SN Computer Science (2026) 7:586
https://doi.org/10.1007/s42979-026-05176-1

ORIGINAL RESEARCH

Joint Static Calibration for Multiview Phase-Measuring Profilometry

Hyeongjun Cho[1] · **Min H. Kim**[1]

Received: 8 October 2025 / Accepted: 1 June 2026

Abstract

Conventional phase-measuring profilometry (PMP) calibration methods are primarily designed for single-view systems. Extending PMP to multiview configurations typically requires separate geometric calibrations for each camera and projector, often relying on repeated manual repositioning of a calibration target. This decoupled, two-stage procedure is labor-intensive and prone to error accumulation and local minima. We propose a unified framework for static calibration of multiview PMP systems that jointly optimizes geometric and phase-to-height parameters in a single formulation. Our approach is built on two key components: (a) a compact multi-planar calibration target with embedded fiducial markers, which removes the need for target motion, and (b) a joint optimization strategy that simultaneously estimates intrinsic, extrinsic, and lens distortion parameters of all cameras and projectors from a single static scene. The proposed pipeline automatically detects planar features, constructs multiview homographies via a target-aware bundle adjustment, and incorporates unwrapped phase measurements to refine system parameters. This fully automated approach significantly reduces user intervention while improving stability across views. Experimental results demonstrate that our method provides accurate calibration and supports consistent fusion of multiview phase measurements.

Keywords 3D imaging · Multiview camera calibration · Phase-measuring profilometry

Introduction

Phase-measuring profilometry (PMP) is a widely used technique for high-accuracy 3D reconstruction of static objects. It estimates surface geometry by projecting phase-shifted sinusoidal patterns onto a scene and capturing the reflected images with one or more cameras. Each camera pixel encodes surface position through a measured phase value, which serves as a proxy for depth. By exploiting the correspondence between camera pixel coordinates and phase measurements, PMP can recover dense height or depth maps with subpixel precision, making it attractive for a wide range of industrial inspection and metrology applications.

Existing PMP calibration approaches can be broadly divided into two categories: phase-to-height models and multiview geometry models. Traditional phase-to-height methods calibrate a single camera-projector pair by fitting an explicit mapping from phase to surface height of stepped calibration target, often under restrictive geometric assumptions such as a shared baseline plane [1]. To relax these assumptions, various linear inverse [2–5] and polynomial [6–11] models have been proposed.

However, relying on such multi-level targets poses a significant challenge in optimization, making the process prone to local minima. Since the phase-to-height mapping is intrinsically coupled with the observed pillar heights at each pixel, the system mandates a strict geometric alignment between the projector and camera. Furthermore, this approach requires establishing a direct phase-to-height mapping for each individual camera-projector pair, without performing explicit calibration of intrinsic and extrinsic parameters. As a result, they do not naturally scale to systems involving multiple cameras and projectors, and typically require an additional geometric calibration step, such as Zhang's method [12], to support multiview setups.

In contrast, multiview geometry-based PMP models treat the projector as an inverse pinhole camera and interpret phase measurements as geometric constraints. By leveraging epipolar geometry, these methods recover 3D

✉ Min H. Kim
minhkim@vclab.kaist.ac.kr

Hyeongjun Cho
hjcho@vclab.kaist.ac.kr

[1] School of Computing, KAIST, Daejeon 34141, South Korea

points through triangulation [13–15] or ray-plane intersection [16, 17]. Because all intrinsic and extrinsic parameters of cameras and projectors are explicitly modeled, multiview geometry formulations naturally support scalable multiview reconstruction and accurate geometric registration.

Despite this advantage, calibrating multiview PMP systems remains challenging in practice. Conventional pinhole calibration methods require capturing multiple images of a checkerboard target at different poses [12], followed by a separate calibration procedure for each projector using projected patterns [18, 19]. Even recent self-calibration approaches based on RGB-D sensors [20] do not fully eliminate the need for repeated target manipulation. These procedures are labor-intensive and difficult to deploy in compact or constrained environments, such as the limited workspace inside a multiview 3D scanning instrument.

In this paper, we address these challenges by introducing a static calibration framework with a joint multiview optimization framework for multiview PMP systems. Our approach is centered on a compact multi-planar calibration target, where each planar surface is embedded with ArUco fiducial markers [21]. The markers enable accurate and metrically consistent estimation of the target geometry, while the diverse orientations of the planar surfaces provide useful geometric constraints to estimate intrinsic, extrinsic, and lens distortion parameters through multi-plane homography extraction.

A key difficulty in static calibration arises from lens distortion nonlinearity. When planar surfaces occupy only a small region in screen space, the resulting homographies provide weak constraints on distortion, making the optimization susceptible to overfitting and local minima. This issue is particularly pronounced in compact multi-planar targets, where increasing the number of surfaces necessarily reduces the screen coverage of each individual plane. To address this problem, we introduce a multiview geometric consensus that couples the poses of all planar surfaces across cameras and projectors. By extending bundle adjustment [22] to the multiview PMP setting, this constraint regularizes lens distortion parameters and stabilizes joint estimation of intrinsics and extrinsics.

Importantly, the shape of our calibration target is flexible and can be adapted to different hardware configurations, as long as a sufficient number of planar surfaces with visible fiducial markers are observed. This allows the target design to accommodate various camera-projector layouts, fields of view, and depth ranges. Moreover, our formulation optimizes a dense objective defined over phase measurements rather than sparse feature correspondences, leading to improved robustness and accuracy in practical multiview PMP systems.

In summary, our main contributions are:

- A customizable static calibration target that enables efficient single-shot calibration for multiview PMP systems without manual target manipulation.
- A joint multiview optimization framework that jointly estimates geometric and distortion parameters, with multiview geometric consensus improving optimization stability.

Related Work

Phase measuring profilometry has been widely adopted for high-precision 3D reconstruction due to its pixel-wise spatial resolution. The typical PMP pipeline consists of three sequential stages: (a) phase extraction to compute the wrapped phase from intensity images, (b) phase unwrapping to resolve the 2π ambiguity, and (c) phase-to-3D mapping to convert the absolute phase into metric coordinates. In this section, we briefly review the foundational methods and recent advancements associated with each of these key components.

Phase Extraction

PMP uses sequence of N sinusoidal structured light images with frequency f encodes a phase that is linearly related to the projector's pixel index in either the vertical or horizontal direction [23]. For $k \in [0, N-1]$, when the k-th structured light image with frequency f is projected onto an object, the reflected light is captured by a camera, resulting in pixel intensity $\mathrm{I}_{fk}(u,v)$ at each pixel location (u, v):

$$\mathrm{I}_{fk}(u,v) = \mathrm{I}_a(u,v) + \mathrm{I}_b(u,v)\cos(\phi_f(u,v) + 2\pi k/N), \qquad (1)$$

where $\phi_f(u,v) \in [-f\pi, f\pi]$ denotes the wrapped phase. The phase $\phi_f(u,v)$ can be recovered using the following arctangent-based formulation:

$$\phi_f(u,v) = \tan^{-1}\frac{\sum_{k=0}^{N-1} I_{fk}(u,v)\sin(2\pi k/N)}{\sum_{k=0}^{N-1} I_{fk}(u,v)\cos(2\pi k/N)}. \qquad (2)$$

Phase Unwrapping

Due to the limited dynamic range of projectors, a single frequency is often insufficient to recover high-precision phase. Therefore, structured light images at multiple frequencies are used to unwrap the phase map. Specifically, the structured light sequences are typically generated by recursively multiplying a base frequency f by a scale factor $\alpha > 1$, generating n different frequency levels.

Given wrapped phase maps ϕ_f where $f = \alpha^0, \ldots, \alpha^{n-1}$, phase unwrapping algorithms compute $\psi_{\alpha f}$ using previously unwrapped ψ_f and newly measured $\phi_{\alpha f}$ [15]. The initial phase is set to $\psi_1(u, v) = \phi_1(u, v)$. The final unwrapped phase is $\psi(u, v) = \psi_f(u, v)$ where $f = \alpha^{n-1}$. The relationship between a wrapped phase $\phi_f(u, v)$ and its corresponding unwrapped phase $\psi_f(u, v)$ is given by:

$$\psi_f(u, v) = \phi_f(u, v) + 2\pi h_f(u, v), \tag{3}$$

where $h_f(u, v)$ is the fringe order, an integer corresponding to frequency *f*.

Given ψ_f and $\phi_{\alpha f}$, the fringe order $h_{\alpha f}$ at frequency αf is estimated by assuming $\psi_{\alpha f} \approx \alpha \psi_f$, leading to:

$$h_{\alpha f}(u, v) = \text{round}\left(\frac{\alpha \psi_f(u, v) - \phi_{\alpha f}(u, v)}{2\pi}\right). \tag{4}$$

The corresponding unwrapped phase $\psi_{\alpha f}(u, v)$ is then computed using Equation (3). In our method, we choose $\alpha = 4$ and $n = 4$, resulting in structured light frequencies of 1, 4, 16, and 64.

Phase-Height Models

Linear models. Traditional calibration models describe the inverse of height *h* as a linear inverse function of $\Delta\psi$, where $\Delta\psi = \psi - \psi_{\text{ref}}$ and ψ_{ref} is a pre-calculated phase corresponding to a reference plane [1–5]. However, these models require a strict geometric constraint on the relative positions of the system components. To relax this constraint, the phase-height function is extended to more complex forms. In particular, when accounting for tilted cameras or projectors, the key difference from the traditional model is that the coefficients in the phase-height mapping are no longer globally consistent; instead, they vary pixel by pixel. Moreover, the calibration of the linear inverse model is highly sensitive to its initial parameter values, and may fail to converge if the initialization is far from the true solution. A fundamental limitation of this approach is its inability to effectively model nonlinear effects such as lens distortion within the phase-height function.

Polynomial models To address the nonlinearity of the phase-height relationship, several approaches have been proposed to extend the inverse linear model to an inverse polynomial formulation [6, 7]. However, these models still suffer from convergence issues when initialized with poor parameter values. With advances in computational power, numerous inverse polynomial models have recently been introduced [8–11]. To mitigate the risk of non-convergence, some methods attempt to approximate the inverse polynomial function using a direct polynomial form [24]. However, selecting the maximum polynomial order requires careful consideration due to Runge's phenomenon [25]. The main limitations of both polynomial and inverse polynomial models are their lack of robustness and inconsistency. These models typically require pixel-wise coefficients for the phase-height function in order to relax geometric constraints, which complicates calibration and reduces generalizability.

Governing equation models The primary distinction between governing equation models and other phase-height models is that the governing equation is defined as a function of u, v, ψ, rather than $\Delta\psi$, where *u* and *v* are the pixel indices of the camera. The basic form of the governing equation is expressed as a rational function, in which the highest order of u, v, ψ is one in both the numerator and denominator [26].

Wang et al. refined the relationship between normalized coordinates and pixel indices by introducing higher-order terms in the governing equation [27]. Lee et al. extended the model from a phase-to-height formulation to a phase-to-3D-point mapping by calibrating separate governing equations for *x*, *y*, and *h*, respectively [28]. By incorporating pixel indices as variables, governing equation models rely on global parameters rather than pixel-wise coefficients [29, 30], which enhances their generalizability. An additional advantage of this formulation is its robustness to lens distortion. Since both the numerator and denominator are polynomial functions of u, v, ψ, lens distortion from both the camera and projector can be absorbed into the governing equation parameters [31].

Multiview Geometry Models

Multiview geometry models are grounded in epipolar geometry. In practice, Zhang's method [12] is widely used, which involves capturing multiple images of a checkerboard target at various poses by physically moving the calibration board.

In structured-light and phase-measuring/fringe-projection profilometry (PMP/FPP), the projector is often not modeled through a fully explicit physical-optics description, but through a practical geometric abstraction that makes calibration and 3D reconstruction tractable. One common choice is to treat the projector in a pinhole-like inverse-camera form, so that phase-derived constraints can be related to projector-domain geometry in a stereo-style formulation. This modeling choice is consistent with a broader class of practical phase-to-geometry approximations in the literature, including reference-plane-based phase-to-height models and calibration-oriented surrogate mappings. In this sense, the inverse-camera interpretation should be understood not as a unique physical description, but as one of several widely used geometric approximations for structured-light system calibration.

3D reconstruction in this framework can be achieved via ray-ray matching [13–15] or ray-plane intersection [16, 17], where phase values are mapped to the normalized coordinates of the projector. A major advantage of multiview geometry models is that they provide explicit calibration of all relevant parameters for both cameras and projectors.

In contrast, phase-height models primarily describe the geometric and physical relationship among a camera, a projector, and a reference plane, and are typically limited to a single camera-projector pair. This makes them inefficient and less scalable for multiview systems. Moreover, such calibration procedures require capturing many checkerboard images while physically rotating the target, which is inconvenient and often impractical within the confined space of a 3D scanning instrument head.

It is also important to note that phase-height models do not account for either implicit or explicit system parameters, nor for lens distortion in cameras and projectors. Therefore, they are not directly applicable to multiview configurations, and must be complemented with a separate calibration method such as Zhang's approach [12]. Although recent methods have attempted to calibrate systems with multiple cameras and projectors without fiducial markers [32], they still suffer from the scale ambiguity inherent to structure-from-motion.

To overcome these limitations, we aim to integrate the strengths of both phase-to-height and multiview geometry models. To this end, we design a novel static calibration target and develop a multiview calibration framework. The following section presents the technical details of our approach.

Bundle adjustment. Bundle adjustment [22] is a widely adopted optimization technique for refining camera parameters to achieve optimal geometric consistency. This method is also applicable to camera-projector systems. Furukawa et al. apply bundle adjustment as the final step in their calibration pipeline to refine focal length parameters in a camera-projector setup [33]. Garrido-Jurado et al. extend this approach to jointly calibrate both camera and projector parameters [34], while Li et al. introduce a weighted objective function to improve optimization stability [35].

These methods generally rely on self-calibration, where the target shape is initially unknown. Consequently, the accuracy of system parameters and reconstructed geometry often suffers when the target is geometrically complex, as the optimization relies on sparse feature correspondences. In contrast, our method utilizes fiducial markers, which enable robust and reliable parameter estimation, while simultaneously establishing an absolute scale for the entire calibration process.

While prior work has explored camera and projector calibration, multiview geometry formulations, and static calibration targets in isolation, existing approaches do not jointly address phase-measuring profilometry calibration in a multiview setting using a single static target. In particular, previous methods typically rely on either movable calibration targets, separate phase-to-height fitting, or sparse geometric constraints. In contrast, our work focuses on unifying geometric calibration and phase-based constraints within a single joint optimization framework tailored to multiview PMP systems.

Static Calibration Target Design

Target design. We present a novel static calibration target specifically designed for multiview phase-measuring profilometry systems. The target comprises multiple planar surfaces arranged with diverse orientations, providing sufficient geometric diversity to estimate intrinsic parameters, extrinsic transformations, and lens distortion coefficients from homographies obtained in a single static capture.

The calibration target is physically realized by attaching a printed paper pattern containing ArUco fiducial markers [21] onto a rigid aluminum substrate. The aluminum body ensures high planarity and geometric stability, while the printed marker pattern supplies reliable fiducial features for automatic plane detection and pose estimation.

Each surface is equipped with fiducial markers whose local coordinates are predefined, and the coordinate system of the bottom surface is designated as the global reference frame. Accordingly, estimating the rotation and translation of each planar surface with respect to the world coordinate system becomes part of the overall calibration process.

Considering the printable resolution of the fiducial markers, our design incorporates 18 polygonal faces, each affixed to the surface of the 3D structure (Fig. 1b). To ensure reliable phase decoding under sinusoidal pattern projection, the target surfaces are printed in middle-gray tones. The physical dimensions of the target are $60 \times 60 \times 12.3$ mm.

Design rationale of the calibration target. The calibration target was designed to balance two competing requirements of the proposed single-shot initialization. On the one hand, each view should observe as many planar surfaces as possible, since multiple visible planes improve the geometric constraints available for camera and target initialization. On the other hand, sufficient usable area must remain on each plane to place fiducial markers for reliable homography estimation. The adopted target geometry was therefore chosen so that multiple planes with distinct orientations could be simultaneously observed while still preserving enough marker-covered regions for robust corner detection and

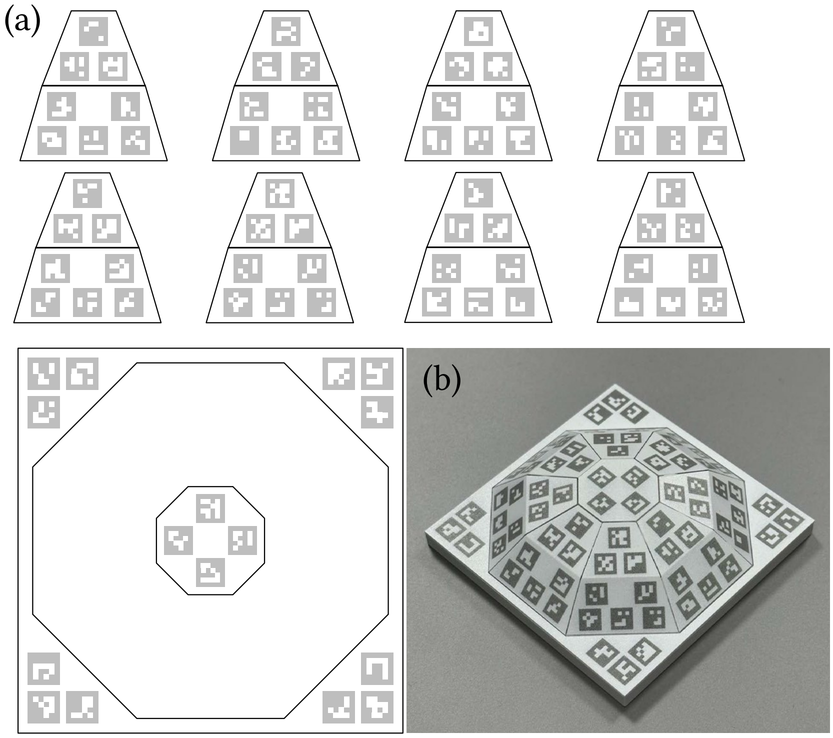


Fig. 1 Our novel PMP calibration target design. **a** Unfolded surfaces with unique fiducial markers, which are attached to the 3D model. **b** A fabricated calibration target. Courtesy of Cho and Kim [36]

homography computation. In this sense, the target design is heuristic and was guided by practical initialization requirements rather than by an explicit conditioning optimization.

Note that the target geometry is not restricted to a fixed configuration. The model shown in Fig. 1 illustrates one possible instantiation that is compatible with the calibration method described in Section Nonlinear Optimization of PMP Calibration.

Target geometry model. Our calibration target is composed of independent planar surfaces, each defined in its own local coordinate system. For convenience, we designate surface index 0 as the reference plane at the bottom, such that its local coordinate system coincides with the world coordinate system.

Each surface is equipped with a sufficient number of ArUco markers, ranging from at least 3 to at most 12, to ensure reliable pose estimation. Each marker has a unique ID, allowing the system to determine the specific surface on which its corners reside.

We parametrize the pose of each surface as a rigid body transformation $(\mathrm{R}_s, \mathrm{t}_s) \in SE(3)$ that maps the predefined local marker coordinates to the global world frame. To ensure numerical stability and valid updates on the geometric manifold during optimization, we utilize the Lie algebra $\mathfrak{se}(3)$ representation provided by the lietorch [37]. This approach allows us to perform gradient descent on the tangent space of the special Euclidean group, avoiding singularities associated with Euler angles.

Given a point $\mathbf{X}_{si} = [x_{si} \quad y_{si} \quad 0]^\mathsf{T} \in \mathbb{R}^3$ be an local coordinate of the i-th corner point in surface s, the corresponding world coordinates are

$$\hat{\mathrm{X}}_w = \begin{matrix} \mathrm{R}_s & \mathrm{t}_s \\ 0 & 1 \end{matrix} \hat{\mathrm{X}}_{si} \, , \tag{5}$$

where the hat symbol denotes that a point is expressed in a homogeneous coordinate notation, and $\mathbf{R}_s$, $\mathbf{t}_s$ are the rotation/translation from local coordinates of surface s to world coordinates. Then given rotation and translation of surface s, we can derive a plane equation in terms of $\mathbf{n}_s$, the normal of the surface and c_s, the constant of the surface as: $\mathbf{n}_s \cdot \mathrm{X}_w + c_s = 0$, where $\mathrm{n}_s = \mathrm{R}_s[0 \quad 0 \quad 1]^\mathsf{T}$ and $c_s = -\mathrm{n}_s \cdot \mathrm{t}_s$.

Camera model. The intrinsic and extrinsic parameters of a camera or projector define how a 3D point is projected onto the image plane. These parameters include the camera (or projector) matrix K, which encodes focal lengths along the *x* and *y* axes and the principal point (u_0, v_0), along with the distortion coefficients $\mathrm{d}_c = (k_1, k_2, k_3)$, and the rotation and translation vectors [12].

Given a point $\mathrm{X}_w = [x_w \quad y_w \quad z_w]^\mathsf{T}$ in world coordinates, it can be transformed into camera coordinates as follows:

$$\mathrm{X}_c = [x_c \quad y_c \quad z_c]^\mathsf{T} = \mathrm{R}_c \mathrm{X}_w + \mathrm{t}_c . \tag{6}$$

This point is then projected onto the normalized image plane by

$$\mathrm{x}_c = [x_n \quad y_n]^\intercal = [x_c/z_c \quad y_c/z_c]^\intercal. \tag{7}$$

Finally, for a camera c (or projector p), the image coordinates are computed using

$$\hat{\mathrm{p}}_c = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathrm{K}_c \hat{\mathrm{x}}_c = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \hat{\mathrm{x}}_c, \tag{8}$$

where the hat symbol indicates homogeneous coordinates.

However, this image coordinate cannot be directly compared to actual pixel measurements due to lens distortion present in the raw image. In our model, radial distortion is corrected in the normalized coordinate space. A distorted pixel location $\hat{\mathrm{p}'}_c$ is first mapped to normalized coordinates by

$$\hat{\mathrm{x}}'_c = [x'_n \quad y'_n \quad 1]^\intercal = \mathrm{K}_c^{-1} \hat{\mathrm{p}'}_c. \tag{9}$$

The distortion is then removed by

$$\begin{aligned} \hat{\mathrm{p}}_c^{\text{undist}} &= \mathrm{K}_c [x_c \quad y_c \quad 1]^\intercal = \mathrm{K}_c \times \mathrm{undist}(\hat{\mathrm{x}'}_c, \mathrm{d}_c) \\ &= \mathrm{K}_c \begin{bmatrix} x'_n (1 + k_1 {r'}^2 + k_2 {r'}^4 + k_3 {r'}^6) \\ y'_n (1 + k_1 {r'}^2 + k_2 {r'}^4 + k_3 {r'}^6) \\ 1 \end{bmatrix}, \end{aligned} \tag{10}$$

where $r' = \sqrt{{x'_n}^2 + {y'_n}^2}$ [38].

A telecentric camera follows an orthographic projection model, whereas a pinhole camera uses a perspective projection model. In the orthographic case, the z-component is ignored [39]. To incorporate this into our model, we modify x_c in Equation (7) as follows:

$$\mathrm{x}_c = [x_n \quad y_n]^\intercal = [x_c \quad y_c]^\intercal. \tag{11}$$

Projector model. The projector model is an inverse pinhole camera model in our method, the only difference is that the projector uses phases in horizontal and vertical axes rather than uv-image coordinates. However, the phases have linear relation with uv-coordinates, $\psi_u = w_u u + b_u$ and $\psi_v = w_v v + b_v$.

$$\begin{bmatrix} \psi_u \\ \psi_v \\ 1 \end{bmatrix} = \begin{bmatrix} w_u & 0 & b_u \\ 0 & w_v & b_v \\ 0 & 0 & 1 \end{bmatrix} \mathrm{K}_p \begin{bmatrix} x_n \\ y_n \\ 1 \end{bmatrix}. \tag{12}$$

Since the relationship between the pixel index of the projector and the phase is unnecessary to our model, we can define a new projector matrix K'_p, which maps normalized coordinates to phases.

$$\begin{aligned} \mathrm{K}'_p &= \begin{bmatrix} w_u & 0 & b_u \\ 0 & w_v & b_v \\ 0 & 0 & 1 \end{bmatrix} \mathrm{K}_p \\ &= \begin{bmatrix} w_u & 0 & b_u \\ 0 & w_v & b_v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} w_u f_x & 0 & w_x u_0 + b_u \\ 0 & w_v f_y & w_y v_0 + b_v \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned} \tag{13}$$

Nonlinear Optimization of PMP Calibration

Overview. We propose a nonlinear joint calibration framework for multiview phase-measuring profilometry systems, in which all camera and projector parameters are refined within a unified optimization process. The overall pipeline consists of four main stages that progressively introduce geometric and phase-based constraints. First, we detect fiducial markers on the calibration target and estimate the plane equations of each planar surface. Using the marker corner observations from multiple cameras, we obtain initial estimates of the intrinsic and extrinsic parameters of each camera. Next, we project phase-shifted structured light patterns with multiple frequencies and compute unwrapped phase maps for each camera–projector pair. Based on these phase measurements, we derive dense per-pixel 3D point estimates across all views. Finally, we jointly refine all camera and projector parameters by minimizing the sum of point-to-plane distances between the reconstructed 3D points and their corresponding target planes through iterative nonlinear optimization.

Initial Camera Calibration Using Target

Camera and projector models. As a first step toward joint optimization, we define the geometric models of both cameras and projectors used in our calibration framework. All devices are represented using variants of the standard pinhole camera model.

Perspective cameras are modeled with a pinhole projection augmented by a sixth-order radial distortion model. Telecentric cameras are handled by omitting the z-axis projection component, reflecting their parallel-ray imaging geometry. Projectors are treated as inverse pinhole cameras, allowing phase measurements to be interpreted as projector ray constraints within the same geometric formulation.

To calibrate camera parameters from a single shot, multiple surfaces with fiducial markers must be visible within a single image. Theoretically, at least three surfaces with distinct poses are required to compute homographies [12].

The optimization process involves two main steps. First, the *i*-th corner on surface s is projected to the undistorted image coordinate p_{csi}. Second, the undistorted pixel location $\mathrm{p}^{\mathrm{undist}}_{csi}$ is obtained by correcting distortion from the detected image corner or pixel index. The loss is then defined as the squared Euclidean distance between p_{csi} and $\mathrm{p}^{\mathrm{undist}}_{csi}$:

$$\begin{aligned}&\mathcal{L}_{csi}(\mathrm{p}'_{csi}, \mathrm{X}_{si})\\&= \|\mathrm{p}^{\mathrm{undist}}_{csi}(\mathbf{p'}_{csi}; \mathbf{K}_c, \mathrm{d}_c) - \mathbf{p}_{csi}(\mathbf{X}_{si}; \mathbf{R}_c, \mathbf{t}_c, \mathbf{K}_c, \mathrm{R}_s, \mathrm{t}_s)\|_2^2.\end{aligned} \tag{14}$$

Optimizing the target shape and camera parameters from a single viewpoint is highly ambiguous, resulting in a solution space with substantial ambiguity and multiple plausible solutions. However, by leveraging the multiview geometric consensus inherent to our target-encircling setup, we enforce a global consensus on the target geometry. This constraint helps reduce the ambiguity of the solution space, reducing the effective variance of the estimate and stabilizing the target shape estimation. Subsequently, this refined geometric structure serves as a stable geometric anchor for the global bundle adjustment, improves stability of the entire system.

Let $\mathrm{S}_{\mathrm{cam}}$ and $\mathrm{S}_{\mathrm{surface}}$ denote the sets of cameras and surfaces, respectively, and let $\mathbf{\Theta}_{\mathrm{cam}}$ and $\mathbf{\Theta}_{\mathrm{surface}}$ represent the parameter sets, where $\theta_c = \{\mathrm{K}_c, \mathrm{R}_c, \mathrm{t}_c, \mathrm{d}_c\}$ for each camera $c \in \mathrm{S}_{\mathrm{cam}}$ and $\theta_s = \{\mathrm{R}_s, \mathrm{t}_s\}$ for each surface $s \in \mathrm{S}_{\mathrm{surface}}$. The overall optimization minimizes the sum of all per-corner losses across all cameras and surfaces:

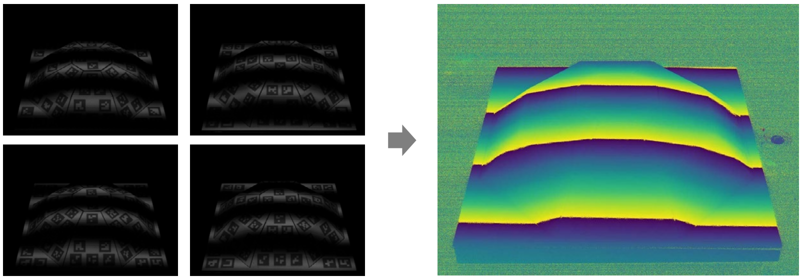
$$\underset{\mathbf{\Theta}_{\mathrm{surface}}, \mathbf{\Theta}_{\mathrm{cam}}}{\mathrm{minimize}} \sum_{c \in \mathrm{S}_{\mathrm{cam}}} \sum_{s \in \mathrm{S}_{\mathrm{surface}}} \sum_{i} \mathrm{L}_{csi}(\mathrm{p}'_{csi}, \mathrm{X}_{si}). \tag{15}$$

During optimization, the world coordinates of all ArUco marker corners are shared globally across views. In practice, this initialization is not obtained by directly optimizing all parameters at once. Instead, we adopt a short staged refinement procedure to improve numerical stability before the full joint optimization. Starting from the homography-based estimates, we first optimize only the surface parameters $\mathbf{\Theta}_{\mathrm{surface}}$ while keeping the camera parameters fixed, so that the coarse target geometry can be adjusted consistently across views. Next, we perform an additional intermediate step in which all parameters except the camera distortion coefficients are optimized jointly. More specifically, after the surface-only refinement, we optimize $\mathbf{\Theta}_{\mathrm{surface}}$ together with the camera intrinsics and extrinsics, while holding the distortion coefficients in $\mathbf{\Theta}_{\mathrm{cam}}$ fixed. The distortion parameters are released only in the final full optimization stage. It is important to note that this initial camera and plane calibration serves strictly as a preliminary initialization step. Its primary purpose is to provide a reasonable starting point (or coarse estimate) for the subsequent joint optimization, rather than serving as the final high-precision calibration method itself (Fig. 2).

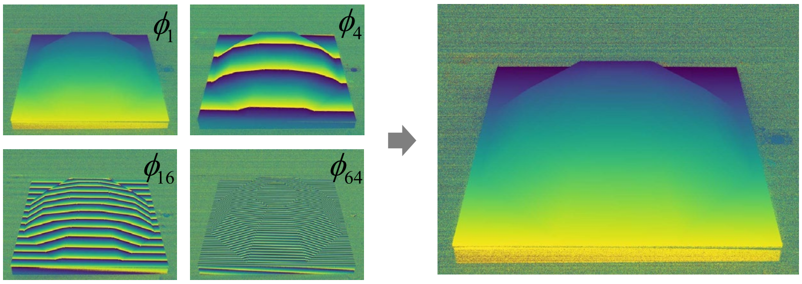
Phase Filtering

To obtain reliable phase measurements for calibration, we apply a two-stage filtering strategy that enforces (i) temporal

Fig. 2 **a** Phase-shifting pattern examples with a frequency ($f = 4$). **b** The reconstructed phase map. **c** Phase maps of four different frequencies. **d** Final unwrapped phase map from camera 1. Courtesy of Cho and Kim [36]



(a) Phase-shifting patterns (ϕ_4) (b) Reconstructed phase map (ϕ_4)



(c) Phase maps of four different frequencies (d) Final unwrapped phase map modulation validity and (ii) local spatial consistency. This preprocessing step removes unreliable phase observations that would otherwise destabilize the subsequent nonlinear optimization.

Modulation-based filtering. We first discard pixels with low signal reliability, such as those affected by specularities, shadows, or dark surfaces. As a reliability metric, we compute the temporal standard deviation of the phase-shifting intensities, denoted by $M(u, v)$. A pixel is retained only if

$$M(u,v) > \tau_{\text{mod}}. \tag{16}$$

Local spatial outlier removal. To suppress impulsive noise and phase unwrapping artifacts, we further apply a local statistical filter to the unwrapped phase map ϕ. Within an $r \times r$ neighborhood, we compute the local mean $\mu_\phi(u,v)$ and standard deviation $\sigma_\phi(u,v)$ using a box filter. We then define a normalized deviation score

$$Z(u,v) = \frac{\phi(u,v) - \mu_\phi(u,v)}{\sigma_\phi(u,v)}. \tag{17}$$

Pixels are retained only if their squared deviation satisfies

$$Z(u,v)^2 < \tau^2_{\text{spatial}}. \tag{18}$$

This criterion effectively removes isolated phase spikes whose values deviate significantly from their local neighborhood, resulting in a smoother and more spatially consistent phase surface for the subsequent bundle adjustment.

Nonlinear Optimization

Depth by intersection. Given a camera pixel location (u_c, v_c) and its corresponding horizontal and vertical unwrapped phases $(\psi_u(u_c,v_c), \psi_v(u_c,v_c))$, we compute the 3D intersection point $\mathrm{X}_{\text{intersect}}$ by intersecting a phase plane derived from the projector with a viewing ray from the camera.

As illustrated in Fig. 3, the normalized projector coordinates $x_{n,p}$ (or $y_{n,p}$) can be computed from the unwrapped phases $(\psi_u(u_c,v_c), \psi_v(u_c,v_c))$ and the projector's intrinsic parameters. Given a line $x = x_{n,p}$ (or $y = y_{n,p}$) on the normalized projector plane and the projector's center of projection (COP) O_p, we compute the normal vector of the phase plane, denoted as n_p (green arrow). The resulting phase plane (green plane) is then defined as the set of points satisfying $\{\mathrm{X} : \mathrm{n}_p \cdot (\mathrm{X} - \mathrm{O}_p) = 0\}$, where X is a 3D point in world coordinates.

The corresponding camera ray (blue arrow) can be expressed as a unit vector $\mathrm{v}_c = \mathrm{x}_n - \mathrm{O}_c$, where x_n is a point on the normalized camera plane corresponding to pixel coordinates (u_c, v_c), and O_c is the camera's COP. The ray can be parametrized as $\{\mathrm{X} : \mathrm{X} = \mathrm{O}_c + \lambda \mathrm{v}_c\}$. Substituting into the phase plane equation yields the intersection point $\mathrm{X}_{\text{intersect}}$ by solving for λ:

$$z_c = \lambda = \frac{\mathrm{n}_p \cdot (\mathrm{O}_p - \mathrm{O}_c)}{\mathrm{n}_p \cdot \mathrm{v}_c}. \tag{19}$$

For telecentric cameras, which follow an orthographic model, the 3D point is expressed as $\{\mathrm{X} : \mathrm{X} = \mathrm{x}_n + \lambda \mathrm{v}_c\}$,

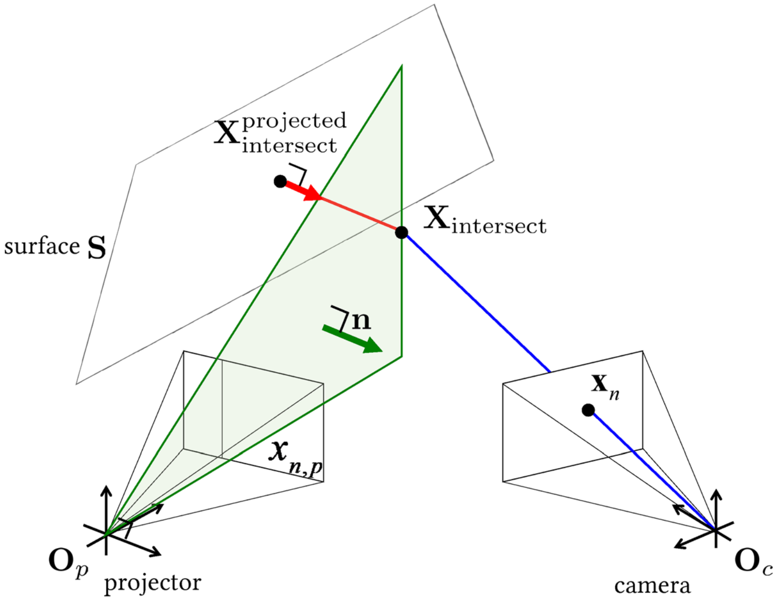


Fig. 3 Visualization of our bundle-adjusting calibration. Once we obtain an unwrapped phase map, we can measure depth by computing an intersection point $\mathrm{X}_{\text{intersect}}$ with a plane formed by $x_{n,p}$ (green plane) and a camera vector x_n (blue arrow). Given $\mathrm{X}_{\text{intersect}}$, our nonlinear optimization updates the camera and the projector's parameters iteratively by minimizing the distance between $\mathrm{X}_{\text{intersect}}$ and $\mathrm{X}^{\text{projected}}_{\text{intersect}}$ (red line) along the surface normal n_s of the target surface s (red arrow). Courtesy of Cho and Kim [36]

where $\mathrm{v}_c = \mathrm{R}_c^{-1}\,[0 \quad 0 \quad 1]^\mathsf{T}$. The scalar λ is then computed as:

$$z_c = \lambda - 1, \qquad \lambda = \frac{\mathrm{n}_p \cdot (\mathrm{O}_p - \mathrm{x}_n)}{\mathrm{n}_p \cdot \mathrm{v}_c}. \tag{20}$$

Thus, the final intersection point is given by $\mathrm{X}_{\text{intersect}} = \mathrm{O}_c + \lambda \mathrm{v}_c$ for a pinhole camera, and $\mathrm{X}_{\text{intersect}} = \mathrm{x}_n + \lambda \mathrm{v}_c$ for a telecentric camera.

Bundle-adjusting optimization. Let S_{proj} and $\boldsymbol{\Theta}_{\text{proj}}$ denote the set of projectors and their corresponding parameters, respectively, where each projector $p \in \mathrm{S}_{\text{proj}}$ has $\theta_p = \{\mathrm{K}_p, \mathrm{R}_p, \mathrm{t}_p, \mathrm{d}_p\}$. For each surface s of the calibration target, we define $A(c, s)$ as the set of pixels in camera c that observe surface s, as identified via the fiducial markers' unique IDs.

Since the poses of all surfaces are already optimized in Section Initial Camera Calibration using Target, we define the bundle adjustment loss as the squared distance between the 3D point $\mathrm{X}_{\text{intersect}}$ and its corresponding surface, measured along the surface normal n_s (i.e., red arrow in Fig. 3):

$$\begin{aligned} &\mathcal{L}_{\text{bundle}}(c, p, s) \\ &= \sum_{(u_c, v_c) \in A(c,s)} \{\mathrm{n}_s \cdot \mathrm{X}_{\text{intersect}}(u_c, v_c, \psi_{u,p}(u_c, v_c), \\ &\psi_{v,p}(u_c, v_c); \theta_c, \theta_p) + c_s\}^2. \end{aligned} \tag{21}$$

Since the calibration of cameras and projectors is interdependent, the entire bundle adjustment must be performed jointly across all views and devices:

$$\underset{\boldsymbol{\Theta}_{\text{cam}}, \boldsymbol{\Theta}_{\text{proj}}}{\text{minimize}} \sum_{p \in \mathrm{S}_{\text{proj}}} \sum_{c \in \mathrm{S}_{\text{cam}}} \sum_{s \in \mathrm{S}_{\text{surface}}} \mathcal{L}_{\text{bundle}}(c, p, s). \tag{22}$$

Unlike traditional solvers that require explicit analytical Jacobians, we leveraged the automatic differentiation (using autograd) to compute the gradients of the loss function with respect to the optimization variables (intrinsics, extrinsics, and surface coefficients). In contrast to prior camera-projector calibration methods that apply bundle adjustment based on reprojection error [33–35], our approach incorporates an additional constraint: the calibration target consists of planar surfaces whose orientations are known from Equation (15). As a result, our objective function minimizes point-to-plane distance errors rather than reprojection errors.

Using this formulation, our method achieves highly accurate estimation of camera and projector parameters. This supports reconstruction of fine-scale surface details without the need for additional uniform color illumination, which is required in some previous methods such as Li et al. [35].

Reconstruction

Before reconstructing the final 3D point cloud, we compute a confidence map associated with each depth map. This confidence map serves two purposes: (1) to filter out unreliable points with low confidence, and (2) to refine the reconstructed 3D geometry through weighted integration.

For each depth map D_c obtained from camera c, we compute the confidence of each pixel based on its geometric consistency across views. Given a 3D point $\mathrm{X}_c(u, v)$ reconstructed from $\mathrm{D}_c(u, v)$, it can be transformed into the coordinate system of another camera c' as:

$$\hat{\mathrm{X}}_{c'} = \begin{matrix} x_{c'} \\ y_{c'} \\ z_{c'} \\ 1 \end{matrix} = \begin{matrix} \mathrm{R}_{c'} & \mathrm{t}_{c'} \\ 0 & 1 \end{matrix} \begin{matrix} \mathbf{R}_c & \mathbf{t}_c \\ 0 & 1 \end{matrix}^{-1} \hat{\mathrm{X}}_c. \tag{23}$$

The transformed point $\hat{\mathrm{X}}_{c'}$ is then projected onto the image plane of camera c' to obtain the pixel location $\mathrm{p}_{c'} = [u' \quad v']^\mathsf{T}$. Based on the discrepancy between the reprojected depth $z_{c'}$ and the observed depth $\mathrm{D}_{c'}(u', v')$, the confidence value $\mathrm{C}_{c,c'}(u, v)$ for the pixel (u, v) in camera c is defined as:

$$\mathrm{C}_{c,c'}(u, v) = \begin{cases} 1, & \text{if } V_{c'}(\mathrm{X}_c(u, v)) = 0, \\ \exp\{-(z_{c'} - \mathrm{D}_{c'}(u', v'))^2\}, & \text{otherwise,} \end{cases} \tag{24}$$

where $V_{c'}$ is a visibility function that returns 1 if $\mathrm{X}_c(u, v)$ is visible to camera c', and 0 otherwise.

The final confidence $\mathrm{C}_c(u, v)$ is aggregated by multiplying consistency scores from all other views:

$$\mathrm{C}_c(u, v) = \prod_{c' \in \mathrm{S}_{\text{cam}},\ c' \neq c} \mathrm{C}_{c,c'}(u, v). \tag{25}$$

Pixels with $\mathrm{C}_c(u, v) < \tau_{\text{conf}}$ are discarded from the reconstruction.

Once the confidence and depth maps are computed for all cameras, we merge them into a unified point cloud by transforming D_c and C_c into world coordinates for every $c \in \mathrm{S}_{\text{cam}}$. Let $\mathrm{S}_{\text{total},c}(u, v)$ denote the set of visible depth-confidence pairs (d, ω) at pixel (u, v) from camera c in the merged point cloud. The refined depth map $\mathrm{D}_c^{\text{refined}}$ is then computed via weighted averaging:

$$\mathrm{D}_c^{\text{refined}}(u, v) = \frac{\sum_{(d,\omega) \in \mathrm{S}_{\text{total},c}(u,v)} \omega \cdot d}{\sum_{(d,\omega) \in \mathrm{S}_{\text{total},c}(u,v)} \omega}. \tag{26}$$

Results and Discussion

Experiment setup. We construct a multiview PMP system consisting of five cameras and four projectors. A telecentric-lens camera is positioned at the top, while four standard perspective cameras are placed at the four cardinal directions on a metallic frame. Each side-facing camera is paired with a projector mounted above it, and each projector emits sinusoidal structured light at four different frequencies, using four-phase shifting patterns with $\pi/2$ intervals. See Fig. 4 for the system configuration.

To maximize illumination coverage, we design the calibration target to have a convex shape, allowing each side surface to effectively receive light from a nearby projector. During the calibration procedure, each projector is paired with all cameras except the one on the opposite side to ensure full surface visibility. In contrast, during the reconstruction process, each projector is paired only with the top-view camera and the side camera on the same face to minimize noise caused by occlusions and shadows, particularly for non-convex objects.

We implement the proposed optimization framework using the Adam optimizer in PyTorch, and optimize rigid-body transformations via Lietorch [37]. Gradient-based optimization on Lie groups via tangent space retraction is now standard in geometric deep learning and bundle adjustment literature [40]. Within this framework, Adam serves as the update rule in the tangent space, where the optimization landscape is locally Euclidean. The use of adaptive learning rates in Adam is particularly beneficial in our setting due to the heterogeneous parameter scales across intrinsic, extrinsic, and distortion groups, which would otherwise require extensive manual tuning of a fixed learning rate.

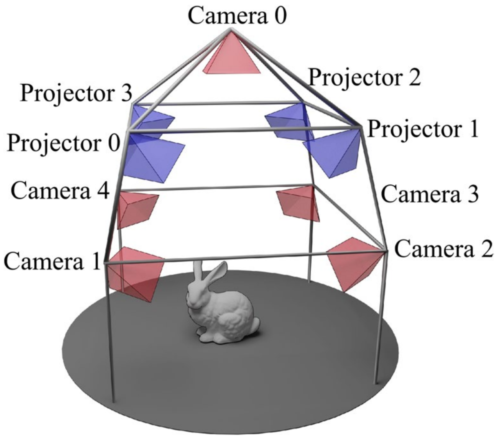


Fig. 4 Schematic diagram of the cross-section view of our multiview PMP imaging setup. Courtesy of Cho and Kim [36]

Synthetic Dataset Generation

To quantitatively evaluate the robustness of the proposed calibration framework under controlled and repeatable conditions, we generate a synthetic dataset using the Mitsuba 3 physically-based renderer [41]. Unlike simple geometric projection models, this ray-tracing environment simulates realistic light transport effects, enabling systematic ablation of specific error sources while preserving exact ground-truth geometry. The synthetic dataset is designed to isolate and analyze key factors that influence calibration stability and accuracy in multiview PMP systems. In particular, the dataset generation protocol focuses on three critical aspects, as detailed below.

Impact of target coverage and distortion zones. To analyze how feature distribution affects lens distortion observability, we design a synthetic experiment that systematically varies the field-of-view (FOV) coverage of the calibration target. The telecentric camera is kept fixed to serve as a stable reference, while the side perspective cameras are progressively retracted along their optical axes. As a result, the projected target region gradually shrinks toward the image center. Because radial distortion is typically negligible near the optical axis and becomes more pronounced toward the image periphery, this setup enables a controlled comparison between two regimes: (i) limited target coverage confined to low-distortion central regions, and (ii) wide coverage that spans the full sensor and captures peripheral distortion effects. This experiment directly probes the sensitivity of the calibration algorithm to distortion under constrained feature coverage, a scenario that commonly arises in compact static calibration targets.

Lens distortion severity. To evaluate the robustness of the proposed calibration algorithm under increasing nonlinear lens effects, we simulate camera configurations with varying levels of radial distortion using the Brown–Conrady model. The distortion parameters are systematically varied to span a wide range of practical imaging conditions. Specifically, the test sets range from mild distortion (e.g., $k_1 \approx -0.05,\ k_2 \approx 0.01,\ k_3 = 0$), which is representative of high-quality optics, to severe distortion (e.g., $k_1 \approx -0.2$, $k_2 \approx 0.05,\ k_3 \approx -0.01$), corresponding to wide-angle or low-cost lenses.

Simulation of optical imperfections and sensor noise. To bridge the gap between idealized synthetic rendering and real-world data acquisition, we incorporate physics-based degradation models that account for both projector optical imperfections and camera sensor noise. These degradations are introduced to evaluate the robustness of the calibration framework under realistic imaging conditions.

- **Projector defocus.** Digital projectors typically exhibit a limited depth of field, which results in slight blurring of high-frequency fringe patterns. We simulate this effect by convolving the ideal rendered images with a Gaussian point spread function (PSF). The degree of defocus is controlled by the standard deviation σ_{defocus} of the Gaussian kernel, effectively modeling the low-pass filtering behavior of an imperfectly focused projector lens.
- **Physics-based sensor noise.** To model the stochastic nature of image formation, we adopt a mixed Poisson–Gaussian noise model. Photon shot noise is first introduced to capture the signal-dependent variability caused by the discrete arrival of photons,

$$I_{\text{shot}} \sim \frac{\mathcal{P}(I_{\text{clean}} \cdot N_{\max})}{N_{\max}}, \tag{27}$$

where I_{clean} denotes the noise-free intensity and $N_{\max}$ represents the maximum photon count, which approximates the full-well capacity of the sensor. We then add signal-independent readout noise, modeled as additive white Gaussian noise with zero mean and standard deviation σ_{read}.

To systematically stress-test the proposed calibration algorithm, we define two representative noise configurations that reflect different hardware qualities:

- **Low-noise setting.** An ideal laboratory scenario with sharp projector focus ($\sigma_{\text{defocus}} = 0.5$ px), sufficient illumination ($N_{\max} = 10{,}000$ photons), and low electronic noise ($\sigma_{\text{read}} = 0.005$).
- **High-noise setting (degraded setup).** A challenging industrial scenario with mild defocus ($\sigma_{\text{defocus}} = 1.0$ px), reduced illumination ($N_{\max} = 5{,}000$ photons), and increased sensor noise ($\sigma_{\text{read}} = 0.01$).

Rendering Configuration

To isolate the factors that influence calibration performance, we employ two distinct rendering configurations, each tailored to a specific experimental objective.

Idealized setup for geometric analysis. For the target coverage and lens distortion experiments, our objective is to analyze the purely geometric effects of feature distribution without interference from photometric or sensor-related factors. We therefore simulate an idealized imaging environment by excluding sensor noise and optical aberrations. The target surface is modeled as a perfect Lambertian diffuser, ensuring that any residual calibration errors are attributable solely to the geometric constraints of the camera and distortion models.

Realistic setup for robustness analysis. In contrast, the optical imperfection and sensor noise experiments are designed to evaluate the stability of the proposed method under challenging real-world conditions. In addition to the defocus and noise models described above, we incorporate non-Lambertian material properties to simulate surface glossiness. Specifically, the target surface is rendered using a rough plastic BSDF with a Beckmann microfacet distribution ($\alpha = 0.1$). This configuration introduces realistic specular reflections and potential intensity saturation, thereby testing the robustness of the calibration algorithm against high-dynamic-range and non-ideal reflectance effects.

Synthesized Calibration Results

In this section, we report synthetic calibration results for both the geometric analysis and the robustness analysis.

Geometric analysis. As illustrated in Fig. 5, we construct six synthetic datasets (Set 1 through Set 6) with systematically varying field-of-view (FOV) coverage. Set 1 represents a configuration in which target features are sparsely distributed near the image center, corresponding to limited coverage, while Set 6 corresponds to a wide-coverage configuration where features span nearly the entire sensor area. The progression from Set 1 to Set 6 is achieved by incrementally retracting the perspective cameras along their optical axes. For each dataset, we evaluate the accuracy of the estimated radial distortion coefficients (k) and extrinsic parameters (R, t).

Table 1 summarizes the calibration results under mild distortion. The extrinsic pose errors decrease as the target moves from the closest configuration (Set 1) to an intermediate range, reaching a minimum around Set 4. This trend indicates that moderate viewing distances provide a favorable balance between triangulation stability and feature observability.

In contrast, the accuracy of distortion estimation exhibits a strong dependence on FOV coverage. When features span the image periphery, the estimated primary radial distortion coefficient is closest to the ground truth. Specifically, in the wide-coverage configuration, the estimated k_1 (≈ -0.08) is nearest to the true value (-0.05). This result confirms that peripheral observations, where distortion effects are most pronounced, are essential for correctly constraining the lens model and preventing parameter drift.

As the camera retracts further and the effective coverage shrinks (Set 6), the quality of intrinsic estimation degrades substantially. In this regime, the estimated k_1 drifts to -0.19 , deviating significantly from the ground truth. This behavior indicates that when features are confined to the low-distortion central region, the optimization lacks sufficient geometric leverage to disentangle distortion from other parameters, leading to numerical coupling and inaccurate intrinsic recovery.

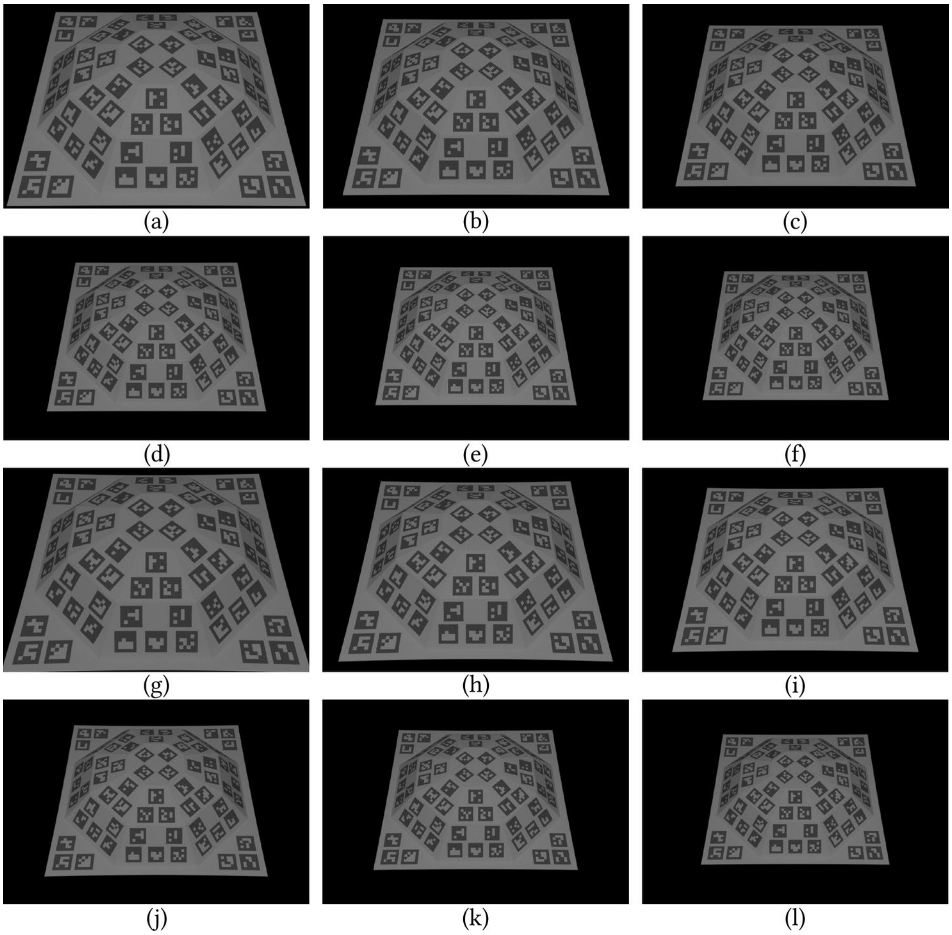


Fig. 5 Representative images from the synthetic geometric dataset used for evaluation. The top row displays the low distortion configuration, while the bottom row corresponds to the high distortion case. The columns labeled (**a–f**) represent the six distinct dataset groups, corresponding to Set 1 through Set 6, respectively.

The results under the high-distortion configuration (Table 2) exhibit trends consistent with the low-distortion case, demonstrating the robustness of the proposed framework. Although the absolute errors increase due to stronger non-linearities, the method remains effective under severe radial distortion ($k_1 = -0.2$). In the optimal configuration, the extrinsic rotation error remains below $0.7°$, indicating that the joint optimization partially seperates distortion effects from the underlying extrinsic geometry.

Robustness analysis. Following the geometric evaluation, we conduct a robustness analysis using synthetic datasets corrupted by physics-based optical imperfections (defocus) and sensor noise. This analysis examines whether the proposed framework can maintain global geometric consistency and reliable pose estimation under challenging conditions involving specular reflections ($\alpha = 0.1$) and low signal-to-noise ratios.

We evaluate four representative configurations that span a range of realistic operating conditions: (1) a baseline setting with reference camera positions, low distortion, a diffuse target, and noise-free measurements; (2) a specular setting that introduces surface glossiness; and (3–4) three noisy settings in which increasing levels of sensor noise (low and high) are injected into the specular configuration.

Table 3 summarizes the calibration performance across these conditions. Comparing the *Baseline* and *Specular* cases reveals stable behavior to non-Lambertian reflectance. Despite the presence of specular highlights ($\alpha = 0.1$), the extrinsic estimates for both cameras and projectors remain stable, with camera translation errors consistently within 7–8 mm. This result indicates that the proposed phase filtering strategy effectively suppresses saturated pixels and specular artifacts without degrading geometric accuracy.

The introduction of sensor noise exposes a fundamental difference between the two sensing modalities. Camera calibration exhibits high robustness: under the *Low Noise* condition, the pose errors are comparable to those of the noise-free baseline. Even in the *High Noise* scenario, although the rotational error increases to approximately $4.4°$ due to feature localization uncertainty, the translational error remains well constrained (below 8 mm), demonstrating the method's practical viability for localization tasks in low-SNR environments.

In contrast, projector calibration is significantly more sensitive to noise. Already at the *Low Noise* level, we observe

Table 1 Calibration results including distortion parameters and pose errors under low distortion configuration

Exp	Device	Distortion			Extrinsic errors (deg / mm)				
		k_1	k_2	k_3	θ_{err}	t_x	t_y	t_z	t_{norm}
Set 1	Cam 1	−0.0855	−0.0345	0.9321	1.3420	0.30	4.47	−5.69	7.24
	Cam 2	−0.0848	−0.0172	0.7148	1.4170	0.42	5.57	−4.79	7.36
	Cam 3	−0.0752	−0.0189	0.8188	1.4079	−0.16	4.86	−5.75	7.53
	Cam 4	−0.0796	−0.0526	0.7766	1.6050	0.75	4.43	−5.45	7.06
	Proj 1	−0.0754	−0.0301	0.0277	2.1029	0.94	4.01	−22.94	23.30
	Proj 2	−0.0430	0.0031	0.0578	2.5680	0.04	5.05	−27.69	28.15
	Proj 3	−0.0782	−0.0309	0.0345	2.7756	−0.43	4.65	−25.95	26.37
	Proj 4	−0.1011	−0.0378	0.0486	2.0008	0.58	5.55	−20.85	21.59
Set 2	Cam 1	−0.1431	0.2717	1.2581	0.9725	0.04	4.58	−4.72	6.57
	Cam 2	−0.1242	0.2080	1.0630	1.2105	0.12	5.37	−4.69	7.13
	Cam 3	−0.1463	0.2779	1.3411	0.8119	−0.16	4.98	−4.09	6.44
	Cam 4	−0.1314	0.2129	1.2551	1.2395	0.47	4.86	−4.61	6.72
	Proj 1	−0.0846	−0.0439	0.0051	1.9620	0.38	2.62	−27.60	27.73
	Proj 2	−0.0686	−0.0374	0.0050	2.3531	−0.46	5.31	−28.86	29.34
	Proj 3	−0.0772	−0.0427	−0.0018	2.2090	0.91	6.95	−28.36	29.22
	Proj 4	−0.0997	−0.0547	−0.0072	1.7693	3.32	4.59	−27.12	27.71
Set 3	Cam 1	−0.1927	0.6347	1.7081	0.6508	−0.02	4.59	−4.00	6.09
	Cam 2	−0.1888	0.6176	1.7238	0.8910	0.09	5.22	−4.02	6.59
	Cam 3	−0.1987	0.6830	1.7410	0.6432	−0.15	4.93	−3.79	6.22
	Cam 4	−0.1810	0.5408	1.7393	0.8753	0.43	5.02	−3.99	6.43
	Proj 1	−0.0999	−0.0658	−0.0255	1.5932	1.09	2.55	−24.86	25.02
	Proj 2	−0.0619	−0.0449	−0.0141	2.3208	−0.24	4.89	−28.46	28.88
	Proj 3	−0.0771	−0.0517	−0.0134	2.2855	0.25	6.62	−27.31	28.11
	Proj 4	−0.0921	−0.0816	−0.0621	1.6964	3.10	4.69	−26.85	27.43
Set 4	Cam 1	−0.2487	1.0747	2.3954	0.3206	0.02	4.69	−2.22	5.19
	Cam 2	−0.2756	1.3627	2.4507	0.4729	0.04	5.12	−2.04	5.51
	Cam 3	−0.2679	1.2586	2.5107	0.2054	−0.07	4.94	−1.43	5.14
	Cam 4	−0.2608	1.2084	2.4620	0.5575	0.29	4.90	−2.48	5.50
	Proj 1	−0.1099	−0.0980	−0.0788	1.6897	0.63	2.70	−23.00	23.16
	Proj 2	−0.0802	−0.0718	−0.0472	2.3356	−0.34	5.31	−25.84	26.38
	Proj 3	−0.0886	−0.0644	−0.0262	2.2739	0.18	6.32	−23.74	24.57
	Proj 4	−0.1001	−0.0794	−0.0496	2.0029	2.13	4.92	−23.12	23.73
Set 5	Cam 1	−0.2698	1.4539	2.6207	0.6114	−0.07	4.91	−3.75	6.18
	Cam 2	−0.2720	1.4740	2.6382	0.6668	0.07	4.98	−3.36	6.00
	Cam 3	−0.2526	1.3827	2.5260	0.6215	−0.05	4.92	−3.84	6.24
	Cam 4	−0.2407	1.1907	2.5159	0.6870	0.46	5.06	−3.82	6.36
	Proj 1	−0.1098	−0.0986	−0.0818	1.8465	0.86	3.14	−23.82	24.04
	Proj 2	−0.0842	−0.0748	−0.0547	2.4926	0.16	4.81	−28.34	28.74
	Proj 3	−0.0788	−0.0930	−0.0905	2.3082	0.32	6.70	−27.75	28.54
	Proj 4	−0.1197	−0.1151	−0.1054	1.6443	2.97	4.48	−25.25	25.81
Set 6	Cam 1	−0.1936	0.8273	2.1331	0.6373	−0.14	3.83	−4.95	6.26
	Cam 2	−0.1379	0.4473	1.5190	1.6402	0.15	6.64	−8.06	10.45
	Cam 3	−0.2298	1.2171	2.3909	0.6792	−1.11	5.01	−4.79	7.02
	Cam 4	−0.1806	0.8015	1.9806	1.3102	1.61	4.87	−6.80	8.52
	Proj 1	−0.1206	−0.1250	−0.1271	1.7334	0.58	2.59	−23.85	24.00
	Proj 2	−0.0925	−0.1087	−0.1143	2.4956	−0.54	5.50	−27.70	28.24
	Proj 3	−0.1204	−0.1105	−0.1011	2.1742	−0.18	7.52	−27.42	28.44
	Proj 4	−0.1244	−0.1220	−0.1189	1.6469	3.32	3.89	−23.74	24.29

The ground truth distortion coefficients are $k_1 = -0.05, k_2 = 0.01, k_3 = 0$

Table 2 Calibration results including distortion parameters and pose errors under high distortion configuration

Exp	Device	Distortion			Extrinsic errors (deg / mm)				
		k_1	k_2	k_3	θ_{err}	t_x	t_y	t_z	t_{norm}
Set 1	Cam 1	−0.1633	−0.0665	0.4711	2.2568	−0.01	3.90	−11.02	11.69
	Cam 2	−0.1654	−0.0566	0.4668	2.4160	−0.09	6.03	−9.34	11.11
	Cam 3	−0.1603	−0.0768	0.4504	2.4711	−0.98	4.91	−11.38	12.43
	Cam 4	−0.1578	−0.0839	0.4742	2.2022	1.11	4.97	−9.04	10.38
	Proj 1	−0.0909	−0.0541	−0.0192	2.4337	0.29	3.10	−35.78	35.92
	Proj 2	−0.0814	−0.0520	−0.0219	3.1399	0.28	4.73	−35.33	35.65
	Proj 3	−0.0732	−0.0489	−0.0212	3.2444	−0.07	4.75	−33.23	33.57
	Proj 4	−0.0983	−0.0588	−0.0246	2.1139	3.17	5.51	−35.59	36.15
Set 2	Cam 1	−0.2184	0.1564	0.9285	1.5314	0.02	3.88	−8.84	9.65
	Cam 2	−0.2177	0.1658	0.8984	1.7170	−0.09	6.01	−7.91	9.94
	Cam 3	−0.2039	0.1090	0.7608	1.8422	−1.02	4.88	−9.77	10.97
	Cam 4	−0.2060	0.1259	0.7741	1.7561	1.24	4.83	−8.43	9.79
	Proj 1	−0.1050	−0.0715	−0.0391	2.3609	0.20	3.00	−33.91	34.04
	Proj 2	−0.0880	−0.0736	−0.0562	3.1701	0.23	5.41	−36.50	36.90
	Proj 3	−0.0847	−0.0605	−0.0332	3.2514	1.08	5.11	−33.89	34.29
	Proj 4	−0.1152	−0.0788	−0.0472	2.0937	3.37	4.03	−36.84	37.22
Set 3	Cam 1	−0.2381	0.3175	1.1221	1.3889	−0.03	3.56	−8.76	9.46
	Cam 2	−0.2336	0.3193	1.0234	1.4301	0.05	6.33	−7.71	9.97
	Cam 3	−0.2404	0.3673	1.0626	1.3792	−1.06	4.88	−8.48	9.84
	Cam 4	−0.2211	0.2591	1.0501	1.5965	1.50	5.07	−8.37	9.90
	Proj 1	−0.1306	−0.1132	−0.0998	1.9640	0.27	1.17	−35.85	35.87
	Proj 2	−0.0928	−0.0750	−0.0542	3.0936	0.58	4.40	−36.43	36.70
	Proj 3	−0.1023	−0.0858	−0.0693	2.7030	−0.34	6.55	−33.88	34.51
	Proj 4	−0.1201	−0.0968	−0.0757	2.2874	2.64	4.26	−32.05	32.44
Set 4	Cam 1	−0.3017	0.7799	1.6935	0.6611	−0.04	4.05	−5.65	6.95
	Cam 2	−0.2589	0.5301	1.4734	1.3122	0.07	6.41	−7.51	9.87
	Cam 3	−0.2850	0.7032	1.6009	0.9225	−0.98	4.83	−6.80	8.40
	Cam 4	−0.2749	0.6605	1.5348	1.1391	1.27	4.94	−6.82	8.52
	Proj 1	−0.1250	−0.1137	−0.1014	2.1586	1.60	2.96	−30.59	30.77
	Proj 2	−0.1022	−0.0999	−0.0931	2.8472	−0.17	5.01	−34.28	34.64
	Proj 3	−0.1159	−0.1069	−0.0995	2.5582	0.23	7.40	−35.52	36.28
	Proj 4	−0.1256	−0.1103	−0.0945	2.3259	1.95	5.15	−27.87	28.41
Set 5	Cam 1	−0.2859	0.7166	1.5961	1.0180	0.02	3.13	−7.91	8.51
	Cam 2	−0.2921	0.7900	1.7087	0.7212	−0.01	6.38	−5.46	8.39
	Cam 3	−0.2667	0.6578	1.5837	1.1153	−1.48	4.78	−8.11	9.53
	Cam 4	−0.2574	0.5726	1.4411	1.1691	1.56	4.82	−7.42	8.99
	Proj 1	−0.1387	−0.1354	−0.1320	2.3848	0.53	2.69	−32.90	33.02
	Proj 2	−0.1251	−0.1171	−0.1083	2.6858	−0.63	5.36	−31.82	32.27
	Proj 3	−0.1194	−0.1213	−0.1197	2.9436	0.08	5.81	−29.09	29.67
	Proj 4	−0.1307	−0.1240	−0.1170	2.6724	1.19	5.32	−29.93	30.42
Set 6	Cam 1	−0.2596	0.4235	1.3192	0.7442	0.03	3.09	−6.61	7.30
	Cam 2	−0.2872	0.7790	1.6748	0.9223	0.09	7.33	−6.93	10.09
	Cam 3	−0.2019	0.0235	0.6330	1.3382	−2.40	4.85	−10.09	11.45
	Cam 4	−0.2971	0.8312	1.7066	0.7206	2.04	4.91	−5.88	7.93
	Proj 1	−0.1472	−0.1540	−0.1609	2.4067	0.95	3.23	−29.26	29.46
	Proj 2	−0.1329	−0.1382	−0.1430	2.6286	−0.87	5.66	−33.19	33.67
	Proj 3	−0.1454	−0.1475	−0.1472	2.4110	−1.05	7.83	−31.69	32.66
	Proj 4	−0.1510	−0.1546	−0.1562	2.1984	2.97	4.80	−28.67	29.22

The ground truth distortion coefficients are $k_1 = -0.2, k_2 = 0.05, k_3 = -0.01$

Table 3 Calibration results including distortion parameters and pose errors to evaluate the robustness of the framework

Exp	Device	Distortion			Extrinsic errors (deg / mm)				
		k_1	k_2	k_3	θ_{err}	t_x	t_y	t_z	t_{norm}
Baseline	Cam 1	−0.0855	−0.0345	0.9321	1.3420	0.30	4.47	−5.69	7.24
	Cam 2	−0.0848	−0.0172	0.7148	1.4170	0.42	5.57	−4.79	7.36
	Cam 3	−0.0752	−0.0189	0.8188	1.4079	−0.16	4.86	−5.75	7.53
	Cam 4	−0.0796	−0.0526	0.7766	1.6050	0.75	4.43	−5.45	7.06
	Proj 1	−0.0754	−0.0301	0.0277	2.1029	0.94	4.01	−22.94	23.30
	Proj 2	−0.0430	0.0031	0.0578	2.5680	0.04	5.05	−27.69	28.15
	Proj 3	−0.0782	−0.0309	0.0345	2.7756	−0.43	4.65	−25.95	26.37
	Proj 4	−0.1011	−0.0378	0.0486	2.0008	0.58	5.55	−20.85	21.59
Baseline + specular	Cam 1	−0.1005	−0.0691	1.3852	1.2632	0.33	4.46	−5.42	7.03
	Cam 2	−0.0323	−0.3236	1.3694	1.9015	−0.08	5.46	−6.48	8.47
	Cam 3	−0.0313	−0.2629	1.0838	2.1771	−0.36	4.87	−8.82	10.08
	Cam 4	−0.0675	−0.1400	1.1666	1.7808	0.45	5.05	−5.80	7.70
	Proj 1	−0.0870	−0.0449	0.0151	2.2134	0.25	3.58	−26.63	26.87
	Proj 2	−0.0633	−0.0192	0.0303	2.6142	0.13	5.31	−30.03	30.50
	Proj 3	−0.0808	−0.0257	0.0285	1.9375	−0.47	7.64	−31.36	32.28
	Proj 4	−0.0739	−0.0265	0.0380	2.3749	1.18	4.05	−28.64	28.95
Baseline + specular + low noise	Cam 1	−0.0843	0.0127	0.9975	1.8107	0.46	4.27	−6.39	7.70
	Cam 2	−0.0708	0.0325	0.9932	1.0851	−0.13	5.89	−5.71	8.21
	Cam 3	0.0070	0.1512	1.0588	1.8981	−0.71	5.22	−6.88	8.67
	Cam 4	−0.0419	0.0065	0.7628	2.3783	0.12	4.88	−4.77	6.82
	Proj 1	−0.0018	−0.0037	0.0279	1.8507	−6.19	−5.87	−35.60	36.61
	Proj 2	0.0141	0.0740	0.0214	20.9801	15.25	165.46	12.90	166.66
	Proj 3	0.1163	0.1988	0.2136	1.8025	30.32	28.62	−41.11	58.55
	Proj 4	0.1178	0.1623	0.1406	8.1631	−21.64	−8.57	−28.86	37.08
Baseline + specular + high noise	Cam 1	−0.1675	0.1374	1.0989	4.3895	0.47	6.29	−3.28	7.11
	Cam 2	−0.1588	0.0786	1.0446	4.4376	0.22	5.26	−2.44	5.80
	Cam 3	−0.1738	0.1812	1.1432	4.7135	0.71	5.91	−4.36	7.38
	Cam 4	−0.1937	0.3406	1.2420	4.8158	−1.14	5.06	−1.53	5.41
	Proj 1	0.2520	−0.0531	−0.3438	1.1155	−53.87	33.52	69.04	93.77
	Proj 2	−0.0645	−0.1153	−0.1631	10.1113	−0.52	137.61	−122.71	184.37
	Proj 3	−0.0202	0.0103	−0.0187	14.3574	207.50	11.56	−38.45	211.35
	Proj 4	−0.3729	−0.0813	0.1210	1.2433	−21.97	−6.89	68.26	72.04

The ground truth distortion coefficients are $k_1 = -0.2, k_2 = 0.05, k_3 = -0.01$

catastrophic outliers (e.g., a translation error exceeding 160 mm for Proj 2). This vulnerability stems from the pixel-wise phase decoding process, where photometric noise directly perturbs phase correspondences, in contrast to feature-based camera calibration, which benefits from spatial averaging and robust corner localization.

Evaluation of intrinsic parameter accuracy. Table 4 reports the estimated intrinsic parameters along with the ground-truth values ($f_{\text{gt}} \approx 1303.7$, $u_{0,\text{gt}} = 540$, $v_{0,\text{gt}} = 360$). Under the low-distortion configuration, Set 4 exhibits the highest estimation accuracy. The estimated focal length ($\bar{f}_x \approx 1301.7$) deviates by less than 0.2% from the ground truth, quantitatively confirming our earlier observation that intermediate target coverage (Set 4) provides an favorable balance between feature resolution and geometric constraints. In this regime, the solver recovers camera parameters that are quantitatively close to the ground-truth parameters.

In contrast, the high-distortion scenarios exhibit a systematic overestimation bias in focal length. Even in the widest-coverage case (Set 1), the estimated focal length ($\bar{f}_x \approx 1429.1$) is significantly larger than the ground truth (by approximately 9.6%). This behavior indicates that under severe radial distortion ($k_1 = -0.2$), the optimization implicitly increases the focal length to compensate for the barrel distortion effect, thereby prioritizing reprojection error minimization over strict physical accuracy.

Across all datasets, the estimated vertical principal point ($\bar{v}_0$) consistently lies below the image center (typically in the range 310–340, compared to the ground truth value of 360). We attribute this systematic shift to the non-uniform spatial distribution of target features along the vertical image

Table 4 Averaged intrinsic parameters (f_x, f_y, u_0, v_0) across four side cameras for each analysis configuration

Analysis	Dataset	$\bar{f}_x \pm \bar{\sigma}_{f_x}$	$\bar{f}_y \pm \bar{\sigma}_{f_y}$	$\bar{c}_x \pm \bar{\sigma}_{c_x}$	$\bar{c}_y \pm \bar{\sigma}_{c_y}$
Geometric (Low Dist.)	Set 1	1361.6 ± 24.9	1348.8 ± 24.6	537.5 ± 2.8	318.7 ± 13.6
	Set 2	1338.0 ± 20.0	1331.9 ± 19.8	538.9 ± 2.4	327.4 ± 10.5
	Set 3	1325.4 ± 20.0	1318.7 ± 19.8	539.6 ± 2.7	333.7 ± 10.1
	Set 4	1301.7 ± 18.5	1297.7 ± 18.5	539.4 ± 2.9	347.0 ± 9.0
	Set 5	1315.2 ± 20.7	1310.4 ± 20.5	539.5 ± 2.9	337.9 ± 10.3
	Set 6	1344.6 ± 27.3	1342.3 ± 27.1	538.4 ± 3.0	333.9 ± 14.4
Geometric (High Dist.)	Set 1	1429.1 ± 28.5	1410.2 ± 27.2	540.1 ± 2.2	284.8 ± 15.8
	Set 2	1397.3 ± 28.6	1381.9 ± 27.5	539.7 ± 2.3	303.0 ± 14.5
	Set 3	1384.7 ± 28.6	1373.1 ± 27.6	539.8 ± 2.4	312.6 ± 14.0
	Set 4	1359.4 ± 27.6	1348.9 ± 26.9	539.5 ± 2.6	324.8 ± 12.8
	Set 5	1361.9 ± 29.0	1351.5 ± 28.3	540.0 ± 2.6	328.2 ± 13.4
	Set 6	1362.5 ± 30.5	1356.0 ± 29.8	539.6 ± 2.7	334.3 ± 14.0
Robustness	Specular	1372.2 ± 22.4	1361.1 ± 22.1	541.9 ± 2.3	308.3 ± 12.9
	Low Noise	1362.6 ± 46.7	1354.2 ± 46.6	541.8 ± 6.5	315.9 ± 27.6
	High Noise	1321.4 ± 17.6	1375.9 ± 17.3	549.5 ± 1.8	336.6 ± 9.0

The uncertainty values denote the averaged local 1σ estimates obtained from the Hessian-based analysis

axis. When feature observations are unevenly distributed, the principal point estimation becomes weakly constrained, leading to a biased solution.

The uncertainty values in Table 4 correspond to averaged local 1σ estimates for the four side cameras, obtained from the Hessian-based observability analysis. They quantify the local sensitivity of the calibrated intrinsic parameters around the converged solution, rather than globally valid confidence intervals; further details are provided in Section Additional Analyses and Ablation Studies.

Depth ambiguity. In pinhole-based calibration, translation along the optical axis (t_z) is inherently coupled with the focal length (*f*). A small perturbation in t_z can be optically compensated by a proportional change in *f*, resulting in nearly identical reprojection errors. This phenomenon is well known in computer vision as projective depth ambiguity.

Although the proposed static calibration target contains multiple planar surfaces, the available depth variation may still be insufficient to fully decouple t_z from *f*. As a result, when intrinsic (*f*) and extrinsic (t_z) parameters are jointly optimized to satisfy the phase-to-pixel mapping, errors in t_z tend to be absorbed by corresponding adjustments in *f*.

This coupling can be understood from a simplified depth relationship, where the reconstructed depth Z is approximately proportional to $Z \propto f \cdot B/d$, with B denoting a baseline and d the measured disparity. If t_z is overestimated, the optimizer increases *f* accordingly to preserve the reprojection consistency. Consequently, when these coupled parameters are used together for 3D reconstruction, their individual errors largely cancel out, preserving the metric accuracy of the reconstructed point cloud.

Additional Analyses and Ablation Studies

Hessian-based local observability and uncertainty analysis.

To examine local identifiability around the converged calibration parameters, we evaluated the residual vector $\mathrm{r}(\boldsymbol{\theta}) \in \mathbb{R}^m$ at the final solution and computed its Jacobian

$$\mathrm{J}(\boldsymbol{\theta}) = \frac{\partial \mathrm{r}}{\partial \boldsymbol{\theta}}. \tag{28}$$

Using the Gauss–Newton approximation, the local Hessian was formed as

$$\mathrm{H}_\theta \approx \mathrm{J}^\top \mathrm{J}, \tag{29}$$

where $\boldsymbol{\theta}$ denotes the calibration parameter vector. The analysis reported here is therefore a *local* study around the converged solution rather than a global characterization of the entire optimization landscape.

Because the parameter groups have heterogeneous units and magnitudes, we additionally analyzed a scale-normalized coordinate system. Let

$$\boldsymbol{\phi} = \mathrm{S}^{-1}\boldsymbol{\theta}, \tag{30}$$

where $\mathrm{S} = \mathrm{diag}(s_i)$ is a diagonal scaling matrix. Each scale factor s_i was chosen from the magnitude of the converged parameter estimate together with group-wise lower bounds to avoid unstable normalization near zero. The corresponding normalized Hessian becomes

$$\mathrm{H}_\phi = \mathrm{S}^\top \mathrm{H}_\theta \mathrm{S}. \tag{31}$$

In the following, we report the scale-normalized results, since they reduce trivial unit-dependent effects while preserving the local coupling structure.

The overall conditioning was summarized by the global condition number

$$\kappa(\mathrm{H}_\phi) = \frac{\lambda_{\max}(\mathrm{H}_\phi)}{\lambda^{+}_{\min}(\mathrm{H}_\phi)}, \tag{32}$$

where $\lambda_{\max}$ is the largest eigenvalue and $\lambda^{+}_{\min}$ is the smallest retained non-degenerate eigenvalue. For readability, we report $\log_{10}\kappa$ rather than κ itself. We also count the number of near-null directions, i.e., eigen-directions whose eigenvalues fall below the adopted observability threshold.

Parameter coupling was quantified using the normalized Hessian coupling

$$C_{ij} = \frac{H_{ij}}{\sqrt{H_{ii}H_{jj}}}, \tag{33}$$

which measures how similarly two parameter directions affect the residuals in the local quadratic approximation. Values close to one indicate that the corresponding parameter directions are difficult to disentangle locally. Finally, local parameter uncertainty was estimated from the Hessian pseudoinverse. In normalized coordinates,

$$\mathrm{Cov}(\boldsymbol{\phi}) \approx \sigma_r^2\,\mathrm{H}^{+}_{\phi}, \tag{34}$$

and the covariance in the original parameterization was mapped back as

$$\mathrm{Cov}(\boldsymbol{\theta}) \approx \mathrm{S}\,\mathrm{Cov}(\boldsymbol{\phi})\,\mathrm{S}^{\top}, \tag{35}$$

where σ_r^2 is the residual variance estimate and $(\cdot)^{+}$ denotes the Moore–Penrose pseudoinverse. The reported ± values should therefore be interpreted as *local* Hessian-based 1σ uncertainties.

Table 5 summarizes a representative side camera (`Cam 1`) and a representative projector (`Proj 1`). Similar tendencies were observed for the remaining devices. Two observations are especially relevant. First, both representative devices remain severely ill-conditioned even after scale normalization, with $\log_{10}\kappa \approx 16.94$ for the side camera and $\log_{10}\kappa \approx 20.44$ for the projector. Second, the strongest coupling in both cases is between the intrinsic and translation groups, which is consistent with the well-known focal/depth-scale ambiguity. The representative side camera still exhibits four near-null directions after normalization, while the projector exhibits eight, indicating that the projector-side subproblem is substantially less observable under the adopted residual geometry.

Table 5 Representative scale-normalized Hessian summary for one side camera and one projector

Device	$\log_{10}\kappa$	Near-null	Strongest coupling	Magnitude
Cam 1	16.94	4	Intrinsic ↔ Translation	0.9979
Proj 1	20.44	8	Intrinsic ↔ Translation	0.9998

The strongest coupling is reported as the maximum off-diagonal entry of the normalized Hessian coupling matrix

The group-wise observability trends are also informative. For `Cam 1`, the dominant intrinsic and translation directions are above the adopted threshold, whereas the dominant rotation and distortion directions remain weak. By contrast, all major projector groups remain below threshold for `Proj 1`, implying that the projector-side parameters are only weakly constrained in the local Hessian sense. This result is compatible with the strong coupling already visible in Table 5, where the projector exhibits not only intrinsic–translation coupling but also comparably large intrinsic–rotation and translation–rotation couplings.

Representative parameter-wise local uncertainties are also summarized in Table 4. Overall, the recovered side-camera intrinsics remain associated with non-negligible local uncertainty, whereas the projector-side uncertainties are numerically smaller in the adopted residual domain. These values should be interpreted only as local Hessian-based uncertainty measures around the converged solution, not as globally valid confidence intervals. Taken together with the conditioning, eigenvalue, and coupling analyses, they indicate that the optimization landscape contains multiple weakly constrained directions, with the projector-side calibration being particularly prone to local ambiguity.

Sensitivity to parameter perturbation. Fig. 6 shows the point-to-mesh RMSE obtained after perturbing the converged calibration parameters along the smallest-eigenvalue direction of the scaled Hessian. The purpose of this experiment is to examine whether the weak Hessian directions identified by the local observability analysis necessarily lead to a practically harmful degradation in the intended PMP reconstruction task. It is important to note that, during calibration, no explicit 3D surface-shape information of the target was provided, except for planar geometric cues such as the marker size and their locations on the planes. In particular, the full 3D structure of the target surface represented by the ground-truth mesh was never used as supervision in the calibration stage. Nevertheless, the calibrated system produced reconstructions that were geometrically very close to the ground-truth mesh, and the reconstruction error remained small over a finite interval even after additional parameter perturbation.

To contextualize the scale of the perturbation, we note that the selected eigenvector primarily encodes a projector-side direction: the camera parameters remain virtually unchanged across all tested coefficients, while the

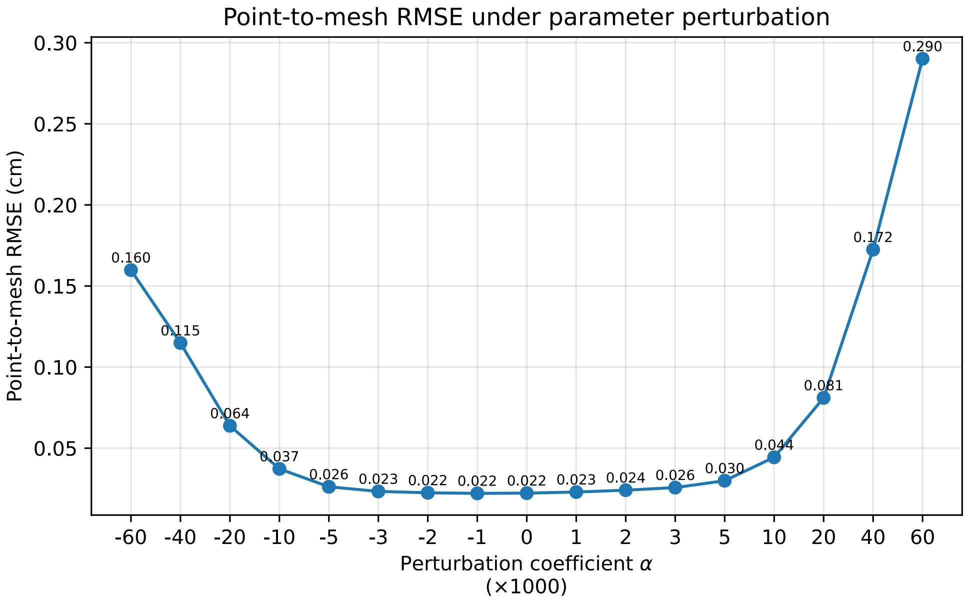


Fig. 6 Point-to-mesh RMSE under parameter perturbation along the smallest-eigenvalue direction of the scaled Hessian. The horizontal axis shows the perturbation coefficient α with a fixed scale factor of $\times 1000$. The reconstruction remains close to the ground-truth mesh over a finite interval around the converged solution, indicating that the selected weak parameter direction is also a low-sensitivity direction for the downstream reconstruction task

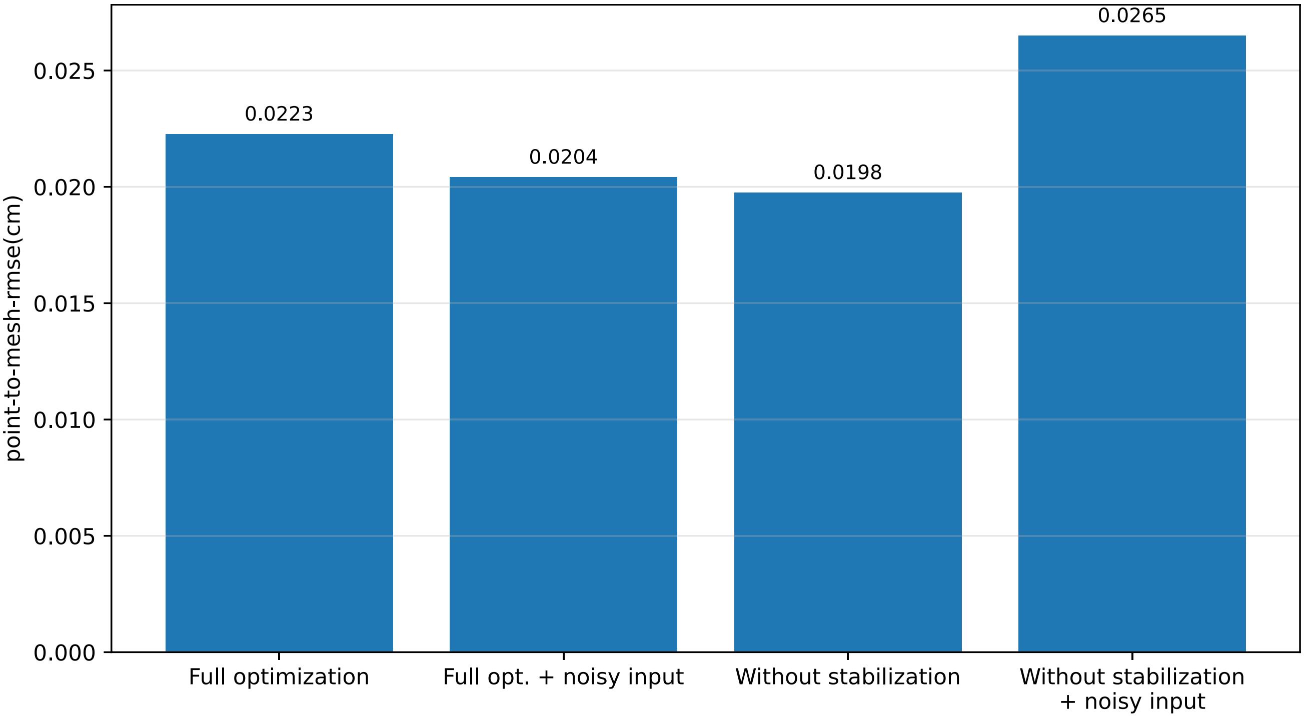


Fig. 7 Initialization ablation evaluated by point-to-mesh RMSE. The proposed full optimization pipeline shows only a small performance change between the clean and noisy settings, whereas the variant without the stabilization steps degrades more noticeably under noisy input. This indicates that the stabilization procedure mainly improves robustness to input degradation rather than merely lowering the nominal clean-case reconstruction error

projector orientation deviates from its converged value by 0.62° at $|\alpha| = 10 \times 10^3$, 1.26° at $|\alpha| = 20 \times 10^3$, and 3.89° at $|\alpha| = 60 \times 10^3$—perturbation magnitudes that are nontrivial from a practical calibration standpoint. Despite these substantial projector orientation changes, the point-to-mesh RMSE remains below 0.030 cm for $|\alpha| \leq 5 \times 10^3$ (corresponding to a projector rotation of $\approx 0.31^\circ$), confirming that this ill-conditioned projector direction is also a low-sensitivity direction for the downstream 3D reconstruction task.

This result suggests that some weak parameter-space directions correspond to low-sensitivity directions for the downstream reconstruction task: although the calibration landscape contains local ambiguity, the recovered system can still preserve accurate target geometry within a nontrivial neighborhood of the converged solution.

Ablation of the stabilization procedure. Fig. 7 compares the reconstruction accuracy obtained with and without the stabilization procedure in the initialization stage, using the point-to-mesh RMSE as the evaluation metric. Overall, the full optimization scheme exhibits only a small performance difference between the clean and noisy settings, indicating that the staged initialization provides a stable starting point even when the input is degraded.

In contrast, when the stabilization steps are removed, the reconstruction error increases more noticeably under noisy input. In particular, the gap between the clean and noisy cases is modest for the full optimization setting, whereas the corresponding gap becomes larger without stabilization. These results suggest that the staged initialization does not necessarily produce the lowest error in every clean synthetic case, but it improves robustness against input degradation and yields a more stable calibration pipeline in the presence of noise.

Sensitivity to phase noise. Fig. 8 shows the reconstruction RMSE as a function of the noise level applied directly to the decoded phase, rather than to the input images. In this experiment, the reconstruction was performed using the first side camera and first projector pair, and Gaussian perturbations were injected in the normalized phase domain before being mapped back to the original phase range. This phase-domain noise is fundamentally different from the input-image degradation considered in Fig. 7. While the full optimization pipeline in Fig. 7 exhibited only a modest performance change under noisy input, Fig. 8 shows that reconstruction error increases rapidly even for small phase-noise levels. This result indicates that the calibrated system is comparatively robust to moderate degradation in the input images, but the downstream 3D reconstruction is substantially more sensitive to direct corruption of the recovered phase itself.

Multiview Reconstruction Results

Moving beyond synthetic evaluation, we now demonstrate the practical applicability of our system through 3D reconstruction results using real-world data.

Real-world object results. Fig. 9 shows the qualitative results of our multiview 3D reconstruction pipeline. From left to right, we display: (a) photographs of real-world objects, (b) multiview phase maps from different viewpoints, (c) the merged depth maps, and (d) the final 3D reconstructions rendered from various angles.

In the coin example, fine geometric details such as the Queen's relief, less than 500 μm in height, are accurately recovered. The side profile of the coin is also clearly captured through side-view phase observations, demonstrating the precision of our multiview alignment. Additionally, fine features such as inscribed letters on a statue are faithfully reconstructed. Other examples, including complex objects like dolls, show accurate reconstruction of both frontal and side surfaces. Notably, the front-facing areas exhibit no visible artifacts, indicating the effectiveness of our multiview registration and calibration.

Multiview reconstruction accuracy. To quantitatively assess the accuracy of our multiview registration, we compare our reconstructed 3D geometry with ground truth data obtained from a commercial laser scanner (NextEngine). For maximal fidelity, we use one-pass scans from the NextEngine device. We evaluate the geometric alignment by computing the Hausdorff distance between our reconstructed point clouds and the laser-scanned ground truth.

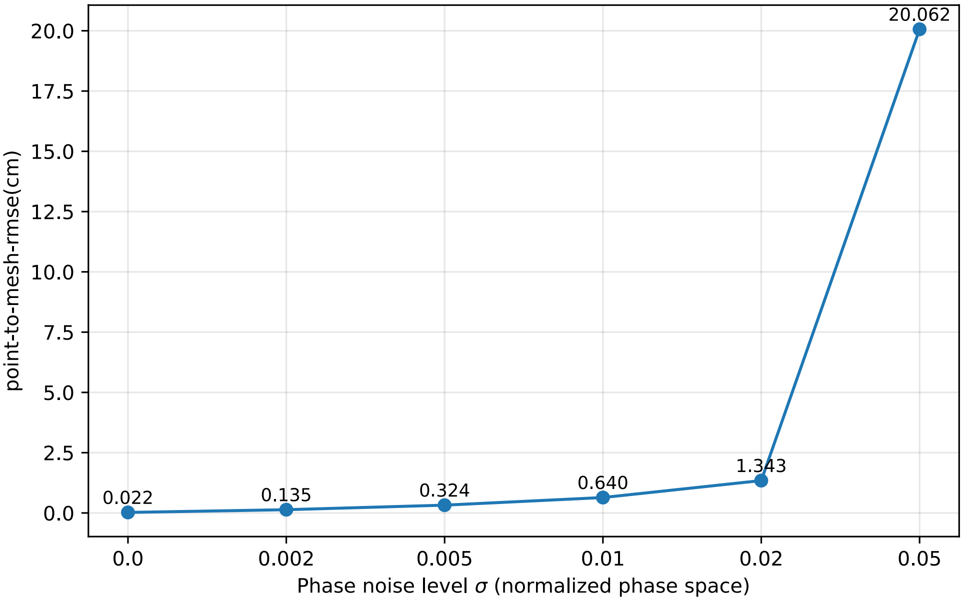
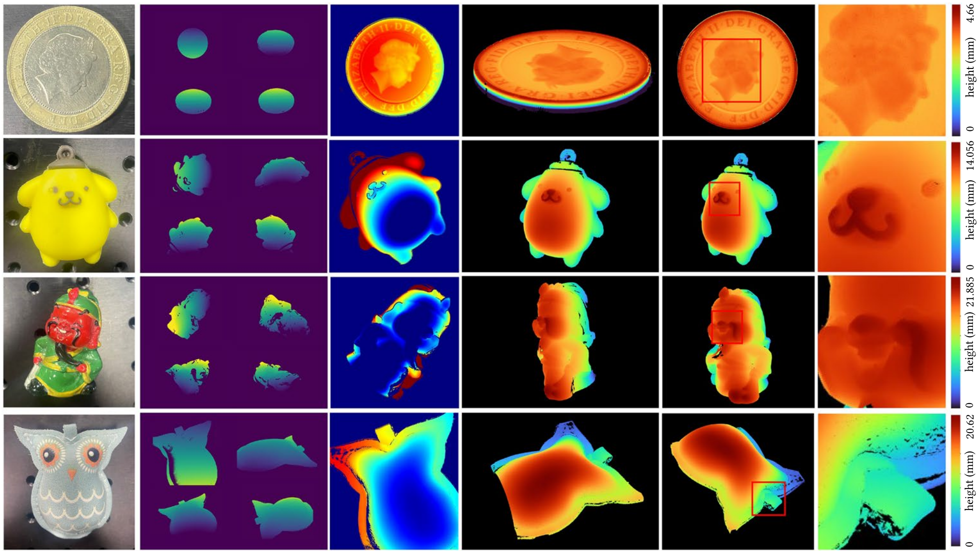
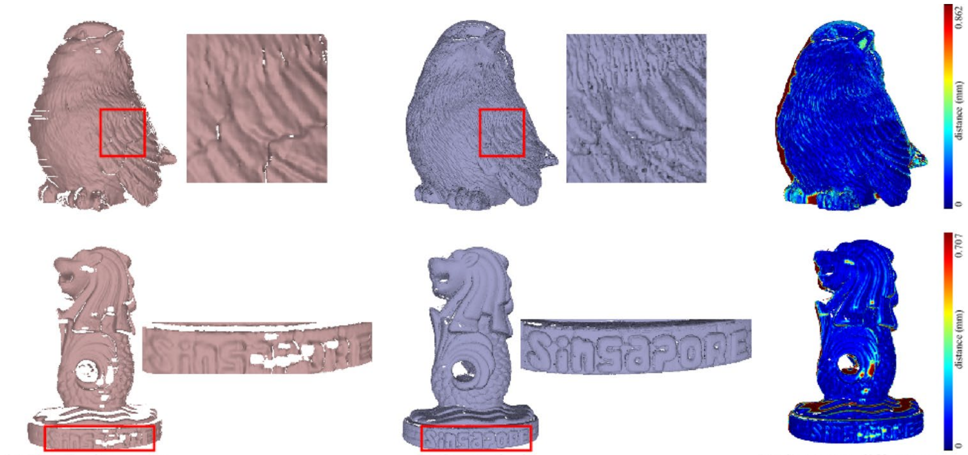


Fig. 8 Reconstruction sensitivity to phase-domain noise. Gaussian perturbations were applied directly to the decoded phase in the normalized phase space, and the resulting reconstruction RMSE was evaluated using the first side camera and first projector pair. Unlike the relatively small change observed when noise was added to the input images in Fig. 7, direct corruption of the recovered phase causes a rapid increase in reconstruction error, indicating that the downstream 3D reconstruction is considerably more sensitive to phase perturbation than to moderate input-image degradation



(a) Photograph (b) Multiview phase (c) Combined depth (d) 3D reconstruction rendered in different angles Closeup

Fig. 9 3D reconstruction results of real-world objects. **a** Photographs of objects. **b** Multiview phase maps. **c** Combined depth maps. **d** 3D reconstruction rendered in different views



(a) NextEngine scanner (b) Our method (c) Geometry differences

Fig. 10 Accuracy evaluation of our registration using the optimized calibration parameters. **a** 3D geometry captured by a commercial scanner. **b** Our multiview reconstruction result. **c** Hausdorff's distance map. Blue means zero error

Figure 10 visualizes the distance map. The results show that most of the overlapping regions exhibit very low errors, while higher errors occur only in non-overlapping or occluded regions absent in the laser scans. These results confirm that our calibration framework, based on static targets and nonlinear bundle adjustment, achieves highly accurate multiview registration, comparable to that of commercial-grade scanning systems.

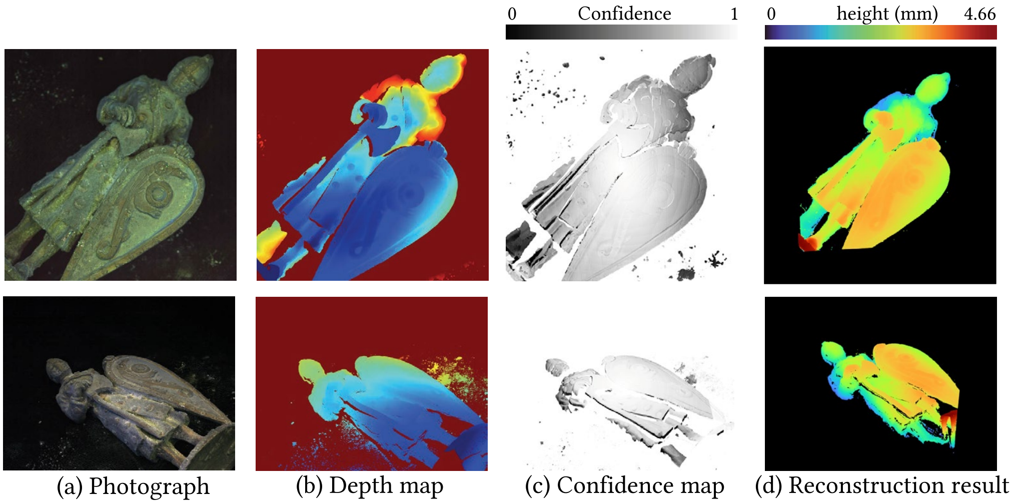
Confidence filtering. Prior to final 3D reconstruction, it is critical to estimate a confidence map based on each depth map. Our framework computes confidence values on a per-pixel basis by comparing the consistency of depth estimates across multiple views. This approach reduces the risk of incorporating incorrect 3D points caused by phase estimation errors.

By leveraging geometric relationships among views, we assess the reliability of each point by checking whether it appears consistently across all depth maps. Inconsistent points, those not observed or mismatched in other views, are considered unreliable and subsequently removed.

As shown in Fig. 11(b), the leg of the knight on the far left produces invalid phase input in one camera due to geometric occlusion. This results in inconsistent reconstruction, which our confidence-based filtering effectively removes. After filtering, only reliable points are retained and fused to produce the final 3D geometry, as illustrated in Fig. 11(d).

Camera Intrinsic Compared with Zhang's Method

To provide a partial reference for the camera-side intrinsic calibration accuracy, we compare the reprojection errors obtained using our method against those from Zhang's standard calibration technique [12]. For Zhang's method, we capture a calibration dataset comprising 15 images of a 9×8 checkerboard pattern taken from various poses. Reprojection errors are then evaluated on the same validation image set for both methods.



(a) Photograph (b) Depth map (c) Confidence map (d) Reconstruction result

Fig. 11 Our confidence filtering results. **a** Object photographs from different views. **b** Estimated depth maps from each view. **c** Confidence maps that we computed through reprojection. **d** Final 3D reconstruction result rendered from each view. Courtesy of Cho and Kim [36]

Table 6 Reprojection errors of the validation sets of images with the standard pinhole camera calibration method and our method

	Zhang		Ours	
	RMSE	MAE	RMSE	MAE
Camera 0	0.4938	0.1892	0.8543	0.3513
Camera 1	0.4028	0.1590	0.5612	0.2116
Camera 2	0.3416	0.1348	0.4058	0.1570
Camera 3	0.4349	0.1664	0.7204	0.2794

Courtesy of Cho and Kim [36]

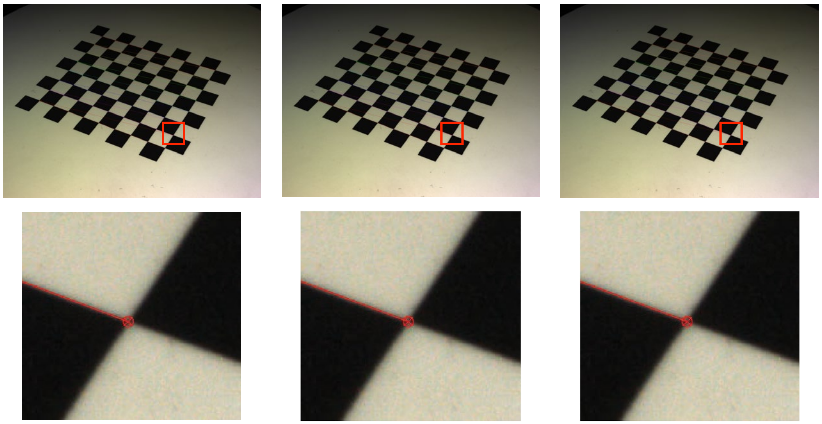
It should be noted, however, that this comparison is necessarily limited in scope. Zhang's method is a camera-only calibration technique, whereas the present work aims at calibrating the coupled camera–projector PMP system. Moreover, in our pipeline, the camera-side initialization stage is not an isolated intrinsic calibration step, but is tied to the subsequent joint optimization through the estimation of target-related geometry. As a result, Zhang's method cannot be directly inserted into the proposed pipeline as a drop-in replacement while preserving the remaining projector and target optimization stages. For this reason, the comparison with Zhang's method is restricted to reprojection error, which serves here as a partial camera-side reference rather than as a full system-level baseline.

Table 6 summarizes the reprojection error statistics, and Fig. 12 shows example images with reprojected corners overlaid using both Zhang's and our calibration results. Although this comparison does not constitute a complete PMP-system comparison, it shows that the proposed method achieves subpixel-level camera-side reprojection accuracy on the evaluated data, while operating within the practical constraints of the intended scanning setup.

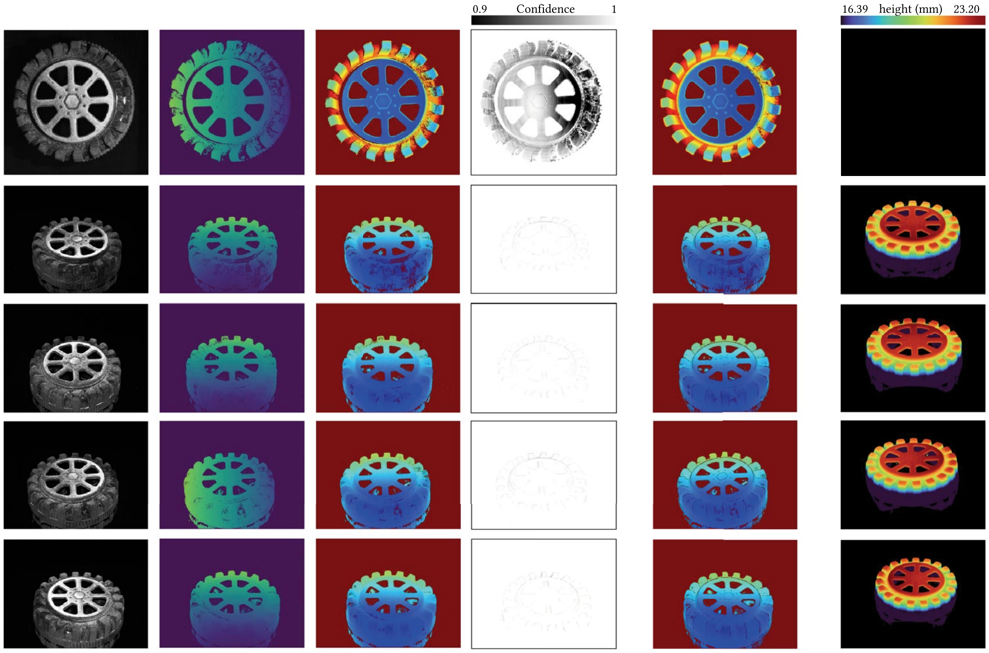
View-pairwise Consistency Results

In this section, we demonstrate the geometric consistency of depth maps across different views prior to final reconstruction. The overall procedure and intermediate results are visualized in Fig. 13.

Fig. 12 The images (top) and their close-ups (bottom). **a** 9×8 checkerboard image marked with detected corners. **b** Corners reprojected by Zhang's method. **c** Corners reprojected by our method. Courtesy of Cho and Kim [36]



(a) Detected (b) Reprojected (Zhang) (b) Reprojected (ours)



(a) Photograph (b) Phase map (c) Depth map (d) Confidence map (e) Combined depth map (f) Reconstruction result

Fig. 13 3D reconstruction results of real-world objects with a top-view (the first row) and 4 side-views (the second to the fifth row). **a** Photographs of objects. **b** Multiview phase maps. **c** Multiview depth maps. **d** Multiview confidence maps. **e** Multiview combined depth maps. **f** 3D reconstruction rendered in different views. Courtesy of Cho and Kim [36]

Table 7 Statistics including mean and standard deviation (SD) of Hausdorff's distance between point clouds generated by four side-views and top view for various real-world objects

		View 0	View 1	View 2	View 3
Owl	Mean	0.0817	0.1149	0.126	0.1197
	SD	0.1077	0.1536	0.1626	0.1489
Pouch	Mean	0.0393	0.0427	0.0493	0.0367
	SD	0.0534	0.0689	0.0772	0.0527
Cat	Mean	0.1168	0.0851	0.1024	0.0945
	SD	0.1504	0.1129	0.1334	0.1304
Tire	Mean	0.1027	0.0795	0.0961	0.0953
	SD	0.1496	0.1228	0.1432	0.1429
Girl statue	Mean	0.0605	0.0565	0.0488	0.0687
	SD	0.0826	0.0853	0.0663	0.1028
Canon holder	Mean	0.1091	0.0691	0.1382	0.1043
	SD	0.1566	0.1205	0.1643	0.1423
Singapore lion statue	Mean	0.0945	0.0947	0.0906	0.0838
	SD	0.1479	0.1416	0.1404	0.1400
Yellow doll	Mean	0.1188	0.1058	0.1825	0.0986
	SD	0.1669	0.1515	0.2125	0.1242

Courtesy of Cho and Kim [36]

A key indicator of successful calibration is that the point clouds generated from individual views align well with each other, without requiring any additional post-processing such as rigid-body registration. This implies that the extrinsic parameters across views are accurately estimated.

All results presented here are computed using point clouds sampled before the final reconstruction stage, using voxel downsampling with a diameter of 0.025 mm.

Hausdorff distance. We first evaluate the geometric consistency across views by computing the Hausdorff distance between point clouds generated from different cameras. Since the top-view camera primarily observes the entire object, we use its point cloud as a reference and compare it against those from the side-view cameras.

Because not all views share complete overlap, we restrict the evaluation to points with Hausdorff distances below 0.5 mm, thereby focusing on the overlapping regions. We compute the mean and standard deviation of these filtered distances to quantitatively assess alignment quality.

Table 7 summarizes the statistical results of the Hausdorff distances across different objects. As shown, the average distances between views lie in the range of 30 μm to 200 μm, indicating strong consistency among the calibrated views.

Figure 14 visualizes the point clouds colored by Hausdorff distances. The results confirm that discrepancies within overlapping regions are extremely small, further validating the accuracy of our multiview calibration.

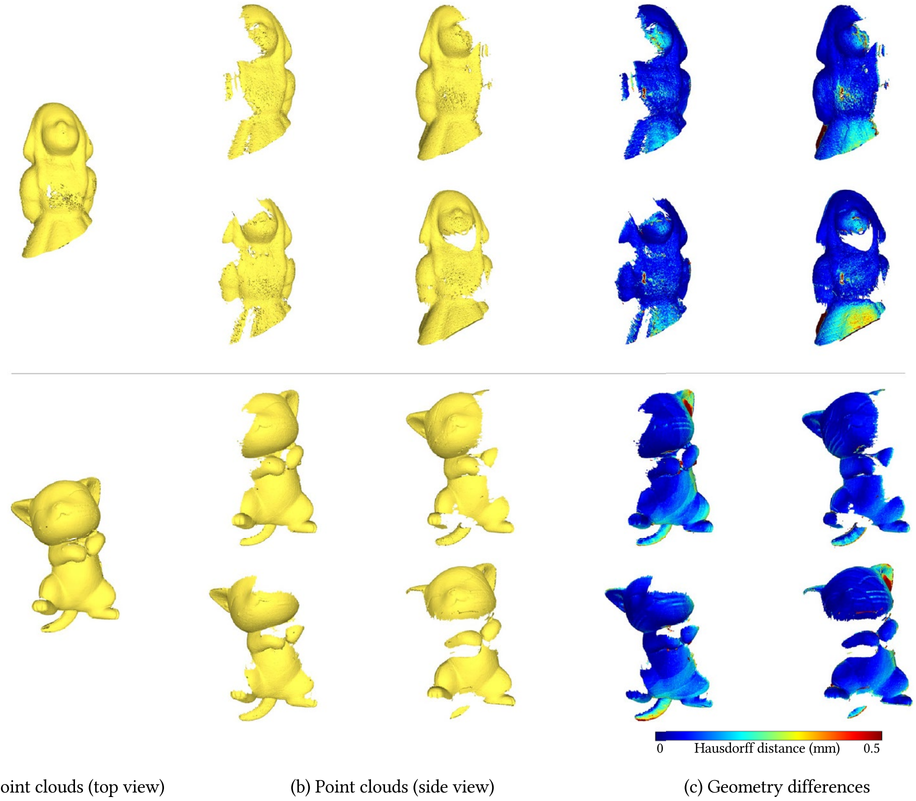


Fig. 14 Point clouds are generated by depth maps of top and side views and their geometry differences. **a** Point clouds generated by top view. **b** Point clouds generated by side view. **c** Geometry differences measured as Hausdorff's distance. Courtesy of Cho and Kim [36]

Discussion

Local conditioning and coupling. The additional analyses in Section Additional Analyses and Ablation Studies show that the proposed calibration problem is locally ill-conditioned and contains multiple weakly constrained directions. Even after scale normalization, large values of $\log_{10}\kappa$ and several near-null directions remain, indicating that the difficulty is not merely caused by heterogeneous parameter units, but is intrinsic to the underlying calibration geometry itself. At the same time, the observed parameter coupling should not be interpreted as a pathology unique to the proposed method. In fact, coupling between intrinsic and extrinsic parameters is a well-known structural issue in camera calibration more broadly, since different parameter changes can induce similar variations in the image-domain residuals. Conventional planar calibration methods mitigate this ambiguity by imposing strong geometric constraints through a known target, for example via the known marker or checker size and the known planar layout of feature points.

In our setting, the marker size and the marker arrangement on each plane play a similar anchoring role and therefore help reduce the intrinsic–translation and related couplings. However, unlike a calibration setup in which the full target geometry is completely prescribed, our formulation does not explicitly supervise the full 3D global shape of the target, such as the complete surface-orientation structure across all planes. As a result, although the known planar marker information alleviates the ambiguity, some parameter-level coupling and weak directions can still remain in the joint camera–projector–surface optimization.

Task-space robustness under parameter perturbation. At the same time, the perturbation experiments show that weak parameter-space directions do not necessarily imply equally severe degradation in the intended downstream task. When the converged solution is perturbed along the smallest-eigenvalue direction of the scaled Hessian, the reconstruction error remains comparatively small over a finite interval around the optimum, even with a large perturbation scale factor. This result is particularly meaningful because, during calibration, the target was not supervised using its full 3D surface shape; only planar geometric information such as marker size and marker locations on each plane was used. Nevertheless, the resulting reconstruction remains geometrically close to the ground-truth mesh, and moderate perturbations of the recovered parameters still preserve a low reconstruction error. In this sense, some of the weak parameter directions revealed by the Hessian analysis should be interpreted as low-sensitivity directions for the system-level reconstruction task rather than as immediately harmful failure modes.

Stabilization and initialization. Global basin-of-attraction guarantees are not available for bundle adjustment in general, as the underlying optimization problem is inherently non-convex and convergence to the global minimum cannot be ensured regardless of the optimizer employed [22, 42]. Consequently, the quality of initial parameter estimates plays a central role in determining whether the optimization converges to a practically useful solution. Rather than addressing this through exhaustive random restarts, our staged initialization progressively releases parameters — surface poses first, then intrinsics and extrinsics, and finally distortion coefficients — constraining the search space at each step and reducing the risk of converging to degenerate solutions under unfavorable starting conditions.

The initialization ablation further supports this interpretation. The staged stabilization procedure does not necessarily guarantee the smallest clean-case reconstruction error in every synthetic setting, but it consistently improves robustness when the input is degraded. In particular, the full optimization pipeline exhibits only a small performance difference between clean and noisy input, whereas the variant without the stabilization steps shows a more noticeable degradation under noisy observations. This suggests that the practical benefit of the staged initialization lies less in nominal error minimization and more in stabilizing the optimization against imperfect observations and unfavorable starting conditions.

Projector model as a practical geometric approximation. In the present work, we adopt the projector model in this practical geometric sense, i.e., as an inverse-camera-style approximation that enables a tractable joint formulation for PMP system calibration. Our intent is not to claim a complete physical model of projector image formation or phase propagation. Rather, the model is used to convert phase-derived constraints into a system-level geometric calibration problem that is directly tied to the intended 3D reconstruction task. We acknowledge that such a formulation does not explicitly account for all radiometric nonlinearities, phase-decoding bias, or the full propagation of phase uncertainty. Accordingly, the proposed model should be interpreted as a calibration-oriented geometric approximation whose validity is assessed primarily through geometric consistency and reconstruction behavior of the coupled PMP system, rather than as a fully exhaustive physical description of the projector.

Geometric objective for joint calibration. A further point that merits clarification is the choice of the final optimization objective. Although the projector is modeled as an inverse pinhole device, this does not imply that a reprojection-error formulation is necessarily the most appropriate objective for the present PMP system calibration problem. In a static-target, single-scene setting with jointly estimated camera, projector, and target-related quantities, image-domain reprojection consistency alone does not guarantee that the recovered 3D geometry will be consistent with the actual target shape. Different parameter combinations may compensate each other in the image domain while still producing geometrically inconsistent reconstructions. This concern is also consistent with the broader calibration literature, where diverse target orientations are typically required to suppress ambiguity even in conventional camera calibration. In our setting, where stronger coupling remains in the joint problem, there is no a priori guarantee that minimizing reprojection residual alone would recover a physically meaningful target geometry. For this reason, we formulate the optimization using a geometry-driven point-to-plane objective, which more directly penalizes inconsistency between the reconstructed 3D points and the target planes. Importantly, although the full ground-truth mesh is never provided as supervision during calibration, the resulting reconstructions remain close to the ground-truth geometry in the synthetic experiments. This supports the view that the point-to-plane objective is better viewed as a practical geometric choice for reprojection loss, but a practical choice for enforcing reconstruction-space consistency in the intended PMP calibration task.

Sensitivity to phase noise. On the other hand, the phase-noise sensitivity analysis reveals a more direct vulnerability of the calibrated system. In contrast to the relatively modest performance change observed under noisy input in the initialization ablation, the reconstruction error increases rapidly when Gaussian perturbations are applied directly to the decoded phase. This distinction is important: the input-image noise considered in the stabilization study and the phase-domain noise considered in the phase-sensitivity study are not equivalent. The latter corrupts the quantity that is used directly for phase-to-3D reconstruction, and therefore propagates much more strongly to the recovered geometry. Consequently, the present results suggest that, for the proposed PMP framework, direct corruption of the recovered phase is a more critical source of downstream reconstruction error than moderate image-level degradation or certain weak parameter perturbations.

Implications for the proposed calibration framework. Taken together, these observations help refine the interpretation of the proposed calibration method. The method should not be understood as producing fully disentangled and uniquely identifiable device parameters in the strictest metrological sense. Rather, its practical objective is to calibrate the PMP system as a whole, i.e., to obtain a camera–projector–target geometry that supports accurate 3D reconstruction. From this perspective, the Hessian analysis confirms the existence of local ambiguity and strong coupling, while the perturbation and ablation results show that part of this ambiguity is task-space benign within a nontrivial neighborhood of the converged solution. At the same time, the phase-noise experiment highlights an important limitation: reconstruction quality remains strongly dependent on the fidelity of the recovered phase. Accordingly, the proposed framework should be viewed as a system-level calibration method that is effective for reconstruction consistency, but whose local parameter identifiability remains limited and whose performance is still sensitive to phase-domain corruption.

Conclusion

In this paper, we presented a practical calibration framework that enables joint estimation of intrinsic and extrinsic parameters for multiview phase-measuring profilometry systems from a single static geometric setup. The key strength of the proposed approach lies in its operational practicality: it eliminates the need for repeated target manipulation, making it particularly suitable for spatially constrained environments such as compact multiview scanning systems.

Despite relying on only a single capture, the proposed method achieves sufficient calibration accuracy to support high-quality 3D reconstruction, providing a viable alternative to conventional multi-pose calibration procedures. Extensive synthetic and robustness evaluations further demonstrate that the framework maintains global geometric consistency under realistic imaging conditions, including surface specularity and moderate levels of sensor noise.

We also acknowledge inherent limitations of the single-shot configuration. The reduced diversity of geometric constraints can lead to an ill-conditioned optimization landscape, resulting in parameter coupling and ambiguities, most notably the focal length–depth ambiguity and systematic bias in the vertical principal point (v_0). As a result, while the reconstructed geometry remains metrically accurate, individual estimated parameters may deviate from their physical ground-truth values.

Future work will focus on incorporating additional geometric or statistical priors to further constrain the optimization, with the goal of resolving these ambiguities while preserving the simplicity and convenience of single-shot calibration.

Author Contributions All authors contributed equally to this work.

Funding Open Access funding enabled and organized by KAIST. Min H. Kim acknowledges the Samsung Research Funding & Incubation Center (SRFC-IT2402-02), the Korea NRF grant (RS-2024-00357548), the MSIT/IITP of Korea (RS-2022-00155620, RS-2024-00398830, RS-2024-00436680, and 2017-0-00072), and Microsoft Research Asia.

Data Availability Not applicable.

Declarations

Conflict of Interest The authors have no conflict of interest to declare that are relevant to the content of this article.

Research Involving Human and/or Animals Not applicable.

Informed Consent Not applicable.

References

1. Takeda M, Mutoh K. Fourier transform profilometry for the automatic measurement of 3-d object shapes. Appl Opt. 1983;22(24):3977–82.
2. Zhang Z, Towers CE, Towers DP. Uneven fringe projection for efficient calibration in high-resolution 3d shape metrology. Appl Opt. 2007;46(24):6113–9.
3. Sansoni G, Carocci M, Rodella R. Calibration and performance evaluation of a 3-d imaging sensor based on the projection of structured light. IEEE Trans Instrum Meas. 2000;49(3):628–36.
4. Xiao Y, Cao Y, Wu Y. Improved algorithm for phase-to-height mapping in phase measuring profilometry. Appl Opt. 2012;51(8):1149–55.
5. Li J-L, Su H-J, Su X-Y. Two-frequency grating used in phase-measuring profilometry. Appl Opt. 1997;36(1):277–80.
6. Li Y, Su X, Wu Q. Accurate phase-height mapping algorithm for pmp. J Mod Opt. 2006;53(14):1955–64.
7. Li W, Su X, Liu Z. Large-scale three-dimensional object measurement: a practical coordinate mapping and image data-patching method. Appl Opt. 2001;40(20):3326–33.
8. Guo W, Wu Z, Xu R, Zhang Q, Fujigaki M. A fast reconstruction method for three-dimensional shape measurement using dual-frequency grating projection and phase-to-height lookup table. Optics & Laser Technology. 2019;112:269–77.
9. Ma Q, Cao Y, Chen C, Wan Y, Fu G, Wang Y. Intrinsic feature revelation of phase-to-height mapping in phase measuring profilometry. Optics & Laser Technology. 2018;108:46–52.
10. Chen L, Huayang L, Xu Z, Huan Z. Coding line structured light based on a line-scan camera and its calibration. Opt Express. 2020;28(17):24799–812.

11. Cai Z, Liu X, Li A, Tang Q, Peng X, Gao BZ. Phase-3d mapping method developed from back-projection stereovision model for fringe projection profilometry. Opt Express. 2017;25(2):1262–77.
12. Zhang Z. A flexible new technique for camera calibration. IEEE Trans Pattern Anal Mach Intell. 2000;22(11):1330–4.
13. Juarez-Salazar R, Diaz-Ramirez VH. Flexible camera-projector calibration using superposed color checkerboards. Opt Lasers Eng. 2019;120:59–65.
14. Jiang C, Lim B, Zhang S. Three-dimensional shape measurement using a structured light system with dual projectors. Appl Opt. 2018;57(14):3983–90.
15. Juarez-Salazar R, Giron A, Zheng J, Diaz-Ramirez VH. Key concepts for phase-to-coordinate conversion in fringe projection systems. Appl Opt. 2019;58(18):4828–34.
16. Li W, Li H, Zhang H. Light plane calibration and accuracy analysis for multi-line structured light vision measurement system. Optik. 2020;207:163882.
17. Feng Z, Man D, Song Z. A pattern and calibration method for single-pattern structured light system. IEEE Trans Instrum Meas. 2019;69(6):3037–48.
18. Lanman D, Taubin G. Build your own 3d scanner: 3d photography for beginners. In: ACM Siggraph 2009 Courses. 2009. p. 1–94.

19" GrobidStyleApplied="Changed. Moreno D, Taubin G. Simple, accurate, and robust projector-camera calibration. In: 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission, 2012 pp. 464–471 . IEEE

20. Ibrahim MT, Gopi M, Majumder A. Self-calibrating dynamic projection mapping system for dynamic, deformable surfaces with jitter correction and occlusion handling. IEEE Int Symp on Mixed and Augmented Reality. 2023;293–302.
21. Munoz-Salinas R. Aruco: a minimal library for augmented reality applications based on opencv. Universidad de Córdoba; 2012.

22" GrobidStyleApplied="Changed. Triggs B, McLauchlan PF, Hartley RI, Fitzgibbon AW. Bundle adjustment – a modern synthesis. In: International Workshop on Vision Algorithms, 1999 pp. 298–372 . Springer

23. Feng S, Zuo C, Zhang L, Tao T, Hu Y, Yin W, et al. Calibration of fringe projection profilometry: A comparative review. Opt Lasers Eng. 2021;143:106622.
24. Zhang Z, Ma H, Zhang S, Guo T, Towers CE, Towers DP. Simple calibration of a phase-based 3d imaging system based on uneven fringe projection. Opt Lett. 2011;36(5):627–9.
25. Guo H, He H, Yu Y, Chen M. Least-squares calibration method for fringe projection profilometry. Opt Eng. 2005;44(3):033603.
26. Du H, Wang Z. Three-dimensional shape measurement with an arbitrarily arranged fringe projection profilometry system. Opt Lett. 2007;32(16):2438–40.
27. Wang Z, Nguyen DA, Barnes JC. Some practical considerations in fringe projection profilometry. Opt Lasers Eng. 2010;48(2):218–25.
28. Lee YB, Kim MH. Integrated calibration of multiview phase-measuring profilometry. Opt Lasers Eng. 2017;98:118–22.
29. Léandry I, Brèque C, Valle V. Calibration of a structured-light projection system: development to large dimension objects. Opt Lasers Eng. 2012;50(3):373–9.
30. Fu Y, Wang Y, Wang W, Wu J. Least-squares calibration method for fringe projection profilometry with some practical considerations. Optik. 2013;124(19):4041–5.
31. Huang L, Chua PS, Asundi A. Least-squares calibration method for fringe projection profilometry considering camera lens distortion. Appl Opt. 2010;49(9):1539–48.
32. Tehrani MA, Gopi M, Majumder A. Automated geometric registration for multi-projector displays on arbitrary 3d shapes using uncalibrated devices. IEEE Trans Visual Comput Graphics. 2019;27(4):2265–79.
33. Furuakwa R, Inose K, Kawasaki H. Multi-view reconstruction for projector camera systems based on bundle adjustment. In: Proc. IEEE CVPR Workshops. IEEE; 2009. p. 69–76.
34. Garrido-Jurado S, Munoz-Salinas R, Madrid-Cuevas FJ, Marin-Jiménez MJ. Simultaneous reconstruction and calibration for multi-view structured light scanning. J Visual Communication and Image Representation. 2016;39:120–31.
35. Li C, Monno Y, Hidaka H, Okutomi M. Pro-cam ssfm: Projector-camera system for structure and spectral reflectance from motion. In: Proc. IEEE ICCV. 2019. p. 2414–23.
36. Cho H, Kim MH. Joint calibration of cameras and projectors for multiview phase measuring profilometry. In: Proc. Int. Conf. Computer Vision, Theory and Applications (VISAPP 2025). Porto, Portugal; 2025.
37. Teed Z, Deng J. Tangent space backpropagation for 3d transformation groups. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. p. 10338–47.
38. Ricolfe-Viala C, Sanchez-Salmeron A-J. Lens distortion models evaluation. Appl Opt. 2010;49(30):5914–28.
39. Chen Z, Liao H, Zhang X. Telecentric stereo micro-vision system: Calibration method and experiments. Opt Lasers Eng. 2014;57:82–92.
40. Teed Z, Deng J. Droid-slam: Deep visual slam for monocular, stereo, and rgb-d cameras. Adv Neural Inf Process Syst. 2021;34:16558–69.

41" GrobidStyleApplied="Changed. Jakob W, Speierer S, Roussel N, Nimier-David M, Vicini D, Zeltner T, Nicolet B, Crespo M, Leroy V, Zhang Z. Mitsuba 3 Renderer. https://mitsuba-renderer.org

42. Zach C. Robust bundle adjustment revisited. In: European Conference on Computer Vision. Springer; 2014. p. 772–87.