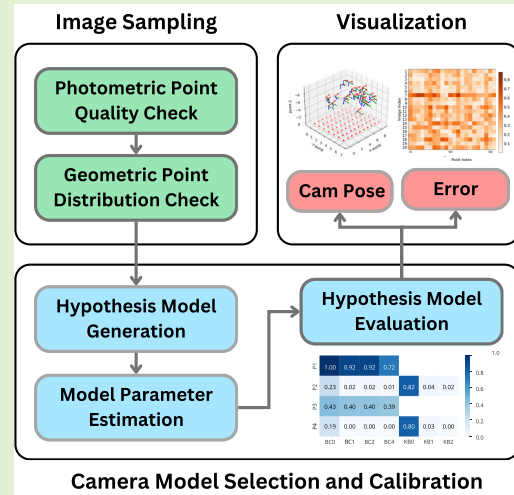


Generalized Camera Calibration: Camera Model Selection and Calibration With Effective Image Sampling

Quy Nguyen Cong^{1b} and Sunglok Choi^{1b}, *Member, IEEE*

Abstract—Camera calibration is essential for accurate optical measurement and computer vision applications. The precision of calibration parameters significantly impacts the performance of subsequent computer vision tasks. However, the process is often complicated by the need to select the most suitable camera projection model, determine the optimal number of images, choose appropriate viewpoints, and define criteria for high-quality dataset construction. To address these challenges, we introduce the *generalized camera calibration* framework, a novel approach that automates dataset creation, considers various camera projection models, and identifies the optimal model and its parameters based on comprehensive model selection criteria. This framework streamlines the calibration process, eliminating the need for manual image and camera model selection before camera calibration. Our method demonstrates outstanding performance on both synthetic and real data. On synthetic datasets, it achieves a remarkable 93.18% accuracy in identifying the correct ground-truth (GT) model using 40 images, employing BIC for model selection. When applied to real datasets, our method maintains a consistent root-mean-square reprojection error (RMSE) of approximately 0.3 pixels across both training and test sets. Extensive validation on synthetic and real data underscores the significant performance enhancements achieved through our approach, making it a powerful tool for simplifying and improving camera calibration in various applications.

Index Terms—Camera calibration, camera model selection, image sampling, model selection criteria.



I. INTRODUCTION

CAMERA calibration (as a part of sensor calibration) is a vital process that involves determining both intrinsic and extrinsic camera parameters. The intrinsic parameters include the focal length, principal point, and distortion coefficients, while the extrinsic parameters consist of the rotation matrix and translation vectors. Obtaining precise camera parameters

is crucial for many computer vision tasks, including measuring object dimensions, object detection [1], [2], 3-D reconstruction [3], [4], or determining the location and size of individuals or vehicles via CCTV [5], [6]. Accurate calibration is essential in industries like robotics, augmented reality, and autonomous vehicles, where precision in visual perception directly impacts performance.

In recent years, numerous camera projection models and calibration methods have been developed [7], [8], [9]. The most widely used approach involves employing a known calibration pattern and capturing multiple images of it from different camera positions [10], [11]. By taking advantage of the relationship between 3-D and 2-D elements in the captured images, the intrinsic and extrinsic parameters of the camera can be determined. Despite significant advancements in camera calibration techniques, current methods often rely on manual model selection, which can be time-consuming and prone to errors, leading to suboptimal results. Researchers continue to face important challenges, including addressing the following key questions.

Received 26 May 2025; accepted 16 June 2025. Date of publication 25 June 2025; date of current version 1 August 2025. This work was supported by the Ministry of Science and ICT (MSIT)/National Research Foundation of Korea (NRF) Grant for the Bridge Convergence Research and Development Program (AI-Based Localization and Path Planning on 3D Building Surfaces) under Grant 2021M3C1C3096810. The associate editor coordinating the review of this article and approving it for publication was Prof. Jun Wang. (Corresponding author: Sunglok Choi.)

The authors are with the Department of Computer Science and Engineering, Seoul National University of Science and Technology (SEOULTECH), Seoul 01811, Republic of Korea (e-mail: sunglok@seoultech.ac.kr).

This article has supplementary downloadable material available at <https://doi.org/10.1109/JSEN.2025.3581377>, provided by the authors.

Digital Object Identifier 10.1109/JSEN.2025.3581377

- 1) Which camera projection model (and model complexity) is most suitable for my camera?
- 2) How many images are necessary for accurate camera calibration?
- 3) What are the optimal viewpoints for capturing images to create a comprehensive dataset?
- 4) What level of calibration image quality is essential for developing a reliable calibration dataset?

Effectively addressing these questions is crucial for achieving accurate and dependable calibration results.

In this study, we introduce the *generalized camera calibration* framework, which encompasses camera model selection and calibration with effective image sampling to address the aforementioned challenges. Our method begins with efficient dataset construction, followed by proposing potential camera models and estimating their parameters, ultimately leading to the selection of the most suitable model based on predefined criteria. This comprehensive method not only enhances calibration accuracy but also ensures that the selected model is the most appropriate for the task. For calibration, we use a chessboard as the calibration pattern because of its efficiency, simplicity, and relative accuracy. To select the best model, we employ two classic and widely recognized criteria, AIC [12] and BIC [13], allowing us to effectively compare and identify the most suitable option. Moreover, we confront the critical issue of image sampling, which has often been overlooked in previous studies, by introducing several methods aimed at enhancing dataset quality, thereby contributing to more accurate calibration results.

Contribution: We introduce the *generalized camera calibration* framework, featuring three key components: image sampling, model selection, and model parameter estimation. This method automates the creation of reliable datasets, calibrates camera parameters across multiple models, and identifies the optimal model. Our image sampling strategy enables the selection of optimal viewpoints for capturing samples while ensuring the quality of each sample. This results in the creation of a dependable dataset that enhances calibration outcomes. Furthermore, the camera model estimation and selection process allows us to identify the most suitable model with characteristics that best align with the constructed dataset, as well as determine the most appropriate camera parameters. Additionally, the framework offers three types of visualization tools: model-wise heatmap, point-wise heatmap, and camera pose visualization to visually display calibration results. We provide an open-source platform to streamline camera calibration, helping researchers quickly identify the most reliable model that balances accuracy and computational efficiency. Our framework shows strong performance on both synthetic and real data, positioning it as a baseline for future research. The code is publicly available at https://github.com/mint-lab/mint_camera_calib

The remainder of this article is structured as follows. Section II reviews related work. Section III details our proposed generalized camera calibration framework. Sections IV and V present experimental results on synthetic and real data. Section VI concludes.

II. RELATED WORKS

A. Camera Calibration

Camera calibration encompasses various approaches, ranging from conventional methods leveraging the geometric relationships between 3-D–2-D correspondences using known markers, lines, or vanishing points [7], [10], to advanced techniques employing complex deep learning architectures [8], [14], [15] for estimating camera parameters. Traditional calibration methods can generally be classified into three categories: the first method, calibration with known targets, involves using a known calibration target, such as a chessboard, matrices of circles, or black squares on a white background [16], which are intentionally placed at various positions within the 3-D scene. The camera captures the target from various angles, and specific feature points (e.g., corners of the chess) are identified to calculate the camera parameters. The second method, geometric-prior-based calibration, utilizes geometric structures to establish correspondences between 3-D and 2-D elements in the scene, including lines and vanishing points [17]. The final method, self-calibration [18], [19], leverages a sequence of images to estimate camera parameters using multiview geometry. In terms of learning-based camera calibration, there are two common approaches: regression-based calibration and reconstruction-based calibration. While regression-based calibration derives parameters from semantic features extracted from uncalibrated input [20], [21], reconstruction-based calibration eliminates parameter regression and directly learns the pixel-level mapping function between the uncalibrated input and the target [22], [23]. Each method has its own advantages and disadvantages; traditional methods require extensive manual effort and lack automated calibration capabilities, whereas deep learning methods involve complex model construction and time-consuming training. The selection of calibration patterns is also crucial and significantly influences the calibration outcome. Therefore, a variety of calibration patterns have been employed in recent years. Patterns can take various forms, including 1-D patterns with points aligned on the same line or vanishing lines [24], [25], [26], 3-D patterns [27], plane calibration patterns [10], and later methods of using indoor or outdoor scenes [28], [29], [30]. Among these methods, particularly the chessboard pattern described in [10] is the most utilized, recognized for its speed, simplicity, and relative accuracy.

B. Camera Model Selection

The selection of camera models for calibration has become increasingly important due to the wide variety of available options. This challenge mirrors the model selection task in machine learning and deep learning, where the goal is to achieve a balance between complexity and accuracy. To address these challenges, numerous criteria have been developed, including AIC [12], CAIC [31], GAIC [32], and BIC [13]. Among the various criteria, AIC and BIC are the most commonly used. AIC, initially introduced in statistical literature as a criterion for model selection, focuses on minimizing information loss without heavily penalizing model

complexity, while BIC, introduced by Schwarz [13], includes a stronger penalty for model size, favoring simpler models as sample size increases. CAIC and GAIC are extensions that introduce more stringent penalties for overfitting, with CAIC placing even greater emphasis on simplicity compared to AIC, and GAIC offering flexibility by adjusting the penalty term based on specific needs. However, a quantitative criterion for selecting the optimal camera model during calibration is notably absent. This gap has motivated previous research to utilize statistical information criteria for camera model selection. These studies have applied such criteria in practical experiments, using AIC to decide between affine and projective camera models [33] during calibration. Additionally, there have been investigations into applying AIC, CAIC, and BIC for the selection of distortion models [34], [35], [36]. Building on these efforts, new methods now leverage AIC, BIC, and GAIC in large-scale structure-from-motion (SfM), with enhanced approaches for automatic camera model selection based on efficient uncertainty evaluation [37]. However, there has yet to be a study proposing a combined criterion for selecting both projection and distortion models; existing research instead focuses on separate criteria for selecting projection models [33] or distortion models [34], [35], [36]. Among the various criteria, as indicated by the findings in [35], BIC is considered the most effective for selecting the optimal camera model. Therefore, we adopt BIC as the primary criterion for model selection in this study. In addition, to enhance objectivity and facilitate comparison, we also incorporate AIC, a fundamental model selection criterion that serves as the foundation for many other criteria. AIC has been widely used in various selection problems, providing a robust framework for evaluating model performance across different contexts.

C. Effective Image Sampling

Carefully selecting data is essential for creating a reliable dataset; yet, it is a factor frequently neglected in prior research. The most relevant works to ours are [38] and [39], which propose an efficient strategy for pose selection in camera calibration. Their approach is rooted in strategy-based methods, where poses are selected from a fixed dataset of predefined options. While this method demonstrates effectiveness in certain scenarios, such as calibrating a stationary camera in a controlled laboratory environment or a standard industrial setup, its reliance on a static dataset poses significant limitations. These limitations become evident when addressing diverse requirements, such as calibrating cameras for autonomous vehicles operating in dynamic outdoor conditions or wearable devices subject to varying user movements. Such an approach may lack the flexibility needed to adapt to real-world scenarios where calibration poses must accommodate a wide range of camera setups and environmental factors. Recognizing these limitations, our work aims to overcome these challenges by developing a more dynamic and flexible method, enabling the generation of a reliable and adaptable dataset that meets a variety of specific needs. A superior dataset is pivotal for favorable calibration outcomes. Calibration datasets are usually created by capturing individual samples or extracting them from videos. However, motion blur

commonly occurs during image capture or video recording, substantially impacting corner detection and, consequently, calibration results. Furthermore, diversifying the dataset by capturing images or recording videos of a known calibration pattern from different angles and aspect ratios is essential for optimal calibration results. Excessive images may not be advantageous as many are nearly indistinguishable, leading to prolonged calibration time and potential negative effects on results. Placing the calibration pattern too far from the camera can also lead to a small pattern-to-image ratio, which may cause significant errors in corner detection, thereby impacting the calibration results. Thus, meticulous dataset creation is essential to achieve the desired outcomes. Eliminating blurry images and those with minimal differences can expedite calibration and enhance results. In this research, we not only focus on camera calibration and model selection but also present strategies for building a comprehensive dataset.

D. Comparison With Existing Calibration Toolboxes

To better contextualize the contributions of our proposed framework, we present a comparative analysis with several widely used camera calibration toolboxes, including both general-purpose and specialized solutions such as OpenCV [40], MATLAB [41], Kalibr [44], and others.

Existing toolboxes typically lack automation in two critical aspects: image selection and camera model selection. Most require users to manually select the camera model (e.g., Pinhole, KB, or BC) as well as configure the parameter complexity. In addition, they typically lack automated mechanisms for evaluating multiple model candidates and selecting the optimal one based on data-driven criteria. This manual configuration process is susceptible to errors and can be inefficient, especially when working with unfamiliar camera systems or deploying at a large scale.

In contrast, our calibration toolbox introduces two key innovations that significantly improve usability and accuracy. First, it supports automatic image selection by analyzing image sequences or video frames to extract an optimal subset of images for calibration. This eliminates the need for manual inspection and reduces the impact of low-quality or redundant images. Second, and more critically, our method automatically selects the best camera model and its corresponding parameter complexity using BIC as a selection metric. This includes evaluating different models (e.g., Pinhole, BC, and KB) and comparing simpler configurations (e.g., focal center and radial distortion only) with more complex ones (e.g., complete intrinsic parameters and higher-order distortion mode), thereby ensuring the most suitable model is chosen for a given dataset.

Table I presents a comprehensive comparison of the functional differences between our toolbox and several representative alternatives. The comparison focuses on four main aspects: image selection strategy, calibration approach, support for model selection, and the diversity of supported camera models. As shown in the table, most existing toolboxes lack automation in image selection and model evaluation. In contrast, our method delivers a fully automated pipeline, offering enhanced flexibility and robustness across diverse scenarios.

TABLE I

SUMMARY AND COMPARISON OF WIDELY USED CAMERA CALIBRATION TOOLBOXES AND OUR PROPOSED FRAMEWORK

Camera Calibration Toolboxes	Automatic Image Sampling	Camera Calibration	Camera Model Selection	Supported Camera Models	Note
OpenCV [40]	✗	✓	✗	Pinhole / BC	The C version of Jean-Yves Bouguet's MATLAB toolbox
MATLAB Computer Vision Toolbox [41]	✗	✓	✗	Pinhole / BC	A commercial version of Jean-Yves Bouguet's MATLAB toolbox
Calibration Toolbox for Generic Lenses [42]	✗	✓	✗	Pinhole / KB	The original MATLAB version for OpenCV fisheye module
MC-Calib [43]	✗	✓	✗	Pinhole / BC, KB	For (non-overlapped) multi-camera system
Kalibr [44]	✗	✓	✗	Pinhole, Omni, ... / BC, KB	For multi-camera and IMU systems
OcamCalib [45]	✗	✓	✗	Pinhole, Omni, ... / BC	For omnidirectional cameras
Ours	✓	✓	✓	Pinhole / BC, KB	A fully automatic pipeline for image sampling and model selection

III. GENERALIZED CAMERA CALIBRATION FRAMEWORK

A. Problem Formulation

Our generalized camera calibration framework consists of two main processes: *camera model selection and calibration* and *effective image sampling*.

1) *Camera Calibration*: Camera calibration estimates a camera's intrinsic and extrinsic parameters using calibration images. Based on the predefined camera projection model M , camera calibration is typically formulated as the least-squares minimization of reprojection errors as follows:

$$\mathcal{C}_M^*, \mathcal{T}^* = \operatorname{argmin}_{\mathcal{C}_M, \mathcal{T}} \sum_{j=1}^m \sum_{i=1}^n \|\mathbf{x}_i^j - \pi_M(\mathbf{X}_i; \mathcal{C}_M, R_j, \mathbf{t}_j)\|_2^2 \quad (1)$$

where \mathcal{C}_M and \mathcal{T} are sets of intrinsic and extrinsic parameters, respectively. A set of intrinsic parameters \mathcal{C}_M can include focal length (f_x, f_y) , principal point (c_x, c_y) , and more as follows:

$$\mathcal{C}_M = \{f_x, f_y, c_x, c_y, k_1, k_2, \dots\} \quad (2)$$

where a series of k_s is distortion coefficients defined in the specific camera model M . A set of extrinsic parameters \mathcal{T} describes each pose of a camera corresponding to each calibration image as follows:

$$\mathcal{T} = \{R_j, \mathbf{t}_j | 1 \leq j \leq m\} \quad (3)$$

where R_j and \mathbf{t}_j are the rotation matrix and translation vector of the j th calibration image, respectively, and m is the total number of calibration images. Each calibration image captures known 3-D points, usually given from a reference object or marker. The total number of reference points is written as n . \mathbf{X}_i is the i th 3-D point and \mathbf{x}_i^j is its projection on the j th image. The camera projection of the camera model M is described as a function π_M , which returns a 2-D point from its given 3-D points and camera parameters.

2) *Camera Model Selection and Calibration*: Camera model selection and calibration identify both the most suitable camera model and its optimal parameters. Its problem includes not only parameter estimation but also model selection. In contrast to camera calibration, a camera model M is no longer predefined, which is one of the optimization variables. Its problem formulation is generalized as

$$M^*, \mathcal{C}_M^*, \mathcal{T}^* = \operatorname{argmin}_{M, \mathcal{C}_M, \mathcal{T}} \mathcal{L}(\mathbf{x}_{1:n}^{1:m}, \mathbf{X}_{1:n}; M, \mathcal{C}_M, \mathcal{T}) \quad (4)$$

where $\mathbf{X}_{1:n}$ is a set of 3-D reference points and $\mathbf{x}_{1:n}^{1:m}$ is its corresponding 2-D projections on calibration images. \mathcal{L} is an objective function to select a camera and estimate its parameters. Continued from camera calibration, a naïve example of

an objective function is the sum of squared reprojection errors (shortly SSE) as follows:

$$\mathcal{L}_{SSE} = \sum_{j=1}^m \sum_{i=1}^n \|\mathbf{x}_i^j - \pi_M(\mathbf{X}_i; \mathcal{C}_M, R_j, \mathbf{t}_j)\|_2^2. \quad (5)$$

However, this naïve objective function tends to fall into the overfitting problem as it does not consider model complexity.

3) *Effective Image Sampling*: Our image sampling method selects calibration images from a video or image sequence, ensuring that only useful images for accurate calibration are retained while filtering out poorly captured ones. This automated process employs quantitative measures to differentiate between good and bad images. Although it could involve large-scale joint optimization with camera model selection and calibration, we simplify the task by decoupling it into two distinct processes to reduce computational load. We define the image sampling step as a preprocessing stage focused on extracting reference points from the selected images. Our sampling step is formulated as

$$I_{1:m}, \mathbf{x}_{1:n}^{1:m} = \operatorname{sample}(\mathcal{I}) \quad (6)$$

where \mathcal{I} is a set of all given images and sample selects a set of images $I_{1:m}$ and extracts reference points $\mathbf{x}_{1:n}^{1:m}$ with reasonable quantitative criteria and conditions.

B. Camera Projection Models

A camera projects a 3-D point as a 2-D point on its image plane. There have been numerous camera projection models for various physical inspirations and applications. For the specific camera model M , its camera projection is mathematically represented as a function π_M . The camera model M involves a different set of intrinsic parameters based on its definition, which is noted as \mathcal{C}_M . The camera projection also needs extrinsic parameters R and \mathbf{t} to transform the 3-D point \mathbf{X} (noted in the world coordinate system) based on the camera coordinate system as follows:

$$\mathbf{X}' = R\mathbf{X} + \mathbf{t} \quad (7)$$

where \mathbf{X} and \mathbf{X}' are the same 3-D point represented in the world and camera coordinate systems, respectively.

1) *Pinhole Camera Model*: The pinhole camera projection is an ideal model that assumes all light rays pass through the pinhole (also referred to as a *focal point*). Therefore, the projected 2-D point \mathbf{x} of the 3-D point \mathbf{X}' (in the camera coordinate system) is proportionally expressed as

$$\mathbf{x} = \operatorname{project}(\hat{\mathbf{x}}) = [f_x \hat{x} + c_x, f_y \hat{y} + c_y]^T \quad (8)$$

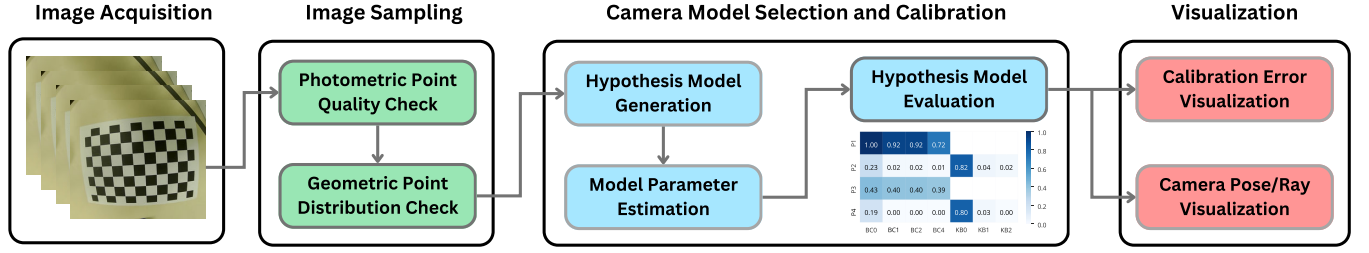


Fig. 1. Process for calibration and model selection. First, a series of chessboard images are captured. Then, various combinations of distortion and projection models are established. Each combination is performed and evaluated by model selection criteria. Finally, the most suitable model will be determined based on the evaluation scores.

where $\hat{\mathbf{x}} = [\hat{x}, \hat{y}]^T$ is an ideally projected 2-D point on the normalized plane ($Z' = 1$) as

$$\hat{\mathbf{x}} = \text{normalize}(\mathbf{X}') = [X'/Z', Y'/Z']^T. \quad (9)$$

Overall, the pinhole camera projection π_P is mathematically expressed as

$$\begin{aligned} \mathbf{x} &= \pi_P(\mathbf{X}; C_P, R, \mathbf{t}) \\ &= \text{project}(\text{normalize}(R\mathbf{X} + \mathbf{t})). \end{aligned} \quad (10)$$

The complexity of the pinhole camera model can be controlled according to its applications and assumptions. Generally, the intrinsic parameters of the pinhole camera model can be enumerated as

$$C_{P4} = \{f_x, f_y, c_x, c_y\} \quad (11)$$

where the subscription number next to acronyms (e.g., P) means its number of parameters for convenience. If the aspect ratio of the image plane is equal, horizontal and vertical focal lengths are also equal, $f_x = f_y$. This assumption is quite popular because the equal aspect ratio is common for recent image sensors. Moreover, if the principal point is assumed to be at the center of the image plane, the principal point is no longer a variable, but constant. Under this assumption, the principal point can be described as the center¹ of an image, $c_x = W/2$ and $c_y = H/2$. Three combinations of two assumptions derived the following simplified sets of the intrinsic parameters:

$$C_{P3} = \{f, c_x, c_y\} \quad (12)$$

$$C_{P2} = \{f_x, f_y\} \quad (13)$$

$$C_{P1} = \{f\} \quad (14)$$

where C_{P3} is from the equal aspect ratio assumption, C_{P2} is from the centered principal point assumption, and C_{P1} is from both assumptions. In contrast, as a more complex form, the pinhole camera projection can include the skew parameter which can describe the nonorthogonality of the image plane. In this work, we do not consider the skew parameter because it has been mostly ignored recently due to the more precise manufacturing of cameras. However, we believe that it is straightforward to extend our framework to support the skew parameter.

¹In the zero-based array indexing, the center of images is calculated as $c_x = (W - 1)/2$ and $c_y = (H - 1)/2$.

2) Geometric Distortion Models: A camera has a lens to collect more light rays, whose nonlinearity causes undesired distortion on the projected images. Even though there are also photometric distortions such as *vignetting*, this article focuses on geometric distortion because we tackle geometric camera calibration, not photometric (or color) calibration.

Geometric distortion models describe how light rays deviate from the ideal pinhole camera projection. Accurate physical modeling of complex geometry and optics of a lens and sensor placement is intractable. Therefore, most distortion models are based on simple equations or approximated models. The distortion models are usually defined on the normalized plane (with corresponding notations $\hat{\mathbf{x}}$) to be independent of other intrinsic parameters such as focal length and principal point.

The *Brown–Conrady model* [46] approximates the geometric distortion using a high-order polynomial equation. That is why it is also known as the *polynomial distortion model*. This model was inspired by Taylor’s theorem, which states that a high-order polynomial equation can approximate any function at a specific point. The Brown–Conrady distortion model considers two types of geometric distortions: *radial* distortion and *tangential* distortion. From the perspective of the polar coordinate system, radial and tangential distortions describe the deviations occurring in the radial and tangential directions, respectively. A distorted point on the normalized plane is mathematically expressed as

$$\begin{aligned} \hat{\mathbf{x}}_d = f_{BC}(\hat{\mathbf{x}}) &= \left(1 + k_1 r^2 + k_2 r^4 + \dots\right) \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} \\ &+ \left(1 + p_3 r^2 + p_4 r^4 + \dots\right) \begin{bmatrix} 2 p_1 \hat{x} \hat{y} + p_2 (r^2 + 2 \hat{x}^2) \\ 2 p_2 \hat{x} \hat{y} + p_1 (r^2 + 2 \hat{y}^2) \end{bmatrix} \end{aligned} \quad (15)$$

where the first term and k_i describe the radial distortion and the second term and p_i represent the tangential distortion. Moreover, the symbol r is the Euclidean distance of the given point $\hat{\mathbf{x}}$ from the origin as

$$r = \sqrt{\hat{x}^2 + \hat{y}^2}. \quad (16)$$

The Brown–Conrady distortion model has been the most common and popular in many theories and applications.

The complexity of the Brown–Conrady model can be controlled with the order of polynomials. Typically, second-order parameters for each distortion are widely used as follows:

$$C_{BC4} = \{k_1, k_2, p_1, p_2\}. \quad (17)$$

It is the default calibration setting in OpenCV [47]. Since the tangential distortion is usually small enough to be negligible, only radial distortion can be considered as

$$\mathcal{C}_{BC2} = \{k_1, k_2\} \quad (18)$$

$$\mathcal{C}_{BC1} = \{k_1\}. \quad (19)$$

If the radial distortion is also ignored, it is no distortion situation and its set of coefficients is

$$\mathcal{C}_{BC0} = \{ \}. \quad (20)$$

The overall camera projection with the Brown–Conrady distortion model, π_{P+BC} , is expressed as

$$\begin{aligned} \mathbf{x} &= \pi_{P+BC}(\mathbf{X}; \mathcal{C}_{P+BC}, R, \mathbf{t}) \\ &= \text{project}(f_{BC}(\text{normalize}(R\mathbf{X} + \mathbf{t}))) \end{aligned} \quad (21)$$

where \mathcal{C}_{P+BC} is a union of two sets of parameters for the pinhole projection and the Brown–Conrady distortion.

The *Kannala–Brandt model* [48] is an alternative to the Brown–Conrady model, especially for wide field-of-view cameras (e.g., fisheye cameras). Wide field-of-view cameras need higher reflection of light rays, and, therefore, their images contain more severe radial distortion. Such severe radial distortion is not adequately described by the Brown–Conrady model, especially near the border of images. Based on the equidistance projection, the Kannala–Brandt distortion model expresses more severe radial distortion as follows:

$$\hat{\mathbf{x}}_d = f_{KB}(\hat{\mathbf{x}}) = \left(1 + k_1\theta^2 + k_2\theta^4 + \dots\right) \frac{\theta}{r} \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} \quad (22)$$

where θ is the projection angle, which is the angle between the principal axis and an incoming light ray as $\theta = \tan^{-1} r$. The original Kannala–Brandt model starts with $(k_0 + k_1\theta^2 + k_2\theta^4 + \dots)$, and this work is based on its modified version with $k_0 = 1$ because the modified version is more common and it becomes the equidistance projection in case of $k_i = 0$. Similarly, the complexity of the Kannala–Brandt model can be controlled by the order of polynomials as follows:

$$\mathcal{C}_{KB2} = \{k_1, k_2\} \quad (23)$$

$$\mathcal{C}_{KB1} = \{k_1\} \quad (24)$$

$$\mathcal{C}_{KB0} = \{ \}. \quad (25)$$

The overall camera projection with the Kannala–Brandt distortion model is

$$\begin{aligned} \mathbf{x} &= \pi_{P+KB}(\mathbf{X}; \mathcal{C}_{P+KB}, R, \mathbf{t}) \\ &= \text{project}(f_{KB}(\text{normalize}(R\mathbf{X} + \mathbf{t}))). \end{aligned} \quad (26)$$

Even though \mathcal{C}_{KB0} contains no distortion coefficient, it still contains radial distortion by the equidistance projection noted as θ/r . The relationship of ground cases with no coefficient can be mathematically expressed as

$$\pi_{P+KB0} \neq \pi_{P+BC0} = \pi_P. \quad (27)$$

C. Overall Procedure

In this study, we present a comprehensive framework called *generalized camera calibration*: camera model selection and calibration with effective image sampling. Fig. 1 illustrates the overall process, which involves a series of distinct steps.

1) Image Acquisition:

- a) Capture a sequence of images using a chessboard pattern for calibration.
- b) Generate a dataset of multiple images (\mathcal{I}) for further processing.

2) Image Sampling:

- a) *Photometric Point Quality Check*: Ensure visual consistency and clarity of reference points in the collected images.
- b) *Geometric Point Distribution Check*: Verifies that points are uniformly distributed and well-spaced, enhancing the dataset's reliability.

3) Camera Model Selection and Calibration:

- a) *Hypothesis Model Generation*: Create multiple candidate models (\mathcal{C}_{P+BC_i} and \mathcal{C}_{P+KB_i}) based on sampled images, combining projection model P_i with either distortion model BC_i or KB_i (see Fig. 2).
- b) *Model Parameter Estimation*: Estimate camera parameters, including intrinsic \mathcal{C}_M and extrinsic matrix \mathcal{T} .
- c) *Hypothesis Model Evaluation*: Evaluate each model by assessing reprojection errors and model complexity to identify the optimal camera model ($M^*, \mathcal{C}_M^*, \mathcal{T}^*$).

4) Visualization:

- a) *Calibration Error Visualization*: Offers point-wise heatmap visualizations of calibration errors in each sample, providing insights for improvement in images with high reprojection errors. Model-wise heatmaps are also included to simplify the comparison and selection of different models.
- b) *Camera Pose/Ray Visualization*: Shows the camera's position and orientation relative to the reference objects, serving as a tool to review the constructed dataset and facilitate the development of strategies for improving data quality.

D. Effective Image Sampling

Image sampling is essential for precise camera calibration. We categorize it into *photometric point quality* and *geometric point distribution*. Photometric point quality focuses on maintaining image clarity to avoid blurriness that could influence corner detection. In contrast, geometric point distribution assesses the spatial arrangement of detected corners, aiding in the selection of optimal views for subsequent images. By enhancing both aspects, we can effectively build the dataset and streamline the calibration process.

- 1) *Photometric Point Quality*: To assess image blurriness, we utilize a simple method based on the Laplacian filter, a widely used edge detection technique. This filter calculates

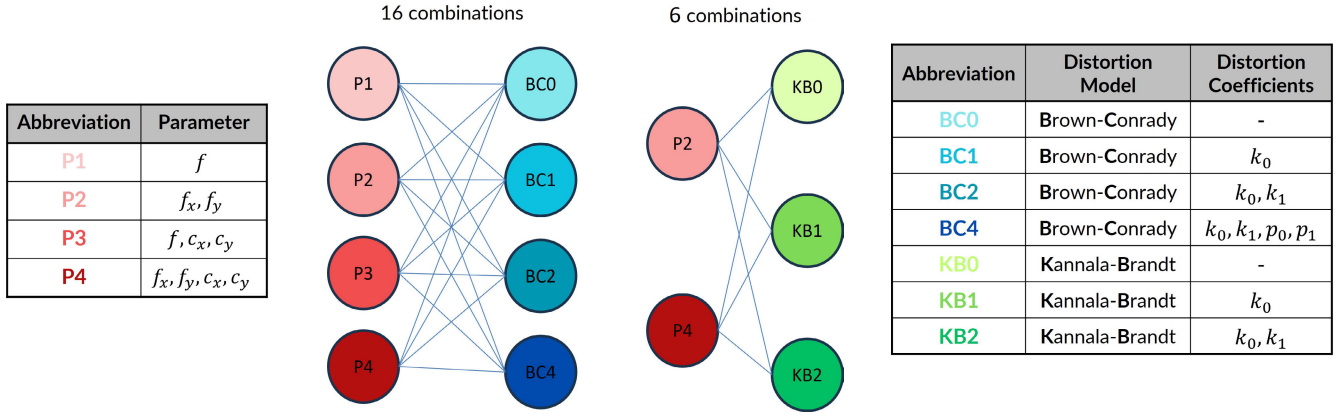


Fig. 2. Model combinations are considered in this study (middle), derived from projection (left), and distortion models (right).

the second spatial derivatives of an image, providing a measure of how the first derivatives change. The Laplacian operator for a function of two variables $f(x, y)$ is defined as

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}. \quad (28)$$

In image processing, the Laplacian ($\nabla^2 f$) can be computed by applying a convolution kernel K

$$K = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \quad (29)$$

To assess the blur level of an image, we first take a single channel of the image I (typically in grayscale) and apply convolution with the kernel K to obtain a new image I'

$$I'(u, v) = \sum_{i=0}^2 \sum_{j=0}^2 I(u+i-1, v+j-1) \cdot K(i, j) \quad (30)$$

where $I(u, v)$ is the original image at position (u, v) and $I'(u, v)$ is the resulting image after convolution at position (u, v) . The resulting response I' is then evaluated for variance, that is, the squared standard deviation, to determine the quality score as

$$Q_s = \text{Var}(I') = \frac{1}{H \cdot W} \sum_{u=1}^H \sum_{v=1}^W (I'(u, v) - \mu_I)^2. \quad (31)$$

Here, μ is defined as the mean of I'

$$\mu_I = \frac{1}{H \cdot W} \sum_{u=1}^H \sum_{v=1}^W I'(u, v). \quad (32)$$

$I'(u, v)$ denotes the elements in the resulting image I' , where h is the height and w is the width of the image. If this variance falls below a predefined threshold τ_{Q_s} , the image is deemed blurry; otherwise, it is considered sharp. The key lies in selecting an appropriate threshold τ_{Q_s} . Setting the threshold too low may falsely mark clear images as blurry, while setting it too high may overlook genuinely blurry ones. Achieving the right balance is crucial for accurate blurriness assessment. Therefore, during the experiment, it is vital to experiment with different τ_{Q_s} thresholds to identify the most appropriate one for the dataset and yield optimal results.

2) Geometric Point Distribution: Choosing the next best view is a vital topic in computer vision and robotics. The strategy for selecting optimal image candidates is essential for enhancing the calibration dataset, which, in turn, improves calibration results. In our approach, we propose an efficient strategy that follows an uncertainty-driven method, maximizing dataset diversity to the greatest extent possible. A good candidate is determined based on the distribution of corners in the image using two parameters: the point distribution score D_s and the distribution distance D_d .

The point distribution score D_s is inspired by concepts from COLMAP [49]. It is determined by assessing the arrangement and visibility of corners in each candidate. A higher D_s score is expected when there are more visible corners and a more uniform distribution. To achieve this, we implement resolution scoring by dividing the image into a fixed-size grid composed of S bins in each dimension. Each cell can exist in one of two states: *empty* or *occupied*. When a detected corner is located within a cell, its state changes to *occupied*, and it returns to *empty* when no corners are present. Consequently, we enhance the method by extending it to a multiresolution score across $l = 1$ to L levels. The point distribution score is expressed as

$$D_s = \sum_{l=1}^L S_l N_o^l = \sum_{l=1}^L 2^l N_o^l \quad (33)$$

where l is the resolution level, $S_l = 2^l$ is the number of bins for both dimensions, and N_o^l is the number of *occupied* cells at resolution l . The next image candidate will be selected if its $D_s > \tau_{D_s}$. Fig. 3 illustrates the application of multiresolution scoring to evaluate an image for selecting the next image candidate, showing different configurations yield varying D_s scores.

The distribution distance D_d is used to ensure diversity in our dataset and prevent duplication or similarity in point distribution among samples. We employ the distribution distance D_d alongside the Hamming distance, a widely used metric in computer science and information theory that measures differences between two strings of equal length by counting the differing symbols at corresponding positions. To effectively incorporate this distance in assessing point distribution distance, we will follow these steps. First, the image I is divided into multiple equally sized cells, each identified by

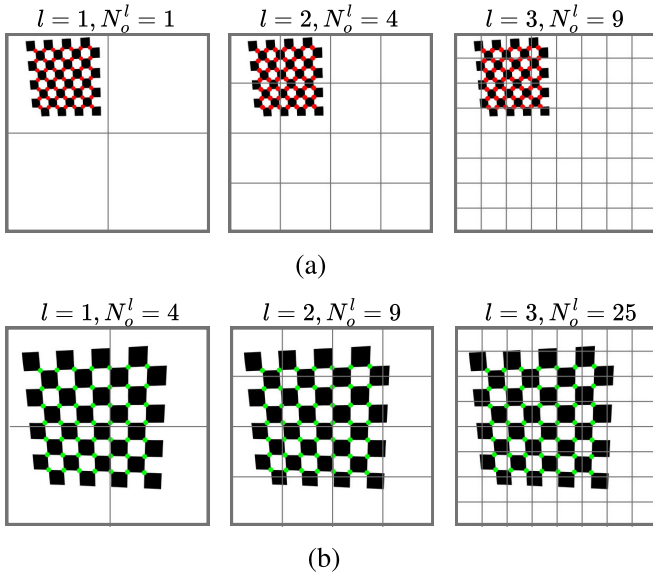


Fig. 3. Example of calculating the point distribution score with multiresolution scoring. (a) $D_S = 90$. (b) $D_S = 224$.

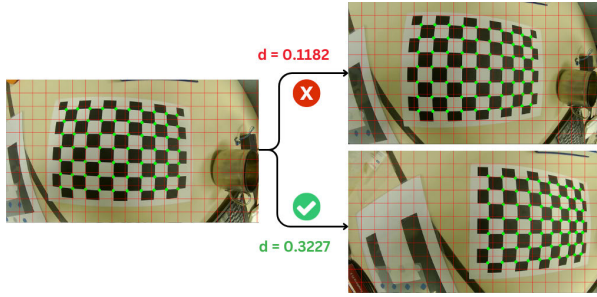


Fig. 4. Illustration of evaluating the distribution distance between two images using multiresolution analysis and Hamming distance.

an adjustable cell size denoted as s_c (refer to Fig. 4). Second, a point distribution vector

$$\mathbf{v}^I = [v_1, v_2, \dots, v_{N_c}] \quad (34)$$

is defined, where N_c is the number of cells in the image when dividing it into equal-sized cells with size s_c . Let C be the set of detected corners in the image. For each cell i ,

$$v_i = \begin{cases} 1, & \text{if } \exists c \in C \text{ such that } c \text{ is in the } i\text{-th cell} \\ 0, & \text{otherwise.} \end{cases} \quad (35)$$

This process yields binary vectors composed of 0s and 1s. The Hamming distance is then used to compare two images based on their respective point distribution vectors. The distribution distance D_d between two input images A and B is determined using the following formula:

$$D_d^{AB} = \frac{\text{Ham}(\mathbf{v}^A, \mathbf{v}^B)}{N_c} \quad (36)$$

where $\text{Ham}(\mathbf{v}^A, \mathbf{v}^B)$ is the Hamming distance between two point distribution vectors \mathbf{v}^A and \mathbf{v}^B . A candidate will be selected if its distribution distance with all the images in the dataset satisfies $D_d > \tau_{D_d}$. Considering the number of cells on the chessboard, along with the image's height and width, various threshold values τ_{D_d} can be used to determine the similarity between two images. The most effective threshold is identified through multiple experiments.

In summary, to enhance the robustness and flexibility of image sampling during calibration, our proposed camera calibration toolbox introduces only three additional parameters: τ_{Q_s} , τ_{D_s} , and τ_{D_d} . These parameters are established to enhance and improve the reliability of dataset construction. Importantly, they are introduced in a minimal yet effective manner that avoids unnecessary complexity in the calibration pipeline. A notable advantage of our design is that these parameters generally do not require manual tuning for different camera models, image resolutions, or checkerboard configurations. This practical insensitivity significantly reduces the user burden during deployment. While popular calibration toolboxes such as OpenCV and MATLAB also rely on several user-defined parameters, most of which are rarely adjusted from their default values, our approach aligns with this practice. The default settings of τ_{Q_s} , τ_{D_s} , and τ_{D_d} have been empirically validated to perform well across a wide range of calibration scenarios. As a result, the added complexity is minimal, while the benefit in terms of adaptability and robustness is substantial, highlighting the practicality and user-friendliness of our toolbox.

E. Camera Calibration and Model Selection

1) *Hypothesis Model Generation*: The camera model combination (M) denotes the integration of various projection models (P) and distortion models (BC or KB). In Section III-B, we discussed both types of projection and distortion models. All camera model combinations evaluated in this study are summarized in Fig. 2, with specific cases for projection models on the left and distortion models on the right. Together, these models generate a comprehensive set of potential camera models for calibration. Specifically, we have a total of 22 camera model combinations, denoted as $\mathcal{C}_{P_i+BC_i|KB_i}$ including (4 projection models \times 4 BC distortion models) + (2 projection models \times 3 KB distortion models), as illustrated in Fig. 2. Since our framework is extensible, additional camera models can be easily integrated. In such cases, the number of hypotheses will vary based on the added models and their complexity.

2) *Model Parameter Estimation and Hypothesis Model Evaluation*: As previously discussed, the camera model selection and calibration problem are formulated as in 4. The primary challenge lies in selecting the optimal camera model and its corresponding parameters. To address this, we define loss functions \mathcal{L}_{AIC} and \mathcal{L}_{BIC} , which are inspired by two widely used criteria (AIC and BIC), to jointly optimize the optimal camera model M^* , intrinsic parameters \mathcal{C}_M^* , and extrinsic parameters \mathcal{T}^* . These criteria guide the evaluation of our candidate models, with lower values of \mathcal{L}_{AIC} and \mathcal{L}_{BIC} indicating a better balance of performance and complexity. Together, they provide a robust framework for effective camera model selection.

Given m images, each containing n detected points, and a camera model with k parameters (e.g., \mathcal{C}_{P4+BC4} : $k = 8$, \mathcal{C}_{P2+KB1} : $k = 3$), \mathcal{L}_{AIC} and \mathcal{L}_{BIC} are defined as follows:

$$\mathcal{L}_{AIC} = mn \ln \left(\mathcal{L}_{RMSE}^2 \right) + 2k \quad (37)$$

$$\mathcal{L}_{BIC} = mn \ln \left(\mathcal{L}_{RMSE}^2 \right) + k \log(mn) \quad (38)$$

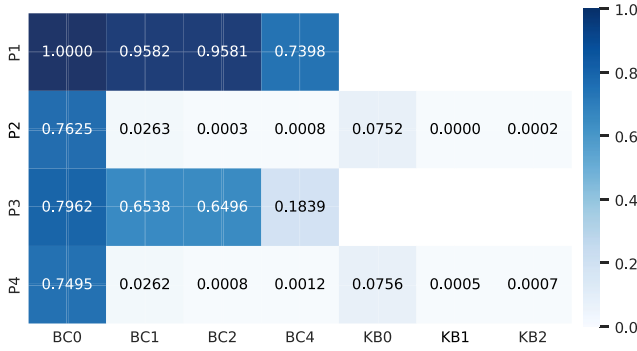


Fig. 5. Model-wise heatmap for model comparison and selection based on the BIC score \downarrow , normalized between 0 and 1.

where $\mathcal{L}_{\text{RMSE}}$ is the root-mean-square projection error, defined as

$$\mathcal{L}_{\text{RMSE}} = \sqrt{\frac{1}{mn} \sum_{j=1}^m \sum_{i=1}^n \|\mathbf{x}_i^j - \pi_M(\mathbf{X}_i; \mathcal{C}_M, R_j, \mathbf{t}_j)\|_2^2}. \quad (39)$$

In terms of computational cost, it is undeniable that our method requires more resources compared to other approaches. The complexity of deep learning methods often varies depending on the model's design and architecture. On the other hand, traditional methods typically have a computational cost of $O(m \times n)$, where m represents the number of samples and n is the number of detected points. Our approach, however, involves estimating parameters for 22 different models, resulting in a complexity of $22 \times O(m \times n)$.

Fortunately, this additional computational burden is mitigated by the nature of the calibration process. Calibration is generally a one-time operation, meaning it is performed infrequently or only once unless there are significant changes to the camera setup. Once the intrinsic and extrinsic parameters are calculated, they can be reused over an extended period. Moreover, calibration is inherently an offline process, outside the scope of real-time system operations. As a result, the computational cost or time required for calibration does not directly affect the performance of real-time tasks, ensuring minimal impact on the system's operational efficiency. Specifically, the execution time of our method is 68.96 s (including both camera calibration and model selection), while the baseline method takes approximately 5.2 s on the same dataset of 40 images. Despite performing calibration and evaluating multiple models, the process takes only seconds rather than hours or days, making the execution time during calibration a relatively minor issue.

F. Calibration Results Visualization

Our framework offers two main types of visualizations to present calibration results: calibration error visualization and camera pose/ray visualization, with the calibration error visualization consisting of model-wise and point-wise heatmaps.

1) *Model-Wise Heatmap Visualization*: It aids in model selection by simplifying comparisons between different models, allowing us to identify the most suitable one (see Fig. 5). Each cell represents a camera model, defined by a combination of distortion (vertical axis) and projection (horizontal axis)

models. Cell values, derived from normalized BIC or AIC scores (0–1), indicate performance, with lower values reflecting better results. This heatmap enables quick identification of well-performing models and those unsuitable for the current dataset. In addition to selecting the model with the lowest score, we can easily choose the top 2 or 3 models, enhancing the selection process to find the most suitable model.

2) *Point-Wise Heatmap Visualization*: In contrast to model-wise visualization comparing different camera models, point-wise visualization analyzes individual samples within the dataset. After calibrating with the best model and obtaining intrinsic and extrinsic parameters, we calculate the reprojection error for each chessboard corner in every image. As depicted in Fig. 6, each row represents an image, while each column corresponds to a chessboard corner in that image. The cell values indicate the normalized reprojection error (0–1) for each corner based on the selected model. For instance, in Fig. 6, image index 7 has several corners with notably high reprojection errors (indicated by significantly darker cells compared to other images). This observation allows us to enhance the overall quality of the dataset.

3) *Camera Pose/Ray Visualization*: This tool effectively illustrates camera positions and orientations relative to captured images, highlighting spatial distribution and alignment with reference objects. It facilitates an intuitive comprehension of how the camera's perspective impacts the captured data, making it essential for analyzing camera configurations and refining strategies aimed at enhancing data collection processes.

IV. EXPERIMENTS WITH SYNTHETIC DATA

In this section, we present experiments to validate the effectiveness of our method for camera model selection and calibration, while assessing the impact of efficient image sampling on the calibration process. We create synthetic datasets to test different model combinations and evaluate our approach's accuracy and robustness under various conditions. The experiments are divided into three parts: 1) camera model selection and calibration; 2) photometric point quality; and 3) geometric point distribution. Detailed configurations and results for each experiment are detailed below.

A. Camera Model Selection and Calibration

1) *Configuration*: We validate our method by testing if it can accurately identify the ground-truth (GT) model used to generate synthetic data. Using 22 different model combinations $\mathcal{C}_{P_i+BC_i|KB_i}$ (as illustrated in Fig. 2), we create 10 different intrinsic matrices and distortion coefficients per combination, resulting in 220 unique camera models (22 combinations \times 10 intrinsic matrices and distortion coefficients). For each model, we synthesize datasets with 5, 10, 20, and 40 images, enabling analysis of how the image count affects calibration accuracy and the minimum images required for reliable results. All images are corrupted with zero-mean Gaussian noise ($\mu = 0, \sigma = 1$) to simulate real-world conditions. During the corner detection process, the initial corner points of the chessboard were identified. Subpixel accuracy was then applied to refine

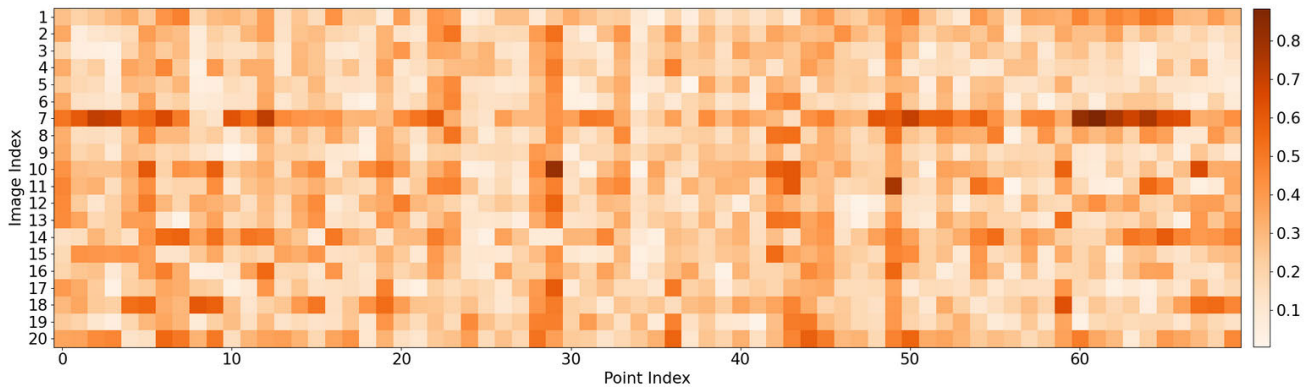


Fig. 6. Point-wise heatmap for comparing reprojection errors of chessboard corners across images in the dataset.

these corner positions. Specifically, after detecting the initial corner points, these positions were further refined using a sub-pixel refinement method. The refinement process was carried out with termination criteria set to a maximum of 30 iterations or until the change in corner positions was less than 0.001.

We use our method to identify the best model among all the proposed options and compare it with the GT. We determine the accuracy of correct GT model selection (Acc [%]) and the ratio of incorrect cases (Ratio [%] = 100 – Acc). Additionally, we assess the best model from our method against the GT model using metrics such as root-mean-square reprojection error (RMSE) $\mathcal{L}_{\text{RMSE}}$, average principal point error \bar{e}_c , and average focal length error \bar{e}_f as follows:

$$\bar{e}_c = \frac{1}{N} \sum_{i=1}^N \|c_{\text{GT}_i} - c_{\text{predict}_i}\|_2 \quad (40)$$

$$\bar{e}_f = \frac{1}{N} \sum_{i=1}^N \|f_{\text{GT}_i} - f_{\text{predict}_i}\|_2. \quad (41)$$

2) Results and Discussions: The results, summarized in Table II, highlight the impact of model selection criteria and dataset size on the accuracy of identifying the correct model used to generate synthetic data. Both AIC and BIC show a progressive increase in selection accuracy as the number of images increases. Notably, BIC slightly outperforms AIC across all dataset sizes, with a particularly noticeable margin at higher image counts (20 and 40 images). These findings highlight the substantial influence of the number of input images on calibration quality, with a higher quantity of images correlating with improved results. Especially, calibrating with 40 images and applying BIC yields optimal results, achieving a 93.18% (205/220 cases) accuracy rate in selecting the correct model used to generate synthetic data. The comparison results between the best model determined by our method and the GT model regarding the RMSE, the average principal point error \bar{e}_c , and the average focal length error \bar{e}_f are also included in Table II. Notably, when using 40 images, both AIC and BIC yield minimal errors in terms of principal point, focal length, and reprojection error.

In cases of misidentification, we observe that while the exact model used to generate the synthetic data is not selected, the chosen model closely approximates the GT. This is evidenced by the minimal reprojection error relative to the GT. The

RMSE for these misidentified models is detailed on the right side of Table II. In practical scenarios, there are cases where, despite not correctly identifying the GT, the error between the chosen model and the GT remains relatively small. To address this, we introduce the concept of a pseudo-GT. Rather than selecting only the best model, we consider models in the top three as pseudo-GT when the projection error between the chosen model and the GT is below a specified threshold of $\tau = 0.5$ pixels. The accuracy of selecting the exact GT model and the exact GT model plus pseudo-GT using AIC and BIC is reported in Table III. Pseudo-GT increases model selection flexibility by offering a broader range of options. In scenarios where a combination of the GT and pseudo-GT models is permitted, the accuracy for both AIC and BIC improves significantly, suggesting that the inclusion of pseudo-GT models may provide a more flexible and robust selection process. For instance, at 5 images, the accuracy jumps from 53.64% (exact GT) to 76.36% (GT + pseudo-GT) for both criteria. This trend is consistent across all dataset sizes, with BIC again showing a slight advantage, particularly at the 20 and 40 image marks. It is important to note that when using large datasets, the concept of pseudo-GT is no longer useful. These findings underscore the importance of choosing the appropriate selection criterion and considering model flexibility based on the dataset size in practical applications.

From the findings in these experiments, both AIC and BIC are effective in identifying the correct model. While our experiments consistently favor BIC over AIC for camera model selection, this preference is also supported by theoretical analysis. Both AIC and BIC aim to balance model accuracy with model complexity, but there is a difference in the way they penalize complexity: AIC applies a constant penalty per parameter ($2k$) without considering the number of data points in the penalty term, whereas BIC's penalty ($k \log(mn)$) increases with the number of observations. In contexts with large datasets, such as ours, BIC imposes a stronger penalty on complex models, thereby favoring simpler models unless the added complexity significantly improves the $\mathcal{L}_{\text{RMSE}}$. This characteristic reduces the risk of overfitting, leading to models that generalize better across varying conditions. From the results shown in Table III, when the dataset is small, AIC and BIC tend to have the same accuracy. However, when the number of images increases, BIC works more efficiently, demonstrating its sensitivity to the size of the dataset. We also

TABLE II

SUMMARY OF THE PERFORMANCE OF THE GT MODEL SELECTION ON SYNTHETIC DATA, SHOWING THE ACCURACY (ACC [%]) FOR CORRECT AND THE RATIO (RATIO [%] = 100 – ACC) FOR INCORRECT MODEL SELECTIONS BASED ON AIC AND BIC. THE TABLE DETAILS THE RMSE (IN PIXELS), AVERAGE FOCAL LENGTH ERRORS \bar{e}_f (IN PIXELS), AND AVERAGE PRINCIPAL POINT ERRORS \bar{e}_c (IN PIXELS) FOR THE MODELS SELECTED BY AIC AND BIC ACROSS DIFFERENT DATASET SIZES (5, 10, 20, 40 IMAGES)

# Images	Correct Model Selection								Incorrect Model Selection							
	AIC				BIC				AIC				BIC			
	Acc	RMSE	\bar{e}_f	\bar{e}_c	Acc	RMSE	\bar{e}_f	\bar{e}_c	Ratio	RMSE	\bar{e}_f	\bar{e}_c	Ratio	RMSE	\bar{e}_f	\bar{e}_c
5	53.64	1.384	21.09	1.74	53.64	1.385	19.96	1.54	46.36	1.382	21.42	9.21	46.36	1.383	20.76	9.77
10	72.73	1.380	14.10	1.80	73.18	1.380	12.83	1.79	27.27	1.377	13.44	10.04	26.82	1.379	12.16	10.72
20	82.73	1.383	10.12	1.55	83.18	1.383	10.00	1.43	17.27	1.381	17.19	6.85	16.82	1.383	16.42	9.58
40	88.18	1.383	9.28	1.34	93.18	1.383	8.99	1.22	11.82	1.385	10.43	2.73	6.82	1.380	10.63	4.91

TABLE III

EVALUATION OF THE ACCURACY IN PRECISELY SELECTING THE MODEL USED TO GENERATE SYNTHETIC DATA. RESULTS ARE SHOWN FOR TWO SCENARIOS: 1) USING ONLY THE GT MODEL, AND 2) COMBINING GT WITH A PSEUDO-GT MODEL. ACCURACY PERCENTAGES (ACC [%]) ARE COMPARED ACROSS DIFFERENT DATASET SIZES FOR BOTH CRITERIA

# Images	GT Model Selection Acc (%)		GT + Pseudo GT Model Selection Acc (%)	
	AIC	BIC	AIC	BIC
5	53.64	53.64	76.36	76.36
10	72.73	73.18	77.27	77.27
20	82.83	83.18	84.55	85.00
40	88.18	93.18	88.18	93.18

observed that AIC may be suitable in scenarios where the dataset is small, or when accuracy is prioritized over model simplicity. In such cases, AIC's lighter penalty can allow for the selection of more complex models that better fit the available data.

Based on the above observations, we will, therefore, employ BIC for all future experiments.

B. Photometric Point Quality

1) *Configuration*: To assess the impact of blurry images on calibration results, we create two synthetic datasets: the *original dataset* and the *blur dataset*. Each dataset is generated from the camera model (C_{P4+BC1}), with the camera model parameters as follows: principal point $c(c_x, c_y) = (650, 450)$, focal length $f_x = 950$, $f_y = 1000$, and distortion coefficients $C_{BC1} = [0.05, 0, 0, 0]$. The *original dataset* consists of 40 images, while the *blur dataset* is created from the *original dataset* by taking 30 images from it and adding vertical or horizontal motion blur to the remaining ten images. All images are corrupted with zero-mean Gaussian noise ($\mu = 0, \sigma = 1$). By applying our calibration process to both, we evaluate how motion blur affects corner detection, ultimately influencing calibration accuracy. After examining the influence of blurry images on the calibration results, we proceed to experiment further to assess the effectiveness of our proposal by filtering out blurry images. Subsequently, we generate two datasets: the *randomly selected dataset*, comprising 30 randomly chosen images from the *blur dataset*, and the *filtered dataset*, created by applying our proposed method on the *blur dataset* to eliminate blurry images with $\tau_{Q_s} = 150$.

2) *Results and Discussions*: The results presented in Fig. 7 provide a comprehensive comparison of calibration results on both the *original* and *blur datasets*, using RMSE and normalized BIC scores for evaluation, with normalized BIC scores being the BIC scores scaled from 0 to 1 for easier display and comparison. Fig. 7(a) and (c) displays the RMSE and normalized BIC scores for the *original dataset*, while Fig. 7(b) and (d) depicts these for the *blur dataset*. RMSE heatmaps visualize the root-mean-square reprojection error for each camera model's parameter estimation. The GT model is highlighted in green while the minimum RMSE value is highlighted in red. Additionally, the normalized BIC score heatmaps offer an additional layer of analysis, emphasizing model complexity alongside accuracy. The models highlighted in red were selected as the best models by BIC, balancing goodness-of-fit with model simplicity. These heatmaps visually compare model performance, with lower values indicating better models for both datasets, helping identify the optimal model through both RMSE and BIC criteria. The slight variations in RMSE and normalized BIC values between the *original* and *blur datasets* reflect the impact of data quality on model performance. The *blur dataset* generally yields slightly higher RMSE and BIC scores, indicating that the presence of blurry images significantly influences the calibration results for parameter estimation. In Fig. 7(a) and (b), we observe that the models with the lowest RMSE (highlighted in red) do not align with the GT models (highlighted in blue). However, when using our camera calibration and model selection framework, the chosen model aligns with the GT model, as demonstrated in Fig. 7(c) and (d). Even though we successfully identified the GT model in both datasets, the calibration results for the parameters can vary. To illustrate these variations and to further assess the effect of blurred images on calibration outcomes, we provide Table IV as a comprehensive comparison of the calibrated parameters for the best model applied to both the *original dataset* and the *blur dataset*. It quantitatively compares these calibration results to the GT using Euclidean distance. Specifically, in the *original dataset*, the L2 distance of the calibrated principal point and focal length relative to the GT is 0.2236 and 1.082, respectively. In contrast, these values increase to 0.9219 and 9.202 in the *blur dataset*.

Table V compares the calibration results obtained from the best BIC-selected camera models using the *randomly selected data* and *filtered dataset*. The evaluation focuses on key

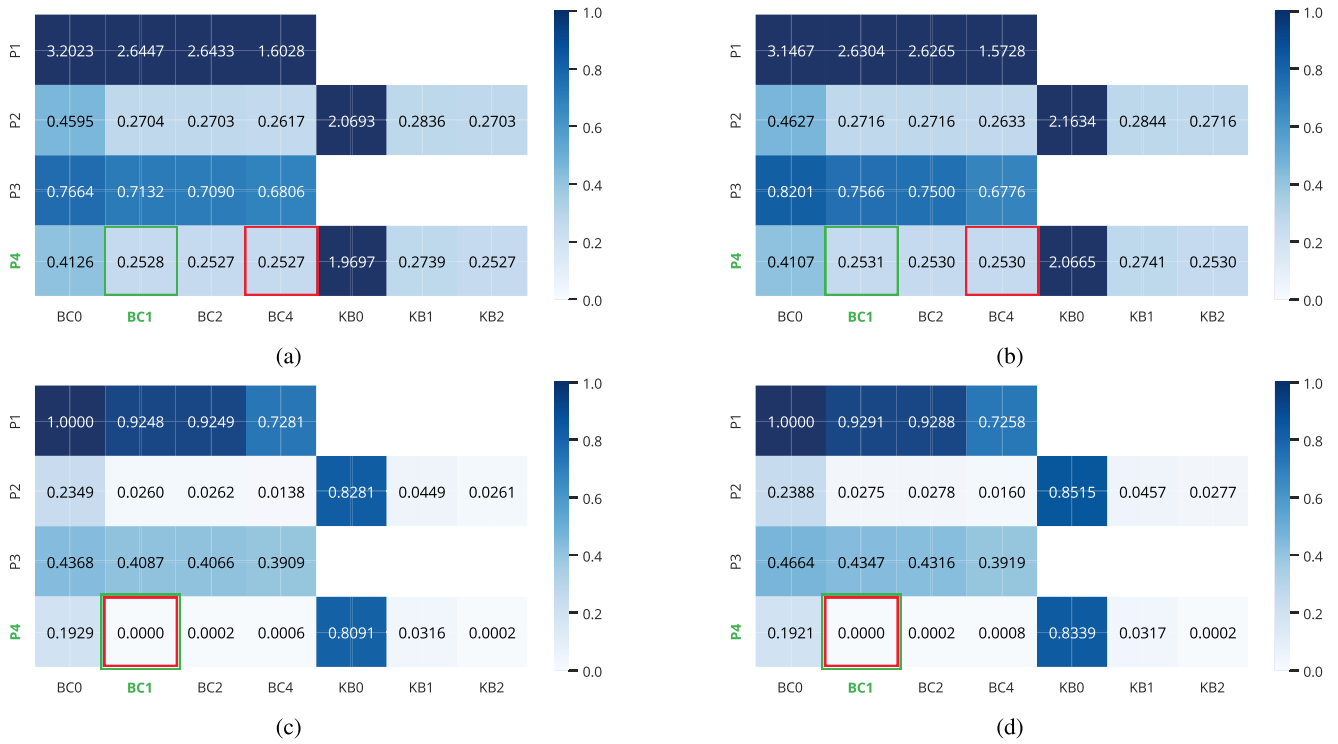


Fig. 7. Results of the photometric point quality experiment. Comparative analysis of model performance on the *original* (left) and the *blur* (right) dataset, illustrated through RMSE (\downarrow) (top) and normalized BIC scores (\downarrow) (bottom), with highlighted cells in green as GT model and in red, showing the minimum RMSE value (top) or the selected model by BIC (bottom). (a) RMSE on the *original* dataset. (b) RMSE on the *blur* dataset. (c) Normalized BIC scores on the *original* dataset. (d) Normalized BIC scores on the *blur* dataset.

TABLE IV

PHOTOMETRIC POINT QUALITY EXPERIMENT RESULTS. THIS TABLE COMPARES CAMERA CALIBRATION RESULTS FOR THE *Original* AND *Blur* Datasets, USING THE BEST MODELS SELECTED BY BIC. THE RESULTS INCLUDE THE CAMERA MODEL, PRINCIPAL POINT, FOCAL LENGTH, AND DISTORTION COEFFICIENTS, WITH L2 DISTANCE (\downarrow) FROM GT

Parameters	GT	Original dataset (Best model)		Blur dataset (Best model)	
		Parameters	L2 Distance	Parameters	L2 Distance
Camera model	P4, BC1	P4, BC1	-	P4, BC1	-
Principal point	(650, 450)	(650.1, 450.2)	0.2236	(650.6, 449.3)	0.9219
Focal length	950, 1000	949.4, 999.1	1.082	943.8, 993.2	9.202
Distortion coefficients	[0.05, 0, 0, 0]	[0.049, 0, 0, 0]	0.001	[0.048, 0, 0, 0]	0.002

TABLE V

PHOTOMETRIC POINT QUALITY EXPERIMENT RESULTS. THIS TABLE COMPARES CALIBRATION RESULTS FOR *Randomly Selected* AND *Filtered* Datasets, USING THE BEST MODELS SELECTED BY BIC. THE RESULTS INCLUDE THE CAMERA MODEL, PRINCIPAL POINT, FOCAL LENGTH, AND DISTORTION COEFFICIENTS, WITH L2 DISTANCE (\downarrow) FROM GT

Parameters	GT	Randomly selected dataset (Best model)		Filtered dataset (Best model)	
		Parameters	L2 Distance	Parameters	L2 Distance
Camera model	P4, BC1	P4, BC1	-	P4, BC1	-
Principal point	(650, 450)	(651.4, 450.2)	1.4142	(650.6, 449.3)	0.9219
Focal length	950, 1000	942.7, 991.9	10.904	943.8, 993.2	9.202
Distortion coefficients	[0.05, 0, 0, 0]	[0.048, 0, 0, 0]	0.002	[0.048, 0, 0, 0]	0.002

camera parameters: principal point, focal length, and distortion coefficients, with the L2 distance providing a quantitative measure of how closely the calibration results align with the GT. The findings reveal that the *filtered dataset* generally yields a closer approximation to the GT compared to the *randomly selected dataset*, as evidenced by the lower L2 distances across most parameters. Specifically, the principal point parameter for the *filtered dataset* shows a significantly lower L2 distance (0.9219) compared to the *randomly selected dataset* (1.4142), indicating a more accurate calibration. This suggests that filtering the data contributes to a more precise estimation of the camera's intrinsic parameters. However, the focal length parameter presents a notable exception, where the L2 distance for the *filtered dataset* (9.202) is slightly

lower than that of the *randomly selected dataset* (10.904), but both values are considerably high compared to the GT. This discrepancy may point to the complexity of accurately estimating focal length, even with data filtering, possibly due to the inherent limitations of the datasets. The distortion coefficients, however, exhibit minimal L2 distances across both datasets, suggesting that both methods perform well in estimating this parameter with high precision.

C. Geometric Point Distribution

1) *Configuration*: To investigate the influence of point distribution on calibration results, we conduct an experiment using two synthetic datasets: the *original dataset* and the *duplicate dataset*. To diversify and differentiate from experiments

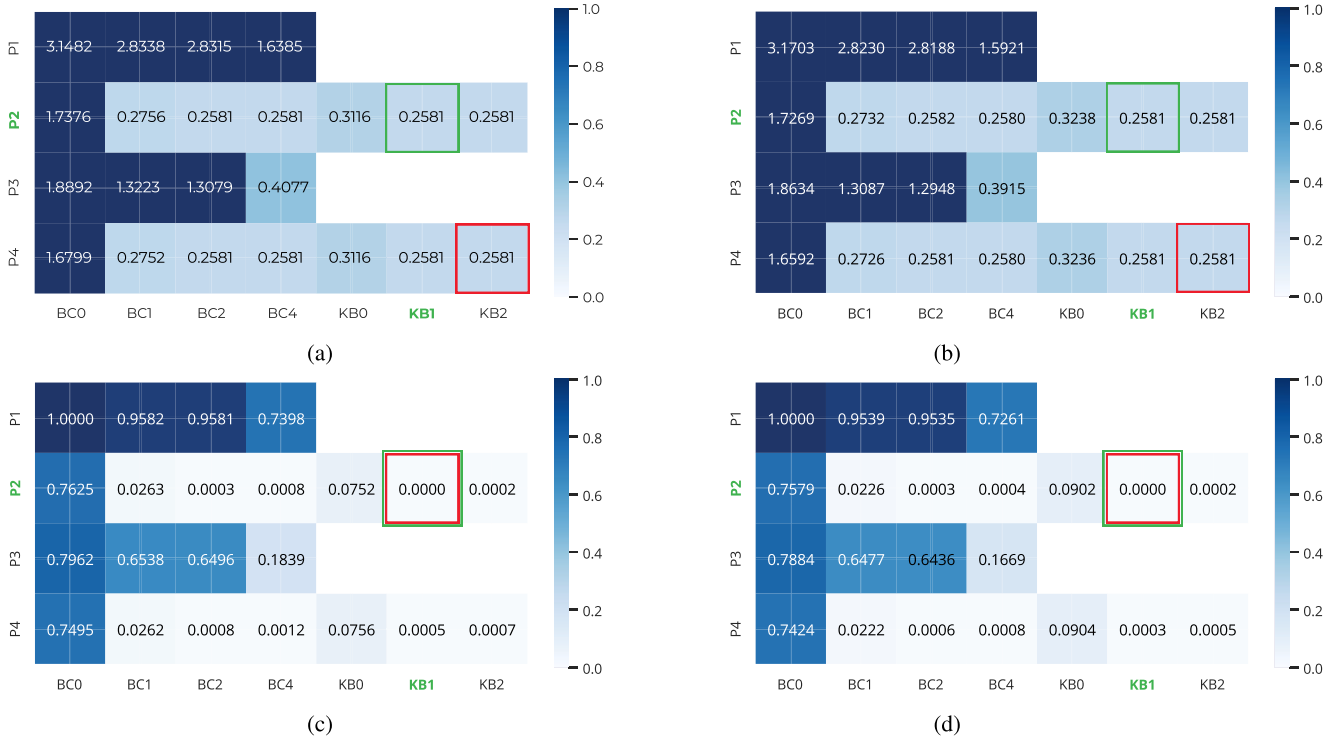


Fig. 8. Geometric point distribution experiment results. Comparative analysis of model performance on the *original* (left) and the *duplicate* (right) datasets, illustrated through RMSE (\downarrow) (**top**) and normalized BIC scores (\downarrow) (**bottom**), with highlighted cells in green as the *GT model* and in red, showing the minimum RMSE value (**top**) or the selected model by BIC (**bottom**). (a) RMSE on the *original dataset*. (b) RMSE on the *duplicate dataset*. (c) Normalized BIC scores on the *original dataset*. (d) Normalized BIC scores on the *duplicate dataset*.

conducted with photometric point quality, we employ a different model (C_{P2+KB1}) to generate the dataset. The model parameters are as follows: the principal point in the center $c(c_x, c_y) = (639.5, 479.5)$, focal length $f_x = 950$, $f_y = 1000$, and distortion coefficients $C_{KB1} = [0.05, 0, 0, 0]$. Additionally, Gaussian noise with mean $\mu = 0$ and standard deviation $\sigma = 1$ is applied. The *original dataset* consists of 40 images, while the *duplicate dataset* contains 50 images. The *duplicate dataset* comprises 40 images from the *original dataset* and ten additional images generated by making slight adjustments to the translation vector and rotation vector of ten random images from the *original dataset*. The objective is to simulate images that could be taken from nearly identical positions or where the chessboard is placed too far from the camera. In these cases, images with similar point distributions will have a minimal distribution distance D_d between the two images, and those with less uniformly distributed points will get a lower distribution score D_s through the multiresolution scoring mechanism.

2) Results and Discussions: Our calibration process was applied to both the *original dataset* and the *duplicate dataset* to investigate whether including additional images with small variations in rotation, translation, and distance from the camera would improve calibration outcomes. This experiment allowed us to explore how sensitive the calibration process is to slight changes in image perspectives and additional data points. The results presented in Fig. 8 provide a comprehensive comparison of model performance on both the *original* and *duplicate datasets*, using RMSE and normalized BIC scores for evaluation. The RMSE and normalized BIC scores for the

original dataset are shown in Fig. 8(a) and (c), while those for the *duplicate dataset* are presented in Fig. 8(b) and (d). In both datasets, our calibration process accurately identified the GT, as demonstrated in Fig. 8(c) and (d). An interesting observation is that, although the *duplicate dataset* has ten more images than the *original dataset*, the RMSE values across all models in Fig. 8(a) and (b) are almost the same. This similarity suggests that simply adding more images, especially if they differ only slightly in rotation and translation or are taken from a greater distance, does not have a significant impact on the accuracy of the calibration results.

Furthermore, Table VI provides a detailed comparison of the calibrated parameters of the best model applied to both datasets. To offer a quantitative perspective, we measured the Euclidean distance from each calibration result to the GT. Notably, although ten images are added to the set, the *duplicate dataset* not only fails to improve the results but also negatively impacts the calibration results. Evidence for this is that the error between the focal length obtained from the *duplicate dataset* and the GT is 1.747, which is relatively larger than the error of 1.047 between the focal length obtained from the original dataset and the GT. To further validate our method, we conduct an experiment using point distribution score and distribution distance score to eliminate images deemed not the best candidates for calibration. We then perform calibration with this new dataset derived from the *duplicate dataset* and apply our method. Despite having fewer images, the new dataset yields results similar to the *duplicate dataset*, closely matching those in Table VI, with no substantial changes.

TABLE VI

GEOMETRIC POINT DISTRIBUTION EXPERIMENTAL RESULTS. THIS TABLE COMPARES CAMERA CALIBRATION RESULTS FOR THE *Original* AND *Duplicate Datasets*, USING THE BEST MODELS SELECTED BY BIC. THE RESULTS INCLUDE THE CAMERA MODEL, PRINCIPAL POINT, FOCAL LENGTH, AND DISTORTION COEFFICIENTS, WITH L2 DISTANCE (\downarrow) FROM GT

Parameters	GT	Original dataset (Best model)		Duplicated dataset (Best model)	
		Parameters	L2 Distance	Parameters	L2 Distance
Camera model	P2, KB1	P2, KB1	-	P2, KB1	-
Principal point	(639.5, 479.5)	(639.5, 479.5)	0	(639.5, 479.5)	0
Focal length	950, 1000	950.74, 1000.74	1.047	951.23, 1001.24	1.747
Distortion coefficients	[0.05, 0, 0, 0]	[0.049, 0, 0, 0]	0.001	[0.049, 0, 0, 0]	0.001

D. Comprehensive Comparison

1) *Configuration*: Based on the results obtained from Table II, which showed the best outcomes with 40 images and BIC for model selection, we proceed with a comprehensive comparison and evaluation against both traditional methods and deep learning-based approaches. We have created a synthetic dataset consisting of 65 images, including images simulated with motion blur, as well as images captured from viewpoints that, according to our criteria, are suboptimal for building a reliable dataset. We compare our method with Zhang's method [10] (equivalent to our C_{P4+BC4} model) and with three models from the GeoCalib [15], denoted as GeoCalib (P), which is a pinhole model using only the focal lengths f_x and f_y without lens distortion (similar to our C_{P2+BC0} model); GeoCalib (R), a simple radial model with weak distortions and a single polynomial distortion parameter $k1$ (similar to our C_{P2+BC1} model); and GeoCalib (D), a simple divisional model (C_{P2+FI1}) with strong fisheye distortions and a single distortion parameter $k1$, as proposed by Fitzgibbon in [50]. Simultaneously, we also perform comparisons between different versions of our method: Ours (PQ), which only uses the point quality sampling method without considering point distribution; Ours (PD), which is the best model utilizing only point distribution while excluding point quality; and Ours, combining both point quality and point distribution. Since the GeoCalib method requires only a single input image, we conduct experiments using all the images in the dataset and calculate the average results.

2) *Results and Discussion*: Table VII compares methods based on RMSE (in pixels), average focal length errors \bar{e}_f (in pixels), and average principal point errors \bar{e}_c (in pixels). Overall, our method demonstrates superior performance compared to the deep learning-based GeoCalib method. This aligns with findings from [15], which highlight a major limitation of deep learning approaches as they are robust but struggle to generalize well to new environments and often require extensive training data, which can be costly to collect. Furthermore, these methods are generally less accurate than traditional geometry-based approaches. Despite these limitations, the GeoCalib method is notable for its robustness, requiring only a single input. Compared to Zhang's method, our approach

TABLE VII

COMPREHENSIVE COMPARISON BETWEEN OUR METHOD AND ZHANG'S [10] METHOD (C_{P4+BC4}), GEOCALIB [15], ALONG WITH VERSIONS (P), (R), AND (D). SPECIFICALLY, (P) REFERS TO A SIMPLE PINHOLE MODEL WITHOUT DISTORTION (C_{P2+BC0}), (R) TO A SIMPLE RADIAL MODEL (C_{P2+BC1}), AND (D) TO A SIMPLE DIVISIONAL MODEL (C_{P2+FI1}) WITH STRONG FISHEYE DISTORTIONS. THE COMPARISON IS BASED ON THE RMSE (\downarrow) IN PIXELS, AVERAGE FOCAL LENGTH ERRORS \bar{e}_f (\downarrow) IN PIXELS, AVERAGE PRINCIPAL POINT ERRORS \bar{e}_c (\downarrow) IN PIXELS, AND AVERAGE DISTORTION COEFFICIENT \bar{e}_d (\downarrow)

Method	Model Selection	Sampling (PQ)	Sampling (PD)	RMSE	\bar{e}_f	\bar{e}_c	\bar{e}_d
Zhang's	-	-	-	0.266	8.54	7.45	0.351
GeoCalib (P)	-	-	-	2.949	1079.62	0.71	0.050
GeoCalib (R)	-	-	-	2.982	1298.00	0.71	0.012
GeoCalib (D)	-	-	-	2.726	1301.30	0.71	0.017
Ours (PQ)	✓	✓	-	0.258	5.09	0.00	0.003
Ours (PD)	✓	-	✓	0.266	3.07	0.00	0.002
Ours	✓	✓	✓	0.256	3.08	0.00	0.002

achieves significant improvements: the principal point error \bar{e}_c is eliminated, the distortion error (\bar{e}_d) is reduced to 0.002 (compared to 0.351 in Zhang's), and the focal length error (\bar{e}_f) is improved by approximately 63%. Among our variants, the full-option version delivers the best results, achieving an RMSE of 0.256, \bar{e}_d of 0.002, and zero principal point error ($\bar{e}_c = 0$).

V. EXPERIMENTS WITH REAL DATA

A. Camera Model Selection and Calibration

1) *Configuration*: Our proposed approach, specifically camera model selection and calibration, was tested on five real datasets collected from five different cameras: ISAW Edge, SJCAM SJ4000, ORBEC Gemini 336L, Intel RealSense T265, and Logitech Webcam. Each dataset contains 40 images, whose sample images are shown in Fig. 9. As with the experiments using synthetic data, we compared our method with Zhang's method [10] and GeoCalib [15] for each dataset. GeoCalib's results are averaged across the dataset, as it operates using only a single input image. Consistent with the synthetic data experiments, subpixel accuracy was maintained for detecting chessboard corners in the real datasets.

2) *Results and Discussion*: The results displayed in Table VIII reveal that, across all five datasets, the models selected by our method consistently outperformed those of other methods. For datasets with high levels of distortion, such as ISAW Edge and Intel RealSense T265, our method tends to select the KB model, which proves highly effective. Similarly, for GeoCalib, the divisional model with one distortion coefficient [GeoCalib (D)] consistently outperforms the radial model [GeoCalib (R)] and the pinhole camera model without distortion [GeoCalib (P)], following the order GeoCalib (D) > (R) > (P). Conversely, for datasets with less severe distortion, such as Logitech Webcam or ORBEC Gemini 336L, our method shows a preference for the BC model, and a similar pattern is observed with GeoCalib, where GeoCalib (P) > (R) > (D). The best models for the ISAW Edge, SJCAM SJ4000, ORBEC Gemini 336L, Intel

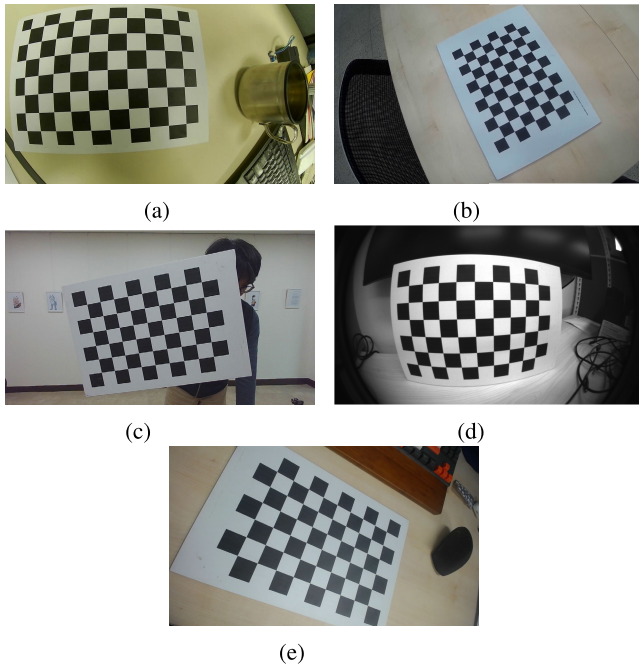


Fig. 9. Examples of images captured from different cameras. (a) ISAW Edge. (b) SJCAM SJ4000. (c) ORBEC Gemini 336L. (d) Intel RealSense T265. (e) Logitech Webcam.

RealSense T265, and Logitech Webcam datasets achieved RMSE values of approximately 0.329, 0.321, 0.389, 0.116, and 0.579 pixels, respectively. Notably, the Intel RealSense T265 dataset exhibits a significant RMSE difference between the model selected by our method and Zhang’s method, with the RMSE of our best model being around 0.2 pixels lower than Zhang’s method. For the Logitech Webcam dataset, the model chosen by our method aligns with the one selected by Zhang’s method. In the ORBEC Gemini 336L dataset, while our best model and Zhang’s method achieve the same RMSE, our model has fewer parameters. Additionally, observing heatmaps is very useful for comprehensive and visual evaluation of all models. As mentioned in the concept of pseudo-GT in synthetic data, through heatmaps, in addition to selecting the top-1 model, we can also consider choosing the top-2 and top-3 models since the differences between these models are relatively small.

B. Camera Model Selection and Calibration With Effective Image Sampling

1) *Configuration*: After performing camera model selection and calibration with a randomly selected dataset, we proceed to test the entire framework. An ISAW Edge camera is used to record a video and collect images. We create four distinct datasets, each containing 30 images: Ours, Ours (PQ), Ours (PD), and Random. The first three were generated using the image sampling module with the parameter $\tau_{Q_s} = 80$, $\tau_{D_d} = 0.12$, and $\tau_{D_s} = 2000$, while the Random dataset is constructed by selecting random frames from the video. To enhance reliability, fairness, and objectivity, we divide the dataset into training and test sets with a ratio of 8:2. The training set is used to estimate camera parameters and select the best model. Subsequently, this selected model is employed

TABLE VIII
CALIBRATION RESULTS ON THREE DIFFERENT REAL DATASETS FROM ISAW EDGE, SJCAM SJ4000, ORBEC GEMINI 336L, INTEL REALSENSE T265, AND LOGITECH WEBCAM CAMERAS, SHOWING RMSE (↓) IN PIXELS. A COMPARISON IS MADE BETWEEN OUR METHOD WITH ZHANG’S METHOD AND GEOCALIB (P), (R), AND (D)

Camera	Method	Model	RMSE
ISAW Edge	Zhang’s	P4 + BC4	0.557
	GeoCalib (P)	P2 + BC0	6.087
	GeoCalib (R)	P2 + BC1	5.904
	GeoCalib (D)	P2 + F11	3.355
	Ours	P4 + KB2	0.329
SJCAM SJ400	Zhang’s	P4 + BC4	0.331
	GeoCalib (P)	P2 + BC0	5.685
	GeoCalib (R)	P2 + BC1	4.953
	GeoCalib (D)	P2 + F11	3.046
	Ours	P4 + KB2	0.321
ORBEC Gemini 336L	Zhang’s	P4 + BC4	0.389
	GeoCalib (P)	P2 + BC0	0.850
	GeoCalib (R)	P2 + BC1	1.203
	GeoCalib (D)	P2 + F11	3.827
	Ours	P4 + BC2	0.389
Intel RealSense T265	Zhang’s	P4 + BC4	0.346
	GeoCalib (P)	P2 + BC0	6.475
	GeoCalib (R)	P2 + BC1	6.369
	GeoCalib (D)	P2 + F11	6.067
	Ours	P4 + KB2	0.116
Logitech Webcam	Zhang’s	P4 + BC4	0.579
	GeoCalib (P)	P2 + BC0	7.581
	GeoCalib (R)	P2 + BC1	10.808
	GeoCalib (D)	P2 + F11	12.332
	Ours	P4 + BC4	0.579

to recalibrate the test set, allowing us to evaluate reprojection errors.

2) *Results and Discussion*: Comparison results of camera model selection and calibration performance on two real datasets collected from the ISAW Edge camera: a random dataset and a dataset constructed by our effective image sampling method are shown in Table IX. According to the results, our framework can effectively perform calibration and model selection. Both the RMSE values for training and testing datasets are reduced when using our method (0.296 and 0.305, respectively) as opposed to the random sampling (0.303 and 0.337). This approximately 10% decrease in RMSE demonstrates that our sampling method enhances overall calibration performance. Upon analyzing the results of Ours (PQ), Ours (PD), and Ours, it is evident that Ours (PQ) demonstrates an improvement over “Random” sampling but yields slightly higher results compared to Ours. Conversely, Ours (PD) closely matches the performance of Ours. This discrepancy can be attributed to the fact that during video recording, we moved the camera slowly, which diminished the effectiveness of the ‘point quality’ sampling method in this context. As a result, our method provides a more reliable estimation of the camera’s intrinsic parameters, which

TABLE IX

CAMERA MODEL SELECTION AND CALIBRATION PERFORMANCE ON TWO REAL ISAW EDGE DATASETS, COMPARING PRINCIPAL POINT, FOCAL LENGTH, DISTORTION COEFFICIENTS, AND RMSE (\downarrow). THE TRAINING PHASE IS USED TO IDENTIFY THE MOST SUITABLE MODEL, WHILE THE TESTING PHASE EVALUATES THE RMSE OF THE SELECTED MODEL

Dataset	Principal Point	Focal Length	Distortion Coefficients	RMSE Train	RMSE Test
Random	(477.90, 288.36)	(434.27, 432.81)	[0.019, 0.005, 0, 0]	0.303	0.337
Ours (PQ)	(477.36, 288.14)	(440.51, 438.97)	[0.011, 0.006, 0, 0]	0.297	0.321
Ours (PD)	(478.71, 288.76)	(433.85, 432.40)	[0.019, 0.006, 0, 0]	0.303	0.319
Ours	(476.33, 288.62)	(441.18, 439.57)	[0.007, 0.009, 0, 0]	0.296	0.305

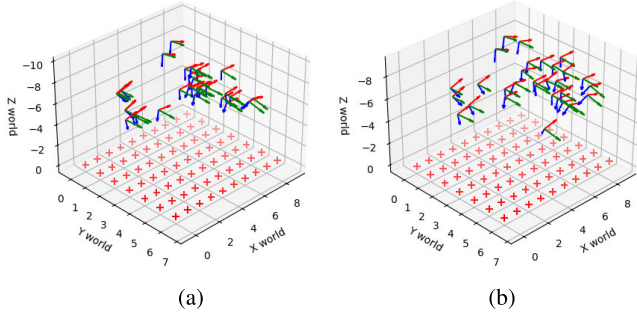


Fig. 10. Camera pose visualization using (a) random and (b) our datasets. Red plus signs (+) mark the chessboard corners; red, green, and blue arrows represent camera X-, Y-, and Z-axes. (a) Random dataset. (b) Our method dataset.

is crucial for high-precision applications. The ability of our model to balance the complexity of the camera further supports its robustness and suitability for advanced computer vision tasks. Overall, the improved accuracy and reliability of our method make it a preferable choice for applications requiring high-quality camera parameter estimation.

The 3-D visualizations in Fig. 10 illustrate a comparison of camera positioning and orientation after calibration using two different datasets: (a) a random dataset and (b) a dataset generated by our proposed method. In the random dataset, the camera positions for capturing calibration images are not widely distributed and tend to cluster in certain areas, which may lead to less comprehensive calibration. In contrast, the dataset created using our method shows a more dispersed distribution of camera positions, resulting in a greater diversity of camera angles and viewpoints. This broader and more varied coverage contributes to a more robust and accurate calibration outcome. The differences between the two datasets underscore the advantages of our method over random sampling, particularly in achieving a broad scatter distribution of camera positions that minimizes potential calibration errors, which is essential for applications requiring precise camera calibration.

VI. CONCLUSION

This work presents a novel and practical method for automatic camera calibration and model selection that streamlines the entire process. Our approach evaluates various potential camera models and uses model selection criteria to identify the optimal balance between accuracy and complexity. Designed to be user-friendly, our method requires no complex data collection or extensive training. Our results

demonstrate superior performance across both synthetic and real data. To further assist users, we provide some intuitive visualizations that simplify model comparison and selection, along with suggestions for effective dataset construction to enhance calibration outcomes. All source codes are openly available, encouraging easy access and contribution from the academic and industrial communities.

While our approach is effective, simple, and convenient, it does have some limitations. First, it does not involve large-scale joint optimization for image sampling and calibration; instead, we decouple this complex task into two simpler processes to reduce computational demands. Second, our method offers passive rather than active guidance, meaning that strategies for creating better datasets are only developed after image sampling and calibration are completed. Additionally, in further work, we aim to explore extensions that incorporate additional camera (distortion) models, broadening the applicability of our approach. We also acknowledge that in dynamic or uncontrolled environments, the image quality and feature distribution may vary significantly. In such cases, static thresholding could be less effective. To address this, future work could explore adaptive methods, such as using real-time quality assessment or learning-based approaches, to improve sampling robustness under varying conditions.

REFERENCES

- [1] Z. Wang, "Review of real-time three-dimensional shape measurement techniques," *Measurement*, vol. 156, May 2020, Art. no. 107624.
- [2] F. Küppers, J. Kronenberger, A. Shantia, and A. Haselhoff, "Multivariate confidence calibration for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1322–1330.
- [3] Z. Ma and S. Liu, "A review of 3D reconstruction techniques in civil engineering and their applications," *Adv. Eng. Informat.*, vol. 37, pp. 163–174, Aug. 2018.
- [4] Z. Kang, J. Yang, Z. Yang, and S. Cheng, "A review of techniques for 3D reconstruction of indoor environments," *ISPRS Int. J. Geo-Information*, vol. 9, no. 5, p. 330, May 2020.
- [5] T. N. Schoepflin and D. J. Dailey, "Algorithms for calibrating roadside traffic cameras and estimating mean vehicle speed," in *Proc. IEEE Intell. Transp. Syst. Conf.*, Nov. 2004, pp. 60–65.
- [6] B. Hardjono, A. Wibisono, A. Nurhadiyatna, I. Sina, and W. Jatmiko, "Virtual detection zone in smart phone, with CCTV, and Twitter as part of an integrated ITS," *Int. J. Smart Sens. Intell. Syst.*, vol. 6, no. 5, pp. 1830–1868, Jan. 2013.
- [7] A. De la Escalera and J. M. Armingol, "Automatic chessboard detection for intrinsic and extrinsic camera parameter calibration," *Sensors*, vol. 10, no. 3, pp. 2027–2044, Mar. 2010.
- [8] J. Lee, H. Go, H. Lee, S. Cho, M. Sung, and J. Kim, "CTRL-C: Camera calibration transformer with line-classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 16208–16217.
- [9] H. Deng, K. Yang, Q. Quan, and K.-Y. Cai, "Accurate and flexible calibration method for a class of visual sensor networks," *IEEE Sensors J.*, vol. 20, no. 6, pp. 3257–3269, Mar. 2020.
- [10] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.
- [11] S. Gasparini, P. Sturm, and J. P. Barreto, "Plane-based calibration of central catadioptric cameras," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Sep. 2009, pp. 1195–1202.
- [12] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.
- [13] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, Mar. 1978.
- [14] O. Bogdan, V. Eckstein, F. Rameau, and J.-C. Bazin, "DeepCalib: A deep learning approach for automatic intrinsic calibration of wide field-of-view cameras," in *Proc. 15th ACM SIGGRAPH Eur. Conf. Vis. Media Prod.*, Dec. 2018, pp. 1–10.

- [15] A. Veicht, P.-E. Sarlin, P. Lindenberger, and M. Pollefeys, "GeoCalib: Learning single-image calibration with geometric optimization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2024, pp. 1–20.
- [16] S. Shah and J. K. Aggarwal, "A simple calibration procedure for fish-eye (high distortion) lens camera," in *Proc. IEEE Int. Conf. Robot. Autom.*, Jul. 1994, pp. 3422–3427.
- [17] F. Bukhari and M. N. Dailey, "Automatic radial distortion estimation from a single image," *J. Math. Imag. Vis.*, vol. 45, no. 1, pp. 31–45, Jan. 2013.
- [18] C. S. Fraser, "Digital camera self-calibration," *ISPRS J. Photogramm. Remote Sens.*, vol. 52, no. 4, pp. 149–159, Aug. 1997.
- [19] R. Hartley, "Self-calibration from multiple views with a rotating camera," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jan. 1994, pp. 471–478.
- [20] S. Workman, C. Greenwell, M. Zhai, R. Baltenberger, and N. Jacobs, "DEEFOCAL: A method for direct focal length estimation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 1369–1373.
- [21] Y. Hold-Geoffroy et al., "A perceptual measure for deep single image camera calibration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2354–2363.
- [22] K. Liao, C. Lin, L. Liao, Y. Zhao, and W. Lin, "Multi-level curriculum for training a distortion-aware barrel distortion rectification model," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4369–4378.
- [23] X. Li, B. Zhang, P. V. Sander, and J. Liao, "Blind geometric distortion correction on images through deep learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4850–4859.
- [24] J. I. Ronda, A. Valdés, and G. Gallego, "Line geometry and camera autocalibration," *J. Math. Imag. Vis.*, vol. 32, no. 2, pp. 193–214, Oct. 2008.
- [25] Z. Zhang, "Camera calibration with one-dimensional objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 7, pp. 892–899, Jul. 2004.
- [26] L. Grammatikopoulos, G. Karras, E. Petsa, and I. Kalisperakis, "A unified approach for automatic camera calibration from vanishing points," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 36, no. 5, 2006.
- [27] E. Trucco, *Introductory Techniques for 3D Computer Vision*, vol. 2. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
- [28] D. Farin, J. Han, and P. H. N. de With, "Fast camera calibration for the analysis of sport sequences," in *Proc. IEEE Int. Conf. Multimedia Expo*, Sep. 2005, pp. 482–485.
- [29] D. Farin et al., "Robust camera calibration for sport videos using court models," in *Storage and Retrieval Methods and Applications for Multimedia 2004*, vol. 5307, Bellingham, WA, USA: SPIE, Oct. 2003, pp. 80–91.
- [30] M.-C. Hu, M.-H. Chang, J.-L. Wu, and L. Chi, "Robust camera calibration and player tracking in broadcast basketball video," *IEEE Trans. Multimedia*, vol. 13, no. 2, pp. 266–279, Apr. 2011.
- [31] H. Bozdogan, "Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions," *Psychometrika*, vol. 52, no. 3, pp. 345–370, Sep. 1987.
- [32] K. Kanatani, "Uncertainty modeling and model selection for geometric inference," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 10, pp. 1307–1319, Oct. 2004.
- [33] K. Kinoshita and L. Lindenbaum, "Camera model selection based on geometric AIC," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Aug. 2000, pp. 514–519.
- [34] R. K. Hamad, B. Hamed, and H. Hassonny, "The automatic selection of radial distortion models," *Int. J. Comput. Appl.*, vol. 179, no. 8887, pp. 9–12, May 2018.
- [35] V. Orekhov, B. Abidi, C. Broaddus, and M. Abidi, "Universal camera calibration with automatic distortion model selection," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2007, pp. 397–400.
- [36] P. H. S. Torr, "An assessment of information criteria for motion model selection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Aug. 1997, pp. 47–52.
- [37] M. Polic, S. Steidl, C. Albl, Z. Kukeleva, and T. Pajdla, "Uncertainty based camera model selection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5990–5999.
- [38] P. Rojberg and A. Kuijper, "Efficient pose selection for interactive camera calibration," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Oct. 2018, pp. 31–36.
- [39] A. Richardson, J. Strom, and E. Olson, "AprilCal: Assisted and repeatable camera calibration," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2013, pp. 1814–1821.
- [40] J.-Y. Bouguet, "Camera calibration toolbox for MATLAB," Microprocessor Res. Labs, Intel Corp., Santa Clara, CA, USA, Tech. Rep., 2000. [Online]. Available: <http://vigor.missouri.edu/Images/Caltech-Toolbox/>
- [41] MathWorks. (2010). *MATLAB Computer Vision Toolbox*. [Online]. Available: <https://www.mathworks.com/help/vision/camera-calibration.html>
- [42] J. Kannala and S. S. Brandt, "A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 8, pp. 1335–1340, 2006.
- [43] F. Rameau, J. Park, O. Bailo, and I. S. Kweon, "MC-calib: A generic and robust calibration toolbox for multi-camera systems," *Comput. Vis. Image Understand.*, vol. 217, Mar. 2022, Art. no. 103353, doi: [10.1016/j.cviu.2021.103353](https://doi.org/10.1016/j.cviu.2021.103353).
- [44] P. Furgale, J. Maye, and J. Rehder, "The Kalibr visual-inertial calibration toolbox," Auton. Syst. Lab, ETH Zurich, Zurich, Switzerland, Tech. Rep., 2013. [Online]. Available: <https://github.com/ethz-asl/kalibr>
- [45] D. Scaramuzza, "OCamCalib: Omnidirectional camera calibration toolbox for MATLAB," Comput. Vis. Robot. Lab (CVLAB), EPFL, Lausanne, Switzerland, Tech. Rep., 2006. [Online]. Available: <https://sites.google.com/site/scarabotix/ocamcalib-omnidirectional-camera-calibration-toolbox-for-matlab>
- [46] A. E. Conrady, "Decentered lens-systems," *Monthly Notices Roy. Astronomical Soc.*, vol. 79, no. 5, pp. 384–390, Mar. 1919.
- [47] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*. Sebastopol, CA, USA: O'Reilly Media, 2008.
- [48] J. Kannala and S. S. Brandt, "A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 8, pp. 1335–1340, Aug. 2006.
- [49] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4104–4113.
- [50] A. W. Fitzgibbon, "Simultaneous linear estimation of multiple view geometry and lens distortion," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. CVPR*, vol. 1, Jul. 2001, pp. 125–132.



Quy Nguyen Cong received the B.S. degree in electronics and telecommunication engineering from The University of Danang—University of Science and Technology, Danang, Vietnam, in 2021. He is pursuing the master's degree with the Department of Computer Science and Engineering, Seoul National University of Science and Technology (SEOULTECH), Seoul, Republic of Korea.

His research interests include machine learning, 3-D computer vision, and autonomous driving.



Sunglok Choi (Member, IEEE) received the B.S. degree in mechanical and aerospace engineering from Seoul National University, Seoul, Republic of Korea, in 2006, and the M.S. and Ph.D. degrees in robotics from KAIST, Daejeon, Republic of Korea, in 2008 and 2019, respectively.

From 2008 to 2021, he was a Research Scientist at ETRI, Daejeon. Since 2021, he has been an Assistant Professor at the Department of Computer Science and Engineering, SEOULTECH, Seoul. His research interests include autonomous navigation and 3-D computer vision.