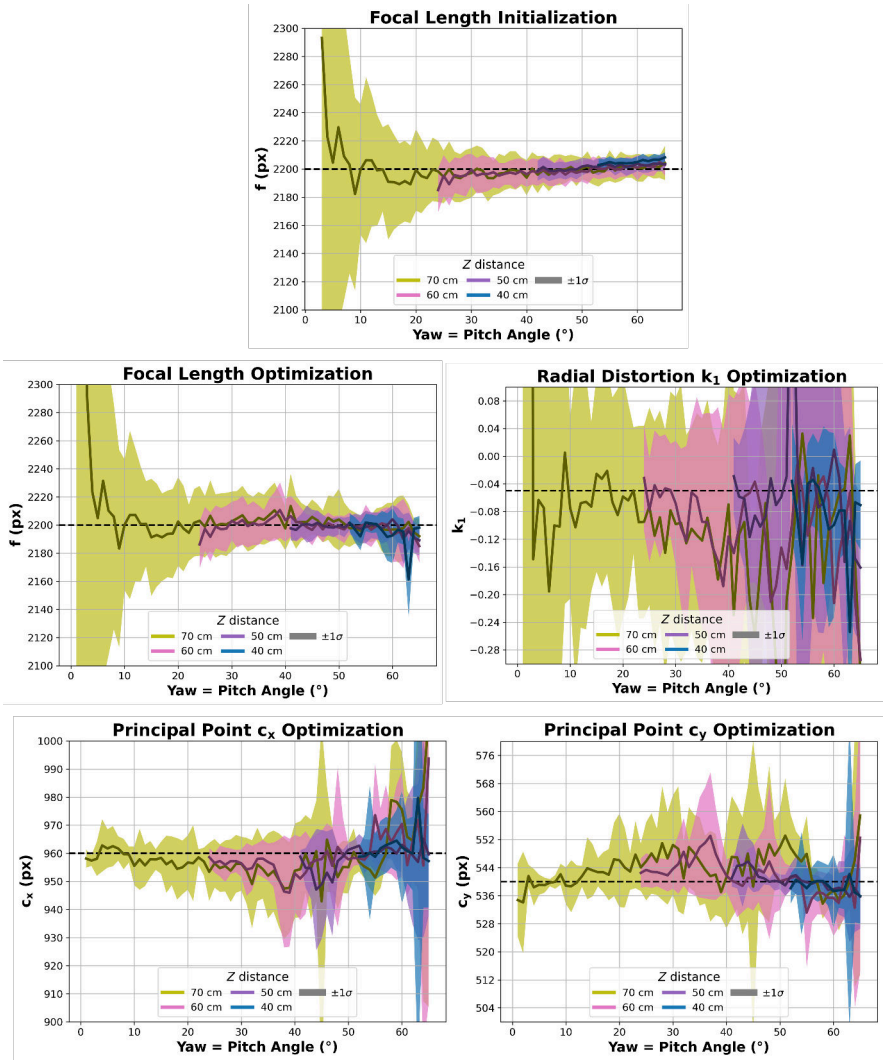


Figure C.3: Calibration results obtained using the same experimental configuration with a narrow-angle camera.

Pitch/Yaw (°)	Z (m)	f (px)	k_1	c_x (px)	c_y (px)
35	0.40	-	-	-	-
35	0.70	2205.13 ± 18.39	-0.12 ± 0.18	952.01 ± 15.18	545.78 ± 12.61
45	0.40	–	–	–	–
45	0.70	2205.74 ± 12.49	-0.26 ± 0.14	942.91 ± 62.40	538.56 ± 41.58
55	0.40	2201.81 ± 6.19	-0.04 ± 0.06	958.73 ± 8.18	538.14 ± 5.75
55	0.70	2199.79 ± 10.81	-0.04 ± 0.33	951.38 ± 6.47	547.92 ± 10.80

Table C.1: Estimated calibration parameters obtained using the minimal Zhang approach for a narrow field-of-view camera.

Wide-angle camera

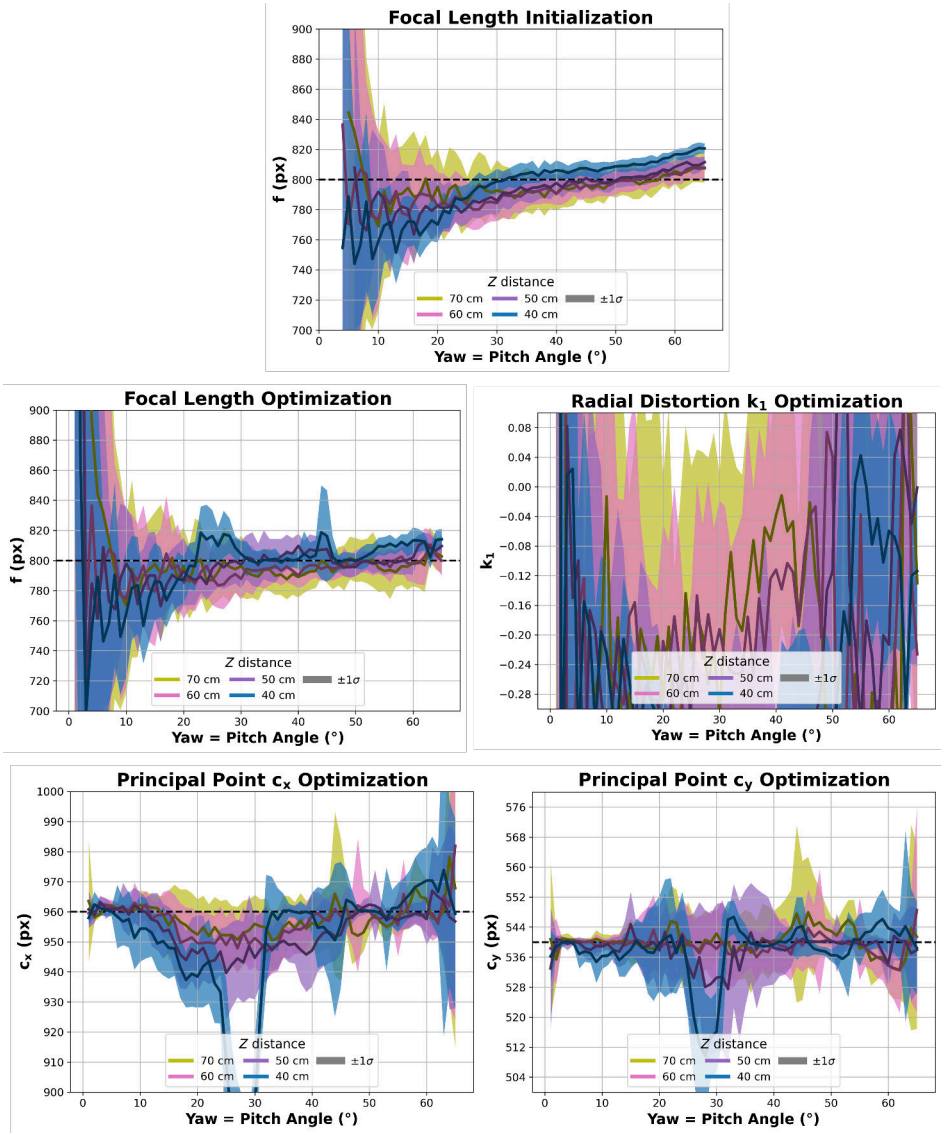


Figure C.4: Calibration results obtained using the same experimental configuration with a wide-angle camera.

Pitch/Yaw (°)	Z (m)	f (px)	k_1	c_x (px)	c_y (px)
35	0.40	800.47 ± 6.26	-0.34 ± 0.03	960.31 ± 5.72	541.63 ± 3.16
35	0.70	794.12 ± 14.42	-0.14 ± 0.24	954.94 ± 10.46	539.46 ± 9.17
45	0.40	816.27 ± 30.42	-0.46 ± 0.27	955.45 ± 21.15	544.24 ± 7.86
45	0.70	797.28 ± 12.11	-0.12 ± 0.29	960.86 ± 23.79	543.10 ± 18.47
55	0.40	809.32 ± 6.29	0.04 ± 0.29	961.37 ± 8.13	540.55 ± 6.97
55	0.70	792.98 ± 13.43	-0.63 ± 1.09	959.22 ± 7.43	538.36 ± 4.31

Table C.2: Estimated calibration parameters obtained using the minimal Zhang approach for a wide field-of-view camera.

C.3 Main point shift

The focal length initialization does not exhibit meaningful variation in this analysis thus is not shown.

Main point shift $(\mathbf{c_x + 15, c_y - 10})$

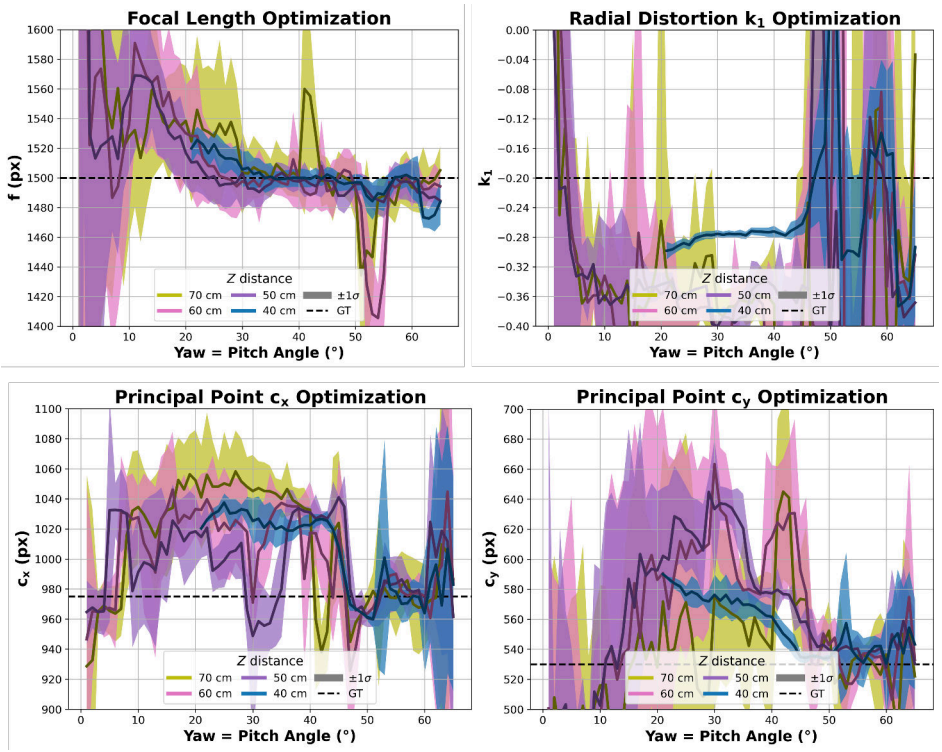


Figure C.5: Calibration results under identical experimental conditions as baseline, with a shifted principal point $(c_x + 15, c_y - 10)$.

Pitch/Yaw (°)	Z (m)	f (px)	k_1	c_x (px)	c_y (px)
35	0.40	1500.88 ± 9.28	-0.28 ± 0.00	1021.61 ± 11.95	569.59 ± 11.54
35	0.70	1499.79 ± 12.97	-0.37 ± 0.03	1040.02 ± 12.52	557.67 ± 55.04
45	0.40	1502.33 ± 5.28	-0.26 ± 0.02	1007.83 ± 11.45	534.81 ± 3.58
45	0.70	1497.96 ± 15.67	-0.41 ± 0.22	1023.99 ± 47.84	573.53 ± 31.73
55	0.40	1488.81 ± 11.13	-0.31 ± 0.11	978.12 ± 12.71	534.25 ± 17.63
55	0.70	1488.04 ± 13.51	-0.60 ± 0.89	974.08 ± 25.83	535.58 ± 9.26

Table C.3: Estimated calibration parameters obtained using the minimal Zhang approach under a moderate principal point shift $(c_x + 15, c_y - 10)$.

Main point shift $(\mathbf{c_x + 25, c_y - 20})$

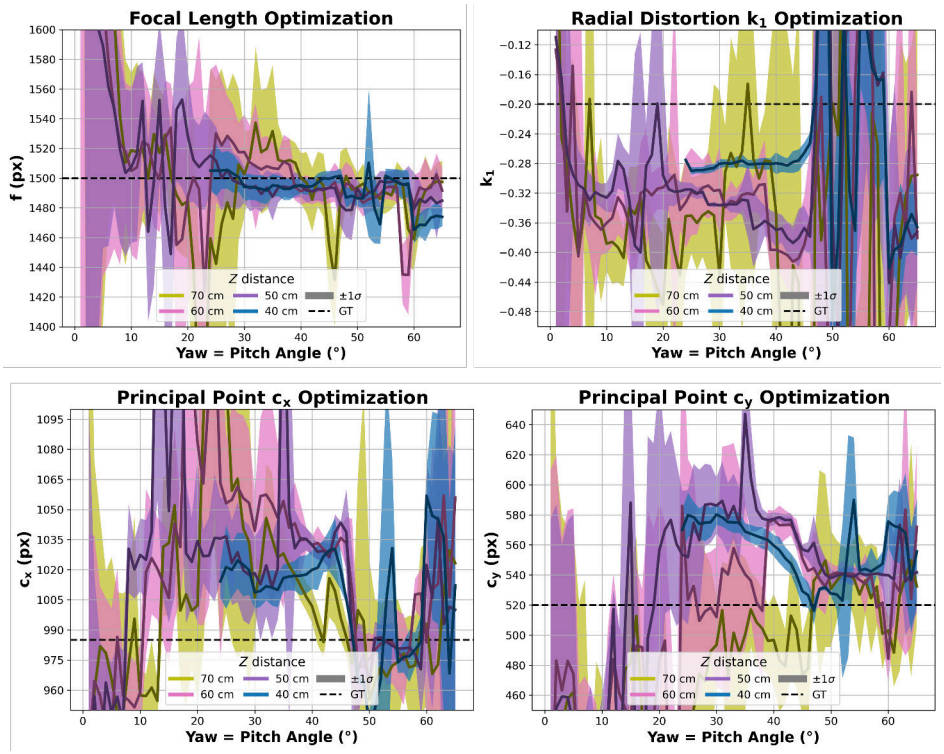


Figure C.6: Calibration results under identical experimental conditions as baseline, with a larger shift on principal point $(c_x + 25, c_y - 20)$.

Pitch/Yaw (°)	Z (m)	f (px)	k_1	c_x (px)	c_y (px)
35	0.40	1495.21 ± 7.45	-0.28 ± 0.01	1015.89 ± 6.72	568.87 ± 5.53
35	0.70	1531.70 ± 44.33	-0.17 ± 0.25	1028.08 ± 14.73	500.23 ± 51.22
45	0.40	1499.55 ± 5.24	-0.27 ± 0.01	1019.79 ± 9.67	529.27 ± 10.29
45	0.70	1455.06 ± 33.88	-1.26 ± 0.67	996.04 ± 10.87	473.53 ± 26.71
55	0.40	1501.57 ± 4.94	-0.05 ± 0.13	973.83 ± 5.26	543.07 ± 2.73
55	0.70	1490.34 ± 13.93	-1.00 ± 0.55	974.08 ± 31.09	527.00 ± 40.03

Table C.4: Estimated calibration parameters obtained using the minimal Zhang approach under a considerable principal point shift.

C.4 Noise robustness

1 px noise

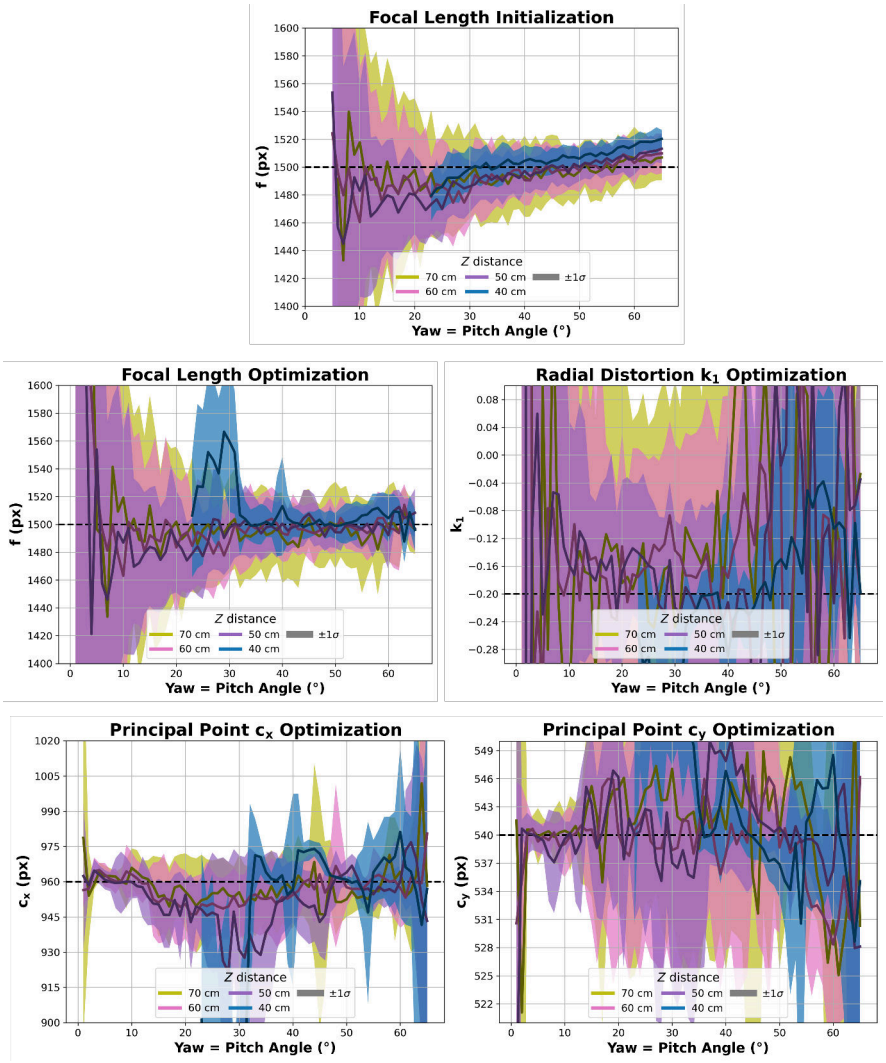


Figure C.7: Calibration results under identical experimental conditions as baseline, with added image noise (1 px).

Pitch/Yaw (°)	Z (m)	f (px)	k_1	c_x (px)	c_y (px)
35	0.40	1499.21 ± 13.25	-0.20 ± 0.03	964.02 ± 4.71	542.93 ± 4.89
35	0.70	1493.81 ± 30.20	-0.16 ± 0.24	952.75 ± 17.63	538.55 ± 14.23
45	0.40	1499.91 ± 9.53	-0.22 ± 0.08	971.95 ± 4.61	539.80 ± 2.45
45	0.70	1504.62 ± 20.74	-0.22 ± 0.18	946.31 ± 49.37	535.74 ± 38.08
55	0.40	1506.37 ± 8.09	-0.05 ± 0.08	963.97 ± 13.74	541.39 ± 8.95
55	0.70	1484.96 ± 34.25	0.09 ± 1.51	956.05 ± 10.85	542.92 ± 16.64

Table C.5: Estimated calibration parameters obtained using the minimal Zhang approach under image noise with a standard deviation of 1 px.

2 px noise

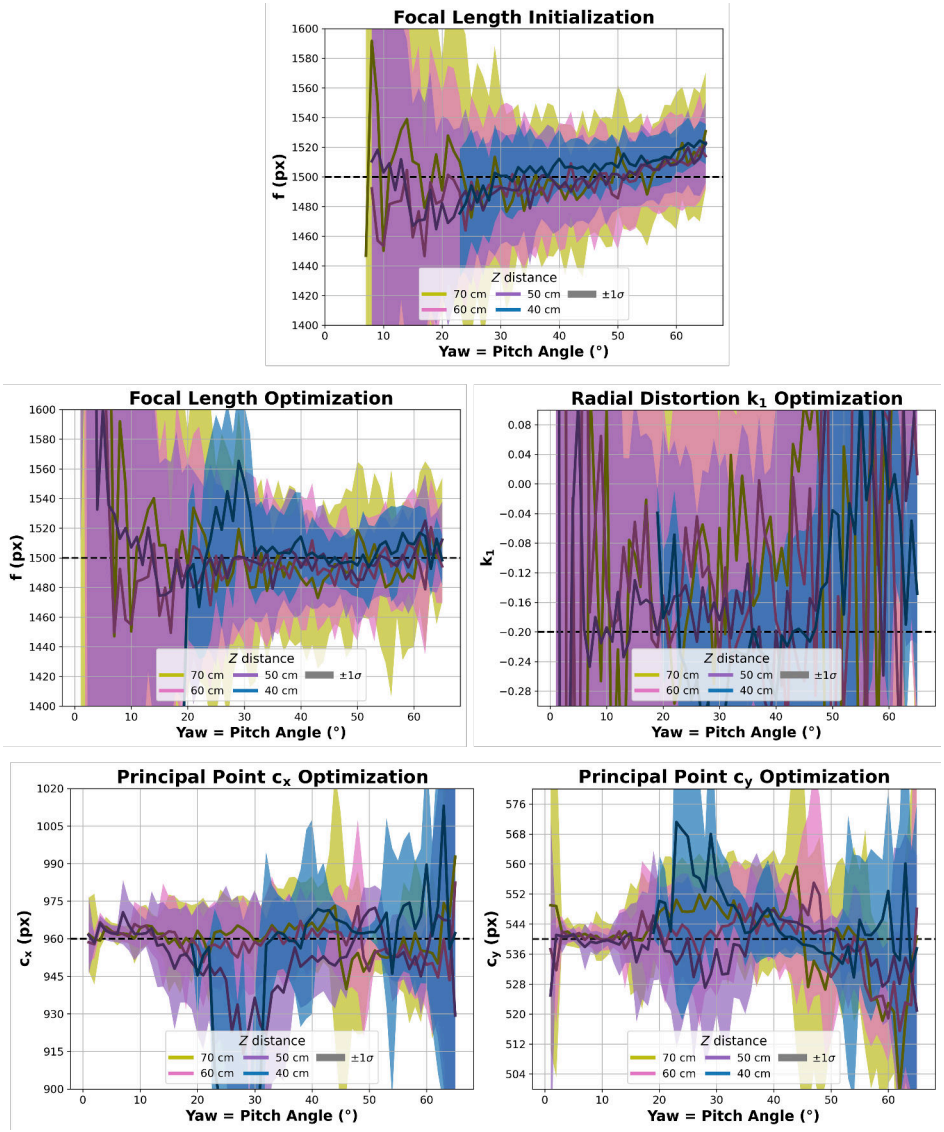


Figure C.8: Calibration results under identical experimental conditions as baseline, with added image noise (2 px).

Pitch/Yaw (°)	**Z (m)**	**f (px)**	**k_1**	**c_x (px)**	**c_y (px)**
35	0.40	1507.37 ± 28.70	-0.23 ± 0.09	961.88 ± 13.86	541.60 ± 8.88
35	0.70	1493.21 ± 45.15	-0.07 ± 0.58	962.19 ± 17.71	544.90 ± 16.25
45	0.40	1496.15 ± 14.50	-0.20 ± 0.14	966.96 ± 5.84	538.54 ± 2.94
45	0.70	1488.63 ± 40.31	0.11 ± 0.60	951.33 ± 59.04	543.23 ± 48.16
55	0.40	1511.08 ± 21.10	0.12 ± 0.32	967.26 ± 23.96	545.13 ± 17.08
55	0.70	1480.14 ± 51.18	-0.53 ± 1.53	953.53 ± 13.11	539.93 ± 18.21

Table C.6: Estimated calibration parameters obtained using the minimal Zhang approach under image noise with a standard deviation of 2 px.

C.5 Camera poses

Only pitch

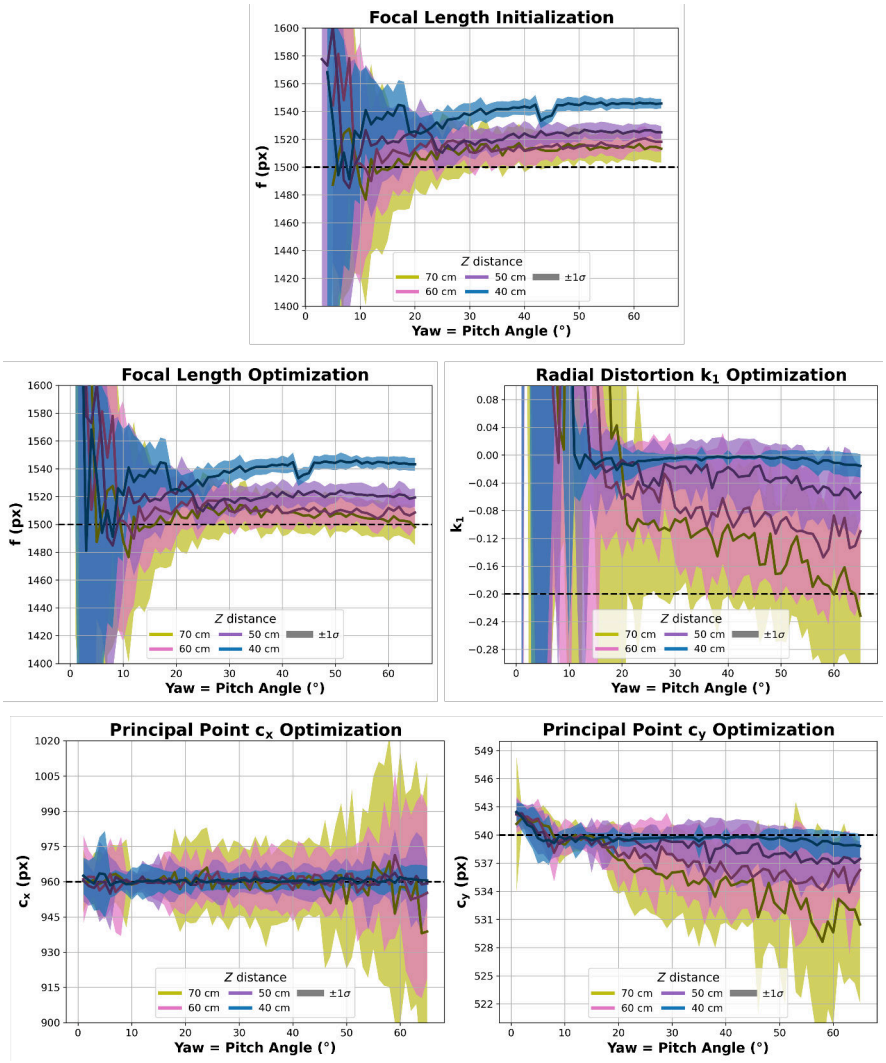


Figure C.9: Calibration results under identical experimental conditions, varying only camera pitch.

Pitch/Yaw (°)	Z (m)	f (px)	k_1	c_x (px)	c_y (px)
35	0.40	1541.26 ± 5.65	-0.01 ± 0.02	960.31 ± 4.82	539.33 ± 1.85
35	0.70	1512.57 ± 14.94	-0.11 ± 0.10	955.97 ± 23.93	534.79 ± 6.75
45	0.40	1536.83 ± 4.33	-0.00 ± 0.00	960.03 ± 2.95	539.75 ± 0.23
45	0.70	1506.66 ± 13.04	-0.11 ± 0.11	960.73 ± 14.73	535.31 ± 4.60
55	0.40	1543.73 ± 4.01	-0.01 ± 0.01	960.92 ± 4.74	539.53 ± 1.10
55	0.70	1505.00 ± 13.53	-0.15 ± 0.10	968.21 ± 36.71	533.24 ± 7.09

Table C.7: Estimated calibration parameters obtained using the minimal Zhang approach with pitch-only camera rotation.

Coupled pitch/yaw

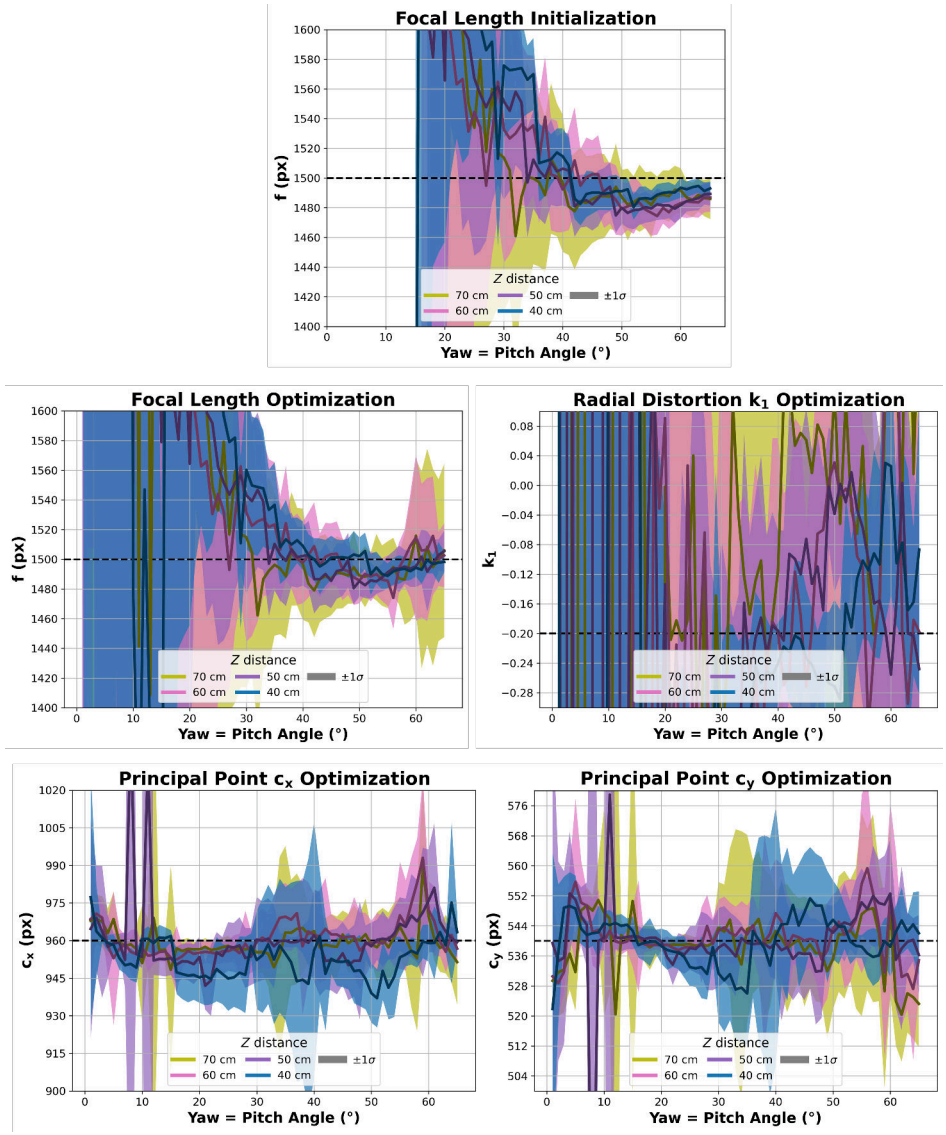


Figure C.10: Calibration results under identical experimental conditions, varying camera pitch and yaw in a coupled way (yaw=pitch).

Pitch/Yaw (°)	Z (m)	f (px)	k_1	c_x (px)	c_y (px)
35	0.40	1536.65 ± 28.61	-0.32 ± 0.11	954.82 ± 26.62	525.98 ± 13.75
35	0.70	1496.31 ± 47.72	-0.17 ± 0.54	963.23 ± 34.45	543.15 ± 25.12
45	0.40	1501.10 ± 14.75	-0.23 ± 0.09	951.53 ± 10.12	547.53 ± 15.83
45	0.70	1486.91 ± 24.68	0.06 ± 0.38	962.43 ± 9.53	537.12 ± 14.63
55	0.40	1488.86 ± 8.23	-0.10 ± 0.11	953.31 ± 8.26	538.47 ± 3.03
55	0.70	1489.64 ± 22.47	0.05 ± 0.74	965.56 ± 11.72	547.05 ± 27.12

Table C.8: Estimated calibration parameters obtained using the minimal Zhang approach with coupled yaw and pitch rotations.

C.6 Stereo Vision

Tilt (°)	Z (m)	Camera	f (px)	k_1	c_x (px)	c_y (px)
35	0.40	Left	1504.98 ± 5.35	-0.21 ± 0.02	970.52 ± 1.81	538.05 ± 0.38
35	0.40	Right	1484.00 ± 6.43	-0.16 ± 0.01	963.05 ± 4.72	550.88 ± 7.62
35	0.70	Left	1502.66 ± 10.52	-0.19 ± 0.08	970.38 ± 7.33	531.52 ± 6.07
35	0.70	Right	1524.15 ± 57.89	-0.36 ± 0.32	966.45 ± 22.78	604.10 ± 65.93
45	0.40	Left	1501.40 ± 4.93	-0.20 ± 0.04	962.54 ± 0.73	538.47 ± 0.56
45	0.40	Right	1465.87 ± 5.91	-0.10 ± 0.02	966.26 ± 2.24	593.16 ± 5.09
45	0.70	Left	1503.34 ± 10.90	-0.19 ± 0.15	963.71 ± 2.91	533.82 ± 5.52
45	0.70	Right	1517.67 ± 44.64	-0.42 ± 0.46	979.27 ± 26.06	588.89 ± 54.81
55	0.40	Left	1509.05 ± 7.28	-0.09 ± 0.07	949.42 ± 4.76	531.29 ± 4.92
55	0.40	Right	1501.16 ± 4.36	-0.19 ± 0.04	959.55 ± 0.13	540.47 ± 0.14
55	0.70	Left	1495.65 ± 19.29	-0.36 ± 0.56	961.04 ± 11.77	531.23 ± 11.55
55	0.70	Right	1501.85 ± 14.82	-0.20 ± 0.18	960.43 ± 11.33	550.97 ± 29.53

Table C.9: Stereo vision calibration results obtained using the standard minimal Zhang approach, with the left camera shown in blue and the right camera shown in red.

Tilt (°)	Z (m)	Camera	f (px)	k_1	c_x (px)	c_y (px)
35	0.40	Left	1523.81 ± 4.88	-0.20 ± 0.01	993.74 ± 2.11	511.03 ± 1.99
35	0.40	Right	1433.71 ± 6.20	-0.13 ± 0.01	949.10 ± 0.57	602.77 ± 3.71
35	0.70	Left	1513.07 ± 10.58	-0.18 ± 0.05	976.02 ± 6.02	517.27 ± 9.79
35	0.70	Right	1475.10 ± 17.07	-0.20 ± 0.04	928.67 ± 6.16	572.14 ± 11.08
45	0.40	Left	1507.01 ± 4.99	-0.19 ± 0.01	968.96 ± 1.01	526.98 ± 1.49
45	0.40	Right	1463.96 ± 5.66	-0.15 ± 0.01	953.40 ± 1.04	609.73 ± 5.48
45	0.70	Left	1510.11 ± 10.78	-0.17 ± 0.06	964.09 ± 2.03	521.02 ± 12.15
45	0.70	Right	1471.87 ± 13.41	-0.11 ± 0.05	957.68 ± 4.65	583.85 ± 14.74
55	0.40	Left	1512.15 ± 6.37	-0.15 ± 0.03	941.55 ± 16.67	516.58 ± 9.07
55	0.40	Right	1501.18 ± 4.47	-0.19 ± 0.04	959.87 ± 0.63	540.20 ± 0.48
55	0.70	Left	1473.87 ± 51.20	-0.66 ± 0.81	946.81 ± 28.55	504.42 ± 33.74
55	0.70	Right	1494.61 ± 19.50	-0.17 ± 0.10	955.87 ± 7.75	564.37 ± 44.96

Table C.10: Stereo vision calibration results obtained using the minimal Zhang approach with a baseline of 10 cm, with the left camera shown in blue and the right camera shown in red.