A NOVEL METHOD FOR FUSION OF GNSS AND VISUAL-INERTIAL-WHEEL ODOMETRY

Yahui Zhang,* Linxuan Wang,* Aimin Li,* Yongsheng Zheng,* and Mingzhuang Wu*

Abstract

Aiming at the problem that the state estimation of visual-inertial-odometre (VIO) is affected by long-term trajectory drift accumulation, a VIO compact coupling system based on optimisation is proposed. To estimate all the initial state variables of the system effectively, a method combining static initialisation and dynamic initialisation is proposed. To realise the transformation between local frames and local Cartesian coordinates coordinate system (ENU) coordinates of the local Cartesian coordinates coordinate system, local state and global navigation satellite system (GNSS) measurement information are fused. A noise reestimation method is used to ensure the simultaneity of inertial measurement unit (IMU) and wheel speedometre. We evaluated the proposed system on a public data set and tested it in a real-world scenario. The experimental results show that the VIO tightly coupled system based on the optimisation can present better positioning effect and improve the precision of global fusion. The system can achieve accurate attitude estimation in the outdoor environment.

Key Words

State estimation, sensor fusion, tight coupling, precise positioning

1. Introduction

Accurate state estimation is a prerequisite for autonomous driving and multi-sensor fusion state estimation is a hot research topic at present [1], [2]. The combination of a camera and an inertial measurement unit (IMU) can provide high-frequency and anomaly-free inertial measurements, often achieving high accuracy and greater robustness in complex environments [3], [4].

However, visual-inertial-odometres (VIO) are accompanied by significant cumulative drift over long-term trajectories [5], [6]. While some applications will perform well in the face of such significant drift, efforts should be made to minimise these effects, but it must also be acknowledged that some degree of drift will be inevitable in the absence of global measurements [7]. The integration of global measurement and visual inertia has been proved to eliminate the drift problem very well [8].

The integration of vision, IMU, and global navigation satellite system (GNSS) can realise accurate global positioning in complex urban scenes [9]. Visual inertial navigation system (VINS) is usually divided into two types of algorithms: 1) filtering-based method and 2) optimisation-based method [10]. This classification naturally applies to GPS assist methods as well. The filter-based method estimates the probability distribution of states, including attitude and signpost [11]. Based on the optimisation method, a nonlinear least-squares minimisation problem is established and solved by constraint adjustment [12]. The filter-based method is usually highly efficient [13]. However, the filter-based approach is very sensitive to time synchronisation [13]. Any late measurement will cause trouble because the state cannot propagate back during the filtration process [14]. Therefore, special ordering mechanisms are needed to ensure that all measurements from multiple sensors are ordered [16]. In an optimisation-based framework, many visual and inertial measurements are kept together [15]. Compared with the filter-based approach, the optimisation-based approach has advantages in this respect [18], optimising the states associated with the observation and measurement. One advantage of the optimisation-based method over the EKF-based method is that the state can be iteratively linearised to improve the accuracy [8], [17].

From the perspective of fusion measurements, tight coupling and loose-coupling methods can be distinguished [19]. The loose-coupling method is used to estimate the state of different sensors, and then the estimated results of different sensors are fused [20]. On the contrary, the tight coupling method combines the measured values from different sensor modes for state estimation [21]. Because the available information is not fully utilised in the loosely coupled method, the loosely coupled method is not optimal [22]. To address the significant cumulative drift associated with VIO over long trajectories, for ground vehicles,

* School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou 221116, China; e-mail: zhangyahui0610@163.com; ts21050174p31@cumt.edu.cn; liaimin@cumt.edu.cn; {7563260, 1872912541}qq.com
Corresponding author: Aimin Li

Recommended by Anmin Zhu
(DOI: 10.2316/J.2024.206-1032)

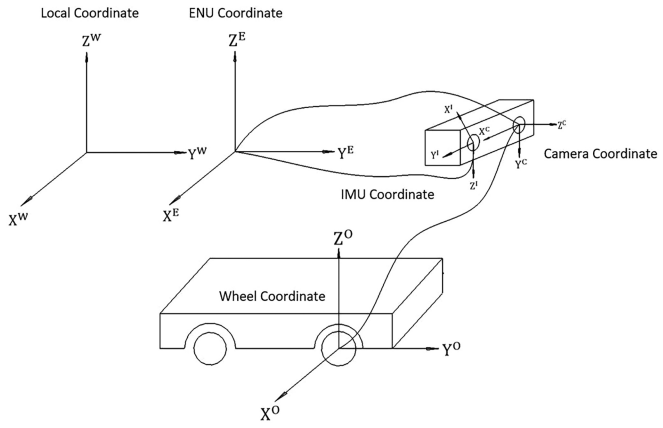

Figure 1. The coordinates between different sensors and external parameters.

[23] tightly-coupled wheel odometres act as compensation sensors, making scale, roll, and pitch angle permanently visible.

To reduce the drift phenomenon in state estimation, more accurate global positioning can be achieved. In this paper, we propose a multi-sensor fusion system based on optimisation that tightly coupled the camera, IMUs, and wheel odometre (VIWO) to achieve accurate state estimation. The specific contributions of this paper are as follows.

1. In the initialisation stage, static initialisation and dynamic initialisation are combined to effectively estimate all initial state variables.
2. A noise reestimation method is used to ensure the simultaneity of IMU and wheel speedometre.
3. To combine GNSS data, vision-IMU-wheel odometre (VIWO) data under the local framework is loosely coupled with the global position information collected in GNSS data, so as to achieve accurate pose estimation in real time.

2. Method

We first define the coordinates and symbols associated with the multi-sensor fusion framework. The ENU coordinate, which is the northeast upper coordinate with symbol $(\cdot)^E$, is defined as the global coordinate system. Use the local world coordinates of $(\cdot)^W$ to represent the pose in the first IMU coordinate system, where the z-axis can be aligned with gravity so that the roll and pitch angles become observable. The absolute Z-axis direction in ENU coordinates aligns with the direction of the local frame. $(\cdot)^C$ is the camera frame. In contrast to fixed world coordinates, the IMU coordinates using $(\cdot)^I$ as the frame of the human body describe the local time pose, changing constantly with the state of motion. $(\cdot)^O$ is the wheel coordinate information with the origin at the centre of the wheels on both sides. T_i^j said $(\cdot)^i$ to $(\cdot)^j$ coordinate system of the external parameters, including i and j said the arbitrary coordinates, can be set up by Fig. 1.

2.1 System Overview

We reserve the key frame pose in the local coordinate system and the key frame pose in the ENU coordinate system, respectively. To estimate the corresponding external parameters, we need to convert the local frame to ENU coordinates.

The whole system is divided into two parts, namely, the multi-sensor fusion module and the semantic mapping module. The system uses four sensors as sources of environmental information perception: camera, IMU, wheel speed encoder, and GNSS. First, the measurement data of the camera, IMU, and wheel speed encoder are integrated for system initialisation. After the system initialisation is completed, several measurement data are coupled together by the tight coupling optimisation algorithm to obtain the real-time position and pose of the current vehicle, as shown in Fig. 2.

Then, the GNSS measurement data is fused by the loose-coupling optimisation method. The semantic mapping module combines the real-time pose and GNSS measurement results of the current vehicle to build a globally consistent semantic map. The schematic diagram of the global pose map structure is shown in Fig. 3. Each attitude is a node in the attitude diagram, containing the position and direction in the world coordinate system.

Figure 3 shows an illustration of the global attitude diagram structure. Each node represents an attitude in the world frame, which contains location and direction. The edge between two continuous nodes is a local constraint derived from the locally estimated VIO tightly coupled system (VIWO). The other edges are global constraints, which come from global sensors. In addition, linear EKF method was used to alleviate the abrupt impact of GNSS rejection area to open area, and the error caused by drift was evenly distributed to positions and poses at different times to smooth the trajectory.

In this paper, the measurement data of tightly coupled vision, IMU, and wheel speed metre are used to complete a robust and high-precision positioning system by using their complementary characteristics. To improve the tracking algorithm of visual feature points, it is proposed to use the measured value of IMU to predict the feature points with depth inside the sliding window and provide an initial value for the feature points in the next frame, and then carry out optical flow tracking on the basis of this initial value, which can effectively improve the tracking robustness of visual feature points in the system and improve the accuracy of the algorithm, as shown in Fig. 4.

2.2 Dynamic and Static Initialisation of Visual-Inertial-Wheel Odometre

During the initialisation of the VINS, the pose estimated by the monocular vision SFM will be affected by the scale uncertainty of the single purpose. Therefore, there is a proportional relationship between the pose estimated by the sliding window and the actual motion state, which is called the scale factor. Since autonomous vehicles are not

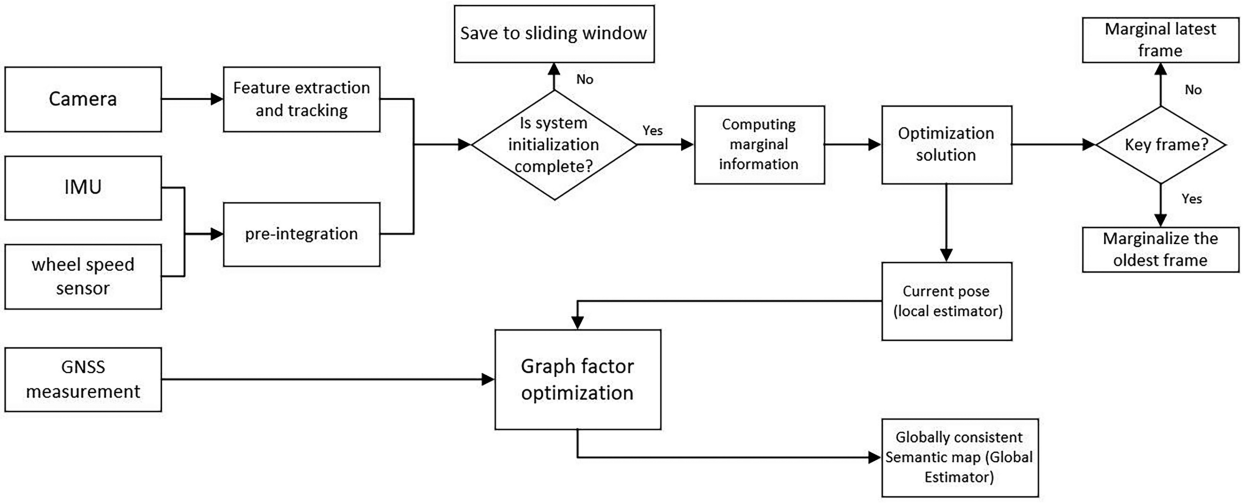

Figure 2. System overview frame diagram.

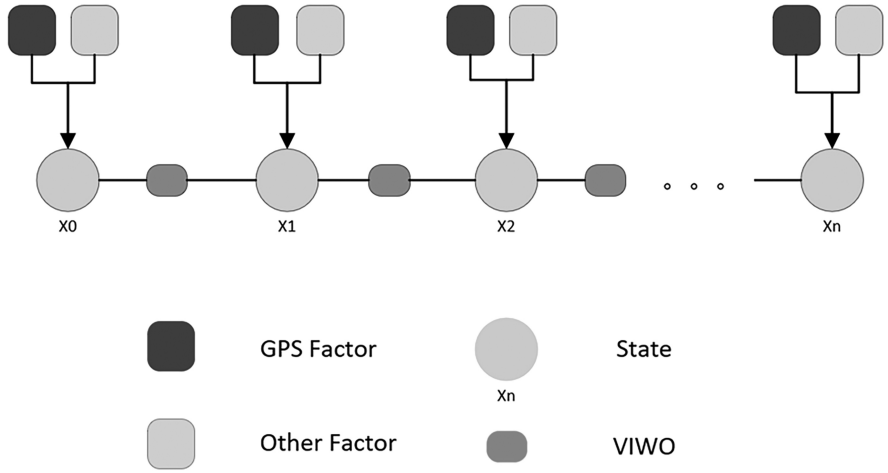

Figure 3. Global attitude structure diagram.

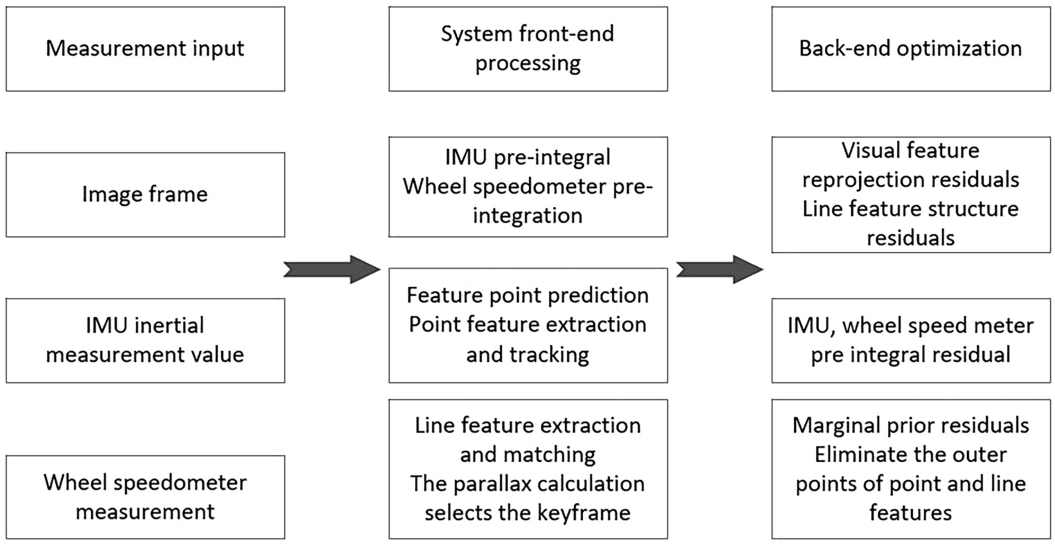

Figure 4. VIWO compact coupled algorithm block diagram.

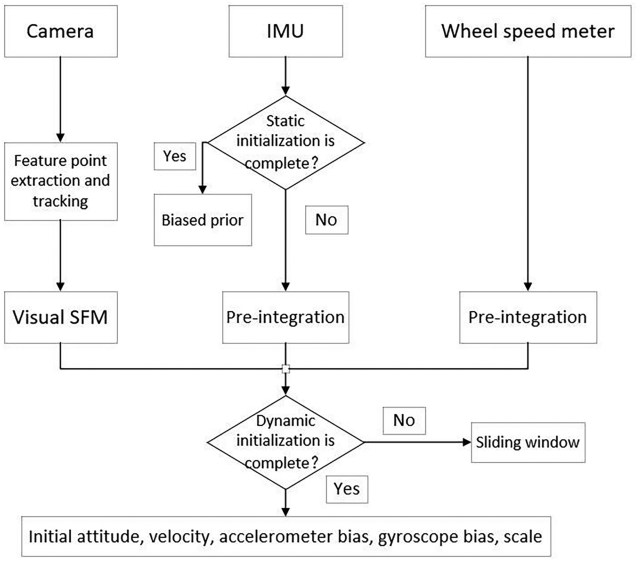

Figure 5. Static and dynamic joint initialisation framework.

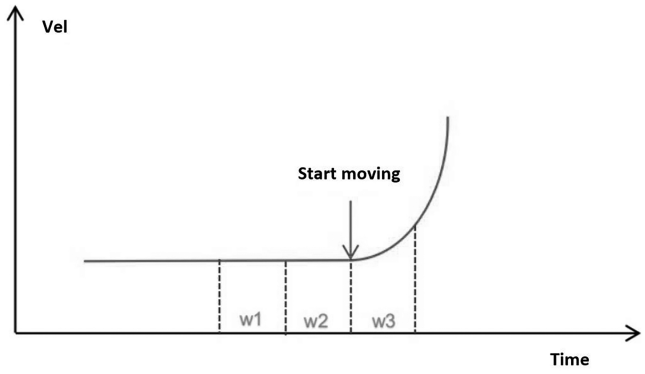

Figure 6. The jumping process diagram of the IMU.

as smooth as mobile robots when they start, IMUs are not sufficiently motivated in this case. Due to the insufficient excitation of IMUs, the single integration of the visual SFM result and the pre-integration result of IMUs will lead to the slow initialisation of the system and, more seriously, the single-purpose scale factor estimation error. Therefore, the pre-integral of the wheel speedometre is integrated during initialisation to provide absolute mileage observation, and the single-purpose scale factor can be estimated effectively. Then, we can get the scaling factor X [24] by solving the linear least-squares problem:

$$\min_{X} \sum_{k \in W} \left\| z_{O_{k+1}}^{O_k} - F_{O_{k+1}}^{O_k} X \right\|^2 \tag{1}$$

Where W indexes all frames in the window. Since the acceleration measurement value of IMU is coupled with the influence of gravity acceleration, the optimisation problem of estimating acceleration zero bias is a sick problem when the dynamic initialisation method aligns the visual SFM result and the IMU pre-integral result. The zero bias of the accelerometre cannot be estimated, and can only be estimated after the completion of system initialisation. As a result, the zero bias estimation of the accelerometre converges slowly, which affects the accuracy of the system.

Therefore, this study proposed a combination of static initialisation and dynamic initialisation. Based on static IMU measurements, the bias of accelerometre and gyroscope was first estimated as the initial value of dynamic initialisation. During dynamic initialisation, the zero bias of accelerometre and gyroscope was optimised and adjusted, which could effectively estimate all initial state variables of the system. That is, initial position, velocity, gravity direction, accelerometre bias, and gyroscope bias of the system, as shown in Fig. 5.

Static initialisation refers to the need for a static state at the beginning of the video sequence. Systems use IMU data at rest to estimate gravity orientation, accelerometre gyroscopic biases. Static initialisation needs to detect the bouncing process of the IMU, which includes two states: 1) static state and 2) bouncing state, as shown in Fig. 6.

A threshold (IMU excite threshold) is set to determine if the IMU is excited enough, and the IMU initialisation parameter is estimated from the IMU measured at rest before the motion excitation is obtained.

First, we need to compute the initial rotation matrix. Set the accelerometre mean of window w2 to a and the gyroscope mean of window w2 to b. Because, under the world system, the acceleration measured at rest is (0,0,g). Divide a of window w2 by its modulus length to obtain the direction of the average acceleration (unit vector), which is the projection of the direction of the Z-axis of the world system on the IMU coordinate system:

$$z = \frac{z}{Z.norm(\cdot)} \tag{2}$$

After determining the direction of z-axis, the unit coordinate system is constructed by Schmidt orthogonalisation. Suppose ρ_1 is the unit direction vector corresponding to X^I-axis of IMU system:

$$e_1 = (1, 0, 0)^T \tag{3}$$

Project the X^I-axis onto the $Z^w - X^w$ plane and find the direction vector of the X^w-axis in the IMU coordinate system:

$$\begin{aligned} e_1 &= e_1 - (z^T e_1) z \\ x &= \frac{e_1}{e_1 norm(\cdot)} \end{aligned} \tag{4}$$

The Y-axis direction is obtained by the cross product of the x and z axes:

$$y = [Z]_\times x \tag{5}$$

Through the above steps, the initial rotation $R^{\text{wi}} = [X, Y, Z]$ from the Cosmological system w to the IMU coordinate system is obtained. Then, the gyroscope bias is calculated. The ideal measurement value (angular velocity) of the gyroscope at the static moment should be zero, so the bias of the gyroscope can be calculated, that is, the average value $\text{av}g_{\text{wm}}$ of the gyroscope data in window w2:

$$\mathrm{av}g_{\mathrm{wm}} = \frac{1}{n}\sum_{1}^{n} w_m \tag{6}$$

Finally, calculate the accelerometre bias, assuming M. Accelerometre bias at rest is the difference between the average acceleration and the actual acceleration due to gravity:

$$M = \mathrm{av}g_{\mathrm{wm}} - R_g \tag{7}$$

Where g is the acceleration of gravity in the world system, and R is the rotation matrix obtained in the first step, projecting the gravity vector from the world system into the IMU coordinate system.

The initialisation of IMU here is to obtain the relative rotation of R^Z-axis (0,0,1) from the world coordinate system W to the IMU coordinate system based on the principle that the measured value of the accelerometre at rest is in the same direction as the gravity vector.

In other commonly used VIO systems, the first frame of vision is usually taken as the world system. Since the relative relationship between the camera and the IMU is known, it is easy to get the pose of the camera system with respect to the world system, and the pose of the IMU system with respect to the world system.

If the platform is not static, then we use dynamic initialisation to try to restore the initial state. It solves the initialisation problem by first creating a linear system for recovering velocity, gravity, and feature position. After the initial recovery, a full optimisation is performed to allow covariance recovery.

2.3 Tight Coupling Optimisation of Visual-Inertial-Wheel Odometre

After system initialisation, several measurement data are coupled together by optimisation algorithm to obtain the real-time position and pose of the current vehicle. To provide more robust and accurate state estimation, a tightly-coupled VIWO based on optimisation is adopted. This paper makes full use of the reliable observation characteristics of linear velocity [26], and takes the velocity along the X-axis into consideration in the construction of the pre-integral. Therefore, we obtained the wheel pre-integral by means of discrete median integral:

$$\widehat{a}_{i+1}^{O_k} = \widehat{a}_i^{O_k} + \overline{\widehat{v}}_i\delta_t, \overline{\widehat{v}}_i = [q_i^0 \cdot s \cdot \widehat{v}_i + q_{i+1}^0 \cdot s \cdot \widehat{v}_{i+1}] \tag{8}$$

$$\widehat{\gamma}_{i+1}^{O_k} = \widehat{\gamma}_i^{O_k} \otimes q_I^0 \otimes \begin{bmatrix} 1 \\ \frac{\overline{\widehat{w}}_i\delta_t}{2} \end{bmatrix}, \overline{\widehat{w}}_i = \frac{(\widehat{w}_i + \widehat{w}_{i+1})}{2} - b_g \tag{9}$$

After the IMU and wheel pre-integral are obtained, the noise reestimation method is proposed to adjust the commensurable noise by adjusting the consistency of the variance of each predicted position [26]. Field tests found that the additional frequency produced greater noise during inertial measurement, which required amplification of 50 $\sim$ 100 times. To balance this, set the repropagation parameter to 10. In this paper, the optimal ratio of IMU to wheel pre-integral is 1. If the ratio is less than 1, it indicates that the wheel measurement is abnormal, and then the set wheel noise will be multiplied by the reciprocal of the ratio and the measurement parameters. If the value is greater than 1, the IMU measured value is considered to have a problem. In this case, the IMU parameter set will be multiplied by the ratio and noise parameter. Finally, the IMU noise caused by motor vibration is taken into account, and the nonlinear optimisation is expressed as:

$$\min_X \{\|\gamma_P - H_P X\|^2 + \sum_{k\in B} \left\|\gamma_I(z_{I_{k+1}}^{I_k}, X)\right\|^2 + \sum_{k\in O} \left\|\gamma_O(z_{O_{k+1}}^{O_k}, X)\right\|^2 + \sum_{(i,j)\in C} \left\|\gamma_C(z_i^{c_j}, X)\right\|^2\} \tag{10}$$

In the above equation, X is the state variable, and the four residuals correspond to the prior information of marginalisation, IMU measurement, wheel measurement, and reprojection error between adjacent frames. Using the Ceres Solver, we can solve this problem and get the attitude of the local view frame.

2.4 Inertial-Wheel Odometre

To improve computing efficiency, we designed a special module for calculating and publishing raw data of IMUs and odometre. It is worth noting that because the data is acquired at different positions, there is a tangential acceleration applied to the rigid connection position during rotation. We also remove the error caused by the raw data received by the gyroscope during attitude estimation. The accelerometre is used to approximate the gyroscope deviation.

Real-time updates of rotation matrix and deviation can be realised by using EKF [27]. Since we have obtained the attitude Angle, we can calculate the left Jacobian matrix relative to the translational decomposition:

$$J = \frac{\sin\theta}{\theta} I_{3\times 3} + \left(1 - \frac{\sin\theta}{\theta}\right) aa^T + \frac{1-\sin\theta}{\theta} \hat{a} \tag{11}$$

Where θ represents the rotation angle, a is the rotation axis represented by the unit vector, and $(\cdot)\hat{}$ is the skew-symmetric matrix. We convert the linear and additional wheel speeds into IMU frames:

$$v^I = R^I{}_0 \cdot w_{0I}\hat{} \times t^I{}_0 + R^I{}_0 \cdot v^0 \tag{12}$$

Where w_{0I} is the angular velocity generated by the rigid connection. The amount of translation in SE (3) is obtained by multiplying the left Jacobian, the velocity, and the nearest time interval.

2.5 Multi-sensor Fusion Positioning

As mentioned earlier, we maintain two key frame poses in the system, one from the local coordinate system and one from the absolute coordinate system UNU. Our goal

is to estimate the corresponding external parameters. The specific idea is to achieve relative pose constraints on VIWO and IWO by converting local frames to ENU frames. Since there is no truth reference, coordinate transformation of factor graph is not considered. We subdivide the graph optimisation into three stages: initialisation, external calibration, and positioning.

The initialisation stage is mainly for external parameter estimation. Our setup is that when the vehicle has completed enough movement in the previous node, we select a new key frame. The key frame is inserted into the factor graph by means of attitude interpolation, and time alignment can be carried out in this way. The way to determine the initialisation is successful is after we have collected 15 key frame nodes.

To solve the coordinate system transformation relation accurately, we construct the ternary edge. Specifically, the state of the key frame is constrained in two frames of images and external parameters. We use the resulting ENU coordinates for the Angle constraint. Every global node has an absolute translation constraint. The number of ternary edge constraints is limited to 52. We believe that if the optimised posture in VIWO does not appear in adjacent key frames, it is complementary to the posture in IWO.

The optimisation variance represented by SE (3) is defined as:

$$X = \{X^E, X_I^E, X^I\} \tag{13}$$

The corresponding question can be expressed as:

$$\arg \min_X \mathrm{r}(X) = Log((X^E \cdot X^I)^{-1} X^E) \tag{14}$$

Where r is defined as the mapping of resists from SE (3) to se (3), the initial guess of which is given by X_I^E aligned with the track of the first 15 nodes.

The following is a detailed explanation of our approach and purpose. The ternary edge is introduced to accurately calibrate the external parameters between two frames and optimise the posture under different frames. To simplify the multi-sensor fusion process, the calibrated external parameters were removed from the variables to be adjusted. In high-frequency positioning, based on the optimised attitude of the previous node, the system will accumulate the attitude information provided by the IWO relative to the current attitude. In addition, linear EKF method was used to alleviate the abrupt impact of GNSS rejection area to open area, and the error caused by drift was evenly distributed to positions and poses at different times to smooth the trajectory.

3. Experiment

Multi-sensor fusion localisation algorithm was tested based on KAIST open data set and outdoor data set collected in the field. The outdoor data set collected in the field was recorded by a homemade path acquisition cart fitted with a camera, IMU, wheel odometre, and GNSS receiver. The dataset was run on a 16 GHz 6-core NVIDIA Carmel ARM®v8.2 64-bit CPU. Wheel encoder (100 Hz), and GPS (10 Hz). The real attitude of the vehicle on the ground is generated by the correlation algorithm.

Table 1
RMSE in KAIST Data Sets and Outdoor Experiments [m]

Sequence	**Length [m]**	**RMSE [m]**	
		VINS-Fusion [8]	**Proposed**
KAIST-32	7100	5.54	**0.57**
KAIST-33	7600	10.68	**0.79**
Outdoor 1	72.03	0.24	**0.12**
Outdoor 2	98.83	0.20	**0.12**
Outdoor 3	59.37	0.18	**0.07**
Outdoor 4	287.30	0.16	**0.09**

3.1 Data Sets of KAIST

The images collected from the camera were simulated on the KAIST data set, and RMSE was used as the evaluation index. In the collected KAIST data set, the sensor information was selected. The main selection included left camera (10 Hz), IMU (100 Hz), wheel encoder (100 Hz), and GPS (5 Hz). We ran two tests, the urban 32 and the urban 33. Among them, the test distance on the urban 32 orbit is 7100 m, and the test distance on the urban 32 orbit is 7600 m. After testing in urban 32, we found that the trajectory of our proposed system was in good agreement with the real trajectory on the ground, while the trajectory of VINS-Fusion [8] was intuitively deviated from the reference.

To quantitatively evaluate our method, RMSE was calculated using a trajectory evaluation tool named EVO in each data set, showing that the performance of our proposed method is significantly better than that of VINS-Fusion [8], as shown in Table 1. This shows that our method has better results.

3.2 Outdoor Data Sets

In the outdoor real experiment, we evaluate the global positioning accuracy on one hand, and verify the convergence of the external parameters and the consistency of the sensor noise transmission on the other hand. Since there is no global direction for real ground values, local frame relocation is evaluated by the relocation error constructed by anchoring Apriltag.

In practical experiments, both the IMU and wheel odometre sampled at 100 Hz, while the camera and GNSS receiver sampled at much lower frequencies of 10 Hz. Outdoor 1 works properly, and Outdoor 2 records the dynamic environment with vehicles moving around. Based on dynamic factors, the Outdoor 3 generated by the car body is stationary for a long time. To test the robustness of the algorithm, two complex data sets, Outdoor 4, were considered, including a variety of challenging scenarios, such as fast motion and rotation, long-term rest, *etc.*

Table 2
Reconstruction Evaluation of Noise Based on Apriltag in Outdoor Experiment

Sequence	Length [m]	Evaluation	VIW	VIW_{NR}	Proposed	$\text{Proposed}_{\text{NR}}$
Outdoor 1	88.34	Angle Error (deg/m)	0.058	0.049	0.017	**0.013**
		Position Error (%)	1.623	1.069	**0.081**	0.083
Outdoor 2	75.09	Angle Error (deg/m)	0.059	0.059	0.018	**0.014**
		Position Error (%)	1.152	0.978	**0.073**	0.087
Outdoor 3	67.37	Angle Error (deg/m)	0.051	0.031	0.019	**0.018**
		Position Error (%)	0.780	0.639	0.098	**0.093**
Outdoor 4	292.54	Angle Error (deg/m)	0.069	0.062	0.009	**0.007**
		Position Error (%)	1.618	1.330	0.025	**0.024**

To discuss the feasibility of the noise reestimation method, an ablation experiment was conducted using the data set collected in the field. $*_{\text{NR}}$ refers to the algorithmic repropagation with noise mentioned in Table 2. This shows that our method has better results. Through the noise propagation method based on Apriltag, the angle error of VIWO under the local frame is reduced, and the system accuracy is improved, indicating that the method can play a role in attitude estimation. However, the global attitude is seriously affected by GNSS, so the accuracy of global positioning has not been significantly improved.

3.3 Discussion

There are some problems with VINS-Fusion [8]. The state estimation of visual inertial odometre is affected by the accumulation of long-term trajectory drift. By introducing wheel odometre, the drift of vehicle body caused by noise and other factors can be solved well, and the accuracy and robustness of state estimation of vehicle body in non-rigid environment can be improved. In addition, the wheel odometre provides a more reliable observation and uses less noise directly to the speed, which is optimised to be more constrained than the IMU measurement of the expected fraction of the speed. Therefore, the VIWO tightly coupled optimisation system proposed by us presents better results when taking RMSE as the index and compared with VINS-Fusion [8].

As shown in Table 1, since the optimisation-based method can realise online initialisation of external parameters, our system can obtain more accurate pose transformation relationship for attitude estimation in ENU coordinate system. In the whole system, the noise reestimation method guarantees the consistent estimation of IMU and wheel odometre. The positioning accuracy is guaranteed by this adaptive adjustment. To make up for the delay effect caused by optimisation, we use short-term pre-integral to realise the decomposition of pose estimation. To sum up, the multi-sensor fusion positioning system proposed by us is superior to VINS-Fusion [8] in RMSE evaluation, and has good localisation performance.

4. Conclusion

To realise drift-free state estimation of ground moving vehicles, a VIO tightly coupled system based on optimisation is proposed, and the local state and GNSS measurement information are fused by factor graph optimisation method.

1. By tightly coupling the VIWO, and by optimising the framework used to fuse the absolute position obtained by the GNSS, and an IWO that can publish the pose at high speed.
2. Adaptive adjustment of the covariance matrix was achieved by the noise reestimation method, and weight variance during optimisation was obtained to ensure the consistency of IMU and wheel.
3. We tested the proposed system in an outdoor experiment, and the experimental results show that the proposed method achieves accurate and robust positioning.

Acknowledgments

The project is supported by the National Key Research and Development Program (2022YFB3403003) and the National Natural Science Foundation of China (51575512, 52074272). This work is also supported by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

[1] M. Dong, G. Yao, J. Li, and L. Zhang, Calibration of low cost IMU's inertial sensors for improved attitude estimation, *Journal of Intelligent and Robotic Systems, 100*(2), 2020, 1015–1029.

[2] C. Campos, R. Elvira, J.J.G. Rodriguez, J.M.M. Montiel, and J.D. Tardos, ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM, *IEEE Transactions on Robotics*, *37*(6), 2021, 1874–1890.

[3] G. Chen and B. Jin, Position-posture trajectory tracking of a six-legged walking robot, *International Journal of Robotics and Automation, 34*(1), 2019, 24–37.
[4] S. Han, F. Deng, T. Li, and H. Pei, Tightly coupled optimization-based GPS-visual-inertial odometry with online calibration and initialization, 2022, *arXiv:2203.02677*.
[5] M. Boujelben, C. Rekik, and N. Derbel, A multi-agent architecture with hierarchical fuzzy controller for a mobile robot, *International Journal of Robotics and Automation, 30*(3), 2015, 289–298.
[6] H.W. Lee, C.L. Shih, and C.-L. Hwang, Design by applying compensation technology to achieve biped robots with stable gait, *International Journal of Robotics and Automation*, *29*(1), 2014.
[7] T. Qin and S. Shen, online temporal calibration for monocular visual-inertial systems. *Proc. IEEE. IEEE*, Madrid, Spain, 2018, 3662–3669.
[8] T. Qin, S. Cao, J. Pan, and S. Shen, A general optimization-based framework for global pose estimation with multiple sensors, 2019, *arXiv:1901.03642*.
[9] J. Liu, W. Gao, and Z. Hu, Optimization-based visual-inertial slam tightly coupled with raw GNSS measurements, 2020, *arXiv:2010.11675*.
[10] S. Cao, X. Lu, and S. Shen, GVINS: tightly coupled GNSS–visual–inertial fusion for smooth and consistent state estimation, *IEEE Transactions on Robotics, 38*(4), 2022, 2004–2021.
[11] S. Lynen, M.W. Achtelik, S. Weiss, M. Chli, and R. Siegwart, A robust and modular multi-sensor fusion approach applied to MAV navigation, *Proc. IEEE. IEEE*, 2013, 3923–3929.
[12] S. Leutenegger, OKVIS2: Realtime scalable visual-inertial slam with loop closure, 2022, *arXiv:2202.09199*.
[13] A.I. Mourikis and S.I. Roumeliotis, A multi-state constraint Kalman filter for vision-aided inertial navigation, *Proc. Robotics and Automation, 2007 IEEE International Conf. on. IEEE, (LVBO),* 2007, 3565–3572.
[14] M. Li and A.I. Mourikis, High-precision, consistent EKF-based visual-inertial odometry, *The International Journal of Robotics Research, 32*(6), 2013, 690–711.
[15] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, Keyframe-based visual-inertial odometry using nonlinear optimization, *The International Journal of Robotics Research, 34*(3), 2014, 314–334.
[16] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, Robust visual inertial odometry using a direct EKF-based approach, *Proc. of the IEEE/RSJ* International *Conf. on* Intelligent Robots and Systems, Hamburg, Germany, 2015, 298–304.
[17] R. Mur-Artal and J.D. Tardós, Visual-inertial monocular SLAM with map reuse, *IEEE Robotics and Automation Letters, 2*(2), 2017, 796–803.
[18] H. Strasdat, J. Montiel, and A.J. Davison, Real-time monocular SLAM: Why filter? *Proc. IEEE. IEEE,* Anchorage, AK, USA, 2010, 2657–2664.
[19] R. Mascaro, L. Teixeira, T. Hinzmann, R. Siegwart, and M. Chli, GOMSF: Graph-optimization based multi-sensor fusion for robust UAV pose estimation, Proc. IEEE International Conf. on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 2018, 1421–1428.
[20] H. Liu, M. Chen, G. Zhang, H. Bao, and Y. Bao, ICE-BA: Incremental, consistent and efficient bundle adjustment for visual-inertial SLAM. *Proc. 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, 1974–1982.
[21] K. Eckenhoff, Y. Yang, P. Geneva, and G. Huang, Tightly-coupled visual-inertial localization and 3D rigid-body target tracking, *IEEE Robotics and Automation Letters, 4*(2), 2019, 1541–1548.
[22] Z. Gong, P. Liu, F. Wen, R. Ying, and W. Xue, Graph-based adaptive fusion of GNSS and VIO under intermittent GNSS-degraded environment, *IEEE Transactions on Instrumentation and Measurement, 70*, 2021, 1–16.
[23] K.J. Wu, C.X. Guo, G. Georgiou, and S.I. Roumeliotis, VINS on wheels. *Proc. 2017 IEEE International Conf. on Robotics and Automation (ICRA), IEEE*, Singapore, 2017, 5155–5162.
[24] R. Kang, L. Xiong, M. Xu, J. Zhao, and P. Zhang, VINS-vehicle: A tightly-coupled vehicle dynamics extension to visual-inertial state estimator, *Proc. 2019 IEEE Intelligent Transportation Systems Conf. - ITSC*, *IEEE*, Auckland, New Zealand, 2019, 3593–3600.
[25] T. Qin, P. Li, and S. Shen, VINS-mono: A robust and versatile monocular visual-inertial state estimator, *IEEE Transactions on Robotics, 34*(4), 2018, 1004–1020.
[26] G. Cioffi and D. Scaramuzza, Tightly-coupled fusion of global positional measurements in optimization-based visual-inertial odometry, 2020, *arXiv:2003.04159*.
[27] C.V. Angelino, V.R. Baraniello, and L. Cicala, UAV position and attitude estimation using IMU, GNSS and camera, *Proc. 2012 15th International Conf. on Information Fusion*, *IEEE,* Singapore, 2012, 735–742.

Biographies

Yahui Zhang is currently pursuing the Ph.D. degree in mechanical engineering with the School of Mechanical and Electrical Engineering, China University of Mining and Technology. His main research interests are driverless path planning technology for engineering vehicles and design and optimisation of inspection robots.

Linxuan Wang is currently studying with the School of Mechanical and Electrical Engineering, China University of Mining and Technology. His research interest covers the design and optimisation of detection robots.

Aimin Li received the Ph.D. degree from the School of Mechanical and Electrical Engineering, China University of Mining and Technology. He is a Professor with the China University of Mining and Technology. His research interests include intelligent fault diagnosis of belt conveyors, intelligent detection and transportation system of coal equipment, driverless path planning technology for engineering vehicles, and design and optimisation of detection robots.

Yongsheng Zheng is currently pursuing the Ph.D. degree in mechanical engineering with the School of Mechanical and Electrical Engineering, China University of Mining and Technology, Xuzhou, China. His main research interests are mining machinery automation design and construction machinery automation system design.

Mingzhuang Wu is currently pursuing the Ph.D. degree in mechanical engineering with the School of Mechanical and Electrical Engineering, China University of Mining and Technology, Xuzhou, China. His research interests include magnetorheological fluid transmission and control and cement shotcrete structure design and control systems.