

Optimal Pose Guidance for Stereo Calibration in 3D Deformation Measurement

Dongcai Tan^{1,2}, Shunkun Liang^{1,2}, Bin Li^{1,2}, Banglei Guan^{1,2*},
Ang Su^{1,2}, Yuan Lin³, Dapeng Zhang^{1,4}, Minggang Wan³,
Zibin Liu^{1,2}, Chenglong Wang³, Jiajian Zhu³, Zhang Li^{1,2},
Yang Shang^{1,2}, Qifeng Yu^{1,2}

¹College of Aerospace Science and Engineering, National University of Defense Technology, Changsha, 410073, Hunan, China.

²Hunan Provincial Key Laboratory of Image Measurement and Vision Navigation, Changsha, 410073, Hunan, China.

³Hypersonic Technology Laboratory, National University of Defense Technology, Changsha, 410073, Hunan, China.

⁴Hunan Key Laboratory of Intelligent Planning and Simulation for Aerospace Missions, Changsha, 410073, Hunan, China.

*Corresponding author(s). E-mail(s): banglei.guan@hotmail.com;

Abstract

Background Stereo optical measurement techniques, such as digital image correlation (DIC), are widely used in 3D deformation measurement as non-contact, full-field measurement methods, in which stereo calibration is a crucial step. However, current stereo calibration methods lack intuitive optimal pose guidance, leading to inefficiency and suboptimal accuracy in deformation measurements.

Objective The aim of this study is to develop an interactive calibration framework that automatically generates the next optimal pose, enabling high-accuracy stereo calibration for 3D deformation measurement.

Methods We propose a pose optimization method that introduces joint optimization of relative and absolute extrinsic parameters, with the minimization of the covariance matrix trace adopted as the loss function to solve for the next optimal pose. Integrated with this method is a user-friendly graphical interface, which guides even non-expert users to capture qualified calibration images.

Results Our proposed method demonstrates superior efficiency (requiring fewer images) and accuracy (demonstrating lower measurement errors) compared to random pose, while maintaining robustness across varying FOVs. In the thermal

deformation measurement tests (500–700°C) on an S-shaped specimen, the results exhibit high agreement with finite element analysis (FEA) simulations in both deformation magnitude and evolutionary trends.

Conclusions We present a pose guidance method for high-precision stereo calibration in 3D deformation measurement. The simulation experiments, real-world experiments, and thermal deformation measurement applications all demonstrate the significant application potential of our proposed method in the field of 3D deformation measurement.

Keywords: Stereo calibration, Optimal pose guidance, 3D deformation measurement, Digital image correlation

1 Introduction

Digital image correlation [1–3], and other optical measurement techniques [4–9] have emerged as standard solutions for 3D reconstruction, deformation analysis, and pose estimation due to their non-contact, full-field measurement, and subpixel accuracy. These techniques are widely adopted in industrial inspection [10, 11], biomechanics [12, 13], and extreme mechanics [14, 15]. Camera calibration is a key step for optical measurement techniques, which is crucial for achieving accurate image measurement. Essentially, camera calibration is the process of establishing a mapping model between the 3D control points and corresponding image points [16–19]. Both MATLAB and OpenCV have integrated calibration toolboxes [20, 21], which require users to manually select the poses of the calibration target and then optimize the camera parameters from a batch of calibration images by estimating the maximum likelihood. However, achieving robust and high-precision camera calibration requires professional calibration experience to precisely control the pose of the calibration target or the camera. It is well known that to accurately estimate the distortion coefficients and camera intrinsic parameters, the calibration target should cover as large a field of view as possible. And to avoid the degraded pose of the camera, Reimar [22] and Zhang [23] recommend varying the relative orientation between the camera and the calibration target as much as possible when acquiring calibration images. However, these experiences and studies do not intuitively guide inexperienced users in acquiring calibration images that can yield high-precision calibration results.

Some scholars have investigated methods for guiding users through the camera calibration process. Tan et al. [24] project virtual patterns onto the screen, thus avoiding the operator from moving the camera or marker, but they did not study how to control the pose relationship between the virtual pattern and the camera for efficient calibration. Tsoy et al. [25] proposed a virtual camera calibration evaluation pipeline, while the study automates pose sampling within the viewing frustum, it lacks explicit optimization of target-camera pose relationships to minimize reprojection errors, akin to the limitation noted by [26] in virtual pattern calibration. Some studies have employed the minimization of the uncertainty of the internal parameter estimation [27], geometric reprojection error [28, 29], and the uncertainty of corner points [29] as loss

functions to search for the next optimal pose, thus guiding the users. However, the methods mentioned above are all applicable only to monocular calibration. Compared to monocular systems, stereo vision systems exhibit superior depth sensitivity and measurement accuracy due to their inherent epipolar geometry and binocular disparity constraints.

In response to these problems, we propose a pose optimization algorithm and a system that guides users to move the calibration target to the optimal pose for high-precision stereo calibration. On one hand, our approach introduces a coupled optimization term for the relative and absolute extrinsic parameters, which enables joint geometric parameter estimation that surpasses the capabilities of monocular methods lacking stereometric constraints. On the other hand, while DIC technology is mature, expertise in optimizing calibration target poses for high-accuracy calibration remains limited. Our method empowers non-specialists to achieve high-precision calibration efficiently, reducing DIC’s application barrier with tangible practical value. Our contributions to this paper can be summarized as follows:

- We propose a method that solves for the next optimal pose by introducing a joint optimization for relative and absolute extrinsic parameters and adopting the minimization of the trace of the covariance matrix as the loss function.
- We present a user-friendly system for generating optimal poses tailored to high-precision stereo calibration, which guides users even without prior expertise to capture qualified calibration images via an intuitive graphical interface.
- Simulation experiments, real-world comparative experiments, and validation in thermal deformation application confirm that the proposed method is practically viable for stereo calibration, characterized by high precision and efficiency.

The remainder of the paper is organized as follows. Section 2 introduces the stereo imaging model. Section 3 describes the optimal pose guidance process and mathematical details of the proposed method. The comparative experiments and thermal deformation experiment are reported in Section 4. And section 5 draws the conclusions.

2 Stereo Imaging Model

As shown in Fig. 1, the stereo imaging model is typically composed of two cameras with a common field of view (FOV), which maps a 3D point in the real world to the left and right camera planes. A 3D control point $\mathbf{P}(X^w, Y^w, Z^w)$ in the world coordinate system can be represented in homogeneous coordinates as $\mathbf{P} = [X^w, Y^w, Z^w, 1]^T$, and it can be projected onto the left and right camera planes by the homogeneous coordinates, denoted as $\mathbf{p}^l = [u^l, v^l, 1]^T$ and $\mathbf{p}^r = [u^r, v^r, 1]^T$, respectively. The projection calculation process is as follows:

$$s^l \begin{bmatrix} u^l \\ v^l \\ 1 \end{bmatrix} = \mathbf{K}^l [\mathbf{R}^l | \mathbf{t}^l] \begin{bmatrix} X^w \\ Y^w \\ Z^w \\ 1 \end{bmatrix} \quad (1)$$

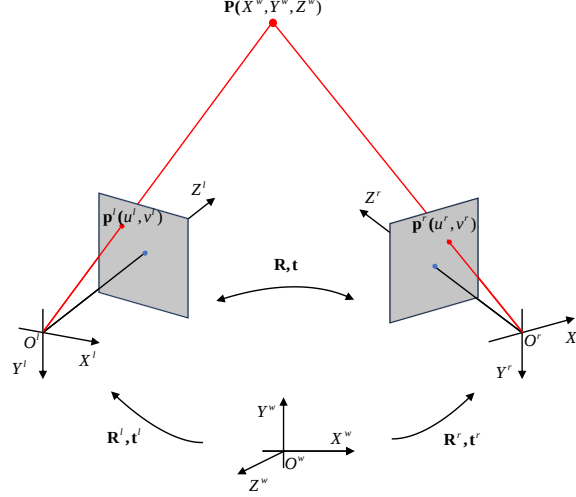


Fig. 1 Stereo imaging model

$$s^r \begin{bmatrix} u^r \\ v^r \\ 1 \end{bmatrix} = \mathbf{K}^r [\mathbf{R}^r | \mathbf{t}^r] \begin{bmatrix} X^w \\ Y^w \\ Z^w \\ 1 \end{bmatrix} \quad (2)$$

where s^l and s^r denote the depth scaling factors, which help to convert pixel coordinates in an image into actual physical units. $\mathbf{K}^l$ and $\mathbf{K}^r$ are the intrinsic parameter matrices of the cameras, which reflect the geometric structure and optical characteristics of the camera, and are used to project 3D camera coordinates onto the 2D pixel coordinates. The rotation matrices $\mathbf{R}^l$ and $\mathbf{R}^r$ and the translation vectors $\mathbf{t}^l$ and $\mathbf{t}^r$ are referred to as the absolute extrinsic parameters, which denote the rigid transformation from the world coordinate system to the camera coordinate system.

It is important to note that the left and right cameras are not entirely independent. Once the stereo measurement system is set up, the relative extrinsic parameters between the left and right cameras are fixed. The absolute extrinsic parameters of the right camera can be expressed through the absolute extrinsic parameters of the left camera and the relative extrinsic parameters:

$$\begin{cases} \mathbf{R}^r = \mathbf{R}\mathbf{R}^l \\ \mathbf{t}^r = \mathbf{R}\mathbf{t}^l + \mathbf{t} \end{cases} \quad (3)$$

where $\mathbf{R} \in SO(3)$ is the rotation matrix and $\mathbf{t} \in \mathbb{R}^3$ is the translation vector.

The stereo imaging model establishes the projective mapping between 3D control points and their corresponding undistorted 2D image points. Meanwhile, the distortion model [30] refers to the geometric distortion of objects in images due to the non-ideal characteristics of the camera lens. Specifically, such distortions are primarily divided

into two major categories: radial distortion [31] and tangential distortion [32]. The radial distortion coefficients k_1 and k_2 describe the degree of distortion along the radial direction from the image center. And the tangential distortion coefficients p_1 and p_2 describe the degree of distortion along the tangential direction.

Let us represent the distortion coefficients by a 4×1 vector $\mathbf{d}$, that is, $\mathbf{d} = [k_1, k_2, p_1, p_2]^T$. Typically, reprojection error is used to quantify calibration accuracy, which is defined as the difference between the projected image points and the actual observed image points. Suppose that we use the left and right cameras to capture the corresponding m pairs of images, and each pair of images has n pairs of corresponding corner points. On this basis, the geometric reprojection error for stereo calibration can be expressed as:

$$\min_{\mathbf{X}_i} \sum_{i=1}^m \sum_{j=1}^n \left[\rho(\|\mathbf{p}_{ij}^l - \pi(\mathbf{K}^l, \mathbf{d}^l, \mathbf{R}_i^l, \mathbf{t}_i^l, \mathbf{P}_j)\|^2) + \rho(\|\mathbf{p}_{ij}^r - \pi(\mathbf{K}^r, \mathbf{d}^r, \mathbf{R}_i^r, \mathbf{t}_i^r, \mathbf{P}_j)\|^2) \right] \quad (4)$$

where $\mathbf{X}_i = \{\mathbf{K}^l, \mathbf{d}^l, \mathbf{R}_i^l, \mathbf{t}_i^l, \mathbf{K}^r, \mathbf{d}^r, \mathbf{R}_i^r, \mathbf{t}_i^r\}$ denotes the set of parameters to be optimized, $\rho(\cdot)$ denotes the robust kernel function, and the projection equation is represented by $\pi(\cdot)$.

3 Methodology

This paper proposes a pose optimization algorithm to compute the next optimal pose, which assists users in achieving high-precision and efficient stereo calibration. Section 3.1 briefly introduces the process for guiding stereo calibration, and Section 3.2 elaborates on the mathematical formulation of the proposed method.

3.1 Stereo Calibration Guidance Process

The stereo calibration guidance process is illustrated in Fig. 2. The stereo calibration workflow is systematically divided into three modules: (1) Initial input module, where parameters of left and right cameras are determined while collecting calibration images to compute initial estimates of relative and absolute extrinsic parameters; (2) Pose output module, employing the established parameters to construct Jacobian matrices, and the optimal pose is determined by searching for the solution that minimizes the trace of the covariance matrix related to relative extrinsic parameters; and (3) Pose guidance module, where optimized poses are graphically rendered to guide the acquisition of calibration images, ultimately completing the stereo calibration.

In the initial input module, after setting up the stereo system, precise monocular calibrations are performed on the left and right cameras using a collimator [33–35], respectively. Thus, the intrinsic parameter matrices and the distortion coefficient vectors of the left and right cameras are obtained. Then, the initial stereo calibration is performed using a few (usually 2) pairs of freely captured calibration images. And the initial values of relative and absolute extrinsic parameters are calculated via PNP. These estimated parameters serve as initial input to the proposed algorithm.

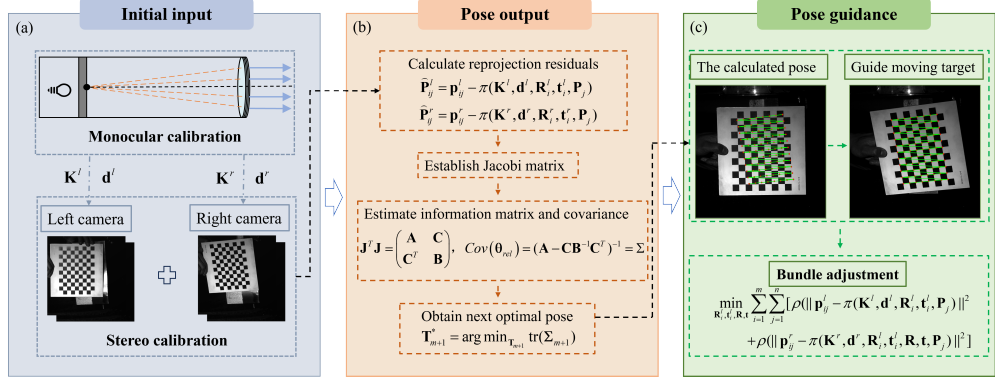


Fig. 2 Flowchart of the process of guiding the stereo calibration. (a) Initial input module, (b) pose output module, and (c) pose guidance module

In the pose output module, the reprojection residuals are computed. These residuals quantify the discrepancy between the projected image points and the observed image points. The Jacobian matrix is established by taking the partial derivatives of the reprojection residuals with respect to the estimated parameters. Additionally, the information matrix and the covariance matrix for estimating relative and absolute extrinsic parameters are constructed. A set of poses is searched in 3D space, and when the trace of the covariance matrix associated with relative extrinsic parameters reaches a threshold, this pose is referred to as the optimal pose.

In the pose guidance module, after calculating the next optimal pose, the system guides the user to complete the stereo calibration through an interactive interface. As shown in Fig. 2(c), the system displays the corner coordinates corresponding to the next optimal pose within the frame, and users move the calibration target to the designated position according to the pose guidance. Once the target is positioned correctly, users capture a synchronized pair of images of the calibration target at this pose. These images are added to the calibration dataset for refining the relative extrinsic parameters. Iterate the aforementioned procedures until the predetermined calibration accuracy is attained.

3.2 Caculate Optimal Pose

The process of calculating optimal pose is summarized in Algorithm 1. First, the intrinsic parameter matrices K^l, K^r and the distortion coefficient vectors d^l, d^r of the left and right cameras are calibrated using a collimator. Consequently, the optimization parameter in Eq. (4) can be simplified to $X_i = \{R_i^l, t_i^l, R_i^r, t_i^r\}$, where $R_i^r = RR_i^l$ and $t_i^r = Rt_i^l + t$. In a well-configured stereo measurement system, the relative extrinsic parameters between the left and right cameras are considered to be constant. The goal of our work is to make the relative extrinsic parameter estimation more accurate by adjusting the pose of the calibration target, which is named the optimal pose.

Subsequently, several pairs of calibration images are captured, and the absolute extrinsic parameters of the left and right cameras are calculated using the PNP. On this

basis, the initial value of relative extrinsic parameters can be solved using the variant of Eq. (3). Except for camera parameters, both the absolute extrinsic parameters of the left camera and the relative extrinsic parameters serve as initial inputs to the algorithm.

Let us construct the Jacobian matrix $\mathbf{J}$ composed of partial derivatives of reprojection residuals with respect to relative and absolute extrinsic parameters. For notational simplicity, we define the reprojection residuals of the left and right cameras as:

$$\begin{cases} \widehat{\mathbf{p}}_{ij}^l = \mathbf{p}_{ij}^l - \pi(\mathbf{K}^l, \mathbf{d}^l, \mathbf{R}_i^l, \mathbf{t}_i^l, \mathbf{P}_j) \\ \widehat{\mathbf{p}}_{ij}^r = \mathbf{p}_{ij}^r - \pi(\mathbf{K}^r, \mathbf{d}^r, \mathbf{R}\mathbf{R}_i^l, \mathbf{R}\mathbf{t}_i^l + \mathbf{t}, \mathbf{P}_j) \end{cases} \quad (5)$$

where $\widehat{\mathbf{p}}_{ij}^l$ and $\widehat{\mathbf{p}}_{ij}^r$ represent the reprojection residuals of the j -th image point of the i -th image, respectively. Inspired by Peng et al. [29], we derive the Jacobian matrices of the reprojection residuals as follows:

$$\mathbf{U}_i = \begin{pmatrix} \frac{\partial \widehat{\mathbf{p}}_{i1}^r}{\partial \mathbf{R}} & \frac{\partial \widehat{\mathbf{p}}_{i1}^r}{\partial \mathbf{t}} \\ \vdots & \vdots \\ \frac{\partial \widehat{\mathbf{p}}_{in}^r}{\partial \mathbf{R}} & \frac{\partial \widehat{\mathbf{p}}_{in}^r}{\partial \mathbf{t}} \end{pmatrix} \quad \mathbf{V}_i = \begin{pmatrix} \frac{\partial \widehat{\mathbf{p}}_{i1}^l}{\partial \mathbf{R}_i^l} & \frac{\partial \widehat{\mathbf{p}}_{i1}^l}{\partial \mathbf{t}_i^l} \\ \vdots & \vdots \\ \frac{\partial \widehat{\mathbf{p}}_{in}^l}{\partial \mathbf{R}_i^l} & \frac{\partial \widehat{\mathbf{p}}_{in}^l}{\partial \mathbf{t}_i^l} \end{pmatrix} \quad (6)$$

Specifically, $\mathbf{U}_i$ denotes the right camera's reprojection residuals with respect to the relative extrinsic parameters, while $\mathbf{V}_i$ represents the left camera's reprojection residuals with respect to its absolute extrinsic parameters. The explicit expressions of matrix blocks $\mathbf{U}_i$ and $\mathbf{V}_i$ are provided in Appendix A.

Based on this formulation, we construct a highly sparse Jacobian matrix that simultaneously encodes the relationships between the absolute extrinsic parameters of the left camera and the relative extrinsic parameters:

$$\mathbf{J} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{V}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{U}_2 & \mathbf{0} & \mathbf{V}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{U}_m & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{V}_m \end{pmatrix} \quad (7)$$

Next, establish the information matrix $\mathbf{J}^T \mathbf{J}$, which can be expressed by block matrix as:

$$\mathbf{J}^T \mathbf{J} = \begin{pmatrix} \sum_{i=1}^m \mathbf{U}_i^T \mathbf{U}_i & \mathbf{U}_1^T \mathbf{V}_1 & \mathbf{U}_2^T \mathbf{V}_2 & \cdots & \mathbf{U}_m^T \mathbf{V}_m \\ \mathbf{V}_1^T \mathbf{U}_1 & \mathbf{V}_1^T \mathbf{V}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{V}_2^T \mathbf{U}_2 & \mathbf{0} & \mathbf{V}_2^T \mathbf{V}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{V}_m^T \mathbf{U}_m & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{V}_m^T \mathbf{V}_m \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{B} \end{pmatrix} \quad (8)$$

Algorithm 1 Optimal pose calculation algorithm

Input: $\{\mathbf{K}^l, \mathbf{K}^r\}$ – camera intrinsic matrix; $\{\mathbf{d}^l, \mathbf{d}^r\}$ – distortion coefficients.**Output:** $\mathbf{R}^l$ – optimal rotation vector and $\mathbf{t}^l$ – optimal translation vector.

- 1: Capture m pairs of calibration images to compute $\{\mathbf{R}_m^l, \mathbf{t}_m^l, \mathbf{R}_m^r, \mathbf{t}_m^r\}$ via PnP.
 - 2: Calculate the initial value of $\{\mathbf{R}, \mathbf{t}\}$.
 - 3: **for** $i = 1$ **to** m **do**
 - 4: Calculate the Jacobian matrix $\mathbf{J}_m$ using Eq.(5) – Eq.(7).
 - 5: **end for**
 - 6: **while** $\Delta RE > 1e-6$ **and** iterations < 500 **do**
 - 7: Generate candidate pose $\{\mathbf{R}_{m+1}^l, \mathbf{t}_{m+1}^l\}$.
 - 8: **if** The target in the candidate pose is in the common FOV **then**
 - 9: Update the Jacobian matrix to $\mathbf{J}_{m+1}$ using Eq.(5) – Eq.(7).
 - 10: Calculate the trace of the covariance matrix $\text{tr}(\Sigma_{m+1})$.
 - 11: Calculate the optimal pose using Eq.(10).
 - 12: **end if**
 - 13: **end while**
 - 14: Return next optimal pose $\mathbf{R}_{m+1}^l, \mathbf{t}_{m+1}^l$.
-

with

$$\begin{cases} \mathbf{A} = \sum_i \mathbf{U}_i^T \mathbf{U}_i \\ \mathbf{B} = \text{diag}(\mathbf{V}_1^T \mathbf{V}_1, \dots, \mathbf{V}_m^T \mathbf{V}_m) \\ \mathbf{C} = (\mathbf{U}_1^T \mathbf{V}_1, \dots, \mathbf{U}_m^T \mathbf{V}_m) \end{cases} \quad (9)$$

The inverse of the information matrix $(\mathbf{J}^T \mathbf{J})^{-1}$ provides an estimate of the covariance matrix for both the absolute and relative extrinsic parameters. In particular, the upper-left submatrix of $(\mathbf{J}^T \mathbf{J})^{-1}$ can be used to estimate the covariance of relative extrinsic parameters, which is expressed as $\text{Cov}(\theta_{\text{rel}}) = (\mathbf{A} - \mathbf{CB}^{-1}\mathbf{C}^T)^{-1} = \Sigma$. Relevant proofs are provided in Appendix B.

From this, we proceed to compute the next optimal pose. Specifically, we first establish a set of absolute extrinsic parameters of the left camera in 3D space (i.e., the pose of the calibration board relative to the left camera) and set the initial value of the relative extrinsic parameters as the ground truth. When a calibration board corresponding to a specific absolute extrinsic parameter lies within the common FOV, this absolute extrinsic parameter is regarded as a candidate for the next optimal pose. This optimization problem can be formulated as:

$$\mathbf{T}_{m+1}^* = \arg \min_{\mathbf{T}_{m+1}} \text{tr}(\Sigma_{m+1}) \quad (10)$$

The above process is executed iteratively, with termination conditions: the number of iterations reaches or exceeds 500, or the relative error of the trace is less than or

equal to 10^{-6} . Thus, we obtain the next optimal pose. Subsequently, a visualization-based guidance interface is employed to assist the user in capturing calibration images. Ultimately, the bundle adjustment approach is applied to jointly optimize the system parameters and the image coordinates, thereby completing the stereo calibration of the binocular camera system.

4 Experiments and Application

We have compared our method with independent monocular calibration and stereo-specialized calibration methods [18]. The experimental procedures and results are presented in Appendix C and Appendix D, respectively. And to quantitatively evaluate the effectiveness of the proposed method, we conducted comparative experiments between the optimal pose derived from our optimization framework and the constrained random pose. Note that the random pose is not completely random, which is carefully designed to avoid degenerate configurations. Inspired by [22, 23, 28], the random pose follows the following principles:

1. **Full-FOV coverage constraint:** Throughout the calibration sequence, the composite projection of all calibration target positions must cover more than 90% of the overlapping FOV between both cameras to ensure complete parameter observability.
2. **Optical axis non-alignment constraint:** The calibration target’s surface normal vector must not be parallel to the optical axes of either camera to ensure adequate depth observability and prevent degenerate configurations in stereo vision.
3. **Sampling constraint:** Random poses are sampled within the SE(3) space, where rotational parameters follow a random distribution within $\pm 30^\circ$, and the calibration target is randomly moved within the overlapping FOV.

4.1 Simulation Evaluations

To verify the robustness of the method proposed in this paper, Gaussian noise with different standard deviations was introduced in the simulation experiments. The experimental hardware platform used in this paper is configured with an NVIDIA GeForce RTX 3050 Laptop GPU, featuring a core frequency of 3.3 GHz and 16 GB of RAM. Here, we briefly introduce the simulation experiment settings.

Specifically, the calibration target adopted in the experiment is a 9×6 2D chessboard with a grid spacing of 5 mm, and the origin of the world coordinate system is set on the chessboard plane. Based on the aforementioned principles, a set of random poses for the absolute extrinsic parameters of the left camera is defined, and the analysis is conducted by comparing the calibration results of the optimal pose and random poses. In addition, the parameters of the left and right cameras are consistent, with specific parameters as follows: resolution of 640×480 pixels, principal point coordinates of (320, 240), focal length of 800, and distortion coefficients of $[0.01, 0.1, 0, 0]^T$, the relative extrinsic parameters are $\mathbf{R} = [-0.003, -0.303, -0.017]^T$ and $\mathbf{t} = [440.3, -6.2, 25.1]^T$.

The experimental process is as follows:

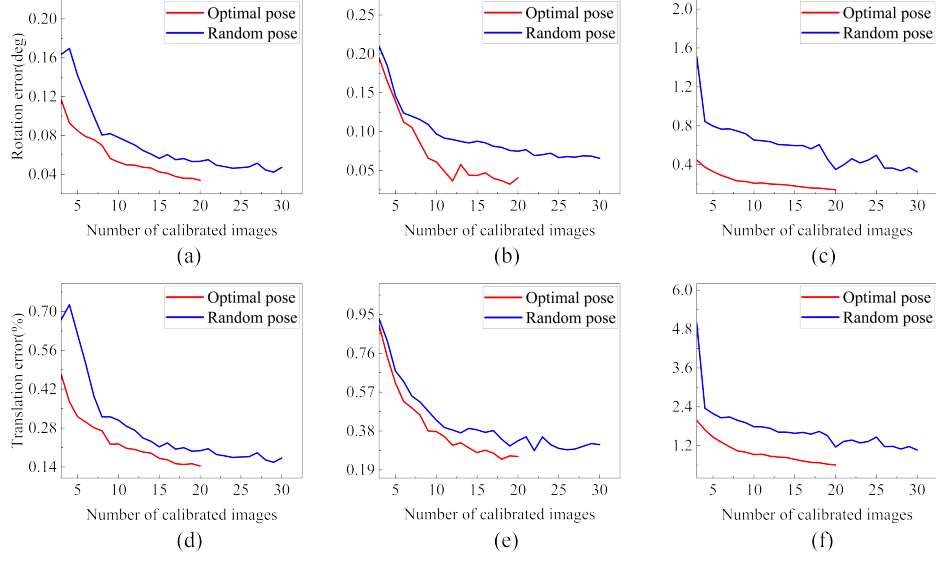


Fig. 3 Comparison of relative extrinsic parameter estimation errors between random poses and optimized poses under Gaussian noise with different standard deviations. (a) Rotation error, and (d) translation error under Gaussian noise with a standard deviation of 0.5 pixel; (b) rotation error, and (e) translation error under Gaussian noise with a standard deviation of 1 pixel; (c) rotation error, and (f) translation error under Gaussian noise with a standard deviation of 2 pixel

First, two pairs of random images in the common FOV are generated. After adding Gaussian noise to the image points, the initial values of the relative extrinsic parameters are solved. These two pairs of random images serve as calibration images for both the random pose method and the optimized pose method. Subsequently, 28 images are generated according to the definition of random poses, resulting in a total of $28 + 2 = 30$ calibration images. Meanwhile, 18 images are generated through optimal pose, yielding a total of $18 + 2 = 20$ calibration images. For Gaussian noise with the same standard deviation in the same group of experiments, three standard deviation levels are set, that is, 0.5 pixel, 1 pixel, and 2 pixels. The relative extrinsic parameters are solved by inputting the calibration images. 100 trials are performed for each experiment.

To evaluate the calibration results, the calculated relative extrinsic parameters are compared with the ground truth for error assessment, and the specific error calculation formulas are as follows:

$$E(\mathbf{R}_{\text{cal}}) = \arccos\left(\frac{\text{tr}(\mathbf{R}_{\text{ref}}\mathbf{R}_{\text{cal}}^{\top}) - 1}{2}\right) \cdot \frac{180}{\pi} \quad (11)$$

$$E(\mathbf{t}_{\text{cal}}) = 2 \left(\frac{\|\mathbf{t}_{\text{ref}} - \mathbf{t}_{\text{cal}}\|}{\|\mathbf{t}_{\text{ref}}\| + \|\mathbf{t}_{\text{cal}}\|} \right) \cdot 100\% \quad (12)$$

where $(\mathbf{R}_{\text{ref}}, \mathbf{t}_{\text{ref}})$ denotes the ground truth, and $(\mathbf{R}_{\text{cal}}, \mathbf{t}_{\text{cal}})$ represents the calculated result. As illustrated in Fig. 3, it can be observed that, when an appropriate number of calibration images are employed, the calculated value of the relative extrinsic parameter gradually converges to the ground truth with the increase in the number of calibration images. In comparison with the random pose strategy, the relative extrinsic parameter calibrated via the optimal pose exhibits a higher consistency with the ground truth and converges at a faster rate, thus validating the effectiveness of the proposed method. Additionally, when Gaussian noise with standard deviations of 0.5 pixel, 1 pixel, and 2 pixels is introduced, the calibration result based on the optimal pose still maintains a rapid convergence speed, further corroborating the robustness of the proposed method.

4.2 Comparative Experiment on Calibration Efficiency

The experimental procedure was designed as follows. Initially, both cameras were rigidly mounted on a stereoscopic rig with adjustable baselines to form a stereo vision system, ensuring that they had an appropriate overlapping FOV. And individual calibration of the left and right cameras was conducted to obtain precise intrinsic parameters and distortion coefficients. Subsequently, an operator manipulated the calibration target in accordance with the above principles of random pose and captured 20 calibration images. We conducted stereo calibration with selected sets of image pairs (e.g., 2, 10, and 20 pairs). And the stereo calibration was completed through the optimization of the minimal reprojection error, which is widely recognized and employed in practical applications. To ensure comparability, we selected the first two images captured with the aforementioned random poses as the initial input for calculating the optimal pose, and then captured the calibration images under the guidance of our proposed method until the reprojection error reached that of random-pose-based calibration.

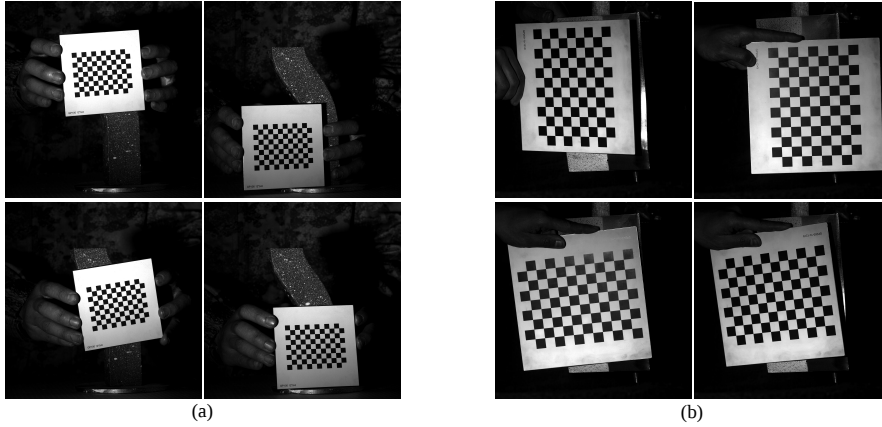


Fig. 4 Sample calibration images in the comparative experiments. **(a)** The calibration target approximately occupied 1/9 of the common FOV, and **(b)** the calibration target approximately occupied 1/2 of the common FOV

Furthermore, to verify the robustness of the proposed method, two distinct experimental conditions were employed to conduct the comparative experiments. The configuration of the stereo vision system varied with different experimental conditions, while the parameters of the two cameras were approximately the same, with the FOVs of the two systems roughly between 20° and 40° . As shown in Fig. 4(a), the difference in the comparative experiments lay in the proportion of the calibration target within the common FOV, which was used to test the performance of the proposed method. In the low-coverage case (*l*-system), the calibration target approximately occupied 1/9 of the overlapping FOV in terms of area (Fig. 4(a)). And in the high-coverage case (*h*-system), the calibration target occupied approximately 1/2 of the overlapping FOV in terms of area (Fig. 4(b)).

Table 1 The experimental results for the *l* system

Calibration scheme	$\mathbf{R}/\text{deg}$	$\mathbf{t}/\text{mm}$	$\mathbf{RE}/\text{pixel}$	$\mathbf{TE}/\text{mm}$
2-random	$(-1.350, -18.698, -0.515)$	$(-376.5, 12.8, 67.7)$	0.6179	0.0973
10-random	$(-0.592, -19.528, -0.517)$	$(-390.9, -1.9, 75.0)$	0.3587	0.0894
20-random	$(-0.382, -19.465, -0.503)$	$(-385.0, -5.8, 72.0)$	0.2808	0.0739
2-random + 2-optimal	$(-0.969, -18.960, -0.512)$	$(-381.2, 5.4, 69.0)$	0.4150	0.0846
2-random + 4-optimal	$(-0.474, -19.127, -0.526)$	$(-384.0, -4.1, 71.3)$	0.3323	0.0690
2-random + 6-optimal	$(-0.388, -19.186, -0.532)$	$(-385.0, -5.8, 72.0)$	0.2715	0.0676

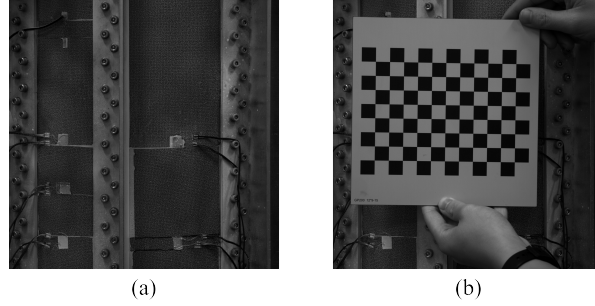
Note: The table presents reprojection errors ($\mathbf{RE}$) in pixels, and triangulation errors ($\mathbf{TE}$) in millimeters for different calibration schemes. For clarity, $\mathbf{R}$ in this paper is expressed in vector

To comprehensively evaluate the calibration accuracy, we employed both the reprojection error and the triangulation error as quantitative metrics. The reprojection error was determined by calculating the Euclidean distance between the image point coordinates predicted by the projection equations and the actual extracted image point coordinates. The triangulation error was calculated by performing spatial intersection using corresponding pixel points from the left and right image planes, based on the calibration results, to compute the 3D point coordinates. These computed 3D coordinates were then compared with the 3D coordinates of the calibration target by calculating their Euclidean distance. This dual-metric approach provides complementary perspectives on calibration accuracy, assessing both the 2D image plane consistency and 3D spatial reconstruction fidelity.

Table 1 presents the experimental results when the calibration target occupies 1/9 of the FOV. It demonstrates that the optimal pose significantly outperforms random pose. When comparing 10-random with 2-random + 4-optimal configurations in the *l*-system, the proposed approach reduced reprojection error by 7.4% (from 0.3587 to 0.3323 pixels) and triangulation error by 22.8% (from 0.0894 to 0.0690 mm). The comparison between 20-random and 2-random + 6-optimal configurations showed slightly diminished but still meaningful gains, with reprojection error improving by 3.3% (from 0.2808 to 0.2715 pixels) and triangulation error by 8.5% (from 0.0739 to 0.0676 mm) in the *l*-system.

Table 2 The experimental results for the h system

Calibration scheme	$\mathbf{R}/\text{deg}$	$\mathbf{t}/\text{mm}$	$\mathbf{RE}/\text{pixel}$	$\mathbf{TE}/\text{mm}$
2-random	(−0.069, 17.578, 1.073)	(438.8, −3.5, 31.2)	0.1582	0.0551
10-random	(0.038, 17.587, 1.061)	(438.9, −6.0, 31.2)	0.1417	0.0487
20-random	(0.061, 17.621, 1.051)	(439.7, −6.5, 31.6)	0.1222	0.0455
2-random + 2-optimal	(−0.002, 17.632, 1.059)	(440.0, −5.0, 31.6)	0.1194	0.0476
2-random + 4-optimal	(0.047, 17.614, 1.059)	(439.5, −6.2, 31.6)	0.1239	0.0474
2-random + 6-optimal	(0.049, 17.615, 1.056)	(439.6, −6.3, 31.6)	0.1215	0.0473

**Fig. 5** Comparative experiment on calibration accuracy. (a) Experimental setup, and (b) sample of the calibration images

Similarly, Table 2 presents the experimental results when the calibration target occupies $1/2$ of the FOV. The data demonstrate that the 2-random + 2-optimal configuration achieves convergence, with the reprojection error reduced to 0.1194 pixels, which is superior to the 20-random approach (0.1222 pixels). Error distribution analysis confirms this convergence state, showing that subsequent increases to 4 or 6 optimized poses produce only minimal variations in both reprojection error (within ± 0.0025 pixels) and triangulation error (within ± 0.0003 mm).

As evidenced by Table 1 and Table 2, the optimized pose achieves comparable calibration accuracy to the random pose, while simultaneously reducing the required number of calibration images. Additionally, the method maintains robustness across different experimental conditions, demonstrating its practical applicability.

4.3 Comparative Experiment on Calibration Accuracy

Furthermore, we quantitatively compared the calibration accuracy between the random pose and the optimal pose using the same number of calibration images, and evaluated the calibration results based on strain measurement data. Figure 5(a) shows the experimental setup, a cylindrical specimen is prepared with randomly distributed water-transfer speckle patterns and strategically mounted strain gauges, whose measurements served as the ground truth for strain verification. Figure 5(b) presents a sample of the calibration images, where the checkerboard occupies approximately $1/2$ of the FOV. We established a rigid stereo vision system through two IDS industrial

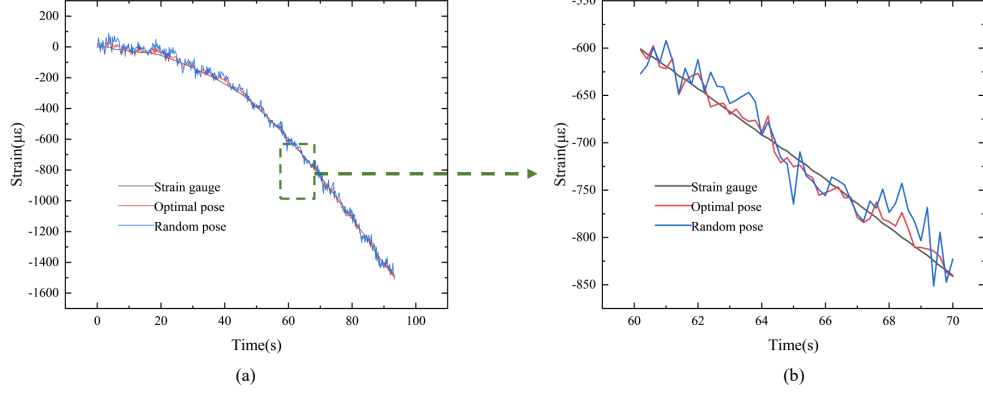


Fig. 6 Comparative analysis of ε_{yy} strain distributions. **(a)** Macroscopic strain distribution under full loading history, and **(b)** local strain concentration

Table 3 The experimental results of stereo calibration

Calibration scheme	R/deg	t/mm	RE/pixel
8-random	(0.407, -17.614, -1.118)	(439.3 - 4.5, 38.2)	0.0531
2-random + 6-optimal	(0.601, -18.164, -1.151)	(438.2, -1.1, 48.4)	0.0471

cameras (resolution: 2048×2048 pixels). Based on the data in Table 2, we compared the stereo calibration results between the 8-random configuration and the 2-random + 6-optimal configuration. The results of stereo calibration are presented in Table 3.

The experiment conducted was an axial compression test, where a constant loading rate was applied to the cylindrical specimen until reaching the target load of 80 metric tons. Throughout the process, the stereo vision system captured deformation images of the specimen at 5 fps. The stereo calibration results obtained from random poses and optimal poses were integrated with DIC software to compute the specimen's strain fields, respectively. After synchronizing the sampling frequencies of the strain measurement device and the stereo vision system, we performed the analysis of their strain data.

Figure 6 presents the longitudinal strain ε_{yy} of the specimen. Figure 6(a) shows the comparative strain diagram during the entire loading process, while Fig. 6(b) provides a segmental strain comparison for enhanced visualization. As shown in Fig. 6(b), a qualitative analysis reveals that, using strain gauge measurements as ground truth, the optimal-pose calibration yields strain values that are closer to the ground truth with less variability compared to random-pose calibration.

As shown in Fig. 7, post-processing was applied to the strain data. The strain error, defined as the differences between the calculated values (obtained from both optimal and random pose calibrations) and the strain gauge measurements, was computed and visualized using a box diagram. The strain error between the calculated values from

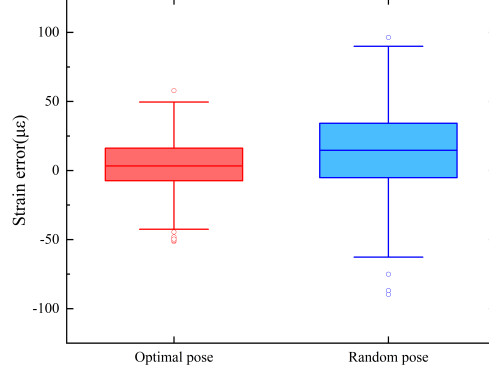


Fig. 7 Box diagram of longitudinal strain ϵ_{yy}

optimal pose calibration and strain gauge measurements had a mean of 3.3 $\mu\epsilon$ with a standard deviation of 18.2 $\mu\epsilon$, whereas the residuals from random pose calibration showed a mean of 14.7 $\mu\epsilon$ and a standard deviation of 30.2 $\mu\epsilon$. The box diagram and local strain analysis further confirmed that optimal pose reduces measurement variance and enhances strain field reconstruction accuracy.

4.4 Application of Thermal Deformation Measurement

Investigating the deformation characteristics of the S-shaped inlet duct of a hypersonic vehicle under high-temperature conditions is of significant importance for its rational design. However, constrained by current experimental conditions (e.g., limited operating space), operators must complete calibration tasks rapidly with a minimal set of calibration images. To address this, we apply the proposed method to this task. We combined the proposed method with DIC to measure the in-plane displacement of the specimen under high-temperature conditions and compared it with the numerical simulation results.

An S-shaped specimen was designed in this experiment. The geometry of the S-shaped specimen was identical to that of the S-shaped bend in the constant cross-sectional intake duct of hypersonic vehicles. A rigid stereo vision system was established by mounting two IDS industrial cameras (resolution: 2048×2048 pixels) on a stabilized tripod. The thermal deformation measurement experiment was conducted using a heating cabinet, with heating controlled by thermocouples. To mitigate thermal radiation at high temperatures, a high-power blue light source was employed, and narrow bandpass filters with a central wavelength of 450 ± 10 nm were installed on the left and right cameras. To overcome the limitations of conventional speckle patterns in high-temperature environments, we applied a specialized high-temperature-resistant speckle pattern using thermally stable ceramic-based spray paint, ensuring reliable pattern stability for accurate DIC measurements throughout the thermal loading cycle.

We completed the stereo calibration using the 2-random + 5-optimal poses and implemented the thermal deformation measurement task in conjunction with DIC software. The calibration results of the stereo vision system are shown in Table 4.

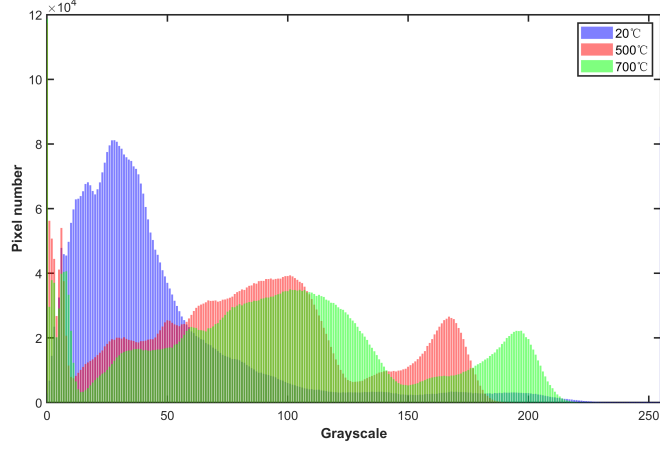


Fig. 8 Comparison of grayscale histograms of speckle patterns at 20°C, 500°C and 700°C

Table 4 Stereo camera calibration results

Parameter	Left camera	Right camera
$(f_u, f_v)/\text{pixel}$	(10221.8, 10191.7)	(10063.6, 10079.5)
$(u_0, v_0)/\text{pixel}$	(965.8, 974.0)	(904.3, 943.1)
$(k_1, k_2)/\text{pixel}$	(0.330, 30.356)	(0.403, -0.313)
$(p_1, p_2)/\text{pixel}$	(-0.012, 0.002)	(-0.012, 0.006)
$\mathbf{R}/\text{rad}$	(-0.174, 8.314, 1.023)	
$\mathbf{t}/\text{mm}$	(-80.0, 1.1, 9.4)	
RE/pixel	0.1011	

(f_u, f_v) represents the focal length, (u_0, v_0) refers to the principal point, which are all camera intrinsic parameters, and the mean reprojection error is merely 0.1011 pixels.

In this experiment, two target temperatures were set as 500°C and 700°C, respectively. To ensure that the measurement results accurately reflect the deformation at each temperature, once the temperature first reached the set point, it was maintained at that level for 20 minutes, during which the camera continuously captured images. For deformation analysis, we selected the frame captured at the 20-minute mark as the reference image for DIC computation, representing the steady-state thermal deformation at each target temperature.

To verify the stability of speckle patterns under the high-temperature conditions established in this study, we evaluated the speckle patterns at 20°C, 500°C, and 700°C using three key metrics [36, 37]. Fig. 8 shows the distribution of pixel numbers across various grayscale values for speckle patterns captured at 20°C, 500°C, and 700°C. The histograms reveal a shift in brightness and the presence of distinct peaks at low and high grayscale ranges, indicating good contrast and texture information suitable for DIC analysis. To further evaluate image quality, we analyze the mean intensity (MI)

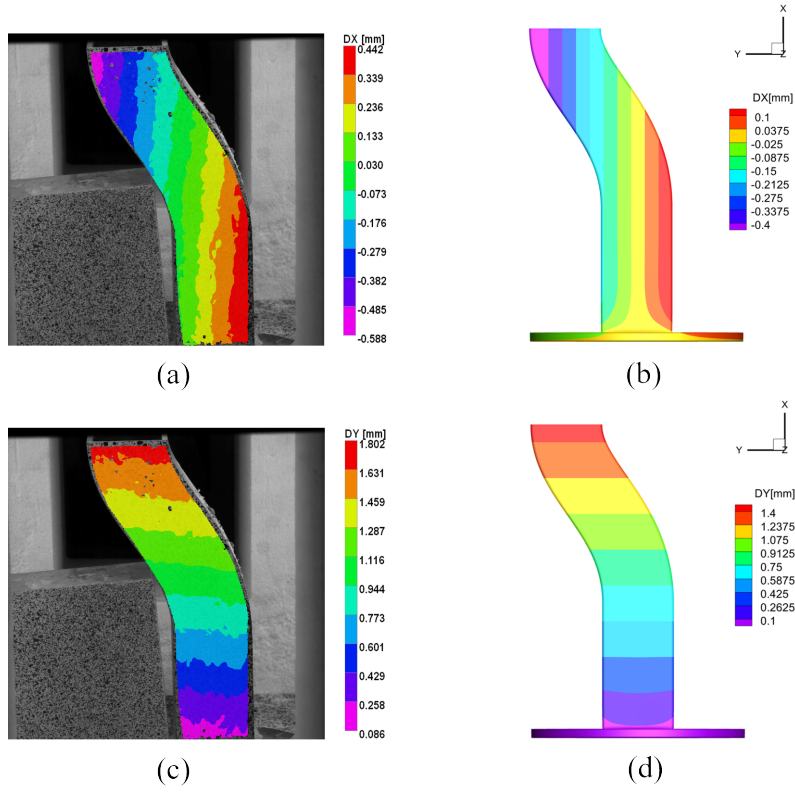


Fig. 9 Thermal deformation measurement results at 500°C. **(a)** Actual displacement U, **(b)** simulated displacement U, **(c)** actual displacement V, and **(d)** simulated displacement V

Table 5 Comparative analysis of MI and MIG at 20°C, 500°C and 700°C

Statistical parameters	20°C	500°C	700°C
MI	50.43	81.48	95.38
MIG	6.45	6.85	6.87

and mean intensity gradient (MIG). As shown in Table 5, MI increases with temperature due to thermal radiation, yet no pixel saturation or overexposure is observed. MIG values remain comparable across all temperatures and slightly increase at 700°C, indicating maintained image quality and suitability for DIC.

To validate the measurement results, we computed the displacement field of the specimen using the open-source FEA software MOOSE, with post-processing visualization performed in Tecplot 360. The S-bend specimen was fabricated from 310S stainless steel, with temperature-dependent material properties carefully selected for the simulation: Young's modulus values are 163 GPa and 146 GPa, and the coefficients of thermal expansion are $17.0 \times 10^{-6} \text{ }^{\circ}\text{C}^{-1}$ and $18.0 \times 10^{-6} \text{ }^{\circ}\text{C}^{-1}$ at 500°C and 700°C,

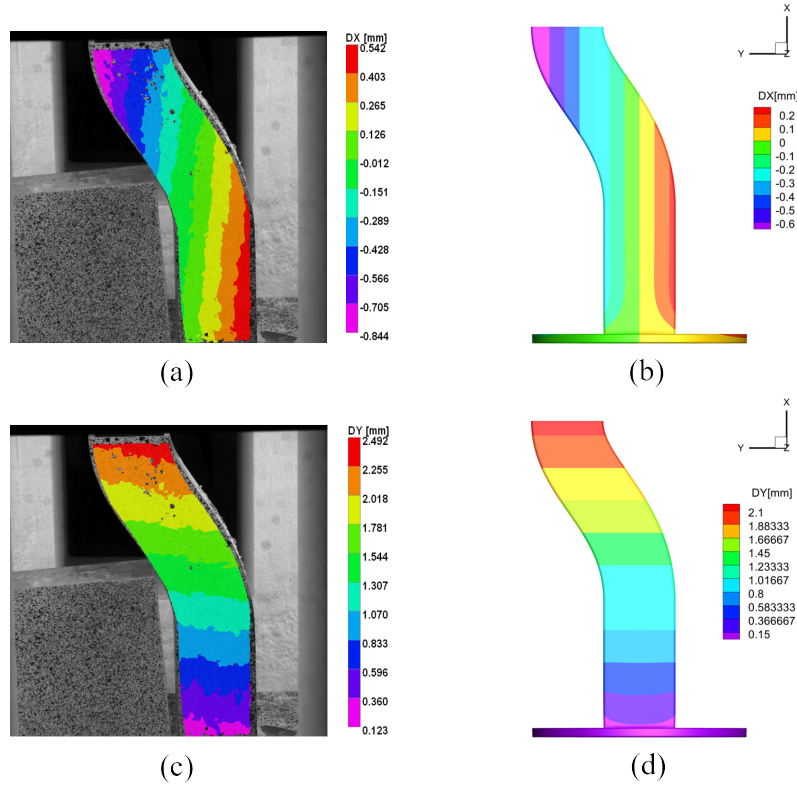


Fig. 10 Thermal deformation measurement results at 700°C. (a) Actual displacement U, (b) simulated displacement U, (c) actual displacement V, and (d) simulated displacement V

respectively. These values, along with all other material parameters, were sourced from established literature [38].

To enable quantitative visual comparison between experimental measurements and numerical simulations, we standardized the colormap ranges and scaling for both datasets. As shown in Fig. 9 and Fig. 10, the quantitative comparison between the experimental measurements and numerical simulations reveals consistent agreement in both deformation magnitude and evolutionary trends. Experimental results demonstrate consistent deformation behavior at both 500°C and 700°C. Figures 9(a) and (b), figures 10(a) and (b) illustrate the horizontal displacement U, showing thermal expansion along the specimen centerline with negative values indicating expansion opposite to the defined coordinate system positive direction. Figures 9(c) and (d), and figures 10(c) and (d) present the vertical displacement V, revealing monotonic increase in upward thermal expansion magnitude with elevation from the fixed base, following a characteristic gradient pattern typical of thermally induced bending. The thermal deformation measurement results of the S-shaped specimen highlight the significant potential of the proposed method in engineering 3D deformation measurement.

5 Conclusion

In this paper, we present a pose guidance method for stereo calibration, which guides users to complete high-precision stereo calibration through a graphical user interface. The method solves for the next optimal pose by introducing a joint optimization for relative and absolute extrinsic parameters and adopting the minimization of the trace of the covariance matrix as the loss function. The experimental results demonstrate that the proposed pose guidance method significantly enhances the precision and efficiency of stereo calibration in 3D deformation measurement.

In experiments, this study systematically evaluates calibration efficiency and accuracy through comparative experiments on pose selection strategies. Simulation experiments have verified the robustness of this method. The calibration efficiency test demonstrates that our optimal pose method achieves comparable accuracy to random pose while requiring fewer calibration images and maintaining robustness across varying FOVs. In the accuracy validation of using a cylindrical specimen under 80-ton axial compression, the calibration of our method (2-random + 6-optimal) outperformed that of random pose (8-random), yielding lower reprojection error (0.0471 vs. 0.0531 pixels) and superior strain measurement precision ($3.3 \pm 18.2 \mu\epsilon$ vs. $14.7 \pm 30.2 \mu\epsilon$). This dual improvement in both precision and efficiency validates the practical efficacy of the proposed optimization framework.

Finally, the proposed method, in conjunction with DIC, has been successfully applied to thermal deformation measurement tasks. By spraying a high-temperature-resistant speckle pattern on the surface of an S-shaped specimen and utilizing the stereo calibration method proposed in this paper, deformation measurement under high-temperature conditions was achieved. The measurement results show good agreement with the simulation results obtained from the software, further highlighting the significant potential in engineering deformation measurement. The image distortion caused by heat haze is regarded as the primary source of error in our experiment. This limitation highlights a key challenge in the current experimental setup. Future work will focus on optimizing the system configuration and developing image processing techniques to enhance measurement stability under high-temperature conditions.

Acknowledgements. This work is supported by Hunan Provincial Natural Science Foundation for Excellent Young Scholars (Grant 2023JJ20045), National Natural Science Foundation of China (Grant 12372189) and the Science and Technology Innovation Program of Hunan Province(2022RC1196).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

- [1] Sutton MA, Orteu JJ, Schreier H (2009) Image correlation for shape, motion and deformation measurements: basic concepts, theory and applications. Springer Science & Business Media, <https://doi.org/10.1007/978-0-387-78747-3>

- [2] Ma K, Liu ZQ, Cai YK, et al (2025) In-situ dic characterization of dislocation density and stored strain energy fields for deformation zones during cutting of ti6al4v alloy. *Measurement* 250:117197. <https://doi.org/10.1016/j.measurement.2025.117197>
- [3] Dong YL, Zhao JQ, Pan B (2020) Ultraviolet 3d digital image correlation applied for deformation measurement in thermal testing with infrared quartz lamps. *Chin J Aeronaut* 33(3):1085–1092. <https://doi.org/10.1016/j.cja.2019.03.038>
- [4] Wang YK, Ye F, Li J, et al (2021) Underwater fuel assembly deformation measurement with multisensor dual-line structured light. *Appl Opt* 60(31):9624–9633. <https://doi.org/10.1364/AO.438074>
- [5] Dong NT, Ninh DG (2023) Stability and resolution of a conventional displacement measuring heterodyne interferometer using a single phase-locked loop. *Exp Mech* 63(6):1015–1032. <https://doi.org/10.1007/s11340-023-00970-x>
- [6] Yuan K, Zhu WD (2022) In-plane operating deflection shape measurement of an aluminum plate using a three-dimensional continuously scanning laser doppler vibrometer system. *Exp Mech* 62(4):667–676. <https://doi.org/10.1007/s11340-021-00801-x>
- [7] Xi CJ, Liu SJ, Chang SP, et al (2025) Enhancing accuracy of line-structured light sensor for narrow weld gap measurement. *IEEE Trans Instrum Meas* 74:1–10. <https://doi.org/10.1109/TIM.2025.3553256>
- [8] Liu ZB, Guan BL, Shang Y, et al (2025) Stereo event-based, 6-dof pose tracking for uncooperative spacecraft. *IEEE Trans Geosci Remote Sens* 63:1–13. <https://doi.org/10.1109/TGRS.2025.3530915>
- [9] Lei TH, Guan BL, Liang MZ, et al (2025) Event-based multi-view photogrammetry for high-dynamic, high-velocity target measurement. *Acta Mechanica Sinica* <https://doi.org/10.1007/s10409-025-25314-x>
- [10] Wei K, Wei G, Shao XX, et al (2024) An experimental method for stereodigital measurement of large-scale thin-film structures. *Exp Mech* 64(7):1175–1181. <https://doi.org/10.1007/s11340-024-01079-5>
- [11] Yuan F, Ren X, Pan R, et al (2023) Tensile construction monitoring and progressive collapse test of suspen-dome structure based on uav-assisted close-range photogrammetry and multi-camera stereo-digital image correlation. *Exp Mech* 63(8):1371–1389. <https://doi.org/10.1007/s11340-023-00993-4>
- [12] Clark JN, Tavana S, Clark B, et al (2021) High resolution three-dimensional strain measurements in human articular cartilage. *J Mech Behav Biomed Mater* 124:104806. <https://doi.org/10.1016/j.jmbbm.2021.104806>

- [13] Lian WJ, Chen C, Wang J, et al (2024) Application of optical coherence tomography in cardiovascular diseases: bibliometric and meta-analysis. *Front Cardiovasc Med* 11:1414205. <https://doi.org/10.3389/fcvm.2024.1414205>
- [14] Luo YX, Dong YL, Yang FQ, et al (2024) Ultraviolet single-camera stereo-digital image correlation for deformation measurement up to 2600°C. *Exp Mech* 64:1343–1355. <https://doi.org/10.1007/s11340-024-01087-5>
- [15] Yin Y, Wu L, Li J, et al (2019) High-temperature fracture mechanics parameter measurement and yielding zone analysis of superalloy gh4169 based on single-lens 3d digital image correlation. *Exp Mech* 59(7):953–962. <https://doi.org/10.1007/s11340-019-00490-7>
- [16] Richardson A, Strom J, Olson E (2013) Aprilcal: assisted and repeatable camera calibration. In: *IEEE/RSJ Int. Conf. Intell. Robots Syst.*, pp 1814–1821, <https://doi.org/10.1109/IROS.2013.6696595>
- [17] Shi ZC, Shang Y, Zhang XF, et al (2021) Dlt-lines based camera calibration with lens radial and tangential distortion. *Exp Mech* 61(8):1237–1247. <https://doi.org/10.1007/s11340-021-00726-5>
- [18] Guan BL, Yu YJ, Su A, et al (2019) Self-calibration approach to stereo cameras with radial distortion based on epipolar constraint. *Appl Opt* 58(31):8511–8521. <https://doi.org/10.1364/AO.58.008511>
- [19] Liu ZB, Guan BL, Shang Y, et al (2024) Lecalib: Line-based event camera calibration. *Measurement* 235:114900. <https://doi.org/10.1016/j.measurement.2024.114900>
- [20] Bradski G (2000) The opencv library. *Dr Dobb's Journal: Software Tools for the Professional Programmer* 25(11):120–123
- [21] Fetić A, Jurić D, Osmanković D (2012) The procedure of a camera calibration using camera calibration toolbox for matlab. In: *International Convention MIPRO*, pp 1752–1757
- [22] Lenz R, Tsai R (1987) Techniques for calibration of the scale factor and image center for high accuracy 3d machine vision metrology. In: *IEEE Int. Conf. Robot. Autom.*, pp 68–75, <https://doi.org/10.1109/ROBOT.1987.1088012>
- [23] Zhang ZY (2000) A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Mach Intell* 22(11):1330–1334. <https://doi.org/10.1109/34.888718>
- [24] Tan L, Wang YN, Yu HS, et al (2017) Automatic camera calibration using active displays of a virtual pattern. *Sensors* 17(4):685. <https://doi.org/10.3390/s17040685>

- [25] Tsoy T, Safin R, Martinez-garcia E, et al (2022) Exhaustive simulation approach for a virtual camera valibration evaluation in gazebo. In: IEEE Int. Conf. Autom. Robot. Appl., pp 233–238, <https://doi.org/10.1109/ICARA55094.2022.9738581>
- [26] Ren YZ, Hu F (2021) Camera calibration with pose guidance. In: IEEE Int. Conf. Acoust. Speech Signal Process., pp 2180–2184, <https://doi.org/10.1109/ICASSP39728.2021.9414468>
- [27] Lei WT, Xu MD, Hou FF, et al (2020) Calibration venus: an interactive camera calibration method based on search algorithm and pose decomposition. Electronics 9(12):2170. <https://doi.org/10.3390/electronics9122170>
- [28] Rojtborg P, Kuijper A (2018) Efficient pose selection for interactive camera calibration. In: IEEE Int. Symp. Mixed Augment. Real., pp 31–36, <https://doi.org/10.1109/ISMAR.2018.00026>
- [29] Peng SY, Sturm P (2019) Calibration wizard: a guidance system for camera calibration based on modelling geometric and corner uncertainty. In: IEEE Int. Conf. Comput. Vis., pp 1497–1505, <https://doi.org/10.1109/ICCV.2019.00158>
- [30] Weng J, Cohen P, Herniou M (1992) Camera calibration with distortion models and accuracy evaluation. IEEE Trans Pattern Anal Mach Intell 14(10):965–980. <https://doi.org/10.1109/34.159901>
- [31] Sun W, Cooperstock JR (2005) Requirements for camera calibration: must accuracy come with a high price. In: IEEE Workshops Appl. Comput. Vis., pp 356–361, <https://doi.org/10.1109/ACVMOT.2005.102>
- [32] Heikkila J, Silven O (1997) A four-step camera calibration procedure with implicit image correction. In: IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp 1106–1112, <https://doi.org/10.1109/CVPR.1997.609468>
- [33] Liang SK, Guan BL, Yu ZB, et al (2024) Camera calibration using a collimator system. In: Eur. Conf. Comput. Vis., pp 374–390, https://doi.org/10.1007/978-3-031-73668-1_22
- [34] Liang SK, Tan DC, Guan BL, et al (2025) High-precision visual navigation device calibration method based on collimator. Opt Lasers Technol 190:113170
- [35] Liu ZB, Liang SK, Guan BL, et al (2025) Collimator-assisted high-precision calibration method for event cameras. Opt Lett 50:4254–4257. <https://doi.org/10.1364/OL.564294>
- [36] Pan B, Lu ZX, Xie HM (2010) Mean intensity gradient: An effective global parameter for quality assessment of the speckle patterns used in digital image correlation. Optics and Lasers in Engineering pp 469–477

- [37] Yu LP, Pan B (2021) Overview of high-temperature deformation measurement using digital image correlation. *Experimental Mechanics* pp 1121–1142
- [38] Heanjia Super-Metals Co., Ltd (2015) Stainless steel 310/310s (uns s31000/ uns s31008). URL <https://super-metals.com/wp-content/uploads/2015/03/SS-310.pdf>, accessed: July 10, 2025
- [39] Selig JM (2004) Lie groups and lie algebras in robotics. In: *Computational Noncommutative Algebra and Applications*, pp 101–125
- [40] Diane VO (1981) Schur complements and statistics. *Linear Algebra and its Applications* 36:187–295. [https://doi.org/https://doi.org/10.1016/0024-3795\(81\)90232-9](https://doi.org/https://doi.org/10.1016/0024-3795(81)90232-9)

Supplementary Material

Appendix A The explicit formulas of $\mathbf{U}_i$ and $\mathbf{V}_i$

In the following, we derive the analytical expression of the Jacobian blocks $\mathbf{U}_i$ and $\mathbf{V}_i$. Initially, we compute the Jacobian block ($\mathbf{V}_i$) of the left camera's reprojection error with respect to its absolute extrinsic parameters, which characterizes the influence of pose perturbations on projective geometry.

Assuming that the 3D coordinates of the control point $\mathbf{P}(X^w, Y^w, Z^w)$ in the left camera coordinate system are $\mathbf{Q}^l(X^l, Y^l, Z^l)$, the rigid-body transformation between the left camera coordinate system and the world coordinate system is given by:

$$\mathbf{Q}^l = \mathbf{R}^l \mathbf{P} + \mathbf{t}^l \quad (\text{A1})$$

where $\mathbf{R}^l$ denotes the rotation matrix, and the rotation vector in the Lie algebra [39]. We employ the following model to characterize lens distortion effects:

$$\begin{cases} x_d^l = x^l(1 + k_1^l(r^l)^2 + k_2^l(r^l)^4) + 2p_1^l x^l y^l + p_2^l((r^l)^2 + 2(x^l)^2) \\ y_d^l = y^l(1 + k_1^l(r^l)^2 + k_2^l(r^l)^4) + p_1^l((r^l)^2 + 2(x^l)^2) + 2p_2^l x^l y^l \end{cases} \quad (\text{A2})$$

with

$$x^l = \frac{X^l}{Z^l}, \quad y^l = \frac{Y^l}{Z^l}, \quad (r^l)^2 = (x^l)^2 + (y^l)^2 \quad (\text{A3})$$

For clarity of presentation, we define $\mathbf{x}^l(x^l, y^l)$ as the normalized coordinate and $\mathbf{x}_d^l(x_d^l, y_d^l)$ as the distorted normalized coordinate. Based on the projection model, $\mathbf{x}_d^l$ can be mapped to the pixel coordinates $\mathbf{p}^l(u^l, v^l)$:

$$\begin{cases} u^l = f_x^l x_d^l + c_x^l \\ v^l = f_y^l y_d^l + c_y^l \end{cases} \quad (\text{A4})$$

where f_x^l, f_y^l, c_x^l and c_y^l are the intrinsic parameters of the left camera. As described in this paper, the reprojection residual $\widehat{\mathbf{p}}^l(p_x^l, p_y^l)$ characterizes the differentials between the observed coordinates $\tilde{\mathbf{p}}^l(\tilde{u}^l, \tilde{v}^l)$ and the projected coordinates $\mathbf{p}^l(u^l, v^l)$:

$$\begin{cases} p_x^l = \tilde{u}^l - u^l \\ p_y^l = \tilde{v}^l - v^l \end{cases} \quad (\text{A5})$$

Our goal is to derive the partial derivative of the left camera's reprojection residual with respect to its absolute extrinsic parameters. According to the chain rule, it can be expressed as:

$$\frac{\partial \widehat{\mathbf{p}}^l}{\partial \mathbf{R}^l} = \frac{\partial \widehat{\mathbf{p}}^l}{\partial \mathbf{x}_d^l} \cdot \frac{\partial \mathbf{x}_d^l}{\partial \mathbf{x}^l} \cdot \frac{\partial \mathbf{x}^l}{\partial \mathbf{Q}^l} \cdot \frac{\partial \mathbf{Q}^l}{\partial \mathbf{R}^l} \quad (\text{A6})$$

and

$$\frac{\partial \widehat{\mathbf{p}}^l}{\partial \mathbf{t}^l} = \frac{\partial \widehat{\mathbf{p}}^l}{\partial \mathbf{x}_d^l} \cdot \frac{\partial \mathbf{x}_d^l}{\partial \mathbf{x}^l} \cdot \frac{\partial \mathbf{x}^l}{\partial \mathbf{Q}^l} \cdot \frac{\partial \mathbf{Q}^l}{\partial \mathbf{t}^l} \quad (\text{A7})$$

The solutions for each component are derived as follows:

$$\frac{\partial \mathbf{p}^l}{\partial \mathbf{x}_d^l} = - \begin{bmatrix} f_x^l & 0 \\ 0 & f_y^l \end{bmatrix} \quad (\text{A8})$$

$$\frac{\partial \mathbf{x}_d^l}{\partial \mathbf{x}^l} = \begin{bmatrix} \frac{\partial x_d^l}{\partial x^l} & \frac{\partial x_d^l}{\partial y^l} \\ \frac{\partial y_d^l}{\partial x^l} & \frac{\partial y_d^l}{\partial y^l} \end{bmatrix} \quad (\text{A9})$$

with

$$\begin{cases} \frac{\partial x_d^l}{\partial x^l} = (1 + k_1^l (r^l)^2 + k_2^l (r^l)^4) + 2(x^l)^2 (k_1^l + 2k_2^l (r^l)^2) + 2p_1^l y^l + 6p_2^l x^l \\ \frac{\partial y_d^l}{\partial y^l} = (1 + k_1^l (r^l)^2 + k_2^l (r^l)^4) + 2(y^l)^2 (k_1^l + 2k_2^l (r^l)^2) + 6p_1^l y^l + 2p_2^l x^l \\ \frac{\partial x_d^l}{\partial y^l} = 2x^l y^l (k_1^l + 2k_2^l (r^l)^2) + 2p_1^l x^l + 2p_2^l y^l \\ \frac{\partial y_d^l}{\partial x^l} = 2x^l y^l (k_1^l + 2k_2^l (r^l)^2) + 2p_1^l x^l + 2p_2^l y^l \end{cases} \quad (\text{A10})$$

$$\frac{\partial \mathbf{x}^l}{\partial \mathbf{Q}^l} = \begin{bmatrix} \frac{1}{Z^l} & 0 & -\frac{X^l}{(Z^l)^2} \\ 0 & \frac{1}{Z^l} & -\frac{Y^l}{(Z^l)^2} \end{bmatrix} \quad (\text{A11})$$

$$\frac{\partial \mathbf{Q}^l}{\partial \mathbf{R}^l} = -[\mathbf{Q}^l]_{\times}, \quad \frac{\partial \mathbf{Q}^l}{\partial \mathbf{t}^l} = \mathbf{I} \quad (\text{A12})$$

Thus, we have rigorously derived the solution for the Jacobian block ($\mathbf{V}_i$) of the left camera's reprojection residual with respect to its absolute extrinsic parameters. While the imaging geometry of the right camera is analogous to the left camera, its absolute extrinsic parameters are represented as a composite transformation of the left camera's absolute extrinsic parameters and their relative extrinsic parameters. Due to the high-precision estimation of the relative extrinsic parameters constitutes a core objective of stereo calibration, we therefore define the Jacobian block $\mathbf{U}_i$ as the partial derivative of the right camera's reprojection residual with respect to the relative extrinsic parameters.

In fact, the calculation of $\mathbf{U}_i$ follows a similar procedure, with the only difference being that the left camera coordinate system is used as the world coordinate system, that is:

$$\mathbf{Q}^r = \mathbf{R}\mathbf{Q}^l + \mathbf{t} = \mathbf{R}(\mathbf{R}^l\mathbf{P} + \mathbf{t}^l) + \mathbf{t} \quad (\text{A13})$$

the solutions for the relevant components are modified accordingly:

$$\frac{\partial \mathbf{Q}^r}{\partial \mathbf{R}} = -[\mathbf{Q}^r]_{\times}, \quad \frac{\partial \mathbf{Q}^r}{\partial \mathbf{t}} = \mathbf{I} \quad (\text{A14})$$

All the parameters of the Jacobian matrix are now determined.

Appendix B Covariance derivation

Let the joint covariance $\begin{pmatrix} \boldsymbol{\theta}_{\text{rel}} \\ \boldsymbol{\theta}_{\text{abs}} \end{pmatrix}$ be defined by the equation matrix:

$$\text{Cov} \begin{pmatrix} \boldsymbol{\theta}_{\text{rel}} \\ \boldsymbol{\theta}_{\text{abs}} \end{pmatrix} = (\mathbf{J}^\top \mathbf{J})^{-1} \quad (\text{B15})$$

where $\boldsymbol{\theta}_{\text{rel}} = [\mathbf{R}, \mathbf{t}]$ and $\boldsymbol{\theta}_{\text{abs}} = [\mathbf{R}_i^l, \mathbf{t}_i^l]$ are positive definite.

According to Eq. (7) in our manuscript, the information matrix $\mathbf{J}^\top \mathbf{J}$ can be determined as:

$$\mathbf{J}^\top \mathbf{J} = \begin{pmatrix} \sum_{i=1}^m \mathbf{U}_i^\top \mathbf{U}_i & \mathbf{U}_1^\top \mathbf{V}_1 & \mathbf{U}_2^\top \mathbf{V}_2 & \cdots & \mathbf{U}_m^\top \mathbf{V}_m \\ \mathbf{V}_1^\top \mathbf{U}_1 & \mathbf{V}_1^\top \mathbf{V}_1 & 0 & \cdots & 0 \\ \mathbf{V}_2^\top \mathbf{U}_2 & 0 & \mathbf{V}_2^\top \mathbf{V}_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{V}_m^\top \mathbf{U}_m & 0 & 0 & \cdots & \mathbf{V}_m^\top \mathbf{V}_m \end{pmatrix} \quad (\text{B16})$$

For clarity of presentation, we construct the block matrices as given in Eqs (8) and (9) of the manuscript. According to the Schur Complement Lemma [40], for a block matrix,

$$\mathbf{M} = \begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{pmatrix} \quad (\text{B17})$$

if $\mathbf{H}$ is invertible, then the Schur complement of $\mathbf{E}$ with respect to $\mathbf{M}$ is well-defined and given by:

$$\boldsymbol{\Sigma} = (\mathbf{E} - \mathbf{F}\mathbf{H}^{-1}\mathbf{G})^{-1} \quad (\text{B18})$$

For our case, the block matrix is defined as:

$$\mathbf{J}^\top \mathbf{J} = \begin{pmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^\top & \mathbf{B} \end{pmatrix} \quad (\text{B19})$$

since $\mathbf{B}$ is block-diagonal and each block $\mathbf{V}_i^\top \mathbf{V}_i$ is invertible, due to the positive definiteness of measurement Jacobians, $\mathbf{B}$ is invertible. By the definition of the joint covariance, the sub-block corresponding to $\boldsymbol{\theta}_{\text{rel}}$ can be described as:

$$\text{Cov}(\boldsymbol{\theta}_{\text{rel}}) = (\mathbf{A} - \mathbf{C}\mathbf{B}^{-1}\mathbf{C}^\top)^{-1} = \boldsymbol{\Sigma} \quad (\text{B20})$$

Appendix C Improvement over independent monocular calibration

To quantify how stereo-specific constraints improve performance over independent monocular calibration, we have conducted relevant comparative experiments for verification. And all calibration images are checkerboard images captured under optimal

pose guidance, and the intrinsic parameters and distortion coefficients are kept consistent. The monocular calibration procedure is defined as: solving the absolute extrinsic parameters of the left and right cameras separately from each corresponding set of calibration images, and then deriving relative extrinsic parameters from the absolute extrinsic parameters of the left and right cameras based on the variant of Eq. (3). Perform a quantitative comparison between the mean value of derived relative extrinsic parameters and the relative extrinsic parameters obtained from our method. Note that the translation vector is calculated using the arithmetic mean method, whereas the rotation matrices need to first be converted into quaternions, averaged in quaternion space, and then transformed back into rotation vectors. The calibration images used are the 2-random + 6-optimal images in Section 4.2, and the result is shown as follows:

Table C1 Quantitative comparison of relative extrinsic parameters between independent monocular calibration and stereo calibration

Calibration scheme	$\mathbf{R}/\text{deg}$	$\mathbf{t}/\text{mm}$	$\mathbf{RE}/\text{pixel}$	$\mathbf{TE}/\text{mm}$
Monocular	(1.181, -19.421, -1.104)	(-370.4, -12.7, 31.7)	0.2502	0.1703
Stereo	(-0.388, -19.186, -0.532)	(-385.0, -5.8, 72.0)	0.2715	0.0676

As shown in Table C1, according to the above procedure, although the reprojection error of stereo calibration realized from monocular calibration results is smaller than that of our method, the triangulation error calculated therefrom is much larger. We consider that there is a certain overfitting phenomenon in its calibration results. Compared with stereo calibration, monocular calibration lacks some geometric constraints between cameras, so it is more prone to overfitting.

Appendix D Compare against stereo-specialized methods

Table D2 Comparison of different stereo calibration methods

Parameter	Our	Guan[10]
$\mathbf{R}$	(-0.388, -19.186, -0.532)	(-0.407, -19.271, -0.509)
$\mathbf{t}$	(-385.0, -5.8, 72.0)	(-390.2, -10.1, 65.3)
$\mathbf{RE}/\text{pixel}$	0.2715	0.2924
$\mathbf{TE}/\text{mm}$	0.0676	0.0734

We have compared our method with the stereo calibration method in [18]. This method estimates intrinsic and extrinsic parameters simultaneously using nine or more image points based on the epipolar constraint between stereo image pairs. Since it is

a self-calibration method, it does not require the use of a checkerboard. To enable an effective comparison between the two methods, we have used the intersection coordinates extracted from the checkerboard as input to the method in [18]. The calibration images used are also the 2-random + 6-optimal images in Sec. 4.2:

As shown in Table D2, compared with the method in [18], our method has slight advantages in both reprojection error and intersection error, which demonstrates the effectiveness of our method. Meanwhile, both stereo calibration methods outperform independent monocular calibration in terms of triangulation error, which also reflects the superiority of the binocular system in measurement.