WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 2, Number 2, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v2n2.27
Journal homepage: https://wepub.org/index.php/IJCSIT/index

An Online Monocular Camera-IMU Calibration Method with Structural Feature Constraints

Yunmao Liao*, Chao Fang

School of Geomatics and Urban Spatial Information, Beijing University of Civil Engineering and Architecture, Beijing, China

*Corresponding Author: Yunmao Liao

ABSTRACT

Spatial configuration and time synchronization technology are the key to achieve high precision positioning of visual inertial system. The premise is to calibrate the internal and external parameters of the camera, the IMU (inertial measurement unit) bias and the external parameters between the camera and the IMU. In the traditional calibration method, the work of external parameter calibration is carried out offline, and it is generally completed in the initialization stage before being put into use, and it is assumed that the external parameters of each sensor will not change during the long-term operation. However, in complex practical working environments, sensor suites consisting of cameras and low-cost IMU often have time delay problems due to device processes, and visual inertia systems may be subjected to physical shocks from outside. Therefore, it is often necessary to face the failure of initial external parameters due to uncertain factors such as external shocks, time delays, mechanical structure adjustments or cumulative deformation under long-term work. In view of the above background, this paper constructs an online calibration model of monocular IMU based on the feature that the online calibration method can correct the external parameter deviation in real time and the constraint factor of the natural environment. Firstly, the external parameters of monocular camera and IMU are initialized based on physical space constraints, and the initial values to be optimized are obtained. Finally, the problem of matching the structural features in the online calibration process is solved by using the method of matching the structural features in the vertical main direction, and the objective optimization function of reprojection error considering the point and structure line is constructed. The Jacobian matrix of the reprojection error is given, and the decreasing direction of the optimization function is determined. The global optimal solution of the external parameters of camera and IMU is obtained in the online calibration.

KEYWORDS

Camera IMU Calibration; Structural Feature

1. INTRODUCTION

The research on location services has very important civil and military value. People's demand for location services is no longer simply satisfied with regional positioning, but more inclined to all-round, free from environmental location constraints, real-time positioning services, and the requirements for positioning accuracy are getting higher and higher. The technology of spatial positioning and time synchronization is the key to realize high precision navigation and positioning system. GNSS (global navigation satellite system) can provide efficient and stable navigation and positioning services in open outdoor environments, but it often fails to work properly indoors or in areas where satellite signals are blocked. Therefore, researchers have begun to explore positioning systems based on natural feature-based vision combined with low-cost IMU (inertial measurement

unit), hoping to achieve high-precision spatial positioning and time synchronization in a variety of environments. The visual inertial system can effectively solve the problems such as low precision, insufficient robustness and weak reliability of a single positioning source, and can realize continuous positioning in a variety of scenes, which is a necessary supplement to GNSS. However, in order to achieve high precision spatial configuration and time synchronization, the visual inertial system must be calibrated accurately.

The purpose of calibration is to solve the relative pose between the camera coordinate system and the IMU coordinate system, including a rotation quantity and a translation quantity, also known as the external parameter of the sensor, referred to as the external parameter. The precision of external parameters directly determines the precision of fusion positioning, so how to solve the accurate external parameters is the key problem in the research of camera and IMU fusion positioning technology. According to the operating cycle of the system, the calibration mode of external parameters between the camera and the IMU is usually offline and online. Usually in the initial stage of system operation, in order to obtain accurate sensor parameters, the involved sensors are calibrated offline. The off-line calibration method can provide more accurate calibration results, but it needs a specific site and has high learning cost. Device process, algorithmic reasoning and external environment are the main sources of error accumulation, which seriously affects the robustness of system operation and has a destructive impact on the system. Therefore, during the operation of the system, most of the online calibration method is used to correct the accumulated errors in real time, so as to realize the dynamic update of the external parameters between the camera and the IMU.

VIO (visual inertial odometry) and VI-SLAM (visual inertial simultaneous localization and mapping) technology not only provide accurate motion estimation, but also real-time correction of sensor parameters. Therefore, the external parameter calibration methods of camera and IMU based on VIO and VI-SLAM have attracted the attention of many researchers. Weiss et al. [1] proposed a UAV (unmanned aerial vehicle) navigation algorithm equipped with monocular camera and IMU, and used EKF (extended kalman filter) to estimate the external parameters of monocular camera and IMU online. The VIO algorithm of MSCKF (multi state constraint kalman filter) 2.0 proposed by Li et al. [2] also uses EKF to perform online estimation of external parameters of monocular camera and IMU. However, there is no systematic analysis of calibration performance in these papers, and the parameters of the sensor need to be known in advance, and the convergence of calibration parameters still depends on the accuracy of the initial value. For this reason, Dong-Si et al. [3] proposed an external parameter calibration method for monocular camera and IMU without prior knowledge. The method first carried out feasibility analysis on the rotation quantity, translation quantity and the order of parameters such as camera scale between calibration sensors, and then constructed each state quantity as a least squares problem about the reprojection error of feature points. However, the IMU bias is not taken into account, and the accumulation of bias over time will cause performance degradation. Based on Dong-Si, Yang et al. [4] adopted the sliding window strategy to linearly calibrate the rotation quantity, and linearly calibrate other parameters during initialization. Finally, all state quantities were optimized using nonlinear methods. In their subsequent VINS-mono [5], IMU bias was added. Huang et al. [6] proposed an online initialization method to calibrate the external parameters of monocular cameras and IMU by constructing linear equations, assuming that IMU bias is an unknown and fixed value. However, the external parameter becomes a fixed value after initialization, and the real-time calibration is not realized, and the time difference of heterogeneous sensor fusion is not considered. Since then, there have been many calibration methods that do not require any mechanical configuration parameters. For example, Yu et al. [7] used the known initial rotation quantity between the camera and the IMU, expressed the alignment model as a polynomial equation set based on Nomogram constraints through a first-order approximation of the rotation quantity, and recovered the rotation alignment parameters. Arbabmir et al. [8] used a hybrid optimization algorithm to carry out automatic calibration of camera internal calibration and camera IMU external calibration.

Through analysis, it can be seen that the above work does not make full use of the scene structure information in the natural environment, and ignores the localization difficulty of the camera in the low-texture environment. Three-dimensional structure information reflects the layout of the scene, the shape of the object and the relationship between the objects, which is very important for the understanding and reconstruction of the scene. As an important feature of scene structure, vanishing point was mainly used for camera position calibration [9], 3D scene reconstruction and target recovery in the early stage. Therefore, an online IMU calibration method for monocular cameras with structural feature constraints is proposed in this paper.

2. EXTRACTION AND MATCHING OF STRUCTURAL FEATURES

The line feature detection method usually uses the edge detection algorithm to extract the edge feature in the image, and then extracts the line segment by least square fitting or Hough transform. This kind of method is a global fitting algorithm, the detection accuracy depends on the accuracy of edge detection algorithm, and it is easy to cause error detection in the dense edge texture region, and the time cost is high. Considering the performance requirements of subsequent working point and line feature matching, this paper adopts LSD (line segment detector), which is real-time, does not need to adjust image parameters, and can control error detection and sub-pixel level detection accuracy [10]. The essence of LSD algorithm is to merge pixels with the same pixel gradient direction in a local area.

After the detection of line segments in the image, the direction of all line segments is classified, that is, the line segments in the same leading direction are parallel to each other and belong to the same vanishing point. Therefore, the vanishing point estimation algorithm mainly relies on the detected line segments in the image, and utilizes the characteristics of the perspective projection of parallel line segments. Usually based on EM (expectation maximization) [11], RANSAC (random sample consistency), ES (exhaustive searching) to estimate the optimal vanishing point results. However, the above vanishing point estimation algorithm has high requirements on time cost and initial vanishing point value, and it is easy to produce local optimal solution.

According to the definition of the vanishing point, the position of the vanishing point in the image is only related to the direction of the corresponding parallel line segment, and has nothing to do with the displacement of the corresponding parallel line segment. Therefore, the vanishing point estimation method in the Manhattan world hypothesis needs to satisfy the following two constraints: (1) the solution of the vanishing point should satisfy the global constraint, that is, it is not a locally optimal solution. (2) The solution of the vanishing point should satisfy the Manhattan world hypothesis, that is, the vanishing point should satisfy the orthogonal constraint. Therefore, this paper adopts the vanishing point real-time estimation algorithm which satisfies both global and orthogonal constraints. This algorithm uses 2 MSS (minimal solution set) to estimate the first vanishing point, relies on the orthogonality of the vanishing point to estimate the second vanishing point on the normal plane of the first vanishing point, and finally adopts the exhaustive search strategy to obtain the global optimal estimation result, and constructs the polar coordinate grid at the intersection of the straight lines. To achieve the purpose of speeding up exhaustive search. The specific flow of the algorithm is to use 2 line segments to estimate the first vanishing point, and then evenly extract the sample of the second vanishing point on the equivalent sphere of the first vanishing point, and finally calculate the third vanishing point through the cross product of the first vanishing point and the second vanishing point [12].

Then all line segments need to be classified for verification, that is, the construction of structural features. As show in Figure 1 , θ is the angle between the vanishing point and the line segment, VP_{img} is the projection position of the vanishing point in the image, P_1 and P_2 are the two endpoints of the line segment, and P_{mid} is the midpoint of the line segment. When $\theta < 6^{\circ}$, the line segment and the vanishing point belong to the same dominant direction.

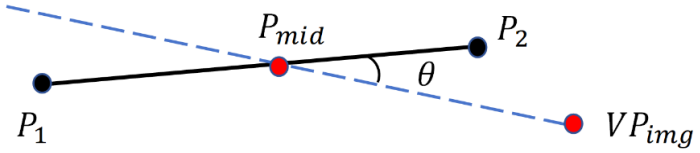

Figure 1. The Position of θ in Image

Finally, in order to accelerate the matching of structural features in nonlinear optimization, a matching method based on vertical principal direction is adopted in this paper. From the definition of structure lines, we can see that structure features are constrained by three orthogonal points, so the matching problem of structure features can be transformed into the matching problem of vanishing points. From the orthogonality of vanishing points, it can be seen that if the direction of a vanishing point is known, it is only necessary to match the direction of another vanishing point to complete the matching of phase structure features between adjacent image frames. The specific matching strategy is as follows: The structural feature values corresponding to continuous image frames I and J are converted to the world coordinate system. When the θ between the main direction of I and J is less than 6°, the matching is considered successful. If the matching fails, the relative rotation between image frames I and J is calculated using pre-integration, and the structural feature values in image frame I are converted to image frame J for re-matching.

3. INITIALIZATION OF MONOCULAR CAMERA – IMU

3.1. Initialization of the Rotation

According to the constraint relationship between monocular camera and IMU, extracting two aligned rotation sequences can obtain constant relative rotation. For monocular cameras, the rotation increment $R_{c_{k+1}}^{c_k}$ between two images can be calculated based on RANSAC's classic five-point method. At the same time, the relative rotation$R_{b_{k+1}}^{b_k}$ in the IMU coordinate system can be obtained by integrating the angular velocity measured by the IMU gyroscope

$$R_{b_{k+1}}^{b_k} \cdot R_b^c = R_b^c \cdot R_{c_{k+1}}^{c_k}. \tag{1}$$

The quaternion expression of formula (1) is

$$q_{b_{k+1}}^{b_k} \otimes q_c^b = q_c^b \otimes q_{c_{k+1}}^{c_k}. \tag{2}$$

Obtained by (2)

$$\left[\mathcal{Q}_1\left(q_{b_{k+1}}^{b_k}\right) - \mathcal{Q}_2\left(q_{c_{k+1}}^{c_k}\right)\right] \cdot q_c^b = Q_{k+1}^k \cdot q_c^b = 0. \tag{3}$$

$$\begin{cases} \mathcal{Q}_1(q) = \begin{bmatrix} q_\omega I_3 + \lfloor q_{XYZ} \times \rfloor & q_{XYZ} \\ -q_{XYZ} & q_\omega \end{bmatrix} \\ \mathcal{Q}_2(q) = \begin{bmatrix} q_\omega I_3 + \lfloor q_{XYZ} \times \rfloor & q_{XYZ} \\ -q_{XYZ} & q_\omega \end{bmatrix} \end{cases}. \tag{4}$$

Where, the matrices $\mathcal{Q}_1(q)$、$\mathcal{Q}_2(q)$ are left and right multiplications of quaternions respectively, $\lfloor q_{XYZ} \times \rfloor$ is the partial symmetric matrix of the first three elements q_{XYZ} of the quaternions, and $\otimes$ is the operator of four-element multiplication.

For the rotation increments between successive images, a linear overdetermined equation can be constructed as

$$\begin{bmatrix} \omega_1^0 \cdot Q_1^0 \\ \omega_2^1 \cdot Q_2^1 \\ \vdots \\ \omega_N^{N-1} \cdot Q_N^{N-1} \end{bmatrix} \cdot q_c^b = Q_N \cdot q_c^b = 0. \tag{5}$$

Where N is the minimum number of images required when the rotation matrix calibration is complete. In order to eliminate outliers in the rotation increment estimated by the five-point method, a weight ω_{k+1}^k is introduced. The rotation increment $\hat{R}_b^c$ estimated by the camera IMU is taken as the initial value and iteratively calculated with the incoming measurement value. Then the residual is weighted by the Huber norm, and the residual rotation matrix is represented by the angular norm of the angular axis

$$r_{k+1}^k = acos\left(\frac{\left(tr\left(\hat{R}_b^{c^{-1}} {R_{b_{k+1}}^{b_k}}^{-1} \hat{R}_b^c R_{c_{k+1}}^{c_k}\right) - 1\right)}{2}\right). \tag{6}$$

definition of weight is

$$\omega_{k+1}^k = \begin{cases} 1, r_{k+1}^k < t_0 \\ \dfrac{t_0}{r_{k+1}^k}, others \end{cases}. \tag{7}$$

Where t_0 is a fixed threshold, and the solution of the linear overdetermined equation is the right singular vector corresponding to the minimum singular value of Q_N. When the camera IMU system obtains enough rotation excitation, it indicates that the calibration work is successful, that is, a high-precision rotation increment R_c^b is obtained, and the null space of Q_N is one-dimensional. Therefore, in order to judge the convergence of the camera IMU rotation calibration algorithm, the second small singular value σ_R^{min2} of Q_N is used as the evaluation index. The calibration process is terminated when the singular threshold is $\sigma_R < \sigma_R^{min2}$.

3.2. Initialization of Scale, Gravity, Translation

After the initialization of the monocular camera-IMU rotation, the scale s of the monocular camera, the gravity vector g^w of the accelerometer, and the translation P_b^c between the monocular camera and the IMU can be further initialized. Since the bias and estimated value of the gyroscope are corrected during the pre-integration, it can be assumed that the gyroscope bias is a constant value, and $J_{\Delta\bar{P}}^g = 0$, $J_{\Delta\bar{v}}^g = 0$. According to the spatial conversion relationship between the camera and the IMU, the position relationship of two consecutive keyframes can be obtained

$$\mathrm{s} \cdot P_{c_{k+1}}^w = \mathrm{s} \cdot P_{c_k}^w + v_{b_k}^w \Delta t_{k,k+1} + \frac{1}{2} g^w \Delta t_{k,k+1}^2 + R_{c_k}^w \cdot \hat{R}_b^c \cdot \Delta\bar{P}_{b_{k+1}}^{b_k} - \left(R_{c_{k+1}}^w - R_{c_k}^w\right) P_b^c. \tag{8}$$

Since the zero bias of the accelerometer is not considered at this stage, s, g^w, P_b^c can be estimated by taking $R_{b_k}^w = R_{c_k}^w \cdot \hat{R}_b^c$.

In order to converge the iterative solution, the parameters can avoid too much mutual interference. For N velocity terms $v_{b_k}^w$ generated by N keyframes, two constraint relations of three consecutive keyframes can be used to construct a linear equation set

$$[\lambda(i) \quad \beta(i) \quad \varphi(i)] \begin{bmatrix} s \\ g^w \\ P_b^c \end{bmatrix} = \Upsilon(i). \tag{9}$$

Define key frames i, i + 1, and i + 2 as 1, 2, and 3 respectively

$$\begin{cases} \lambda(i) = \left(P_{c_2}^w - P_{c_1}^w\right)\Delta t_{23} - \left(P_{c_3}^w - P_{c_2}^w\right)\Delta t_{12} \\ \beta(i) = \frac{1}{2}(\Delta t_{12}\Delta t_{23}^2 + \Delta t_{12}^2 \Delta t_{23}) I_{3\times 3} \\ \varphi(i) = \left(R_{c_2}^w - R_{c_3}^w\right)\Delta t_{12} - \left(R_{c_1}^w - R_{c_2}^w\right)\Delta t_{23} \\ \Upsilon(i) = R_{c_1}^w \hat{R}_b^c \Delta\bar{P}_{b_2}^{b_1} \Delta t_{23} - R_{c_2}^w \hat{R}_b^c \Delta\bar{P}_{b_3}^{b_2} \Delta t_{12} - R_{c_1}^w \hat{R}_b^c \Delta\bar{v}_{b_2}^{b_1} \Delta t_{12}\Delta t_{23} \end{cases}. \tag{10}$$

For N consecutive keyframes, a linear constrained system of $3(N-2)$ equations with 7 unknowns can be generated

$$A_{3(N-2)\times 7} \cdot x_{7\times 1} = B_{3(N-2)\times 1}. \tag{11}$$

In the above formula, at least 5 key frames are needed to solve the scale $\hat{s}$, the gravity vector $\hat{g}^w$, and the translation vector $\hat{P}_b^c$ using the SVD (singular value decomposition) method.

4. ONLINE CALIBRATION OF MONOCULAR CAMERA-IMU

4.1. Reprojection Error of Line Segment

The external parameters obtained during initialization inevitably have errors. Therefore, after obtaining the initial values of the external parameters of the camera and IMU, a cost function needs to be constructed to optimize the external parameters online, and finally converge to a set of external parameters that meet the maximum likelihood estimation.

$$E = \text{argmin}\frac{1}{2}\left(\sum\left\|e_p\right\|^2 + \sum\|e_l\|^2\right). \tag{12}$$

Where e_p is the reprojection error of the point feature, and e_l is the reprojection error of the line feature.

4.2. The Jacobian Matrix of Reprojection Error

It is considered that the extremum of the cost function cannot be obtained directly by the general derivation method, and the optimal solution of the cost function is equivalent to the maximum likelihood estimation of the state quantity. Therefore, it is necessary to take the derivative of each state quantity and deduce the Jacobian matrix of each state quantity. For the reprojection error of characteristic points, there is

$$\frac{\partial e_p}{\partial P_k^C} = -\begin{bmatrix} \frac{f_x}{Z_C} & 0 & -\frac{f_x x_C}{Z_C^2} \\ 0 & \frac{f_y}{Z_C} & -\frac{f_y y_C}{Z_C^2} \end{bmatrix}. \tag{13}$$

Reprojection error for feature lines

$$\frac{\partial e_l}{\partial l'} = \begin{bmatrix} \frac{u_1 l_2^2 - l_1 l_2 v_1 - l_1 l_2}{(l_1^2 + l_2^2)^{\frac{3}{2}}} & \frac{v_1 l_2^2 - l_1 l_2 u_1 - l_2 l_3}{(l_1^2 + l_2^2)^{\frac{3}{2}}} & \frac{1}{(l_1^2 + l_2^2)^{\frac{1}{2}}} \\ \frac{u_1 l_2^2 - l_1 l_2 v_2 - l_1 l_3}{(l_1^2 + l_2^2)^{\frac{3}{2}}} & \frac{v_1 l_2^2 - l_1 l_2 v_2 - l_2 l_3}{(l_1^2 + l_2^2)^{\frac{3}{2}}} & \frac{1}{(l_1^2 + l_2^2)^{\frac{1}{2}}} \end{bmatrix}. \tag{14}$$

Where, (u_1, v_1), (u_2, v_2) are the endpoint coordinates of the characteristic line segment.

In order to optimize the external parameters between the camera and the IMU, according to the pose constraints between the camera and the IMU, there are

$$P_k^C = R_B^C R_B^{W^{-1}}(P_k^W - p_B^W) + p_B^C. \tag{15}$$

Assume that the translation vector of the monocular camera and the external parameter of the IMU has an attitude increment δp

$$P_k^C(\delta p) = R_B^C R_B^{W^{-1}}(P_k^W - p_B^W) + p_B^C + R_B^C \delta p = P_k^C + R_B^C \delta p. \tag{16}$$

Therefore, according to the chain rule, the Jacobian matrix of the reprojection error $e_{p,k}$ of the feature points with respect to the pose increment δp of the translation vector of the camera IMU external parameter is

$$\frac{\partial e_{p,k}}{\partial \delta P} = \frac{\partial e_{p,k}}{\partial P_k^C} \cdot \frac{\partial P_k^C}{\partial \delta P} = -\begin{bmatrix} \frac{f_x}{Z_C} & 0 & -\frac{f_x x_C}{Z_C^2} \\ 0 & \frac{f_y}{Z_C} & -\frac{f_y y_C}{Z_C^2} \end{bmatrix} R_B^C. \tag{17}$$

Similarly, assume that the rotation matrix of the camera IMU external parameter has a rotation increment $\delta\theta$

$$P_k^C(\delta\theta) = R_B^C Exp(\delta\theta) R_B^{W^{-1}}(P_k^W - p_B^W) + p_B^C. \tag{18}$$

The above formula is processed

$$P_k^C(\delta\theta) = Exp(R_B^C \delta\theta) R_B^C R_B^{W^{-1}}(P_k^W - p_B^W) + p_B^C. \tag{19}$$

To approximate $Exp(R_B^C \delta\theta)$, there is

$$Exp(R_B^C \delta\theta) \approx (I + (R_B^C \delta\theta)^\wedge) = R_B^C R_B^{W^{-1}} + (R_B^C \delta\theta)^\wedge. \tag{20}$$

Substitute the formula (88) into the formula (87) to get

$$P_k^C(\delta\theta) = p_k^C - \left[R_B^C R_B^{W^{-1}}(P_k^W - p_B^W)\right]^\wedge R_B^C \delta\theta. \tag{21}$$

Therefore, according to the chain rule, the Jacobian matrix of the reprojection error $e_{p,k}$ of the feature points with respect to the rotation matrix $\delta\theta$ of the camera IMU external parameter is

$$\frac{\partial e_{p,k}}{\partial \delta\theta} = \frac{\partial e_{p,k}}{\partial P_k^C} \cdot \frac{\partial P_k^C}{\partial \delta\theta} = \begin{bmatrix} \frac{f_x}{Z_C} & 0 & -\frac{f_x x_C}{Z_C^2} \\ 0 & \frac{f_y}{Z_C} & -\frac{f_y y_C}{Z_C^2} \end{bmatrix} \left[R_B^C R_B^{W^{-1}}(P_k^W - p_B^W)\right]^\wedge R_B^C. \tag{22}$$

Based on the line segment parameterization, the Jacobian matrix of the reprojection error $e_{l,k}$ of the feature line with respect to the pose increment δp of the external parameter translation vector of monocular IMU can be easily written

$$\frac{\partial e_{l,k}}{\partial \delta P} = \frac{\partial e_{l,k}}{\partial l'} \cdot \frac{\partial l'}{\partial \delta P} = \begin{bmatrix} \frac{u_1 l_2^2 - l_1 l_2 v_1 - l_1 l_2}{(l_1^2 + l_2^2)^{\frac{3}{2}}} & \frac{v_1 l_2^2 - l_1 l_2 u_1 - l_2 l_3}{(l_1^2 + l_2^2)^{\frac{3}{2}}} & \frac{1}{(l_1^2 + l_2^2)^{\frac{1}{2}}} \\ \frac{u_1 l_2^2 - l_1 l_2 v_2 - l_1 l_3}{(l_1^2 + l_2^2)^{\frac{3}{2}}} & \frac{v_1 l_2^2 - l_1 l_2 v_2 - l_2 l_3}{(l_1^2 + l_2^2)^{\frac{3}{2}}} & \frac{1}{(l_1^2 + l_2^2)^{\frac{1}{2}}} \end{bmatrix} R_B^C. \tag{23}$$

Similarly, the Jacobian matrix for the reprojection error $e_{l,k}$ of the feature line with respect to the pose increment $\delta\theta$ of the monocular IMU external parameter rotation matrix is

$$\frac{\partial e_{l,k}}{\partial \delta\theta} = \frac{\partial e_{l,k}}{\partial l'} \cdot \frac{\partial l'}{\partial \delta\theta} = \begin{bmatrix} \frac{u_1 l_2^2 - l_1 l_2 v_1 - l_1 l_2}{(l_1^2 + l_2^2)^{\frac{3}{2}}} & \frac{v_1 l_2^2 - l_1 l_2 u_1 - l_2 l_3}{(l_1^2 + l_2^2)^{\frac{3}{2}}} & \frac{1}{(l_1^2 + l_2^2)^{\frac{1}{2}}} \\ \frac{u_1 l_2^2 - l_1 l_2 v_2 - l_1 l_3}{(l_1^2 + l_2^2)^{\frac{3}{2}}} & \frac{v_1 l_2^2 - l_1 l_2 v_2 - l_2 l_3}{(l_1^2 + l_2^2)^{\frac{3}{2}}} & \frac{1}{(l_1^2 + l_2^2)^{\frac{1}{2}}} \end{bmatrix} \left[R_B^C R_B^{W^{-1}} (P_k^W - p_B^W)\right]^{\wedge} R_B^C \delta\theta. \tag{24}$$

5. EXPERIMENTAL VERIFICATION

5.1. Experimental Specification

The EuRoC [13] data set is a UAV as a carrier platform, carrying a binocular camera and an IMU flight measurement, which can provide kalibr tools to calibrate the external parameters between the monocular camera and the IMU. The EuRoC dataset consists of 11 data series, which are classified as easy, medium, and difficult based on drone flight speed, light, and environmental texture characteristics. The sampling frequency of binocular camera and IMU is 20 Hz and 200 Hz respectively, and time synchronization is realized by hardware. All the experiments are carried out with an Intel CPU i7-11800H laptop computer with 8GB RAM. Since the object of the study is to calibrate the external parameters between monocular camera and IMU online, only one set of images are taken out for experiment in this paper.

5.2. Experimental Analysis

First, this paper calibrates V2_01_easy sequence in EuRoC data set. The calibration results of V2_01_easy sequence by kalibr tool are shown in Table 1 as reference values.

Table 1. Calibrated Result by Kalibr

	Rotation (degree)			Translation (m)		
V2_01_easy	yaw 89.147953	pitch 1.476930	roll 0.215286	x -0.021640	y -0.064677	z 0.00981

Figure 2 shows the calibration results of the rotation estimation between the monocular camera and the IMU.

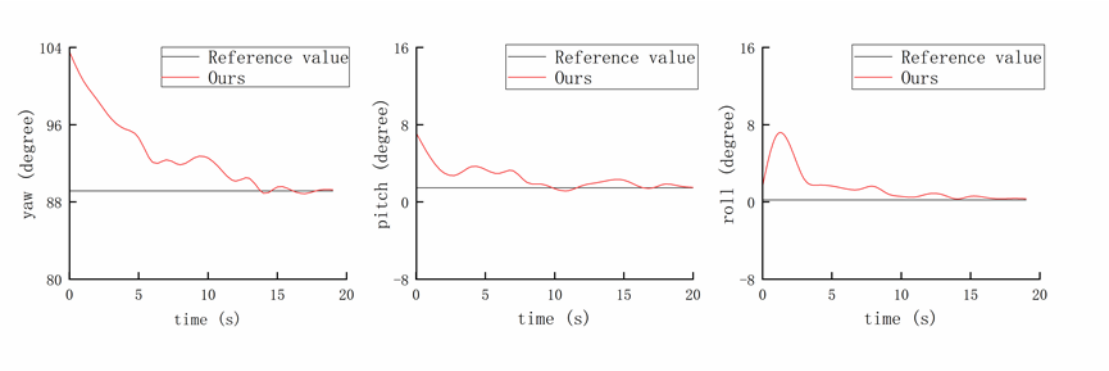

Figure 2. Rotation Initialization Result

As can be seen from Figure 2, for the visual inertial system without any prior conditions and relying only on motion excitation, the convergence time of the calibrated rotation matrix is linearly related to the eigenvalues. When sufficient eigenvalues are obtained, the rotation estimation is basically consistent with the reference value, and the convergence time is about 20s. Figure 3-5 shows the calibration results of translation vector, camera scale, and gravity vector between monocular camera and IMU.

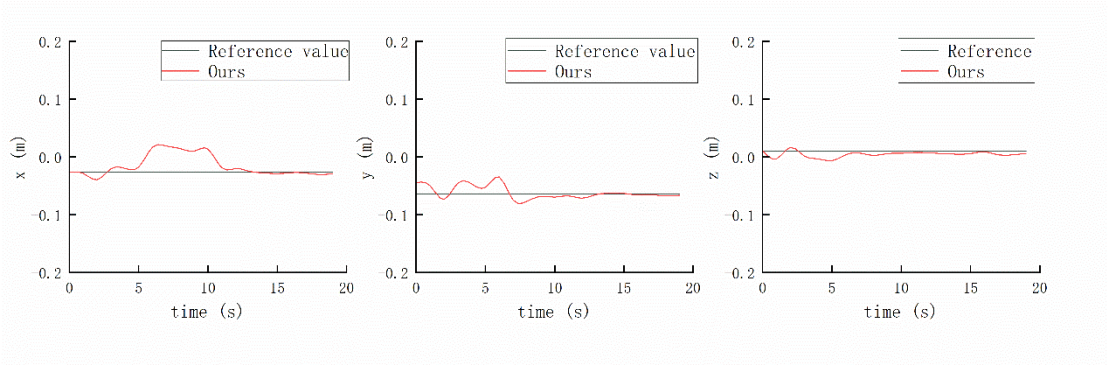

Figure 3. Translation Initialization Result

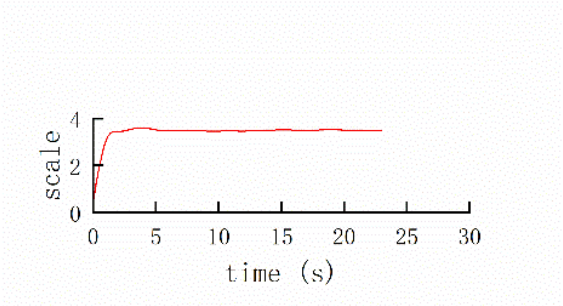

Figure 4. Scale Initialization Result

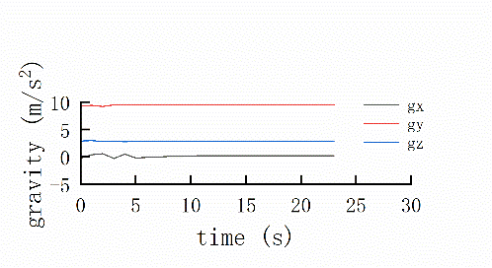

Figure 5. Gravity Initialization Result

As can be seen from Figure 3, the convergence time of translation estimation is about 10s, and there is a large error between the estimated value and the reference value.

In this paper, the estimated value of each state calibrated by the VINS-mono algorithm is taken as the ratio to verify the feasibility of the proposed algorithm. The VMS-MONO algorithm firstly initializes the rotation and translation parameters linearly, and then introduces the zero bias of accelerometer and gyroscope to optimize all state variables. On the basis of V2_01_easy sequence, 10 calibration experiments have been carried out for both the proposed algorithm and the VINS-mono algorithm, and the calibration results are shown in Figure 6.

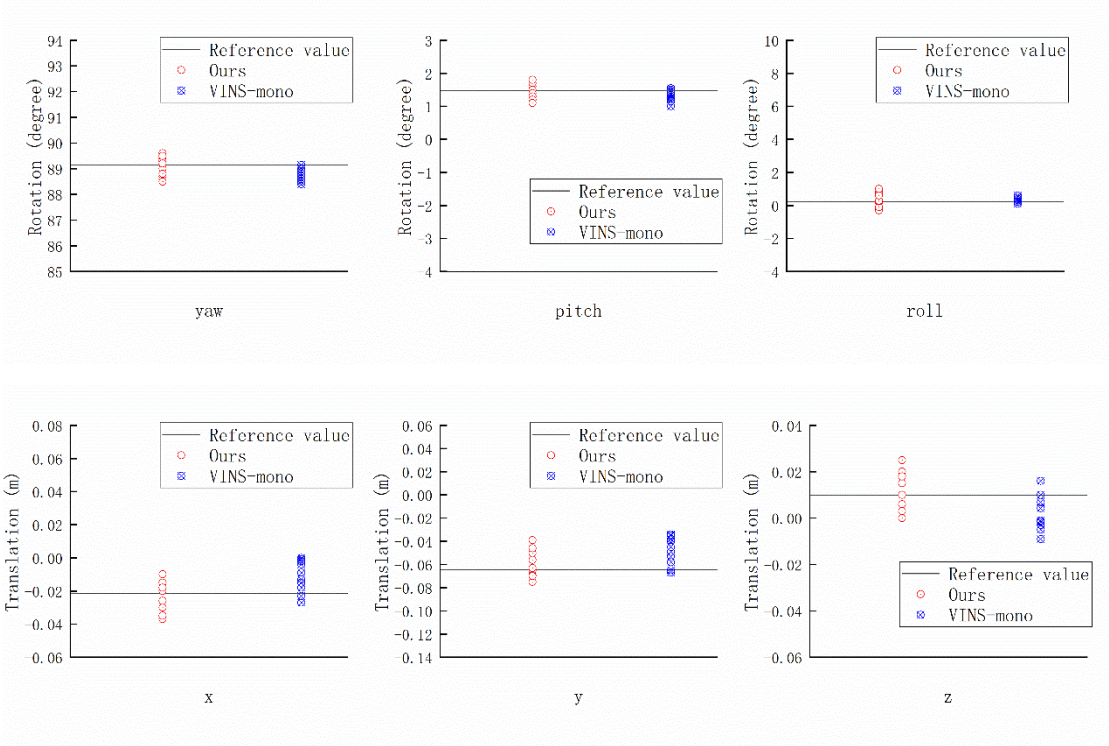

Figure 6. Consistency Analysis of Different Methods

As can be seen from Figure 6, for the rotation external parameter calibration of monocular camera and IMU, the variation trend of rotation external parameter solved by the proposed algorithm is basically consistent with that of the VINS-mono algorithm. Because the influence of zero bias of gyroscope is not taken into account, the calibration reference value of external rotation participation has a greater degree of dispersion. In the process of calibration of translational external parameters, the zero deviation of gyroscope is calibrated, so a good calibration accuracy of translational external parameters can be obtained. Finally, the online estimation of external parameters in this paper

achieves a high precision, and the estimation error of rotational external parameters is within 0.8 degree, and that of translational external parameters is within 0.04m.

VINS-mono algorithm can obtain good estimation accuracy without any prior knowledge, so this paper uses various accuracy indexes estimated by VINS-mono algorithm as reference values to verify the feasibility of the proposed method. Table 2 shows the rotation direction error, translation error and RMSE (root mean square error) of the keyframe trajectory of the three sequences in the EuRoC data set. All experimental results are the average of 5 runs, and loop detection and global nonlinear optimization are not used. The ideal scale factor is obtained by measuring the alignment of the estimated keyframe trajectory to the most suitable real trajectory. The scale factor error can be defined as $|s^* - \hat{s}|/|\hat{s}| \times 100\%$.

Table 2. Extrinsic Calibration Error

Ours									VINS-mono
	Rotation Error (degree)			Translation Error (m)			RMSE (m)	Scale Error (%)	RMSE (m)
	yaw	pitch	roll	x	y	z			
V1_01_easy	-0.312	-0.267	0.088	-0.008	-0.034	0.025	0.064	1.3	0.070
V2_01_easy	-0.304	-0.054	-0.076	-0.014	0.033	0.065	0.072	1.5	0.068
Mh_01_easy	-0.072	0.600	0.031	-0.043	0.015	0.021	0.050	0.56	0.135

As can be seen from Table 2, the RMSE error of the method in this paper is basically consistent with that of VINS-mono when unknown external parameters also exist. In MH_01_easy, the proposed method shows more accurate performance. The analysis shows that the scene structure of MH_01_easy basically conforms to the triorthogonality of Manhattan world hypothesis, and adding the prior knowledge of global constraints can further improve the accuracy of the system. In the two scenarios V1_01_easy and V2_01_easy, the precision of keyframe pose estimated by the sliding window strategy in the nonlinear optimization of VINS-mono is slightly lower than that estimated by the proposed method, and the estimated external parameter results of the proposed method are more reasonable. According to the analysis, in the face of the environment without significant structural features, the structural features constructed by the parameter threshold can provide more constraints, and the local constraint relationship still plays a key role in maintaining the robustness of the system. The data in the table again show that the proposed method can accurately calibrate the external parameters between the monocular camera and the IMU, and the error of the rotating external parameters is up to 0.6 degree, and the error of the translating external parameters is up to 0.07 m.

6. CONCLUSION

In order to improve the positioning accuracy and stability of the visual inertial system in the natural environment, this paper proposes an online calibration method with external parameters constrained by the characteristics of natural environment. The effectiveness, accuracy and robustness of the proposed method are tested on EuRoC data set. Experiments show that the proposed method can improve the on-line calibration accuracy and stability of visual inertia system without any mechanical configuration parameters, and can meet the basic application requirements of long-term operation of visual inertia system in indoor and outdoor environment.

REFERENCE

[1] S. Weiss, M.W. Achtelik, S. Lynen, M. Chli, R. Siegwart, Real-time onboard visual-inertial state estimation and self-calibration of MAVs in unknown environments, in Proc. IEEE Int. Conf. Robot. Autom., Saint Paul, MN, USA, May 2012, pp. 957–964.

[2] M. Li, A.I. Mourikis, High-precision, consistent EKF–based visual–inertial odometry,,Int. J. Robot. Res., May. 2013, pp.690-711.

[3] T.C. Dong-Si, A.I. Mourikis, Estimator initialization in vision-aided inertial navigation with unknown camera-IMU calibration, in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Vilamoura, Portugal, Oct. 2012, pp. 1064–1071.

[4] Y. Zhenfei, S. Shaojie, Monocular visual–inertial state estimation with online initialization and camera–IMU extrinsic calibration, IEEE Transactions on Automation Science and Engineering 14.1 (2016) 39-51.

[5] Q. Tong, L. Peiliang, S. Shaojie, Vins-mono: A robust and versatile monocular visual-inertial state estimator, IEEE Transactions on Robotics 34.4 (2018) 1004-1020.

[6] H. Weibo, L. Hong, Online initialization and automatic camera-IMU extrinsic calibration for monocular visual-inertial SLAM, 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018, pp.5182-5189.

[7] Y. Yingjian, G. Banglei, S. Xiangyi, L. Zhang, F. Friedrich, Rotation alignment of a camera–IMU system using a single affine correspondence, Applied Optics 60.24 (2021) 7455-7465.

[8] M. Arbabmir, M. Ebrahimi, Visual–inertial state estimation with camera and camera–IMU calibration, Robotics and Autonomous Systems 120 (2019) 103249.

[9] F.M. Mirzaei, S.I. Roumeliotis, Optimal estimation of vanishing points in a manhattan world, 2011 International Conference on Computer Vision. IEEE, 2011, pp.2454-2461.

[10] V.G. Rafael Grompone; J. Jeremie, M. Jean-Michel, R. Gregory, LSD: A fast line segment detector with a false detection control, IEEE transactions on pattern analysis and machine intelligence 32.4 (2008) 722-732.

[11] M.E. Antone, T. Seth, Automatic recovery of relative camera rotations for urban scenes, Proceedings IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2000 (Cat. No. PR00662). Vol. 2. IEEE, 2000, pp.282-289.

[12] L. Xiaohu, Y.Y. Jian, L. Haoang, L. Yahui, Z. Xiaofeng, 2-line exhaustive searching for real-time vanishing point estimation in manhattan world, 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2017, pp.345-353.

[13] M. Burri, J. Nikolic, P . Gohl, T. Schneider, J. Rehder, S. Omari, M.W. Achtelik, and R. Siegwart, The EuRoC micro aerial vehicle datasets, Int. J. Robot. Res., 2016, pp. 1157–1163.