Accepted Manuscript

Compass Aided Visual-Inertial Odometry

Yandong Wang, Tao Zhang, Yuanchao Wang, Jingwei Ma, Yanhui Li,
Jingzhuang Han

PII: S1047-3203(18)30355-9
DOI: https://doi.org/10.1016/j.jvcir.2018.12.029
Reference: YJVCI 2394

To appear in: *J. Vis. Commun. Image R.*

Received Date: 15 November 2018
Revised Date: 13 December 2018
Accepted Date: 14 December 2018

Please cite this article as: Y. Wang, T. Zhang, Y. Wang, J. Ma, Y. Li, J. Han, Compass Aided Visual-Inertial Odometry, *J. Vis. Commun. Image R.* (2018), doi: https://doi.org/10.1016/j.jvcir.2018.12.029

Compass Aided Visual-Inertial Odometry

Yandong Wang[1,2],Tao Zhang[1], Yuanchao Wang[1,2*] , Jingwei Ma[1] Yanhui Li[1], Jingzhuang Han[1],

1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
2. University of Chinese Academy of Sciences, Beijing 100049, China

There is no conflict of interest.

Abstract: With the development of vision and optimization techniques, visual-inertial odometry (VIO) has shown the capability of motion estimating in the GNSS-denied condition. The VIO can provide absolute pitch and roll angles estimating value, but no the absolute azimuth. In the paper, we proposed a VIO aided by compass, which can obtain the azimuth with respect to the north direction in the geographic frame. Moreover, aided by compass, the yaw angle estimating error was reduced to a greater degree, due to the measurement of azimuth. Furthermore, the consistency of the VIO backend estimator is improved as well, while the accuracy of the estimated pose states was also wholly improved. The aiding approach is a tightly-couple information fusion system of camera, IMU and magnetoresistive sensors. The optimization method is based on the pre-integration and bundle adjustment. In the paper, we derived the compass residual model based on the pre-integration model, and then its Jacobian and covariance formation were deduced to solve the nonlinear equations. The compass aided VIO software was implemented based on the Nvidia Jetson Tx2. The system was fully tested based on hardware-in-the-loop simulation and vehicle test in the real physical environment. The pose errors of VIOs with and without compass aiding were compared in the above tests. The simulation results showed that the position was and yaw errors were improved obviously; the compass aided VIO was still consistent, but the pure VIO was consistent not. The consistency character is evaluated by average NEES by Monte-Carlo in simulation. The vehicle test showed that the position error was reduced by 23%; the yaw error was reduced by 21%. As a result, the compass aided VIO not only improved the pose estimated accuracy, especially position and yaw, but also improved the consistency of VIO system.

Keywords: visual-inertial odometry (VIO), compass, sliding window estimator, inconsistency, pre-integration, minimum cost function

I. Introduction

Visual odometry (VO) uses the feature matching information between successive image sequences to estimate the position and orientation increment value in real time, which has been gradually applied to the robot navigation system in the condition of GNSS-denied environment [1]. However, the performance of VO is dependent on the illumination and texture of the scene. Strapdown Inertial Navigation System (SINS) has been used as a standard configuration for any navigation system [2] [3], which is the base system for general applications, such as plane, UAV, vehicle marine and smart phone. The adaptability of SINS is better than other navigation systems due to ego-motion measured by IMU. However, the main error source of SINS is the IMU which is divergent with time. For example, the position error is proportional to cubic of travel time and the scale factor is the gyro bias. As a result, SINS with low precision grade IMU cannot be used independently, it must be integrated with other sensors, such as GNSS receiver and VO. The accuracy of VO mainly depends on the resolution of the camera. It can be demonstrated that camera with millions of pixels is enough providing higher accuracy than SINS with consumer degrade IMU. Therefore, the fusion of VO and INS system is able to improve the robustness and precision of navigation system, which can be applied to motion estimation in GPS-denied conditions, such as indoor environments and urban canyons.

The motion estimation system which is integrated of VO and INS is called visual-inertial odometry[44-46]. Generally, it is divided into three parts according to its function: VO frontend, SINS calculating and VIO backend. The VO frontend is used for image feature extraction and matching as well as pose calculation. In accordance with feature extraction and matching, VO is usually divided into two types—direct method [4] [5] and indirect method [6] [7]. The pose calculation method is classified according to monocular vision and binocular vision. The main methods are: Nister's 5-point method for monocular vision, Iterative Closest Point (ICP) for binocular vision and Perspective-n-Points (PnP) for the both types [8]. The SINS calculate the position, velocity and orientation of the vehicle with respect to the navigation reference coordinate system based on the specific force and angular velocity measured by the IMU [9]. The VIO backend is used for motion state and sensor error estimating[47,48]. The main methods are extend Kalman filtering [10] , sliding window smoothing estimator [7] and recursive global smoothing estimator [12] , etc.

The system integrated IMU and magnetic compass often called Dead-Reckoning system, which has been widely used in motion state estimation in GPS-denied conditions [11] , overcoming the error divergence problem of pure inertial navigation systems [13] [14] . The inertial navigation system is able to provide global measurements of the pitch and roll angles referring to the gravitation vector, while the yaw angle measurement is local. However, compass can provide the absolute azimuth with reference to the geomagnetic vector. Therefore, the Dead-Reckoning system make the yaw direction of SINS global observable. There are many benefits by the fusion of the compass and VIO. Firstly, it provides azimuth of earth frame instead of yaw angle of ground frame. Secondly, where the environment is low illumination, texture duplication or motion blur, compass aided VIO system will still work. Lastly, the global observability of the backend estimator is changed from 2 to 3 by aiding. The yaw direction is observable, which can overcome inconsistency of the estimator to avoid the system degenerating to suboptimal.

The main contributions in this paper are summarize: 1) The system scheme and computing framework of compass aided VIO aided are presented; 2) The minimum cost function of the compass aided VIO is deduced based on the pre-integration theory, while its Jacobian and covariance iteration formation are derived as well. 3) The hardware-in-the-loop simulation platform is implemented based on Airsim for comparing the performance of compass aided VIO and the classic VIO.

II. Notation and Definitions

1. Coordinate frame

The compass aided VIO motion estimation system contains four kinds of sensors, which can be classified into external information sensors and ego-motion sensors according to the sources of measured information. The external information sensors are cameras and geomagnetic sensors, which are used to acquire motion scene images and measure geomagnetic field intensity respectively, and the ego-motion sensor is IMU, comprised of gyroscope and accelerometer, which are used to measure the vehicle angular velocity and specific force respectively. The estimation information of system is the rotation and translational incremental value and sensor errors. Therefore, the coordinate systems must be unified.

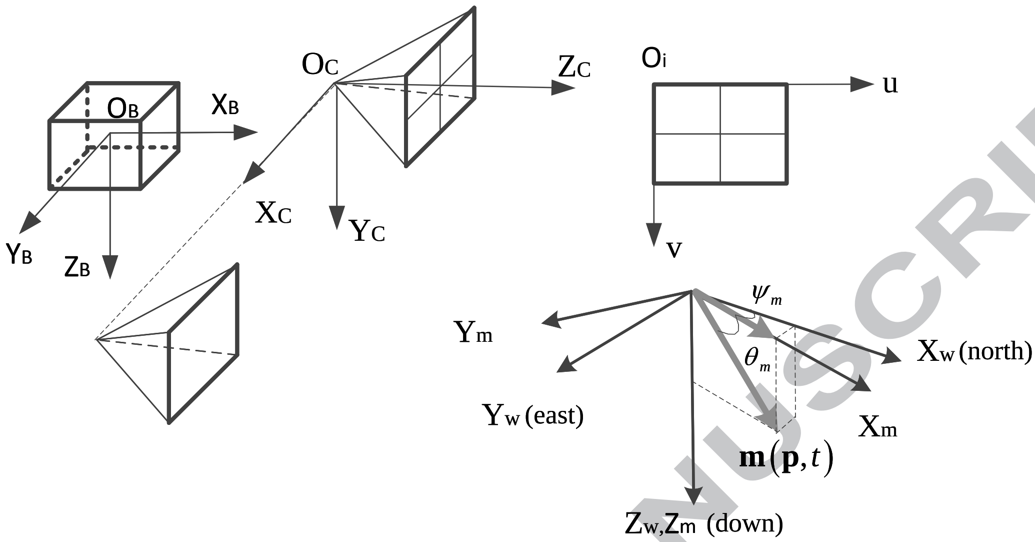


Fig.1 Coordinate frames involved in the compass aided VIO system

Generally, compass aided VIO system mainly involves five kinds of coordinate frames shown in Fig.1: body frame (B frame), camera frame (C frame), image frame (I frame), world frame (W frame) and geomagnetic frame (M frame). World frame is NED frame in this paper, three axis point to the north, east and downward direction, respectively. The Z-axis of geomagnetic frame is the same to the one of world frame. Geomagnetic field intensity vector is depicted in world frame in terms of the declination angle ψ_m and inclination angle θ_m. IMU frame is coincided to body frame, which is commonly defined in most vehicles. The camera is fixed on the vehicle, which is described by camera frame, the principle axis of camera is along to the $O_B X_B$ axis defined by $O_C Z_C$; $O_C X_C$ is the same to $O_B Y_B$. The image frame is used to indicate the feature location in pixel. The origin is at the left-up corner. The axis is along to the two sides.

2. Translation and Rotation

In this paper, the rotational motion is presented in manifold structure of SO(3), and the translational motion is presented in manifold structure of SE(3). It is because that motion computed by manifold structure can avoid the "Gimbal lock" phenomenon [17]. As the compass is applied to the backend optimization, and the azimuth is defined in the Euclidean space, so the Euler angle is still used. The azimuth involves only one degree of freedom, because it is computed after aligning the roll and pitch orientation, so it won't lead to the problem of "Gimbal lock". The definition of the Special Orthogonal Group is $\mathrm{SO}(3) \doteq \left\{ \mathbf{R} \in \mathbb{R}^3 : \mathbf{R}^T\mathbf{R} = \mathbf{I}, \det(\mathbf{R}) = 1 \right\}$ .The tangent space of the manifold is called the lie algebra, which is denoted by $\mathfrak{so}(3)$ .The conversion between the $\mathfrak{so}(3)$ and $\mathrm{SO}(3)$ is Rodrigues formula, which is used exponential and logarithm to realize, as shown in (1) and (2). In (2), $\mathbf{a}$ and φ denote rotation axis and rotation angle, $\log(R)^{\vee} = \mathbf{a}\varphi$. The symbol " $\wedge$ " and " $\vee$ ", which are used to denote the conversion between vector and its skew symmetric matrix, as shown in (3).

$$\exp\left(\phi^{\wedge}\right) = \mathbf{I} + \frac{\sin\left(\|\phi\|\right)}{\|\phi\|}\phi^{\wedge} + \frac{1-\cos\left(\|\phi\|\right)}{\|\phi\|^2}\left(\phi^{\wedge}\right)^2 \tag{1}$$

$$\log\left(R\right) = \frac{\varphi \cdot \left(\mathbf{R} - \mathbf{R}^T\right)}{2\sin\left(\varphi\right)}, \varphi = \cos^{-1}\left(\frac{tr\left(\mathbf{R}\right)-1}{2}\right) \tag{2}$$

$$\mathbf{b}^{\wedge} = \begin{bmatrix} 0 & -b_3 & b_2 \\ b_3 & 0 & -b_1 \\ -b_2 & b_1 & 0 \end{bmatrix} = \mathbf{B}, \mathbf{B}^{\vee} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \mathbf{b} \tag{3}$$

In Euclidean space, $\mathbf{R}$ is the orientation matrix, $\mathbf{R}_{WB}$ denotes the rotation of B frame about W frame, shown in (4).

$$\mathbf{R}_{WB} = \begin{bmatrix} \cos\theta\cos\psi & -\cos\phi\sin\psi + \sin\phi\sin\theta\cos\psi & \sin\phi\sin\theta + \cos\phi\sin\theta\cos\psi \\ \cos\theta\sin\psi & \cos\phi\cos\psi + \sin\phi\sin\theta\sin\psi & -\sin\phi\sin\theta + \cos\phi\sin\theta\sin\psi \\ -\sin\theta & \sin\phi\cos\theta & \cos\phi\cos\theta \end{bmatrix} \tag{4}$$

In this paper, 6 degrees of freedom rigid body dynamics is depicted by Special Euclidean Group SE(3), its definition is $\mathrm{SE}(3) \in \left\{ \left(\mathbf{R}, \mathbf{p}\right) : \mathbf{R} \in \mathrm{SO}(3), \mathbf{p} \in \mathbb{R}^3 \right\}$.

III. SINS and Magnetic Compass

1. SINS and IMU

In the SINS, Inertial Measurement Unit (IMU) is used to measure the angular velocity and the specific force of the vehicle respectively, and calculates position, velocity and orientation of vehicle. IMU is consisted of accelerometer and gyroscope. Accelerometer measures specific force $\mathbf{a}_f$, and the definition of specific force is the resultant acceleration except gravitational acceleration. In NED frame, accelerometer measures the Coriolis acceleration adapted from earth rotation, $\boldsymbol{\omega}_e \times \mathbf{v}$. However, IMU the paper adopted is consumer grade, so Coriolis acceleration can be neglected comparing the high noise level. In W frame, SINS basic equation is (5) and (6). For expression convenience, we neglected subscript of $\mathbf{a}_f$ in (5), denoting $\mathbf{a}$.

$$_W\dot{\mathbf{v}} = \mathbf{R}_{WB}\mathbf{a}_B + g_W \tag{5}$$

$$_W\dot{\mathbf{p}} = {}_W\mathbf{v} \tag{6}$$

The gyroscope measures the angular velocity relative to the inertial system. However, the gyroscope adopted is a MEMS gyro, of which the noise level is high and the cost is low, so the influence of the acceleration caused by transfer velocity and earth rotation velocity are ignored. The differential equation for solving orientation is (7). In navigation computing scheme, SINS is calculated by recursive form, which is derived in SO(3) seen in (8). The pipeline of SINS calculating can be seen in Fig.2.

$$\dot{\mathbf{R}}_{WB} = \mathbf{R}_{WB}\left[{}_B\boldsymbol{\omega}_{WB}\right]^{\wedge} \tag{7}$$

$$\begin{cases} \mathbf{R}_{WB}\left(t+\Delta t\right) = \mathbf{R}_{WB}\left(t\right)Exp\left({}_B\boldsymbol{\omega}_{WB}\Delta t\right) \\ _W\mathbf{v}\left(t+\Delta t\right) = \mathbf{v}_{WB}\left(t\right) + \mathbf{g}_W\Delta t + \mathbf{R}_{WB}\left(t\right)\mathbf{a}_B\Delta t \\ _W\mathbf{p}\left(t+\Delta t\right) = {}_W\mathbf{p}\left(t\right) + {}_W\mathbf{v}\left(t\right)\Delta t + \frac{1}{2}\mathbf{g}_W\Delta t^2 + \frac{1}{2}\mathbf{R}_{WB}\left(t\right)\mathbf{a}_B\Delta t^2 \end{cases} \tag{8}$$

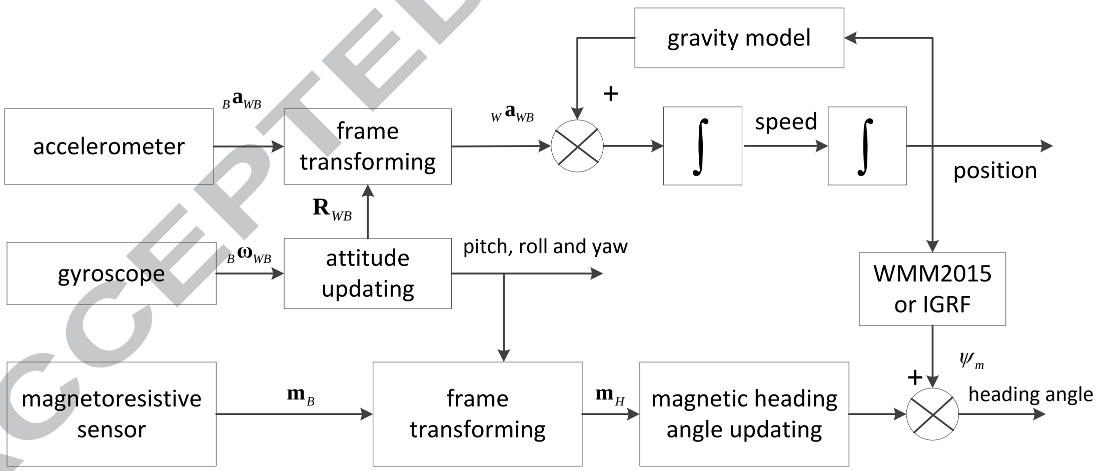


Fig.2 Pipeline of SINS and Compass calculation

There are several errors in gyro and accelerometer measurement value, such as bias, proportional factor error, cross coupling error and white noise, etc. Most of errors can be compensated by the calibration. However, bias is changed every time when IMU is powered on, and it affects motion estimation accuracy distinctly. The nonlinearity error and the proportion factor error are poorly observable in the estimator, so they are not estimated in the VIO backend. The model of IMU is (9). In (9), $\tilde{\mathbf{a}}_B$ and $_B\tilde{\boldsymbol{\omega}}_{WB}$ denote the measurement of specific force and gyro respectively. The vector $\mathbf{b}_a$ and $\mathbf{b}_g$ are the bias of accelerometer and gyro, $\boldsymbol{\eta}_{ad}$ and $\boldsymbol{\eta}_{gd}$ are white noise process.

$$\begin{cases} \tilde{\mathbf{a}}_B = \mathbf{a}_B + \mathbf{b}_a + \boldsymbol{\eta}_{ad} \\ {}_B\tilde{\boldsymbol{\omega}}_{WB} = {}_B\boldsymbol{\omega}_{WB} + \mathbf{b}_g + \boldsymbol{\eta}_{gd} \end{cases} \tag{9}$$

2. Magnetic Compass and Geomagnetic Sensors

In the interior of the earth, the direction of geomagnetic field is from the geomagnetic North Pole to the geomagnetic South Pole, which is opposite exteriorly. The intensity of geomagnetic field near the equator is about 30nT, and that near the geomagnetic pole is about 60nT. The magnetic field intensity varies with geographical position. The variation of the geomagnetic field with time is very slow. The magnetic field intensity in most regions of the earth can be calculated according to World Magnetic Model (WMM) [18] or International Geomagnetic Reference Field (IGRF) [20], as well as the bearing angle of the direction of the magnetic field in the earth frame. In Fig.1, the geomagnetic field intensity is denoted by $\mathbf{m}(\mathbf{p},t)$ and can be calculated according to the WMM model or the IGRF model. θ_m and ψ_m are called inclination and declination respectively. These three physical parameters define the magnitude and direction of the geomagnetic field intensity in the NED system, the formula is shown in (10). In short distance motion, the geomagnetic can be considered invariant.

$$\mathbf{m}_w = \begin{bmatrix} \cos\psi_{wm}\cos\theta_{wm} \\ \sin\psi_{wm}\cos\theta_{wm} \\ \sin\theta_{wm} \end{bmatrix} m(p,t) \tag{10}$$

The intensity of the local geomagnetic field can be measured by the geomagnetic sensor, and the heading angle can be calculated [21]. The geomagnetic from world frame to body frame is shown in (10). A transitional coordinate frame is defined called horizontal frame, and the heading angle can be calculated referred to it, denoted by H frame in (12). The rotation of H frame with respect to H frame is about pitch and roll angle. The pitch and roll angle can be obtained by SINS. Furthermore, the heading angle is determined, shown in (13) and (14). In (14), ψ_m is declination, and it is obtained by WMM or IGRF model. The pipeline of heading angle calculation can be seen in Fig.2.

$$\mathbf{m}_B = \mathbf{R}_{BW}\mathbf{m}_W \tag{11}$$

$$\mathbf{m}_H = \begin{bmatrix} \cos\theta & \sin\phi\sin\theta & \cos\phi\sin\theta \\ 0 & -\cos\phi & \sin\phi \\ -\sin\theta & \sin\phi\cos\theta & \cos\phi\cos\theta \end{bmatrix}\mathbf{m}_B \tag{12}$$

$$\psi_{mb} = \arctan\frac{-m_{H_y}}{m_{H_x}} = \arctan\frac{m_{B_y}\cos\phi - m_{B_z}\sin\phi}{m_{B_x}\cos\theta + m_{B_y}\sin\phi\sin\theta + m_{B_z}\cos\phi\sin\theta} \tag{13}$$

$$\psi_{wb} = \psi_{mb} + \psi_m \tag{14}$$

IV. The Front-End of Visual Inertial Odometry

The visual inertial odometry front-end is derived from visual odometry. The main functions are these: image features extraction and matching, land marks sifting, calculation of pose increments, and triangulation, shown in Fig.3. The camera sensor can be either monocular or binocular.

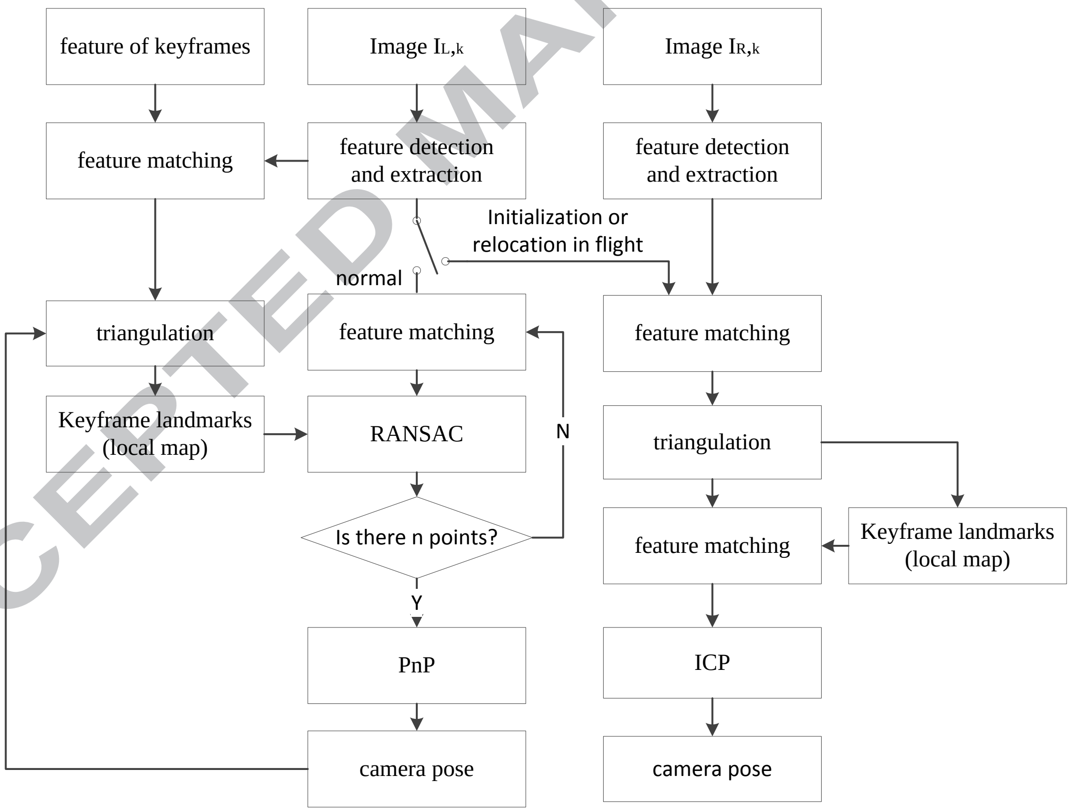


Fig.3 The pipeline of VO Frontend

The algorithm of VIO feature extraction and matching is expected to be of high accuracy, robustness, and repeatability. With the development of image processing and computer vision technology, Harris [23], FAST [24], SIFT [25] and SURF [26] algorithms have been developed for VO front-end. Because of motion estimation, feature detection algorithms should have well geometric properties, such as rotation invariance, scale invariance and affine invariance, while real-time is also an unavoidable constraint [27] characteristic in engineering. Based on the above requirements of motion estimation for feature extraction and matching, the Oriented FAST and Rotated BRIEF (ORB) algorithm was selected in this paper. The ORB feature improves the problem that original FAST detection algorithm which cannot be oriented and uses the binary descriptor BRIEF to make the image feature extraction greatly accelerated. Compared with SIFT algorithm which is high precision and strong robustness, and FAST algorithm which has better real-time performance, ORB algorithm is a better compromise between those performance. It has been well applied in SLAM technology [28]. ORB descriptor is binary, so we can achieve feature matching based on Hamming distance by FLANN method [30]. Therefore, the real-time performance of matching is guaranteed.

Feature mismatching often happens due to noise, occlusion, blur, and illumination changing, which will cause the error of pose increased obviously, even divergence. To avoid the fault, RANSAC has been used as a standard algorithm for visual navigation to reject outliers [31].

The mathematical basis for the pose estimation in terms of visual information is the epipolar geometry [31]. The method is based on the coordinates of the matching features. f_{k-1} and f_k are matched feature sets of two successive images I_{k-1} and I_k. The method can be classified based on monocular or binocular of f_{k-1} and f_k, shown in Tab.1. In our project, binocular vision is implemented, so we use 3D-to-2D method for general pose incremental calculation, and we applied 3D-to-3D method to initialize and remap when the carrier is in motion, as shown in Fig.3.

Tab.1 Compare the algorithms of pose incremental calculation

Method	f_{k-1} feature frame	f_k feature frame	Algorithm	Mono/Stereo VO
2D-to-2D	image frame	image frame	Nister's 5-points [32]	Mono VO
3D-to-2D	world frame	image frame	PnP [33]	Both
3D-to-3D	world frame	world frame	ICP [34]	Stereo VO

For 3D-to-2D and 3D-to-3D methods, it is necessary to triangulate the 2D correspondences as the landmarks. Triangulating 3-D points are determined by intersecting back-projected rays from 2D image correspondences of at least two frame images or the binocular images. It is not as much landmarks as possible. Too much landmarks not only make computation complex, but also not all of them work [35]. For stereo VO, landmarks distance more than 40 times baseline length that cannot provide enough information for translational calculation, it is only appropriate for rotational calculation. It is because that if the landmarks distance is too far, the binocular vision would degenerate to monocular vision. As a result, the landmarks which are less than the 40 times baseline length was adopted in the paper [27].

V. The Backend of Visual Inertial Odometry

The backend of visual Inertial Odometry is used to optimize the motion state and IMU error. In our project, according to source of the measurement value for objective function, estimator is divided to three parts: bundle adjustment, IMU pre-integration and the variation in yaw direction.

1. State equation of estimation

The referenced information for optimization is obtained from the measurements of world frame, such as gravity, geomagnetic intensity and the feature of environment. For the compass aided VIO, the state should be divided into two parts, $\mathbf{x}_V$ and $\mathbf{x}_L$ for simplification. $\mathbf{x}_V$ is consisted of position, velocity, orientation, gyro bias, accelerometer bias and magnetoresistive sensor bias [38]. It can be further divided into two parts: INS related part $\mathbf{x}_V^I$ and compass related part $\mathbf{x}_V^M$. The state error $\delta\mathbf{x}$ is chosen as state variables, because it will make the state equation linear. If the $\delta\mathbf{x}$ is estimated, it is esteemed the compensation to correct the state updated by VO, INS and compass. The symbol $\mathbf{x}_L$ denotes the landmarks position in the world frame. The state variables are (15) and (16).

$$\delta\mathbf{x} = \begin{bmatrix} \delta\mathbf{x}_V & \delta\mathbf{x}_L \end{bmatrix}^T \tag{15}$$

$$\begin{cases} \delta\mathbf{x}_V = \begin{bmatrix} \delta\mathbf{x}_V^I & \delta\mathbf{x}_V^M \end{bmatrix}^T \\ \delta\mathbf{x}_V^I = \begin{bmatrix} \delta\mathbf{\Phi}_{WB} & {}_W\delta\mathbf{p} & {}_W\delta\mathbf{v} & \mathbf{b}^g & \mathbf{b}^a \end{bmatrix}^T \\ \delta\mathbf{x}_V^M = \begin{bmatrix} \delta\psi_w & \mathbf{b}^m \end{bmatrix}^T \end{cases} \tag{16}$$

State transition matrix $\mathbf{F}_{k+1,k}$ represents the propagation relationship of state variables in discrete time. In estimating process, it is used to update covariance matrix. According to the state variables, the matrix is divided into two parts, INS part and compass part which are not correlated, as shown in (17)~(19). In (19), The Jacobian $\mathbf{J}_{\psi_w}^{\mathbf{b}_m}$ is the first-order partial derivatives of ψ_w to $\mathbf{b}_m$.

$$\mathbf{F}_{k+1,k} = \begin{bmatrix} \mathbf{F}_{k+1,k}^{I} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{F}_{k+1,k}^{M} \end{bmatrix} \tag{17}$$

$$\mathbf{F}_{k+1,k}^{I} = \begin{bmatrix} \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & -\mathbf{R}_k\Delta t & \mathbf{0}_{3\times3} \\ -\frac{1}{2}\mathbf{R}_k\mathbf{a}_k^{\wedge}\Delta t^2 & \mathbf{I}_{3\times3} & \mathbf{I}_{3\times3}\Delta t & \mathbf{0}_{3\times3} & \frac{1}{2}\mathbf{R}_k\Delta t^2 \\ -\mathbf{R}_k\mathbf{a}_k^{\wedge}\Delta t & \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{R}_k\Delta t \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} \end{bmatrix} \tag{18}$$

$$\mathbf{F}_{k+1,k}^{M} = \begin{bmatrix} 1 & \mathbf{J}_{\psi_w}^{\mathbf{b}_m} \\ \mathbf{0}_{3\times1} & \mathbf{I}_{3\times1} \end{bmatrix} \tag{19}$$

In (13), $\mathbf{m}_B$ is the true of value of magnetoresistive sensors ideal measurements. Practically, we can only get measurements $\tilde{\mathbf{m}}_B$, so the true value is considered as $\mathbf{m}_B = \tilde{\mathbf{m}}_B - \mathbf{b}_m$. According the chain rule of derivation (20), $\mathbf{J}_{\psi_w}^{\mathbf{b}_m}$ can be obtained from (13), the parameters of partial derivation as shown in (21) and (22).

$$\mathbf{J}_{\psi_w}^{\mathbf{b}_m} = \frac{\partial \psi}{\partial \mathbf{b}^m} = \frac{\partial \psi}{\partial m_{H_x}}\frac{\partial m_{H_x}}{\partial \mathbf{b}^m} + \frac{\partial \psi}{\partial m_{H_y}}\frac{\partial m_{H_y}}{\partial \mathbf{b}^m} \tag{20}$$

$$\begin{cases} \dfrac{\partial \psi}{\partial m_{H_x}} = \dfrac{m_{H_y}}{\left(m_{H_x}\right)^2 + \left(m_{H_y}\right)^2} \\ \dfrac{\partial \psi}{\partial m_{H_y}} = \dfrac{-m_{H_x}}{\left(m_{H_x}\right)^2 + \left(m_{H_y}\right)^2} \end{cases} \tag{21}$$

$$\begin{cases} \dfrac{\partial m_{H_x}}{\partial \mathbf{b}^m} = \begin{bmatrix} \dfrac{\partial m_{H_x}}{\partial b_x^m} & \dfrac{\partial m_{H_x}}{\partial b_y^m} & \dfrac{\partial m_{H_x}}{\partial b_z^m} \end{bmatrix} \\ \qquad = -\left[\cos\theta \quad \sin\phi\sin\theta \quad \cos\phi\sin\theta\right] \\ \dfrac{\partial m_{H_y}}{\partial \mathbf{b}^m} = \begin{bmatrix} \dfrac{\partial m_{H_y}}{\partial b_x^m} & \dfrac{\partial m_{H_y}}{\partial b_y^m} & \dfrac{\partial m_{H_y}}{\partial b_z^m} \end{bmatrix} \\ \qquad = -\left[0 \quad \cos\phi \quad \sin\phi\right] \end{cases} \tag{22}$$

2. Bundle Adjustment

The unique function of camera is that mapping the 3D world to 2D image. However, there are lots of errors, such as noise and distortion of lens, make the image not correspond to the ideal mapping. Bundle Adjustment (BA) is the process that "artificially" adjusts rays to focus on camera center to overcome the errors. The method of adjustment is Minimum Least Estimation (MLE) theory which minimizes the reprojection error of landmarks. It is assumed that there is a landmark $\mathbf{X}_w$ in W frame and its measurement position in the image frame is $\mathbf{x}=\begin{pmatrix}u & v\end{pmatrix}^T$. However, the projection of $\mathbf{X}_w$ by the extrinsic and intrinsic matrix is $\mathbf{x}^{'}=(u^{'},v^{'})^T$. The residual is expected to be eliminated by BA, shown in (26). The projection method is shown in (23) or (24). The subscript "m" in (23) denotes monocular camera and the subscript "s" in (24) denotes binocular camera. In (23) and (24), the X_C is the mapping of $\mathbf{X}_w$ in the camera frame by the extrinsic matrix, which is constructed of current pose. And $c=(c_u,c_v)^T$ is center camera in image frame; $f=(f_x,f_y)^T$ is equivalent focus length. Both parameters construct the intrinsic matrix, which can be determined by camera calibration.

$$\mathbf{x}_m^{'}=\left(f_x\cdot X_C/Z_C+c_x,\ f_y\cdot X_C/Z_C+c_y\right)^T \tag{23}$$

$$\mathbf{x}_s^{'}=\left(f_x\cdot X_C/Z_C+c_x,\ f_y\cdot X_C/Z_C+c_y,\ f_x\cdot\left(X_C-b\right)/Z_C+c_x\right)^T \tag{24}$$

The objective function of BA is established according to the reprojection error, and then the nonlinear optimization method [39] is used to optimize the parameters of the extrinsic matrix, shown in (27). In (27), ρ is kernel function to guarantee the robustness of optimization, Σ_k is the information matrix, **K** is the total number of the chosen landmarks.

$$\mathbf{X}_C=\mathbf{R}_{CB}\mathbf{R}_{BW}\left(\mathbf{X}_w-{}_{w}\mathbf{p}_B\right)+{}_{C}\mathbf{p}_B \tag{25}$$

$$\mathbf{e}_{proj}=\mathbf{x}-\mathbf{x}^{'} \tag{26}$$

$$E_{proj}\left(k\right)=\sum_{k}^{K}\rho\left(\mathbf{x}_k-\mathbf{x}_k^{'}\right)^T\Sigma_k\left(\mathbf{x}_k-\mathbf{x}_k^{'}\right) \tag{27}$$

3. IMU pre-integration and SINS objective function

The cost function of INS term is constructed by the residual of the actual pose increment and IMU pre-integration in the same optimization period. IMU pre-integration base on the INS principle and it is used to synchronize IMU and camera acquisition period. The period of camera acquisition, VO pose calculation and optimization are the same in the paper. If the optimization period is $\left[t_i,t_j\right)$ and IMU sampling time is denoted by k, the pre-integration of IMU is (28). For convenience, we neglect the subscript like (8), $\mathbf{R}$ substituting for $\mathbf{R}_{WB}$, $\mathbf{p}$ substituting for ${}_{W}\mathbf{p}$, $\mathbf{v}$ substituting for ${}_{W}\mathbf{v}$ in (28).

$$\begin{cases} \Delta\tilde{\mathbf{R}}_{ij} = \prod_{k=i}^{j-1} Exp\left(\left(\tilde{\boldsymbol{\omega}}_k - \mathbf{b}_i^g\right)\Delta t\right) \\ \Delta\tilde{\mathbf{p}}_{ij} = \sum_{k=i}^{j-1}\left(\tilde{\mathbf{v}}_k \Delta t + \frac{1}{2}\Delta\tilde{\mathbf{R}}_{ik}\left(\tilde{\mathbf{a}}_k - \mathbf{b}_i^a\right)\Delta t^2\right) \\ \Delta\tilde{\mathbf{v}}_{ij} = \sum_{k=i}^{j-1}\Delta\tilde{\mathbf{R}}_{ik}\left(\tilde{\mathbf{a}}_k - \mathbf{b}_i^a\right)\Delta t \end{cases} \tag{28}$$

IMU model was shown in (9), bias error of IMU is considered varying slowly with time. The IMU bias is considered being invariant, in an optimization period of $\left[t_i, t_j\right)$, so it is existing that $\bar{\mathbf{b}}_{i_0} = \bar{\mathbf{b}}_{i_k} = \cdots = \bar{\mathbf{b}}_{j-1}$, k is the index of IMU sampling value. The $\delta\mathbf{b}_i$ is the variant value between t_i and t_j. So, it is existing an equation that $\mathbf{b}_j = \bar{\mathbf{b}}_i + \delta\mathbf{b}_i$, and the IMU pre-integration form includes the bias variation. Furthermore, the SINS solution considering the bias variation is shown in (29).

$$\begin{cases} \Delta\tilde{\mathbf{R}}_{ij} = \Delta\tilde{\mathbf{R}}_{ij}\left(\bar{\mathbf{b}}_i^g\right) Exp\left(\mathrm{J}_{\Delta R}^{b_g}\,\delta\mathbf{b}_i^g\right) \\ \Delta\tilde{\mathbf{v}}_{ij} = \Delta\tilde{\mathbf{v}}_k\left(\bar{\mathbf{b}}_i^g, \bar{\mathbf{b}}_i^a\right) + \mathbf{J}_{\Delta v}^{b^g}\delta\mathbf{b}_i^g + \mathbf{J}_{\Delta v}^{b^a}\delta\mathbf{b}_i^a \\ \Delta\tilde{\mathbf{p}}_{ij} = \Delta\tilde{\mathbf{p}}_k\left(\bar{\mathbf{b}}_i^g, \bar{\mathbf{b}}_i^a\right) + \mathbf{J}_{\Delta p}^{b^g}\delta\mathbf{b}_i^g + \mathbf{J}_{\Delta p}^{b^a}\delta\mathbf{b}_i^a \end{cases} \tag{29}$$

In the period of optimization $\left[t_i, t_j\right)$, the incremental position and orientation can be derived from visual odometry front-end, while incremental velocity is derived from SINS, as shown in (30). The residual of SINS related is (31), in which the first three equations are the residual of position, orientation and velocity, and the last one is the IMU bias residual. Above all, the objective function of IMU related in VIO backend is (32). The Jacobians of the objective function are derivation in detail in Christian Forster's paper [40].

$$\begin{cases} \Delta\tilde{\mathbf{R}}_{t_{ij}} = \mathbf{R}_i^T\mathbf{R}_j \\ \Delta\tilde{\mathbf{v}}_{t_{ij}} = \mathbf{R}_i^T\left(\mathbf{v}_j - \mathbf{v}_i - g\Delta t_{ij}\right) \\ \Delta\tilde{\mathbf{p}}_{t_{ij}} = \mathbf{R}_i^T\left(\mathbf{p}_j - \mathbf{p}_i - \mathbf{v}_i\Delta t_{ij} - \frac{1}{2}g\Delta t_{ij}^2\right) \end{cases} \tag{30}$$

$$
\begin{cases}
\mathbf{e}_{\Phi} = Log\left(\left(\Delta\mathbf{R}_{ij} Exp(\mathrm{J}_{\Delta R}^{b_g}\mathrm{b}_g^j)\right)^T \mathbf{R}_i^T \mathbf{R}_j\right) \\
\mathbf{e}_p = \mathbf{R}_i^T\left(\mathbf{p}_j - \mathbf{p}_i - \mathbf{v}_i\Delta t_{ij} - \frac{1}{2}g\Delta t_{ij}^2\right) \\
\qquad -\left(\sum_{k=i}^{j-1}\left(\tilde{\mathbf{v}}_k\Delta t + \frac{1}{2}\Delta\tilde{\mathbf{R}}_{ik}\left(\tilde{\mathbf{a}}_k - \overline{\mathbf{b}}_i^a\right)\Delta t^2\right) + \mathbf{J}_{\Delta p}^{b^g}\delta\mathbf{b}_i^g + \mathbf{J}_{\Delta p}^{b^a}\delta\mathbf{b}_i^a\right) \\
\mathbf{e}_v = \mathbf{R}_i^T\left(\mathbf{v}_j - \mathbf{v}_i - g\Delta t_{ij}\right) - \left(\sum_{k=i}^{j-1}\Delta\tilde{\mathbf{R}}_{ik}\left(\tilde{\mathbf{a}}_k - \overline{\mathbf{b}}_i^a\right)\Delta t + \mathbf{J}_{\Delta v}^{b^g}\delta\mathbf{b}_i^g + \mathbf{J}_{\Delta v}^{b^a}\delta\mathbf{b}_i^a\right) \\
\mathbf{e}_b = \mathbf{b}_j - \mathbf{b}_i
\end{cases}
\tag{31}
$$

$$
\mathbf{E}_{IMU}(i,j) = \rho\left[\left(\mathbf{e}_{\Phi}^T\mathbf{e}_p^T\mathbf{e}_v^T\right)\mathbf{\Sigma}_I\left(\mathbf{e}_{\Phi}^T\mathbf{e}_p^T\mathbf{e}_v^T\right)^T + \mathbf{e}_b^T\mathbf{\Sigma}_b\left(\mathbf{e}_{\Phi}^T\right)^T\right] \tag{32}
$$

4. Compass Objective function

The compass objective function is constructed according to the difference of incremental yaw angle calculated by IMU pre-integration and incremental heading angle calculated by compass. In the optimization period $\left[t_i, t_j\right)$, the incremental yaw angle $\Delta\psi_{BW}$ is derived from the first equation of (28), then the Rodrigues formula (2) is applied to compute the incremental angle. The incremental heading angle is obtained by $\Delta\psi_m = \psi_{m_j} - \psi_{m_i}$ from formula (13). The bias error of magnetoresistive sensor is also changing slowly. In the optimization period $\left[t_i, t_j\right)$, the bias at time j is $\mathbf{b}_j^m = \mathbf{b}_i^m + \delta\mathbf{b}_i^m$ like IMU. Above all, the residual function is (33), and the objective function is (34).

$$
e_m(i,j) = \Delta\psi_{BW_{ij}} - \left(\Delta\psi_{m_{ij}} + \mathbf{J}_{\psi}^{\mathbf{b}_m}\delta\mathbf{b}_i^m\right) \tag{33}
$$

$$
E_m(i,j) = \rho e_m^2(i,j)\Sigma_m \tag{34}
$$

In nonlinear optimization process, it is necessary to derive the Jacobian of the compass objective function. For the single Euler angle, the actual incremental yaw angle can be written as $\Delta\psi_{ij} = \psi_j - \psi_i$. For the partial derivative with respect to state, “lifting” step is used firstly [40]. “Lifting” is the method that substitutes the “$x + \Delta x$” for “x”, like (35). And then, Jacobian of the objective function is partial derivative with respect to “Δx”. The result of Jacobian of compass objective function is (36).

$$\begin{cases} \psi_i \leftarrow \psi_i + \delta\psi_i \\ \psi_i \leftarrow \psi_i + \delta\psi_i \\ \delta\mathbf{b}_i^m \leftarrow \delta\mathbf{b}_i^m + \hat{\delta}\mathbf{b}_i^m \end{cases} \tag{35}$$

$$\begin{cases} \dfrac{\partial r_\psi}{\partial\delta\psi_j} = 1 \\ \dfrac{\partial r_\psi}{\partial\delta\psi_i} = -1 \\ \dfrac{\partial e_\psi}{\partial\hat{\delta}\mathbf{b}_m} = -\dfrac{\partial\psi_m}{\partial\mathbf{b}_m} \end{cases} \tag{36}$$

As a consequence, the optimization problem of VIO aided by compass can be modeled as (37).

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \frac{1}{2}\rho\left(E_{proj}(k,i) + E_{IMU}(i,j) + E_m(i,j)\right) \tag{37}$$

5. Sliding window smoothing estimator

For VO or VIO system, there are three kinds of estimator, nonlinear filter (such as extended Kalman filter), full smoothing estimator and sliding window smoothing estimator. The nonlinear filter just estimates current state using measurements [41], while smoothing estimator uses history states as prior information and measurements to estimate current state, so the accuracy of smoother is better than filter. However, the dimensions of smoothing estimator are much larger than filter due to the history states, the computational complexity is the obvious burden for computer. Full smoothing estimator saves all the key frame poses and landmarks as the Gaussian prior, while sliding window smoothing estimator just saves many latest keyframes poses and landmarks and some sequent non-keyframes just before current temporal point, which is a local map. In the paper, the sliding window smoothing estimator is chosen as the approach of optimization, compromising the accuracy and computational complexity [33]. There are several algorithms to solve the optimization problem in real-time, such as Gauss-Newton algorithm and Levenberg-Marquardt algorithm. The Gauss-Newton algorithm is selected, the calculation steps are in Fig.2.

The landmarks of keyframes construct the local map, which also keep the co-visible relationship of the frames in the windows. Non-keyframes reserve the latest information of recent frames. Keyframes are decided based on the features which have low matching ratio with respect to the existing landmarks saved. It is because that keyframes should cover more feature of environment. If the number of current local map landmarks is N, and the number of which are detected is S, while the number of features matching to the landmarks of local map is T. If $S/\mathrm{N} \le \mathrm{s}_1$ or $S/\mathrm{T} \le \mathrm{s}_2$, the current frame is considered a keyframe [42]. The symbols s_1 and s_2 are threshold for keyframe judging.

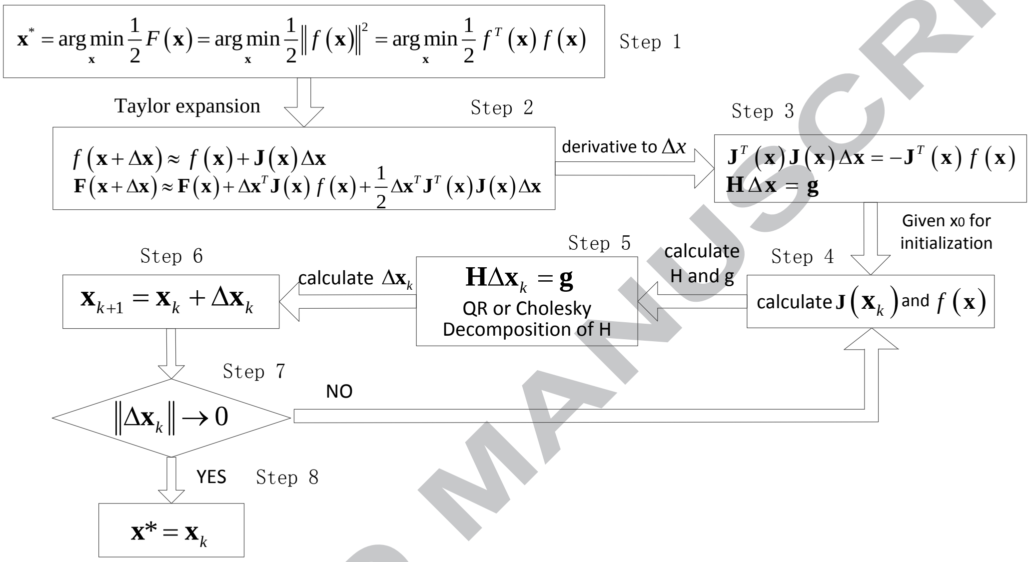


Fig. 4 The calculation step of Gauss-Newton algorithm

The size of sliding window is fixed, including M keyframes and N non-keyframes. The estimation should drop old frame when new frame is added to the window. But the process of dropping is not simply deleting the information of the oldest motion states and visible landmarks. It will lose much useful information, and it will make the estimator underdetermined. The method that "dropping" the oldest frame is called marginalization which algorithm is Schur complement.

In the process of solving the nonlinear optimization equation using Gauss-Newton algorithm in Fig.4, the equation is linearized based on Jacobian. The matrix H is esteemed Hessian. Both Jacobian and Hessian reserve the sparsity characteristic. Sparsity makes VO and VIO technology realize in onboard computer, using the same method to marginalization—Schur complement. For the linearized equation $\mathbf{H}\Delta\mathbf{x} = \mathbf{b}$ in Fig.4 (step 5), it is can be expressed in partitioned matrix like (38). The vector $\Delta\mathbf{x}_m$ means the state to be marginalized, while $\mathbf{x}_r$ is state to be reserved in the smoothing window. The Schur complement is the matrix $\begin{bmatrix} \mathbf{I} & -\mathbf{C}\mathbf{B}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$, which is left multiplied to both sides of the equation (38), and the result is (39). The $\Delta\mathbf{x}_r$ can be solved easily from the first equation in (39), and the $\Delta\mathbf{x}_m$ which is to be marginalized just left there with no more concern. With the Schur complement, the information which is older for current state is dropped, but the constraint information for current state is reserved in the estimator.

$$\begin{bmatrix} \mathbf{A} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x}_r \\ \Delta\mathbf{x}_m \end{bmatrix} = \begin{bmatrix} \mathbf{b}_r \\ \mathbf{b}_m \end{bmatrix} \tag{38}$$

$$\begin{bmatrix} \mathbf{A} - \mathbf{C}\mathbf{B}^{-1}\mathbf{C}^T & \mathbf{0} \\ \mathbf{C}^T & \mathbf{B} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x}_r \\ \Delta\mathbf{x}_m \end{bmatrix} = \begin{bmatrix} \mathbf{b}_r - \mathbf{C}\mathbf{B}\mathbf{b}_m \\ \mathbf{b}_m \end{bmatrix} \tag{39}$$

6. Inconsistency of VIO

In the process of solving optimization problem, the error of objective function linearization will result in the estimator drifting and inconsistency. The system will be over-confident due to inconsistency, and it will cause the estimator degenerate to sub-optimization. The paper [37] has compared with observability analysis for the navigation systems based on visual/inertial/magnetic sensors. The reason of inconsistency for VIO is that the degree of global unobservability is 4 dimensions, including 3 position dimensions and 1 yaw dimension, while the linearization value at different estimation which is used to calculate the Jacobians will lead the yaw direction to be observable. In the iterative computation, the spurious error is added to the yaw direction as Gaussian prior. The First Estimation Jacobian (FEJ) is an approximated method to solve the inconsistency problem [44]. The method is that Jacobian is not updated during $\mathbf{x}_k$ iteration in Gauss-Newton algorithm in Fig.4, which is calculated by the first estimation $\mathbf{x}_0$. The deficiency of FEJ is that the estimating state is approximated. Aided by compass, the yaw direction of VIO is global observable, so the dimensions of observability are not changed. The inconsistency of VIO system will be improved in theory.

VI. Simulation of Compass Aided VIO

1. Simulation Platform

Compass aided VIO system is tested by the Airsim [23]. It is a simulator for drones, cars and more built on Unreal Engine developed by Microsoft. The simulation system is consisted of the simulation computer and navigation on-board computer. The simulation computer based on

Airsim calculates UAV dynamic model and simulates the avionic sensor information. The Airsim supports hardware-in-loop with flight controllers PIXHAWK [14] for physically and visually realistic simulations. And in the VIO simulation system, the simulation computer is needed to generate the first-person-view (FPV) binocular images. AirSim is developed as an Unreal plugin that can simply be dropped in to any Unreal environment. As a result, the binocular images and the UAV flight scene can be simulated.

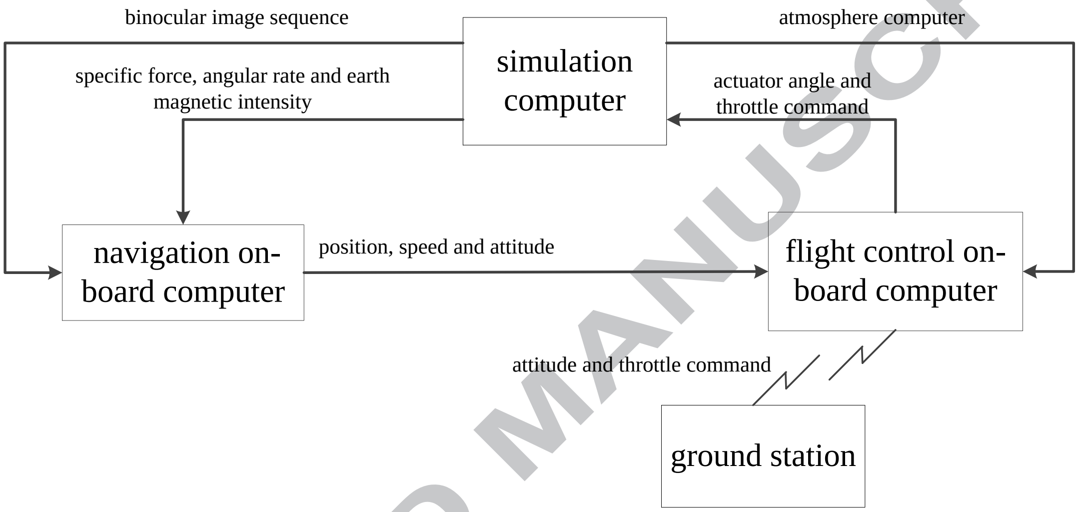


Fig.5 Simulation Platform Composition

The avionic sensor information can be simulated by Airsim. There are lots of software modular for simulating avionic sensor information according to UAV flighting state including specific force and angular rate which are measured by IMU and geomagnetic intensity of body frame which are measured by magnetoresistive sensors. The UAV flighting state is calculated in real-time by dynamics model. The demonstration of simulation system and its data flow is shown in FIG.5. The simulation system can be regarded as the hardware-in-the-loop simulation for the aerial rotor vehicle system.

From Fig.5, the simulation computer calculates the rotor vehicle dynamics and kinematics equations, of which the period is 5ms. The output information are specific force, angular rate, geographic intensity and binocular images. The specific force, angular rate, geographic intensity generation period is 5ms. The binocular images simulation period is also 20ms. Navigation onboard computer receives the sensor information and implements the compass aided VIO algorithm. The VIO send position, speed and orientation information to the flight controller computer, and transfer specific force and angular rate to it. The flight controller computer calculates the actuator angle and throttle command feedback to simulation for dynamic calculation according the VIO information and barometric pressure height also generated by simulation computer. The flight command is sent to flight control computer by ground station via wireless data transfers. Furthermore, the simulation is a closed loop system.

There is a flight scene developed by Airsim using Unreal4 vision engine, which is called “Neighborhood”, shown in Fig.6. The Neighborhood scene is consisted of low houses, trees and live facilities, which provide abundant vision information for VIO.

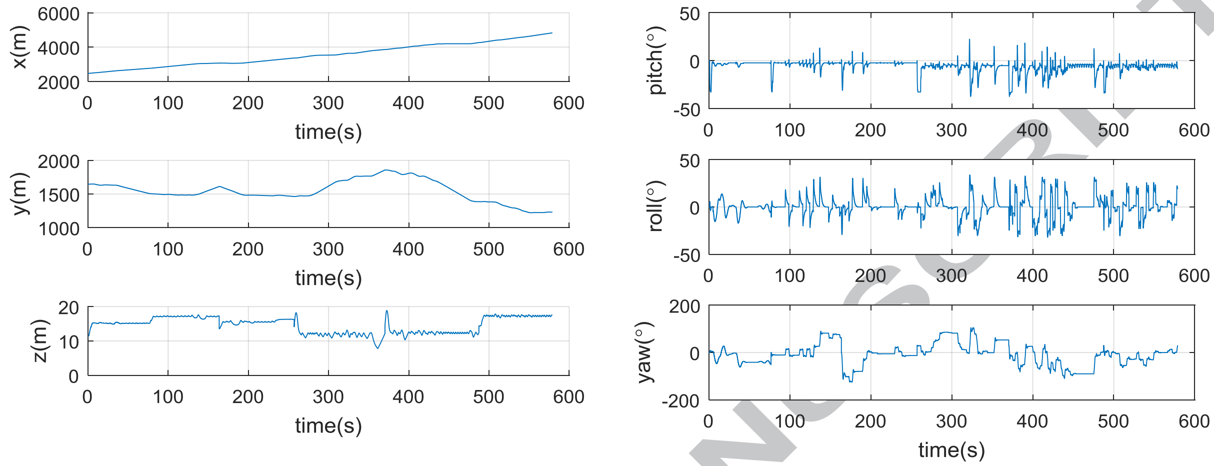
Fig. 6 The demonstration of Neighborhood scene

Tab.1 The sensor parameters of UAV list

Parameters of sensor	camera	image size	1024×512
		field view angle	90°
		auto exposure time range	0.016~20ms
		auto exposure brightness range	0.03~0.64
	IMU	gyroscope bias	10° /h
		angular random walk	0.05° $/h^{1/2}$
		accelerometer bias	0.4mg
		velocity random walk	0.04 $m/s/h^{1/2}$
	magnetoresistive sensor	zero bias	0.05nT
		white noise RMS	8.0×10^{-3} mGauss
	atmosphere computer	airspeed measurement error	2m/s
		baro altimeter measurement error	4m

2. Pose Error Result of Simulation

The trajectory of the simulation test is shown in FIG.7. The FIG. 7a) is the position of the ground truth, and the FIG. 7b) is the orientation of the ground truth. The ORB matching between current image and the latest local map image is demonstrated in Fig.8.



a) The position of simulation

b) The orientation of simulation

Fig.7 The trajectory of the simulation

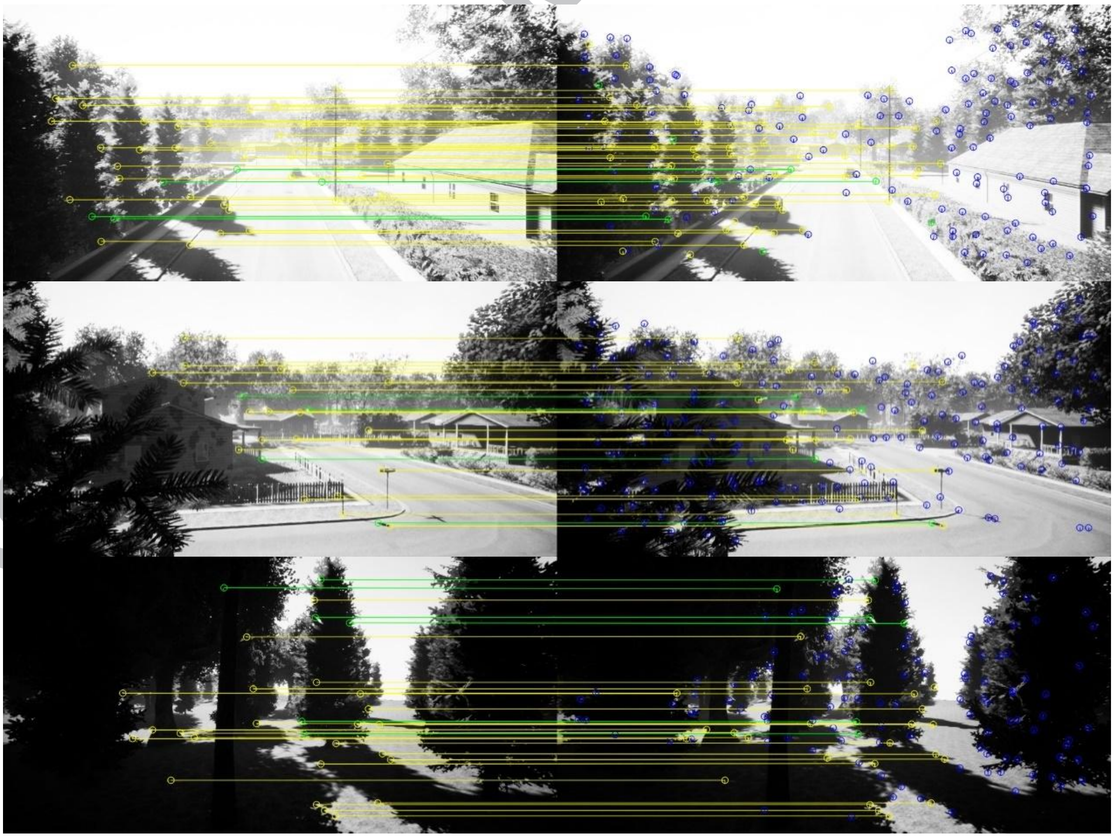
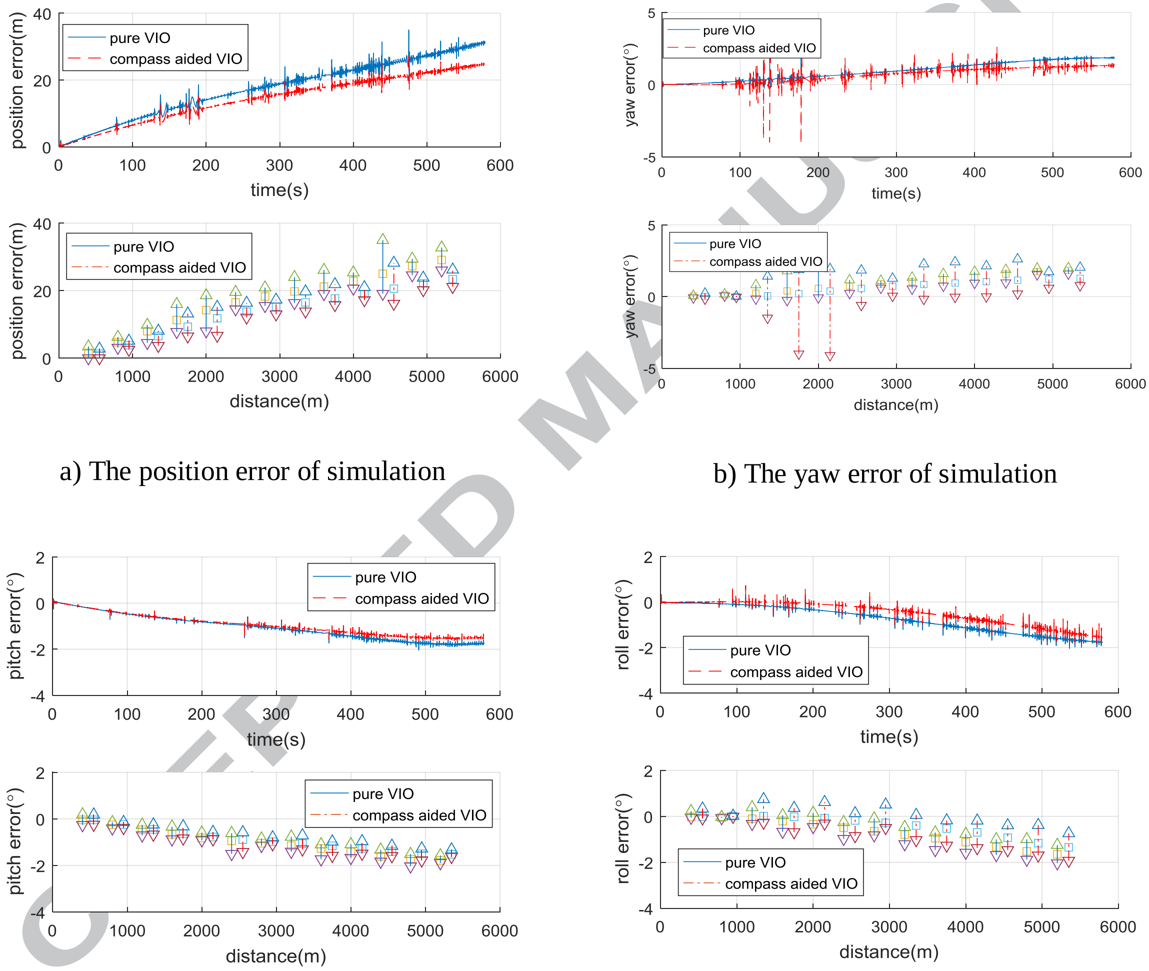


Fig.8 The simulation test matching images of VIO front-end demonstration

The simulation test can be repeated any times in the same condition according to the same trajectory and sensor parameters. The pure VIO and compass aided VIO can be simulated in the manner of the input errors controllability and the environment repeatability. The sensor parameters are listed in Tab.1 Because of the repeatability and the error controllability, the comparison of both VIO systems is meaningful. The simulation results are shown in Fig.9. The pose errors are calculated from the simulation ground truth recorded by the simulation and VIO result.



a) The position error of simulation

b) The yaw error of simulation

c) The pitch error of simulation

d) The roll error of simulation

Fig.9 The results of the simulation

The Fig.9a) is the position errors of both VIO systems. It can be concluded that the position estimated accuracy is improved by the compass aiding. The position error is denoted by the absolute position error of the VIO dividing the motion distance, which is 0.56%(5km) for pure VIO and 0.45%(5km) for compass aided VIO. The yaw angle is improved obviously because of the compass aiding. The pure VIO yaw error is 0.35° /km(5km) and the compass aided yaw error is 0.24° /km(5km), the yaw accuracy is increased by 31.4%. The improvement of position accuracy is mainly due to the yaw error decreasing. The results of the pitch and roll errors are slightly changed by the compass aiding. The pitch error of pure VIO is 0.34° /km(5km), which is 0.29° /km(5m) for the compass aided VIO. The roll error of pure VIO is 0.32° /km(5km), and the compass aided VIO is 0.26° /km(5km). Although the pitch and roll angle are estimated by the absolute measured value, the improvement of smoothing estimator inconsistency make the estimator more precise.

3. Result of VIO Inconsistency

In the process of solving nonlinear optimization problem, Jacobian is updated at different estimating states during iteration will cause the yaw direction observable which is unobservable originally in theory. Compass provides global observability for the yaw direction, the logic for inconsistency of VIO is invalid. Consistency of estimation system is evaluated by average Normalized Estimation Error Squared (NEES) [47], shown in (40). In (40), $\boldsymbol{\varepsilon}$ represents the error of position and orientation with respect to ground truth, and $\hat{\mathbf{P}}_k$ is the covariance of error of estimator. The average NEES is (41), it is statistically calculated by Monte-Carlo method, $\boldsymbol{\eta}_k^i$ represents the i-th running. It is assumed that VIO estimator is consistent, and $N \cdot \bar{\boldsymbol{\eta}}_k$ should obey χ_n^2 distribution. Symbol n in χ_n^2 is freedom of degree, it is obtained by the product of $\boldsymbol{\varepsilon}_k$ dimension and Monte-Carlo running times, shown in (42).

$$\boldsymbol{\eta}_k = \boldsymbol{\varepsilon}_k^T \hat{\mathbf{P}}_k \boldsymbol{\varepsilon}_k \tag{40}$$

$$\bar{\boldsymbol{\eta}}_k = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\eta}_k^i \tag{41}$$

$$n = \dim\left(\boldsymbol{\varepsilon}_k\right) \cdot N \tag{42}$$

Because of the repeatability of the simulation test, it is available to compare the NEES value of both VIO systems. According to χ_n^2 distribution, the significance level of distribution is α =2.5%, and the times Monte-Carlo running is 30. As a result, the degree of freedom is 180, the confidence interval of bilateral probability is $\bar{\boldsymbol{\eta}}_k \in [4.67, 7.51]$. If the average NEES curve exceeds the upper bound, it means that system is inconsistent. If the average NEES curve is lower than the lower bound, it means that system is conservative.

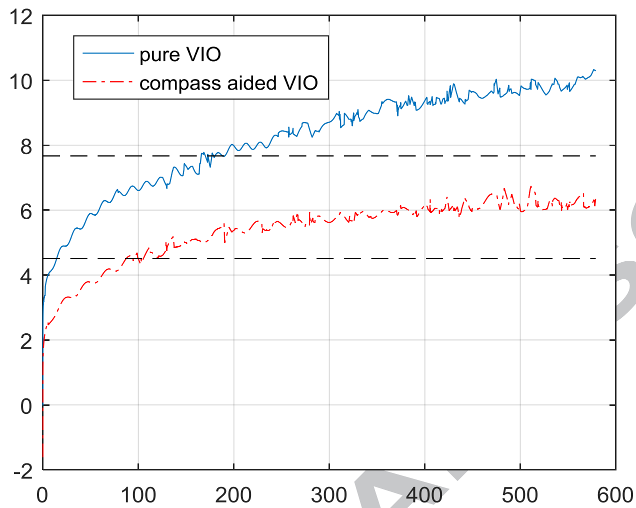


Fig.10 Average Normalized Estimation Error Squared value of the simulation test

The NEES of the simulation test is shown in Fig.9. It can be concluded that both NEES curves are divergent during the 5km distance trajectory. The pure VIO system is inconsistent with the Jacobian updating at different estimated states, which makes the yaw direction observable. However, the compass VIO system is consistent in the same condition to the pure VIO. Although the NEES value tend to divergence, the curve is still in the consistency range. It is because that the yaw direction is global observable with the compass aiding, the spurious error will not add to the yaw direction as the Gaussian prior. In addition to the observability, the yaw and position errors are decreased by compass aiding, while the covariance is tending to convergence, which make the compass aided VIO keep consistent during the simulation flight.

VII. Vehicle Test of Compass Aided VIO

1. The Configuration of the Vehicle Test System

For testing the accuracy of the compass aided VIO, the devices of compass aided VIO are installed on the testing vehicle, including cameras, magnetoresistive sensors and IMU. For evaluating the accuracy, the precision of integrated navigation system is taken for comparing, which is consisted of optic-fiber gyroscopes and GPS receiver. The integrated navigation system we adopted is SPAN-CPT produced by NovAtel. The position precision can attain to 1m (horizon) and 0.6m (vertical). The Euler angle precision can attain to 0.02° of roll and pitch, 0.06° of azimuth. In view of the precision of SPAN-CPT is more than 10 times than compass aided VIO, it can be used as the ground truth for evaluating VIO systems. The configuration of compass aided VIO test system configuration is shown in Fig.11. The test vehicle and its installation are demonstrated in Fig.12.

In Fig.11, the stereo vision is generated by two same cameras, whose version is GS3-U3-41C6C-C produced by Point Grey. The lens of the camera is Cinegon produced by Schneider. The resolution of the camera is 2048×2048, the pixel size is 5.5μm, and the image size is 1". And the focus of the lens is 10mm. The resolution of the camera is configured to be 2048×1024, it because that the occlusion is existed in FOV of pitch direction because of the space of vehicle, and the computational complexity of VIO will be reduced as well.

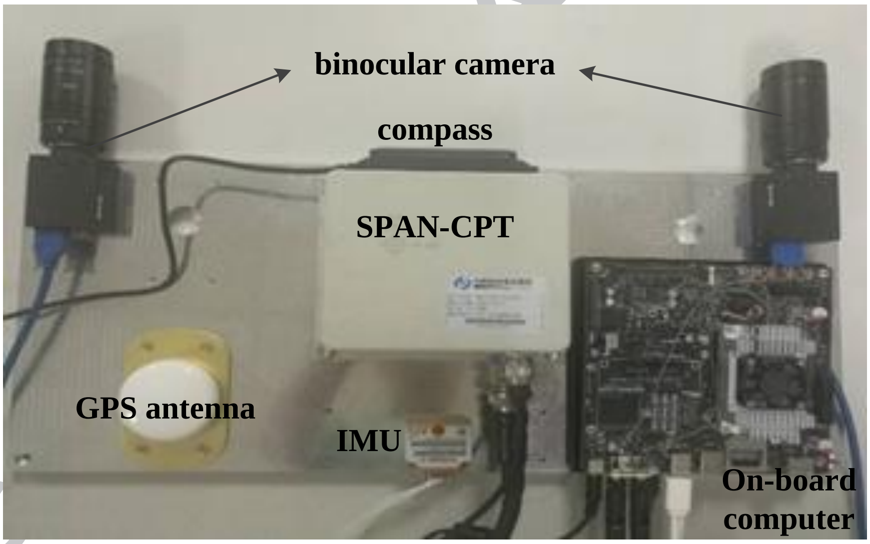


Fig.11 The test device of compass aided VIO configuration

a) The test vehicle

b) The installation scheme of the test device

Fig.12 The test vehicle and the installation of the test device

The IMU of the compass aided VIO adopted is the STIM300 produced by Sensonor. The IMU is consisted of three-axis gyro and accelerometer, the 6 degree of freedom can be calculated by the SINS. The Bias error over temperature gradients is 10° /h, and the bias repeatability of the gyro 4° /h. The angular random walk is 0.15° /h$^{1/2}$, which is used as the process noise of the orientation term of the estimator. The measurement range of accelerometer is adjustable. Because of the load of the vehicle is small, the range is configured to be ±5g. In condition of the range configuration, the maximum bias repeatability is 0.38mg; the bias error over temperature is ±1mg. The velocity random walk is 0.04 m/s/h$^{1/2}$, which is used as the state of position process noise of the estimator.

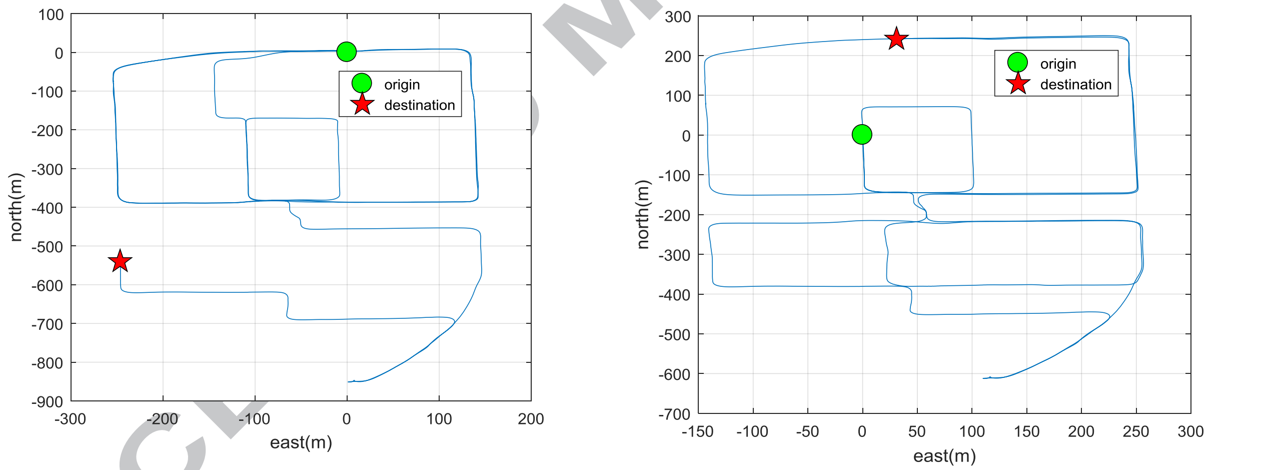
The compass of the system we adopted is HMC5883L produced by Honeywell, the heading angle accuracy can attain be to 1~2° , and the cost is low. The total size of the compass including its power and peripheral circuit is 71×63×6mm, and the shell is manufactured by wave transparent material shown in Fig.7. Because of the interference of the electromagnetic and other exterior magnetic field, the digital compass should be calibrated before working. In the condition without any magnetic inference, the vector of the orthogonal three-axis magnetoresistive sensor measurement can construct a sphere. However, the interference make the sphere degenerate to a ellipsoid. The principal of the calibration method is to fit sphere using the least square algorithm by multi-direction rotation of the magnetoresistive sensors[35] .

2. The Process of Vehicle Test

The device of compass aided VIO and test reference device SPAN-CPT are installed on the top of the test vehicle. Firstly, the SPAN-CPT start to initial alignment. When the visible satellites is more than 4 for the GPS receiver, the test vehicle begin to drive. Once the software of the ground software suggested that “alignment completed”, the vehicle stop to wait for VIO starting. Secondly, the compass aided VIO begin to initialize. The initialization process according to accelerometer and compass calculate the orientation; construct the initial local map; and the estimator convergence. While the initialization accomplished, the VIO received the GPS time from SPAN-CPT as the origin point of VIO time, which can be considered as the synchronization of SPAN-CPT and VIO synchronization. Thirdly, the vehicle begins to drive. When driving, the position, orientation, velocity of both compass aided VIO and SPAN-CPT are recorded for comparing. Besides the data, the intermediate matched images are saved by sampling. Once the compass aided VIO test ends, the pure VIO without compass aiding is tested according to the same process.

3. Result

The distance for driving to evaluate the performance of the compass aided VIO is 5km. The place of the test in the district of CIOMP, where the position of LLA coordinate frame is (43.849092°, 125.401490°, 2.22m). For convenience of calculating the position error, the position of LLA coordinate frame is converted to ground coordinate frame. The position, Euler angle and speed errors of the both systems are compared in the chapter to analyze the performance of the compass aided VIO. The trajectories of both tests are shown in Fig.13.



a) The trajectory of pure VIO vehicle test

b) The trajectory of compass aided VIO vehicle test

Fig.13 The trajectory of the both vehicle tests

The immediate matching images in VIO are shown in Fig.14. The left image in Fig.14 is the latest of the local map; the right image is the current image for solving the pose.

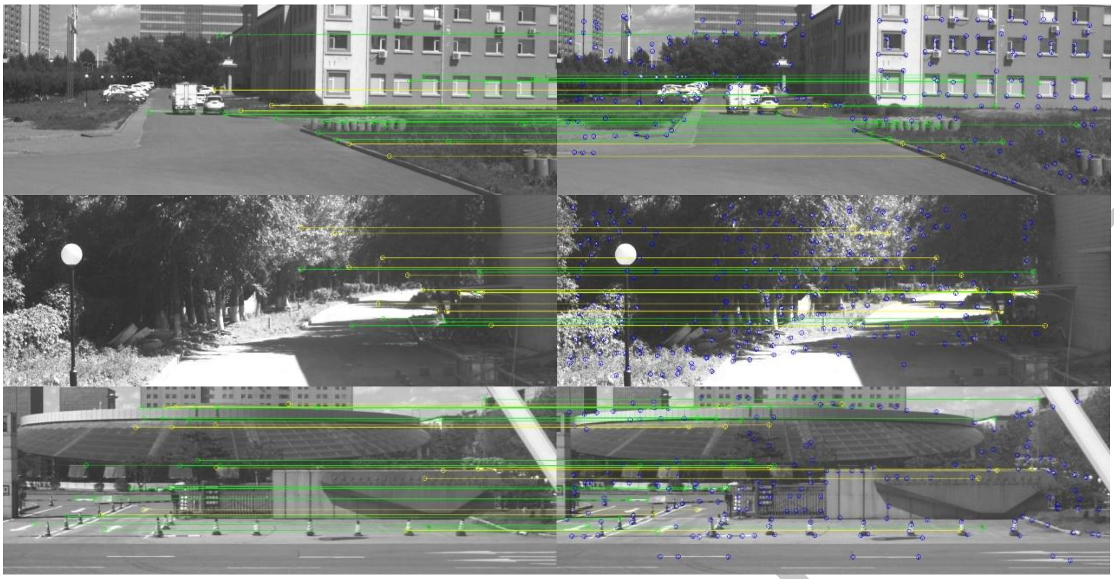
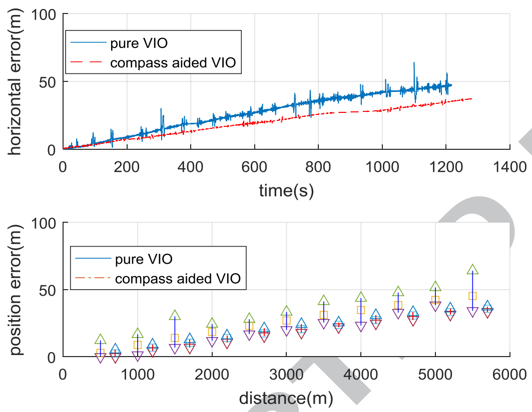
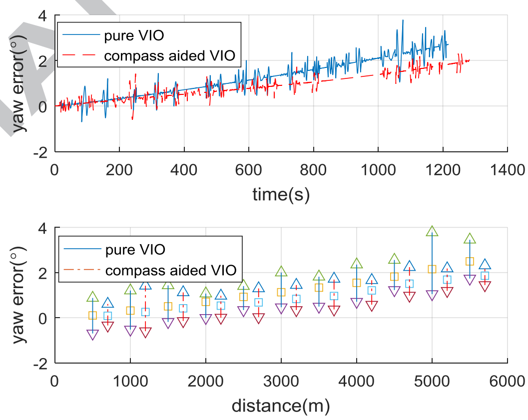


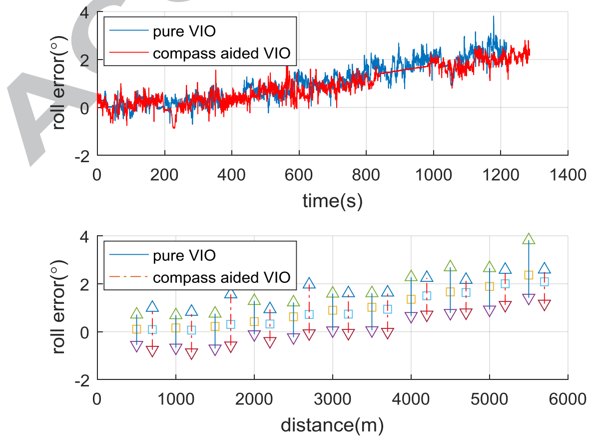
Fig.10 The matching images of VIO front-end demonstration



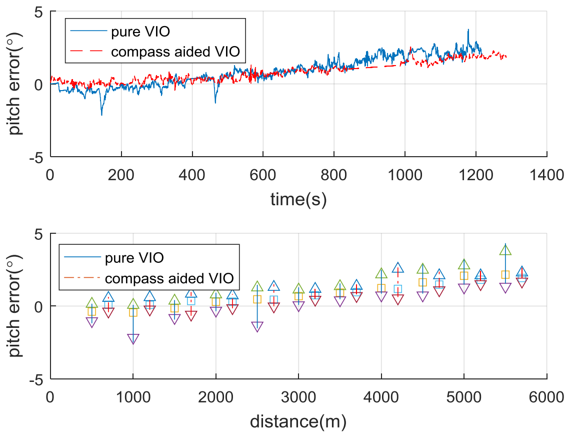
a) position error



b) yaw error



c) roll error



d) pitch error

Fig.12 Comparison of Eular angle error of pure VIO and compass aided VIO

The position, Euler angle and speed error curves of pure VIO and compass aided VIO are plotted in Fig11 and Fig.12. From Fig.11, we can conclude that position error of the compass aided VIO is 0.64%(5km), which of the pure VIO is 0.84%(5km). The horizontal position error is improved more obviously. The azimuth error is 0.34° /km(5km). which is improved depending on the compass aiding, while the pure VIO azimuth error is 0.43° /km (5km). The pitch and roll error are almost equal. The pitch and roll error of pure VIO are 0.42° /km (5km).and 0.40° /km (5km); the both angle errors of compass aided are 0.37° /km (5km). and 0.37° /km (5km).

VIII. Conclusion

In this paper, a method of compass aided VIO has been demonstrated and the motion estimation system with tightly coupled by the sensors of magnetoresistive sensor, IMU and camera is established. Firstly, the calculation method of magnetic heading is introduced and the design process of the front-end of the visual odometry is summarized. Then, based on the sliding window smoothing estimator, the objective function of the yaw angle with Compass and its Jacobian calculation form were deduced. After that, the Airsim is used to simulate the VIO aided by compass and compared with the VIO and compass aided VIO systems. It shows that the yaw and position accuracy of the latter is significantly improved. The average NEES value of pure VIO and VIO aided by compass is compared using Monte Carlo simulation, which shows that the latter significantly improves the inconsistency of the estimator, and then solves the problem that the estimator degrades to be sub-optimal. Finally, the compass aided VIO system is further tested by vehicle test in the real environment, the position and yaw accuracy is indeed improved, in which the increasement amplitude of the yaw accuracy is 21% and the position accuracy is 23%.

IX. Acknowledgement

The authors are grateful for the comments and suggestions of the reviewers and the Editor that helped to improve the paper significantly.

References

[1] D. Nistér, O. Naroditsky, J. Bergen, Visual odometry, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1 (1) (2004) I-652-659.

[2] D. Titterton, J. Weston, Strapdown inertial navigation technology, IEEE

Aerospace & Electronic Systems Magazine. 20 (7) (2005) 33-34.

[3] X. M. Liu, Z. Y. Chen, W. C. Chen, et al., Multiple optical flow sensors aiding inertial systems for UAV navigation, 2016 UKACC 11th International Conference on Control. (2016) 1-7.

[4] J. Engel, T. Schöps, D. Cremers, LSD-SLAM: Large-scale direct monocular SLAM, European Conference on Computer Vision. 8690 (2014) 834-849.

[5] V. Usenko, J. Engel, J. Stückler, et al., Direct visual-inertial odometry with stereo cameras, IEEE International Conference on Robotics and Automation. (2016) 1885-1892.

[6] R. Mur-Arta, J. M. M. Montiel, J. D. Tardós, ORB-SLAM: A versatile and accurate monocular SLAM system, IEEE Transactions on Robotics. 31 (5) (2017) 1147-1163.

[7] S. Leutenegger, S. Lynen, M. Bosse, et al., Keyframe-based visual-inertial odometry using nonlinear optimization, International Journal of Robotics Research. 34 (3) (2015) 314-334.

[8] D. Scaramuzza, F. Fraundorfer, Visual odometry: Part I: The first 30 years and fundamentals, IEEE Robotics & Automation Magazine. (2011).

[9] P. G. Savage, Strapdown inertial navigation integration algorithm design Part 1: Attitude algorithms, J. dyn. syst. meas. control. 21 (2) (1998) 384-384.

[10] A. I. Mourikis, S. I. Roumeliotis, A multi-state constraint Kalman filter for vision-aided inertial navigation, IEEE International Conference on Robotics & Automation. 22 (2007) 3565-3572.

[11] Y. L. Song, B. Xian, Y. Zhang, et al, Towards autonomous control of quadrotor unmanned aerial vehicles in a GPS-denied urban area via laser ranger finder, Optik. 126 (2015) 3877-3882.

[12] M. Kaess, H. Johannsson, R. Roberts, et al.,iSAM2: Incremental smoothing and mapping using the bayes tree, International Journal of Robotics Research. 31 (2) (2012) 216-235.

[13] R. Mahony, T. Hamel, J. M. Pflimlin, Nonlinear complementary filters on the special orthogonal group, IEEE Transactions on Automatic Control. 53 (5) (2008) 1203-1218.

[14] L. Meier, P. Tanskanen, F. Fraundorfer, et al., PIXHAWK: A system for autonomous flight using onboard computer vision, IEEE International Conference on Robotics and Automation. 19 (6) (2011) 2992-2997.

[15] D. L. Yuan, J. G. Yan, X. M. Wang, et al., A study of information fusion for UAV based on RBF neural network, 2007 IEEE International Conference on Control and Automation. (2007) 2839-2842.

[16] G. P. Huang, A. I. Mourikis, S. I. Roumeliotis, A first-estimates Jacobian EKF for improving SLAM consistency, Springer Tracts in Advanced Robotics.54 (2009) 373-382.

[17] M. Moakher, Means and averaging in the group of rotations, Siam Journal on Matrix Analysis & Applications. 24 (1) (2006)1-16.

[18] S. Maus, S. Mcmillan, S. Mclean, et al., The US / UK world magnetic model for 2010-2015. (2010).

[19] Zhang L, Song M, Liu Z, Liu X, Bu J, Chen C (2013) Probabilistic graphlet cut: exploiting spatial structure cue for weakly supervised image segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, Portland, pp. 1908–1915

[20] S. Maus, S. Mclean, D. Dater, et al., NGDC/GFZ candidate models for the 10th-generation international geomagnetic reference field, Earth Planets & Space 57 (12) (2005)1135-1140.

[21] D. Gebre-Egziabher, G. H. Elkaim, J. D. Powell, et al., A gyro-free quaternion-based attitude determination system suitable for implementation using low cost sensors, Position Location and Navigation Symposium. (2000) 185-192.

[22] Zhang, L., Yang, Y., Gao, Y., Yu, Y., Wang, C., & Li, X. (2014). A probabilistic associative model for segmenting weakly supervised images. IEEE Transactions on Image Processing, 23(9), 4150-4159.

[23] C. G. Harris, J. M. Pike, 3D positional integration from image sequences, Image & Vision Computing. 6 (2) (1988)87-90.

[24] E. Rosten, T. Drummond, Machine learning for high-speed corner detection, European Conference on Computer Vision. 3951 (2006) 430-443.

[25] D. G. Lowe, Distinctive image features from scale-invariant key points, International Journal of Computer Vision. 60 (2) (2004) 91-110.

[26] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, Speeded up robust features (SURF), Computer Vison & Image Understanding. 110 (3) (2008) 346-359.

[27] F. Fraundorfer, D. Scaramuzza, Visual odometry: Part II: Matching, robustness, optimization, and applications, IEEE Robotics & Automation Magazine. 19 (2) (2012) 78-90.

[28] R. Mur-Artal, J. D. Tardós, ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras, IEEE Transactions on Robotics. 33 (5) (2017) 1255-1262.

[29] G. Bradski, A. Daebler, Learning OpenCV: computer vision with OpenCV library, University of Arizona Usa Since. (2012).

[30] M. A. Fischler, R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Readings in Computer Vision. 24 (6) (1987) 726-740.

[31] R. Hartley, A. Zisserman, Multiple view geometry in computer vision, Cambridge University Press. (2000) 1865 - 1872.

[32] D. Nistér, An efficient solution to the five-point relative pose problem, IEEE Transactions on Pattern Analysis & Machine Intelligence. 26 (6) (2004) 756-777.

[33] H. Strasdat, J. M. M. Montiel, A. J. Davison, Real-time monocular SLAM: Why filter?, International Conference on Robotics & Automation. 58 (8) (2010) 2657-2664.

[34] D. Chetverikov, D. Stepanov, P. Krsek, Robust Euclidean alignment of 3D point sets: the trimmed iterative closest point algorithm, Image & Vision Computing. 23 (3) (2005) 299-309.

[35] H. Zhu, L. Z. Deng, A landmark-based navigation method for autonomous aircraft, Optik. 127 (2016) 3572-3575.

[36] Zhang L, Hong R, Gao Y, Ji R, Dai Q, Li X (2016) Image Categorization by Learning a Propagated Graphlet Path. IEEE T-NNLS 27(3):674–685

[37] G. H. Feng, X. S. Huang, Observability analysis of navigation system using point-based visual and inertial sensors, Optik. 125 (2014) 1346-1353.

[38] C. Y. Chiang, J. T. Jeng, B. L. Lai, et al., Tri-axis magnetometer with in-plane giant magnetoresistance sensors for compass application, Journal of Applied Physics. 117 (17) (2015) 631-649.

[39] T. Song, F. B. Zhang, D. F. Lin, Nonparametric frequency response function estimates for switching piecewise linear systems, Signal Processing. 129 (2016) 150-165.

[40] C. Forster, L. Carlone, F. Dellaert, et al., IMU preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation,Georgia Institute of Technology. (2015).

[41] L. Z. Cui, S. Y. Gao, H. G. Jia, et al., Application of Neural network aided Kalman filtering to SINS/GPS, Optics and Precision Engineering. 22 (5) (2014) 1304-1311.

[42] S. Leutenegger, P. Furgale, V. Rabaud, et al., Keyframe-based visual-inertial SLAM using nonlinear optimization, Robotics: Science & Systems. 34 (11) (2013) 789–795.

[43] Zhang L, Gao Y, Ji R, Dai Q, Li X (2014) Actively Learning Human Gaze Shifting Paths for Photo Cropping. IEEE T-IP 23(5):2235–2245

[44] D. G. Kottas, J. A. Hesch, S. L. Bowman, et al., On the consistency of vision-aided inertial navigation. Springer International Publishing. 88 (2013) 303-317.

[45] R. Kuemmerle, G. Grisetti, H. Strasdat, et al., G2o: A general framework for graph optimization, IEEE International Conference on Robotics & Automation. 7 (8) (2011) 3607–3613.

[46] M. Burri, J. Nikolic, P. Gohl, et al., The EuRoC micro aerial vehicle datasets, International Journal of Robot Research. 35 (10) (2016) 1157–1163.

[47] Y. Bar-Shalom, X. R. Li, T. Kirubarajan, Estimation with applications to tracking and navigation, John Wiley and Sons. 34 (2) (2001) 727-736.

[48] Zhang, L., Xia, Y., Ji, R., & Li, X. (2015). Spatial-aware object-level saliency prediction by learning graphlet hierarchies. IEEE Transactions on Industrial Electronics, 62(2), 1301-1308.