Pisco Med Publishing

doi: 10.18282/iac.v6i1.578

Original Research Article

A fast and accurate camera-IMU calibration method for localization system

Xiaowen Tao, Pengxiang Meng, Bing Zhu*, Jian Zhao

Jilin University, Changchun 130012, Jilin Province, China

*** Corresponding author:** Bing Zhu, zhubing@jlu.edu.cn

Abstract: Autonomous driving has spurred the development of sensor fusion techniques, which combine data from multiple sensors to improve system performance. In particular, a localization system based on sensor fusion, such as Visual Simultaneous Localization and Mapping (VSLAM), plays a crucial role in environment perception and serves as the foundation for decision-making and motion control in intelligent vehicles. The accuracy of extrinsic calibration parameters between the camera and IMU is of utmost importance for precise positioning in VSLAM systems. However, existing calibration methods are often time-consuming, rely on complex optimization techniques, and are sensitive to noise and outliers, leading to potential degradation in system performance. To address these challenges, this paper presents a fast and accurate camera-IMU calibration method based on space coordinate transformation constraints and SVD (Singular Value Decomposition) tricks. The method involves constructing constraint equations by ensuring the equality of rotation and transformation matrices between camera frames and IMU coordinates at different time instances. Subsequently, the external parameters of the camera-IMU system are solved using quaternion transformation and SVD techniques. To validate the proposed method, experiments were conducted using the ROS (Robot Operating System) platform, where camera images and velocity, acceleration, and angular velocity data from the IMU were recorded in a ROS bag file. The results demonstrate that the proposed method achieves reliable camera-IMU calibration parameters, requiring less tuning time and exhibiting reduced uncertainty.

Keywords: camera-IMU calibration; autonomous vehicles; sensor fusion; slam; localization system

Received: 13 May 2023; **Accepted:** 28 June 2023; **Available online:** 12 July 2023

1. Introduction

The rapid development of artificial intelligence has sparked a growing interest in autonomous driving, which holds the promise of transforming transportation systems by enhancing safety, improving efficiency, and alleviating traffic congestion[1]. Within the field of autonomous driving, the localization system plays a crucial role by utilizing a diverse range of sensors for high-precision positioning. It provides the foundation for decision-making, planning, and control execution in intelligent vehicles. Improved localization precision has significant implications for the Intelligent Transportation System (ITS), enabling more informed route selection and dynamic path guidance[2].

Various positioning methods are employed in intelligent vehicles, including GPS, UWB positioning, LiDAR positioning, RFID positioning, visual positioning, RTK positioning, IMU positioning, and DR map matching[1–3]. However, each method has its limitations and lacks generalization capabilities. To address these limitations, multi-sensor fusion technology is utilized, integrating data from different sensors to enhance the accuracy, stability, and generalization capabilities of positioning systems. Fusion techniques such as GPS/IMU,

vision/IMU, GPS/vision, GPS/DR, LiDAR/IMU, LiDAR/GPS, and LiDAR/GPS/RTK are commonly employed to leverage the strengths of individual sensors.

IMU positioning involves obtaining velocity, position, and rotation information using the built-in three-axis gyroscopes and three-axis accelerometers[3]. This method relies on internal sensors, providing high positioning accuracy in a short time without relying on external environments. However, IMU sensors, including accelerometers and gyroscopes, have inherent biases, and measurement errors accumulate over time, making them unsuitable for long-term positioning requirements. In contrast, visual positioning does not suffer from drift and can directly measure rotation and translation. It is commonly used to calibrate the errors in IMU measurements using low-cost camera sensors. Furthermore, IMU sensors can estimate absolute scale and effectively respond to rapid motions and rotations, compensating for the limitations of visual positioning, such as the inability to measure scale and the loss of measurement information during fast movements[4]. As a result, the camera-IMU localization system has become a mainstream sensor-fusion positioning method due to its advantages of low cost, high efficiency, high accuracy, convenience, and speed.

In the camera-IMU localization system, camera and IMU calibration plays a crucial role in achieving accurate data integration. It enables the alignment of visual information from cameras with inertial measurements from IMUs, ensuring precise and synchronized perception for robust decision-making in autonomous systems. The calibration process encompasses camera intrinsic parameters, IMU intrinsic parameters, and camera-to-IMU extrinsic calibration. Camera intrinsic parameters define the projection relationship between the three-dimensional real world and the camera coordinate system. These parameters, including focal length, principal point, and distortion coefficients, are essential for maintaining the accuracy of the camera's data source. They directly impact the quality and accuracy of the captured images. IMU intrinsic parameter calibration focuses on parameters related to deterministic errors and random errors. Deterministic errors refer to systematic biases or scale factors that consistently affect the IMU measurements. They can arise from factors such as sensor misalignments, temperature variations, and nonlinearity in sensor response. Calibration helps estimate and compensate for these deterministic errors, thereby improving the accuracy and reliability of the IMU measurements. Random errors, on the other hand, are typically caused by noise and uncertainties in the IMU sensors. These errors have statistical properties, such as a zero mean and a Gaussian distribution. The calibration of the IMU aims to characterize and minimize these random errors, enhancing the accuracy and precision of the IMU measurements. Camera-to-IMU extrinsic parameters establish the transformation between the camera and IMU measurements, aligning them to a common coordinate frame. These parameters are particularly crucial for Visual Simultaneous Localization and Mapping (VSLAM) systems. Even a slight deviation of 1° to 2° between the calibrated camera sensor coordinate system and the IMU coordinate system can significantly degrade the localization accuracy of the VSLAM system[4,5]. Therefore, the calibration of cameras and IMUs is a critical step in sensor fusion-based positioning systems. It ensures the accuracy and effectiveness of the fusion process, ultimately improving the performance and reliability of the overall positioning system.

Currently, several research efforts are focused on ensuring accurate calibration between cameras and IMUs. In terms of intrinsic calibration, Zhang et al.[6] introduced a widely used technique known as Zhang's method. This method utilizes a calibration pattern with known geometry and uses multiple images to estimate the camera's intrinsic parameters. Advancements in camera calibration techniques have also been made, including self-calibration methods[7] and techniques based on pattern recognition and optimization algorithms[8]. For extrinsic calibration, a common approach involves the use of a calibration target and controlled movements. Zhang et al.[9] proposed a method that utilizes a planar calibration target to estimate the

transformation matrix between the camera and IMU. They employed bundle adjustment techniques to optimize the calibration results. Similarly, Brink et al.[10] used a multi-view calibration target to estimate the relative pose between the camera and IMU. They employed a Kalman filter-based optimization approach for accurate extrinsic calibration. These research efforts highlight the importance of developing accurate calibration techniques to ensure precise alignment and synchronization between camera and IMU measurements in sensor fusion-based positioning systems.

Despite these advancements, it is important to acknowledge that camera-IMU calibration techniques still have limitations. Sünderhauf et al.[11] have highlighted several challenges associated with these techniques, including sensitivity to initialization, the requirement for precise sensor synchronization, and the influence of environmental factors on calibration accuracy. These factors can introduce errors into the calibration process and subsequently impact the accuracy of the sensor fusion system. Furthermore, existing methods often involve time-consuming processes that rely on complex optimization methods, and they can be sensitive to noise and outliers, which may arise due to calibration errors. These issues can have a negative impact on the overall performance of the system.

In this paper, we propose a fast and accurate camera-IMU calibration method based on space coordinate transformation constraints and SVD (Singular Value Decomposition) techniques. Firstly, we construct constraint equations by considering the equality of rotation and transformation matrices between camera frames and IMU coordinates at different time instances. Secondly, we solve for the external parameters of the camera-IMU using quaternion transformation and SVD techniques. The structure of this paper is as follows: Section 1 introduces the calibration process for the camera's parameters. Section 2 provides details on the calibration of the IMU's parameters. Section 3 presents the innovative calibration method for the camera-IMU extrinsic parameters. We present the experimental results in Section 4, and finally, Section 5 concludes the paper.

2. Calibration of camera intrinsic parameters based on pinhole model

2.1. Pinhole model

Before extracting image features for positioning, it is crucial to model the camera sensor accurately and consistently to ensure reliable input image frames. An image frame represents a two-dimensional image plane consisting of pixels with pixel coordinates and brightness information. Meanwhile, corresponding landmarks exist in three-dimensional space. Therefore, a camera model is employed to establish the mapping between three-dimensional landmarks and the two-dimensional image plane. Various camera models are available, including the pinhole camera model[2], stereo camera model[12], and RGB-D camera model[5]. Among these models, the pinhole camera model is widely used due to its simplicity and effectiveness, making it suitable for camera sensors in visual-inertial odometry systems. Therefore, this paper adopts the pinhole camera model to describe the entire mapping process.

The imaging principle of a camera sensor is illustrated in **Figures 1** and **2**.

In the pinhole camera model (**Figure 1**), a camera coordinate system denoted as *Oxyz* is established, with the camera's optical center *O* as the origin. The *x*-axis aligns with the right side of the camera sensor, the *y*-axis aligns with the vertical direction of the camera sensor, and the *z*-axis aligns with the front direction of the camera sensor. Similarly, a camera image plane coordinate system denoted as *O'x'y'* is established. Assuming a point $P = [X, Y, Z]^{\mathrm{T}}$ in the three-dimensional world projects through the pinhole *O* onto the image plane as point $P' = [X', Y', Z']^{\mathrm{T}}$, and considering the distance between the camera plane and the physical image plane

as the focal length *f*, we can derive the following equation based on the principle of similar triangles in the pinhole camera model (**Figure 2**).

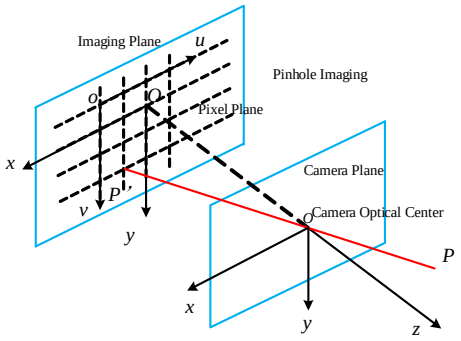


Figure 1. Schematic diagram of the pinhole camera model.

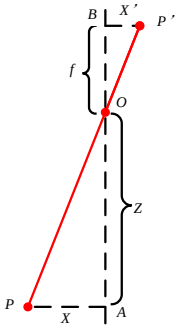


Figure 2. Schematic diagram of similar triangles.

$$\frac{Z}{f} = -\frac{X}{X'} = -\frac{Y}{Y'} \tag{1}$$

The negative sign in Equation (1) traditionally indicates that the resulting image is inverted. However, in practice, the image captured by a camera sensor is not inverted. To better represent the pinhole camera model and align with the actual scenario, this paper adopts a convention where the camera image plane is symmetrically placed in front of the camera sensor, as illustrated in **Figure 3**. This convention ensures that the resulting image is not inverted and corresponds to the observed scene in a more intuitive manner.

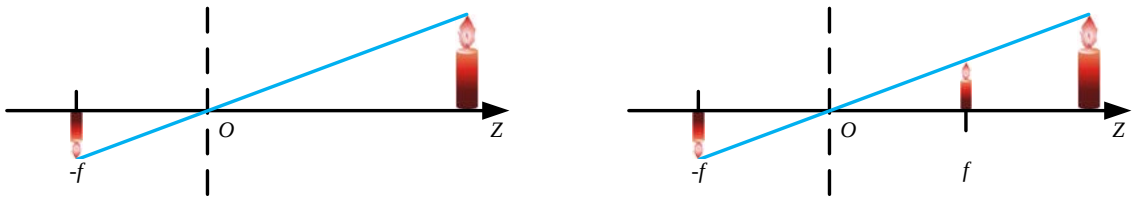


Figure 3. Schematic diagram of both the actual image plane and the symmetric image plane in the pinhole camera model.

By removing the symbols in Equation (1) and rearranging the terms, we can obtain the Equation (2):

$$\begin{cases} X' = f\dfrac{X}{Z} \\ Y' = f\dfrac{Y}{Z} \end{cases} \tag{2}$$

Equation (2) describes the three-dimensional spatial relationship between the 3D landmark point *P* and its image point *P'*. However, when working with camera sensors, the information is acquired in the form of individual pixels. Therefore, it is necessary to perform translation and scaling on the image plane coordinate system *O'x'y'* to convert it into the pixel coordinate system *ouv*, as depicted in **Figure 1**. By assuming that the pixel coordinate system is scaled by α and β along the *u* and *v* axes, respectively, and translated by c_x and c_y, we can establish the relationship between the image point *P'* and its pixel coordinates $[u, v]^{\mathrm{T}}$ as follows:

$$\begin{aligned} u &= \alpha X' + c_x \\ v &= \beta Y' + c_y \end{aligned} \tag{3}$$

Let $f_x = \alpha f$ and $f_y = \beta f$. By rearranging the equation, we can obtain Equation (4):

$$\begin{cases} u = f_x \dfrac{X}{Z} + c_x \\ v = f_y \dfrac{Y}{Z} + c_y \end{cases} \tag{4}$$

By expressing Equation (4) in matrix form and simplifying, we have:

$$Z\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \triangleq \boldsymbol{KP} \tag{5}$$

where the matrix composed of the intermediate quantities is the camera's intrinsic parameter matrix $\boldsymbol{K}$. Based on coordinate transformation knowledge, the coordinates in the camera coordinate system and the coordinates in the world coordinate system $\boldsymbol{P}_w$ have the following relationship:

$$\boldsymbol{P} = \boldsymbol{RP}_w + \boldsymbol{t} \tag{6}$$

where $\boldsymbol{R}$ and $\boldsymbol{t}$ are the rotation matrix and translation matrix that transform coordinates from the world coordinate system to the camera coordinate system. The corresponding coordinates $\boldsymbol{P}_{uv}$ in the pixel coordinate system can be related to the world coordinate system, camera coordinate system, and pixel coordinate system as follows:

$$Z\boldsymbol{P}_{uv} = Z\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \boldsymbol{K}(\boldsymbol{RP}_w + \boldsymbol{t}) = \boldsymbol{KTP}_w \tag{7}$$

where $\boldsymbol{T} = [\boldsymbol{R}, \boldsymbol{t}]$ is the transformation matrix. In Equation (7), it specifically represents the transformation matrix from the world coordinate system to the pixel coordinate system. To facilitate the derivation of the camera sensor's intrinsic parameters, it is common to normalize Equation (7) by setting Z = 1 and perform calculations based on the projection on the normalized plane, as illustrated in **Figure 4**.

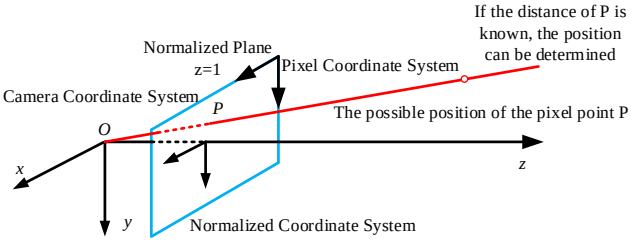


Figure 4. Schematic diagram of the normalized plane.

This paper utilizes two constraints to solve the intrinsic parameters. Firstly, when rotating the camera, the two rotation vectors along the camera center are orthogonal to each other. Secondly, the magnitude of the rotation vector is 1. Based on these constraints, the values of the parameters in the intrinsic parameter matrix can be obtained, as shown below.

$$\begin{cases} f_x = \sqrt{\lambda / B_{11}} \\ f_y = \sqrt{\lambda B_{11} / (B_{11}B_{22} - B_{12}^2)} \\ c_x = \gamma c_y / f_y - B_{13} f_x^2 / \lambda \\ c_y = (B_{12}B_{13} - B_{11}B_{23}) / (B_{11}B_{22} - B_{12}^2) \\ \gamma = -B_{12} f_x^2 f_y / \lambda \\ \lambda = B_{33} - \left[B_{13}^2 + c_y(B_{12}B_{13} - B_{11}B_{23})\right] / B_{11} \end{cases} \tag{8}$$

The $\boldsymbol{B}$ matrix is a user-defined matrix, and its definition is as follows:

$$\boldsymbol{B} = (\boldsymbol{K}^{-1})^T \boldsymbol{K}^{-1} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} \tag{9}$$

2.2. Distortion model

Additionally, the lens on the camera sensor introduces distortion during the projection from the three-dimensional world to the two-dimensional image plane[13]. The distortion is caused by the optical properties of the lens. Therefore, this paper also incorporates a distortion model to accurately describe the projection process from the three-dimensional world to the two-dimensional image plane.

Typically, a lens is placed in front of the camera, resulting in radial distortion as shown in **Figure 5**. Moreover, during the hardware assembly process, it is common for the lens and the imaging plane to have slight misalignment, leading to tangential distortion as depicted in **Figure 6**.

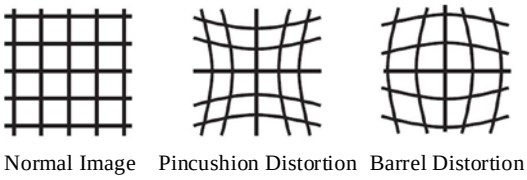


Figure 5. Schematic diagram of radial distortion.

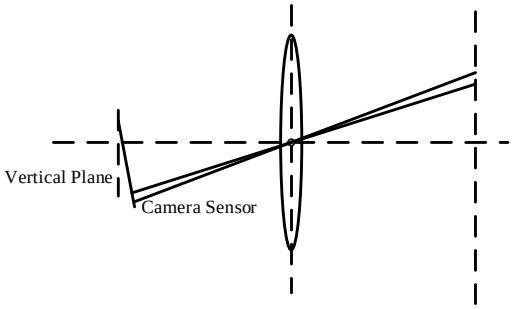


Figure 6. Schematic diagram of tangential distortion source.

Therefore, to improve the imaging results, it is necessary to perform distortion correction on the camera sensor. Assuming that tangential distortion and radial distortion follow a polynomial relationship, for any point $p = [x, y]^{\mathrm{T}}$ on the normalized plane, the corrected normalized coordinates $[x_{\mathrm{dis\_r}}, y_{\mathrm{dis\_r}}]^{\mathrm{T}}$ after radial distortion correction are related as follows:

$$x_{\mathrm{dis\_r}} = x(1 + k_1 r^2 + k_2 r^4) \tag{10}$$

$$y_{\mathrm{dis\_r}} = y(1 + k_1 r^2 + k_2 r^4) \tag{11}$$

where k_1 and k_2 are radial distortion coefficients, and r is the distance between point p and the coordinate origin in polar coordinates. Similarly, for tangential distortion, the corrected normalized coordinates $[x_{\mathrm{dis\_t}}, y_{\mathrm{dis\_t}}]^{\mathrm{T}}$ of point p can be related as follows:

$$x_{\mathrm{dis\_t}} = x + 2p_1 xy + p_2(r^2 + 2x^2) \tag{12}$$

$$y_{\mathrm{dis\_t}} = y + p_1(r^2 + 2y^2) + 2p_2 xy \tag{13}$$

where p_1 and p_2 are tangential distortion coefficients, and r is also the distance of point p from the origin in polar coordinates. Therefore, by combining the operations of radial and tangential distortion correction on points in the normalized plane, we obtain the distortion correction parameter matrix $\boldsymbol{U_d} = \begin{bmatrix} k_1 & k_2 \\ p_1 & p_2 \end{bmatrix}$. Consequently, we can obtain the normalized coordinates of the undistorted points $[x_{\mathrm{distorted}}, y_{\mathrm{distorted}}]^{\mathrm{T}}$ as follows:

$$x_{\text{distorted}} = x(1 + k_1 r^2 + k_2 r^4) + 2p_1 xy + p_2(r^2 + 2x^2) \tag{14}$$

$$y_{\text{distorted}} = y(1 + k_1 r^2 + k_2 r^4) + p_1(r^2 + 2y^2) + 2p_2 xy \tag{15}$$

For the distortion correction coefficients, rearranging them into the distortion correction parameter matrix, and substituting Equations (14) and (15) into Equation (4), we can derive the expression for projecting the undistorted points onto the pixel plane using the intrinsic parameter matrix. It is given as follows:

$$\begin{cases} u = f_x x_{\text{distorted}} + c_x \\ v = f_y y_{\text{distorted}} + c_y \end{cases} \tag{16}$$

By utilizing the pinhole camera model and distortion model, the three-dimensional landmarks can be projected onto the camera's two-dimensional image plane, enabling the estimation of the camera's intrinsic parameters.

3. IMU intrinsic parameters calibration

3.1. IMU measurement model

The measurement errors of an IMU mainly consist of Gaussian white noise and biases[1,14]. Therefore, this paper considers the random walk of biases and noise in modeling the IMU's measured acceleration $\widehat{\boldsymbol{a}}_t^b$ and angular velocity $\widehat{\boldsymbol{\omega}}_t^b$:

$$\widehat{\boldsymbol{a}}_t^b = \boldsymbol{a}_t^b + \boldsymbol{R}_{bw}\boldsymbol{g}^w + \boldsymbol{b}_{a_t} + \boldsymbol{n}_a \tag{17}$$

$$\widehat{\boldsymbol{\omega}}_t^b = \boldsymbol{\omega}_t^b + \boldsymbol{b}_{\omega_t} + \boldsymbol{n}_\omega \tag{18}$$

where $\boldsymbol{R}_{bw}$ is the rotation matrix from the world coordinate system to the IMU coordinate system, $\boldsymbol{g}^w = [0, 0, g]^T$ is the gravity vector in the world coordinate system, $\boldsymbol{b}_{a_t}$ is the acceleration bias, $\boldsymbol{n}_a$ is the acceleration noise, ; $\boldsymbol{b}_{\omega_t}$ is the angular velocity bias, and n_ω is the angular velocity noise. This paper assumes that the acceleration noise n_a and the angular velocity noise $\boldsymbol{n}_\omega$ follow a Gaussian distribution, i.e., $\boldsymbol{n}_a \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\sigma}_a{}^2)$, $\boldsymbol{n}_\omega \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\sigma}_\omega{}^2)$.

It is further assumed that the acceleration bias $\boldsymbol{b}_{a_t}$ and the angular velocity bias $\boldsymbol{b}_{\omega_t}$ are modeled as a bounded random walk and a random walk, respectively, with their first derivatives as follows:

$$\dot{\boldsymbol{b}_{a_t}} = -\frac{1}{\tau}\boldsymbol{b}_{a_t} + \boldsymbol{n}_{b_a} \tag{19}$$

$$\dot{\boldsymbol{b}_{\omega_t}} = \boldsymbol{n}_{b_\omega} \tag{20}$$

where τ is the time constant, $\boldsymbol{n}_{b_a}$ and $\boldsymbol{n}_{b_\omega}$ are the noise of the acceleration bias and the noise of the gyroscope bias, respectively, which follow a Gaussian distribution, i.e., $\boldsymbol{n}_{b_a} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\sigma}_{b_a}{}^2)$, $\boldsymbol{n}_{b_\omega} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\sigma}_{b_\omega}{}^2)$.

3.2. IMU errors

IMU intrinsic parameter calibration mainly involves parameters related to random errors and deterministic errors. The parameters related to random errors include noise and random walk. In the actual process of IMU sampling, discretization is required. This paper derives the relationship between the continuous and discrete representations of random errors[3,15].

Integrating white Gaussian noise over a fixed time interval Δt, we can write:

$$n_d(k) = n(t + \Delta t) \approx \frac{1}{\Delta t}\int_t^{t+\Delta t} n(\tau)\, dt \tag{21}$$

where the covariance is:

$$E\left(n_d^2(k)\right) = E\left(\frac{1}{\Delta t^2}\int_{t_0}^{t_0+\Delta t}\int_{t_0}^{t_0+\Delta t} n(\tau)n(t)\, d\tau dt\right) = \frac{\sigma^2}{\Delta t} \tag{22}$$

It can be observed from Equation (22) that the variance of the discrete Gaussian white noise can be obtained by dividing the continuous Gaussian white noise variance by a time interval. Similarly, for the integral of a bias random walk over a fixed time interval Δt, we can express it as follows:

$$b(t_0 + \Delta t) = b(t_0) + \int_{t_0}^{t_0+\Delta t} n_{b_\omega}(t)\, dt \tag{23}$$

The covariance can be expressed as:

$$E\left(b^2(t_0 + \Delta t)\right) = E\left(\left[b(t_0) + \int_{t_0}^{t_0+\Delta t} n_{b_\omega}(t)\, dt\right]\left[b(t_0) + \int_{t_0}^{t_0+\Delta t} n_{b_\omega}(\tau) d\tau\right]\right) = \sigma_b^2 \Delta t \tag{24}$$

Similarly, it can be seen from Equation (22) that the variance of the discrete Gaussian white noise can be obtained by simply multiplying the continuous Gaussian white noise variance by a time interval.

The parameters related to deterministic errors are mainly bias, scale, and misalignment coefficients. In this paper, the parameters to be optimized are obtained by constructing loss functions for the accelerometer and gyroscope, respectively.

$$L(\boldsymbol{\theta}_{acc}) = \sum_{k=1}^{M}\left(\|\boldsymbol{g}\|^2 - \left\|h(\boldsymbol{a}_k^b, \boldsymbol{\theta}_{acc})\right\|^2\right)^2 \tag{25}$$

$$L(\boldsymbol{\theta}_{gyro}) = \sum_{k=2}^{M}\left\|\boldsymbol{u}_{a,k} - \boldsymbol{u}_{g,k}\right\|^2 \tag{26}$$

where $\boldsymbol{g}$ is the gravity vector; $h(\boldsymbol{a}_k^b, \boldsymbol{\theta}_{acc})$ describes the relationship between the IMU's measured acceleration $\boldsymbol{a}_k^b$ and the true acceleration in the world, given the estimated parameters $\boldsymbol{\theta}_{acc}$ for the accelerometers. $\boldsymbol{\theta}_{gyro}$ represents the estimated parameters for the gyroscopes. M is the number of intervals considered for computation. $\boldsymbol{u}_{a,k}$ is the initial gravity estimated from the calibrated accelerometer, and $\boldsymbol{u}_{g,k}$ is the calculated gravity, as shown below.

$$\boldsymbol{u}_{g,k} = \psi[\boldsymbol{\omega}_k^b, \boldsymbol{u}_{a,k-1}] \tag{27}$$

4. Camera-IMU calibration

4.1. IMU pre-intergration model

Since the sampling frequency of the IMU is significantly higher than that of the camera sensor, this paper performs IMU pre-integration by integrating the IMU measurements between two consecutive image frames. This enables the alignment of the sampling frequencies of different sensors (camera and IMU). The schematic diagram of the IMU pre-integration model can be seen in **Figure 7**.

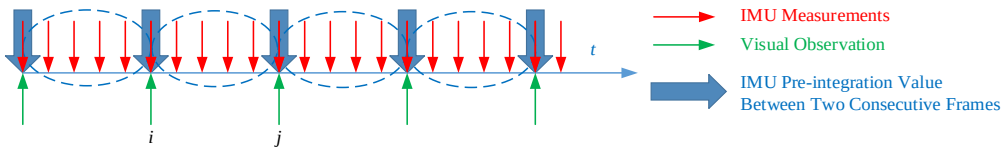


Figure 7. Schematic diagram of IMU pre-integration model.

Based on the IMU measurement model equations presented in Section 2.1, the position, velocity, and rotation information in the world coordinate system can be calculated as follows:

$$\boldsymbol{p}_{wb_j} = \boldsymbol{p}_{wb_i} + \boldsymbol{v}_i^w \Delta t_i + \iint_{t\in[i,j]} (\boldsymbol{R}_{wb_t}(\hat{\boldsymbol{a}}_t^b - \boldsymbol{b}_{a_t} - \boldsymbol{n}_a) - \boldsymbol{g}^w)\, dt^2 \tag{28}$$

$$\boldsymbol{v}_{b_j}^w = \boldsymbol{v}_{b_i}^w + \int_{t\in[i,j]} (\boldsymbol{R}_{wb_t}(\hat{\boldsymbol{a}}_t^b - \boldsymbol{b}_{a_t} - \boldsymbol{n}_a) - \boldsymbol{g}^w)\, dt \tag{29}$$

$$\boldsymbol{q}_{wb_j} = \int_{t\in[i,j]} \boldsymbol{q}_{wb_t} \otimes \begin{bmatrix} 0 \\ \frac{1}{2}\boldsymbol{\omega}^{b_t} \end{bmatrix} dt \tag{30}$$

where $\boldsymbol{p}_{wb_j}$ and $\boldsymbol{p}_{wb_i}$ represent the translation from the IMU coordinate system to the world coordinate system at time *j* and time *i*, respectively. $\boldsymbol{v}_{b_j}^w$ and $\boldsymbol{v}_{b_i}^w$ represent the IMU's velocity information in the world coordinate system at time *j* and time *i*, respectively. $\boldsymbol{q}_{wb_j}$ and $\boldsymbol{q}_{wb_i}$ are the rotation quaternions from the IMU coordinate system to the world coordinate system at time *j* and time *i*, respectively. Δt_i is the time interval between the *i*-th and *j*-th image frames. The $\otimes$ symbol denotes the multiplication operation of quaternions.

By applying the rotation relationship of the coordinate systems, this paper rotates Equations (28)–(30) from the world coordinate system to the IMU coordinate system of the *i*-th frame. As a result, the continuous-time IMU pre-integration formula can be obtained as follows:

$$\boldsymbol{R}_{b_i w}\boldsymbol{p}_{wb_j} = \boldsymbol{R}_{b_i w}\left(\boldsymbol{p}_{wb_i} + \boldsymbol{v}_i^w \Delta t_i - \frac{1}{2}\boldsymbol{g}^w \Delta {t_i}^2\right) + \boldsymbol{\alpha}_{b_i b_j} \tag{31}$$

$$\boldsymbol{R}_{b_i w}\boldsymbol{v}_{b_j}^w = \boldsymbol{R}_{b_i w}(\boldsymbol{v}_{b_i}^w - \boldsymbol{g}^w \Delta t_i) + \boldsymbol{\beta}_{b_i b_j} \tag{32}$$

$$\boldsymbol{q}_{b_i w} \otimes \boldsymbol{q}_{wb_j} = \boldsymbol{\gamma}_{b_i b_j} \tag{33}$$

The symbols in the above equations are represented as follows:

$$\boldsymbol{\alpha}_{b_i b_j} = \iint_{t\in[i,j]} \boldsymbol{R}_{b_i b_t}(\hat{\boldsymbol{a}}_t^b - \boldsymbol{b}_{a_t} - \boldsymbol{n}_a)\, dt^2 \tag{34}$$

$$\boldsymbol{\beta}_{b_i b_j} = \int_{t\in[i,j]} \boldsymbol{R}_{b_i b_t}(\hat{\boldsymbol{a}}_t^b - \boldsymbol{b}_{a_t} - \boldsymbol{n}_a)\, dt \tag{35}$$

$$\boldsymbol{\gamma}_{b_i b_j} = \int_{t\in[i,j]} \frac{1}{2}\boldsymbol{\Omega}(\hat{\boldsymbol{\omega}}_t^b - \boldsymbol{b}_{\omega_t} - \boldsymbol{n}_\omega)\boldsymbol{\gamma}_{b_i b_t}\, dt \tag{36}$$

For $\boldsymbol{\Omega}(\boldsymbol{\omega})$ in the Equation (36), the definition in this paper is as follows:

$$\boldsymbol{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix} -[\boldsymbol{\omega}]^\wedge & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^T & \boldsymbol{0} \end{bmatrix} \tag{37}$$

$$[\boldsymbol{\omega}]^{\wedge} = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \tag{38}$$

where ω_x, ω_y, and ω_z are the components of $\boldsymbol{\omega}$ on the corresponding coordinate axes, respectively.

4.2. Camera-IMU extrinsic parameters solving

After obtaining the camera intrinsics and IMU intrinsics through separate calibrations, the joint calibration is performed using the measurements from both sensors to determine their extrinsic parameters. The extrinsic parameters of the camera and IMU refer to the transformation matrix between the camera coordinate system and the IMU coordinate system. The external parameter calibration of the camera and the IMU is very important for the VSLAM system. Even if there is only a 1–2 deviation between the final camera coordinate system and the IMU coordinate system, the positioning accuracy of the VSLAM system will be significantly affected. Therefore, it is necessary to perform calibration of the external parameters of the camera and IMU. The coordinate relationship between the camera and the IMU is shown in **Figure 8**.

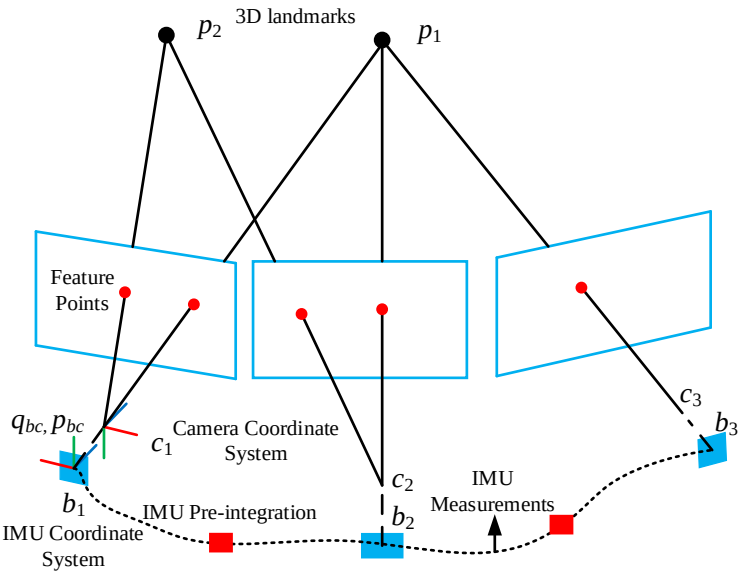


Figure 8. Schematic diagram of coordinate relationship between the camera and the IMU.

According to the IMU pre-integration model in Section 3.1, the IMU rotation matrix $\boldsymbol{R}_{b_k b_{k+1}}$ from time $k + 1$ to time k is obtained, and the camera frame at any moment satisfies the following equation:

$$\boldsymbol{R}_{b_k b_{k+1}} \boldsymbol{R}_{bc} = \boldsymbol{R}_{bc} \boldsymbol{R}_{c_k c_{k+1}} \tag{39}$$

where $\boldsymbol{R}_{bc}$ is the rotation matrix transformed from the camera coordinate system to the IMU coordinate system; $\boldsymbol{R}_{c_k c_{k+1}}$ is the camera rotation matrix from time $k + 1$ to time k. Equation (39) can be written in the form of quaternions as follows:

$$\boldsymbol{q}_{b_k b_{k+1}} \otimes \boldsymbol{q}_{bc} = \boldsymbol{q}_{bc} \otimes \boldsymbol{q}_{c_k c_{k+1}} \tag{40}$$

According to the properties of quaternion operations, Equation (40) can be transformed into the form of quaternion left multiplication matrix and right multiplication matrix, as shown below.

$$\left(\left[\boldsymbol{q}_{b_k b_{k+1}} \right]_L - \left[\boldsymbol{q}_{c_k c_{k+1}} \right]_R \right) \boldsymbol{q}_{bc} = \boldsymbol{Q}_{k+1}^{k} \boldsymbol{q}_{bc} = 0 \tag{41}$$

where $\left[\boldsymbol{q}_{b_k b_{k+1}}\right]_L$ and $\left[\boldsymbol{q}_{c_k c_{k+1}}\right]_R$ are the left multiplication matrix and right multiplication matrix of quaternion respectively. For *N* pairs of measurements, the following system of equation can be constructed:

$$\begin{bmatrix} w_1^0 \boldsymbol{Q}_1^0 \\ w_2^1 \boldsymbol{Q}_2^1 \\ \vdots \\ w_N^{N-1} \boldsymbol{Q}_N^{N-1} \end{bmatrix} \boldsymbol{q}_{bc} = 0 \tag{42}$$

where w_{k+1}^k is the Huber robust kernel function, which is used to represent the weight of removing outliers, expressed as follows:

$$w_{k+1}^k = \begin{cases} 1, & r_{k+1}^k < r_{thr} \\ \dfrac{r_{thr}}{r_{k+1}^k}, & \text{others} \end{cases} \tag{43}$$

where r_{thr} is the angle residual threshold of the rotation matrix, which is set to 5° in this paper; r_{k+1}^k is the angle residual of the rotation matrix, defined as follows:

$$r_{k+1}^k = arccos\left(\frac{tr\left(\widehat{\boldsymbol{R}}_{bc}^{-1} \boldsymbol{R}_{b_k b_{k+1}}^{-1} \widehat{\boldsymbol{R}}_{bc} \boldsymbol{R}_{c_k c_{k+1}}\right) - 1}{2}\right) \tag{44}$$

So far, the extrinsic parameter $\boldsymbol{q}_{bc}$ of the camera and IMU can be obtained by SVD decomposition of Equation (44).

5. Experiments

In this paper, a notebook computer MateBook D is used as the carrier, and its computer configuration is shown in **Table 1**. The experimental operating environment utilizes the Robot Operating System (ROS)—Kinetic[3] version on the Ubuntu 16.04 system. The proposed camera-IMU calibration method is implemented through C++ programming[1].

Table 1. Test computer equipment parameters.

Configuration type	Configuration information
CPU	i5-1135G7 processor
CPU graphics card	Intel integrated graphics
Processor base frequency	2.4 GHz
Running memory	16 GB

The MYNT EYE Camera D1010-50 integrates both the camera and IMU, as illustrated in **Figure 9**. In this paper, only the left-eye camera and IMU sensor of the MYNT EYE camera are utilized. The definition of the camera sensor and IMU coordinate systems can be found in **Figure 10**.

The coordinate systems of the camera and IMU are defined as follows:

1) Camera: The *x*-axis corresponds to the right-right direction of the camera, the *y*-axis corresponds to the direction directly behind the camera, and the *z*-axis corresponds to the direction directly above the camera. This coordinate system follows the right-handed convention.

2) The *x*-axis corresponds to the direction directly to the right of the camera, the *y*-axis corresponds to the direction directly in front of the camera, and the *z*-axis corresponds to the direction directly below the camera. This coordinate system also follows the right-handed convention.

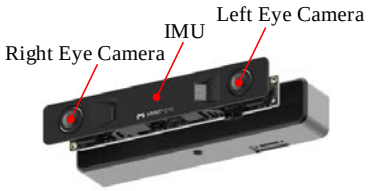


Figure 9. Schematic diagram of the pinhole camera model.

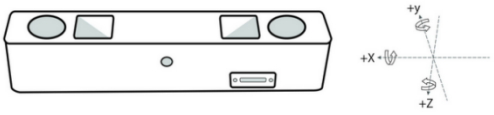


Figure 10. Schematic diagram of similar triangles.

The performance parameter information of MYNT EYE Camera D1010-50 is shown in **Table 2**.

Table 2. Performance Parameters of MYNT EYE Camera D1010-50.

Configuration type	Configuration information
Camera frame rate	10 Hz
Resolution	1280 × 720
Focal length	3.9 mm
IMU frequency	200 Hz
Output data format	YUYV/MJPG
Output accuracy	≤2.5%

5.1. Experiment of camera intrinsic parameter calibration

In this paper, Aprilgrid calibration boards are utilized for calibrating the internal parameters of the camera. The calibration board consists of a grid of Apriltags with a size of 6 × 6. Each Apriltag has a tagSize of 0.021 m, and there is a tagSpacing of 0.0063 m between adjacent Apriltags. During the calibration process, the camera is kept stationary while the calibration board is continuously moved and rotated in various poses. The camera captures images of the calibration board from different angles and orientations, as illustrated in **Figure 11**.

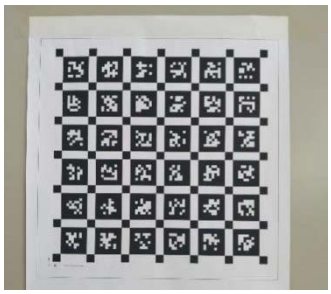


Figure 11. Aprilgrid calibration boards.

The internal parameter matrix $\boldsymbol{K}$ and distortion correction parameter matrix $\boldsymbol{U}_d$ of the Xiaomi camera are calculated by the method proposed in this paper, as follows:

$$\boldsymbol{K}=\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}=\begin{bmatrix} 1091.635505127837 & 0 & 615.7646844724167 \\ 0 & 1094.097509334247 & 336.08607722962415 \\ 0 & 0 & 1 \end{bmatrix} \quad (45)$$

$$\boldsymbol{U_d} = \begin{bmatrix} k_1 & k_2 \\ p_1 & p_2 \end{bmatrix} = \begin{bmatrix} 0.0158121998824731 & -0.04906947859369595 \\ -0.007932332725861788 & -0.0036593828274275953 \end{bmatrix} \tag{46}$$

After the camera is calibrated, it is necessary to verify the obtained parameter matrix, and its polar angle and azimuth angle errors are shown in **Figure 12**.

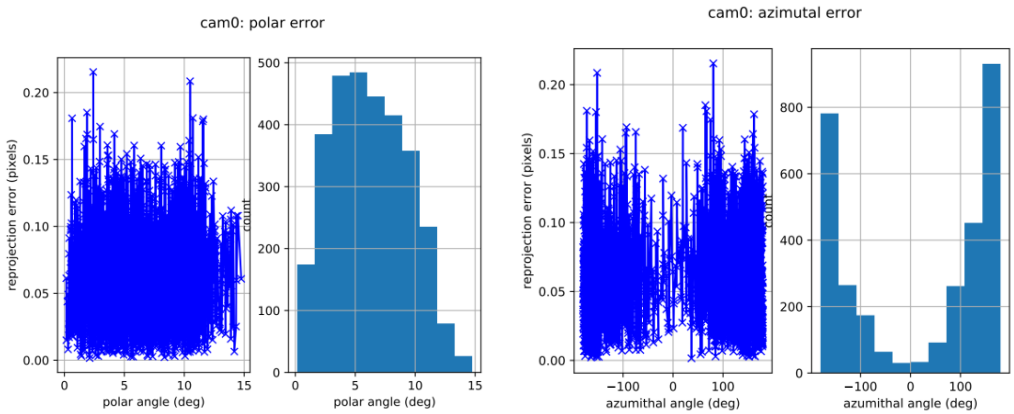
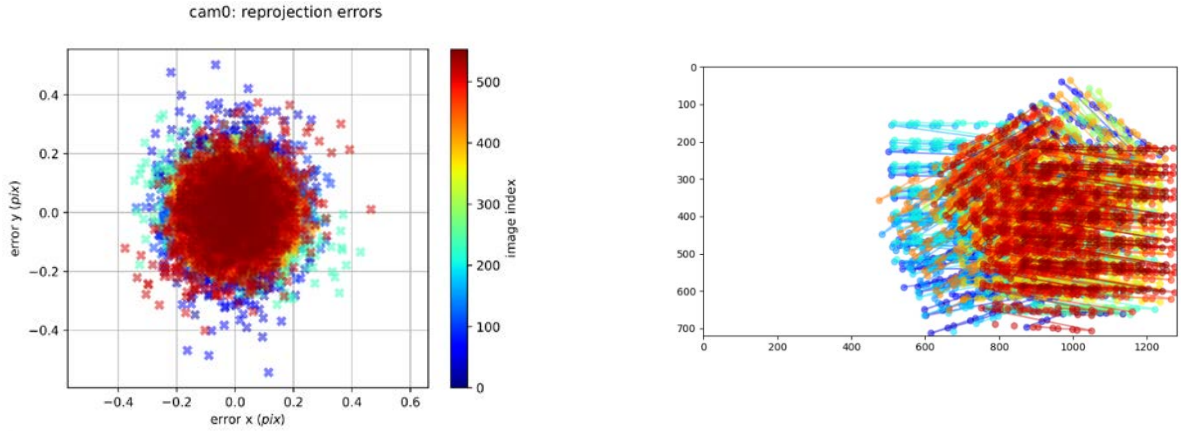


Figure 12. Error diagram of polar angle and azimuth angle.

The reprojection error of the camera intrinsics is shown in **Figure 13**. It can be seen from the figure that under this calibration internal reference, the maximum reprojection error does not exceed 0.5 pixels, which is within the allowable range of 1 pixel, so this calibration is valid and it is used as the internal reference of the MYNT EYE Camera D1010-50.



(1) Reprojected residuals diagram. (2) Camera movement diagram.

Figure 13. Camera reprojection error diagrams.

5.2. Experiment of IMU intrinsic parameter calibration

In this paper, the MYNT camera was left still and recorded data packets for 120 min, in which the data volume of the recorded accelerometer and angular velocity meter on the *xyz* axis was 1,358,672. The calculated IMU internal parameters are shown in **Table 3**.

The average value of gyroscope noise, random walk, accelerometer noise, and random walk can be calculated from the above table, as shown in **Table 4**.

Table 3. IMU random error calibration results.

Category	Axis	Values
Gyro noise (unit: rad/s)	*x*	$1.7150288432418979 \times 10^{-3}$
	y	$2.2810285332151279 \times 10^{-3}$
	z	$1.4803156902749898 \times 10^{-3}$
Accelerometer noise (unit: m/s^2)	*x*	$2.0263473443483032 \times 10^{-3}$
	y	$1.9452407902823307 \times 10^{-2}$
	z	$2.0800966656033096 \times 10^{-2}$
Gyro random walk (unit: rad/s)	*x*	$3.0555539333903460 \times 10^{-5}$
	y	$4.6905078690435216 \times 10^{-5}$
	z	$9.4998484143845586 \times 10^{-6}$
Accelerometer random walk (unit: m/s^2)	*x*	$3.3002849804958574 \times 10^{-4}$
	y	$3.5835234450158800 \times 10^{-4}$
	z	$3.3851545364446103 \times 10^{-4}$

Table 4. The average value of IMU calibration results.

Category	Parameters	Values
Gyro (unit: rad/s)	Gyro Noise	$1.8254576889106717 \times 10^{-3}$
	Gyro Random Walk	$2.8986822146241081 \times 10^{-5}$
Accelerometer (unit: m/s^2)	Accelerometer Noise	$2.0172282667446476 \times 10^{-2}$
	Accelerometer random walk	$3.4229876539854489 \times 10^{-4}$

The Allan variance method, introduced by David Allan of the National Bureau of Standards in the 1960s, is a time domain analysis technique[16–18]. It offers a convenient way to comprehensively characterize and identify various sources of error and their contributions to the overall statistical properties of noise. In this paper, the gyroscope and accelerometer Allan variances obtained during the IMU calibration process are depicted in **Figures 14** and **15**, respectively.

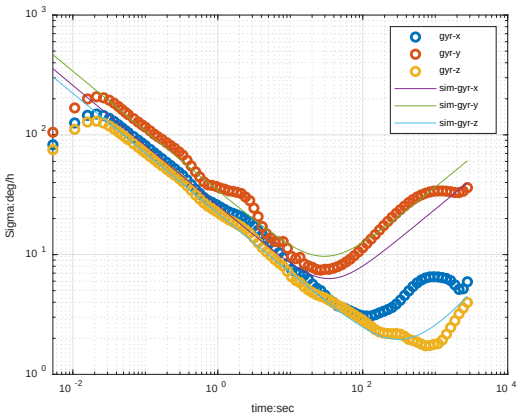


Figure 14. Diagram of Allan Variance for gyroscope.

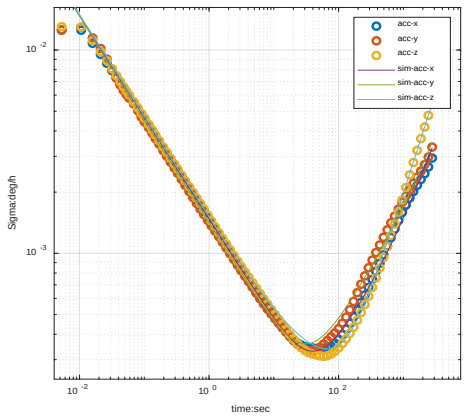


Figure 15. Diagram of Allan Variance for accelerator.

For the calibration of IMU deterministic errors, the process employed in this paper involves several steps: (1) placing the IMU on a table and allowing it to remain stationary for 50 sec; (2) moving and rotating the IMU in the air; (3) placing the IMU back on the desktop for 1 sec. This sequence is repeated approximately 50 times to complete the calibration process. **Figure 16** illustrates the recorded acceleration values and corresponding time intervals along the *x*, *y*, and *z* axes during the calibration procedure.

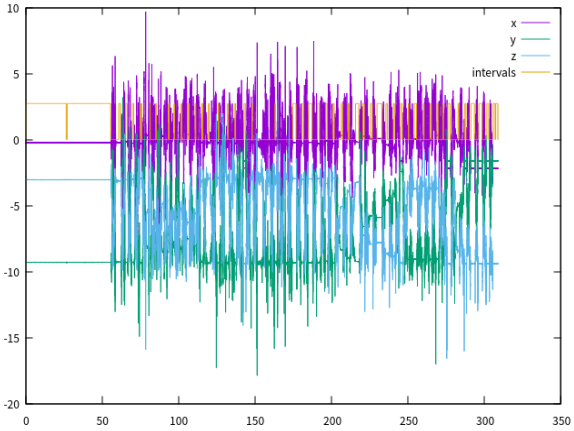


Figure 16. IMU *xyz* three-axis acceleration and time interval diagram.

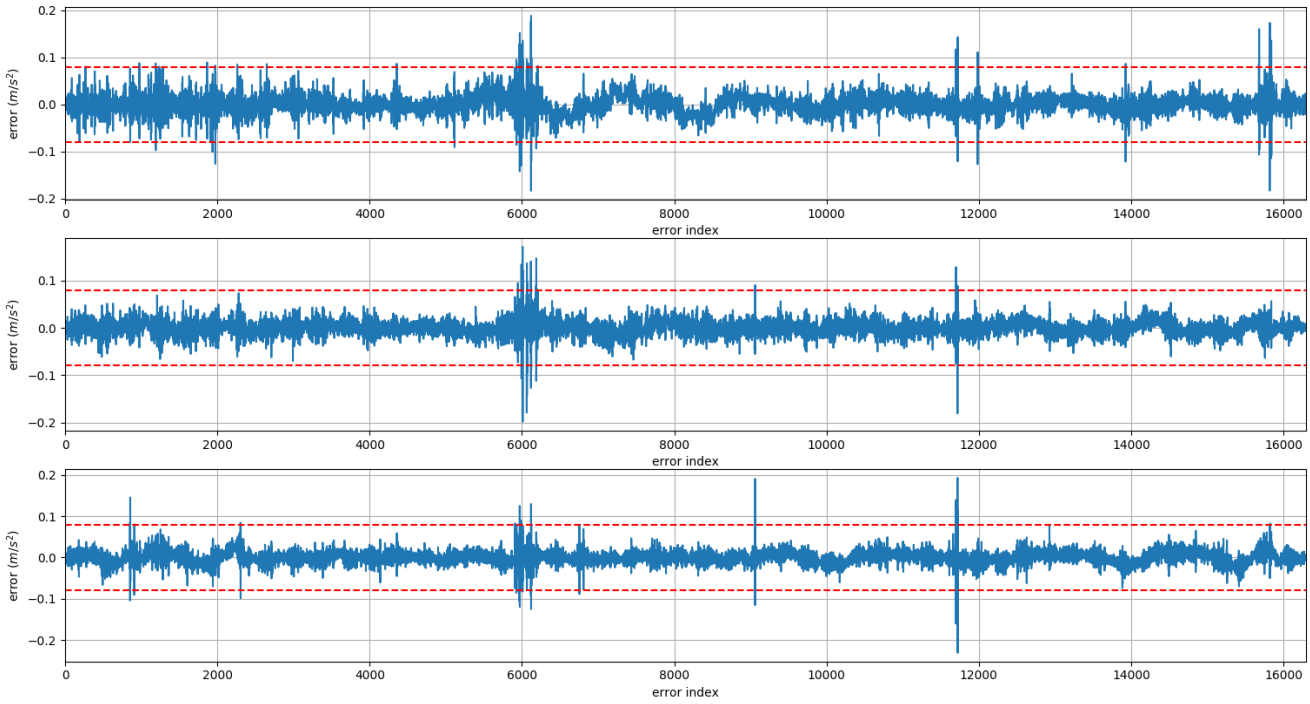
The computed parameters associated with the deterministic errors of the IMU are presented in **Table 5**.

Table 5. IMU deterministic error calibration results.

Category	**Parameters**	**Values**
Accelerometer bias	*x*-axix	14.1412
	y-axix	7.51009
	z-axix	4.37088
Gyroscope bias	*x*-axix	14.1412
	y-axix	7.51009
	z-axix	4.37088
Accelerometer scale matrix	$\boldsymbol{S}_{acc} = \begin{bmatrix} 0.636032 & 0 & 0 \\ 0 & 0.592957 & 0 \\ 0 & 0 & 0.666332 \end{bmatrix}$	
Accelerometer offset coefficient matrix	$\boldsymbol{M}_{acc} = \begin{bmatrix} 1 & -0.744004 & -0.3843 \\ 0 & 1 & -0.297786 \\ 0 & 0 & 1 \end{bmatrix}$	
Gyroscope scale matrix	$\boldsymbol{S}_{acc} = \begin{bmatrix} 0.636032 & 0 & 0 \\ 0 & 0.592957 & 0 \\ 0 & 0 & 0.666332 \end{bmatrix}$	
Gyroscope offset coefficient matrix	$\boldsymbol{M}_{acc} = \begin{bmatrix} 1 & -0.744004 & -0.3843 \\ 0 & 1 & -0.297786 \\ 0 & 0 & 1 \end{bmatrix}$	

5.3. Experiment of camera-IMU extrinsic parameter calibration

To validate the proposed calibration method for camera and IMU extrinsic parameters, this paper conducted an experiment by fixing the Aprilgrid calibration board. The MYNT camera recorded data packets of the calibration board's movement, rotation, and "8" pattern through ROS, resulting in a total of 1858 image data collected for this experiment. The acceleration error and angular velocity error during the calibration process are illustrated in **Figure 17** and **Figure 18**, respectively. The red dashed line represents the acceptable range of error.



Figuer 17. Acceleration error graph.

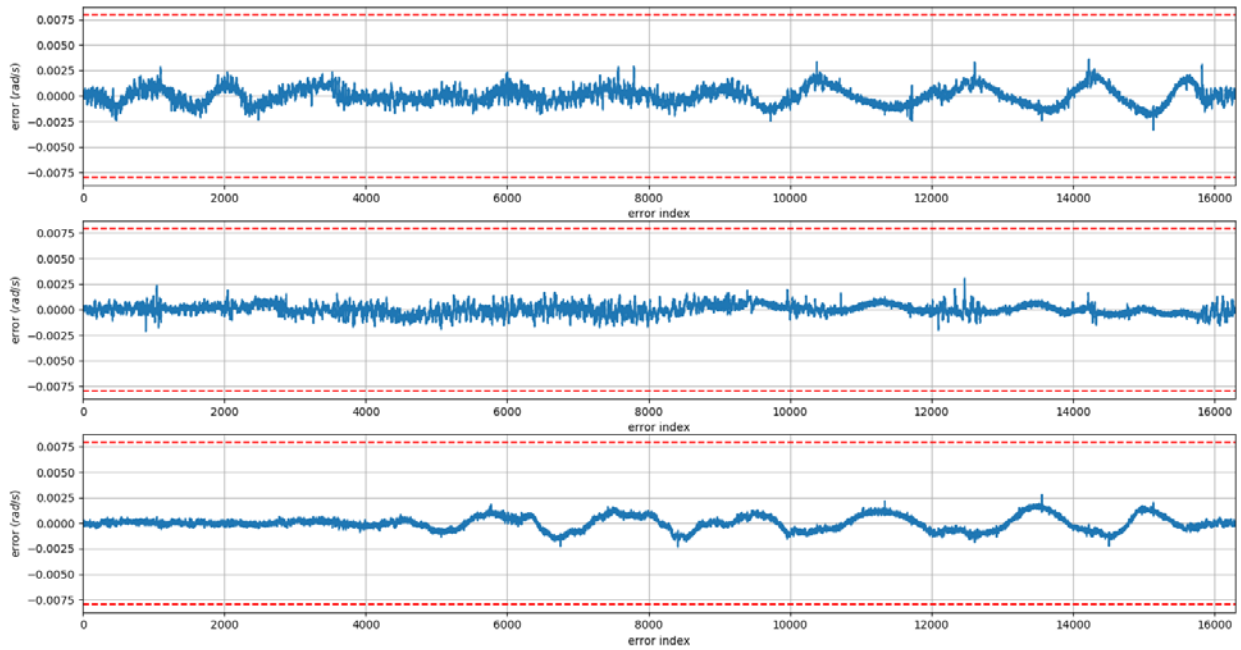


Figure 18. Angular velocity error graph.

From the figures, it is evident that the acceleration error and angular velocity error remain within the acceptable range. Furthermore, the estimated accelerometer biases and gyroscope biases are presented in **Figure 19** and **Figure 20**, respectively.

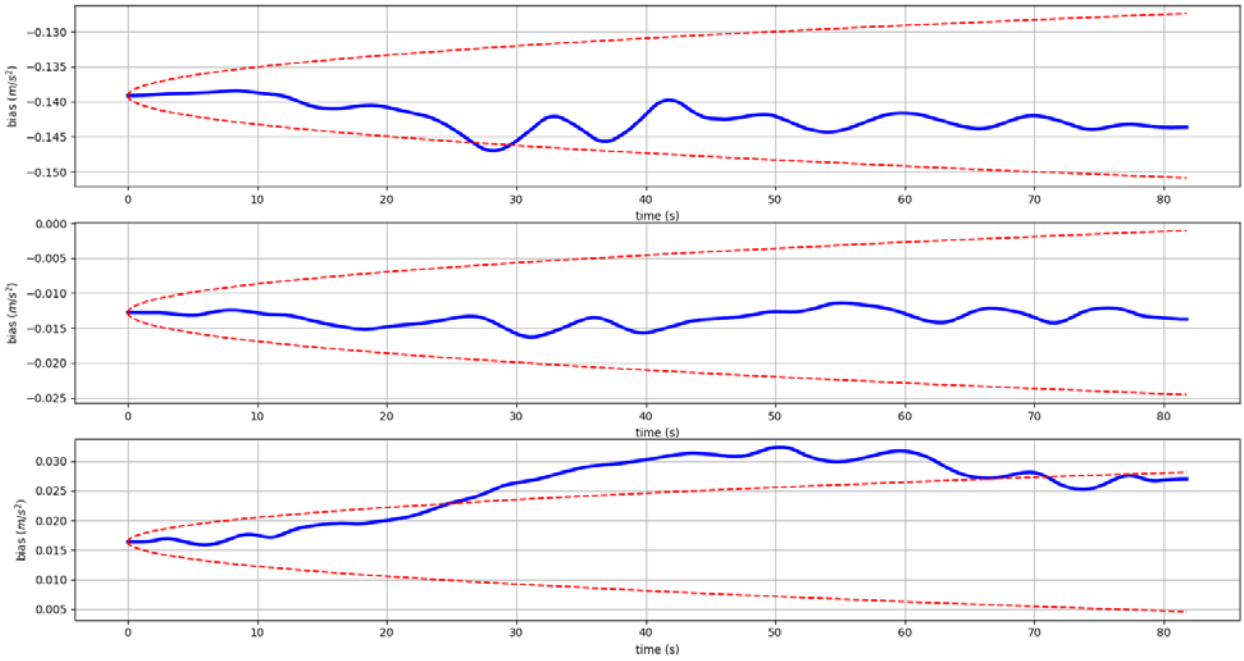


Figure 19. Accelerometer bias estimation graph.

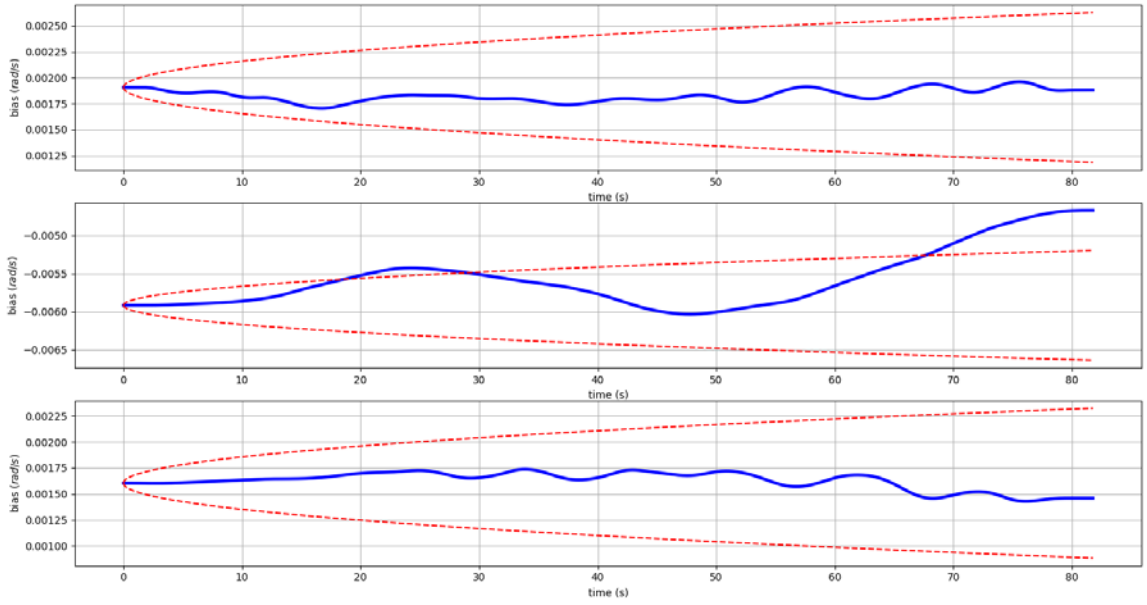


Figure 20. Gyroscope bias estimation graph.

In our study, we used the MYNT camera, which is a low-cost sensor priced at only $15. This choice of sensor is based on its affordability, but it also comes with lower precision in its IMU. It is important to note that while there are a few instances where the estimated values of the accelerometer and gyroscope bias exceed the predefined boundary, overall, these values converge and fall within a reasonable range. In addition, **Figure 21** and **Figure 22** present a comparison between the predicted and measured values of specific force and angular velocity, while **Figure 23** illustrates the reprojection error of the camera.

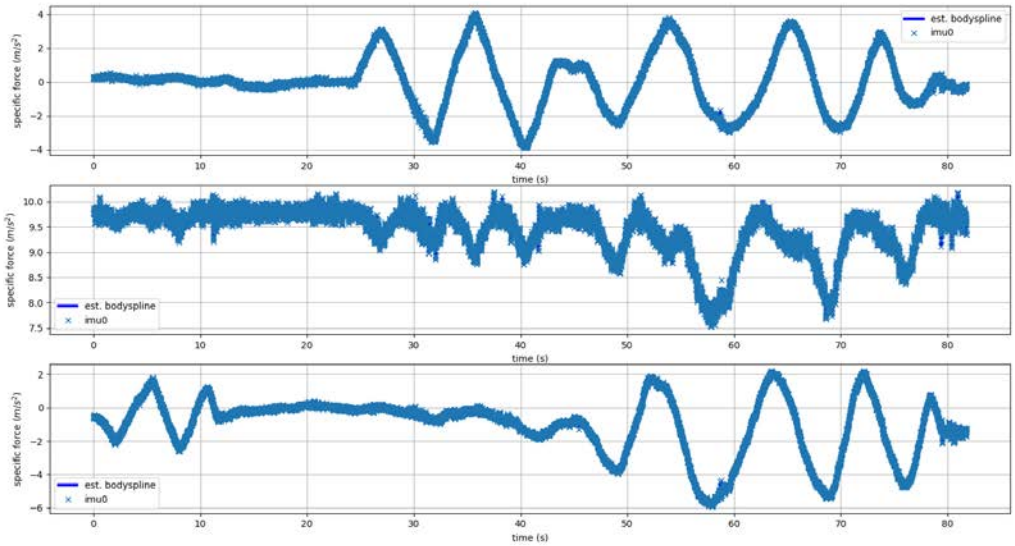


Figure 21. Comparison graph of predicted and measured specific force.

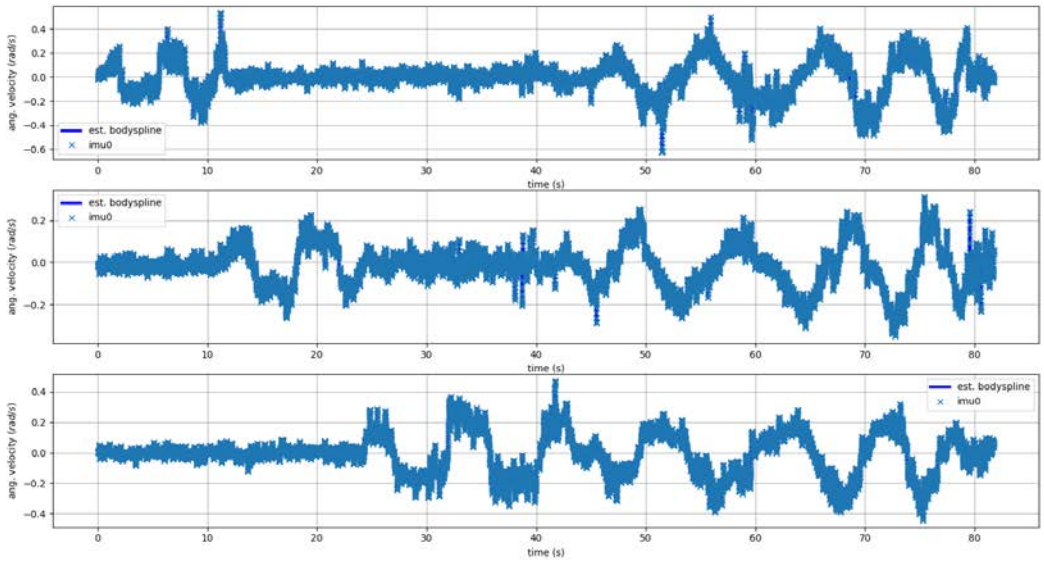


Figure 22. Comparison graph of predicted and measured angular velocity.

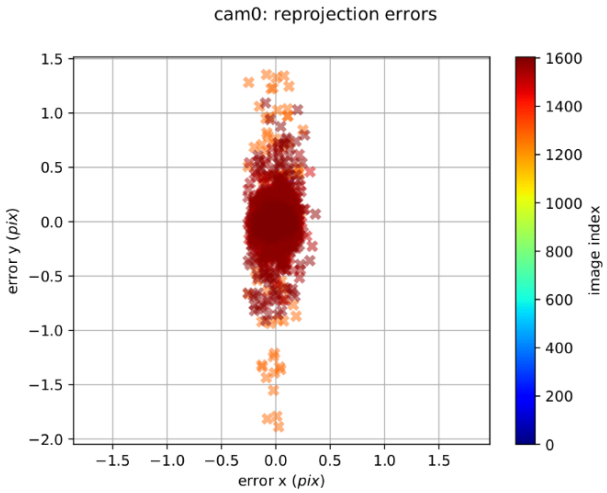


Figure 23. Camera reprojection error diagram.

The figures demonstrate a close agreement between the predicted and measured values of specific force and angular velocity, indicating good consistency. Additionally, the camera reprojection error remains within the acceptable range of 1 pixel. Consequently, the calibrated extrinsic parameters are deemed accurate, and the proposed methods in this paper are considered reliable. The rotation matrix, $\boldsymbol{R}_{bc}$, and the translation matrix, $\boldsymbol{t}_{bc}$, from the calibrated camera coordinate system to the IMU coordinate system are as follows.

$$\boldsymbol{R}_{bc} = \begin{bmatrix} 0.99976023688788251 & 0.012861567340825 & -0.0177214227256832 \\ 0.0137597595321047 & -0.998576606647093 & -0.0515308613822358 \\ -0.017033430526573 & 0.0517623486978128 & -0.99851415688601164 \end{bmatrix} \tag{47}$$

$$\boldsymbol{t}_{bc} = [0.05715507571041868 \quad 0.254279370393975 \quad 0.01424385318401530]^{\mathrm{T}} \tag{48}$$

6. Conclusion

In this paper, we presented a fast and accurate camera-IMU calibration method based on space coordinate transformation constraints and SVD (Singular Value Decomposition) tricks. First, we constructed constraint equations based on the equality of rotation and transformation matrices between camera frames and IMU coordinates at different moments. Secondly, we solved the external parameters of the camera-IMU using quaternion transformation and SVD techniques. The experimental results demonstrated the performance of our proposed camera-IMU calibration method. In the future, we will consider the characteristics of VSLAM systems to further model the measurement and noise of camera and IMU sensors.

Author contributions

Methodology: XT; software: XT; validation: XT; formal analysis: PM; investigation: PM; writing—review and editing: XT; visualization: BZ; supervision: BZ and JZ; project administration: BZ and JZ.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. Zhu B, Tao X, Zhao J, et al. An integrated GNSS/UWB/DR/VMM positioning strategy for intelligent vehicles. *IEEE Transactions on Vehicular Technology* 2020; 69(10): 10842–10853. doi: 10.1109/TVT.2020.3014516
2. Zhu B, Zhang P, Zhao J, et al. Hazardous scenario enhanced generation for automated vehicle testing based on optimization searching method. *IEEE Transactions on Intelligent Transportation Systems* 2021; 23(7): 7321–7331. doi: 10.1109/TITS.2021.3068784
3. Li X, Tao X, Zhu B, et al. Research on a simulation method of the millimeter wave radar virtual test environment for intelligent driving. *Sensors* 2020; 20(7): 1929. doi: 10.3390/s20071929
4. Tao X, Zhu B, Xuan S, et al. A multi-sensor fusion positioning strategy for intelligent vehicles using global pose graph optimization. *IEEE Transactions on Vehicular Technology* 2021; 71(3): 2614–2627. doi: 10.1109/TVT.2021.3139006
5. Zhu B, Tao X, Zhao J, et al. Two-stage UWB positioning algorithm of intelligent vehicle. *Journal of Traffic Engineering* 2021; 21(2): 256–266. doi: 10.19818/j.cnki.1671-1637.2021.02.022
6. Zhang Z. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2000; 22(11): 1330–1334. doi: 10.1109/34.888718
7. Heikkila J, Silvén O. A four-step camera calibration procedure with implicit image correction. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 17–19 June 1997; San Juan, USA. pp. 1106–1112.
8. Salvi J, Armangué X, Batlle J. A comparative review of camera calibrating methods with accuracy evaluation. *Pattern Recognition* 2002; 35(7): 1617–1635. doi: 10.1016/S0031-3203(01)00126-1

9. Zhang Z, Scaramuzza D. A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 1–5 October 2018; Madrid, Spain. pp. 7244–7251.
10. Brink K, Soloviev A. Filter-based calibration for an IMU and multi-camera system. In: Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium; 23–26 April 2012; Myrtle Beach, USA. pp. 730–739.
11. Sünderhauf N. *An Outlook on Robust Optimization for Sensor Fusion and Calibration: Switchable Constraints for Robust Simultaneous Localization and Mapping and Satellite-Based Localization*. Springer International Publishing; 2023. pp. 167–170.
12. Zhao J, Li Y, Zhu B, et al. Method and applications of LiDAR modeling for virtual testing of intelligent vehicles. *IEEE Transactions on Intelligent Transportation Systems* 2020; 22(5): 2990–3000. doi: 10.1109/TITS.2020.2978438
13. Zhu B, Han J, Zhao J, et al. Combined hierarchical learning framework for personalized automatic lane-changing. *IEEE Transactions on Intelligent Transportation Systems* 2020; 22(10): 6275–6285. doi: 10.1109/TITS.2020.2990787
14. Han J, Zhao J, Zhu B, et al. Adaptive steering torque coupling framework considering conflict resolution for human-machine shared driving. *IEEE Transactions on Intelligent Transportation Systems* 2021; 23(8): 10983–10995. doi: 10.1109/TITS.2021.3098466
15. Zhang P, Zhu B, Zhao J, et al. Performance evaluation method for automated driving system in logical scenario. *Automotive Innovation* 2022; 5(3): 299–310.
16. Jiang Y, Zhu B, Yang S, et al. Vehicle trajectory prediction considering driver uncertainty and vehicle dynamics based on dynamic Bayesian network. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 2022; 53(2): 689–703. doi: 10.1109/TSMC.2022.3186639
17. Zhao J, Du J, Zhu B, et al. Intelligent vehicle longitudinal cruise control based on adaptive dynamic sliding mode control. *Automotive Engineering* 2022; 44(1): 8–16. doi: 10.19562/j.chinasae.qcgc.2022.01.002
18. Zhu B, Wang Z, Zhao J, Tao X. Automatic parking space recognition method based on laser radar boundingboxes. In: Proceedings of the 2021 Annual Conference of the China Society of Automotive Engineers; 19–21 October 2021; Shanghai, China.