Online Extrinsic Parameters Calibration of On-board Stereo Cameras based on Certifiable Optimization

Abstract:
The extrinsic parameters of on-board stereo cameras can be slightly altered due to temperature fluctuations, vibrations, and accidental impacts during automobile driving, leading to significant performance loss in dense stereo matching. In this paper, we propose an online calibration method based on certifiable optimization to address this issue. Initially, sparse feature points are collected based on plane distribution and disparity. A robust optimization model is then developed to minimize the epipolar error, utilizing iterative local optimization to eliminate outliers and determine the 5DOF extrinsic parameters. Subsequently, a relaxation problem is constructed using inliers, and global optimization is performed to certify that the locally optimal results are indeed globally optimal. Comparative experimental results demonstrate that the proposed method offers high accuracy and reliability. Additionally, the quality of the disparity map generated by our calibration method is comparable to that achieved through offline calibration.
Key words: stereo camera; self calibration; pose estimation; epipolar geometry; certifiable optimization;

1 Introduction

Stereo cameras utilize two cameras—one on the left and one on the right—to simultaneously capture images from two perspectives. Through pixel-level matching and disparity calculation between these two images, three-dimensional depth information of the scene can be obtained. In the process of pixel matching, the algorithm typically searches for corresponding pixels within the same row. Therefore, stereo rectification is a necessary prerequisite, which crucially involves precise calibration of the intrinsic and extrinsic parameters of the camera. This serves as the fundamental prerequisite for most stereo matching algorithms, enabling the horizontal alignment of pixels in the images and simplifying the correspondence problem of feature points into an efficient one-dimensional search along the epipolar lines.

Traditionally, stereo cameras are fixed in place using sophisticated materials and structures. However, changes in temperature, mechanical vibrations, and unforeseen impacts while driving can cause the extrinsic parameters, initially set during offline calibration, to shift over time. Even minor changes in these parameters can result in substantial reductions in the number of matched pixels in the stereo disparity map. Consequently, the capability for stereo cameras to conduct online self-calibration is crucial for maintaining long-term autonomy in self-driving cars that rely on stereo vision.

There have been extensive studies in the field of camera calibration. Traditional calibration methods require specific calibration targets, such as 1D straight lines[1][2], 2D flat plates[3][4][5]. For example, Zhang's calibration method[6] involves capturing images of a chessboard. While these methods offer high accuracy and robustness[7][8][9][10], it is cumbersome to re-calibrate after any changes to the camera parameters. The recalibration process often renders the camera and vehicle temporarily unavailable, significantly impacting efficiency in practical applications. The self-calibration method does not limit to any known spatial structure or camera motion information, and can complete the calibration in an unknown scene directly from the image, which is very flexible to use. These methods are often based on mapping relationships between images to estimate the parameters of the camera. Faugeras[11] first derived the Kruppa equation based on the principle of epipolar line transformation to solve for camera parameters, which is applicable to any scenarios with overlapping fields of view. Refs[12][13][14][15] recovered the essential matrix from matching points in two views and use SVD decomposition to obtain the relative pose. Tan[16] and Guo[17] pointed out that compared

to using clustered feature point data, using feature points evenly distributed across the image can better utilize epipolar geometric constraints to obtain more accurate poses. Li[18] proposed an unconstrained method based on image gridding, using a minimal solver method, the eight-point algorithm[19], to calibrate the extrinsic parameters. The minimal solver method is often combined with the RANSAC[20] algorithm to enhance robustness against incorrect matching points. Guan[21] estimated the fundamental matrix using at least nine points from stereo images based on epipolar geometric constraints, and then simultaneously estimated the intrinsic and extrinsic parameters of the stereo camera with radial distortion by minimizing epipolar errors. Though these minimal solver methods are convenient to use, the results are often influenced by noise in the matched feature point data, often yielding suboptimal solutions. Briales[22] and Zhao[23] considered utilizing more matched feature points and employed non-minimal solvers for camera pose estimation, achieving more stable and accurate results. However, the solutions obtained by directly solving the relaxed problem may be approximate solutions of the primal problem and are not guaranteed to satisfy the epipolar constraints.

With known camera intrinsic parameters, Schonberger[24] turned the camera pose solving problem into a PnP(Perspective-n-Point) problem by establishing a correspondence between 2D image points and 3D spatial coordinates. PnP problem can be solved using linear methods or nonlinear methods. Linear methods, which include DLT[25], EPNP[26], UPNP[27], offer faster computation speeds but can be unstable when there are fewer point correspondences. Nonlinear methods formulate the PNP problem as a Bundle Adjustment(BA) problem and obtain the camera's extrinsic parameters by minimizing the reprojection error. Tian[28] and Xing[29] developed a three-view cross-reprojection optimization method in a trinocular camera system to estimate the relative poses of the cameras. There are also camera extrinsic calibration methods that are achieved through SLAM[30][31][32], enabling self-calibration of extrinsic parameters even when cameras do not have overlapping fields of view. However, these methods often require maintaining the temporal relationship of matched feature points, limiting their use in dynamic road scenarios, and tend to consume more performance and time when processing the same observational data.

Some studies have also proposed using neural network models for self-calibration of camera extrinsic parameters. Huo[33] built a convolutional neural network for phase extraction of calibration pattern. Kendall[34] introduced an end-to-end neural network called PoseNet for detecting the 6-DOF pose of a camera. Omid[35] proposed a neural network model that can compute the essential matrix without requiring any matched data points. Yotam[36] established consistency between the outputs of pre-trained binocular and monocular depth estimation networks when the binocular camera's extrinsic parameters are unknown, enabling implicit calibration of the binocular camera. However, it can be challenging to assess the generalizability of these neural network-based self-calibration techniques because their efficacy is frequently directly correlated with the size of the dataset. Moreover, their substantial computational overhead renders them unsuitable for industrial practical self-calibration jobs.

Addressing these limitations, this paper proposes an online calibration method for onboard stereo cameras using certifiable optimization with epipolar constraints, achieving high-precision extrinsic calibration. The main contributions are:

(1) A method for collecting and filtering points through image gridding, fully utilizing epipolar constraints for more accurate extrinsic parameter estimation.
(2) A robust M-estimator optimization model that improves accuracy under noise and outliers compared to minimum solvers.
(3) By relaxing the primal problem and solving the SDP, the global optimality of local optimization results is certified, ensuring solutions satisfy epipolar constraints.

2 Stereo vision model

A stereo camera typically comprises two cameras positioned horizontally with overlapping fields of view, as illustrated in Fig. 1.

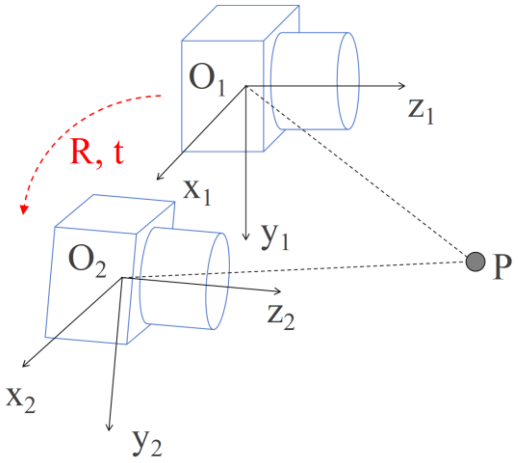


Fig.1 Extrinsic parameters of stereo cameras.

The coordinate transformation between the left and right camera coordinate systems can be represented by a 3×3 matrix rotation $\mathbf{R}$ and a 3×1 translation vector $\mathbf{t}$, which are the extrinsic parameters to be calibrated for the stereo system. In the world coordinate system, the projection of a point $\mathbf{P}(x, y, z)^T$ onto the images of the left and right cameras can be represented by the homogeneous forms of image coordinate points, denoted as $\mathbf{p}_1(u_1, v_1, 1)^T$ and $\mathbf{p}_2(u_2, v_2, 1)^T$ respectively. The projection process begins by transforming the coordinate point from the world coordinate system to the camera coordinate system using the camera's extrinsic parameters. It then converts the point from the camera coordinate system to the image coordinate system utilizing the intrinsic parameters. The specific calculation process is as

$$s_1 \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = \mathbf{K}_1[\mathbf{I} \mid \mathbf{O}] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \tag{1}$$

$$s_2 \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = \mathbf{K}_2[\mathbf{R} \mid \mathbf{t}] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}. \tag{2}$$

where s_1 and s_2 are scaling factors. $\mathbf{K}_1$ and $\mathbf{K}_2$ are the intrinsic parameter matrices of the left and right cameras. $\mathbf{I}$ is a 3×3 identity matrix, and $\mathbf{O}$ represents the zero vector, indicating that the world coordinate system aligns with the left camera's coordinate system. $\mathbf{R}$ and $\mathbf{t}$ represent the rotation and translation needed to convert a point from the left camera's coordinate system to the right camera's coordinate system.

The points $\mathbf{p}_1$ and $\mathbf{p}_2$ in the images of the left and right cameras should satisfy the epipolar constraint in multi-view geometry, as shown in Eq. (3). Epipolar geometry is an important concept in multi-view geometry of computer vision, describing the visual geometric relationship between two images of the same scene.

$$\mathbf{p}_2{}^T \mathbf{F} \mathbf{p}_1 = \mathbf{p}_2{}^T \mathbf{K}_2{}^{-T} \mathbf{E} \mathbf{K}_1{}^{-1} \mathbf{p}_1 = 0. \tag{3}$$

where $\mathbf{F}$ and $\mathbf{E}$ represent the fundamental matrix and essential matrix, both of which are in the form of

3×3 matrices. They specifically describe the linear transformation of objects in the left and right images. The extrinsic parameters $\mathbf{R}$ and $\mathbf{t}$ of the stereo camera can be obtained by performing an SVD(Singular Value Decomposition) on the essential matrix, as shown in Eq. (4).

$$\mathbf{E} = [\mathbf{t}]_{\times}\mathbf{R}. \tag{4}$$

where $[\mathbf{t}]_{\times}$ is used to represent the skew-symmetric matrix form of the translation vector $\mathbf{t} = [t_1, t_2, t_3]^T$ as shown in Eq. (5).

$$[\mathbf{t}]_{\times} = \begin{pmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{pmatrix}. \tag{5}$$

It is important to note that the true scale of the unit vector $\mathbf{t}$ obtained from SVD is unknown. Fortunately the error of the scale is very small and does not affect the number of valid matched pixels in the stereo disparity map, because the baseline length of the stereo camera does not change significantly except in extreme cases, and the depth error of the stereo camera is linearly related to the change in baseline length[37]. Thus, this paper can utilize the offline calibrated baseline length to recover the scale of $\mathbf{t}$.

3 Online calibration method

Eq. (4). provides a way to obtain the extrinsic parameters of a stereo camera system. Specifically, it calculates the essential matrix between the left and right image using the corresponding points. Then, by performing SVD decomposition on the matrix, the relative pose between the left and right cameras can be determined. The prerequisite for using this method is that the intrinsic parameters of the left and right cameras and the accurate correspondences of points are known. In practice, the camera's intrinsic parameters are often obtained accurately through offline calibration methods, such as Zhang's checkerboard calibration method. On-board stereo cameras are often calibrated during manufacturing and do not change easily, so in this paper, the camera's intrinsic parameters are considered known constants. The correspondences between feature points can be achieved through feature point extraction and matching. However, due to the noise and outliers in the matched feature point data, traditional methods often fail to obtain precise calibration results, and the optimized results tend to fall into local optima or cannot guarantee that the results meet the internal constraints. The uncertainties and measurement noise in the camera system during online calibration significantly compromise its accuracy [38][39][40], as the detection of feature points is inherently subjected to uncertainties and noise, and the camera's internal parameters contain minor errors even after factory calibration. Furthermore, the iterative optimization process of online calibration can potentially converge to suboptimal solutions. Therefore, this paper proposes a robust and optimally verifiable optimization method for self-calibration of stereo cameras.

As shown in Fig. 2, the proposed method mainly comprises four parts: data collection, robust optimization of the essential matrix, global optimal certification, and pose recovery. Firstly, ORB feature points are extracted from the left and right camera images, and feature point matching and mismatch removal are performed. After image gridding, the feature point density of each grid is calculated based on the distance from the grid to the image center, and the extracted feature point data is filtered by combining the disparity size of the feature points. Subsequently, a robust optimization model that minimizes epipolar errors is established. Through gradually non-convex optimization iterations of the essential matrix and data weights, a local optimal result is obtained. Then, by relaxing the primal problem and solving the SDP (Semi-Definite Programming) problem using inlier data, a lower bound of the dual problem is obtained to globally certificate the local optimal result. Finally, the verified essential matrix is decomposed using SVD to recover the camera's extrinsic parameters, thus completing the online extrinsic calibration of the stereo cameras.

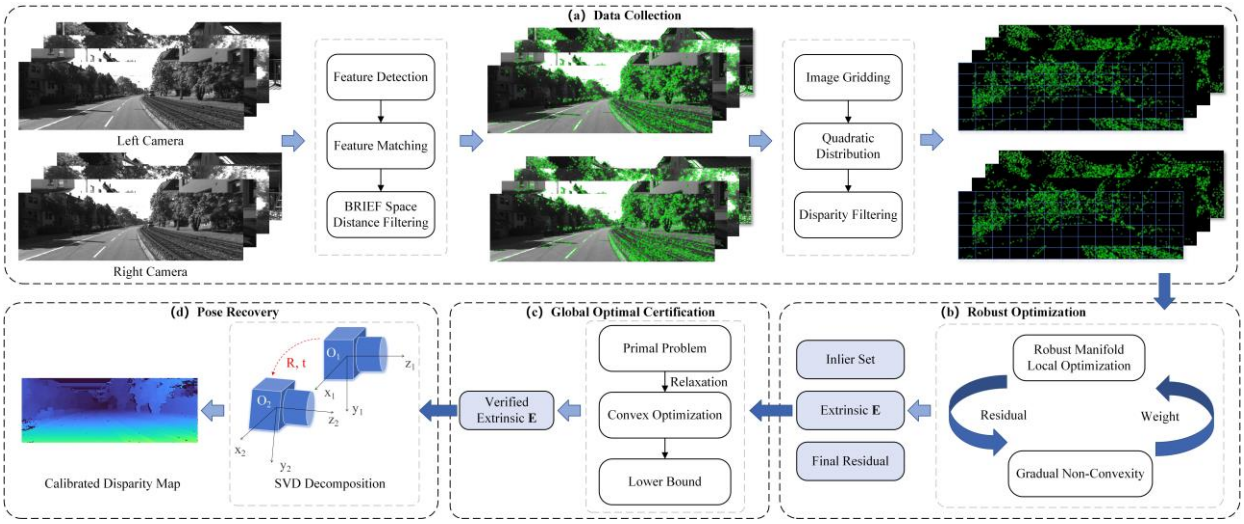


Fig.2 Process of self calibration method for on-board stereo cameras.

3.1 Date collection

To obtain the relative pose between the left and right cameras, it is first necessary to collect feature points data on the images. In the self-calibration of stereo cameras, feature points tend to cluster in areas with richer textures in road scenes, resulting in uneven spatial distribution of feature points. Typically, all corresponding image point pairs in an image should satisfy epipolar constraints, but a better distribution of feature points can provide better geometric constraints, thus improving the reliability and accuracy of matrix $\mathbf{E}$. Therefore, the selection of feature points is also crucial for improving the accuracy of self-calibration. In this paper, the description and extraction of feature points rely on the ORB method. ORB has a significant speed advantage, but the feature points extracted by this method tend to cluster on the image plane. To address this issue, we propose a feature point selection method based on image plane and disparity. The image plane-based selection method involves filtering image feature points to ensure that the final collected image feature points satisfy specific distribution characteristics on the image plane. To guarantee that, the image is first split into grids, with the number of feature points in each grid being maintained. Considering that the image edges have larger distortions due to aberrations, we adopt a quadratic distribution method for feature point selection, as shown in Fig.3. Specifically, after dividing the image into grids, we calculate the feature point density of each grid based on the distance from the center of the grid to the center of the image using Eq. (6).

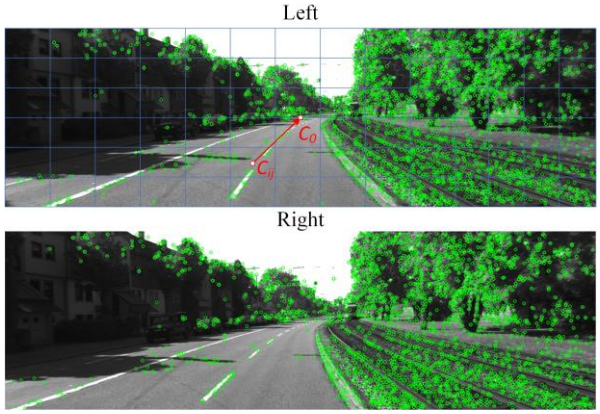


Fig.3 Density calculation of feature points per grid.

$$\rho_{ij} = \rho_0 - \alpha \cdot \left\| \mathbf{C}_{ij} - \mathbf{C}_0 \right\|^2 . \tag{6}$$

where ρ_{ij} represents the density of feature points in the grid located at the *i*-th row and *j*-th column, $\mathbf{C}_{ij}$

denotes the central coordinate point of the grid， and $\mathbf{C}_0$ represents the center of the image. The values of ρ_0 and α are determined by the following boundary conditions: the feature point density in the grid at the image edge is set to 0, and the feature point density in the grid at the image center is set to 1.

Considering that in a stereo camera system, the feature points only fall within the overlapping regions of the images, rather than the entire image, this paper introduces a certain pixel offset $\mathbf{C}_0' = \mathbf{C}_0 + [x_0, 0]^T$ to the image center $\mathbf{C}_0$. This adjustment ensures that more feature points in the overlapping field of view, rather than those in a single view, better satisfy the quadratic distribution.

When detected feature points fall into the divided grids, they are added or removed based on the following criteria:

(1) If the number of points has not reached the upper limit in a grid, all matching feature points are retained.

(2) If the number of points reaches the upper limit in a grid, the matching feature points with larger disparities are retained.

(3) If the number of points reaches the upper limit in a grid and the matching feature points have similar disparity values, the latest matching feature points are retained.

The proposed feature point data collection method in this paper is completed only in the image space, without considering the temporal relationship of matching feature points. This allows for efficient data collection in dynamic road scenarios. The reason for retaining matching feature points with larger disparities in the process is that these data contain more information about camera translation and rotation perpendicular to the epipolar plane, which can better reduce self-calibration errors and improve calibration accuracy.

3.2 Robust optimization

In practical scenarios, the acquired observation data often unavoidably contains numerous outliers and noise. Outliers greatly influence the precision of the calibration. If they are not removed or their influence is not reduced, the precision of the optimization outcome will be greatly diminished, and even completely incorrect results may arise. A commonly used method to remove outliers is RANSAC algorithm, which is often combined with a least-squares solver. The RANSAC-based method is effective when the proportion of outliers is high, but its performance is limited in cases where the data contains a significant amount of noise. To address the issue of insufficient accuracy due to the presence of significant noise and outliers in the data, this paper introduces an optimization method based on the M-estimator. This approach can achieve more precise results in the presence of noisy data and eliminate outliers using robust functions to complete the optimization of the camera's extrinsic parameters.

In this paper, the robust solution for the essential matrix is achieved by minimizing the squared error of epipolar geometry over a set of essential matrices, denoted as $\mathbf{M}_\mathbf{E}$. The translation $\mathbf{t}$ can be parameterized on a two-dimensional sphere, while the rotation matrix $\mathbf{R}$ belongs to the Lie group SO(3). Therefore, the set of essential matrices can be expressed in the following form:

$$\mathbf{M}_\mathbf{E} = \{\mathbf{E} \mid \mathbf{E} = [\mathbf{t}]_\times \mathbf{R},\ \exists \mathbf{R} \in \mathrm{SO}(3),\ \mathbf{t} \in \mathrm{S}^2\}. \tag{7}$$

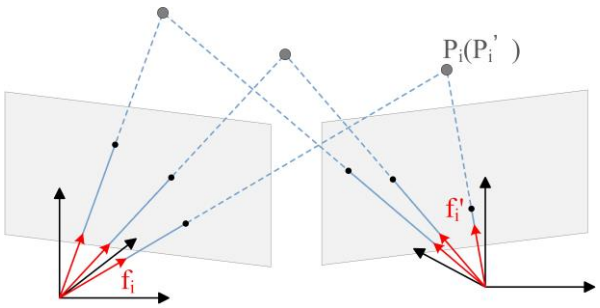


Fig.4 Construct optimization problems through N pair-wise correspondences.

$(\mathbf{P}_i, \mathbf{P}_i^{'})$ represent the left and right camera coordinates of the *i*-th point in the world coordinate. Each coordinate point can be transformed into a representation as a vector passing through the origin of the camera coordinate system, as shown in Fig. 4. The transformation formula is given as

$$\mathbf{f}_i = \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|}, \quad \mathbf{f}_i^{'} = \frac{\mathbf{P}_i^{'}}{\|\mathbf{P}_i^{'}\|}. \tag{8}$$

$(\mathbf{f}_i, \mathbf{f}_i^{'})$ represents the observation data of the same 3D point from different perspectives by the left and right cameras. In the absence of noise, the epipolar constraint can be expressed as

$$\mathbf{f}_i^T \mathbf{E} \mathbf{f}_i' = 0. \tag{9}$$

In practical scenarios, the observation data often contains noise, and the equality in Eq. (9). does not always hold strictly. Therefore, a joint optimization function can be constructed to minimize the epipolar error.

$$\min_{\mathbf{E} \in \mathbf{M}_\mathbf{E}} \sum_{i=1}^{N} (\mathbf{f}_i^T \mathbf{E} \mathbf{f}_i')^2. \tag{10}$$

The objective function in problem (10) can be rewritten as a quadratic programming problem.

$$\min_{\mathbf{E} \in \mathbf{M}_\mathbf{E}} \sum_{i=1}^{N} (\mathbf{f}_i^{\mathbf{T}} \mathbf{E} \mathbf{f}_i^{'})^2 = \min_{\mathbf{E} \in \mathbf{M}_\mathbf{E}} \mathrm{vec}(\mathbf{E})^T \mathbf{C} \mathrm{vec}(\mathbf{E}), \tag{11}$$

$$\mathbf{C} = \sum_{i=1}^{N} (\mathbf{f}_i' \otimes \mathbf{f}_i)(\mathbf{f}_i' \otimes \mathbf{f}_i)^T. \tag{12}$$

where "$\otimes$" represents the Kronecker product, and $\mathbf{C}$ is a positive semidefinite matrix.

To robustly and accurately estimate the camera pose and minimize the impact of outliers and noise, this paper utilizes the robust loss function strategy from the M-estimator, incorporating the scalable Welsch robust function $\rho(\cdot)$ into the optimization process.

$$\min_{\mathbf{E} \in \mathbf{M}_E} \sum_{i=1}^{N} \rho(r_i). \tag{13}$$

where r_i represents the epipolar error resulting from the *i*-th pair of matched feature points. The scalable Welsch robust function is defined as

$$\rho(r, \tau) = \frac{\tau^2}{2}(1 - e^{\frac{-r^2}{\tau^2}}). \tag{14}$$

where $\tau \in (0, +\infty)$ represents the scaling factor of the robust function, which can be used to control the shape of the function.

To facilitate the removal of outliers, variable weights $w_i \in [0,1]$ are further introduced, corresponding to the *i*-th matched feature point pair, along with an outlier handling term $\Psi(w_i)$. The problem can be expressed in the following form:

$$\min_{\substack{E \in \mathbf{M}_E, \\ w_i, i=1,\ldots N}} \sum_{i=1}^{N} w_i r_i^2 + \Psi(w_i). \tag{15}$$

The term $\Psi(w_i)$ is used to determine whether to ignore the *i*-th matched point pair. The closer it is to 0, the more valid the matched point pair data is considered. The specific expression is given as

$$\Psi(w_i) = \frac{\tau^2}{2}(1 + w_i \log(w_i) - w_i). \tag{16}$$

Problem (15) represents a non-convex optimization function, and its shape is controlled by the scaling factor τ. To facilitate the optimization of this problem, this paper adopts the idea of Gradually Non-Convex (GNC) optimization[32]. By controlling the size of τ, the optimization can start from a more tractable convex

optimization problem and gradually transition to a harder-to-optimize non-convex problem. The initial convex problem is solved to obtain an initial value, and then the non-convexity of the problem is gradually increased while values are computed and iterated, ultimately leading to the solution of the primal problem.

During the actual problem-solving process, the iterative updates of $\mathbf{E}$ and w_i are performed separately. In one optimization cycle, $\mathbf{E}$ is first optimized while fixing the w_i values obtained from the previous optimization. Then, through a closed-form calculation, $\mathbf{E}$ is substituted into Eq. (17). to update the weights w_i. In practical applications, a weight threshold can be set, and only the matched point pairs with weights higher than this threshold will be considered as inliers and included in subsequent global optimization calculations, while those below the threshold will be treated as outliers and excluded.

$$w_i = e^{-\frac{r^2}{\tau^2}}. \tag{17}$$

To solve the problem, one can choose to adopt a manifold-based optimization method to perform unconstrained local optimization of the essential matrix on the essential matrix manifold. The solver selected is the RTR[33], which offers an excellent balance between convergence rate and speed.

3.3 Global optimal certification

In Section 3.2, through GNC iterative optimization, a local optimal solution satisfying the intrinsic constraints of epipolar geometry for the self-calibration of stereo camera's extrinsic parameters, along with an inlier set after outlier removal, is obtained. To ensure the precision of the results, this paper further verifies that the local optimal solution is the global optimal solution. First, the primal problem is relaxed by removing one of the epipolar constraints. Then, an SDP solver is used to calculate the dual problem and obtain a lower bound for the primal problem, which is used to globally certificate the local optimal result.

In epipolar geometry, an essential matrix $\mathbf{E}$ should satisfy the following constraints:

$$\mathbf{E}\mathbf{E}^T = [\mathbf{t}_\times][\mathbf{t}_\times]^T \text{ and } \mathbf{t}^T\mathbf{t} = 1, \tag{18}$$

The primal problem (9) can be reformulated into a quadratic programming form with the constraints (18).

$$\begin{aligned} &\min_{\mathbf{E},\mathbf{t}} \ \mathrm{vec}(\mathbf{E})^T \mathbf{C}_{\text{inlier}} \mathrm{vec}(\mathbf{E}) \\ &\text{s.t. } \mathbf{E}\mathbf{E}^T = [\mathbf{t}_\times][\mathbf{t}_\times]^T, \\ &\quad \mathbf{t}^T\mathbf{t} = 1. \end{aligned} \tag{19}$$

where $\mathbf{C}_{\text{inlier}}$ represents the data matrix composed of inliers after the robust optimization model in Section 3.2. In this optimization problem, there are 12 variables and 7 constraints, with the specific expressions of the constraints as

$$\begin{cases} h_1 = \mathbf{e}_1^T\mathbf{e}_1 - (t_2^2 + t_3^2) = 0, \\ h_2 = \mathbf{e}_2^T\mathbf{e}_2 - (t_1^2 + t_3^2) = 0, \\ h_3 = \mathbf{e}_3^T\mathbf{e}_3 - (t_1^2 + t_2^2) = 0, \\ h_4 = \mathbf{e}_1^T\mathbf{e}_2 + t_1t_2 = 0, \\ h_5 = \mathbf{e}_1^T\mathbf{e}_3 + t_1t_3 = 0, \\ h_6 = \mathbf{e}_2^T\mathbf{e}_3 + t_2t_3 = 0, \\ h_7 = \mathbf{t}^T\mathbf{t} - 1 = 0. \end{cases} \tag{20}$$

where $\mathbf{e}_i$ denotes the i-th column vector of the matrix $\mathbf{E}$. The optimization objective function and feasible region in the primal problem (9) and the problem (19) are the same, therefore, the problems are equivalent. To optimize problem (19), this paper relaxes the constraint of the problem by removing one h_i constraint (i = 2, …, 7), resulting in a relaxed problem, which can be specifically expressed as:

$$\begin{aligned} &\min_{x} \mathbf{x}^T \mathbf{Q} \mathbf{x} \\ &\text{s.t. } \mathbf{x}^T \mathbf{A}_i \mathbf{x} = b_i, i = 1,...,6. \end{aligned} \tag{21}$$

In this problem, $\mathbf{x} = [\text{vec}(\mathbf{E})^T, \mathbf{t}^T]^T$ is a 12-dimensional vector. $\mathbf{Q}$ is a 12×12 symmetric matrix, with the upper-left corner composed of the $\mathbf{C}_{\text{inlier}}$ matrix, and the rest filled with zeros. $\mathbf{A}_i$ and b_i represent the i-th remaining constraint among the six remaining constraints after relaxing the constraints in (19). Empirically, even though the constraints on the essential matrix $\mathbf{E}$ are relaxed, the solutions obtained in practice often still satisfy all the constraints of the essential matrix. Therefore, this paper considers the relaxation in problem (21) for problem (19) to be a tight relaxation. Furthermore, the dual form of the relaxed problem is expressed using Lagrangian duality.

$$\begin{aligned} &\max_{\boldsymbol{\lambda}} \lambda_1 \\ &\text{s.t.} \mathbf{Q}(\boldsymbol{\lambda}) = \mathbf{Q} - \sum_{i=1}^{6} \lambda_i \mathbf{A}_i \pm 0. \end{aligned} \tag{22}$$

where $\mathbf{Q}(\boldsymbol{\lambda})$ represents the Lagrangian Hessian matrix, $\boldsymbol{\lambda}$ denotes the Lagrangian multipliers.

According to the theorems related to duality, this dual problem gives a lower bound for the primal problem (19), which allows us to verify whether a given locally optimal feasible solution $\hat{\mathbf{x}}$ is globally optimal. Specifically, if the $\hat{\lambda}_1 = \hat{\mathbf{x}}^T \mathbf{Q} \hat{\mathbf{x}}$ equation holds (within a certain precision range), the duality is considered strong, and the solution $\hat{\mathbf{x}}$ is the global optimal solution. The relaxed problem (22) is an SDP problem that can be resolved with SDP solvers, like SDPT3[34] and SeDuMi[35].

In the specific certification of the optimality of the solution, two conditions need to be met: (a) when the size of the duality gap is less than a certain value; (b) the minimum eigenvalue of $\mathbf{Q}(\boldsymbol{\lambda})$ is within a certain range close to 0, the solution can be considered certified.

4 Experiments and results

In this section, to validate the proposed method, we elaborate on the specific experimental methods and results. Firstly, we use generated data to verify the robustness of the optimization model to noise and outliers. Then, we conduct data collection experiments and online self-calibration experiments based on the KITTI real stereo images, comparing the results obtained by different methods with the results of offline calibration. Finally, we perform disparity map calculation experiments to demonstrate the effectiveness and reliability of the online calibration. The experimental hardware platform used in this paper is an Intel i5-7500 CPU with 3.4 GHz clock speed and 16GB of RAM.

4.1 Robust testing

To validate the robustness of the optimization model to noise in the data and outliers caused by incorrect matches or other reasons, this paper refers to the scene generation method in [36] by artificially adding different levels of pixel noise to the data points and testing the robust optimization model through randomly generated scenes containing different percentages of outliers. In addition, we further conducted robustness experiments in real-world scenarios where the vehicle's movement caused motion blur in the images.

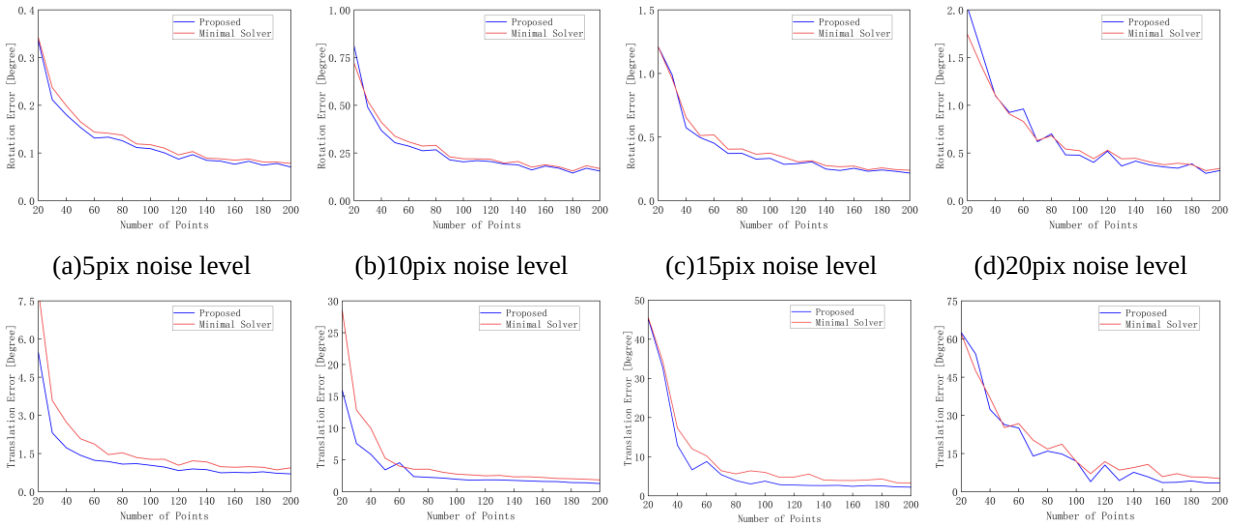
In the generated scenes, the world coordinate system is fixed on the left camera, while the right camera is randomly translated within a range of 0.3m and randomly rotated within a range of 10°. The camera's focal length is set to 800 pixels, and the scene points are positioned at distances ranging from 1m to 15m from the camera. The field of view of the camera is set to 100°. Different levels of pixel noise are added, including specific noise levels of 5, 10, 15, and 20 pixels, and each scene is repeated 100 times under each condition to calculate the average error. This paper uses the results previously artificially generated or obtained from offline

calibration as a reference value and compares different methods with it to verify the algorithm's accuracy. The error calculation can be divided into two parts: relative rotation error and translation error.

$$E(\mathbf{R}_{cal}) = \arccos\left(\frac{\mathrm{trace}(\mathbf{R}_{cal}\mathbf{R}_{ref}^{T}) - 1}{2}\right) \cdot \frac{180^{\circ}}{\pi}, \tag{23}$$

$$E(\mathbf{t}_{\mathrm{cal}}) = \arccos\left(\frac{\mathbf{t}_{ref}^{T}}{\|\mathbf{t}_{ref}\|} \cdot \frac{\mathbf{t}_{cal}}{\|\mathbf{t}_{cal}\|}\right) \cdot \frac{180^{\circ}}{\pi}. \tag{24}$$

where $\mathbf{R}_{cal}$ and $\mathbf{t}_{cal}$ represent the rotation and translation parameters of the camera's extrinsic parameters obtained from the self-calibration method, while $\mathbf{R}_{ref}$ and $\mathbf{t}_{ref}$ represent the reference rotation and translation parameters. Since the true scale of the translation vector cannot be obtained, for the sake of convenience in calculation, the translation error in this paper is calculated as the angular error between unit vectors.



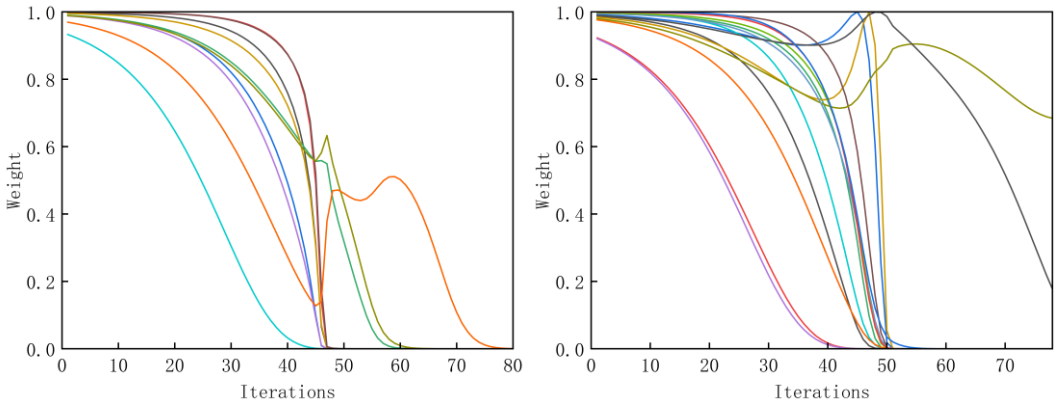
(a)5pix noise level (b)10pix noise level (c)15pix noise level (d)20pix noise level

(e)5pix noise level (f)10pix noise level (g)15pix noise level (h)20pix noise level

Fig.5 Rotation and translation error with different noise levels. Figures in the first row show the rotation error, and the second row show the translation error.

It can be found in Fig. 5, as the number of generated points increases, the method proposed in this paper can estimate the camera's rotation and translation more accurately compared to the least-squares solver under different noise levels, when the number of matched point pairs exceeds 20. Moreover, the higher the noise level, the greater the gap between the two methods, especially when predicting translation parameters.

To validate the detection effect of the robust optimization model on outliers, we generated 100 points with a noise level of 5 pixels, some of which are outliers. We then observed the effect of iteratively reducing the weights of outliers under different outlier ratios.



（a）10% outliers　　（b）15% outliers

Fig.6 Outliers rejection experiments with different occupancy ratios. Each line of different color represents a outlier correspondence.

As shown in Fig. 6, under the conditions of 10% and 15% outlier ratios, the weights of most outliers are reduced to 0 within the first 50 iterations of the GNC method. In the case of 15% outliers, while the weights of some remaining outliers decrease to a certain extent after additional iterations, the outliers are still successfully detected.

The choice of the robust function has a significant impact on the performance of the optimization model. To compare the performance of different robust functions under varying outlier ratios, this paper compares the Welsch function with two other robust functions, GM and Tukey. The generated scene conditions are set with a noise level of 0.2 pixels and 500 points. Different robust functions are used to detect outliers, with 500 repetitions of the scene experiment, recording the number of successful detections of all outliers. Tab. 1 shows the success probabilities of different robust functions in detecting outliers. When the outlier ratio is 10%, the GM robust function and the Welsch robust function have similar detection effects on outliers, and both can successfully detect outliers and converge the optimization results in most cases. However, as the outlier ratio increases, the detection effect of the GM robust function declines significantly. In contrast, the Welsch robust function maintains a higher success rate in detecting outliers even when the outlier ratio reaches 20%, outperforming both the Tukey and GM robust functions. This indicates that the Welsch robust function is more stable in detecting outliers under different outlier ratios.

Tab.1 Comparison of success rates in detecting outliers with different robust functions

	GM	Tukey	Welsch
10% outliers	**85.0%**	71.0%	84.7%
15% outliers	**70.6%**	65.4%	70.3%
20% outliers	50.8%	52.8%	**55.0%**

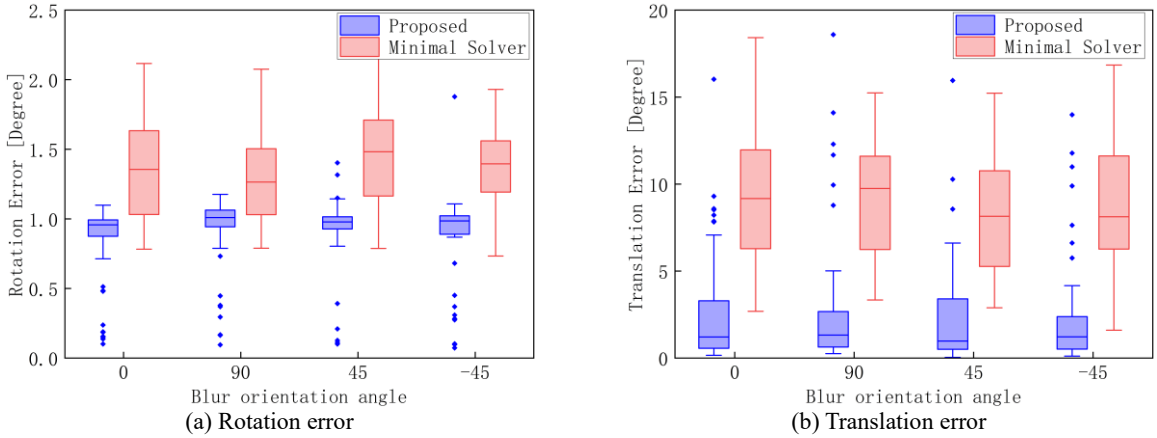
In actual scenarios, due to the vehicle's fast movement such as turning or a bad road condition, the on-board camera may capture images with motion blur, which may lead to noise in the coordinates of feature points, resulting in unstable calibration results. Therefore, we have further conducted experiments on real images with motion blur to validate the robustness of the proposed method. To simulate different directions of motion, this paper generates motion-blurred images on the KITTI stereo dataset by using different motion blur filters with orientation angles. Fig. 7 illustrates the images with different directions of motion blur, which may be encountered during driving. By performing feature point detection and matching on a set of motion-blurred images, 100 randomly selected matching feature points are used as input to calculate the camera's rotation and translation error.

(a) Horizontal motion blur in 0° (b) Vertical motion blur in 90°

(c) Bottom-left to top-right motion blur in 45° (d) Bottom-right to top-left motion blur in -45°

Fig.7 Motion blur images in different direction with a kernel size equal to 19.

It can be observed from Figure 8 that under the influence of motion blur noise at different angles, the rotation and translation results calculated by the method proposed in this paper are more stable compared to the minimal solver method, demonstrating the robustness of the proposed approach under challenging conditions.



(a) Rotation error (b) Translation error

Fig.8 Rotation and translation error with respect to different blur orientation angles.

4.2 Data collection experiment

In order to validate the practical application of the self-calibration algorithm, this paper selects real stereo camera images from the KITTI driving raw data as the experimental data source for online calibration of on-board stereo cameras. This dataset contains only natural road scenes without any artificial calibration references. The driving scenes include cities, residential areas, and structured roads. There are numerous dynamic objects in the scenes, which closely resemble actual driving conditions. The baseline length of the stereo camera used to collect data is approximately 54 cm, and the data is collected at a frequency of 10 Hz. The stereo images in the dataset have been temporally synchronized based on the timestamp files in the dataset, but they have not undergone distortion removal or rectification processing, simulating the complete self-calibration process of an actual on-board stereo camera.

The dataset also includes calibration data, as shown in Tab. 2. The experiments will use the camera intrinsics, distortion coefficients, and baseline length from the calibration data files as known quantities, and the camera extrinsics as the reference values for the self-calibration algorithm presented in this paper.

Tab.2 Stereo camera parameters.

Parameters	Left camera	Right camera
Focal length $[f_u,\ f_v]$	[2061.94, 2060.67]	[2061.27, 2060.78]
Principle point $[u_0,\ v_0]$	[971.90, 579.79]	[968.46, 586.84]
Skew	0.55	-0.60
Distortion coefficients	[-0.091386, 0.112783, 0.000165, 0.000090]	[-0.092928, 0.127109, -0.000662, -0.000495]
Rotation quaternion	$[0, 0, 0, 1]^T$	$[0.003200, -0.011919, -0.010843, 0.999865]^T$
Translation vector	$[0, 0, 0]^T$	$[-0.544915, -0.000155, -0.00226]^T$

To verify that the proposed data acquisition strategy is effective, this paper compares four types of feature point distributions, as shown in Fig. 9, which are: (1) grid-based quadratic distribution of feature points, while retaining the feature points with the largest disparity; (2) grid-based evenly distributed feature points after filtering, with randomly retained feature points; (3) continuously collecting feature point pairs until reaching an upper limit of 2000, resulting in a random distribution of feature points; (4) retaining all feature points from a single self-calibration process.

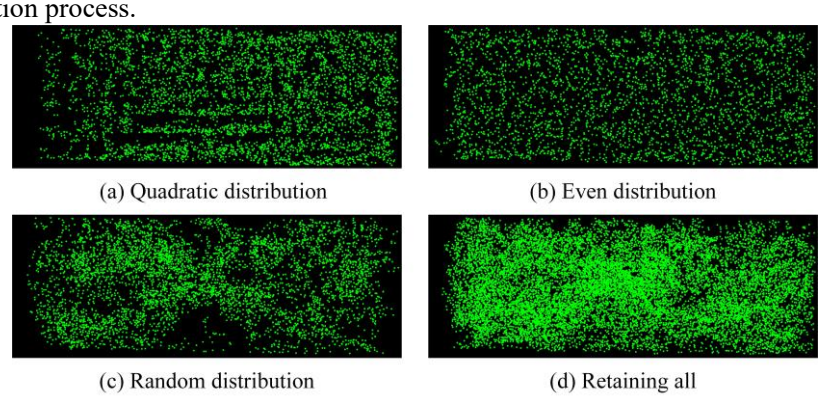


Fig.9 Different distribution forms of feature points.

For the feature point data with the different distribution patterns, the essential matrix is computed for each set. Then, the epipolar error for each feature point is calculated using Eq. (3). Finally, the squared errors are accumulated and averaged. The resulting values are compared, as shown in Tab. 3. Based on the analysis of image and table data, it can be concluded that random distribution of feature points tends to concentrate data points in areas with richer textures in the image, leading to unstable performance in different scenarios and often resulting in significant average epipolar errors. While retaining all data points can slightly improve accuracy by increasing the number of points, the redundant data points are numerous (exceeding 30,000 matched pairs), and the improvement is quite limited, only slightly better than the random distribution strategy. In contrast, the quadratic distribution can achieve the lowest average epipolar geometric error, resulting in a better pose estimation.

Tab.3 Comparison of errors in different distribution forms.

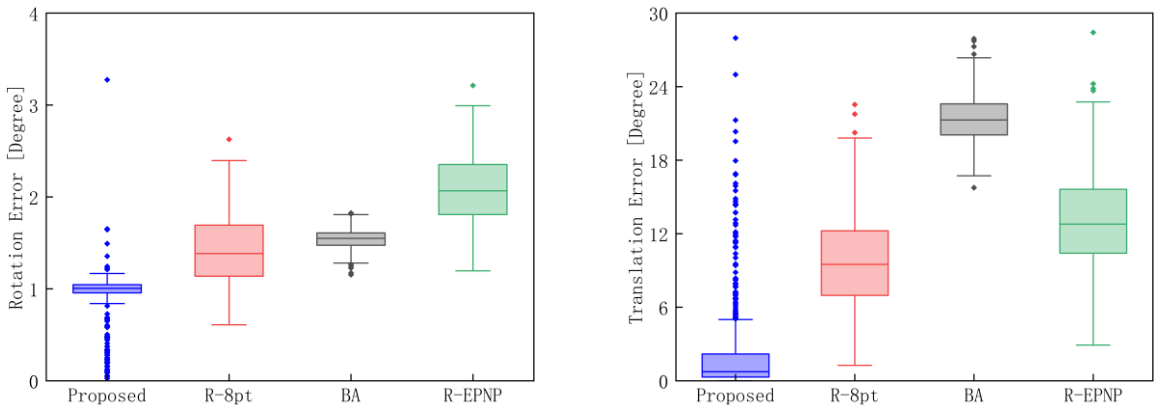
Distribution	Average epipolar error（pixel）
Quadratic distribution	**0.68092**
Even distribution	1.29723
Random distribution	2.07459
Retaining all	1.68465

4.3 Online calibration experiment

Online calibration can be divided into two parts: pose initialization and continuous estimation. To validate the effectiveness of the proposed method, this paper compares it with different 2D-2D and 2D-3D pose estimation methods on the KITTI driving dataset.

The result of pose initialization has a significant impact on the accuracy of subsequent estimation results in calibration. A good initial value can lead to better calibration results and faster convergence speed. In the pose initialization part, this paper compares the proposed method with eight-point method, bundle adjustment(BA) method, and EPNP method. The data for the self-calibration experiment comes from a randomly selected driving segment in the KITTI dataset, and the initial data is collected using the data acquisition strategy proposed in Section 3.1 of this paper. Among them, the R-8pt and R-EPNP pose estimation algorithms represent the use of RANSAC for outlier removal. Each method is repeated 500 times, and the results are plotted as the error box plot shown in Fig. 10.

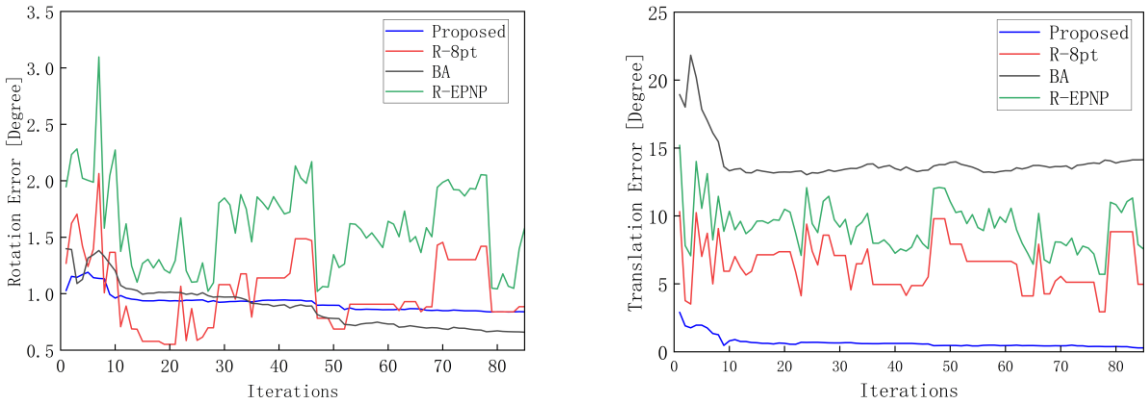
Under the same data distribution, the 2D-2D pose estimation methods, the proposed method and the R-8pt method, outperform the 2D-3D methods such as BA and EPNP. Compared to the R-8pt method, the proposed method exhibits significantly lower average and median errors. Specifically, the median rotation error and median translation error of the optimized pose initialization method proposed in this paper are 1.05° and 0.74°, respectively. In contrast, the median rotation errors of R-8pt, BA, and R-EPNP are 1.38°, 1.54°, and 2.07°, and the median translation errors are 9.50°, 21.82°, and 12.78°, respectively.



(a)Rotation error (b)Translation error

Fig.10 Initialization errors using different methods.

In the continuous estimation part, in order to validate the effectiveness of proposed method in the process of driving, the system continues to estimate and tracking parameters. According to the criteria proposed in Section 3.1 for data collection, the system updates the data every additional 50 images captured. It then uses the previous optimization results as the initial values for an iterative optimization. It simultaneously tracks the variations in the dual gap parameters and minimal eigenvalue related to the optimal certification.

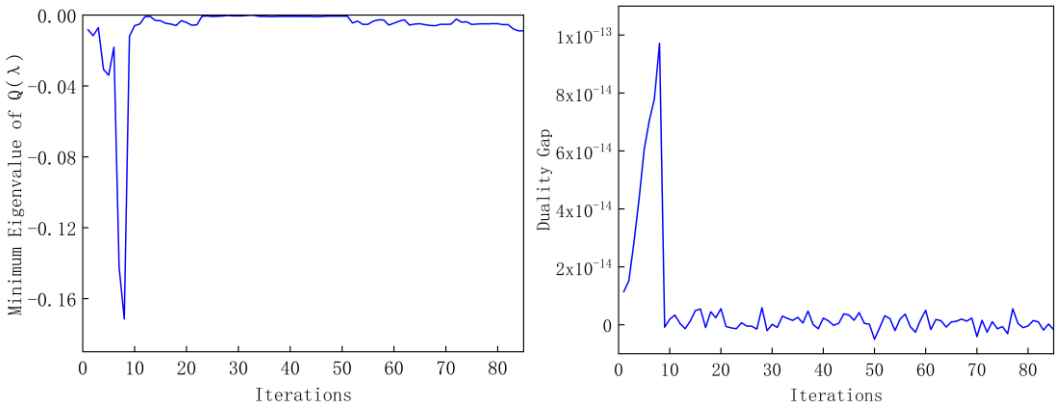


(a) Rotation error (b)Translation error

Fig.11 Changes in rotation and translation errors in continuous self calibration.

As can be seen in Fig. 11a and 11b, the BA method steadily reduces the error in estimating rotation with increasing iterations. Although its estimation results for rotation are better than the other three methods, the BA method performs poorly in estimating translation, with errors significantly higher than the other three methods. The R-8pt and R-EPNP methods, due to falling into local optima, exhibit fluctuations in their error curves, resulting in unstable pose estimation results and large errors in estimating translation. In contrast, the prediction results of the proposed method in this paper are very stable, with a very smooth error curve.

Fig. 12 shows the tracking results of the optimal certifier parameters during a continuous self-calibration process. It can be observed that after the 10th iteration, the parameters remain stable within a certain threshold, indicating that the optimization has reached a global optimal result.



（a）Minimum eigenvalue （b）Duality gap

Fig.12 Certification related parameters tracking.

When considering the overall error estimation for both rotation and translation comprehensively, the similarity between essential matrices $\mathbf{E}_{cal}$ and $\mathbf{E}_{ref}$ can be calculated using Eq. (25). The essential matrix error results during the continuous estimation process are shown in Fig. 13.

$$E(\mathbf{E}_{cal}) = \arccos\left(\frac{\mathbf{E}_{ref}{}^{T}\mathbf{E}_{cal}}{\|\mathbf{E}_{ref}\| \cdot \|\mathbf{E}_{cal}\|}\right) \cdot \frac{180^{\circ}}{\pi}. \tag{25}$$

It can be observed that there is a clear correspondence between the minimum eigenvalue and dual gap parameters used to certify the optimization results in Fig. 12 and the essential matrix errors shown in Fig. 13. After 10 iterations, the minimum eigenvalue of matrix $\mathbf{Q}(\boldsymbol{\lambda})$ converges to 0, and the dual gap lies within a range near 0, indicating that the optimization result at this point is the global optimal solution for the problem. This corresponds to the decrease in essential matrix errors after the 10th iteration in Fig. 13, demonstrating that the optimal verifier method proposed in this paper can effectively validate the stereo online calibration

optimization results. The comparison of error curves in Fig. 13 reflects that the proposed method exhibits smaller and more stable errors in estimating the essential matrix during continuous estimation compared to the other three methods, enabling more accurate extrinsic parameters of the stereo camera to be obtained.

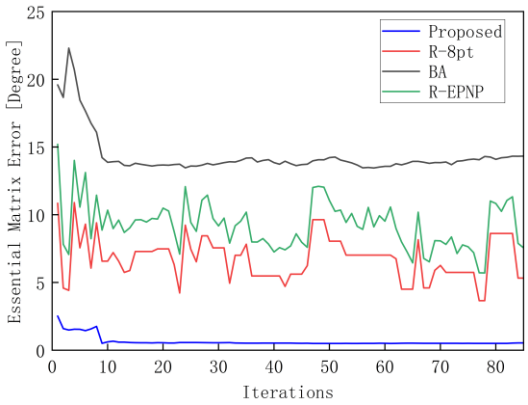


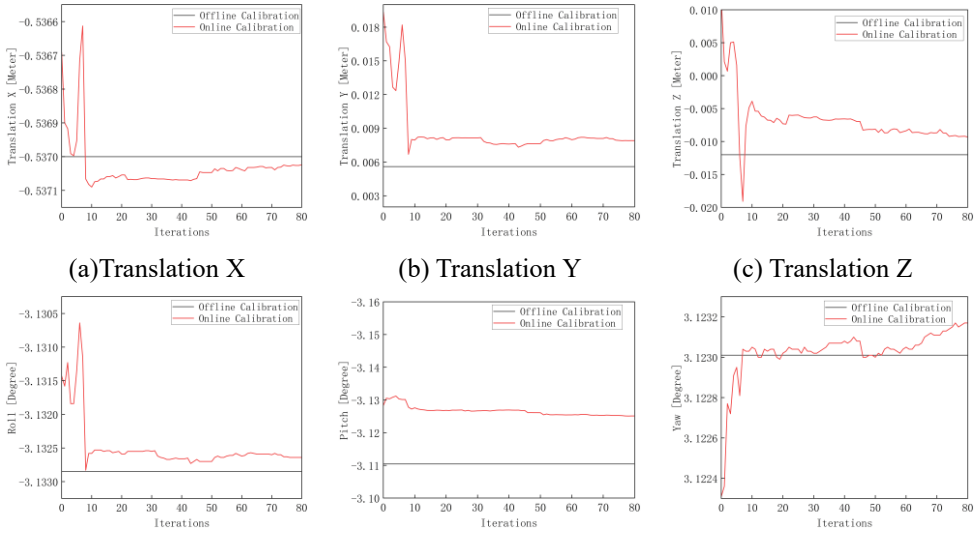
Fig.13 Changes in essential matrix error in continuous online calibration.

We compared the proposed method with the other three methods in handling the running time of each iteration, and the efficiency comparison results are shown in Tab. 4. It can be seen that the proposed algorithm is at least five times faster in running efficiency than other methods, and its excellent real-time performance can enable vehicles to perform online calibration.

Tab.4 Efficiency comparison with other methods

Method	Runtime	Norm.runtime
BA	**2.5 s**	625
R-EPNP	**22 ms**	5.5
R-8pts	20 ms	5
Proposed	4 ms	1

Fig. 14 illustrates the temporal variations of the specific 6 parameters for rotation and translation during the continuous self-calibration process. It can be observed that the translation components X and Z gradually converge towards the reference values as more data is collected, while the error in the translation component Y remains stable at around 2mm. The two rotation roll and yaw are nearly equal to the true values with very small errors, and the maximum pitch error is maintained within a very narrow range.



(a)Translation X (b) Translation Y (c) Translation Z

(d) Roll (e) Pitch (f) Yaw

Fig.14 Extrinsic parameters tracking.

4.4 Disparity map experiment

In this section, a comparative experiment on disparity map calculation is conducted out in order to better display the effectiveness of the online self-calibration algorithm. The experiment uses both offline calibration results and online calibration results to rectify the original image pairs captured by the stereo camera, aligning the unaligned images onto the same plane. Then, the standard SGBM stereo matching algorithm with parameterized settings is employed to compute the disparity map. The experimental results are shown in Figure 15.

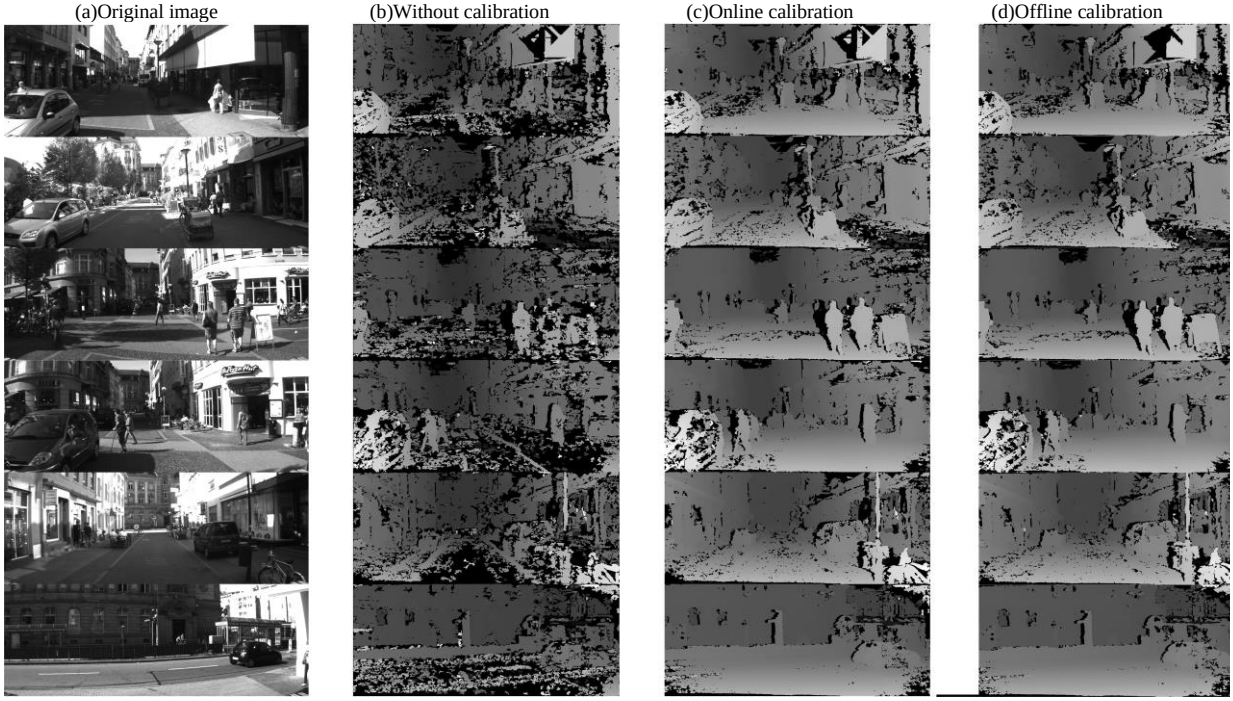


Fig.15 Comparison of disparity map.

The second column of images in Fig. 15 shows the disparity map calculation results after the camera's extrinsic parameters have changed. It can be observed from the disparity map that the number of successfully matched points is very sparse. By comparing the second and third columns of images, it can be seen that the proposed online self-calibration algorithm recover the correct structure of the stereo camera, significantly improving the density and reliability of the disparity map. When comparing the disparity maps obtained from offline calibration and online calibration in the fourth column, it is obvious that the disparity map results obtained using online calibration are comparable in reliability and density to those computed using parameters from offline calibration, without the need for manual calibration targets, thus demonstrating the reliability of online self-calibration.

5 Conclusion

This paper proposes an online extrinsic parameters calibration method for on-board stereo cameras through certifiable optimization using a specific feature points distribution. In the data collection process, sparse feature points are gathered based on plane distribution and disparity, leading to a smaller epipolar error. A robust optimization model is then developed to minimize the epipolar error, performing iterative local optimization to reject outliers and determine the 5DOF extrinsic parameters. Finally, a relaxation problem is constructed using inliers, and global optimization is performed to ensure that the locally optimal results are globally optimal. Comparative experimental results demonstrate that the proposed method offers higher accuracy and reliability compared to R-8pt, BA, and R-EPNP methods, with a maximum rotation error of 0.09° and a translation (unit vector) error of 0.25°. Additionally, the quality of the disparity map generated by the proposed online calibration

method is comparable to that achieved through offline calibration, meeting the practical requirements for visual navigation of on-board stereo cameras and simplifying the cumbersome re-calibration process.

6 Acknowledgements

This work was supported by Guangxi Science and Technology Major Program (Grant No. AA24206036), Guangxi Science and Technology Major Program (Grant No. AA22068108), Guangxi Science and Technology Major Program (Grant No. AA23062088), and the State Key Laboratory of Mechanics and Control for Aerospace Structures at Nanjing University of Aeronautics and Astronautics (Grant No. MCAS-E-0224K01).

References

[1] P. Ge, Y. Wang, H. Wang, G. Li, M. Zhang, Multivision Sensor Extrinsic Calibration Method With Non-Overlapping Fields of View Using Encoded 1D Target, IEEE Sens J 22 (2022) 13519–13528. https://doi.org/10.1109/JSEN.2022.3178950.

[2] Z. Liu, B. Guan, Y. Shang, Z. Yu, Y. Bian, Q. Yu, LECalib: Line-based event camera calibration, Measurement 235 (2024) 114900. https://doi.org/https://doi.org/10.1016/j.measurement.2024.114900.

[3] L. Yang, F. Zhou, W. Zhang, Y. Liu, A novel camera calibration method based on circle projection model, Measurement 222 (2023) 113651. https://doi.org/https://doi.org/10.1016/j.measurement.2023.113651.

[4] J. Cao, X. Zhang, D. Tu, G. Zhou, A visual guidance calibration method for out-of-focus cameras based on iterative phase target, Measurement 218 (2023) 113104. https://doi.org/https://doi.org/10.1016/j.measurement.2023.113104.

[5] J. Wan, X. Zhang, W. Yang, C. Zhang, M. Lei, Z. Dong, A calibration method for defocused cameras based on defocus blur estimation, Measurement 235 (2024) 115045. https://doi.org/https://doi.org/10.1016/j.measurement.2024.115045.

[6] Z. Zhang, A flexible new technique for camera calibration, IEEE Trans Pattern Anal Mach Intell 22 (2000) 1330–1334. https://doi.org/10.1109/34.888718.

[7] W. Ding, W. Tan, G. Liu, H. Zhang, W. Wang, Adaptive adjustment of brightness and blur of the camera for high precision internal parameter calibration, Measurement 231 (2024) 114637. https://doi.org/https://doi.org/10.1016/j.measurement.2024.114637.

[8] Q. Ou, Q. Xie, F. Chen, J. Peng, B. Xiong, Reinforcement learning-based calibration method for cameras with large FOV, Measurement 202 (2022) 111732. https://doi.org/https://doi.org/10.1016/j.measurement.2022.111732.

[9] J. Huo, Z. Meng, H. Zhang, S. Chen, F. Yang, Feature points extraction of defocused images using deep learning for camera calibration, Measurement 188 (2022) 110563. https://doi.org/https://doi.org/10.1016/j.measurement.2021.110563.

[10] S. Yang, J. Wen, S. Wu, T. Yang, Y. Wu, F. Liu, Camera calibration with active standard Gaussian stripes for 3D measurement, Measurement 233 (2024) 114793. https://doi.org/https://doi.org/10.1016/j.measurement.2024.114793.

[11] Q.-T. and M.S.J. Faugeras O. D. and Luong, Camera self-calibration: Theory and experiments, in: G. Sandini (Ed.), Computer Vision — ECCV'92, Springer Berlin Heidelberg, Berlin, Heidelberg, 1992: pp. 321–334.

[12] Multiple View Geometry in Computer Vision, Kybernetes 30 (2001) 1333–1341. https://doi.org/10.1108/k.2001.30.9_10.1333.2.

[13] C. Sweeney, T. Sattler, T. Hollerer, M. Turk, M. Pollefeys, Optimizing the viewing graph for structure-from-motion, in: Proceedings of the IEEE International Conference on Computer Vision, 2015: pp. 801–809.

[14] S. Zhu, R. Zhang, L. Zhou, T. Shen, T. Fang, P. Tan, L. Quan, Very large-scale global sfm by distributed motion

averaging, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: pp. 4568–4577.

[15] P. Moulon, P. Monasse, R. Marlet, Global fusion of relative motions for robust, accurate and scalable structure from motion, in: Proceedings of the IEEE International Conference on Computer Vision, 2013: pp. 3248–3255.

[16] X. Tan, C. Sun, X. Sirault, R. Furbank, T.D. Pham, Feature matching in stereo images encouraging uniform spatial distribution, Pattern Recognit 48 (2015) 2530–2542. https://doi.org/https://doi.org/10.1016/j.patcog.2015.02.026.

[17] X. Guo, X. Cao, Good match exploration using triangle constraint, Pattern Recognit Lett 33 (2012) 872–881. https://doi.org/https://doi.org/10.1016/j.patrec.2011.08.021.

[18] H. Li, K. Wang, K. Yang, R. Cheng, C. Wang, L. Fei, Unconstrained self-calibration of stereo camera on visually impaired assistance devices, Appl. Opt. 58 (2019) 6377–6387. https://doi.org/10.1364/AO.58.006377.

[19] O.D. Faugeras, S. Maybank, Motion from point matches: Multiplicity of solutions, Int J Comput Vis 4 (1990) 225–246. https://doi.org/10.1007/BF00054997.

[20] T. Botterill, S. Mills, R. Green, Refining essential matrix estimates from RANSAC, in: Proceedings Image and Vision Computing New Zealand, 2011: pp. 1–6.

[21] B. Guan, Y. Yu, A. Su, Y. Shang, Q. Yu, Self-calibration approach to stereo cameras with radial distortion based on epipolar constraint, Appl. Opt. 58 (2019) 8511–8521. https://doi.org/10.1364/AO.58.008511.

[22] J. Briales, L. Kneip, J. Gonzalez-Jimenez, A certifiably globally optimal solution to the non-minimal relative pose problem, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: pp. 145–154.

[23] J. Zhao, An Efficient Solution to Non-Minimal Case Essential Matrix Estimation, IEEE Trans Pattern Anal Mach Intell 44 (2022) 1777–1792. https://doi.org/10.1109/TPAMI.2020.3030161.

[24] J.L. Schonberger, J.-M. Frahm, Structure-from-motion revisited, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: pp. 4104–4113.

[25] A. Ansar, K. Daniilidis, Linear pose estimation from points or lines, IEEE Trans Pattern Anal Mach Intell 25 (2003) 578–589. https://doi.org/10.1109/TPAMI.2003.1195992.

[26] V. Lepetit, F. Moreno-Noguer, P. Fua, EPnP: An Accurate O(n) Solution to the PnP Problem, Int J Comput Vis 81 (2009) 155–166. https://doi.org/10.1007/s11263-008-0152-6.

[27] A. Penate-Sanchez, J. Andrade-Cetto, F. Moreno-Noguer, Exhaustive Linearization for Robust Camera Pose and Focal Length Estimation, IEEE Trans Pattern Anal Mach Intell 35 (2013) 2387–2400. https://doi.org/10.1109/TPAMI.2013.36.

[28] X. Tian, Q. Gao, Q. Luo, J. Feng, Trinocular camera self-calibration based on spatio-temporal multi-layer optimization, Measurement 217 (2023) 113003. https://doi.org/https://doi.org/10.1016/j.measurement.2023.113003.

[29] C. Xing, J. Huang, Z. Wang, Q. Duan, Z. Li, M. Qi, A high-accuracy online calibration method for structured light 3D measurement, Measurement 210 (2023) 112488. https://doi.org/https://doi.org/10.1016/j.measurement.2023.112488.

[30] E. Dexheimer, P. Peluse, J. Chen, J. Pritts, M. Kaess, Information-Theoretic Online Multi-Camera Extrinsic Calibration, IEEE Robot Autom Lett 7 (2022) 4757–4764. https://doi.org/10.1109/LRA.2022.3145061.

[31] G. Carrera, A. Angeli, A.J. Davison, SLAM-based automatic extrinsic calibration of a multi-camera rig, in: 2011 IEEE International Conference on Robotics and Automation, 2011: pp. 2652–2659. https://doi.org/10.1109/ICRA.2011.5980294.

[32] K. Eckenhoff, P. Geneva, J. Bloecker, G. Huang, Multi-Camera Visual-Inertial Navigation with Online Intrinsic and Extrinsic Calibration, in: 2019 International Conference on Robotics and Automation (ICRA), 2019: pp.

3158–3164. https://doi.org/10.1109/ICRA.2019.8793886.

[33] J. Huo, Z. Meng, H. Zhang, S. Chen, F. Yang, Feature points extraction of defocused images using deep learning for camera calibration, Measurement 188 (2022) 110563. https://doi.org/https://doi.org/10.1016/j.measurement.2021.110563.

[34] A. Kendall, M. Grimes, R. Cipolla, Posenet: A convolutional network for real-time 6-dof camera relocalization, in: Proceedings of the IEEE International Conference on Computer Vision, 2015: pp. 2938–2946.

[35] O. Poursaeed, G. Yang, A. Prakash, Q. Fang, H. Jiang, B. Hariharan, S. Belongie, Deep fundamental matrix estimation without correspondences, in: Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018: p. 0.

[36] Y. Gil, S. Elmalem, H. Haim, E. Marom, R. Giryes, Online Training of Stereo Self-Calibration Using Monocular Depth Estimation, IEEE Trans Comput Imaging 7 (2021) 812–823. https://doi.org/10.1109/TCI.2021.3098927.

[37] T. Dang, C. Hoffmann, C. Stiller, Continuous Stereo Self-Calibration by Camera Parameter Tracking, IEEE Transactions on Image Processing 18 (2009) 1536–1550. https://doi.org/10.1109/TIP.2009.2017824.

[38] C. Yang, Y. Liu, Multi-objective optimization for robust attitude determination of satellite with narrow bound theory, Advances in Space Research 74 (2024) 3273–3283. https://doi.org/https://doi.org/10.1016/j.asr.2024.06.002.

[39] C. Yang, Z. Fang, H. Ren, W. Lu, Y. Xia, Interval uncertainty-oriented impedance force control for space manipulator with time-dependent reliability, Acta Astronaut 222 (2024) 207–218. https://doi.org/https://doi.org/10.1016/j.actaastro.2024.06.002.

[40] C. Yang, Z. Fan, W. Lu, H. Gao, Uncertain Iterative Optimal Attitude Control Method for Periodic Satellite With Reliability Constraint, IEEE Trans Aerosp Electron Syst (2024) 1–15. https://doi.org/10.1109/TAES.2024.3404915.

[41] H. Yang, P. Antonante, V. Tzoumas, L. Carlone, Graduated Non-Convexity for Robust Spatial Perception: From Non-Minimal Solvers to Global Outlier Rejection, IEEE Robot Autom Lett 5 (2020) 1127–1134. https://doi.org/10.1109/LRA.2020.2965893.

[42] P.-A. Absil, R. Mahony, R. Sepulchre, Optimization algorithms on matrix manifolds, Princeton University Press, 2008.

[43] M.J.T. K. C. Toh, R.H. Tütüncü, SDPT3 — A Matlab software package for semidefinite programming, Version 1.3, Optim Methods Softw 11 (1999) 545–581. https://doi.org/10.1080/10556789908805762.

[44] F.S.J.U. SeDuMi, 1.02, a MATLAB toolbox for optimization over symmetric cones", Optimisation Methods and Software. 1999, 11 (1): 625-653,< $http, sedumi.mcmaster.ca$ (n.d.).

[45] L. Kneip, S. Lynen, Direct optimization of frame-to-frame rotation, in: Proceedings of the IEEE International Conference on Computer Vision, 2013: pp. 2352–2359.