

# Geometry-Aware Super-Resolution Fusion Calibration for Binocular Structured Light 3D Reconstruction

Yijie Shen

[yijie.shen@ntu.edu.sg](mailto:yijie.shen@ntu.edu.sg)

Nanyang Technological University <https://orcid.org/0000-0002-6700-9902>

HONGYAN CAO

Northwestern Polytechnic University

Dayong Qiao

Mengya Han

Wangke Yu

Benquan Wang

---

## Article

## Keywords:

**Posted Date:** January 16th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-8481671/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** There is **NO** Competing Interest.

---

**Version of Record:** A version of this preprint was published at Communications Physics on May 21st, 2026. See the published version at <https://doi.org/10.1038/s42005-026-02666-9>.

# Geometry-Aware Super-Resolution Fusion Calibration for Binocular Structured Light 3D Reconstruction

Hongyan Cao,<sup>1,2,3</sup> Dayong Qiao,<sup>1,2,\*</sup> Mengya Han,<sup>4</sup> Wangke Yu,<sup>3</sup> Benquan Wang,<sup>3,†</sup> and Yijie Shen<sup>3,5,‡</sup>

<sup>1</sup>Ministry of Education Key Laboratory of Micro/Nano Systems for Aerospace,  
Key Laboratory of Micro- and Nano-Electro-Mechanical Systems of Shaanxi Province,  
School of Mechanical Engineering, Northwestern Polytechnical University, 127 Youyi West Road, Xi'an, 710072, China.

<sup>2</sup>Ningbo Institute, Northwestern Polytechnical University, Ningbo, Zhejiang, 315103, China.

<sup>3</sup>Centre for Disruptive Photonic Technologies, School of Physical and Mathematical Sciences,  
Nanyang Technological University, Singapore 637371, Republic of Singapore

<sup>4</sup>National Engineering Research Center for Multimedia Software, School of Computer Science,  
Wuhan University, and Hubei LuoJia Laboratory, Wuhan, 430071, China.

<sup>5</sup>School of Electrical and Electronic Engineering,  
Nanyang Technological University, Singapore 639798, Singapore

(Dated: December 30, 2025)

High-precision binocular camera calibration serves as the cornerstone for 3D reconstruction. Despite existing advances in calibration methods, the inherent sub-pixel accuracy of feature points remains compromised by various degradations, such as noise, blur, and distortion, which fundamentally constrain the accuracy and practical performance in uncontrolled real-world scenarios. This paper proposes a novel geometric-fusion calibration framework that introduces saddle point preservation and unbiased cell centroids based on the binocular epipolar line constraints to maintain the flexibility and robustness of global feature points for the first time, substantially advancing structured light 3D reconstruction accuracy. Extensive experiments demonstrate that the proposed method enhances the robustness of corner coordinates against noise, blur, and distortion and reduces reprojection error by 17%. More importantly, we validate significant improvements in point cloud accuracy for real-world 3D reconstruction tasks, establishing the practical value of our approach for precision-critical applications like medical endoscopy, AR/VR systems, and embodied AI platforms.

## I. INTRODUCTION

Binocular camera calibration<sup>1–3</sup> is foundational to computer vision and 3D sensing, as it determines the intrinsic parameters of each camera and the stereo geometry that links the two views. Accurate calibration is therefore central to a broad range of downstream tasks<sup>4–9</sup>, and is particularly critical for structured-light 3D reconstruction (SL3D)<sup>10–14</sup>. In SL3D, known patterns are projected onto a scene and the deformed observations are captured by a calibrated stereo pair<sup>15–18</sup>; the recovered surface depends not only on the coding and decoding strategy<sup>19,20</sup>, but also on the fidelity of the underlying stereo model. Calibration errors propagate through phase decoding and stereo correspondence, ultimately setting the accuracy ceiling of the reconstructed 3D geometry<sup>21–23</sup>.

Despite sustained progress, binocular calibration remains brittle in practical deployments. A persistent limitation is the reliability of sub-pixel feature localization<sup>24,25</sup>. Image noise, illumination variations, optical blur, and imperfections or deformation of the calibration target can introduce systematic biases in detected features, which are amplified during geometric estimation. Moreover, binocular systems require accurate recovery

of the relative pose between cameras<sup>26,27</sup>, and calibration accuracy often degrades toward the periphery of the field of view<sup>28</sup>, yielding spatially non-uniform errors across the measurement volume. Together, these effects undermine both the stability of estimated parameters and the transferability of calibration to real SL3D acquisitions.

Two lines of recent advances motivate a more robust formulation. The first leverages geometric structure: planar calibration targets such as checkerboards embed strong priors—including orthogonality, parallelism, and equidistance—that can regularize feature extraction and suppress noise-induced variability<sup>29,30</sup>. The second arises from learning-based image restoration: super-resolution methods can recover sharper edges and richer local detail from degraded observations, improving the signal quality available for subsequent geometric processing<sup>31–33</sup>. However, existing approaches typically exploit these cues in isolation, and most calibration pipelines still rely predominantly on corner measurements, leaving them vulnerable when corners are blurred, distorted, or locally ambiguous<sup>34–38</sup>.

Here we present a geometry-driven, two-stage binocular calibration framework, GeoRobustCalib, which models each checkerboard cell as a geometric primitive and introduces an unbiased cell centroid as a complementary constraint. As shown in Fig. 1, we enhance feature observability using geometry-aware super-resolution restricted to the checkerboard region, then localize cor-

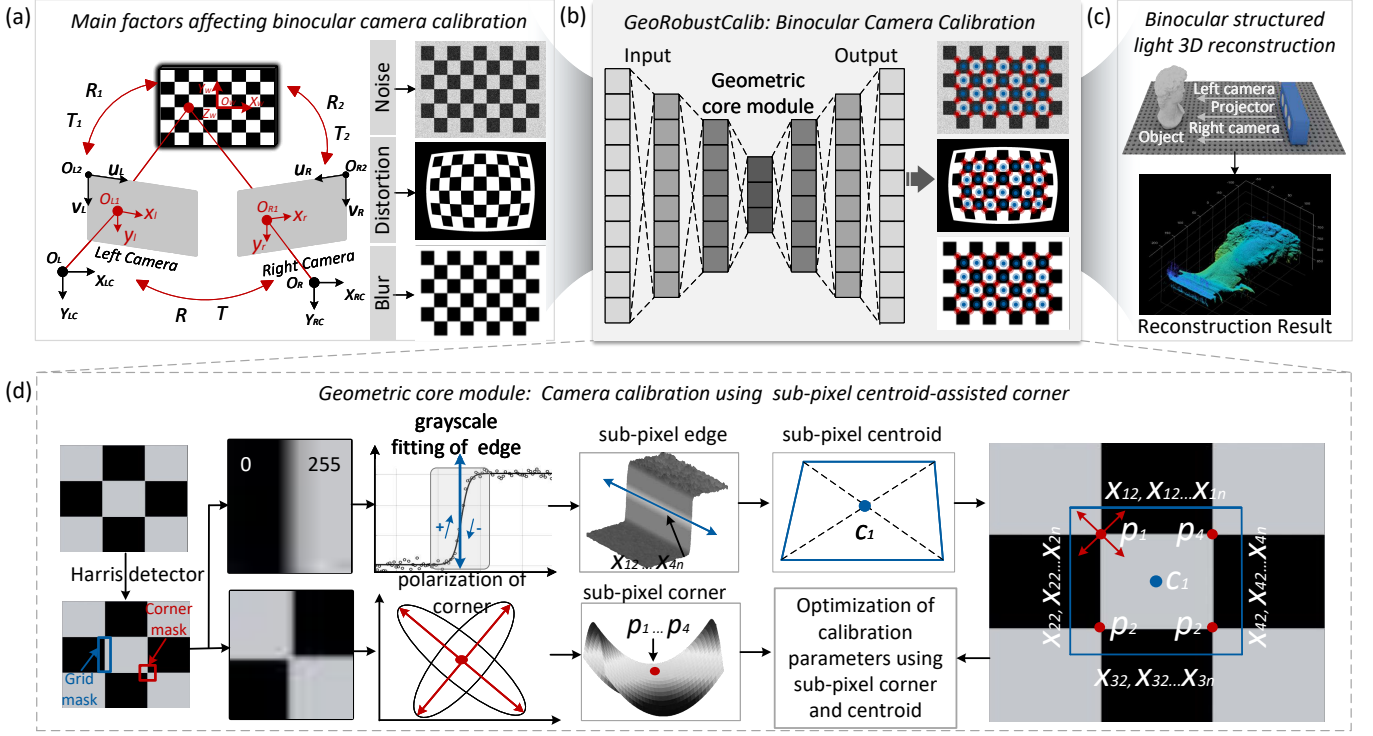


FIG. 1: A new binocular camera calibration framework based on geometric prior information. (a) The main purpose is to improve the calibration robustness under noise, blur, and distortion. (b) The main method is to improve the calibration accuracy by calculating the sub-pixel corners and centroids of the calibration plate. (d) Geometric prior calibration framework process. (c) Using the calculated binocular calibration parameters to achieve higher-precision structured light 3D reconstruction.

ners at sub-pixel precision via saddle-point fitting while estimating cell centroids that remain stable under common degradations. Building on these features, we adopt a two-stage optimization in which corners provide strong constraints and centroids act as soft geometric priors with lightweight epipolar regularization, improving global consistency without sacrificing flexibility. Experiments on both synthetic data and real structured-light systems demonstrate reduced reprojection error (up to 17%) and consistently improved point-cloud accuracy, indicating a practical route to robust, precision-grade calibration for SL3D and general binocular 3D sensing.

## II. RESULTS

### Experimental setup

All experiments are performed on a binocular SL3D measurement platform consisting of two synchronized grayscale cameras and a MEMS-based projection module, as shown in Fig. E1. The cameras share identical imaging settings (native 4 : 5 aspect ratio, fixed aperture F/2.5, and a 6 mm fixed-focus lens with the focus set to 2.6 m) and operate in global-shutter mode to mitigate motion-induced artifacts during pattern projection and capture. The projector uses a single-axis MEMS micro-vibration mirror to display coded stripe patterns

for SL3D reconstruction.

Our geometry-aware checkerboard super-resolution and centroid-assisted calibration pipeline is implemented in Python, using OpenCV for image processing and calibration, and PyTorch for the super-resolution backbone. Training is conducted on a workstation with four NVIDIA GPUs (16 GB memory each) under CUDA 11.8. Matlab is used for auxiliary visualization and SL3D reconstruction analysis.

To evaluate GeoRobustCalib comprehensively, we construct a hybrid dataset with roughly equal portions of synthetic and real calibration images, as shown in Fig. E2. For real data, a physical checkerboard is mounted on a six-axis robotic arm or adjusted manually; for each configuration, about 30 images are captured under diverse poses and viewpoints to emulate practical variations in geometry and illumination. For synthetic data, a high-fidelity checkerboard is modeled in Blender and rendered through a simulated binocular imaging chain with controlled noise, blur, and distortion, providing ground truth for corners, centroids, and camera parameters.

We report results at three levels: (i) feature-level localization accuracy for corners and cell centroids; (ii) calibration-level metrics including reprojection RMS error, centroid reprojection bias and vertical disparity af-

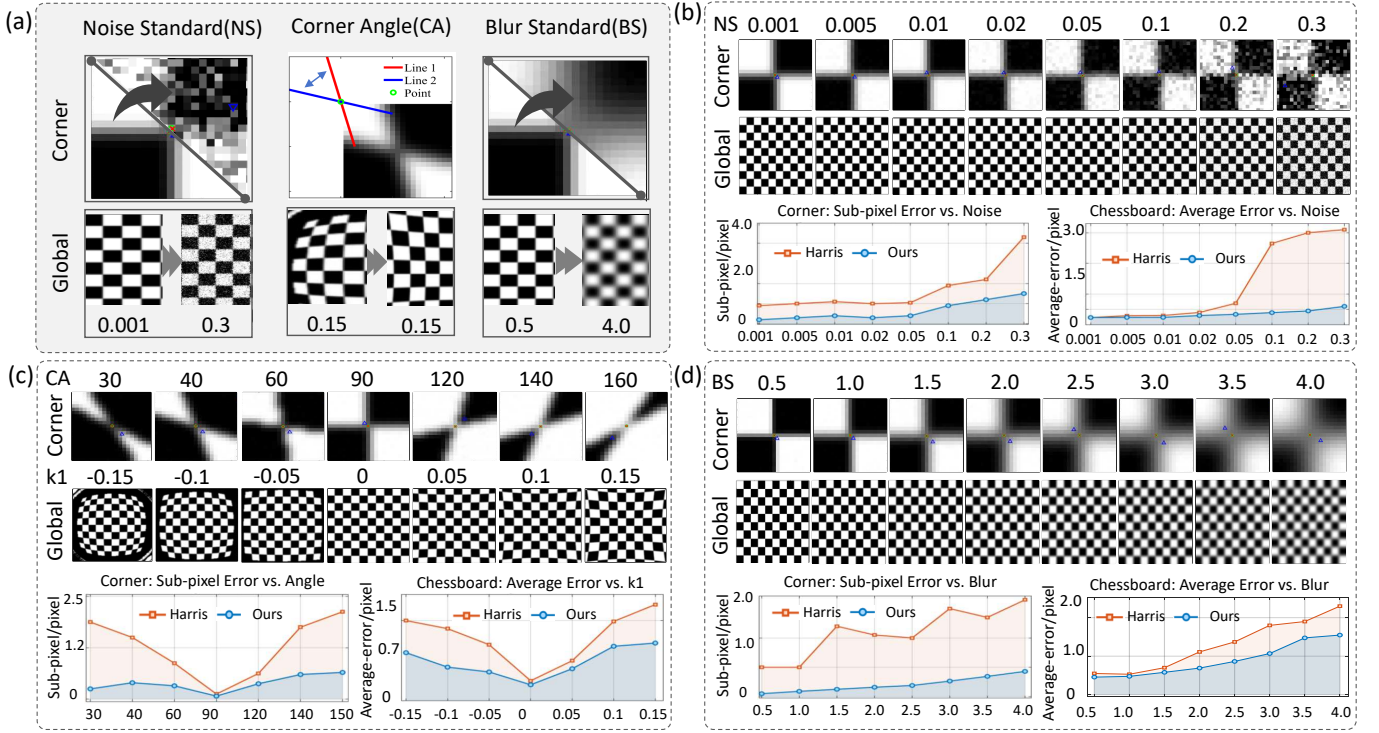


FIG. 2: **Robust sub-pixel corner localization under controlled degradations.** (a) Three representative perturbation factors used to stress-test corner extraction on checkerboard images: additive Gaussian noise with standard deviation  $NS$ , corner-geometry variation quantified by the corner opening angle  $CA$  (intersection of two edges), and Gaussian blur with standard deviation  $BS$ . For each factor we show a zoomed corner patch (*Corner*, top) together with the corresponding global checkerboard appearance (*Global*, bottom). (b–d) Robustness sweeps over (b) increasing noise levels ( $NS$ ), (c) varying corner angles ( $CA$ , with additional radial-distortion variations indicated by  $k_1$ ), and (d) increasing blur levels ( $BS$ ). Each panel reports representative *Corner* and *Global* visual results and two quantitative metrics: the sub-pixel localization error at the evaluated corner (left plots) and the average error aggregated over the full checkerboard (right plots). Across all degradations, the proposed pipeline (*Ours*) preserves the checkerboard lattice and maintains stable corner positions, whereas the Harris baseline exhibits pronounced error growth under strong noise and blur, leading to degraded corner accuracy and reduced global consistency.

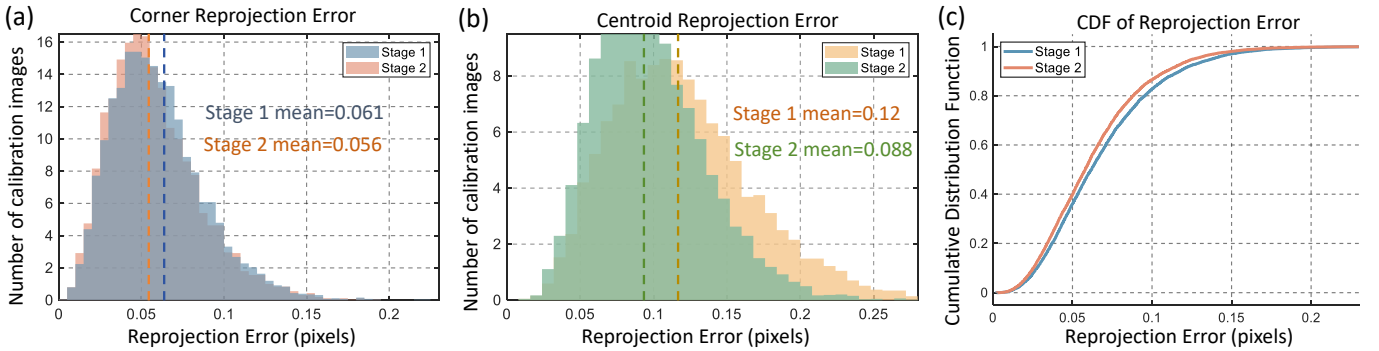


FIG. 3: **Two-stage calibration consistently reduces reprojection error.** (a) Histogram of per-image corner reprojection RMS error over all calibration views for Stage 1 (initial optimisation) and Stage 2 (refined optimisation). Dashed vertical lines indicate the mean error of each stage, showing a clear shift toward lower residuals after Stage 2. (b) The same analysis for centroid-based reprojection error, where each checkerboard cell is treated as a geometric primitive and its unbiased centroid is reprojected. Stage 2 yields a pronounced reduction in both the central tendency and the long-error tail. (c) Empirical cumulative distribution functions (CDFs) of corner reprojection error across all images. The Stage 2 curve rises faster and reaches unity earlier, confirming that the refinement improves not only the average case but also suppresses larger-error outliers across views.

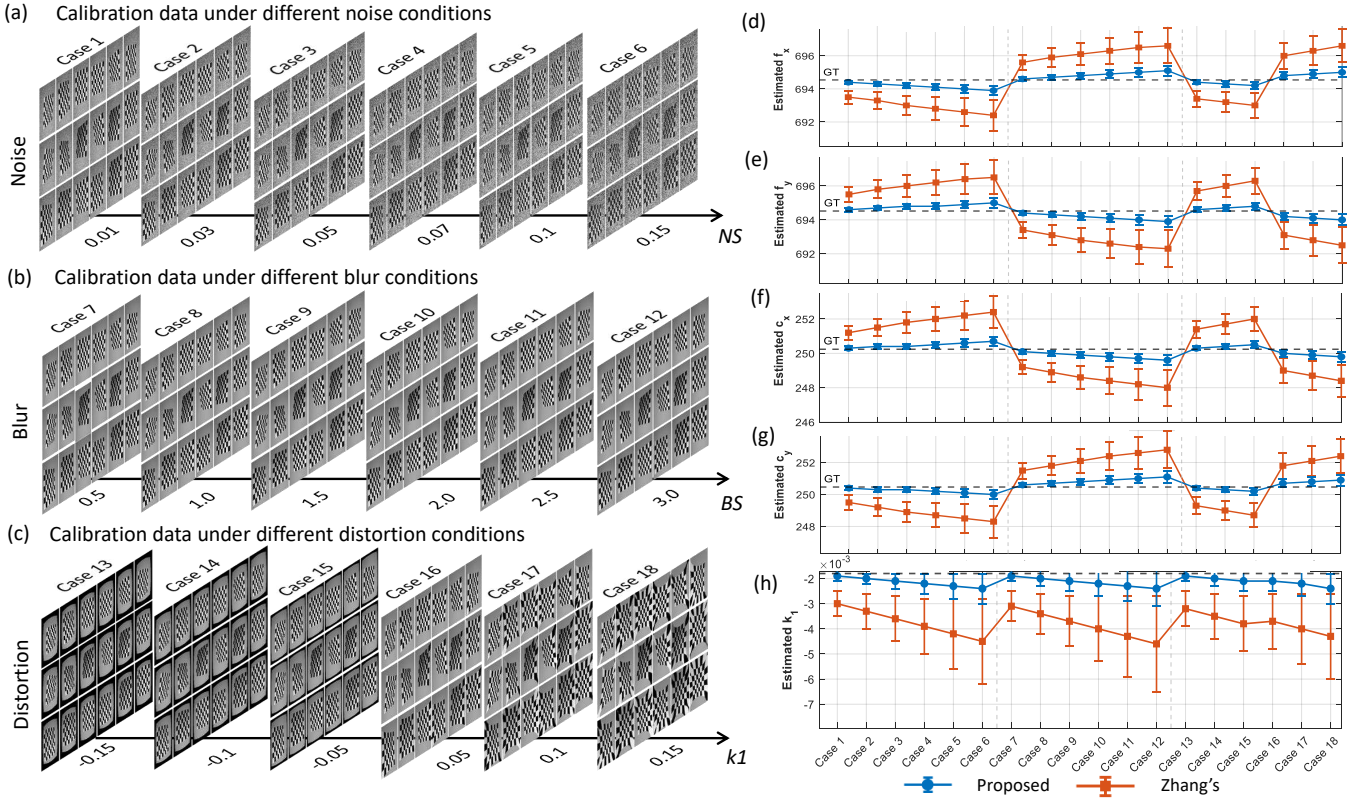


FIG. 4: **Parameter stability under controlled image degradations.** (a–c) Perturbation protocols used to generate 18 degraded calibration sets. We independently sweep (a) additive Gaussian noise with standard deviation NS, (b) Gaussian blur with standard deviation BS, and (c) radial distortion by varying the first-order coefficient  $k_1$ , and show representative degraded checkerboard observations for each case. (d–h) Estimated intrinsic parameters and the first radial distortion coefficient obtained by calibrating each degraded set, comparing the proposed two-stage method (*Proposed*, blue) against a standard Zhang-style implementation (*Zhang's/MATLAB*, red). The black dashed line indicates the reference value (GT) obtained from clean data; error bars show the variability across images within each set. Across all perturbation sweeps, our estimates remain tightly clustered around the reference with minimal drift, whereas the baseline exhibits increasing bias and variance under stronger blur and distortion, indicating reduced robustness of the conventional pipeline in degraded acquisition conditions.

ter rectification; and (iii) reconstruction-level point-cloud RMSE with respect to reference geometry. Unless otherwise stated, reprojection errors are in pixels and 3D errors in millimeters.

#### Sub-pixel corner and centroid localization under degradation

We first assess feature robustness under controlled degradations using synthetic checkerboard images with known ground-truth corners. Following the settings summarized in Fig. 2(a), we vary the Gaussian noise standard deviation (NS), Gaussian blur standard deviation (BS) and radial distortion coefficient  $K_1$  over wide ranges, and compare our geometry-aware SR plus saddle-point refinement against the classical Harris detector.

As shown qualitatively in Fig. 2(b)–(d), Harris progressively exhibits localization drift as degradation increases, with corners shifting along edges and failures emerging

at severe interference. In contrast, our method maintains coherent grid structure and stable corner positions, even at  $NS = 0.3$ ,  $BS = 4.0$ , or  $K_1 = \pm 0.15$ . This behavior reflects two complementary design choices: geometry-aware SR enhances only the checkerboard ROI to recover edge contrast where it matters, and saddle-point fitting explicitly enforces the four-quadrant intensity configuration of an ideal checkerboard corner. The same pipeline further supports the construction of unbiased cell centroids from polygon moments, providing higher-level geometric primitives that remain consistent with the underlying grid across all tested degradations and can be used as soft constraints in the subsequent refinement stage.

#### Centroid-assisted two-stage calibration accuracy

We quantify how the centroid-assisted Stage 2 refinement improves upon the standard corner-only initialization (Stage 1). Fig. 3 reports reprojection error statis-

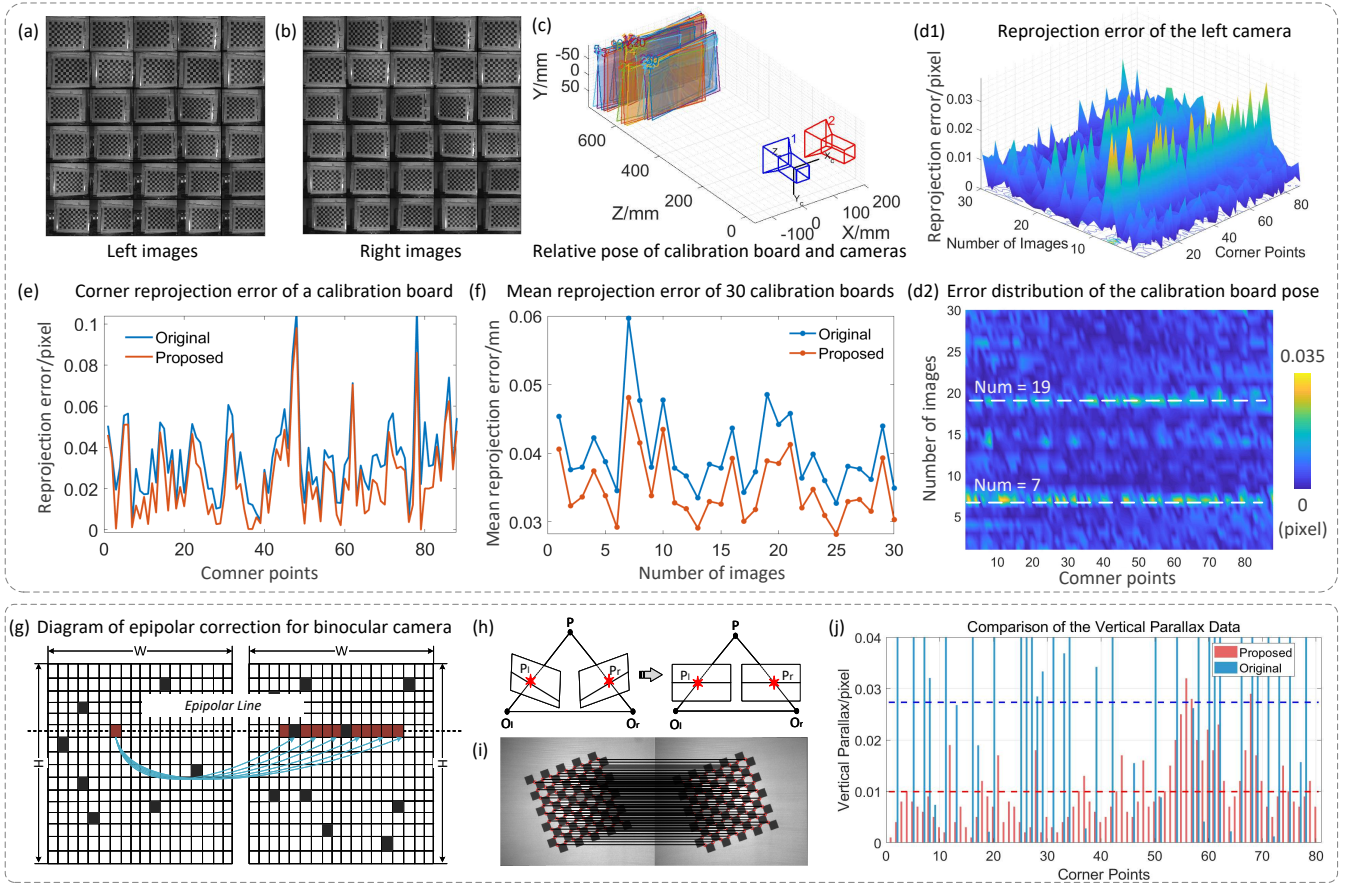


**FIG. 5: Calibration accuracy and epipolar alignment evaluation of the proposed two-stage binocular calibration.** (a,b) Representative left/right calibration images of the checkerboard target. (c) Estimated relative poses of the calibration board over 30 views together with the stereo rig geometry in the rig coordinate frame (axes in mm). (d1) View- and corner-wise reprojection error of the left camera visualized as a 3D surface (error in pixels), showing the non-uniform error distribution across calibration views and corner indices. (e) Corner-wise reprojection error for a representative board, comparing the baseline calibration (*Original*, MATLAB calibration toolbox) and our method (*Proposed*). (f) Mean reprojection error over all corners for each of the 30 calibration views; the proposed method consistently reduces the error across the sequence. (d2) Heatmap of reprojection error distribution across corner indices as the number of calibration images varies; dashed lines highlight two representative settings (Num= 7 and Num= 19). (g) Schematic of epipolar correction for a binocular system: after rectification, ideal correspondences should lie on the same horizontal epipolar line, and vertical offsets are measured and compensated. (h) Geometric illustration of vertical-parallax reduction before/after correction. (i) Overlay of rectified corner correspondences demonstrating improved horizontal alignment. (j) Per-corner vertical parallax magnitude  $|y_L - y_R|$  after rectification; the proposed calibration substantially suppresses vertical disparity compared with the baseline, indicating tighter epipolar consistency under real calibration conditions.

tics across a large set of calibration images. Stage 2 reduces the mean corner reprojection RMS from 0.061 px to 0.056 px (Fig. 3(a)) while preserving a tight distribution, indicating that refinement does not introduce outliers. The improvement is more pronounced for centroids: the mean centroid reprojection error drops from 0.12 px to 0.088 px (Fig. 3(b)). The CDF in Fig. 3(c) shows that Stage 2 yields lower errors for nearly all images, supporting the role of unbiased centroids as effective priors that regularize distortion and extrinsic estimation without weakening the corner constraints.

We further evaluate generalization against three representative baselines: OpenCV toolbox<sup>39</sup>, Wang’s

checkerboard-based method<sup>40</sup>, and the Matlab calibration toolbox<sup>41</sup>, using three additional calibration sets captured under distinct conditions. Table I summarizes reprojection RMS errors. Across all sets, GeoRobust-Calib (Stage 1 + Stage 2) achieves the lowest reprojection RMS. On Set-1, the reprojection RMS decreases from 0.1241 px (Matlab toolbox) to 0.0931 px, corresponding to a reduction of about 25%. On Set-2 and Set-3, our method improves upon the strongest baseline by approximately 10%–15%. These results indicate that fusing SR-enhanced features with saddle-point refinement and centroid priors improves calibration beyond the training distribution and remains effective under unseen poses and

imaging conditions.

Beyond feature-level robustness, we test whether the complete pipeline yields stable camera parameters when calibration images are perturbed. Using the synthetic dataset, we generate 18 degraded calibration sets by sweeping noise, blur and radial distortion according to Fig. 2, and estimate binocular parameters using our full pipeline and the Matlab baseline. Fig. 4 tracks the evolution of focal length, principal point and  $k_1$  as degradations intensify. Across all three perturbation types, parameters estimated by GeoRobustCalib remain nearly invariant and close to the ground-truth reference, suggesting that residual feature bias is effectively suppressed at the optimization level. By contrast, the baseline exhibits noticeable drift, particularly in the principal point and distortion coefficient, under strong blur or geometric distortion. These results demonstrate that combining SR-enhanced observability with saddle-point refinement and centroid-guided Stage 2 optimization improves not only local feature accuracy but also the global stability of the calibrated camera model under adverse imaging conditions.

### Ablation analysis

To disentangle the contribution of each component, we perform an ablation study on our binocular dataset. Starting from a Zhang-style baseline that uses raw corners only (no SR, no saddle-point refinement, no centroid-assisted Stage 2), we progressively add: (i) geometry-aware checkerboard SR (SR), (ii) saddle-point corners and unbiased centroids (SPC), and (iii) centroid-assisted two-stage refinement (2-Stage). Quantitative results are reported in Table II.

The baseline yields a corner RMS of 0.165 px, a centroid bias of 0.118 px, a reprojection RMS of 0.181 px, and a 3D RMSE of 0.208 mm. SR alone provides consistent but modest gains, indicating that improved local contrast benefits downstream geometry. Adding SR + SPC (corner-only calibration) produces a larger improvement, reducing reprojection RMS to 0.145 px and 3D RMSE to 0.171 mm, confirming that saddle-point refinement and centroid construction reduce degradation-induced localization bias. The full system (SR + SPC + 2-Stage) achieves the best overall performance (reprojection RMS 0.136 px; 3D RMSE 0.157 mm), corresponding to reductions of about 24.9% and 24.5% relative to the baseline, respectively. These results support a complementary effect: SR enhances feature observability, SPC reduces local bias, and Stage 2 further removes systematic residuals through centroid regularization.

Complementary to Table II, Fig. 5 provides a view-level and point-level diagnosis of the binocular calibration and the resulting epipolar consistency. We first show representative checkerboard observations from the

left and right cameras (Fig. 5a,b) and the recovered distribution of calibration-board poses together with the stereo rig geometry (Fig. 5c), illustrating the diversity of viewpoints used in the optimisation. The reprojection behaviour of the baseline solution is further visualised by a per-view/per-corner error surface (Fig. 5d1) and a corresponding error map across images and corner indices (Fig. 5d2), where two representative settings (Num= 7 and Num= 19) are highlighted to indicate how error patterns vary with the number of views. We then directly compare the corner reprojection error between the MATLAB toolbox (*Original*) and our full configuration (*Proposed*). For an individual board, our method lowers the error across most corners (Fig. 5e), and this improvement persists over all 30 calibration views when plotting the mean reprojection error per image (Fig. 5f), indicating a more stable and accurate calibration across viewpoints. Finally, we evaluate epipolar alignment after stereo rectification. The principle of epipolar correction and the associated vertical-parallax metric are illustrated schematically (Fig. 5g,h), and a qualitative overlay of rectified corner correspondences confirms tighter alignment (Fig. 5i). Quantitatively, the per-corner vertical parallax is markedly reduced by our method (Fig. 5j), yielding an average vertical disparity of  $\sim 0.01$  px, compared with  $\sim 0.037$  px for the MATLAB toolbox. This near-elimination of residual vertical mismatch indicates substantially improved epipolar consistency, which directly benefits downstream stereo matching and structured-light 3D reconstruction under real calibration conditions.

### Structured-light 3D reconstruction

Finally, we test whether the calibration improvements translate into downstream SL3D measurement gains. We integrate the estimated parameters into a standard phase-based SL3D pipeline and evaluate four representative scenes with varying geometric complexity (Fig. 6). Using the proposed calibration, we decode absolute phase (Fig. 6(b)) and triangulate dense point clouds (Fig. 6(c)). Reconstructions obtained with GeoRobustCalib are visually smoother, exhibit fewer holes and preserve sharper geometric details, especially on curved surfaces and fine structures.

For quantitative evaluation, we compute surface RMSE against high-quality reference reconstructions. As summarized in Fig. 6(d, e), GeoRobustCalib consistently outperforms the Matlab calibration toolbox across all scenes. RMSE decreases from about 0.26–0.33 mm with the baseline calibration to about 0.14–0.20 mm with GeoRobustCalib, corresponding to an error reduction of roughly 30%. These results indicate that improving calibration robustness yields tangible benefits in practical SL3D reconstruction quality under realistic imaging conditions.

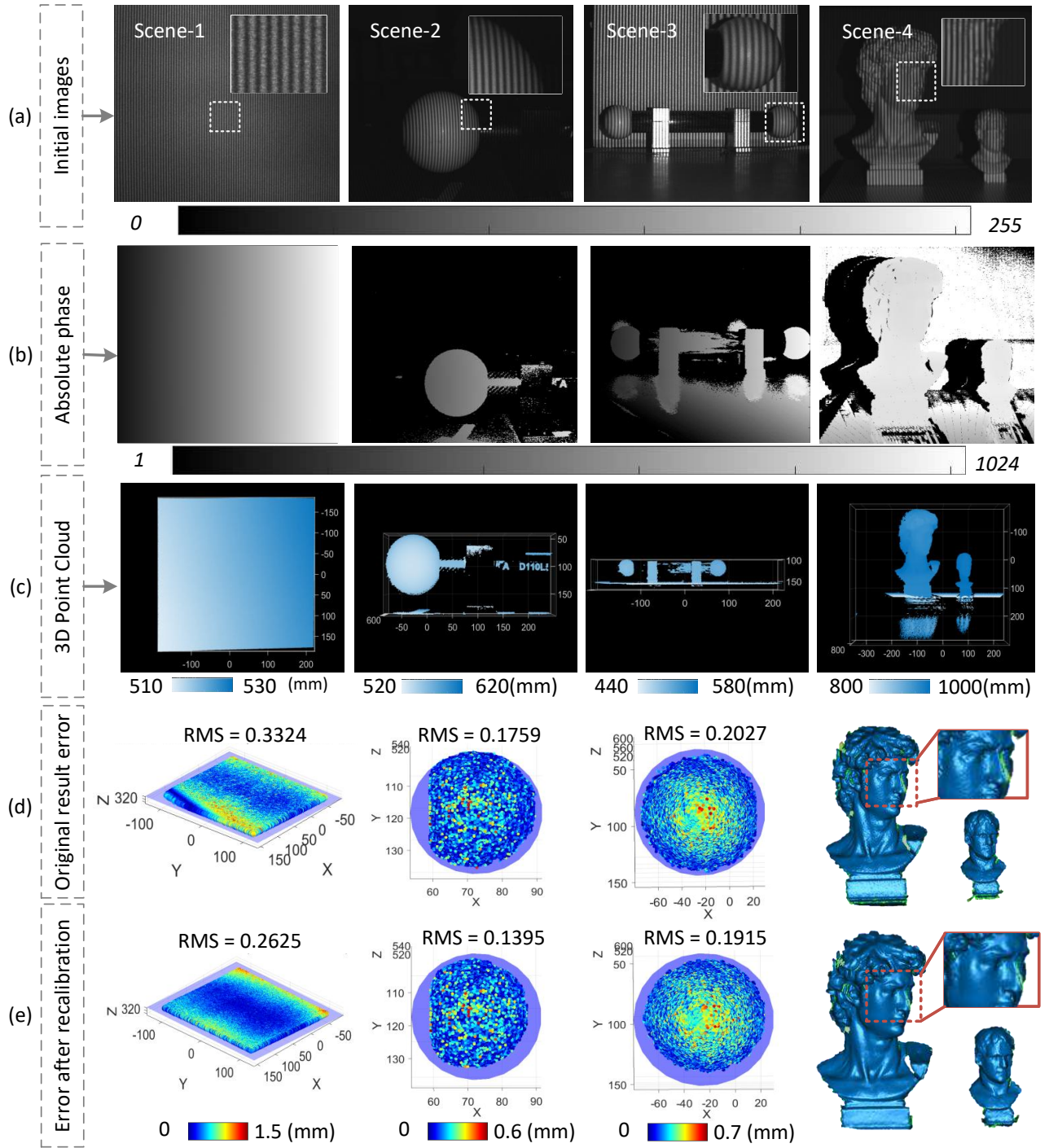


FIG. 6: **Improved binocular calibration translates to higher-fidelity structured-light 3D reconstruction.** (a) Representative captured fringe patterns from four test scenes (Scene 1–4); insets show zoomed regions highlighting fine stripe details and challenging surface reflectance/geometry. (b) Absolute phase maps recovered by phase decoding from the captured fringes (phase range encoded by the colour bar). (c) Dense 3D point clouds obtained by stereo triangulation using the corresponding calibration parameters. (d,e) Reconstruction error visualisations on a planar target (Scene 1), a sphere (Scene 2), a complex industrial object (Scene 3), and a textured sculpture (Scene 4). Errors are computed by fitting the appropriate geometric primitives (plane/sphere) or by surface-to-surface deviation for complex shapes, and are colour-coded in millimetres with the RMS reported above each plot. Using the proposed calibration (e) consistently reduces the RMS error relative to the baseline MATLAB toolbox result (d), yielding cleaner planar depth, less distortion on curved surfaces, and sharper recovery of fine structures (see zoom-in on Scene 4).

TABLE I: Cross-dataset evaluation on additional calibration sets. Best results are in **bold**, second best are underlined.

| Method                       | Set-1 Reproj. RMS ↓ | Set-2 Reproj. RMS ↓ | Set-3 Reproj. RMS ↓ |
|------------------------------|---------------------|---------------------|---------------------|
| Opencv-toolbox <sup>39</sup> | 0.2240              | 0.1075              | 0.0831              |
| Wang’s method <sup>40</sup>  | 0.1742              | 0.0736              | 0.0325              |
| Matlab-toolbox <sup>41</sup> | <u>0.1241</u>       | <u>0.0668</u>       | <u>0.0293</u>       |
| Ours Stage 1+2               | <b>0.0931</b>       | <b>0.0590</b>       | <b>0.0269</b>       |

TABLE II: Ablation study of the three proposed modules on our binocular dataset. SR: geometry-aware checkerboard super-resolution; SPC: saddle-point corners and unbiased centroids; 2-Stage: centroid-assisted two-stage calibration. Best results are in **bold**, second best are underlined.

| Method                               | Corner RMS ↓<br>(px) | Centroid bias ↓<br>(px) | Reproj. RMS ↓<br>(px) | 3D RMSE ↓<br>(mm) |
|--------------------------------------|----------------------|-------------------------|-----------------------|-------------------|
| matlab (baseline, no SR/SPC/2-Stage) | 0.165                | 0.118                   | 0.181                 | 0.208             |
| + SR only                            | 0.152                | 0.111                   | 0.170                 | 0.196             |
| + SR + SPC (corner-only calib.)      | <u>0.131</u>         | <u>0.086</u>            | <u>0.145</u>          | <u>0.171</u>      |
| Full (SR + SPC + 2-Stage, ours)      | <b>0.124</b>         | <b>0.072</b>            | <b>0.136</b>          | <b>0.157</b>      |

### III. DISCUSSION

GeoRobustCalib improves binocular calibration by coupling feature observability, local geometric fidelity and global stereo consistency. The key idea is to move beyond corner-only constraints by treating each checkerboard cell as a geometric primitive: saddle-point fitting stabilizes corner localization at sub-pixel precision, while unbiased cell centroids provide complementary, degradation-resistant cues that regularize distortion and extrinsic estimation under epipolar consistency. The resulting two-stage formulation preserves the strong constraints of corners and uses centroids as soft priors to remove residual systematic bias, yielding both lower re-projection error and improved parameter stability under noise, blur and distortion.

The improved calibration is consequential for structured-light metrology. Because SL3D reconstruction couples phase decoding and triangulation to the calibrated stereo model, parameter drift and spatially non-uniform calibration errors directly translate into surface noise, holes and biased geometry. By reducing feature-level bias and stabilizing intrinsic and distortion parameters, GeoRobustCalib tightens epipolar alignment and improves reconstruction fidelity across diverse scenes, consistent with the observed reduction in point-cloud RMSE.

Several limitations motivate future work. First, super-resolution performance may depend on the appearance and capture conditions of the calibration target; extending the training strategy to broader target materials and severe illumination regimes may further improve robustness. Second, while centroids regularize the solution effectively, the optimal weighting schedule can vary with imaging quality and target coverage; adaptive weighting driven by uncertainty estimates could provide a principled alternative. Finally, exploring alternative planar targets and extending the geometric-fusion formulation to multi-camera rigs may broaden applicability to emerg-

ing compact 3D systems.

Overall, GeoRobustCalib offers a unified calibration route that explicitly fuses learning-enhanced observability with geometric consistency. Starting from geometry-aware super-resolution restricted to the checkerboard ROI, the method refines sub-pixel corners via saddle-point fitting and derives unbiased cell centroids as degradation-resistant geometric primitives. These complementary measurements are then integrated in a centroid-assisted two-stage optimization, where corners act as hard constraints and centroids (together with lightweight epipolar regularization) serve as soft priors to suppress residual bias in distortion and extrinsics. By coupling local feature fidelity with global stereo consistency, the proposed framework yields stable, precision-grade binocular calibration and translates these gains into improved structured-light 3D reconstruction accuracy under realistic degradations.

#### Data availability

The data that support the findings of this study are available within the paper and the Extended Data. Other relevant data are available from the corresponding authors on reasonable request.

#### Acknowledgements

Authors thank the National Natural Science Foundation of China (Grant No. U21B2035, 62074128) and Key-Area Research and Development Program of Guangdong Province (2021B0101410001).

#### Author Contributions

D. Q and Y. S. supervised the project; H. C. developed the algorithm development, theoretical simulation, and experiments; all authors analyzed the data and prepared the manuscript.

#### Conflict of Interest

The authors declare no conflict of interest.

- \* Electronic address: [dyqiao@nwpu.edu.cn](mailto:dyqiao@nwpu.edu.cn)
- † Electronic address: [benquan001@e.ntu.edu.sg](mailto:benquan001@e.ntu.edu.sg)
- ‡ Electronic address: [yijie.shen@ntu.edu.sg](mailto:yijie.shen@ntu.edu.sg)
- <sup>1</sup> Zhang, Z. *et al.* Structured light meets machine intelligence. *eLight* **5**, 1–35 (2025).
  - <sup>2</sup> Gong, X. *et al.* Industrial binocular vision system calibration with unknown imaging model. *IEEE Transactions on Industrial Informatics* **20**, 7370–7379 (2024).
  - <sup>3</sup> Zhang, H. *et al.* Study on calibration method for binocular colour camera. *Optics and Lasers in Engineering* **188**, 108894 (2025).
  - <sup>4</sup> Lin, H., Wang, L., Qu, X. *et al.* A high-precision calibration and evaluation method based on binocular cameras and lidar for intelligent vehicles. *IEEE Transactions on Vehicular Technology* (2025).
  - <sup>5</sup> Hua, Y. *et al.* Accurate lidar-camera calibration using feature edges. *Image and Vision Computing* **155**, 105394 (2025).
  - <sup>6</sup> Sun, X., Song, K., Wen, X., Wang, Y. & Yan, Y. Sdd-detr: surface defect detection for no-service aero-engine blades with detection transformer. *IEEE Transactions on Automation Science and Engineering* (2024).
  - <sup>7</sup> Li, Z. *et al.* Enhancing smt quality and efficiency with self-adaptive collaborative optimization. *IEEE Transactions on Cybernetics* (2024).
  - <sup>8</sup> Forbes, A., De Oliveira, M. & Dennis, M. R. Structured light. *Nature Photonics* **15**, 253–262 (2021).
  - <sup>9</sup> Dorrah, A. H. & Capasso, F. Tunable structured light with flat optics. *Science* **376**, eabi6860 (2022).
  - <sup>10</sup> Li, Y. *et al.* Sl3d-bf: A real-world structured light 3d dataset with background-to-foreground enhancement. *IEEE Transactions on Circuits and Systems for Video Technology* (2025).
  - <sup>11</sup> Cui, B., Tao, W. & Zhao, H. High-precision 3d reconstruction for small-to-medium-sized objects utilizing line-structured light scanning: A review. *Remote Sensing* **13**, 4457 (2021).
  - <sup>12</sup> Liang, J., Ye, Y., Wu, D., Chen, S. & Song, Z. High-efficiency automated triaxial robot grasping system for motor rotors using 3d structured light sensor. *Machine Vision and Applications* **35**, 132 (2024).
  - <sup>13</sup> Choi, E. *et al.* 360° structured light with learned metasurfaces. *Nature Photonics* **18**, 848–855 (2024).
  - <sup>14</sup> Li, X. *et al.* Adaptive structured-light 3d surface imaging with cross-domain learning. *Laser & Photonics Reviews* 2401609 (2025).
  - <sup>15</sup> Hsu, W.-C., Chang, C.-H., Hong, Y.-H., Kuo, H.-C. & Huang, Y.-W. Metasurface-and pscel-based structured light for monocular depth perception and facial recognition. *Nano letters* **24**, 1808–1815 (2024).
  - <sup>16</sup> Tan, J., Niu, H., Su, W. & He, Z. Structured light 3d shape measurement for translucent media base on deep bayesian inference. *Optics & Laser Technology* **181**, 111758 (2025).
  - <sup>17</sup> Yang, T. & Gu, F. Overview of modulation techniques for spatially structured-light 3d imaging. *Optics & Laser Technology* **169**, 110037 (2024).
  - <sup>18</sup> Wang, B. *et al.* Single-shot super-resolved fringe projection profilometry (sssr-fpp): 100,000 frames-per-second 3d imaging with deep learning. *Light: Science & Applications* **14**, 70 (2025).
  - <sup>19</sup> Li, Y., Qian, J., Feng, S., Chen, Q. & Zuo, C. Deep-learning-enabled dual-frequency composite fringe projection profilometry for single-shot absolute 3d shape measurement. *Opto-Electronic Advances* **5**, 210021–1 (2025).
  - <sup>20</sup> Chen, W. *et al.* Deep-learning-enabled temporally super-resolved multiplexed fringe projection profilometry: high-speed khz 3d imaging with low-speed camera. *Photonix* **5**, 25 (2024).
  - <sup>21</sup> Yin, W. *et al.* High-resolution, wide-field-of-view, and real-time 3d imaging based on spatial-temporal speckle projection profilometry with a vcsel projector array. *ACS Photonics* **11**, 498–511 (2024).
  - <sup>22</sup> Zuo, C. *et al.* Deep learning in optical metrology: a review. *Light: Science & Applications* **11**, 39 (2022).
  - <sup>23</sup> Li, Y. *et al.* mrpca-pp: Robust motion-resistant phase-shifting profilometry based on principal component analysis. *Laser & Photonics Reviews* 2401938 (2025).
  - <sup>24</sup> Wang, S. *et al.* Accurate projector calibration based on an improved sub-pixel homography mapping method of circular edge points. *Optics and Lasers in Engineering* **186**, 108768 (2025).
  - <sup>25</sup> Zhao, W., Shi, X., Ni, F. & Tian, Y. Semi-dense sub-pixel displacement measurement for structural health monitoring: A framework of deep learning-based detector-free feature matching. *Measurement* 117899 (2025).
  - <sup>26</sup> Luo, J. *et al.* A binocular camera synchronization calibration method using closed-loop hybrid optimization for the froth flotation process. *IEEE Sensors Journal* (2025).
  - <sup>27</sup> Wu, J. *et al.* Displacement monitoring and modal parameter identification of cable net structure based on feature optical flow and binocular stereo vision. In *Structures*, vol. 76, 108914 (Elsevier, 2025).
  - <sup>28</sup> Sheng, B. *et al.* A markless 3d human motion data acquisition method based on the binocular stereo vision and lightweight open pose algorithm. *Measurement* **225**, 113908 (2024).
  - <sup>29</sup> Li, Q. *et al.* A review of building extraction from remote sensing imagery: Geometrical structures and semantic attributes. *IEEE Transactions on Geoscience and Remote Sensing* **62**, 1–15 (2024).
  - <sup>30</sup> Zou, L., Huang, Z., Gu, N. & Wang, G. Learning geometric consistency and discrepancy for category-level 6d object pose estimation from point clouds. *Pattern Recognition* **145**, 109896 (2024).
  - <sup>31</sup> Tian, Y., Chen, H., Xu, C. & Wang, Y. Image processing gnn: Breaking rigidity in super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 24108–24117 (2024).
  - <sup>32</sup> Xu, J. *et al.* Super-resolution 3d reconstruction from low-dose biomedical images based on expertized multi-layer refining. *Expert Systems with Applications* **281**, 127474 (2025).
  - <sup>33</sup> Dan, J. *et al.* Pirn: Phase invariant reconstruction network for infrared image super-resolution. *Neurocomputing* **599**, 128221 (2024).
  - <sup>34</sup> Cao, Z. *et al.* Aberration-robust monocular passive depth sensing using a meta-imaging camera. *Light: Science & Applications* **13**, 236 (2024).
  - <sup>35</sup> Li, Z.-S. *et al.* Real-time holographic camera for obtaining real 3d scene hologram. *Light: Science & Applications* **14**, 74 (2025).
  - <sup>36</sup> Zhou, Z. *et al.* Electrically tunable planar liquid-crystal

- singlets for simultaneous spectrometry and imaging. *Light: Science & Applications* **13**, 242 (2024).
- <sup>37</sup> Shanker, A. *et al.* Quantitative phase imaging endoscopy with a metalens. *Light: Science & Applications* **13**, 305 (2024).
- <sup>38</sup> Zhang, L. *et al.* Real-time machine learning-enhanced hyperspectro-polarimetric imaging via an encoding metasurface. *Science advances* **10**, eadp5192 (2024).
- <sup>39</sup> Wu, A., Xiao, H. & Zeng, F. A camera calibration method based on opencv. In *Proceedings of the 4th International Conference on Intelligent Information Processing, ICIIP '19*, 320–324 (Association for Computing Machinery, New York, NY, USA, 2020). URL <https://doi.org/10.1145/3378065.3378127>.
- <sup>40</sup> Wang, G., Zheng, H. & Zhang, X. A robust checkerboard corner detection method for camera calibration based on improved yolox. *Frontiers in Physics* **9**, 819019 (2022).
- <sup>41</sup> Xue-Jun, W., Dong-Yuan, G. & Xi-Fan, Y. Application of matlab calibration toolbox for camera's intrinsic and extrinsic parameters solving. In *2019 International Conference on Smart Grid and Electrical Automation (ICSGEA)*, 106–111 (2019).
- <sup>42</sup> Jasani, B. A., Lam, S.-K., Meher, P. K. & Wu, M. Threshold-guided design and optimization for harris corner detector architecture. *IEEE Transactions on Circuits and Systems for Video Technology* **28**, 3516–3526 (2017).
- <sup>43</sup> Song, C., Lee, D., Lim, J. & Kim, A. Camera calibration via circular patterns: A comprehensive framework with measurement uncertainty and unbiased projection model. *arXiv preprint arXiv:2506.16842* (2025).
- <sup>44</sup> Zhang, Z. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**, 1330–1334 (2000).

#### IV. METHOD

Fig. E3 illustrates the overall pipeline of our geometry-driven binocular calibration framework. Fig. E4 shows the coordinate system and a schematic diagram of the unbiased centroid during the calibration process. Given binocular images of a checkerboard, we first perform *geometry-aware super-resolution* restricted to the checkerboard region, using Harris-based masks of the global ROI, corners, and grid structure to enhance only the pattern while preserving the background. On the super-resolved checkerboard, we then apply *saddle-point fitting* to obtain reliable sub-pixel corners and treat each checkerboard cell as a geometric primitive with an unbiased centroid. Finally, we conduct a *centroid-assisted two-stage binocular calibration*, where Stage 1 performs conventional corner-only calibration and Stage 2 refines the intrinsics and extrinsics by keeping corners as hard constraints and introducing cell centroids as soft priors, together with lightweight epipolar-line regularization.

Throughout the paper, we use a standard four-coordinate formulation: the world coordinate system  $\{W\}$  with 3D points  $\mathbf{P}_W$ , the left/right camera coordinate systems  $\{C^L\}, \{C^R\}$  with points  $\mathbf{P}_C^s$ , the normalized image coordinate systems  $\{r^s\}$  with points  $\mathbf{P}_r^s$ , and the pixel coordinate systems  $\{v^s\}$  with points  $\mathbf{P}_v^s$ , where  $s \in \{L, R\}$ . The rigid transformations  $(\mathbf{R}^s, \mathbf{t}^s)$ , intrinsic matrices  $\mathbf{K}^s$ , and distortion parameters  $\mathbf{k}^s$  follow the classical pinhole model, detailed definitions and notation for corners and cell centroids are provided in the Supplementary Material (Sec. S1).

##### Geometry-Aware Checkerboard Super-Resolution

The raw quality of calibration images has a direct impact on the reliability of feature extraction and, consequently, on the final calibration accuracy. Instead of blindly enhancing the entire image, we explicitly exploit the projective geometry of the checkerboard and perform super-resolution (SR) only on the checkerboard region. This design both reduces computational cost and, more importantly, enforces geometric consistency between the enhanced image and the underlying grid structure, as illustrated in Fig. E3.

**Geometry-Driven ROI and Grid Mask Extraction.** Given a pair of raw calibration images, we first perform light-weight preprocessing (denoising and contrast normalization) and then detect initial corner candidates using the Harris detector<sup>42</sup>. For a local displacement  $(u, v)$ , the Harris energy is defined as

$$E(u, v) = \sum_{x, y} w(x, y) [I(x + u, y + v) - I(x, y)]^2, \quad (1)$$

where  $I(x, y)$  denotes the image intensity and  $w(x, y)$  is a local window function. Although the Harris response is not sufficient for final sub-pixel localization under chal-

lenging conditions, it provides robust and dense corner hypotheses that are well suited as geometric seeds.

Instead of treating these corners as isolated points, we exploit the regular grid structure of the checkerboard. Let  $\mathbf{p}_g = [i, j, 1]^\top$  denote the homogeneous coordinate of the ideal grid point at integer row-column index  $(i, j)$ , and  $\mathbf{p}_v = [u, v, 1]^\top$  denote its corresponding image point in the pixel coordinate system  $\{v^s\}$ . We estimate a homography  $\mathbf{H}$  between the ideal grid and the image by robustly matching detected Harris corners to their nearest grid indices and solving

$$\mathbf{p}_v \sim \mathbf{H} \mathbf{p}_g, \quad (2)$$

where  $\sim$  denotes equality up to scale. This projective mapping allows us to propagate from a sparse set of corner detections to a complete checkerboard layout. Using the estimated homography  $\mathbf{H}$ , we generate three types of masks: (i) a global checkerboard ROI mask obtained from the convex hull of all projected grid points, (ii) an internal-corner mask that refines the corner neighborhood, and (iii) a per-cell polygon mask for each checkerboard square. These masks are shared by the left and right views after appropriate cropping and are used in subsequent stages to restrict SR reconstruction, guide saddle-point fitting, and support unbiased centroid computation. All implementation details for homography estimation, outlier rejection, and mask rasterization are provided in the Supplementary Material (Sec. S2.1).

**Checkerboard Super-Resolution with Geometric Priors.** With the geometry-driven ROI and grid masks, we perform super-resolution only within the checkerboard region, while leaving the background unchanged. This design preserves the original photometric context of the scene, avoids artifacts on irrelevant areas, and concentrates the network capacity on the structures that matter for calibration. Concretely, we adopt a learning-based SR backbone (IPG reconstruction<sup>31</sup>) and augment it with geometric priors derived from the masks. The input to the network consists of: (i) the cropped checkerboard patch, and (ii) auxiliary mask channels encoding the global ROI, the internal corner neighborhoods, and the per-cell grid structure. These mask channels act as soft geometric guidance, encouraging the network to sharpen edges along the projected grid lines and to preserve local contrast around corners and cell boundaries. The network is trained with a combination of pixel-wise reconstruction losses and a lightweight geometric consistency term that penalizes deviations between the enhanced checkerboard edges and the ideal grid predicted by the homography in Eq. (2). This geometry-aware enhancement significantly improves the signal-to-noise ratio for both corner and cell-boundary structures, providing a robust basis for the subsequent saddle-point-based corner localization and unbiased cell centroid estimation.

### Saddle-Point Corners and Unbiased Cell Centroids

We refine the initial checkerboard corners and construct unbiased cell centroids as geometric primitives. Different from moment-based dot methods<sup>43</sup>, which rely on intensity segmentation of circular blobs, our checkerboard formulation starts from high-precision corners and exploits the global grid structure (including cross-ratio invariants), leading to more accurate polygon moments and centroids.

**Second-Stage Saddle-Point Corner Refinement.** For each internal grid intersection in image  $n$  of camera  $s \in \{L, R\}$ , we first obtain an initial sub-pixel corner  $\hat{\mathbf{q}}_{v,0}^{s,(n,i,j)} = [u_0^{s,(n,i,j)}, v_0^{s,(n,i,j)}]^\top$  using a standard corner detector on the super-resolved image. Around this location we define a local coordinate system  $(x, y)$  such that

$$u = u_0^{s,(n,i,j)} + x, \quad v = v_0^{s,(n,i,j)} + y. \quad (3)$$

On the local SR patch  $\hat{I}_{\text{SR}}(x, y)$ , we fit a quadratic intensity surface

$$\hat{I}_{\text{SR}}(x, y) \approx ax^2 + bxy + cy^2 + dx + ey + f, \quad (4)$$

where the coefficients  $\{a, b, c, d, e, f\}$  are estimated by least squares. The gradient and Hessian are

$$\nabla \hat{I}_{\text{SR}}(x, y) = \begin{bmatrix} 2ax + by + d \\ bx + 2cy + e \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} 2a & b \\ b & 2c \end{bmatrix}. \quad (5)$$

A checkerboard corner is modeled as a *saddle point* of this surface: the gradient vanishes and the Hessian has eigenvalues of opposite signs. The refined sub-pixel offset  $(x^*, y^*)$  is therefore defined by

$$\nabla \hat{I}_{\text{SR}}(x^*, y^*) = \mathbf{0}, \quad \lambda_1(\mathbf{H}) \cdot \lambda_2(\mathbf{H}) < 0, \quad (6)$$

which yields the closed-form solution

$$\begin{bmatrix} x^* \\ y^* \end{bmatrix} = -\mathbf{H}^{-1} \begin{bmatrix} d \\ e \end{bmatrix}. \quad (7)$$

The second-stage refined corner in the pixel coordinate system  $\{v^s\}$  is then given by

$$\hat{\mathbf{q}}_v^{s,(n,i,j)} = \begin{bmatrix} u^{s,(n,i,j)} \\ v^{s,(n,i,j)} \end{bmatrix} = \begin{bmatrix} u_0^{s,(n,i,j)} + x^* \\ v_0^{s,(n,i,j)} + y^* \end{bmatrix}. \quad (8)$$

Compared with directly using the first-stage detector, this saddle-point refinement explicitly enforces the four-quadrant intensity pattern of a checkerboard corner on SR images and leverages the local second-order structure of the grayscale field, resulting in more accurate and robust corner coordinates. These refined corners are subsequently used for both corner-based calibration and cell centroid construction.

**Unbiased Cell Centroids from Polygon Moments.** Each checkerboard cell provides a 2D region whose centroid is well defined under the physical imaging model. We

follow the moment-based idea<sup>43</sup>, but replace circular dots by quadrilateral cells defined by four refined corners and further regularized by the global checkerboard geometry, as shown in Figure ??.

**Undistorted quadrilaterals on the normalized plane.** Consider the  $k$ -th cell in image  $n$  of camera  $s$ , bounded by four refined corners in the pixel coordinate system  $\{v^s\}$ ,  $\hat{\mathbf{q}}_v^{s,(n,k,\ell)} = [u_\ell, v_\ell]^\top$ ,  $\ell \in \{1, 2, 3, 4\}$ , ordered counter-clockwise. We first map these points to distorted normalized coordinates via the intrinsic matrix:

$$\tilde{\mathbf{p}}_d^{s,(n,k,\ell)} = \mathbf{K}^{s^{-1}} \begin{bmatrix} u_\ell \\ v_\ell \\ 1 \end{bmatrix}, \quad \mathbf{p}_d^{s,(n,k,\ell)} = (x_{d,\ell}, y_{d,\ell})^\top. \quad (9)$$

Given current distortion parameters  $\mathbf{k}^s$ , we apply inverse radial distortion to obtain undistorted normalized coordinates on the plane  $\{r^s\}$ :

$$\mathbf{p}_d^{s,(n,k,\ell)} = \mathcal{D}(\mathbf{p}_r^{s,(n,k,\ell)}; \mathbf{k}^s) \Rightarrow \hat{\mathbf{p}}_r^{s,(n,k,\ell)} = (x_\ell, y_\ell)^\top. \quad (10)$$

These four vertices define a quadrilateral region  $A_n^{s,(n,k)}$  in the undistorted normalized plane. Let  $\mathbf{v}_\ell = (x_\ell, y_\ell)^\top$  and  $\mathbf{v}_5 = \mathbf{v}_1$ . We define

$$\Delta_\ell = x_\ell y_{\ell+1} - x_{\ell+1} y_\ell, \quad \ell = 1, \dots, 4, \quad (11)$$

and obtain the area and centroid of  $A_n^{s,(n,k)}$  by the standard polygon formulas:

$$|A_n^{s,(n,k)}| = \frac{1}{2} \sum_{\ell=1}^4 \Delta_\ell, \quad \bar{x}_n^{s,(n,k)} = \frac{1}{6|A_n^{s,(n,k)}|} \sum_{\ell=1}^4 (x_\ell + x_{\ell+1}) \Delta_\ell, \quad (12)$$

$$\bar{y}_n^{s,(n,k)} = \frac{1}{6|A_n^{s,(n,k)}|} \sum_{\ell=1}^4 (y_\ell + y_{\ell+1}) \Delta_\ell. \quad (13)$$

**Moment-based centroid under radial distortion.** On the normalized plane, we adopt a polynomial radial distortion model

$$\mathbf{p}_d = k(s) \mathbf{p}_r, \quad s = x^2 + y^2, \quad k(s) = \sum_{i=0}^{n_d} d_i s^i, \quad (14)$$

where  $\mathbf{p}_r = (x, y)^\top$  and  $\mathbf{p}_d = (x_d, y_d)^\top$ . Following the moment-based derivation in<sup>43</sup>, one can show that the centroid of the distorted region  $A_d^{s,(n,k)} = \mathcal{D}(A_n^{s,(n,k)})$  can be written as a rational function of a finite set of polygon moments of  $A_n^{s,(n,k)}$ :

$$\bar{\mathbf{p}}_d^{s,(n,k)} = \begin{bmatrix} \bar{x}_d^{s,(n,k)} \\ \bar{y}_d^{s,(n,k)} \end{bmatrix} = \mathbf{F}_{\text{centroid}}(\{M_n^{a,b}(s, n, k)\}, \mathbf{k}^s), \quad (15)$$

where  $M_n^{a,b}(s, n, k)$  denotes normalized monomial moments of  $A_n^{s,(n,k)}$  and  $\mathbf{F}_{\text{centroid}}(\cdot)$  is an explicit, distortion-dependent mapping. The detailed expressions and derivation are provided in the Supplementary Material

(Sec. S3), and are omitted here for brevity. Under the assumed polynomial model,  $\bar{\mathbf{p}}_d^{s,(n,k)}$  is an *unbiased* centroid of the distorted cell in the normalized plane. Mapping back to the pixel coordinate system  $\{v^s\}$  via the intrinsic matrix yields the *theoretical* pixel centroid

$$\tilde{\mathbf{c}}_v^{s,(n,k)} = \begin{bmatrix} \tilde{u}_d^{s,(n,k)} \\ \tilde{v}_d^{s,(n,k)} \end{bmatrix} = \mathbf{K}^s \begin{bmatrix} \bar{x}_d^{s,(n,k)} \\ \bar{y}_d^{s,(n,k)} \\ 1 \end{bmatrix}. \quad (16)$$

Geometric centroid observation and residual. In parallel, we obtain a purely geometric centroid *observation* in the pixel plane by applying the polygon-centroid formula directly to the four refined corners  $\hat{\mathbf{q}}_v^{s,(n,k,\ell)}$  (similar to Eq. (13), but in  $\{v^s\}$ ); we denote this by

$$\hat{\mathbf{c}}_v^{s,(n,k)} = \begin{bmatrix} u_c^{s,(n,k)} \\ v_c^{s,(n,k)} \end{bmatrix}. \quad (17)$$

The difference between the theoretical and observed centroids,

$$\mathbf{e}_{\text{centroid}}^{s,(n,k)} = \tilde{\mathbf{c}}_v^{s,(n,k)} - \hat{\mathbf{c}}_v^{s,(n,k)}, \quad (18)$$

measures the combined effect of residual calibration error and local imaging degradation. Aggregated over all cells, images and cameras, these centroid residuals act as *soft constraints* in our centroid-assisted two-stage calibration, where corners remain as hard constraints.

Compared to the conic-based circular-dot formulation<sup>43</sup>, our checkerboard-based centroid construction benefits from (i) second-stage saddle-point corners that provide highly accurate polygon vertices, and (ii) a globally regular corner grid whose projective invariants (cross ratios along rows and columns) further stabilize the undistorted layout. These additional geometric priors lead to more accurate polygon moments and centroids, which in turn translate into improved calibration precision in the subsequent two-stage optimization.

### Centroid-Assisted Two-Stage Binocular Calibration

With refined saddle-point corners and unbiased cell centroids available for both cameras, we perform binocular calibration in two stages. Corners serve as *hard* geometric constraints, while centroids provide *soft* auxiliary constraints that regularize distortion and extrinsics without degrading corner accuracy. Detailed optimization and Jacobian derivations are given in the Supplementary Material (Sec. S4).

**Stage 1: Corner-Only Binocular Initialization.** In the first stage, we perform a conventional binocular calibration using only refined checkerboard corners. For camera  $s \in \{L, R\}$ , image index  $n$ , and internal corner index  $i$ , we denote the world coordinate on the checkerboard plane  $\{W\}$  as  $\mathbf{Q}_W^{(i)}$ ; refined pixel coordinate in  $\{v^s\}$  as  $\hat{\mathbf{q}}_v^{s,(n,i)}$  (Eq. (8)).

Using the projection chain defined in Sec. 3.1, the model prediction is

$$\tilde{\mathbf{q}}_v^{s,(n,i)} = \Pi(\mathbf{K}^s, \mathbf{k}^s, \mathbf{R}^{s,(n)}, \mathbf{t}^{s,(n)}, \mathbf{Q}_W^{(i)}), \quad (19)$$

and the corner reprojection residual is

$$\mathbf{e}_{\text{corner}}^{s,(n,i)} = \hat{\mathbf{q}}_v^{s,(n,i)} - \tilde{\mathbf{q}}_v^{s,(n,i)}. \quad (20)$$

Stage 1 estimates intrinsics  $\mathbf{K}^s$ , distortion  $\mathbf{k}^s$  and per-image extrinsics  $\{\mathbf{R}^{s,(n)}, \mathbf{t}^{s,(n)}\}$  by minimizing

$$\mathcal{L}_{\text{corner}} = \sum_{s \in \{L, R\}} \sum_n \sum_i \|\mathbf{e}_{\text{corner}}^{s,(n,i)}\|_2^2. \quad (21)$$

This yields a stable binocular initialization, equivalent in spirit to a Zhang-style calibration<sup>44</sup> but operating on our refined saddle-point corners.

**Stage 2: Centroid-Assisted Refinement.** Starting from the Stage 1 solution, we refine the parameters using both corners and cell centroids. For the  $k$ -th cell in image  $n$  of camera  $s$ : the *observed* geometric centroid in the pixel plane  $\{v^s\}$ ,  $\hat{\mathbf{c}}_v^{s,(n,k)}$ , is computed from the four refined corners using the polygon-centroid formula (Eq. (13) applied in  $\{v^s\}$ ); the *theoretical* unbiased centroid  $\tilde{\mathbf{c}}_v^{s,(n,k)}$  is obtained from polygon moments on the normalized plane  $\{r^s\}$  and the polynomial radial distortion model, then projected by  $\mathbf{K}^s$  (Eq. (16)).

The centroid residual is

$$\mathbf{e}_{\text{centroid}}^{s,(n,k)} = \tilde{\mathbf{c}}_v^{s,(n,k)} - \hat{\mathbf{c}}_v^{s,(n,k)}, \quad (22)$$

and the corresponding loss

$$\mathcal{L}_{\text{centroid}} = \sum_{s \in \{L, R\}} \sum_n \sum_k \|\mathbf{e}_{\text{centroid}}^{s,(n,k)}\|_2^2. \quad (23)$$

To further stabilize the binocular geometry, we optionally add a standard epipolar consistency term  $\mathcal{L}_{\text{epi}}$  based on the essential (or fundamental) matrix implied by the relative pose between  $\{C^L\}$  and  $\{C^R\}$ ; its exact expression is provided in the Supplement (Sec. S4).

The Stage 2 objective combines these three components:

$$\mathcal{L}_{\text{stage2}} = \mathcal{L}_{\text{corner}} + \lambda_c \mathcal{L}_{\text{centroid}} + \lambda_e \mathcal{L}_{\text{epi}}, \quad (24)$$

where  $\lambda_c$  and  $\lambda_e$  control the influence of centroid and epipolar constraints. In practice, we adopt a progressive schedule that gradually increases  $\lambda_c$  starting from a small value, while monitoring the RMS corner reprojection error and preventing it from degrading beyond a small tolerance with respect to Stage 1. This implements our design principle that corners act as hard constraints, whereas centroids provide soft yet informative regularization.

Overall, the two-stage strategy preserves the excellent local accuracy of saddle-point corners (Stage 1) and then reduces systematic bias in distortion and extrinsics through unbiased cell centroids and binocular geometry (Stage 2), which directly benefits downstream 3D reconstruction accuracy.

### Implementation Details

All components are implemented in Python: image processing, initial corner detection, and classical calibration build on OpenCV, while the geometry-aware checkerboard super-resolution module is realized in PyTorch. For each calibration image, we first locate the checkerboard ROI using a Harris-based mask and apply a fixed  $\times 2$  SR model only inside this ROI. Initial sub-pixel corners on the SR images are refined by the quadratic saddle-point fitting, after which valid cells are selected, quadrilaterals are formed from the refined corners, and unbiased centroids are computed via polygon moments on the normalized planes and Supplement Sec. S3.

Stage 1 performs a Zhang-style binocular initialization with bundle adjustment on refined corners only, minimizing  $\mathcal{L}_{\text{corner}}$ ; Stage 2 starts from this solution and optimizes the combined loss  $\mathcal{L}_{\text{stage2}}$  with a progressive increase of the centroid weight  $\lambda_c$  and a small fixed epipolar weight  $\lambda_e$ . We use a damped Gauss-Newton solver with auto-differentiated Jacobians and simple corner-RMS safeguards; typical hyperparameters, weight schedules, and runtime statistics are summarized in Supplement Sec. S5.

### Downstream Application: SL3D

SL3D reconstruction stands as a high-precision, high-efficiency active vision measurement technique. The core of this technology involves a projection system casting structured light patterns (such as Gray codes or phase-shifting codes). An imaging system (typically binocular cameras) then captures the deformed patterns, followed by phase decoding and 3D coordinate reconstruction to recover the target surface’s precise geometry.

## EXTENDED DATA

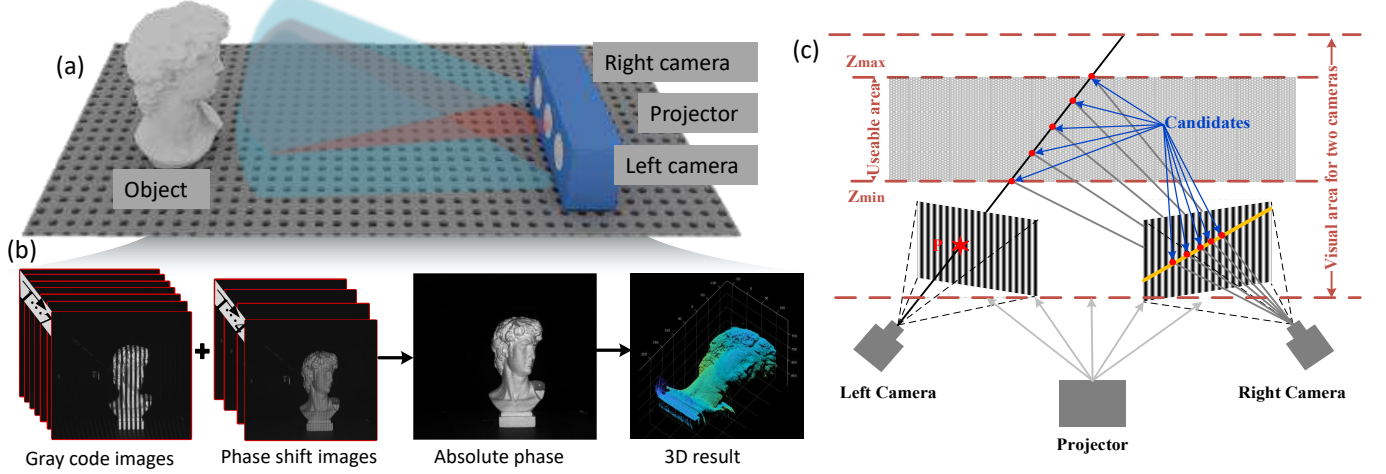


FIG. E1: **Dataset acquisition system and structured-light 3D reconstruction pipeline.** (a) Experimental setup of the binocular structured-light system, consisting of a left camera, a right camera, and a projector observing the common measurement volume. (b) Typical projection-capture sequence, including Gray-code images and phase-shifting fringe patterns, together with the decoded depth/absolute-phase maps. (c) Final 3D reconstruction result obtained from the absolute phase, illustrating the complete processing chain used throughout our experiments.

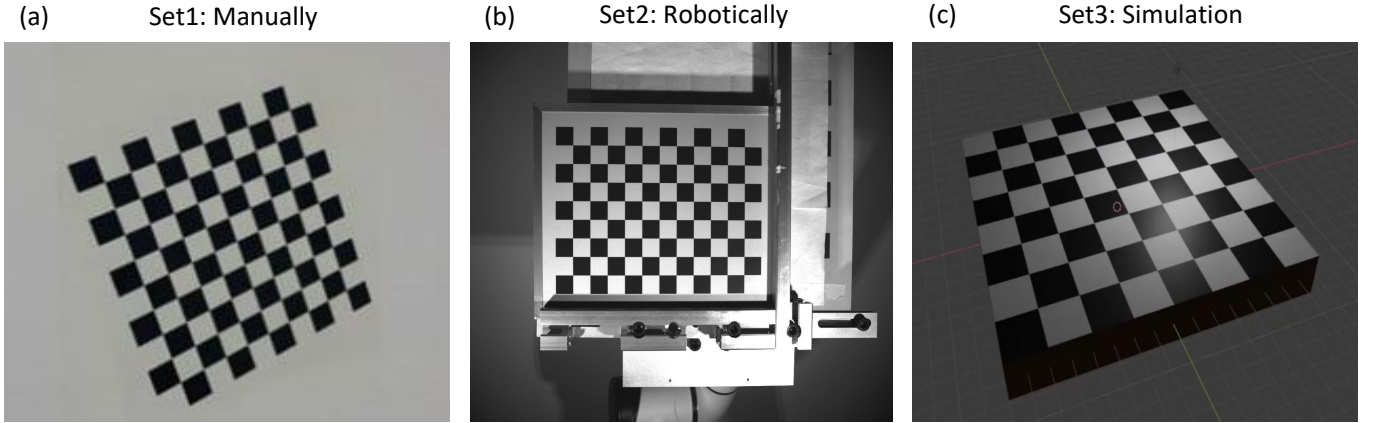


FIG. E2: **Overview of the calibration dataset acquisition.** (a) Set 1 (Manually): checkerboard images captured with manually adjusted target poses to provide diverse viewpoints. (b) Set 2 (Robotically): automated pose sampling by mounting the checkerboard at the end-effector of a six-axis robotic arm, enabling repeatable and dense pose coverage. (c) Set 3 (Simulation): rendered a synthetic checkerboard dataset with controllable geometry and imaging conditions for systematic evaluation.

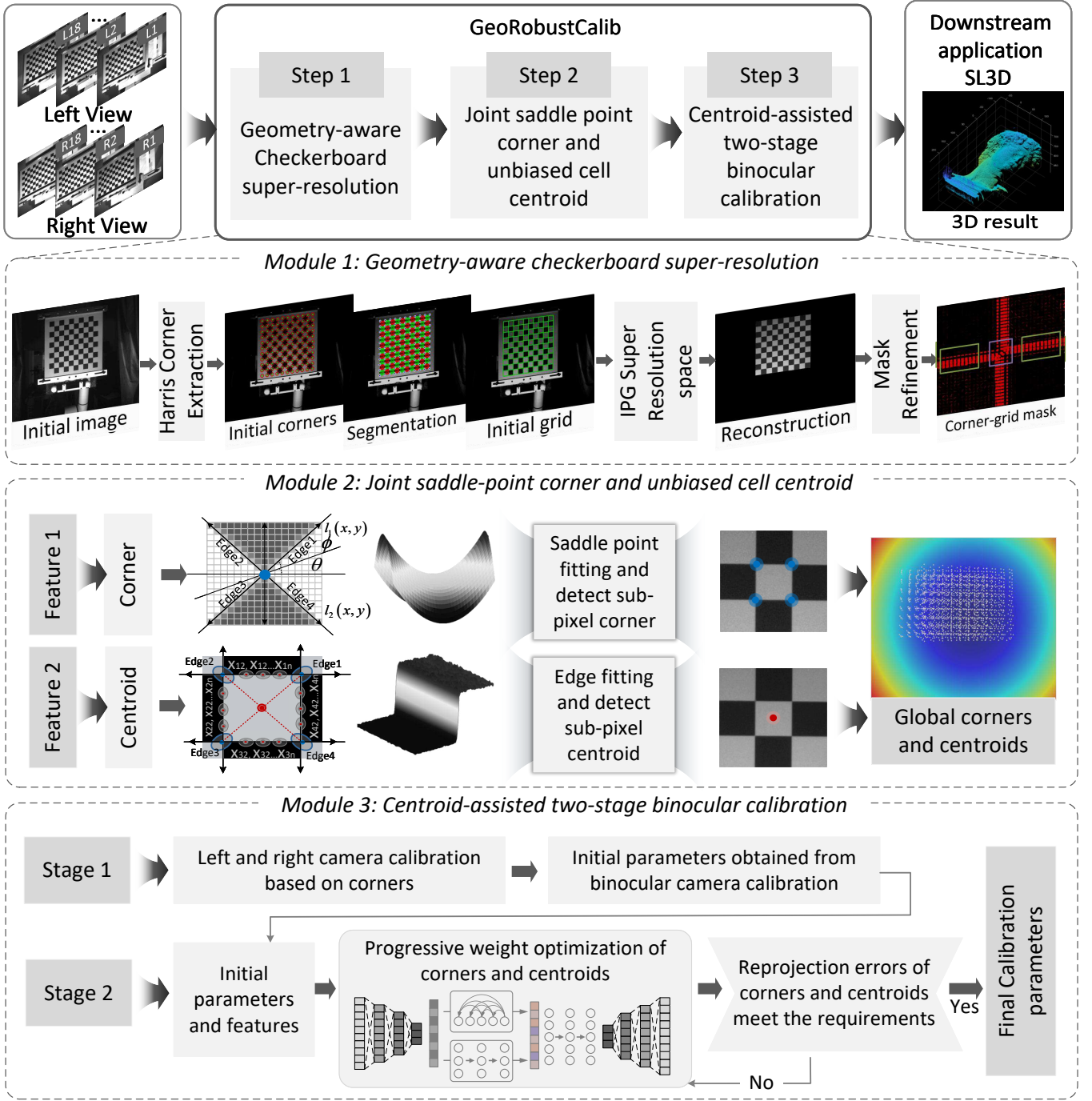


FIG. E3: **Overview of the generation pipeline.** The process begins by using the Harris detector to find initial corner points on calibration board images from both left and right views. Subsequently, geometric information is used to calculate the ROI and mask. Then, sub-pixel corners, boundaries, and centroids are precisely computed based on the geometric properties of the calibration board and stereo epipolar constraints to obtain global feature points. Finally, calibration parameters are calculated and optimized using these feature points, significantly improving calibration accuracy and 3D measurement precision.

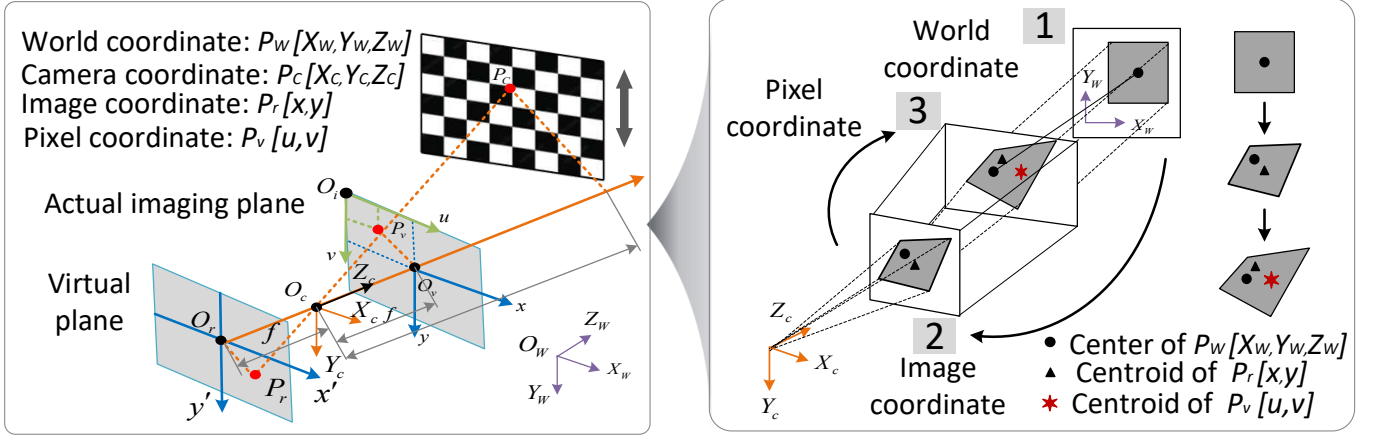


FIG. E4: **Checkerboard coordinate systems and centroid computation pipeline.** Left: Coordinate definitions for the checkerboard calibration target and the imaging geometry, including the world coordinate  $\mathbf{P}_w[X_w, Y_w, Z_w]$ , camera coordinate  $\mathbf{P}_c[X_c, Y_c, Z_c]$ , image-plane coordinate  $\mathbf{P}_r[x, y]$ , and pixel coordinate  $\mathbf{P}_v[u, v]$ . The schematic also illustrates the relationship between the actual imaging plane and a virtual reference plane used for geometric interpretation. Right: Centroid computation across coordinate spaces. Starting from the checkerboard cell in the world coordinate (1), the corresponding region is projected to the image coordinate (2) and discretized in the pixel grid (3). The centroids in world (black dot), image (black triangle), and pixel (red star) coordinates are indicated, illustrating how the cell centroid is propagated and measured for centroid-assisted calibration.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterialforGeometryAwareSuperResolutionFusionCalibrationforBinocularStructuredLight3DReconstruction.pdf](#)