



# A low-cost visual inertial odometry for mobile vehicle based on double stage Kalman filter<sup>☆</sup>

Ruping Cen, Tao Jiang\*, Yaoyao Tan, Xiaojie Su, Fangzheng Xue

College of Automation, Chongqing University, Chongqing 400044, China

## ARTICLE INFO

### Article history:

Received 13 November 2021

Revised 27 January 2022

Accepted 11 March 2022

Available online 15 March 2022

### Keywords:

Double stage Kalman filter

Sliding window

Schur complements

Visual inertial odometry

## ABSTRACT

When the low-cost inertial measurement units are used in actual vehicle localization, the measurements suffer from large noise density and unstable bias. Therefore, the integral gyroscope error becomes larger, which results in visual-inertial odometry system degradation or even failure. To this end, this paper proposes a robust visual-inertial odometry method based on a double-stage Kalman filter for a low-cost visual-inertial system, which consists of two Kalman filters. The first filter is a complementary Kalman filter, which uses the accelerometer to correct the gyroscope bias, and then an accurate initial pose estimation is calculated. The second filter is a multi-state observation-constrained Kalman filter, in which the re-projection error of features is calculated based on the multi-state observation constraint strategy to update the system states. Additionally, a Schur complement model is used for the sliding window to marginalize the oldest camera pose of the system states, avoiding the loss of associated information between images and improving the accuracy of the camera pose. Finally, the EuRoC dataset and a home-made low-cost visual-inertial hardware system are used to evaluate the performance of the proposed algorithm. The results show that the accuracy of the low-cost gyroscope bias estimation will decrease when the visual observation is inaccurate in the classic VIO, the proposed algorithm corrects the gyroscope bias with accelerometer measurements, which significantly improves the accuracy and robustness of the vehicle pose estimation when low-cost inertial measurement units are used.

© 2022 Elsevier B.V. All rights reserved.

## 1. Introduction

Simultaneous localization and mapping (SLAM) is the key technology of autonomous mobile robots, which can be used for robot localization [1] and 3D reconstruction [2]. The vision SLAM system [3,4] fails when the scene is poorly textured, the camera moves quickly, or the image contains large noise [5,6]. Therefore, an inertial measuring unit (IMU) is used to improve both the robustness and accuracy of the system. Visual and IMU measurements are typically fused using one of two types of models: loosely-coupled or tightly-coupled. The loosely-coupled model based visual inertial

odometry (VIO) [7] uses IMU and visual measurements to calculate the rotation and translation increments, respectively. Then, an extended Kalman filter (EKF) [8] is used to fuse the IMU and visual calculation results. The tightly-coupled model adds the IMU states into the visual measurement model and jointly optimizes the IMU parameters and camera pose, which can be calculated by EKF [9] and graph optimization [10].

The optimization-based approaches use a tightly coupled model to jointly optimize IMU bias, position, velocity and pose. Leutenegger et al. [11] proposed an open-source image keyframe-based binocular visual-inertial odometer (OKVIS). Mur-Artal et al. [12] proposed a VI-ORB-SLAM system based on the co-visibility graph optimization model, introduced IMU pre-integration [13], closed-loop detection and relocation modules, and proposed a new IMU initialization method. Additionally, Qin et al. [10] proposed an open-source monocular visual-inertial SLAM system (VINS-Mono) [14]. In the optimization-based VIO systems such as OKVIS and VINS-Mono, the pre-integration model [13] is used to integrate the IMU measurements between image intervals directly.

EKF is widely used in signal processing [15,16] and sensor fusion [8,9]. For the filtering-based VIO, Mourikis et al. [17] proposed a multi-state constraint Kalman filter (MSCKF), which uses

<sup>☆</sup> This work was partially supported by Key-Area Research and Development Program of Guangdong Province under Grant (2020B0909020001), the Graduate Research and Innovation Foundation of Chongqing, China (CYB21064), the National Natural Science Foundation of China (62173051, 62106027), Chongqing Science Fund for Outstanding Young Scholars (cstc2019jcyjX0015), Postdoctoral Science Foundation of China (2021M693718) and Postdoctoral Science Fund Project of Chongqing Natural Science Foundation (cstc2020jcyjbsX0053).

\* Corresponding author.

E-mail addresses: [rupingcen@cqu.edu.cn](mailto:rupingcen@cqu.edu.cn) (R. Cen), [jiangtao\\_1992@cqu.edu.cn](mailto:jiangtao_1992@cqu.edu.cn) (T. Jiang), [tanyaoyao@cqu.edu.cn](mailto:tanyaoyao@cqu.edu.cn) (Y. Tan), [suxiaojie@cqu.edu.cn](mailto:suxiaojie@cqu.edu.cn) (X. Su), [xuefangzheng@cqu.edu.cn](mailto:xuefangzheng@cqu.edu.cn) (F. Xue).

the left null space projection method to eliminate features in the system state to reduce the computation cost and improve robustness. From this, a series of derived studies were published to improve the robustness of the MSCKF. Huang et al. [18] proposed the first estimation Jacobian (FEJ) model to improve the consistency of the EKF-SLAM system by selecting the appropriate linearization point. Then, Li et al. [19] proposed MSCKF2.0, which uses the FEJ model to improve the observable consistency of the system and calibrate the external parameters and the time delay parameters between the IMU and the camera online. Sun et al. [20] used the observability-constrained EKF (OC-EKF) [21] model to improve the observable consistency of the system and extend the MSCKF with the use of a stereo camera to estimate the pose of an unmanned aerial vehicle (UAV). In addition, there are variants of the Kalman filter used in the VIO system. Paul et al. [22] proposed a square root inverse filter (SR-ISWF) based on the inverse square root sliding window filter. Bloesch et al. proposed ROVIO [23], which is a tightly coupled VIO based on the photometric error of feature points, and the iterated extended Kalman filter (I-EKF) [24] is used to update the camera pose. Geneva et al. [25] proposed the Schmidt-MSCKF, which adds an image keyframe to the state vector and uses the Schmidt-KF model to update the position of the keyframe. A hybrid visual-inertial navigation algorithm [26] consists of the MSCKF and nonlinear visual-inertial graph optimization. Recently, Geneva et al. [27] proposed a general open-source VIO algorithm, which calibrates the time delay of IMU and camera, as well as internal and external parameters of the camera online. These VIO algorithms add the gyroscope bias to the system states and only use the re-projection errors of the features to update the system states.

On the other hand, both the filtering-based and optimization-based VIO directly integrate the IMU measurements between image intervals to obtain the attitude, speed, and position. Then, the camera pose is augmented with the IMU states. This results in most VIO algorithms obtaining ideal trajectory accuracy on the benchmark because the expensive vision and microelectromechanical system (MEMS) IMU sensors are used to record the dataset. These IMU sensors have a relatively stable noise density and random walk factor, and integrating the IMU measurement can obtain an accurate poster pose [13]. Once a low-cost IMU sensor is used, the measurements of the gyroscope contain large noise, which makes the attitude and position increment calculated by the IMU pre-integration deviate from the truth value. It reduces the performance of nonlinear pose optimization and further deprives the accuracy of the VIO. This paper is inspired by the pose estimation in UAVs [28,29] and inertial navigation systems [30] in which the accelerometer and gyroscope are complementary sensors. In general, gyroscopes are sensitive to high-frequency motion and accelerometers have good static stability. Therefore, when the IMU moves smoothly, the measurements of the accelerator can be used to correct the bias of the gyroscope.

In this paper, we propose a double-stage Kalman filter (DS-KF) for the low-cost visual-inertial odometry that improves the accuracy of vehicle pose estimation. The DS-KF consists of a complementary Kalman filter and an MSCKF. The complementary Kalman filter fuses the accelerometer measurements to correct the gyroscope bias and provides an accurate initial pose for the MSCKF. Then, the MSCKF is used to calculate the re-projection error of the features, which is used to update the camera status and IMU status. Additionally, a Schur complement update model is applied to further improve the accuracy of the VIO and preserve useful information while marginalizing the oldest camera states. The main contributions of this paper are listed as follows:

- (1) A DS-KF model is proposed to provide a more accurate pose for a low-cost VIO system, which consists of a comple-

mentary Kalman filter and an MSCKF. The complementary Kalman filter corrects the gyroscope bias and calculates an accurate initial pose estimation of the IMU.

- (2) A Schur complement model is used to update the covariance matrix which avoids losing useful information when marginalizing the useless camera states in the sliding window.
- (3) A robust stereo VIO system based on the DS-KF is proposed to estimate the vehicle pose, improving the accuracy and robustness of the vehicle pose.

## 2. System overview

The DS-KF-based VIO system, see an overview in Fig. 1, is mainly divided into three parts: the front-end, stage I, and stage II of the back-end. The front-end tracks feature between images. The assumption of constant speed and the gyroscope measurements are used to predict the feature position to increase the robustness of the system. Next, the Lucas-Kanade (LK) [31] algorithm is used to track the features. To avoid feature degradation, new features [32] are added when the interior features are less than 100. In addition, if feature tracking fails, the system is triggered to re-initialize.

A complementary Kalman filter and an MSCKF are used in stage I and stage II of the back-end, respectively. Stage I aims to estimate the gyroscope bias. First, the error  $\mathbf{e}_a$  between the acceleration measurements and the gravity vector as the innovation updates the gyroscope bias with EKF. Then, the gyroscope measurements are used to calculate the IMU attitude, velocity and position by the kinematics model.

Visual measurements are fused with IMU estimates in Stage II, which receives the features and the IMU pose from the front end and Stage I, respectively. First, the initial pose of the camera is estimated by the external parameters and the IMU pose. Then, the historic camera states and the current IMU state are added to the EKF state. The re-projection error of features is calculated by the camera with the same view in the process of the multi-state constraints. Also, the re-projection error is used for the MSCKF to update the historic camera pose and IMU pose. Finally, the Schur complement model is used to update the covariance matrix when marginalizing the oldest camera state in the sliding window. It is an efficient way to limit the length of system states to improve computational efficiency. The detailed steps of the robust VIO based on the DS-KF model are described in Algorithm 1.

## 3. System model

### 3.1. IMU bias model

The IMU sensor consists of a 3-axis gyroscope and 3-axis accelerometer, which provide 3D linear acceleration and angular velocity in the body frame. Usually, the bias of the accelerometer and gyro based on MEMS are changed over time. Thus, the IMU measurements are modeled as follows:

$$\begin{cases} \mathbf{w}_m(t) = \mathbf{w}_b(t) + \mathbf{b}^g(t) + \mathbf{n}^g \\ \mathbf{a}_m(t) = \mathbf{R}_{wb}^T(\mathbf{a}_w(t) - \mathbf{g}) + \mathbf{b}^a(t) + \mathbf{n}^a \end{cases} \quad (1)$$

where  $\mathbf{n}^g$  and  $\mathbf{n}^a$  are Gaussian white noises with zero mean of the gyroscope and accelerometer, respectively.  $\mathbf{w}_m(t) \in \mathbb{R}^3$  and  $\mathbf{a}_m(t) \in \mathbb{R}^3$  are the measurements of angular velocity and linear acceleration, respectively.  $\mathbf{g}$  is the gravity vector.  $\mathbf{R}_{wb} \in \mathbb{R}^{3 \times 3}$  is the rotation matrix, which maps the points from the body coordinate  $B$  to the earth coordinate  $W$ .

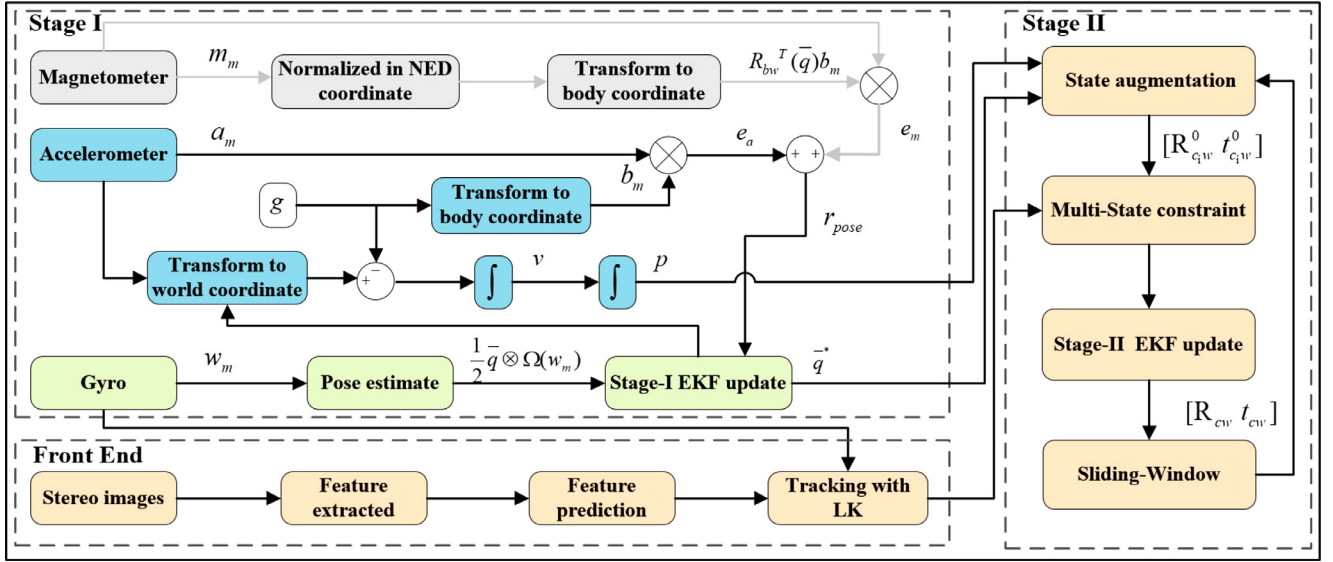


Fig. 1. The structure of the double stage Kalman filter

**Algorithm 1** The robust VIO Based on DS-KF.

**Input:** The left image  $I_{c,0}$ ; right image  $I_{c,1}$ ;  
IMU measurements  $\mathbf{w}_m(t)$ ,  $\mathbf{a}_m(t)$  and  $\mathbf{m}_m(t)$ .  
**Output:** The camera pose  $[\hat{\mathbf{q}}_{cw} \ \mathbf{p}_{cw}]$ ;

- 1: Track features between current and last images;
- 2: **for** IMU measurements between images period **do**
- 3:   Calculate the IMU initial pose by integral  $\mathbf{w}_m(t)$ ;
- 4:   Calculate accelerometer error  $\mathbf{e}_a = \mathbf{a}_m - \mathbf{R}_{wb}^T \mathbf{g}$ ;
- 5:   **if** the  $\mathbf{m}_m(t)$  is available **then**
- 6:     Calculate magnetometer error  $\mathbf{e}_m = \mathbf{m}_m - \mathbf{b}_m$ ;
- 7:     The residual  $\mathbf{r}_{pose} = [\mathbf{e}_a \ \mathbf{e}_m]^T$ ;
- 8:   **else**
- 9:     The residual  $\mathbf{r}_{pose} = \mathbf{e}_a$ ;
- 10:   **end if**
- 11:   Update the IMU attitude  $\hat{\mathbf{q}}_{bw}$  and gyroscope bias  $\mathbf{b}^w$  by equation (10);
- 12: **end for**
- 13: Augment the new camera pose by equation (18);
- 14: Triangulate features with large disparity are triangulated;
- 15: Build the reprojection error  $\mathbf{r}_{i,j}$  for all 3D point  $P_j$  observed by the cameras in buffer;
- 16: Update the IMU and cameras states  $\mathbf{x}$  by equation (31);
- 17: Margin the old camera pose based on the Schur complement model;
- 18: **return** The camera pose;

The bias of the gyroscope and accelerometer are modeled as the time-varying bias  $\mathbf{b}^g(t)$  and  $\mathbf{b}^a(t)$ , given as

$$\begin{cases} \mathbf{b}^g(t+1) = \mathbf{b}^g(t) + \mathbf{n}^{wg}(t) \\ \mathbf{b}^a(t+1) = \mathbf{b}^a(t) + \mathbf{n}^{wa}(t) \end{cases} \quad (2)$$

where  $\mathbf{n}^{wa}(t)$  and  $\mathbf{n}^{wg}(t) \in \mathbb{R}^3$  are Gaussian white noises with zero mean.

The IMU state is defined as a  $16 \times 1$  vector  $\mathbf{x}_{imu}$ , which consist of quaternion  $\hat{\mathbf{q}}_{wb}$ , position  $\mathbf{p}_w$ , velocity  $\mathbf{v}_w$ , and bias  $\mathbf{b}^a$  and  $\mathbf{b}^g$ .

$$\mathbf{x}_{imu} = [\hat{\mathbf{q}}_{wb} \ \mathbf{p}_w \ \mathbf{v}_w \ \mathbf{b}^a \ \mathbf{b}^g]^T \in \mathbb{R}^{16} \quad (3)$$

Similarly, the error state of the IMU is defined as

$$\tilde{\mathbf{x}}_{imu} = [\tilde{\mathbf{q}}_{wb} \ \tilde{\mathbf{p}}_w \ \tilde{\mathbf{v}}_w \ \tilde{\mathbf{b}}^a \ \tilde{\mathbf{b}}^g]^T \in \mathbb{R}^{16} \quad (4)$$

where  $\tilde{\mathbf{p}}_w$  and  $\tilde{\mathbf{v}}_w$  are the position and velocity errors of the IMU in the global coordinate  $W$ .  $\tilde{\mathbf{b}}^a$  and  $\tilde{\mathbf{b}}^g$  are the errors of the IMU bias.

The quaternion error is given by a small angle rotation  $\delta\theta$ .

$$\delta\tilde{\mathbf{q}} = \tilde{\mathbf{q}} \otimes \hat{\mathbf{q}}^{-1} \simeq \begin{bmatrix} \frac{1}{2}\delta\theta^T & 1 \end{bmatrix}^T \quad (5)$$

Thus, the error state of the IMU can be rewritten as:

$$\tilde{\mathbf{x}}_{imu} = [\delta\theta \ \tilde{\mathbf{p}}_w \ \tilde{\mathbf{v}}_w \ \tilde{\mathbf{b}}^a \ \tilde{\mathbf{b}}^g]^T \in \mathbb{R}^{15} \quad (6)$$

**3.2. Complementary Kalman filter**

The attitude state of IMU  $\mathbf{x}_{pose}$  is defined as a  $7 \times 1$  vector, which includes the attitude represented by the quaternion and the gyroscope bias.

$$\mathbf{x}_{pose} = [\tilde{\mathbf{q}}_{wb} \ \mathbf{b}^g]^T \quad (7)$$

where  $\tilde{\mathbf{q}}_{wb} = [q_x \ q_y \ q_z \ q_w]^T$  is a quaternion vector. The residual between the gravity measurements and the true value is defined as

$$\mathbf{e}_a = \mathbf{a}_m - \mathbf{R}_{wb}^T \mathbf{g} \quad (8)$$

where  $\mathbf{a}_m$  is the measurement of the accelerometer,  $\mathbf{R}_{wb}$  is the matrix corresponding to quaternions  $\tilde{\mathbf{q}}_{wb}$ . and  $\mathbf{r}_{pose} = \mathbf{e}_a$ . Hence, the observation matrix  $\mathbf{H}_{pose}$  can be described as the derivative of the residual  $\mathbf{r}_{pose}$  relative to the attitude  $\mathbf{x}_{pose}$

$$\mathbf{H}_{pose} = \frac{\partial \mathbf{r}_{pose}}{\partial \mathbf{x}_{pose}} \quad (9)$$

Finally, the attitude and bias of the IMU is updated with the measurements of the accelerometer with standard measurement update procedure of the EKF.

$$\begin{cases} \mathbf{x}_{pose \ k|k} = \mathbf{x}_{pose \ k|k-1} + \mathbf{K}^0 \mathbf{r}_{pose} \\ \mathbf{K}^0 = \mathbf{P}^0_{k|k-1} \mathbf{H}_{pose}^T (\mathbf{H}_{pose} \mathbf{P}^0_{k|k-1} \mathbf{H}_{pose}^T + \mathbf{R}^0)^{-1} \\ \mathbf{P}^0_{k|k} = (\mathbf{I} - \mathbf{K}^0 \mathbf{H}_{pose}) \mathbf{P}^0_{k|k-1} \end{cases} \quad (10)$$

By fusing the measurements of the accelerometer, the gyroscope bias and an accurate IMU attitude  $\tilde{\mathbf{q}}_{wb}$  is computed. In addition, the attitude  $\tilde{\mathbf{q}}_{wb}$  will have a positive effect on the speed and position in Eq. (11). Furthermore, if the magnetometer is available, the residual of the magnetometer is also can be added into  $\mathbf{r}_{pose}$ ,

which described in Madgwick et al. [33]. It can also improve the unobservable directions of the global yaw angle [18]. But, the magnetometer must be calibrated off-line and away from the magnetic interference source. And the use of the accelerometer to observe the gravity vector is based on the assumption that the movement of the IMU is smooth. The accuracy of the fusion algorithm decreases when the IMU moves with a large acceleration on the three axes.

### 3.3. IMU propagation

The continuous kinematic model of the IMU are:

$$\begin{cases} \dot{\bar{\mathbf{q}}}_{bw} = \frac{1}{2} \boldsymbol{\Omega}(\mathbf{w}(t)) \bar{\mathbf{q}}_{bw}(t) \\ \dot{\mathbf{p}}_w = \mathbf{v}_w \\ \dot{\mathbf{v}}_w = \mathbf{R}_{wb}(\bar{\mathbf{q}}) \mathbf{a}(t) \\ \dot{\mathbf{b}}^a = \mathbf{n}^{wa} \\ \dot{\mathbf{b}}^g = \mathbf{n}^{wg} \end{cases} \quad (11)$$

where  $\mathbf{w}(t)$  and  $\mathbf{a}(t)$  are the true values of the angular velocity and linear acceleration.

$$\begin{cases} \mathbf{a}(t) = \mathbf{R}_{wb}(\mathbf{a}_m(t) - \mathbf{b}^a(t) - \mathbf{n}^a) + \mathbf{g} \\ \mathbf{w}(t) = \mathbf{w}_m(t) - \mathbf{b}^g(t) - \mathbf{n}^g \end{cases} \quad (12)$$

$\mathbf{b}^g(t)$  is the gyro bias estimated in the complex Kalman filter.  $\boldsymbol{\Omega}(\cdot)$  is a function that changes a vector  $\mathbf{w}$  into a skew-symmetric matrix

$$\boldsymbol{\Omega}(\mathbf{w}) = \begin{bmatrix} -\mathbf{w}^\times & \mathbf{w} \\ -\mathbf{w}^T & 0 \end{bmatrix}_{4 \times 4}, \text{ And } \mathbf{w}^\times = \begin{bmatrix} 0 & -w_z & w_y \\ w_z & 0 & -w_x \\ -w_y & w_x & 0 \end{bmatrix} \quad (13)$$

The propagation model of the IMU states is obtained by the linearized continuous kinematics for the IMU error state.

$$\hat{\mathbf{x}}_{imu}(k+1) = \mathbf{F} \hat{\mathbf{x}}_{imu}(k) + \mathbf{G} \mathbf{n}_{imu} \quad (14)$$

where  $\mathbf{n}_{imu}$  is the process noise.  $\mathbf{n}^a$  and  $\mathbf{n}^g$  are the noise densities of the accelerometer and gyroscope, respectively.  $\mathbf{n}^{wa}$  and  $\mathbf{n}^{wg}$  are the random walk rate of the accelerometer and gyroscope biases, respectively.

$$\mathbf{n}_{imu} = [\mathbf{n}^a \quad \mathbf{n}^{wa} \quad \mathbf{n}^g \quad \mathbf{n}^{wg}]^T \quad (15)$$

The matrices  $\mathbf{F}$  and  $\mathbf{G}$  are derived from literature [17]. The discrete time state transition matrix and noise covariance matrix are

$$\begin{cases} \Phi_{imu}(k) = \exp\left(\int_{t_k}^{t_{k+1}} \mathbf{F} d\tau\right) \\ \mathbf{Q}_{imu}(k) = \exp\left(\int_{t_k}^{t_{k+1}} \Phi_{imu}(\tau) \mathbf{G} \mathbf{n}_{imu} \mathbf{G}^T \Phi_{imu}(\tau)^T d\tau\right) \end{cases} \quad (16)$$

The 4th order Runge-Kutta is used to integrate Eq. (7) for  $\mathbf{x}_{imu}$ .

### 3.4. MSCKF update

The covariance matrix of the VIO system at the  $k$ th step can be described as

$$\mathbf{P}_{k+1|k}^1 = \begin{bmatrix} \Phi_{imu} \mathbf{P}_{l_{k|k}} \Phi_{imu}^T + \mathbf{Q}_{imu} & \Phi_{imu} \mathbf{P}_{l_{k|k}} \\ \mathbf{P}_{l_{k|k}}^T \Phi_{imu}^T & \mathbf{P}_{CC_{k|k}} \end{bmatrix} \quad (17)$$

Once a new image is received, the initial pose of the new camera at time  $k$  can be calculated from the IMU pose.

$$\begin{cases} \mathbf{R}_{c_k w} = \mathbf{R}_{bc}^T \mathbf{R}_{bw}(\bar{\mathbf{q}}_{bw}) \\ \mathbf{t}_{wc_k} = \mathbf{p}_w(k) + \mathbf{R}_{bw}^T(\bar{\mathbf{q}}_{bw}) \mathbf{t}_{bc} \end{cases} \quad (18)$$

The system error state is defined as follows and consists of the IMU state and  $N$  camera pose:

$$\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_{imu} \mid \tilde{\mathbf{x}}_{c_1} \quad \tilde{\mathbf{x}}_{c_2} \quad \dots \quad \tilde{\mathbf{x}}_{c_i}]^T \quad (19)$$

where

$$\tilde{\mathbf{x}}_{c_i} = [\delta \boldsymbol{\theta}_{c_i w} \quad \delta \tilde{\mathbf{p}}_{c_i w}]^T \quad (20)$$

Suppose a 3D map point  $\mathbf{P}_j = [X_j, Y_j, Z_j]^T$  is observed by a stereo camera.  $[\mathbf{R}_{c_i w} \mathbf{t}_{c_i w}]$  are the poses of the left camera. The re-projected points  $\mathbf{p}_j^l$  and  $\mathbf{p}_j^r$  in the normalized coordinate system can be obtained according to the camera model as

$$\begin{cases} \mathbf{p}_{c_i,j}^l = \begin{bmatrix} u_{c_i,j}^l \\ v_{c_i,j}^l \\ 1 \end{bmatrix} = \frac{1}{Z_{c_i,j}^l} \mathbf{P}_{c_i,j}^l \\ \mathbf{p}_{c_i,j}^r = \begin{bmatrix} u_{c_i,j}^r \\ v_{c_i,j}^r \\ 1 \end{bmatrix} = \frac{1}{Z_{c_i,j}^r} \mathbf{P}_{c_i,j}^r \end{cases} \quad (21)$$

where

$$\begin{cases} \mathbf{P}_{c_i,j}^l = [X_{c_i,j}^l \ Y_{c_i,j}^l \ Z_{c_i,j}^l]^T = \mathbf{R}_{c_i w}(\mathbf{P}_j - \mathbf{t}_{wc_i}) \\ \mathbf{P}_{c_i,j}^r = [X_{c_i,j}^r \ Y_{c_i,j}^r \ Z_{c_i,j}^r]^T = \mathbf{R}_{c_i c_r}^T \mathbf{P}_{c_i,j}^l + \mathbf{t}_{c_i c_r} \end{cases} \quad (22)$$

And  $[\mathbf{R}_{c_i c_r} \ \mathbf{t}_{c_i c_r}]$  are external parameters of the stereo camera. Hence, the measurements of the map point  $\mathbf{P}_j$  in the current frame  $c_i$  can be described as

$$\hat{\mathbf{z}}_{c_i,j} = \begin{bmatrix} u_{c_i,j}^l \\ v_{c_i,j}^l \\ u_{c_i,j}^r \\ v_{c_i,j}^r \end{bmatrix} \quad (23)$$

The re-projection error of the 3D point  $\mathbf{P}_j$  is given by

$$\mathbf{r}_{i,j} = \mathbf{z}_{c_i,j} - \hat{\mathbf{z}}_{c_i,j} \quad (24)$$

Then, linearizing the measurement model, the measurements can be expressed as

$$\mathbf{r}_{i,j} = \mathbf{H}_{X_{i,j}} \tilde{\mathbf{x}}_{c_i} + \mathbf{H}_{f_{i,j}} \tilde{\mathbf{p}}_j + \mathbf{n} \quad (25)$$

$\mathbf{n}$  is Gaussian white noises with zero mean. Applying the chain derivation rule to Eq. (24), the Jacobian matrix is described as:

$$\begin{cases} \mathbf{H}_{X_{i,j}} = \frac{\partial \mathbf{r}_{c_i,j}}{\partial \mathbf{P}_{c_i,j}^l} \frac{\partial \mathbf{P}_{c_i,j}^l}{\partial \tilde{\mathbf{x}}_{c_i}} + \frac{\partial \mathbf{r}_{c_i,j}}{\partial \mathbf{P}_{c_i,j}^r} \frac{\partial \mathbf{P}_{c_i,j}^r}{\partial \tilde{\mathbf{x}}_{c_i}} \\ \mathbf{H}_{f_{i,j}} = \frac{\partial \mathbf{r}_{c_i,j}}{\partial \mathbf{P}_{c_i,j}^l} \frac{\partial \mathbf{P}_{c_i,j}^l}{\partial \tilde{\mathbf{p}}_j} + \frac{\partial \mathbf{r}_{c_i,j}}{\partial \mathbf{P}_{c_i,j}^r} \frac{\partial \mathbf{P}_{c_i,j}^r}{\partial \tilde{\mathbf{p}}_j} \end{cases} \quad (26)$$

where

$$\begin{cases} \mathbf{H}_{X_{i,j}} = [\mathbf{H}_{\theta_{c_i w,j}} \ \mathbf{H}_{p_{c_i w,j}}] \\ \mathbf{H}_{f_{i,j}} = (\mathbf{J}^l + \mathbf{J}^r \mathbf{R}_{c_i c_r}^T) \mathbf{R}_{c_i w} \end{cases} \quad (27)$$

$\mathbf{H}_{X_{i,j}}$  are the Jacobian of rotation and translation with respect to the system error state  $\tilde{\mathbf{x}}_{c_i}$ .

$$\begin{cases} \mathbf{H}_{\theta_{c_i w,j}} = (\mathbf{J}^l + \mathbf{J}^r \mathbf{R}_{c_i c_r}^T) [\mathbf{P}_{c_i,j}^l]^\times \\ \mathbf{H}_{p_{c_i w,j}} = -(\mathbf{J}^l + \mathbf{J}^r \mathbf{R}_{c_i c_r}^T) \mathbf{R}_{c_i w} \end{cases} \quad (28)$$

$\mathbf{J}^l$  and  $\mathbf{J}^r$  are the Jacobian matrices of the re-projection error with respect to 3D map points  $\mathbf{P}_j$ .

$$\begin{cases} \mathbf{J}^l = \frac{\partial \mathbf{r}_{c_i,j}}{\partial \mathbf{P}_{c_i,j}^l} = \frac{1}{Z_{c_i,j}^l} \begin{bmatrix} f_x^l & 0 & -\frac{f_x^l X_j}{Z_{c_i,j}^l} \\ 0 & f_y^l & -\frac{f_y^l Y_j}{Z_{c_i,j}^l} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ \mathbf{J}^r = \frac{\partial \mathbf{r}_{c_i,j}}{\partial \mathbf{P}_{c_i,j}^r} = \frac{1}{Z_{c_i,j}^r} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ f_x^r & 0 & -\frac{f_x^r X_j}{Z_{c_i,j}^r} \\ 0 & f_y^r & -\frac{f_y^r Y_j}{Z_{c_i,j}^r} \end{bmatrix} \end{cases} \quad (29)$$



To eliminate the influence of the feature position on the system states [17], the residual  $\mathbf{r}_{i,j}$  is projected onto the left null space of matrix  $\mathbf{H}_{f_i,j}$ .

$$\mathbf{r}_{i,j}^o = \mathbf{A}^T \mathbf{r}_{i,j} = \mathbf{H}_{\mathbf{x}_{i,j}}^o \tilde{\mathbf{x}}_i + \mathbf{n}^o \quad (30)$$

$\mathbf{A}$  is the left null space matrix of  $\mathbf{H}_{f_i,j}$ , satisfying  $\mathbf{A}^T \mathbf{H}_{f_i,j} = \mathbf{0}$ . When a new image is received, the IMU state is updated by the measurements of the gyroscope and acceleration. Then, the new camera's pose is estimated by Eq. (18), and the new camera state is added into the system state  $\mathbf{X}$ . Finally, the system states and covariance matrix are updated with the re-projection error of the 3D point in the stereo camera.

$$\begin{cases} \mathbf{x}_{k+1|k+1} = \mathbf{x}_{k+1|k} + \mathbf{K}^1 \mathbf{r} \\ \mathbf{K}^1 = \mathbf{P}_{k+1|k}^1 \mathbf{H}^T (\mathbf{H} \mathbf{P}_{k+1|k}^1 \mathbf{H}^T + \mathbf{R}^1)^{-1} \\ \mathbf{P}_{k+1|k+1}^1 = \mathbf{P}_{k+1|k}^1 - \mathbf{K} \mathbf{H} \mathbf{P}_{k+1|k}^1 \end{cases} \quad (31)$$

where  $\mathbf{H}$  and  $\mathbf{r}$  are the observation matrix and the re-projection error, respectively, for all 3D map points that can be seen in the current frame.  $\mathbf{H}$  consists of  $\mathbf{H}_{\mathbf{x}_{i,j}}^o$ , and  $\mathbf{r}$  consists of  $\mathbf{r}_{i,j}^o$ .

### 3.5. The Schur complement model

In general, a sliding window is used for bundle adjustment(BA) [34,35] optimization at the back-end of a visual SLAM [4]. It is an effective way to improve computational efficiency by limiting the length of the system states. To further improve the robustness of the MSCKF algorithm [17], which constructs the re-projection errors from multi-cameras that observe the same feature. However, with the continuous movement of the robot, the length of the state variable of the VIO system is continuously increasing, which leads to an increased computational cost to estimate the camera pose. Therefore, to limit the computational cost to achieve real-time system operation, it is necessary to discard the oldest image frames in the system state and maintain the length of the system state inconsistent. However, directly deleting the oldest state of the system [17] and its associated landmark points results in the loss of information. Thus, the Schur complement model is used to marginalize the oldest state in this paper.

Fig. 2 (a) shows the difference between the system variables before and after marginalization. Suppose that  $\mathbf{x}_1$  is the oldest frame that should be marginalized. Fig. 2(a) shows that there is a connection between  $\mathbf{x}_1$ ,  $\mathbf{x}_0$ , and  $\mathbf{x}_2$ , which indicates that there is an indirect connection between  $\mathbf{x}_0$  and  $\mathbf{x}_2$ . If the  $\mathbf{x}_1$  variable is directly removed, the connection represented by the magenta arrow between  $\mathbf{x}_0$  and  $\mathbf{x}_2$  is lost. Fig. 2(b) shows the covariance matrix during the process of the marginalization of state  $\mathbf{x}_1$ . First, the elements related to the state variable  $\mathbf{x}_1$  in the covariance matrix  $\mathbf{P}_{k+1|k+1}^1$  are moved to the left side. Then, the system covariance matrix is split into four matrices  $\mathbf{\Lambda}_{BB}$ ,  $\mathbf{\Lambda}_{AB}$  and  $\mathbf{\Lambda}_{AA}$ .

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_{BB} & \mathbf{\Lambda}_{AB}^T \\ \mathbf{\Lambda}_{AB} & \mathbf{\Lambda}_{AA} \end{bmatrix} \quad (32)$$

Where  $\mathbf{\Lambda}_{BB}$  is the covariance matrix of state  $\mathbf{x}_1$ .  $\mathbf{\Lambda}_{AA}$  is the covariance of the retained states and,  $\mathbf{\Lambda}_{AB}$  is the covariance between state  $\mathbf{x}_1$  and the retained states. In the process of marginalization, state  $\mathbf{x}_1$  moves to the top of the covariance matrix first. Then, by using Schur's complement model, the new covariance is given as

$$\mathbf{\Lambda}_{AA}^{new} = \mathbf{\Lambda}_{AA} - \mathbf{\Lambda}_{AB} \mathbf{\Lambda}_{BB}^{-1} \mathbf{\Lambda}_{AB}^T \quad (33)$$

$\mathbf{\Lambda}_{AA}^{new}$  is the new covariance matrix of the VIO system, which maintains the prior information of state variable  $\mathbf{x}_1$ . This indicates that

**Table 1**

the average root-mean-square error (RMSE) of VINS-Mono, MSCKF-VIO and the proposed algorithm On EuRoC datasets in meters.

Sequence	VINS-Mono	VINS-Mono no loop	MSCKF-VIO	DS-KF
V101	<b>0.068</b>	0.089	0.090	0.077
V102	<b>0.084</b>	0.111	0.126	0.117
V103	0.193	0.187	0.217	<b>0.145</b>
V201	0.081	0.086	0.087	<b>0.061</b>
V202	0.166	0.158	<b>0.143</b>	0.144
V203	<b>0.225</b>	0.278	1.663	0.233
MH01	0.120	0.156	0.114	<b>0.098</b>
MH02	<b>0.129</b>	0.178	0.183	0.169
MH03	<b>0.138</b>	0.220	0.228	0.203
MH04	<b>0.181</b>	0.347	0.298	0.304
MH05	<b>0.218</b>	0.302	0.350	0.412

a new constraint is added between states  $\mathbf{x}_0$  and  $\mathbf{x}_2$ , as shown in Fig. 2(a) with the magenta arrow.

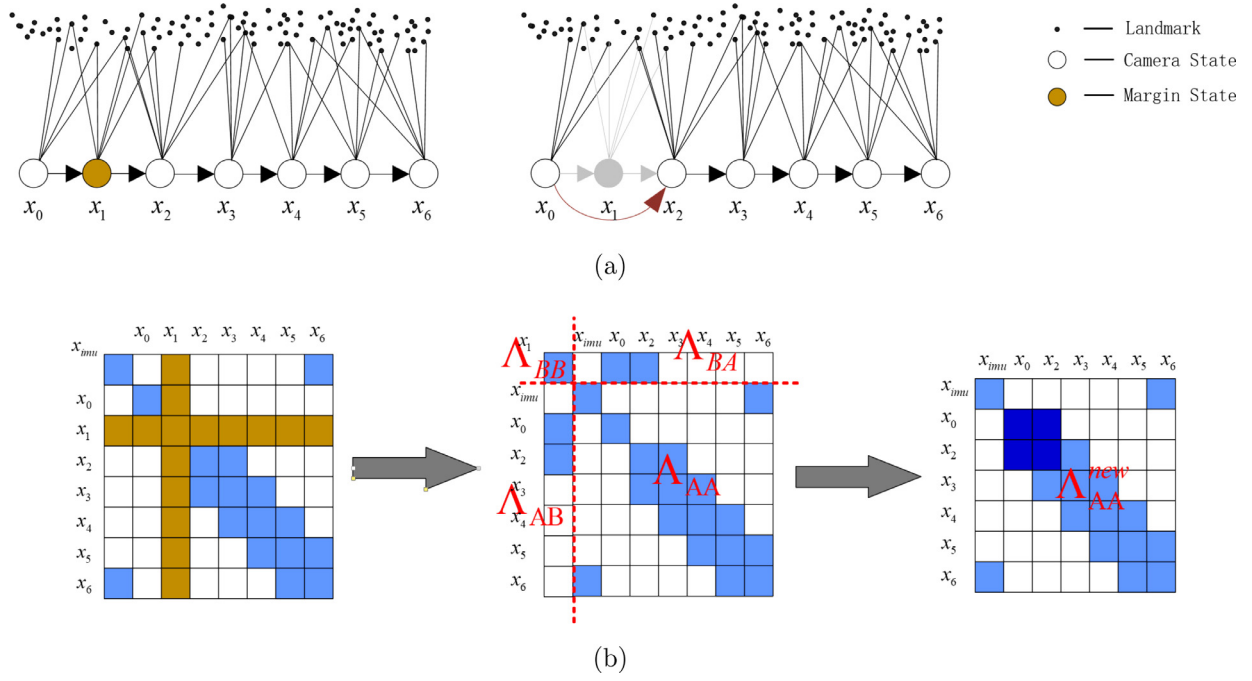
## 4. Experiments

In this section, the proposed DS-KF algorithm is compared with state-of-art VIO algorithms, both optimization-based VINS-Mono [10] and EKF-based MSCKF-VIO [20]. First, the pose estimate performance of the VIO system is evaluated with the EuRoC dataset [36]. The EuRoC dataset provides stereo image pairs with a resolution of  $640 \times 480$  at 20 Hz and IMU measurements with ADIS16448 at 200 Hz. The camera and IMU are synchronized with a hardware device. This dataset is usually used to evaluate the performance of the SLAM or VIO system. Second, the robustness of the proposed algorithm is evaluated in a campus environment with a homemade camera-IMU system, which uses a ZED camera with a  $1280 \times 720$  resolution at 20 Hz, Xsens MIT-30 IMU at 100 Hz, and low-cost IMU BMI-160 at 100 Hz. The camera and two IMU sensors are mounted on the top of a mobile vehicle. Fig. 5 shows the platform. A laptop with an i7-4710MQ (4 cores with 2.50 GHz) CPU and 16 GB RAM is used.

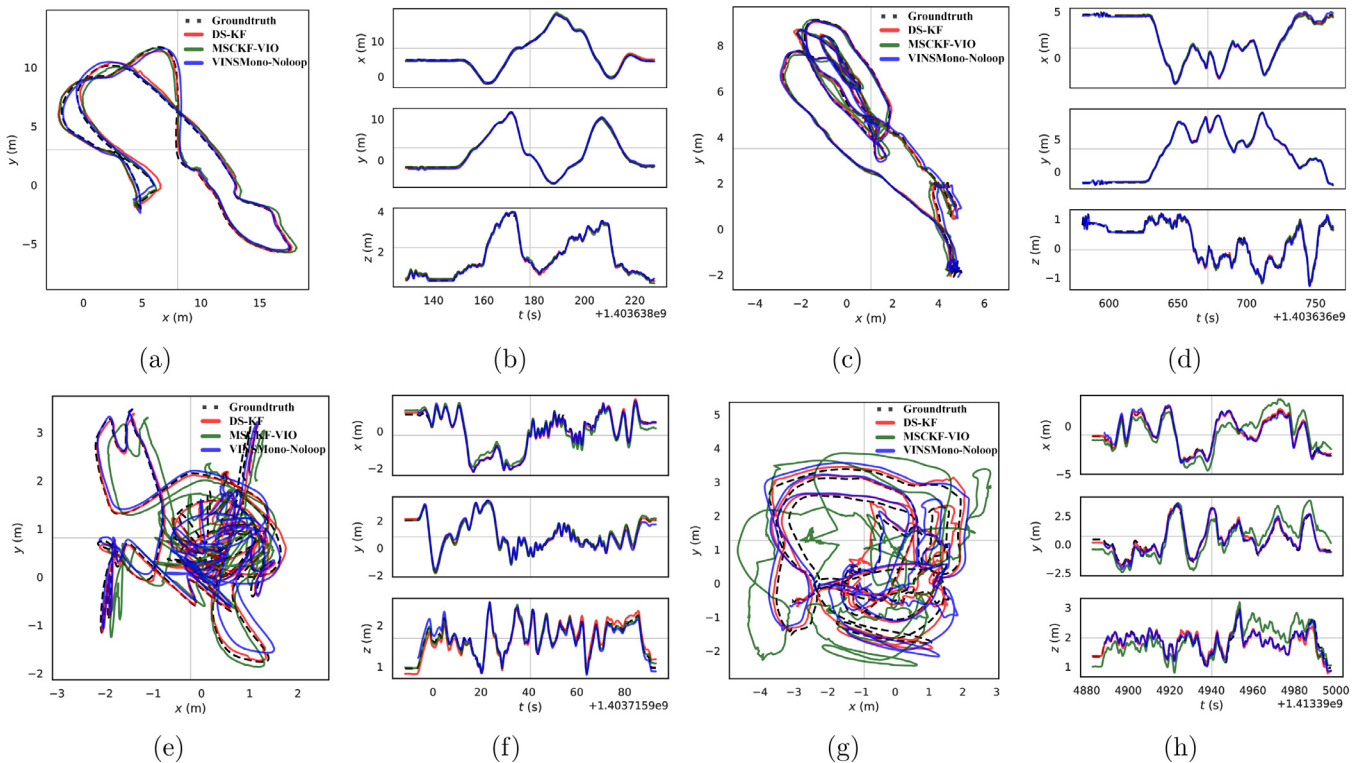
### 4.1. Performance on EuRoC benchmark

The EuRoC dataset is used to test the performance of the algorithm, and the absolute trajectory error (ATE) is the evaluation criterion. All algorithms are run ten times on the EuRoC dataset, and the root means square error (RMSE) of ATE is the final result. In addition, the EuRoC dataset lacks magnetometer measurements, so only an accelerometer is used to calibrate the gyro bias.

Fig. 3 illustrates the trajectory of VINS-Mono, MSCKF-VIO, and the proposed method on the EuRoC dataset. The red, green, and blue lines represent the trajectories of the proposed algorithm, MSCKF-VIO, and VINS-Mono algorithm, respectively, and the black line represents the ground truth. Fig. 3(a) and (c) show the trajectories of the three algorithms for the MH01 and MH04 sequences, respectively. Fig. 3(b) and (d) show the RMSE of the algorithms for the three axes, respectively. As can be observed from Fig. 3(a) to (d), the RMSE of the proposed method is slightly better than VINS-MONO and MSCKF-VIO. Fig. 3(e) and (g) show the trajectories of the three algorithms on the sequence of V103 and V203, respectively. Fig. 3(f) and (h) show the trajectory of the algorithms along three axes, respectively. As can be seen from Fig. 3(e) to (h), the RMSE of the proposed method is better than that of MSCKF-VIO. A more detailed RMSE of the VINS-MONO MSCKF-VIO and the proposed method are listed in Table 1. As can be observed from Table 1, VINS-Mono with loop closure is still the most robust algorithm on the EuRoC dataset with a medium precision IMU sensor. The accuracy of the proposed algorithm is the same as VINS-Mono without loopback detection mode for the EuRoC dataset,



**Fig. 2.** The schematic of the marginalization process. (a) The framework of marginalized states  $x_1$ . The black dot represents the landmarks in 3D space, and the circle represents the state of the camera. (b) The covariance change process at the marginalized state  $x_1$ .



**Fig. 3.** The Trajectory of the DS-KF, VINS-Mono and MSCKF-VIO on the EuRoC datasets MH01, MH04, V103, and V203 sequences.

and it is slightly better than VINS-Mono in VIO mode for most sequences.

The main difference between the MH01, MH04, V103 and V203 sequences is that the light of the sequences MH01 and MH04 is relatively uniform and the light change is only a small part, whereas the light of the sequences V103 and V203 is very low, significantly reducing the feature point tracking performance. In

the traditional VIO, the error of the gyroscope is added into the system state variables and corrected by the visual measurements. When the visual measurements are inaccurate, the gyroscope bias cannot be corrected accurately, causing the integration error to become larger and the pose estimation to deviate further from the true value. This situation is particularly evident in a dark environment, as shown in Fig. 3(e)–(h). However, our algorithm can

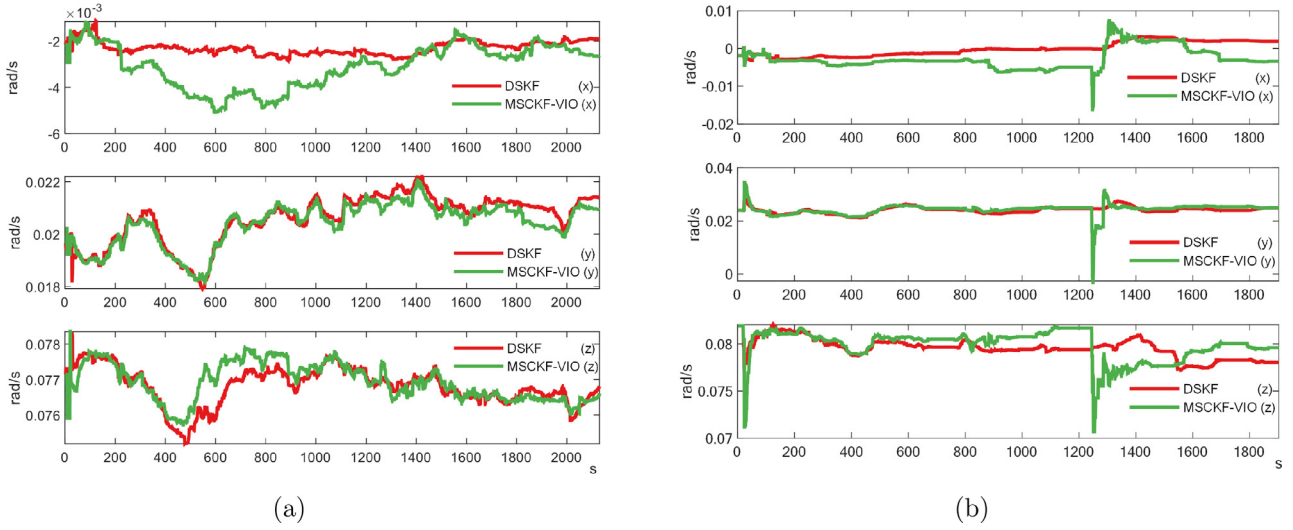


Fig. 4. The gyroscope bias estimated on the EuRoC datasets V103, and V203 sequences.

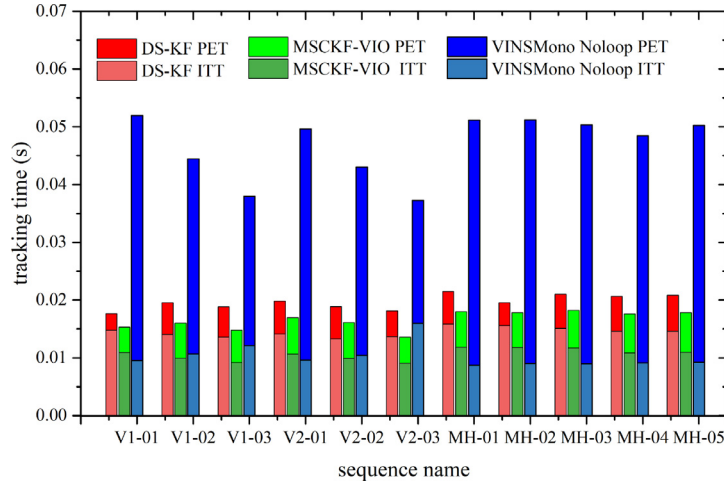


Fig. 5. The average tracking time of VINS-Mono, MSCKF-VIO and the proposed algorithm on the EuRoC datasets.

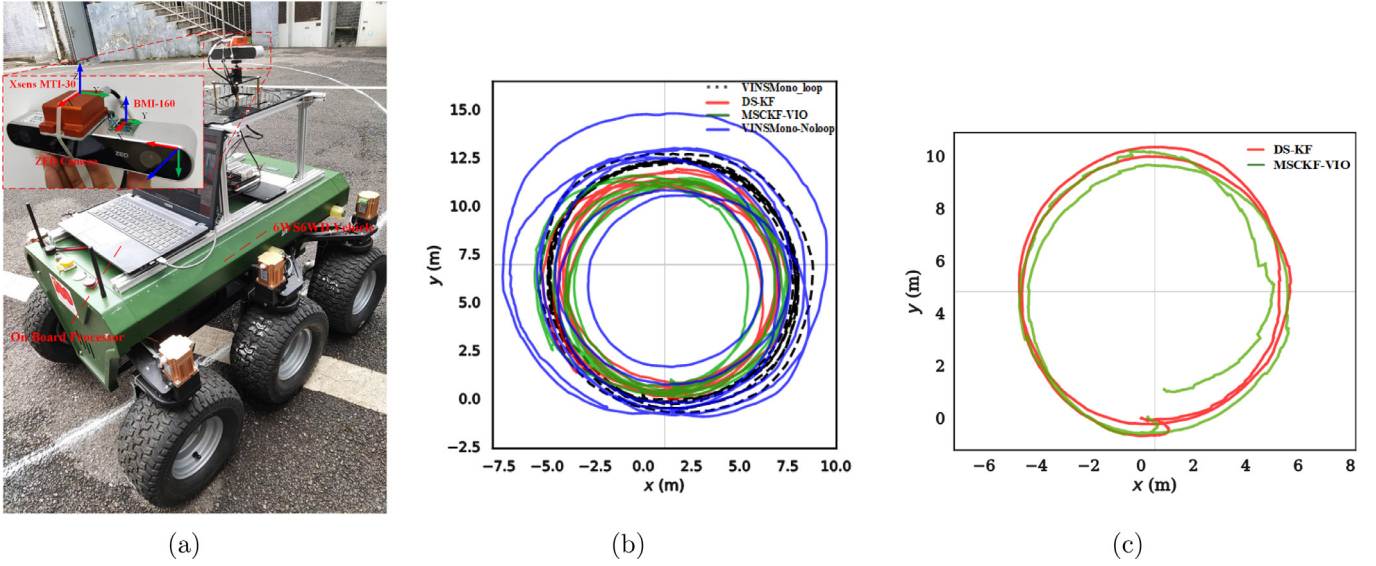
correct the gyroscope error with accelerometer measurements in this case and improve the accuracy of the IMU attitude calculation. The gyroscope bias estimated on the V103 and V203 sequences is listed in Fig. 4. Therefore, correcting the gyroscope deviation value through the accelerometer can improve the accuracy of VIO pose estimation.

Fig. 5 shows the average image tracking time for the DS-KF, MSCKF-VIO, and VINS-Mono (VIO Mode). In these VIO systems, image tracking is usually divided into two parts, feature point tracking and pose optimization, and two threads are used to run these two modules separately. In Fig. 5, the red, green, and blue bars represent the average image tracking time for the DS-KF, MSCKF-VIO, and VINS-Mono, respectively. In addition, the dark red bar represents the pose estimation time, and the pale color indicates the feature tracking time. It can be seen from Fig. 4 that the average image tracking time for the DS-KF, MSCKF-VIO, and VINS-Mono on the entire EuRoC dataset is 14.36 ms, 10.75 ms, and 10.3 ms, respectively. The average pose estimation time for the DS-KF, MSCKF-VIO, and VINS-Mono on the whole EUROC dataset is 5.18 ms, 5.93 ms, and 36.58 ms, respectively. The pose estimation speed of the DS-KF and MSCKF-VIO is 6 times that of VINS-Mono. This is due to the proposed algorithm and MSCKF-VIO using the multi-state observation constraint Kalman filter (MSCKF) structure based on the EKF.

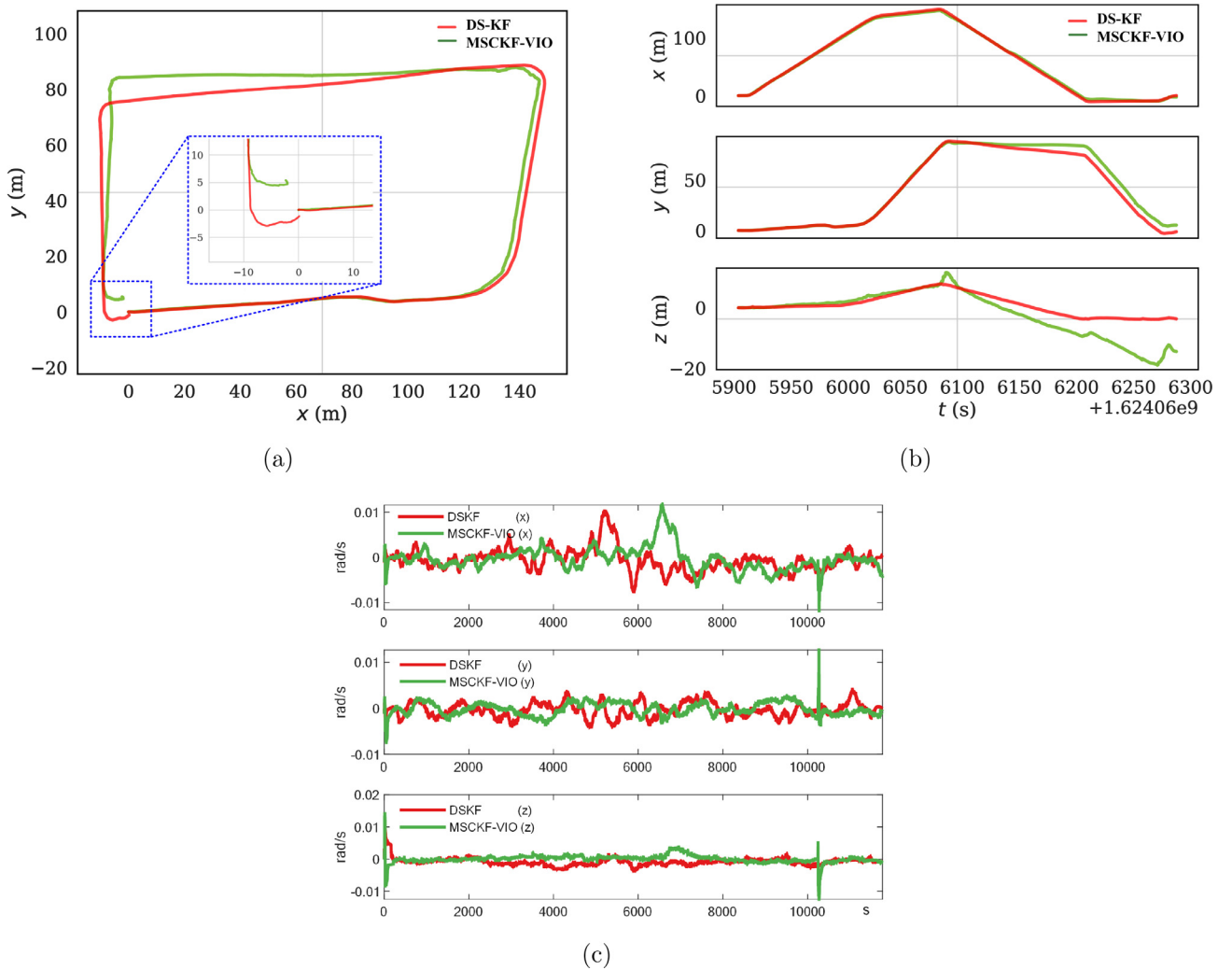
#### 4.2. Performance on low-cost VIO system

We further evaluated the performance of the proposed algorithm on a low-cost VIO hardware system. The image pairs with a resolution of  $1280 \times 720$  are provided by the ZED camera at 20 Hz. Two IMU sensors, Xsens MTI-30 and BMI-160 are used to measure the linear acceleration and angular velocity. Xsens MTI-30 is a medium precision IMU sensor, and BMI-160 is a low precision and low-cost IMU sensor. The camera and IMUs are rigidly connected, as shown in Fig. 6(a). The external parameters between the camera and IMUs are calibrated offline with the Kalibr [37] calibration toolbox.

The steps of the evaluation began with first drawing a circle with a radius of 5 m on the ground and marking it with white flour. Then the mobile vehicle runs around the circle. Fig. 6(b) shows the trajectory of the DS-KF, MSCKF-VIO, and VINS-Mono with the Xsens MTI-30 sensor as the red, green, and blue lines, respectively. The black line is the trajectory of the VINS-Mono with a loop closure, which is set as the ground truth. It can be observed from Fig. 6(b) that all systems can obtain accurate trajectory on IMU with a medium precision, and the trajectory of the DS-KF is closer to the ground truth, which indicates that the use of the accelerometer to dynamically correct the gyroscope bias can improve the accuracy of the initial pose and the VIO system results.



**Fig. 6.** The trajectory of mobile vehicle on the circle, the radius is almost 5 m. (a) The home-made low-cost VIO system with stereo camera, Xsens MTI-30 and BMI-160. (b) The trajectory with Xsens MTI-30. (c) The trajectory with low-cost IMU BMI-160.



**Fig. 7.** The trajectory of mobile vehicle with low-cost IMU BMI-160 on the large scale scene. (a) The trajectories of the DS-KF and MSCKF-VIO. (b) The trajectory in three axes. (c) The gyroscope bias estimated by the DS-KF and MSCKF-VIO in three axes.



Fig. 6 (c) shows the trajectory of the DS-KF and MSCKF-VIO when using a low-cost IMU BMI-160. In this case, VINS-mono cannot successfully align the measurements of visual and inertial sensors, causing the vehicle pose estimation to fail. The main reason is that the BMI-160 sensor has a large gyro noise and unstable bias, which leads to the IMU error factor not being estimated in the initialization step of VINS-Mono. Therefore, the IMU pose of pre-integration and visual SFM is misaligned. However, the DS-KF and MSCKF-VIO can still robustly estimate the vehicle pose in this situation, thanks to the structure of the Kalman filter used in both algorithms. In summary, we can see from Fig. 6(b) and (c) that the DS-KF can still obtain ideal accuracy when using a low-cost IMU compared to MSCKF-VIO.

Fig. 7 further confirms this conclusion. Fig. 7(a) shows the trajectory of the vehicle running along the road in the campus scene when using the BMI-160. The trajectory distance and time are almost 0.48 km and 383 s, respectively. Fig. 7(c) shows the gyroscope bias estimated by the DS-KF and MSCKF-VIO. As can be seen from Fig. 7, the DS-KF using the accelerometer to calibrate the gyroscope bias can obtain a more accurate trajectory. This is because an accurate IMU attitude is computed by the corrected gyroscope measurements, which have a positive effect on the speed and position estimation in Eq. (14). This indicates that our proposed method can improve the accuracy of the algorithm for the low-cost VIO system.

## 5. Conclusion

In this paper, we proposed a double-stage Kalman filter model (DS-KF) for low-cost visual-inertial odometry that consists of a complementary Kalman filter and an MSCKF. The complementary Kalman filter fuses accelerometer measurements to correct the gyroscope bias and calculate an accurate initial pose estimation. The MSCKF uses the re-projection error of features to update the system states. In addition, a Schur complement update model is used for the sliding window to marginalize the old camera poses, preserving the association information between images. In the experiment, the EuRoC dataset and a homemade low-cost VIO hardware system are used to evaluate the performance of the proposed algorithm. The results show that the accuracy of the low-cost gyroscope bias estimation will decrease when the visual observation is inaccurate in the classic VIO, such as in dark environments and motion blur. Our algorithm corrects the low-cost gyroscope bias through accelerometer measurements, which significantly improves the accuracy and robustness of the low-cost VIO system. Additionally, the pose estimation speed of the proposed algorithm is 2 times that of the VINS-Mono, which is beneficial to the size and power constraints of mobile vehicles. This is attributed to the structure of the Kalman filter used in the proposed algorithm. To some extent, this study is an extension of the filtering-based VIO and can be extended to applications regarding autonomous robots.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRediT authorship contribution statement

**Ruping Cen:** Writing – original draft, Conceptualization, Methodology. **Tao Jiang:** Methodology, Software. **Yaoyao Tan:** Formal analysis, Methodology. **Xiaojie Su:** Supervision, Project administration. **Fangzheng Xue:** Writing – review & editing.

## References

- [1] J. Liu, L. Sun, J. Pu, Y. Yan, Hybrid cooperative localization based on robot-sensor networks, *Signal Process.* 188 (2021) 108242.
- [2] R. Munoz-Salinas, H. Sarmadi, D. Cazzato, R. Medina-Carnicer, Flexible body scanning without template models, *Signal Process.* 154 (2019) 350–362.
- [3] J. Engel, T. Schöps, D. Cremers, LSD-SLAM: large-scale direct monocular slam, in: *European Conference on Computer Vision*, 2014, pp. 834–849.
- [4] R. Mur-Artal, J.D. Tardós, Orb-slam2: an open-source slam system for monocular, stereo, and RGB-d cameras, *IEEE Trans. Robot.* 33 (5) (2017) 1255–1262.
- [5] T. Hiraoka, Y. Tsurunari, Generation of maze-like images by bilateral filter and unsharp mask in HLS color space, *ICIC Express Lett.* 15 (10) (2021) 1053–1059.
- [6] T. Hiraoka, Generation of moire-like images smoothing stepwise changes using non-local bilateral filter, *ICIC Express Lett.* 15 (8) (2021) 829–835.
- [7] S. Weiss, R. Siegwart, Real-time metric state estimation for modular vision-inertial systems, in: *IEEE International Conference on Robotics and Automation*, 2011, pp. 4531–4537.
- [8] M. Faessler, F. Fontana, C. Forster, E. Mueggler, M. Pizzoli, D. Scaramuzza, Autonomous, vision-based flight and live dense 3D mapping with a quadrotor micro aerial vehicle, *J. Field Robot.* 33 (4) (2016) 431–450.
- [9] M. Bloesch, S. Omari, M. Hutter, R. Siegwart, Robust visual inertial odometry using a direct EKF-based approach, in: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2015, pp. 298–304.
- [10] T. Qin, P. Li, S. Shen, Vins-mono: a robust and versatile monocular visual-inertial state estimator, *IEEE Trans. Robot.* 34 (4) (2018) 1004–1020.
- [11] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, P. Furgale, Keyframe-based visual-inertial odometry using nonlinear optimization, *Int. J. Robot. Res.* 34 (3) (2015) 314–334.
- [12] R. Mur-Artal, J.D. Tardós, Visual-inertial monocular slam with map reuse, *IEEE Robot. Autom. Lett.* 2 (2) (2017) 796–803.
- [13] C. Forster, L. Carlone, F. Dellaert, D. Scaramuzza, On-manifold preintegration for real-time visual-inertial odometry, *IEEE Trans. Robot.* 33 (1) (2016) 1–21.
- [14] T. Qin, S. Shen, Online temporal calibration for monocular visual-inertial systems, in: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2018, pp. 3662–3669.
- [15] J. Yang, Y. Liu, Z. Li, Unconstrained continuous control set model predictive control based on Kalman filter for active power filter, *Int. J. Innov. Comput., Inf. Control* 17 (5) (2021) 1705–1716.
- [16] A. Kanchanaharuthai, E. Mujjalinvimut, Finite-time command filtered backstepping control design for power systems with superconducting magnetic energy storage system, *Int. J. Innov. Comput., Inf. Control* 17 (3) (2021) 873–885.
- [17] A.I. Mourikis, S.I. Roumeliotis, A multi-state constraint Kalman filter for vision-aided inertial navigation, in: *IEEE International Conference on Robotics and Automation*, 2007, pp. 3565–3572.
- [18] G.P. Huang, A.I. Mourikis, S.I. Roumeliotis, Observability-based rules for designing consistent EKF SLAM estimators, *Int. J. Robot. Res.* 29 (5) (2010) 502–528.
- [19] M. Li, A.I. Mourikis, High-precision, consistent EKF-based visual-inertial odometry, *Int. J. Robot. Res.* 32 (6) (2013) 690–711.
- [20] K. Sun, K. Mohata, B. Pfrommer, M. Watterson, S. Liu, Y. Mulgaonkar, C.J. Taylor, V. Kumar, Robust stereo visual inertial odometry for fast autonomous flight, *IEEE Robot. Autom. Lett.* 3 (2) (2018) 965–972.
- [21] J.A. Hesch, D.G. Kottas, S.L. Bowman, S.I. Roumeliotis, Observability-Constrained Vision-Aided Inertial Navigation, Technical report, University of Minnesota, Department of Computer Science & Engineering, MARS Lab, 2012.
- [22] K. Wu, A. Ahmed, G.A. Georgiou, S.I. Roumeliotis, A square root inverse filter for efficient vision-aided inertial navigation on mobile devices, *Robotics* 2 (2015) 2.
- [23] M. Bloesch, M. Burri, S. Omari, M. Hutter, R. Siegwart, Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback, *Int. J. Robot. Res.* 36 (10) (2017) 1053–1072.
- [24] Y. Zhang, S. Cao, Initial alignment for large azimuth misalignment angle with iterated extended Kalman filter, *ICIC Express Lett., Part B* 11 (11) (2020) 1037–1043.
- [25] P. Geneva, K. Eickenhoff, G. Huang, A linear-complexity EKF for visual-inertial navigation with loop closures, in: *International Conference on Robotics and Automation*, 2019, pp. 3535–3541.
- [26] S. Heo, J. Cha, C.G. Park, EKF-based visual inertial navigation using sliding window nonlinear optimization, *IEEE Trans. Intell. Transp. Syst.* 20 (7) (2018) 2470–2479.
- [27] P. Geneva, K. Eickenhoff, W. Lee, Y. Yang, G. Huang, Openvins: a research platform for visual-inertial estimation, in: *International Conference on Robotics and Automation*, 2020, pp. 4666–4672.
- [28] S. Sabatelli, M. Galgani, L. Fanucci, A. Rocchi, A double-stage Kalman filter for orientation tracking with an integrated processor in 9-d imu, *IEEE Trans. Instrum. Meas.* 62 (3) (2012) 590–598.
- [29] L. Chang, B. Hu, K. Li, Iterated multiplicative extended Kalman filter for attitude estimation using vector observations, *IEEE Trans. Aerosp. Electron. Syst.* 52 (4) (2016) 2053–2060.
- [30] Y. Zhao, Performance evaluation of cubature Kalman filter in a GPS/IMU tightly-coupled navigation system, *Signal Process.* 119 (2016) 67–79.
- [31] S. Baker, I. Matthews, Lucas-Kanade 20 years on: a unifying framework, *Int. J. Comput. Vis.* 56 (3) (2004) 221–255.
- [32] E. Rosten, T. Drummond, Machine learning for high-speed corner detection, in: *European Conference on Computer Vision*, 2006, pp. 430–443.

- [33] S.O. Madgwick, A.J. Harrison, R. Vaidyanathan, Estimation of IMU and MARG orientation using a gradient descent algorithm, in: IEEE International Conference on Rehabilitation Robotics, 2011, pp. 1–7.
- [34] K. Konolige, M. Agrawal, FrameSLAM: from bundle adjustment to real-time visual mapping, IEEE Trans. Robot. 24 (5) (2008) 1066–1077.
- [35] J. Xiao, D. Xiong, Q. Yu, K. Huang, H. Lu, Z. Zeng, A real-time sliding window based visual inertial odometry for MAVs, IEEE Trans. Ind. Inf. 16 (6) (2019) 4049–4058.
- [36] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M.W. Achtelik, R. Siegwart, The EuRoC micro aerial vehicle datasets, Int. J. Robot. Res. 35 (10) (2016) 1157–1163.
- [37] J. Rehder, J. Nikolic, T. Schneider, T. Hinzmann, R. Siegwart, Extending kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes, in: IEEE International Conference on Robotics and Automation, 2016, pp. 4304–4311.