CHINESE-RUSSIAN "NAVIGATION AND MOTION CONTROL" YOUTH FORUM

CONFERENCE PROCEEDINGS

Table of Contents

2024

Chinese-Russian
"Navigation and Motion Control"
Youth Forum

Topic 1: Inertial Sensors, Navigation and Positioning

Systems

Self-Compensation Algorithm for Nonuniform Damping Drift of HRG Inertial Navigation System

Zhang Haoqian[1], Zhang Zhiqiang[2], Song Chao[1], Yang Songpu[1], Chen Gang[1]

(1. Tianjin Navigation Instrument Research Institute, Tianjin 300131, China;

2. Unit 92730 of PLA, Sanya 572016, China)

Abstract: Aiming at the problem that the navigation accuracy of hemispherical resonant gyroscope inertial navigation system (HRG-INS) is greatly affected by the nonuniform damping drift (NDD) of hemispherical resonant gyroscope (HRG), a self-compensation algorithm for NDD of HRG-INS is proposed. Based on the HRG-INS structure with four gyroscopes in conical configuration, the dynamic time-sharing working method is designed according to the motion state of the carrier. The angular velocity output of HRG-INS in navigation calibration mode and full navigation mode are determined respectively, and the NDD model of HRG is established and error compensation is performed in the navigation calibration mode. Finally, the self-compensation for NDD of HRG-INS is realized by switching the working mode on time. Experimental results prove that the algorithm proposed in this paper can effectively reduce the NDD error of gyroscopes and significantly improve the navigation accuracy of the HRG-INS during navigation without relying on external information.

Key words: hemispherical resonant gyroscope; conical configuration; nonuniform damping drift; self-compensation;

1 Introduction

Unevenly distributed internal stress remain inside the resonator of HRG, resulting in inconsistent damping at different positions of the resonator, which causes the output of HRG to contain NDD error. Reference [1] improved the accuracy of gyro by modeling and compensating for the NDD error of HRG, further research on drift compensation in navigation process is needed.

Aiming at the above problem, this paper proposes a self-compensation algorithm for NDD of HRG-INS which can be used in navigation process. This method is based on the HRG-INS structure with four gyroscopes in conical configuration. A dynamic time-sharing working method of HRG-INS is designed according to the motion state of the carrier. When the system is in the navigation calibration mode, the NDD of four gyroscopes are estimated in sequence to realize the self-compensation for the NDD of the HRG-INS while the system continuously outputs navigation information, significantly improving the navigation accuracy.

2 Four gyroscopes optimal configuration scheme

HRG has two working modes [2]: whole angle (WA) mode and force to rebalance (FTR) mode. The WA mode has a large dynamic range and is suitable for high-dynamic situations of the carrier. It is the mainstream solution of HRG-INS. The range and bandwidth of gyro in the FTR mode are limited, which is suitable for low-dynamic situations of carrier. However, it can stabilize the standing wave at a specific vibration angle by applying force, reducing the error of NDD. The

traditional three-gyro orthogonal configuration scheme requires all three gyros work in WA mode, and cannot work in the FTR mode which can compensate the NDD, so INS does not have the drift self-compensation function. Therefore, a four-gyro redundant configuration scheme is designed to solve the contradiction between the self-compensation and the continuity of navigation output.

2.1 Optimization criteria for redundant configuration scheme

For a strapdown inertial navigation system using n single-degree-of-freedom gyroscopes, the measurement input of the gyroscope can be expressed by formula (1)

$$Y = HX + \eta \tag{1}$$

Where Y is an $n \times 1$ dimensional measurement vector; H is an $n \times 3$ dimensional configuration matrix; X is an 3×1 dimensional vector to be determined (carrier input angular velocity); η is an $n \times 1$ dimensional measurement noise vector.

Assume that the measurement noise η is independent Gaussian white noise with a mean of 0 and a variance of σ^2. According to the least squares estimation theory, the estimated value of the angular velocity and the corresponding estimated covariance matrix C can be obtained

$$\hat{X} = \left(H^T H\right)^{-1} H^T Y = GY$$
$$C = Var\left(X - \hat{X}\right) = \left(H^T H\right)^{-1} \sigma^2 \tag{2}$$

The optimal criterion for the carrier angular velocity estimation accuracy is defined as

$$F = \sqrt{\left|\left(H^T H\right)^{-1}\right|} \tag{3}$$

The smaller the value of F, the smaller the error caused by the noise η. When $\left(H^T H\right)^{-1} = \dfrac{3}{n} I_3$, F takes the minimum value.

For the four-gyro redundant configuration scheme selected in this paper, its estimated covariance is $\sqrt{0.75}$ times of the three-gyro orthogonal configuration scheme, which can provide the basis for high precision self-compensation of gyroscopic NDD error.

2.2 Comparison of four-gyro redundant configuration schemes

The two common four-gyro redundant configuration schemes are the oblique configuration scheme (OCS) and the conical configuration scheme (CCS). The OCS is shown in Figure 1 (a), three gyros are installed orthogonally, and the measurement axis of the remaining gyro is installed at an equal angle of 54.73° to the other gyroscopes. The CCS is shown in Figure 1 (b), four gyros are evenly distributed on the conical surface, the semi-cone angle of the cone is 54.73°.

The measurement equation of OCS and CCS can be respectively obtained in formula (4).
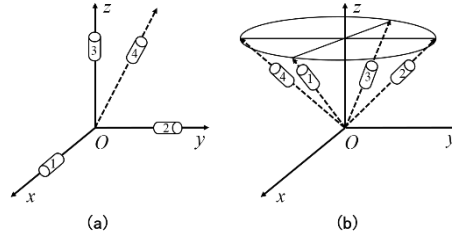
Fig.1 Four-gyro redundant configuration scheme

$$Y_a = H_a \omega = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \cos 54.73° & \cos 54.73° & \cos 54.73° \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}$$

$$Y_b = H_b \omega = \begin{bmatrix} \sin 54.73° & 0 & \cos 54.73° \\ 0 & \sin 54.73° & \cos 54.73° \\ -\sin 54.73° & 0 & \cos 54.73° \\ 0 & -\sin 54.73° & \cos 54.73° \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} \tag{4}$$

According to the optimal criterion for the carrier angular velocity estimation accuracy defined in formula (3), it can be calculated the F values of the two configuration schemes are 0.7071 and 0.6495, which proves CCS can achieve higher accuracy angular velocity estimation.

When a gyro fails in the four-gyro redundant configuration, the remaining three gyros can be used to continue angular velocity measurement. The measurement accuracy of different gyro combinations is directly related to the configuration matrix. Table 1 compares the measurement accuracy of different gyro combinations under each configuration scheme.

Tab.1 Measurement accuracy of different gyro combinations under each configuration scheme

Combination	Gyro number	OCS accuracy(%)	CCS accuracy(%)
a	①②③	100	76.98
b	①②④	57.74	76.98
c	①③④	57.74	76.98
d	②③④	57.74	76.98

As can be seen from Table 1, different gyro combinations of CCS have the same angular velocity measurement accuracy and better consistency. Based on the above analysis, CCS is the optimal solution for the four-gyro redundant configuration.

3 Dynamic time-sharing self-compensation algorithm for NDD

Set a reasonable angular velocity threshold and divide the working mode of HRG-INS into navigation calibration mode and full navigation mode based on the motion state of the carrier. By determining the mode switching method, drift compensation can be achieved during navigation.

3.1 Navigation calibration working mode

The core of the navigation calibration working mode is to use three navigation gyros operating in WA mode to measure the angular velocity of the carrier for navigation and project it to the direction of another redundant gyro. Combined with the output of the redundant gyro in FTR mode,

the NDD of the redundant gyro can be modeled and compensated during navigation.

As shown in Figure 1 (b), the base coordinate system $Oxyz$ of the carrier is determined, and its basis unit vectors are $\vec{e}_X, \vec{e}_Y, \vec{e}_Z$. The four gyros numbered ①②③④ are evenly distributed on the conical surface, and the semi-cone angle of the cone is $\alpha = 54.73°$. The angular velocity output vectors are represented by $\vec{g}_1, \vec{g}_2, \vec{g}_3, \vec{g}_4$, the conversion relationship between the corresponding unit vector $\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4$ and vector $\vec{e}_X, \vec{e}_Y, \vec{e}_Z$ is shown in formula (5).

$$\begin{bmatrix} \vec{e}_1 \\ \vec{e}_2 \\ \vec{e}_3 \\ \vec{e}_4 \end{bmatrix} = H_{4\times3} \begin{bmatrix} \vec{e}_X \\ \vec{e}_Y \\ \vec{e}_Z \end{bmatrix} = \begin{bmatrix} \sin\alpha & 0 & \cos\alpha \\ 0 & \sin\alpha & \cos\alpha \\ -\sin\alpha & 0 & \cos\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \vec{e}_X \\ \vec{e}_Y \\ \vec{e}_Z \end{bmatrix} \tag{5}$$

Since the four gyros are consistent, the gyro numbered ④ is used as a redundant gyroscope as an example to introduce the drift compensation scheme.

According to formula (5), the coordinate of angular velocity $\bar{\omega}_0$ in $Oxyz$ is expressed as

$$\bar{\omega}_0 = \vec{g}_1 + \vec{g}_2 + \vec{g}_3 = \left[\left(|\vec{g}_1| - |\vec{g}_3| \right)\sin\alpha \quad |\vec{g}_2|\sin\alpha \quad \left(|\vec{g}_1| + |\vec{g}_2| + |\vec{g}_3| \right)\cos\alpha \right] \tag{6}$$

$\bar{\omega}_0$ is the angular velocity of HRG-INS in the navigation calibration mode, which can be used for navigation. The projection size $|\bar{\omega}_1|$ of $\bar{\omega}_0$ on the redundant gyro can be expressed as

$$|\bar{\omega}_1| = \bar{\omega}_0 \cdot \vec{e}_4 = -|\vec{g}_2|\sin^2\alpha + \left(|\vec{g}_1| + |\vec{g}_2| + |\vec{g}_3| \right)\cos^2\alpha \tag{7}$$

The HRG can calculate angular velocity according to the magnitude of the applied control force in the FTR mode, as shown in formula (8).

$$K_{SF} C_P = \dot{\theta} + \alpha'\Omega + \Omega_b \tag{8}$$

Where K_{SF} is the force scale, θ is the gyro vibration angle, α' is the Bryan coefficient, Ω is the external input angular velocity, and Ω_b is the NDD of gyro.

The gyro is controlled to stabilize at the set vibration angle, and there is $\dot{\theta} = 0$. According to formula (8), the angular velocity measurement value of the redundant gyro can be calculated as

$$|\bar{\omega}_2| = \Omega = \frac{K_{SF} C_P - \Omega_b}{\alpha'} \tag{9}$$

Due to $|\bar{\omega}_1| = |\bar{\omega}_2|$, combined with formula (7) and formula (9), the NDD of the redundant gyro at different vibration angles can be calculated.

According to the error model of HRG [3], the NDD of HRG can be expressed as $\varepsilon(\theta) = \Delta\tau\cos(4\theta - 4\theta_\tau)$. Where $\varepsilon(\theta)$ represents the NDD that changes with the vibration angle, $\Delta\tau$ represents the amplitude of the NDD.

It can be seen the NDD is a symmetrical drift, which can be modeled as

$$\varepsilon(\theta) = a_0 + a_1 \cos 4\theta + b_1 \sin 4\theta \tag{10}$$

With θ and corresponding Ω_b, the NDD of redundant gyro can be modelled with the least squares method to calculate the parameters a_0, a_1, a_2. According to model (10), the compensation of the HRG's NDD can be achieved by subtracting the drift from the original gyro output.

3.2 Full navigation working mode

The core of the full navigation working mode is the four gyros are all in WA mode for navigation, which can achieve higher precision measurement of the carrier angular velocity [4].

According to formula (4), the transfer matrix H_b can be determined. Since the matrix $H_{4\times3}$ is irreversible, the least squares method is used to convert the four-gyro information to the three-axis orthogonal coordinate system.

$$\tilde{\omega}_{3\times1}^{out} = \tilde{H}_{3\times4} g_{4\times1}^{out} \tag{11}$$

Where $\tilde{H}_{3\times4}$ is the transfer matrix calculated by the least squares method, $g_{4\times1}^{out}$ is the angular velocity information output by the four gyros; $\tilde{\omega}_{3\times1}^{out}$ is the equivalent angular velocity information in the coordinate system $Oxyz$.

3.3 Working mode timing switching design

Firstly, according to the angular velocity of the carrier measured by the gyros, the carrier dynamic can be judged and the working mode of HRG-INS can be selected. When the carrier is in a low-dynamic environment, HRG-INS switches to navigation calibration working mode, select three navigation gyros in the order shown in Table 1, model and compensate for the NDD of the remaining redundant gyro during navigation. When the carrier is in a high-dynamic environment, it switches to the full navigation working mode. According to Table 1, the self-compensation for NDD error of HRG-INS can be realized by cyclic compensation of the NDD of the four gyros.

4 Experimental verification

In order to verify the effectiveness of the proposed algorithm, a turntable accuracy test was carried out. The test device includes a HRG-INS with four gyros in CCS, a three-axis turntable, a power supply, and a computer. The internal structure of the HRG-INS is shown in Figure 2.
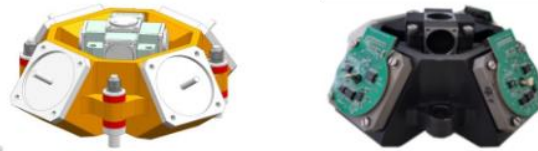
Fig.2 Internal structure of HRG-INS

In the experiment, from the point of view of verifying the algorithm correctness, the turntable was rotated in three axes at a rate of 5°/s for 5 hours. The redundant gyro in navigation calibration mode was set to control the gyro vibration angle to precess from 0° to 90° at an interval of 3°, and maintain 60 seconds at each vibration angle. When the redundant gyro completed two complete

vibration angle switching processes, the NDD model of this gyro was considered to be established. The experimental results are shown in Figure 3. Figure 3(a) is the NDD models of four gyros established in the experiment, Figure 3 (b) and Figure 3 (c) are the attitude errors and positioning errors of HRG-INS before and after the NDD compensation.

(a)   (b)   (c)
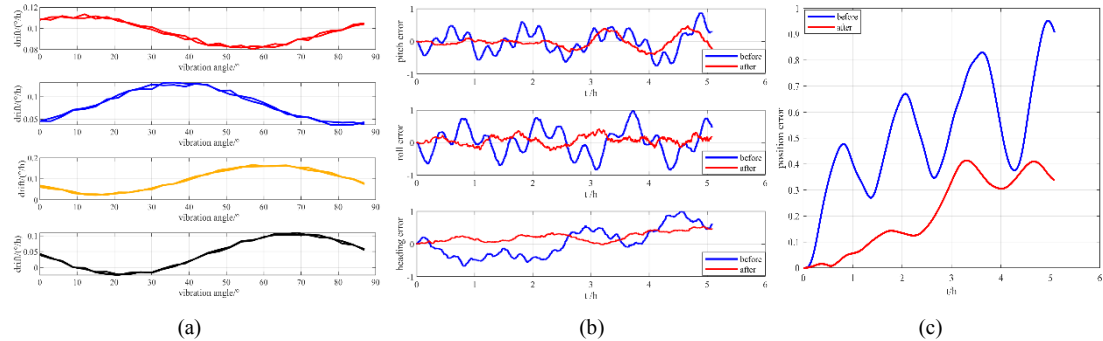
Fig.3   Experimental results

According to Figure 3, for 5-hour pure inertial navigation, the maximum attitude error and the maximum positioning error after the self-compensation are reduced by 54% and 57%, compared with the results before the compensation of NDD error of HRG-INS. The results show that the algorithm can effectively improve the navigation accuracy of HRG-INS.

5   Conclusion

This paper has proposed a self-compensation algorithm for NDD of HRG-INS. Based on the HRG-INS structure with four gyros in conical configuration, designed a dynamic time-sharing working method, established the NDD models of HRG. The experimental results prove that this algorithm can compensate the NDD error of HRG and effectively improve the navigation accuracy of the HRG-INS during navigation without relying on external information.

References:

[1]   Delhaye F. SpaceNaute® the HRG based inertial reference system of Ariane 6 European space launcher[J]. Gyroscopy and Navigation, 2019, 10(1): 1-6

[2]   Foloppe Y, Lenoir Y. HRG Crystal™ DUAL CORE: Rebooting the INS revolution[C]//2019 DGON Inertial Sensors and Systems (ISS). IEEE, 2019: 1-24.

[3]   Lynch D D. Vibratory gyro analysis by the method of averaging[C]//Proc. 2nd St. Petersburg Conf. on Gyroscopic Technology and Navigation, St. Petersburg. 1995: 26-34.

[4]   Wang L, Wang W, Zhang Q, et al. Self-calibration method based on navigation in high-precision inertial navigation system with fiber optic gyro[J]. Optical Engineering, 2014, 53(6): 064103.

Dynamic modelling of an imperfect hemispherical resonator with nonuniformities coupling

Cheng Wei[1], Xi Boqi[1], Ren Shunqing[1], Wang Jie[2], Ning Youhuan[1], Tian Zhen[1]

(1. School of Astronautics, Harbin Institute of Technology, Harbin 150001, China;

2. School of Materials Science and Engineering, Harbin Institute of Technology, Harbin 150001, China)

Abstract: Hemispherical Resonator Gyroscopes (HRG) represents a new type of inertial sensor. A hemispherical resonator serves as the core element of a HRG and its vibrational properties critically influence the output accuracy of the HRG. This paper proposes, for the first time, a novel dynamic model for an imperfect hemispherical resonator. This model accounts for the coupling between mass nonuniformity and residual stress nonuniformity. Firstly, equation of motion for the resonator, are derived based on the thin shell theory with Lagrange equation. Then, the frequency split and the azimuth of stiffness axis of the resonator are obtained by using a technique called "polar decomposition". Finally, test systems for the vibration parameter and the residual stress are designed. The experimental results show good consistency with theoretical predictions, confirming the correctness of the dynamic model. This work could lay a solid foundation for advancing the development of high-precision HRGs.

Key words: Hemispherical resonator; Dynamic modelling; Frequency split; Stiffness axis; Residual stress.

1 Introduction

The Hemispherical Resonator Gyroscope (HRG) has garnered significant interest in recent years due to its exceptional attributes such as small size, low noise, long lifetime, high performance and no wear-out [1]. Its applications span multiple domains, including land, sea, aviation, and space exploration [2,3]. Gradually, HRG is turning into one of the most promising inertial sensors in the 21st century.

Given the significance of the hemispherical resonator, the vibration behaviors of the resonator have been investigated from perfect shell to imperfect shell [4-7]. However, the hemispherical resonator actually used in the HRG is coated with a thin metal film on the inner surface of the resonator, which will additionally introduce residual stress [8] into the resonator.

To date, literatures lack dynamic models that account for the effects of the metal film layer. Furthermore, there is no existing theoretical framework addressing the coupling of mass nonuniformity and residual stress nonuniformity. Motivated by the above facts, this paper develops a dynamic model for an imperfect hemispherical resonator that considers the coupling of these nonuniformities. Then, based on the dynamic model, the frequency split and the azimuth of stiffness axis of the resonator, two important vibrational parameters, are obtained. Finally, test systems for

---

the vibration parameter and the residual stress are designed to verify the correctness of this model.

2 Theoretical analysis

The flat-electrode HRG is shown in Fig. 1. The inner spherical surface and the end surface adjacent to the resonator's lip are coated with a thin metal film layer. In order to facilitate the mathematical modeling, the hemispherical resonator is simplified as a hemispherical shell with uniform wall thickness, ignoring the influence of its support rod.
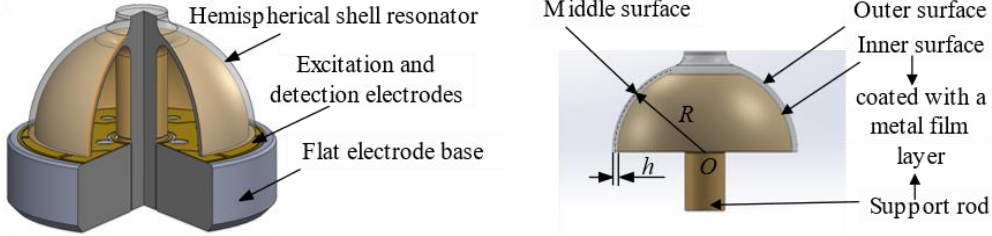
Fig. 1 Structure diagram of flat-electrode HRG and the hemispherical resonator.

2.1 Equation of motion

In order to characterize the uneven distribution of mass in the imperfect hemispherical resonator, the density is expanded into a Fourier series with respect to the circumferential angle β. This equivalent method greatly simplifies the problem. Similarly, due to the layer of metal thin film, the residual stresses are expanded into a Fourier series with respect to the circumferential angle β. So, these imperfections are assumed to occur as a small additional portion in the strain.

Based on the thin shell theory [9] and Lagrange equation, the equation of motion for the imperfect hemispherical resonator is obtained as

$$\begin{cases} (\rho_0 c_1 + \rho_4 c_2 \cos 4\beta_4)p'' + \rho_4 c_2 \sin 4\beta_4 q'' + (k_0 + k_1 - 2k_{\Delta f1})p - 2k_{\Delta f2}q = 0 \\ (\rho_0 c_1 - \rho_4 c_2 \cos 4\beta_4)q'' + \rho_4 c_2 \sin 4\beta_4 p'' + (k_0 + k_1 + 2k_{\Delta f1})q - 2k_{\Delta f2}p = 0 \end{cases} \tag{1}$$

where $p(t)$ and $q(t)$ are the physical coordinates that represent the second-order operating mode of the resonator. The symbol " $'$ " denotes the derivative with respect to time t. ρ_0 is the average density. ρ_4 is the fourth Fourier harmonic amplitude of the density. The specific expressions for coefficients k_0, k_1, $k_{\Delta f1}$, and $k_{\Delta f2}$ are calculated, as follows:

$$k_0 = R^2 \int_{-\frac{h}{2}}^{\frac{h}{2}} \int_0^{\frac{\pi}{2}} \left\{ \frac{E\gamma^2}{1-\mu^2}[c_{\chi_1^2}(\alpha) + c_{\chi_2^2}(\alpha) + 2\mu c_{\chi_1\chi_2}(\alpha)] + \frac{2E\gamma^2}{(1+\mu)}c_{\chi_{12}^2}(\alpha) \right\} \sin\alpha \, d\alpha \, d\gamma$$

$$k_1 = R^2 \int_{-\frac{h}{2}}^{\frac{h}{2}} \int_0^{\frac{\pi}{2}} [\sigma_{1\_0}^0 c_{\bar{\varepsilon}_1}(\alpha) + \sigma_{2\_0}^0 c_{\bar{\varepsilon}_2}(\alpha)] \sin\alpha \, d\alpha \, d\gamma \tag{2}$$

$$k_{\Delta f1} = c_{f14}\sigma_{1\_4}^0 \cos 4\beta_{1\_4} + c_{f24}\sigma_{2\_4}^0 \cos 4\beta_{2\_4} - c_{f\tau4}\tau_{12\_4}^0 \sin 4\beta_{12\_4}$$

$$k_{\Delta f2} = c_{f14}\sigma_{1\_4}^0 \sin 4\beta_{1\_4} + c_{f24}\sigma_{2\_4}^0 \sin 4\beta_{2\_4} + c_{f\tau4}\tau_{12\_4}^0 \cos 4\beta_{12\_4}$$

where $\sigma_{1\_0}^0$ and $\sigma_{2\_0}^0$ are the average residual normal stresses, and $\tau_{12\_0}^0$ is average residual shear stress in the meridional and circumferential directions, respectively. $\sigma_{1\_4}^0$ and β_{1_4} represent the amplitude and the initial phase of the fourth Fourier harmonic of residual normal stress $\sigma_1^0(\beta)$, respectively. $\sigma_{2\_4}^0$ and β_{2_4} represent the amplitude and the initial phase of the fourth Fourier

harmonic of residual normal stress σ_2^0 $\sigma_2^0(\beta)$, respectively. $\tau_{12\_4}^0$ and β_{12_4} represent the amplitude and the initial phase of the fourth Fourier harmonic of residual shear stress $\tau_{12}^0(\beta)$, respectively. In addition, the expressions for coefficients c_1, c_2, $c_{\chi_1^2}$, $c_{\chi_2^2}$, $c_{\chi_1\chi_2}$, $c_{\chi_{12}^2}$, $c_{\bar{\varepsilon}_1}$, $c_{\bar{\varepsilon}_2}$, $c_{f_{14}}$, $c_{f_{24}}$ and $c_{f_{r4}}$ are integral constant values, which are complex but do not affect the subsequent analysis, so their specific expressions are not provided here.

It can be concluded that the amplitude and the initial phase of the fourth harmonic for both mass and residual stresses contribute to the equation of motion. Upon examination of Eq. (1), the imperfect hemispherical resonator can be represented by an equivalent mechanical model of a two-dimensional mass-spring oscillator, as shown in Fig.2.
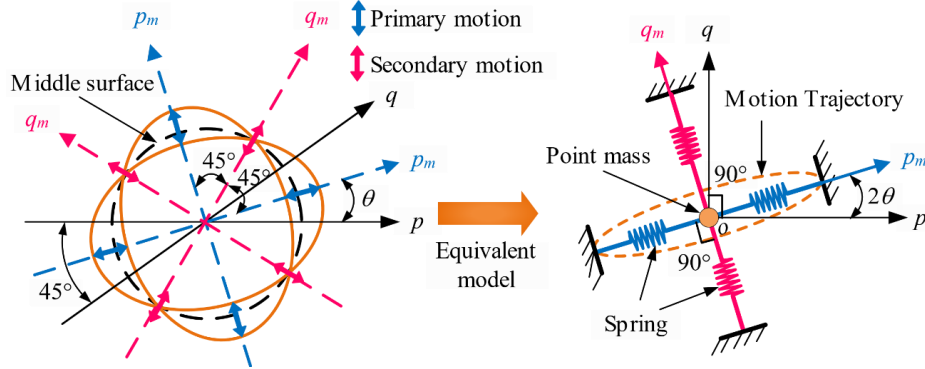
Fig. 2   Equivalent mechanical model.

The angle between the axis on which physical coordinate p (or q) is located and the axis on which principal coordinate p_w (or q_w) is located, is called the azimuth of stiffness axis θ.

2.2   Frequency Split

In order to facilitate the subsequent analysis, Eq. (1) is written in matrix form, as follows:

$$MX'' + KX = 0 \tag{3}$$

where X denotes the physical coordinates vector, M and K denote the mass matrix and the stiffness matrix, respectively, which are given as

$$M = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}, \quad K = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix}, \quad X = \begin{bmatrix} p \\ q \end{bmatrix} \tag{4}$$

Here, coefficients m_{11}, m_{12}, m_{21}, m_{22}, k_{11}, k_{12}, k_{21} and k_{22} are written as

$$\begin{aligned} m_{11} = \rho_0 c_1 + \rho_4 c_2 \cos 4\beta_4, \quad m_{12} = m_{21} = \rho_4 c_2 \sin 4\beta_4, \quad m_{22} = \rho_0 c_1 - \rho_4 c_2 \cos 4\beta_4 \\ k_{11} = k_0 + k_1 - 2k_{\Delta f1}, \quad k_{12} = k_{21} = -2k_{\Delta f2}, \quad k_{22} = k_0 + k_1 + 2k_{\Delta f1} \end{aligned} \tag{5}$$

According to the vibration properties for a two-degree-of-freedom system, the characteristic equation for the equivalent model is $|K-\omega^2 M|=0$. So, two natural frequencies ω_1 and ω_2 are obtained as

$$\omega_1 = \sqrt{\frac{k_{11}m_{22} - 2k_{12}m_{12} + k_{22}m_{11} + val}{2(m_{11}m_{22} - m_{12}^2)}}, \quad \omega_2 = \sqrt{\frac{k_{11}m_{22} - 2k_{12}m_{12} + k_{22}m_{11} - val}{2(m_{11}m_{22} - m_{12}^2)}} \tag{6}$$

where coefficient *val* can be expressed as

$$val = \sqrt{(k_{11}m_{22} - k_{22}m_{11})^2 + 4(k_{11}m_{22} - k_{22}m_{12})(k_{12}m_{11} - k_{11}m_{12})} \tag{7}$$

Then, the frequency split Δf is defined as

$$\Delta f = \frac{|\Delta\omega|}{2\pi} = \frac{|\omega_1 - \omega_2|}{2\pi} \tag{8}$$

2.3 Azimuth of stiffness axis

According to the vibration properties of a two degree of freedom system, its eigenvectors form a modal matrix $\boldsymbol{C}$, which is obtained as

$$\boldsymbol{C} = \begin{bmatrix} \dfrac{k_{22}m_{11} - k_{11}m_{22} - val}{2k_{11}m_{12} - 2k_{12}m_{11}} & \dfrac{k_{22}m_{11} - k_{11}m_{22} + val}{2k_{11}m_{12} - 2k_{12}m_{11}} \\ 1 & 1 \end{bmatrix} \tag{9}$$

Here, the eigenvectors in the modal matrix contain information about the azimuth of stiffness axis. Therefore, it is necessary to extract a pure rotation matrix $\boldsymbol{R}$ from the modal matrix $\boldsymbol{C}$. By using a technique called "polar decomposition" [10], the rotation matrix $\boldsymbol{R}$ is can be calculated by continuously averaging the transpose of the inverse of matrix $\boldsymbol{C}$ until convergence, as follows:

$$\boldsymbol{C}_{i+1} = \frac{1}{2}[\boldsymbol{C}_i + (\boldsymbol{C}_i^{\mathrm{T}})^{-1}] \tag{10}$$

At this point, $\boldsymbol{C}_i = \boldsymbol{R}$, and the form of the rotation matrix $\boldsymbol{R}$ is

$$\boldsymbol{R} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} = \begin{bmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{bmatrix} \tag{11}$$

Finally, the azimuth of stiffness axis θ can be obtained as

$$\theta = \frac{1}{2}\arctan\frac{-R_{12}}{R_{11}} \pm n \times \frac{\pi}{4}, \ (n=0, 1, 2, \cdots\cdots) \tag{12}$$

3 Experimental verification

The physical and geometrical parameters of resonator #01 are listed in Tab.1.

Tab.1 Physical and geometrical parameters of the resonator.

Symbol	Parameter	Value [Unit]
ρ_0	Average density	2200 kg/m^3
E	Young's modulus	7.67×10^{10} Pa
μ	Poisson's ratio	0.17
h	Thickness	0.85 mm
R	Middle surface radius	15 mm

As shown in Fig. 3 and 4, test systems for vibration parameter and residual stress are built. Here, the vibrational parameters refer to the frequency split Δf and the azimuth of stiffness axis θ.
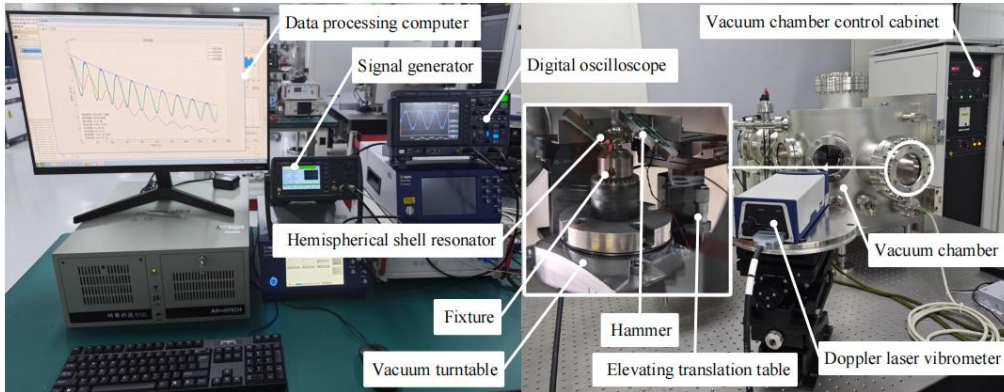
Fig. 3. Test system for vibration parameter.
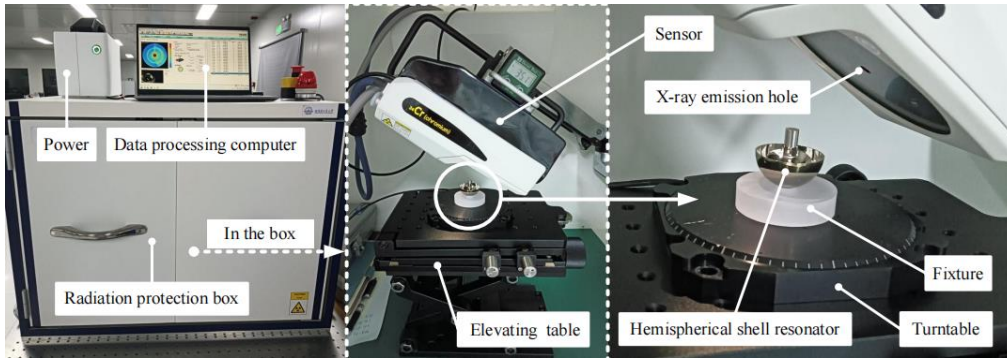
Fig. 4. Test system for residual stress.

First, based on the test system for the vibration parameter, the test results of the resonator #01 without metal film layer are: the frequency split is 0.0289 Hz, and the rigid axis angle is 7.72°, which reflects the influence of only mass nonuniformity. Then, the resonator is coated with a very thin metal film layer on its inner surface. In the test system for the residual stress, the turntable is controlled to rotate the resonator at intervals of 15 °, and 24 points in total are tested around the circumference of the resonator. Next, the error parameters for these three stresses are identified by the least squares method, and the identification results are shown in Tab.2.

Tab.2   Identification results of error parameters for three residual stresses.

Residual normal stress 1			Residual normal stress 2			Residual shear stress		
$\sigma_{1\_0}^{0}$	$\sigma_{1\_4}^{0}$	β_{1_4}	$\sigma_{2\_0}^{0}$	$\sigma_{2\_4}^{0}$	β_{2_4}	$\tau_{1\_0}^{0}$	$\tau_{1\_4}^{0}$	β_{12_4}
415MPa	356Pa	21.62°	-1023MPa	-819Pa	5.38°	362MPa	371Pa	10.89°

Finally, according to the error parameters in Tab.2, the frequency split and the azimuth of stiffness axis of resonator #01 with metal film layer are calculated by Eqs. (8) and (12), and the experimental results are obtained by the test system for the vibration parameters mentioned above, as listed in Tab.3.

Tab.3   Theoretical and experimental values of vibrational parameters.

Vibrational parameter	Theoretical value	Experimental value	Difference value
Δf	0.0422 Hz	0.0418 Hz	0.0004 Hz
θ	9.55°	9.49°	0.06°

Comparing the theoretical and experimental values in Tab.3, it can be concluded that the difference value of frequency split is 0.0004 Hz and the difference value of rigid axis angle is 0.06°, which are within a reasonable range for engineering applications. The experimental results show good agreement with the theoretical analysis. Thus, it proves the correctness of the dynamical model established in this paper.

4   Conclusions

This study presents a new dynamic model for an imperfect hemispherical resonator featuring a thin metal film layer based on the thin shell theory and Lagrange equation. This model incorporates the effects of mass nonuniformity and residual stress nonuniformity, which involves the amplitude and the initial phase of the fourth harmonic for both mass and residual stress. It proves that these nonuniformities coupling influence the frequency split and the azimuth of stiffness axis of the resonator. Finally, the test systems for the vibration parameter and the residual stress are designed, and successfully applied in analyzing an imperfect hemispherical resonator. The experimental results show the good consistency with the theoretical predictions. Thus, it proves the correctness of the dynamical model and error analysis in this paper.

References:

[1]   Jeanroy A, Grosset G, Goudon J C, et al. HRG by Sagem from laboratory to mass production[C]//2016 IEEE International Symposium on Inertial Sensors and Systems. Laguna Beach, CA, 2016: 1-4.

[2]   Delhaye F. HRG by SAFRAN: The game-changing technology[C]//2018 IEEE international symposium on inertial sensors and systems (INERTIAL). Italy, 2018: 173-176.

[3]   Delhaye F. Space Naute® the HRG based inertial reference system of Ariane 6 European space launcher[J]. Gyroscopy and navigation, 2019, 10: 1-6.

[4]   Ning Y, Wang C, Cheng W, et al. Anchor loss improvement in hemispherical resonators with ion beams[J]. International Journal of Mechanical Sciences, 2024, 277: 109413.

[5]   Xu Z, Xi B, Yi G, et al. A novel model for fully closed-loop system of hemispherical resonator gyroscope under force-to-rebalance mode[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(12): 9918-9930.

[6]   Deng K, Pan Y, Jia Y, et al. Separation of mass and gap errors of hemispherical resonator gyroscopes by varying bias voltage[J]. Measurement, 2024, 228: 114308.

[7]   Zheng L, Li F, Wu Z, et al. Dynamic modelling and quality factor evaluation of hemispherical shell resonators[J]. Journal of Sound and Vibration, 2024, 571: 118046.

[8]   Yang D, Zhao L, Chen M, et al. A novel method to characterize the residual stress on the fused silica surface based on the evolution of the atomic point defects[J]. Applied Surface Science, 2023, 640: 158323.

[9]   Huo Y, Wei Z, Ren S, et al. High precision mass balancing method for the fourth harmonic of mass defect of fused quartz hemispherical resonator based on ion beam etching process[J]. IEEE Transactions on Industrial Electronics, 2022, 70(9): 9601-9613.

[10]   Shoemake K, Duff T. Matrix animation and polar decomposition[C]//Proceedings of the conference on Graphics interface. UNIV BRIT COLUMBIA, VANCOUVER, CANADA , 1992, 92: 258-264.

Complete Calibration Cycle for Assembled Strapdown Inertial Navigation Systems

A.V. Kozlov, F.S. Kapralov, G.O. Barantsev, S.A. Fedorov

(Lomonosov Moscow State University, Moscow, Russia)

Abstract: The report provides an extended list of error sources of inertial sensors and their corresponding calibration procedures. Altogether, they constitute a complete calibration cycle, which every navigation system should undergo through its development and production. We discuss standard sensor error model, as well as measurement timing skew, size effect, angular misalignment, elastic deformations, and other parameters. Quantitative assessment of the produced navigation errors allows one to evaluate whether particular effects should be addressed when developing a new system and its calibration methods.

1   Introduction

Most publications and reports about calibration of strapdown inertial navigation systems (INS) focus on specific methods for estimating null biases, scaling errors and axial misalignments of inertial sensors which measure angular rate (gyroscopes) and specific force (accelerometers). All other systematic error sources are considered already eliminated at that point. Indeed, one have to estimate these listed above parameters for each individual instrument, and their accuracy strongly affects navigation solution [1]. However, complete set of calibration and test experiments intended to assess whether the actual sensor errors conform to the models accepted for them, appears to be substantially wider, and especially so for systems under development. The report provides a more complete list of inertial sensor error sources, based on the expertise of our laboratory in developing and calibrating a wide variety of navigation instruments — from micromechanical low-grade gyrocompasses to strapdown airborne gravimeters. Authors have developed some of reported calibration techniques and their combinations, which allow reducing the time taken by calibration experiments, while some methods are listed for completeness. We also study mutual influence of error sources onto the estimation of one another, which naturally occurs in practice while working with systems not yet fully calibrated in production and/or development. We also provide quantitative relations and examples.

2   Error classes

For the sake of this report, we will further consider the following classes of sensor errors based on the amount and types of calibration procedures needed to produce necessary compensation models:

- parameters identical for all systems of the same type;
- parameters different for each individual instrument;
- temperature variations;
- calibration experiment parameters affecting the calibration of INS.

Let us further discuss these classes one after another.

3 Parameters identical for all systems of the same type

The below list of error sources are usually calibrated (or tested for) once for all INS of the same model while in development:

- synchronization of gyroscope measurements;
- timing skew between accelerometer and gyroscope measurements;
- spatial separation of accelerometer proof masses (size effect).

The significance of precise synchronization of gyroscope measurements is often not recognized at development stage, mainly due to most tests involving rotations only around single instrumental axis. However, when INS undergoes harmonic oscillations around an angle bisector between two instrumental axes, the timing skew between corresponding gyroscope measurements results in average attitude drift around orthogonal axis roughly proportional to the square of amplitude times frequency squared. For example, when amplitude is 90° and period is 5 seconds (active maneuvering), or when amplitude is 3′ and frequency is 320 Hz (vibrations) each microsecond of the timing skew produces average attitude drift of approximately 0.15°/hr, which is unacceptable for navigation grade systems. To estimate the magnitude of this effect in a particular INS the authors have developed a dedicated experiment [2]. We designed the method to be resilient to other residual INS calibration errors.

The timing skew between accelerometer and gyroscope measurements appears to be of lower importance for navigation solution; however, as airborne gravimeter flight tests have shown, it produces considerable variations in gravity anomaly estimates. The dynamic calibration referenced below allows the estimation of the magnitude of the effect.

Spatial separation of accelerometer proof masses, better known as size effect, had been studied well in literature. Nevertheless, since gravimetry requires extreme precision in measuring acceleration while on a plane flying airborne, it implies sub-millimeter accuracy in estimating the separation parameters (see [3] for reference). This, in turn requires special methods of calibration which also should tolerate residual calibration errors for other parameters.

3 Parameters different for each individual instrument

We divide here error sources that require estimation of their model parameters individually for each INS instrument (excluding for now their temperature variations) into the following groups:

- errors produced by electronic circuits;
- angular alignment parameters;
- standard model of inertial sensor errors;
- extended inertial sensor error models: elastic deformations of axes of sensitivity, etc.

Errors produced by electronic circuits sometimes require separated calibration procedures, and sometimes they are mitigated using an "auto-calibration" in real time.

Angular alignment parameters allow one to align instrumental axes of the INS or inertial measurement unit (IMU) with those of the carrier object [4]. Three components of Euler rotation vector (often being small) may serve this purpose. The rotation itself may be either constant (preferably), or variable. For example, if the elastic torsion coefficients of the IMU damping system are larger than 0.5′/g in magnitude, then under 6-g loads alignment errors will exceed threshold of 3′, which may be undesirable.

Sensor bias, scaling and misalignment (referred here to as standard errors) calibration is widely covered in literature, though often separated from other error sources. The presence of these errors should be taken into account in every other calibration procedure. We consider two methods of calibration for the standard errors, as well as for other INS error parameters: static and dynamic calibration [5].

Extended inertial sensor error models include non-linear terms. They must be small enough to neglect multiplicative noises appearing in their compensation. Their parameter estimability in different calibration procedures, and especially how these parameters are separated from other error sources and how they are conditioned in estimation, is poorly studied. Known error sources of this type are: g-sensitivity of mechanical gyroscopes, non-linear characteristics of inertial sensors (e.g., their scaling asymmetry) and elastic deformations of their axes of sensitivity. Their calibration may be included in the procedures for standard sensor error model [6]. While developing a new INS, one should test it for the above effects, and then decide whether to neglect or compensate them n calibration.

4   Temperature variations

Traditionally, the approach to calibrating temperature variations is to perform calibration experiments in several temperature points. As research [7] indicates, linear temperature variations coefficients may be estimated along with other parameters under changing temperature considerably reducing the time required for calibration. For micromechanical sensors, there exist cases of signal dependency on not only on temperature itself, but also on its time profile.

5   Parameters inherent to calibration experiment

There exist quantities that do not affect sensor errors themselves, but still may affect calibration or skew the desired estimates if not accounted for. They include:

- turntable inclination from the plumb line and declination from reference azimuth;
- INS spatial separation from the center of rotation on a turntable [8];
- IMU damping system parameters mentioned above in regards to its axial misalignment.

For these parameters, full observability is not required, however taking them into account and their separation from estimated sensor error parameters are necessary.

6   Conclusion

We have presented an extended list of INS error sources, which may guide one through development and testing of newly developed inertial navigation systems. An overview of typical error manifestation cases and methods for their calibration and mitigation has been given.

References:

[1]   Kuznetsov A., Molchanov A., Fomichev A., Zheleznov V., Kozlov A. On Guaranteeing Tolerances for Strapdown INS Instrumental Errors. 30th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS), Saint Petersburg, Russia, 2023, pp. 1-6. DOI: 10.23919/ICINS51816.2023.10168385.

[2]   Kozlov A., Kapralov F., Fomichev A. Calibration of a Timing Skew between Gyroscope Measurements in a Strapdown Inertial Navigation System. 26th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS), St. Petersburg, Russia, 2019, pp. 1-5. DOI: 10.23919/ICINS.2019.8769417.

[3]   Kozlov A., Kapralov F. Millimeter-level calibration of IMU size effect and its compensation in navigation grade systems. 2019 DGON Inertial Sensors and Systems (ISS), Braunschweig, Germany, 2019, pp. 1-12. DOI: 10.1109/ISS46986.2019.8943630.

[4] Kozlov A.; Kapralov F. Angular Misalignment Calibration for Dual-Antenna GNSS/IMU Navigation Sensor. Sensors 2023, 23, 77. DOI: 10.3390/s23010077.

[5] Vavilova N., Vasineva I., Golovan A., Kozlov A., Papusha I., Parusnikov N. The Calibration Problem in Inertial Navigation. Journal of Mathematical Sciences, 253. DOI: 10.1007/s10958-021-05272-y.

[6] Barantsev G., Kozlov A., Shaimardanov I., Nekrasov A. A Model of the Elastic Dynamic Torsion of a Ring Laser Gyroscope Mechanical Dither and a Method for its Calibration. 29th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS), Saint Petersburg, Russian Federation, 2022, pp. 1-4. DOI: 10.23919/ICINS51784.2022.9815435.

[7] Kozlov A., Tarygin I., Golovan A., Shaymardanov I., Dzuev A. Calibration of an inertial measurement unit at changing temperature with simultaneous estimation of temperature variation coefficients: A case study on BINS-RT. 24th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS), St. Petersburg, Russia, 2017, pp. 1-3. DOI: 10.23919/ICINS.2017.7995635.

[8] Kozlov A., Sazonov I., Vavilova, N. IMU calibration on a low grade turntable: Embedded estimation of the instrument displacement from the axis of rotation. 2014 International Symposium on Inertial Sensors and Systems (INERTIAL), Laguna Beach, CA, USA, 2014, pp. 1-4. DOI: 10.1109/ISISS.2014.6782525.

Wideband Angular Rate Measurement Enhancement: Adaptive Strong-Tracking Kalman Filtering for MEMS and MHD Gyro Data Fusion

Yue Ji, Cui Guo, Leyun Liang

(Tianjin Key Laboratory of Intelligent Control of Electrical Equipment, Tiangong University, Tianjin 300387, China)

Abstract: This paper presents a comparative study of broadband angular rate measurement methods utilizing MEMS and MHD gyroscopes. We propose an enhanced adaptive strong-tracking Kalman filter fusion method to develop a broadband virtual gyroscope. Initially, the MHD and MEMS gyroscopes are modeled in the frequency domain. Subsequently, we design an adaptive Kalman filter algorithm based on this frequency domain analysis to effectively fuse the data from these two sensors, compensating for the low-frequency errors associated with MHD-ARS. The fusion filter not only estimates the state but also adapts the noise matrix through a noise estimator, maintaining stability even as sensor characteristics evolve during operation. Furthermore, we introduce an attitude solution approach leveraging a conic error compensation algorithm that capitalizes on gyro frequency domain features, culminating in a dual-sensor hybrid system that achieves a wide-band virtual gyroscope. This fusion method harnesses the strengths of MEMS gyroscopes in the low-frequency range (0-1 Hz) while taking advantage of the optimal high-frequency performance of MHD-ARS, thus providing effective compensation for its low-frequency limitations.

Key words: broadband angular rate measurement, MHD gyroscope, data fusion, adaptive Kalman filtering;

1 Introduction

In today's high-tech era, the demand for precision angular velocity measurement technology is escalating, particularly in aerospace, marine navigation, satellite stabilization, and precision instrumentation. As a critical component of inertial navigation systems (INS), the performance of angular velocity sensors significantly impacts the overall effectiveness of these systems[1]. Increasing requirements for high accuracy, wide dynamic range, and long-term stability challenge traditional sensors like fiber optic gyroscopes and microelectromechanical systems (MEMS) gyroscopes[2,3]. While these sensors perform well in certain applications, their adaptability and stability in harsh environments remain limited. Furthermore, their poor performance in the low-frequency range constrains their utility in long-duration stable measurements.

Magnetohydrodynamic (MHD) angular velocity sensors (ARS) have garnered attention for their ultra-low noise[4], shock resistance, and low cost across an ultra-wide bandwidth. Notably, MHD-ARS eliminate mechanical moving parts, allowing for rapid recovery from large-angle rotations while maintaining low measurement noise across a broad frequency range. However, MHD-ARS also exhibit limitations in the low-frequency range, particularly below 1 Hz[5,6], which

restricts their application in sustained, stable measurements.

To address these challenges, this study investigates fusion algorithms that combine MHD and MEMS gyroscope signals. Conventional adaptive Kalman filters often face issues such as signal-to-noise ratio degradation and fusion frequency jitter, primarily due to deviations in the measurement transfer matrix, rendering them unsuitable for fast dynamic systems. This paper proposes an enhanced adaptive Kalman filter that connects measurement covariance to signal frequency, allowing for corrections based on frequency domain characteristics. The filter is designed to rely primarily on the MEMS gyroscope in the low-frequency band, fuse data from both sensors in the mid-frequency range, and depend predominantly on the MHD gyroscope in the high-frequency band[7].

Iwata et al. used Kalman filters for MHD and MEMS gyroscopes and recommended exploring adaptive Kalman complementary filters for better data fusion[8]. Adaptive estimation methods have improved and are now applied to real-time state estimation. The Sage-Husa Adaptive Kalman Filter (SHAKF) helps estimate unknown model parameters in real time by recursively calculating the measurement noise covariance matrix[9]. This reduces the need for large amounts of historical data, making data processing more efficient. However, SHAKF's accuracy is sensitive to the noise covariance matrix, which can lead to issues like filter divergence. To address these challenges, researchers have proposed various modifications. For instance, Samrat L. Sabat et al. adjusted SHAKF to reduce noise in fiber optic gyroscopes[10]. Du et al. proposed enhancing adaptive Kalman filters with digital low-pass filtering for better noise estimation[11], but MHD and MEMS gyroscopes often have mixed signals that cannot be filtered effectively in the frequency domain. Arsalan et al. applied adaptive bandpass filters for heart rate signal extraction, which work well for narrow bandwidth signals but are not suitable for real-time measurements across a wide range[12].

This paper introduces a new adaptive strong tracking Kalman filter with enhanced spectral analysis to fuse MHD and MEMS gyroscope data. A dual-sensor hybrid system is developed, incorporating an attitude solution method based on a conic error compensation algorithm that leverages gyro frequency domain features. The organization of the paper is as follows: Section II presents the adaptive strong tracking Kalman filter fusion method and attitude solution technique; Section III validates the algorithm through simulations; Section IV verifies its feasibility through experimental results; and Section V concludes the findings.

2 Data processing

2.1 Fusion theory for adaptive strong tracking Kalman filtering

In the angular velocity measurement system, due to the inherent principle error of the MHD angular velocity sensor, a large amount of noise is mixed in the output signal of the low-frequency band, and the existence of this noise directly affects the accurate measurement of the low-frequency angular velocity information. The system extracts the low-frequency band angular velocity information in the input signal through the MEMS gyroscope, the MHD angular velocity transmission measures the information of the high-frequency band angular velocity, and finally applies the adaptive strong-tracking Kalman filter fusion algorithm to derive the fused angular velocity value[13].

The adaptive strong tracking Kalman filter fusion model adds an adaptive asymptotic

cancellation factor to the state covariance of the standard Kalman filter fusion model, which increases the influence of the present data on the state covariance, thus making the state covariance more dependent on the present data, and realizes real-time correction of state covariance in the Kalman filter fusion algorithm, and establishes the adaptive strong tracking Kalman filter[14].

$$
\begin{cases}
\dot{\mathbf{X}}_L = \mathbf{A}_L X_L + \mathbf{B}_L \mathbf{x}_{true} + \boldsymbol{\eta}_L \\
\mathbf{Z}_L = \mathbf{C}_L X_L + \mathbf{v}_L
\end{cases}
\tag{1}
$$

$$
\begin{cases}
d_k = \dfrac{1}{k} \\
E_k = Z_k - CX_{k,k-1} \\
R_k = (1-d_k)R_{k-1} + d_k \left(E_k E_k^T - CP_{k,k-1}C^T \right) \\
Q_k = (1-d_k)Q_{k-1} + d_k \left(K_k E_k E_k^T K_k^T + P_k - AP_k A^T \right)
\end{cases}
\tag{2}
$$

$$
\mathbf{P}_{k|k-1} = \lambda_{k-1} \mathbf{A}_{k,k-1} \mathbf{P}_{k-1} \mathbf{A}_{k,k-1}^T + \mathbf{Q}_{k-1}
\tag{3}
$$

where $\lambda_k \geq 1$ is denotes the adaptive fading factor. It can be calculated according to the following algorithm:

where, $\quad \lambda_k = \begin{cases} \lambda_{0,k}, \lambda_{0,k} \geq 1 \\ 1, \lambda_{0,k} < 1 \end{cases}$, $\quad \lambda_{0,k} = \dfrac{Tr(\mathbf{N}_k)}{Tr(\mathbf{M}_k)}$, $\quad \mathbf{N}_k = \mathbf{V}_k - \mathbf{Q}_{k-1} - j\mathbf{R}$, $\quad y(k) = \mathbf{Z}_k - \mathbf{X}_k$

$$
\mathbf{V}_k = \begin{cases} y^2(1), k = 1 \\ \dfrac{\rho \mathbf{V}_k + y^2(k)}{1+\rho}, k \geq 2 \end{cases} \quad M_k = A_{k,k-1}^2 P_{k-1}
$$

The variables in the formula are: $0 < \rho \leq 1$ for the forgetting factor, generally taken $j \geq 1$ for the weakening factor, the value of which is specifically adjusted according to the test, the formulas are combined as an adaptive strong-tracking Kalman filter fusion algorithm, and the strong-tracking factor is adjusted to the state covariance at all times, and the fusion error is smaller , and the results are more accurate.

2.2 Attitude Resolution

In response to the increase in error accumulation that practically leads to the iterative process of conventional equivalent rotation vectors, a conic error compensation algorithm based on gyro frequency-domain features is used that we propose. The method constructs the angular velocity input function matching the gyro frequency characteristics, and then calculates the difference between the amplitude-frequency rotation vector increment and the ideal rotation vector, and solves the cone coefficients by expanding the eigenfunctions into Taylor series in the frequency domain, so that the coefficients of the frequency domain characteristic terms correspond to the equal, and

the compensation coefficients can be regarded as the corrected compensation amount of the ideal coefficients after taking the frequency-domain characteristics into account. Simulation results show that the proposed amplitude-frequency feature compensation algorithm reduces the drift rate of the taper axis from $10^{-5}°/h$ to $10^{-11}°/h$. This algorithm has a higher computational accuracy.

3 Experiments

3.1 Adaptive strong tracking Kalman filter fusion experiments

In the use of TES-3V low-frequency vibration rotary table for the test, the USB-6361 data acquisition card produced by National Instruments (NI) was selected to collect the angular velocity signal. This data acquisition card is connected and controlled through the USB interface of the computer, and the acquisition frequency can be set flexibly. In the experiment, in order to analyze the vibration characteristics at different frequencies, frequency sweeping experiments were conducted to cover the frequency range from 0.01 Hz to 50 Hz. The data acquisition card will be responsible for acquiring the voltage signals generated by the vibration rotary table, and these signals are subsequently transferred to the industrial control computer for further processing and analysis. In the frequency sweeping experiment, the signals are selected as low frequency 0.2 Hz, mid-frequency 10 Hz, and high frequency 50 Hz three frequency points for the fusion algorithm verification. The turntable signal was used as the real signal and the experiment yielded the fused signal output, two angular velocity sensor signals and the turntable signal.
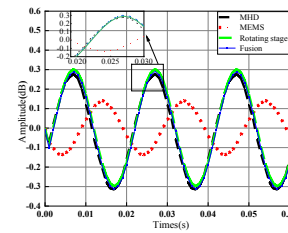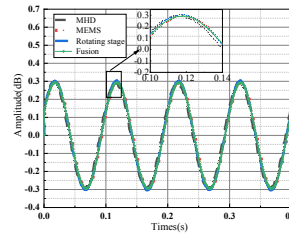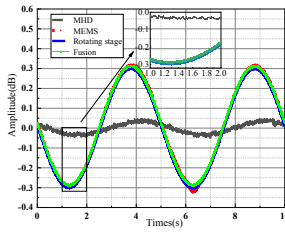
Fig.1 0.2Hz signal fusion results   **Fig.2** 10Hz signal fusion results   **Fig.3** 50Hz signal fusion results

At an input signal of 0.2 Hz, it can be seen from the frequency domain characterization of the MHD angular velocity sensor that the performance of the sensor is poor in the low-frequency range, resulting in the measurement of the signal amplitude obtained is extremely weak, with a large deviation from the real signal, which introduces an obvious measurement error. In contrast, when the same signal is measured using a MEMS gyroscope, the amplitude attenuation of the signal is only 0.02 dB. When the fusion method is used to fuse the signals from the MHD angular velocity sensor and the MEMS gyroscope, the amplitude attenuation of the signal obtained is reduced to 0.014 dB, which indicates that the fusion technique effectively improves the measurement accuracy of the low-frequency signals and reduces the error. In the case where the input signal frequency is 10Hz, both sensors are able to measure the signal accurately as the frequency lies within the effective operating band of both the MHD angular velocity sensor and the MEMS gyroscope. Through detailed data analysis, it is observed that the signals measured by the MHD angular velocity sensor and the MEMS gyroscope, the fused signal and as a real turntable signal, both show a small attenuation in amplitude, with the attenuation level roughly at 0.01 dB. Under the input of 50Hz signal, the MEMS gyroscope in the inertial measurement unit has serious distortion, and there will

be amplitude attenuation and phase deviation is relatively large, which seriously affects the accuracy of the angular velocity signal measurement. The amplitude attenuation of the MHD angular velocity sensor is 0.025dB, and the amplitude attenuation of the fused signal is 0.012dB, and the fused signal is reduced by 0.013dB in amplitude attenuation compared with that of the MHD angular velocity sensor. It proves that the fused signal has a small attenuation of 0.013dB. 0.013dB. It is proved that the fused output can track the turntable signal well in all three frequency bands, and the fusion effect is good.

3.2 Attitude solving experiments

Verify the effect of the angular velocity signal in the inertial measurement unit before fusion and the angular velocity signal after fusion on the attitude solution. The MEMS gyroscope, MHD angular velocity sensor and the fused signal are plotted against the turntable signal to calculate the magnitude of the attitude error angle, respectively. The corresponding magnitude of the attitude error angle is shown in Fig. 4.
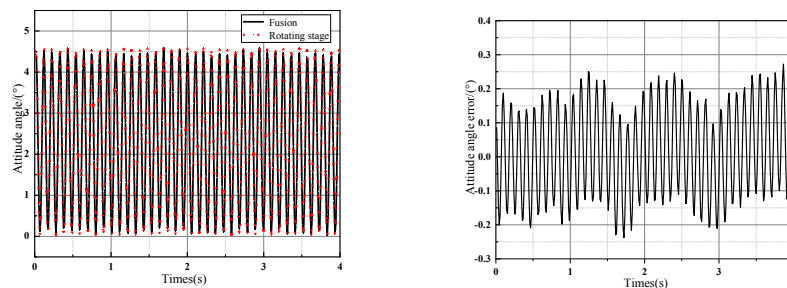
Fig.4 10Hz Fusion Signal and Turntable Signal Attitude Resolution Comparison

Fig.5 Errors in 10Hz fused signal and attitude solving

Under the frequency of 10 Hz test, since the fusion output is the result of Kalman filter fusion of MHD angular velocity sensor and MEMS gyroscope, it is concluded by comparing the results shown in Fig. 4 and Fig. 5 that the maximum MEMS attitude angle solving error is 0.285°, the maximum MHD angular velocity sensor attitude angle solving error is 0.396°, and the maximum post-fusion attitude angle solving error is 0.25°, and the fused attitude angle solution error is reduced by 0.035° compared to the MEMS gyroscope[15].

6 Conclusion

In this paper, we have used a dual sensor hybrid system in a wide bandwidth virtual gyroscope. The proposed ASTKF is used for sensor fusion to improve the stability margin. The intrinsic forgetting factor is reconstructed to enhance the effect of the latest measured data. Then, the covariance of the recursive state-space equations is used to estimate R. To further validate the estimation, the tracking effect is evaluated based on the attitude solution. Finally, the experimental results demonstrate that ASTKF-based sensor fusion can effectively improve the performance of a wide-band virtual gyroscope. When Q and R are estimated simultaneously, there is a coupling relationship between them. Therefore, if there are outliers in the observations, the estimation accuracy will be affected. How to design the strap down inertial measurement unit, solve the cross-coupling error, and improve the measurement accuracy of the three-axis inertial device is the next

research direction.

References:

[1] Ji Yue,Yan Guozhong,Li Xingfei, et al. Research on low frequency extension design of the magnetohydrodynamic angular velocity sensor with multiple magnetohydrodymamic pump[J]. Chinese Journal of Scientific Instrument, 2021, 42(11): 54-61.

[2] Ji Y, Li X, Wu T, et al. Theoretical and Experimental Study of Radial Velocity Generation for Extending Bandwidth of Magnetohydrodynamic Angular Rate Sensor at Low Frequency[J]. Sensors, 2015, 15(12): 31606-31619.

[3] High Sensitivity MEMS Piezoresistive Pressure Sensor[D]. Chinese Master's Theses Full-text Database, 2024.

[4] JI Yue,ZHANG Yan-peng,LI Xing-fei, et al. Research on Solid Liquid Contact Resistance Characteristics of Magnetohydrodynamic Angular Velocity Sensor[J]. Navigation Positioning and Timing, 2022, 9(03): 153-158.

[5] Li Xingfei,Zhou Xinli,Wu Tengfei, et al. A Method of Signal Fusion for Combination Measurement System of MHD Angular Rate Sensor and MEMS Gyro[J]. Journal of Tianjin University(Science and Technology), 2018, 51(02): 159-166.

[6] ZHANG Shaoqiang,LI Xingfei,WU Tengfei, et al. Error Analysis in Low Frequency Region and Correction Method of MHD Angular Rate Sensor[J]. Chinese Journal of Sensors and Actuators, 2015, 28(11): 1634-1639.

[7] YANG Le,WU Tengfei,JI Yue, et al. Cross High-Order Spectrum Method of Frequency Estimation for Weak Signal of MHD Angular Rate Sensor[J]. Chinese Journal of Sensors and Actuators, 2019, 32(02): 217-224.

[8] LIU Chun,WANG Zhining,DAI Lei, et al. Application of Improved Allan Variance Adaptive Filtering in Attitude Solution[J]. Chinese Journal of Sensors and Actuators, 2020, 33(05): 682-687.

[9] Qi Shuai,Jia Jichao,Kou Demin, et al. An Improved Adaptive Filter Algorithm of SINS/GNSS Navigation Based on Allan Variance[J]. Chinese Journal of Sensors and Actuators, 2024, 37(05): 818-824.

[10] ZHOU Xian-lin,ZHANG Hui-jun,HE Tao, et al. Research on adaptive Kalman filter algorithm for GPS/INS loosely coupled integrated navigation[J]. Journal of Time and Frequency, 2020, 43(03): 222-230.

[11] Beijing University of Chemical Technology Beijing 100029 China) ZHAO LiQiang LUO DaCan WANG JianLin YU Tao(College of Information Science and Technology. An adaptive strong tracking cubature Kalman filter[J]. Journal of Beijing University of Chemical Technology(Natural Science Edition), 2013, 40(03): 98-103.

[12] Zhou Xinli,Li Xingfei,Wu Tengfei, et al. Compensation Algorithm of MHD Angular Rate Sensor at Low Frequency Based on Low-Pass Filter[J]. Nanotechnology and Precision Engineering, 2018, 1(01): 79-85.

[13] Ji Y, Du Y, Yan G, et al. Adaptive Kalman Filter Enhanced With Spectrum Analysis for Wide-Bandwidth Angular Velocity Estimation Fusion[J]. IEEE Sensors Journal, 2020, 20(19): 11527-11536.

[14] LIU Yicai,FAN Zhixian,WANG Xiangyu, et al. Road Bank Estimation Based on Curve Adaptive Strong Tracking Kalman Filter[J]. Journal of Tongji University(Natural Science), 2021, 49(S1): 148-154.

[15] YANG Lu,LI Xingfei,JI Yue, et al. Signal De-Noising Method of Eddy Current Sensor Based on EMD-DFA-NLM[J]. Chinese Journal of Sensors and Actuators, 2019, 32(09): 1318-1323.

Adaptive Zero-velocity Detection for Pedestrian Navigation Based on Transformer Model Gait Classification

Qian Yunong, Xiong Zhi[1], Li Xiaodong, Cui Yan, Pan Jie

(College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210096, China)

Abstract: Aiming at the problem of poor positioning accuracy of traditional fixed-threshold zero-velocity correction algorithms for pedestrians in mixed gaits, an adaptive zero-velocity thresholding method for pedestrian navigation based on the gait classification of the Transformer model is proposed. Firstly, the raw pedestrian foot data are classified by the Transformer model. Then, the thresholds are adjusted for different gaits to detect more accurate zero-velocity intervals. Finally, the zero-velocity correction is performed by Kalman filtering, and the final localization results are output. The experimental results show that the overall average recognition rate of this method for the four gaits of walking, running, going upstairs, and going downstairs has reached 92.91%, and the 2D spatial localization accuracy of pedestrians in a wide range of multi-movement gaits reaches 3.56% of the total mileage. Our method effectively solves the problem of low detection accuracy and poor localization performance of fixed-threshold zero-velocity detection in complex gaits.

Keywords: inertial navigation; zero-velocity detection; Transformer; gait classification

1 Introduction

The current pedestrian dead reckoning (PDR) methods can estimate the pedestrian's walking trajectory and other information by measuring and counting the pedestrian's steps, step length, heading, etc. But it only applies to the walking gait, and has a large error for running and walking up and down stairs across floors. To reduce the error accumulated by inertial devices over time, the zero-velocity correction method is currently one of the main techniques for the strapdown inertial navigation system (SINS) to improve navigation and positioning accuracy.

In 2010, Skog *et al.*[1] proposed a zero-velocity detector based on generalized likelihood ratio detection, which improves the localization accuracy of pedestrians under low dynamics. However, the diversity of the pedestrian's gaits affects the accuracy of the zero-velocity detection. To adapt to zero-velocity interval detection under complex gait conditions, researchers have continuously improved the classical zero-velocity detection methods, and in recent years, with the development of artificial intelligence algorithms, machine learning or deep learning algorithms have also been applied to zero-velocity detectors, such as the support vector machine adaptive detector[2], the histogram gradient-enhanced detector[3], and random forest detector[4], etc..

In summary, this paper proposes a pedestrian dead reckoning method based on the Transformer model with multi-gait classification and adaptive threshold zero-velocity detection, which inputs

---

1 Corresponding author, XIONG Zhi: xiongzhi@nuaa.edu.cn.

the inertial data into the trained Transformer model for training and classification, and adjusts the thresholds under the gaits of walking, running, and walking up and down stairs to determine the zero-velocity intervals under different gaits. It is experimentally proved that this method improves the deficiency of the traditional fixed-threshold zero-velocity detection method and can effectively improve the accuracy of the zero-velocity correction algorithm.
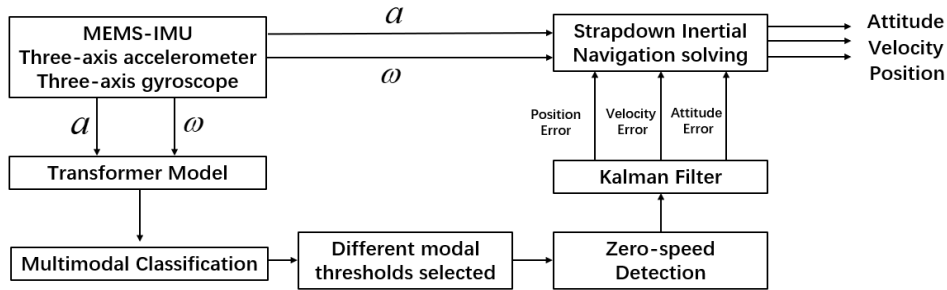
Fig.1   Overall technical flow chart

2 Gait Classification Method Based on the Transformer Model

The inputs to the Transformer model are six-dimensional data, which are the three-axis acceleration of the accelerometer and the three-axis angular velocity of the gyroscope. The Transformer model consists of an input layer, encoder, decoder, and output layer, in which the multi-head attention mechanism is extended to multiple attention heads, each of which can learn different attention weights to better capture different types of relationships. The Transformer model constructed in this paper has 4 multi-head attention heads and 4 Transformer layers. The maximum iteration period for training this model is 150 times. The structure of the model is shown in Figure 2.
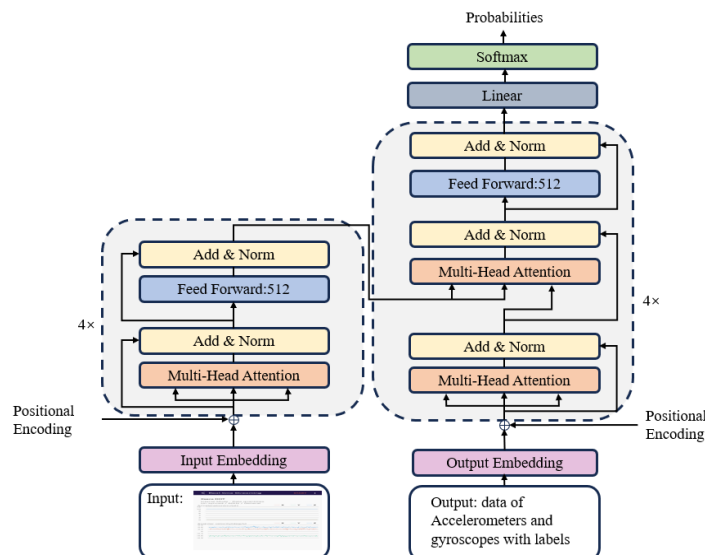
Fig.2   Transformer model structure

3 Adaptive Threshold Zero-velocity Correction for Pedestrian Localization Based on Gait Classification

3.1 Zero-velocity Interval Detection

On the basis of accurate gait classification, it is analyzed that the three-axis acceleration and angular velocity are relatively smooth and close to zero when the foot is in actual contact with the ground. The traditional zero speed detection method is through the three-axis accelerometer modal value a_1 and three-axis standard deviation σ_1, the gyroscope three-axis modal value g_1 and three-axis standard deviation σ_2.

Comprehensive use of three-axis accelerometer and gyroscope output characteristics, the traditional zero-velocity interval detection for mixed gait only a set of thresholds, so there is a leakage of false detection, this paper combines the basis of gait classification for each modality to adjust a set of thresholds most suitable for this modality.

$$\lambda_{up} = \begin{cases} 1 & (a_1 < \varepsilon_{a1,up}) \ \& \&(\sigma_1 < \varepsilon_{a2,up}) \ \& \&(g_1 < \varepsilon_{\omega1,up}) \ \& \&(\sigma_2 < \varepsilon_{\omega2,up}) \\ 0 & others \end{cases} \qquad (1)$$

$$\lambda_{run} = \begin{cases} 1 & (g_1 < \varepsilon_{\omega1,run}) \ || \ (\sigma_2 < \varepsilon_{\omega2,run}) \\ 0 & others \end{cases} \qquad (2)$$

$$\lambda_{walk} = \begin{cases} 1 & (a_1 < \varepsilon_{a1,walk}) \ \& \&(\sigma_1 < \varepsilon_{a2,walk}) \ \& \&(g_1 < \varepsilon_{\omega1,walk}) \ \& \&(\sigma_2 < \varepsilon_{\omega2,walk}) \\ 0 & others \end{cases} \qquad (3)$$

Where, a_1 and σ_1 are the accelerometer three-axis modulus and standard deviation, g_1 and σ_2 are the gyroscope three-axis modulus and standard deviation, and the zero speed interval is successfully detected when the zero speed detection flag is 1. Where, $\varepsilon_{a1,up}, \varepsilon_{a2,up}, \varepsilon_{\omega1,up}, \varepsilon_{\omega2,up}$ are thresholds for walking up and down stairs, $\varepsilon_{\omega1,run}, \varepsilon_{\omega2,run}$ are thresholds for running, $\varepsilon_{a1,walk}, \varepsilon_{a2,walk}, \varepsilon_{\omega1,walk}, \varepsilon_{\omega2,walk}$ are thresholds for walking. Threshold parameter determination is obtained by comparing and analyzing the accelerometer three-axis mode value curve and three-axis standard deviation curve, the gyroscope three-axis mode value curve and three-axis standard deviation curve with the corresponding cycle of the human walking zero-velocity intervals based on the accelerometer three-axis mode value curve and three-axis standard deviation curve from multiple sets of experiments.

3.2 Zero Speed Correction Algorithm

The zero-velocity update algorithm (ZUPT) used in this paper takes the three-dimensional velocity as the observed quantity of the system, and uses Kalman filtering for estimating the 9-dimensional state error. In this paper, the system is modeled with the error quantities of the nine basic parameters of navigation as the state quantities.

$$X = [\phi_E \quad \phi_N \quad \phi_U \quad \delta v_E \quad \delta v_N \quad \delta v_U \quad \delta L \quad \delta\lambda \quad \delta h]^T \qquad (4)$$

where, ϕ is the platform error angle, δv is the velocity error, $\delta L, \delta\lambda, \delta h$ is the longitude, latitude and altitude error, indicating the three directions of the 'North-East-Up'.

State modeling is formed as:

$$\dot{X}(t) = F(t)X(t) + G(t)W(t) \qquad (5)$$

where $X = [\phi_E \quad \phi_N \quad \phi_U \quad \delta v_E \quad \delta v_N \quad \delta v_U \quad \delta L \quad \delta\lambda \quad \delta h]^T$ is the system state vector, F

is the state transfer matrix, G is the system noise matrix, and W is the white noise random error.

With zero velocity as the virtual observation, the measurement model is formed as:

$$Z = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} v_E \\ v_N \\ v_U \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = HX + V \tag{6}$$

where H is the observation matrix, $H = [0_{1\times3} \ \ 1 \ \ 1 \ \ 1 \ \ 0_{1\times3}]^T$, $V = [\delta v_E \ \ \delta v_N \ \ \delta v_U]^T$ is the observation noise matrix.

4 Experimental Results and Analysis

4.1 Analysis of the Transformer Model Classification Results

4.1.1 Data Collection and Evaluation of Classification Results

In daily life, in the indoor environment of large buildings, the main motion states involved in pedestrians are standing still, walking, running, going upstairs, and going downstairs[5].

The sensors used in this paper are IMU sensors (Xsens Dot), whose sampling data include three-axis acceleration, three-axis angular velocity information, with a sampling frequency of 60Hz, as shown in Figure 3. According to the analysis of human kinematics and biological characteristics, as well as the portability of the IMU installation, the back of the foot was finally chosen as the site of data acquisition[6]. The experimenters performed the movement with walking, running, going up and down the stairs. A total of 7755 data samples were collected, and the number of samples for each motion state is shown in Table 1.

Tab.1   Sample number of dataset

Gaits of movement	Label	Sample number
walk	0	2342
run	1	1982
upstairs	2	1883
downstairs	3	1548

Data Acquisition Interface

Foot-mounted IMU

Fig.3   IMU Wearing Styles and Data Acquisition Interface

The CNN[7] model, and the LSTM[8] model in this section are trained using the dataset taken in Section 2.1, and the results are compared with the Transformer model in this paper. The resulting

confusion matrix is shown in Figure 4. The accuracy of each training model is shown in Table 2.
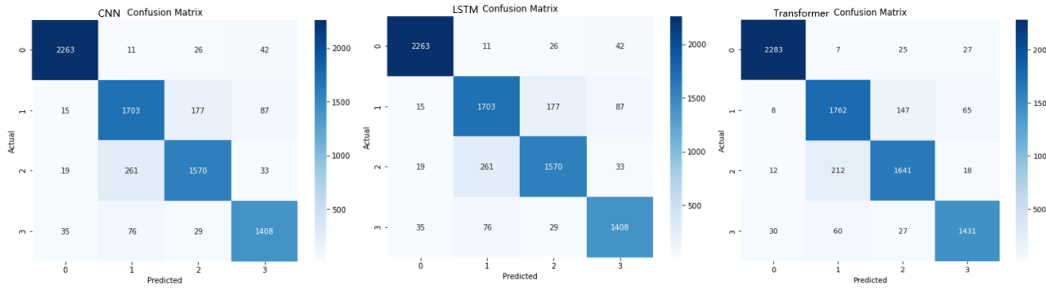
Fig.4 Confusion matrixes for each deep learning model

Tab.2 Classification accuracy rate of each deep learning model

Gaits of movement	CNN	LSTM	**Transformer**
walk	98%	97%	**98%**
run	86%	83%	**86%**
upstairs	87%	87%	**89%**
downstairs	92%	90%	**93%**
average accuracy rate	91.15%	91.36%	**92.91%**

As can be seen from Table 2, the three deep learning models constructed have better accuracy for walking and going downstairs, but poorer accuracy for running and going upstairs. The Transformer model has the highest classification accuracy in all motion gaits, so the Transformer model is used for subsequent predictions.

4.1.2 Analysis of Integrated Gait Prediction Results

The experimental site is chosen in the laboratory building of the school, and the starting point of the movement is chosen at the stairway on the 1st floor of the laboratory building. Firstly, we walk a 56m corridor from the stairway on the 1st floor to the west, and stop for 10s at the other end of the corridor, and then switch the gait to go upstairs, go up to the 2nd floor, and stop for 10s, and then run through the corridor of the same length to the east, and then run to the stairway, and stop for 10s, and then go down the stairs to return to the origin point finally. The classification results obtained from the prediction of the Transformer model trained in Section 2.2 are shown in Figure 5.

In Figure 5, the blue line is the Y-axis angle data, and the red line is the classification result. The label 4 is given to the long-time stationary state. The experimental results show that excluding the misjudged labels within 20 sampling points from the predicted label sequence gives reliable gait prediction results. The gait classification method based on the deep learning model is effectively verified.
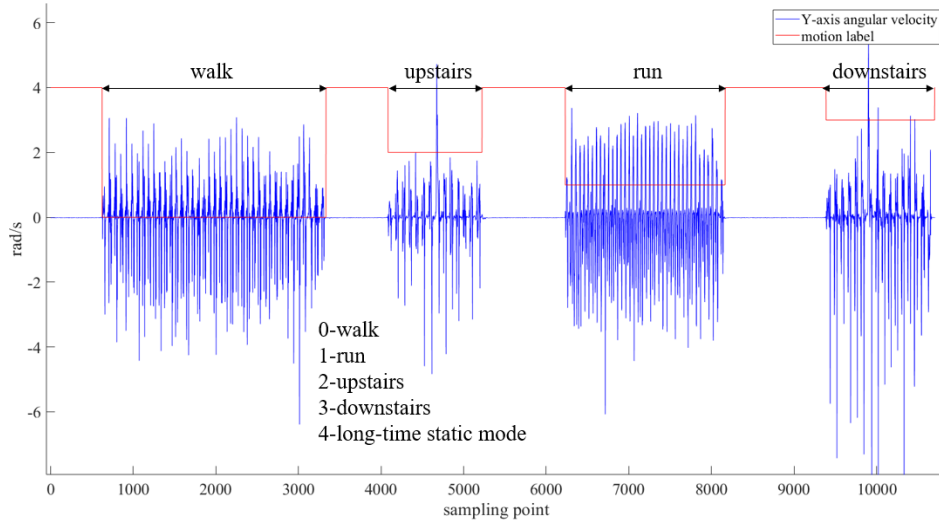
Fig.5 Gait classification result

4.2 Analysis of Zero Velocity Detection Results for Different Gaits

The zero-velocity detection threshold parameters for different gaits are adjusted according to the data taken from the experiments as shown in Table 3 below, and ε_{a1}, ε_{a2}, $\varepsilon_{\omega1}$, $\varepsilon_{\omega2}$ are the thresholds for acceleration modulus, acceleration standard deviation, angular velocity modulus, and angular velocity standard deviation, respectively.

Tab.3   Zero-velocity detection thresholds for different gaits

Gaits of motion	$\varepsilon_{a1}(m\ /\ s^2)$	$\varepsilon_{a2}(m\ /\ s^2)$	$\varepsilon_{\omega1}(rad\ /\ s)$	$\varepsilon_{\omega2}(rad\ /\ s)$
walk	12.54	5.69	0.90	1.22
run	—	—	0.57	0.78
upstairs	10.03	4.55	0.72	0.98
downstairs	10.03	4.55	0.72	0.98

The experimental results are shown in Figure 6. It can be seen that the pedestrian is walking in 41 steps, going upstairs in 14 steps, running in 38 steps, and going downstairs in 16 steps. Our method based on the Transformer model gait classification has a 100% detection rate in all kinds of gait. With a fixed threshold for detection, normal walking, going up and down the stairs can be detected efficiently, but the zero-velocity interval can not be detected during running gaits.
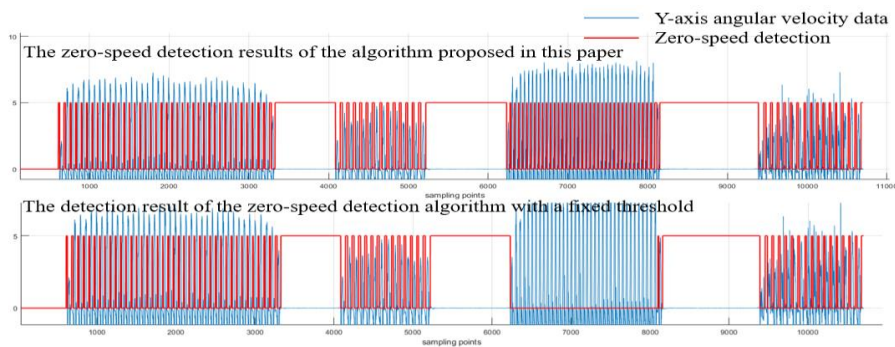
Fig.6 Zero-velocity detection result

29

4.3 Analysis of Zero-velocity Correction Algorithm Localization Results

Figure 7 shows the classification results of the four motion states of the adaptive zero-velocity detection algorithm based on the Transformer model classification and the pedestrian trajectory map after the zero-velocity correction by Kalman filtering. The green dot is the starting point, and the blue dot is the ending point. The black dots are the truth points of the pedestrians on their way. It can be seen in Figures (b) and (c) that the fixed-threshold zero-velocity detection method for walking up and down stairs cannot detect the zero-velocity interval in the running state, so the trajectory of the pedestrian is rapidly dispersed when running, and the fixed-threshold zero-velocity detection method for walking has omission and misdetection for zero-velocity detection of walking up and down stairs.

The experimental results show that through the adaptive zero-velocity detection method based on the Transformer model gait classification proposed in this paper, the localization results in the horizontal two-dimensional space start and end point localization error is 4.21m, which is 3.56% of the total mileage. The vertical direction localization error is 0.78m, which is 23.5% of the total mileage, which effectively avoids the situation that a certain motion state can not be detected and the trajectory is drifting under a single fixed threshold value.
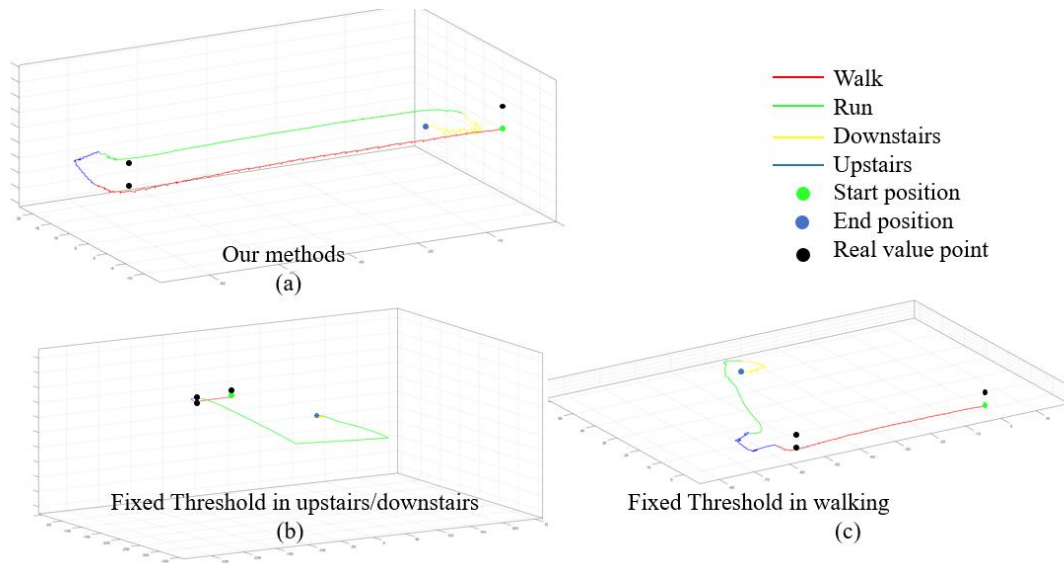
Fig.6   Three-dimensional motion trajectory display

4 Conclusion

In this paper, for the problem of poor positioning accuracy of traditional fixed-threshold zero-velocity correction algorithms for pedestrians in mixed motion gaits, we propose an adaptive zero-velocity thresholding method for pedestrian navigation based on gait classification of the Transformer model. The raw pedestrian foot data collected are classified by the Transformer model, and then the thresholds are adjusted for different classification gaits, to detect more accurate zero-velocity intervals, and the zero-velocity correction is performed by Kalman filtering, and the final localization results are output. The experimental results show that the overall average recognition rate of this method for the four gaits of walking, running, going upstairs and going downstairs has reached 92.91%, and the 2D spatial localization accuracy of pedestrians in a wide range of multi-

movement gaits reaches 3.56% of the total mileage. It is experimentally proved that this method improves the deficiency of the traditional fixed-threshold zero-velocity detection method and can effectively improve the accuracy of the zero-velocity correction algorithm.

References:

[1]   Skog, I Handel, P Nilsson, J-O, et al. Zero-velocity detection—an algorithm evaluation[J]. IEEE Transactions on Biomedical Engineering, 2010, 57(11):2657-2666.

[2]   Yang M, Zhu R, Xiao Z, et al. Symmetrical-Net: Adaptive Zero Velocity Detection for ZUPT-Aided Pedestrian Navigation System[J]. IEEE Sensors Journal, 2021, PP(99):1-1.DOI:10.1109/JSEN.2021.3094301.

[3]   Zhi Nong, Miao Hongsheng, Li Haijun, et al. A pedestrian navigation algorithm based on adaptive zero-velocity detection using SVM[J].Journal of Chinese Inertial Technology,2021,29(5):604-610

[4]   Fei Chengyu, Su Zhong, Li Qing. Zero velocity detection algorithm for pedestrian inertial navigation[J].Sensors and Microsystems,2016,35(3): 147-150

[5]   Niu X, Liu T, Kuang J, et al. Pedestrian trajectory estimation based on foot-mounted inertial navigation system for multistory buildings in postprocessing mode[J]. IEEE Internet of Things Journal, 2021, 9(9): 6879-6892.

[6]   Zhang L, Wu J, Jiang C, et al. Learning-based stance-phase detection for a pedestrian dead-reckoning system with dynamic gait speeds[J]. Measurement Science and Technology, 2021, 32(10): 105108.

[7]   Ashry S, Ogawa T, Gomaa W. CHARM-deep: Continuous human activity recognition model based on deep neural network using IMU sensors of smartwatch[J]. IEEE Sensors Journal, 2020, 20(15): 8757-8770.

[8]   Chen Z, Pan X, Wu M, et al. A novel adaptive zero-velocity detector for inertial pedestrian navigation based on optimal interval estimation[J]. IEEE Access, 2020, 8: 191888-191900.

Pedestrian Inertial Navigation Algorithm with Learnable Time-Series State-Space-Model

Zheming Tu, Xianfei Pan, Wenqi Wu, Changhao Chen[*]

(College of Intelligence Science and Technology, National University of Defense Technology,
Changsha, 410073, China)

Abstract: The deep learning-based inertial navigation method has emerged as a prominent research focus due to its ability to eliminate cumulative errors and leverage learned experiences from data. However, existing approaches often rely on RNNs or CNNs, which do not fully exploit the temporal characteristics of long inertial sequences. To address this limitation and further mitigate error divergence in learning-based micro inertial navigation systems, we propose a novel inertial navigation algorithm featuring a learnable time-series state space model. This model incorporates a selective implicit state space framework to estimate motion displacement and uncertainty within a fixed time window. The estimated displacement serves as the observation input for an extended Kalman filter algorithm to evaluate the state of the inertial navigation system. Experiments conducted on wearable devices for pedestrian navigation demonstrate that our method enhances the accuracy of learned displacement estimation from inertial data, effectively suppresses error drift in the inertial system, and achieves reliable pedestrian navigation.

Key words: Inertial navigation; Autonomous navigation; Deep Learning; Pedestrian Navigation; Time-Series State-Space-Model

1 Introduction

Pedestrian navigation is essential in various application scenarios, including emergency rescue, supermarket shopping, and augmented reality. Recent advancements in Micro Inertial Measurement Units (MIMUs) have enabled miniaturization, reduced power consumption, and lower costs. However, the inherent noise in MIMUs can lead to rapid divergence in strapdown navigation algorithms due to the double integration of this noise, thereby limiting their application in navigation tasks.

To address this issue, increasing attention has been given to utilizing pedestrian motion states as observation information to mitigate navigation error divergence. Recent research in Pedestrian Dead Reckoning (PDR) technology has primarily focused on fusing PDR with external information sources, such as GNSS[1], UWB positioning[2], and barometers or magnetometers[3]. However, these methods often rely on precise parameter tuning and constraints on pedestrian motion states, which can be difficult to achieve in complex motion scenarios requiring high-precision navigation. Additionally, sensor noise in micro-inertial devices poses challenges in accurate modeling and exhibits poor repeatability after power-up, resulting in rapid error divergence in traditional physics-based navigation methods.

Consequently, adopting a data-driven approach to enhance pedestrian navigation algorithms

has become a crucial research direction. IONet[4] first introduced a learning framework based on LSTM to estimate relative pose and motion displacement in polar coordinates from segmented inertial data windows. Similarly, the RoNIN[5] method employs residual networks for motion estimation and positioning using IMU sensor data, but its reliance on direction estimation limits its effectiveness. TLIO[6] utilizes ResNet to learn three-dimensional displacement and covariance from gravity-aligned inertial data sequences, integrating these as observation states in an extended Kalman filter. However, these approaches have not fully explored the temporal information inherent in inertial data, leaving room for improvements in positioning accuracy.

In recent years, deep learning has shown considerable promise in time-series signal processing, with common architectures including RNNs (e.g., LSTM), convolutional networks (e.g., WaveNet[7], Temporal Convolutional Networks[8]), and attention-based models (e.g., Transformers[9]). Notably, the selective state space model Mamba[10] has been successfully applied across various domains, such as speech processing[11], demonstrating superior feature learning and time-series modeling capabilities. Given that inertial data represents long sequence information, it is particularly well-suited for learnable state space models. However, the potential for modeling temporal motion characteristics using state space models remains an area ripe for exploration.

To further suppress error divergence in learning-based micro inertial navigation systems, we propose a novel inertial navigation algorithm that incorporates a learnable time-series state space model. This selective implicit state space model captures temporal inertial data, estimates motion displacement and uncertainty within a fixed time window, and utilizes the estimated displacement as observation input for an extended Kalman filter algorithm to evaluate the state of the inertial navigation system. By leveraging motion background features derived from the deep learning model, our approach effectively mitigates the divergence of the inertial navigation algorithm. Experimental results from pedestrian navigation wearable devices demonstrate that our method significantly enhances displacement estimation accuracy from inertial data and effectively suppresses error drift. Our approach surpasses existing state-of-the-art learning-based models, such as RoNIN and TLIO, achieving an average absolute trajectory error of 0.87 meters over tests with an average pedestrian working distance of approximately 130 meters.

2　Method

In response to the insufficient exploration of temporal information and the limited positioning accuracy in current deep learning-based inertial navigation methods, we propose an inertial navigation algorithm that leverages a learnable temporal state model to enhance the positioning accuracy of pedestrian inertial navigation systems. This algorithm includes a displacement estimation module based on the temporal state learning model. The overall architecture of the algorithm is illustrated in Figure 1 and consists of two main components: 1) a displacement estimation neural network model grounded in the temporal state learning framework, and 2) a displacement learning extended Kalman filter.

The displacement estimation neural network model takes the raw IMU measurement data sequence as input and produces both displacement estimates and associated uncertainty for that sequence. This model primarily comprises three key elements: feature encoding, temporal modeling, and feature decoding. The temporal modeling module is constructed using the temporal state learning model, which employs a learnable state space approach to effectively capture and represent

the temporal information inherent in the data.
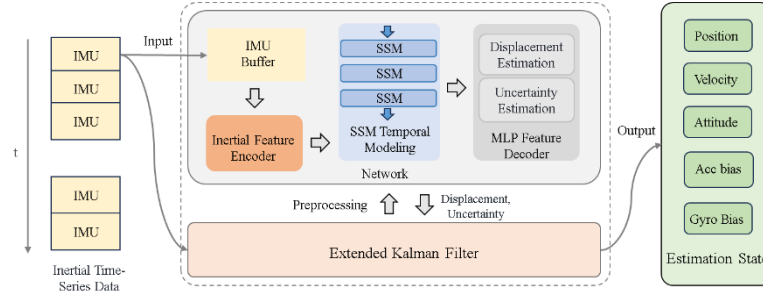
Fig.1 Inertial Navigation Algorithm Framework for Temporal Learning State Space Models

2.1 Time-Series State-Space-Model

To further enhance the accuracy of neural network-based displacement estimation, the inertial navigation state space learning model utilizes a state space model (SSM) to capture temporal information. The fundamental structure of the state space model can be represented in three forms: continuous-time, recursive, and convolution representations. The state space model maps the inertial motion feature vector $\mathbf{U}_t$ extracted from the neural network to an N-dimensional variable $\mathbf{X}_t$ in the implicit state space through Equation (1), and then projects it onto the output variable $\mathbf{y}_t$, with the calculation process defined as follows:

$$\mathbf{X}_t = \overline{\mathbf{A}}\mathbf{X}_{t-1} + \overline{\mathbf{B}}\mathbf{U}_t$$
$$\mathbf{y}_t = \mathbf{C}\mathbf{X}_t$$

(4)

Where, $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ is the discrete state matrix, input matrix. $\mathbf{C}$ is the output matrix. The model can perform inference and prediction in the form of a linear recursive loop as expressed in Equation (3).
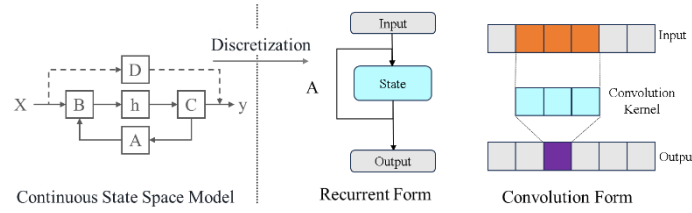
Fig. 2 State Space Model Network Representation

To make it suitable for parallelized training processes, the multiplier can be regarded as a convolution kernel $\overline{\mathbf{K}} = \mathbf{C}\overline{\mathbf{B}} \cdot (\overline{\mathbf{A}}^0, \overline{\mathbf{A}}^1, \overline{\mathbf{A}}^2, ..., \overline{\mathbf{A}}^L)$, allowing for parallelized training in the form of convolution.

2.2 Learning Location Displacement

The inertial navigation state-space learning model mainly consists of a feature encoding module, a temporal modeling module, and a feature decoding module. The feature encoding module uses the EfficientNet-B0 network structure[12] to perform feature encoding on inertial data and map it to the implicit state space. After Inertial data $\boldsymbol{I} \in \mathbb{R}^{6 \times L}$ undergoes feature encoding, the inertial feature vector $\boldsymbol{U} \in \mathbb{R}^{M \times D}$ in the implicit space is obtained, which is expressed as:

$$U = EfficientNet(\mathbf{I}) \tag{5}$$

Here, L represents the length of the IMU sequence, M is the length of the feature vector after feature encoding, and D is the dimensionality of the feature vector.

The temporal modeling module employs a learnable state-space model, constructed based on selective bidirectional state-space model units. Its basic structure is shown in Figure 3, with the utilization of IMU data temporal information enhanced through the stacking of three learnable state-space models. It performs temporal modeling on the feature vector U after feature encoding, which can be expressed as:

$$Y = SSM(U) \tag{6}$$

Finally, a fully connected layer is used for feature decoding, outputting displacement estimation d and uncertainty u, as shown in the following equations:

$$\mathbf{d}, \mathbf{u} = MLP(\mathbf{Y}) \tag{7}$$

During the training process, two types of loss functions are used: root mean square error loss and Gaussian maximum likelihood loss. The root mean square error loss on the training set is defined as follows:

$$\mathcal{L}_{\mathrm{MSE}}(\Delta \mathbf{l}, \Delta \hat{\mathbf{l}}) = \frac{1}{L} \sum_{j=1}^{n} (\Delta \mathbf{l}_j - \Delta \hat{\mathbf{l}}_j)^2 \tag{8}$$

Wherein, $\Delta \hat{\mathbf{l}} = \{\Delta \hat{\mathbf{l}}_j\}_{j \le L}$ is the 3-dimensional displacement output value of the network model, and $\Delta \mathbf{l} = \{\Delta \mathbf{l}_j\}_{j \le L}$ is the corresponding ground truth label.

The maximum likelihood loss is defined as the negative log-likelihood of the displacement according to the regression Gaussian distribution:

$$\mathcal{L}_{\mathrm{ML}}(\Delta \mathbf{l}, \Sigma, \Delta \hat{\mathbf{l}}) = \frac{1}{L} \sum_{j=1}^{L} -\log \left(\frac{1}{\sqrt{8\pi^3 \det(\Sigma_j)}} e^{-\frac{1}{2} \|\Delta \mathbf{l}_j - \Delta \hat{\mathbf{l}}_j\|_{\Sigma_j}^2} \right)$$
$$= \frac{1}{L} \sum_{j=1}^{L} \left(\frac{1}{2} \log \det(\Sigma_j) + \frac{1}{2} \Delta \mathbf{l}_j - \Delta \hat{\mathbf{l}}_{i\Sigma_j}^2 \right) + \mathrm{Cst} \tag{9}$$

Wherein, $\hat{\Sigma} = \{\hat{\Sigma}_j\}_{j \le L}$ is the 3×3 covariance matrix corresponding to the uncertainty $\hat{\mathbf{u}}_j$ of the j-th displacement estimate. $\hat{\Sigma}_j$ has 6 degrees of freedom, and there are various covariance parameters for neural network uncertainty estimation. To enhance numerical stability, the covariance matrix can be simplified to a diagonal matrix parameterized by 3 coefficients, thereby eliminating singular values near the zero points of the loss function, denoted as:

$$\Sigma_j(\mathbf{u}_j) = \mathrm{diag}(e^{2\mathbf{u}_{xj}}, e^{2\mathbf{u}_{yj}}, e^{2\mathbf{u}_{zj}}) \tag{10}$$

The training process uses the Adam optimizer for optimization, with a learning rate of 1e-4. In the first ten epochs of training, the loss function is set to $\mathcal{L}_{\mathrm{MSE}}$, and then it is set to $\mathcal{L}_{\mathrm{ML}}$ until the end of training.

2.3 Extended Kalman Filter for Learning Based Inertial Navigation

The extended Kalman filter is used to fuse the displacement learning results from the neural network model with inertial navigation. The filter maintains a window of k pose vectors to construct relative displacement observations. The navigation state of the system can be expressed as:

$$\mathbf{E} = (\epsilon_1, \ldots, \epsilon_k, \boldsymbol{c}) \tag{11}$$

Wherein, $\epsilon_j = ((\boldsymbol{R}_{iw})_j, \boldsymbol{p}_j), j = 1, \ldots, k$ is the past motion state, which only includes the pose state variables for optimization. $\boldsymbol{c} = (\boldsymbol{R}_{iw}, \boldsymbol{v}, \boldsymbol{p}, \boldsymbol{b}_\omega, \boldsymbol{b}_a)$ is the current motion state, $\boldsymbol{R}_{iw} \in SO(3)$ represents the rotation matrix transforming from the inertial-frame i to the world-frame w, $\boldsymbol{v}, \boldsymbol{p}$ represents the velocity and position of the body in the world coordinate system, and $\boldsymbol{b}_\omega, \boldsymbol{b}_a$ is the bias of the inertial sensor's gyroscope and accelerometer. The error state linear recursive equation can then be expressed as:

$$\tilde{\mathbf{c}}_{k+1} = \mathbf{F}_k^s \tilde{\mathbf{c}}_k + \mathbf{G}_k^s \mathbf{n}_k \tag{12}$$

Where, $\mathbf{n}_k = [\mathbf{n}_{\omega k}, \mathbf{n}_{ak}, \boldsymbol{\eta}_{\omega dk}, \boldsymbol{\eta}_{adk}]$ represents the system noise, $\tilde{\mathbf{c}}_{k+1}$ represents the system state at time k+1, $\mathbf{F}_k^s$ represents the state transition matrix at time k, $\tilde{\mathbf{c}}_k$ represents the system state at time k, and $\mathbf{G}_k^s$ represents the system noise matrix at time k.

The measurement is the three-dimensional displacement which can be expressed as:

$$h(\mathbf{E}) = \mathbf{R}_\gamma^T(j)(\mathbf{p}_k - \mathbf{p}_j) = \mathbf{d}_{jk} + \mathbf{n}_{jk} \tag{13}$$

Where, $\mathbf{d}_{kj}$ is the three-dimensional displacement between states k and j output by the inertial navigation learning state space model, $\mathbf{R}_\gamma^T(k)$ is the rotation matrix corresponding to the heading angle at time k, $\mathbf{n}_{kj}$ is the zero-mean Gaussian white noise of the measurement, whose covariance $\boldsymbol{\Sigma}_j$ is given by the uncertainty estimated by the neural network

3 Experiment

3.1 Dataset

To validate the algorithm, we utilized the public pedestrian navigation inertial dataset[6]. This dataset features the Bosch BMI055 IMU, which has an accelerometer noise density of 150 and a gyroscope bias stability of 1.6. The sensor is mounted on the pedestrian's head, collecting inertial data at a frequency of 200 Hz. The complete dataset comprises over 60 hours of pedestrian movement data, including more than 400 sequences from 5 pedestrians engaged in various activities such as walking, standing still, arranging a kitchen, playing billiards, and ascending or descending stairs. The data collection was conducted using a wearable device, with the MSCKF visual-inertial odometry[13] serving as the reference for ground truth positioning. For training, 80% of the dataset was randomly selected, while 10% was designated for validation and 10% for testing.

3.2 Results

The test dataset includes four groups of pedestrian navigation experiments, each lasting approximately 3 to 7 minutes. Both training and testing were conducted on an Nvidia RTX 3090 computing platform. The experimental results were compared with two deep learning-based pedestrian inertial navigation methods, RoNIN[5] and TLIO[6]. The primary evaluation criteria were the positioning results, assessed through two metrics: Absolute Trajectory Error (ATE) and Drift Rate (DR).
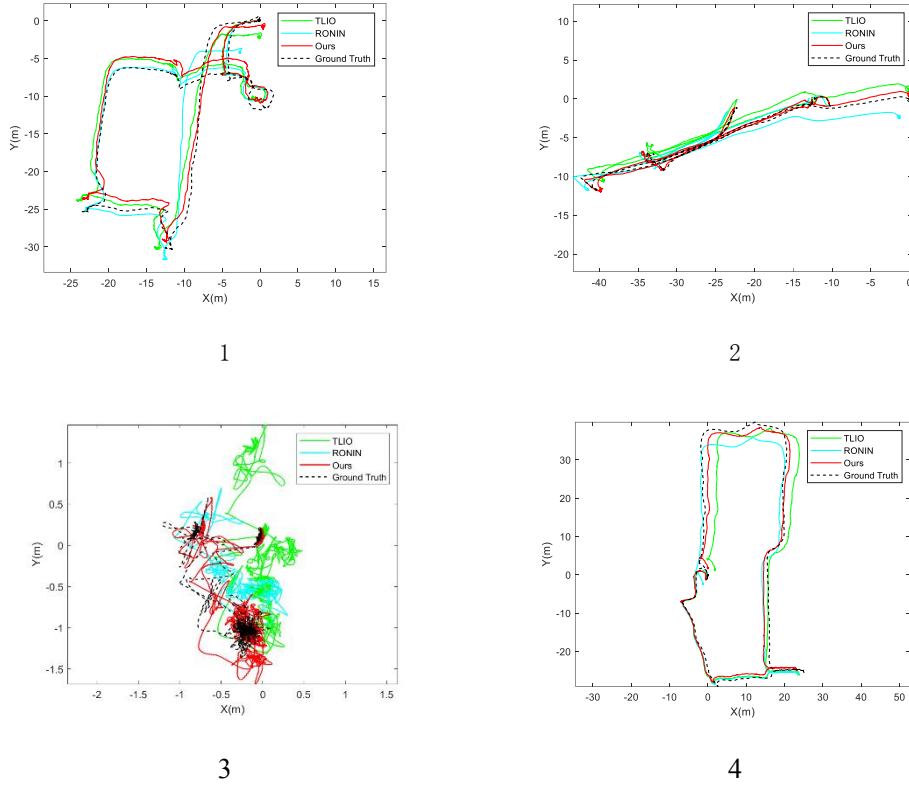
Fig. 4   Comparison of Positioning Results

The positioning results for selected sequences in the test dataset are presented in Figure 4. The results indicate that the pedestrian inertial navigation algorithm based on the temporal state learning model exhibits smaller trajectory drift errors compared to the TLIO and RoNIN methods, providing more accurate position estimations during movement. The mean values of the absolute trajectory error and displacement drift rate from the experiments are summarized in Table 1. Quantitative analysis reveals that the average absolute trajectory error of our proposed algorithm is 0.87 m, with a displacement drift rate of 0.63%. Notably, in Experiment 3, where the pedestrian moved within a confined range, both TLIO and RoNIN algorithms exhibited significant deviations, while our algorithm effectively estimated the pedestrian's movement state.

Tab. 1 Positioning Error Summary in Test Datasets

	Trajectory Length (m)	ATE/m			DR/%		
		RoNIN	TLIO	ours	RoNIN	TLIO	ours
1	141.49	2.34	1.55	1.36	3.51	1.58	0.72
2	59.80	0.58	0.72	0.20	0.59	2.16	0.35
3	213.90	2.59	2.38	1.36	1.23	1.05	0.84
4	129.21	1.38	1.36	0.56	2.08	1.11	0.62
Average	136.1	1.7225	1.5025	0.87	1.8525	1.475	0.63

4   Conclusion

This paper presents a pedestrian inertial navigation algorithm using temporal state learning model. The algorithm employs a selective implicit state space model to capture temporal inertial data, estimating movement displacement and uncertainty within a fixed time window. These displacement estimates are then used as observations in the extended Kalman filter algorithm to assess the state of the inertial navigation system. Experimental validation demonstrates that the positioning accuracy of the odometry model, enhanced through temporal modeling, surpasses that of TLIO. Future work may build on these contributions by focusing on lightweight optimization and acceleration of inference for the network model, thereby facilitating the deployment and application of neural network-based inertial odometry methods in practical systems such as smartphones and VR glasses.

References

[1] Jiang C, Chen Y, Chen C, et al. Smartphone PDR/GNSS Integration via Factor Graph Optimization for Pedestrian Navigation[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 12.

[2] Guo S, Zhang Y, Gui X, et al. An Improved PDR/UWB Integrated System for Indoor Navigation Applications[J]. IEEE Sensors Journal, 2020, 20(14): 8046-8061.

[3] Liu H, Gao Z, Wang L, et al. Reliable Positioning Model of Smartphone Sensors and User Motions Tightly Enhanced PDR[J]. IEEE Internet of Things Journal, 2024, 11(19).

[4] Chen C, Lu X, Markham A, et al. Ionet: Learning to cure the curse of drift in inertial odometry[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2018, 32(1).

[5] Herath S, Yan H, Furukawa Y. RoNIN: Robust Neural Inertial Navigation in the Wild: Benchmark, Evaluations, & New Methods[C]//2020 IEEE International Conference on Robotics and Automation (ICRA). 2020: 3146-3152Paris, France: IEEE, 2020: 3146-3152.

[6] Liu W, Caruso D, Ilg E, et al. TLIO: Tight Learned Inertial Odometry[J]. IEEE Robotics and Automation Letters, 2020, 5(4): 5653-5660.

[7] Oord A Van Den, Dieleman S, Zen H, et al. Wavenet: A generative model for raw audio[J]. arXiv preprint arXiv:1609.03499, 2016.

[8] Bai S, Kolter J Z, Koltun V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling[Z]. arXiv, 2018(2018-04-19).

[9] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[10] Gu A, Dao T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces[Z]. arXiv, 2024(2024-05-31).

[11] Li K, Chen G. SPMamba: State-space model is all you need in speech separation[Z]. arXiv, 2024(2024-04-02).

[12] Tan M, Le Q V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks[Z]. arXiv, 2020(2020-09-11).

[13] Mourikis A I, Roumeliotis S I. A Multi-State Constraint Kalman Filter for Vision-Aided Inertial Navigation[C]//Proc IEEE Int Conf Rob Autom. 2007: 3565-3572.

Failure Detection for A Redundant Sensor Unit

E.V. Shestakova

(Lomonosov Moscow State University, Faculty of Mechanics and Mathematics, Moscow)

Abstract: We solve the problem of failure detection in a redundant unit of angular velocity sensors. Four methods are investigated: the method of "zero" linear combinations, the least squares method, the least absolute deviations method, and the guaranteeing approach.

1 Introduction

To increase the reliability of the navigation system, the redundant inertial sensor units (angular velocity sensors) are used. Instead of the commonly employed three mutually orthogonal sensors, six sensors are used. The failure detection problem in a redundant angular velocity sensor unit is considered. Four methods are studied: the method of "zero" linear combinations, the least squares method, the least absolute deviations method, and the guaranteeing approach.

2 Problem statement

Let the measurements delivered by the angular velocity sensor unit (at a given time instant) have the form

$$z = Gq + \rho \,, \tag{1}$$

where $z = \left(z_1, \cdots, z_6\right)^T \in R^6$, R is a vector of angular velocity sensor unit measurements, G is a given matrix, $q \in R^3$ is a vector to be estimated, it consists of angular velocity components in projections onto the instrumental frame, $\rho = \left(\rho_1, \cdots, \rho_6\right) \in R^6$ is a vector of the measurement error.

Assume two failures at most can occur at a certain time instant. It is required to detect possible failures in the measurements. In other words, we need to detect the presence of failures and to determine which measurement channels they occur in.

We consider several methods for detecting the failures.

3 Zero linear combinations method

When using this method, we assume that no more than one failure can occur in the measurement channels (at a chosen time instant). The essence of the method is as follows. We find vectors x such that the desired signal is excluded:

$$G^T x = 0 \tag{2}$$

This way we achieve "direct access" to the error:

$$x^T z = x^T \varrho.$$

By solving the system (2), we obtain six linear combinations $x[1],\ldots,x[6]$. Each of them contains exactly one zero entry. To detect a singular failure, we need to calculate the six inner products of z

and $x[i]$. The failure occurs in the i-th channel if and only if the combination number I equals zero (is close to zero) and the other combinations are non-zero. When there are no failures, all combinations are equal to zero. Thus, the presence of a single failure can be determined by one "zero" (small in magnitude) combination and the rest non-zero combinations.

4   Least absolute deviations and least squares

We investigate classical estimation methods: least absolute deviations and least squares [1]. First, we find the estimate $\hat{q}$ the parameter vector q using the least deviations method (or least squares method). Then by the obtained estimate $\hat{q}$ we construct the estimate of the measurement error using the formulas

$$\varrho_i = z_i - G_i\hat{q}, \quad i=1,\ldots,6. \tag{3}$$

Then the absolute values of this estimates are compared to the prescribed threshold value Δ. We assume that the failure occurs in the channel, where the absolute value of the estimate is greater than the threshold value Δ.

Numerical testing shows that the classical methods do not provide new advantages with such small amount of measurements (six measurements in our case). The least squares method does not always detect even a singular failure. The least absolute deviations method does detect one failure but requires more computational cost than the "zero" combinations method. Both methods are unable to detect two failures reliably.

5   Guaranteeing estimation method

Consider a new method for determining failures, which is based on the guaranteeing approach [2-4]. The algorithm is as follows. For each row of the matrix G we construct the estimate of the scalar quantity $\widehat{G_1 q}$ using the guaranteeing approach. The problem is set as an optimal estimation problem, where we minimize the maximal estimation error under all uncertain factors. At the same time, we assume that the absolute values of the errors are bounded by a known constant σ:

$$|\varrho_i| \leq \sigma, \quad i=1,\ldots,6 \tag{4}$$

Then using the formulas

$$\hat{\rho}_i = z_i - \hat{G}_i\hat{q}, \quad i=1,\ldots,6 \tag{5}$$

we obtain the estimate for the estimation error, which is compared to the threshold value Δ.

In order to calculate $\hat{\varrho}$ we need to solve a set of variational problems, which can be reduced to linear programs. Due to the low dimension of the variational problems, the corresponding computational load is very moderate. At the same time, the guaranteeing estimation problem can be a reference for evaluating the quality of other simpler implemented algorithms. The advantage of the guaranteeing approach is that the problem of detecting failures is formulated as an optimal estimation problem. Therefore, unlike traditional approaches, all available information is used for estimation. Moreover, it is possible to calculate the optimal guaranteed estimation errors for the parameters. Extensive numerical testing shows that the guaranteeing approach allows us to detect two failures in an angular rate sensor unit.

6 Conclusion

We proposed the solution to the problem for failure detection in a redundant unit of angular velocity sensors. Four methods are investigated. Numerical testing showed that the guaranteeing approach provides the most accurate results.

References:

[1] V.I. Mudrov, V.L. Kushko. Least Absolute Deviations Method. Moscow: URSS, 2013.

[2] A.B. Kurzhanskii. Control and Estimation under Uncertainty Conditions. Moscow: Nauka, 1977.

[3] M.L. Lidov. Minimax Estimation Methods. Preprint No.71, Keldysh Institute of Applied Mathematics of RAS, 2010.

[4] A. Matasov. Guaranteeing Estimation Method. Moscow: Moscow University Press, 2009.

Inertial Autonomous In-Flight Alignment Method Based on Velocity Model Constraint in Strong Disturbance Environment

Jinwen Wang, Hongxuan Che, Jiachao Wu, Qingjie Liu

(School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: It is difficult for in-flight alignment of projectile inertial navigation system in the condition of strong wind disturbance and GNSS-denied. Thus, an inertial autonomous in-flight alignment method based on velocity model constraint is proposed, and the dynamic alignment model of inertial navigation system is constructed. With the constraint of ballistic kinetics model, the velocity estimation state equation is derived. According to the inertial navigation specific force equation, the measurement equation is obtained. The velocity is estimated online via UKF for nonlinear filtering model. The simulation results show that the yaw angle alignment accuracy is improved by 80%.

Key words: In-flight alignment, Inertial navigation, GNSS-denied, Velocity model constraint, Strong disturbance environment

1 Introduction

Initial alignment of inertial navigation is the basis of achieving inertial precise navigation[1]. For guided projectiles with high rotation ($\geq$20 r/s) and high overload ($\geq$10000 g), it is necessary to activate the inertial navigation system after launch. Thus, the in-flight alignment of the projectile inertial navigation system is very important, as it is directly related to the hitting accuracy for projectiles[2].

The initial alignment includes two stages: coarse alignment and fine alignment. Coarse alignment only determines the attitude matrix approximately. Fine alignment can obtain an accurate attitude matrix[3]. Many attitude estimation methods based on strapdown inertial navigation system (SINS) and global navigation satellite system (GNSS) information have been proposed for some special applications[4][5], and many scholars have carried out GNSS-denied in-flight alignment of projectile inertial navigation[6][7]. However, in practical applications, guided projectiles face challenges such as strong wind disturbance and GNSS-denied, leading to difficulties for in-flight alignment. Thus, it is necessary to find another way to obtain reference information that can assist the SINS in-flight alignment.

In recent years, research on kinetics model-aided SINS has gained significant attention. For land vehicles, Binhan Du et al. proposed a three-layer structure of a vehicle kinetics model-aided inertial navigation system, which can improve the navigation system accuracy[8]. For aerial vehicles, Xiaoyau Ye et al. elaborated on the progress of using UAV pneumatic/kinetics models to assist

---

② Corresponding author, WANG Jinwen : javedwjw@163.com.

navigation systems in improving navigation accuracy[9]. For underwater vehicles, A. Ramezanifard et al. used the rotating kinetics model of underwater vehicles to assist gyros, which enhanced navigation algorithm performance[10]. Thus, following this approach, we can estimate velocity information using the projectile kinetics model.

This paper proposes an inertial autonomous in-flight alignment method based on velocity model constraint. Based on the fundamental principles of inertial navigation, an inertial navigation in-flight alignment model is constructed. The kinematic model within the ballistic model is utilized for differential operations, and the system state equation is derived with the three-axis velocity as the state variable. According to the specific force equation, the accelerometer output is used as the observation to construct the measurement equation. The Unscented Kalman Filter (UKF) is employed to solve the nonlinear filtering model, enabling online estimation of the projectile's velocity. By integrating the velocity results with the in-flight alignment model, inertial autonomous in-flight alignment is achieved. The specific approach is illustrated in Fig. 1.
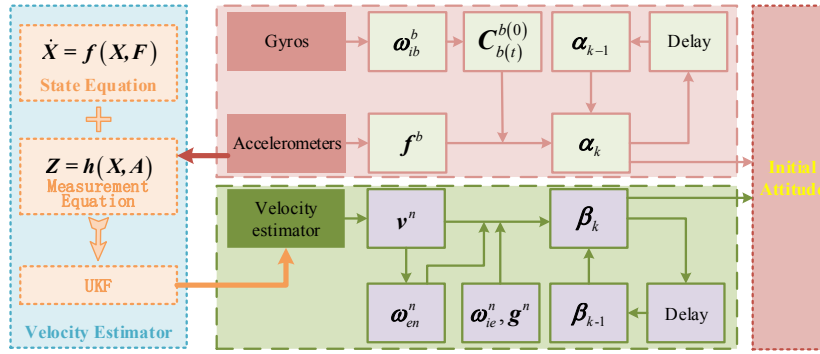
Fig.1 Core idea

The remainder of this paper is organized as follows. In Section II, establish an in-flight alignment model. In Section III, a velocity estimation method with model constraints is proposed. In Section IV, simulation results are analyzed. Section V draws some conclusions.

2 In-flight Alignment Model

The key problem of projectiles in-flight alignment for SINS is to determine attitude matrix $C_{b(0)}^{n(0)}$ from body frame to navigation frame. According to reference [6] as follows.

$$C_{b(0)}^{n(0)} \tilde{\boldsymbol{\alpha}} = \tilde{\boldsymbol{\beta}} \tag{1}$$

where $\begin{cases} \tilde{\boldsymbol{\alpha}} = \int_0^t C_{b(t)}^{b(0)} \boldsymbol{f}^b dt \\ \tilde{\boldsymbol{\beta}} = \int_0^t \dot{\boldsymbol{v}}^n + \left(2\boldsymbol{\omega}_{ie}^n + \boldsymbol{\omega}_{en}^n\right) \times \boldsymbol{v}^n - \boldsymbol{g}^n dt \end{cases}$. $\boldsymbol{v}^n = \begin{bmatrix} v_E^n & v_N^n & v_U^n \end{bmatrix}^T$ denote velocity, with denoting

east, north, and up velocities, respectively. $\boldsymbol{f}^b$ denotes triaxial specific force measured by accelerometer. Other parameters are shown in reference [6].

The specific recursive solutions of $\tilde{\boldsymbol{\alpha}}$ and $\tilde{\boldsymbol{\beta}}$ are as follows

$$\begin{cases} \tilde{\boldsymbol{\alpha}}_k = \int_0^{t_{k-1}} C_{b(t)}^{b(0)} \boldsymbol{f}^b dt + \int_{t_{k-1}}^{t_k} C_{b(t)}^{b(0)} \boldsymbol{f}^b dt = \tilde{\boldsymbol{\alpha}}_{k-1} + \int_{t_{k-1}}^{t_k} C_{b(t)}^{b(0)} \boldsymbol{f}^b dt \\ \tilde{\boldsymbol{\beta}}_k = \int_0^{t_{k-1}} \dot{\boldsymbol{v}}^n + \left(2\boldsymbol{\omega}_{ie}^n + \boldsymbol{\omega}_{en}^n\right) \times \boldsymbol{v}^n - \boldsymbol{g}^n dt + \int_{t_{k-1}}^{t_k} \dot{\boldsymbol{v}}^n dt + \int_{t_{k-1}}^{t_k} \left(2\boldsymbol{\omega}_{ie}^n + \boldsymbol{\omega}_{en}^n\right) \times \boldsymbol{v}^n - \boldsymbol{g}^n dt \\ = \tilde{\boldsymbol{\beta}}_{k-1} + \boldsymbol{v}_{t_k}^n - \boldsymbol{v}_{t_{k-1}}^n + \int_{t_{k-1}}^{t_k} \left(2\boldsymbol{\omega}_{ie}^n + \boldsymbol{\omega}_{en}^n\right) \times \boldsymbol{v}^n - \boldsymbol{g}^n dt \end{cases} \tag{2}$$

The solution of Equation (1) is Wahba problem[6], and it is solved by MKF algorithm[7]. In the

strong disturbance environment, there are challenges such as GNSS-denied and strong wind interference, and the velocity information in Equation (2) is unknown. Thus, ballistics model parameters are introduced to design an online velocity estimation method.

3 On-line Velocity Estimation Method with Model Constraint

Kinematics equation in ballistic model is as follows.

$$\begin{cases} \dot{x} = v_x = v\cos\psi_2\cos\theta_a \\ \dot{y} = v_y = v\cos\psi_2\sin\theta_a \\ \dot{z} = v_z = v\sin\psi_2 \end{cases} \tag{3}$$

where, v is the projectile resultant velocity. $\begin{bmatrix} \theta_a & \psi_2 \end{bmatrix}$ is the velocity elevation angle and velocity direction angle, which can be obtained according to the set ballistic model.

According to Equation (3), the triaxial acceleration of the projectile is inversed, and it can be obtained as follows.

$$\begin{cases} a_x = \ddot{x} = \dot{v}\cos\psi_2\cos\theta_a - v\dot{\psi}_2\sin\psi_2\cos\theta_a - v\dot{\theta}_a\cos\psi_2\sin\theta_a \\ \quad = \dfrac{F_{x_2}}{m}\cos\psi_2\cos\theta_a - \dfrac{F_{z_2}}{m}\sin\psi_2\cos\theta_a - \dfrac{F_{y_2}}{m}\sin\theta_a \\ a_y = \ddot{y} = \dot{v}\cos\psi_2\sin\theta_a - v\dot{\psi}_2\sin\psi_2\sin\theta_a + v\dot{\theta}_a\cos\psi_2\cos\theta_a \\ \quad = \dfrac{F_{x_2}}{m}\cos\psi_2\sin\theta_a - \dfrac{F_{z_2}}{m}\sin\psi_2\sin\theta_a + \dfrac{F_{y_2}}{m}\cos\theta_a \\ a_z = \ddot{z} = \dot{v}\sin\psi_2 + v\dot{\psi}_2\cos\psi_2 \\ \quad = \dfrac{F_{x_2}}{m}\sin\psi_2 + \dfrac{F_{z_2}}{m}\cos\psi_2 \end{cases} \tag{4}$$

where $\begin{bmatrix} F_{x_2} & F_{y_2} & F_{z_2} \end{bmatrix}$ denote triaxial component of exterior forces in ballistic coordinate system, and m denotes the projectile mass, which can be obtained according to the set ballistic model.

Combining Equation (3) and Equation (4), we can obtain as follows.

$$\begin{cases} \dot{v}_x = \dfrac{F_{x_2}}{m}\cos\psi_2\cos\theta_a - \dfrac{F_{z_2}}{m}\sin\psi_2\cos\theta_a - \dfrac{F_{y_2}}{m}\sin\theta_a \\ \dot{v}_y = \dfrac{F_{x_2}}{m}\cos\psi_2\sin\theta_a - \dfrac{F_{z_2}}{m}\sin\psi_2\sin\theta_a + \dfrac{F_{y_2}}{m}\cos\theta_a \\ \dot{v}_z = \dfrac{F_{x_2}}{m}\sin\psi_2 + \dfrac{F_{z_2}}{m}\cos\psi_2 \end{cases} \tag{5}$$

where

$$\begin{cases} v = \sqrt{v_x^2 + v_y^2 + v_z^2} \\ \psi_2 = \mathrm{ac}\sin(v_z/v) \\ \theta_a = \mathrm{ac}\tan(v_y/v_x) \end{cases} \tag{6}$$

Let $\boldsymbol{X} = \begin{bmatrix} v_x & v_y & v_z \end{bmatrix}^{\mathrm{T}}$ can simplify Equation (5) to as follows.

$$\dot{\boldsymbol{X}} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{F}) \tag{7}$$

where $\boldsymbol{f}(*)$ is the nonlinear mapping relation in Equation (5). $\boldsymbol{F} = \begin{bmatrix} F_{x_2} & F_{y_2} & F_{z_2} \end{bmatrix}$.

According to the specific force equation, it can be obtained as follows.

$$a = \dot{v}^n = C_b^n f^b - \left(2\omega_{ie}^n + \omega_{en}^n\right) \times v^n + g^n \tag{8}$$

where $\quad a = \begin{bmatrix} a_z & a_x & a_y \end{bmatrix}^{\mathrm{T}} \quad$ denote triaxial acceleration.

$v^n = \begin{bmatrix} v_E^n & v_N^n & v_U^n \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} \dot{z} & \dot{x} & \dot{y} \end{bmatrix}^{\mathrm{T}}$

$\quad = \begin{bmatrix} v\sin\psi_2 & v\cos\psi_2\sin\theta_a & v\cos\psi_2\cos\theta_a \end{bmatrix}^{\mathrm{T}}$ denote velocity, with denoting east, north, and

up velocities, respectively.

$C_n^b = \left(C_b^n\right)^{-1} =$

$\begin{bmatrix} \cos\gamma\cos\varphi_2 + \sin\gamma\sin\varphi_2\sin\varphi_a & -\cos\gamma\sin\varphi_2 + \sin\gamma\cos\varphi_2\sin\varphi_a & -\sin\gamma\cos\varphi_a \\ \sin\varphi_2\cos\varphi_a & \cos\varphi_2\cos\varphi_a & \sin\varphi_a \\ \sin\gamma\cos\varphi_2 - \cos\gamma\sin\varphi_2\sin\varphi_a & -\sin\gamma\sin\varphi_2 - \cos\gamma\cos\varphi_2\sin\varphi_a & \cos\gamma\cos\varphi_a \end{bmatrix}$.

$\begin{bmatrix} \varphi_a & \varphi_2 & \gamma \end{bmatrix}$ denote projectile-axis elevation angle, projectile-axis direction angle, roll angle.

According to Equation (8), the specific force output by the accelerometer is inversed, and it can be obtained as follows.

$$f^b = \left(C_b^n\right)^{-1} \bullet \left(a + \left(2\omega_{ie}^n + \omega_{en}^n\right) \times v^n - g^n\right) \tag{9}$$

Let measurement $Z = f^b$.

$$Z = h(X, A) \tag{10}$$

where $h(*)$ is the nonlinear mapping relation in Equation (9). $A = \begin{bmatrix} \varphi_a & \varphi_2 & \gamma \end{bmatrix}$.

Via combining Equation (7) and Equation (10), the filtering model of on-line velocity estimation can be obtained as follows.

$$\begin{cases} \dot{X} = f(X, F) \\ Z = h(X, A) \end{cases} \tag{11}$$

Aiming at the strongly nonlinear filtering model Equation (11), the UKF method is used to estimate the projectile velocity.

4 Method Verification and Result Analysis

4.1 Simulation Condition

Method evaluation indices see reference [6]. Taking the 155-type projectile as the simulation object, the 6-DOF rigid-body ballistic model is adopted to simulate the full ballistic attitude data in standard meteorological conditions. Initial rotation rate is 300r/s and sampling frequency is 1000Hz. The shooting point coordinates are (118°E, 31°N, 2m), the initial shooting angle is 51°, and the initial shooting direction is 0°. The sensor parameters are shown in Table 1.

Table 1 Sensor parameters

Sensors	Parameters	Values
Gyroscope	Bias	10°/h
	Bias stability	1°/h
Accelerometer	Bias	50mg
	Bias stability	5mg

Different wind velocities are used to simulate the strong disturbance flight environment for projectiles, and the alignment data are generated under the wind velocities of 10 m/s, 15 m/s and 20

m/s respectively, and the ballistic data with the wind velocity of 0 m/s is used as the initial set ballistic. At different wind velocities, the velocity of set ballistic is used as the reference for alignment, and the methods are called Set-10, Set-15 and Set-20, for short. UKF with model velocity constraint is used to estimate the velocity, and the methods are called UKF-10, UKF-15 and UKF-20 for short.

4.2 Simulation Results and Analysis

The results of one alignment with different methods in different environments and RMSE results with Monte Carlo 100 alignments are shown in Fig. 2.
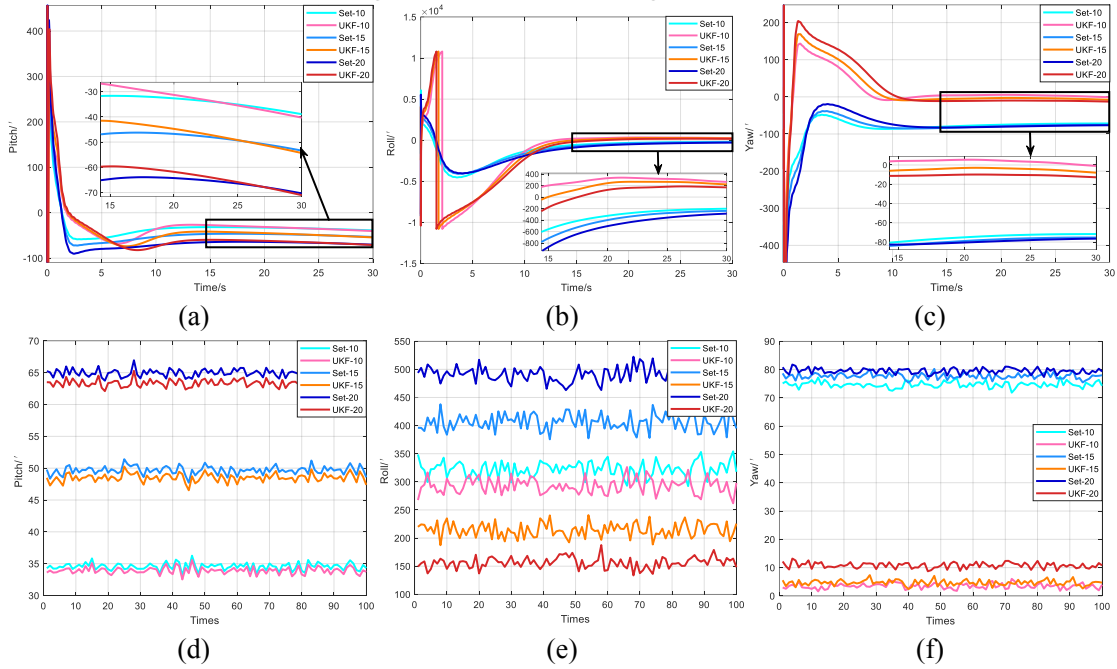
Fig .2 Alignment results. (a)Pitch. (b)Roll. (c)Yaw. (d)Pitch RMSE. (e)Roll RMSE. (f) Yaw RMSE.

It can be seen from Fig.2, in different wind velocities, the greater the wind velocity, the greater the error when the set ballistic velocity is used for alignment. The in-flight alignment accuracy based on velocity model constraint proposed in this paper is better than that of the set ballistic method, especially for yaw angle, and the alignment accuracy is improved by nearly one order of magnitude, which fully shows that the in-flight alignment method based on velocity model constraint has the characteristics of high accuracy, full autonomy and robustness. Further statistics of ARMSE, the results are shown in Table 2.

Table 2 ARMSE

Methods / Axial	Set-10	UKF-10	IP-10	Set-15	UKF-15	IP-15	Set-20	UKF-20	IP-20
Pitch/'	34.5675	33.9273	1.85%	49.6870	48.5091	2.37%	64.9481	63.3457	2.47%
Roll/'	322.7762	292.9210	9.25%	406.1806	214.5949	47.17%	489.9753	156.1671	68.13%
Yaw/'	74.6927	3.4861	**95.33%**	77.6986	4.6819	**93.97%**	79.5362	10.7916	**86.43%**

It can be seen from Table 2 that in the conditions of strong wind disturbance and GNSS-denied, the online velocity estimation method based on the ballistic model proposed in this paper can effectively estimate the velocity and realize inertial autonomous in-flight alignment. In particular, the yaw angle alignment accuracy is improved by more than 80%.

5 Conclusion

Aiming at the problem that there is no information reference for inertial navigation in-flight alignment in the condition of GNSS-denied, this paper proposes an inertial autonomous in-flight alignment method based on velocity model constraint. We construct a filtering model for online velocity estimation, adopts UKF to realize online velocity estimation, and assists inertial navigation system to complete in-flight alignment. The simulation results show that the inertial in-flight alignment method with velocity model constraint proposed in this paper can realize autonomous in-flight alignment without relying on external sensor reference. Our next work goal is to improve the velocity estimation method and improve the accuracy of in-flight alignment.

References:

[1] X. Xu, Y. Sun, Y. Yao and T. Zhang. A Robust In-Motion Optimization-Based Alignment for SINS/GPS Integration[J]. IEEE Transactions on Intelligent Transportation Systems,2022, 23(5):4362-4372.

[2] J. Peng, X. Zhang, X. Peng, P. Yao, J. Chen and G. Yang. RODM-Alignment: An Inertial Alignment Algorithm for Sliding Window Reverse Optimization Delayed Marginalization[J]. IEEE Transactions on Instrumentation and Measurement, 2024,73: 1-12, 2024, 8506112.

[3] Huang Y, Zhang Y, Chang L . A New Fast In-motion Coarse Alignment Method for GPS-aided Low-cost SINS[J]. IEEE/ASME Transactions on Mechatronics, 2018:1-1.

[4] Pei F , ang S , Yin S . In-Motion Initial Alignment Using State-Dependent Extended Kalman Filter for Strapdown Inertial Navigation System[J]. IEEE Transactions on Instrumentation and Measurement, 2022,70:1-12.

[5] Xu X , Sun Y , Yao Y , et al. A Robust In-Motion Optimization-Based Alignment for SINS/GPS Integration[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, PP(99):1-11.

[6] J. Wang, Z. Deng and Y. Bo. Relaying Fast In-Flight Alignment Method Based on Adaptive Multi constraints[J]. IEEE Sensors Journal, 2024,24(2):1264-1274.

[7] J. Wang, Z. Deng, X. Liang and Z. Miao. Functional Iteration in-Flight Alignment Method for Projectiles MSINS[J]. IEEE/ASME Transactions on Mechatronics, 2022,27(5):2887-2896.

[8] B. Du, B. Yang, H. Wang, G. Ren and Z. Shi. Multilayer Model-Aided INS Based on Real-Time Corrected Vehicle Dynamics Model[J]. IEEE Transactions on Instrumentation and Measurement,2023,72(1):1-15, 2023.

[9] X. Ye, F. Song, Z. Zhang and Q. Zeng. A Review of Small UAV Navigation System Based on Multisource Sensor Fusion[J]. IEEE Sensors Journal, 2023,23(17): 18926-18948.

[10] A. Ramezanifard, M. Hashemi, H. Salarieh and A. Alasty. Enhancing the Robustness of INS-DVL Navigation Using Rotational Model of AUV in the Presence of Model Uncertainty[J]. IEEE Sensors Journal, 2022,22(11):10931-10939.

Phase-Delay-Error Calibration Method for Control Circuit of Rate Integrating Hemispherical Resonator Gyro

Wei Zhennan, Wu Hongbo, Guan Zeyuan, Yi Guoxing, Wang Changhong

(Space Control and Inertial Technology Research Center, Harbin Institute of Technology, Harbin 150001, China)

Abstract: The performance of the Rate Integrating Hemispherical Resonator Gyro (RI-HRG) is limited by the anisotropic drift error of resonator standing wave. After the resonator mechanical defects and the unbalance error of control circuit signal chain, the Phase-Delay-Error (PDE) of control circuit is the primary factor affecting the standing wave drift error. In this paper, the source of control circuit PDE is analyzed, the standing wave drift error model involving PDE is established, and a PDE calibration method of control circuit based on dual Direct Digital Synthesizers (DDS) is proposed. The effectiveness of the proposed calibration method is demonstrated with experiment in kind. After PDE compensation, the drift error of the standing wave in different directions of RI-HRG is significantly suppressed.

Key words: rate integrating mode; hemispherical resonator gyro; phase delay error; standing wave drift error; phase compensation;

1 Introduction

The circumferential distribution characteristics of resonator standing wave drift error are the key factors affecting the performance of RI-HRG. In addition to improving the resonator mechanical properties by updating manufacturing processes, precise identification and compensation of the error characteristics for gyro control circuit are also effective means of reducing standing wave drift error [1-3].

The main circuit error parameters that affect the standing wave control performance in rate integrating mode include Channel Unbalance Error, Signal Coupling Error, and Phase Delay Error [4]. Channel Unbalance Error and Signal Coupling Error are reflected in the detection of resonator vibration signals and the application of control signals, which are usually equivalent to gain unbalance error and standing wave vector detection & drive angle error. These two equivalent errors can be corrected by applying feedforward controller separately in the signal chain [5-9]. The Phase Delay Error, PDE, is mainly caused by the open-loop part of the gyro control circuit, resulting in an unexpected phase delay between the driving signal applied to the resonator and its vibration signal. The PDE causes additional standing wave drift error by inducing standing wave vector control error.

In this paper, the sources of PDE in RI-HRG control circuit are discussed, and its impact on standing wave drift error is analyzed. A PDE identification method based on observed orthogonal

control quantity and a compensation method based on dual DDS are proposed. The experimental results show that the proposed method can effectively suppress the influence of PDE and significantly reduce the gyro standing wave drift error.

2 Sources and Effects of PDE

2.1 Sources of PDE

There are three main sources of control circuits PDE, including buffer amplifier, multistage linear amplifier, and signal converter.

Buffer amplifier is used to convert the capacitance signal caused by resonator vibration into voltage signal [10]. PDE in this section includes nominal phase error $\delta\varphi_n$ caused by the circuit structure and residual phase error $\delta\varphi_r$ caused by non-ideal component parameters.

Multistage linear amplifiers are used for amplifying vibration signals to meet signal transmission requirements. The PDE $\delta\varphi_a$ in this section is mainly affected by amplifier parameters such as the Gain-Bandwidth Product and Slew Rate. The frequency of resonator vibration signal is about 7kHz [11-12], which is a low-frequency signal, and $\delta\varphi_a$ is very small usually.

A/D and D/A signal converter are used to collect resonator vibration signals and reconstruct vibration drive signals, respectively. The PDE $\delta\varphi_t$ is due to several driving clock cycles of the converters. The sampling frequency of the signal converter for RI-HRG control circuit is usually a few MHz, and $\delta\varphi_t$ is also very small [13-14].

As for the digital signal processing part between the A/D and D/A converter in control circuit, it belongs to the standing wave closed-loop control part, and the frequency tracking loop PI controller can ensure that this part does not generate PDE.

In summary, the PDE $\delta\varphi$ of RI-HRG control circuit can be expressed as

$$\delta\varphi = \delta\varphi_n + \delta\varphi_r + \delta\varphi_a + \delta\varphi_t \tag{1}$$

2.2 Effects of PDE

The orthogonal reference signal of gyro control circuit with PDE $\delta\varphi$ can be expressed as

$$\begin{cases} u_{rc} = \cos(\omega_0 t - \delta\varphi) \\ u_{rs} = \sin(\omega_0 t - \delta\varphi) \end{cases} \tag{2}$$

Where, ω_0 is the resonator inherent angular frequency with the action of orthogonal control.
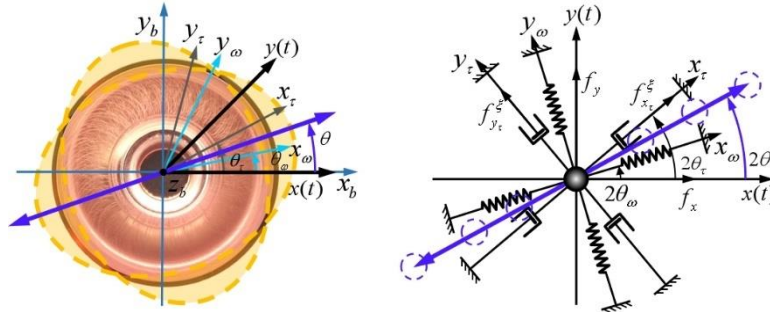
Fig.1 Coordinate system definition of hemisphere resonator

Assuming that the detection and drive unbalance error compensation for gyro control circuit have been completed [5-6]. Let f_a and f_q represent the amplitude and orthogonal control force of resonator, respectively, and 2θ be the azimuth angle of standing wave. As shown in Fig.1, when applying amplitude and orthogonal control loops, the control forces in the orthogonal direction , x

and y, to the resonator can be expressed as

$$\begin{cases} f_x = f_a \cos \delta\varphi \cos 2\theta \cos \omega_0 t + f_a \sin \delta\varphi \cos 2\theta \sin \omega_0 t \\ \quad - f_q \cos \delta\varphi \sin 2\theta \sin \omega_0 t + f_q \sin \delta\varphi \sin 2\theta \cos \omega_0 t \\ f_y = f_a \cos \delta\varphi \sin 2\theta \cos \omega_0 t + f_a \sin \delta\varphi \sin 2\theta \sin \omega_0 t \\ \quad + f_q \cos \delta\varphi \cos 2\theta \sin \omega_0 t - f_q \sin \delta\varphi \cos 2\theta \cos \omega_0 t \end{cases} \tag{3}$$

Equation (3) indicates that PDE will cause amplitude and phase control forces to project simultaneously in space and time, forming four driving forces that are coincident or orthogonal to the standing wave azimuth or reference signal phase, $\tilde{f}_{ac}$, $\tilde{f}_{as}$, $\tilde{f}_{qc}$, and $\tilde{f}_{qs}$, respectively.

$$\begin{cases} \tilde{f}_{ac} = f_a \cos \delta\varphi \\ \tilde{f}_{as} = f_a \sin \delta\varphi \\ \tilde{f}_{qc} = -f_q \sin \delta\varphi \\ \tilde{f}_{qs} = f_q \cos \delta\varphi \end{cases} \tag{4}$$

Where, $\tilde{f}_{qs}$ represents the actual orthogonal control force exerted on resonator at this time.

The changing rate of resonator orthogonal error Q can be obtained by applying averaging method [15].

$$\dot{Q} = -\frac{2}{\tau}Q - \Delta\omega \sin 4(\theta - \theta_\omega)\sqrt{E^2 - Q^2} + \frac{a}{\omega_0}f_{qs} - \frac{q}{\omega_0}f_{ac} \tag{5}$$

Where, $\Delta\omega$ is Frequency Split, θ_ω is Rigid Axis Angle, a is Primary Wave Amplitude, and q is Orthogonal Wave Amplitude. The standing wave energy is $E=a^2+q^2$, and the orthogonal error is $Q=2aq$. Under the action of control circuit, Q and its change rate are zero, q is zero, and E and a are constants. Equation (4) shows the required orthogonal control force for the resonator

$$f_{qs} = a\omega_0\Delta\omega \sin 4(\theta - \theta_\omega) \tag{6}$$

Orthogonal control loop will force $\tilde{f}_{qs}$ and f_{qs} to be equal. According to equations (4) and (6), the actual orthogonal control force output by the orthogonal circuit will be

$$f_q = \cos^{-1}\delta\varphi \cdot a\omega_0\Delta\omega \sin 4(\theta - \theta_\omega) \tag{7}$$

Then, according to equation (4), the force $\tilde{f}_{qc}$ applied to the resonator that is orthogonal to the standing wave direction and in phase with the reference signal is

$$\tilde{f}_{qc} = -a\omega_0\Delta\omega \sin 4(\theta - \theta_\omega)\tan \delta\varphi \tag{8}$$

According to the averaging method [15], the standing wave drift error caused by force $\tilde{f}_{qc}$ is

$$\dot{\theta}_{qc} = \frac{1}{4\omega_0 a}\tilde{f}_{qc} = \frac{1}{4}\Delta\omega \tan \delta\varphi \sin 4(\theta - \theta_\omega) \tag{9}$$

Equation (9) indicates that the amplitude of standing wave drift error caused by PDE is positively correlated with the resonator Frequency Split and the tangent value of PDE, and the direction coincides with the resonator Rigid Axis Angle.

3 Identification and Compensation of PDE

3.1 Identification Method of PDE

On the basis of amplitude and orthogonal control loop, when initiative driving control force f_v is involved to drive standing wave precession, the control force in the resonator orthogonal directions, x and y, can be expressed as

$$\begin{cases} f_x = f_a \cos \delta\varphi \cos 2\theta \cos \omega_0 t + f_a \cos \delta\varphi \cos 2\theta \cos \omega_0 t \\ \qquad - \left(f_q \cos \delta\varphi + f_v \sin \delta\varphi \right) \sin 2\theta \sin \omega_0 t + \left(f_q \sin \delta\varphi - f_v \cos \delta\varphi \right) \sin 2\theta \cos \omega_0 t \\ f_y = f_a \cos \delta\varphi \sin 2\theta \cos \omega_0 t + f_a \sin \delta\varphi \sin 2\theta \sin \omega_0 t \\ \qquad + \left(f_q \cos \delta\varphi + f_v \sin \delta\varphi \right) \cos 2\theta \sin \omega_0 t - \left(f_q \sin \delta\varphi - f_v \cos \delta\varphi \right) \cos 2\theta \cos \omega_0 t \end{cases} \quad (10)$$

Equation (10) indicates that the orthogonal control force applied to resonator at this time is

$$\widetilde{f}_{qs} = f_q \cos \delta\varphi + f_v \sin \delta\varphi \quad (11)$$

Similarly, the orthogonal control loop will force $\widetilde{f}_{qs}$ and f_{qs} to be equal. According to equations (6) and (11), the actual orthogonal control force f_q output by orthogonal control loop is

$$f_q = -f_v \tan \delta\varphi + \cos^{-1} \delta\varphi \cdot a\omega_0 \Delta\omega \sin 4\left(\theta - \theta_\omega\right) \quad (12)$$

Equation (12) indicates that the PDE can couple the force f_v to the orthogonal control force, and there is a bias in the circumferential distribution of the orthogonal control force. Moreover, the sign of the PDE is opposite to that of the bias. Let the gain from gyro excitation electrode to electrostatic force be G_c, and the orthogonal control quantity and initiative driving control quantity be d_q and d_v, respectively. Then, equation (12) can be corrected as

$$d_q = -d_v \tan \delta\varphi + \frac{a\omega_0 \Delta\omega}{G_c \cos \delta\varphi} \cdot \sin 4\left(\theta - \theta_\omega\right) \quad (13)$$

Equation (13) indicates that PDE can be identified by driving standing wave precession with constant initiative driving control quantity. When the standing wave is driven to under go precession at a small constant rate, the bias of orthogonal control quantity can be identified with the least squares method, and then the PDE can be calculated and compensated. Considering the presence of other control errors, multiple calibration iterations for PDE is required.
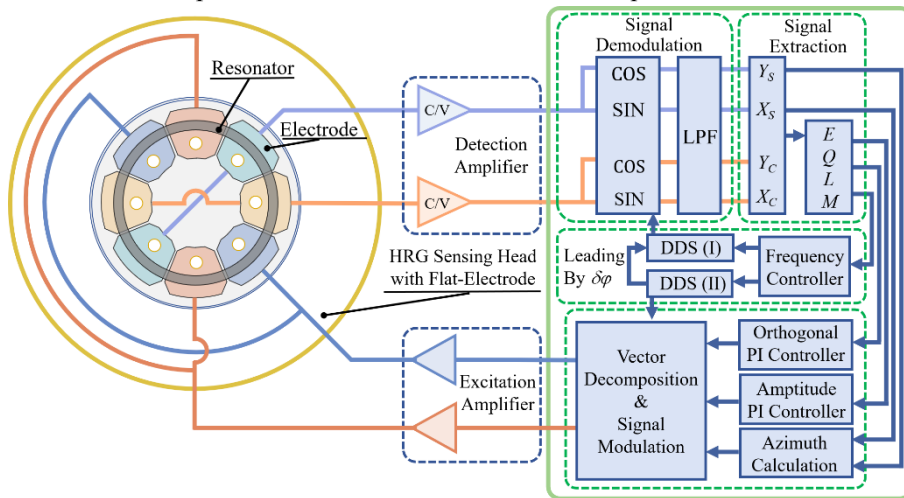
Fig.2 RI-HRG control system block diagram using PDE compensation technology

3.2 Compensation Method of PDE

As shown in Fig.2, a gyro control system utilizing PDE compensation technology is presented,

with the core being dual synchronous DDS. DDS (I) is used for signal demodulation, DDS (II) is used for signal modulation, and the phase of DDS (II) ahead of DDS (I) is $\delta\varphi$.

4 Experiment

Fig.3 and Fig.4 show the RI-HRG experimental results in kind. Fig.3 shows the circumferential distribution of standing wave orthogonal control quantity before and after phase error compensation, Fig.4 shows the test results of standing wave drift error before and after phase error compensation, and the phase error compensation value is -1.3125°. It can be seen that there is a significant bias in the orthogonal control quantity before compensation, while the bias is eliminated after compensation. Meanwhile, the amplitude of standing wave drift error before and after compensation is 0.001372°/s and 0.000347°/s, respectively. After compensation, the drift error is significantly further suppressed, which will help improve the scale factor nonlinearity and zero bias stability of RI-HRG.
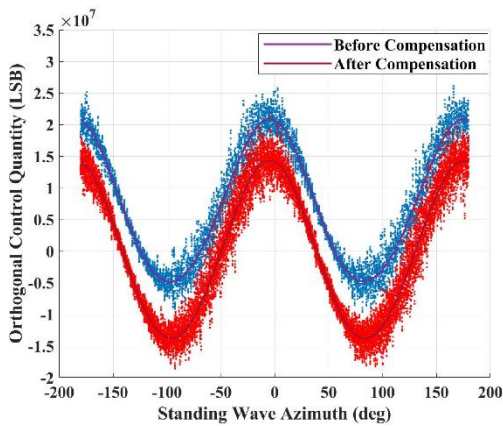
Fig.3 Distribution of orthogonal control quantity **Fig.4** Standing wave drift error
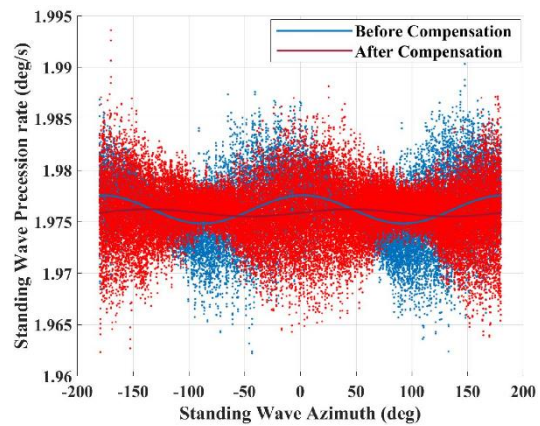
5 Conclusion

In this paper, a method for correcting PDE in RI-HRG control circuit is proposed, and the amplitude of the standing wave drift error is significantly suppressed. Firstly, the sources of PDE in RI-HRG control circuit are analyzed, and a standing wave drift error model under the influence of PDE is established. Secondly, the circumferential distribution model of standing wave orthogonal control force under amplitude, orthogonal and initiative driving control is studied, and a PDE identification method based on bias detection of orthogonal control force is proposed. Thirdly, a PDE compensation method for RI-HRG control circuit based on dual DDS is proposed, and the compensation of control circuit PDE is implemented by adjusting the phase difference between the standing wave modulation and demodulation reference signals. Finally, the proposed PDE identification and compensation methods are validated with experiments in kind.

References:

[1] Ragot V, Remillieux G. A New Control Mode Greatly Improving Performance of Axisymmetrical Vibrating Gyroscopes[J]. Gyroscopy and Navigation, 2011, 2(4): 229-238.

[2] Remillieux G. Delhaye F. Sagem Coriolis Vibrating Gyros: a vision realized[C]//2014 DGON Inertial Sensors and Systems (ISS), Karlsruhe, Germany, 2014, pp. 1-13.

[3] Wei Zhen-nan, Sun Yi-wei, Yi Guo-xing, Xi Bo-qi, Wang Chang-hong. Standing Wave Vector Control Error Analysis and Suppression Method for Rate Integrating Hemispherical Resonator Gyro[C]//2023 30th Saint

Petersburg International Conference on Integrated Navigation Systems (ICINS), Saint Petersburg, Russian Federation, 2023, pp. 1-5.

[4] Yan Kai-chen, Wang Xiao-xu, Zou Kang, Qu Tian-liang, Che Chi-cheng, Lu Qian-bo. Self-Excitation Enabled Decoupling, Calibration, and Compensation of Errors for Whole-Angle Hemispherical Resonator Gyroscope[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1001613.

[5] Sun Yi-wei, Yi Guo-xing, Wei Zhen-nan, Wang Qi. Identification and compensation method for unbalance error in signal detection chain of Rate-Integrating hemispherical resonator gyro[J]. Measurement, 2025, 240: 115599.

[6] Sun Yi-wei, Wei Zhen-nan, Yi Guo-xing, Wang Ning. Identification and Compensation Method of Unbalanced Error in Driving Chain for Rate-Integrating Hemispherical Resonator Gyro[J]. Sensors, 2024, 24(13): 4328.

[7] Zhang Yong-meng, Miao Tong-qiao, Guo Ke-chen, Sun Jiang-kun, Wu Xue-zhong, Xiao Ding-bang. Electronic Gain Error Compensation For Whole-Angle Mode Coriolis Vibrating Gyroscopes With High Q Factor[C]//2020 IEEE International Symposium on Inertial Sensors and Systems (INERTIAL), Hiroshima, Japan, 2020, pp. 1-4.

[8] Sun Jiang-kun, Yu Sheng, Zhang Yong-meng, Shi Yan, Lu Kun, Xi Xiang, Wu Xue-Zhong, Xiao Ding-bang. Characterization and Compensation of Detection Electrode Errors for Whole-Angle Micro-Shell Resonator Gyroscope[J]. Journal of Microelectromechanical Systems, 2022, 31(1): 19-28.

[9] Wang Qi, Xie Wei-nan, Xi Bo-qi, Sun Yi-wei, Yi Guo-xing. Rate Integrating Hemispherical Resonator Gyroscope Detection Error Analysis and Compensation[J]. IEEE Sensors Journal, 2023, 23(7): 7068-7076.

[10] Shtek S G, Zheglov M A, Isaev M M. Development of a converter unit for reading signals from a wave solid-state gyroscope with a quartz resonator[J]. Journal of «Almaz － Antey» Air and Space Defence Corporation, 2020, (2): 65-73.

[11] Jeanroy A, Bouvet A, Remillieux G. HRG and Marine Applications[J]. Gyroscopy and Navigation, 2014, 5(2): 67-74.

[12] Hu Qiang, Yue Ya-zhou, Cao Shi-yin, Chen Tong-yang, Yang Yu-heng, Teng Lin. A method for balancing the fourth harmonic mass of hemispherical resonator gyro[J]. Journal of Chinese Inertial Technology, 2023, 31(6): 601-610.

[13] Lee J, Yun S W, Rhim J. Design and Verification of a Digital Controller for a 2-Piece Hemispherical Resonator Gyroscope[J]. Sensors, 2016, 16(4): 555.

[14] Zhuravlev V P, Perelyaev S E, Bodunov B P, Bodunov S B. New-generation small-size solid-state wave gyroscope for strapdown inertial navigation systems of unmanned aerial vehicle[C]//2019 26th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS), St. Petersburg, Russia, 2019, pp. 1-3

[15] Lynch D D. Vibratory gyro analysis by the method of averaging[C]//2nd St. Petersburg Conference on Gyroscopic Technology and Navigation, St. Petersburg, Russia, 1995, pp. 26-34.

A Novel Differential Frequency Modulation Hemispherical Resonator Gyroscope with 0.5356ppm Scale Factor Nonlinearity, 0.2°/s Angular Velocity Noise, and 0.0094°/h Bias Instability[1]

Kaichen Yan[1], Xiaoxu Wang[1], Xiquan Wang[1], Tianliang Qu[2], Qianbo Lu[3]

(1. School of Automation, Northwestern Polytechnical University, Xi'an 710072, China;

2. Huazhong Optoelectronics Technology Research Institute, Wuhan 430233, China;

3. Institute of Flexible Electronic, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: Differential frequency modulation whole-angle (DFM-WA) hemispherical resonator gyroscope (HRG) has both the advantages of large dynamics as same as amplitude modulation whole-angle (AM-WA) HRG, and the superiority of high resolution and signal-to-noise ratio (SNR) outputs based on the DFM-WA scheme. This paper realizes the measurement, control, and output of the HRG based on the DFM-WA scheme, and constructs the slow variable model of the DFM-WA HRG based on Lynch average method. The experiment results demonstrate that the DFM-WA HRG has the performance of large dynamics, high precision and low noise. The scale factor (SF) nonlinearity, angular velocity noise (AVN), and bias instability (BI) of the DFM-WA HRG are 0.5356ppm, 0.2°/s, and 0.0094°/h, which reduce 47%, 90%, and 27% compared with the same HRG based on the AM-WA scheme.

Key words: Hemispherical Resonator Gyroscope (HRG); differential frequency modulation whole-angle (DFM-WA) scheme; amplitude modulation whole-angle scheme (AM-WA) scheme; output resolution and signal-to-noise ratio (SNR)

1 Introduction

Hemispherical resonator gyroscope (HRG) represents the way forward for inertial devices. The amplitude modulation whole-angle (AM-WA) scheme[1-5] is the mainstream measurement and control scheme. However, the AM-WA scheme encounters the challenge of poor angular velocity output resolution and signal-to-noise (SNR).

The differential frequency modulation whole-angle (DFM-WA) scheme was proposed by Tsukamoto and Tanaka in 2017[6-7]. The DFM-WA scheme can acquire the angle and angular velocity by frequency information. Frequency information belongs to the intrinsic properties of the resonator, and is less noisy and higher resolution than the amplitude information produced by the control forces. The properties of the DFM-WA scheme render it feasible to combine with HRG. The DFM-WA HRG has the performance potential of large dynamics, high precision and low noise.

Tsukamoto and Tanaka[6-7] built the first validation system for the DFM-WA scheme using a ring resonator, two UHF-LI from Zurich Instruments, and a rate turntable. They[8-14] further focused

on eliminate the effects of the Q-factor and frequency mismatch of the ring resonator based on the DFM-WA scheme. In order to jump out of the rate dead zone of the DFM-WA gyroscope, they[8] proposed a method to generate virtual rotation angular velocity by changing the phase-locked target values of the clockwise and counter clockwise (CW/CCW) modes. They[9] analyzed the relationship of the virtual rotation angular velocity with the temperature, and compensated the virtual rotation angular velocity with temperature modeling using the resonant frequency as temperature information. Tsukamoto and Tanaka[10] further built the principle prototype of the DFM-WA gyroscope by a low Q-factor, high Q-factor mismatch, and high frequency mismatch ring resonator. In order to eliminate the CW/CCW mode coupling due to the Q-factor and frequency mismatch, they observed the coupling amplitude of the unexcited mode by single mode excitation of the resonator, and suppressed the CW/CCW mode coupling by adjusting the amplitude and phase of the control voltages required for the excited mode. They[11-14] analyzed the effects of Q-factor and frequency mismatch at the theoretical level, and proposed various methods for Q-factor and frequency mismatch compensation based on DFM-WA scheme.

The related studies mentioned above pushed the development of the DFM-WA gyroscope. However, the resonator errors are not the major limitations for combining DFM-WA scheme with a high Q-factor, low Q-factor mismatch, and low frequency mismatch hemispherical resonator. The slow variable model of the DFM-WA HRG need to be built based on Lynch average method[15] for the optimized design of the measurement and control scheme, and the analysis, modeling, identification, and compensation of the detection, drive, phase, and resonator errors.

This paper realizes the measurement, control, and output of the HRG based on the DFM-WA scheme, and constructs the slow variable model of the DFM-WA HRG based on Lynch average method. The performances of the same HRG based on the DFM-WA and AM-WA schemes are compared. The DFM-WA HRG has obvious advantages in scale factor (SF) nonlinearity, angular velocity noise (AVN), and bias instability (BI). The measurement and control scheme, the slow variable model, the experiment, and the conclusion of the DFM-WA HRG are presented in Sections 2 - 5.

2   The Measurement and Control Scheme of The DFM-WA HRG

The measurement and control scheme of the DFM-WA HRG mainly includes two automatic gain control (AGC) loops, two phase-locked loops (PLLs), and a CW/CCW mode separator, as shown in Fig.1. The CW/CCW mode separator is the key to realize the measurement, control, and output of the CW/CCW traveling waves. The CW/CCW mode separator is a modulation and low-pass filter module, and generates the amplitude amplification voltages a_{cw}, a_{ccw} and phase-locked errors $\delta_{cw}, \delta_{ccw}$ of the CW/CCW traveling waves. The AGC loops and PLLs control $a_{cw} = a_{ccw} = a$ and $\delta_{cw} = \delta_{ccw} = 0$, and realize the output of CW/CCW traveling waves' resonant frequencies $\omega_{cw}, \omega_{ccw}$ and phases ϕ_{cw}, ϕ_{ccw}. a is the target value of the CW/CCW traveling waves' amplitude amplification voltages, which contributes to the synthesis of the standing wave by the CW/CCW traveling waves. The output of the angular velocity Ω and the standing wave azimuth θ by DFM-WA HRG can be expressed as:

$$\begin{cases} \Omega = \left(\omega_{ccw} - \omega_{cw} \right) / \left(-4K \right) \\ \theta = \left(\phi_{ccw} - \phi_{cw} \right) / 4 \end{cases}, \tag{1}$$

where K is the precession factor of the standing wave.
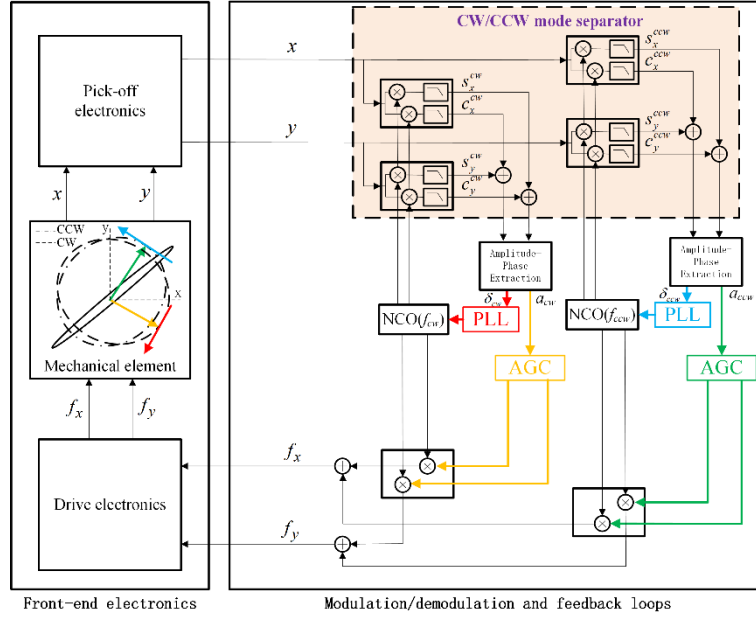
Fig.1 The measurement and control scheme of the DFM-WA HRG with a CW/CCW mode separator, two AGC loops, and two PLLs.

3 The Slow Variable Model of The DFM-WA HRG

Under the measurement and control scheme of the DFM-WA HRG shown in Fig.1, the vibration displacements x, y and control forces f_x, f_y can be expressed as:

$$z = \begin{bmatrix} x \\ y \end{bmatrix} = \mathrm{Re} \begin{bmatrix} a_{cw} e^{i\phi_{cw}} + a_{ccw} e^{i\phi_{ccw}} \\ i a_{cw} e^{i\phi_{cw}} - i a_{ccw} e^{i\phi_{ccw}} \end{bmatrix} / K_s, \tag{2}$$

$$f = \begin{bmatrix} f_x \\ f_y \end{bmatrix} = \mathrm{Re} \begin{bmatrix} \left(V_a^{cw} - i V_\theta^{cw} \right) e^{i\phi'_{cw}} + \left(V_a^{ccw} + i V_\theta^{ccw} \right) e^{i\phi'_{ccw}} \\ \left(i V_a^{cw} + V_\theta^{cw} \right) e^{i\phi'_{cw}} + \left(-i V_a^{ccw} + V_\theta^{ccw} \right) e^{i\phi'_{ccw}} \end{bmatrix} K_d, \tag{3}$$

where ϕ'_{cw}, ϕ'_{ccw} are the phases of the control forces, ϕ_{cw}, ϕ_{ccw} are the phases of the vibration displacements, ϕ'_{cw}, ϕ'_{ccw} exceed ϕ_{cw}, ϕ_{ccw} by 90° in theory; K_s, K_d are the gains of the pick-off electronics and the drive electronics, the loop gain is K_{ds}, $K_{ds} = K_d K_s$; V_a^{cw}, V_a^{ccw} are the quadrature control voltages of the CW/CCW modes, $V_\theta^{cw}, V_\theta^{ccw}$ are the in-phase control voltages of the CW/CCW modes.

The dynamic model of the HRG can usually be expressed as:

$$\begin{aligned} &\ddot{z} + 2 \left[2 \left(-i\sigma_2 \right) K\Omega + 1/\tau + \Delta\left(1/\tau\right)\left(\sigma_3 \cos 4\theta_\tau + \sigma_1 \sin 4\theta_\tau \right)/2 \right] \dot{z} \\ &+ \left[\omega^2 - \omega\Delta\omega\left(\sigma_3 \cos 4\theta_\omega + \sigma_1 \sin 4\theta_\omega \right) \right] z = f \end{aligned}, \tag{4}$$

where $\Delta\left(1/\tau\right), \theta_\tau$ and $\Delta\left(\omega\right), \theta_\omega$ are the anisodamping and anisoelasticity efficients of the hemispherical resonator, ω and τ are the resonant frequency and decay time constant of the hemispherical resonator, and $\sigma_1, \sigma_2, \sigma_3$ are the Pauli rotation matrices.

Substituting (2) and (3) into (4), the slow variable model of the CW/CCW traveling waves

based on Lynch average method can be expressed as:

$$
\left\{
\begin{array}{l}
-\dot{\phi}_{cw}^2 a_{cw} + 4K\Omega\dot{\phi}_{cw}a_{cw} - \Delta\left(\dfrac{1}{\tau}\right)\dot{\phi}_{ccw}a_{ccw}\sin 4\left(\theta - \theta_\tau\right) + \omega^2 a_{cw} - \omega\Delta\omega a_{ccw}\cos 4\left(\theta - \theta_\omega\right) = K_{ds}V_\theta^{cw} \\[2mm]
2\dot{\phi}_{cw}\dot{a}_{cw} + \dfrac{2}{\tau}\dot{\phi}_{cw}a_{cw} + \Delta\left(\dfrac{1}{\tau}\right)\dot{\phi}_{ccw}a_{ccw}\cos 4\left(\theta - \theta_\tau\right) - \omega\Delta\omega a_{ccw}\sin 4\left(\theta - \theta_\omega\right) = K_{ds}V_a^{cw} \\[2mm]
-\dot{\phi}_{ccw}^2 a_{ccw} - 4K\Omega\dot{\phi}_{ccw}a_{ccw} + \Delta\left(\dfrac{1}{\tau}\right)\dot{\phi}_{cw}a_{cw}\sin 4\left(\theta - \theta_\tau\right) + \omega^2 a_{ccw} - \omega\Delta\omega a_{cw}\cos 4\left(\theta - \theta_\omega\right) = -K_{ds}V_\theta^{ccw} \\[2mm]
2\dot{\phi}_{ccw}\dot{a}_{ccw} + \dfrac{2}{\tau}\dot{\phi}_{ccw}a_{ccw} + \Delta\left(\dfrac{1}{\tau}\right)\dot{\phi}_{cw}a_{cw}\cos 4\left(\theta - \theta_\tau\right) + \omega\Delta\omega a_{cw}\sin 4\left(\theta - \theta_\omega\right) = K_{ds}V_a^{ccw}
\end{array}
\right.
$$

$$(5)$$

The effects of $V_a^{cw}, V_\theta^{cw}, V_a^{ccw}, V_\theta^{ccw}$ on $a_{cw}, \phi_{cw}, a_{ccw}, \phi_{ccw}$ are represented in (5). When $V_\theta^{cw} = V_\theta^{ccw} = 0$, (5) can be further solved as:

$$
\left\{
\begin{array}{l}
\dot{\phi}_{cw} = \omega + 2K\Omega \\
\dot{\phi}_{ccw} = \omega - 2K\Omega
\end{array}
\right.
$$

$$(6)$$

The synthesis of the standing wave by the CW/CCW traveling waves has the primary and quadrature modes. The amplitude amplification voltages of the primary and quadrature modes are a and q, where $a = a_{ccw} + a_{cw}$ and $q = a_{ccw} - a_{cw}$. The slow variable model of the standing wave can be solved by (5) as:

$$
\left\{
\begin{array}{l}
\dot{a} = -\left[\dfrac{1}{\tau} + \dfrac{1}{2}\Delta\left(\dfrac{1}{\tau}\right)\cos 4\left(\theta - \theta_\tau\right)\right]a + K_{ds}\dfrac{\left(V_a^{cw} + V_a^{ccw}\right)}{2\omega} \\[3mm]
\dot{q} = -\dfrac{1}{2}\Delta\omega\sin 4\left(\theta - \theta_\omega\right)a + K_{ds}\dfrac{\left(V_a^{ccw} - V_a^{cw}\right)}{2\omega} \\[3mm]
\dot{\theta} = -K\Omega + \dfrac{1}{4}\Delta\left(\dfrac{1}{\tau}\right)\sin 4\left(\theta - \theta_\tau\right) + K_{ds}\dfrac{\left(V_\theta^{cw} + V_\theta^{ccw}\right)}{4a\omega}
\end{array}
\right.
$$

$$(7)$$

The slow variable model of the DFM-WA HRG are represented in (5) and (7), which can provide theoretical support for the optimized design of the measurement and control scheme, and the analysis, modeling, identification, and compensation of the detection, drive, phase, and resonator errors.

4 The Performance Comparison of the DFM-WA HRG and AM-WA HRG

The HRG is installed on a signal-axis turntable, and its sensitive axis points to the sky. The Q-factor and frequency difference of the HRG are more than 10 million and less than 1mHz. The resonant frequency of the HRG is about 5267Hz.

The test of the SF nonlinearities of the same HRG based on the DFM-WA and AM-WA schemes require the input angular velocities at 0°/s, ±0.1°/s, ±0.2°/s, ±0.5°/s, ±1°/s, ±2°/s, ±5°/s, ±10°/s, ±20°/s, ±50°/s, ±100°/s, ±200°/s, ±500°/s in turn. The sample frequency of the HRG is 100Hz. The validity data in each rate is 30s, and the final output angular velocity at each rate is the average of the validity data. The SF of the HRG can be obtained by the linear fitting of the input and final output angular velocity. The SF nonlinearity at each rate can be calculated by the difference between the final out angular velocity and its fitting value. The SF nonlinearity of the

HRG is the maximum value of the differences at all rates divided by the range of the input angular velocities. The SF nonlinearity of the DFM-WA HRG and AM-WA HRG are 0.5356ppm and 1.0095ppm as shown in Fig.2(a). The 100Hz static and 100°/s dynamic angular velocity outputs are shown in Fig.2(c) and (d). The static and 100°/s dynamic AVNs of the DFM-WA HRG are about 0.2°/s, and no rate-dependent. The static AVN of the AM-WA HRG is about 2°/s, which is a magnitude larger than DFM-WA HRG. The 100°/s AVN of the AM-WA HRG is about 4°/s, two times larger than its static AVN.

The BI of the same HRG based on the DFM-WA and AM-WA schemes can be obtained by Allan variance analysis. The standing wave is released from 0°, and the angular velocity drift of the HRG is acquired by 1Hz sampling. The BI of the DFM-WA HRG and AM-WA HRG are 0.0094°/h and 0.0129°/h as shown in Fig2.(b).

(a)   (b)

(c)   (d)
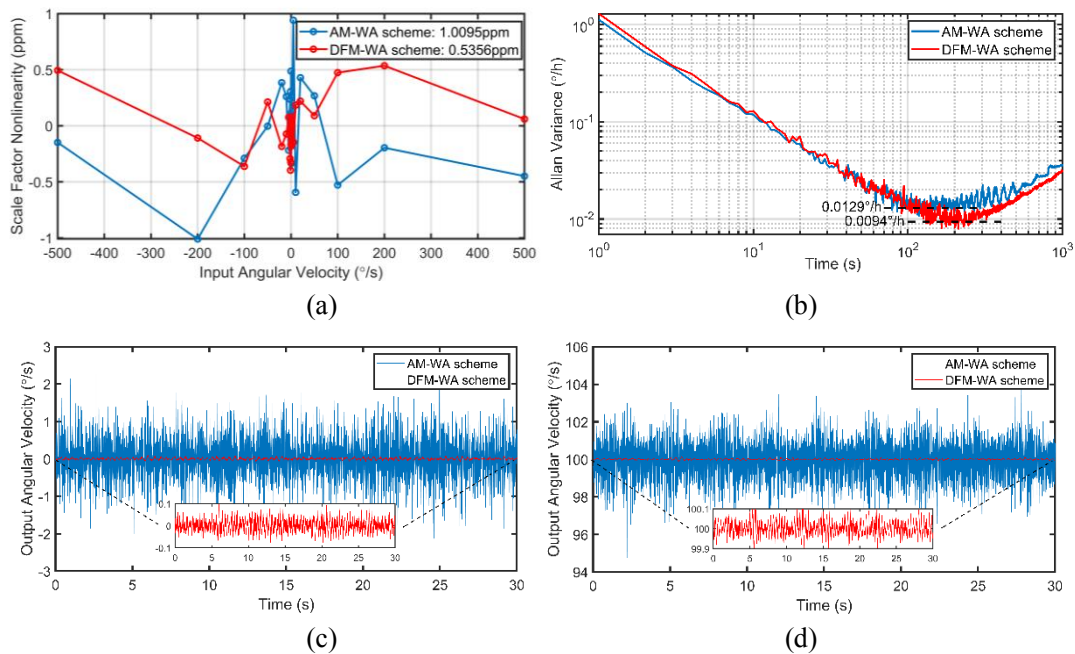
Fig.2   The performances of the same HRG based on the DFM-WA and AM-WA schemes, (a) the SF nonlinearities, (b) the BIs, (c) the static AVNs, (d) the 100°/s dynamic AVNs.

5   Conclusion

This paper realizes the measurement, control, and output of the HRG based on the DFM-WA scheme, and constructs the slow variable model of the DFM-WA HRG based on Lynch average method. The experiment results demonstrate that the DFM-WA HRG has the performance of large dynamics, high precision and low noise, the SF nonlinearity and BI both meet the requirement of the inertial-level. The SF nonlinearity, AVN, and BI of the DFM-WA HRG are 0.5356ppm, 0.2°/s, and 0.0094°/h, which reduce 47%, 90%, and 27% compared with the same HRG based on the AM-WA scheme. More importantly, the slow variable model of the DFM-WA HRG can guide the optimized design of the measurement and control scheme, and the analysis, modeling, identification, and compensation of the detection, drive, phase, and resonator errors.

References:

[1]   K. Yan, X. Wang, K. Zou, X, Zhang, T. Qu, C. Che, and Q. Lu. Self-excitation enabled decoupling, calibration, and compensation of errors for whole-angle hemispherical resonator gyroscope[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-13.

[2] R. Wang, G. Yi, W. Xie, and C. Wang. High-precision identification and compensation of nonlinear error in hemispherical resonator gyro[J]. Measurement, 2024, 237: 1-14.

[3] Z. Gao, R. Xu, F. Nan, Y. Han, D. Li, and Y. Zhang. Identification and calibration of the detection signal distortion in whole angle mode hemispherical resonator gyro[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-9.

[4] W. Chen, X. Ding, Z. Qin, and H. Li. Modeling, characterization, and compensation of detection and actuation coupling errors for whole-angle micro hemispherical resonant gyroscope[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-14.

[5] J. Sun, K. Liu, S. Yu, Y. Zhang, X. Xi, K. Lu, Y. Shi, X. Wu, and D. Xiao. Identification and correction of phase error for whole-angle micro-shell resonator gyroscope[J]. IEEE Sensors Journal, 2022, 22(20): 19228-19236.

[6] T. Tsukamoto and S. Tanaka. FM/Rate integrating MEMS gyroscope using independently controlled CW/CCW mode oscillations on a single resonator[C]. 4th IEEE International Symposium on Inertial Sensors and Systems, 2017: 1-4.

[7] T. Tsukamoto and S. Tanaka. Fully-differential single resonator FM/whole angle gyroscope using CW/CCW mode separator[C]. 30th IEEE International Conference on Micro Electro Mechanical Systems, 2017: 1118-1121.

[8] T. Tsukamoto and S. Tanaka. Virtually rotated MEMS whole angle gyroscope using independently controlled CW/CCW oscillations[C]. 5th IEEE International Symposium on Inertial Sensors and Systems, 2018: 89-92.

[9] T. Tsukamoto and S. Tanaka. MEMS rate integrating gyroscope with temperature corrected virtual rotation[C]. 6th IEEE International Symposium on Inertial Sensors and Systems, 2019: 1-4.

[10] T. Tsukamoto and S. Tanaka. Rate integrating gyroscope using independently controlled CW and CCW modes on single resonator[J]. Journal of Microelectromechanical Systems, 2020, 30(1): 15-23.

[11] T. Tsukamoto and S. Tanaka. Theoretical consideration of mismatch compensation for MEMS resonator having unaligned principle axes[C]. 8th IEEE International Symposium on Inertial Sensors and Systems, 2021: 1-4.

[12] T. Tsukamoto and S. Tanaka. Real time Q-factor mismatch detection for rate integrating gyroscope using amplitude modulated driving signal[C]. 9th IEEE International Symposium on Inertial Sensors and Systems, 2022: 1-4.

[13] J. Chen, T. Tsukamoto, G. Langfelder, and S. Tanaka. Virtually Rotated Multiple Mass Resonator Enabled by Electrostatic Frequency and Q-factor Tuning[C]. 10th IEEE International Symposium on Inertial Sensors and Systems, 2023: 1-4.

[14] T. Tsukamoto, F. Miyazaki, Y. Tomizawa, and S. Tanaka. Real Time Mismatch Monitoring Using Software Defined Frequency Modulation and Rate Integrating Gyroscope[C]. 37th IEEE International Conference on Micro Electro Mechanical Systems, 2024: 1-4.

[15] D. D. Lynch. Vibratory gyro analysis by the method of averaging[C]. 2th Saint Petersburg International Conference on Integrated Navigation Systems, 1995: 26-34.

A Method for Polar Transfer Alignment Based on Inertial Vector Matching

Li Xiangyuan[1,2], Zhao Yingwei[1,2], Tan Wenfeng[1,2], Zheng Jiaxing[1,2]

(1. College of Advanced Interdisciplinary Studies, National University of Defense Technology, Changsha 410073, China; 2. Nanhu Laser Laboratory, National University of Defense Technology, Changsha 410073, China)

Abstract: Due to the rapid convergence of Earth meridians in polar regions, the conventional transfer alignment algorithm, which uses the measurement information expressed in the local geographic frame, will encounter difficulties in providing accurate orientation. In this manuscript, a transfer alignment method using inertial vectors matching is proposed. The measurement equation is established by differentiating the inertial vectors of the two inertial navigation systems (INS), including angular velocity and specific force. This approach circumvents the need for transformation into the geographic frame, thereby addressing the challenges associated with traditional methods. To validate the proposed polar transfer alignment method, a virtual polar region is constructed through the simulated data in nonpolar regions, based on the grid attitude and velocity invariant principle. Simulation results demonstrate that the proposed transfer alignment method is effective in polar regions.

Key words: polar transfer alignment; virtual polar region; inertial navigation system; inertial vectors matching;

1 Introduction

The polar regions, with their abundant resources, potential for voyage reduction, and significant scientific research value, have garnered increasing attention from researchers[1-4]. Accurate navigation information is crucial for ensuring the safe operation of vessels in a harsh polar environment. In contrast to nonpolar regions, some navigation approaches are challenged by the complex geographical and severe climatic conditions of polar regions. The global navigation satellite system (GNSS) and magnetic-based instruments are unreliable in a polar region, owing to the active ionosphere and geomagnetic field variations[5]. The performance of a celestial navigation system (CNS) is affected by the polar day or polar night[6]. An inertial navigation system (INS) can autonomously provide attitude, velocity, and position information at a high rate, making it well-suited for polar navigation[7].

To enhance robustness, vessels are usually equipped with more than two INSs[8]. The master INS (MINS) incorporates a high-precision inertial measurement unit (IMU), while the slaver INS (SINS) employs a lower-cost and less accurate IMU. A transfer alignment (TA) technique is used to estimate the misalignment angle between them. In a nonpolar region, the traditional TA method

uses the attitude and velocity differences of two INSs as the observation, which is expressed in the local geographic frame[9]. However, the unique polar characteristic of convergent meridians will lead to an inaccurate TA. Specifically, north orientation defined in the local geographic frame will be rapidly changing in a polar region, particularly at the poles where the north orientation becomes indeterminate[10,11]. To address the issue associated with the local geographic frame, an inertial navigation mechanism in the grid or transverse geographic frame has been constructed[10-14].

The polar TA methods fundamentally employ attitude or attitude plus velocity matching equations as the observation within the grid geographic frame[15-17]. The grid inertial navigation error mechanism is different with the nonpolar scenario, which thus complicates the transformation of inertial navigation algorithms during the transition into or out of polar regions[1]. In this manuscript, a polar TA method using inertial vectors matching is proposed. Inertial vectors, including angular velocity and specific force, can be directly obtained from INS outputs and represent the angular and linear motions between the body frame and inertial frame. Since the inertial vectors matching does not involve projection onto the geographic frame, the TA method based on this approach is applicable to both nonpolar and polar scenarios, eliminating the need for back-and-forth switching like grid TA does.

Due to the limitations of geography, climate, and financial resources, it is hard to evaluate the performance of the polar TA methods in an authentic polar environment[18]. A simulated polar trajectory can be converted from the actual data in a region of low or medium latitude. Traditional polar trajectory generator operates according to the principle of unchanged attitude, velocity, and height, which cannot traverse the poles due to the singular values of yaw and position errors at the poles[11]. To facilitate the effective verification of polar TA methods, a virtual polar region technique has been developed. Similarly, the construction of virtual polar regions in the grid frame and the transverse frame adheres to the principles of invariable grid attitude and velocity for the former, and invariable transverse attitude and velocity for the latter[19,20]. In this manuscript, a grid virtual polar region is constructed through the simulated data in nonpolar regions.

2 Inertial Vectors Matching Method

Due to the misalignment between the MINS and SINS, the inertial vectors of them will be inconsistent. The inertial vectors matching equations can be written as[8]:

$$
\begin{cases}
\Delta\boldsymbol{\omega} = \left(\boldsymbol{\omega}_{is}^{s}\times\right)\boldsymbol{\varphi} + \dot{\boldsymbol{\varphi}} + \boldsymbol{\varepsilon} \\
\Delta\boldsymbol{f} = \left(\boldsymbol{f}_{is}^{s}\times\right)\boldsymbol{\varphi} + \left(\boldsymbol{\omega}_{im}^{m}\times\right)\left(\boldsymbol{\omega}_{im}^{m}\times\right)\boldsymbol{r} + \left(\dot{\boldsymbol{\omega}}_{im}^{m}\times\right)\boldsymbol{r} + 2\left(\boldsymbol{\omega}_{im}^{m}\times\right)\dot{\boldsymbol{r}} + \ddot{\boldsymbol{r}} + \nabla
\end{cases}
\tag{1}
$$

Where $\Delta\boldsymbol{\omega} = \boldsymbol{\omega}_{is}^{s} - \boldsymbol{\omega}_{im}^{m}$ is the angular velocity difference, $\boldsymbol{\omega}_{is}^{s}$ and $\boldsymbol{\omega}_{im}^{m}$ are the angular velocity outputs of the SINS and MINS, $\Delta\boldsymbol{f} = \boldsymbol{f}_{is}^{s} - \boldsymbol{f}_{im}^{m}$ is the specific force difference, $\boldsymbol{f}_{is}^{s}$ and $\boldsymbol{f}_{im}^{m}$ are the specific force outputs of the SINS and MINS, $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}_{s} - \boldsymbol{\varepsilon}_{m}$ is the gyroscope bias difference between the SINS and MINS, and $\nabla = \nabla_{s} - \nabla_{m}$ is the accelerometer bias difference between the SINS and MINS. $\boldsymbol{\varphi}$ is the misalignment angle between the SINS and MINS, while $\boldsymbol{r}$ is the lever arm. Considering that $\boldsymbol{\varphi}$ is a small angle, the misalignment matrix $\boldsymbol{C}_{s}^{m}$ can be approximated as

$C_s^m = I + (\boldsymbol{\varphi} \times)$, where I is a three-dimensional identity matrix.

The KF state vector is defined as $X(k) = \left[\boldsymbol{\varphi}(k), \boldsymbol{\varphi}(k-1), \boldsymbol{\varepsilon}(k), \boldsymbol{r}(k), \boldsymbol{r}(k-1), \boldsymbol{r}(k-2), \nabla(k) \right]^T$. $\boldsymbol{\varphi}$, $\boldsymbol{r}$, $\boldsymbol{\varepsilon}$, and ∇ are modelled as a combination of a random constant and white Gaussian noise (WGN). The KF state equation can then be written as:

$$X(k) = AX(k-1) + GW(k-1) \tag{2}$$

where k is the sampling time epoch, the process noise matrix W is written as $W = \left[w_{\varphi}, w_{\varepsilon}, w_r, w_{\nabla} \right]^T$, the state transition matrix A and the process noise coupling matrix G are expressed as follows.

$$A = \begin{bmatrix} I_{3\times3} & 0_{3\times3} & 0_{3\times15} \\ & 0_{3\times21} & \\ 0_{3\times6} & I_{3\times3} & 0_{3\times12} \\ 0_{3\times9} & I_{3\times3} & 0_{3\times9} \\ & 0_{6\times21} & \\ 0_{3\times9} & 0_{3\times9} & I_{3\times3} \end{bmatrix} \quad G = \begin{bmatrix} I_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & I_{3\times3} & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & I_{3\times3} & 0_{3\times3} \\ 0_{6\times3} & 0_{6\times3} & 0_{6\times3} & 0_{6\times3} \\ 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & I_{3\times3} \end{bmatrix} \tag{3}$$

Referring to (1), the KF measurement equation is written as:

$$Z(k) = H(k)X(k) + V(k) \tag{4}$$

where $Z(k) = \left[\Delta\boldsymbol{\omega}(k), \Delta\boldsymbol{f}(k) \right]^T$, V is the measurement noise while its covariance matrix $E(VV^T)$ is expressed as R. $H(k)$ is the measurement matrix which is expressed as:

$$H(k) = \begin{bmatrix} F_s I + \boldsymbol{\omega}_{is}^s(k) \times & -F_s I & I & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\ \boldsymbol{f}_{is}^s(k) \times & 0_{3\times3} & 0_{3\times3} & H_1(k) & H_2(k) & F_s^2 I & I \end{bmatrix}$$

$$H_1(k) = \left[\boldsymbol{\omega}_{im}^m(k) \times \right]\left[\boldsymbol{\omega}_{im}^m(k) \times \right] + \left\{ \left[\boldsymbol{\omega}_{im}^m(k) \times \right] - \left[\boldsymbol{\omega}_{im}^m(k-1) \times \right] \right\} F_s + 2\left[\boldsymbol{\omega}_{im}^m(k) \times \right] F_s + F_s^2 I \tag{5}$$

$$H_2(k) = -2\left[\boldsymbol{\omega}_{im}^m(k) \times \right] F_s - 2F_s^2 I$$

where F_s is the sampling frequency.

3 Results and Discussion

In this section, a virtual polar region is constructed based on the principle of invariant grid attitude and velocity from the simulated nonpolar data. The SINS data is converted from the MINS data through the configuration of the misalignment angle and lever arm, which is further used to verify the proposed TA method.

3.1 Virtual Polar Region Construction

The simulated nonpolar data is used to generate a virtual polar region trajectory based on the virtual polar region technique as presented in [19]. The attitude and velocity expressed in the n frame are transformed into the grid attitude and velocity. The initial position of the virtual polar

region is set to the north pole with the same longitude as the actual initial position. The polar position is then constructed by integrating the grid velocity and initial polar position. The IMU data is reconstructed by replacing the Earth's rotational angular velocity, positional angular velocity, and gravitational acceleration from the non-polar regions with their corresponding values in the virtual polar regions. The trajectories in the virtual polar region and actual nonpolar region are compared in Fig.1.
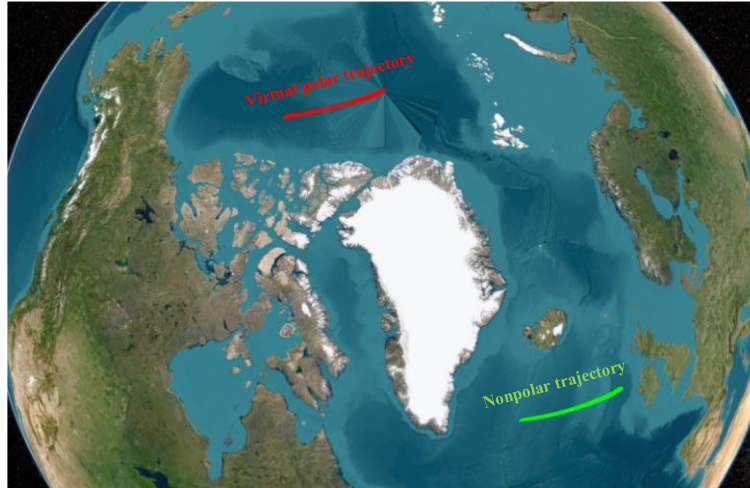
Fig.1   Trajectories in the virtual polar region and actual nonpolar region

Fig.1 demonstrates that trajectories in virtual polar and actual nonpolar regions are coincident, which validates the virtual polar region technique.

3.2   Simulation

The trajectory data includes the navigation information and IMU data of the MINS. The misalignment angle and lever arm are configured to simulate the data for SINS, which is detailed in Tab.1. The sensor error parameters are listed in Tab.2.

Tab.1   Specifications of misalignment angle and lever arm

Type	Constant	Standard deviation of WGN
Misalignment angle	[0.1°,0.1°,0.1°]	0.001°
Lever arm	[0,20m,0]	0.001m

Tab.2   Inertial sensor error parameters

Type	Instrument	Bias	Noise
MINS	Gyroscope	0.003°/h	0.0005°/√h
	Accelerometer	20mGal	10mGal/√Hz
SINS	Gyroscope	0.01°/h	0.0015°/√h
	Accelerometer	100mGal	20mGal/√Hz

The misalignment angle errors of the simulated nonpolar and polar regions are shown in Fig.2. The upper subplots present the comprehensive results, while the lower subplots display the detailed results within the 0-5min interval.

(a) Nonpolar region;

(b) Virtual polar region;
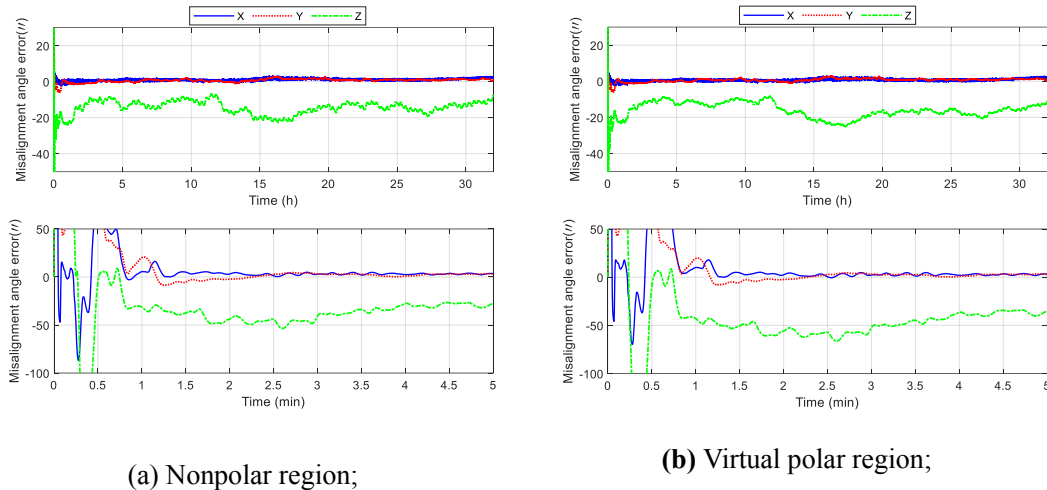
Fig.2   Misalignment angle error

Fig.2 indicates that the misalignment angle errors in polar regions can achieve the similar level as the nonpolar region, verifying the effectiveness of the proposed TA method. The horizontal misalignment angles errors are about a few arc-seconds, whereas the azimuth misalignment angle error is approximately 20 arc-seconds. Due to the cross-effect of gravitational acceleration in the specific force matching as illustrated in (1), the horizontal misalignment angles are more observable than the azimuth. As can be seen from the lower subplots, the transfer alignment results can be convergent within 2 minutes, which satisfies the requirement for rapidity.

4   Conclusion

In order to address the issue that traditional transfer alignment methods will suffer from the convergent meridians in polar regions, a transfer alignment method using inertial vectors matching is proposed in this manuscript. The inertial vectors relationships between the MINS and SINS are modelled to estimate the misalignment angle between them. The proposed method eliminates the need for the transformation of geographic frames, making it well-suited for both nonpolar and polar regions. A virtual polar region is constructed to validate the proposed transfer alignment method. The simulation results demonstrate that the estimation accuracy of misalignment angle in the virtual polar region are similar to that in the actual nonpolar region.

References:

[1]   Bian Hong-wei, Wen Zhe, Ma Heng, Wang Rong-ying. Error model of IMU data conversion in polar navigation evaluation at middle and low latitudes [J]. Journal of Chinese Inertial Technology, 2021, 29(3): 406-413.

[2]   Liu Chan, Wu Wen-qi, Feng Guo-hu,Wang Mao-song. Comprehensive calibration algorithm for rotational inertial navigation system based on n-vector model in polar regions[J]. Journal of Chinese Inertial Technology, 2023, 31(2): 107-113.

[3]   Cheng Jian-hua, Liu Jia-xin, Zhao Lin. Survey on polar marine navigation and positioning system[J]. Chinese Journal of Ship Research, 2021, 16(5): 16-29.

[4]   Peshekhonov V.G. Navigation aids for submarine polar missions[J]. Gyroscopy and Navigation, 2022, 13: 120–127.

[5]   Cheng Jian-hua, Wang Tong-da, Wang Lu, Wang Zhen-min. A new polar transfer alignment algorithm with the aid of a star Sensor and based on an adaptive Unscented Kalman Filter[J]. Sensors, 2017, 17(10): 2417.

[6]   Cai Jing, Cheng Jian-hua, Liu Jia-xin, Wang Zhen-min, and Xu Yue-hang. A polar rapid transfer alignment assisted by the improved polarized-light navigation[J]. IEEE Sensors Journal, 2022, 22(3): 2508-2517.

[7]   Xu Xiang, Ning Xiu-li, Yao Yi-qing, Li Kui. In-Motion coarse alignment method for SINS/GPS integration in polar region[J]. IEEE Transactions on Vehicular Technology, 2022, 71(6): 6110–6118.c

[8]   Li Xiang-yuan, Qin Shi-qiao, Wang Xing-shu, Tan Wenfeng, Zheng Jia-xing, Zhao Ying-wei. Multi inertial navigation system fusion method considering ship deformation[C]. 2023 30th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS), Saint Petersburg, Russian Federation: IEEE, 2023: 1–10.

[9]   Tosoni L, Pham M.T, Massioni P, Broussard E. Rapid transfer alignment for large and time-varying attitude misalignment angles[J]. IEEE Control Systems Letters, 2023, 7: 1981-1986.

[10] Zhou Qi, Yue Ya-zhou, Zhang Xiao-dong, Tian Yu. Indirect grid inertial navigation mechanization for transpolar aircraft[J]. Journal of Chinese Inertial Technology, 2014, 22(1): 18-22, 66.

[11] Li Qian, Sun Feng, Ben Yue-yang, Yu Fei. Polar navigation of strapdown inertial navigation system based on transversal frame in polar region[J]. Journal of Chinese Inertial Technology, 2014, 22(3): 288-295.

[12] Yao Yi-qing, Xu Xiao-su, Li Yao, Liu Yi-ting, Sun Jin, and Tong Jin-wu. Transverse navigation under the ellipsoidal Earth model and its performance in both polar and non-polar areas[J]. The Journal of Navigation, 2016, 69(2): 335-352.

[13] Qin Fang-jun, Chang Lu-bin, Li An. Improved transversal polar navigation mechanism for strapdown INS using ellipsoidal Earth model[J]. The Journal of Navigation, 2018, 71(6): 1460–1476.

[14] Bian Hong-wei, Lin Xiu-xiu, Wang Rong-ying, Ma Heng. Ellipsoid model navigation method based on unified transverse coordinate system in polar region[J]. Journal of Chinese Inertial Technology, 2018, 26(5): 579-584.

[15] Cheng Jian-hua, Cai Jing, Wang Zhen-min, and Liu Jia-xin. A Novel polar rapid transfer alignment for shipborne SINS under arbitrary misalignments[J]. IEEE Access, 2020, 8: 197567–197580.

[16] Cai Jing, Y. Xu, Cheng Jian-hua, and Liu Jia-xin. Application of Unscented Kalman Filter with neural network in the polar rapid transfer alignment[C]. 2022 29th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS), Saint Petersburg, Russian Federation: IEEE, 2022: 1–4.

[17] Wang Yong-jun, Xu Jingshuo, Yang Bo. A new polar rapid transfer alignment method based on grid Frame for shipborne SINS[J]. IEEE Sensors Journal, 2022, 22(16): 16150-16163.

[18] Wu Feng, Shao Tian-yi, Gu Cong, Fu Qiang-wen, and Xu Ya-fen. Virtual polar region method based on the Earth's transverse ellipsoid model[J]. Mathematical Problems in Engineering, 2021, 1: 1-9.

[19] Fu Qiang-wen, Zhou Qi, Yan Gong-min, Li Si-hai, Wu Feng. Unified all-earth navigation mechanization and virtual polar region technology[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 8501211.

[20] Huang Ling, Xu Xiang, Ge Hao-ran, Zhao He-ming. Robust Unscented quaternion Kalman Filter for polar in-motion alignment and virtual polar region technology[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 8502510.

The Methods for Measuring the Number Density and Polarization of Alkali Atoms in Vapor Cells

Huang Chenyu[1,2], Lei Xing[1,2], Cao Yaohui[1,2], Chen Yuanliang[1,2], Yue Yazhou[1,2]

(1. AVIC Xi'an Flight Automatic Control Research Institute, Xi'an 710065, China;

2. National Key Laboratory of Science and Technology on Aircraft Control, Xi'an 710065, China)

Abstract: Alkali vapor cells are core units in nuclear magnetic resonance gyroscopes, which measure angular rate by probing the resonant frequency drift of noble gas in a magnetic field. The number density and polarization of alkali atoms, both temperature-dependent, affect noble gas polarization and hence gyroscope performance. Optimal temperature yields high signal-to-noise ratio. This paper introduces modified spectral absorption method for number density measurement and paramagnetic absorption method for polarization measurement, both based on quantum properties of alkali atoms, achieving high precision and reliability.

Keywords: nuclear magnetic resonance gyroscope; number density; polarization; optical depth; electron paramagnetic resonance

1 Introduction

Inertial navigation systems, independent of external signals, are widely used in aviation navigation. Compared with traditional optical gyroscopes, quantum gyroscopes are theoretically no longer constrained by volume [1,2]. They, especially nuclear magnetic resonance gyroscopes (NMRGs) with theoretically infinite dynamic range [3], represent a significant development direction. The core sensitive unit of NMRGs is the alkali vapor cell [4], where polarization of alkali atoms is transferred to noble gas nuclei through spin-exchange collisions. Optimal temperature maximizes noble gas polarization [5] and gyroscope signal-to-noise ratio. Measuring noble gas polarization directly is difficult. Thus, measuring alkali atom number density and polarization is crucial for evaluating NMRGs performance.

Existing methods for measuring alkali atom number density include empirical formula, optical rotation detection, and spectral absorption method. Empirical formula overestimates due to temperature gradients and anti-relaxation coating effects [6]. Optical rotation detection is based on the Faraday rotation angle [7]. To achieve a noticeable angle change, it requires high-intensity magnetic fields and long optical paths, unsuitable for miniaturized NMRGs. An alternative is Faraday modulation on the probe laser [8], though it complicates the system. Spectral absorption method offers in-situ, high-precision measurement [9] but needs improvement in fitting function. For polarization measurement, empirical formula and optical rotation detection are used [10], with optical rotation detection more common but suitable only for long optical paths.

This paper proposes modified spectral absorption method (MSAM) for number density and

---

paramagnetic absorption method (PAM) for polarization measurement, supporting further NMRGs performance improvement.

2 Measurement Methods

2.1 Modified Spectral Absorption Method

MSAM measures alkali atom absorption of linearly polarized laser at different frequencies to fit number density. By measuring laser detuning and intensity attenuation, the optical depth (OD) spectrum is fitted. OD is defined as

$$\text{OD} = -\ln \frac{I_{\text{out}}}{I_{\text{in}}} = n_{\text{Rb}} \sigma(v) l \tag{1}$$

where I_{in} and I_{out} are incident and outgoing laser intensities, respectively. The relationship involves number density n_{Rb}, absorption cross-section $\sigma(v)$, and optical path length l.

There are Gauss, Lorentz and Voight line patterns of cross-section. They depend on natural linewidth, pressure broadening, and Doppler broadening. When there is less pressure broadening, Voigt tends towards Gauss. On the contrary, Voigt tends towards Lorentz. Due to high internal pressure from alkali vapor, noble gas, and buffer gas, the absorption spectrum can be simplified from Voigt to Lorentz. The full width at half maxima (FWHM) of Lorentz line pattern Γ_L is [11]

$$\Gamma_L = \Gamma_{\text{nat}} + \Gamma_{\text{pre}} = \left(\frac{p}{p_0} \right) \left(\frac{T_0}{T} \right)^d \Gamma_0 \tag{2}$$

where p_0=101325Pa, T_0=273.15K. It depends on buffer gas pressure p, actual temperature T, and FWHM Γ_0 induced by per amg gas. In N_2, FWHM of D_1 line and D_2 line are $\Gamma_{0,D1}$=17.8GHz/amg, $\Gamma_{0,D2}$=18.1GHz/amg at 353K. d is the temperature dependence, which is taken as d=0.3 for matching with temperature 353K. The Lorentz line pattern function is

$$L(v - v_0) = \frac{\Gamma_L / 2\pi}{(v - v_0)^2 + (\Gamma_L / 2)^2} \tag{3}$$

where v_0 is the resonance frequency of ^{87}Rb. The cross-section $\sigma(v)$ is expressed as

$$\sigma_L(v) = \pi r_e c f_{\text{res}} L(v - v_0) \tag{4}$$

where v is the laser frequency, also considering classical electron radius r_e, speed of light c, and resonance intensity f_{res}. For ^{87}Rb, $f_{D1} \approx 0.332$, $f_{D2} \approx 0.668$.

For known OD, the commonly used fitting function is unimodal. However, it's actually caused by power broadening [12]. Due to hyperfine splitting of alkali atomic energy levels, the OD spectrum exhibits a multimodal form [13] rather than unimodal. For ^{87}Rb, the D_2 transition ground state splits into 2 sublevels with a spacing of 6.8GHz, and the excited state splits into 4 sublevels. Considering pressure broadening limitations, only the influence of 2 ground state levels is significant. Thus, OD is modified to the superposition of 2 Lorentz functions as

$$\text{OD} = n_{\text{Rb}} K \left[\alpha \frac{\Gamma_L / 2\pi}{(v - v_1)^2 + (\Gamma_L / 2)^2} + (1 - \alpha) \frac{\Gamma_L / 2\pi}{(v - v_2)^2 + (\Gamma_L / 2)^2} \right] + I_{\text{loss}} \tag{5}$$

where $K = \pi l r_e c f_{\text{res}}$ is the product of known constants. I_{loss} represents the optical intensity loss. α

is the total relative intensity of transition from ground state $F=2$ to excited state. v_1 and v_2 are center frequencies. For bimodal OD, n_{Rb} is obtained by fitting the above equation.

The optical path for measuring n_{Rb} is shown in Fig. 1. In the figure, HWP is a half-wave plate, PBS is a polarized beam-splitter, PD is a photodetector, and DAS is a data acquisition system.
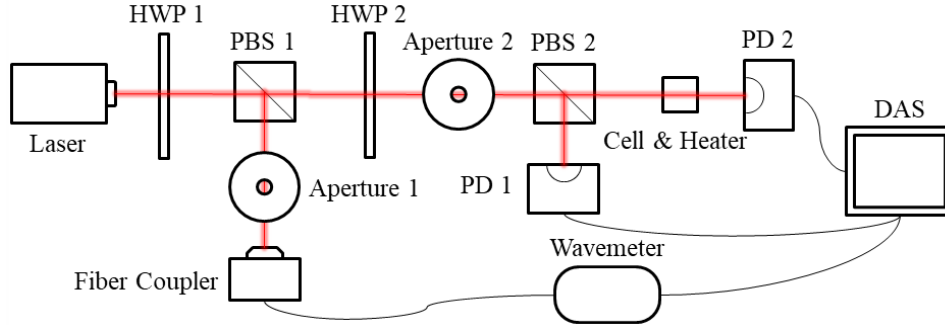
Fig.1 The optical path for MSAM

2.2 Paramagnetic Absorption Method

PAM's theoretical basis lies in Electron Paramagnetic Resonance (EPR) [14]. In an external magnetic field B, ^{87}Rb experiences Zeeman splitting $\Delta E = g\mu_B B$. When a pump laser with frequency v matches the Zeeman energy level spacing $hv = g\mu_B B$, resonance occurs. EPR refers to the transition between the Zeeman energy levels in ground state of unpaired electrons [15].

According to the selection rule, using a D1 left-handed circularly polarized pump laser at 794.98nm, atoms concentrate in the ground state $5S^2_{1/2}$. It's the ^{87}Rb polarization process. During stable polarization, laser intensity attenuation is fixed. During EPR, atoms on $\left| F=2, m_F=2 \right\rangle$ rapidly decay to other Zeeman levels and are repumped, causing greater laser intensity attenuation. Polarized atom number density n_P is calculated using laser intensities at EPR I_{EPR} and stable polarization I_P, and the D_1 cross-section value at v_0, $\sigma_1(v_0)$. n_P can be calculated by

$$I_{EPR} = I_P \exp\left[-n_P \sigma_1(v_0) l\right] \tag{6}$$

Measuring n_{Rb} by pump laser using MSAM, the polarization is

$$R_P = \frac{n_P}{n_{Rb}} \tag{7}$$

The optical path for measuring R_P is shown in Fig. 2. QWP is a quarter-wave plate.
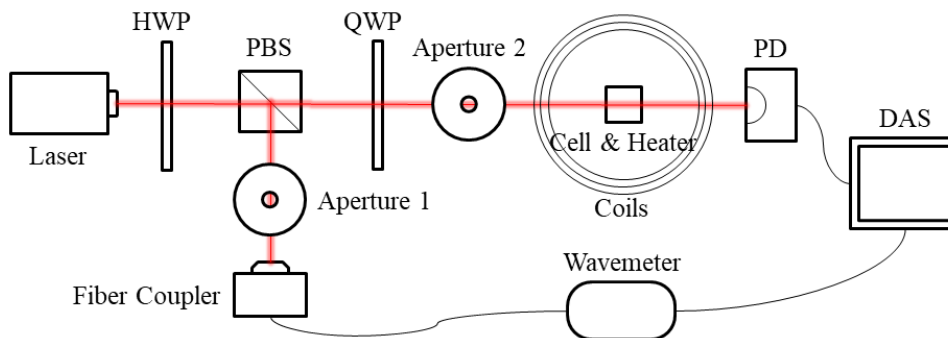
Fig.2 The optical path for PAM

3 Experiment

3.1 Number Density Measurement

MSAM measured n_{Rb} in a cell over a temperature range of 70°C to 130°C, in steps of 10°C, with a probe laser power of 350μW.
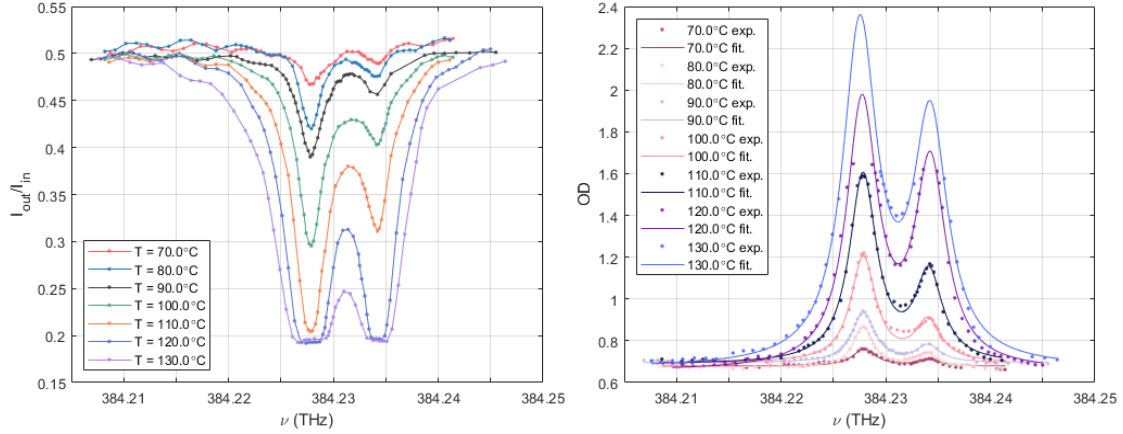
Fig.3 D$_2$ absorption spectrums, OD and fitting curves

The D$_2$ absorption spectrum line pattern was a superposition of 2 Lorentz functions (bimodal form). At 120°C and 130°C, saturation absorption at resonance frequency resulted in incomplete spectra. OD was calculated excluding saturation points. The absorption spectrums, OD and fitting curves are shown in Fig. 3.

Fig. 4 displays the fitting results. FWHM conforms to equation (2), and demonstrates good agreement with theoretical expectations. The center frequency difference of the bimodal spectrum, approximately 6.45GHz (±0.1GHz), aligns with the theoretical value of 6.8GHz, taking into account the excited state hyperfine energy levels and experimental errors.
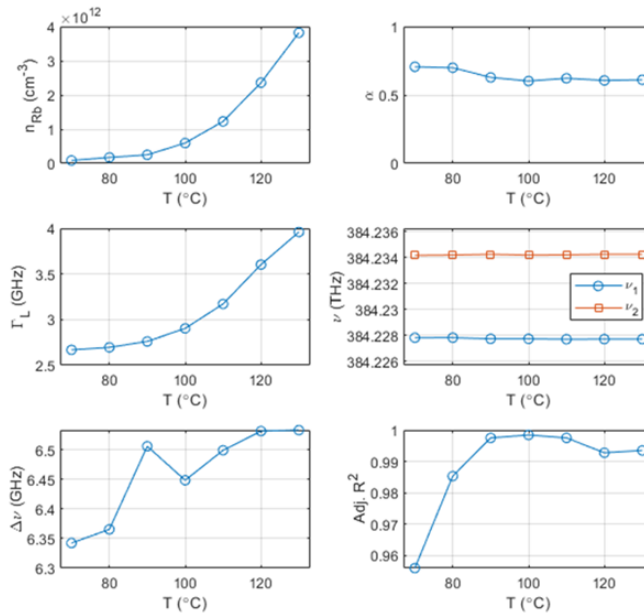
Fig.4 Coefficient fitting results

We randomly selected 3 vapor cells from the same batch to conduct independent, repeated

experiments. The error range of the obtained number density didn't exceed 2.1%, which proves the effectiveness and reliability of MSAM.

3.2 Polarization Measurement

PAM measured R_p in the same cell over a temperature range of 70°C to 110°C, in steps of 10°C, with a magnetic field intensity of a periodic triangular wave.

Taking the example of 110°C as shown in Fig. 5, outgoing pump laser intensity variation showed flat parts I_p during stable polarization and deep peaks I_{EPR} during EPR. Additionally, due to the presence of a small amount of ^{85}Rb in the cell, the laser intensity also displayed periodic shallow peaks.
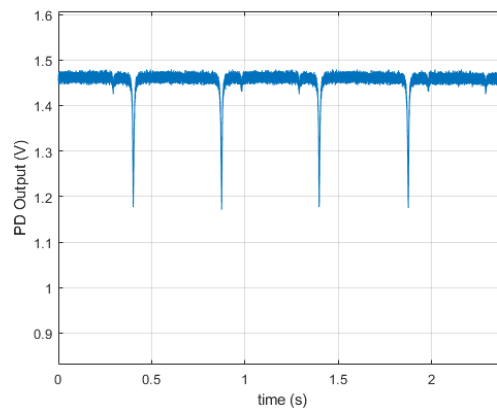
Fig.5   Outgoing pump laser intensity variation at 110°C

The results at various temperatures are shown in Fig. 6. It indicates that as temperature increases, n_p increases, whereas R_p first increases and then decreases due to varying efficiencies of the pump laser in polarizing the atoms.
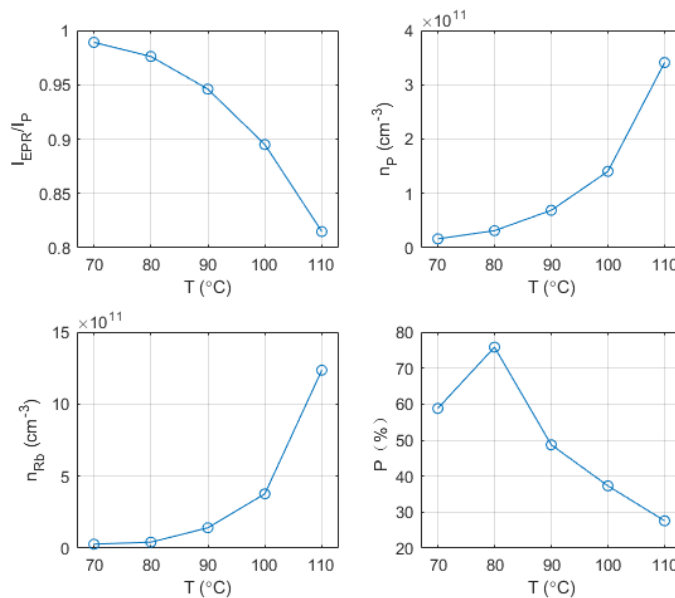
Fig.6   Measurement results

We randomly selected 3 vapor cells from the same batch to conduct independent, repeated experiments. The error range of the obtained polarization didn't exceed 2.7%, which proves the

effectiveness and reliability of PAM.

4 Conclusion

We have designed the modified spectral absorption method for number density measurement and the paramagnetic absorption method for polarization measurement, leveraging the quantum properties of alkali atoms to overcome the limitations of commonly used methods. The results obtained using the modified bimodal fitting function not only met theoretical expectations but also exhibited experimental agreement, achieving high precision and reliability. In random sampling tests, MSAM and PAM errors didn't exceed 2.1% and 2.7%, respectively. The proposed methods are beneficial for determining the optimal temperature for NMRG, suppressing noise-induced bias, and enhancing overall performance.

References:

[1] Chen Y, Liu Z, Liu G. Advances in Nuclear Magnetic Resonance Gyroscope [J]. Control Theory & Applications, 2019, 36(7): 1017-1023.

[2] Savoie D, Altorio M, Fang B, et al. Interleaved Atom Interferometry for High Sensitivity Inertial Measurements [C]. 2019 Joint Conference of the IEEE International Frequency Control Symposium and European Frequency and Time Forum. IEEE, 2019.

[3] Zhang Y, Wang Z, Jiang Q. Signal Simulation and Noise Analysis of Nuclear Magnetic Resonance Gyroscope [J]. Navigation and Control, 2020, 19(1): 18-26.

[4] Yue Y, Lei X, Cao Y, et al. Measuring Method of Rubidium Atom Transverse Relaxation Time Based on Alkali Metal Magnetometer [J]. Journal of Chinese Inertial Technology, 2021, 29(4): 496-501.

[5] Chen L, Zhou B, Lei G, et al. Effects of Temperature on Rb and ^{129}Xe Spin Polarization in a Nuclear Magnetic Resonance Gyroscope with Low Pump Power [J]. Aip Advances, 2017, 7(11): 115101.

[6] Eklund E J. Microgyroscope Based on Spin-Polarized Nuclei [D]. University of California, Irvine, 2008.

[7] Rosenberry M A, Reyes J P, Tupa D, et al. Radiation Trapping in Rubidium Optical Pumping at Low Buffer-gas Pressures [J]. Physical Review A, 2007, 75(2): 1-6.

[8] Han S. System of Extremely Small Optical Rotation Angles Detection [D]. Harbin Engineering University, 2012.

[9] Chi H, Wang X, Quan W. Pressure Measurement of Each Gas in Alkali-Metal Vapor Cell with a Mixed Gas Based on Saturated Absorption Spectrum [J]. Spectroscopy and Spectral Analysis, 2018, 38(3): 948-952.

[10] Shang H, Quan W, Chen Y, et al. The Measuring Method of Atomic Polarization of Alkali Metal Vapor Based on Optical Rotation and the Analysis of the Influence Factors [J]. Spectroscopy and Spectral Analysis, 2016, 36(2): 305-309.

[11] Pitz G A, Sandoval A J, Tafoya T B, et al. Pressure Broadening and Shift of the Rubidium D1 Transition and Potassium D2 Transitions by Various Gases with Comparison to Other Alkali Rates [J]. Journal of Quantitative Spectroscopy & Radiative Transfer, 2014, 140: 18-29.

[12] Wei X. Coherent Control of the Inhomogeneous Broadened Absorption Spectrum in Rubidium Vapor [D]. Jilin University, 2006.

[13] Han Y, Yan X, Deng L, et al. Study on Intensity Difference Mechanism of Rubidium Atom Hyperfine Spectra [J]. Journal of Northwest Normal University (Natural Science), 2021, 57(2): 41-47.

[14] Zhou S, Chen H. Quantum Mechanics Tutorial [M]. Beijing: Higher Education Press, 2022.

[15] Zhang Q, Zou S, Zhang H. Application and Progress of Residual Magnetometry Based on Electron Paramagnetic Resonance Spectroscopy [J]. Spectroscopy and Spectral Analysis, 2024, 44(1): 22-28.

A SINS/GNSS fault tolerant algorithm based on improved smooth bounded layer width[1]

Wang Jinbao，Zhao Guiling

(1. School of Geomatics, Liaoning Technical University, Fuxin, 123000, China)

Abstract: In the operational process of the Global Navigation Satellite System/Strapdown Inertial Navigation System (GNSS/SINS) integrated navigation system, the Global Navigation Satellite System (GNSS) signal is vulnerable to external interference, resulting in abnormal system measurement data, and system faults. These faults will reduce the navigation and positioning performance of the system and reduce the measurement accuracy of the system. Aiming at this problem, a GNSS/SINS fault detection and robust adaptive algorithm based on sliding average smooth bounded layer width is proposed. The algorithm evaluates the system measurement data based on the innovation residual and incorporates the sliding average filter to design the fault detection function based on the smooth bounded width layer. Accurate detection of system faults using fault detection function. The fault detection function value is used to construct the robust cofactor matrix to correct the measurement error in real-time, to improve the accuracy and robustness of the state estimation. The experimental results show that: The proposed algorithm in the paper compares with two traditional robust adaptive algorithms based on smooth bounded layer fault detection and residual chi-square fault detection. The detection sensitivity and filtering estimation accuracy of the proposed algorithm are improved. Throughout the fault, the filtering accuracy demonstrates an increase of more than 30.96% and 16.90%, respectively.

Key words: GNSS/SINS; smooth bounded layer width; sliding average filter; small step fault; slowly varying fault;

1 Introduction

The Global Navigation Satellite System (GNSS) and Strapdown Inertial Navigation System (SINS) have emerged as the predominant technologies in the advancement of integrated navigation systems due to their robust complementarity and reliability. The GNSS/SINS integrated navigation system enhances positioning accuracy in complex environments and finds extensive application in various sectors including UAV measurement and vehicle movement measurement [1,2]. During the operation of the GNSS/SINS integrated navigation system, the GNSS signal is highly susceptible to the interference of houses, trees, electromagnetism and other external factors resulting in abnormal system measurement data, and system faults. These faults affect the performance of the filter and reduce the navigation and positioning accuracy and robustness of the system [3]. Therefore, how to effectively enhance the fault-tolerant of the system becomes the key to improving the

navigation and positioning performance of the GNSS/SINS integrated navigation system [4].

In recent years, The residual chi-square fault detection algorithm is currently a commonly used method in engineering for fault detection [5-6]. This algorithm utilizes the innovation residual and its covariance matrix to create a fault detection function. However, it cannot effectively detect small step faults or slowly varying faults with small initial fault values and lacks the ability to effectively tolerate faults. To improve the efficiency of fault detection and enhance the fault-tolerant of the system. In [7], a fault detection algorithm is suggested using the sequential probability ratio test, where the likelihood of fault occurrence is assessed through a log-likelihood ratio function to determine the system's fault condition. Additionally, reference [8] introduces a fading factor to improve the sequential probability ratio fault detection method, addressing the issue of undetected fault endings and enhancing the algorithm's fault detection capabilities. The Sequential Probability Ratio Fault Detection algorithm exhibits a significant time delay in detecting the end of a fault. With prolonged fault duration, the algorithm's introduced fault isolation adaptive adjustment mechanism causes the accumulation of errors in the inertial navigation solution, leading to decreased system estimation accuracy. The smooth bounded layer width introduced in the Smooth Variable Structure Filter (SVSF) algorithm is gradually applied to the field of integrated navigation. [9]. Reference [10] examined the role of smooth bounded layer width on fault detection and introduced a robust adaptive algorithm for smooth bounded layer fault detection. Nevertheless, the fault detection model generated by the algorithm lacks a precise comparison standard, leading to significant missed detections that hinder effective system fault identification.

In view of the above problems, this paper studies the smooth bounded layer width model, and proposes a SINS/GNSS fault tolerance algorithm based on improved smooth bounded layer width combined with moving average filtering. The algorithm evaluates the prior quality of the system measurement data based on the innovation residual, selects a part of the fault-free period data to solve the overall standard deviation of the smooth bounded layer width, and combines the moving average filter to establish a fault detection function based on the smooth bounded layer width, so as to accurately detect the GNSS signal fault and improve the fault detection efficiency. The algorithm uses the fault detection function value to construct the robust cofactor matrix, and combines the robust estimation to correct the error in real time, reduce the influence of the fault on the filter, and improve the accuracy and robustness of the system state estimation.

2   Analysis of smooth bounded layer width

The concept of smooth bounded layer width is proposed in the gain matrix K_k of SVSF algorithm[9].Smooth bounded layer width can reflect the average accuracy of integrated navigation system estimation and reflect the interference of filter. The gain matrix K_k is expressed as :

$$K_k = H_k^{-1} \begin{bmatrix} diag(|e_{k/k-1}| + \gamma |e_{k-1/k-1}|) \\ \circ sat(\varphi_k^{-1} diag(e_{k/k-1}) \end{bmatrix} \left[diag(e_{k/k-1}) \right]^{-1} \tag{1}$$

where γ represents the memory factor, and $0 < \gamma < 1$, $e_{k/k-1} = Z_k - H_k X_{k/k-1}$ represents the residual of prior innovation at the time k , $e_{k-1/k-1} = Z_{k-1} - H_{k-1} X_{k-1}$ denotes the posterior innovation residual at the time $k-1$, ' $\circ$ ' represents the Hadmard product, φ_k represents the smooth bounded layer width matrix at the time k , $sat()$ is the saturation operation of each

element of the column vector in '()', The calculation formula of the smooth bounded layer width matrix $\boldsymbol{\varphi}_k$ can be obtained by simplifying formula (1):

$$\boldsymbol{\varphi}_k = \left\{ \left[diag(|\boldsymbol{e}_{k/k-1}| + \gamma |\boldsymbol{e}_{k-1/k-1}|) \right]^{-1} \boldsymbol{H}_k \boldsymbol{P}_{k/k-1} \boldsymbol{H}_k^{\mathrm{T}} \boldsymbol{T}_k^{-1} \right\}^{-1} \tag{2}$$

It can be seen from formula (2) that the width of the smoothed bounded layer is a function composed of the prior innovation residual at the current moment, the measurement matrix, the prior state estimation covariance matrix, the innovation residual covariance matrix and the posterior innovation residual at the previous moment. The smooth bounded layer width is closely related to the system noise and measurement noise estimated by the system model, which can reflect the level of interference in the filter. The stability of the system model and system noise is generally considered to be good, while the measurement noise is more susceptible to external environmental factors and has a greater impact on the filter. As a result, the smooth bounded layer width offers a novel approach for the construction of the GNSS/SINS fault detection function.

3 The GNSS/SINS fault-tolerant algorithm design

3.1 The fault detection based on improved smooth bounded layer width

The innovation residual $\boldsymbol{e}_k$ and the innovation residual covariance matrix $\boldsymbol{T}_k$ are normalized to obtain $\boldsymbol{E}_k$. Select some fault-free period data to obtain the overall standard deviation of the smooth bounded layer width of the fault-free period data.

$$\tilde{\sigma} = \sqrt{\sum_{l=1}^{B} (\boldsymbol{\varphi}_1 - \bar{\boldsymbol{\varphi}})^2 (\boldsymbol{\varphi}_2 - \bar{\boldsymbol{\varphi}})^2 \dots (\boldsymbol{\varphi}_L - \bar{\boldsymbol{\varphi}})^2 \Big/ B} \tag{3}$$

where B represents the number of selected fault-free period data, in this paper B is 300. $\bar{\boldsymbol{\varphi}}$ represents the average of the smoothed bounded layer width. The smooth bounded layer width $\boldsymbol{\varphi}_k$ is compared with the overall standard deviation $\tilde{\sigma}$ of the smooth bounded layer width in real-time:

$$\boldsymbol{\vartheta}_k = \boldsymbol{\varphi}_k / \tilde{\sigma} \tag{4}$$

$$\boldsymbol{\mu}_k = \boldsymbol{\vartheta}_k^{\mathrm{T}} \boldsymbol{\vartheta}_k \tag{5}$$

In the process of utilizing the innovation residual to initially screen the measurement data and calculate the overall standard deviation of the smooth bounded layer width, it is important to consider the impact of random errors. To minimize the effect of these errors and enhance the efficiency of fault detection, a fault detection function based on moving average filtering is introduced in conjunction with the formula (5).

$$\lambda_k = \begin{cases} \mu_k, & k < N \\ \dfrac{1}{N} \displaystyle\sum_{n=k+1-N}^{k} \mu_k, & k \geq N \end{cases} \tag{6}$$

where N is the length of the sliding window, After many experiments, in this paper N is taken as 5.

The threshold Z is selected, and the fault detection function value λ_k is compared with the threshold Z. When the fault detection function λ_k is less than or equal to the threshold Z, it indicates that the measurement data is normal and the system has no fault. When the fault detection function λ_k exceeds the threshold Z, it indicates that the measurement data is abnormal and the system fails. This paper Z is taken to be 150.

3.2 The robust adaptive based on improved smooth bounded layer width

The equivalent measurement information is abnormal. After the system detects the fault, the fault detection function value is used to construct the robust cofactor matrix α_k.

$$\alpha_k = \begin{cases} 1, & \lambda_k \leq Z \\ \dfrac{Z}{\lambda_k}, & \lambda_k > Z \end{cases} \tag{7}$$

Combined with the robust estimation theory, the weight of the measurement noise in the Kalman filter gain matrix K_k is adjusted in real-time by the robust cofactor matrix to reduce the influence of the fault on the system accuracy.

$$K_k = P_{k/k-1} H_k^{\mathrm{T}} (H_k P_{k/k-1} H_k^{\mathrm{T}} + \alpha_k^{-1} R_k)^{-1} \tag{8}$$

$$X_k = X_{k/k-1} + K_k (Z_k - H_k X_{k/k-1}) \tag{9}$$

The corresponding state estimation covariance matrix P_k is:

$$P_k = (I - K_k H_k) P_{k/k-1} (I - K_k H_k)^{\mathrm{T}} + K_k \alpha_k^{-1} R_k K_k^{\mathrm{T}} \tag{10}$$

When the system detects a fault with abnormal equivalent measurement information, the measurement noise covariance matrix R_k is adaptively adjusted by the robust cofactor matrix α_k. This adjustment reduces observation gain and enhances the accuracy of the integrated navigation system.

4 Experimental verification and analysis

In order to assess the effectiveness and reliability of the proposed algorithm, a GNSS/SINS integrated navigation experimental platform was established for conducting simulation experiments and analysis. Two types of typical GNSS faults, small step and slowly varying, were introduced into the simulation data of the GNSS/SINS integrated navigation experimental platform. The traditional robust adaptive algorithm (M1) based on smooth bounded layer fault detection, the robust adaptive algorithm (M2) based on residual chi-square fault detection, and the SINS/GNSS fault tolerant algorithm(M3) based on improved smooth bounded layer width proposed in this paper are used for processing. the simulation running trajectory is shown in figure 2. The equipment parameters of the GNSS/SINS integrated navigation experimental platform are shown in table 1.
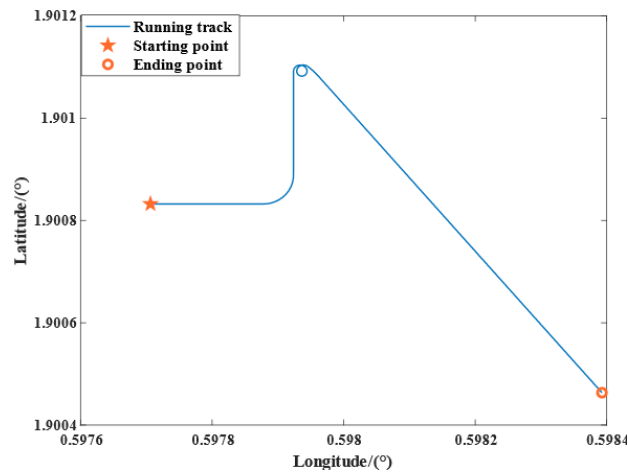
Fig.1. Running track diagram

Tab.1   Equipment parameters of GNSS/SINS integrated navigation experimental platform

Equipment	Equipment parameter type	Parameter setting	Frequency / Hz
SINS	Gyro bias instability	$0.3\,°/h$	200
	Gyro noise	$0.15\,°/\sqrt{h}$	200
	Accelerometer bias instability	$0.05\,mg$	200
	Accelerometer noise	$0.06\,m/s/\sqrt{h}$	200
GNSS	Velocity error	$0.1\,m/s$	10
	Position error	$10\,m$	10

A small step fault is assumed to occur between 301s and 400s with a location fault size of 15m, while the slowly varying fault occurs between 601s and 700s with a position retardation pattern of $\delta_r = 0.3*(k-600)$. Three fault detection algorithms such as M1, M2 and M3 are used for fault detection respectively, and the fault detection effect of the three fault detection algorithms is obtained as shown in figure 2. The robust adaptive algorithms of M1 and M2 and the M3 robust adaptive algorithm proposed in this paper are used for processing. The position error curves of the three algorithms are shown in figure 3.
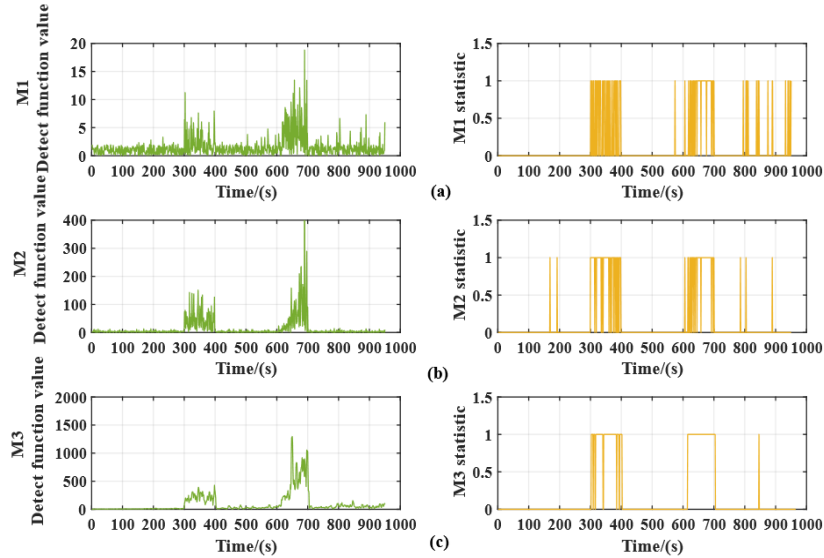
Fig.2   Fault detection effect of three fault detection algorithms
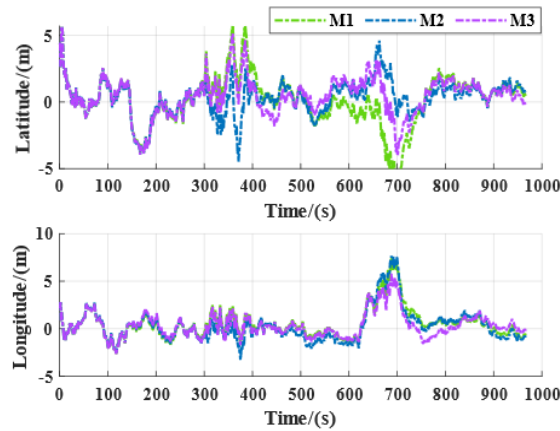
Fig.3 Position error curves of three algorithms

Figure 2 illustrates that as the error value increases, the detection effectiveness of the three methods improves. The traditional M1 and M2 fault detection algorithms have low sensitivity to

small step faults and slow-varying faults, and there are multiple detections, and the fault detection effect is not good. The M3 fault detection algorithm in this paper is highly sensitive to small step faults and slow-varying faults, and there are almost no common cases, and the fault detection effect is good.

The results presented in figure 3 demonstrate that the M3 fault-tolerant algorithm proposed in this study yields a smaller root mean square error and higher accuracy in parameter estimation compared to the traditional M1 and M2 robust adaptive algorithms. Specifically, the parameter estimation accuracy is enhanced by approximately 30.90% and 16.90%, respectively.

5 Conclusion

The GNSS signal is susceptible to external interference, which leads to abnormal information in the system and reduces the navigation and positioning performance of the system.The sensitivity and accuracy of fault detection of SINS / GNSS fault tolerance algorithm based on improved smooth bounded layer width proposed in this paper are greatly improved, which effectively improves the accuracy and robustness of SINS / GNSS integrated navigation system.

References:

[1] Wang Qiu-ying, Diao Ming, Gao Wei, et al. Integrated navigation method of a marine strapdown inertial navigation system using a star sensor[J]. Measurement Science and Technology, 2015,26, 115101.

[2] Pan Chen, Gao Jing-xiang, Li Zeng-ke, et al. Multiple fading factors-based strong tracking variational Bayesian adaptive Kalman filter[J]. Measurement, 2021,176, 109139.

[3] Wen Ze-yang, Yang Gong-liu, Cai Qing-zhong. An improved SINS/NHC integrated navigation algorithm based on Ackermann turning geometry [J]. Measurement, 2022, 192, 110859.

[4] Dai Hai-fa, Bian Hong-wei, Wang Rong-ying, et al. An INS/GNSS integrated navigation in GNSS denied environment using recurrent neural network[J]. Defence Technology, 2020, 16:334-340.

[5] Li Bo, Chen Wen, Yu Peng, et al. Robust Kalman filtering based on chi-square increment and its application[J] Remote Sensing, 2020, 12, 732.

[6] Liu Zi-long, Lou Jia-qi Fault detection of integrated navigation system based on orthogonality principle[J]. Control Engineering of China, 2022,02, 97.

[7] Zhang Hua-qing, Li Dong-xing, Zhang Guo-qiang. Application of hybrid chi-square test method in fault detection of integrated navigation system[J]. Journal of Chinese Inertial Technology 2016,24: 696-700.

[8] Gao Guang-le, Gao Du-sheng, Peng Xu, et al. Fading SPRT method for soft fault diagnosis in SINS/CNS/SRS integrated navigation system[J].Journal of Chinese Inertial Technology,2020,28(6): 834-840.

[9] Prabhu G, Bhashyam S, Gopalan A, Sundaresan R. Sequential multi-hypothesis testing in multi-armed bandit problems: an approach for asymptotic optimality[J]. IEEE Transactions on Information Theory, 2022,68:4790-4817.

[10] Habibi S. The smooth variable structure filter[J]. Proceedings of the IEEE, 2007,95:1026-1059.

Identification of Fiber Optic Gyroscope Temperature Drift Models

Zlygosteva M.P., Kolevatov A.P.1, Ulianovskaia T.A., Ivshina I.V., Sergeev A.M.

(Perm Scientific-Industrial Instrument Making Company, Perm)

Abstract: Models of fiber-optic gyroscope temperature drift are investigated. Test results of fiber-optic gyroscope defects arising from changes in ambient temperature are presented. It is shown that these defects deteriorate temperature compensation efficiency. Based on the study results classification of these defects of fiber-optic gyroscope is proposed, it allows to divide inertial navigation systems created on their basis by accuracy classes. Methods have been developed for identifying models of fiber-optic gyroscope temperature drift, including using fuzzy logic method.

1 Introduction

The paper studies behavior of fiber-optic gyroscope (FOG) included in strapdown inertial navigation systems (SINS) produced by Perm Scientific-Industrial Instrument Making Company. Based on long-term observations of the FOG-based SINS temperature calibration results classification and identification of the FOG defects is proposed, which allows to divide the manufactured navigation systems by accuracy classes.

2 Fiber optic gyroscope temperature drift models

It is known that the FOG principle of operation is based on Sagnac effect [1, 2]: "the phase difference of two light waves propagating along a closed contour in opposite directions when the contour rotates around an axis normal to its plane is proportional to the angular velocity of rotation". Despite apparent relative simplicity of the sensor, it is very sensitive to various external influences that cause side drifts distorting its signal and, therefore, leading to increase of the measurement error of absolute angular velocity projection on the object sensitivity axis [4-6, 33-36]. Change in the FOG ambient temperature has a significant influence on the FOG characteristics. Therefore, one of the main issues arising in the FOG-based SINS development is the FOG drift compensation caused by temperature changes inside a strapdown measurement module (SMM).

Direct measurement of temperature gradients and finding correlation dependences of the FOG zero offset is difficult, since multiple temperature sensors placement inside a multilayer fiber coil is a difficult task. It is difficult to develop a numerical model of this effect, because it is necessary to take into account many parameters, both design and technological: frame material and geometry, properties of a particular type of optical fiber, protective and hardening coating material, impregnation compound, coil filling factor, fiber tension force, etc. Therefore, the researchers propose simplified models of the FOG temperature drift, assuming that by the efforts of designers and process specialists the FOG temperature error gradient component will be minimized, ensuring

the proposed models relevancy. Models of the FOG temperature drift are presented in [8-25]. All these works declare successful results of thermal compensation, so the authors of this report find it difficult to recommend any method, but suggest focusing on the FOG defects that worsen quality of thermal compensation.

3 FOG defects classification

To calculate the thermal compensation model coefficients, the FOG readings (as part of the SINS) obtained at thermal transitions when calibrating the system on Acutronic turn tables were used. The SINS calibration is performed at static temperatures and thermal transitions. Thermal transitions of the following types were used:

(1) transitions with temperature change by $20 \div 25$ºC between calibration temperature points, the rate of temperature change averages $2.5 \cdot 10\text{-}3 \div 5.0 \cdot 10\text{-}3$ ºC/sec;

(2) additional transitions by $40 \div 70$ºC, the rate of temperature change averages $7.0 \cdot 10\text{-}3 \div 0.01$ ºC/sec;

(3) self-heating - the system cold start and keeping it switched on until it reaches a stable temperature mode without external heat/cooling sources, the temperature change rate is up to $1.5 \cdot 10\text{-}3$ ºC/sec.

At thermal transitions the SINS is installed in the plane of horizon with a given direction relative to the meridian plane. Thermal drift is evaluated by averaging the FOG initial readings at 5 minutes using moving average method. The chosen value of the FOG readings averaging interval is determined by the results of calculation of D. Allan variations of the FOG investigated batch output readings [30]. The temperature derivative is calculated from the temperature sensor readings of the FOG. The detected defects of optical units according to the results of thermal compensation are classified as follows:

1) change of steady-state temperature zero offset;

2) increased intensity of the FOG drift noise component;

3) S-shape of the FOG thermal drift;

4) irregular shape of the FOG drift;

5) drift advance/delay relative to the rate of temperature change;

6) presence of emissions.

The concept of the FOG controlled temperature drift is introduced, for which the value of the permissible residual drift is limited to the range of ± 0.03°/hour.

4 Defect Identification

As a rule, reviewing graphs with test results of a product takes a lot of time and it is a manual and labor-intensive process. Therefore, in the present work we proposed algorithms that allow automatic identification of defects of a fiber-optic gyroscope during thermal compensation. To identify defects, methods of statistical processing of measurements [31] are used, with the help of which basic scales and membership functions are formed for application of the fuzzy set theory [32]. About 60 reports on thermal compensation of FOGs were analyzed. The results of automatic identification were compared with manual defect analysis. According to the study sample about 72%

of the reports completely conforms with the visual inspection results, this demonstrates effectiveness of the proposed identification methods. Note that all the FOG described defects are not amenable or little amenable to algorithmic thermal compensation, however, they may indicate certain structural or technological defects in the FOG sensitive elements, which in the future will be able to help FOG developers to understand the defects cause and get the opportunity to eliminate them.

References:

[1] Lefevre, H.C., The Fiber Optic Gyroscope, Boston, MA, USA, Artech House, 2014, p. 343.

[2] W.K. Burns, P.F. Liao, P. Kelley. Optical fiber rotation sensing. Academic press, 1994, p. 390

[3] Shupe D.M. Thermally induced non-reciprocity in the fiber-optic interferometer // Applied optics 1980, vol.19, No. 5, pp. 654-655. DOI: 10.1364/AO.19.000654

[4] Mohr F., Schadt F. Bias error in fiber optic gyroscopes due to elasto-optic interactions in the sensor fiber. Proc. SPIE, 2004, vol. 5502, pp. 410 – 413. DOI: 10.1117/12.566654 Available at: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=848587

[5] Schadt F., Mohr F. Error signal formation in FOGs through thermal and elastooptical environment influence on the sensing coil // Proc. Inertial Sensors and Systems. 2011. P. 2.1 – 2.13.

[6] Hocker G.B. Fiber-optic sensing of pressure and temperature. Appl. Opt., 1979, Vol. 18, No. 9, pp. 1445 – 1448. DOI: 10.1364/AO.18.001445

[7] Ling W., Li X., Xu Z., Zhang Z., Wei Y. Thermal effects of fiber sensing coils in different winding pattern considering both thermal gradient and thermal stress. Optics Communications, 2015, vol. 356, pp. 290 – 295. DOI: 10.1016/j.optcom.2015.08.002.

[8] IEEE Standard Specifying and Testing Single-Axis Interferometric Fiber Optic Gyros, IEEE Std 952- 2020 (Revision of IEEE Std 952-1997)

[9] V. E. Dzhashitov, V. M. Pankratov. Mathematical Models of Thermal Drift of Inertial Systems Gyroscopic Sensors. St. Petersburg: SSC RF Central Research Institute "Electropribor", 2001. 150 pages.

[10] V. E. Dzhashitov, V. M. Pankratov, A. V. Golikov, S. G. Nikolaev, A. P. Kolevatov, A. D. Plodnikov, K. V. Koffer. Hierarchical Thermal Models of FOG-based Strapdown Inertial Navigation System. Gyroscopy and Navigation, July 2014, Volume 5, Issue 3, pp. 162-173.

[11] V. E. Dzhashitov, V. M. Pankratov, A. V. Golikov, S. G. Nikolaev, A. P. Kolevatov, A. D. Plodnikov, K. V. Koffer. Temperature Calibration of a Strapdown Navigation System Based on Signals from Distributed Thermal Sensors. Mechatronics, Automation, Control No 7, 2013, pp. 42-47.

[12] V. E. Dzhashitov, V. M. Pankratov. Mathematical Models of Thermal Drift of Inertial Systems Gyroscopic Sensors. St. Petersburg: SSC RF Central Research Institute "Electropribor", 2001. 150 pages.

[13] A. P. Kolevatov, S. G. Nikolaev, А. G. Andreev, V. S. Ermakov, О. L. Kel, D. I. Shevtsov. Fiber-Optic Gyroscope for Strapdown Navigation Systems. Development, Thermal Compensation, Testing. Gyroscopy and Navigation No 3(70) 2010, pp. 49-60.

[14] Investigation of the Output Signal Dependence of a Fiber-Optic Gyroscope on Temperature as part of a Gimballess Inertial Measurement Unit // E. V. Dranitsyna, D. A. Egorov. (https://textarchive.ru/c-1987880.html)

[15] G. I. Yemelyantsev, A. P. Stepanov. Integrated Navigation System. St. Petersburg: SSC RF Central Research Institute "Electropribor", 2016. – 394.

[16] Dranitsyna, E., Gontar, D.A., Compensation for the Temperature Dependence of the Fog Output Signal, International Workshop Navigation and Motion Control (NMC 2021), IOP Conference Series: Materials Science and Engineering, 2022, vol. 1215 (2022), no. 012003, pp. 1 – 5.

[17] Esipenko I. A., Lykov D. A., Mathematical Model of Fiber-Optic Gyroscope Thermal Drift and its Experimental Verification // Bulletin of the Bauman Moscow State Technical University. Ser. Instrumentation. 2017. No 5, pp. 31 – 46. DOI: 10.18698/0236-3933-2017-5-31-46.

[18] Vakhrameev E. I., Galyagin K. S., Ivonin A. S., Oshivalov M. A., Ulrich T. A. Fiber-Optic Gyroscope Thermal Drift // News of Higher Educational Institutions. Instrumentation. 2011. Volume 54, No 1. pp. 32—37.

[19] Vakhrameev E. I., Galyagin K. S., Ivonin A. S., Oshivalov M. A. Forecast and Correction of Fiber-Optic Gyroscope Thermal Drift // News of Higher Educational Institutions. Instrumentation. 2013. Volume 56, No 5, pp. 79—84

[20] Vakhrameev E. I., Galyagin K. S., Oshivalov M. A., Savin M. A. Methods of Numerical Forecasting and Correction of Fiber-Optic Gyroscope Thermal Drift // News of Higher Educational Institutions. Instrumentation. 2017. Volume 60, No 1, pp. 32—38

[21] Frolov A. V., Mikhailov Yu. V., Smirnov S. V. Development and Validation of a Methodology for Modeling Thermal and Deformed State of Strapdown Inertial Navigation System Parts. Bulletin of the Bauman Moscow State Technical University. Ser. Instrumentation, 2022, No 1 (138), pp. 32 – 48. DOI: https://doi.org/10.18698/0236-3933-2022-1-32-48

[22] Research and Improvement of the Fiber-Optic Gyroscopes Accuracy Parameters. / A.S. Aleynik. Dissertation for the Degree of Candidate of Technical Sciences. St. Petersburg 2012. St. Petersburg National Research University of Information Technologies, Mechanics and Optics at the Department of Physics and Optical Communication Technology.

[23] Pylaev Yu. K. Temperature Disturbances of a Strapdown Inertial Navigation System with Fiber-Optic Gyroscopes. Degree of Candidate of Technical Sciences. Institute of Problems of Precision Mechanics and Control of the Russian Academy of Sciences, NPP Antares LLC, Saratov 2004. 263 pages.

[24] Tarygin E. I. Calibration Method of Thermal Model of Sensitive Elements Unit Consisting of Three Angular Velocity Sensors. Gyroscopy and Navigation. Volume 27. No 4 (107), 2019, pp. 88 – 102.

[25] Kozlov A. V., Tarygin I. E., Golovan A. A., Shaimardanov I. H., Dzuev A. A. Calibration of Inertial Measurement Units with Assessment of Temperature Dependencies Based on Variable Temperature Experiment: Calibration Results of the БИНС-РТ // XXIV St. Petersburg International Conference on Integrated Navigation Systems. Digest. St. Petersburg: JSC "Concern "Central Research Institute "Electropribor", 2017, pp. 225 – 228.

[26] Nikiforovsky D. A., Deineka I. G., Sharkov I. A., Meshkovsky I. K. Method for Compensating Fiber-Optic Gyroscope Temperature Drift Using Correlations between the Gyroscope and Several Temperature Sensors Readings. Gyroscopy and Navigation. Volume 30. No 2 (117), 2022. pp. 71-80.

[27] Meshkovsky I. K., Rupasov A. V., Strigalev V. E., Sharkov I. A. (ITMO Research Institute, St. Petersburg, Russia) Investigation of the Effect of Thermal Influences on Fiber-Optic Gyroscope Operation. XXI St. Petersburg International Conference on Integrated Navigation Systems. 2014.

[28] Klimkovich B. V. Influence of Temperature Sensors Random Error on the Quality of Temperature Compensation of the Zero Offset by a Neural Network // Gyroscopy and Navigation. 2020. Т. 28. №4 (111), pp. 53 – 70. DOI 10.17285/0869-7035.0049.

[29] Antonova M. V., Matveev V. A. A Model of Fiber-Optic Gyroscope Error under the Influence of Thermal and Magnetic Fields // Bulletin of the Bauman Moscow State Technical University. Ser. Instrumentation. 2014. No 3, pp. 73 – 80.

[30] On the Issue of Evaluating Noise Components of the Measuring Path using Allan Variation Method / D. A. Kutova, O. I. Maslova, S. Y. Perepelkina, A. A. Fedotov // Gyroscopy and Navigation. – 2015. – № 2 (89). – pp. 30 – 39.

[31] Pugachev V. S. Probability Theory and Mathematical Statistics. М.: ФИЗМАТГИЗ. 2002. – 496 pages.

[32] Averkin A. N., Batyrshin I. Z., Blishun A. F., Silov V. B., Tarasov V. B., Pospelova D. A. Fuzzy Sets in Control and Artificial Intelligence Models. М.: Книга по требованию, 2024. – 312 pages.

[33] Rui S., Xiyuan C., Chong S., Hong Z. Modeling FOG drift using back-propagation neural network optimized by artificial fish swarm algorithm // Journal of Sensors, 2014, Volume 2014, Article ID 273043, 6 pages, DOI : http://dx.doi.org/10.1155/2014/273043

[34] Haotian Y., Bin Z., Lixin W., Rong Z. A Novel tri-axial MEMS gyroscope calibration method over a full

temperature range // Sensors, 2018, 18, 3004; DOI: 10.3390/s18093004, www.mdpi.com/journal/sensors

[35] Xuyou L., Chen Z., Zhou H., Zhen Z. Temperature errors of IFOG and its compensation in engineering application // The Ninth International Conference on Electronic Measurement & Instruments, 2009, ICEMI' 2009

[36] Chun-Lei Z., Qi Z., Shu-Hua Y., Lei G., Guo-Chao W. Modeling and compensation for temperature errors of interferometric fiber optic gyroscope // International Conference on Information and Automation, 2009

Attitude Determination Using Least Information From Incomplete Two-Vector Measurements

Tan Caiming[1,2]

(1. College of Automation & College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210023, China; 2. Robotics Information Sensing and Control Research Institute, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract: Attitude determination is essential for spacecraft control and robot navigation. Attitude determination methods from vector measurements, including the most famous methods, TRIaxial Attitude Determination (TRIAD) and QUaternion ESTimator (QUEST), normally require at least two vector measurements. This paper proposes the SIMplest Attitude Determination (SIMAD) method, a complete solution for attitude determination using the minimum vector measurements. Three available components from the two vector measurements are required to obtain four potential solution candidates, and part of the remaining three components are required to determine the unique correct solution. Theoretical and experimental results demonstrate that SIMAD can outperform TRIAD and QUEST when some of the components of either vector have lower precision. Thus, SIMAD can be used for attitude determination when part of the two-vector measurements is unavailable or contains more measurement noise due to sensor error and disturbances. Moreover, SIMAD can be used for fault detection and calibration when two-vector measurements are available.

Key words: Attitude determination; Minimum vector measurements; Measurement uncertainty; TRIAD; QUEST.

1　Introduction

The TRIaxial Attitude Determination (TRIAD) algorithm is a well-known deterministic attitude determination solution using two vector measurements [1]. Some accuracy is lost because part of the measurement from the two vector observations is discarded [1,2]. The optimal attitude estimation problem using any number of vector measurements was defined by Wahba [3] and is referred to as the Wahba problem. The Wahba problem uses weights to describe the significance of different vector measurements. Many effective optimal algorithms have been proposed to solve the Wahba problem [1,4-7], including the famous QUaternion ESTimator (QUEST)[1]. An easily overlooked premise for Wahba problem-based methods is that the three components in each vector should have the same precision. Neither the TRIAD nor QUEST algorithms apply to cases with incomplete two-vector measurements.

Sinister [8] first deduced that four possible solution candidates are obtained if three components of the two vector measurements are available. However, the additional required measurements to

---

determine the unique solution were not examined, and it is not a complete solution. A deterministic attitude estimation method for a sun-pointing satellite was proposed for when one component of either vector is not measured, such as in the Earth's magnetic field measurement vector. However, its application is limited to attitude estimation only when the *x*-axis points in a known direction [9]. The borehole attitude determination problem using a gyroscope inclinometer without a sensor measuring the axial angular rate component has been solved [10]. The attitude is obtained based on the relationship between the Euler angles, three-axis accelerometers, and two-axis gyroscopes. Partly due to the focus on solving practical applications, the information used for attitude determination from incomplete two-vector measurements in those works [9-13] is not the minimum.

The author's previous work has verified that only three components of the two vector measurements are required to calculate four potential attitude solution candidates in the land vehicle experiment [14-17], where accelerometers and gyroscopes were used for attitude determination. The geometrical method proposed in [14] requires two of the three available components to be coaxial, while later proposed methods do not have this requirement [15-17]. A few examples show that some properties from the remaining outputs, such as the sign and the approximate value, can determine the unique solution [14-17]. However, the exact minimum requirement for determining the unique solution among the potential solution candidates is unknown, and all three of the remaining components were used to determine the unique solution.

This paper proposes the Simplest Attitude Determination (SIMAD) using least information from incomplete two-vector measurements, including the minimum requirement for determining the unique solution, measurement uncertainty, and experimental verification.

2   Problem Statement

The reference frame is denoted r. The two non-parallel vectors referenced in the r-frame are denoted r_1 and r_2, which are known and error-free. The two vector measurements denoted $\boldsymbol{a}$ and $\boldsymbol{b}$ are obtained after the three-axis sensors measure vectors r_1 and r_2, respectively. The sensor frame is denoted s. The rotation matrix from the r-frame to the s-frame is denoted $\boldsymbol{C}_r^s$, and the vectors satisfy

$$\begin{cases} \boldsymbol{a} = \boldsymbol{C}_r^s \boldsymbol{r}_1 \\ \boldsymbol{b} = \boldsymbol{C}_r^s \boldsymbol{r}_2 \end{cases} \tag{1}$$

All the vectors are assumed to have been normalized for ease of analysis. The attitude determination from the two vector measurements is achieved by obtaining $\boldsymbol{C}_r^s$ from Eq. (1). TRIAD and QUEST are the two well-known methods for addressing this problem. However, TRIAD and QUEST do not work if $\boldsymbol{a}$ and $\boldsymbol{b}$ are not completely known. Moreover, if the accuracies of each component in $\boldsymbol{a}$ or $\boldsymbol{b}$ are not equal, then QUEST, known as the most accurate method, will no longer be the most accurate.

Because Cartesian coordinate transformation does not change the norm of a vector or the dot product between vectors, Eq. (2) can be obtained from Eq. (1):

$$\begin{cases} a_1^2 + a_2^2 + a_3^2 = 1 \\ b_1^2 + b_2^2 + b_3^2 = 1 \\ a_1 b_1 + a_2 b_2 + a_3 b_3 = \cos\theta \end{cases} \tag{2}$$

where $\boldsymbol{a} = [a_1, a_2, a_3]^{\mathrm{T}}$, $\boldsymbol{b} = [b_1, b_2, b_3]^{\mathrm{T}}$, and $\cos\theta = \boldsymbol{r}_1^{\mathrm{T}} \boldsymbol{r}_2$, and where θ is the angle

between $\boldsymbol{a}$ and $\boldsymbol{b}$, and $0 < \theta < \pi$. The unavailable components can be regarded as unknown quantities in Eq. (2), which consists of three equations that may be used to compute three unknown quantities. It is assumed that a_i is unavailable, the other two components in vector $\boldsymbol{a}$ are available, b_j is available, and the other two components in vector $\boldsymbol{b}$ are unavailable, where i and j are 1, 2, or 3. More than one potential solution exists for the unavailable components, so more information is required to determine the unique solution. SIMAD is achieved if the minimum requirements for obtaining $\boldsymbol{a}$ and $\boldsymbol{b}$ are obtained.

3 Method

It is defined specifically for $i1$, $j1$, $i2$ and $j2$ that

$$i1, j1 = \begin{cases} 2 & i,j = 1 \\ 3 & i,j = 2, \\ 1 & i,j = 3 \end{cases} \quad i2, j2 = \begin{cases} 3 & i,j = 1 \\ 1 & i,j = 2 \\ 2 & i,j = 3 \end{cases}$$

As shown in Eq. (2), a_i, b_{j1}, and b_{j2} are unavailable components to be computed using the three available components a_{i1}, a_{i2}, and b_j. The four solutions for a_i, b_{j1}, and b_{j2}, labeled (A), (B), (C), and (D), are obtained as

$$
\begin{array}{ll}
a_i = \sqrt{1 - a_{i1}^2 - a_{i2}^2} & a_i = \sqrt{1 - a_{i1}^2 - a_{i2}^2} \\
\text{(A)} \quad b_{j1} = \frac{a_{j1}(\cos\theta - a_j b_j) + a_{j2}\sqrt{\Delta}}{a_{j1}^2 + a_{j2}^2}, \text{(B)} \quad b_{j1} = \frac{a_{j1}(\cos\theta - a_j b_j) - a_{j2}\sqrt{\Delta}}{a_{j1}^2 + a_{j2}^2} \\
\left\lfloor b_{j2} = \frac{a_{j2}(\cos\theta - a_j b_j) - a_{j1}\sqrt{\Delta}}{a_{j1}^2 + a_{j2}^2} \right. & \left\lfloor b_{j2} = \frac{a_{j2}(\cos\theta - a_j b_j) + a_{j1}\sqrt{\Delta}}{a_{j1}^2 + a_{j2}^2} \right. \\
a_i = -\sqrt{1 - a_{i1}^2 - a_{i2}^2} & a_i = -\sqrt{1 - a_{i1}^2 - a_{i2}^2} \\
\text{(C)} \quad b_{j1} = \frac{a_{j1}(\cos\theta - a_j b_j) + a_{j2}\sqrt{\Delta}}{a_{j1}^2 + a_{j2}^2}, \text{(D)} \quad b_{j1} = \frac{a_{j1}(\cos\theta - a_j b_j) - a_{j2}\sqrt{\Delta}}{a_{j1}^2 + a_{j2}^2} \\
\left\lfloor b_{j2} = \frac{a_{j2}(\cos\theta - a_j b_j) - a_{j1}\sqrt{\Delta}}{a_{j1}^2 + a_{j2}^2} \right. & \left\lfloor b_{j2} = \frac{a_{j2}(\cos\theta - a_j b_j) + a_{j1}\sqrt{\Delta}}{a_{j1}^2 + a_{j2}^2} \right.
\end{array}
\tag{3}
$$

where $\Delta = \sin^2\theta - a_j^2 - b_j^2 + 2a_j b_j \cos\theta$. More information is required to determine the unique correct solution.

The quantitative relations between the four solutions shown in Eq. (3) are analyzed to find the minimum additional requirements for determining the unique solution. It is apparent that

$$a_i(1) = a_i(2) = -a_i(3) = -a_i(4) \geq 0$$

where $\chi(k)$ is the kth solution of χ shown in Eq. (3), and χ represents a_i, b_{j1}, and b_{j2}, etc. Based on the equations for b_{j1} and b_{j2} shown in Eq. (3), the four solutions for b_{j1} and b_{j2} satisfy the relations shown in Tab. 1.

Tab. 1 Quantitative relations of the four solutions for b_{j1} and b_{j2}.

	b_{j1}	b_{j2}
$i = j1$	$b_{j1}(1) = -b_{j1}(4), b_{j1}(2) = -b_{j1}(3)$	$b_{j2}(1) = b_{j2}(4), b_{j2}(2) = b_{j2}(2)$
$i = j2$	$b_{j1}(1) = b_{j1}(4), b_{j1}(2) = b_{j1}(3)$	$b_{j2}(1) = -b_{j2}(4), b_{j2}(2) = -b_{j2}(3)$

Based on the above conclusions, the criteria for determining the unique solution with the minimum additional requirements are shown in Tab. 2. When $i = j$, all the solutions for either b_{j1} or b_{j2} are different. Therefore, the solutions for either b_{j1} or b_{j2} are compared to their measured

values, and the closest should be the correct solution. Criteria 1 and 2 use the approximate values of b_{j1} and b_{j2}, respectively. When $i \neq j$, two of the solutions for one of b_{j1} and b_{j2} are equal. Therefore, the sign of the measured value of a_i is also required. Because the measured values of b_{j1} or b_{j2} are used solely for determination, the measured value can be approximate.

Tab. 2 Criteria to determine the unique solution with minimum additional measurement requirements.

		a_i	b_{j1}	b_{j2}
$i = j$	Criterion 1	/	approximate value	/
	Criterion 2	/	/	approximate value
$i = j1$	Criterion 1	/	approximate value	/
	Criterion 2	sign	/	approximate value
$i = j2$	Criterion 1	sign	approximate value	/
	Criterion 2	/	/	approximate value

After the components of the two vectors have all been obtained using SIMAD, the attitude is computed using the TRIAD method.

4 Measurement uncertainty

The misalignment angles using SIMAD are denoted $\boldsymbol{\phi}$, where $\boldsymbol{\phi} = [\phi_x \quad \phi_y \quad \phi_z]^{\mathrm{T}}$. The covariance matrices of $\boldsymbol{\phi}$ have been given in [8]. The norm $\|\boldsymbol{\phi}\|_2$ can be used as a scalar value indicating the attitude error with the feasibility explained in [18]. The variance of $\|\boldsymbol{\phi}\|_2$, denoted $\mathrm{var}(\|\boldsymbol{\phi}\|_2)$, satisfies

$$
\mathrm{var}(\|\boldsymbol{\phi}\|_2) = \mathrm{trace}\big(\mathrm{cov}(\boldsymbol{\phi})\big)
$$

$$
= \begin{cases}
\dfrac{a_j^2 b_{i1}^2 + a_{i1}^2 r}{\tau^2}\mathrm{var}(a_{i1}) + \dfrac{a_j^2 b_{i2}^2 + a_{i2}^2 r}{\tau^2}\mathrm{var}(a_{i2}) + \dfrac{\mathrm{var}(b_j)}{s^2}, i = j \\[3mm]
\dfrac{\mathrm{var}(a_{i1})}{a_i^2} + \dfrac{a_j^2 b_i^2 + a_i^2 r}{\tau^2}\mathrm{var}(a_{i2}) + \dfrac{\mathrm{var}(b_j)}{s^2}, i = j1 \\[3mm]
\dfrac{a_j^2 b_i^2 + a_i^2 r}{\tau^2}\mathrm{var}(a_{i1}) + \dfrac{\mathrm{var}(a_{i2})}{a_i^2} + \dfrac{\mathrm{var}(b_j)}{s^2}, i = j2
\end{cases}
\tag{4}
$$

where $\mathrm{trace}(\cdot)$ returns the sum of the diagonal components, $r = b_{j1}^2 + b_{j2}^2$, $s = a_{j2}b_{j1} - a_{j1}b_{j2}$, and $\tau = a_i s$. The variance of $\|\boldsymbol{\phi}\|_2$ increases with decreasing $|\tau|$.

5 Numerical experiment and discussion

A simulation experiment based on the Monte Carlo method is conducted to test SIMAD compared to the two classical attitude determination methods, TRIAD and QUEST. Reference vectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are set to $\boldsymbol{r}_1 = [1 \quad 0 \quad 0]^{\mathrm{T}}$, $\boldsymbol{r}_2 = [\cos \pi/4 \quad \sin \pi/4 \quad 0]^{\mathrm{T}}$, where the angle θ between $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ is $\pi/4$. Euler angles representing the attitude between the reference and sensor frames are generated randomly from the continuous uniform distributions on their domain intervals. The two vector measurements $\boldsymbol{a}$ and $\boldsymbol{b}$ are generated given $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, and the Euler angles. The error of each component of the two vector measurements is Gaussian randomly distributed. The standard deviations of the errors of a_{j1}, a_{j2}, and b_j are 0.001, and the standard deviations of the errors of a_i, b_{j1}, and b_{j2} are 0.01. i and j are given by a random function. One thousand samples of $\boldsymbol{a}$ and $\boldsymbol{b}$ are generated randomly from the normal distribution for each case.

Vector $\boldsymbol{a}$ is used as the anchor vector for TRIAD [2]. The weights for vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are equal for QUEST. Statistical results were obtained given the 1000 samples. The above procedure is for one test, and a total of 50 tests were conducted.

The 50 tests are reordered from large to small according to the order of $|\tau|$ in Eq. (4). The attitude accuracy is shown in Fig. 1 using this new order. The angular distance [18] between the computed attitude and ground truth is used as the scalar indicator of the attitude error. The experimental accuracy is expressed by the root-mean-square error (RMSE). The TRIAD and QUEST accuracies are markedly lower than for SIMAD when $|\tau|$ is not close to zero. SIMAD uses the three components with the highest accuracy among the two vectors to compute the four potential solution candidates. The remaining three components with lower accuracies are used to determine the unique solution.
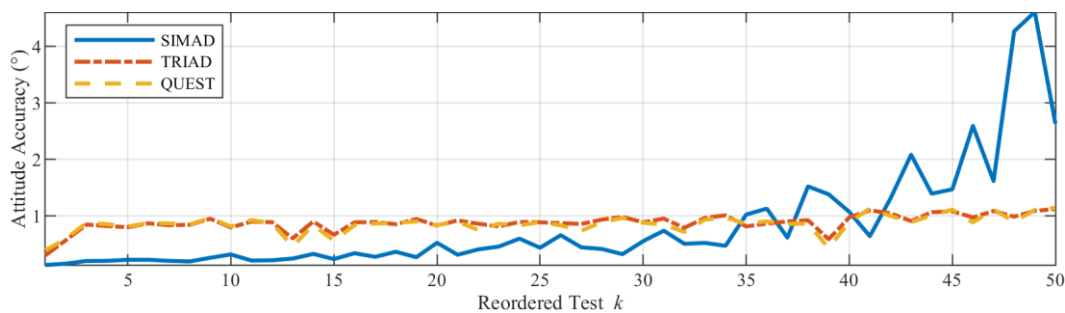
Fig. 1 Attitude accuracy using SIMAD, TRIAD, and QUEST, shown after sorting the 50 tests from largest to smallest $|\tau|$ [see Eq. (4)].

6   Conclusions

SIMAD, a complete attitude determination solution using the minimum vector measurements, is proposed. First, three available components from two vector measurements are required to compute four potential solution candidates. Then, information from the remaining three components is required to determine the unique solution. The minimum required information for determination can be the approximate value of only one component or the approximate value of one component and the sign of another component. Theoretical and experimental results demonstrated that SIMAD could outperform TRIAD and QUEST when some of the components of either vector have lower accuracy.

SIMAD can be used for attitude determination when part of the two-vector measurements is unavailable or contains measurement noise due to sensor error and disturbances. Moreover, when two vector measurements are available, SIMAD can be used for fault detection and calibration.

References:

[1]   Shuster Malcolm David, Oh S D. Three-axis attitude determination from vector observations [J]. Journal of guidance and Control, 1981, 4(1): 70-77.

[2]   Bar-Itzhack Itzhack Y, Harman Richard R. Optimized TRIAD algorithm for attitude determination [J]. Journal of Guidance, Control, and Dynamics, 1997, 20(1): 208-211.

[3]   Wahba Grace. A least squares estimate of satellite attitude [J]. SIAM review, 1965, 7(3): 409-409.

[4]   Keat J. Analysis of least-squares attitude determination routine DOAOP [R]. Technical Report CSC/TM-

77/6034, Comp. Sc. Corp, 1977.

[5]   Markley F Landis. Attitude determination using vector observations and the singular value decomposition [J]. Journal of the Astronautical Sciences, 1988, 36(3): 245-258.

[6]   Mortari Daniele. ESOQ: A closed-form solution to the Wahba problem [J]. Journal of the Astronautical Sciences, 1997, 45: 195-204.

[7]   Ghadiri Hossein, Esmaelzadeh Reza, Zardashti Reza. Robust optimal attitude determination using interval analysis [J]. Advances in Space Research, 2022, 69(6): 2611-2617.

[8]   Sinister Malcolm D. Deterministic three-axis attitude determination [J]. Journal of the Astronautical Sciences, 2004, 52: 405-419.

[9]   Ahn Hyo-Sung, Lee Seon-Ho. Gyroless attitude estimation of sun-pointing satellites using magnetometers [J]. IEEE Geoscience and Remote Sensing Letters, 2005, 2(1): 8-12.

[10]   Binder Yakov I, Paderina Tatyana V, Litmanovich Yury A. Advanced borehole attitude determination without measuring axial angular rate component[C]//Proceedings of IEEE/ION PLANS 2006. 2006: 963-968.

[11]   Litmanovich Yu A. On one approach to the use of redundant information in attitude determination from two vector observations [J]. Gyroscopy and Navigation, 2012, 3(4): 280-285.

[12]   Lee Seon-Ho, Ahn Hyo-Sung, Yong Ki-Lyuk. Three-axis attitude determination using incomplete vector observations [J]. Acta Astronautica, 2009, 65(7-8): 1089-1093.

[13]   Ahn Hyo-Sung, Lee Seon-Ho. Geometrical Attitude Determination Algorithm Based on Vector Measurements [J]. Transactions of the Japan Society for Aeronautical and Space Sciences, 2010, 53(179): 8-18.

[14]   Tan Caiming, Zhu Xinhua, Su Yan, et al. A new analytic alignment method for a SINS [J]. Sensors, 2015, 15(11): 27930-27953.

[15]   Tan Caiming, Wang Yu, Zhu Xinhua, et al. Improved alignment method for a SINS using two vector measurements[C]//2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC). IEEE, 2015: 1674-1678.

[16]   Tan Caiming. Research on the initial alignment of SINS employed in land moving vehicle [D]. Ph. D. dissertation, Nanjing University of Science & Technology, Nanjing, China, 2016.

[17]   Tan Caiming, Gao Xiang, Xu Guozheng, et al. Three-axis Attitude Determination from Incomplete Two Vector Observations [J]. Chinese Journal of Scientific Instrument, 2018, 39(11): 140-146.

[18]   Hartley Richard, Trumpf Jochen, Dai Yuchao, et al. Rotation averaging [J]. International Journal of Computer Vision, 2013, 103: 267-305.

Purcell-enhanced sensitivity of NV-based gyroscope using plasmonic waveguide

Ji Lumin[1], Zhao Liye[1,*], Ding Xukai[1] and Wang Yuhai[1]

(1. Key Laboratory of Micro-Inertial Instrument and Advanced Navigation Technology, Department of Instrument Science and Engineering, Southeast University, Nanjing 210096, China)

Abstract: Nitrogen-vacancy (NV) color center atomic spin systems serve as valuable tools for inertial measurement and other sensing applications that necessitate miniaturization. NV-based quantum sensing encounters the challenge of reducing the spin readout noise to approach the standard quantum limit (SQL). In this work, we propose a fluorescence enhancement structure based on a plasmon waveguide. The Purcell effect induced by the metallic structure, coupled with the light-guiding facilitated by the dielectric structure, enhances fluorescence through dual mechanisms: radiation and collection. The finite element simulation results show that the fluorescence can be enhanced by 70 times in theory, and the limit sensitivity can be optimized by 8.37 times. This fluorescence-enhanced scheme, which holds potential for miniaturization and integration, introduces a novel concept for quantum sensor devices, including on-chip NV-based gyroscope scopes.

 Key words: NV center; spin readout; Purcell effect;

1 Introduction

As a controllable solid-state quantum platform operable at room temperature, the nitrogen-vacancy (NV) color center spin system has garnered significant attention for its potential to detect various physical quantities, including magnetic fields, angular velocity, force, and temperature fields[0,0]. Research aimed at optimizing its sensing performance predominantly concentrates on minimizing magnetic and thermal noise to enhance the characteristic time of the NV color center [0-0]. In the domain of inertial measurement, a significant challenge in angular velocity sensing using NV centers is achieving reliable spin state readout that approaches the standard quantum limit (SQL) of quantum measurement. However, regarding the reduction of spin readout noise, the predominant focus of researchers has been on enhancing fluorescence collection efficiency after fluorescence generation[0]. In contrast, there is a scarcity of studies addressing the improvement of the spontaneous emission rate during the fluorescence generation process.

Aiming at achieving the accurate extraction of the angular velocity signal of the NV color center gyroscope, this paper presents a spin state measurement scheme. We focus on optimizing fluorescence excitation and collection to improve both the total internal reflection effect in the diamond and the spontaneous emission lifetime. In this study, we propose a compatible fluorescence

radiation scheme that leverages plasmonic enhancement and introduces a dielectric waveguide to enhance fluorescence collection efficiency. The impact on sensing performance is evaluated through finite element simulations and numerical calculations. The results show the readout signal-to-noise ratio (SNR) and limit sensitivity can be improved 8.37 times through this plasmon waveguide enhanced spin readout technique. In the future, this scheme can be combined with bottom-up processing techniques, which offers new perspectives for realizing on-chip and scalable NV centers based gyroscope.

2 Principle and Scheme
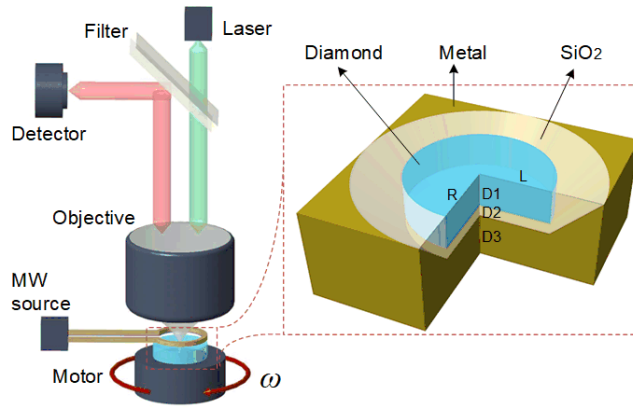
Fig.1 Schematic of the rotation sensing system based on NV color center. The inset shows a profile of the plasmon waveguide structure.

Since the concept of the NV center gyroscope with miniaturization potential was proposed in 2012[0,0], researchers have demonstrated that the rotation velocity ω of the diamond can be reflected in the fluorescence emitted by the NV center (as illustrated in Fig.1) through the phase accumulated in a single measurement time t. The population difference between 0 spin state and 1 spin state of the two-level systems can be expressed as $\langle S_z \rangle \sim cos(\omega t)$. The measurement signal of this quantum system consists of two variables with different weights, and they have Poisson noise and binomial distribution noise respectively. We represent the spin readout noise in terms of a dimensionless quantity with the photon averaging readout technique, which can be expressed as follows.

$$\sigma_r = \sqrt{1 + \frac{2}{SNR^2}}$$

(1)

Where $SNR = \frac{a-b}{\sqrt{a+b}}$ is the spin readout signal-to-noise ratio of one measurement, which reflects the differentiation degree to the spin state of the quantum system[0]. The parameters a and b represent the number of measured photons in spin state 0 and 1 of NV center, respectively. Considering the low contrast ($V = \frac{a-b}{a+b} < 0.15$) of NV based measurements[0], the spin readout noise

can be simplified to $\sigma_r \approx \frac{\sqrt{2}}{SNR}$. We optimize the SNR of NV center gyroscope by optimizing the fluorescence generation and collection process of the system, leading to an improvement in its sensing performance.

In the classical model, the fluorescence emission process by the transition of NV center two-level system can be equivalent to an oscillating dipole. The designed plasmon waveguide structure is shown in the inset of Fig.1. When the NV center is coupled with the electromagnetic environment under the action of this structure, its fluorescence radiation rate will be modified as:

$$P_e = \frac{\gamma}{\gamma_0} = \frac{P}{P_0} \tag{2}$$

Here, P_e is the Purcell factor, γ and P are the fluorescence radiation rate and far-field radiation power after fluorescence enhancement, respectively. γ_0 and P_0 with zero notations are the corresponding quantities before fluorescence enhancement. It should be noted that the fluorescence quenching effect of NV center and metal structure is significant in a close range, so the dielectric structure is laid between diamond and metal to improve the radiation quantum efficiency [0]. To mitigate the total internal reflection (TIR) effect caused by the high refractive index of diamond, the SiO_2 waveguide is designed to reflect the light. The ratio of collection efficiency before and after the improvement is denoted by the enhancement coefficient C_e.

3 Simulation and Analysis

The fluorescence rate enhancement effect was simulated using the finite element simulation software COMSOL. Considering that the fluorescence peak appears around 670nm, we optimize the diamond thickness D1, the dielectric thickness D2, and the metal thickness D3 to ensure an appropriate resonance frequency. The parameters are taken as D1=140nm, D2=10nm and D3=500nm, respectively. Without loss of generality [0,0], the diamond radius, the right-angled side length of the prism whose cross-section is a right-angled isoscentic triangle, and the height of the NV center from the diamond surface are taken as R=50nm, L=D1+D2, and H=10nm, respectively. The numerical aperture of the objective is taken as NA=0.9.
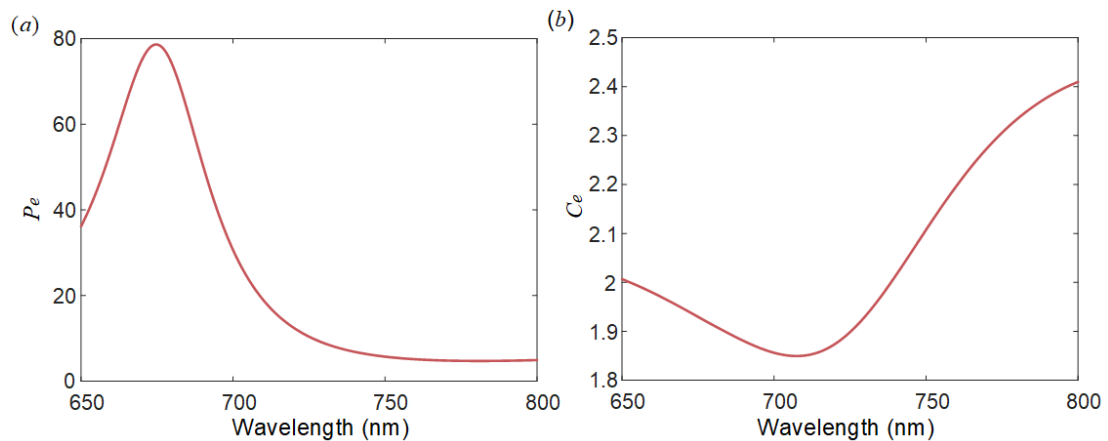
Fig.2   Schematic representation of the Purcell factor and collection efficiency enhancement coefficient

as a function of wavelength under the effect of the enhancement structure.

As shown in Fig.2, the P_e and C_e change as a function of the fluorescence wavelength λ are obtained. As can be seen from Fig.2 (a), the peak value of P_e appears at 670nm. The resonance wavelength almost coincides with the fluorescence peak to maximize the number of radiated photons. From Fig.2 (b), it can be obtained that the collection efficiency is greater than 1, indicating that the overall collection efficiency is improved in this fluorescence band.
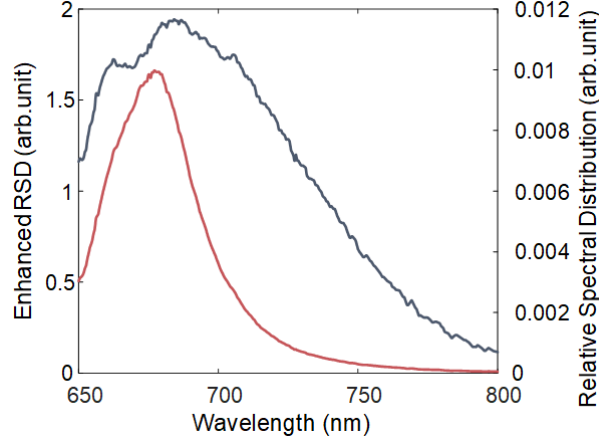
Fig.3   Comparison of relative spectral distribution (RSD) before (right axis) and after (left axis) fluorescence enhancement.

The NV center spin state fluorescence band is concentrated in 650nm-800nm. Finally, we normalized the NV fluorescence spectra in this range and combined the simulation results to compare the collected fluorescence rate before and after the enhancement as shown in Fig.3. The integration calculation result shows that the total collected fluorescence can be increased $F_e = P_e C_e = 70.1$ times. The sensitivity of an NV center gyroscope is defined as the magnitude of the smallest rotation velocity that can be detected per unit time. Its expression is as follows:

$$\eta = \frac{\pi \hbar \sqrt{T_2 + T_o}}{2g\mu_B T_2 \sqrt{N}} \cdot \sigma_r = \frac{\pi \hbar \sqrt{T_2 + T_o}}{\sqrt{2N} g\mu_B T_2} \cdot \frac{1}{SNR} \tag{3}$$

Where η is the sensitivity. $\hbar$, g and μ_B represent Planck's constant, gyromagnetic ratio, and Bohr magneton. N is the number of NV color centers. T and T_o are the coherence time and overhead time, respectively. The relationship between the enhanced SNR and the initial SNR_0 can be expressed as $SNR = \sqrt{F_e} \cdot SNR_0$ under the assumption that the spin mixing is nonradiative and the excitation optical power is sufficiently high. It can be seen that the limit sensitivity can be optimized 8.37 times.

4   Conclusion

In this work, we designed a structure for fluorescence enhancement based on the Purcell effect. The enhancement of the fluorescence excitation process is achieved by coupling the plasmon under the action of metal with the spontaneous emission of the NV center. At the same time, we introduce dielectric materials to suppress the fluorescence quenching effect and the compensation of the diamond total internal reflection effect. Simulation results show that the fluorescence rate and

sensitivity can be theoretically optimized by 70.1 and 8.37 times, respectively. In the future, the sensing performance of this scheme with scalability potential can be further optimized by bottom-up processing.

References:

[1] M. Radtke, E. Bernardi, A. Slablab, R. Nelz, and E. Neu, "Nanoscale sensing based on nitrogen vacancy centers in single crystal diamond and nanodiamonds: Achievements and challenges," Nano Futures, vol. 3, no. 4, 2019, doi: 10.1088/2399-1984/ab5f9b.

[2] L. Zhao, X. Shen, L. Ji, and P. Huang, "Inertial measurement with solid-state spins of nitrogen-vacancy center in diamond," 2022, Taylor & Francis. doi: 10.1080/23746149.2021.2004921.

[3] G. Wang et al., "Emulated nuclear spin gyroscopescope with 15 N-V centers in diamond," Phys Rev Appl, vol. 22, no. 4, p. 1, 2024, doi: 10.1103/PhysRevApplied.22.044016.

[4] A. Jarmola et al., "Demonstration of diamond nuclear spin gyroscopescope," Sci Adv, vol. 7, no. 43, pp. 3–8, Oct. 2021, doi: 10.1126/sciadv.abl3840.

[5] M. Hirose and P. Cappellaro, "Coherent feedback control of a single qubit in diamond," Nature, vol. 532, no. 7597, pp. 77–80, 2016, doi: 10.1038/nature17404.

[6] D. Le Sage et al., "Efficient photon detection from color centers in a diamond optical waveguide," Phys Rev B, vol. 85, no. 12, pp. 1–4, 2012, doi: 10.1103/PhysRevB.85.121202.

[7] M. P. Ledbetter, K. Jensen, R. Fischer, A. Jarmola, and D. Budker, "Gyroscopescopes based on nitrogen-vacancy centers in diamond," Phys Rev A (Coll Park), vol. 86, no. 5, p. 052116, Nov. 2012, doi: 10.1103/PhysRevA.86.052116.

[8] A. Ajoy and P. Cappellaro, "Stable three-axis nuclear-spin gyroscopescope in diamond," Phys Rev A, vol. 86, no. 6, pp. 1–7, 2012, doi: 10.1103/PhysRevA.86.062104.

[9] D. Hopper, H. Shulevitz, and L. Bassett, "Spin Readout Techniques of the Nitrogen-Vacancy Center in Diamond," Micromachines (Basel), vol. 9, no. 9, p. 437, Aug. 2018, doi: 10.3390/mi9090437.

[10] J. F. Barry et al., "Sensitivity optimization for NV-diamond magnetometry," Rev Mod Phys, vol. 92, no. 1, p. 015004, Mar. 2020, doi: 10.1103/RevModPhys.92.015004.

[11] N. Kongsuwan et al., "Suppressed Quenching and Strong-Coupling of Purcell-Enhanced Single-Molecule Emission in Plasmonic Nanocavities," ACS Photonics, vol. 5, no. 1, pp. 186–191, 2018, doi: 10.1021/acsphotonics.7b00668.

[12] X. Meng, S. Liu, J. I. Dadap, and R. M. Osgood, "Plasmonic enhancement of a silicon-vacancy center in a nanodiamond crystal," Phys Rev Mater, vol. 1, no. 1, p. 015202, Jun. 2017, doi: 10.1103/PhysRevMaterials.1.015202.

[13] J. T. Choy et al., "Enhanced single-photon emission from a diamond–silver aperture," Nat Photonics, vol. 5, no. 12, pp. 738–743, Dec. 2011, doi: 10.1038/nphoton.2011.249.

Time-resolved Fluorescence Detection Depending on Bias Magnetic Field in Three-axis Quantum Gyroscope

Wang Yuhai1, Zhao Liye[1,*] , Xukai Ding[1] and Ji Lumin[1]

(1. Key Laboratory of Micro-Inertial Instrument and Advanced Navigation Technology, Department of Instrument Science and Engineering, Southeast University, Nanjing 210096, China)

Abstract: A variety of solid-state gyroscope schemes based on the negatively charged nitrogen vacancy (NV) center in diamond have been proposed for rotation measurement, including a three-axis gyroscope scheme. However, it is difficult to solve the signal under a weak magnetic field because of the superposition of four axial signals. Here, we pay attention to the high bias magnetic field, under which the mixing of electron spins will change the spin dynamics of NV center. Firstly, we explain the scheme of measuring angular velocity by investigating the accumulation of geometric phase. We have established a seven-level model of PL intensity and bias magnetic field. The model quantitatively characterizes the dependence of laser polarization efficiency and PL intensity on bias magnetic field. This provides a theoretical basis for quantum spin measurement and control of three-axis gyroscope under a strong magnetic field.

 Key words: bias magnetic field; NV center; three-axis gyroscope; PL intensity.

1　Introduction

　　Diamond nitrogen vacancy (NV) center solid-state quantum spin gyroscope combines the interaction of electron spin and nuclear spin, and uses the quantum spin structure in NV center and the spin information caused by various fluorescent radiation systems to read the rotation information. It has many advantages, such as high solid-state spin density, stable ground state and excited state energy level structure, polarization detection in visible light band, long spin relaxation time at atmospheric pressure and room temperature, and can be controlled by microwave and radio frequency signals. Based on the principle of Berry phase of the quantum states, gyroscopes of using NV center electronic spins and ^{14}N nuclear spins have been proposed[1][2][3]. In 2020, the A.V.Akimov[4] team released a nuclear spin gyroscope based on pseudopotential magnetic field dynamics phase, with angular velocity measurement sensitivity of 52 [°/s]/HZ$^{1/2}$. In this paper, based on the influence of bias magnetic field in three-axis gyroscope on the mixing of spin eigenstates, a seven-level model of energy level transition under different magnetic field strengths and directions is established. The influence of bias magnetic field on laser polarization efficiency, photoluminescence fluorescence and ODMR contrast is analyzed.

2　GYROSCOPES BASED ON BERRY PHASE OF SPINS IN DIAMOND

　　The diamond NV center inertial measurement technology relies on the geometric phase

accumulation of spin states during inertial rotation. The NV center is an atomic scale defect composed of displaced nitrogen near vacancies in the diamond lattice. Suppose the diamond is rotated with an angular velocity Ω about an axis z oriented at an angle θ relative to the NV symmetry axis, and define the spherical coordinates $\phi(t)=\Omega t$ [Fig 1(a)]. The Hamiltonian of the rotating NV center can be obtained by a rotational transformation $R_{NV} = e^{-i\phi S_z}e^{-i\theta S_y}$ on the stationary Hamiltonian H_0. The Hamiltonian $H(t)$ and the eigenstates are written in the spin basis of the inertial (laboratory) coordinate system, which is simply given by equation (1) and (2).

$$H(t) = R_{NV}H_0R_{NV}^{\dagger} = D\begin{pmatrix} \dfrac{\cos^2\theta+1}{2} & \dfrac{\cos\theta\sin\theta}{\sqrt{2}}e^{-i\phi} & \dfrac{1-\cos^2\theta}{2}e^{-2i\phi} \\ \dfrac{\cos\theta\sin\theta}{\sqrt{2}}e^{i\phi} & \sin^2\theta & \dfrac{\cos\theta\sin\theta}{\sqrt{2}}e^{-i\phi} \\ \dfrac{1-\cos^2\theta}{2}e^{2i\phi} & \dfrac{\cos\theta\sin\theta}{\sqrt{2}}e^{i\phi} & \dfrac{\cos^2\theta+1}{2} \end{pmatrix} \tag{1}$$

$$\psi_1' = \begin{pmatrix} \cos^2\dfrac{\theta}{2} \\ \dfrac{\sqrt{2}}{2}\sin\theta e^{i\phi} \\ \sin^2\dfrac{\theta}{2}e^{2i\phi} \end{pmatrix} \psi_0' = \begin{pmatrix} -\dfrac{\sqrt{2}}{2}\sin\theta e^{-i\phi} \\ \cos\theta \\ \dfrac{\sqrt{2}}{2}\sin\theta e^{i\phi} \end{pmatrix} \psi_{-1}' = \begin{pmatrix} \sin^2\dfrac{\theta}{2}e^{-2i\phi} \\ -\dfrac{\sqrt{2}}{2}\sin\theta e^{-i\phi} \\ \cos^2\dfrac{\theta}{2} \end{pmatrix} \tag{2}$$

The geometric phase, after the crystal has been rotated along a trajectory, defined as rotation about the z axis with fixed θ, is given by equation (3).

$$\gamma_m(t) = i\int_0^t \psi_m' |\frac{\partial}{\partial t'}\psi_m' dt = m\Omega(1-\cos\theta)t \tag{3}$$
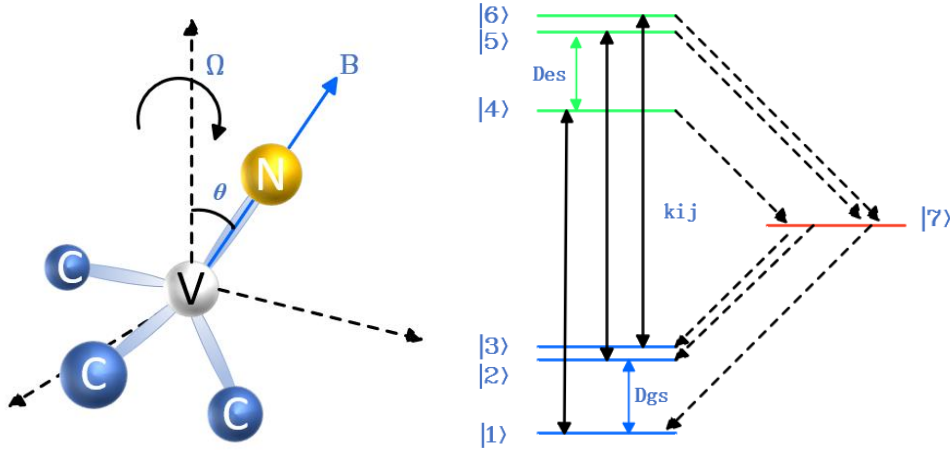
Fig.1 (a) Schematic diagram of angular velocity measurement principle based on NV center **(b)**Seven-level energy level structure considered for the NV center

This phase-like dependence of the interaction Hamiltonian on the NV orientation can be eliminated via a gauge transformation, of the form

$$\psi_m'(t) \rightarrow e^{-im\Omega t}\psi_m'(t) \tag{4}$$

The accumulated geometric phase can be read out by using a Ramsey scheme. the fluorescence intensity and obtained Sensitivity is given by

$$F = \eta N \left[1 - R\cos\left(\Omega\cos\theta t\right) \right] \tag{5}$$

$$S = 1 / \left(R\sqrt{\eta N T_2^* \tau} \right) \tag{6}$$

Here, R is the is the detection contrast for eigenstates between ms=0 and ms=±1, η is the collection efficiency, N is the number of participating NV centers, T_2^* is the dephasing time, τ is the total integration time. Similar to the measurement of three-dimensional vector magnetic fields[5], the bias magnetic field applied to a three-dimensional gyroscope does not need to be along the NV axis of the diamond. The literature[6] provides a scheme for detecting the nonlinear phase using spin-echo interferometry of a single nitrogen-vacancy qubit in a diamond rotating at 200 000 rpm with a magnetic field at an angle of 54.7 ° to the NV axis.

3 Photodynamics of NV centers in a biased magnetic field

3.1 Model

The ground level of NV center is a spin triplet $^3A_2$(S=1), while excited level is a spin triplet ^{3}E(S=1). Once optically excited to the ^{3}E level, the NV center can relax either through the same radiative transition producing a broadband red PL, or through a secondary path involving non-radiative intersystem crossing (ISC) to singlet states. Therefore, the radiative fluorescence of ms=0 state is brighter than that of ms=±1 states, which can be optically polarized to ms=0 state and read out. The ground state spin Hamiltonian

$$H_{gs} = D_{gs}S_z^2 + \gamma_e B\cos\theta S_z + \gamma_e B\sin\theta S_x \tag{7}$$

where D_{gs}=2.87GHz correspond to zero field splitting, γe is the spin magnetic ratio, θ is the angle between the magnetic field and the NV axis. The excited level spin Hamiltonian is thus simply given by equation (7) while replacing D_{gs} by D_{es}=1.42GHz. We note that in the absence of a magnetic field, the eigenstates of ground level and excited level are labelled $|i\rangle$ .In this work, the single metastable level is labelled $|7\rangle$. The transition rate from $|i\rangle$ to $|j\rangle$ is represented by k_{ij} [Fig1(b)]. When the applied bias magnetic field is not along the NV axis, the spin eigenstates under a transverse magnetic field can be represented by a linear combination of zero field eigenstates. The coefficients $k_{ij}(B, \theta)$ of the seven level model under the influence of a bias magnetic field are determined by k_a, k_b, k_{47}, k_{57}, k_{71}, k_{72}[7].

$$|i'\rangle = \sum_{j=1}^{7} c_{ij}\left(B,\theta\right)|j\rangle \tag{8}$$

$$k_{ij}\left(B,\theta\right) = \sum_{m=1}^{7}\sum_{n=1}^{7} \left|c_{im}\left(B,\theta\right)\right|^2 \left|c_{jn}\left(B,\theta\right)\right|^2 k_{mn} \tag{9}$$

The parameters simulated in the seven level model are obtained from literature[8][9], the corresponding parameter parameters may change due to factors such as laser power and diamond.

Tab.1 Transition Rates in Numerical Simulation

k_{ij}	Transition Rate(ns⁻¹)	k_{ij}	Transition Rate(ns⁻¹)
k_a	0.628	k_b	0.4396
k_{47}	0.0314	k_{71}	0.0207
k_{57}	0.1884	k_{72}	0.0069

After laser polarization of spin states, the population of each excited state decays exponentially. Here, n_i and τ_i represents the population and lifetime of each excited state. By summing up the three excited states, the final time-resolved PL signal obtained is:

$$R(t,B,\theta) = \eta \sum_{i=4}^{6} \sum_{j=1}^{3} k_{ij}(B,\theta) n_i(0) e^{-t/\tau_i}$$

(10)

3.2 Experimental setup

The inertial measurement device system based on diamond NV color center includes an optical system, a three-dimensional static magnetic field system, a spin manipulation system, a timing pulse system, and a data acquisition and processing system. The excitation source is a solid-state laser emitting at a wavelength $\lambda = 532$ nm. The objective lens is mounted on an XYZ scanning stage to allow fast laser scanning. High Pressure High Temperature (HPHT) synthesis process, purchased from Element Six, which has {100} oriented faces. The three-dimensional static magnetic field system consists of permanent magnets and a three-axis displacement platform, providing an arbitrary directional bias magnetic field B. A 20-μm copper wire is spanned on the surface of the diamond. Installing a square magnet with a diameter of two centimeters on an XYZ three-axis displacement table can provide a constant strong magnetic field with a size of 0-0.3T.

3.3 Results and discussion

In diamond ensemble samples, NV centers have four axes, which is the design principle of three-axis vector magnetometers and gyroscopes. By using permanent magnets fixed on the XYZ three-axis displacement table to apply magnetic fields of different strengths and directions, it is possible to make the magnetic field direction parallel to one of the NV axes, forming a fixed angle of 109°28′ with the other three NV axes[see Fig2 (a)].
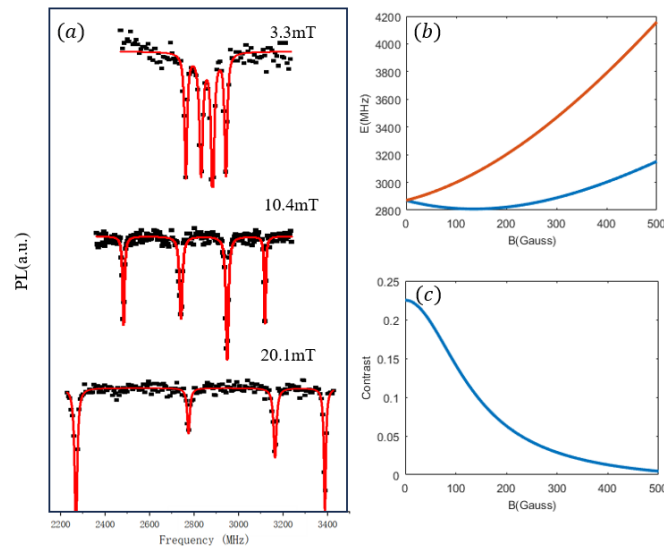
Fig.2 (a) ODMR image of magnetic field increasing along a certain NV axis (b) The required microwave frequency for driving the magnetic field at an angle of 109.5° (c) The variation of fluorescence signal with the density of magnetic field

In optical magnetic resonance experiments, the frequency of the microwave field corresponds to the frequency of electron spin transitions, and the number of fluorescence is determined by the population of excited states and the rate of transition from excited states to ground states. Two pairs of resonance peaks can be observed from Figure 2(a). One group is strictly symmetrical by Dgs=2870MHz, while the other group shows nonlinear changes with the increase of the magnetic field [see Fig2 (b)]. The fluorescence contrast decrease and disappear with the increase of the magnetic field [see Fig2 (c))]. Angular velocity detecting through ODMR is therefore inefficient in the regime of a 'high' off-axis magnetic field. It hard to achieve three-dimensional inertial measurement because of multi-peak signals overlap in low magnetic fields.
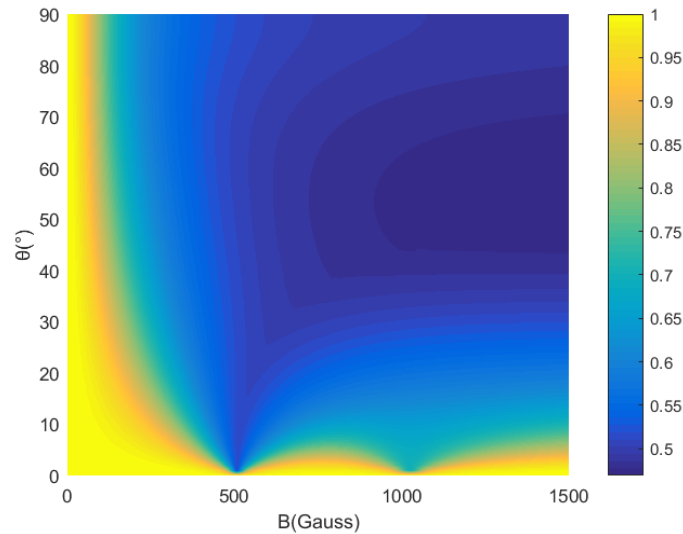
Fig.3   Calculated PL intensity varies with the density and direction of the magnetic field

Finally, we discussed the angle dependence of NV centers optical properties on B field. The increase in magnetic field density leads to a population change in the intrinsic state of laser polarization to the ground state energy level. Figure3 shows the relationship between calculated PL intensity varies with the density and direction of the magnetic field, where B varies from 0 to 1500G and θ varies from 0 to 90 degrees. Regardless of the magnitude of magnetic field B, PL always decreases with increasing angle θ. This effect results from a mixing of electron spin states both in the ground and excited levels. When the Larmor precession frequency wsinθ caused by transverse magnetic field component is much lower than the D-wcosθ, the component can be calculated as a perturbation and has a weak impact on the electron spin state, with PL decreases slowly for B<100G. We observed that the PL intensity rapidly decreases around B=510G and 1024G, corresponding to spin mixing caused by the ground state (GSLAC) and excited state level crossing (ESSAC) with only a small angle change. This phenomenon can be used to achieve polarization of nuclear spin. In fact, spin mixing in the ±1 state can also occur near zero field, but this cannot be distinguished by fluorescence observation.

4   Conclusion

In this article, we present a scheme for a gyroscope based on diamond spin BERRY phase and investigate the effect of bias magnetic field settings on the optical properties of diamond. We have

developed a seven-level model related to magnetic field density and direction, which demonstrates the decrease in laser polarization efficiency, PL intensity, and ODMR contrast when the magnetic field angle is not parallel to the NV axis. We analyzed the application of these effects in the design of three-axis diamond quantum gyroscopes and nuclear spin polarization manipulation. The high off-axis magnetic field reduces the sensitivity of gyroscope angular velocity measurement in terms of fluorescence intensity and ODMR contrast. This study is beneficial for avoiding signal superposition and coupling of NV signals in four axes under small magnetic fields, as well as for signal attenuation and disappearance under strong magnetic fields.

References:
[1] Maclaurin D, Doherty M W, Hollenberg L C L, Martin A M. Measurable Quantum Geometric Phase from a Rotating Single Spin[J]. Physical Review Letters, 2012, 108(24): 240403.
[2] Ledbetter M P, Jensen K, Fischer R, Jarmola A, Budker D. Gyroscopes based on nitrogen-vacancy centers in diamond[J]. Physical Review A, 2012, 86(5): 052116.
[3] Ajoy A, Cappellaro P. Stable three-axis nuclear-spin gyroscope in diamond[J]. Physical Review A, 2012, 86(6): 062104.
[4] Soshenko V V, Bolshedvorskii S V, Rubinas O, Sorokin V N, Smolyaninov A N, Vorobyov V V, Akimov A V. Nuclear Spin Gyroscope based on the Nitrogen Vacancy Center in Diamond[J]. Physical Review Letters, 2021, 126(19): 197702.
[5] Schloss J M, Barry J F, Turner M J, Walsworth R L. Simultaneous Broadband Vector Magnetometry Using Solid-State Spins[J]. Physical Review Applied, 2018, 10(3): 034044.
[6] Wood A A, Hollenberg L C L, Scholten R E, Martin A M. American Physical Society, 2020. Observation of a Quantum Phase from Classical Rotation of a Single Spin[J]. Physical Review Letters, 2020, 124(2): 020401.
[7] Tetienne J-P, Rondin L, Spinicelli P, Chipaux M, Debuisschert T, Roch J-F, Jacques V. IOP Publishing, 2012. Magnetic-field-dependent photodynamics of single NV defects in diamond: an application to qualitative all-optical magnetic imaging[J]. New Journal of Physics, 2012, 14(10): 103033.
[8] Song Y, Tian Y, Hu Z, Zhou F, Xing T, Lu D, Chen B, Wang Y, Xu N, Du J. Optica Publishing Group, 2020. Pulse-width-induced polarization enhancement of optically pumped N-V electron spin in diamond[J]. Photonics Research, 2020, 8(8): 1289–1295.
[9] Gupta A, Hacquebard L, Childress L. Efficient signal processing for time-resolved fluorescence detection of nitrogen-vacancy spins in diamond[J]. Journal of the Optical Society of America B, 2016, 33(3): B28.

Comparative Analysis of Algorithms
for Increasing Fault Tolerance of Complexes
with Redundant Strapdown Inertial Navigation Systems

E.V. Lukoyanov, E.V. Dranitsyna

(Concern CSRI Elektropribor, ITMO University, St. Petersburg, Russia)

Abstract: Three algorithms for increasing fault tolerance of complexes—integrated systems with redundant strapdown inertial navigation systems (SINS)—are analyzed and compared from the standpoint of their efficiency. The proposed algorithms are based on the method of pairwise difference calculation and its various modifications. Comparison is supported by semi-natural simulation, in which the experimental data are complemented with the simulated faults.

1 Introduction

Like any equipment, SINS are prone to faults and can generate navigation solutions with errors that significantly higher than the stated accuracy. This can be caused by hardware or software faults, as well as by operating conditions that differ from those specified for these systems. In order to increase fault tolerance of SINS-based integrated systems (IS) [1,2], it is necessary to implement methods for fault detection and isolation (FDI) [3-7].

FDI methods for SINS can be implemented at the levels of either inertial sensors or IS [8]. The following methods can be are used for sensors: checking for exceeding the ranges of measured values, based on the expected object's motion dynamics; integrity monitoring by direct checks of the consistency between the values calculated from different combinations of measurements, or with the model of the object's motion. Different FDI methods are not mutually exclusive and are often combined.

If a fault is detected in an inertial sensor, its output data cannot be used any further. The absence of redundant inertial sensors means a failure of the SINS as a whole. For this reason, redundant SINS are often provided to increase fault tolerance at the IS level [9]. In this case, three identical systems are commonly used, wherein the faulty system is determined by comparing pairwise differences of the generated navigation parameters of the same type. As a result, the faulty SINS is excluded from generation of a general navigation solution. This approach is used not only in marine navigation, but also in aviation [10-12] and satellite navigation systems [13, 14].

The purpose of this study is to compare the efficiency of three methods for detecting faulty SINS caused by abnormal readings of inertial sensors when using a triple system. The methods we propose are based on the analysis of pairwise differences (PD) for the parameters measured by inertial sensors, namely:

- detection by the weighted sum of PD;
- detection using an adaptive threshold for PD;
- detection using an adaptive threshold for the principal components [15] of PD.

2 FDI problem statement

In comparison with the existing algorithms, in which the FDI problem is solved at the IS level, so that the values of the SINS output navigation and dynamic parameters are used, the authors propose to implement the isolation of faults at the sensor level, which will reduce the time from the occurrence of a fault to its detection. In addition, this approach will make it possible, due to redundancy, to replace unreliable measurements of sensors, and thereby ensure isolation of faults and the uninterrupted output of the navigation solution.

The measurements used to solve the diagnostics problems of a redundant system are a composite vector of output signals from similar sensors of different SINS, for example, accelerometers, gyroscopes or temperature sensors:

$$S_i = \begin{bmatrix} X_{1,i} & Y_{1,i} & Z_{1,i} & X_{2,i} & Y_{2,i} & Z_{2,i} & X_{3,i} & Y_{3,i} & Z_{3,i} \end{bmatrix}^{\mathrm{T}}, \tag{1}$$

where $X_{r,i}$, $Y_{r,i}$, $Z_{r,i}$ is a signal from the r-th SINS ($r = 1$ is the main SINS, $r = 2$, $r = 3$ are redundant) along the corresponding measuring axis at the i-th moment in time.

It is necessary to solve the problem of detecting the fact of the fault at the i-th moment in time and to determine the number of the sensor whose data are not reliable.

3 Experimental data

The comparison of the algorithms was carried out by means of semi-natural simulation of the three SINS operation under conditions of a vibrating stationary base. Then, the obtained data were complemented with the simulated faults. In particular, a class of multiple (from 1 to 5 cycles) additive faults of medium and low levels was simulated, occurring at arbitrary moments in the gyroscope readings, along different axes in three identical SINS. The fault amplitude was set randomly in the range from 100 to 1000 °/h.

4 Brief description of the compared algorithms

The first algorithm is based on the analysis of the weighted sum of pairwise differences for the values measured by inertial sensors along the same axes. For each measurement S_i, we calculate the vector of pairwise differences, after which the weighted sum of its components is divided by the average variance of the vector S_i components (it is /assumed that it is identical for the sensors installed on different axes and in different systems). In this case, the function of making a decision on the fault occurrence is quite simple: if the value of the sum exceeds a certain specified threshold (in accordance with the expert data on the system), then the system is faulty.

In the second algorithm, unlike the first one, the operation threshold is calculated individually for each pairwise difference, which allows increasing the algorithm's sensitivity to faults. In order to determine the fact of a fault, it is sufficient to determine that the individual threshold is exceeded at least for one of the pairwise differences.

The third algorithm is, to a certain extent, a variation of the second approach in terms of formation of an adaptive detection threshold. The difference is that in the calculation of the threshold value, we no longer use the values of the pairwise differences themselves as a parameter, but their transformations to the principal components. To determine the fact of a fault, it is sufficient to determine that the individual threshold is exceeded at least for one of the principal components of

the pairwise differences.

5 Comparative analysis of the algorithm's performance

For each algorithm, the confusion matrix was calculated, which is classical in binary classification problems [16], a special case of which is the FDI problem.

Analysis of the simulation results has shown that the algorithms using the adaptive threshold have a higher detection recall in the case of small- and medium-level faults (~ 30-55%), while the classical pairwise difference method provides a solution to the problem at a level of (~ 2-37%). For large-level faults, all algorithms have shown approximately the same result (~ 99-100%). For the problem of finding a failed sensor, we calculated similar metrics, which showed a 100% result in all experiments. It follows that to solve diagnostics problems at a high level, it is necessary to provide, first of all, a solution to the problem of detecting small-level faults, while solution to the problem of finding faults after the detection stage does not present difficulties any longer.

6 Conclusion

A comparative analysis of the efficiency of three algorithms for increasing the fault tolerance of IS with redundant SINS based on the comparison of pairwise differences in the parameters measured by inertial sensors has been carried out. The comparison of the methods was performed by means of semi-natural simulation of the three SINS operation under the same conditions but with different faults in the readings of inertial sensors. Based on the simulation results, it can be concluded that the most effort-consuming faults to detect are small-amplitude faults. In this case, methods with an adaptive threshold solve this problem more effectively. One of the possible lines of research is integration of the proposed algorithms for solving FDI problems of inertial sensors to achieve the greatest efficiency.

References:

[1] IMO Resolution MSC.252(83). Adoption of the revised performance standards for integrated navigation systems (INS).

[2] Standard IEC 61924-2:2021. Maritime navigation and radiocommunication equipment and systems - Integrated navigation systems (INS) - Part 2: Modular structure for INS - Operational and performance requirements, methods of testing and required test results.

[3] Chen, J. and Patton R.J., Robust model-based fault diagnosis for dynamic systems. Springer Science & Business Media, 2012. – T. 3.

[4] Dmitriev, S.P., Kolesov, N.V., and Osipov, A.V., Informatsionnaya' nadezhnost' upravlenie i diagnostika navigatsionnykh system (Information reliability, monitoring and diagnosis of navigation systems), St. Petersburg: OAO Kontsern TsNII, 2003.

[5] Kolesov, N.V., Tolmacheva, M.V., and Yukhta, P.V., Real-time systems. Planning, analysis, and diagnosis. St. Petersburg: OAO Kontsern TsNII, 2014.

[6] Wang R., Xiong, Z., Liu, J., Xu, J., and Shi, L. Chi-square and SPRT combined fault detection for multisensor navigation, IEEE Transactions on Aerospace and Electronic Systems, 2016, vol. 52, no. 3, pp. 1352-1365.

[7] Koshaev, D.A., Kalman filter-based multialternative method for fault detection and estimation, Automation and Remote Control, 2010, vol. 71, no. 5, pp. 790-802.

[8] Groves, P.D., Principles of GNSS, inertial, and multisensor integrated navigation systems, [Book review], IEEE Aerospace and Electronic Systems Magazine, 2015, vol. 30, no. 2, pp. 26-27.

[9] Dubrova E. Hardware redundancy, Fault-Tolerant Design, 2013, pp. 55-86.

[10] Zemlyanyi, E.S. and Tektov, M.V., Quorum-element for determining the parametric fault of heading and vertical

reference systems, Vestnik Tul'skogo gosudarstvennogo universiteta. Tekhnicheskiye nauki, 2023, no. 6, pp. 104-109.

[11] Dai, Y., Lai, J., Zhang, Q., Li, Z., and Shen, Y., An Improved Fault Detection and Isolation Method for Airborne Inertial Navigation System/Attitude and Heading Reference System Redundant System, Aerospace, 2023, vol. 10, no. 12, p. 1024.

[12] Slim, M., Saied, M., Mazeh, H., Shraim H., Francis C. Fault-Tolerant Control Design for Multirotor UAVs Formation Flight, Gyroscopy and Navigation, 2021, vol. 12, no. 2, pp. 166-177. DOI:10.1134/S2075108721020061

[13] Humphreys, T.E., Ledvina, B.M., Psiaki, M.L., O'Hanlon, B.W., and Kintner, P.M., Assessing the spoofing threat: Development of a portable GPS civilian spoofer, Proc. 21st International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2008), 2008, pp. 2314-2325.

[14] Veremeyenko, K.K. and Antonov, D.A., Detection of Satellite Navigation System Faults in the Integrated Navigation System, XXIII St. Petersburg International Conference on Integrated Navigation Systems, 2016.

[15] Abdi, H. and Williams L.J., Principal component analysis, Wiley interdisciplinary reviews: computational statistics. 2010, vol. 2, no.   4, pp. 433-459.

[16] Li, B., Wang, H., Mu, L., Shi, Z., & Du, B.   A configuration design method for a redundant inertial navigation system based on diagnosability analysis //Measurement Science and Technology, 2020, vol. 32, no. 2 p. 025111.

Research on Magnetic Shielding of High-Precision Fiber Coil Module

Zang Zerun[1], Hu Xudong[1], Ma Fengyuan[1], Liu Pan[2], Hong Wei[1], Zhou Yuan[1]

(1.Xi'an Aerospace Precision Electromechanical Institute, Xi'an 710100, China; 2.Beijing Institute of Control and Electronic Technology, Beijing 100038, China)

Abstract: Fiber optic gyroscope is a key device in inertial navigation systems. The magnetic field environment it is in is complex and diverse, including geomagnetic field. The external magnetic field causes optical errors in the fiber coil through the Faraday effect, thus affecting the output of the fiber optic gyroscope. This paper combined with the photoelectric separation structure of high-precision fiber optic gyroscopes, finite element software is used to conduct simulation analysis on high-precision fiber coil modules. The analysis results of this paper can provide quantitative references for the errors and protection of high-precision fiber optic gyroscopes in magnetic field environments.

 Key words: fiber optic gyroscope; fiber coil; error; magnetic shielding

1　Introduction

Fiber optic gyroscopes are widely used in the field of inertial navigation. The external magnetic field will cause the magneto-optic effect of the optical fiber material and trigger the non-reciprocal phase difference of the fiber coil[1], resulting in errors in the gyroscope output.

In order to reduce the influence of the external magnetic field on the fiber optic gyroscope, The common method is to install a magnetic shielding cover on the outside of the fiber optic gyroscope. Through special material and structural design, the external magnetic field is guided and dispersed, thereby reducing the influence on the fiber optic gyroscope. Some studies have pointed out that magnetic sensors can be installed to measure the magnetic field in the environment, and magnetic field compensation can be carried out through control circuits or software algorithms to eliminate the errors caused by the magnetic field[2]. There are also studies that point out that through new material research and development, active exploration of the application of diamagnetic optical fiber materials[3, 4] can reduce the influence of the external magnetic field on the fiber coil. The existing magnetic shielding structure of fiber optic gyroscopes is relatively simple. And with the development of high-precision fiber optic gyroscopes, the structure needs to be updated.

In this paper, by using three typical geometric relationships of fiber coils in a magnetic field, a relatively comprehensive magnetic field error model of fiber optic gyroscopes is established. Finite element analysis is carried out to determine the influences of factors such as thickness, complex surface, and double-layer shielding of the module shielding on the shielding effect.

2　Magnetic field error model of fiber coil

2.1　Magnetic field error theory of fiber optic gyroscope

The fiber optic gyroscope uses the Sagnac effect[5] of the fiber coil to sense the external angular

rate. Its formula is as follows:

$$\Phi_S = \frac{4\pi RL}{\lambda c}\Omega \qquad (1)$$

Where, Ω is the rotational angular velocity of the fiber coil, R is the radius of the fiber coil, L is the total length of the optical fiber, λ is the wavelength of light, and c is the speed of light propagation.

The Jones vector and Jones matrix are commonly used to describe the polarization state of light and the polarization characteristics of polarization elements. Through the derivation of multiplying the Jones vector of incident light by the Jones matrix of the polarization element[6], it is obtained that the phase difference generated after the clockwise and counterclockwise light waves of the fiber coil merge due to the influence of the magnetic field is:

$$\Phi = \sum_{i=1}^{n}\frac{4\xi_i t_i}{\Delta\beta}dz$$

(2)

Where, z is the length of the i-th section in the optical fiber, with the unit of rad/m; ξ_i is the circular birefringence caused by the Faraday effect, with the unit of rad/m; t_i is the circular birefringence caused by the torsion of the optical fiber, with the unit of rad/m; $\Delta\beta$ is the linear birefringence of the optical fiber itself. Based on this formula and according to different types of external magnetic fields, the following three types of magnetic field error models are constructed.

2.2 Radial and axial magnetic field error model

The fiber coil in a radially uniform magnetic field is shown in Figure 1. The plane of the fiber coil is parallel to the XOY plane. B_R represents the intensity of the radial magnetic field, and its angle with the X-axis is θ. Let z be the distance of any point A on the fiber coil relative to the starting point of the fiber coil. The total length of the optical fiber is L, and the radius of the fiber coil is R.
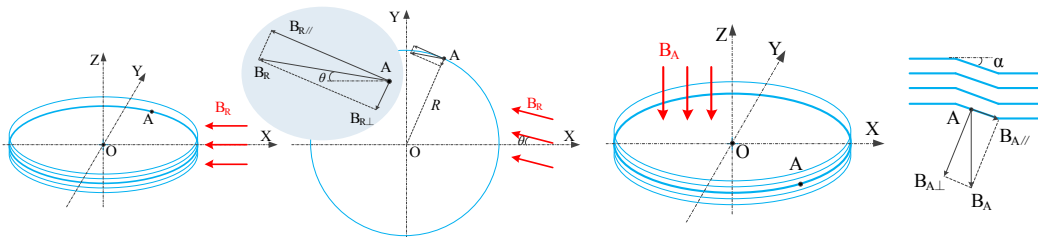
Fig.1 Fiber coil under radial magnetic field   **Fig.2** Fiber coil under axial magnetic field

According to the geometric relationship in Figure 1, the distribution of circular birefringence caused by the Faraday effect on the optical fiber is:

$$\xi(z) = VB_{R\parallel} = VB_R \sin(z/R + \theta) \qquad (3)$$

Where, V is the Verdet constant, with the unit of rad/m/mT. $B_\parallel$ is the component of the external magnetic field intensity in the direction of light propagation.

Bringing formula (3) into formula (2) and formula (1), the phase difference generated after the clockwise and counterclockwise light waves of the fiber coil merge due to the influence of the radial magnetic field can be obtained as follows:

$$\Omega_R = \frac{\lambda c VB_R}{\pi RL\Delta\beta}\int_0^L t(z)\sin(\frac{z}{R}+\theta)dz \qquad (4)$$

If the distribution of fiber torsion in the fiber length direction $t(z)$ is periodically distributed and the period is the same as that of $\sin(z/R+\theta)$, then $\int_0^L t(z)\sin(z/R+\theta)dz$ will have a maximum value, that is, the angular velocity error generated by the fiber coil under the influence of the radial magnetic field is the largest.

When $t(z)=t_0\sin(z/R)$, formula (4) is simplified as:

$$\Omega_{R1}=\frac{\lambda cVB_Rt_0}{2\pi R\Delta\beta}\cos\theta$$

（5）

The fiber coil in an axially uniform magnetic field is shown in Figure 2. The direction of the external magnetic field is perpendicular to the plane of the fiber coil.

According to the geometric relationship in Figure 2, the distribution of circular birefringence caused by the Faraday effect on the optical fiber is:

$$\xi(z)=\begin{cases}VB_AD(p)\sin\alpha, Turn-changing\ bending\ part\\0,\ Parallel\ winding\ part\end{cases}$$

（6）

Where, p is the number of layers of the fiber coil; $D(p)$ is a function of the number of fiber coil layers, characterizing the bending direction of the fiber turn-changing part on each layer. Its value is: 1 for upward bending and -1 for downward bending.

Bringing formula (6) into formula (2) and formula (1), the phase difference generated after the clockwise and counterclockwise light waves of the fiber coil merge due to the influence of the axial magnetic field is as follows:

$$\Omega_A=\frac{\lambda cVB_A}{\pi RL\Delta\beta}\sum_{j=1}^{pq}\left[\int_{L_j}^{L_j+s}t(z)D(p)\sin\alpha dz\right]$$

（7）

Where, q is the number of turns per layer of the fiber coil, and $p\cdot q=N$. j represents the ordinal number of the turn-changing bending part of the fiber coil. L_j and L_j+s are the starting and ending positions of the turn-changing position within the fiber coil. s is the length of the turn-changing part.

It can be obtained that if the distribution of fiber torsion in the fiber length direction $t(z)$ is periodically distributed and the period is the same as that of $D(p)$, then formula (9) will have a maximum value, that is, the angular velocity error generated by the fiber coil under the influence of the axial magnetic field is the largest.

When $t(z)=t_0D(p)$, formula (7) is simplified as:

$$\Omega_{A1}=\frac{\lambda cVB_At_0s}{2\pi^2R^2\Delta\beta}\sin\alpha=\frac{\lambda cVB_At_0D_{\text{fiber optic}}}{2\pi^2R^2\Delta\beta}$$

（8）

2.3　Error model of short-distance spherical magnetic field

As shown in Figure 3, it is a schematic diagram of the fiber coil being affected by a spherical divergent magnetic field.

According to the geometric relationship in Figure 3, the distribution of circular birefringence caused by the Faraday effect on the optical fiber is:

$$\xi(z)=VB_B{'}_{\parallel}=VB_B{'}\sin\delta=VB_B\sin\delta\cos\gamma\times10^{-2}/l_1$$

（9）

Bringing formula (9) into formula (2) and formula (1), the angular velocity error of the fiber coil in the spherical magnetic field is as follows:

$$\Omega_B = \frac{\lambda c V B_B}{100\pi RL\Delta\beta}\int_0^L \frac{t(z)d\sin(z/R)}{h^2+d^2-2dR\cos(z/R)+R^2}dz \qquad (10)$$
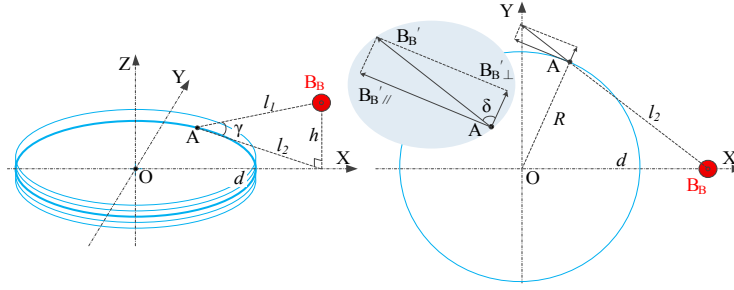
Fig.3   Fiber coil under spherical magnetic field

According to the geometric relationship in Figure 3, the distribution of circular birefringence caused by the Faraday effect on the optical fiber is:

$$\xi(z) = VB_B{}'_{\parallel} = VB_B{}'\sin\delta = VB_B\sin\delta\cos\gamma\times10^{-2}/l_1 \qquad (9)$$

Bringing formula (9) into formula (2) and formula (1), the angular velocity error of the fiber coil in the spherical magnetic field is as follows:

$$\Omega_B = \frac{\lambda c V B_B}{100\pi RL\Delta\beta}\int_0^L \frac{t(z)d\sin(z/R)}{h^2+d^2-2dR\cos(z/R)+R^2}dz \qquad (10)$$

If the distribution of fiber torsion in the fiber length direction is $t(z)=t_0/\sin(z/R)$, and formula (10) is scaled, then formula (10) will have a maximum value as follows:

$$\Omega_{B1} = \frac{\lambda c V B_B t_0 d}{100\pi R\Delta\beta\left[h^2+(d-R)^2\right]} \qquad (11)$$

3   Magnetic shielding simulation based on high-precision fiber coil module

High-precision fiber optic gyroscopes generally adopt photoelectric separation layout. Magnetic shielding is mainly carried out for the optical path part of the fiber optic gyroscope. The optical path part mainly refers to the fiber coil module. Its structure is shown in Figure 4.
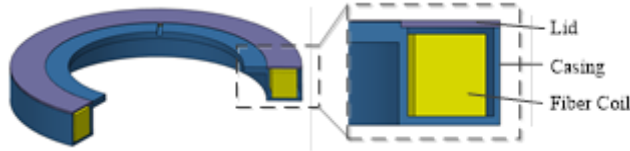
Fig.4   Structure of fiber coil module

In order to analyze the influence of different factors on the magnetic shielding effect, the three-dimensional model of magnetic shielding for the fiber coil module is classified, including a total of seven types of models: material, thickness, opening, complex surface, and double-layer shielding of the casing.

In the finite element simulation software ANSYS, the above seven types of models are used in turn to simulate radial magnetic fields, axial magnetic fields, and spherical magnetic fields. Taking the simulation of Model #2 using permalloy material for magnetic shielding as an example, its simulation result is shown in Figure 5 and Figure 6.

According to the above magnetic field strength data inside the fiber coil, take it as the magnetic field strength value in formulas (5), (8), and (11) in Section 2. Then set the following typical values of fiber coil parameters: light source wavelength $\lambda=1550\text{nm}$, speed of light $c=3.0\times10^8\text{m/s}$, Verdet constant of optical fiber $V=6\times10^{-4}\text{rad/m/mT}$, radius of fiber coil $R=0.06\text{m}$, diameter of

optical fiber is taken as $150\mu m$, fiber torsion amplitude is taken as 0.5rad/m, and the linear birefringence of polarization-maintaining fiber itself is $\Delta\beta = 2200\text{rad/m}$. After calculation, the magnetic field error values of different fiber coil models in different magnetic field situations can be obtained.

(a) Radial magnetic field   (b) Axial magnetic field   (c) Apherical magnetic field
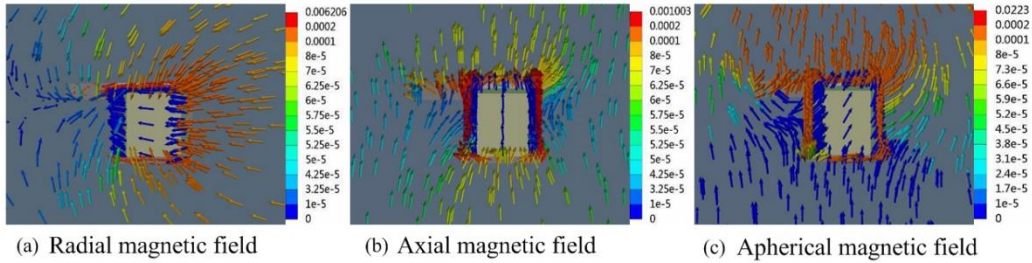
Fig.5   Vector diagrams of magnetic field cross-sections of modules

(a) Radial magnetic field   (b) Axial magnetic field   (c) Apherical magnetic field
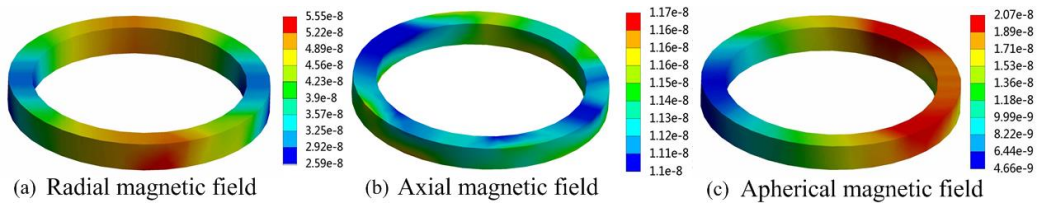
Fig.6   Distribution of magnetic field intensity of fiber coil

Since Models #2 to #7 all use permalloy with high magnetic permeability, and Model #1 using ordinary steel material has a relatively low magnetic permeability, which is very different from the simulation and calculation results of the other models. Therefore, the simulation and calculation results of Models #2 to #7 are compared as shown in Figure 7.
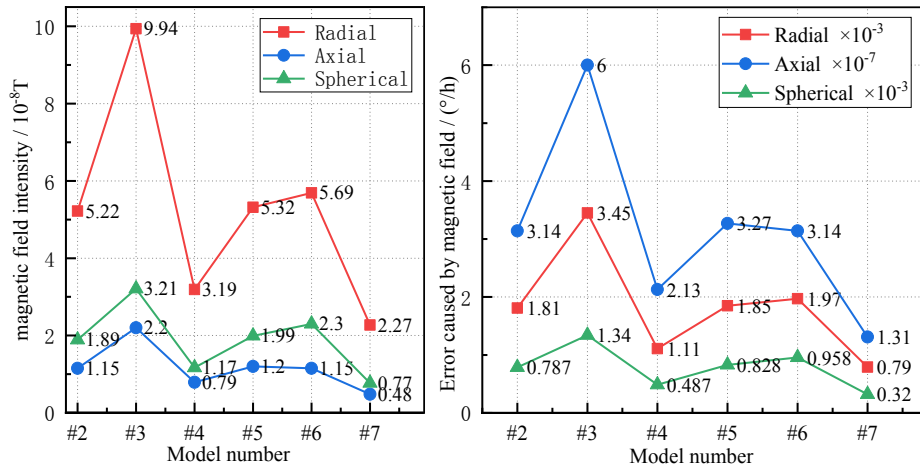
Fig.7   Comparison of magnetic shielding effects and errors

At the same time size and thickness, materials with high magnetic permeability have a greater improvement in the shielding effect of the shielding shell, and the shielding effect can differ by several orders of magnitude.

Referring to the magnetic field strength inside the module in the radial magnetic field situation in Figure 7, also using permalloy material, the thickness of the fiber coil casing of Model #4 is doubled compared to Model #2, and the internal magnetic field strength is reduced from 5.22×10^{-8}T to 3.19×10^{-8}T, a reduction of 39%. This shows that the thickness of the magnetic shielding shell and the magnetic field shielding effect are not in a linear relationship. As the thickness of the

shielding shell increases, the benefit of the magnetic shielding effect will become lower and lower.

And from the simulation results, it can be seen that the two cases of Model #5 and Model #6 are similar to the results of Model #2, with a change within 9%. Therefore, the shape and openings of the fiber coil module have little impact on the final magnetic shielding effect.

Referring to the magnetic field strength inside the module in the radial magnetic field situation in Figure 7, Model #7 adopts a double-layer shielding scheme. Compared to Model #2, the internal magnetic field strength is reduced from 5.22×10^{-8}T to 2.27×10^{-8}T, a reduction of 56%. Model #4 is a single-layer shielding with doubled thickness. The radial magnetic field error corresponding to Model #7 is smaller than that of #4, proving that double-layer shielding is better than single-layer thickened shielding of the same thickness. Therefore, adopting double-layer magnetic shielding can effectively improve the overall magnetic field protection ability of the fiber coil module without adding too much weight.

4 Conclusion

In this paper, according to the change in the polarization state of light waves in the optical fiber, using three typical geometric relationships of the fiber coil in the magnetic field, a relatively comprehensive fiber optic gyroscope magnetic field error model is established. Based on the fiber coil module of high-precision fiber optic gyroscopes, and three-dimensional finite element analysis is carried out to determine the influence of factors such as material, thickness, opening, complex surface, and double-layer shielding of the fiber coil module shielding on the shielding effect. In the practical application of fiber optic gyroscopes, using high-permeability materials for magnetic shielding of high-precision fiber coil modules can effectively reduce the influence of the external magnetic field on the fiber coil. Adopting double-layer magnetic shielding can take into account size and weight and better reduce the influence of the external magnetic field on the fiber coil. This paper provides a quantitative reference for the error and protection of high-precision fiber optic gyroscopes in a magnetic field environment.

References:

[1] Wang J, Feng L, Wang Q, et al. Reduction of angle random walk by in-phase triangular phase modulation technique for resonator integrated optic gyro[J]. Optics Express, 2016, 24(5): 5463-5468.

[2] Olson M A, Lange C H. Magnetic compensators for fiber optic gyroscopes[P]. U.S. Patent, 2007,7298491.

[3] Kim H K, Digonnet M J F, Kino G S. Air-core photonic-bandgap fiber-optic gyroscope[J]. Journal of Lightwave Technology, 2006,24(8):3169.

[4] Li J, Zhou Y R, Xue L Y. Magnetic error suppression in polarization- maintaining fiber optic gyro system with orthogonal polarization states[J]. Optical Fiber Technology, 2022.

[5] Wang Wei. High-Precision Fiber Optic Gyroscope Technology[M]. National Defense Industry Press, 2021.

[6] Liu Pan. Research on the mechanism analysis and suppression measures of magnetic field error of fiber optic gyroscope[D]. Harbin Engineering University, 2018.

Variation Analysis and Nonlinearity Optimization of Scale Factor of Fiber Optic Gyro Based on SLD Light Source

Yang Xiong, Hu Xudong, Zhou Yuan, Jia Xinliang, Liang Kaixu, Pan Zijun

(Xi'an Aerospace Precision Electromechanical Institute, Xi'an, 710100, China)

Abstract: In order to reduce the nonlinearity of the scale factor of the fiber optic gyro based on SLD light source under different input angular rates, we analyzed and tested the trend and mechanism of the scale factor at different rate points of the fiber optic gyro in the cross-stripe operating mode, and propose a reasonable and effective compensation scheme according to the change characteristics of the fiber optic gyro, and established a compensation model for the nonlinearity of the scale factor to realize the engineering application of the compensation scheme. The test samples are tested and analyzed to prove the feasibility of the compensation scheme, which is now widely used in the development of fiber optic gyro models.

Key words: fiber optic gyro; cross-stripe; scale factor nonlinearity; compensation scheme;

1 Introduction

At present, the fiber optic gyro in navigation system commonly adopts the cross-stripe technology to extend its measurement range, but due to the existence of the inherent spectral width of the SLD light source, the sinusoidal response curve of the interference light intensity and Sagnac phase difference is distorted, and the instantaneous summation of feedback phase and Sagnac phase difference in the closed-loop system is not zero, and the modulation error generated causes deterioration of the scale factor nonlinearity[1].In this paper, a scalar factor nonlinearity compensation scheme is proposed by software algorithm. After the test and validation of the gyro, it is proved that the scheme can effectively reduce the scalar factor nonlinearity.

2 Mechanism of Scale Factor Change in Cross-Stripe Mode

The scale factor of fiber optic gyro is the ratio the output value of gyro to the input angular rate, which can be expressed by the slope of a particular straight line, this ratio is calculated by least squares fitting of the data measured in the range of input angular rates, which is directly related to the sensitivity and accuracy of the gyro[2].In the range of input angular rate, the maximum deviation and the maximum output value by least squares fitting straight line of the fiber optic gyro is called the scalar factor nonlinearity, which determines the confidence of the fitted data. The scale factor integrally reflects the test accuracy of the fiber optic gyro and is one of the key indicators for evaluating the performance of the fiber optic gyro.

The response curve of the Sagnac effect is sinusoidal, and the rate signal required by the fiber optic gyro should be linear. The demodulated bias signal is used as an error signal during the processing of digital closed-loop signal, which is fed back into the system to generate an additional feedback phase difference $\Delta\phi_{FB}$, which is opposite to the phase difference $\Delta\phi_S$ caused by the rotation. The total phase difference $\Delta\phi_T = \Delta\phi_{FB} + \Delta\phi_S$ is zero, so that the system is always operated around a high slope point. Thus there exists a monotonic phase measurement interval

centered at zero for the fiber optic gyro, so the maximum range of the input angular rate can be expressed as:

$$\Omega_\pi = \frac{\lambda_0 c}{2LD} \tag{1}$$

As the fiber optic gyro precision index requirements increase, the fiber-optic coil L increases, the maximum input angular rate Ω_π within a single stripe decreases accordingly, and the expansion of the measurement range needs to be realized through the cross-stripe mode. In the ideal case, the relationship between the interference light intensity and the phase difference is shown in Fig.1(a), when the phase difference generated by the input angular rate is exceeded $\pm\pi$, the interference light intensity shows a periodic same-amplitude change with the increase of the phase difference, a and the gyro is able to maintain normal operation.

In practice, due to the wide spectral width of the SLD light source, when the phase difference generated by the input angular rate is exceeded $\pm\pi$, the interference light intensity shows a periodic non-uniform amplitude change with the increase of the phase difference as shown in Fig.1(b) [4,5].

(a)   (b)
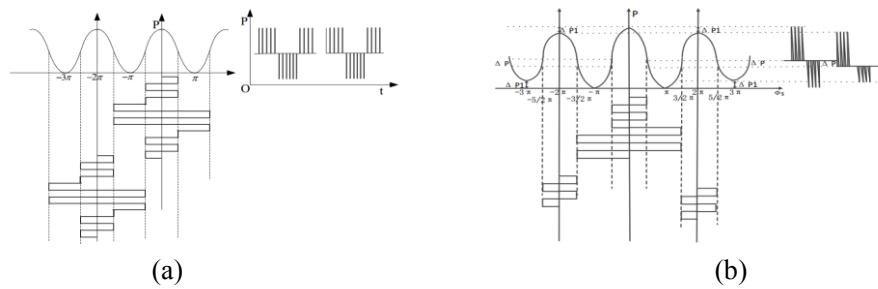
Fig.1   Cross-stripe modulation schematic

As can be seen in Fig.1(b), the interference light intensity error Δp generated by the fiber optic gyro in the cross-stripe mode, which will lead that the instantaneous summation of feedback phase and Sagnac phase difference in the closed-loop system is not zero, which can generate the modulation error.

The gyro sample operating in cross-stripe mode is taken for the scale factor test and the results are shown in Fig. 2.
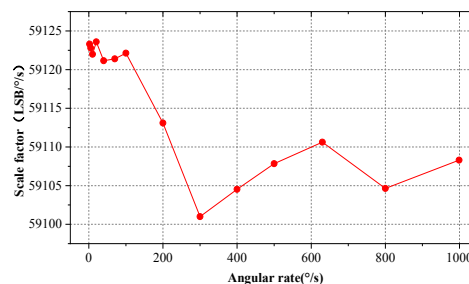
Fig.2   The trend of the scalar factor of the gyro sample in different input angular rates

As can be seen in Fig. 2, the first characteristic change of the gyro occurs when the angular

rate range is 100°/s ~ 200°/s, the second characteristic change of the gyro occurs when the angular rate range is 200°/s ~ 400°/s, which causes deterioration of the scale factor nonlinearity. The change in the scale factor in the test is due to a shift in the output value of the gyro with a steady increase in the input angular rate.

In order to clarify the specific rate point corresponding to the two characteristic changes of the gyro scale factor, a rate point is set every 10°/s when the angular rate is 100°/s~400°/s, and the rate test is performed with the model of this group of rate points, which is a mapping test of the change characteristics of the gyro scale factor in this rate range, and the results are shown in Fig.3.
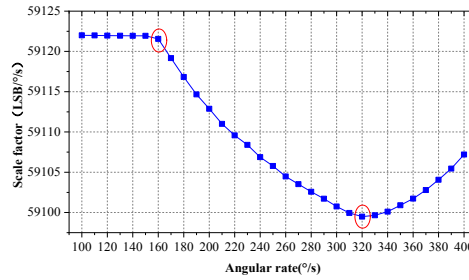
Fig.3 Mapping tests of the scale factor characterization

In conclusion, the rate point corresponds the first segmentation point of the scale factor at about 160°/s, and the rate point corresponds the second segmentation point corresponds to the rate point of the scale factor at about 320°/s. The two segmentation rate points divide the input angular rate range into three segments, with the first segment being the low-rate segment, in which the scale factor remains relatively stable; the second segment being the medium-rate segment, in which the scale factor decreases linearly; and the third segment being the high-rate segment, in which the scale factor appears to be sinusoidally changing.

3 Program design

3.1 Program conception

According to the characteristics of the gyro's scale factor changes in different rate ranges, a three-segment scale factor nonlinearity compensation scheme realized by software is proposed.

The gyro scale factor in the first rate segment can be used as a benchmark for compensation, so that the gyro in the second and third rate bands can maintain the linear fitting relationship output, which can achieve the purpose of the optimization of the nonlinear degree of the scale factor. The design flow of the scale factor nonlinearity compensation scheme are shown in Fig.4.
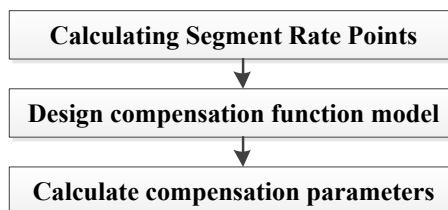
Fig.4 Design process of the program

3.2 Establish scale factor compensation model

3.2.1 Calculating Segment Rate Points

A linear fit is made to the rate points and the corresponding gyro outputs within the first rate segment to obtain the fit function:

$$f_1 = k_1 \times v + b_1 \tag{2}$$

At this point, the input angular rate v satisfies $v \leqslant v_c$.

Similarly a linear fit is made to the rate points and the corresponding gyro outputs within the second rate segment to obtain the fit function:

$$f_2 = k_2 \times v + b_2 \tag{3}$$

At this point, the input angular rate satisfies $v_c < v \leqslant v_{2c}$.

The intersection of the above fitted functions f_1 and f_2, which is taken as the first segmented rate point v_c, is obtained by the following equation:

$$v_c = \frac{(b_1 - b_2)}{(k_2 - k_1)} \tag{4}$$

As can be seen from Fig. 3, the scale factor K is minimized at the intersection of the second and third rate segments, and a quadratic polynomial is fitted to the fit function for the rate points with input angular rates in the range of $300\sim350°$/s and the corresponding scale factor K. The fit function is then calculated as follows:

$$K = av^2 + bv + c \tag{5}$$

At this point, the input angular rate v satisfies $300°$/s $\leqslant v \leqslant 350°$/s.

The extreme point of the above quadratic function as the second segmented rate point v_{2c} can be calculated by the following equation:

$$v_{2c} = -\frac{b}{2a} \tag{6}$$

Since the gyro output value is used as the characterization of the input angular rate in the internal logic algorithm of the software, the first segmented rate point v_c is substituted into Eq. (2) to obtain the corresponding gyro output value $f_c = k_1 \times v_c + b_1$, which leads to the first parameter of the compensation model: the first segmented point f_c. The second segmented rate point is

substituted into Eq. (3) to obtain the corresponding gyro output value $f_{2c} = k_2 \times v_{2c} + b_2$, which leads to the second parameter of the compensation model: the second segmented point f_{2c}.

3.3.2 Design compensation function model

Take the fitting function derived from the first rate segment as the reference function, and substitute it into the rate points in the second and third rate segments, respectively, to get the gyro fitting output reference, as shown in Fig. 5.
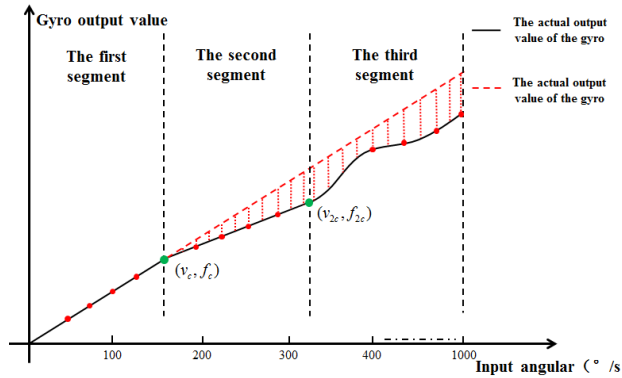
Fig.5 Gyro fitting output reference

In the third rate segment, the gyro output difference is $\Delta f_{2i} = f_1 - f_{2i} = k_1 \times v + b_1 - f_{2i}$ ($v_c < v \leqslant 2v_c$), and f_{2i} is the actual gyro output value corresponding to the rate points of the second rate segment.

In the third rate segment, the gyro output difference is $\Delta f_{3i} = f_1 - f_{3i} = k_1 \times v + b_1 - f_{3i}$ ($v > 2v_c$), and f_{3i} is the actual gyro output value corresponding to the rate points of the second rate segment.

The output difference between the actual gyro output and the gyro fitted output is used as the dependent variable, and the gyro output value is used as the independent variable, and the above results is linearly fitted to obtain the difference compensation model as shown in Fig.6, where the difference is the amount of compensation.
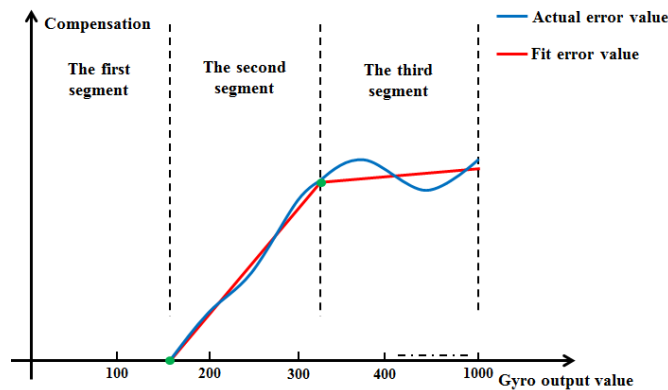
Fig.6 Gyro fitting output reference

3.3.3 Calculate the parameters of compensation

In the second rate segment, the compensation amount of the functional relationship is obtained by of the linear fitting, which about the output difference of gyro Δf_{2i} and the output value of gyro actually: $\Delta f_2 = k_{com2} \times f + b_{com2}$, which leads to the third parameter of the compensation model: the second segment slope k_{com2}, the fourth parameter of the compensation model: the second segment constant b_{com2}.

The final gyro output is obtained by adding the amount of compensation and the output value of gyro actually: $f_{com2} = f + \Delta f_2$ ($f_c < f \leq f_{2c}$).

In the third rate segment, the compensation amount of the functional relationship is obtained by of the linear fitting, which about the output difference of gyro Δf_{3i} and the output value of gyro actually: $\Delta f_3 = k_{com3} \times f + b_{com3}$, which leads to the fifth parameter of the compensation model: the third segment slope k_{com3}, the fifth parameter of the compensation model: the third segment constant b_{com3}.

The final gyro output is obtained by adding the amount of compensation and the output value of gyro actually: $f_{com3} = f + \Delta f_3$ ($f > f_{2c}$).

4 Conclusion

After the scale factor compensation scheme is implemented in the software, the software reads out the six parameters derived in the previous section from the hardware device substitutes them into the model, the gyro is tested according to the same sate points, and the results show that the scale factor nonlinearity was reduced from the previous 274ppm to 55ppm. The comparison of the tests results before and after the compensation scheme as shown in Fig.7.
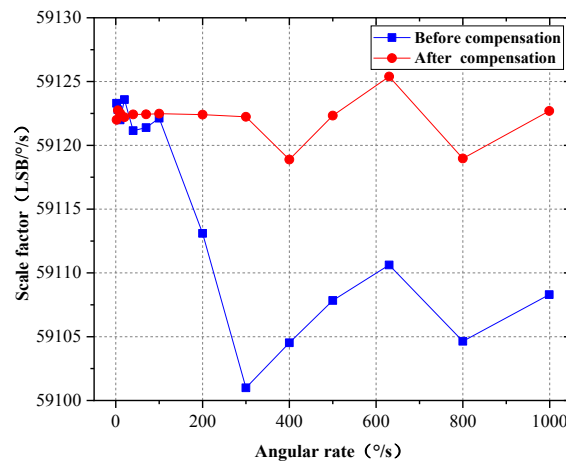
Fig.7   The comparison of the tests results before and after the compensation scheme

After physical verification, the measure of automatically adjusting the gyro output value by software algorithm can effectively reduce the scale factor nonlinearity of the fiber optic gyro based on SLD light source. Since the scale factor compensation scheme is based on a one-time fitting function development and design, the compensation effect is related to the linearity of the gyro output value and input angular rate.

References:

[1] Zhang Guicai. Fiber optic gyro principle and technology [M]. Beijing: National Defense Industry Press, 2008.

[2] Wang Wei. Interferometric fiber optic gyroscope technology [M]. Beijing: China Aerospace Press, 2010.

[3] Zhang Guicai, Yang Ye. Principles and Technology of Fiber Optic Gyro [M]. Beijing: National Defense Industry Press, 2020.

[4] Wu Xunfeng. Research on nonlinear error compensation method of fiber optic gyro based on curve fitting [D]. Harbin: Harbin Engineering University, 2014.

[5] Chen H.Y., Zheng Y., Wang X.X., et al. Influence of average wavelength on the scale factor of fiber gyro[J]. China Laser, 2019(3).

Technological Advances in Fiber Optic Gyroscopes for Space Missions

Ma Fengyuan, Hong Wei, Hu Xudong, Zang Zerun, Zhang Mianzhi, Zhao Zeheng

(Xi'an Aerospace Precision Electromechanical Institute, Xi'an 710100, China)

Abstract: Interferometric fiber optic gyroscopes, renowned for their exceptional precision and stability, are increasingly utilized in space applications. This paper provides an overview of the current status of fiber optic gyroscopes in space applications both domestically and internationally, with a special focus on several models extensively employed in the space sector. It then discusses the technological advancements in fiber optic gyroscopes specific to space missions, particularly emphasizing recent developments in critical technical areas such as radiation resistance, miniaturization and integration.

Key words: Fiber Optic Gyroscope; Space Applications; Radiation Resistance; Miniaturization and Integration;

1　Introduction

The Interferometric Fiber Optic Gyroscope (IFOG), an inertial sensor based on the Sagnac effect, has become the device of choice for inertial navigation and attitude-orbit control systems in spacecraft. It features a solid-state design, low energy consumption, high dynamic performance, and longevity, making it exceptionally suitable for complex space missions[00]. With the increasing global emphasis on strategic space deployment and rapid advancements in technologies related to low Earth orbit, miniaturization, and high-throughput satellites, the performance requirements for space-based fiber optic gyroscopes are becoming more stringent. The trend toward miniaturization in satellite systems, favoring low-power, lightweight, and cost-effective standalone devices, is especially pronounced in the domain of microsatellites, where resource limitations are acute. Therefore, the development of miniaturization technologies for fiber optic gyroscopes is critical, not only enhancing their competitiveness in applications requiring mid-to-low precision but also meeting the stringent demands for space resource efficiency and cost-effectiveness.

In satellite applications, fiber optic gyroscopes play a pivotal role by providing high-precision angular velocity measurements that significantly support and enhance satellite attitude control and positioning accuracy, which are crucial for tasks such as Earth observation, communication, and scientific experiments. The high dynamic performance of these gyroscopes also facilitates real-time, accurate navigation information during satellite orbital maneuvers or docking operations, ensuring the efficiency and safety of these processes. Thus, fiber optic gyroscopes have become an indispensable component within modern satellite technology frameworks and are expected to continue playing a central role in future space exploration and technological applications[0].

2　Fiber Optic Gyroscopes in Space Applications

2.1　Development Abroad

FOGs based on their performance metrics, are utilized in various space applications. Currently, leading fiber optic gyroscope manufacturing countries including the USA, France, Japan, and Russia, have extensively applied FOG technology across diverse space systems. These technologies have been successfully integrated into the attitude control systems of rockets, spacecraft, satellites, and space telescopes.

Since their debut, significant progress has been made in the field of fiber optic gyroscopes for space applications globally. Table 1 provides a comprehensive list of the core performance parameters of major fiber optic gyroscope products used in space applications by several international companies, reflecting the efforts of different manufacturers to enhance precision and stability.

Tab.1 Parameters of Fiber Optic Gyroscopes for Space Applications by Foreign Companies

Fiber Optic Gyroscope Manufacturers	Key Products	Angular Random Walk（$°/\sqrt{h}$）	Bias Stability（$°/h$）	Scale Factor Error（ppm）
Northrop Grumman	LN-200S	0.07	0.1	100
Honeywell	Spirit IMU	0.001	0.01	110
Exail	Astrix 200	0.0001	0.0005	10
Advanced Navigation	BoreasX90	---	0.001	100
Optlink	VOBIS	0.005	0.03	---
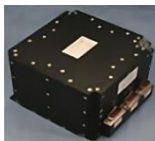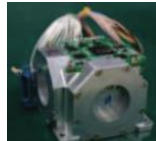
2.2 International Research Status

In recent years, China has made significant advancements in the research and development of fiber optic gyroscope (FOG) technology, with many research achievements successfully applied in the aerospace field. In 2011, China launched its first space laboratory, Tiangong-1, marking the country's initial use of fiber optic gyroscopes for attitude measurement and control of a spacecraft. Since then, FOG technology in space applications has rapidly developed domestically[0].Since then,the Shenzhou series spacecraft and Chang'e series spacecraft have all used FOGs to measure attitude.

The Institute of Aerospace Science and Technology No. 13 has extensively applied its developed fiber optic gyroscopes in various Chinese satellites, with more than 120 sets in orbit and accumulating over 9 million hours-axis of powered operation[0]. All products have operated stably in orbit without any failures, and during the "Twelfth Five-Year Plan" period, the zero bias stability of the delivered products reached $0.0003°/h$（1σ）.

Table 2 lists typical space-grade fiber optic gyroscopes developed by Beijing University of Aeronautics and Astronautics, tailored for different space mission requirements along with their main performance parameters. These series of products, known for their miniaturization, high reliability, and technical maturity, are operating smoothly on more than 60 in-orbit satellites, including Jilin-1, Taijing-2, and Zhuhai-1[0].

Tab.2   Typical Fiber Optic Gyroscope Products from the Institute of Optoelectronics at Beijing University of Aeronautics and Astronautics

Products	Miniaturized Four-Axis Dual Light Source Fiber Optic Gyroscope	Miniaturized Three-Axis Fiber Optic Gyroscope	Time-Division Multiplexed Three-Axis Fiber Optic Gyroscope
Structural Diagram			
Bias Stability / （°/ h ）	0.02	0.05	0.3
Random Walk / （°/$\sqrt{h}$ ）	0.005	0.02	0.1
Scale Factor / （ppm）	50	100	200
Weight /(kg)	1.7	0.4	0.18

3 Key Technical Solutions

3.1 Radiation Resistance Technology

During on-orbit operations, satellites are subjected to harsh space radiation, which adversely affects the performance of photonic components in IFOGs, significantly reducing overall system performance. Consequently, studying the radiation resistance of fiber optic gyroscopes is crucial.

In radiation environments, particularly under the influence of high-energy particles such as protons and heavy ions, color centers form and changes occur in the silica matrix structure within the fiber. These microscopic changes directly impact the optical properties of the fiber. Exail has conducted extensive research on these issues, suggesting in the literature that radiation-induced attenuation and photobleaching are two primary responses of optical fibers under radiation. These findings are significant for understanding and improving the design and application of fiber optic gyroscopes in extreme environments[0].

One method to enhance radiation resistance involves altering the composition of the fiber. Russian scientists have developed fluorine-doped fibers, which experiments have shown to possess certain radiation-resistant properties[0]. It was found that reducing the drawing tension or speed during fiber production can significantly decrease radiation-induced attenuation in these radiation-resistant fibers made from fluorine-doped silica glass. However, this type of fiber has limitations; on one hand, it may alter the mechanical strength and durability of the fiber, which is not conducive to long-term operation in space. On the other hand, experiments have shown that fluorine doping might affect the optical performance of the fiber, necessitating precise control over the amount of fluorine added.

Compared to traditional polarization-maintaining fibers, Polarization Maintaining Photonic Crystal Fibers (PM-PCF) demonstrate significant radiation resistance in high-radiation environments due to their unique microstructure design. This design optimizes the fiber's microstructure, such as air holes and the fiber core, not only enhancing its polarization-maintaining

capabilities but also significantly reducing radiation-induced optical attenuation, improving the fiber's overall optical performance and environmental stability[0]. On December 22, 2020, the 3S high-precision photonic crystal fiber gyroscope module, equipped by Beijing University of Aeronautics and Astronautics, underwent its inaugural flight test on the New Technology Verification Satellite No.7, successfully implementing high-precision satellite attitude measurement and control. Attitude data indicated that the satellite's three-axis attitude stability achieved using the 3S high-precision photonic crystal fiber gyroscope module ranged from 0.000046 to 0.000055°/s （3σ）[0]. This achievement thoroughly validates the significant potential of high-precision photonic crystal fiber gyroscopes in satellite attitude control applications.

PM-PCF have demonstrated excellent environmental adaptability in short-term applications, their high manufacturing costs limit their widespread deployment in cost-sensitive areas. Future research should focus on developing more cost-effective manufacturing techniques or exploring more economical alternative materials. Additionally, while PM-PCF has shown promising initial performance, further research is needed to understand its long-term stability in extreme environments.
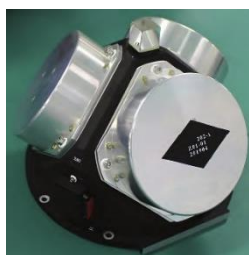
(a) 3S inertial core unit (b) PCF coil

Fig.1 High-Precision Photonic Crystal Fiber Gyroscope Components

3.2 Miniaturization and Integration Technologies

In the context of high reliability and longevity requirements, the design of a dual light source four-axis integrated fiber optic gyroscope represents a typical and efficient configuration. However, the redundancy of the sensitive axis components results in the fiber optic gyroscope modules based on this design being notably heavier and more power-consuming. With the focus on applications for low-orbit micro and nano satellites in recent years, there are increasingly stringent demands for the lightweight and low power consumption of fiber optic gyroscopes. This trend underscores the need for more optimized design solutions to meet the growing challenges of space missions.

To address the issues mentioned above, Figure 2(a) presents a lightweight three-axis integrated fiber optic gyroscope configuration, where the orthogonally arranged three-axis gyroscopes are connected to the light source module via a coupler. Moreover, the adoption of time-division multiplexing technology further enables the miniaturization of fiber optic gyroscopes. As shown in Figure 2(b), the design based on time-division multiplexing technology represents the most simplified configuration for a three-axis integrated fiber optic gyroscope. This approach further simplifies the structural composition of the fiber optic gyroscope from that in Figure 2(a) by using a single coupler and photodetector, reducing the complexity of the optical path and the number of fiber splicing points, and employing a single set of signal processing circuitry, significantly reducing

the mass and power consumption of the gyroscope module[0].

(a) Lightweight Miniature Three-Axis Integrated Fiber Optic Gyroscope

(b) Time-Division Multiplexed Three-Axis Integrated Fiber Optic Gyroscope
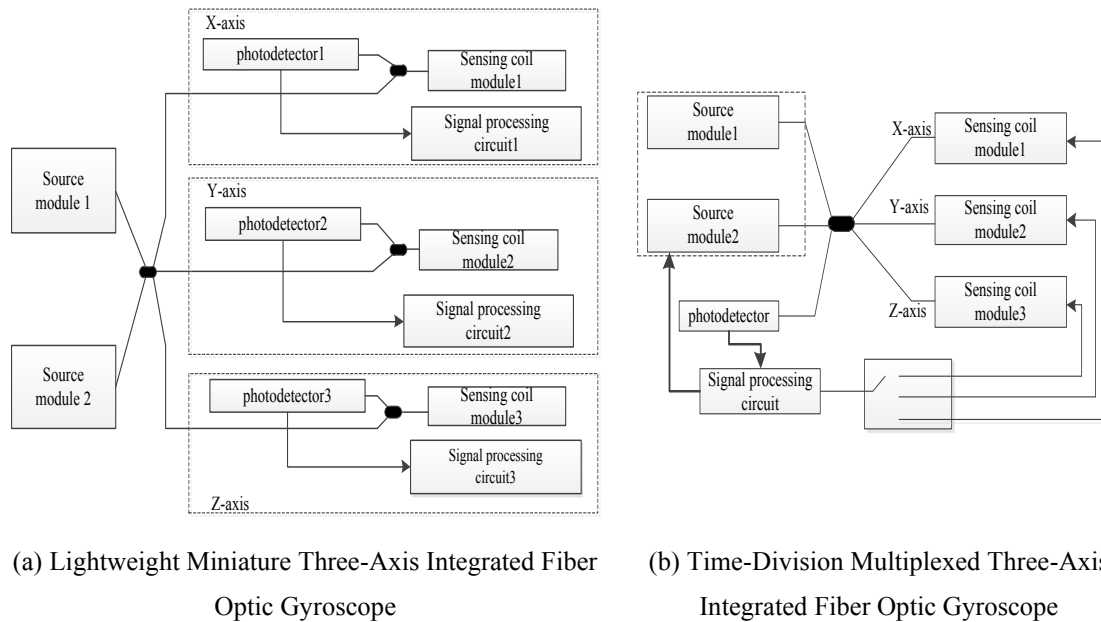
Fig. 2   Miniaturized Fiber Optic Gyroscope Structural Diagram

In addition to these innovations, the miniaturization of optoelectronic devices plays a crucial role in reducing the size and weight of fiber optic gyroscopes. By employing thin fibers, miniaturized integrated optical circuits, integrated PIN-junction field-effect transistor photoreceiver assemblies, and optimized compact digital processing circuits, the volume and weight of fiber optic gyroscopes can be significantly reduced[0]. Future research could further explore the application of cutting-edge materials, such as graphene and other two-dimensional materials, which have attracted widespread attention due to their outstanding electronic and optical properties. The potential of these materials in devices such as photodetectors, modulators, and lasers is being actively investigated[0]. Additionally, the combination of carbon dots with certain wide-bandgap materials has shown excellent luminescent properties, opening new possibilities for the manufacture of efficient optoelectronic devices[0]. However, these highly integrated and miniaturized solutions often come with higher costs, making the study of more economical and efficient production methods to reduce the manufacturing and integration costs of fiber optic gyroscopes a key research direction for the future.

4 Conclusion

As a Sagnac-based inertial sensor, the FOG has been widly adopted in space applications due to its unique advantages .Both domestic and international FOG technologies have reached a high level of maturity in this field, resulting in the development of several singnificant model series. Additionally, significant achievements have been made in related research both at international and abroad, including enhancements in radiation resistance, breakthroughs in miniaturization and integration technologies. However, cost control, spatial complexity, and sustained stability in extreme environments remain focal points for future research. As key technologies continue to advance, the application scenarios for fiber optic gyroscopes are expected to expand further.

References:

[1] Wang Wei. Apolication of GNSS real-time satellites clock offset estimation in seismic monitoring[J]. Navigation Positioning and Timing, 2020, 19(Z1): 18-28.

[2] Xue Lianli, DAI Min, GE Yuetao, et al. 2018 Review and Development of Foreign Inertial Technology[J]. Cruise Missile, 2019, (04):16-21 .

[3] Jin Jing, LI Tiezhi, WANG Xiaowei, et al. Fiber Optic Gyroscope Space Application Technology[J]. Chinese Journal of Laser, 2022, 49(19): 255-262.

[4] Gao Zhongxing. Research on the Mechanism of Environmental Errors in Fiber Optic Gyroscopes and Methods of Suppression[D]. Harbin. Harbin Engineering University, 2017 .

[5] Paveau, A. et al. Robustness of Astrix Fiber Optic Gyros in space radiative environment[J]. CEAS Space Journal 2019, (11): 219-227.

[6] Egorov, D. A. et al. Ensuring Radiation Resistance of Fiber Optic Gyroscopes and Ways to Improve It[J]. Gyroscopy and Navigation 2018, (9): 314 - 324.

[7] Zhang Chuxi, Zhang Zuchen, Gao Fuyu. Photonic Crystal Fiber Gyroscope Technology [J]. Acta Optica Sinica , 2022, 42(17):121-126 .

[8] Xu Xiaobin, Wang Xiaoyang, Gao Fuyu, et al. Photonic Crystal Fiber Gyroscope Technology and Its First Space Experiment[J]. Chinese Journal of Inertial Technology, 2021, 29(01): 1-7.

[9] Yen-Chieh Wang, Sin-Yun Lu, Tzu-Hsiang Yen, et al. Silicon Photonics Multi-Function Integrated Optical Circuit for Miniaturized Fiber Optic Gyroscope[J]. Journal of Lightwave Technology. 2023, 41(19): 6324-6332.

[10] Akinwande D, Huyghebaert C, Wang CH, Serna MI, et al. Graphene and two-dimensional materials for silicon technology[J]. Nature. 2019, 573(7775) :507-518.

[11] Zhao M, Hao Y, Zhang C, Zhai R, Liu B, et al. Advances in Two-Dimensional Materials for Optoelectronics Applications[J]. Crystals. 2022; 12(8):1087.

Influence of Curing on Stability of Fiber Coil and Optimization of Adhesive Parameters

Hu Xudong[1], Zang Zerun[1], Liu Pan[2], Wang Yihua[1], Liang Kaixu[1], Yang Xiong[1]

(1. Xi'an Aerospace Precision Mechatronics Institute Xi'an 710100, China; 2. Beijing Institute of Control and Electronic Technology, Beijing, 100038)

Abstract: The application of multipolar symmetrical winding has significantly reduced the Shupe effect of FOG. However, the residual stress generated by high-temperature curing and the thermal stress caused by temperature changes have become new major factors affecting temperature performance of FOG. The influence of adhesive on the fiber coil under high-temperature curing is studied by finite element simulation. A method for improving the temperature performance of fiber coils with matched adhesive parameters is proposed and experiments are conducted. The results indicate that optimizing adhesive parameters can be a new approach to improve the temperature performance of fiber coils.

 Key words: fiber coil; adhesive; curing; parameter optimization

1 Introduction

The Shupe effect [1] and the photoelastic effect [2] are one of the main factors affecting the stability of FOGs. Due to the mismatch in thermal expansion coefficient between the fiber core and the cladding and adhesive, stress will be generated on the core during the curing process of fiber coils. In this paper, the influence of curing on the fiber coil is simulated by finite element method according to the formula for calculating the angular rate error induced by heat of fiber coil, and the technical idea of improving the temperature performance of the fiber coil by optimizing the parameters of the coating adhesive is proposed.

2 Curing mechanism of fiber coil and drift caused by temperature

2.1 Curing mechanism of adhesive

Generally, curing is carried out after the fiber coil is wound; Fig. 1 describes the five stages experienced by resin materials in a typical curing process [3].

(1) Flow stage: At this stage, the resin is in a liquid state before the Gelation, the curing shrinkage and thermal expansion of the resin do not contribute to the curing deformation and residual stress of fiber.

(2) Gel-glass transition stage: At this stage, the temperature is greater than the Tg (glass transition temperature) of resin. The resin undergoes curing shrinkage. But due to its low modulus, the curing shrinkage strain and thermal expansion do not cause obvious residual stress [4].

(3) Temperature holding stage: Since the degree of curing of the resin does not increase much at this stage, and it has little impact on curing deformation and residual stress.

(4) Cooling stage: The resin curing reaction ends before the cooling starts. Due to the mismatch in thermal expansion coefficients between the resin and the fiber coating and the fiber core, the curing deformation and residual stress formation are mainly concentrated at this stage.
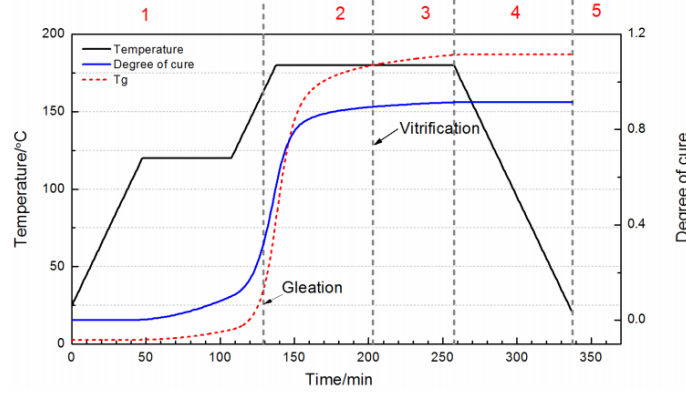
Fig.1   Curing process of resin material

If ignoring the influence of the curing chemical reaction of the adhesive on the fiber coil, the stress effect generated by curing shrinkage is the same as the stress effect generated by the thermal expansion and contraction of the adhesive. In the subsequent simulation analysis, the stress effect of curing shrinkage on the fiber coil can be comprehensively considered by adjusting the thermal expansion coefficient of the adhesive.

2.2   Theory of gyro zero bias change caused by heat and stress

The reason is that the changing temperature gradient inside the fiber coil and the thermally induced stress generated by temperature change the length and refractive index of the optical fiber [5]. According to our previous analysis, the influence of the Shupe effect and photoelastic effect on the bias can be expressed as the following formula [6]:

$$\Omega_E(t) = \frac{n}{DL} \int_0^L \left((n\alpha + \frac{\partial n}{\partial T}) \dot{T}(l,t) + \left(\frac{\partial n}{\partial P_x} + n\frac{\partial l}{\partial P_x} \right) \dot{P}_x(l,t) + \left(\frac{\partial n}{\partial P_y} + n\frac{\partial l}{\partial P_y} \right) \dot{P}_y(l,t) + \left(\frac{\partial n}{\partial P_z} + n\frac{\partial l}{\partial P_z} \right) \dot{P}_z(l,t) \right) (L - 2l) dl \quad (1)$$

Where: n is the refractive index of the optical fiber. $\dot{P}_x$, $\dot{P}_y$, and $\dot{P}_z$ are the stress change rates in the fast axis, slow axis, and length direction of the optical fiber respectively. $\partial l / \partial P_x$, $\partial l / \partial P_y$, and $\partial l / \partial P_z$ are the influence coefficients of stress in three directions on the length of the optical fiber respectively. $\partial n / \partial P_x$, $\partial n / \partial P_y$, and $\partial n / \partial P_z$ are the influence coefficients of stress in three directions on the refractive index of the optical fiber respectively. $\partial n / \partial T$ is the influence coefficient of temperature on the refractive index.

To simplify the calculation, it is assumed that the temperature of each layer of optical fiber on the fiber coil is equal, and the three-dimensional model of the fiber coil is simplified to a two-dimensional model. The total length L is divided into M layers, and each layer has N turns. Therefore, the total number of turns of the fiber coil is MN, as shown in Fig. 2. Formula (1) can be discretized into formula (2):

$$\Omega_E(t) = \frac{n}{DL} \sum_{i=1}^{MN} \left((n\alpha + \frac{\partial n}{\partial T}) \dot{T}(l_i,t) + \left(\frac{\partial n}{\partial P_x} + n\frac{\partial l}{\partial P_x} \right) \dot{P}_x(l_i,t) + \left(\frac{\partial n}{\partial P_y} + n\frac{\partial l}{\partial P_y} \right) \dot{P}_y(l_i,t) + \left(\frac{\partial n}{\partial P_z} + n\frac{\partial l}{\partial P_z} \right) \dot{P}_z(l_i,t) \right) (L - 2l_i) dl_i \quad (2)$$

Where: l_i is the starting coordinate of the i-th turn of optical fiber, dl_i is the length of the i-th turn of optical fiber, and $\dot{T}(l_i,t)$ $\dot{P}_x(l_i,t)$ $\dot{P}_y(l_i,t)$ $\dot{P}_z(l_i,t)$ are the temperature change rate and stress change rates in three directions of the i-th turn of optical fiber respectively.
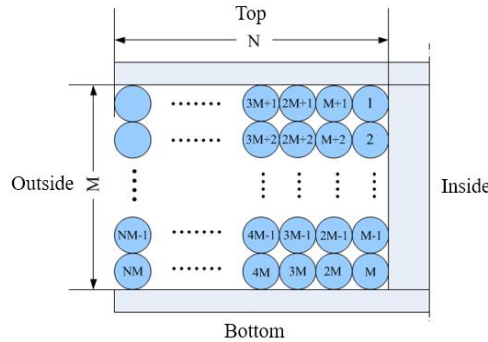
Fig.2 Schematic diagram of the discretized model of the fiber coil

3 Simulation and analysis

Using the size in Tab. 1 and the material parameters in Tab. 2, a finite element model of fiber coil was established as shown in Fig. 3(a). In the simulation, it is assumed that the temperature and stress of each turn of fiber are the same; the lower side of the fiber coil is set as rigid connection; and uniform temperature excitation is applied around the fiber coil.

Tab.1 Size of the fiber coil model for simulation

Parameter	value
Inner diameter /m	0.113
Number of layers	32
Number of turns per layer	20
Fiber diameter /mm	0.135
Winding method	Sixteen-pole symmetry

Tab.2 Material parameters of the fiber coil model for simulation

Parameter	Fiber core	Cladding	Adhesive
Density kg/m^3	2203	1990	970
Specific heat J/(kg·K)	703	1400	1600
Thermal conductivity coefficient [W/(K·m^2)]	1.38	0.21	0.21
Coefficient of thermal expansion K^{-1}	0.55×10^{-6}	67.1×10^{-6}	110×10^{-6}
Young's modulus Pa	80×10^9	0.156×10^9	1.434×10^9
Poisson's ratio.	0.17	0.4	0.499
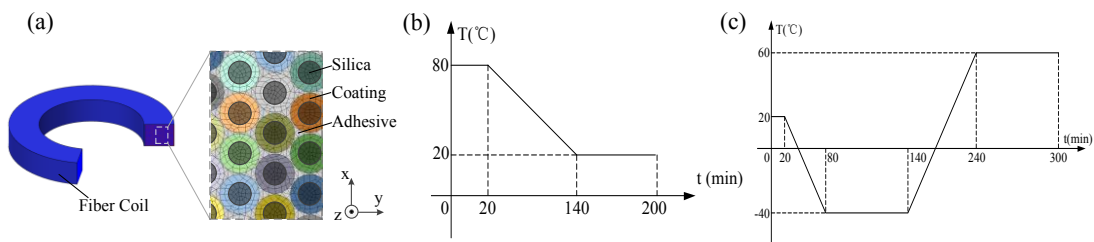
Fig.3 Fiber coil model (a) and temperature curve for simulation. (b) Curing curve; (b) bias test curve

125

Taking Fig. 3(b) as the temperature curve for simulation, the reference temperature of the adhesive is 80°C, the reference temperature of the optical fiber core and cladding is still set at normal temperature of 20°C.

Fig. 4 shows the simulation result of the stress distribution at the stage of completion of curing (Fig. 3(b), 200 min). It can be seen that when curing is completed and the temperature drops to 20°C, the fiber is affected by the stress generated by the contraction of the adhesive.
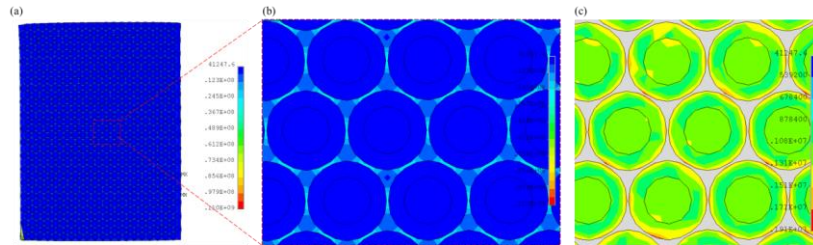
Fig.4　Stress distribution of the fiber coil at the end of curing (Fig. 3(b), 200 min)

Taking the stress of the fiber coil after curing is completed as the initial stress state and taking Fig. 3(c) as the temperature condition, the simulation of the fiber coil is carried out.

Fig. 5 shows the stress on each turn of the fiber coil at 20°C (Fig. 3(c), 18 min), -40°C (Fig. 3(c), 120 min), and 60°C (Fig. 3(c), 300 min). It can be seen that the core is subjected to a stress of about 7×10^5 Pa due to the curing stress at 20℃ , and because the fiber coil is cured at 80℃, the residual stress is kept at a high temperature, so that the stress on the fiber core is greater at -40°C, and the stress at 60°C is lower than 20℃ . In contract, if there is on curing stress, 20℃ is in an unstressed state, and the stress values at -40°C and 60°C are close.
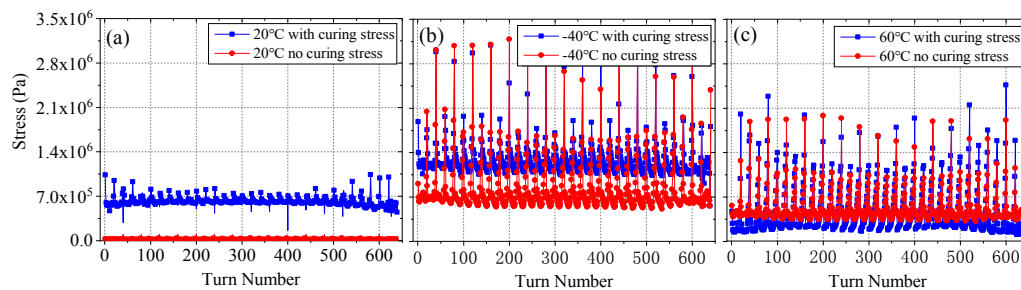
Fig.5　Stress magnitudes of each turn of optical fiber inside the fiber coil at 20°C (a), -40°C (b), and 60°C (c) when considering and not considering curing stress

Substitute the temperature and stress as boundary conditions into formula (2). The bias drift can be calculated as shown in Fig. 7. It can be seen that when there is curing stress, because of the difference in thermal expansion coefficient between the optical fiber and the adhesive, the bias of the simulation model is greater in the high and low temperature.

Therefore, different thermal expansion coefficients of the adhesive are set to simulate on the fiber coil model. The scheme is shown in Tab. 3.

The stress on each turn of the fiber coil at -40°C (Fig. 4(b), 120 min), and 60°C (Fig. 4(b), 300 min) are shown in Fig. 6. The calculation results of bias are shown in Fig. 7.

Tab.3   Simulation scheme for thermal expansion coefficient of adhesive.

Scheme	Thermal expansion coefficient of adhesive (K^{-1})	Numerical basis
a	110×10^{-6}	true value
b	220×10^{-6}	true value×2
c	55×10^{-6}	true value÷2
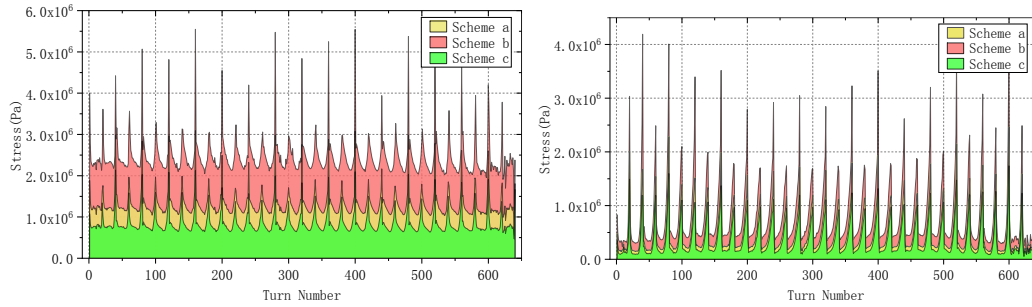
Fig.6   Simulation results of stress of each turn of optical fiber inside the fiber coil at -40°C(a) and 60°C(b) for different thermal expansion coefficient schemes
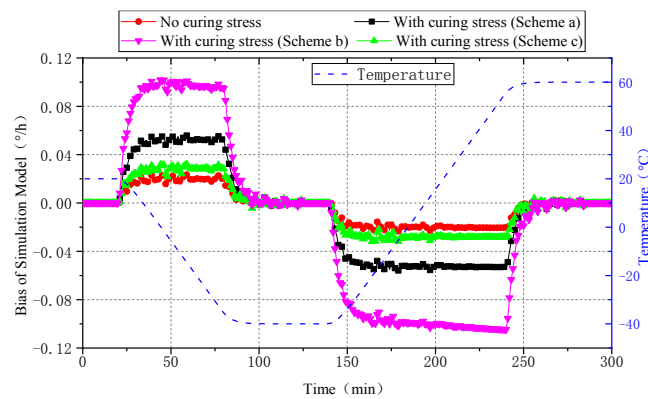
Fig.7   Simulation results of bias drift with different thermal expansion coefficient

The change of the thermal expansion coefficient of the adhesive affects the stress of the fiber core, and the change is particularly significant at -40°C. Combined with Fig. 7, it can be seen that when the thermal expansion coefficient is doubled (scheme b), the stress on the fiber core increases, and its bias also increases accordingly; on the contrary, when the thermal expansion coefficient is halved (scheme c), the stress and drift on the fiber core decrease. And the thermal expansion coefficient of the coating is close to the adhesive in scheme c. It can be known that if the thermal expansion coefficient of the adhesive with the best performance can be found to match the optical fiber, the bias of the fiber coil can be effectively improved.

4 Experiment

Two types of adhesives (thermal expansion coefficient of 231×10^{-6} (adhesive A) and 136×10^{-6} (adhesive B)) are used to wind two fiber coils for performance testing and verification.
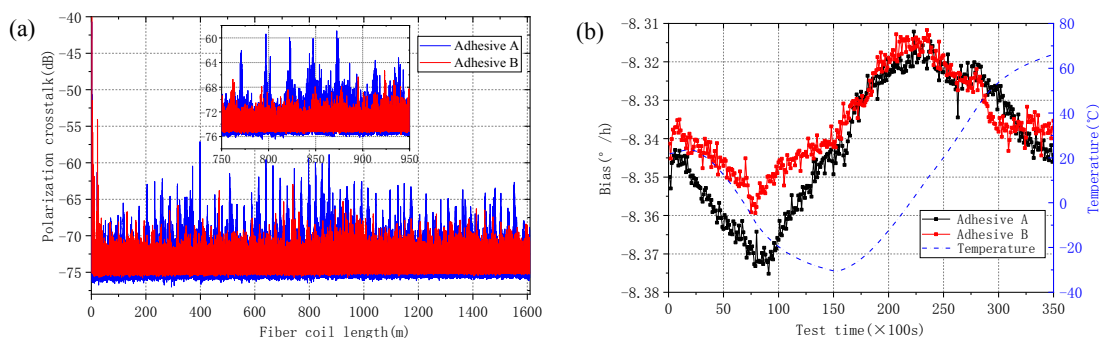
Fig.8 Polarization crosstalk test (a) and bias test (b) of fiber coils using different adhesives

The polarization crosstalk of two fiber coils is test (Fig. 8(a)). It can be seen that when thermal expansion coefficient is small, the noise floor of the polarization crosstalk is flatter, and it is also lower at the layer change position (the spikes in the figure).

A laboratory system is built to test the full-temperature performance of the two coils. The results are shown in Fig. 8(b) and Tab. 4. It can be seen that the bias stability is improved to 0.0119 when the thermal expansion coefficient is small, with an improvement of about 41%. The bias range is improved to 0.0257, with an improvement of about 16%.

Tab.4 Stability and range of fiber coils using different adhesives

Fiber coil adhesive model	bias stability (100s) (°/h)	bias range (100s) (°/h)
Adhesives A	0.0168	0.0298
Adhesives B	0.0119	0.0257

5 Discussion

The influence of the residual stress of the adhesive generated by curing on the fiber coil is analyzed through finite element simulation. And through parametric optimization and simulation of the thermal expansion coefficient of the adhesive, the effectiveness of improving the temperature performance of the FOG is proved.

However, the fiber coil involves different materials, and different parameters of each material will affect the stress distribution of the fiber coil. In the future, analysis and research can be carried out in aspects such as other parameters of the adhesive, fiber cladding parameters, and coupling of different parameters to study the influence of fiber coil curing and material parameters on its performance in depth.

References:

[1] D. M. Shupe, Thermally induced nonreciprocity in the fiber-optic interferometer, Appl. Opt. 19, 654–655 (1980).

[2] H. C. Lefèvre, The Fiber-Optic Gyroscope, 3rd ed. (Artech House, 2022), Chap. 6.

[3] Ding Anxin. Numerical simulation and theoretical research on curing deformation of thermosetting resin matrix composites [D]; Wuhan University of Technology. 2016

[4] Wisnom M R, Potter K D, Ersoy N. Shear-lag Analysis of the Effect of Thickness on Spring-in of Curved Composites [J]. Journal of Composite Materials, 2007, 41(11): 1311-24.

[5] Z. Zhang and F. Yu, Quantitative analysis for the effect of the thermal physical property parameter of adhesive on the thermal performance of the quadrupolar fiber coil, Opt. Express 25, 30513–30525 (2017)

[6] Wei Hong, Xudong Hu, Accurate measurement and enhancement of fiber coil symmetry. Applied Optics, 62,

Optimizing the Radio-Frequency Field Amplitude When Monitoring the Vapor Cell Parameters in Different Quantum Sensors

M.A. Vlasova[1], A.N. Shevchenko[1], S.V. Ermak[2]

(1.Concern CSRI Elektropribor, JSC, St. Petersburg, Russia; 2. St. Petersburg Polytechnic University, St. Petersburg, Russia)

Abstract: The paper studies the dependence of the resonance curve of an optically pumped quantum magnetometer on the amplitude of the radio-frequency (RF) field exciting the precession of magnetic moments. Numerical modeling of the resonance slope vs. RF field amplitude has been performed for the paraffin-coated and buffer vapor cells. Modeling has been experimentally verified on a cell with high buffer gas pressure.

1 Introduction

A quantum sensor is a high-accuracy measurement device using the quantum effects to detect and measure different physical magnitudes: magnetic field intensity, angular velocity, time, frequency. The key component of the quantum sensors based on the effect of double radio-optical resonance is a vapor cell [1-3] containing the vapors of alkali metals used to spin polarize the atoms by optical pumping. The quality of manufacturing and filling the vapor cell mainly determines the accuracy and sensitivity of the quantum sensor [2, 4]. This quality can be estimated by analyzing the magnetic resonance line [5], the width of which can be used to determine the alkali gas relaxation rate, which directly affects the sensor sensitivity. In the cells with different mechanisms of reducing the alkali metal relaxation rate (with antirelaxation coating [6] and buffer gas [7]) the characteristic relaxation rates may differ by two to three orders of magnitudes, whereas the technological tolerance for one type cells does not exceed 10%. In practice, the width of magnetic resonance line depends not only on the parameters of the measured vapor cell but also on the measurement conditions [8-10], the most important being the amplitude of the alternating RF magnetic field, which excites the precession of magnetic moments. In experimental samples of quantum magnetometers, the RF field amplitude is optimized for the specific cell, however, in serial production it is too labor-consuming. This paper focuses on the optimization of RF field amplitude in characterization of vapor cells of different types and of one type.

2 Optimization of RF field

The magnetic resonance line is observed using a setup consisting of an M_x-magnetometer on the tested cell with a magnetic shield and a coil system providing the constant bias field [11]. M_x-signal determining the sensitivity of the quantum M_x-magnetometer (minimum detectable variation of the magnetic field) can be presented as [12]:

$$\Delta\omega = \omega_{rf} - \omega_{0,} \qquad (1)$$

where M_0 is the longitudinal magnetization caused by the circular component of the pumping light along the magnetic field; γ is the gyromagnetic ratio of the alkali metal; B_1 is the RF field amplitude; $\Delta\omega = \omega_{rf} - \omega_0$ is the mismatch between the frequency ω_{rf} and the resonance frequency ω_0; $\Gamma = \dfrac{1}{T}$ is the alkali metal relaxation rate; $W = \sqrt{\Gamma^2 + (\gamma \cdot B_1)^2}$ is the width of magnetic resonance line.

The M_x-signal is primarily determined by the amplitude of RF field used to excite the resonance transitions between the atomic energy levels. Increasing the RF field amplitude, on the one hand, increases the signal amplitude and thus the sensor sensitivity, on the other hand, widens the resonance line so that it becomes visually unobservable. Therefore, it is needed to select the optimal RF field amplitude ensuring the maximum signal amplitude and acceptable resonance line width.

The optimal amplitude is selected using the resonance slope, which is theoretically calculated as the signal first derivative with respect to frequency mismatch at the point with zero mismatch (i.e., when the RF field frequency equals the Larmor frequency, or at the resonance center) [13]:

$$\mathrm{K} = \left. \frac{dM_x(\omega)}{d\Delta\omega} \right|_{\Delta\omega=0} = M_0 \frac{\gamma \cdot B_1}{\Gamma^2 + \gamma \cdot B_1{}^2} = M_0 \frac{\gamma \cdot B_1}{W^2}, \tag{2}$$

3   Numerical Modeling

We provide the results from numerical modeling of the resonance slope vs. the magnitude $\gamma \cdot B_1$ in Hz with the constant gyromagnetic ratio γ for the volume of alkali metal cesium used within the cell volume. Multiplication by B1 in $\gamma \cdot B_1$ is due to the fact that the width of magnetic resonance lines used to determine the alkali metal relaxation rates are usually calculated in Hz. The modeling results are provided for two types of cells: with antirelaxation coating (characteristic relaxation rate $\Gamma_{char} = 1$ Hz [6]) and with buffer gas ($\Gamma_{char} = 1000$ Hz [7]). Figure 1 presents the plots of resonance slopes calculated by (3) normalized to its maximum vs. $\gamma \cdot B_1$ for two types of cells. It is seen that the acceptable values of $\gamma \cdot B_1$ (the slope is max 80% of the maximum achievable value) are within 0.5–2 Hz for the cells with antirelaxation coating and within 500–2000 Hz for buffer gas cells. For both types, the maximum slope is achieved at $\gamma \cdot B_1 = \Gamma$.
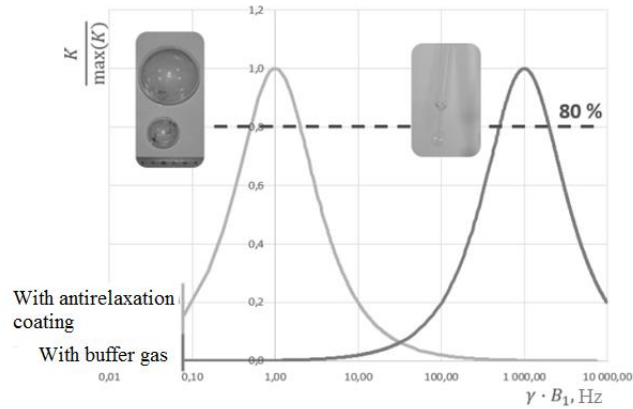
Fig. 1 Modeling the resonance slope for paraffin-coated and buffer gas cells

4 Experiment

To experimentally verify the modeling results we used the buffer gas cell. Different amplitudes of the resonance RF field $\gamma \cdot B_1$ were set with the coils, and for each one the width W and amplitude S were calculated and used further to experimentally find the resonance slope [13]:

$$K = \frac{S}{W}, \tag{3}$$

The resonance slope K vs. $\gamma \cdot B_1$ is shown in Fig.2. For the studied buffer gas cell the resonance slope is maximum at the RF field amplitude $\gamma \cdot B_1 \approx 1000$ Hz, which is close to the characteristic relaxation rate Γ_{char}. The resonance slope decreases by no more than 20% under the amplitudes within 500–1800 Hz.

Therefore, setting the RF field amplitude B_1 so that $\gamma \cdot B_1 \approx 1000$ Hz, we can test the buffer gas cells with the characteristic relaxation rate Γ_{char} within 500–1800 Hz. This range is wider than the technological tolerance of characteristic relaxation rates for the cells of one type, which is max 10%. This allows measuring the cells with the same relaxation rate reduction mechanism without changing the setup parameters.
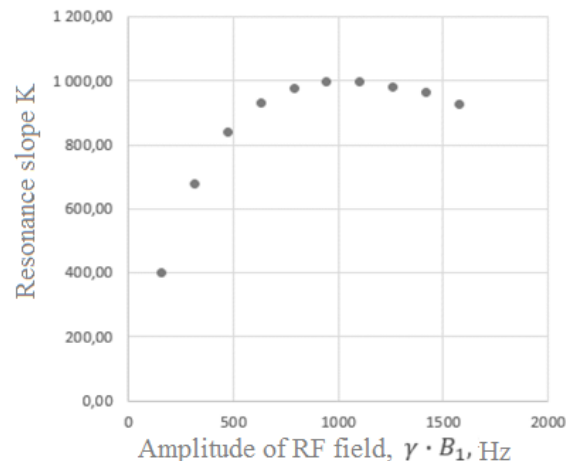
Fig.2 Resonance slope vs. RF field amplitude

5 Conclusions

The paper studies the dependence of sensitivity of a quantum optically pumped magnetometer scheme on the RF field amplitude. Numerical modeling has revealed that different optimal amplitudes provide the maximum resonance slope for different types of cells. For a buffer gas cell, the experimentally determined range of RF field amplitudes ensuring the maximum resonance slope is shown to agree with the numerically obtained range.

References:

[1] Walker T. G., Larsen M. S. Spin-exchange-pumped NMR gyros. Advances in Atomic, Molecular, and Optical physics. 2016. Vol. 65. P. 377.

[2] Vershovskii A.K., Litmanovich Yu.A., Pazgalev A.S., Peshekhonov V.G. Nuclear magnetic resonance gyroscope: The ultimate parameters. Gyroscopy and Navigation. 2018. Vol. 9. No. 3. P. 162.

[3] Knapkiewicz P. Technological assessment of MEMS alkali vapor cells for atomic references. Micromachines. 2018. Vol. 10. No. 1. P. 25.

[4] Vlasova M. A., Shevchenko A. N. Development of alkali vapor cell for miniature nuclear magnetic gyroscope. IOP Conference Series: Materials Science and Engineering. IOP Publishing. 2022. Vol. 1215. No. 1. P. 012013.

[5] Alexandrov E.B., Balabas M.V., Pazgalev A.S., Vershovskii A.K., Yakobson N.N. Double-resonance atomic magnetometers: From gas discharge to laser pumping. Laser Physics. 1996. Vol. 6. No. 2. P. 244-251.

[6] Alexandrov E.B., Balabas M.V., Pasgalev A.S., Vershovskii A.K., Yakobson N.N. Light-induced desorption of alkali-metal atoms from paraffin coating. Physical Review A. 2002. Vol. 66. No. 4. P. 042903.

[7] Budker D., Romalis M. Optical magnetometry. Nature Physics. 2007. Vol. 3. No. 4. P. 227-234.

[8] Appelt S., Ben-Amar Baranga A., Young A. R., and Happer W. Light narrowing of rubidium magnetic-resonance lines in high-pressure optical-pumping cells. Physical Review A. 1999. Vol. 59. No. 3.

[9] Bison G., Wynands R., Weis A. Optimization and performance of an optical cardiomagnetometer. JOSA B. 2005. Vol. 22. No. 1. P. 77-87.

[10] Vlasova M.A., Shevchenko A.N., Dement'ev M.M. Development of thermostabilization systems for quantum sensor vapor cell testing. Almanakh nauchnykh rabot molodykh uchenykh Universiteta ITMO (Almanac of Papers by Young Researchers of ITMO University). 2022. Vol. 2. Pp. 15-18.

[11] Pomerantsev N.M., Ryzhkov V.M., Skrotskii G.V., Fizicheskie osnovy kvantovoi magnitometrii (Physical Foundations of Quantum Magnetometry), Moscow: Nauka, 1972.

[12] Bloch F., Nuclear induction, Physics Review, 1946. Vol. 70. No. 7. Pp. 460-474.

[13] Vershovskii A.K. and Pazgalev A.S. Optimization of the Q factor of the magnetic Mx resonance under optical pump conditions. Technical Physics. 2008. Vol. 53. No. 5. Pp. 646-654.

A Cascade-control Strategy for Whole-angle Hemispherical Resonator Gyroscope

Nan Fangbo[1,2], Wang Gang[2], Du Dan[2], Gao Zhongxing[1,2], Zhang Yonggang[1]

(1. College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China; 2. Qingdao Innovation and Development Center of Harbin Engineering University, Qingdao 266000, China)

Abstract: In this paper, a cascade-control mechanism is proposed for the vibration mode of whole-angle hemispherical resonator gyroscope (WA HRG). The advantages of WA HRG include a wide dynamic range and high reliability. However, the lack of bandwidth in the control system results in the decrease of control gain and the phase lag. The control loop model of WA HRG is examined and a cascade-control strategy is proposed. The simulation results reveal that at 400°/s, the principal vibration wave and quadrature vibration wave decrease by 78.8% and 86.4%, respectively, significantly improving the anti-interference capabilities of WA HRG.

Key words: Hemispherical resonator gyroscope; whole-angle; control circuit; cascade-control.

1 Introduction

Hemispherical resonator gyroscope (HRG) is a kind of inertial device based on Coriolis effect, which is used to obtain carrier angular rate [1]. It is widely used in space exploration, ship navigation and geological exploration, and has shown excellent performance. The working modes of HRG include whole-angle (WA) mode, force-to-balance (FTR) mode and whole-angle-tracking (WAT) mode [2]. The vibration mode of the hemispherical shell resonator (HSR) is an important factor that determines the performance of gyroscope. The damping anisotropy and stiffness anisotropy are the main factors affecting the vibration state of HSR. The researches on accuracy improving focus on the identification and compensation of the errors [3]-[5]. There are residual errors in the process of identification, and the methods are relatively complicated. Therefore, it is necessary to exert control over the gyroscope to maintain the vibration state. The control of WA HRG now relies on proportional-integral (PI) control. However, the control forces struggle to track the external high angular rate because of the control bandwidth constraint, which makes it difficult for the vibration mode to retain its functioning state. In summary, a vibration mode control method that successfully eliminates the control bandwidth limitation and improves the anti-interference performance of WA HRG at high precession rates is required. In this paper, the bandwidth limitation of the control loop of WA HRG is studied. Firstly, the energy control loop and quadrature control loop models are established, and the transfer function is derived. Then, the amplitude-frequency characteristics are

analyzed. Then, a cascade-control method of is proposed, and the control model is analyzed. Then, a cascade-control method of is proposed, and the control model is analyzed. Finally, the simulation experiment of gyroscope control loop is carried out to evaluate the realization effect.

2 Hemispherical Resonator Gyroscope

2.1 Basic Dynamics

The non-ideal HRG motion equations are as follows [6]:

$$
\begin{bmatrix} \ddot{x} \\ \ddot{y} \end{bmatrix} + \mathbf{C}\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} + \mathbf{K}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f_x \\ f_y \end{bmatrix} + \begin{bmatrix} 0 & 4\gamma\,\Omega \\ -4\gamma\,\Omega & 0 \end{bmatrix}\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} \tag{1}
$$

Where

$$
\mathbf{C} = \begin{bmatrix} 2/\tau + \Delta(1/\tau)\cos 4\theta_\tau & \Delta(1/\tau)\sin 4\theta_\tau \\ \Delta(1/\tau)\sin 4\theta_\tau & 2/\tau - \Delta(1/\tau)\cos 4\theta_\tau \end{bmatrix} \tag{2}
$$

$$
\mathbf{K} = \begin{bmatrix} \omega^2 - \omega\Delta\omega\cos 4\theta_\omega & -\omega\Delta\omega\sin 4\theta_\omega \\ -\omega\Delta\omega\sin 4\theta_\omega & \omega^2 + \omega\Delta\omega\cos 4\theta_\omega \end{bmatrix} \tag{3}
$$

x and y are the displacements along $0°$ and $45°$ electrode axis. $\mathbf{C}$ is the damping matrix. $\mathbf{K}$ is the stiffness matrix. f_x and f_y are the control forces along x and y. γ is scale factor. Ω is the precession angular rate. ω is the resonant frequency of HRG. $\Delta\omega$ is anisotropy of stiffness. θ_ω is the primary axis angle of stiffness. τ is time constant. $\Delta(1/\tau)$ is anisotropy of damping. In the elliptic coordinate system, x and y are changed as:

$$
\begin{cases} x = a\cos 2\theta\cos(\omega t + \varphi_0) - q\sin 2\theta\sin(\omega t + \varphi_0) \\ y = a\sin 2\theta\cos(\omega t + \varphi_0) + q\cos 2\theta\sin(\omega t + \varphi_0) \end{cases} \tag{4}
$$

Where a is the principal vibration axis. q is the quadrature vibration axis. θ is the precession angle. φ_0 is the phase of the gyroscope vibration. In order to maintain the motion state of HRG, four parameters need to be controlled. The control equations are as follows [7]:

$$
\begin{cases} \dot{E} = -(2/\tau)E - \Delta(1/\tau)\cos 4(\theta - \theta_\tau)\sqrt{E^2 - Q^2} - \dfrac{\sqrt{E}}{\omega}f_{as} \\[2mm] \dot{Q} = -(2/\tau)Q - \Delta\omega\sin 4(\theta - \theta_\omega)\sqrt{E^2 - Q^2} + \dfrac{\sqrt{E}}{\omega}f_{qc} \\[2mm] \dot{\theta} = -\gamma\Omega + \dfrac{E}{2\sqrt{E^2 - Q^2}}\Delta(1/\tau)\sin 4(\theta - \theta_\tau) + \dfrac{Q}{2\sqrt{E^2 - Q^2}}\Delta\omega\cos 4(\theta - \theta_\omega) - \dfrac{f_{qs}}{2\omega\sqrt{E}} \\[2mm] \dot{\varphi}_0 = -\dfrac{Q}{2\sqrt{E^2 - Q^2}}\Delta(1/\tau)\sin 4(\theta - \theta_\tau) - \dfrac{E}{2\sqrt{E^2 - Q^2}}\Delta\omega\cos 4(\theta - \theta_\omega) + \dfrac{f_{ac}}{2\omega\sqrt{E}} \end{cases} \tag{5}
$$

Where $E = a^2 + q^2$ is the energy of HRG. $Q = 2aq$ is the quadrature component of HRG. f_{as} is the energy control force used to maintain constant vibration mode. f_{ac} is the frequency control force which usually set to 0. f_{qc} is the quadrature control force to suppress quadrature component. f_{qs} is the angle control force.

2.2 Control Circuit Model

135

In the normal working state, PI control is used to maintain the vibration mode stability of HRG. The energy control loop is shown in Fig.1.
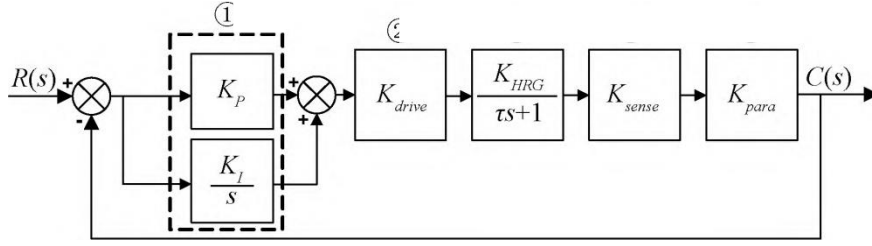
Fig.1   Schematic of PI control loop

Where K_P is the proportional gain. K_I is the integral gain. K_{drive} is drive gain. K_{HRG} is amplitude gain of HRG; K_{sense} is the detection gain. K_{para} is the gain of the elliptic parameter calculation. when $K = K_{HRG}K_{drive}K_{sense}K_{para}$, the close-loop transfer function of the control loop $\Phi_R(s)$ is expressed as:

$$\Phi_R(s) = \frac{G(s)}{1+G(s)} = \frac{K_P K s + K_I K}{\tau s^2 + (K_P K + 1)s + K_I K} \tag{6}$$

The amplitude-frequency characteristic curve of (6) is shown in Fig. 2. It is evident that there are phase lag and decrease in the control force's gain as the external angular rate increases, which is equal to system frequency increase.
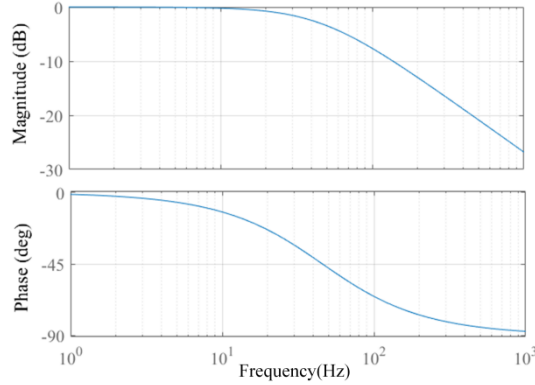
Fig.2   Bode diagram of PI control circuit

3   Elliptical Parameter Cascade Control

3.1   Influence of Anisotropy

The working state of the non-ideal HRG is affected by anisotropy of damping and stiffness. When the gyroscope operates normally, the principal mode amplitude is set to $a\_setting$ and the control interval time is Δt, the energy control loop from (5) is rewritten as:

$$a(\Delta t) = a\_setting = a_0 e^{-\frac{1}{2}[2/\tau + \Delta(1/\tau)\cos 4[\theta(\Delta t) - \theta_\tau]]\Delta t}$$
$$+ \frac{f_{as}(\Delta t)}{\omega}[2/\tau + \Delta(1/\tau)\cos 4[\theta(\Delta t) - \theta_\tau]]^{-1} \tag{7}$$

The energy control of WA HRG is affected by damping anisotropy, which causes the principal

wave amplitude to attenuate at a different rate. The control circuit must also track the mode's precession. However, it is evident from the examination of the amplitude-frequency characteristics of PI control in 2.2 that phase lag and reduction in energy control gain result from an increase in the external input angular rate. As shown in Fig.3, a harmonic component with 4θ is introduced into a. The amplitude of the harmonics will increase as the external speed rises until it tends to a steady value.
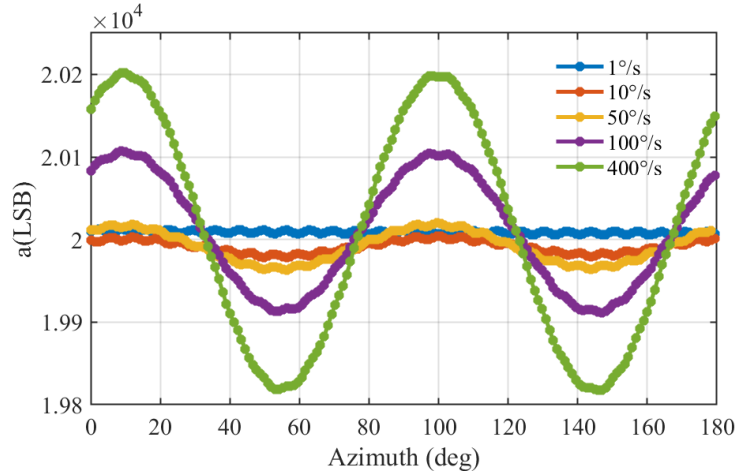
Fig.3 Fluctuation curve of a at different angular rate

Similarly, the quadrature control loop from (5) is rewritten as:

$$
\begin{aligned}
q(\Delta t) = {} & a(\Delta t)\Delta\omega\sin 4[\theta(\Delta t)-\theta_\omega][2/\tau-\Delta(1/\tau)\cos 4[\theta(\Delta t)-\theta_\tau]]^{-1}e^{-\frac{1}{2}[2/\tau-\Delta(1/\tau)\cos 4[\theta(\Delta t)-\theta_\tau]]\Delta t} \\
& + \left[\frac{f_{qc}(\Delta t)}{\omega}-a(\Delta t)\Delta\omega\sin 4[\theta(\Delta t)-\theta_\omega]\right][(2/\tau)-\Delta(1/\tau)\cos 4[\theta(\Delta t)-\theta_\tau]]^{-1}=0
\end{aligned}
$$

(8)

The quadrature control force is applied to suppress the creation of the quadrature wave, as the amplitude of the quadrature wave is influenced by the stiffness anisotropy and the coupling energy of the principal wave as indicated by (7) and (8). However, the control force is influenced by phase lag and gain loss as a result of the control bandwidth limitation, and higher harmonics are injected as an irregular quadrature wave.

3.2 Cascade-Control Strategy

According to the analysis in 3.1, there are variances in the circumferential damping and stiffness of HRG, and the energy control force and the quadrature control force required at different angles is different. Therefore, θ is regarded as dependent value of a and q. A closed-loop cascade-control circuit with the elliptic parameter control as the main loop and vibration mode angle control as the secondary loop is established. The control loop shown in Fig.1 is optimized to the cascade-control loop shown in Fig.4.
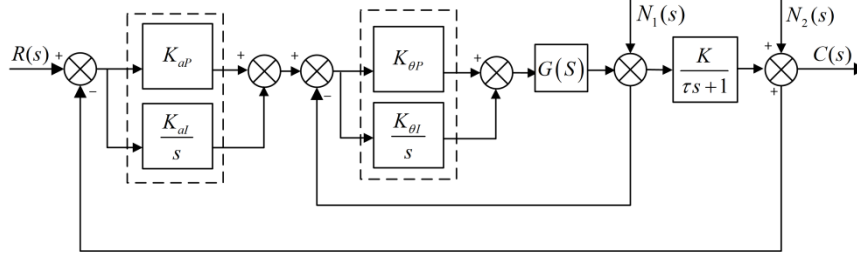
Fig.4 Schematic of cascade-control loop

K_{aP} and K_{aI} are the main loop control parameters. $K_{\theta P}$ and $K_{\theta I}$ are the secondary loop control parameters. $N_1(s)$ is the perturbation term of the secondary loop. $N_2(s)$ is the perturbation term of the main loop. The transfer function in the energy control loop $G_a(s)$ is expressed as:

$$G_a(s) = \frac{A(0)}{s + 1/\tau + \triangle(1/\tau) \cdot \frac{s\,e^{-\theta_r}}{8(s^2 + 16)}} \tag{9}$$

Where $A(0)$ is the set value of a. Similarly, the transfer function $G_q(s)$ in the quadrature control loop is expressed as:

$$G_q(s) = \left[\frac{-a \cdot \triangle\omega \cdot e^{-\theta_\omega s}}{2(s^2 + 16)} + \frac{Q(0)}{s} \right] \tag{10}$$

Because of the existence of the secondary loop, the disturbance of the precession angle to the elliptic parameters is controlled in advance, the influence on the main loop is reduced, and the tracking ability of the control loop to the external input angular rate is improved. At the same time, in the face of channel gain, electrode mismatch and other interference terms that affect the gyroscope vibration mode, it also has a certain adaptive ability.

3.3 Simulation of WA HRG

In order to verify the effectiveness of the cascade-control loop, a simulation platform is built to compare the elliptic parameter fluctuation of PI control and cascade PI (C-PI) control at the same angular rate. The model of WA HRG is established. The elliptic parameter calculated by gyroscope is added to the control loop as input, θ as the interference in the secondary loop, and the damping and stiffness anisotropy between x and y channels as the interference in the main loop. In order to verify the effectiveness of the C-PI control, it is necessary to keep the main loop control parameters of cascade-control consistent with the PI control parameters. An angular rate of 400°/s is applied to the simulation platform. The simulation results are shown in Fig. 5 and Fig. 6. Before and after the cascade-control is applied, the fluctuation of a decreases by 78.8% and the fluctuation of q decreases by 86.4%.
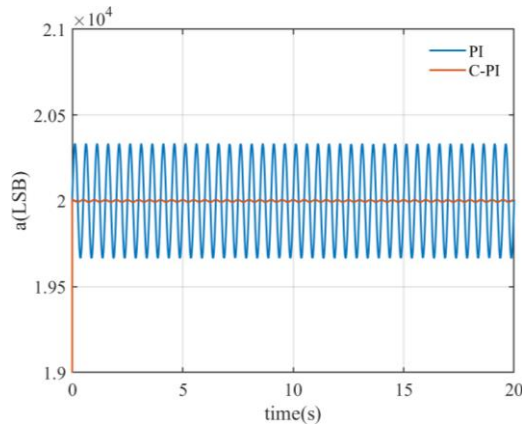
Fig.5 Fluctuation diagram of a at 400°/s   **Fig.6** Fluctuation diagram of q at 400°/s
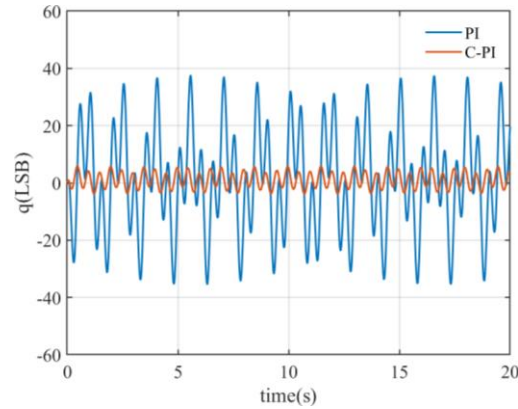
It is found that both the principal wave amplitude a and the quadrature wave amplitude q have higher harmonics related to the theta period under PI control, and the fluctuation of principal wave is a regular fourth harmonic, and quadrature wave is an irregular higher harmonic, which verifies the analysis in 3.1. When the elliptic parameter fluctuation decreases under high precession angular rate conditions, HRG is better able to sustain stable vibration mode.

4 Conclusion

In this paper, the influence of bandwidth limit on energy control loop and quadrature control loop is proposed. The amplitude-frequency analysis shows that the gain of the control loop decreases and the control lag occurs with the increase of angular rate. Through the analysis of the energy control equations, it is found that damping anisotropy and stiffness anisotropy lead to the existence of higher harmonics in the principal wave amplitude and the quadrature wave amplitude respectively. A closed-loop cascade-control circuit with elliptic parameter control as the main loop and angle control as the secondary loop is designed, and its effectiveness is verified by simulation experiments. The simulation results show that the principal wave amplitude is reduced by 78.8% and the quadrature wave amplitude is reduced by 86.4%, which greatly improves the anti-interference ability of WA HRG.

References:

[1] Rozelle D. The Hemispherical Resonator Gyro: From Wineglass to the Planets[J]. 2009(134): 1157-1178.

[2] Anthony M. The operation and mechanization of the hemispherical resonator gyroscope[J]. 2018:7-14.

[3] Yi X, Yang G, Cai Q, et al. A Novel Self-Modulation Method for Whole-Angle Resonator Gyroscope Based on High Harmonic Compensation[J]. IEEE Sensors Journal, 24[2024-11-16].

[4] Nan F, Gao Z, Xu R, et al. Stiffness Tuning of Hemispherical Shell Resonator Based on Electrostatic Force Applied to Discrete Electrodes[J]. IEEE Transactions on Instrumentation and Measurement, 73[2024-11-16].

[5] Xiao D, Yu D, Zhou X, et al. Frequency Tuning of a Disk Resonator Gyroscope via Stiffness Perturbation[J]. IEEE Sensors Journal, 2017 (15):1-1.

[6] Friedland B, Hutton M F. Theory and error analysis of vibrating-member gyroscope[J]. IEEE Transactions on Automatic Control, 1978, 23(4):545-556.

[7] Ruan Z, Ding X, Pu Y, et al. In-Run Automatic Mode-Matching of Whole-Angle Micro-Hemispherical Resonator Gyroscope Based on Standing Wave Self-Precession[J] IEEE Sensors Journal, 2022(22): 13945-13957.

Design and Optimization of Thin-Film Lithium Niobate Electro-Optic Modulator

Yun Jinming, Li Ailun, Mao Yuzheng, Wan Xun, Xie Liangping

(Xi'An Flight Automatic Control Research Institute, Xi'An 710065, China)

Abstract: Fiber optic gyroscope (FOG), as a highly sensitive angular rate sensor, is widely used in inertial navigation systems in aviation, aerospace and other fields. However, the core device of fiber optic gyro, the lithium niobate Y-waveguide, has some problems such as large volume, expensive price and difficult integration, which seriously hinders the further development of fiber optic gyro. Therefore, a high integration and CMOS process-compatible thin film lithium niobate electro-optic modulator was designed. By optimizing the waveguide and electrode structure, the half-wave voltage length product of 3 V•cm and the electro-optic modulation bandwidth of 80 GHz are realized, which can greatly improve the performance of the gyroscope and reduce the production cost.

 Key words: fiber optic gyroscope; Electro-optic modulator; Thin-film lithium niobate; Inertial navigation;

1   Introduction

High-speed electro-optic modulator is the core device of optical sensing system, microwave photonic system and optical communication system, which can realize the conversion of information from electrical domain to optical domain. As an encoder, it changes the refractive index of the material through the applied electric field, so as to regulate[1] the phase and intensity of the transmitted light in the waveguide. Lithium Niobate (LN), as an excellent electro-optic crystal, has a large linear electro-optic coefficient ($\gamma_{33} \approx 30.8$ pm/V); Wide electromagnetic transparent window, showing very low light absorption loss in visible to mid-infrared band ($0.4 \sim 5\mu m$); At the same time, its physical and chemical properties are extremely stabile[2], can be relatively easy to grow large size single crystals, is currently one of the main materials for commercial electro-optical modulators.

The traditional bulk material lithium niobate electro-optical modulator is prepared by titanium diffusion[3] or proton exchange[4] process, and the optical waveguide structure is formed by changing the refractive index of the doped region of the material, which has been widely used in fiber optic gyroscope. It's production method is incompatible with the CMOS process, the device consistency is poor, and it is difficult to integrate with other devices, and the production cost is high, and the key performance improvement suffers a bottleneck, which seriously hinders the further development of fiber optic gyroscope. Thin film lithium niobate prepared by "smart cut" and bonding process provides an effective way to solve the above problems. The optimized dry etching process (Ar^+) can be used to prepare strong limited waveguides with great refractive index difference ($\Delta n \approx 0.8$) on thin films of lithium niobate, which greatly improves the optical field binding ability, greatly reduces the mode spot size (about $1\mu m$), significantly improves the device integration, and effectively reduces the gyro volume. At the same time, the CMOS-compatible

140

production mode can further reduce the cost of the gyroscope and improve product consistency.

At present, the domestic research of on-chip electro-optic modulator with low half-wave voltage, large bandwidth and high linearity for fiber optic gyro is not mature, and the supporting design scheme needs to be further improved[5]. Based on the above problems, the design and optimization simulation of electro-optic modulator on thin film lithium niobate sheet are carried out in this paper. This work can provide guidance for the subsequent processing and preparation of thin film lithium niobate modulator and promote it's engineering application in fiber optic gyro, and promote the fiber optic gyro to the integration direction.

2　Electro-optic modulator principle

2.1　Pockels electro-optical effect

Lithium niobate crystal has unparalleled electro-optical effect, and it's electrooptical effect is mainly first-order linear electro-optical effect (Pockels effect). Electro-optical modulators based on this effect have been widely used in fiber optic gyroscope and optical communication industry. When the applied electric field of lithium niobate crystal is E, it's refractive index change can be expressed as:

$$\Delta \frac{1}{n^2} \equiv \sum_{j=1}^{3} \gamma_{ij} E_j \tag{1}$$

Where E_j is the component of the applied electric field E in all directions, Ej includes Ex, Ey, Ez, γ_{ij} is the electro-optical coefficient of the corresponding direction. At 1550 nm, γ_{13}=8.6 pm/V, γ_{22}=3.4 pm/V, γ_{51}=28 pm/V, γ_{33}=30.8 pm/V[6]. According to the data, γ_{33} is the largest in the electro-optical coefficient of lithium niobate, and the maximum refractive index change will be produced when the applied electric field direction is parallel to the Z direction.

2.2　Mach-Zehnder modulator

Electro-optic modulator usually includes phase modulator and intensity modulator, etc. Mach-Zehnder (MZ) modulator in intensity modulator is the most widely studied and applied at present. It's basic structure is shown in Fig.1. The light enters the waveguide through the left side and divides the beam. By applying voltage to the electrodes placed on both sides of the waveguide, the waveguide refractive index is changed, so as to change[7] the phase of the transmitted light in the waveguide.
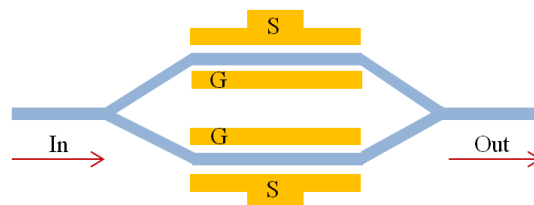
Fig.1　Schematic diagram of thin film lithium niobate MZ electro-optic modulator

The refractive index change and mode phase change caused by light passing through one straight waveguide can be expressed as[8]:

$$\Delta n_{TE} = -\frac{n_e^3 \gamma_{33} V}{2g} \Gamma \tag{2}$$

$$\Delta \varphi_{TE} = \frac{\pi n_e^3 \gamma_{33} VL}{\lambda g} \Gamma \tag{3}$$

Where, V is the applied voltage, g is the distance between the signal electrode and the ground electrode on both sides of the direct waveguide, L is the length of the modulating electrode, Γ is the overlap factor between the applied electric field $E(x,y)$ and the mode field $E_o(x,y)$ of the optical mode in the waveguide, expressed as[8]:

$$\Gamma = \frac{g}{V} \frac{\iint E_o^2(x,y) E(x,y) \, dxdy}{\iint E_o^2(x,y) \, dxdy} \tag{4}$$

When the photosynthetic beam in the two arms interferes, the output light intensity changes periodically, and the difference between the applied voltage corresponding to the maximum light intensity and the minimum light intensity is the half-wave voltage V_π of the modulator[8]:

$$V_\pi = \frac{\lambda g}{n_e^3 \gamma_{33} \Gamma L} \tag{5}$$

By using the push-pull electrode structure, the modulation electric field of the upper and lower two modulation arms can be opposite, so as to achieve twice the phase difference change, and the half-wave voltage of the MZ modulator can be reduced by half.

Another key performance indicator of an electro-optic modulator is the electro-optic modulation bandwidth, which is used to evaluate the high frequency response characteristics of the modulator and is usually defined as the frequency value when the electro-optic modulation curve drops to half (−3 dB). In MZ modulators, the electro-optic modulation curve usually declines as the frequency increases, and the higher the bandwidth indicates that the modulator has less high-frequency attenuation and can be used for higher rate signal transmission. The frequency response curve can be expressed as[8]:

$$M(f) = 20\log_{10}\left\{ \frac{2\sqrt{Z_0 Z_{in}}}{Z_0 + Z_{in}} e^{-\frac{\alpha L}{2}} \left[\frac{\sinh^2(\frac{\alpha L}{2}) + \sin^2(\frac{bL}{2})}{(\frac{\alpha L}{2})^2 + (\frac{bL}{2})^2} \right]^{\frac{1}{2}} \right\} \tag{6}$$

Where, Z_o is the characteristic impedance of the planar waveguide electrode, Z_{in} is the input

142

impedance of the external system, when the two impedances are equal, the microwave reflection in the electrode transmission process is zero, which can improve the electro-optical modulation performance; α is the transmission loss factor of the planar waveguide electrode, b is used to describe the degree of dispersion of the light wave and microwave signal, when the microwave refractive index in the transmission plate electrode matches the refractive index of the light wave group in the waveguide, the degree of dispersion b falls to zero, which can effectively improve the electro-optic 3 dB bandwidth.

In summary, the half-wave voltage of the modulator is determined by the overlap factor Γ of the light wave field and the applied electric field, and the modulation bandwidth is determined by the microwave transmission loss factor, impedance matching degree and refractive index dispersion degree. By optimizing the optical waveguide and electrode structure of the M-Z modulator by simulation, the half-wave voltage can be effectively reduced and the modulation bandwidth can be improved.

3 Design and simulation of modulator structure

The model studied in this paper is based on 0.3 μm thick x-cut lithium niobate film with a bottom SiO_2 thickness of 4.7 μm and a substrate Si thickness of 0.525 mm. The end-face structure of the M-Z modulator made by it is shown in Fig.2. The waveguide is defined by direct etching process in the thin film lithium niobate layer, where the waveguide width is set to 1 μm and the slope of the waveguide side wall is set to 70 degrees. In addition, in order to solve the problem of mutual restriction between low half-wave voltage and large modulation bandwidth, the planar waveguide traveling-wave electrode structure is adopted. S and G in the figure are signal electrode and ground electrode, respectively. In the figure, the x direction represents the Z axis of the crystal, and the y direction represents the X axis of the crystal. When the direction of the applied electric field is along the Z axis of the lithium niobate crystal, the optical waveguide can produce the maximum refractive index change and the corresponding phase change of the transmitted light.
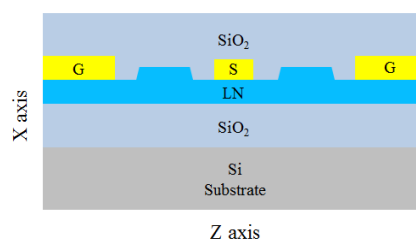
Fig.2　Schematic diagram of cross section structure of x-cut thin film lithium niobate modulator

3.1 Optical waveguide structure design

While considering the low half-wave voltage, M-Z modulator also needs to minimize the absorption loss of metal electrode. In this paper, Comsol Multiphysics simulation software was used to construct the structural model as shown in Fig.3. The waveguide structure and electrode spacing were analyzed and optimized under the condition of low frequency electrostatic field, so as to achieve the dynamic balance between half-wave voltage and loss. Among them, Fig.3(a) is a typical M-Z modulator structure, including two Y branches of input and output as well as the modulation arm region, to achieve beam splitting, modulation and interference of light waves; Fig. 3(b) shows

the cross section structure of a single modulation arm region. The upper and lower parts of the lithium niobate film are SiO_2 layers, and the electrodes on both sides are gold. The thin film lithium niobate material is set to anisotropic in the simulation, and the x, y and z directions of the material at 1550 nm are (2.1376) n_e, (2.2111) n_o and (2.2111) n_o respectively.
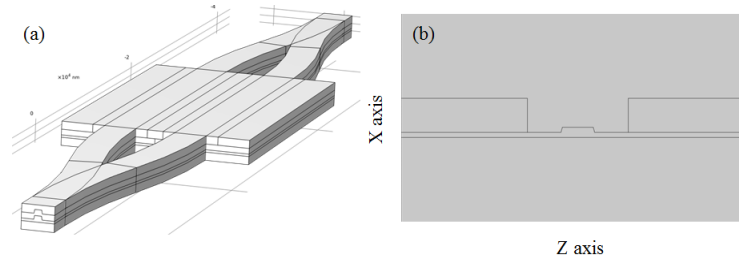
Fig.3 M-Z modulator (a) overall model construction diagram, (b) modulation arm cross section diagram in Comsol

By setting the frequency domain electromagnetic field in the program, the optical field transmission mode in the waveguide can be analyzed. Firstly, the mode distribution in the waveguide when the upper cladding is SiO_2 is studied. The fundamental mode is shown in Fig.4(a), where the white arrow represents the electromagnetic field distribution in the waveguide. At this time, the transmission mode is TE mode, the mode field distribution is uniform and the transmission loss is negligible, and obvious electric field discontinuity can be seen on both sides of the waveguide. This is because the electric displacement vector here is continuous and the relative dielectric constant of SiO_2 ($\varepsilon \approx 3.8$) is different from that of lithium niobate material ($\varepsilon \approx 27.9$). When the upper cladding is air, it's waveguide transmission mode is shown in Fig.4(b), the attenuation above the waveguide is significantly larger, the mode field distribution is uneven, and the internal electromagnetic field is distorted, which is because the upper and lower cladding material is different, the relative dielectric constant gap between air ($\varepsilon \approx 1$) and lithium niobate is larger, so that the electric field produces more serious attenuation. In addition, when the upper cladding is SiO_2, the effective refractive index of the fundamental mode is much larger than that when the upper cladding is air, as shown in Fig.4(c), and the effective refractive index increases with the increase of the width of the waveguide. After the thin film lithium niobate waveguide is etched, evaporation of SiO_2 layer on the surface of the waveguide can effectively improve the mode refractive index, reduce the electromagnetic field attenuation and effectively protect the waveguide structure. In addition, selecting the appropriate waveguide width can effectively avoid multi-mode transmission.
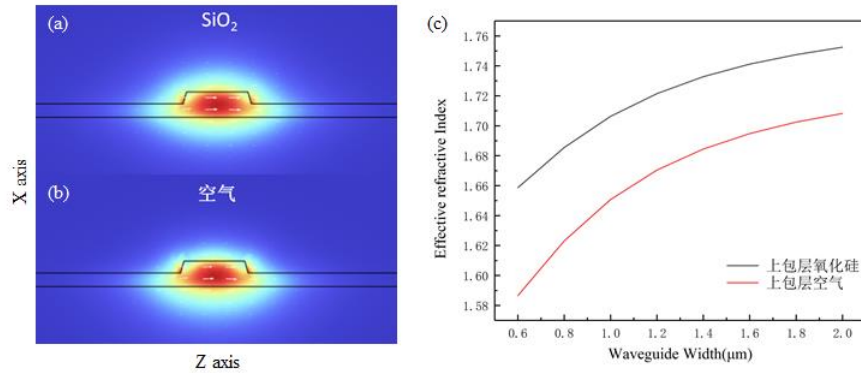
Fig.4 The basic mode pattern when the cladding is (a)SiO2 and (b) air; (c) effective refractive index at different waveguide widths

Then, the upper cladding material is selected as SiO_2, and the waveguide width is selected as 1 μm under the condition of relatively large effective refractive index and no multi-mode transmission. The influence of the etching depth d of the lithium niobate film on the TE mode in the optical waveguide is studied by changing the etching depth d. As can be seen from Fig.5(a), when the etching depth is shallow, the mode spot size is extremely large, indicating that the ridge waveguide has weak binding ability to the light field at this time. As the etching depth increases, the pattern size decreases gradually. The sidewall of the waveguide plays an important role in limiting the lateral diffusion of the light mode, and the effective refractive index of TE mode also decreases. Furthermore, the size of the mode spot also affects the intensity of electro-optical interaction, which is a key parameter index affecting the half-wave voltage.
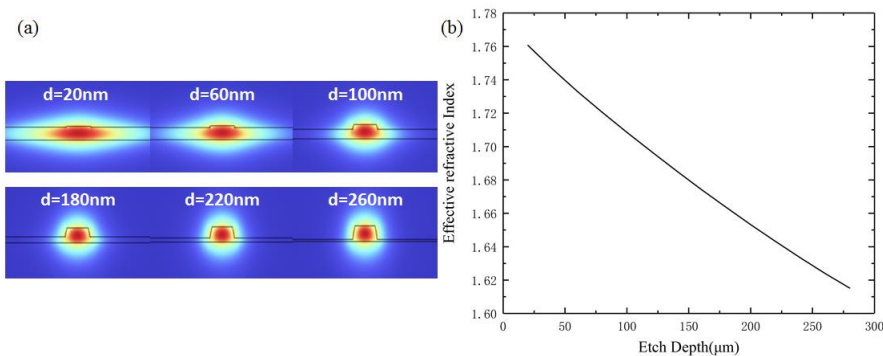
Fig.5 (a) Basic pattern pattern and (b) effective refractive index at different etching depths

3.2 Photoelectric hybrid simulation

After the optical waveguide mode calculation, add 1 V voltage to the gold electrode on the left side of the waveguide, and its potential distribution is shown in Fig.6(a), forming a zero potential point in the middle of the waveguide. The electric field distribution is shown in Fig.6(b), and the direction of the electric field is from the left electrode to the right electrode, as shown by the white arrow in the figure. The maximum electric field strength occurs at the end of the gold electrode, which is caused by the "hot spot effect" caused by the accumulation of charges in the metal. The relationship between the central electric field strength and electrode spacing is shown in Fig.6(c). With the increase of electrode spacing, the central electric field strength of the waveguide gradually

decreases, indicating that the electro-optical modulation effect will be weakened accordingly. With the decrease of electrode spacing, the photoelectric mixing effect is significantly enhanced, but it will also cause strong light absorption loss. Therefore, the two important indexes of etching depth and electrode spacing should be considered comprehensively, and the optimal parameters should be selected in dynamic balance.
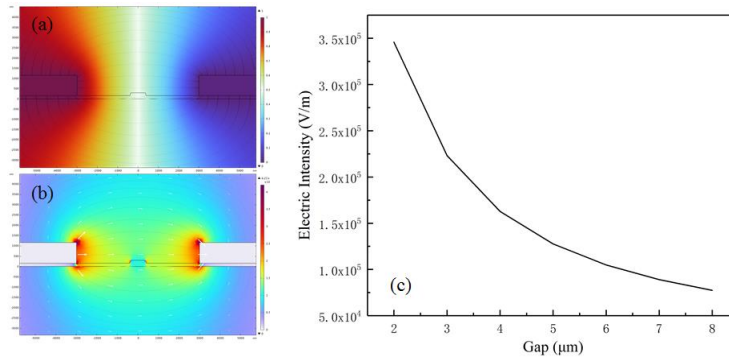
Fig.6 (a) Potential and (b) electric field distribution after electrode charging; (c) the electric field intensity in the center of the waveguide at different electrode spacing

At this time, the electric field and the optical waveguide mode exist at the same time, and the two interact with each other. The overlap factor Γ between the electric field and the optical mode in the lithium niobate region can be calculated by integrating the electric field and the optical mode in COMSOL. According to formula (5), the half-wave voltage length product $V_\pi L$ is inversely proportional to overlap factor Γ. To obtain a small $V_\pi L$, Γ needs to be large enough, and overlap factor Γ is related to electrode spacing, electric field intensity in the center of the waveguide and photoelectric interaction degree. By calculating the half-wave voltage length product and metal absorption loss under different etching depth and electrode spacing, the appropriate etching depth and electrode spacing parameters were selected by dynamic equilibrium.

In the simulation, the etching depth is increased from 20 nm to 280 nm, and the interval is 20 nm; Electrode spacing was increased from 2 μm to 8 μm with 1 μm interval. The calculated results of half-wave voltage length product and absorption loss are shown in Fig.7(a) and 7(b) respectively. As the etching depth gradually increases, the waveguide binding ability becomes stronger, the optical mode spot size decreases, and the degree of photoelectric interaction decreases, resulting in an increase in $V_\pi L$, but the metal absorption loss of optical energy decreases significantly. When the electrode spacing continues to shrink, the electric field in the center of the waveguide gradually becomes stronger, resulting in a continuous reduction of $V_\pi L$, but an exponential increase in metal loss. In all the calculated results, considering the half-wave voltage and loss, the etching depth is 140 nm and the electrode spacing is 6 μm. At this time, the half-wave voltage length product of the MZ modulator is about 3 V•cm, and the metal absorption loss is about 0.1 dB/cm. This index can meet the application requirements of the integrated fiber optic gyroscope system. It provides a basis for the development of the subsequent fabrication process of the modulator.
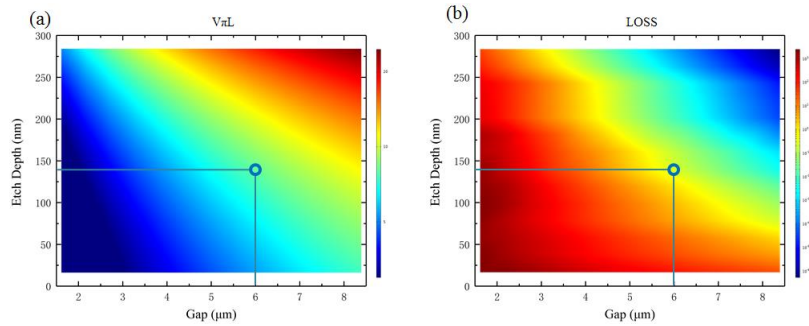
Fig.7 (a) half-wave voltage length product and (b) absorption loss at different etching depths and electrode spacing

3.3 Electrooptic high frequency simulation

The modulation bandwidth is an important parameter used to evaluate the high-frequency characteristics of the modulator. According to formula (6), it is determined by the transmission loss factor of the planar waveguide electrode, the dispersion degree of the effective refractive index of the light wave and the microwave refractive index, the characteristic impedance of the electrode and the impedance dispersion degree of the load, etc. In this paper, the high frequency transmission model of thin film lithium niobate is constructed in HFSS software, and the high frequency performance of the modulator is optimized.
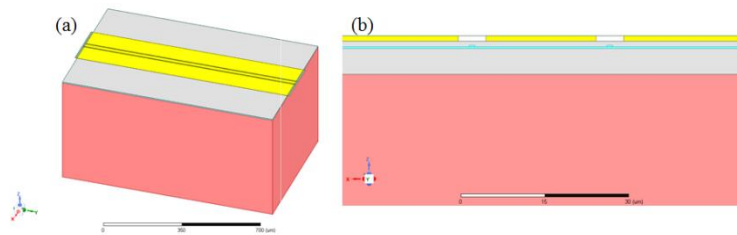
Fig.8 Thin film lithium niobate modulator (a) high frequency simulation model; (b) schematic diagram of the end face structure

The model constructed in this paper is shown in Fig.8, where the red is Si substrate, the gray is SiO_2 layer, the blue is lithium niobate film, and the yellow is gold electrode structure. The terminal impedance is set to 50 Ω, and then high-frequency ports are set at the input and output of microwave signals. Under the above waveguide structure and electrode spacing, the electrode structure parameters were changed to study its response to microwave loss, microwave refractive index, impedance and other parameters. When the width of the signal electrode S increases, the overall capacitance of the electrode increases and the inductance decreases, resulting in the overall microwave refractive index of the electrode decreases, and the characteristic impedance of the electrode decreases significantly. When the overall electrode thickness increases, the microwave transmission loss decreases rapidly, but the characteristic impedance and microwave refractive index decrease significantly.

In order to reduce the microwave loss, match the characteristic impedance with the 50 Ω load impedance, and match the microwave refractive index with the effective refractive index of the light wave, the signal electrode width is finally set to 20 μm, the ground electrode width is 100 μm, the overall electrode thickness is 1 μm, and the electrode spacing is 6 μm. The simulation results are

shown in Fig.9. The results show that the overall reflection parameter S11 is less than -15 dB, showing good characteristic impedance matching characteristics. In addition, due to the reduction of microwave loss at the same time, the modulator 3 dB bandwidth exceeds 80 GHz.
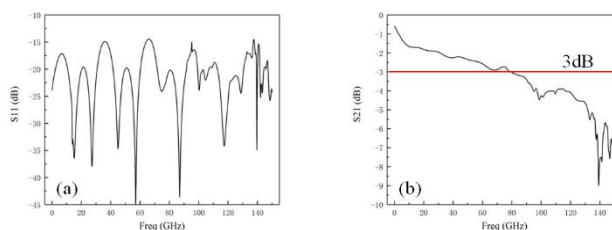
Fig.9　(a) reflectance parameter S11 of electrode structure; (b)3dB bandwidth S21

In summary, the waveguide and electrode structure of thin film lithium niobate electro-optic modulator are optimized by photoelectric hybrid simulation of multi-physical fields. The key structural parameters are selected and determined according to the half-wave voltage and loss characteristics of the modulator, which provides reference for subsequent processing. In addition, the designed modulator half-wave voltage length product is less than 3 V•cm and the bandwidth is more than 80 GHz, which is far superior to the traditional bulk material lithium niobate modulator, and is expected to significantly improve the performance of fiber optic gyro and reduce the production cost of gyro. The next step will continue to optimize the design process and develop the preparation process according to the simulation results.

4　Conclusion

On-chip conversion of electro-optic modulator is an inevitable trend in the development of fiber optic gyroscopes. In this paper, COMSOL multi-physical field simulation software and HFSS RF simulation software are used to design an on-chip electro-optic modulator based on 300 nm thick lithium niobate film, which can greatly improve the modulation performance while reducing the device volume and improve the consistency of gyro products. The performance of the designed on-chip electro-optic modulator is better than that of the traditional bulk material lithium niobate Y waveguide, which provides a new idea for the production and development of gyro.

References:

[1]　Reed G, Mashanovich G, Gardes F, et al. Silicon optical modulators[J]. Nature Photonics, 2010, 4: 518-526.

[2]　Zhang San-mei, Meng Xing. Use of an inertial/magnetic sensor module for pedestrian tracking during normal walking[J]. IEEE Transactions on Instrumentation and Measurement, 2015, 64(3): 776-783.

[3]　Ohmachi Y，Noda J. Electro-optic light modulator with branched ridge waveguide[J]. Applied Physics Letters, 1975, 27(10): 544-546.

[4]　Jackel J L, Rich C E, Veselka J J. Proton exchange for high-index waveguides in LiNbO3[J]. Applied Physics Letters, 1982, 41(7):607-608.

[5]　Cheng B，Zeng M，Liu Y，et al. Simulation research of PIGA based on ADAMS[C].2016 China International Conference on Inertial Technology and Navigation, 2016.

[6]　Liu Ye. Research on High-Speed Thin-Film Lithium Niobate Mach-Zehnder Modulators [D]. Huazhong University of Science and Technology, 2021.

[7]　Giovanni G. Semiconductor devices for high-speed optoelectron-ics[M].Cambridge University Press, UK, 2009.

[8]　Jian J, Xu M, Liu L, et al. High modulation efficiency lithium niobate Michelson interferometer modulator[J].

Optics Express, 2019, 27(13): 18731-18739.

[9] Li Jinyang, Lu Danfeng, Qi Zhimei. Wavelength Dependence Characteristics Analysis of Electro-Optic Overlap Integral Factors in Lithium Niobate Waveguides [J]. Acta Physica Sinica, 2014, 63(7): 077801-1-077801-7.

Wireless Implementation of Surface Acoustic Wave Accelerometers

M. A. Sorvina

(SPbGETU "LETI", St. Petersburg)

Abstract: The operating principle and development trends of surface acoustic wave accelerometers are described. Four methods for implementing wireless data transmission from a sensitive element are proposed. Their advantages and disadvantages are described. A simplified multiphysical model is presented, allowing the proposed methods to be simulated.

1 Introduction

Surface acoustic wave (SAW) accelerometers have attracted considerable attention from research groups due to their robustness, impact resistance, high sensitivity, and low power consumption [1-2]. These accelerometers use the propagation of acoustic waves along the surface of a piezoelectric material to estimate the acceleration values of the object on which they are installed. With the development of wireless technologies [3], the integration of SAW accelerometers with wireless data transmission systems is becoming increasingly feasible, making them ideal for various applications, such as industrial monitoring [4].

This paper examines ways to organize the sensitive element of a SAW accelerometer for subsequent data transmission.

2 Wireless Surface Acoustic Wave Accelerometer

The report includes a description of the operating principle of a SAW accelerometer, variants of passive connection schemes, and modeling in Comsol Multiphysics of a sensitive element with two-way signal transmission.

2.1 Configuration of the sensitive element of the SAW accelerometer

The sensing element of the SAW accelerometer includes a piezoelectric substrate 1 equipped with interdigital transducers (IDT) 2 and 3 on its surface (Fig. 1a). The input IDT 3 converts the electrical signal into an acoustic wave, which propagates along the substrate to the output IDT 2, which performs the inverse transformation [5]. The acceleration measurement process consists of the following steps: first, acceleration causes deformation of the material; second, internal mechanical stresses cause the density to change; third, the change in the material density affects the wave propagation velocity, and, consequently, the wave propagation time along the substrate changes. It is the time delay that can be used to judge the change in the acceleration of the object. Typical design configurations of these sensors include delay lines and resonators, which are structured to maximize sensitivity and measurement accuracy.

2.2 Implementation of wireless data transmission

The basic setup includes a transmitter connected to a SAW accelerometer that converts an electrical signal into a radio frequency signal. The receiver receives this signal and processes it to interpret the data [6]. There are several ways to implement data transmission from the sensing element.

The first method is shown in Figure 1a, when the signal is fed to antenna 4, from which it enters the exciting IDT 3, which generates a SAW that propagates to the receiving IDT, where the reverse conversion into an electrical signal occurs and the data is transmitted via antenna 5. The described method is closest to the wire pickup method. However, its disadvantage is possible interference between the antennas.

The second method (Fig. 1b) involves using one antenna for both receiving and transmitting the signal and placing two reflector blocks, the difference in the distance between their paths can be found by subtracting one from the other: r2-r1. In this way, it will be possible to avoid interference between the antennas. But the disadvantage of this will be the potential excessive weakening of the SAW amplitude, which will have to pass through the first reflector twice. The next two methods allow you to avoid this.

The third method (Fig. 1c) suggests placing the reflectors on different sides of the exciting IDT. The disadvantage may be the subsequent passage of the wave front through the IDT on the return path and the superposition of the waves on each other, which will cause an echo and interfere with the identification of the useful signal.

The fourth method suggests using 2 delay lines (Fig. 1.d). However, a disadvantage may be the mutual influence of the fronts of waves simultaneously propagating along the surface, as well as the occurrence of an electromechanical connection between two IDTs.
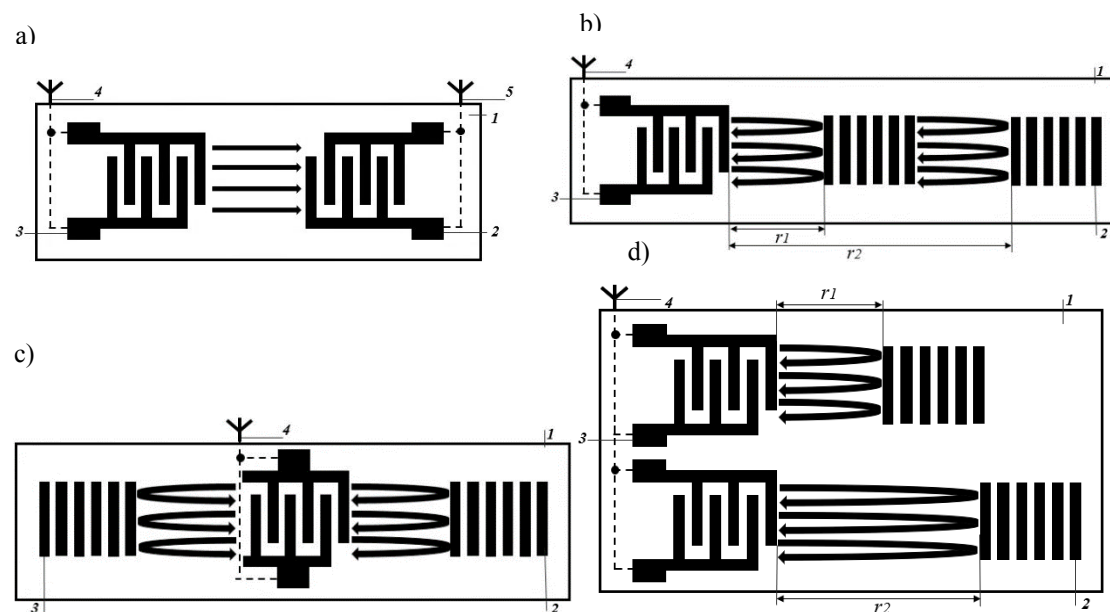
Fig. 1   Topologies of the sensitive element for data transmission

2.3   COMSOL Modeling for Bidirectional Signal Transmission

In order to investigate the methods described above, a simplified model was built in the Comsol Multiphysics environment (Fig. 2a). The dimensions of the model were $3.75\lambda \times 0.25\lambda \times 3\lambda$, where λ is the wavelength, which is 0.796 mm. The model specifies two electrodes with a height of 0.01λ, one of them is grounded, and the other is assigned the Terminated condition with the Circuit type. A sinusoidal signal with an amplitude of 1 V and a period of 0.2 μs is fed to this electrode.
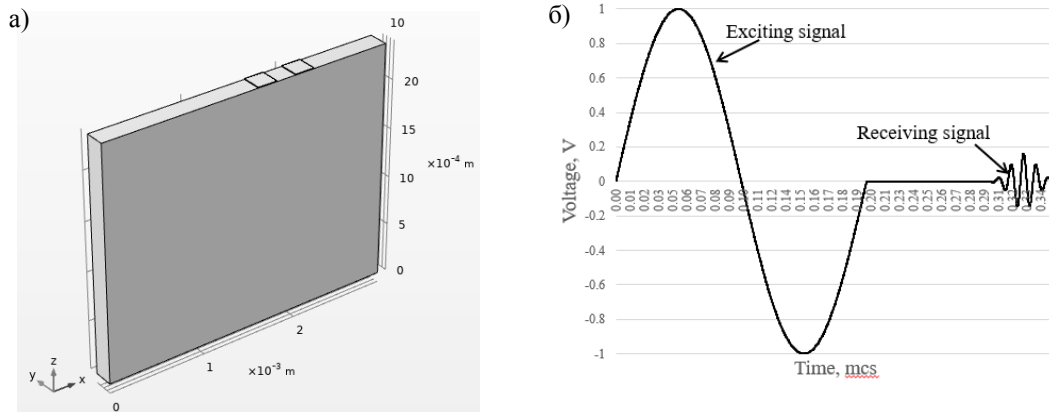
Fig. 2   Simplified model of the sensitive element and graph of voltage change on the IDT

The voltage change graph is shown in Figure 2b. As can be seen, the signal amplitude has decreased by 5 times due to losses. This method of setting the conditions for signal feed and pickup from one IDT is promising for further research.

3   Conclusion

The report examined the principles of wireless data transmission for a SAW accelerometer. Four methods of signal transmission and reading were proposed. A multiphysical model was developed for calculating the parameters and voltage graphs were obtained.

References:

[1]   Yihan Zhao et al. Equal-strength beam design of acoustic wave accelerometers. 2023 Phys. Scr. 98

[2]   Juan Zhang, Wen Hua Cheng, Hao Wang, Lei Zhang, Xiang Rong Li, Hong Shuai Ma, Jun Qi Pang & QiuLin Tan. Langasite-based SAW high-temperature vibration sensor with temperature decoupling. Science China Technological Sciences. Volume 67, pages 1946–1956, 2024

[3]   L. Reindl, I. Shrena, S. Kenshil and R. Peter. Wireless measurement of temperature using surface acoustic waves sensors, IEEE International Frequency Control Symposium and PDA Exhibition Jointly with the 17th European Frequency and Time Forum, 2003. Proceedings of the 2003, Tampa, FL, USA, 2003, pp. 935-941

[4]   Debdyuti Mandal, Sourav Banerjee. Surface Acoustic Wave (SAW) Sensors: Physics, Materials, and Applications. Sensors 2022, 22(3), 820

[5]   Shevchenko S. et al. Surface-Acoustic-Wave Sensor Design for Acceleration Measurement. Sensors 2018. 18(7), 2301

[6]   Fabian Lurz et al. Reader Architectures for Wireless Surface Acoustic Wave Sensors. Sensors 2018, 18(6), 1734

Effect of Installation Noncoincidence of Shearer Inertial Navigation on Straightness Deviation of the Fully Mechanized Mining Face

Shang Ziqin[1], Chen Shaohua[2,3], Li Guannan[1], Zheng Yi[4], Hu Huizhu[1]

(1.College of Optical Science and Engineering, Zhejiang University, Hangzhou, China; 2.Hangzhou Applied Acoustics Research Institute, Hangzhou, China; 3.Navigation Research Center, College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China; 4. Center for Novel Computing Sensing and Intelligent Processing, Zhejiang Lab, Hangzhou, China.)

Abstract: The straightness of the fully mechanized mining face is crucial for ensuring safety and efficiency in underground mining, as it indicates how much the mining track deviates from a straight line. This deviation is assessed through shearer positioning, which utilizes Fiber Optic Gyroscopes (FOGs) known for their large dynamic range and resilience to shock and vibration. FOGs are essential in harsh environments characterized by dust and low visibility. However, installation noncoincidence between the inertial measurement unit (IMU) containing the FOG and the shearer carrier negatively impacts straightness measurements. Despite its importance, there is limited literature on this topic. This study presents an error modeling approach to analyze the relationship between installation noncoincidence and straightness measurement. Using a dead reckoning model as our inertial navigation algorithm, we constructed a test rail to collect data from FOG-IMU/odometer systems under simulated shearer conditions, introducing installation noncoincidence angles of 1°, 3°, 5°, and 10° for analysis. Our findings reveal that pitch direction noncoincidence significantly affects straightness deviation, while heading and roll directions have minimal impact. This error model offers valuable insights for correcting straightness deviations in inertial navigation systems on shearers and serves as a reference for addressing other linear motion errors in underground settings.

 Key words: inertial navigation; straightness deviation; installation noncoincidence;

1 INTRODUCTION

Automation of longwall mining has become an inevitable trend in the development of the coal mining industry[1]. It aims to enable automated systems to complete mining-related tasks, such as straightness measurement, adjustment and equipment maintenance, to reduce the number of underground workers and keep them away from dust and risk of roof collapse and other harsh environments[2], [3], [4].

Monitoring the working conditions of fully mechanized mining face is important, especially the straightness of the mining face, which is closely related to the safety and efficiency of longwall mining[5]. Straightness reflects the degree of deviation of the trajectory of the mining face, and subsequent adjustments are made by hydraulic supports. The conventional methods typically

involve manual wire pulling, which presents significant safety risks in underground environments. Due to the unsuitability of Global Navigation Satellite System (GNSS) in underground settings, vehicle-mounted GNSS-based positioning technology [6] cannot be utilized in mines. Currently, shearer positioning technology is employed to achieve this objective, with straightness data being derived from converting the shearer's motion trajectory under spatial constraints imposed by the shearer, scraper conveyor and hydraulic support at the mining face. The prevailing positioning technologies for coal shearers primarily encompass wireless transceiver methods [7], [8], [9] and autonomous positioning methods [10], [11]. For instance, Yang et al. [7] proposed an enhancement method for a strapdown inertial navigation/ultra-wideband combined system to enhance accuracy. Although they utilize different auxiliary sensors, all are reliant on Strapdown Inertial Navigation Systems (SINS).

However, angular deviations may occur during the installation process, impacting the measurement accuracy of the inertial system due to the double-integral calculation of the SINS. Despite the initial calibration, these angles persist as the device wears and deforms during the coal mining. Some researchers have analyzed this impact on the shearer positioning accuracy and proposed some calibration schemes. Ge et al. [12] found that the installation deviation of heading and pitch angle has a significant impact, and further [13] proposed a calibration method based the similarity of dead reckoning trajectory and the actual trajectory. Wu et al. [14] also simulated this impact on positioning accuracy and assessed the influence of the lever arm. However, current error analysis and calibration solutions primarily focus on absolute errors in geographical coordinates, overlooking the impact of straightness conversion on trajectory deviations in actual mining operations.

To address this gap, this study conducts error modeling to explore the relationship between installation noncoincidence and straightness deviation measurements in Section Two. Section Three details simulation experiments and results. The subsequent analysis confirms that the test results are consistent with the error modeling findings.

2 INSTALLATION NONCOINCIDENCE AND ERROR ANALYSIS

2.1 Installation Noncoincidence

The shearer body coordinate system b (right-forward-up) and IMU coordinate system m are defined. Let ψ, θ, γ be the heading angle, pitch angle, and roll angle respectively.

The installation noncoincidence angle can be expressed by the transformation matrix. The original coordinate system b is rotated three times around each axis as follows:

$$\boldsymbol{C}_m^b = \boldsymbol{C}_j^b \boldsymbol{C}_h^j \boldsymbol{C}_m^h \tag{1}$$

where c is the transformation matrix and the subscript represents the coordinate system. A $\Delta\psi$ rotation around the $Z-$axis yields intermediate system j, followed by a $\Delta\theta$ rotation around the $X-$axis to obtain coordinate system h, and finally a $\Delta\gamma$ rotation around $Y-$axis

results in system m. Assumption: The motion state of the shearer is similar to the vehicle motion state and satisfies a nonholonomic constraint.

2.2 Error analysis

The coordinate systems involved are as follows: navigation system n, shearer system b, IMU system m and relative working surface system r.

The velocity measured by the odometer under ideal conditions represents:

$$\boldsymbol{v}^n = \boldsymbol{C}_b^n \boldsymbol{v}^b \tag{2}$$

where $\boldsymbol{C}_b^n$ is a three by three dimensional transform matrix from system b to system n, and the forward speed of the shearer is not zero, $\boldsymbol{v}^b = \begin{bmatrix} 0 & v & 0 \end{bmatrix}^T$, the odometer installation is in ideal condition, $\boldsymbol{C}_s^b = \boldsymbol{I}$, s represent the odometer system.

When there is an installation deviation angle, the velocity in system n, $\tilde{\boldsymbol{v}}^n$:

$$\tilde{\boldsymbol{v}}^n = \boldsymbol{C}_b^n \boldsymbol{C}_m^b \boldsymbol{v}^b \tag{3}$$

where $\boldsymbol{C}_m^b$ is the rotation matrix representing the installation noncoincidence angle, as described in part A. Then subtract (37) and (38) to get the velocity error $\Delta \boldsymbol{v}^n$.

$$\Delta \boldsymbol{v}^n = \boldsymbol{v}^n - \tilde{\boldsymbol{v}}^n = \boldsymbol{C}_b^n (\boldsymbol{I} - \boldsymbol{C}_m^b) \boldsymbol{v}^b \tag{4}$$

The position error integrated over time:

$$\Delta \boldsymbol{p}^n(t) = \int_0^t \Delta \boldsymbol{v}^n(\tau) d\tau = \int_0^t \boldsymbol{C}_{b(\tau)}^n (\boldsymbol{I} - \boldsymbol{C}_{m(\tau)}^{b(\tau)}) \boldsymbol{v}^b(\tau) d\tau \tag{5}$$

where $\boldsymbol{C}_{m(t)}^{b(t)}$ is transform matrix in time t, which follows the chain rule.

$$\boldsymbol{C}_{m(t)}^{b(t)} = \boldsymbol{C}_{b(0)}^{b(t)} \boldsymbol{C}_{m(0)}^{b(0)} \boldsymbol{C}_{m(t)}^{m(0)} \tag{6}$$

Assuming that the installation deviation angle remains unchanged since the initial position, $\boldsymbol{C}_{b(0)}^{b(t)} = \boldsymbol{C}_{m(0)}^{m(t)}$, then:

$$\boldsymbol{C}_{m(t)}^{b(t)} = \boldsymbol{C}_{m(0)}^{b(0)} = \boldsymbol{C}_m^b \tag{7}$$

The straightness deviation:

$$\begin{bmatrix} \Delta \boldsymbol{s}^r \\ 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{C}_n^r & -\boldsymbol{p}_1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{p}^n \\ 1 \end{bmatrix} \tag{8}$$

$$C_n^r = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{8}$$

where α represents the angle between the $X-$ axis in system n and the line connecting the starting point and the end point of the trajectory, p_1 is the starting point of each trajectory, simplify it to zero for explanation, $p_1 = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T$, C_n^r is a transform matrix. By substituting (36) and (40) into (43), we have:

$$\Delta s^r = C_n^r \int_0^t C_{b(\tau)}^n (I - C_{m(\tau)}^{b(\tau)}) v^b(\tau) d\tau = \int_0^t (C_{b(\tau)}^r - C_{b(\tau)}^r C_{m(\tau)}^{b(\tau)}) v^b(\tau) d\tau \tag{9}$$

During the straightness deviation conversion process, the trajectory error caused by part of the heading angle error is corrected, i.e. $\alpha \approx -\psi$.

$$C_b^r C_j^b \approx I \tag{10}$$

Substitute (46) into (45):

$$\Delta s^r \approx \int_0^t (C_{b(\tau)}^r - C_h^j C_m^h) v^b(\tau) d\tau \tag{11}$$

In that way, the deviation in straightness is influenced by both the pitch and roll angle. Analysis of shearer motion reveals minimal movement in the roll direction, which has a limited impact on overall motion. Ultimately, the straightness deviation is primarily affected by the pitch angle in theory.

3　EXPERIMENTS AND RESULTS

The experiment is conducted on a custom-built test platform, which consists of a monorail mobile carrier, a Inertial Navigation Systems (INS), an odometer and a GNSS-RTK mobile station, as depicted in Fig. 1(a). The INS consists of a FOG-IMU and an embedded motherboard. Tab. 1 gives the key parameters of the FOG-IMU used and the data rate is 100Hz as same as the odometer. The GPS-RTK accuracy is less than 15 mm and 8 mm in height and level, respectively. And Fig. 1(b) gives the true value of straightness deviation measured by GPS-RTK as a reference and this information does not participate in positioning.

The data collection process is as follows. To simulate the back-and-forth coal mining movement of the shearer, the monorail mobile carrier runs back and forth on the experimental rail at a speed of about 3m/min to collect FOG-IMU/Odometer data. Collected sensor data is segmented from the stable phase of each power-on, including truncating the initial 5-minute break and truncating the 1-minute pause after each segment is completed. During the simulation phase, the original noncoincident angle is set to zero and subsequently one of 1°, 3°, 5° and 10° are added as shown in the Tab. 2. Note that the one of 10° is deliberately chosen to emphasize its impact on the

results.

3.1 Results

The Tab. 2 indicates the straightness deviation error after introducing 1°, 3°, 5° and 10° installation noncoincidence angle. For example, the Mean Absolute Error (MAE) of straightness deviation caused by 1° of pitch angle is 2.69 mm, which accounts for 97.8% of the MAE of 2.78 mm caused by 1° of three angles, while the ratio of roll angle is only 0.1%. In addition, in the pitch angle, the Maximum Error (ME) is 7.98 mm at 1° and increases by 0.138m at 10°, while the ME caused by roll angle change only increases by less than 1 mm. As a result, compared with the angle of the heading and the roll, the pitch one has the greatest impact on the straightness deviation, and the straightness deviation error increases with the increase of the installation angle. Among these angles, the roll angle installation deviation has the smallest impact. When there is only roll angle installation deviation, the straightness deviation error approaches zero.

Fig. 2 plots the impact of the one cutting data of the trajectory on the straightness deviation errors in different directions and sizes of the installation angle, which is given as a supplement to tab. 2. Overall, it can be seen from the Fig. 2(a)-(d) that the changing trends of straightness deviation errors are similar despite different set angles. For instance, for the 10° angles in Fig. 2(d), the straightness deviation error increases at the position where the straightness deviation changes drastically, approximately in the displacement range of 20-40 m, with the maximum value around 30m, which corresponds to Fig. 1(b), and the analysis of this relationship is expanded in the discussion section. Furthermore, the installation angle trajectories in the pitch direction and the three directions exhibit close proximity, while errors of the heading and roll consistently hover near zero. This suggests that the primary influencing factor of measurement error lies in the installation deviation of the pitch angle.

Tab.1   FOG-IMU Performance Parameters

	Parameter	Value
Gyroscope	In-run bias stability($°/h$)	0.05
	Angular random walk($°/\sqrt{h}$)	≤0.012
Accelerometer	In-run bias stability(mg)	0.05
	Velocity random walk($mg/\sqrt{h}$)	0.05
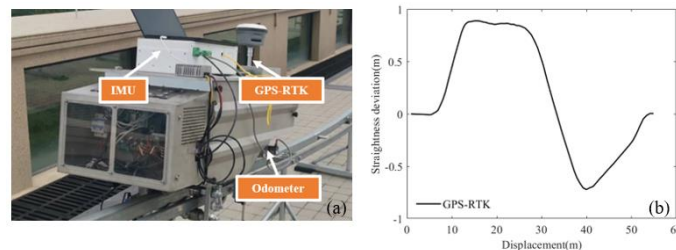
Fig.1   Experimental setup and straightness deviation (a) Experimental setups for simulating the mining process (b) The true value of straightness deviation
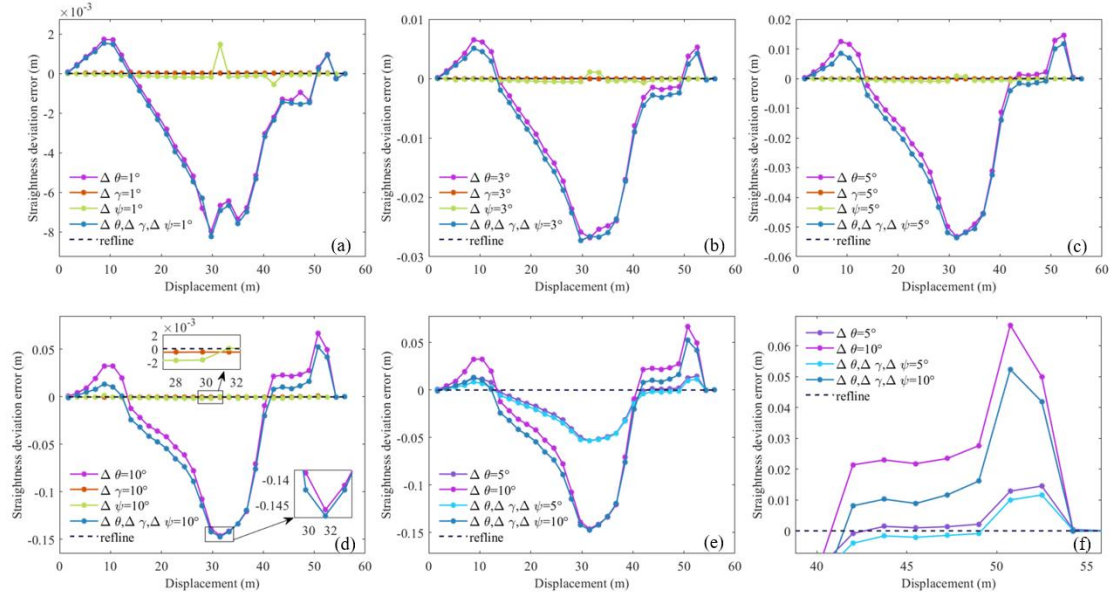
Fig.2 Compare diagrams of the shearer straightness deviation errors (a) 1° (b) 3° (c) 5° (d) 10° (e) Comparison diagram of shearer straightness deviation (f) Partial magnification of (e)

3.2 Discussion

To further analyze the error influence mode, Fig. 2(e)-(f) plots the relationship between the two installation angles (5° and 10°) and the straightness deviation error. We speculate that the impact of one angle installation deviation will be smaller than the impact of installation angle deviation in all three directions. However, in fact, it is just the opposite at some positions (e.g. at the interval 40-50 m), and becomes obvious as the angle increases, as shown in the Fig. 2(f). Assuming that the measured value on the right side of the forward direction is positive, it is not difficult to understand that the installation angle has an impact on the change of installation deviation. As shown in Fig. 1(b) and 4, the straightness deviation in the interval 10-15 m and 40-50 m gradually increases, while the interval 20-40 m gradually decreases, and the corresponding positive and negative signs of the deviation error are opposite. Therefore, the influence of the installation noncoincidence angle is related to the changing trend of the straightness deviation.

Tab.2   The Installation Noncoincidence Angle's Affects

Heading (°)	Pitch (°)	Roll (°)	Max error (m)	Mean absolute error (m)
1	0	0	1.46E-03	1.54E-04
3	0	0	1.13E-03	3.46E-04
5	0	0	9.16E-04	4.93E-04
10	0	0	1.72E-03	7.90E-04
0	1	0	7.98E-03	2.69E-03
0	3	0	2.68E-02	8.96E-03
0	5	0	5.33E-02	1.70E-02

0	10	0	1.46E-01	4.86E-02
0	0	1	7.30E-06	4.31E-06
0	0	3	2.17E-05	1.27E-05
0	0	5	9.92E-05	6.37E-05
0	0	10	6.50E-04	3.55E-04
1	1	1	8.25E-03	2.78E-03
3	3	3	2.73E-02	9.39E-03
5	5	5	5.35E-02	1.74E-02
10	10	10	1.47E-01	4.71E-02

4 CONCLUSION

As a crucial element of intelligent longwall mining, the straightness of fully mechanized mining faces has garnered attention from researchers in tandem with the advancement of coal mining. This study conducts error analysis based on dead reckoning and collects data from a custom-built test platform to carry out installation noncoincidence angle simulation experiments. The findings indicate that the pitch installation deviation angle exerts the most significant influence on straightness deviation, and that straightness deviation error increases with greater

installation angle deviation. This error model offers a theoretical rationale for rectifying straightness deviation through the use of an INS installed on the shearer, and serves as a point of reference for other linear motion scenarios underground. Moving forward, we aim to further target this feature and compensate for errors to advance unmanned and intelligent development within the mining industry.

References:

[1] J. C. Ralston, C. O. Hargrave, and M. T. Dunn, "Longwall automation: trends, challenges and opportunities," International Journal of Mining Science and Technology, vol. 27, no. 5, Art. no. 5, 2017.

[2] W. Dai, S. Wang, and S. Wang, "Longwall Mining Automation—The Shearer Positioning Methods between the Longwall Automation Steering Committee and China University of Mining and Technology," Applied Sciences, vol. 13, no. 22, p. 12168, Nov. 2023.

[3] J. C. Ralston, D. C. Reid, M. T. Dunn, and D. W. Hainsworth, "Longwall automation: Delivering enabling technology to achieve safer and more productive underground mining," International Journal of Mining Science and Technology, vol. 25, no. 6, Art. no. 6, 2015.

[4] J. Wang and Z. Huang, "The Recent Technological Development of Intelligent Mining in China," Engineering, vol. 3, no. 4, pp. 439–444, Aug. 2017.

[5] D. C. Reid, D. W. Hainsworth, J. C. Ralston, R. J. McPhee, and C. O. Hargrave, "Inertial Navigation: Enabling technology for longwall mining automation," 2003.

[6] Q. Chen, H. Lin, J. Kuang, Y. Luo, and X. Niu, "Rapid Initial Heading Alignment for MEMS Land Vehicular GNSS/INS Navigation System," IEEE Sensors J., vol. 23, no. 7, pp. 7656–7666, Apr. 2023.

[7] H. Yang, T. Luo, W. Li, L. Li, Y. Rao, and C. Luo, "A Stable SINS/UWB Integrated Positioning Method of Shearer Based on the Multi-Model Intelligent Switching Algorithm," IEEE Access, vol. 7, pp. 29128–29138, 2019.

[8] B. Cao, S. Wang, S. Ge, and W. Liu, "Improving Positioning Accuracy of UWB in Complicated Underground NLOS Scenario Using Calibration, VBUKF, and WCA," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–13, 2021.

[9] M.-G. Li, H. Zhu, S.-Z. You, and C.-Q. Tang, "UWB-Based Localization System Aided With Inertial Sensor

for Underground Coal Mine Applications," IEEE Sensors J., vol. 20, no. 12, pp. 6652–6669, Jun. 2020.

[10] S. Wang, B. Zhang, S. Wang, and S. Ge, "Dynamic Precise Positioning Method of Shearer Based on Closing Path Optimal Estimation Model," IEEE Trans. Automat. Sci. Eng., vol. 16, no. 3, pp. 1468–1475, Jul. 2019.

[11] D. Song, J. Yang, and J. Song, "Shearer-Positioning Method Based on Non-Holonomic Constraints," Applied Sciences, vol. 12, no. 19, p. 10050, Oct. 2022.

[12] S. Hao, A. Li, S. Wang, Z. Ge, Z. Zhang, and S. Ge, "Effects of shearer inertial navigation installation noncoincidence on shearer positioning error," Journal of China Coal Society, vol. 40, no. 8, pp. 1963–1968, 2015.

[13] B. Zhang, S. Wang, and S. Ge, "Effects of initial alignment error and installation noncoincidence on the shearer positioning accuracy and calibration method," Journal of China Coal Society, vol. 42, no. 3, pp. 789–795, 2017.

[14] G. Wu, X. Fang, Y. Song, M. Liang, and N. Chen, "Study on the Shearer Attitude Sensing Error Compensation Method Based on Strapdown Inertial Navigation System," Applied Sciences, vol. 12, no. 21, p. 10848, Oct. 2022.

Sensorless Control of Induction Motor

A. K. Golubev, A. A. Pyrkin, V. M. Fomin

(ITMO University, Saint Petersburg)

Abstract: In this paper, solves the problem of sensorless control of induction motor. A control algorithm able to operate under parametric and signal uncertainty is proposed. The proposed algorithm allows to expand the application area of induction motors in robotic technical systems with changing load and incomplete information about the parameters of the model. The performance is confirmed by the results of computer modeling.

1 Introduction

Induction motors are used in automation systems because of their efficiency and reliability. However, when controlled by traditional methods, under conditions of changing load, external and internal uncertainty, problems arise, in which case systems using adaptive control algorithms have a high relevancy index.

In this paper, we propose a sensorless vector control algorithm for a discrete model of an induction motor. The algorithm works in a wide range of parametric and signal uncertainty, which makes it promising for use in modern robotic systems, with increased requirements for accuracy and reliability, as well as when there is a lack of a priori information about the modes of operation and operating conditions.

2 Results

The dynamic model of an induction motor in a stationary frame of reference related to the stator in continuous time is defined by the following equations [1]:

$$\dot{\lambda} = -\left(\frac{R_r}{L_r}\boldsymbol{I} - n_p\boldsymbol{J}\omega\right)\lambda + \frac{R_r L_m}{L_r}i,$$

$$\frac{di}{dt} = -\left(\frac{R_s}{L_s\sigma} + \frac{R_r L_m^2}{\sigma L_s L_r^2}\right)i + \frac{L_m R_r}{\sigma L_s L_r}\left(\frac{R_r}{L_r}\boldsymbol{I} - n_p\boldsymbol{J}\omega\right)\lambda + \frac{1}{\sigma L_s}v, \qquad (1)$$

$$\dot{\omega} = \frac{1}{D}(\tau - \tau_L),$$

where: λ – rotor flux; i – stator current; v – stator voltage; ω – rotor velocity; R_r и L_r – rotor resistance and inductance, respectively; R_s и L_s – stator resistance and inductance, respectively; L_m – mutual induction; $\sigma = 1 - \frac{L_m^2}{L_s L_r}$ – inductance leakage factor; n_p – number of pole pairs; D – inertia torque; τ – electromagnetic torque; τ_L – load torque; $\boldsymbol{I}$ – 2x2 identity matrix; $\boldsymbol{J} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

As can be seen, the model (1) is nonlinear, which largely complicates the synthesis of regulators. In practical implementation, a regulator with a large gain is used in the current loop:

$$v = \frac{1}{\epsilon}(i^d - i), \qquad (2)$$

where: i^d – desired stator current value; ϵ – small positive coefficient.

In the limiting case, the coefficient in equation (2) tends to 0 and the motor model (1) is simplified. Performing the transition to the frame of reference rotating with the rotor and sampling the motor model, the following is obtained [2, 3]:

$$\lambda_{k+1} = a\lambda_k + (1-a)u_k,$$
$$\tau_k = bu_k^T J\lambda_k,$$
$$\omega_{k+1} = \omega_k + c\tau_k - d\tau_L,$$

(3)

where: $a = e^{\frac{-R_r}{L_r}T}$; $b = \frac{n_p}{L_r}$; $c = \frac{(1-a)L_r}{DR_r}$; $d = \frac{T}{D}$; T – sampling period; u – stator voltage in the rotating reference frame.

Assuming that the rotor inductance L_r and the lower bound of the inertia torque D are known, the equations of the adaptive discrete-time regulator are obtained:

$$u_k = e^{J\rho_k^d}\left(\frac{\Psi_{k+1}^d}{1-\hat{a}}\cos(\beta_k) - \frac{\hat{a}}{1-\hat{a}}\Psi_k^d\frac{L_r}{\Psi_k^d n_p}\tau_k^d\right),$$
$$\rho_{k+1}^d = \rho_k^d + \beta_k,$$
$$\beta_k = \arcsin\left(\frac{L_r(1-\hat{a})}{n_p\Psi_{k+1}^d\Psi_k^d}\tau_k^d\right),$$
$$\tau_k^d = C(z)(\omega^d - \omega_k),$$
$$C(z) = K_P + K_I\frac{z}{1-z},$$

(4)

where: Ψ^d – desired value of amplitude of rotor flux; τ^d – desired value of torque; ω^d – desired value of rotor velocity; $a = e^{\frac{-\hat{R}_r}{L_r}}$; $\hat{R}_r$ – rotor resistance estimation value, obtained, for example, using the DREM method [4]; K_P, K_I – values of proportional and integral coefficients of the torque regulator.

Replacing in equations (4) the value of rotor velocity by its estimation obtained using an observer, the algorithm of which is given in [4, 5], it turns out that the algorithm (4) does not require direct information about the rotor velocity, and thereby becomes sensorless.

In Figure 1 shows the results of computer modeling in MATLAB/Simulink environment of the algorithm of sensorless control of induction motor. The following values of motor parameters were used: $L_s = L_r = 420$ mH, $L_m = 117$ mH, $R_s = R_r = 2.76\,\Omega$, $D = 0.06$ kg*m², $\Psi^d = 0.0455$ Wb, $\omega^d = 50$ rad*s⁻¹, $n_p = 2$. The sampling period T was taken as 1 second and the load torque as 0.1 N*m. The modeling was performed in scenarios with parametric and signaling uncertainties.

3 Conclusion
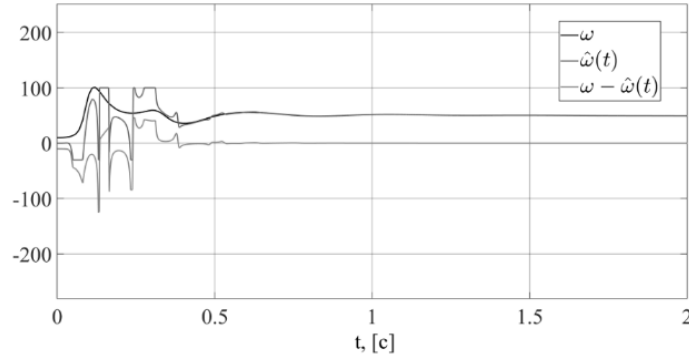
Fig.1 Modeling results in sensorless mode

In the result of this work, a sensorless adaptive control algorithm for a discrete induction motor model was synthesized. The proposed algorithm has demonstrated good performance under conditions of different kinds of uncertainties of the motor model. Various applications in robotic tasks, where inaccurate knowledge of the values of the model parameters is possible, under noise of the measured electrical signals, as well as the impossibility of using the rotor position sensor, are highlighted as prospects for the use of the algorithm. In the future it is planned to tighten some of the introduced assumptions, as well as experimentally verify the presented results on real technical systems.

References:

[1] Pyrkin A., Bobtsov A., Vedyakov A., Ortega R., Vediakova A., Sinetova M. A flux and speed observer for induction motors with unknown rotor resistance and load torque and no persistent excitation requirement // Intern. Journal of Adaptive Control and Signal Processing. 2021. Vol. 35, N 8. P. 1578–1593.

[2] Romeo O., Damia T. On Discrete-time Control of Current-fed Induction Motors // System & Control Letters. – 1996. – Т. 28, №3. – С. 123-128.

[3] Ngyen Q. C., Pyrkin A. A., Golubev A. K. Sensorless control algorithm for induction motor using adaptive observers. Journal of Instrument Engineering. 2024. Vol. 67, N 7. P. 551–558 (in Russian). DOI: 10.17586/0021-3454-2024-67-7-551-558.

[4] Pyrkin, A., Q. C. Nguyen, M. Sinetova, R. Ortega, and A. Bobtsov. 2024. "Adaptive Field-Oriented Control of Current-Fed Induction Motors: A Discrete-Time Implementation." International Journal of Control, July, 1–11. doi:10.1080/00207179.2024.2368056.

[5] Pyrkin A., Bobtsov A., Vedyakov A., Ortega R., Vedyakova A., Sinetova M.M. DREM-Based Adaptive Observer for Induction Motors // Proceedings of the IEEE Conference on Decision and Control (CDC), 2019, pp. 648-653.

Experience in Developing A Technology of Precision Spherical Units Manufacture

M.A. Tit[1], O.S. Yulmetova[1,2], A.G. Scherbak[1]

(1.Concern CSRI Elektropribor, JSC, St. Petersburg,2.ITMO University, St. Petersburg)

Abstract: The article presents the experience of developing the advanced technological solutions for correcting imbalance and applying a contrast pattern to the surface of precision spherical units used in gyroscope instrument making.

1　Introduction

Spherical units are the basis for the most promising products in gyro instrument making [1,2]. Strict requirements for shape accuracy and imbalance are imposed on these units. An equally important technological task is the process of applying contrasting patterns that determine the operation of the optoelectronic data readout system [1]. The originality of the approaches presented in the paper consists in solving the above tasks through the use of laser processing methods.

2　Spherical units manufacturing technology

A spherical unit is a precision beryllium sphere with a pronounced moment of inertia created by reinforcing elements formed in the unit body [3], with an imbalance rated by hundredths of a micrometer, and a contrasting pattern marked on the surface of the unit. The spherical units manufacturing technology includes a number of successive operations that determine the redistribution of masses and the structural and phase modification of the material. These processes are described rather completely using the chemical thermodynamics approach [4, 5].

At the same time, traditional balancing methods based on the use of directed finishing [6] do not always provide the required imbalance and accuracy of the shape of spherical units.
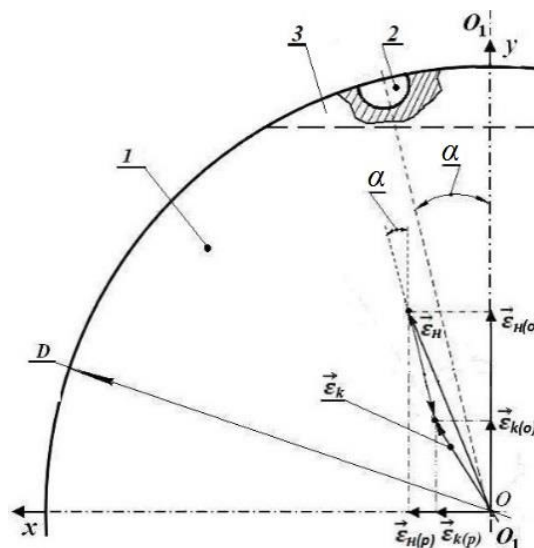
Fig.1 Imbalance correcting scheme: 1 – spherical unit, 2 – recess, 3 – spherical segment where the recess is made.

In this paper, the imbalance correction was performed by local evaporation of a point mass from the surface of a spherical unit [7] until the required final imbalance $\overrightarrow{\varepsilon_k}$ was obtained, at an angle α of the recess axis tilt to the symmetry axis O_1O_1 of the spherical unit 1 (Fig. 1). In this case, the conditions are provided: laser evaporation of a given mass m and formation of a point recess 2 are performed outside the raster pattern zone, and the mass m is determined from the relationship

$$m = 2M \frac{\varepsilon_н(o) - \varepsilon_k(o)}{D \cos \alpha}$$, where M is the mass and D is the diameter of the spherical unit, $\vec{\varepsilon}_н(o)$

$\overrightarrow{\varepsilon_{H(0)}}$ and $\vec{\varepsilon}_k(o)$ – are the axial components of the initial $\vec{\varepsilon}_н$ and final $\vec{\varepsilon}_k$ imbalance vectors, respectively. An important factor is the determination of the recess position specified by the

inclination angle α, defined by the relation: $\alpha = \text{arctg} \frac{\varepsilon_н(p) - \varepsilon_k(p)}{\varepsilon_н(o) - \varepsilon_k(o)}$, where $\overrightarrow{\varepsilon_{H(p)}}$ and $\vec{\varepsilon}_н(p)$

are the radial components of the initial and final imbalance vectors, respectively.

The final operation in the beryllium spherical unit manufacturing cycle is the generation of a contrast pattern. The evolution of this process can be represented as follows: the electrochemical method (before 2012) [8, 9], limited in the attainable optical parameters of the pattern and reducing the accuracy of the spherical unit shape, and method of laser marking (from 2012 to the present day) [10, 11].

Laser technology has provided a number of advantages in terms of the quality of the applied pattern, including uniformity of the contrast of the resulting pattern, as well as minimization of the spherical unit deformations, as compared to the electrochemical method [7].

In this case the generation of a pattern on the beryllium surface is well described by the chemical thermodynamics approach [3], which can serve as the basis for a descriptive model of the laser treatment process. Modeling is associated with the calculation of the Gibbs energy values for possible interactions in the "metal surface – gas environment components" system:

$$\Delta Z_T^0 = \Delta H_{298}^0 - T \cdot \Delta S_{298}^0 + \int_{298}^T C_p dt - T \cdot \int_{298}^T 1/T \Delta C_p dt + RT \ln P_g^{-k} \qquad (1)$$

where C_p – heat capacity; T – temperature; P_g – relative partial pressure of the gas environment component g taking part in the reaction, R – absolute gas constant, k – stoichiometric factor.

Comparison of the Gibbs energy values and selection of the chemical reaction with the lowest value of this factor allows us to determine the phase composition of the compound that forms a contrast pattern in the process of laser treatment on the spherical unit surface [2, 3]. Varying the partial pressures of the gas environment components during laser treatment allows us to change the surface conductivity of the spherical unit in the pattern area, which is illustrated by the differential conductivity curves obtained by scanning probe microscope investigation (NT-MDT Nanoeducator) (Fig. 2). The analysis of the Gibbs energy values shows that after laser marking of beryllium in a nitrogen environment, the composition of the laser raster is determined by the Be_3N_2 nitride [5], which is completely consistent with the X-ray phase analysis data (Fig. 3).
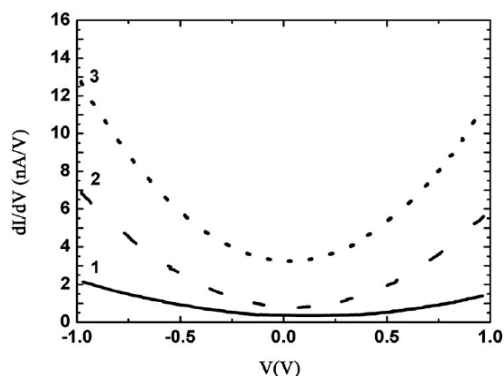
Fig.2 Curves of differential conductivity of a laser pattern generated in the air environment (1), in nitrogen environment (2), and on the base surface of beryllium (3)
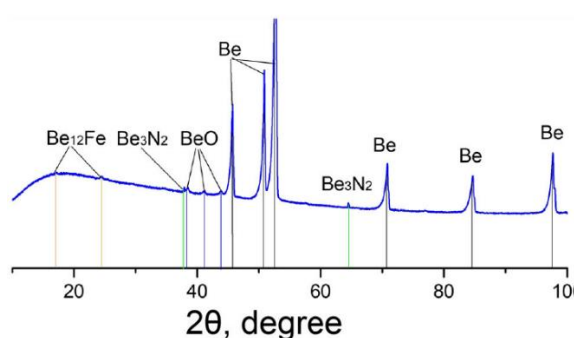
Fig.3 X-ray diffractograms of a laser pattern obtained in the process of treatment of the beryllium spherical unit in nitrogen environment

3 Conclusion

The example of a precision spherical unit shows the efficiency of applying the laser treatment methods in generation of the most important functional parameters of the unit during its manufacture. Mathematical models describing the relationship between the parameters of the spherical unit and the geometrical arrangement of the evaporated mass are used as tools for modeling of technological processes for correcting the imbalance. The use of a chemical thermodynamics approach for studying contrast patterns obtained by means of laser surface treatment is proposed.

References:

[1] Landau, B.E., Elektrostaticheskii giroskop so sploshnym rotorom (Electrostatic Gyroscope with a Solid Rotor), St. Petersburg: CSRI Elektropribor, 2020, 150 p.

[2] Landau, B.E. et al., 2021 J. Phys.: Conf. Ser. 1864012008

[3] Belyaev, S.N., Landau, B.E., Butsyk, A.Ya., et al., Method of ball gyro rotor manufacturing, Patent RU 2286535, 2006.

[4] Yulmetova, O.S., Scherbak, A.G. Contrast image formation based on thermodynamic approach and surface laser oxidation process for optoelectronic read-out system, *Optics and laser technology*. 2018. Vol. 101. pp. 242–247.

[5] Yulmetova, O.S., Scherbak, A.G., Composition analysis of thin films formed on beryllium surfaces under pulsed laser action by the method of chemical thermodynamics, *International Journal of Advanced Manufacturing Technology*. 2018. Vol. 97. No. 9–12. pp. 3231–3236.

[6] Filippov, A.Yu., Fedorovich, S.N., Shcherbak, A.G. Integrated approach to the technology of manufacturing electrostatic gyro rotors, *Journal of Physics: Conference Series,* 13th Multiconference on Control Problems, MCCP *2020*, 2021, p. 012023

[7] Yulmetova, O.S., Scherbak, A.G., Filippov, A.Yu., Landau, B.E., et al., Method of electrostatic gyro rotor manufacturing, Patent RU 2592748, 2016.

[8] Agroskin, B.N., Gallai, V.I., et al., Comparative assessment of electrochemical and photochemical methods of forming a light-contract image on the rotor of a gimballess electrostatic gyro, *Giroskopiya i navigatsiya*, 1996, no. 3, pp. 39–45.

[9]   Yulmetova O.S., Tit M.A., Sisyukov A.N., Yulmetova R.F. Comparative analysis of electrochemical and laser marking for gyroscopic application. *Proc. of the 2020 IEEE Conference of Russian Researchers in Electrical and Electronic Engineering* (EIConRus), 2020, pp. 1053–1055.

[10]   Yulmetova, O.S., Scherbak, A.G., Butsyk, A.Ya., Landau, B.E., Veiko, V.P., and Valetov, V.A., Method of electrostatic gyro rotor manufacturing, Patent RU 2498224, 2013.

[11]   Fedorovich, S.N., Filippov, A.Yu., Shcherbak, A.G., and Leonova, T.G., Method of electrostatic gyro rotor manufacturing, Patent RU 2812538, 2024.

Development of A Modular Stand for Testing Quantum Sensor Optical Cells Based on A Reference Standard

M.A. Vlasova, V.V. Chalkov

(Concern CSRI Elektropribor, JSC, St. Petersburg)

Abstract: The paper presents the results of development of a modular stand for testing optical cells used in quantum sensors. The test stand development is aimed at providing the high accuracy of positioning and efficient testing of the cells during the laboratory research. The testing results show the effectiveness of the proposed approach and its potential for further research in the field of quantum technologies.

1 Introduction

Quantum sensors include both advanced [1 - 2] and proven units [3 - 4] designed to perform high-accuracy measurements of various physical quantities, such as magnetic field, angular rate, time and frequency. These units include quantum magnetometers, atomic frequency standards and gyroscopes, which use the quantum properties of atoms to achieve the exceptional accuracy and stability. Although some quantum sensors are still under development, their present characteristics show significant advantages in the dedicated applications [6, 7]. However, a number of quantum sensors still require further research. The key element of quantum sensors is the cells with a working agent, in which atoms interact with electromagnetic radiation. The quality of such cells directly affects the accuracy and stability of measurements that makes their testing an important stage in the development of quantum sensors. At present, the cells production technology and selection of vapors, optimal for particular applications, remain experimental. Parameters of cells in different batches can vary significantly, and this requires strict quality control and evaluation of their characteristics. One of the efficient non-destructive methods for the cells quality control and assessment of their characteristics is the measurement of the optical resonance width [8, 9]. The use of classical optical equipment [10, 11] for testing such cells can be expensive and difficult, especially in conditions of unstable or non-standard cell geometry.

The purpose of this work is to develop a test stand for comparative analysis of optical cells using a modular structure. The proposed stand allows for non-destructive quality control of cells manufacturing and filling by comparing the reference and tested cells [12]. Additive technologies in developing the stand made it possible to apply a concept of modular design thus ensuring flexibility of adapting the stand to different types of cells. If it is necessary to test a non-standard cell, a module conforming to its size and shape can be quickly developed and printed, providing independent temperature control for the cell thus minimizing the costs and time for preparing to the experiment not decreasing stability of measurements.

2 Spectroscopic analysis of vapor cells

The spectroscopic analysis of the alkali metal absorption line is an efficient method for non-destructive testing of vapor cell parameters [13], because it allows us to identify defects that occur

during their filling, such as the absence of buffer gas, its insufficient amount, or damage to the cell integrity, while leaving the cell suitable for use [14]. In this study, the cell is tested using a laser with a wavelength of 895 nm, corresponding to an optical transition in cesium 133 atoms. When light interacts with cesium atoms, resonant absorption of radiation occurs, associated with transitions between hyperfine levels [15-17].

Energy transitions in cesium 131 atoms have a hyperfine splitting of 9.2 GHz for the ground state and 1.2 GHz for the excited state, which creates four clear resonance lines. The increased buffer gas pressure in the test cell as compared to the reference cell reduces the time of cesium atoms relaxation due to collisions with buffer gas atoms, which in turn leads to a broadening of the resonance line.

Evaluation of the resonance line broadening relative to the reference cell allows us to judge the quality of filling the cell under study. The use of a reference cell makes it possible to carry out such measurements without the use of expensive optical spectrum analyzers.

3 Test stand description

For the approach described above, a stand was designed, which included two optical cells: one serves as a reference cell, and the other is a test sample. Both cells are illuminated by a single laser source, which ensures identical conditions for measurements. The light flux is divided by a beam splitting cube with a factor of 90/10, where 10% of the light is directed to the reference cell, and the remaining 90% – to the test cell. Such splitting was chosen because the shape of each test cell can vary, increasing the light loss.

Due to the expected differences in gas composition, the cells require different temperature conditions to achieve a similar concentration of alkali metal. This is achieved by the modular design of the stand: each cell is located in its own module with independent temperature control, which allows the required temperature to be set for each cell. A vertical-cavity surface-emitting laser (VCSEL laser L895VH1) is used as a light source. A lasing wavelength can be controlled by changing both the temperature and the current. To simplify setting the wavelength of the laser, it is also located in a module with independent temperature control, which allows us to change the wavelength by varying only the strength of the injection current.

It should be noted that the VCSEL family of light sources have low power and light beam concentration compared to emitters with an external resonator, which restricts the accuracy of positioning the cells, laser cube and photodetector on the same optical axis, with an acceptable deviation of no more than 0.5 mm.

A photodetector is installed behind each optical cell, which records the intensity of the transmitted light. The signals received from the photodetectors are transmitted to the data analysis device. A schematic representation of the stand is shown in Figure 1.

Since the concentration of the alkali metal in the cell directly depends on its temperature, and the lasing wavelength is subject to temperature changes, the testing process requires reliable temperature protection against the environment and easy access to all elements of the stand. Therefore, a material with high heat capacity was chosen for manufacturing the stand. The stand is designed with provision for testing cells of various shapes without the need for additional modification of the whole stand; to test non-standard cells, it is enough to replace the module for

the cell being tested with a suitable one. To ensure such functionality, separate modules are integrated into the main body of the stand, each of the modules having independent temperature control.
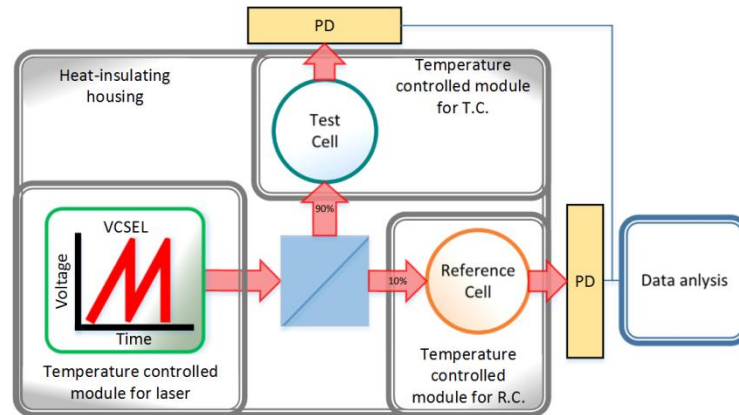
Fig.1   Schematic view of the stand for testing optical cells: R.C. – reference cell; T.C. – tested cell; VCSEL - vertical-cavity surface-emitting laser; PD - photodetector

4 Experiment procedure

Injection current, changing according to a sawtooth law within the range of 1.2-1.4 volts, is supplied to the laser heated to the operating temperature. As the cells heat up, the concentration of alkali metal vapors increases, and when the absorption peak is reached, equal to 60-100% of the amplitude of the laser scanning signal for the tested cell and 160-190% for the reference cell, the signals from the photodetectors are recorded. These values are valid only for the checked cells and should be determined experimentally based on the saturation of cesium in the cells, which indicates the termination of the absorption peak increasing during resonance broadening.

The signals recorded by the photodetectors represent the dependence of the intensity of light, which has passed through the cell and interacted with the substance in it, on the laser injection current (on the laser radiation frequency). An example of such signals is shown in Figure 2. These signals are used to evaluate the characteristics of the cells being tested. First, the stand allows us to find the cells that were not filled, which indicates a manufacturing defect. For cells successfully filled with the necessary gases, the stand measures the magnitude of the shift of absorption peaks and the half-width of the resonances in the cell being tested relative to the reference cell. These characteristics are subsequently taken as parameters when using the cells.
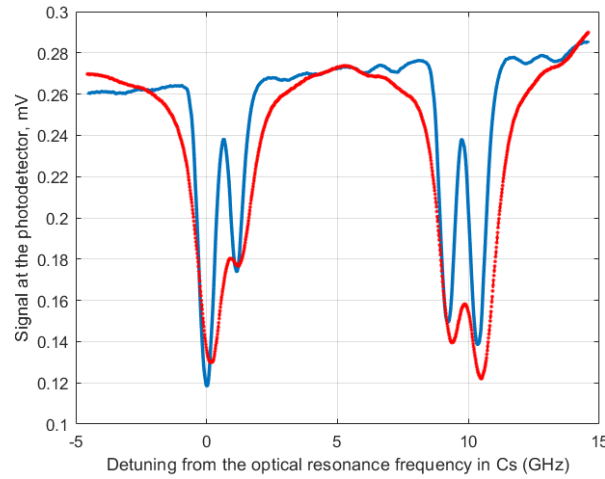
Fig.2 Results of testing the cells: solid curve — a signal obtained when scanning through the reference cell; dashed curve — a signal received from the cell being tested

The processing of signals, an example of which is shown in Fig. 2, is performed using approximation by four Lorentz profiles [18, 19], taking into account the dependence of the intensity of light emitted by the laser on the injection current:

$$S(v) = V + K_P L(v - v_0) - K_P K_M \left(\frac{\Gamma_{2,1} A_1}{(v-v_1)^2 + \Gamma_1{}^2} + \frac{\Gamma_{2,2} A_2}{(v-v_2)^2 + \Gamma_2{}^2} + \frac{\Gamma_{2,3} A_3}{(v-v_3)^2 + \Gamma_3{}^2} + \frac{\Gamma_{2,4} A_4}{(v-v_4)^2 + \Gamma_4{}^2} \right) + V_{ns}$$

（1）

where v is the laser injection voltage; V is the base voltage of the laser injection; K_p is the power/lasing wavelength coefficient; L is the signal slope coefficient; K_m is the coefficient of the laser injection current conversion into the laser radiation frequency; v_n are currents at which the appropriate resonances are achieved; $\Gamma_{2,n}$ are the relaxation constants of transverse magnetization; Γ_n are sought-for half-widths of the corresponding resonances; A_n are amplitudes of the corresponding resonances; V_{ns} is the noise component, which is a white noise; n = 1, 2, 3, 4.

5 Benefits of additive manufacturing

The accuracy of positioning the optical elements of the stand, taking into account the described restrictions on tolerances, is a complex and high-technology task when using subtractive manufacturing methods, especially without additional adjustment during the experiment. In this context, photopolymer printing provides the required accuracy, reaching values within 0.1 mm, which allows for an efficient and fast implementation of the modular design of the stand.

6 Conclusions

The paper presents a modular stand for testing optical cells used in quantum sensors, which provides adjustment flexibility and quick replacement of components. An experiment for determining the broadening of the resonance line in a vapor cell relative to the reference cell is described. It is noted that due to the modular approach, the stand maintains high positioning accuracy of optical elements (up to 0.1 mm), and remains cost-effective regardless of the cell geometry and simplifies the stand setup for different experimental conditions.

References:

[1] Peshekhonov, V.G., The outlook for gyroscopy, Gyroscopy and Navigation, 2020, vol. 11, no. 3, pp. 193–197. DOI 10.17285/0869-7035.0028.

[2] Popov, E.N., Barantsev, K.A., and Litvinov, A.N., Theoretical simulation of a signal for a scheme of an atomic spin gyroscope with optical detection, Kvantovaya Elektronika, vol. 49, no. 2, 2019, pp. 169–177.

[3] Ding, M. Development and prospect of quantum sensing technology, Proceedings of the 20th Conference of Young Scientists with International Participants "Navigation and Motion Control", Eds.: Prof. Peshekhonov, V.G., and Prof. Stepanov O.A., St. Petersburg: Concern CSRI Elektropribor, 2018, 443 p.

[4] Petrov, A.A., Methods for improving accuracy characteristics of quantum frequency standards, RIRT, 2021.

[5] Basov, N.G. and Belenov, E.M., Ultra-narrow spectral lines and quantum frequency standards, Priroda, 1972, no. 12.

[6] Basov N.G. and Gubin M.A. Quantum frequency standards, IEEE Journal of Selected Topics in Quantum Electronics, 2000, vol. 6, no. 6, pp. 857-868.

[7] Alexandrov, E.B. et al., Gradiometric study of a quantum magnetometer on cesium vapors, Zhurnal tekhnicheskoi fiziki, 2007, vol. 77, no. 3, pp. 100-102.

[8] Nwaboh, J.A. et al., Optical path length calibration: a standard approach for use in absorption cell based IR spectrometric gas analysis, International Journal of Spectroscopy, 2014, vol. 2014, no. 1, p. 132607.

[9] Yulmetova, O.S. et al., Analysis of chemical interactions during filling a cesium vapor cell for a quantum magnetometer, Nauchno-tekhnicheskii vestnik informatsionnykh tekhnologii, mekhaniki i optiki, 2024, vol. 24, no. 3, pp. 438-447.

[10] US20160349481A1. Monolithic optical mounting element. United States Patent and Trademark Office. Published on 2016-12-01

[11] Thorlabs, Inc. (2024). Method of making a pre-aligned optical mount (U.S. Patent No. US11982861B2). U.S. Patent and Trademark Office

[12] Đujić, M. et al. Comparative analysis of light storage in antirelaxation-coated and buffer-gas-filled alkali vapor cells, Scientific Reports, 2024, vol. 14, no. 1, p. 14467.

[13] Bulgakova, V.V. et al., Experimental stand for optical pump-terahertz probe diagnostics for studying ultra-fast dynamics of charge carriers in photoconductive media, ITMO University Conference "Nevsky Photonics-2023", St. Petersburg, 2023.

[14] Shevchenko, A.N., Kuzmin, A.G., and Titov, Yu.A., Mass-spectrometric measurement of the gas mixtures composition in the cells of a quantum rotation sensor, Nauchnoe priborostroenie, 2018, vol. 28, no. 2, pp. 62-68.

[15] Boldyrev, K.N. et al., The first observation of the hyperfine level's anticrossing in the optical spectra, 13th International Workshop on Quantum Optics (IWQO-2019), 2019, pp. 49-50.

[16] Sargsyan, A.G. et al., Investigation of atomic transitions of cesium in strong magnetic fields by an optical half-wavelength cell, Pis'ma v zhurnal eksperimental'noi i teoreticheskoi fiziki, 2013, vol. 98, no. 8, pp. 499–503.

[17] Alexandrov, E.B., Vershovskiy, A.K., and Pazgalev, A.S., Magnetometer based on a pair of symmetric transitions in the 87Rb hyperfine structure, Technical Physics, 2006, vol. 76, no. 7, pp. 103-107.

[18] Vershovskiy, A.K., Dmitriev, S.P., and Pazgalev, A.S., Optically pumped quantum magnetometer with combined advantages of Mx and Mz devices, Technical Physics, 2013, vol. 58, no. 10, pp. 1481–1488.

[19] Vlasova, M.A. et al., Special aspects of the VCSEL lasers use in an installation for monitoring the parameters of quantum sensors gas, 2022.

Smoothing Filter for Micromechanical Gyroscope Unstable Bias Estimation in Dynamic Calibration Experiment

S.A. Fedorov, F.S. Kapralov[1]

(Lomonosov Moscow State University, Moscow)

Abstract: Smoothing filter in Rauch-Tung-Striebel form is applied to calibration of a 6-axis low-precision microelectromechanical inertial sensor to obtain more accurate estimates of unstable low-frequency stochastic gyroscope error component. Real experimental data processing indicates that estimates of the unstable gyroscope bias are close to those obtained with the direct method of averaging gyroscope measurements in pre-defined identical static positions. The proposed method, however, does not require any static positions, thus simplifying the calibration experiment while retaining possibility to analyze low-frequency sensor error variation.

1 Introduction

Microelectromechanical inertial sensors (MEMS) have many advantages such as low power consumption, compactness and low cost. However, accuracy characteristics of such sensors are relatively low in most cases, and use of MEMS sensors in applications requires more complex bias model than a constant value. This applies to calibration experiments as well. The goal of calibration is to estimate parameters of the a priori model of the systematic sensor errors. In 2005, N.A. Parusnikov (Lomonosov Moscow State University) proposed a method for calibrating inertial sensors measurement unit with an algorithm based on the estimation of a linear dynamic system state vector using Kalman filter [1,2]. As in other techniques, quality of calibration parameter estimates can be affected by unstable low-frequency component of gyroscope measurements. This paper considers a modified version of the Parusnikov's method [1,2], based on the application of the Rauch-Tung-Striebel (RTS) smoother [3], which uses the entire set of measurements to estimate the state vector at each time instant. It is expected to provide more accurate estimates for the unstable component of gyroscope measurements over time, as well as less biased estimates for other parameters affected by it.

2 Estimation problem for calibration parameters

We use the following a priori models for instrumental errors of accelerometers and gyroscopes:

$$f' - f \equiv \Delta f = \Delta f^0 + \Gamma f + \Delta f^s, \qquad \omega' - \omega \equiv \nu = \nu^0 + \Theta\omega + \nu^s \qquad (1)$$

where f' is a triplet of accelerometer measurements, f contains true components of the measured specific force, Δf^0 is an accelerometer bias, Γ is a matrix of scaling coefficients and sensitive axis misalignments, Δf^s is a stochastic component of accelerometer measurements, ω' is a triplet of gyroscope measurements, ω contains true components of the absolute angular rate, ν^0 is a gyroscope bias independent of angular velocity, Θ is a matrix of scaling coefficients and sensitive axis misalignments, ν^s is a stochastic component of gyroscope measurements.

We introduce linear dynamic system with the state vector x:

$$x = (\beta_{x^0}, \nu^0, \Delta f^0, \Gamma, \Theta) \in R^{24}, \dot{\beta}_{x^0} = \hat{u}_{x^0}\beta_{x^0} + L^T\nu, \Delta\dot{f}^0 = q_{\Delta f}, \dot{\nu}^0 = q_{\nu^0}, \dot{\Gamma} = 0, \dot{\Theta} = 0 \tag{2}$$

The calibration problem is stated as the estimation of the state vector x using measurements y, which are:

$$y \equiv L'^T f' - \begin{pmatrix} 0 \\ 0 \\ g \end{pmatrix} \approx \begin{pmatrix} 0 & g & 0 \\ -g & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \beta_{x^0} + L'^T \Delta f \tag{3}$$

where L is an attitude matrix of the instrumental frame, fixed to the accelerometers sensitivity axes, relative to the geodetic frame (East-North-Up), L' is a computed value of L, obtained by integrating kinematic Poisson attitude equation with gyroscope measurements as angular rate components, and a certain initial orientation $L'(0)$, β_{x^0} is a small rotation vector of the orientation error, g is a reference value of local gravity acceleration, $q_{\Delta f}, q_{\nu^0}$ — white noises with a priori defined variances.

The estimate of the system state vector can be obtained using either standard forward time Kalman filter [3] or smoothing filter in the RTS form [3]. The RTS form of the optimal smoother is preferred because it does not require the Kalman filter to be implemented in reverse time.

3 Results of experimental data processing
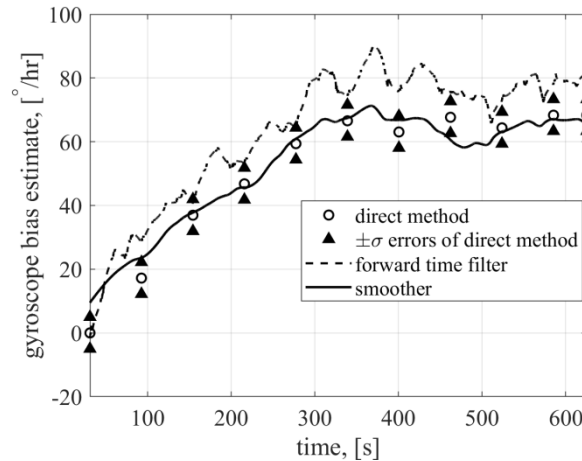
Fig.1 Gyroscope unstable bias estimates

A calibration experiment for an inertial measurement unit (STMicroelectronics, model ISM330DCHX) was performed using a single axis turntable. The experiment scheme consisted of sequential rotations around the horizontal turntable axis by $90°$, with stops at each angular position for 10 seconds. Such experiment scheme allows a direct method of estimating unstable gyroscope bias by averaging gyroscope measurements in identical static positions. Figure 1 shows a graph of gyroscope bias estimates produced by different methods. The graph indicates that the estimate obtained by the RTS smoother approximates the low-frequency component of the bias variation more accurately than the estimate obtained by the Kalman filter. Moreover, for the RTS smoother to work, it is not necessary to follow a special experiment scheme with static positions. According to simulation results, the smoothing filter also shows good performance, for example, in case of rotations with constant angular velocity.

4 Conclusion

The paper considers the problem of a 6-axis MEMS inertial sensor calibration using a smoothing filter in Rauch-Tung-Striebel form to estimate the low-frequency unstable component of gyroscope errors. The results of real and simulated data processing show the advantage of this approach over both the conventional forward-time Kalman filter and the direct method of bias estimation using averaged gyroscope measurements in pre-defined identical static positions.

References:

[1] Parusnikov N.A., Tikhomirov V.V., Trubnikov S.A. Opredeleniye instrumental'nykh pogreshnostey inertsial'noy navigatsionnoy sistemy na nepodvizhnom osnovanii [In Russian] – Fundamental'naya i prikladnaya matematika, Vol. 11, N. 7., P. 159 – 166, 2005.

[2] Vavilova N.B, Vasineva I.A., Golovan A.A., Kozlov A.V., Papusha I.A., Parusnikov N.A. – The Calibration Problem In Inertial Navigation. – Journal of Mathematical Sciences, Vol. 253, N.6, P. 818-836, 2021

[3] Dan Simon – Optimal State Estimation: Kalman, H∞, and Nonlinear Approaches. P. 123-124, 286-294

Method for The Reconstruction of Nanosatellite Angular Motion Based on Magnetometer Measurements: A Case Study of SamSat-ION

A. S. Espinoza Valles, I. V. Belokonov, I. A. Lomaka, P. N. Nikolaev

(Samara National Research University, Samara)

Abstract: This study presents a method for reconstructing the angular motion of a nanosatellite using magnetometry measurements. The core concept of the method is based on using Earth's magnetic field measurements for approximate estimation of orientation quaternions and angular velocities of the nanosatellite. The application of the proposed method made it possible to reconstruct the angular motion of the SamSat-ION nanosatellite during flight tests.

1   Introduction

The reconstruction of spacecraft attitude motion is essential for ensuring operational functionality, particularly for missions involving scientific observations and communication. Numerous approaches have been proposed for attitude determination, utilizing data from solar panel current and sun sensors [1, 2]. Using magnetometer data for in-flight attitude estimation is especially relevant for small spacecraft with limited power and computational resources, as confirmed by multiple studies [3, 4]. This paper is devoted to a technique for nanosatellite angular motion reconstruction based on magnetometer measurements. As a result, the angular motion of SamSat-ION was successfully reconstructed during all communication sessions with the nanosatellite.

2   Methodology for the Reconstruction of Nanosatellite Angular Motion.

The behavior of nanosatellite angular motion can be approximately reconstructed by using the direction of the local magnetic field vector measured by the onboard magnetometer. The process of reconstructing the angular motion can be divided into two stages. The first stage involves preliminary correction of magnetometry measurements, while the second stage includes the estimation of orientation parameters: initial quaternion conditions and angular velocity.

3   Correction of Magnetometer Measurement Errors

Before using magnetometer measurements for angular motion estimation, it is necessary to eliminate caused by the magnetometer bias vector bmag. To estimate the bias, the difference between the magnitude of the magnetometer measurement vector, after subtracting the bias, and the magnitude of the corresponding reference measurement vector is minimized. The calculated magnetic field measurements are derived from the IGRF-13 model, corresponding to the orbital trajectory. The orbital motion is estimated using the SGP4 model with TLE data available for the nanosatellite under consideration.

4   Estimation of the Initial Conditions for Angular Motion

It is necessary to align the reference magnetic field measurements with the nanosatellite's orientation. Based on the initial conditions being determined, the angular motion is predicted according to the dynamic Euler equations and kinematic equations. The calculated reference

measurements are then transformed into the nanosatellite's body-fixed coordinate system (BFCS) using a rotation matrix composed of the predicted quaternions. Taking into account the estimated magnetometer bias vector, the initial conditions of the satellite's attitude motion are determined by minimizing the following objective function:

$$\Phi(\mathbf{q}_0, \boldsymbol{\omega}_0) = \sum_{i=1}^{N}\left(1 - \frac{\left(\mathbf{B}_{meas_i} - \hat{\mathbf{b}}_{mag}\right)\cdot\mathbf{B}_{ref_i}(\mathbf{q}_0, \boldsymbol{\omega}_0)}{\left\|\mathbf{B}_{meas_i} - \hat{\mathbf{b}}_{mag}\right\|\left\|\mathbf{B}_{ref_i}(\mathbf{q}_0, \boldsymbol{\omega}_0)\right\|}\right) \rightarrow \min_{\mathbf{q}_0, \boldsymbol{\omega}_0} \quad (1)$$

where $\mathbf{q}_0$ and $\boldsymbol{\omega}_0$ are the initial conditions for the orientation quaternion vector and angular velocity vector; $\mathbf{B}_{meas}$ is the measurement vector from the onboard magnetometer; $\mathbf{b}_{mag}$ is the bias vector of the magnetometer; $\mathbf{B}_{ref}$ is the vector of calculated magnetometer reference measurements in the BFCS; N is the number of measurements available in a communication session.

The optimal solution to the objective function (1) can be found using methods such as the differential evolution algorithm, which offers a high degree of convergence.

5   Application of the Methodology for the Reconstruction of Nanosatellite Angular Motion
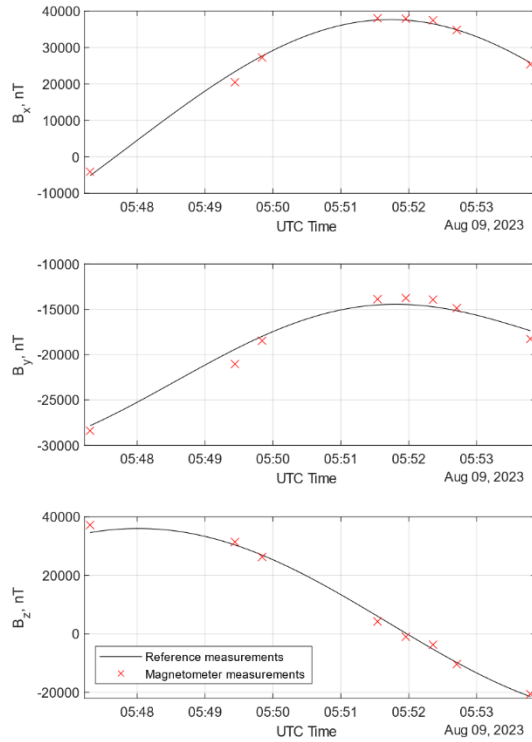
Fig. 1. Attitude estimation   **Fig. 2**. Angular velocity estimation
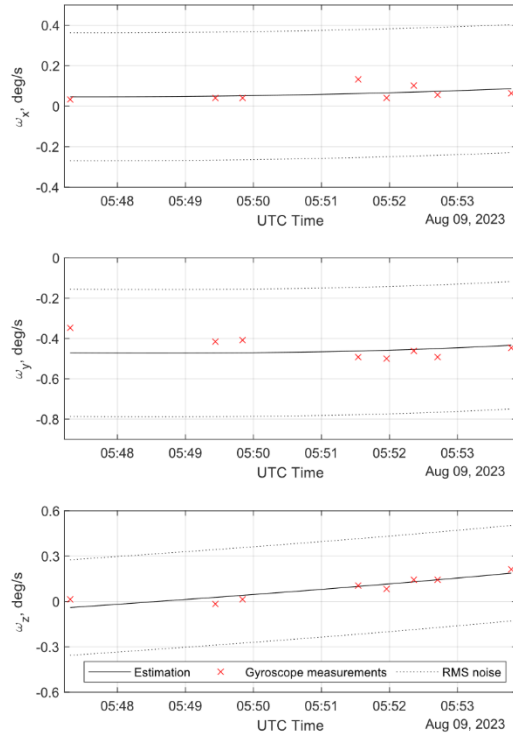
SamSat-ION is a CubeSat 3U developed at Samara University and launched in June 2023 to study the Earth's magnetosphere and ionosphere. The SamSat-ION attitude determination and control system is based on active magnetic stabilization and includes several measurement instruments, such as MMC5883 magnetometers, MPU-9255 and ADXRS290 gyroscopes, and TCS3472 sun sensors. The proposed methodology was applied to reconstruct the relative motion of SamSat-ION using scientific telemetry obtained during flight testing. The initial conditions of the

orientation quaternion and angular velocity were estimated for all communication sessions. Fig. 1 shows the orientation estimation results as the alignment between the magnetometer measurements and the calculated values. Fig. 2 presents the absolute angular velocity estimation results, showing the alignment between gyroscope measurements and the reconstructed angular velocity.

As seen in the graphs, the sensor measurements align well with the calculated values. Since the calculated sensor values are directly related to the nanosatellite's orientation, it can be concluded that the angular motion has been reconstructed with high accuracy. The angular velocity magnitude of the SamSat-ION remained within the range of 0.41–0.55 deg/s during all analyzed communication sessions.

4 Conclusion

This work presents an approach to reconstructing nanosatellite angular motion based on magnetometer measurements. Using the proposed approach, the angular motion of the SamSat-ION nanosatellite was successfully reconstructed during flight tests.

References:

[1] Belokonov, I.V., Kramlikh, A.V., Lomaka, I.A., Nikolaev, P.N. Reconstruction of a spacecraft's attitude motion using the data on the current collected from solar panels [J]. Journal of Computer and Systems Sciences International, 2019, 58(2): 286–296.

[2] Belokonov, I.V., Lomaka, I.A. Postflight recovery of the rotational motion of a small space vehicle from solar sensor information [J]. Journal of Computer and Systems Sciences International, 2023, 62(2): 214–224.

[3] Ivanov, D.S., et al. The results of flight tests of an attitude control system for the Chibis-M microsatellite [J]. Cosmic Research, 2014, 52(3): 205–215.

[4] Ivanov, D.S., et al. Angular motion of the TNS-0#2 nanosatellite after launch from the International Space Station [J]. Cosmic Research, 2019, 57(4): 272–288.

Parameter Selection for Magnetic Attitude Control of Small Satellites in Sun-Synchronous Orbit

T.O. Berger, I. A. Lomaka

(Samara National Research University named after Academician S. P. Korolev, Samara)

Abstract: A methodology for selecting parameters of passive magnetic attitude control systems to ensure the required orientation of small spacecraft in sun-synchronous orbits is presented. This methodology is implemented in dedicated software that evaluates the angular velocity damping time, optimal placement and quantity of magnetic materials, and models the attitude dynamics of small spacecraft with a passive magnetic stabilization system.

1 Introduction

The rapid growth in the number of small satellites, especially those built to the CubeSat standard, over the past decade is driven by their numerous advantages, such as shorter development timelines, cost-effective manufacturing and launch options, and a wide range of potential uses. These characteristics have captured the attention of not only research and educational institutions but also private companies, with the latter now offering services such as Earth remote sensing and satellite-based communication using small satellites. As applications expand and mission requirements become more sophisticated, demands on satellite functionalities and subsystems are higher. This particularly affects attitude control systems, as the majority of scientific research with nanosatellites requires specific spatial orientation [1].

For missions where a pointing accuracy of 10–15 degrees is sufficient, passive control systems that exploit the natural interaction between the spacecraft and external environmental torques are often the optimal choice. Among those, passive magnetic attitude control systems (PMACS) are particularly popular for small satellite missions. By incorporating permanent magnets and hysteresis dampers on board, these systems generate torque through interaction with Earth's magnetic field, aligning the spacecraft with the geomagnetic field vector. With no moving parts, PMACS are highly compatible with the size, weight, and power constraints of small satellites, providing a reliable and simple implementation [2].

Despite its numerous advantages, the complexity of magnetic forces and their behavior in space conditions result in comprehensive modeling of spacecraft angular motion essential yet challenging. Although the literature provides several cases of successful PMACS implementations, it offers limited insight into the criteria guiding key decisions, such as the sizing, positioning, and selection of magnetic materials. Understanding how these parameters influence spacecraft attitude dynamics is crucial; however, the lack of comprehensive methodologies limits the ability to design these systems optimally.

This work presents a methodology developed at the Department of Space Research at Samara University for selecting parameters of magnetic control systems that ensure the required orientation within the desirable deviation range. The methodology is applied to sun-synchronous orbits, where

179

the near-polar trajectory introduces both control challenges and leverageable characteristics. The methodology is then implemented through dedicated software designed to model the attitude dynamics and estimate the angular velocity damping of small spacecraft equipped with magnetic materials in different configurations.

2 Passive magnetic attitude control systems

In PMACS, two types of magnetic materials are used: permanent magnets and hysteresis dampers. Permanent magnets generate a torque that aligns their axes with Earth's magnetic field lines, thus providing a restoring force. However, due to changes in the magnetic field direction during orbital motion, oscillations can arise. To dampen these oscillations, the hysteresis rods are installed in the spacecraft. These changes are particularly pronounced in polar and near-polar orbits, where the satellite passes over nearly all latitudes, from the North to the South Pole, exposing it to a broad range of magnetic field intensities and directions. This results in more dynamic magnetic conditions compared to equatorial or mid-latitude orbits, where field variations are less extreme.

Sun-synchronous orbits, a specific type of near-polar orbit, are designed to pass over the same region of Earth at approximately the same local solar time each day. This timing consistency provides stable and predictable conditions for Earth observation, scientific research, and other tasks that require regular and uniform lighting conditions. Furthermore, the near-polar nature of these orbits enhances the interaction between magnetic materials and Earth's magnetic field, contributing to the effectiveness of PMACS compared to orbits with lower inclinations.

The dipole moments of permanent magnets must be selected so that their restorative action effectively counteracts external torques [3]. This is especially crucial for small satellites, whose low inertial properties make them highly susceptible to gravitational gradient effects and external forces. Additionally, at low Earth orbits, atmospheric drag must be considered, as the ballistic coefficient of nanosatellites is significantly higher than that of larger satellites with comparable volumetric density [4].

The developed methodology takes into account the dynamics of both near-polar orbits and small satellites. In the initial stage, the parameters of PMACS are determined, including the characteristics of the permanent magnet, the required attitude accuracy, and the material of the hysteresis rod. An analysis of the hysteresis loop is performed to compare the properties of various materials, after which the minimum magnetic moment is calculated, and resonance conditions are checked. The optimal placement and quantity of magnetic materials along each axis are then assessed. The methodology is implemented in specialized software that includes an internal library of key hysteresis materials commonly used in small satellite missions, while also allowing for customizable parameters where available.

The complex nature of hysteresis functions, combined with the interaction of onboard magnetic materials with Earth's magnetic field in a sun-synchronous orbit, underscores the need for an accurate methodology for effectively modeling angular motion. Improper placement of these materials can lead to ineffective attitude control, potentially jeopardizing mission success. The developed methodology provides a theoretical foundation for understanding the motion of small satellites under external forces, with onboard magnetic materials to ensure attitude control and stabilization. The specialized software serves as a tool for evaluating effectiveness and selecting

optimal system parameters, thus ensuring precise attitude control necessary for mission objectives.

References:

[1] Belokonov, I.V., Timbai, I.A., Barinova, E.V. Selection of design parameters for CubeSat-format nanosatellites with passive stabilization systems [J]. Gyroscopy and Navigation, 2020, 28(1): 1–7.

[2] Ovchinnikov, M.Yu., Roldugin, D.S. A survey on active magnetic attitude control algorithms for small satellites [J]. Progress in Aerospace Sciences, 2019, 109: Article 100546.

[3] Ovchinnikov, M., Pen'kov, V., Norberg, O., Barabash, S. Attitude control system for the first Swedish nanosatellite "MUNIN" [J]. Acta Astronautica, 2000, 46: 1–7.

[4] Belokonov, I.V., Timbai, I.A. Motion of a nanosatellite relative to its center of mass in near-Earth orbits: Study guide [C]. Samara: Samara University Press, 2020: 128.

2024

Chinese-Russian

"Navigation and Motion Control"

Youth Forum

Topic 2: Integrated Navigation and Motion Control System

Suppression of Synchronous Vibration of Magnetically Suspended Control Moment Gyroscope Rotor Based on Algebraic Derivative Frequency Estimation Method

Chen Xi[1,2], Huang Xiaoguang[1,2]

(1. Shanghai Aerospace Control Technology Institute, Shanghai 201109, China; 2. Shanghai Key Laboratory of Aerospace Intelligent Control Technology, Shanghai 201109, China)

Abstract: The high-speed rotor of magnetic levitation control moment gyroscope is inevitably unbalanced due to factors such as material processing accuracy. The unbalanced mass distribution of the rotor will produce co-frequency vibration, especially when rotating at high speed. The transmission of high-frequency vibration to the spacecraft platform through the base will seriously affect the pointing accuracy and stability of the spacecraft platform. Traditional methods for suppressing synchronous vibrations heavily rely on rotor speed signals provided by Hall sensors. However, reliability issues with Hall speed sensors can compromise the vibration suppression accuracy of magnetically suspended rotors. Therefore, the reliable acquisition of rotor speed signals is particularly crucial for low-vibration control of magnetically suspended rotors. This paper proposes a method for acquiring rotor speed in magnetically suspended rotors based on algebraic derivative frequency estimation. This method leverages signals from radial displacement sensors of the rotor for frequency estimation, thereby providing crucial speed information for an adaptive notch filter to eliminate synchronous currents generated by the control system. Simulation results demonstrate that this method does not depend on the orthogonality of displacement signals, is simple and easy to implement, and, in combination with the notch filter, can effectively suppress synchronous vibrations caused by imbalance in the rotor of a magnetically suspended control moment gyroscope.

 Key words: Frequency estimation; Unbalance control; Synchronous vibration suppression; Magnetically suspended rotor;

1 Introduction

Control Moment Gyroscope (CMG) is a key actuator for achieving three-axis attitude stabilization and maneuver control in large spacecraft such as space stations and agile maneuvering satellites[1]. Disturbances arising from factors such as mass imbalance in the high-speed rotor during gyroscope operation become the primary source of disturbance affecting the high-precision stability of the spacecraft platform and its payloads[2][3]. Magnetically Suspended Control Moment Gyroscope (MSCMG), based on magnetic bearing technology, enables non-contact support for high-speed rotors, featuring no friction, no lubrication requirements, and long service life. In particular, the adjustable stiffness and damping of magnetic bearings can actively suppress disturbances such as mass imbalance in high-speed rotors, significantly enhancing the micro-vibration level of CMGs and making them ideal actuators for achieving high-precision attitude stabilization control of spacecraft[4][5].

The disturbance signals of MSCMGs are mainly introduced by the synchronous high-speed rotor imbalance vibration related to the rotational speed. Researchers have conducted many effective studies on the suppression of unbalanced vibration in magnetically suspended high-speed rotors. Among them, sliding mode control[6], Least Mean Squares (LMS) algorithm[7], repetitive control[8], integral adaptive observers[9], and notch filters[10]-[12] have been used to weaken the control effect of magnetic bearing forces on the rotor rotating around its geometric axis, enabling the rotor to rotate around its inertia axis and eliminating unbalanced vibration forces. Currently, the notch filter is the most widely used solution, and in notch filter-based unbalanced vibration suppression, the speed information of the high-speed rotor is indispensable, and the measurement accuracy of the speed directly affects the vibration suppression effect of the magnetically suspended high-speed rotor.

The rotational speed of the gyroscope's high-speed rotor can be directly obtained by the bearing controller based on the switching Hall sensor or provided by the high-speed motor controller through a communication interface. However, both methods rely on pre-embedded Hall sensors, posing risks of speed information failure due to hardware faults, which may lead to vibration suppression failure or even magnetic bearing instability. To achieve unbalanced vibration suppression of MSCMGs in cases of speed sensor or communication failures, scholars have proposed methods such as Fourier transform, phase-locked loops (PLL), angular position differentiation, and recursive control to estimate the rotational speed[13][14]. The Fourier transform method has a large calculation volume, which is not conducive to implementation. The phase-locked loop has a slower convergence speed, affecting the overall dynamic response performance. Angular position differentiation can easily introduce noise. Recursive control is closely related to the initial given value of the rotational speed frequency, and its effect is greatly influenced by the given value. Chen et al. proposed a rotational frequency estimation method based on the Adaptive Notch Filter (ANF)[15], which extracts the synchronous displacement signals of the magnetic bearing's radial X and Y axes through the ANF and estimates the rotational frequency information based on these signals. This method is based on the assumption that the rotor displacement signals in the radial X and Y directions are completely orthogonal with the same amplitude and offset. However, in reality, due to sensor errors, there may be errors in the phase, amplitude, and offset of the rotor displacement signals in the radial X and Y directions. Li et al. proposed an improved adaptive frequency estimation method that can adaptively estimate the rotor speed using the characteristics of rotor radial displacement sensor signals[16], but it still relies on the orthogonality of radial displacement signals. Zhang et al. proposed using Hilbert transform for empirical mode decomposition and spectral analysis to obtain the rotational speed based on the high-frequency components in the displacement sensor signals[17], which has high accuracy. However, this scheme cannot be applied at all rotational speeds due to contamination of the harmonic components at low frequencies. Liu et al. proposed a method for estimating rotational speed based on a BP neural network model[18]. By pre-extracting rotor displacement signals and rotational speed signals to train the neural network, the trained neural network module can estimate real-time rotational speed based on rotational displacement signals. However, this method requires pre-offline training and has high computational overhead requirements for the controller.

To address the above issues, this paper adopts a method for acquiring the rotational speed

frequency of magnetically suspended rotors based on algebraic derivative frequency estimation using linear system model theory[19]. This method does not rely on the orthogonality of displacement signals and uses rotor radial displacement sensor signals for frequency estimation, thereby providing critical rotational speed information for filters to eliminate synchronous currents generated by the control system. It is simple and easy to implement and, in combination with a notch filter, can achieve synchronous vibration suppression of unbalanced rotor masses in MSCMGs.

The organization of this paper is as follows. In the first part, a detailed analysis of the method for acquiring the rotational speed of magnetically suspended rotors based on algebraic derivative frequency estimation is presented. In the second part, a method for suppressing synchronous vibrations of the rotor based on a notch filter using the estimated rotational speed is provided. In the third part, the effectiveness and adaptability of the proposed scheme are demonstrated through numerical and displacement signal simulations. The fourth part presents the conclusions.

2 Frequency Estimation Method Based on Algebraic Derivative

2.1 Analysis of Displacement Signals of Magnetically Suspended Rotors

The MSCMG high-speed rotor magnetic bearing suspension system consists of two pairs of X, Y radial magnetic bearings at both ends A and B, as well as a pair of axial magnetic bearings. Among them, the radial magnetic bearings are responsible for radial translation and torsional control, while the axial magnetic bearings are responsible for axial translation control. The magnetic bearing system can achieve active suspension control in five degrees of freedom. The structure of the gyroscope high-speed rotor system is illustrated in **Fig. 1**.
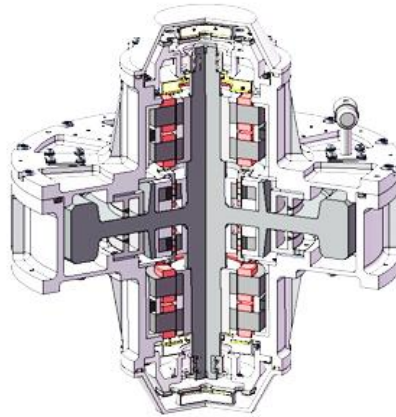
Fig.1 Structural Diagram of the High-Speed Rotor System for Magnetically Suspended Control Moment Gyroscope

The gyroscope's magnetic bearings are equipped with radial and axial displacement sensors, which can detect the displacement information of the high-speed rotor in real time. Based on the fed-back displacement information, the magnetic bearing controller completes closed-loop control to achieve stable suspension of the rotor.
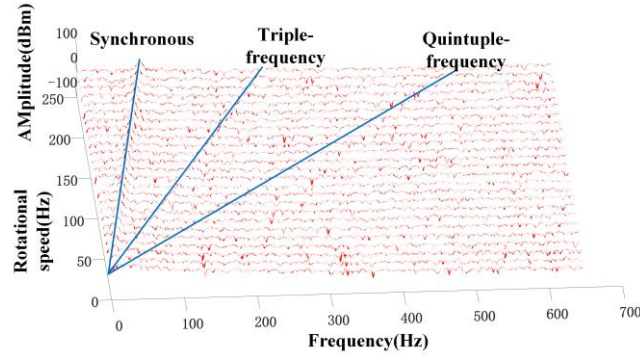
Fig.2 Waterfall Plot of the Frequency Spectrum of Magnetically Suspended Rotor Displacement Signals

Fig.2 presents the waterfall plot of the displacement signal spectrum for a certain radial channel.

It can be observed from the figure that, due to the residual imbalance in the magnetically suspended high-speed rotor, the signal spectrum is dominated by components at the same frequency as the rotational speed. Additionally, due to factors such as uneven sensor detection surfaces and uneven magnetic fields, the signal spectrum also contains harmonic components that are multiples of the rotational speed, with the third harmonic being the most prominent, and harmonics of higher orders being less significant. The output signal of a single radial displacement channel sensor can be expressed as:

$$y(t) = K + A_1 \sin(\omega t + \varphi) + A_3 \sin(3\omega t + \phi) + \xi(t) \tag{1}$$

In Equation (1), K represents the DC component of the displacement signal, A_1 represents the amplitude of the signal at the rotational speed synchronous frequency, ω represents the rotational frequency of the rotor, φ is the phase of the displacement signal at the synchronous frequency, A_3 represents the amplitude of the signal at third rotational frequency harmonic, ϕ is the phase of the signal at third rotational frequency harmonic, and $\xi(t)$ represents the sum of higher-frequency signals and noise above the third harmonic. According to Equation (1), the displacement sensor signal contains rich information about the rotor's rotational speed. To facilitate the extraction of rotational speed information, the original signal can be passed through a low-pass filter to remove high-frequency noise, while ignoring the lower-amplitude harmonics above the third harmonic. The simplified displacement signal can be expressed as:

$$y(t) = K + A_1 \sin(\omega t + \varphi) + A_3 \sin(3\omega t + \phi) \tag{2}$$

2.2 Mathematical Principles of Algebraic Derivative Frequency Estimation

A sine signal parameter estimation algebraic method based on the modal theory of linear systems is presented in reference [19]. This theory is extended and applied to the problem of extracting rotational speed from the displacement signal of a magnetically suspended rotor. Considering the displacement signal without noise interference in Equation (2), its Laplace transform can be expressed as:

$$y(s) = \frac{A_1\omega\cos\varphi}{s^2+\omega^2} + \frac{sA_1\sin\varphi}{s^2+\omega^2} + \frac{A_3 3\omega\cos\phi}{s^2+9\omega^2} + \frac{sA_3\sin\phi}{s^2+9\omega^2} + \frac{K}{s} \tag{3}$$

Without loss of generality, it is assumed that all phases lie between 0 and 2π. To obtain the rotational speed ω, it is necessary to eliminate the other unknown parameters in Equation (3) through integration, integral convolution, and algebraic operations.

For ease of calculation, define:

$$\omega^2=p_1, A_1\sin\varphi=p_2, A_1\omega\cos\varphi=p_3, A_3\sin\phi=p_4, A_3 3\omega\cos\phi=p_5 \tag{4}$$

The rotational speed can be uniquely determined as: $\omega=+\sqrt{p_1}$.

Based on Equation (4), Equation (3) can be rewritten as:

$$y(s) = \frac{p_3}{s^2+p_1} + \frac{sp_2}{s^2+p_1} + \frac{p_5}{s^2+9p_1} + \frac{sp_4}{s^2+9p_1} + \frac{K}{s} \tag{5}$$

By multiplying both sides of Equation (5) by $\left(s^2+p_1\right)\left(s^2+9p_1\right)s$, and rearranging, we obtain:

$$s^5 y(s) + 10p_1 s^3 y(s) + 9p_1^2 sy(s) = s^4(p_2+p_4+K) + s^3(p_3+p_5) +$$
$$s^2(9p_1 p_2 + p_1 p_4 + 10p_1 K) + s(9p_1 p_3 + p_1 p_5) + 9p_1^2 K \tag{6}$$

From Equation (6), it can be seen that the highest power of the unknown parameter unrelated to the rotational speed is the fourth power of s. Obviously, if Equation (6) is differentiated five times relative to s, the unknown parameters can be eliminated, ultimately yielding:

$$p_1^2 R + p_1 L + T = 0 \tag{7}$$

Among them,

$$R = 45\frac{d^4 y(s)}{ds^4} + 9s\frac{d^5 y(s)}{ds^5} \tag{8}$$

$$L = 600\frac{d^2 y(s)}{ds^2} + 600s\frac{d^3 y(s)}{ds^3} + 150s^2\frac{d^4 y(s)}{ds^4} + 10s^3\frac{d^5 y(s)}{ds^5} \tag{9}$$

$$T = 120 y(s) + 600s\frac{dy(s)}{ds} + 600s^2\frac{d^2 y(s)}{ds^2} + 200s^3\frac{d^3 y(s)}{ds^3} + 25s^4\frac{d^4 y(s)}{ds^4} + s^5\frac{d^5 y(s)}{ds^5} \tag{10}$$

By multiplying both sides of Equation (7) by s^5, all derivative operations represented by

positive powers of *s* can be eliminated in the time domain. Consequently, Equations (8)-(10) transform successively into:

$$M = Rs^{-5} = 45s^{-5}\frac{d^4 y(s)}{ds^4} + 9s^{-4}\frac{d^5 y(s)}{ds^5} \tag{11}$$

$$N = Ls^{-5} = 600s^{-5}\frac{d^2 y(s)}{ds^2} + 600s^{-4}\frac{d^3 y(s)}{ds^3} + 150s^{-3}\frac{d^4 y(s)}{ds^4} + 10s^{-2}\frac{d^5 y(s)}{ds^5} \tag{12}$$

$$E = 120s^{-5}y(s) + 600s^{-4}\frac{dy(s)}{ds} + 600s^{-3}\frac{d^2 y(s)}{ds^2} + 200s^{-2}\frac{d^3 y(s)}{ds^3} + 25s^{-1}\frac{d^4 y(s)}{ds^4} + \frac{d^5 y(s)}{ds^5} = Ts^{-5} \tag{13}$$

At this point, a quadratic linear equation (14) regarding the rotational speed can be obtained. Solving this equation yields Equation (15).

$$p_1^2 M + p_1 N + E = 0 \tag{14}$$

$$p_1 = \frac{-N \pm \sqrt{N^2 - 4ME}}{2M} \tag{15}$$

By performing the Laplace inverse transform on the coefficients of Equation (14), we obtain the coefficients of the equation in the time domain, as shown in Equations (16)-(18). In the time domain, it can be observed that the rotational speed is obtained through a linear combination of iterative convolution integrals based on the signal *y(t)*.

$$M = 45\int^{(5)} t^4 y(t) - 9\int^{(4)} t^5 y(t) \tag{16}$$

$$N = 600\int^{(5)} t^2 y(t) - 600\int^{(4)} t^3 y(t) + 150\int^{(3)} t^4 y(t) - 10\int^{(2)} t^5 y(t) \tag{17}$$

$$E = 120\int^{(5)} y(t) - 600\int^{(4)} ty(t) + 600\int^{(3)} t^2 y(t) - 200\int^{(2)} t^3 y(t) + 25\int t^4 y(t) - t^5 y(t) \tag{17}$$

The continuous integral operation can be obtained through multiple applications of the following single integral:

$$M = M_1, \quad \dot{M}_1 = M_2, \quad \dot{M}_2 = M_3, \quad \dot{M}_3 = M_4, \quad \dot{M}_4 = M_5 - 9t^5 y(t), \quad \dot{M}_5 = 45t^4 y(t) \tag{18}$$

$$N = N_1, \quad \dot{N}_1 = N_2, \quad \dot{N}_2 = N_3 - 10t^5 y(t), \quad \dot{N}_3 = N_4 + 150t^4 y(t),$$
$$\dot{N}_4 = N_5 - 600t^3 y(t), \quad \dot{N}_5 = 600t^2 y(t) \tag{19}$$

$$E = E_1 - t^5 y(t), \quad \dot{E}_1 = E_2 + 25 t^4 y(t), \quad \dot{E}_2 = E_3 - 200 t^3 y(t),$$
$$\dot{E}_3 = E_4 + 600 t^2 y(t), \quad \dot{E}_4 = E_5 - 600 t y(t), \quad \dot{E}_5 = 120 y(t)$$

(20)

Based on the coefficients calculated in the time domain, *p1* can be determined from Equation (15), and subsequently, the rotational speed can be obtained.

3   Suppression of Rotor Synchronous-Frequency Vibration Based on a Notch Filter

3.1   Method for Suppressing Synchronous-Frequency Current Based on a Notch Filter

The notch filter is currently the most widely employed method in magnetic suspension rotor control systems to achieve zero-current control for suppressing vibrations at the synchronous and harmonic frequencies of rotational speed. As stated in reference [15], zero-current control can only be implemented by incorporating the notch filter into the system when the rotational speed exceeds a certain fixed threshold. To ensure the stability of the magnetic bearing system within the operating speed range after the introduction of the notch filter, this paper proposes a novel rotational speed-adaptive synchronous notch filter with phase shift. The continuous-time domain block diagram of this filter is illustrated in **Fig.3**.
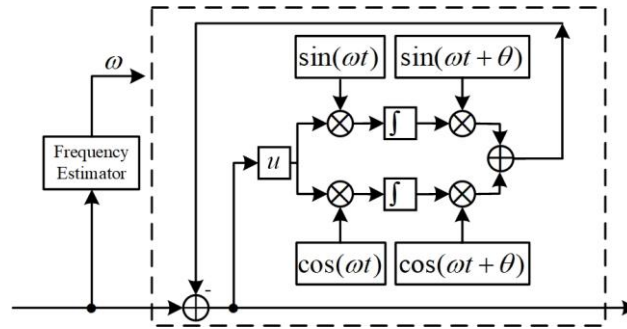
Fig.3   Structure of Rotational Speed Adaptive Notch Filter

The rotational speed-adaptive notch filter comprises three parts: rotational speed frequency estimation, synchronous frequency extractor, and synchronous signal negative feedback. Among them, the rotational speed frequency estimation module estimates the rotational speed frequency information in real-time based on the rotor displacement signal. The synchronous frequency extractor calculates the frequency components in the signal by performing correlation operations between the original displacement error signal and the sine and cosine signals at the synchronous frequency. It then generates the synchronous frequency component in the original displacement error signal based on the frequency components and the set phase shift. By superimposing the extracted synchronous component with negative feedback onto the original displacement error signal, a displacement error signal with the synchronous frequency component filtered out can be obtained. The transfer function between the output signal and the input signal of this notch filter is:

$$G_n(s) = \frac{s^2 + \omega_0^2}{s^2 + u\cos\theta \cdot s + \omega_0^2 - u\sin\theta \cdot \omega_0}$$

(21)

Fig.4 presents the frequency response when the notch filter is configured with different parameters.
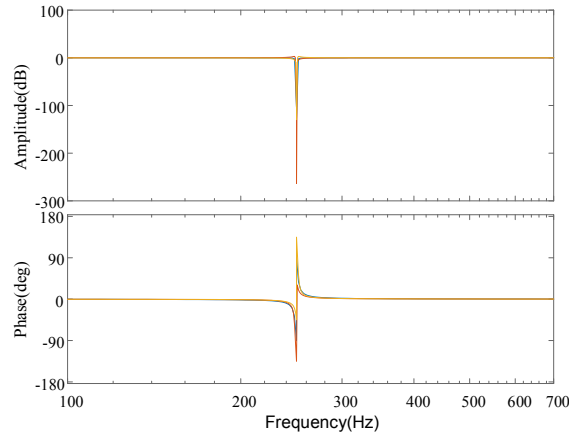
Fig.4 Frequency Response of the Notch Filter

When the estimated frequency equals the rotor's rotational frequency, the synchronous control current is completely suppressed. Consequently, the displacement caused by unbalanced mass cannot elicit a corresponding current response, and the synchronous vibration force decreases accordingly. Additionally, the phase angle of the notch filter can be altered by adjusting the value of parameter θ. When θ is negative, the phase angle becomes positive.

The controller for the magnetic suspension high-speed rotor system employs decentralized PID control $G_c(s) = K_P + \dfrac{K_I}{T \cdot s} + \dfrac{K_D T \cdot s}{K_f T \cdot s + 1 - K_f}$, with the power amplifier system having a transfer function of $G_p(s) = \dfrac{k_{amp} k_u}{Ls + R + k_{amp} k_u k_{ico} k_c}$, and the displacement sensor's proportional amplification stage having a transfer function of $G_s(s) = k_s$.

The structural block diagram of the magnetic bearing control system incorporating the synchronous frequency suppression algorithm of the rotational speed-adaptive harmonic notch filter proposed in this paper is shown in **Fig.5**.
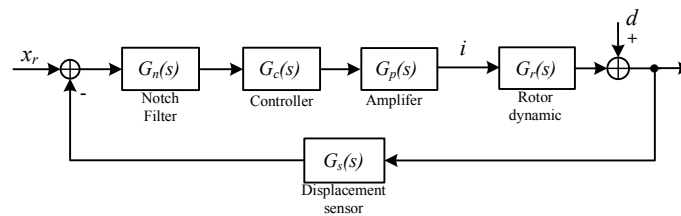
Fig.5 Block Diagram of the Magnetic Bearing Control System with a Notch Filter

3.2 Suppression of Unbalanced Vibrations and Algorithm Stability

Based on **Fig.5**, the transfer function from the disturbance variable d to the control current i can be derived as:

$$\frac{i}{d} = \frac{-G_s(s) \cdot G_n(s) \cdot G_c(s) \cdot G_p(s)}{1 + G_s(s) \cdot G_n(s) \cdot G_c(s) \cdot G_p(s) \cdot G_r(s)} \tag{22}$$

Substituting the transfer function of the notch filter, given by Equation (22), into the previous equation, we obtain:

$$\frac{i}{d} = \frac{-(s^2 + \omega_0{}^2) \cdot G_s(s) \cdot G_c(s) \cdot G_p(s)}{Q(s) + (s^2 + \omega_0{}^2) \cdot G_s(s) \cdot G_c(s) \cdot G_p(s) \cdot G_r(s)} \tag{23}$$

Wherein, $Q(s) = s^2 + \omega_0{}^2 + u\left(\cos\theta \cdot s - \sin\theta \cdot \omega_0\right)$.

When $s = j\omega_0$ and $u \neq 0$, we have $i = 0$. That is, when the rotational speed is ω_0, the synchronous frequency current is completely suppressed.

According to Equation (24), the closed-loop characteristic equation of the system can be described as:

$$Q(s) + (s^2 + \omega_0{}^2) \cdot G_s(s) \cdot G_c(s) \cdot G_p(s) \cdot G_r(s) = 0 \tag{24}$$

Equation (25) can be transformed into:

$$\frac{u(\cos\theta \cdot s - \sin\theta \cdot \omega_0)}{1 + G_s(s) \cdot G_c(s) \cdot G_p(s) \cdot G_r(s)} + s^2 + \omega_0{}^2 = 0 \tag{25}$$

Defining $S(s)$ as:

$$S(s) = \frac{1}{1 + G_s(s) \cdot G_c(s) \cdot G_p(s) \cdot G_r(s)} \tag{26}$$

Substituting Equation (27) into Equation (26) simplifies to:

$$u(\cos\theta \cdot s - \sin\theta \cdot \omega_0) \cdot S(s) + s^2 + \omega_0{}^2 = 0 \tag{27}$$

When $u = 0$, there exists $s = \pm j\omega_0$. Taking u as the independent variable and s as the dependent variable, by taking the partial derivative of Equation (28) at $u = 0$ and $s = j\omega_0$, we obtain:

$$\left.\frac{\partial s}{\partial u}\right|_{u=0, s=j\omega_0} = \frac{j(j\omega_0 \cos\theta - \omega_0 \sin\theta)}{2\omega_0} S(j\omega_0) \tag{28}$$

To ensure that all the roots of the closed-loop characteristic equation lie in the left half-plane,

the following condition must be satisfied:

$$-\frac{\pi}{2} < \theta + \arg[S(j\omega_0)] < \frac{\pi}{2} \tag{29}$$

During the speed-up process of the gyro rotor, the rotational speed ω_0 is constantly changing. The controller needs to appropriately adjust the phase angle parameter based on the estimated rotational speed to ensure that the entire magnetic bearing system remains stable at all times.

4 Simulation

To verify the feasibility and effectiveness of the proposed method, modeling and simulation were conducted on a MSCMG high-speed rotor system. **Tab.1** and **Tab.2** respectively present the structural and simulation control parameters related to the magnetic suspension rotor system used.

Tab.1 Structural Parameters of the Magnetically Suspended High-Speed Rotor System

Symbol	Variable	Value
m	Rotor mass	16.7 kg
ls	Distance between sensor and rotor center of mass	0.111 m
lm	Distance between bearing and rotor center of mass	0.0725 m
Jr	Equatorial moment of inertia of rotor	0.08286 kg·m^2
Jz	Polar moment of inertia of rotor	0.1302 kg·m^2
ki	Radial current stiffness	600 N/A
kx	Radial displacement stiffness	2.4×10^6 N/m
R	Coil resistance	3.4 Ω
L	Coil inductance	40 mH

Tab.2 Simulation Controller Parameters for Magnetically Suspended High-Speed Rotors

Symbol	Variable	Value
Fs	Sampling frequency	6.667kHz
Kp	Proportional coefficient	4.3
Ki	Integral coefficient	0.000375
Kd	Differential coefficient	11.5
Kf	Incomplete differential coefficient	0.6
$kamp$	Current loop proportional coefficient	1.4
$kico$	Current loop feedback coefficient	3.8
kc	Current sensor coefficient	0.495
ks	Sensor coefficient	8333

4.1 Numerical Simulation Analysis

Firstly, the effectiveness of the frequency estimation algorithm is verified based on numerical

simulation methods. The relevant parameters in Equation (1) are set as follows: K=1, A_1=10, $\omega = 150 \times 2\pi$, φ=0, A_3=1, ϕ=4, A_5=0.2, and $\xi(t)$ represents random high-frequency white noise.
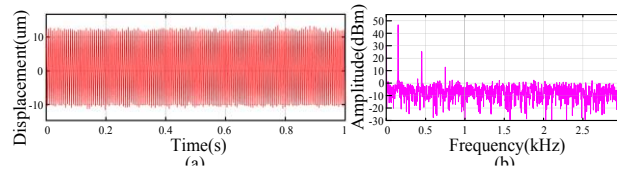
Fig.6 Numerical Simulation Signals: (a) Time-Domain Waveform; (b) Spectrum Plot

The simulated signal is shown in **Fig.6**, where (a) represents the time-domain waveform of the signal, and (b) represents the signal spectrum. The spectrum shows that the signal mainly contains DC components, a 150Hz synchronous component, triple-frequency components, quintuple-frequency components, and high-frequency noise components.

Fig.7 presents the estimation results of the frequency estimation algorithm. Due to the presence of interference noise, there is a certain steady-state error in the estimation curve. When the estimated frequency is 150Hz, the steady-state error is approximately 0.6Hz. The algorithm's estimation results converge within 30ms.
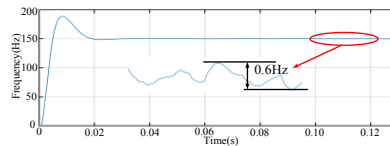
Fig.7 Convergence Plot of Frequency Estimation Results for Numerical Simulation Signals

4.2 Simulation Analysis of Magnetically Suspended Rotor Displacement Signal

In this section, the effectiveness of the frequency estimation algorithm is verified using displacement signals from magnetic suspension rotor dynamics simulations. The simulation frequency is set to the gyroscope's operating speed of 250Hz. The displacement signal is shown in **Fig.8**, where (a) represents the time-domain waveform of the signal, and (b) represents the signal spectrum. The spectrum shows that the signal mainly contains a 250Hz synchronous component, triple-frequency components, a small amount of quintuple-frequency components, and high-frequency noise.
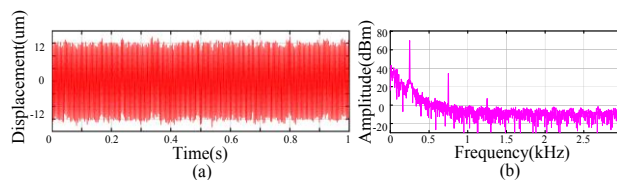
Fig.8 Displacement Simulation Signals: (a) Time-Domain Waveform; (b) Spectrum Plot

Fig.9 presents the estimation results of the frequency estimation algorithm. When the estimated frequency is 250Hz, the steady-state error is approximately 0.8Hz, and the algorithm's estimation results converge within 30ms.
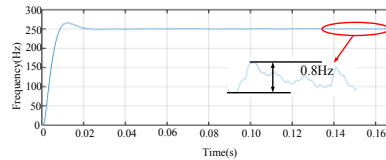
Fig.9   Convergence Plot of Frequency Estimation Results for Displacement Simulation Signals

4.3   Suppression of Unbalanced Vibrations Based on a Notch Filter

This section verifies the suppression effect of synchronous-frequency current and vibration at different rotational speeds based on the rotational speed frequency estimation results and the notch filter.

The parameters are set as $\theta = 0$, u=15. When the rotational speed is 150Hz, the rotational speed frequency estimation and notch filter algorithm proposed in this paper are applied. **Fig.10** shows the change waveforms of displacement, current, and vibration force from top to bottom, respectively. It can be observed that the displacement decreases from approximately 15um to about 8um, the current converges to nearly 0A, and the synchronous-frequency vibration force decreases from about 9N to about 4N, which is a reduction of nearly 55.6%.
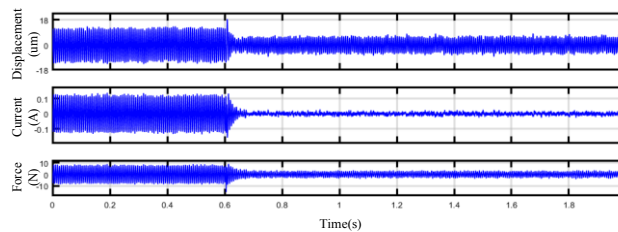
Fig.10   Suppression of Synchronous-Frequency Vibration at a Rotational Speed of 150Hz

The current signal spectra before and after applying vibration suppression are shown in **Fig.11**. It is evident that after vibration suppression, the synchronous-frequency current component at the rotational speed decreases from 8.31dBm to -34.35dBm, and the synchronous-frequency current is almost completely filtered out.
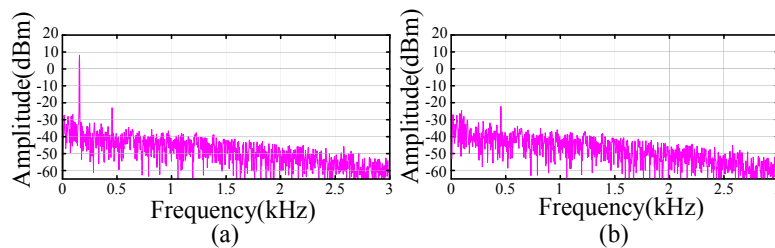
Fig.11   Spectrum of 150Hz Current Signal: (a) Before Suppression; (b) After Suppression

At the gyroscope's operating speed of 250Hz, the rotational speed frequency estimation and notch filter algorithm proposed in this paper are applied. As shown in **Fig.12**, the displacement decreases from approximately 10um to about 7um, the current converges to nearly 0A, and the synchronous-frequency vibration force decreases from about 11N to about 4N, which is a reduction of nearly 63.6%.
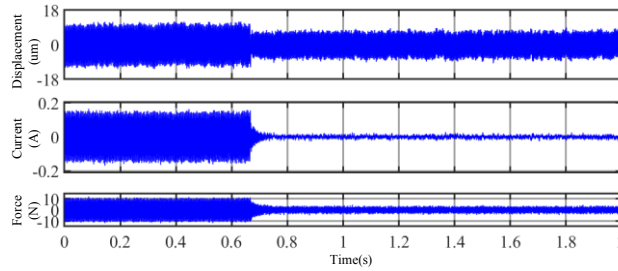
Fig.12 Suppression of Synchronous-Frequency Vibration at a Rotational Speed of 250Hz

The current signal spectra before and after applying vibration suppression at 250Hz are shown in **Fig.13**. It is evident that after vibration suppression, the synchronous-frequency current component at the rotational speed decreases from 6.05dBm to -37.24dBm, and the synchronous-frequency current is almost completely filtered out.
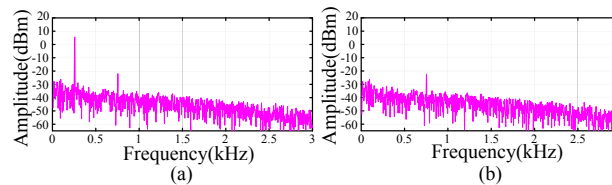
Fig.13 Spectrum of 250Hz Current Signal: (a) Before Suppression; (b) After Suppression

Based on the above simulation results, it is concluded that the rotational speed frequency estimation method proposed in this paper does not depend on the orthogonal characteristics of the displacement signal, has fast convergence speed, and is simple and easy to implement. Combined with the notch filter system, it can achieve suppression of synchronous-frequency current and vibration caused by the imbalance of the MSCMG rotor.

5 Conclusion

Addressing the issues of failure and significant errors in the rotational speed information of MSCMG rotors, this paper proposes a rotational speed estimation method for magnetically suspended rotors based on algebraic derivative frequency estimation. Based on this method, an adaptive notch filter with phase compensation for rotational speed is utilized to suppress the synchronous-frequency current and vibration in the magnetically suspended rotor system. The rotor speed estimation method requires only one displacement signal and does not rely on the orthogonal characteristics of the X and Y displacement signals, enabling fast and accurate estimation of the actual rotor speed. The notch filter system, based on the estimated rotational speed frequency, can suppress the synchronous-frequency current and vibration caused by the imbalance of the MSCMG rotor.

This method has a simple structure and is easy to implement, with the limitation that it only suppresses the vibration force generated by the synchronous-frequency current in the rotor system. To achieve micro-vibration suppression of the rotor, future research should focus on aspects such as synchronous-frequency displacement vibration force compensation and suppression of multi-harmonic vibrations.

References:

[1] Ye Dong. Research on Agile Satellite Attitude Rapid Maneuvering and Stability Control Method[D]. Harbin: Harbin Institute of Technology, 2013

[2] Meng Guang, Zhou Xubin. Progress in Satellite Micro Vibration and Control Technology[J]. Journal of Aeronautics, 2015, 36 (08): 2609-2619

[3] Cao Honglei, Zhu Qinghua, Sun Jiandang, Yu Changli. Precise attitude pointing control of large flexible structure spacecraft[J]. Flight Control and Detection, 2019, 2 (5): 17-24

[4] Shu Shi, Fang Jiancheng, Zhang Wei, Liu Gang, etc. High precision attitude control method based on MSCMG large remote sensing satellite[J]. Chinese Journal of Inertial Technology, 2017, 25 (04): 421-431

[5] Lv Qichao, Lv Dongyuan, Li Yanbao, Liu Pingfan. Feedforward compensation and experiment of dynamic frame effect of small maglev CMG high-speed rotor[J]. Flight Control and Detection, 2019, 2 (1): 49-55

[6] Oguchi K., Tomigashi Y. Sliding Mode Control for Single-Axis Magnetic Suspension Devices[J]. IEEE Transactions on Industry Applications, 2008, 110(11): 1184-1192.

[7] Ihn Y. S., Lee J. K., and Oh D. H. et al. Active Correction of Dynamic Mass Imbalance for A Precise Rotor[J]. IEEE Transactions on Magnetics, 2009, 45(11): 5088-5093.

[8] P. Cui, W. Li, J. Li, Hybrid Odd Repetitive Controller for Magnetically Suspended Rotor System[J]. IEEE Sensors Journal, 2019, 19(22): 10281-10288.

[9] Darbandi S. M., Behzad M., Salarieh H., et al. Harmonic disturbance attenuation in a three-pole active magnetic bearing test rig using a modified notch filter[J]. Journal of Vibration and Control, 2015, 11(4): 46-51.

[10] Herzog R，Bühler P，Gühler C，et al. Unbalance compensation using generalized notch filters in the multivariable feedback of magnetic bearings[J]. IEEE Transactions on Control Systems Technology, 1996, 45(5):580-586

[11] Zhang Jiyang, Chen Zongji, Liu Hu. Active Control of Harmonic Disturbance in Magnetic Suspension Flying Displacement Sensor[J]. Journal of Astronautics, 2015, 36 (11): 1289-1295

[12] C. Peng, M. Zhu, K. Wang, Y. Ren, et al. A two-stage synchronous vibration control for magnetically suspended rotor system in the full speed range[J]. IEEE Transactions on Industrial Electronics, 2020, 67(1):480-489.

[13] W. L. Lee, W. Schumacher, and W. R. Canders, Unbalance compensation on AMB system without a rotational sensor[J]. JSME Int. J. C, Mech. Syst., Mach. Elements Manuf., 2003, 46(2):423–428.

[14] M. Karimi-Ghartemani, A. Ziarani, Performance characterization of a non-linear system as both an adaptive notch filter and a phase-locked loop[J]. Int. J. Adapt. Control Signal Process., 2004, 18: 23–53.

[15] Q. Chen, G. Liu, B. Han, Suppression of Imbalance Vibration in AMB-Rotor Systems Using Adaptive Frequency Estimator[J]. IEEE Trans. Ind. Electron, 2015, 62(12):7696-7705.

[16] J. Li, G. Liu, C. Peng, S. Zheng, Synchronous Vibration Suppression of Magnetically Suspended Rotor System Using Improved Adaptive Frequency Estimation[J]. IEEE Sensors Journal, 2020, 20(19):11212-11220.

[17] Zhang Ran, Liu Hu, Fan Yahong. A method for estimating the displacement signal and speed of a magnetic levitation flywheel rotor[J]. Journal of Astronautics, 2017, 38 (12): 1314-1323

[18] Liu Hu, Yang Zhenpeng, Wu Dengyun. Speed estimation of magnetic levitation flywheel based on displacement signal[J]. Optical Precision Engineering, 2020, 28 (5): 1117-1123

Problems of Angular Motion of Cubesat 6U Nanosatellites in Low Earth Orbits

E.V. Barinova, I. V. Belokonov, I. A. Timbai

(Samara National Research University, Samara)

Abstract: This paper study the passive angular motion of the 6U CubeSat nanosatellites relative to its center of mass in low circular orbits under the influence of aerodynamic and gravitational moments, taking into account the features of the mathematical model of the aerodynamic drag force. A system of equations for determination of the nanosatellite equilibrium positions is obtained. It is shown the possibility of the occurrence of resonant motion modes.

1　Introduction

When performing a number of target tasks, it is necessary to maintain a certain orientation of the nanosatellite (NS), while in order to reduce energy consumption, preference is given to passive or combined stabilization systems. When designing passive stabilization systems, information on stable equilibrium positions under the action of external moments and an analysis of the uncontrolled motion of the NS relative to the center of mass are used. Therefore, an important task is to determine the equilibrium positions, as well as to study the motion of the NS in their vicinity under the action of gravitational and aerodynamic moments, which are predominant in low circular orbits. Since CubeSat NS have the shape of a rectangular parallelepiped, which entails changes in the mathematical model of the aerodynamic drag force, a new class of problems arises related to the search for equilibrium positions and the study of the possibility of the emergence of resonant motion modes.

Previously, the authors obtained a system of equations for determining the equilibrium positions of the angular motion of the CubeSat 1U–3U NS in a circular orbit under the action of aerodynamic and gravitational moments and found its analytical solution for two special cases [2, 3]. In development of the previously conducted studies, this paper analyzes the dynamics of CubeSat 6U nanosatellites. A system for determining the equilibrium positions in the orbital coordinate system under the action of aerodynamic and gravitational moments was obtained. For the special case of the displacement of the center of pressure from the center of mass along one of the main axes of inertia, analytical expressions were obtained. For the case of using aerodynamic stabilization along the velocity vector of the center of mass, the possibility of occurrence of resonant motion modes was shown and recommendations for their prevention were given.

2　Passive angular motion of CubeSat 6U nanosatellites relative to the center of mass under the action of aerodynamic and gravitational moments

Euler angles are used to orient the NS-related coordinate system relative to the trajectory: α is the spatial angle of attack, ψ is the precession angle, φ is the angle of proper rotation.

At the considered altitudes of motion, it is assumed that the flow around the nanosatellite is free-molecular and the impact of gas molecules is absolutely inelastic. In this case, the aerodynamic force is the drag force, which is determined by the area of the projection of the NS on the plane perpendicular to the velocity vector of the incoming flow [1]. Unlike the CubeSat 1U-3U NS, the CubeSat 6U NS has all three different linear dimensions (lx>lz>ly), and, therefore, the expression for the area of the projection of the NS on the plane perpendicular to the velocity vector of the incoming flow is also different:

$$S_x \cdot \tilde{S}(\alpha_n, \varphi) = S_x \cdot \left(|\cos\alpha_n| + \sin\alpha_n \left(\frac{l_x}{l_y} |\sin\varphi| + \frac{l_x}{l_z} |\cos\varphi| \right) \right),$$ (1)

where $S_x = l_y l_z$ is the characteristic area of the NS (in this case, the area of the smallest face).

The equations of spatial motion of a NS relative to the center of mass under the action of aerodynamic and gravitational moments in a circular orbit can be written in the following form:

$$\mathbf{I}\dot{\boldsymbol{\omega}} + \boldsymbol{\omega} \times \mathbf{I}\boldsymbol{\omega} = \mathbf{M}_a + \mathbf{M}_g,$$ (2)

where **I** is the NS inertia tensor; **ω** is the NS absolute angular velocity vector; $\mathbf{M}_a$ is the aerodynamic moment; $\mathbf{M}_g$ is the gravitational moment.

3   Equilibrium positions of the CubeSat 6U nanosatellite under the action of aerodynamic and gravitational moments.

In [2, 3], the equilibrium positions of the angular motion of a dynamically symmetric CubeSat 1U – 3U nanosatellite in a circular orbit under the action of aerodynamic and gravitational moments are determined for two special cases: displacement of the center of pressure of a dynamically symmetric NS from its center of mass along three coordinates and a NS with three different principal moments of inertia when the center of pressure is displaced from the center of mass along one of the principal axes of inertia.

In this paper, a system of equations is obtained for determining the equilibrium positions of a CubeSat 6U nanosatellite under the action of aerodynamic and gravitational moments, taking into account the features of the aerodynamic moment model. Its analytical solution is also found for the particular case of a CubeSat 6U NS with three different principal moments of inertia when the center of pressure is displaced from the center of mass along one of the principal axes of inertia.

3   Resonance Motion Modes

When using uniaxial aerodynamic stabilization of the longitudinal axis of the NS relative to the velocity vector, it is necessary to take into account the possibility of occurrence of resonant motion modes. Resonant motion modes are manifested themselves in a sharp change in the amplitude of oscillations along the spatial angle of attack. In this paper, the resonant motion modes are studied in accordance with the algorithm described in [4], considering the differences caused by

the features of the 6U format. It was found that, compared to the case of CubeSat 1U-3U [4], the CubeSat 6U NS has a greater number of resonant frequency ratios generated exclusively by the shape of the nanosatellite. For the case of "direct" precession, there are 7 ratios compared to 5, for the case of "reverse" - 3 ratios compared to 1. In addition, ratios were obtained for determining the critical values of the longitudinal angular velocity of the NS, at which the conditions for occurrence of resonant motion modes are met.

4 Conclusion

The results presented in this work may be useful for developers of CubeSat 6U nanosatellite, which continue to gain popularity at present.

References:

[1] Beletskii V.V. Dvizhenie iskusstvennogo sputnika otnositel'no tsentra mass (Motion of an Artificial Satellite Relative to the Center of Mass), Moscow: Nauka, 1965.

[2] Barinova E.V., Timba, I.A. Relative equilibria of dynamically symmetric CubeSat nanosatellite under the action of aerodynamic and gravitational torques, Vestnik of Samara University. Aerospace and Mechanical Engineering, 2019, Vol. 18, No. 2, pp. 21-32 (in Russian).

[3] Barinova E.V., Timbai I.A. Determining of Equilibrium Positions of CubeSat Nanosatellite under the Influence of Aerodynamic and Gravitational Moments. in 27th Saint Petersburg International Conference on Integrated Navigation Systems. ICINS 2020. Proceedings. 2020.

[4] Barinova E.V., Belokonov I.V., Timbai I.A. Preventing Resonant Motion Modes for Low-Altitude CubeSat Nanosatellites. Gyroscopy and Navigation. 2021. Vol. 12. № 4. pp. 350–362.

Research on Method of Assisted Inertial Positioning Based on Environmental Magnetic Field Under Particle Filter Framework

Ning Zhiwen[1], Pan Xianfei[1], Xu Meilin[1]

(1. College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China)

Abstract: This paper introduces a method for assisting micro inertial navigation systems based on environmental magnetic fields. Different from the classic geomagnetic field method of correcting course, this method closely couples the original output of gyroscopes and accelerometers with environmental magnetic field measurement information, and uses the magnetic field to correct attitude changes. Considering the nonlinear and non-Gaussian characteristics of the system, particle filtering is used to estimate the state. We verified this method through a set of outdoor experiments. The results show that in about 384 seconds, the RMSE value of the horizontal positioning error is below 3m, and the accuracy of the particle filter is equivalent to that of the classical extended kalman filter. To a certain extent, this method can suppress the error accumulation of micro inertial navigation systems.

 Key words: environmental magnetic field; MIMU; particle filtering; error suppression;

1 Introduction

Due to the natural existence of magnetic fields and the low cost of magnetic sensors, magnetic field-based positioning technology has attracted widespread attention from researchers in recent years [1]. At present, most positioning in indoor or high-rise roadways uses environmental magnetic field information, that is, the superposition of the geomagnetic field and the external magnetic field during travel.

Currently, positioning methods using environmental magnetic fields can be mainly divided into magnetic field matching and magnetic field SLAM. Methods based on matching [2] often require high-precision magnetic maps to be drawn in advance, which consumes a lot of time and computing costs. In the absence of a magnetic field map, magnetic field SLAM can allow the carrier to determine its own position during movement and build an incremental geomagnetic map to achieve autonomous navigation. Reference [3] is the first to realize magnetic field SLAM in the two-dimensional plane, and magnetic maps are constructed in hexagonal tiles of different size. The "magnetic loop-back" mechanism was introduced in reference [4], which uses magnetic field information under the same attitude to detect the closed loop as a position constraint. As long as this scheme can achieve frequent closed-loop, it can achieve long-term and large-scale positioning without drift.

In addition, magnetic field odometer is a new idea proposed in recent years. One implementation scheme is to use a magnetometer array to provide odometer information based on

local magnetic field characteristics. This method does not require the construction of a magnetic field map [5]. Reference [6] uses dual magnetometers to estimate motion speed, and realizes speed estimation by matching the magnetic field waveforms sampled by the front and back magnetometers, which enhances the robustness of the positioning system.

In this paper, we use a nine-axis sensor to tightly couple the original output information of the inertial sensor with the environmental magnetic field information. For the nonlinear and non-Gaussian characteristics of the system, we use particle filtering (PF) [7] to achieve state estimation. System modeling and state estimation methods will be briefly described in sections 2 and 3 of this paper. In section 4, the verification is carried out through a set of outdoor unmanned vehicle tests, and finally the full text is summarized.

2　System Modeling

Combined with Maxwell's Law of Electromagnetic [8], the parametric model of the magnetic field in space can be expressed in the following form:

$$B(p;m) = \mathrm{M}(p)m \tag{1}$$

Where $\mathrm{M}(p)$ represents the regression matrix of the magnetic field coefficients at position p [9], and m is the parameter vector of the polynomial model. For an a-order model, its order satisfies: $\dim(m)=l=a^2+4a+3$ [10]. Model (1) can be defined in any frame. In this paper, in order to make the calculation simple and avoid additional projection errors, we set $a=1$, $l=8$, and deduce them under the carrier frame (b-frame).

Assuming that the carrier frame is transformed from b_k to b_{k+1}, the pose change during this process can be expressed as:

$$\xi_k = \begin{bmatrix} \Delta p_k^{b_k} & \Delta \theta_k \end{bmatrix}^T \tag{2}$$

Here, we use Δp and $\Delta \theta$ to represent the changes in displacement and attitude respectively. Combining formula (1), the change in the magnetic field model can be expressed as follows:

$$B^{b_{k+1}}\left(p^{b_{k+1}};m_{k+1}\right) = C_{b_k}^{b_{k+1}} B^{b_k}\left(p^{b_k};m_k\right) \tag{3}$$

Considering that the sensor we use is a micro inertial measurement unit (MIMU), its rotation matrix and displacement change vector can be expressed as:

$$C_{b_k}^{b_{k+1}} = \left(\exp\left(\left[\Delta\theta_k\right]\times\right)\right)^T = \left(\exp\left(\left[\omega_k\Delta t\right]\times\right)\right)^T, \quad \Delta p_k^{b_k} = C_n^{b_k}\left(v_k^n\Delta t + \left(C_{b_k}^n f_k + g^n\right)\frac{\Delta t^2}{2}\right) \tag{4}$$

We use ω_k and f_k to represent the output data of the gyroscope and accelerometer at the current moment (b-frame) respectively, Δt to represent the sampling interval time, and $[\cdot]\times$ to represent the antisymmetric matrix.

Based on the above analysis, we can transform the magnetic field model in formula (1) as

follows:

$$\mathrm{M}\left(p^{b_{k+1}}\right)m_{k+1} = C_{b_k}^{b_{k+1}}\mathrm{M}\left(p^{b_k}\right)m_k \tag{5}$$

At this time, formula (5) contains three unknown variables: $p^{b_{k+1}}$, m_k, and ξ_k. The dimension of m_{k+1} is greater than 3, so multiple position vectors can be used to solve the least-squares equations. When the number of position vectors is j, its least-squares equations can be expressed as follows:

$$\begin{bmatrix} \mathrm{M}\left(p_1^{b_{k+1}}\right) \\ \vdots \\ \mathrm{M}\left(p_j^{b_{k+1}}\right) \end{bmatrix}m_{k+1} = \begin{bmatrix} C_{b_k}^{b_{k+1}}\mathrm{M}\left(p_1^{b_k}\right) \\ \vdots \\ C_{b_k}^{b_{k+1}}\mathrm{M}\left(p_j^{b_k}\right) \end{bmatrix}m_k \Rightarrow Am_{k+1} = B\left(\xi_k\right)m_k \tag{6}$$

If the A matrix column has full rank, setting $A^{\dagger}$ represents the pseudo-inverse of the A matrix, the following relationship can be obtained:

$$m_{k+1} = A^{\dagger}B\left(\xi_k\right)m_k \tag{7}$$

We combine the output value of the magnetometer with the magnetic field parametric model in formula (1), and the output of the magnetometer at time k can be expressed as:

$$ymag_k = \mathrm{M}\left(p_m\right)m_k + e_k \tag{8}$$

In the above formula, p_m represents the installation position of the magnetometer relative to the center of mass of the carrier. In addition, e_k contains measurement noise and magnetic field modeling errors.

To summarize the analysis, combining the magnetic field parametric model with the inertial navigation equation, we combine the three-dimensional attitude (ϕ_k), velocity (v_k), position (p_k), zero bias of the gyroscope and accelerometer (o_k^ω, o_k^a) [11], and the magnetic field parameter vector (m_k) to form a complete system state space model:

$$x_k = \begin{bmatrix} \phi_k^T & v_k^T & p_k^T & \left(o_k^\omega\right)^T & \left(o_k^a\right)^T & m_k^T \end{bmatrix}^T = \begin{bmatrix} \left(x_k^{ins}\right)^T & m_k^T \end{bmatrix}^T, w_k = \begin{bmatrix} \left(w_k^{ins}\right)^T & \left(w_k^m\right)^T \end{bmatrix}^T \tag{9}$$

In formula (9), w_k represents process noise.

3   State Estimation Based on Particle Filter

As shown in figure 1, we estimate the error state δx_k through particle filter and use δx_k to correct x_k.
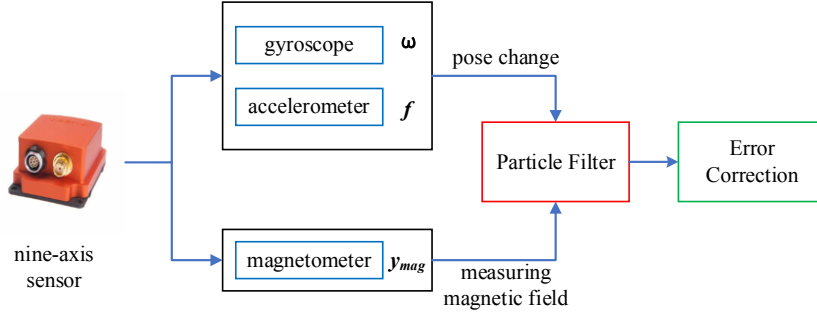
Fig.1 The flow of the principle of algorithm

3.1 Inertial navigation error status

The system error state can be expressed as follows:

$$\delta x_k = \left[\delta\phi_k^T \quad \delta v_k^T \quad \delta p_k^T \quad \left(\delta o_k^\omega\right)^T \quad \left(\delta o_k^a\right)^T \quad \delta m_k^T \right]^T = \left[\left(\delta x_k^{ins}\right)^T \quad \delta m_k^T \right]^T \tag{10}$$

For the inertial navigation error model, we ignore the influence of the earth's rotation and obtain a simplified form of the inertial error model [12]: $\delta\dot{x}^{ins} = M\delta x^{ins} + Nw^{ins}$. Then discretize the M and N matrices to obtain Φ_k^{ins} and G_k^{ins} at times k under the discrete system.

3.2 Magnetic field error status update

First, we solve for the partial derivative of the equation in formula (7):

$$\delta m_{k+1} = A^\dagger \left[B\left(\hat{\xi}_k\right) \quad d\left(B\left(\hat{\xi}_k\right)m_k\right)\Big/d\xi_k \right] \begin{bmatrix} \delta m_k \\ \delta\xi_k \end{bmatrix} + w_k^m \tag{11}$$

Transform formula (4) from the error state in formula (10):

$$\hat{C}_n^{b_k} = \left(I_3 - \left([\delta\phi_k]\times\right)\right)^T C_n^{b_k}, \quad \Delta\hat{p}_k^{b_k} = \hat{C}_n^{b_k}\Delta t\left(\hat{v}_k^n + g^n\Delta t/2\right) + \hat{f}_k\Delta t^2/2 \tag{12}$$

By connecting formula (4) and formula (12), we can deduce the position change error:

$$\delta\Delta p_k = \Delta p_k^{b_k} - \Delta\hat{p}_k^{b_k} \approx \hat{C}_n^{b_k}\delta v_k^n\Delta t + \left(\left[\varpi\left(\hat{C}_n^{b_k},\hat{v}_k^n\right)\right]\times\right)\delta\phi_k, \quad \hat{C}_n^{b_k}\Delta t\left(\hat{v}_k^n + g^n\Delta t/2\right) = \varpi\left(\hat{C}_n^{b_k},\hat{v}_k^n\right) \tag{13}$$

Similarly, the attitude change error can be expressed as:

$$\delta\Delta\theta_k = \Delta\theta_k - \Delta\hat{\theta}_k = -\left(\delta o_k^\omega + w_k^\omega\right)\Delta t \tag{14}$$

After obtaining the expression for attitude change error, formula (11) can be solved: $\delta m_{k+1} = \Phi_k^m \delta\hat{x}_k + G_k^m w_k$ 。

$$\Phi_k^m = A^\dagger \left[B\left(\hat{\xi}_k\right) \quad d\left[B\left(\hat{\xi}_k\right)m_k\right]\Big/d\Delta p_k \quad d\left[B\left(\hat{\xi}_k\right)m_k\right]\Big/d\Delta\theta_k \right] J(\hat{x}_k)$$

$$G_k^m = \left[-A^\dagger \cdot d\left[B\left(\hat{\xi}_k\right)m_k\right]\cdot\Delta t^2\Big/2\cdot d\Delta p_k \quad -A^? \cdot d\left[B\left(\hat{\xi}_k\right)m_k\right]\cdot\Delta t\Big/d\Delta\theta_k \quad 0 \quad 0 \quad I_l \right] \tag{15}$$

$$J(\hat{x}_k) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & I_l \\ 0 & \hat{C}_n^{b_k}\Delta t & \left(\left[\varpi\left(\hat{C}_n^{b_k},\hat{v}_k^n\right)\right]\times\right) & -\Delta t^2\cdot I_3/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & -I_3\Delta t & 0 \end{bmatrix}$$

At this time, the complete systematic error model can be expressed as:

$$\begin{cases} \delta x_{k+1} = \Phi_k\delta x_k + G_k w_k \\ \delta ymag_k = H\delta x_k + e_k \end{cases}, \quad \Phi_k = \begin{bmatrix} \Phi_k^{ins},0 \\ \Phi_k^m \end{bmatrix}, G_k = \begin{bmatrix} G_k^{ins},0 \\ G_k^m \end{bmatrix}, H = \begin{bmatrix} 0 & M(p_m) \end{bmatrix} \tag{16}$$

Considering the nonlinear and non-Gaussian characteristics of the system, we use PF to estimate the state to correct the error. Algorithm 1 describes the operation at time k (n represents the number of particles).

Algorithm 1	PF Algorithm

1:     **Input**: $x_{k-1}, \omega_k, f_k, ymag_k$

2:     **Output:** x_k

3:     **for** $i = 1$ to n **do**

4:         State predicte: $x_{k,i} = MAGINS\_model(\omega_k, f_k, x_{k-1,i})$

5:         Particle weight update: $q_{k,i} = update\_particle(x_{k,i}, ymag_k, q_{k-1,i})$

6:         Magnetic parameter vector update: $m_{k,i} = update\_mag(ymag_k, x_{k,i,} m_{k-1,i})$

7:     **End for**

8:     Normalized weight: $q_{k,i} = q_{k,i} / [sum(q_{k,i})]$, $i = 1{:}n$

9:     Determine whether resampling is needed

10:    **Return** x_k

11:    **End**

4   Experimental Analysis

We use unmanned vehicles for outdoor verification and use the EKF algorithm to compare it with the particle filter method in this paper. The relevant configuration of unmanned vehicles is shown in figure 2.
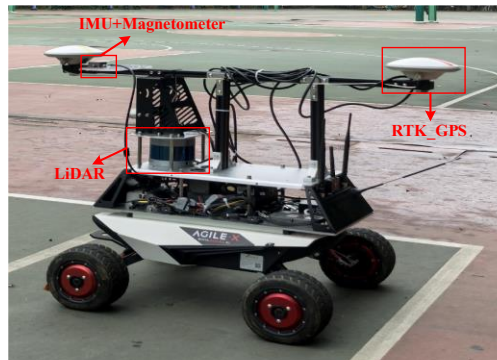
Fig.2   Sensor configuration for unmanned vehicles

We use dual-antenna RTK-GPS to provide attitude benchmarks with accuracy ranging from decimeters to centimeters. IMU and magnetometer are integrated into a 9-axis attitude heading system (sampling frequency 300Hz, minimum angular velocity resolution of 0.01°/s, and magnetometer measurement range of ±1.3Gauss).

During the test, the unmanned vehicle was operated to circle around a rockery for two times. After completing the data collection, the calculation is carried out under the EKF and PF frameworks respectively. The trajectory comparison and heading comparison of different algorithms are shown in figure 3 and figure 4 respectively.
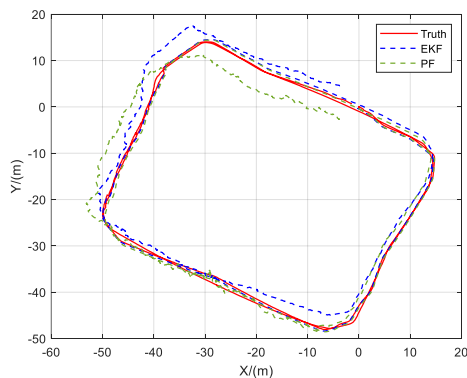
Fig.3   Trajectory comparison (EKF and PF)   **Fig.4**   Heading angle comparison (EKF and
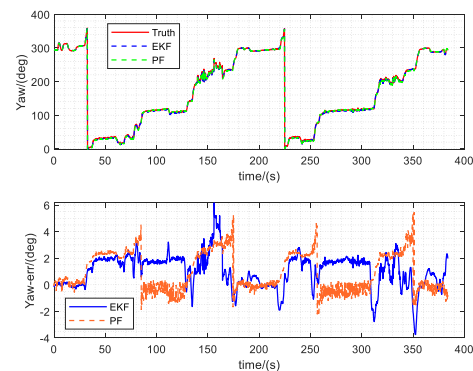
PF)

It can be seen from the above figure that its trajectory and heading changes can be fitted to the reference values, and the maximum heading error is about 5-6°, which shows that the algorithm is feasible. The extreme point of course error mainly occurs before and after turning, taking into account the uneven road surface. The change in position error is shown in figure 5.
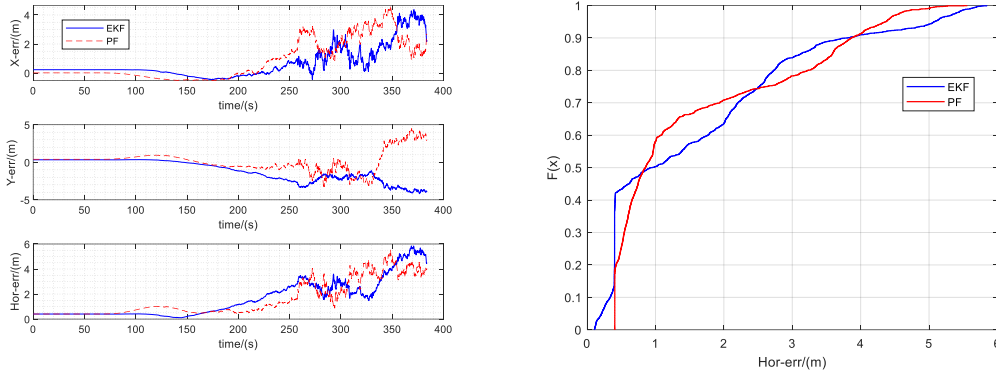
Fig.5 Position error comparison and CDF curve

It can be seen from figure 5 that the horizontal positioning errors under the two algorithm frameworks are both below 6m, and the CDF curve can also reflect that the horizontal errors are of the same order of magnitude. On this basis, we further analyzed the impact of different particle numbers on positioning accuracy and set the particle numbers to 100, 150, and 200 respectively. The comparison between trajectory tracking and horizontal positioning errors is shown in figure 6 and figure 7. Table 1 analyzes the statistical characteristics of horizontal positioning errors.
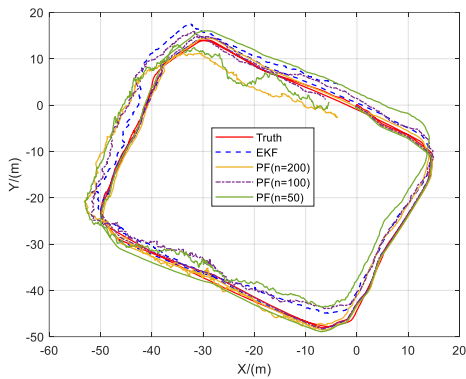
Fig.6 Comparison of trajectories for different particles
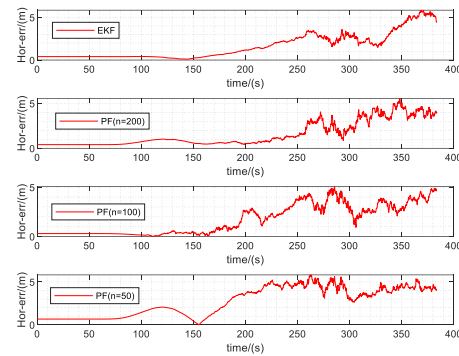
Fig.7 Comparison of horizontal errors for different particles

Tab.2   Statistical characteristics of horizontal error (m)

	MEAN	STD	RMSE	MAX
EKF	1.5963	1.4949	2.1870	5.8465
PF, n=200	1.5547	1.3857	2.0826	5.5569
PF, n=100	1.8623	1.6379	2.4801	5.2628
PF, n=50	2.3836	1.6828	2.9178	5.9142

Although the horizontal positioning error obtained by PF is reduced compared with EKF, the two are still in the same order and have not achieved significant advantages. Within a certain range, as the number of particles decreases, the horizontal error will increase. When the number of particles

is reduced from 200 to 100, although the maximum horizontal error value decreases, the RMSE value increases. The auxiliary method based on the environmental magnetic field described in this paper can effectively suppress the divergence of micro inertial navigation errors and improve positioning accuracy.

5 Conclusion

In this paper, under the framework of particle filtering, environmental magnetic field information and inertial information are closely coupled, inertial positioning errors are suppressed through changes in environmental magnetic fields at different locations, and the experimental verification is carried out in an outdoor environment. The results show that within a time range of about 384 seconds, the maximum horizontal positioning error is below 6m, and the RMSE value is below 3m. In addition, due to the strong electromagnetic interference of the unmanned vehicle itself, the compensation issue of magnetic sensors is not considered in this paper. On the basis of realizing magnetic field interference compensation, its positioning accuracy is expected to be further improved.

References:

[1] Wang Y, Li X, Zou JH. A Foot-Mounted Inertial Measurement Unit (IMU) Positioning Algorithm Based on Magnetic Constraint[J]. Sensors, 2018, vol.18(3): 741.

[2] X. Niu, L. Ding, Y. Wang, et al. MGINS: A Lane-Level Localization System for Challenging Urban Environments Using Magnetic Field Matching/GNSS/INS Fusion[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, vol. 25, no. 10, pp. 14890-14904.

[3] Vallivaara I, Haverinen J, Kemppainen A, et al. Simultaneous localization and mapping using ambient magnetic field[C]. //Proceedings of 2010 IEEE Conference on Multi-sensor Fusion and Integration, 2010:14-19.

[4] Jongdae J, Taekjun O, Hyun M. Magnetic field constraints and sequence-based matching for indoor pose graph SLAM[J]. Robotics &Autonomous Systems, 2015, 70:92-105.

[5] C. Huang, G. Hendeby, H. Fourati, et al. MAINS: A Magnetic-Field-Aided Inertial Navigation System for Indoor Positioning[J]. IEEE Sensors Journal, 2024, vol. 24, no. 9, pp. 15156-15166.

[6] Tisheng Z, Linfu W, Jian K, et al. Mag-ODO: Motion speed estimation for indoor robots based on dual magnetometers[J]. Measurement, 2023, 222.

[7] Y. Cui, Z. Xiong, X. Li, et al. A Pedestrian SLAM Scheme Grounded in Inertial-Based Map and Magnetic Field Map[J]. IEEE Sensors Journal, 2024, vol. 24, no. 5, pp. 6500-6514.

[8] J. D. Jackson. Classical Electrodynamics[M]. Hoboken, NJ, USA: Wiley, 1999.

[9] I. Skog, G. Hendeby, and F. Gustafsson. Magnetic odometry--A model-based approach using a sensor array[C]. in Proc. 21st Int. Conf. Inf. Fusion (FUSION), Jul. 2018, pp. 794-798.

[10] I. Skog, G. Hendeby, and F. Trulsson. Magnetic-field based odometry-An optical flow inspired approach[C]. in Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN), Nov. 2021, pp. 1-8.

[11] Z. Fan and L. Zhao. Autonomous Integrity of Multisource Information Resilient Fusion Navigation System: State-of-the-Art and Open Challenges[J]. IEEE Transactions on Instrumentation and Measurement, 2024, vol. 73, pp. 1-17.

[12] J. Tang, H. Bian, H. Ma, et al. SINS/GNSS Integrated Navigation Based on Invariant Error Models in Inertial Frame[J]. IEEE Sensors Journal, 2024, vol. 24, no. 4, pp. 4290-4303.

An Autonomous Navigation Method Based on Astronomical Spectrum Three-dimensional Velocity Measurement for Deep Space Missions

Huang Qinglong[1,2], Chu Yingzhi[1,2], You Wei[1,2], Chen Xiao[1,2]

(1. Shanghai Institute of Satellite Engineering, Shanghai 200240, China;

2. Shanghai Key Laboratory of Deep Space Exploration Technology, Shanghai 200240, China)

Abstract: Autonomous navigation is one of the key technologies to ensure the successful implementation of deep space exploration missions. Focusing on the demand for astronomical autonomous navigation capabilities in actual deep space exploration tasks, this paper proposes a combined autonomous navigation method based on three-dimensional vector velocity measurement using astronomical spectrum. The observability of the combined navigation method is analyzed, and a simulation of the combined navigation system is provided based on the UKF (Unscented Kalman Filter) algorithm. The simulation results indicate that compared to traditional ground-based radio navigation or autonomous angular navigation, the proposed method has better observability, higher navigation accuracy, and superior real-time performance. Additionally, the development of the solar-spectrum velocity meter and its in-orbit performance are introduced. Moreover, the future development of the three-dimensional spectrum velocity meter is presented.

Key words: Deep space exploration, astronomical spectrum, velocity measurement navigation, combined navigation, spectrum velocity meter

1 Introduction

Nowadays, deep space exploration is one of the most important space activities in the world. It is a reflection of a country's comprehensive national strength and innovative capabilities. Compared to near-Earth satellites, deep space exploration has its unique characteristics such as large communication delays between the probe and Earth, unknown flight environments, only opportunity for obit insertion maneuver and high requirements for the autonomous operations. Therefore, continuous, real-time and high-precision navigation has become one of the key technologies that must be solved in the implementation of deep space exploration missions.

At present, deep space exploration probes mainly rely on ground-based radio for navigation. Radio navigation makes full use of ground resources, such as ground stations, large computers and human participation, which can meet the needs of current space navigation tasks. However, due to geographical limitations, ground radio navigation has its drawbacks such as decreasing positioning accuracy with increasing distance and discontinuous navigation data due to unobservable arcs. Autonomous navigation methods such as angle and range measurement are currently more commonly used for deep space missions. However, due to constraints on the observation conditions of target celestial bodies, it is hard to obtain long-term real-time and continuous navigation information [1-3].

The astronomical spectrum contains the optical information of celestial bodies in space and the frequency shift caused by their motion relative to the probe. In 2013, Chinese research team proposed a high-precision autonomous navigation method for spacecraft using the Doppler frequency shift of the sun and other stars [4-8]. In collaboration with the Institute of Precision Measurement Science and Technology Innovation, Chinese Academy of Sciences, they developed the first one-dimensional solar velocity meter based on atomic frequency discriminator [9], which was successfully verified on the CHASE Mission launched in Oct. 2021.

Based on the aforementioned research, this paper extends one-dimensional spectral velocity measurement to three dimensions and integrates astronomical velocity measurements with angular measurements to form a combined navigation system. Based on observability analysis and UKF filtering algorithm simulation, navigation performance evaluation is provided. The results show that combined navigation can effectively solve the problem of low instantaneous velocity estimation accuracy in angular measurement navigation and the divergence of position estimation accuracy over time in velocity measurement navigation. The combined navigation method can support future deep space missions especially for outer planet explorations.

2   Basic principle

2.1   Principle of astronomical spectral velocity measurement

Astronomical spectrums can be detected by probes during flight. If the probe is stationary relative to the spectrum source, the wavelength of the spectrum received by the probe will be a constant. If the probe is non-stationary relative to the spectrum source, the received spectrum will drift[10, 11].

According to the Doppler principle, the ratio of the wavelength drift proportional to the relative velocity along line-of-sight, i.e:

$$\frac{v_r}{c} = \frac{\Delta\lambda}{\lambda_{rest}} \tag{1}$$

where v_r is the relative velocity, c is the speed of light in vacuum and $\Delta\lambda$ is the wavelength drift. Therefore, the relative velocity can be derived if wavelength drift can be detected by on-board instrument.

Assuming that the motion speed of the probe and the star in the heliocentric ecliptic inertial coordinate system is v and v_{Star} respectivly, and the line-of-sight direction of the probe relative to the star is l_{Star}, the velocity of the probe relative to the star v_{Spe} is:

$$v_{Spe} = \left(v_{Star} - v\right)^T l_{Star} \tag{2}$$

Usually, v_{Star} and l_{Star} can be obtained from the star catalogue such as Hipparcos Catalogues.

2.2   Navigation system state model

The state model of an asteroid probe's autonomous navigation system is established here as an example for numerical simulation. According to orbital dynamics, solar gravitational perturbations, solar radiation pressure perturbations, and probe thrust forces should be considered in state model. The position vector of the probe in the $J2000$ heliocentric ecliptic inertial coordinate system can be expressed as $r = [x, y, z]^T$, and the velocity vector can be expressed as $v = \left[v_x, v_y, v_z \right]^T$, then the state variables can be written as $X(t) = \left[x, y, z, v_x, v_y, v_z \right]^T$, and the orbital dynamics equations can be expressed as:

$$\ddot{r} = -\frac{u_s}{r^3} r + \eta P_{SR} A U^2 C_R \left(\frac{A_R}{m} \right) \frac{r}{r^3} + a_T \tag{3}$$

where the first term on the right hand represents the gravitational acceleration term caused by the sun and μ_s is the solar gravitational constant. The second term represents the perturbation acceleration term of solar light pressure and η is the shadow factor, P_{SR} is the light pressure at 1 AU from the sun, C_R is the surface reflection coefficient of the detector, A_R is the cross-sectional area of the probe perpendicular to the direction of solar light, m is the mass of the probe. The last term represents the thrust acceleration term of the probe.

In general, the state model of the probe can be expressed as:

$$\dot{X}(t) = \begin{bmatrix} \dot{r} \\ \dot{v} \end{bmatrix} = f(X(t), t) + W(t) \tag{4}$$

where $W(t)$ is the state model noise, which is generally treated as Gaussian white noise

3 Simulation analysis

3.1 Observability analysis

The observability of a navigation system reflects the system's ability to determine its state through a finite amount of measurement data within a certain period of time. The observability matrix is an important basis for the observability analysis. Because of the strong nonlinear characteristic of the deep space exploration autonomous navigation system, the observability matrix of the system can be solved using Lie derivatives.

The navigation system equations can generally be expressed as:

$$\sum : \begin{cases} \dot{X} = f(X) \\ Z = h(X) \end{cases} \tag{5}$$

where the state vector $X \in X^n \subset R^n$. The state equation f and the observation equation h are both smooth analytical functions in C^n.

Based on the theory of differential geometry, the Lie derivatives of h along f of various orders can be expressed as:

$$L_f^0 h(X) = h(X) \tag{6}$$

$$L_f^k h(X) = \frac{\partial\left(L_f^{k-1} h\right)}{\partial X} f(X), k = 1, 2, \dots \tag{7}$$

where

$$L_f^k h(X) = \frac{\partial\left(L_f^{k-1} h\right)}{\partial X} f(X), k = 1, 2, \dots \tag{8}$$

$$\mathrm{d}L_f^k h(X) = \frac{\partial\left(L_f^k h\right)}{\partial X}, k = 0, 1, 2, \dots \tag{9}$$

Therefore, the observability space $\boldsymbol{H}$ for the nonlinear system's can be obtained:

$$\boldsymbol{H}^n = \left\{ h, L_f h, \cdots, L_f^{n-1} h \right\} \tag{10}$$

The observability matrix $Q(X)$ for a nonlinear system can be defined by $\mathrm{d}\boldsymbol{H}^n$ as[12]:

$$Q(X) = \begin{bmatrix} \mathrm{d}L_f^0 h(X) \\ \mathrm{d}L_f^1 h(X) \\ \vdots \\ \mathrm{d}L_f^{n-1} h(X) \end{bmatrix} \tag{11}$$

The condition number γ of the observability matrix can represent the observability of the navigation system [13]:

$$\gamma = \frac{1}{cond(Q)} = \frac{1}{\|Q\| \cdot \|Q^{-1}\|} \tag{12}$$

where σ_Q are the singular values of Q. When $\gamma = 0$, $rank(Q) < n$, which means the system is unobservable. This method does not depend on observation data and the result is only related to the state model of the navigation system.

To analyze the observability of the system, this paper defines the system observability order as the minimum number of observations required for the system's observability matrix $Q(k)$ to meet the observability conditions. The system observability order reflects the integral relationship between the system's state and its measurements, and embodies the convergence speed of the navigation estimation results as well as the navigation system's ability to resist interference.

The observability performance for both angle-only and angle-velocity combined navigation systems are as shown in Table 1. It is shown that the line-of-sight vectors of the sun and asteroid are second-order observable schemes, with an observability degree one order of magnitude lower than that of the angle-velocity combined measurement scheme .

Tab. 1   Observability analysis of different navigation systems

Navigation System	Measurement Scheme	Observability Order	Observability Degree
Angular Navigation	Sun and asteroid line-of-sight vectors	2	8.97×10^{-11}
Angle-Velocity Combined Navigation	Sun and asteroid line-of-sight vectors / Sun and star line-of-sight velocities	1	8.24×10^{-10}

3.2   Navigation Accuracy Analysis

Using the asteroid exploration mission as an example, a numerical simulation analysis is carried out on the angle-velocity combined navigation method proposed in this paper. The nominal orbit data used in simulation is generated by Satellite Tool Kit (STK). The results of the angle navigation are shown in Figure 5, where the position estimation accuracy obtained from the navigation simulation is 254.82 km (3σ) and the velocity estimation accuracy is 4.32 m/s (3σ). It can be seen that with the angle measurement accuracy guaranteed to be 2 arcsec, the position estimation accuracy is relatively high and the velocity estimation accuracy is good. Figure 6 shows the combined navigation simulation results after adding the velocity measurements of three stars, with the position estimation accuracy approaching 114.54 km (3σ) and the velocity estimation accuracy approaching 1.31 m/s (3σ), which show a significant improvement over the angle-only navigation.
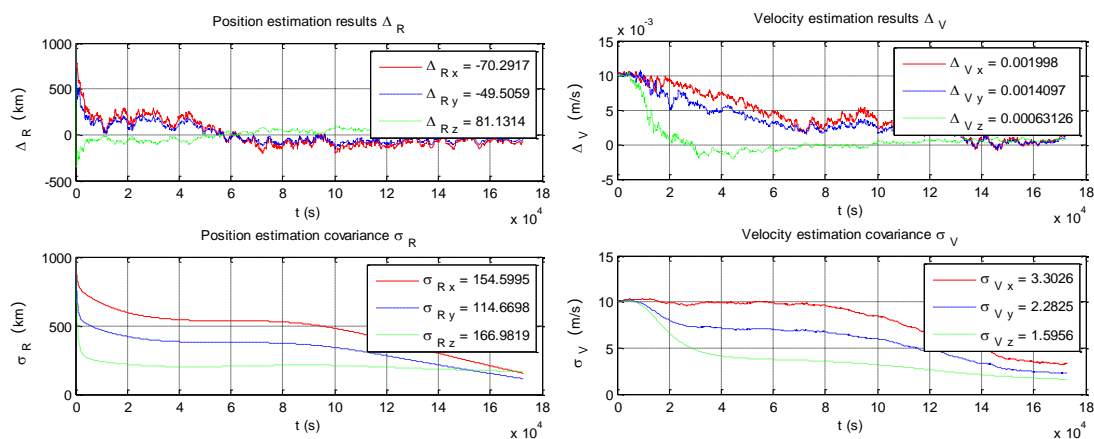
Fig.1   Navigation accuracy analysis of angle-only navigation (left: position error, right: velocity error)
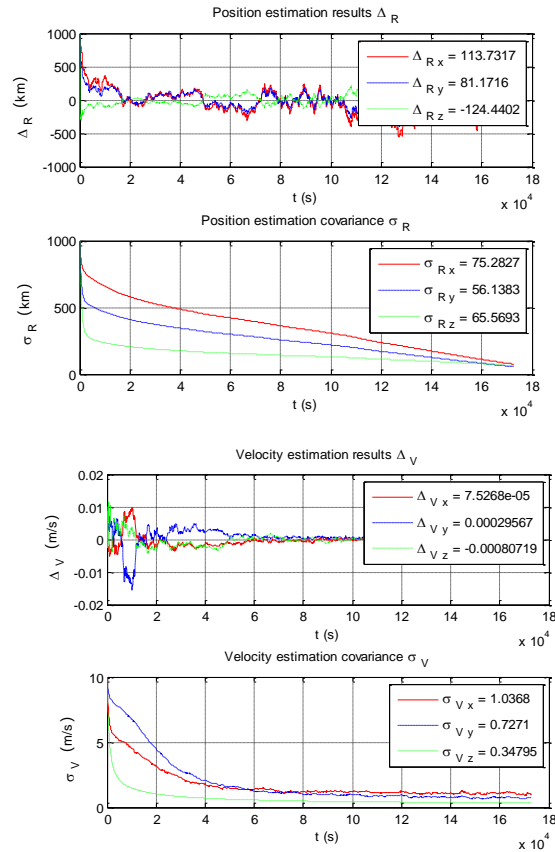
Fig.2 Analysis results of angle-velocity combined navigation accuracy (left: position error, right: velocity error)

Tab.2   Comparison of Navigation Accuracy

Navigation Method	Position error（3σ, km）	Velocity error（3σ, m/s）
Angular Navigation	254.82	4.32
Angle-Velocity Combined Navigation	114.54	1.31

Results show that due to the introduction of astronomical spectrum velocity measurement, of position and velocity estimations are both more accurate and converge faster in the combined navigation system, which makes the system more precise and reliable.

4   Astronomical spectral velocity measurement instrument

The astronomical spectral velocity measurement instrument is a key equipment for the implementation of angle-velocity combined navigation. Here we introduce an instrument scheme based on atomic frequency discriminator. The instrument consists three main units: atomic frequency discrimination unit, photoelectric detection unit and control unit. The instrument utilizes the atomic frequency discrimination mechanism to measure the Doppler frequency shift of solar spectra to obtain the line-of-sight velocity relative to the Sun. When a Doppler shift occurs the red wing signal strength and the blue wing signal strength from the solar spectrum are detected and transformed to speed value.
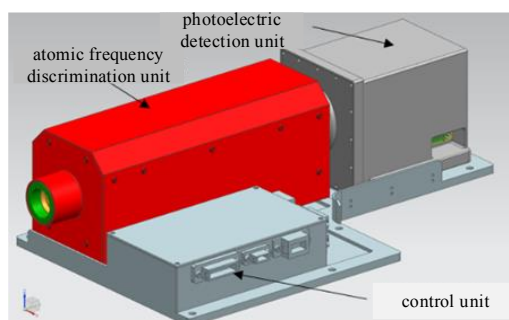
Fig. 7   Astronomical Spectral Velocity Navigation Instrument

In Oct. 2021, the astronomical spectral velocity meter was launched with the CHASE mission and conducted its first test in orbit successfully. The velocity measurement root mean square error was found to be 1.93 m/s, effectively validating the feasibility of the astronomical spectral velocity measurement method and technology. In future, a three-dimensional vector spectral velocity meter will be studied and developed to obtain inertial velocities in three directions simultaneously, which will make the astronomical spectral velocity measurement method more practical.

5 Conclusion

Deep space explorations demand higher levels of autonomous navigation capabilities for spacecraft. Current astronomical autonomous navigation methods primarily rely on angular or range information from celestial body, which cannot fully meet the continuous, real-time and high-precision autonomous navigation requirements. This paper combines the astronomical spectral vector velocity measurement method with angle measurement method to form a combined autonomous navigation system. The observability and navigation accuracy of the combined system are simulated and analyzed. Results show that the introduction of astronomical spectral velocity measurement information can significantly improve the position and velocity estimation accuracy. This method is expected to provide a novel autonomous navigation approach for future deep space exploration missions.

References:

[1]   Zhang Wei, Chen Xiao, You Wei, et al. New autonomous navigation method based on redshift [J]. Aerospace Shanghai, 2013(2): 32-33.

[2]   W. Wang, B. D. Fang, W. Zhang, Deceleration options for a robotic interstellar spacecraft[C]. 64th International Astronautical Congress. Beijing, 2013.

[3]   X. Chen, W. Zhang, W. Wang. Preliminary Research of Mars Local Navigation Constellation[C]. 64th International Astronautical Congress. Beijing, 2013.

[4]   W. Zhang. Celestial Navigation Theory and Method for Deep Space Exploration[M]. Science Press, 2017.

[5]   CHEN Xiao ZHANG Wei PENG Yuming. Mars Entry Real-time Navigation Based on Orbiter Tracking Data[J]. SPACECRAFT RECOVERY & REMOTE SENSING, 2012, 33(6): 17-23.

[6]   YOU Wei, ZHANG Wei, MA Guang-fu. Analysis on error propagation in velocity vector synthesis of deep-space celestial autonomous navigation based on radial velocity measurement[J]. Journal of Chinese Inertial Technology, 2017, 25(3): 338-342.

[7]   Zhang W, Huang Q L, Chen X. Autonomous celestial navigation method of asteroid probe based on angle measurement and velocity measurement [J]. Sci Sin-Phys Mech Astron, 2019, 49: 084510.

[8]   Huang Q L, Chen X, You W., Error analysis of the autonomous celestial navigation based on the spectrum velocity measurement, The 6th CSA/IAA Conference on Advanced Space Technology, Shanghai, China, 2015.

[9] Xiao Chen, Zhaowei Sun, Qinglong Huang, Ming Liu, Wei Zhang. Hardware in-the-loop Simulation of Celestial Angle and Velocity Measurement Integrated Navigation System[J]. The 37th Chinese Control Conference, Wuhan, China, 2018.

[10] J.Harlander, R.J.Reynolds, and F.L.Roesler. Spatial Heterodyne Spectroscopy for the Exploration of Diffuse Interstellar Emission Lines at Far-Ultraviolet Wavelengths[J], The Astrophysical Journal, 1992: 730-740.

[11] Christoph R. Englert, David D. Babcock and John M. Harlander. Doppler asymmetric spatial heterodyne spectroscopy(DASH): concept and experimental demonstration[J], Applied Optics, 2007, Vol. 46, No. 29, 7297-7307.

[12] Hermann R, Krener A J. Nonlinear controllability and observability[J]．IEEE Trans Autom Control, 1977, AC-22(5) : 728-740.

[13] Lee E B, Markus L. Foundations of optimal control theory [M]. New York: Wiley, 1967: 27-39.

A Tightly Coupled GNSS/SINS Navigation Algorithm Assisted by Zenith-Oriented Fisheye Camera

Zhou Ruiyang[1], Meng Fanchen[1*], Sun Junren[2], Nan Zihan[1], Bo Fan[1], Feng Huishuo[1]

(1. Beijing Institute of Aerospace Control Devices, Beijing 100039, China;

2. School of Artificial Intelligence, China University of Mining and Technology-Beijing, Beijing, 100091, China)

Abstract: To address the challenge of reduced positioning accuracy in GNSS/SINS navigation systems due to non-line-of-sight (NLoS) satellite signals, particularly in scenarios with temporary GNSS denial, this paper proposes a tightly coupled GNSS/SINS navigation algorithm augmented by a zenith-oriented fisheye camera. The method leverages the camera's wide-angle view to capture images of the satellite signal reception area, then uses a neural network model to segment regions of signal blockage, detect and filter potential NLoS signals, so as to enhance robustness of tightly coupled GNSS/SINS algorithm. Experimental results in an urban canyon environment demonstrate that the proposed algorithm effectively detects and excludes NLoS satellites, improving convergence speed by 44.3% and stability during environmental transitions by 37.7% compared to conventional GNSS/SINS methods.

Key words: Tightly coupled GNSS/SINS; GNSS denial; NLoS; Sky segmentation net

1 Introduction

The development of multi-source autonomous navigation technology has further promoted the development of unmanned systems such as unmanned vehicles and unmanned aerial vehicles. Since the application scenarios and environments of carriers such as unmanned vehicles and unmanned aerial vehicles are complex and diverse, an accurate and reliable navigation system is an important link for them to perform tasks autonomously. In an open area. However, in practical applications, due to the existence of harsh physical environments and the decline of satellite navigation subjective interference signals, the signal strength of autonomous navigation terminals is weak, causing signal loop unlocking, and then phenomena such as reduced navigation accuracy or even failure occur. The current mainstream research on mobile carrier positioning schemes has the following problems: (1) In urban high-rise areas, satellite signals are blocked, there are multipath effects, non-line-of-sight (NLOS) signals, etc., and the positioning accuracy of GNSS is not sufficient; (2) RTK technology can greatly improve positioning accuracy, but not all areas have accurate RTK signal infrastructure; (3) Urban vehicles have problems such as severe degradation of navigation and positioning performance after entering underground/above-ground parking lots and other scenarios, long cold start positioning time of the navigation system after returning to the ground, and weak

positioning ability in occluded environments such as urban canyons.

With the development of machine vision technology, vision sensors are widely used in the detection of wide-angle NLOS signals. Using vision sensors such as fisheye cameras to collect the environment and sky images where the carrier is located, after performing image segmentation algorithms and satellite star map projections, the satellite navigation signal is classified in terms of quality. The research focus of visual-aided autonomous navigation systems mainly lies in image key frame acquisition and segmentation strategies, and multi-source information fusion of vision-inertia-satellite. In terms of image key frame acquisition and segmentation, literature [1] proposes a dynamic visual aided navigation method based on semantic segmentation and geometric constraints. It uses a semantic segmentation network and dynamic trust degree to generate prior dynamic masks and extract feature points, formulates a feature point elimination strategy for elimination, and uses the remaining static feature points for tracking and positioning; literature [2] uses Canny edge detection and FloodFill algorithm to segment sky images; literature [3] based on the factor graph model tightly combined navigation system, uses different weights to process NLOS and LOS measurements. In terms of multi-source information fusion of vision-inertia-satellite, literature [4] combines semantic information and depth information to identify the true dynamics of feature points, and combines inertial information to fine-tune camera pose estimation; literature [5] addressed the NLoS signal, proposing a sky-view image segmentation algorithm based on a fully convolutional network (FCN); literature [6] processes the pose, velocity and IMU bias errors through graph optimization methods to improve the estimation accuracy of carrier pose and three-dimensional spatial point position.

Aiming at the problem that non-line-of-sight satellite signals reduce the positioning performance of satellite navigation receivers in the GNSS/INS tightly integrated system, this paper uses a fisheye camera to sense the wide-angle image in the zenith direction of the satellite signal receiving antenna, uses the Sky-Segmentation network to segment the signal shielding area, improve the accuracy and robustness of semantic segmentation, detect and process possible non-line-of-sight satellite signals, and finally realize anti-non-line-of-sight GNSS/SINS tightly integrated navigation solution calculation, and proposes a zenith wide-angle visual aided GNSS/SINS tightly integrated algorithm.

2 Overall Framework of the System

The overall framework of the tightly integrated GNSS/SINS system with wide-angle vision in the zenith direction is shown in Fig 1. This framework adds a vision-aided module on the basis of the tightly integrated GNSS/SINS framework to reduce the impact of non-line-of-sight(NLOS) satellite signals on the positioning results. A fisheye camera is used to collect visual images in the zenith direction of the receiver. After pre-processing, the images are sent to the sky segmentation model network(SSMNet) for pixel-level semantic segmentation processing to divide the range where satellite signals are occluded in the image. Combined with the known spatial orientation of the satellite relative to the receiver, the serial number of NLOS satellites is determined, and the observation information of the satellite navigation is reconstructed accordingly. Finally, the newly reconstructed observation set is fused with the observation information predicted by the INS to estimate and compensate the measurement errors and solution errors of the INS, and the positioning

result is output.
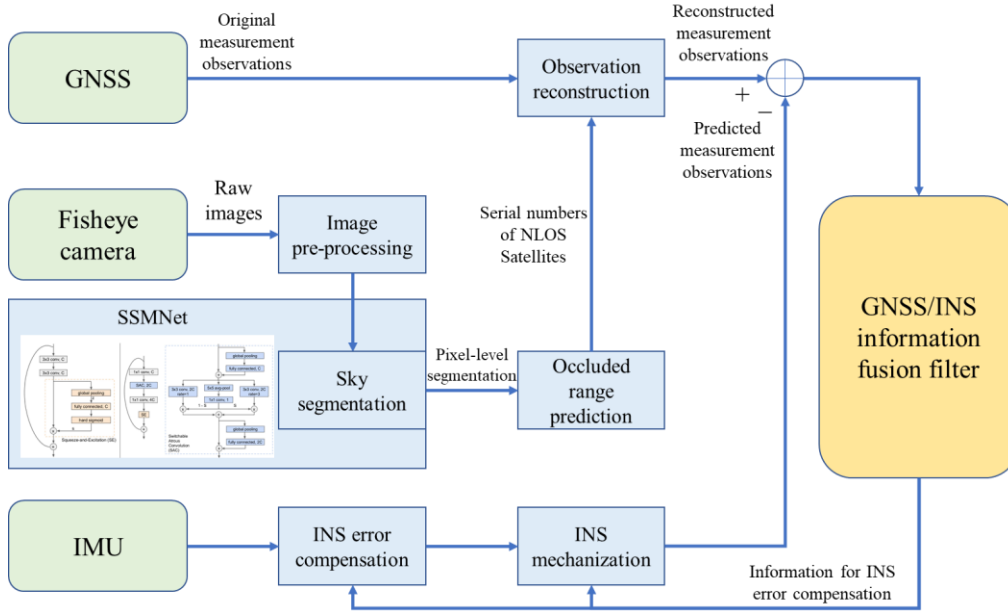
Fig.1 The overall framework of the system

3 Mathematic Model and Algorithm

3.1 Information Fusion Model of the Tightly Integrated GNSS/INS System

In the information fusion model of the tightly integrated GNSS/INS system, the system state vector at the k-th epoch is a 15-dimensional vector composed of INS parameter errors and GNSS clock errors:

$$\delta\boldsymbol{\varphi} = [\delta\varphi_E \ \delta\varphi_N \ \delta\varphi_U]^{\mathrm{T}} \tag{1}$$

Among them, $\delta\boldsymbol{\varphi} = [\delta\varphi_E \ \delta\varphi_N \ \delta\varphi_U]^{\mathrm{T}}$ represents the error vector of pitch angle, roll angle and heading angle. $\delta\boldsymbol{v}^n = [\delta v_E \ \delta v_N \ \delta v_U]^{\mathrm{T}}$ represents the velocity error vector in the east, north and up directions. $\delta\boldsymbol{p} = [\delta L \ \delta\lambda \ \delta h]^{\mathrm{T}}$ represents the position error vector composed of latitude, longitude and height. ε^b and $\varDelta^b$ respectively represent the gyroscope and accelerometer bias vectors. δb_{clk} and δd_{clk} are the clock error and clock drift error of the satellite navigation reference system respectively.

After linearization, the state equation and measurement equation of the tightly integrated GNSS/INS system are expressed as:

$$\boldsymbol{\Phi} \tag{2}$$

where $\boldsymbol{\Phi}$ represents the state transition matrix, and $\boldsymbol{W}_k$ is the system process noise vector at the k-th epoch. $\boldsymbol{Z}_k$ is the measurement vector, $\boldsymbol{H}_k$ is the measurement matrix, and the measurement

noise vector is expressed as V_k.

For the measurement equation, the measurement vector is composed of pseudorange errors and doppler errors. They can be expressed as the differences between the satellite navigation measurements and observations predicted by the inertial navigation system. Z_k is shown as follows:

$$
\begin{aligned}
\mathbf{Z}_k &= [\delta\boldsymbol{\rho}, \delta\dot{\boldsymbol{\rho}}]_k^T \\
&= \left[\delta\rho_1, \delta\rho_2, \ldots, \delta\rho_N, \delta\dot{\rho}_1, \delta\dot{\rho}_2, \ldots, \delta\dot{\rho}_N\right]_k^T
\end{aligned}
\tag{3}
$$

where $\delta\rho_i$ and $\delta\dot{\rho}_i$ respectively represent the pseudorange error and Doppler error of the i-th channel input at the k-th epoch. The measurement matrix H_k is expressed as follows:

$$
H_k = \begin{bmatrix} O_{N\times 6} & H_{\rho 1} & O_{N\times 6} & H_{\rho 2} \\ O_{N\times 3} & H_{\dot\rho 1} & O_{N\times 9} & H_{\dot\rho 2} \end{bmatrix}_k
\tag{4}
$$

where $O_{M\times N}$ is a $M\times N$ zero matrix. $H_{\rho 1}$, $H_{\dot\rho 1}$, $H_{\rho 2}$ and $H_{\rho 2}$ are expressed as follows:

$$
H_{\rho 1} = -\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}\cdot C_1, \quad H_{\rho 2} = \begin{bmatrix} 1_{N\times 1} & O_{N\times 1} \end{bmatrix}, \quad H_{\dot\rho 1} = -\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}\cdot C_2, \quad H_{\dot\rho 2} = \begin{bmatrix} O_{N\times 1} & 1_{N\times 1} \end{bmatrix}
\tag{5}
$$

$$
C_1 = \begin{vmatrix} -(R_N+h)\cos\lambda\sin L & -(R_N+h)\cos L\sin\lambda & \cos L\cos\lambda \\ -(R_N+h)\sin\lambda\sin L & (R_N+h)\cos L\cos\lambda & \cos L\sin\lambda \\ R_N(1-e^2)+h & 0 & \sin L \end{vmatrix}, \quad C_2 = \begin{bmatrix} -\sin\lambda & -\sin L\cos\lambda & \cos L\cos\lambda \\ \cos\lambda & -\sin L\sin\lambda & \cos L\sin\lambda \\ 0 & \cos L & \sin L \end{bmatrix}
\tag{6}
$$

3.2 NLOS Satellite Signal Detection and Processing Algorithm

We use fisheye camera facing the sky to analyze and determine which areas are non-occluded, and then exclude the NLOS signal on the basis of the elevation angle and azimuth angle of satellites. Based on traditional methods for identifying non-occluded areas, there exist phenomena of false detection and missed detection. Therefore, machine learning techniques are commonly used to address the segmentation problem [7]. In this paper a deep learning solution is adopted. This method uses the SSMNet to detect sky areas. The model structure is based on scaling wide residual networks and adopts the Squeeze-and-Excitation(SE) structure, effectively reducing the complexity of the model.

The SSMNet can distinguish the sky and non-sky areas within the image area and effectively

identify obstacles such as buildings and tree crowns. After partitioning the fisheye camera image according to whether the satellite signal is blocked or not, the coordinates of the partitioned image need to be transformed from the local image coordinate system to the body coordinate system. Then, using the azimuth and elevation angle information, the satellite azimuths of all available signals are mapped within the partitioned image to identify the serial numbers of NLOS satellite. Finally, the measurements from the NLOS satellites are excluded to reconstruct a reliable set of measurements. The specific algorithm steps are as follows:

1. Transform the coordinate system of the collected images.

The fisheye camera and the receiver are rigidly connected, and the fisheye camera is placed towards the zenith direction. According to the heading angle information of the receiver antenna, the partitioned image of the fisheye camera after semantic segmentation is transformed into the receiver body coordinate system.

2. Project and map the satellite azimuths of all available signals onto the partitioned image.

Assume that the heading angle provided by the INS is φ_{yaw}. Each satellite corresponds to a set of azimuth and elevation angles, and the i-th set can be represented as (A_i, E_i). The distance between the center of the projection of the i-th satellite in the partitioned image and the center point (x_0, y_0) of the image is represented as r_i, and the calculation method is as follows:

$$r_i = 2f \tan(\frac{\beta_i}{2}) \tag{7}$$

where f represents the camera focal length, β_i satisfies $\beta_i + E_i = \pi/2$, then the coordinate (x_i, y_i) of the center point of the satellite projection in the partitioned image can be calculated by the following formula:

$$\begin{cases} x_i = x_0 + r_i \cos\left(\varphi_{yaw} + A_i\right) \\ y_i = y_0 - r_i \sin\left(\varphi_{yaw} + A_i\right) \end{cases} \tag{8}$$

3. Calculate the proportion τ_i of the shaded area within the projection range.

Assume that the image collected by the fisheye lens is circular, and assume that this satellite is a completely visible satellite ($\tau_i = 0$), the pixel value of the satellite signal shaded area in the partitioned image is 1, and the non - shaded sky area is 0. Taking the coordinates (x_i, y_i) of the center point of the projection as the center and a certain fixed value as the radius, count the pixel values of m pixel points within the circle, and calculate the proportion τ_i of the shaded area within the projection range accordingly:

$$\tau_i = \frac{\sum\limits_{j=1}^{m} I\left(x_{i,j}, y_{i,j}\right)}{m} \tag{9}$$

where $I\left(x_{i,j}, y_{i,j}\right)$ represents the pixel value of the point with coordinates $\left(x_{i,j}, y_{i,j}\right)$ in the search area of the i-th satellite.

4. Reconstruct the observation set.

According to the satellite signal occlusion situation, the measurements of the satellite corresponding to $\tau_i = 1$ are removed, and for the reconstructed observation set, the adaptive measurement noise weighted matrix Ω_k of the pseudorange at the k-th epoch is constructed:

$$\Omega_k = diag\{\omega_1, \omega_2, ..., \omega_N\}$$
(10)

where ω_i is calculated as:

$$\omega_i = -\frac{1}{\tau_i - 1}$$
(11)

The pseudorange adaptive measurement noise weighted matrix Ω_k directly participates in the kalman filtering calculation process of the integration filter to adaptively adjust the measurement noise covariance matrix:

$$\mathbf{R}_{k,adp} = \begin{bmatrix} \Omega_k & \mathbf{0}_{N \times N} \\ \mathbf{0}_{N \times N} & \mathbf{I}_{N \times N} \end{bmatrix} \mathbf{R}_k$$
(12)

4 Experiments

4.1 Experimental platform

The navigation algorithm proposed in this study was designed for a multi-sensor system integrating GNSS, INS, and visual sensors, and was tested on a modified UGV experimental platform. As shown in Fig. 2, the platform comprises a data processing unit (Nvidia Jetson Orin), a stereo depth camera (D435i), a GNSS receiver with an integrated IMU (FX20), a 408 mmWave radar, a Hikvision camera, a fisheye camera, and a four-wheel-drive mobile chassis (Agilex Scout Mini). The Scout Mini is a high-speed, all-terrain UGV featuring four-wheel differential drive, independent suspension, and a maximum speed of 10 km/h, providing a versatile and robust platform for navigation experiments in complex environments.

Fig.2 Configuration of modified UGV as experimental platform

4.2 Experimental Verification

The fisheye-camera-assisted tightly coupled GNSS/INS algorithm proposed in this study aims to improve navigation system performance during short periods of GNSS signal loss. To evaluate the algorithm's effectiveness, experiments were conducted both inside and outside an underground garage, where the UGV followed a predefined route, as illustrated in Fig. 3.
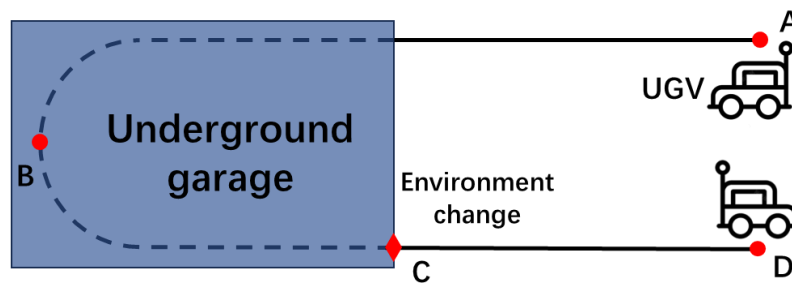
Fig.3 Experimental setup

The experiment process was as follows:

(1) Preparation: sensors calibration, network connectivity check and module synchronization;

(2) The UGV begins route at an outdoor starting point A, moving towards a designated point B within an underground garage at constant speed of 5km/h, remains stationary on point B for 60 seconds before returning to the outdoor starting point A.

(3) Repeat step 2 with a constant UGV speed of 10km/h.

(4) Data processing and algorithm verification, analyze positioning accuracy and reliability.

After experiencing short-term GNSS denial in the underground garage, traditional GNSS/INS navigation system exhibited a noticeable delay in convergence, as no GNSS signals were available to correct accumulated positioning errors. As illustrated in Fig. 4, the GNSS/INS trajectory significantly deviated from the ground truth once the UGV entered the garage. At the transition point where environment changed, the proposed fisheye-camera-assisted GNSS/INS algorithm adaptively detected NLoS satellites and utilized the SSMNet model to filter unreliable satellite signals. This adaptive capability allowed the system to maintain better positional accuracy and avoid error divergence during GNSS-denied phases.
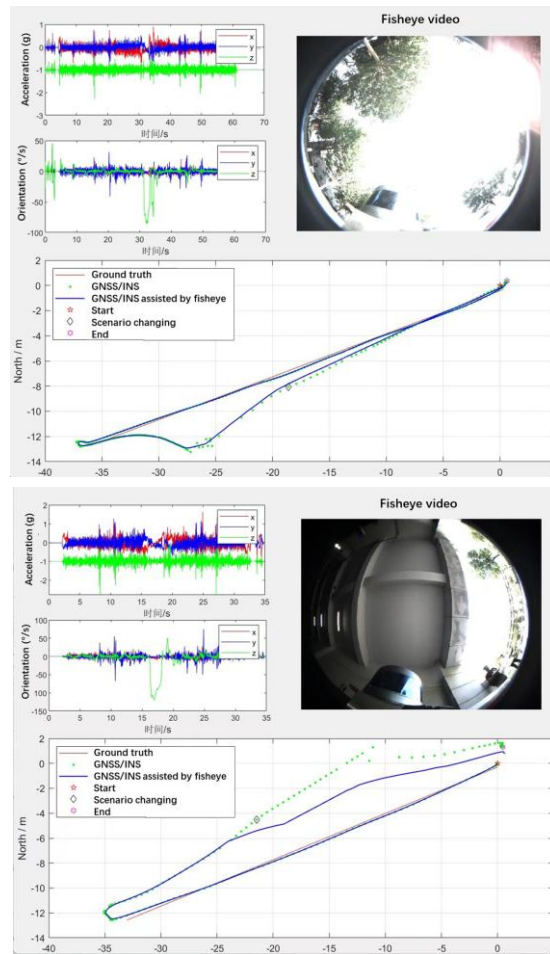
(a) UGV moving speed 5km/h　　　　　　　(b) UGV moving speed 10km/h

Fig.4　Fisheye-assisted GNSS/INS navigation performance during GNSS denial

　　Tab. 1 presents the results of experiments conducted at different UGV velocities. The proposed GNSS/SINS algorithm demonstrated a significant improvement in convergence speed, with gains of up to 66.67%, as well as a notable reduction in positioning error covariance after environmental changes, achieving improvements of up to 53.93%. These results validate the effectiveness of the proposed algorithm in reducing error covariance and enhancing the resilience of the navigation system in challenging GNSS-denied environments.

Tab.1　Experiment results

Velocity	Algorithm	Convergence time	Improvement	Covariance	Improvement
5km/h	GNSS/INS	1.2s	66.667%	1.593m^2	21.497%
	GNSS/INS/Fisheye	0.4s		1.250m^2	
10km/h	GNSS/INS	10s	22%	11.331m^2	53.929%
	GNSS/INS/Fisheye	7.8s		5.220m^2	
Average improvement		44.3%		37.7%	

5　Conclusion

　　This study addresses the challenge of reduced positioning accuracy of GNSS/SINS systems in

GNSS-denied environments, such as urban area and indoor environment, where non-line-of-sight (NLoS) signals are common. To mitigate this issue, we propose an enhanced tightly coupled GNSS/SINS algorithm that incorporates zenith-oriented visual assistance, leveraging fisheye camera and intelligent techniques. We presented the overall framework, theoretical model, and solution approach for this GNSS/SINS integration and developed an algorithm to detect NLOS satellite signals. Experimental results indicate that the proposed algorithm effectively identifies and mitigates the impact of NLoS signals in GNSS-denied environments, resulting in up to a 44.3% improvement in convergence speed and 37.7% in stability during environment changes over traditional tightly coupled GNSS/SINS methods. This advancement demonstrates the potential of visual-aided algorithms to significantly enhance navigation reliability in challenging environments.

References:

[1] Campos C, Elvira R, Rodríguez J J G, et al. Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.

[2] GAKNEP, O'KEEFE K. Tightly-coupled GNSS/vision using a sky-pointing camera for vehicle navigation in urban areas [J]. Sensors, 2018, 18(4): 1244.

[3] WEN Weisong, BAI Xiwei, KAN Y C, et al. Tightly-coupled GNSS/INS integration via factor graph and aided by fish-eye camera[J]. IEEE Transactions on Vehicular Technology,2019,68(11):10651-10662.

[4] Yuan X, Chen S. Sad-slam: A visual slam based on semantic and depth information[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2020: 4930-4935.

[5] Wang, J., Xu, B., Liu, J., Gao, K., & Zhang, S. (2024). Sky-GVIO: Enhanced GNSS/INS/Vision navigation with FCN-based sky segmentation in urban canyon. Remote Sensing, 16(20), 3785.

[6] W. Wen, G. Zhang and L. -T. Hsu, "Object-Detection-Aided GNSS and Its Integration With Lidar in Highly Urbanized Areas," in IEEE Intelligent Transportation Systems Magazine, 2020, 12(3): 53-69.

[7] P. Xu, G. Zhang, B. Yang and L. -T. Hsu, "Machine Learning in GNSS Multipath/NLOS Mitigation: Review and Benchmark," in IEEE Aerospace and Electronic Systems Magazine, 2024, 39(9): 26-44.

A Low-Cost Combined GNSS/MEMS IMU in-Vehicle Navigation Method Based on HKC

Li Leilei, Wang Yang, Huang Qihai

(College of Aerospace Engineering, Chongqing University, Chongqing 400044)

Abstract: Aiming at the problem that the positioning accuracy of low-cost GNSS / MEMS IMU vehicle integrated navigation system decreases rapidly in GNSS signal denial environment, a set of low-cost GNSS / IMU vehicle integrated navigation algorithm based on holonomic kinematic constraints (HKC) is proposed by exploring the motion characteristics of vehicles. In the stationary state, zero position, zero speed and zero heading constraints are constructed to suppress the drift of the navigation state at rest. In the motion state, the non-holonomic constraint (NHC) measurement equation including the installation angle error and the lever arm error is established to realise the online estimation of the installation angle and the lever arm. At the same time, the forward speed predicted by the speed estimation network based on the long short-term memory neural network is used to complete the NHC into a complete speed constraint, thereby improving the positioning accuracy under the condition of long-term GNSS signal denial environment. The real road test results show that compared with the traditional NHC algorithm, the plane positioning accuracy is improved by about 47.3% after estimating the installation parameters, and the plane positioning accuracy is improved by about 28.1% after adding the LSTM speed estimation network.

 Key words: vehicle integrated navigation; mico electro mechanical system inertial measurement unit; kinematic constraint; neural network

An inertial navigation system (INS) employs real-time data collected from accelerometers and gyroscopes to integrate and calculate the navigation status of the carrier, including position, velocity, attitude, and other relevant variables. It is characterised by full autonomy, strong concealment, and high short-term accuracy. However, during the recursive process, the system is significantly affected by noise, resulting in long-term poor accuracy and an increased susceptibility to cumulative errors[1]. The Global Navigation Satellite System (GNSS) employs a network of satellites with known orbital positions to achieve real-time positioning of objects on the Earth's surface through the principle of ranging intersection. It offers the benefits of high-precision positioning, extensive coverage, reliable performance in all weather conditions, and the absence of cumulative errors. However, the frequency of update for GNSS data is relatively low, which limits its ability to provide accurate attitude information. In the event of obstruction of satellite signals, the positioning accuracy will rapidly decrease or become entirely unusable. Consequently, the superior complementarity of performance makes the GNSS/INS integrated navigation system one of the most widely used

navigation integration methods. However, although traditional platform-based IMUs and high-end tactical-grade IMUs have high accuracy, their high cost and large installation space requirements severely limit their application in the field of civilian vehicle navigation.

With the development of micro-electro-mechanical technology, the accuracy of MEMS IMU is constantly improving, while the cost is also lower, and the required installation space and carrier load are smaller, making it more suitable for the field of in-car navigation. Due to the high noise level of MEMS IMU, the positioning accuracy of GNSS/MEMS IMU integrated navigation system will rapidly decrease due to the lack of GNSS measurement in GNSS lock-out environment. Therefore, how to improve the positioning accuracy of GNSS/MEMS IMU integrated navigation system in satellite lock out environment has important practical significance and technical value [3].

Although adding redundant sensors can greatly improve the positioning accuracy of integrated navigation systems under GNSS lock failure conditions, adding redundant sensors means an increase in cost or infrastructure construction, and it is not realistic to add redundant sensors in many engineering environments. Therefore, constructing specialised constraints based on specific application scenarios is one of the most practical and feasible methods to improve the positioning accuracy of the system. For in-car scenarios, special constraint conditions can be constructed by studying the motion state and the motion model of the vehicle. The stationary state is a very special state of motion during the operation of a vehicle. Theoretically, all navigation states of the vehicle should remain unchanged. However, in actual navigation, due to the high noise and high offset characteristics of low-cost MEMS IMUs, the stationary state of the carrier, especially after a long period of standstill, will cause significant offset in various navigation parameters. The Zero Velocity Update (ZUPT) and Zero Heading Update (ZIHR) algorithms effectively suppress the accumulation of velocity and heading errors during the stationary period by using the stationary velocity and heading at the moment before stationary as virtual observations throughout the stationary period[4]. In the field of pedestrian positioning, the Zero Angular Rate Update (ZARU) algorithm is also commonly used to limit the divergence of the gyroscope bias in a stationary state [5]. Vehicle NHC is also a navigation assistance method that does not require the introduction of additional sensors. Reference [6] investigated the influence of the lever arm between the IMU and the NHC operating point on the performance of NHC. However, in the actual vehicle environment, in addition to the influence of the NHC lever arm, the installation error between sensors and supports is also an important factor that affects whether the NHC performance can be fully utilised. Therefore, the calibration research of sensor installation error also has important practical significance [7].

In the actual motion process of vehicles, the occurrence, duration and termination of the stationary state of the vehicle are difficult to predict, and motion is always the most important working state of the vehicle. However, NHC will gradually reduce the positioning accuracy of the integrated navigation system due to the lack of forward velocity observation under long-term GNSS lock conditions. Therefore, maintaining the low-cost accuracy of integrated navigation under long-term GNSS lock conditions has important practical significance and technical value. In recent years, neural networks (NNs) have attracted much attention due to their powerful nonlinear modelling capabilities. The complex non-linear relationship between INS and vehicle motion state in the in-vehicle environment provides new opportunities for low-cost accuracy maintenance of integrated

navigation under long-term GNSS lock conditions. Reference [8] proposed two architectures based on multi-layer perceptron (MLP) for position update (PUA) and position velocity update (PVUA), and verified through actual vehicle experiments that the positioning accuracy of the two architectures is superior to traditional combined navigation modes. Compared with MLP neural network, radial basis function (RBF) neural network has the characteristics of simple structure and simple learning rules, and is more suitable for real-time processing. Reference [9] uses RBF neural network to predict and correct the cumulative error of INS under satellite loss conditions, thereby improving the positioning accuracy under satellite loss conditions. However, the above-mentioned neural networks often do not pay enough attention to the information of the past time when processing time-series data, and the process of in-car navigation is a typical time-series process. Therefore, recurrent neural networks (RNNs) and convolutional neural networks (CNNs) have attracted much attention from researchers due to their excellent performance in processing time-series data. Reference [10] proposed a GNSS/INS integrated navigation method based on RNN neural network, which estimates the error of the inertial navigation system through the excellent memory function of RNN, and improves the positioning accuracy by 60% compared to the classical EKF scheme. Reference [11] used CNN neural network to quickly extract multi-dimensional sequence features to predict the position increment during GNSS interruption. The high accuracy and training efficiency of the algorithm were verified in real road tests. However, the gradient vanishing and exploding problems that exist during the training process of RNN and CNN neural networks greatly affect the stability and reliability of the networks. Therefore, based on the RNN neural network, researchers have proposed an innovative gate structure Long Short Term Neural Network (LSTM), which can greatly improve the long-term dependency problem of RNN and make the modelling of sequence data more accurate [12].

Therefore, based on the GNSS/INS loose combination model, this article integrates the static constraints and NHC of vehicles to improve the positioning accuracy of vehicles under satellite-free conditions, and realises the online estimation of sensor installation parameters. At the same time, the LSTM neural network is combined with the NHC model to use the forward velocity predicted by the LSTM velocity estimation network to complete the NHC into complete velocity constraints, thereby improving the positioning accuracy under long-term satellite lockout conditions.

1　Low cost GNSS/MEMS IMU vehicle integrated navigation algorithm

1.1　Algorithm framework

The overall results of the integrated navigation system in this article are shown in Figure 1. Based on the integrated GNSS/INS navigation, GNSS-assisted INS navigation is used in open-air outdoor environments, and parameters such as zero bias, installation angle and arm between IMU and vehicle motion centre are estimated online. After GNSS lock loss, when the vehicle is detected to be in a stationary state, static constraints are constructed to suppress the divergence of related navigation states. When the vehicle is moving, the sensor output and navigation state are converted in real time to the motion centre using the online estimated installation angle and NHC arm. The forward velocity predicted by the LSTM forward velocity estimation network is then used to

complete the NHC into integrity velocity constraints, thereby improving positioning accuracy after GNSS lock loss.
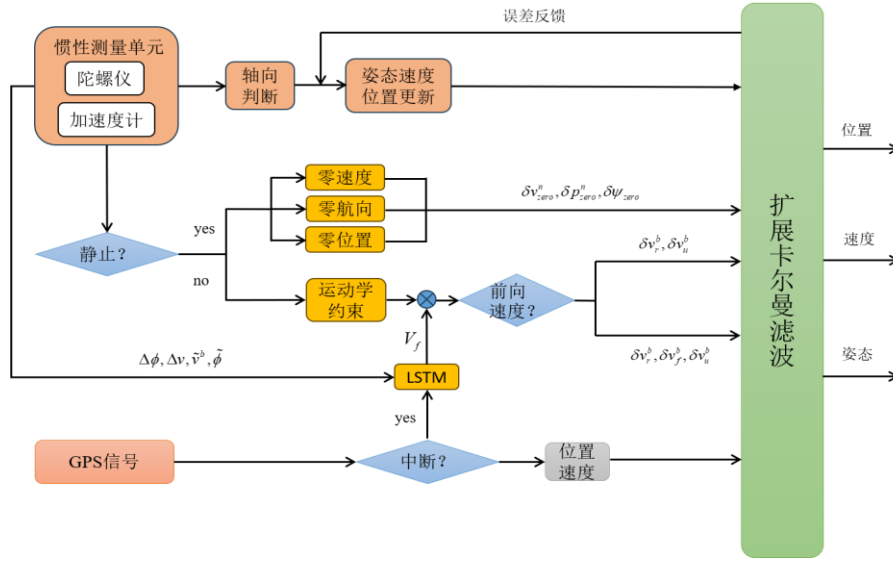
Fig.1 Integrated navigation frame

1.2 GNSS/INS Pine Combination System

According to reference [13], the error transfer equation of the Strapdown Inertial Navigation System (SINS) in the Northeast Sky Navigation Coordinate System n is as follows：：

$$\dot{\boldsymbol{\phi}} = \boldsymbol{F}_{aa}\boldsymbol{\phi} + \boldsymbol{F}_{av}\delta\boldsymbol{v}^n + \boldsymbol{F}_{ap}\delta\boldsymbol{p} - \boldsymbol{C}_b^n\delta\boldsymbol{b}_g$$
$$\delta\dot{\boldsymbol{v}}^n = \boldsymbol{F}_{va}\boldsymbol{\phi} + \boldsymbol{F}_{vv}\delta\boldsymbol{v}^n + \boldsymbol{F}_{vp}\delta\boldsymbol{p} + \boldsymbol{C}_b^n\delta\boldsymbol{b}_a \qquad (13)$$
$$\delta\dot{\boldsymbol{p}} = \boldsymbol{F}_{pv}\delta\boldsymbol{v}^n + \boldsymbol{F}_{pp}\delta\boldsymbol{p}$$

Among them，$\boldsymbol{\phi}$ is the misalignment angle，$\delta\boldsymbol{v}^n$ is the northeast sky velocity error,，$\delta\boldsymbol{p}$ is the position error，$\delta\boldsymbol{b}_g$ is the gyroscope bias，$\delta\boldsymbol{b}_a$ is the accelerometer bias，$\boldsymbol{F}_{**}$ is the coupling coefficient between various error terms，$\boldsymbol{C}_b^n$ is the rotation matrix from the navigation system n-frame to the car system b-frame.

Therefore, based on the error transfer equation of SINS, the system state equation of GNSS/INS loose combination system is established：

$$\dot{\boldsymbol{X}}(t) = \boldsymbol{F}(t)\boldsymbol{X}(t) + \boldsymbol{G}(t)\boldsymbol{w}(t) \qquad (14)$$

Finally, three-dimensional misalignment angle $\boldsymbol{\phi}$, velocity error $\delta\boldsymbol{v}^n$, position error $\delta\boldsymbol{p}$, gyroscope bias $\delta\boldsymbol{b}_g$, and accelerometer bias $\delta\boldsymbol{b}_a$ were selected as system state variables $\boldsymbol{X}$. $\boldsymbol{G}$ Assign a matrix to the noise，$\boldsymbol{w}$ is the system noise, assuming that the noise follows a Gaussian distribution with zero mean.

In the loose combination model, the position and velocity calculated by the GNSS system are

used as fixed measurements for the integrated navigation system, and the system measurement equation is as follows：

$$Z_k = H_k X_k + v_k \tag{15}$$

Among them, Z_k is a 6×1-dimensional GNSS position velocity measurement vector, H_k is a measurement coefficient matrix, v_k is a noise matrix that follows a Gaussian distribution with zero mean.

$$Z_k = \begin{bmatrix} p_{INS} - p_{GNSS} \\ v_{INS} - v_{GNSS} \end{bmatrix} \tag{16}$$

$$H_k = \begin{bmatrix} 0_{3\times3} & 0_{3\times3} & I_{3\times3} & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & I_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \end{bmatrix} \tag{17}$$

2　Kinematic constraints based on vehicle motion characteristics

2.1　Virtual measurement based on static conditions

The motion state of land vehicles can be simply divided into two types: stationary and moving. The stationary state is a relatively special state of motion in vehicle movement. In an ideal, error-free situation, all navigation states of the vehicle, such as attitude, speed, position, etc., should remain unchanged. However, in the actual vehicle environment, due to high sensor noise and environmental influences, the stationary state, especially the navigation state under long-term standstill, will continue to deviate, resulting in a decrease in the positioning accuracy of the integrated navigation system.

Zero Speed Correction Technique (ZUPT) is a method that can effectively suppress the accumulation of speed errors at rest. After detecting that the vehicle is stationary, it can be assumed that the northeast sky speed of the vehicle is zero speed. At this time, a zero speed Kalman measurement can be constructed, and the measurement vector is:

$$\delta Z_v = v_{INS}^n - \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T \tag{18}$$

The measurement coefficient matrix for zero velocity observation is：

$$H_v = \begin{bmatrix} 0_{3\times3} & I_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \end{bmatrix} \tag{19}$$

Zero Position Constraint Technique (ZPCT) can effectively suppress the displacement of position in a stationary state. When the vehicle is detected to be stationary, the position before the stationary moment is saved. The position deviation during the stationary period is caused by sensor noise or vehicle parking noise. At this time, a virtual measurement of the static zero position can be constructed, and the measurement vector is：

$$\boldsymbol{\delta Z}_p = \boldsymbol{p}_{INS} - \boldsymbol{p}_{zero} \tag{20}$$

The measurement coefficient matrix for zero position observation is：

$$\boldsymbol{H}_p = \begin{bmatrix} \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} \end{bmatrix} \tag{21}$$

Similarly, the accumulation of heading error after the vehicle comes to a standstill is also caused by sensor noise or vehicle parking noise. Therefore, by storing the heading at the moment before the vehicle comes to a standstill, heading measurement during the stationary period can be constructed to suppress heading drift caused by prolonged stationary time. At this point, the measurement vector is：

$$Z_\psi = \psi_{INS} - \psi_{zero} \tag{22}$$

Due to the coupling relationship between attitudes, the measurement coefficient matrix for zero heading observation is no longer an identity matrix. Considering：

$$\begin{cases} \boldsymbol{C}_b^{n'} = \left(\boldsymbol{I} - \boldsymbol{\phi} \times\right) \boldsymbol{C}_b^n \\ \psi = -\arctan\left(C_{12} / C_{22}\right) \end{cases} \tag{23}$$

In the formula, C_{**} represents the corresponding element in the rotation matrix $\boldsymbol{C}_b^n$, By expanding and simplifying the elements in the above formula, we can obtain:

$$\tilde{\psi} = -\arctan\left(\frac{C_{12} + C_{22}\tilde{\psi} - C_{32}\tilde{\gamma}}{-C_{12}\tilde{\psi} + C_{22} + C_{32}\tilde{\theta}}\right) \tag{24}$$

By simultaneously differentiating the left and right sides of the above equation, simplification can be achieved：

$$\partial \tilde{\psi} = \frac{\delta f}{\delta \tilde{\theta}} \partial \tilde{\theta} + \frac{\delta f}{\delta \tilde{\gamma}} \partial \tilde{\gamma} + \frac{\delta f}{\delta \tilde{\psi}} \partial \tilde{\psi} \tag{25}$$

Among them, the coefficients in equation (13) are expressed as follows：

$$\frac{\delta f}{\delta \tilde{\theta}} \approx \frac{\hat{C}_{12}\hat{C}_{32}}{\left(\hat{C}_{12}\right)^2 + \left(\hat{C}_{22}\right)^2}$$

$$\frac{\delta f}{\delta \tilde{\gamma}} \approx \frac{\hat{C}_{22}\hat{C}_{32}}{\left(\hat{C}_{12}\right)^2 + \left(\hat{C}_{22}\right)^2} \tag{26}$$

$$\frac{\delta f}{\delta \tilde{\psi}} \approx -1$$

2.2 NHC based on steering model

The motion of land vehicles on the earth's surface can be approximated as planar motion. This means that in its normal state of motion, the vehicle will generally move close to the ground without sudden bouncing or sliding. Therefore, the vehicle's lateral and celestial velocities can be close to zero during normal driving.

In actual vehicle scenarios, the carrier's lateral and celestial velocities after inertial integration are not strictly zero, which can be considered as velocity errors caused by sensor measurement errors and environmental disturbances. Therefore, the zero lateral and celestial velocities are special constraints based on the unique motion mode of land vehicles. Unlike the static constraints in section 2.1, the zero lateral and celestial velocities are a velocity constraint that still holds in motion. Due to the fact that the entire velocity constraint of the vehicle in three-dimensional space should be three-dimensional, this constraint is also known as the NHC.

The two most common motion states for vehicles in real motion are straight ahead and turning. The angular motion during straight-line motion is often small, so that all points inside the vehicle can fully satisfy the conditions of the NHC. However, during turning motion, due to the large angular motion of the vehicle itself, the lateral and celestial velocities in most parts of the vehicle space are no longer strictly zero due to the Coriolis effect, which affects the performance of NHC. Considering that land vehicles usually have relatively small rotational movements in the roll direction, which can be ignored, the following will focus on analysing the motion characteristics of vehicles in horizontal and pitch control in order to fully utilise the performance of NHC.

The simplified schematic diagram of the motion of a vehicle during horizontal turning is shown in Figure 2. The most commonly used steering model for passenger cars is the Ackermann steering model, which means that the front wheels turn while the rear wheels do not deflect during the turning process. Therefore, the trajectory of all wheels is a concentric arc and the centre of rotation is on the extension of the rear wheel line [14]. As shown in the figure, when the vehicle is turning, the speed direction of each point inside the vehicle is the tangent direction of the arc centred on point O. Except for the speed on the AB line of the rear wheel of the vehicle, the speed at other points can be decomposed into forward speed and lateral speed. However, since the IMU cannot be installed directly on the AB line, additional lateral velocity errors will be introduced when using NHC, which will affect the performance of NHC.
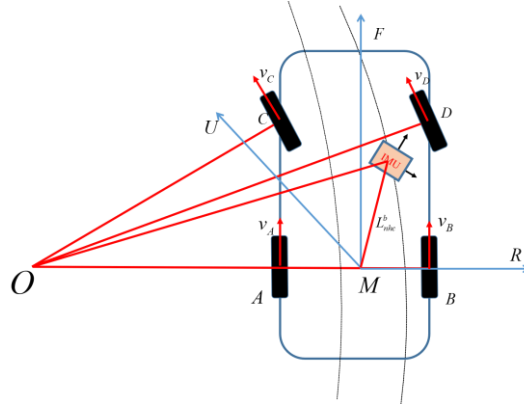
Fig.2 Horizontal turn

Similarly, a schematic diagram of the vehicle's motion during a pitch turn is shown in Figure 3. The vehicle has a pitch motion in the UBF plane. Assuming that the points of contact between the front and rear wheels of the vehicle and the curve Q are at points A and B, and that the centre of rotation is at point O above the centre of the line connecting the front and rear wheels, the direction of speed of the front and rear wheels of the vehicle is tangent to the arc AB. Obviously, there is no airspeed at points A and B. Therefore, in pitch-turning motion, the airspeed of the points in the plane passing through the centre of rotation and coinciding with the longitudinal axis of the vehicle is always zero.
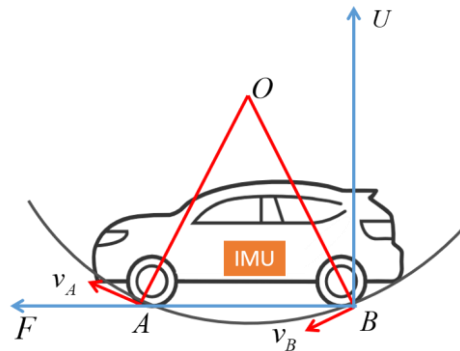
Fig.3 Pitch turn

In summary, the centre point of the vehicle's rear wheel line theoretically satisfies the constraint condition of zero lateral and celestial velocities in both horizontal and pitch steering, making it the most ideal location for vehicle NHC. It can be considered as the centre of motion of the vehicle, and therefore satisfies the constraint condition of zero lateral and celestial velocities at the centre of motion：

$$\begin{cases} v_{right} = 0 \\ v_{up} = 0 \end{cases} \tag{27}$$

The IMU can only measure the specific force and angular velocity information perceived by itself. If the IMU is rigidly attached to the centre of motion of the vehicle, it can be assumed that the IMU measures the specific force and angular velocity information of the vehicle. However, in

the actual on-board environment, there must be installation errors between the IMU and the vehicle body, and the IMU cannot be installed directly at the centre of motion in the actual installation. Therefore, the influence of the lever arm and installation errors on the speed must be considered. The formula for calculating the carrier velocity taking into account the lever arm effect can be obtained from the Coriolis effect:

$$v^b = C_s^b C_n^s v^n + \omega_{nb}^b \times L^b \tag{28}$$

Assuming that there is no error in the sensor coordinate system s, and assuming that there is a small error between the real carrier coordinate system b and the calculated carrier coordinate system b', and assuming that there is also a small error δL^b between the real NHC lever arm and the calculated NHC lever arm, then:

$$\begin{aligned} C_{b'}^b &= I + \alpha \times \\ \tilde{L}^b &= L^b + \delta L^b \end{aligned} \tag{29}$$

Substituting equation (17) into equation (16) and considering both attitude error and velocity error, the formula for calculating the vehicle speed is as follows:

$$\begin{aligned} v^b &= (I + \alpha \times) \tilde{C}_s^b \tilde{C}_n^s (I - \phi \times)(\tilde{v}^n - \delta v^n) \\ &\quad + \omega_{nb}^b \times (\tilde{L}^b - \delta L^b) \end{aligned} \tag{30}$$

Among them，$\omega_{nb}^b = \omega_{ib}^b - \omega_{in}^b$，due to the correlation between ω_{in}^b and the angular velocity of the Earth and the n-series rotation caused by the motion of the carrier, it is much smaller than the angular motion of the carrier itself, Therefore $\omega_{nb}^b \approx \omega_{ib}^b$，expand equation (18) and omit the second-order higher-order terms to obtain the measurement equation with NHC:

$$\begin{aligned} &\tilde{C}_s^b \tilde{C}_n^s \tilde{v}^n + \tilde{\omega}_{ib}^b \times \tilde{L}^b - v^b \\ &= H_\phi \phi + H_v \delta v^n + H_{\delta L} \delta L^b + H_\alpha \alpha \end{aligned} \tag{31}$$

Among them, the coefficients on the right side of the equation describe the coupling relationship between different parameters, and the calculation method is as follows:

$$\begin{cases} H_\phi = -\tilde{C}_s^b \tilde{C}_n^s (\tilde{v}^n \times) \\ H_v = \tilde{C}_s^b \tilde{C}_n^s \\ H_{\delta L} = \omega_{ib}^b \times \\ H_\alpha = (\tilde{C}_s^b \tilde{C}_n^s \tilde{v}^n) \times \end{cases} \tag{32}$$

Therefore, in the GNSS/INS loose combination state space model, three-dimensional installation angle error and three-dimensional NHC lever arm error are added as state variables to achieve online estimation of installation parameters under good satellite positioning quality. The extended state variables are:

$$X = \begin{bmatrix} \boldsymbol{\phi} & \delta \boldsymbol{v}^n & \delta \boldsymbol{p} & \delta \boldsymbol{b}_g & \delta \boldsymbol{b}_a & \delta \boldsymbol{L}^b & \boldsymbol{\alpha} \end{bmatrix}^T \tag{33}$$

The complete measurement matrix of the upper right front velocity can be obtained from equation (19) as follows:

$$H = \begin{bmatrix} H_{\boldsymbol{\phi}} & H_v & \mathbf{0}_{3\times9} & H_{\delta L} & H_{\boldsymbol{\alpha}} \end{bmatrix}^T \tag{34}$$

When there is a lack of forward velocity measurement, equation (22) only retains the elements of the first and third rows, which is a NHC measurement. When there is a forward velocity measurement, equation (22) is fully retained, and the NHC is completed as an integrity velocity constraint.

3　LSTM assisted INS/NHC integrated navigation algorithm

3.1　Long Short Term Neural Networks

Recurrent Neural Networks are widely used in the field of sequence data processing and their outstanding feature is the network structure with cyclic connections. The update of the current state is not only related to the current input, but also to the state at past moments. However, when dealing with long-term dependent sequence data, RNN is difficult to capture long-term dependency relationships due to information decay, and when the interval between input data is too large, it is prone to gradient explosion and vanishing problems. Therefore, many scholars have made improvements based on RNN. Among them, long short-term neural networks effectively improve the long-term dependency problem of RNN by introducing innovative gate structures to filter and forget prior information. The standard LSTM neural network structure is shown in the following figure:
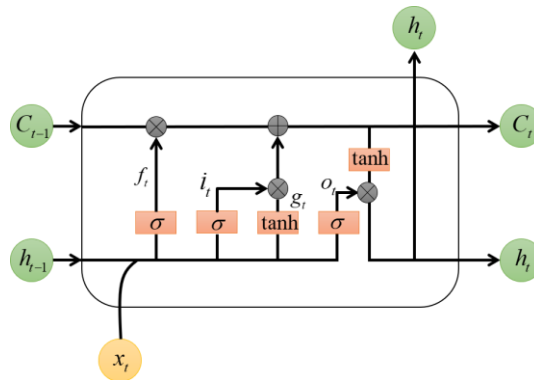
Fig.4　LSTM structure

The update formula for LSTM is as follows：

$$
\begin{aligned}
i_t &= \sigma\left(W_{ii}x_t + W_{hi}h_{t-1} + b_i\right) \\
f_t &= \sigma\left(W_{if}x_t + W_{hf}h_{t-1} + b_h\right) \\
g_t &= \tanh\left(W_{ig}x_t + W_{hg}h_{t-1} + b_g\right) \\
o_t &= \sigma\left(W_{io}x_t + W_{ho}h_{t-1} + b_o\right) \\
c_t &= f_t \odot c_{t-1} + i_t \odot g_t \\
h_t &= o_t \odot \tanh\left(c_t\right)
\end{aligned}
\tag{35}
$$

In the equation, i_t, f_t, g_t, o_t are the input gate, forget gate, cell gate, and output gate of the network, σ is the sigmoid function, and $\odot$ is the Hadamard product. During the entire network update process, the parameters of input gate i_t, forget gate f_t, and cell gate g_t at the current time are calculated based on the output h_{t-1} of the previous hidden layer and the input x_t at the current time. Input gate i_t determines which new information to store in the cell state, while forget gate f_t determines which information to discard from past states. By adding the information from the input gate and forget gate, the cell state c_t at the current time can be calculated. Finally, the value of the current hidden layer is calculated based on the current cell state. The LSTM network continuously loops through the above process, and through the mutual cooperation between gates, ultimately achieves reasonable retention of long-term and short-term information, improving the common long-term dependency problem in standard RNNs.

3.2 LSTM-assisted full speed restriction

Although the accuracy of NHC during steering motion has been improved after installation error compensation, the lack of forward velocity measurement during prolonged GNSS lock conditions can lead to the accumulation of forward velocity errors, ultimately causing the navigation system to diverge. Since the vehicle navigation and positioning process is a typical temporal process, this paper aims to use long short-term neural networks to establish a mapping relationship between IMU output, partial navigation states and vehicle forward velocity. Then, under satellite lock conditions, the forward velocity predicted by the neural network is used to complete the incomplete constraint and make it a complete three-dimensional velocity constraint, thereby improving the positioning accuracy of the navigation system under satellite lock conditions.

According to the derivation process of NHC theory, it can be known that in actual motion, vehicles can only satisfy the constraint condition of zero lateral and celestial velocities at the centre of motion. Therefore, the IMU output and calculated navigation state can be converted to the centre of motion using the estimated installation angle and NHC lever arm under good satellite observation conditions, and the converted parameters can be used as input parameters for the LSTM velocity estimation network. The forward velocity of the vehicle at the current moment is obviously related to the attitude and velocity of the previous moment, and the changes in attitude and velocity over a period of time can be represented by the output of the inertial sensor. Therefore, the angle increment, the velocity increment and the vehicle speed and attitude angle of the previous moment are finally

selected as the input parameters for the network. The specific structure of the network is as follows:
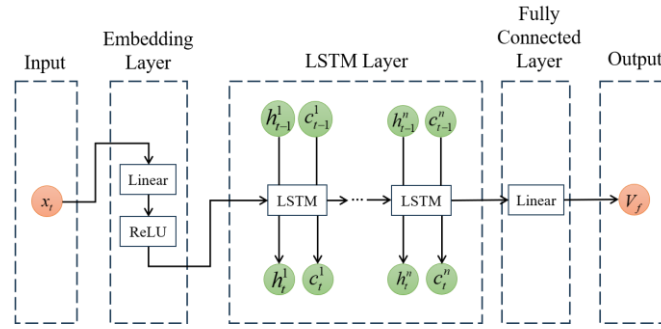
Fig.5 Forward speed estimation network

The input IMU data is normalised with some of the combined navigation states and sequentially passed through the embedding layer, the LSTM hidden layer and the fully connected layer to finally output the predicted forward velocity of the vehicle.

3.3 Input feature preprocessing

Due to the high noise level of the MEMS IMU itself and the high output frequency of IMU data, using IMU data at every instant as input to the velocity estimation network will obviously introduce excessive measurement noise, which will affect the estimation accuracy of the network. Therefore, an integral method can be used to eliminate some of the noise of the MEMS IMU, and the velocity and angle increments within 100ms can be used as network input parameters:

$$
\begin{aligned}
\Delta\boldsymbol{\phi} &= \int \boldsymbol{f}^b \\
\Delta\boldsymbol{v} &= \int \boldsymbol{\omega}^b
\end{aligned}
\tag{36}
$$

Due to the fact that the acceleration and acceleration information measured by the sensors is based on the sensor coordinate system s, it is necessary to use the online estimated installation angle for conversion:

$$
\begin{aligned}
\tilde{\boldsymbol{f}}^b &= \tilde{\boldsymbol{C}}_s^b \tilde{\boldsymbol{f}}^s \\
\tilde{\boldsymbol{\omega}}^b &= \tilde{\boldsymbol{C}}_s^b \tilde{\boldsymbol{\omega}}^s \tilde{\boldsymbol{C}}_b^s
\end{aligned}
\tag{37}
$$

Similarly, in navigation calculations, the north-east sky speed is often the speed at the inertial navigation point, so it is necessary to convert the estimated installation angle and NHC lever arm to the upper right-hand speed at the centre of the vehicle's motion:

$$
\tilde{\boldsymbol{v}}^b = \tilde{\boldsymbol{C}}_s^b \tilde{\boldsymbol{C}}_n^s \tilde{\boldsymbol{v}}^n + \tilde{\boldsymbol{\omega}}^b \times \tilde{\boldsymbol{L}}^b
\tag{38}
$$

Finally, the attitude and heading of the vehicle have a significant effect on the forward speed, so attitude and heading should also be used as input features for the network. After the input features are selected, directly inputting them into the network training will affect the training effect and even fail to converge due to the difference in magnitude between different input features. Therefore, it is

necessary to normalise the input features. In this paper, the min-max normalisation method is used to normalise the input features, and the processing formula is as follows:

$$x^*[i] = \frac{x[i] - \min(x)}{\max(x) - \min(x)} \tag{39}$$

4   Experiment and Analysis

4.1   Experimental platform construction

The experimental test platform of this article is based on Geely Haoyue Automobile, and the equipment it is equipped with mainly includes a GNSS antenna located on the roof of the car, a self-developed vehicle terminal integrated with MEMS IMU and GNSS board in the trunk of the car, and NovAtel's SPAN system used as the reference truth. The layout of the experimental platform is shown in the figure below:

Fig.6   Experimental platform

The SPAN system includes an inertial navigation system consisting of a navigation-grade fibre-optic gyroscope and accelerometer, and a GNSS receiver capable of providing centimetre-level positioning results. The vehicle terminal integrates the ADIS16505 MEMS inertial sensor and Xin Xing Tong's GNSS board. Due to the nanosecond timing accuracy of GNSS Pulse Per Second (PPS), time synchronisation between different devices on this experimental platform is based on PPS. The following table shows the basic parameters of the inertial sensors used in this article.

Tab.1   IMU parameters

	gyroscope	accelerometer
range	±500 °/s	±8 g
Zero bias	2.7(1.4Y) °/h	2.7(4.4Z) ug

stability		
Zero bias repeatability	0.14 °/s	2 mg
Random walk	0.15(0.2Z) °/√h	0.009(0.012Z) m/s/√h

4.2 Experiment 1: Verification of Kinematic Constraints Experiment

The test site is located in Dadukou District, Chongqing. The overall experimental route is shown in the figure below. First, the vehicle is driven along the open area of Linjiang Road, and after sufficient manoeuvring to correct the filter parameters, the vehicle enters the garage. The total driving time in the garage is about 270 seconds, with a total distance of 600 metres. There are four parking spaces and seven turns in the garage. The experimental trajectory is shown in the following figure:

Fig.7 The trajectory of the experiment

There are three ways of processing data in the experiment:

(1) No estimation of sensor installation parameters, using NHC with uncompensated installation parameters in the garage;

(2) Use GNSS to estimate sensor installation parameters online in open environments, and use INS/NHC navigation mode to compensate installation parameters after entering the garage;

(3) Add static constraints based on condition (2).

The trajectories inside the garage under different processing methods are shown in Figure 8. Among them, the green trajectory is the SPAN trajectory used as the reference truth value, the purple trajectory is the INS/NHC result without estimated installation parameters, the blue trajectory is the INS/NHC result with estimated installation parameters, and the red trajectory is the INS/NHC result with estimated installation parameters combined with static constraints. From the figure it can be seen that the purple trajectory is prone to an obvious turn back at the turning point, while the blue and red trajectories are obviously smoother and closer to the SPAN trajectory. This indicates that after estimating the sensor installation parameters, the influence of the Coriolis effect on the vehicle at the turning point is significantly reduced and the overall positioning performance is improved.

Fig.8   Garage trajectory under different processing method

The root mean square error statistics table and velocity error chart of the combined navigation in the garage are as follows. According to the velocity error graph and the root mean square error statistics table, compared with the INS/NHC navigation mode without estimated installation parameters, the INS/NHC navigation mode with sensor installation angle and NHC lever arm compensation has significantly improved the accuracy of the three directional velocities. The eastward velocity error increased by approximately 71.3%, the northward velocity error increased by approximately 57.9%, and the celestial velocity error increased by approximately 52.1%. The significant improvement in speed accuracy is due to the presence of a large number of horizontal turns in the underground car park during the test. If the sensor installation angle and the NHC lever arm are not compensated, the horizontal turns will weaken the zero lateral velocity constraint, resulting in a decrease in the NHC observation performance and an increase in the velocity error. At the same time, due to the presence of parking in all directions in this experiment, the addition of static constraints suppressed the speed error in the stationary state. Therefore, the addition of static constraints further improved the speed accuracy. Compared to the INS/NHC navigation mode without estimated installation parameters, the eastward velocity error increased by about 75%, the northward velocity accuracy increased by about 63.2%, and the celestial velocity accuracy increased by about 52.1%.

Tab.2 Velocity root mean square error table in experiment 1

algorithm	Speed RMSE (m/s)			
	Eastward direction	Northward direction	Up direction	plane
Unestimated installation parameters	0.268	0.185	0.048	0.326
Estimate installation parameters	0.077	0.078	0.023	0.110
Estimate installation parameters+static constraints	0.067	0.068	0.023	0.096
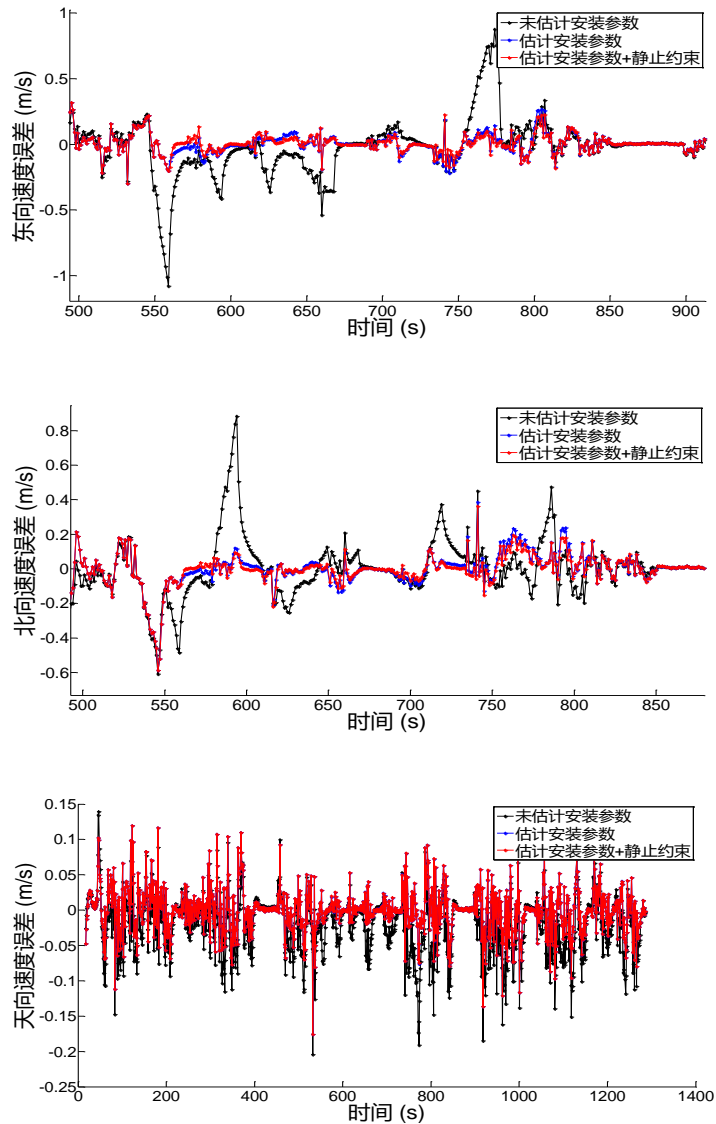
Fig.9   Velocity error diagram in experiment 1

The Root Mean Square Error Statistics Table and the Positioning Error Chart of the integrated navigation in the garage are as follows. From the error chart and the error statistics table, it can be seen that after compensating for the installation parameters of the sensor, the INS/NHC algorithm showed a slight decrease in latitude accuracy in the case of total satellite loss, but a significant improvement in longitude and elevation positioning accuracy. The longitude positioning accuracy increased by approximately 57.8% and the elevation positioning accuracy increased by approximately 95.2%. The improvement in longitudinal positioning accuracy is mainly due to the presence of more horizontal turns in the underground garage. After compensating for the installation angle and the NHC lever arm, the positioning error caused by the higher lateral speed is improved. At the same time, due to the need to drive through a long downhill section when entering the garage, the vehicle can be considered to have undergone a significant pitch turn with an arc. The speed error graph also shows that the vertical speed error increases, resulting in a decrease in vertical position accuracy. At the same time, there were four parking stops in this experiment, so adding static constraints helps to suppress position error under static conditions, further improving positioning

accuracy. Compared to the INS/NHC navigation mode without estimated installation parameters, the planar positioning accuracy is improved by about 47.3%, including an improvement of about 7% in the latitude direction accuracy, an improvement of about 76% in the longitude direction accuracy and an improvement of about 96.3% in the elevation direction accuracy.

Tab.3   Position root mean square error table in experiment 1

algorithm	Position RMSE (m)			
	latitude	longitude	altitude	plane
Unestimated installation parameters	2.171	3.546	3.031	4.158
Estimate installation parameters	2.538	1.495	0.147	2.945
Estimate installation parameters+static constraints	2.018	0.851	0.112	2.190
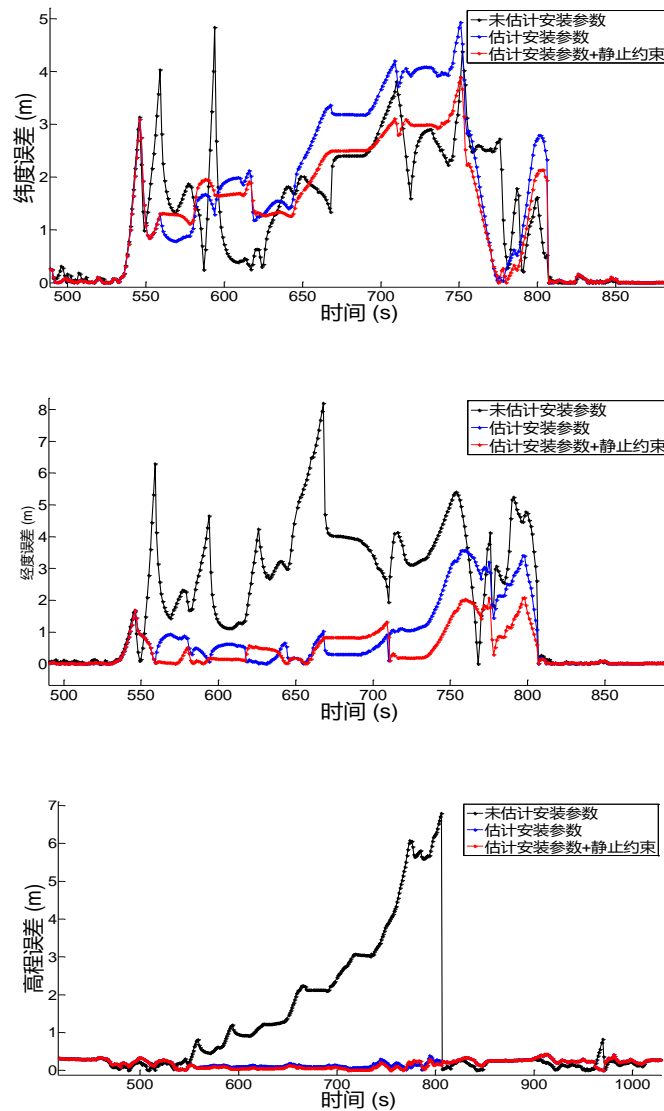
Fig.10   Position error diagram in experiment 1

The online estimated NHC lever arm and installation angle are shown in the following figure. As the terminal is installed on the platform built into the boot of the vehicle, it is located approximately above the centre of the rear wheel joint, and the sensor is installed as flush as possible with the vehicle body. Therefore, in the upper right front coordinate system with the sensor as the origin, the online estimation of the NHC right arm by the filter is about -0.05m, the forward arm is about 0.38m, the sky arm is about -0.77m, the pitch installation angle is about -0.67 degrees, the roll installation angle is about 0.06 degrees, and the heading installation angle is about 0.48 degrees, which is basically consistent with the actual installation of the sensor.
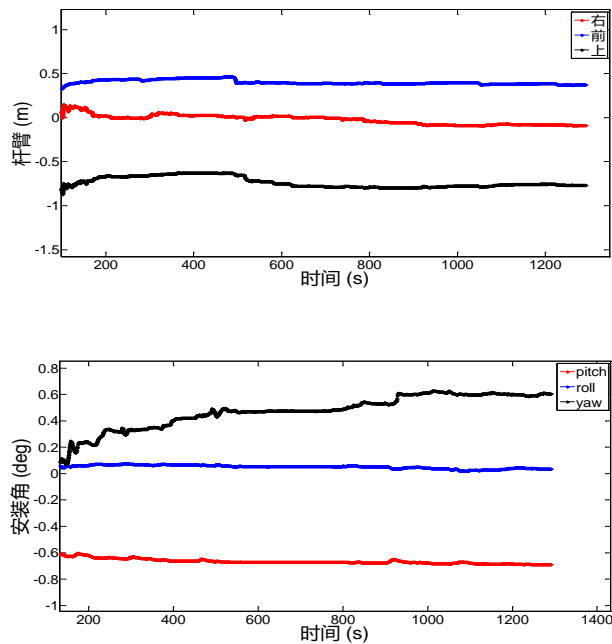
Fig.11　Installation parameter estimation

Due to the fact that the attitude of the reference truth values provided by SPAN is the vehicle body attitude, it is necessary to use the online estimated installation angle to convert the attitude angle calculated by the unestimated installation parameters to the vehicle system before calculating the attitude error. Therefore, the attitude error statistics table and attitude error graph of the integrated navigation are shown below. According to the attitude error statistics table and error curve graph, compared to the INS/NHC navigation mode without estimated installation parameters, the attitude angle of the integrated navigation system after estimating the installation parameters is generally more stable, without significant attitude angle changes due to horizontal or pitch steering. At the same time, the elevation error is significantly improved after installation parameter estimation, with an increase in elevation accuracy of approximately 66%. As the elevation angle is an important parameter for projecting gravity forwards, the improvement in elevation angle accuracy has a significant effect on the improvement in positioning accuracy. The trend for roll and heading errors is the same for the different processing methods. When static constraints are added, it can be seen that the elevation and heading errors are further reduced.

Tab.4   Attitude root mean square error table in experiment 1

algorithm	Attitude RMSE (deg)		
	pitch	Roll	yaw
Unestimated installation parameters	0.144	0.644	0.630
Estimate installation parameters	0.049	0.640	0.891
Estimate installation parameters+static constraints	0.047	0.649	0.634
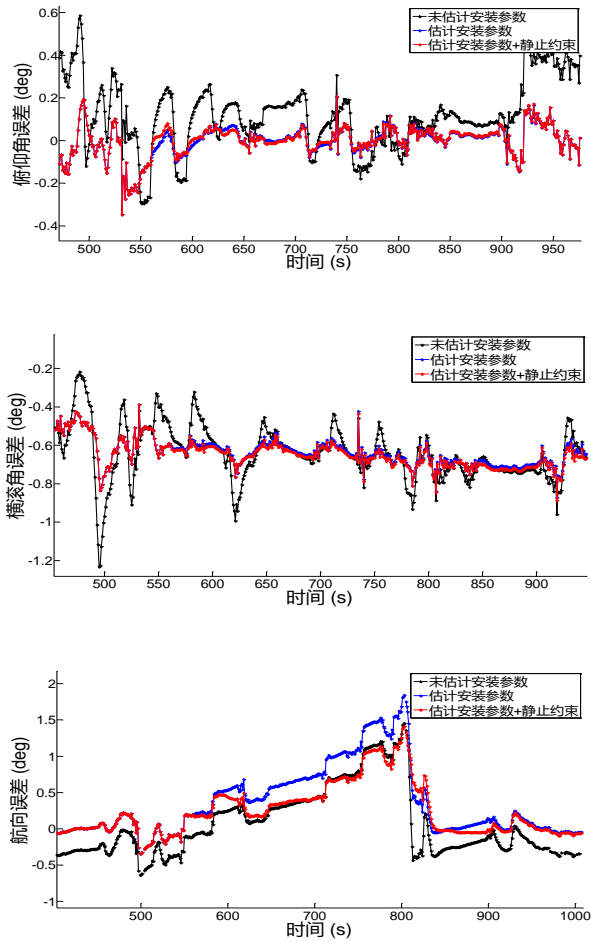
Fig.12   Attitude error diagram in experiment 1

4.3 Experiment2: LSTM/NHC complete speed constraint test

The data used to train the neural network in this article was collected through a self-developed terminal in Chongqing, including various scenarios such as open areas, urban canyons, tree-lined areas, underground spaces, tunnels, etc., with a total duration of about 200 hours. The input feature parameters for the network training are the velocity pre-intergration, angle pre-intergration, right front upper velocity and attitude heading angle of the vehicle body. Using the root mean square error

of the forward velocity of the validation set as a reference, the performance of the model under different numbers of hidden units is examined. The RMS error of the forward velocity under different parameters is summarised in the table below and the comparison between the predicted and actual velocity values is shown in the figure below.

Tab.5　Forward velocity estimation error of different number of hidden units

Number of hidden layer units	Forward velocity error RMSE (m/s)
32	0.0610
64	0.0564
128	0.0510
256	0.0521
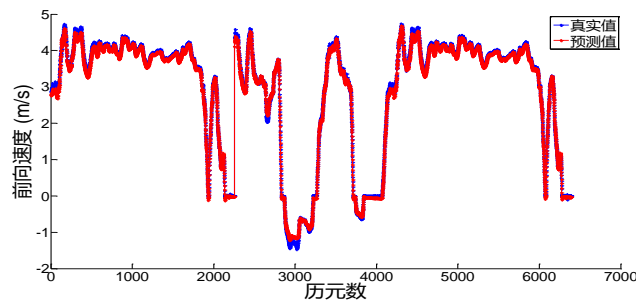
Fig.13　Forward velocity comparison diagram

From the root mean square error of predicted speed, it can be seen that when the number of hidden layer units is small, as the number of hidden layer units increases, the complexity of the model increases and the expressiveness increases. Therefore, the prediction speed error gradually decreases. If the number of hidden layer units is too large, overly complex models may lead to overfitting of the prediction speed. Therefore, when using LSTM networks, an appropriate number of hidden layer units should be selected to avoid overfitting. In this article, 128 is the ideal number of hidden layer units.

To verify the actual positioning effect of the algorithm, a section of real road test data was collected. The test site was located in Dadukou District, Chongqing. After full manoeuvring and correction of the filter parameters in an open-air scene, the vehicle was driven into an underground car park without parking for approximately 220 seconds. The overall test route is as follows:
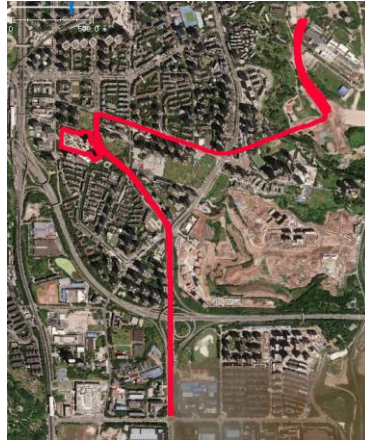
Fig.14 Motion trajectory

The trajectory in the garage is as follows. The green trajectory in the figure is the SPAN reference trajectory, the blue trajectory is the INS+NHC mode and the red trajectory is the INS+NHC+LSTM mode. From the trajectory diagram, it can be seen that the red trajectory is significantly closer to the green reference trajectory than the blue trajectory, indicating that the positioning performance of the integrated navigation system has been significantly improved after adding the forward speed constraint of LSTM.

Fig.15 Garage trajectory

The root mean square error statistics and forward velocity error plots for the different modes are shown below. The graph and table show that the forward velocity error predicted by the LSTM network is reduced by 27% compared to the INS+NHC mode. However, after using the forward velocity of the LSTM to assist the integrated navigation, the forward velocity error is reduced by 21.7% in the INS+NHC+LSTM mode. From a trajectory point of view, the car's trajectory in the garage mainly follows an east-west direction, so reducing the forward velocity error should significantly improve the longitudinal positioning accuracy of the integrated navigation system.

Tab.6 Forward velocity root mean square error

algorithm	LSTM	INS+NHC	INS+NHC+LSTM
Forward velocity error RMSE (m/s)	0.135	0.185	0.145
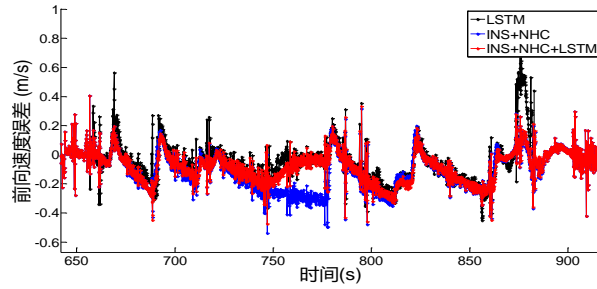
Fig.16 Forward velocity error diagram

The positioning error statistics table and error graph of the integrated navigation system are as follows. From the positioning error graph and statistics table, it can be seen that the addition of the LSTM forward velocity constraint has significantly improved the planar positioning accuracy, with an increase in positioning accuracy of approximately 28.1%. Among them, the latitudinal root mean square error decreased by 16.7% and the longitudinal root mean square error decreased by 37.6%. The elevation positioning accuracy of the two modes is basically equivalent. Obviously, increasing the LSTM forward velocity assistance results in a better accuracy improvement in the longitude direction than in the latitude direction, which is consistent with the actual situation described earlier, where vehicles mainly move in an east-west direction. Therefore, increasing the LSTM's forward velocity assistance is beneficial for improving positioning accuracy under long-term satellite loss of lock.

Tab.7   Position root mean square error in experiment 2

algorithm	Position RMSE (m)			
	latitude	longitude	altitude	plane
INS+NHC	2.335	2.815	0.428	3.657
INS+NHC+LSTM	1.951	1.761	0.400	2.628

5 Conclusion

This article proposes a low-cost GNSS/MEMS IMU integrated vehicle navigation algorithm based on integrity kinematic constraints to solve the problem of rapid decline in positioning accuracy in GNSS lockout environments. The main innovative points are as follows:

(1) The integration of vehicle statics and NHC to achieve high-precision maintenance of the vehicle in full motion without satellite conditions, and the use of an extended filter to achieve online estimation of IMU installation angle and NHC lever arm, thus fully exploiting the performance of NHC.

(2) A forward velocity estimation network based on LSTM is proposed to estimate the forward velocity of vehicles in real time, thereby completing NHC into complete 3D velocity constraints and improving the positioning performance of integrated navigation systems under long-term GNSS lock conditions. Compared to traditional neural networks that directly use sensor data and calculated navigation states as inputs, this paper uses the online estimated IMU installation angle and NHC

lever arm to convert sensor data and calculated navigation states to the carrier motion centre, and combines neural networks with kinematic models to improve the accuracy of vehicle forward velocity estimation.

In order to verify the accuracy of the algorithm proposed in the article, actual road tests were conducted using a self-developed terminal platform. The test results of Experiment 1 verified the effectiveness of estimating the IMU installation angle and NHC lever arm under good GNSS conditions. Compared to the traditional INS/NHC algorithm, compensating the installation parameters and adding static constraints improved the plane positioning accuracy by about 47.3% and the elevation direction accuracy by about 96.3%. The test results of Experiment 2 showed that the addition of the LSTM-assisted rear vehicle reduced the root mean square error of the forward velocity by about 21.7% and improved the plane positioning accuracy by about 28.1%.

References:

[1] Bai Shiyu, Lai Jizhou, Lyu Pin. Improved Preintegration Method for GNSS/IMU/In-Vehicle Sensors Navigation Using Graph Optimization. IEEE Transactions on Vehicular Technology. 2021. 70(11): 11446-11457.

[2] Chiang Kai-Wei, Chang Hsiu-Wen, Li Yu-Hua. Assessment for INS / GNSS / Odometer / Barometer Integration in Loosely-Coupled and Tightly-Coupled Scheme in a GNSS-Degraded Environment. IEEE Sensors Journal. 2020. 20(6): 3057-3069.

[3] Li Zengke, Liu Zan, Zhao Long. Improved robust Kalman filter for state model errors in GNSS-PPP/MEMS-IMU double state integrated navigation. Advances in Space Research. 2021. 67(10): 3156-3168.

[4] Shin Eun Hwan. Estimation techniques for low-cost inertial navigation.[D]. University of Calgary (Canada). 2005.

[5] Zhang Wenchao, Li Xianghong, Wei Dongyan. A foot-mounted PDR system based on IMU / EKF + HMM + ZUPT + ZARU + HDR + compass algorithm[C]//. 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2017: 1-5.

[6] Zhang Quan, Hu Yuanqian, Niu Xiaoji. Required Lever Arm Accuracy of Non-Holonomic Constraint for Land Vehicle Navigation. IEEE Transactions on Vehicular Technology. 2020. 69(8): 8305-8316.

[7] Chen Yingqiu, Kuang Jian, Niu Xiaoji. Vehicle Integrated Navigation Based on Inertial Measurement Units Mounted on Wheels. Journal of Chinese Inertial Technology, 2018, 26(06): 799-804.
Chen Yingqiu, Kuang Jian, Niu Xiaoji. Vehicle integrated navigation based on wheel mounted MEMS-IMU[J]. Journal of Chinese Inertial Technology, 2018, 26(06): 799-804.

[8] El-Sheimy N, Chiang Kai-Wei, Noureldin A. The Utilization of Artificial Neural Networks for Multisensor System Integration in Navigation and Positioning Instruments. IEEE transactions on instrumentation and measurement. 2006. 55(5): 1606-1615.

[9] Sharaf R., Noureldin A. Sensor integration for satellite-based vehicular navigation using neural networks. IEEE Trans Neural Netw. 2007. 18(2): 589-594.

[10] Dai Hai-fa, Bian Hong-wei, Wang Rong-ying. An INS/GNSS integrated navigation in GNSS denied environment using recurrent neural network. Defence Technology. 2020. 16(2): 334-340.

[11] Zhao Shuai, Zhou Yilan, Huang Tengchao. A Novel Method for AI-Assisted INS/GNSS Navigation System Based on CNN-GRU and CKF during GNSS Outage. Remote Sensing. 2022. 14(18): 4494.

[12] Chen Huaiyu, Yin Dayi, Zhang Quan. Analysis and Verification of LSTM Networks for Improving MEMS Inertial Navigation Positioning Accuracy. Journal of Chinese Inertial Technology, 2018, 26(05): 610-615.
Chen Huaiyu, Yin Dayi, Zhang Quan. Analysis and verification on improving MEMS navigation accuracy based on LSTM network[J]. Journal of Chinese Inertial Technology. 2018, 26(05): 610-615.

[13] Yan Gongmin. Research on Vehicle Autonomous Positioning and Orientation System [D]. Xi'an: Northwestern Polytechnical University, 2006.
Yan Gongmin. Research on vehicle autonomous positioning and orientation system[D]. Xian: Northwestern Polytechnical University, 2006.

[14] Melendez-Pastor Carlos, Ruiz-Gonzalez Ruben, Gomez-Gil Jaime. A data fusion system of GNSS data and on-vehicle sensors data for improving car positioning precision in urban environments. Expert Systems with Applications. 2017. 80: 28-38.

A SINS/DVL/PS Fault-Tolerant Tightly Integrated Navigation Method Under Beam Failure Conditions

Shao Qiantong[1], Liu Xixiang[1]

(1. School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China;)

Abstract: The strapdown inertial navigation system (SINS)/Doppler velocity log (DVL) integrated navigation method is the major navigation method for autonomous underwater robot (AUV). However, in complex underwater environments, the system is unable to maintain accurate positioning when the DVL beam fails. The SINS/DVL tightly integrated method based on the DVL bottom tracking mode provides a solution to this problem for continuous navigation, but such methods fail when the DVL is switched to the water tracking mode. To address these problems, this article proposes a novel SINS/DVL/Pressure Sensor (PS) fault-tolerant tightly integrated navigation method under beam failure conditions, which can ensure that the system can still work when the DVL is switched to water tracking mode. The simulation results are designed to verify the performance of the proposed method.

Key words: AUV; SINS/DVL tightly integrated system; virtual beam construction;

1 INTRODUCTION

As an important tool to explore and exploit the ocean, accurate positioning and navigation of Autonomous Underwater Vehicles (AUVs) are crucial to mission realization [1],[2]. AUVs are often equipped with Strapdown inertial navigation system (SINS) during missions because SINS can provide the navigation system with attitude, velocity, and position information needed for positioning. However, SINS usually requires other external sensors to correct the localization results due to the measurement errors present in the inertial sensors and integration flaws that can cause the navigation results to diverge over time [3]-[5]. The SINS/DVL integrated navigation system is inherently highly autonomous and does not depend on prior deployment, which makes it the most popular underwater navigation method [6]-[7].

Conventional SINS/DVL integrated navigation methods generally rely on the dead reckoning techniques of SINS and DVL or may adopt a loosely coupled structure [8]. Unfortunately, the complexities of underwater environments often result in unpredictable scenarios during long endurance voyages, resulting in DVL beam failures that adversely impact the precision of navigation and positioning. Several researchers have studied the integrating mechanisms of SINS/DVL, exploring the SINS/DVL integrated navigation methods under DVL failure. Tal and Klein first proposed an extended loose coupling (ELC) approach. The four sub-methods proposed in [9] comprehensively elaborate how partial beam measurements, SINS information, or motion assumptions can be utilized to reconstruct DVL velocity output during DVL partial failures. Liu et al. [10] discussed the different beam distribution scenarios of DVL under different configurations from the perspectives of a tightly integrated and designed virtual beam (VB) construction method in combination with the Pressure Sensor (PS).

However, all of the currently existing methods are dependent on the bottom tracking velocity of the DVL, but when the bottom tracking velocity is unavailable the DVL switches to water tracking mode, at which point all of the above methods will fail. If the system is degraded to SINS at this stage, the effect caused by ocean currents will inevitably be ignored. Therefore, how to solve the problem of beam failure of DVL working in water tracking mode is the key to improving the accuracy of SINS/DVL integrated navigation. Therefore, this paper combines the ocean current estimation to design a SINS/DVL/PS fault-tolerant tightly integrated navigation method under beam failure conditions.

2 Proposed SINS/DVL-WT/PS Tightly Integrated Navigation Method

In this article, the East-North-Up (ENU) geographic coordinate frame is selected as the navigation coordinate frame, which is denoted by the n-frame. The calculated reference navigation coordinate frame is denoted by the n'-frame. The Right-Forward-Up (RFU) coordinate frame is selected as the AUV's body coordinate frame, which is denoted by the b-frame. Similarly, the RFU coordinate frame of the DVL instrument is denoted by the d-frame, while the DVL's original four-beam frame is denoted by the beam-frame.
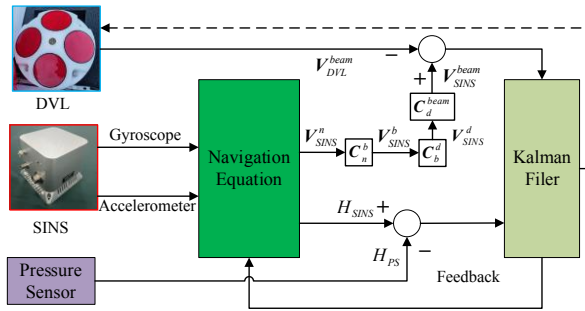
Fig. 1 SINS/DVL-WT/PS tightly integrated navigation system

Fig. 1 shows the structure of the SINS/DVL-WT/PS tightly integrated navigation system, which mainly consists of SINS, DVL, and PS. Note that the installation angle error and lever arm between SINS and DVL were calibrated and compensated before the experiment. The residual installation angle error and lever arm error are negligible. V_{cE} , V_{cN} , and V_{cU} are introduced into the state vector of the system model for real-time estimation. The 20-dimensional state vector X can be expressed as follows:

$$X = \begin{bmatrix} \phi^n & \delta V^n & \delta P^n & \nabla^b & \varepsilon^b & K_D & \delta b_{ps} & V_c^n \end{bmatrix} \tag{1}$$

where ϕ^n represent the misalignment angle; δV^n represent the velocity errors; δP^n represent the position errors; ∇^b represent the accelerometer biases; ε^b represent the gyroscope biases; K_D represent the scale factor of the DVL; δb_{ps} represents PS bias；V_c^n represent the ocean current velocities in the three-axis directions.

Before establishing the measurement vector, it is essential to clarify the relationship of velocities under different coordinate frames, the relationship between these velocity vectors are as follows:

$$V_{DVL}^{beam} = V_{SINS}^{beam} = C_d^{beam} C_b^d C_n^b V_{SINS}^n \tag{2}$$

Since the installation angle error and the lever arm are calibrated in advance, it can be assumed that the b-frame and d-frame are nearly equivalent, that is, V_{SINS}^b is equivalent to V_{SINS}^d, and C_b^{beam} is shown below:

$$C_b^{beam} = C_d^{beam} = \begin{bmatrix} -\sin\alpha\sin\beta & \sin\alpha\cos\beta & -\cos\alpha \\ -\sin\alpha\cos\beta & -\sin\alpha\sin\beta & -\cos\alpha \\ \sin\alpha\sin\beta & -\sin\alpha\cos\beta & -\cos\alpha \\ \sin\alpha\cos\beta & \sin\alpha\sin\beta & -\cos\alpha \end{bmatrix} \tag{3}$$

where α is the angle between the vehicle and the beams, typically 70°. β is the Janus configuration installation angle; 0° for the '+' configuration and 45° for the '×' configuration.

The error models for DVL and PS are expressed as follows:

$$\tilde{V}_{DVL\_C}^{beam} = (1 + K_D) V_{DVL}^{beam} - C_b^{beam} C_n^b V_c^n + \delta b_D + \omega_D \tag{4}$$

$$\tilde{H}_{ps} = H_{ps} + \delta b_{ps} + \omega_{ps} \tag{5}$$

where $\tilde{V}_{DVL\_C}^{beam}$ represents the four-beam true measurement of the DVL, δb_D represents the DVL bias, $\tilde{H}_{ps}$ represents the true measurement of the PS, ω_D and ω_{ps} represent the white noise of the measurement.

$\hat{V}_{SINS}^{beam}$ is calculated by the formula:

$$\hat{V}_{SINS}^{beam} = C_d^{beam} C_b^d \hat{C}_n^b \hat{V}_{SINS}^n = V_{SINS}^{beam} + C_b^{beam} C_n^b \delta V_{SINS}^n - C_b^{beam} C_n^b V_{SINS}^n \times \phi \tag{6}$$

Take the difference between $\hat{V}_{SINS}^{beam}$ and $\tilde{V}_{DVL\_C}^{beam}$ as the observation vector Z_1:

$$Z_1 = \hat{V}_{SINS}^{beam} - \tilde{V}_{DVL\_C}^{beam} = C_b^{beam} C_n^b \delta V_{SINS}^n - C_b^{beam} C_n^b V_{SINS}^n \times \phi - V_{DVL}^{beam} \delta K_D - \delta b_D - \omega_D + C_b^{beam} C_n^b V_c^n \tag{7}$$

Take the difference between SINS altitude and $\tilde{H}_{ps}$ as the observation vector Z_2:

$$Z_2 = \hat{H}_{SINS} - \tilde{H}_{ps} = \delta h - \delta b_{ps} \tag{7}$$

Therefore, the observation vector Z of SINS/DVL-WT/PS tightly integrated method can be expressed as follows:

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \tag{8}$$

3 Fault-Tolerant Processing Scheme

In practical scenarios, the utilization of the "×" configuration typically exhibits superior fault tolerance and efficiency compared to the "+" configuration; therefore, this article focuses on the "×" configuration.

Existing methods for beam failure situations are mostly based on motion constraints, which often introduce new errors when the vehicle dynamics assumptions are not met. Therefore, it is essential to design a constraint-free fault-tolerant processing scheme.

$$\begin{cases} V_{xv}^b = V_1^{beam} \cos\alpha \cos\beta \\ V_{yv}^b = V_1^{beam} \cos\alpha \sin\beta \\ V_{zv}^b = V_1^{beam} \sin\alpha \end{cases} \tag{9}$$

When a beam is available, it can be projected onto the b-frame based on the principle of projection (in the case of beam 1), the projection regulations can be expressed as (9):

Other fault beams can be reconstructed through the above three vectors, and the virtual beams are constructed as follows:

$$\begin{cases} V_2^{beam} = -\sqrt{V_{xv}^2 + V_{yv}^2 + V_{zv}^2} \\ V_3^{beam} = \sqrt{V_{xv}^2 + V_{yv}^2 + V_{zv}^2} \\ V_4^{beam} = -\sqrt{V_{xv}^2 + V_{yv}^2 + V_{zv}^2} \end{cases} \tag{10}$$

This method is not affected by the number of fault beams and can be employed as long as at least one valid beam is present. To sum up, the virtual beam construction method based on the principle of projection can construct the missing beams in most situations.

4 Simulation

The simulation time lasts for the 1200s and involves most of the motion states, including uniform acceleration, deceleration, and turning motion. To better evaluate the proposed algorithm, the three-axis swaying of the vehicle in the ocean environment is simulated and the vehicle attitude is shown below:

$$\theta = 3\sin\left(\frac{2\pi}{15}t + \theta_0\right), \gamma = 4\sin\left(\frac{2\pi}{10}t + \gamma_0\right), \psi = 3\sin\left(\frac{2\pi}{8}t + \psi_0\right) \tag{11}$$

where the initial phases θ_0, γ_0, ψ_0 are all 0. For the IMU, the constant biases and random walk noises of the gyroscopes are set as $0.02°/h$ and $0.0005°/\sqrt{h}$,while those of the

accelerometers are set as $50ug$ and $25ug/\sqrt{Hz}$, and the update rate of the IMU is set as 100 Hz. The scale factor error of DVL is 0.2%, and the update rate is set as 5Hz.

The corresponding motion trajectories and 3D attitudes and velocities in the simulation are shown in Fig. 2.
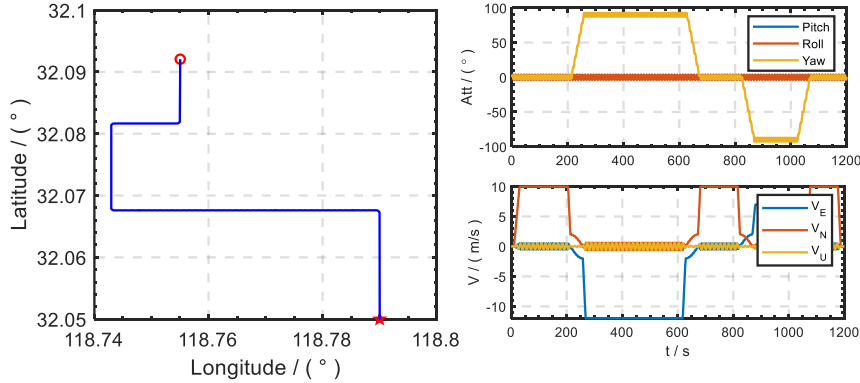
Fig. 2 Moving status and trajectory of the vehicle

The initial latitude and longitude of the vehicle are $32.05°$ and $118.79°$ .The initial velocities are $0m/s$ .The initial attitudes are $\theta = 0°$, $\gamma = 0°$, $\psi = 0°$.

Loosely integrated and tightly integrated methods were compared under DVL beam failure conditions, with the loosely integrated navigation method with ocean current estimation (LI-OC) marked in blue line and the VB-based fault-tolerant tightly integrated navigation method with ocean current estimation (TI-OC-VB) marked in red line. The ocean current is set to [0.5;0.5;0]. Different DVL beam failure periods and failure scenarios are set up and listed in Table 1.

Tab.1   Distribution of DVL beam failures

Outage Num.	Time Interval / s	Outage Beam
#1	200s~225s	Beam 1, Beam 2
#2	285s~310s	Beam 3, Beam 4
#3	500s~550s	Beam 2, Beam 3, Beam 4
#4	650s~700s	Beam 1, Beam 2, Beam 4
#5	930s~950s	Beam 2, Beam 3

The results of the ocean current estimation are given in Fig. 3, where the TI-OC-VB results are closer to the true value. The comparison of the velocity errors is shown in Fig. 4, left. The velocity error of the loosely integrated navigation method increases rapidly during the failure of the DVL beam measurement because the DVL device cannot output the vehicle's velocity information when the DVL effective beam is less than 3. The navigation system switches to the SINS, which ultimately leads to a rapid increase in the velocity error. The tightly integrated method ensures a stable DVL output, so the velocity error of the system remains stable. Fig. 4, right. shows the positional error curve, the horizontal positional error of the two methods keeps growing but the method proposed in this article has a smaller positional error.
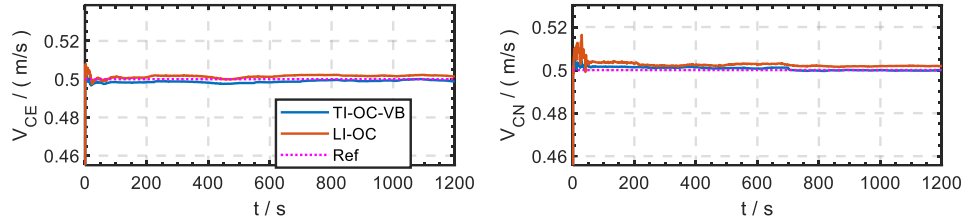
Fig. 3   Estimated curves of ocean currents velocity
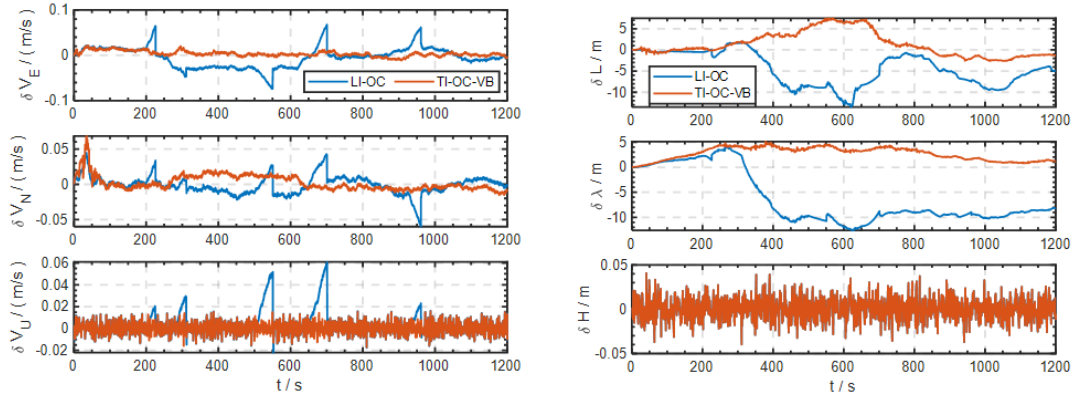
Fig. 4   Velocity error and Position error comparison curves for the two methods

The root mean square error (RMSE) and standard deviation (STD) of the position errors of the two methods are shown in Table 2. The STD and RMSE on depth information are consistent, which indicates that the depth information obtained by PS is accurate and stable. The RMSE value of the latitude error of the proposed TI-OC-VB is 2.0910m less than 2.9902m of the LI-OC method, and the RMSE value of the longitude error is 2.4199m less than 7.1647m of the LI-OC method which is reduced by 30.1% and 66.2%, respectively. The STD of the TI-OC-VB is reduced by 36.37% and 76.62% compared with that of LI-OC, which indicates that TI-OC-VB not only has better accuracy but also better stability than LI-OC.

Tab.2   STD and RMSE for the different methods

	LI-OC		TI-OC-VB	
	STD	RMSE	STD	RMSE
$\delta L / m$	2.9290	2.9902	1.8636	2.0910
$\delta \lambda / m$	4.3833	7.1647	1.0247	2.4199
$\delta h / m$	0.0101	0.0101	0.0101	0.0101

5   CONCLUSION

Aiming at the problem of beam failure of DVL operating in water tracking mode. a SINS/DVL/PS fault-tolerant tightly integrated navigation method under beam failure conditions is proposed in this article. First, a new system model is established based on the error model. Second, to address the limitations of the current reconstruction strategies, a fault-tolerant processing technique based on VB is designed, which can reconstruct the DVL failure beams, and further improve the fault-tolerant capability of the navigation system. Finally, simulation results show that

the proposed algorithm can cope with complex beam failure situations and improve the positioning accuracy of the system in complex underwater environments.

References:

[1] L. Paull, S. Saeedi, M. Seto, and H. Li, "AUV Navigation and Localization: A Review," IEEE J. Oceanic Eng., vol. 39, no. 1, pp. 131–149, Jan. 2014.

[2] L. Stutters, Honghai Liu, C. Tiltman, and D. J. Brown, "Navigation Technologies for Autonomous Underwater Vehicles," IEEE Trans. Syst., Man, Cybern. C, vol. 38, no. 4, pp. 581–589, Jul. 2008.

[3] B. Xu and Y. Guo, "A Novel DVL Calibration Method Based on Robust Invariant Extended Kalman Filter," IEEE Trans. Veh. Technol., vol. 71, no. 9, pp. 9422–9434, Sep. 2022.

[4] L. Chang, J. Di, and F. Qin, "Inertial-Based Integration With Transformed INS Mechanization in Earth Frame," IEEE/ASME Trans. Mechatron., vol. 27, no. 3, pp. 1738–1749, Jun. 2022.

[5] X. Xu, X. Ning, Y. Yao, and K. Li, "In-Motion Coarse Alignment Method for SINS/GPS Integration in Polar Region," IEEE Trans. Veh. Technol., vol. 71, no. 6, pp. 6110–6118, Jun. 2022.

[6] Y. Huang, X. Liu, Q. Shao, and Z. Wang, "Virtual Metrology Filter-Based Algorithms for Estimating Constant Ocean Current Velocity," Remote Sensing, vol. 15, no. 16, p. 4097, Aug. 2023.

[7] L. Chang, H. Tang, G. Hu, and J. Xu, "SINS/DVL Linear Initial Alignment based on Lie Group $SE_3 (3)$," IEEE Trans. Aerosp. Electron. Syst., pp. 1–17, 2023.

[8] G. Sheng, X. Liu, Y. Sheng, X. Cheng, and H. Luo, "Cooperative Navigation Algorithm of Extended Kalman Filter Based on Combined Observation for AUVs," Remote Sensing, vol. 15, no. 2, p. 533, Jan. 2023.

[9] A. Tal, I. Klein, and R. Katz, "Inertial Navigation System/Doppler Velocity Log (INS/DVL) Fusion with Partial DVL Measurements," Sensors, vol. 17, no. 2, p. 415, Feb. 2017.

[10] P. Liu, B. Wang, Z. Deng, and M. Fu, "INS/DVL/PS Tightly Coupled Underwater Navigation Method With Limited DVL Measurements," IEEE Sensors J., vol. 18, no. 7, pp. 2994–3002, Apr. 2018.

Deviation Analysis and Programming Design for Chase and Rendezvous Task of Spacecraft

Su Yan[1,2], HU Haixia[1,2], CAI Biao[1,2], Chen Changqing[1,2], Wang Min

(1. Beijing Control Engineering Research Institute, Beijing 100190, China; 2. State Key Laboratory of Intelligent Space Control Technology, Beijing 100190, China)

Abstract: Space rendezvous technology is an important research and development direction in the space field, which has high economic and strategic value. With the development of space mission and the progress of space mission related technology, the need of integrated strategy of fully autonomous navigation and guidance on board is becoming more and more prominent. The formulation of onboard autonomous guidance strategy is closely related to the navigation deviation model. In this paper, the coupling of trajectory and radial deviation in orbit plane is analyzed, the coupling effect of navigation deviation on guidance deviation under multi-orbit combination strategy is analyzed, and the influence of guidance deviation on nominal trajectory is analyzed. Based on the deviation characteristics, a real-time space pursuit planning method based on navigation guidance deviation constraints is completed.

Key words: Total autonomy on board; Deviation analysis; Real-time planning.

1 Introduction

Space pursuit and space rendezvous belong to the category of spacecraft orbit control and attitude control. Near-earth space and deep space involve tracking spacecraft to formulate multiple orbital maneuvering strategies, and finally complete the pursuit and rendezvous with the target spacecraft in the established orbit. Space rendezvous technology is an important research and development direction in the space field, which has high economic and strategic value.

By using the physical motion characteristics of fast low orbit and slow high orbit, the trajectory pursuit of the spacecraft relative to the target spacecraft can be realized through the combination strategy of multiple different altitude orbits under time constraints. Arzelier et al. studied the linear pulse optimal rendezvous problem by using polynomial optimization algorithm. [1]Meng Zhanfeng studied the multi-circle multi-pulse lunar orbit rendezvous guidance strategy based on measurement and control constraints, established the initial value guessing algorithm based on the two-body orbit prediction model, and established the optimal solution method of rendezvous guidance strategy under the strong constraints of measurement and control of orbit change points. [2]Aiming at the orbital control planning problem of rendezvous and docking mission, Li Gefei et al proposed the problem of near-Earth space rendezvous mission planning based on the ground TT&C system.[3][4]

On the basis of the growing demand for all-autonomous mission on board, the on-board

navigation devices have been developed rapidly, and the coupling algorithm of on-board autonomous orbit measurement has come into being. It includes the inertial navigation algorithm based on the inertial navigation measurement component and the optical navigation algorithm based on the optical measurement sensor. The rapid development of on-board navigation devices has promoted the emergence of the integrated strategy of on-board autonomous navigation and guidance. The all-autonomous navigation and guidance integration strategy can be applied to the close-earth pursuit and rendezvous mission, the near-moon pursuit and rendezvous mission, the deep space pursuit and rendezvous mission and other mission scenarios. The formulation method of guidance strategy is closely related to the navigation accuracy that can be achieved by the navigation coupling algorithm and the deviation model introduced by the navigation accuracy. Especially in the case of poor navigation accuracy, the guidance strategy needs to be subversive and targeted design, and the design of guidance strategy must meet the robustness requirements of different navigation accuracy.

In this paper, the deviation of the integrated guidance strategy and the real-time planning method based on the deviation constraint are studied, which can be used to support the subsequent related tasks.

2   Space pursuit planning method

In this paper, the deviation of the integrated guidance strategy and the real-time planning method based on the deviation constraint are studied, which can be used to support the subsequent related tasks.

The process of space pursuit for target spacecraft involves the convergent pursuit of distance relative position step by step and the precise pursuit of controlled position speed at close distance. In this paper, the accuracy conditions of the control position and speed precise pursuit of the entrance are set, and the planning is carried out for the space relative position pursuit of the distance. The deviation research based on the on-board navigation system and the fully autonomous pursuit planning are completed by means of step-by-step convergence and space iteration, and the accurate delivery of the entrance accuracy conditions is completed.

Assuming that the spacecraft is controlled by the launch end to complete the orbit entry, the periapsis, orbital inclination, orbital ascendant right ascension, and argument of periapsis are all known and determined information. Orbital element $\sigma(t_0)$ include the apoapsis, orbital inclination, ascendant right ascension, and latitude Angle $\{h_a, i, \Omega, u\}$. For a long distance, the relative position information in and out of orbit plane needs to be adjusted. The orbital factors involved in orbit plane include the height of the far sphere point and the spatial phase of the relative target spacecraft, and the orbital factors involved in orbit plane include the deviation of the orbital inclination and the right ascension of the orbital node of the spacecraft and the target spacecraft. Since the out-of-plane orbital deviation has the orbital characteristics of periodic regression, it is only necessary to adjust the orbital inclination and the right ascension of the orbital node at a specific orbital position to meet the accuracy requirements of the actual mission. This paper mainly studies the deviation analysis and step planning of the in-plane orbit adjustment.

It is assumed that the control position of the close distance speed terminal time pursuit entrance track is a known quantity $\sigma(t_f)$, and the accuracy condition is a known quantity $\Delta\sigma(t_f)$, t_f is the terminal pursuit entrance time.

During orbit flight, different orbital characteristics determine the change of the space phase relative to the target spacecraft. The change of the space phase and the height of the far ball point is simplified as follows:

$$\theta_i = \left(u_i - u_{i-1}\right)\left(1 - \sqrt{\left(\frac{h_{ai} + h_p}{2a_T}\right)^3}\right) \qquad (1)$$

Where, h_p is the height of the spacecraft near the celestial point, and a_T is the semi-major axis information of the target spacecraft.

Design the ladder orbit, the number of steps is n, to achieve the following space phase tracking:

$$\theta = \sum_{i=1}^{n}\left(u_i - u_{i-1}\right)\left(1 - \left(\frac{h_{ai} + h_p}{2a_T}\right)^{\frac{3}{2}}\right) \qquad (2)$$

Each step track evaluates the arrival accuracy index $\Delta\sigma(t_i)$, and the deviation term decomposition considers the guidance deviation introduced by the navigation deviation and the kinematic deviation introduced by the guidance deviation.

The space pursuit planning flow chart is shown as follows:
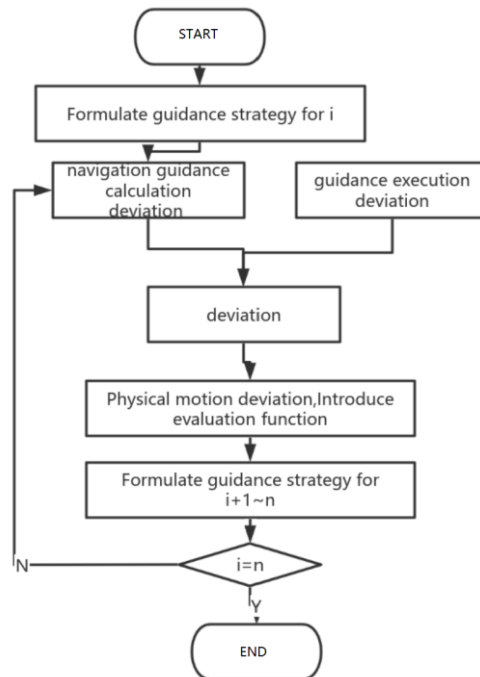
Fig.1　Flow chart of pursuit planning

3　Deviation analysis of space pursuit

Inertial navigation sensor, optical sensor and other measurement components are used on board to complete the on-board orbit autonomous calculation during space rendezvous and pursuit. On-board orbit autonomous calculation relying on inertial navigation sensor to measure components introduces measurement error and long-term stability deviation of inertial components, and on-

board orbit autonomous calculation relying on optical sensor to measure components introduces optical measurement deviation.

The guidance error caused by last orbit change and the navigation error of this time should be considered in the formulation of guidance strategy, and the coupling effect of δh_a and δu should be considered in the calculation process of the guidance error caused by last orbit change. Considering that the in-plane deviation introduced by the navigation deviation at the time of guidance strategy is as follow:

$$\delta\alpha_{i\_nav} = \left\{ \delta h_{ai\_nav}, \delta u_{i\_nav} \right\} \tag{3}$$

The pursuit nominal phase from the current orbit change point to the next orbit change point is as follow:

$$\theta_{iTOi+1} = \left(u_{i+1} - u_i \right) \left(1 - \left(\frac{h_{ai} + h_p}{2a_T} \right)^{\frac{3}{2}} \right) \tag{4}$$

The calculation deviation of the far moon point guidance introduced by the navigation deviation at the current orbit change time is as follows:

$$\delta h_{ai1\_g} = -\delta h_{ai\_nav} + 2a_T \left(\frac{\theta_{iTOi+1}}{u_{i+1} - u_i} \right)^{\frac{2}{3}} \left(\frac{2\delta u_{i\_nav}}{3\theta_{iTOi+1}} - \frac{\delta u_{i\_nav}^2}{9\theta_{iTOi+1}^2} \right) \tag{5}$$

The guidance execution deviation introduced by guidance takes into account the engine installation, engine execution index and spacecraft attitude control deviation is δh_{ai2_g}.[5]

After guidance, due to the characteristics of physical movement, the phase angle deviation from the current time to the next orbit change is introduced as follows:

$$\delta u_{i+1\_gnav} = -\left(u_{i+1} - u_i \right) \left(\frac{h_{ai} + h_p}{2a_T} \right)^{\frac{3}{2}} \left(\frac{3}{2} \frac{\delta h_{ai1\_g} + \delta h_{ai2\_g}}{h_{ai} + h_p} - \frac{1}{9} \left(\frac{\delta h_{ai1\_g} + \delta h_{ai2\_g}}{h_{ai} + h_p} \right)^2 \right) \tag{6}$$

Consider the navigation deviation at the next orbit change time as follows:

$$\delta\alpha_{i+1\_nav} = \left\{ \delta h_{ai+1\_nav}, \delta u_{i+1\_nav} \right\} \tag{7}$$

i+1 in-plane evaluation function is introduced as follows:

$$\begin{cases} \delta h_{ai+1} = -\delta h_{ai+1\_nav} + \delta h_{ai1\_g} + \delta h_{ai2\_g} \\ \delta u_{i+1} = \delta u_{i+1\_gnav} + \delta u_{i+1\_nav} \\ \Delta\sigma_{in\_i+1} = f\left(\delta h_{ai+1}, \delta u_{i+1} \right) \end{cases} \tag{8}$$

4   Real-time planning method based on deviation constraint

From the analysis of deviation term, it can be seen that the deviation introduced into the next guidance increases linearly with the time interval of two guidance. Guidance strategy formulation should reduce the guidance strategy interval as far as possible when the design conditions are satisfied.

At the same time, considering that the guidance and guidance deviation of each orbit change is a cumulative amplification effect on the terminal, a stepwise real-time planning algorithm should be designed in the planning process to avoid the situation that the terminal accuracy evaluation index introduced by the amplification effect of guidance and guidance deviation does not meet the accuracy requirements.

The first objective constraint function is time(or phase), which is often a strong constraint condition for the actual task.

$$\sum_{i=1}^{n}(u_i - u_{i-1}) \in (u_{\min}, u_{\max}) \tag{9}$$

The second objective constraint function is the precision condition of the pursuit point, which is also the strong constraint condition of the practical task.

$$\Delta\sigma(t_f) \in (\Delta\sigma_{\min}, \Delta\sigma_{\max}) \tag{10}$$

The third objective constraint function is the accuracy condition of the step trajectory guidance time, which is the design constraint condition of the actual task.

$$\Delta\sigma(t_i) \in (\Delta\sigma_{i\_\min}, \Delta\sigma_{i\_\max}) \tag{11}$$

In the process of planning the staircase orbit, we need to consider both the arrival accuracy index $\Delta\sigma(t_i)$ and the adaptability of the follow-up mission. The ladder track design function is expressed as follows:

$$h_{ai} = f\left(\Delta\sigma_{in\_i+1}, \theta - \sum_{j=1}^{i-1}\theta_i\right) \tag{12}$$

Step1: According to the time target constraint and the pursuit point accuracy condition target constraint, the nominal step track is designed $h_{a1} \sim h_{an}$;

Step2: Design evaluation function $\Delta\sigma(t_i) = g(h_{ai}, \theta_i)$;

Step3: Reach orbit change,use the evaluation function to evaluate the track information. If the evaluation function index is satisfied, then proceed to the next pulse calculation. If the evaluation function is not satisfied, the current track information $\theta - \sum_{j=1}^{i-1}\theta_i$ is evaluated twice, and the step track design $h_{ai} \sim h_{an}$ of the evaluation results is completed.

5 Simulation examples

Based on space pursuit deviation analysis, the nominal step orbit design is completed considering of time target constraint and pursuit point accuracy condition target constraint.

Based on the real-time planning method, complete the real-time planning, design and simulation of the fully autonomous on-board pursuit orbit.

The simulation results show that the real-time planning and design method based on space pursuit deviation analysis can complete the rendezvous task, and the planning strategy in this paper is feasible.

According to time target constraint and pursuit point accuracy condition target constraint, the nominal ladder track is designed as follows:

Tab.1   Initial design track parameters

	ha (km)	ha precision constraint (km)	θ (deg)	θ Precision constraint (deg)
1	3200+RE	16.5	115.325	11.325
2	500+RE	11.3	42.526	8.255
3	350+RE	7.5	25.123	5.325
4	262+RE	4.5	13.221	2.855
5	242+RE	3.5	6.885	2.015

RE is the centrosome radius(km).

According to the combination of actual mission navigation guidance calculation deviation and guidance execution deviation, the flight track and nominal design value are planned in real time. The simulation results are as follows:

Tab.2   Actual flight track

	ha (km)	θ (deg)
1	3190.458+RE	109.219
2	486.235+RE	37.258
3	353.457+RE	22.985
4	263.872+RE	12.005
5	240.434	6.010

6 Conclusion

The rapid development of navigation and measurement system has put forward the demand for fully autonomous pursuit and rendezvous in the multi-mission scenario on board. The integrated design of fully autonomous navigation and guidance on board needs to have high reliability and high robustness.

In this paper, the problem of in-orbit deviation of space pursuit rendezvous mission is analyzed. The coupling effect of navigation deviation and guidance deviation under the multi-orbit

combination strategy is analyzed. The coupling effect of guidance deviation on nominal trajectory is analyzed. Based on the on-orbit physical characteristics of deviation diffusion, the real-time planning method of space pursuit based on the constraint of navigation guidance deviation is completed. With the orbit characteristics of the actual mission as input,considering the time constraint function, the accuracy condition constraint function of the pursuit point and the accuracy condition constraint function of the step orbit guidance moment as constraint conditions, the spacecraft high-precision pursuit rendezvous task is realized.

References:

[1] ARZELIER D,KARA-ZAITRI M,LOUEMBET C,et al.Using polynomial optimization to solve the fueloptimal linear impulsive rendezvous problem[J].Journal of Guidance,Control and Dynamics,2011,34(5)：1567-1576.

[2] Meng Zhanfeng,Gao Shan,Sheng Ruiqing. Design of Chang 'e-5 lunar orbit rendezvous guidance strategy[J].Acta Aeronautica ET Astronautica Sinica, 2023,44(5): 326584-1-326584-16.

[3] Li Gefei.Rendezvous and docking orbit control planning[M].National Defence Industry Press,2016

[4] Li Gefei,Song Jun,Liu Chengjun.Rendezvous and docking mission orbit control planning, design and implementation[J].Manned Spaceflight,2014,20(1)1-8.

[5] GERGAUD J,HABERKORN T. Orbital transfer:some links between the low-thrust and the impulse cases[J].Acta Astronautica,2007,60(8-9):649-657

Multi-Agent Real-Time Operating System for Collaborative Robot Control

I.S.Arkhipov

(Saint Petersburg State University, Saint Petersburg)

Abstract: The article presents a multi-agent real-time operating system designed for collaborative robot control. It covers the system architecture, current functionalities, and plans for further development, as well as key features that distinguish this solution from existing analogues.

1 Introduction

In recent years, multi-agent technologies have attracted increasing attention, especially in the context of collaborative robot control. Researchers are actively working on the development of integrated systems that combine multiple end-devices with AI-agents into a single ecosystem. Specialised technological solutions are needed to address such challenges. Although current approaches are often based on general-purpose real-time operating systems such as FreeRTOS, the use of specialised software can greatly simplify the process. This paper will present a developed solution focused on real-time multi-agent management.

2 Review of existing solutions

As an example, consider one of the most popular operating systems on the market, FreeRTOS (https://github.com/FreeRTOS). FreeRTOS is open source with an MIT licence and is supported by Amazon. FreeRTOS supports over 35 microprocessor architectures. The Task Manager supports both preemptive and cooperative multitasking, as well as process priorities, semaphores, and queues. It has several narrowly focused variants, such as Amazon FreeRTOS, SafeRTOS, OpenRTOS, and ESP-IDF FreeRTOS. However, the operating system does not have a specialised API for collaborative robot control, such as a network interface. This interface is usually implemented as libraries by hardware manufacturers.

3 Proposed solution

This paper presents a specialised operating system called Martos (https://github.com/IvanArkhipov1999/Martos), designed with a focus on multi-agent technology and collaborative robot control. Martos includes a non-preemptive task manager that provides efficient resource allocation, minimising the overhead of context switching. The system also features memory management mechanisms to optimise the use of available resources and prevent memory leaks.

In addition, Martos provides timer functionality, which is critical for synchronising agent actions and performing time-sensitive tasks. A network interface allows agents to communicate with each other and share data in real time, enabling more efficient, collaborative problem solving.

The operating system is developed in the Rust programming language, which is known for its

securitybeen successfully ported to the ESP32 and ESP32-C6 microcontrollers, which extends its scope and allows the OS to be used in a variety of devices and scenarios requiring high-performance multi-agent control.

4   Planned functionality

In the future, it is planned to develop an application programming interface (API) that will allow searching for the average value of different values residing on different agents in the system. This functionality could be extremely useful in several ways. Firstly, it will allow time synchronisation across different nodes, which in turn will ensure consistency of process execution between different agents. This is especially important for tasks that require precise coordination, such as distributed sensor networks or collaborative robotic operations.

Secondly, this API can be used for distributed computation of neural network weights in a multi-agent system consisting of inexpensive microcontrollers. In such systems, it is often necessary to average the weights received from different agents to improve the quality of model training and increase the overall performance of the neural network.

To implement the search for the mean value within a multi-agent system, we propose to use the local voting algorithm [1]. This algorithm allows each agent to exchange information with its neighbours and make decisions based on the received data, which contributes to a more efficient and faster consensus on the mean value. This approach not only reduces the load on the network, but also increases the resilience of the system to failures of individual agents, which makes it particularly suitable for use in resource-constrained environments.

5   Conclusion

To conclude the paper, the current development of a real-time multi-agent operating system for collaborative robot control shows significant progress in the areas of task, memory and network management. The implementation of a task manager and timer handling functionality already lays the foundation for an efficient system. Future plans include the implementation of an API to search for averages across different agents using a local voting protocol, which will improve the coordination and synchronisation of robot actions. These steps are aimed at creating a more robust and efficient multi-agent system capable of solving complex problems in real time.

References:

[1]   Amelina N. et al. Approximate consensus in stochastic networks with application to load balancing //IEEE Transactions on Information Theory. – 2015. – T. 61. – №. 4. – C. 1739-1752.

Global Finite-Time Stabilization of First-Order Integral Systems With Input Saturation by Gain Feedback Adjustment Bounded Control

Wenyu Ma, Guangyu Liu*, Jian Wang

(School of Automation, Hangzhou Dianzi University, Hangzhou, Zhejiang, 310018, China)

Abstract: Real-world dynamical control systems are generally subject to nonlinearities, such as saturation. To deal with the performance degradation caused by input saturation constraints, we propose a new gain feedback adjustment bounded control method for an integral system with input saturation. The structure of a bounded controller consists of the nested inner and outer layers. The controller in the inner layer uses finite-time states as feedback signals to achieve the stability in finite time. The control gain is adjusted by a constructed nonlinear function with a dead zone characteristic about input saturation to improve transient performance. The outer layer utilizes a saturation function to satisfy the input saturation constraints. Then, it is proved that the closed-loop system is globally finite-time stable. Finally, a tuning algorithm of the control parameters is given through the tradeoff between the convergence time and steady-state oscillation in the closed-loop control system. The effectiveness of the proposed controller is verified in comparison with the existing ones for an integral system with input saturation.

Keywords: input saturation, finite-time stability, bounded control

1 Introduction

Nonlinear phenomena are common. Saturation nonlinearity usually occurs in the fields of aerospace, robotics, machinery [1,2]. Recently, the performance of finite-time convergence has attracted extensive attention. However, it is lacking of an effective control method to achieve finite-time convergence and improve transient performance under saturation constraints, although excellent methods such as robust control and adaptive control have been proposed.

The settling time is an important performance specification of the control system, which characterizes the corresponding convergence rate. Finite-time stability means that the settling time of the system is a finite value [3]. The methods to achieve finite-time stability can be divided into finite-time control for unconstrained systems and saturated systems. A lot of work has been done on finite-time stabilization for better performance and faster convergence [4,5]. Recently, a fixed-time stabilization controller was proposed for a class of integrator systems to make the closed-loop system globally stable [6,7]. A global fast finite-time stabilization of a class of uncertain nonlinear

systems with an adaptive finite-time controller was established by selecting the Lyapunov function [8,9]. However, these finite-time control methods can not consider the input saturation constraints. Therefore, a bounded linear time-varying feedback control is proposed by solving parametric Lyapunov equation [10]. However, it discusses the semi-global finite-time stabilization problem of linear time-invariant systems. The global finite-time stabilization problem of a single-input linear system with input saturation was solved by a bounded controller using smooth saturation functions [11]. A super-twisting algorithm converging in finite time was proposed for a linear system with disturbance and input saturation [12]. But this method only discusses the first-order single-input single-output linear systems. A generic higher-order sliding mode with bounded integral control was proposed for a nonlinear affine system with input saturation to stabilized closed-loop system in a finite time [13]. This method generalizes the studied system to nonlinear systems, but only applies to single-input systems. At present, finite-time control problems considering transient performance have become a hot issue. It is interesting how to solve the problem of faster finite-time convergence for systems under input saturation constraints.

In this paper, the objective is to deal with the problem of the performance degradation caused by input saturation constraints. A gain feedback adjustment bounded control framework is established, where the control structure and parameters are designed respectively. The controller structure is designed by combining a finite-time state feedback and a variable gain function with saturation constraints. A new gain feedback adjustment function is constructed to shorten the settling time as much as possible. A tuning algorithm of control parameters is given to make a tradeoff between the settling time and the steady-state oscillation generated in the case of the imperfections of the practical control system. It is shown that the closed-loop system is globally stable in a finite time. The effectiveness of the proposed control method is shown compared with the existing ones in a numerical simulation, in terms of the transient performance improvement, such as shorter rise time and settling time. The novel results improve the existing ones from several aspects. For example, the finite time stabilization without performance improvement of single-input linear systems with input saturation [11] is extended to the finite time stabilization with performance improvement of single-input linear systems with input saturation. The control method and parameter tuning algorithm are verified in a benchmark.

Notation: Throughout the paper, common use of the abbreviation $[x]^p = |x|^p \mathrm{sign}(x)$ for all $x \in \mathrm{R}$. The absolute value is defined by $|x| = (|x_1|, |x_2|, \cdots, |x_n|)^\mathrm{T}$ for all $x \in \mathrm{R}^n$.

2 Problem formulation

Consider the following first-order integral system with input saturation,

$$\dot{x}(t) = \sigma\big(u(t)\big) \tag{1}$$

where $x(t) \in \mathrm{R}$ is the state, the initial condition of the state is $x(0) = x_0$. $u(t) \in \mathrm{R}$ is the control input. It is assumed that all states are measurable.

The magnitude of the control $u(t)$ is constrained by the saturation characteristic $\sigma(\cdot)$ since any practical actuators can only implement bounded control signals [14]. The saturation function $\sigma(\cdot) : \mathrm{R} \to \mathrm{R}$ is defined as

$$\sigma(u(t)) := \begin{cases} u(t), & \text{if } |u(t)| \le \bar{u} \\ \mathrm{sign}(u(t))\bar{u}, & \text{if } |u(t)| > \bar{u} \end{cases} \tag{2}$$

where $\text{sign}(\cdot)$ is a sign function, and $\bar{u}$ represents the maximum attainable value of the absolute value of the input signal of $\sigma(\cdot)$.

The objective is to design a bounded controller so that the controlled system in Eq. (1) is stabilized in a finite time as short as possible. Considering the system in Eq. (1) with input saturation, design a bounded controller to stabilized the state $x(t)$ of the system in a finite time. At the same time, minimize the settling time and steady-state oscillation in the practical control system to improve the transient and steady-state performance, i.e., $t_s \leq \mathrm{T}$, $c_{ss} \leq \varepsilon$. The following problem will be addressed.

2.1   Problem 1

Consider the system in Eq. (1) with the input saturation in Eq. (2). Design a bounded controller to stabilized globally the state of the system in a finite time as short as possible, i.e., for $u(t) \leq \bar{u}$, there is a real number $\mathrm{T} \in [0, \infty)$, such that

$$\lim_{t \to \mathrm{T}} x(t) = 0, \forall\, x(t) \in \mathrm{R}. \tag{3}$$

At the same time, make a tradeoff between the convergence rate of finite-time stability and the steady-state oscillation in the practical control system to improve the transient and steady-state performance.

Remark: Problem 1 is different from the existing ones in terms of transient performance improvement and tradeoff with steady-state performance. For example, transient performance improvement is ignored in finite-time stabilization problem [11]. Therefore Problem1 is novel and challenging.

3   Novel control framework

A gain feedback adjustment bounded control framework is proposed for an integral system with input saturation in Figure 1. In this framework, the P is used to represent the system in Eq. (1). The C is used to represent the bounded controller.
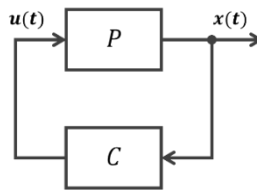
Figure 1. The control scheme diagram of closed-loop system

Task 1. Consider the system P in Eq. (1) with the input saturation in Eq. (2). Design a bounded controller to stabilized the system P in a finite time.

Task 2. Consider the closed-loop system consisting of the system P in Eq. (1) and the bounded controller designed in Task 1. Given a tuning algorithm to make a tradeoff between the convergence rate and the steady-state oscillation to improve transient and steady-state performance.

3.1   Controller structure design.

To fulfill Task 1, we need to design a bounded controller considering the input saturation in Eq. (2). The bounded controller is described as follows,

$$\begin{aligned} u(t) &= \sigma(v(t)), \\ v(t) &= -k_1 \phi(q(t)) \lfloor x(t) \rfloor^{p_1}, \end{aligned} \tag{4}$$

where $k_1 > 0$ and $p_1 \in (0,1)$ are real numbers. A new gain feedback adjustment function to adjust the control gain is constructed as follows,

$$\phi(q(t)) = 1 + \frac{1}{|q(t)| + k_2}. \tag{5}$$

where $q(t) = v(t) - \sigma(v(t))$ is a dead zone characteristic function about input saturation, and $k_2 \in (0,1)$ is a real number.

The gain of the bounded controller in Eq. (4) is adjusted by the constructed gain feedback adjustment function in Eq. (5). In order to reduce settling time, the gain is enlarged to $k_1(1 + 1/k_2)$ when the control input is unsaturated, that is $q(t) = 0$. The gain is kept no less than k_1 when the control input is saturated, that is $q(t) \to \infty$.

3.2 Stability analysis

The investigation of the stability properties of the closed-loop system is performed under input saturation constraints.

The closed-loop system consisting of the system in Eq. (1) with the input saturation in Eq. (2) and the controller in Eq. (4) with the proposed gain feedback adjustment function in Eq. (5) is given by Figure 1.

Theorem 1. Consider the system in Eq. (1) with the input saturation in Eq. (2) and the bounded controller in Eq. (4) with the proposed gain feedback adjustment function in Eq. (5). If the saturation threshold $\bar{u} > 0$, then the equilibrium $x(t) = 0$. of the closed-loop system is globally finite-time stable.

Proof: Consider Lyapunov function candidate $V(t) = \frac{1}{2}x^{\mathrm{T}}(t)x(t)$ whose time derivative along the trajectories of the system in Eq. (1) is derived as follows,

$$\begin{aligned}
\dot{V}(x(t)) &= x^{\mathrm{T}}(t)\dot{x}(t) = x^{\mathrm{T}}(t)\sigma(u(t)) \\
&= -x^{\mathrm{T}}(t)\bar{k}k_1(1 + \frac{1}{|q(t)| + k_2})[x(t)]^{p_1} \\
&\leq -\beta(q(t), \bar{k})V^p(x(t)),
\end{aligned} \tag{6}$$

where $\bar{k} \in (0,1]$ according to $\bar{u} > 0$, $\beta(q(t), \bar{k}) = 2^p\bar{k}k_1(1 + \frac{1}{|q(t)| + k_2})$, and $p = \frac{p_1+1}{2}$.

Therefore, the equilibrium $x(t) = 0$ of the closed-loop system is globally finite-time stable according to Theorem 4.2 in the reference [3] for all $x(t) \in \mathrm{R}$ under $\bar{u} > 0$.

When control input is unsaturated, i.e., $q(t) \equiv 0$, the convergence time is estimated by $\mathrm{T}_1 \leq \frac{1}{\beta(0,1)(1-p)}V^{1-p}(x_0)$. When control input is saturated, the convergence time is estimated by,

$$\mathrm{T} \leq T_w(x_0) + \frac{1}{\beta(0,1)(1-p)}V^{1-p}(x_w), \tag{7}$$

where x_w is the state of the closed-loop system when once $v(t) = -\bar{u}\mathrm{sign}(x_w)$ under $q(t) \equiv 0$, $|x_w| = \left(\frac{k_2\bar{u}}{k_1(1+k_2)}\right)^{\frac{1}{p_1}}$, $T_w(x_0)$ is the time that control input is saturated, which is estimated by the time from $|x_0|$ to $|x_w|$ of the system in Eq. (1) with the controller $u(t) = -\bar{u}\mathrm{sign}(x_0)$.

3.3 Tuning strategy of control parameters

To determine the parameters of the controller in Eq. (4), we need to select three parameters: p_1, k_1, and k_2.

Step one, determine two parameters for p_1 and k_2. In order to ensure the finite-time stability of the closed-loop system, we choose $p_1 \in (0,1)$ according to Theorem 4.2 in the reference [3]. In order to reduce the convergence time T in Eq. (7), we choose $k_2 \in (0,1)$.

Step two, select a parameter for k_1. When the closed-loop system is in a steady state, i.e., $q(t) = 0$ is satisfied. In order to obtain the trajectory $x(t)$ of the closed-loop system, the absolute value $|x(t)|$ of the trajectory $x(t)$ is studied firstly. We obtain

$$\frac{d|x(t)|}{dt} = \text{sign}(x(t))\dot{x}(t) = -\beta_1 |x(t)|^{p_1}, \tag{8}$$

where $\beta_1 = k_1(1 + \frac{1}{k_2})$. The following trajectory of $|x(t)|$ is solved when $|x(t)| > 0$,

$$|x(t)|^{1-p_1} = |x_0|^{1-p_1} - \beta_1(1 - p_1)t. \tag{9}$$

Let $F(|x(t)|) = |x(t)|^{1-p_1}$ to obtain

$$\frac{dF(|x(t)|)}{dt} = -\beta_1(1 - p_1). \tag{10}$$

The amplitude of the steady-state oscillation c_{ss} is proportional to the slope $S = \beta_1(1 - p_1)$, due to the imperfections of the devices in the practical control system.

In order to minimize the convergence time T in Eq. (7) and the steady-state oscillation c_{ss}. The objective function of the tradeoff optimization is built as $F_O(T, c_{ss}) = \lambda T + (1 - \lambda)c_{ss}$, where the convergence time T is replaced with the phased convergence time $T_1 \leq \frac{1}{\beta(0,1)(1-p)} V^{1-p}(x_0)$ to eliminate control saturation, and the steady-state oscillation c_{ss} is replaced with the equivalent slope $S = \beta_1(1 - p_1)$. Therefore, the objective function of the inner loop optimization is given as follows,

$$F_O(T_1, S) = (1 - \lambda)\beta_1(1 - p_1) + \frac{\lambda}{\beta(0,1)(1 - p)} V^{1-p}(x_0), \tag{11}$$

where $\beta_1 = k_1(1 + \frac{1}{k_2})$, $\beta(0,1) = 2^p k_1(1 + \frac{1}{k_2})$, $p = \frac{p_1 + 1}{2}$, and $\lambda \in (0,1)$ is the weight.

We propose the following parameter optimization model, under the parameters p_1 and k_2 are determined by Step one. The associated algorithm is given in Algorithm 1.

$$\min_{\lambda \in (0,1)} \min_{k_1} F_O(T_1, S) \tag{12}$$

To sum up, $p_1 \in (0,1)$, $k_2 \in (0,1)$, and k_1 is given by Algorithm 1.

Algorithm 1. Parameter optimization to minimize the convergence time and steady-state oscillation

---

Input: Given an initial value $V(x_0)$

Output: Optimal k_1

1 Set p_1 and k_2 according to Step one

2 Get optimal k_1 by solving the optimal solution to the problem in Eq. (12)

---

4 Simulation

The functionality of the proposed controller has been tested through simulation on a personal computer with a CPU of 2.00 GHz. The simulation language is C language.

4.1 System description

In order to demonstrate the effectiveness of the proposed method, a simulation study is performed for the integral system in Eq. (1) with input saturation threshold $\bar{u} = 2$.

4.2 The controller

To draw a comparison, different controllers are derived by several control design methods.

WFTC: That generates a windup finite-time controller (WFTC), which is the bounded controller in Eq. (4) without the gain feedback adjustment function in Eq. (5), i.e., $\phi(q(t)) = 1$;

GAFTC: That generates a gain adjustment finite-time controller (GAFTC), which is the bounded controller in Eq. (4) with the gain feedback adjustment function in Eq. (5), but without the parameter tuning in Algorithm 1;

GFABC: That generates the proposed gain feedback adjustment bounded controller (GFABC) in Eq. (4) with the gain feedback adjustment function in Eq. (5). Moreover, the control parameters are determined by the parameter tuning with Algorithm 1;

BFTC [11]: The bounded finite-time control (BFTC) [11] law is $u(t) = \sigma(v_c(t))$, where

$$v_c(t) = -\alpha \tanh^{\frac{p_{c1}}{p_{c2}}}(k_c e(t)) - \beta \tanh^{\frac{p_{c3}}{p_{c4}}}(k_c e(t)), \tag{12}$$

where parameters $\alpha = 1, \beta = 1, k_c = 1, p_{c1} = 7, p_{c2} = 3, p_{c3} = 3$, and $p_{c4} = 5$ are chosen according to Theorem 1 in reference [11].

The parameters of the WFTC and the GAFTC are chosen as $p_1 = 0.5, k_1 = 5$, and $k_2 = 0.5$. The parameters of the proposed GFABC are chosen as $p_1 = 0.5, k_2 = 0.5$, and $k_1 = 3$ determined by Algorithm 1 with $\lambda \in (0.1, 0.9)$ to make a tradeoff between the convergence time and steady-state oscillation.

The proposed GFABC method is compared to the WFTC method, the GAFTC method and the existing BFTC given by the reference [11]. The initial condition of the system is $x_0 = 5$. The simulation step is set to $h = 0.001$ (which represents sampling period of the practical control system).

For the comparison of the transient performance in different closed-loop systems, we define some performance indexes that include the rise time t_r when $|x(t)| \le 0.1 x_0$ for all $t > t_r$, the settling time t_s when $|x(t)| \le 0.02 x_0$ for all $t > t_s > t_r$, and steady-state oscillation $c_{ss} := |x(t)|$ for $t \to \infty$.

4.3 Simulation result

The simulation results of the response of the system in Eq. (1) under the action of different controllers are shown in Figure 2. The corresponding response performance comparisons are shown in Table 1.

(a) The state of the system

(b) The controller

(c) The unconstrained sub-controller

(d) The dead zone characteristic
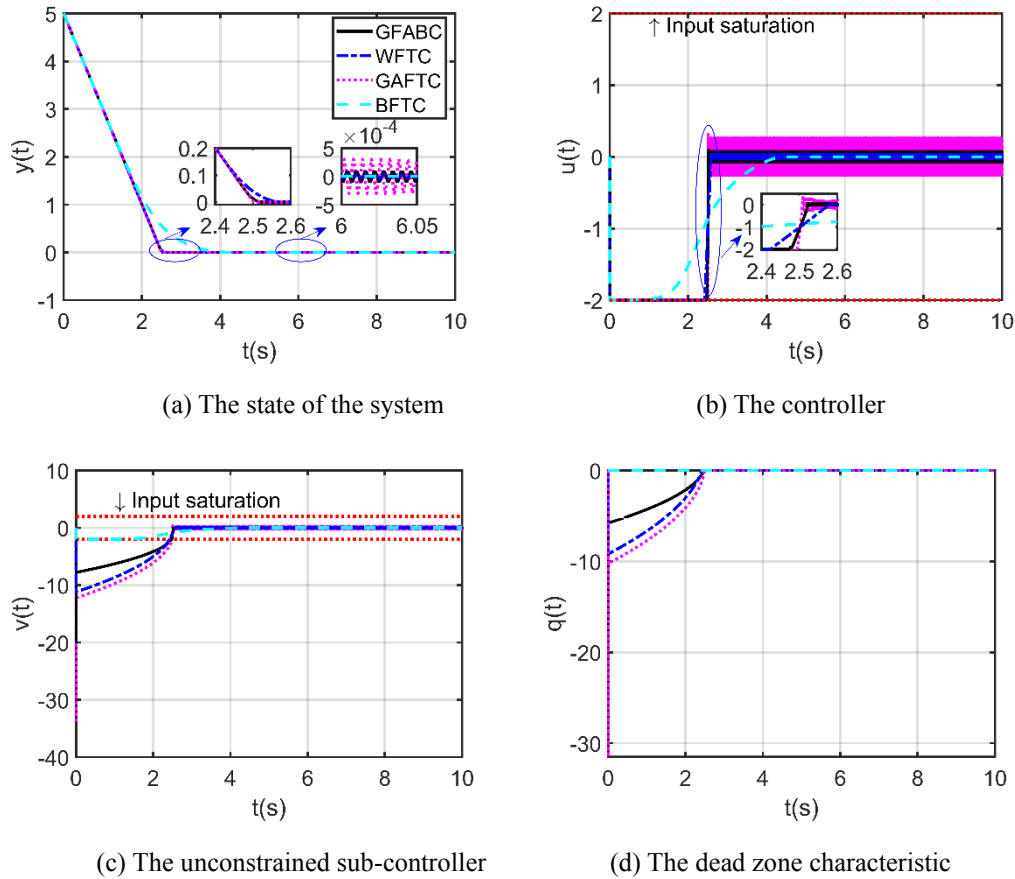
Figure 2. The simulation results of the system with input saturation under different controllers

Table 1. The simulation performance comparison for different methods

Index	BFTC [11]	WFTC	GAFTC	GFABC
t_r (s)	2.5710	2.2510	2.2510	2.2510
t_s (s)	3.3993	2.4542	2.4510	2.4510
c_{ss} (10^{-4})	2.7×10^{-4}	0.3348	3.0135	1.0849

4.4 Analysis and discussion

The control effects of different methods are compared by the simulation. Displayed in Figure 2 and Table 1, $t_s = 2.4510$ (s) of GAFTC is smaller than $t_s = 2.4542$ (s) of WFTC; $c_{ss} = 3.0135 \times 10^{-4}$ of GAFTC is greater than $c_{ss} = 0.3348 \times 10^{-4}$ of WFTC. Thus, the settling time is improved effectively by the proposed gain feedback adjustment function, but increases the steady-state oscillation caused by the fractional power of the state and the sign function. $t_s = 2.4510$ (s) of GFABC is equal to $t_s = 2.4510$ (s) of GAFTC; $c_{ss} = 1.0849 \times 10^{-4}$ of GFABC is smaller than $c_{ss} = 3.0135 \times 10^{-4}$ of GAFTC. In the sub-graphs (c) and (d) of Figure 2, the unconstrained sub-controller of GFABC is less saturated than that of WFTC and GAFTC with the performance is similar. Thus, the proposed control parameter tuning strategy with Algorithm 1 effectively makes a tradeoff between the settling time and the steady-state oscillation. $t_r = 2.2510$ (s) and $t_s = 2.4510$ (s) of GFABC is smaller than $t_r = 2.5710$ (s) and $t_s = 3.3993$(s) of BFTC; but $c_{ss} = 1.0849 \times 10^{-4}$ of GFABC is greater than $c_{ss} = 2.7 \times 10^{-8}$ of BFTC. Thus, the settling time is significantly improved by the proposed control method, although a slight increase in steady-state oscillation is caused by fractional powers of the state and the sign function. However, in terms of

practical control system, a high heat losses is caused by the control switching, in the sub-graph (b) of Figure 2. Therefore, it is worthwhile to solve the chattering phenomenon considering the imperfections of the practical control system in the controller design, in the later research.

5 Conclusions

The gain feedback adjustment bounded control framework is constructed to improve the transient performance under input saturation of the controlled system, where the control structure and parameters are designed respectively. The controller structure is designed by utilizing finite-time state feedback and variable gain function with saturation constraints. the gain feedback adjustment function is constructed to improve the settling time. A parameters tuning strategy is given to make a tradeoff between the settling time and the steady-state oscillation caused by the imperfections of the practical control system. It is shown that the closed-loop system is globally stable in a finite time. The effectiveness of the proposed control method is testified compared with the existing ones in a numerical simulation, in terms of the transient performance improvement, such as rise time and settling time.

References

[1] Q. Hu, X. Shao, Y. Zhang, and L. Guo, "Nussbaum-type function-based attitude control of spacecraft with actuator saturation," International Journal of Robust and Nonlinear Control, vol. 28, no. 8, pp. 2927–2949, 2018.

[2] T. Zhao, Y. Liu, Z. Li, C.-Y. Su, and Y. Feng, "Adaptive control and optimization of mobile manipulation subject to input saturation and switching constraints," IEEE Transactions on Automation Science and Engineering, vol. 16, no. 4, pp. 1543–1555, 2019.

[3] S. P. Bhat and D. S. Bernstein, "Finite-time stability of continuous autonomous systems," SIAM Journal on Control and Optimization, vol. 38, no. 3, pp. 751–766, 2000.

[4] H. Xu, "Finite-time stability analysis: A tutorial survey," Complexity, vol. 2020, pp. 1–12, 2020.

[5] Y. Liu, H. Li, Z. Zuo, X. Li, and R. Lu, "An overview of finite/fixed-time control and its application in engineering systems," IEEE/CAA Journal of Automatica Sinica, vol. 9, no. 12, pp. 2106–2120, 2022.

[6] B. Tian, Z. Zuo, X. Yan, and H. Wang, "A fixed-time output feedback control scheme for double integrator systems," Automatica, vol. 80, pp. 17–24, 2017.

[7] S. Zekraoui, N. Espitia, and W. Perruquetti, "Finite/fixed-time stabilization of a chain of integrators with input delay via PDE-based nonlinear backstepping approach," Automatica, vol. 155, p. 111095, 2023.

[8] Z.-Y. Sun, Y. Shao, and C.-C. Chen, "Fast finite-time stability and its application in adaptive control of high-order nonlinear system," Automatica, vol. 106, pp. 339–348, 2019.

[9] C. Liu and Y. Liu, "Adaptive finite-time stabilization for uncertain nonlinear systems with unknown control coefficients," Automatica, vol. 149, p. 110845, 2023.

[10] B. Zhou, "Finite-time stabilization of linear systems by bounded linear time-varying feedback," Automatica, vol. 113, p. 108760, 2020.

[11] Z. Zuo, X. Li, B. Ning, and Q.-L. Han, "Global finite-time stabilization of first-order systems with bounded controls," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 70, no. 7, pp. 2440–2444, 2023.

[12] R. Seeber and M. Reichhartinger, "Conditioned super-twisting algorithm for systems with saturated control action," Automatica, vol. 116, p. 108921, 2020.

[13] A. Russo, G. P. Incremona, and A. Cavallo, "Higher-order sliding mode design with bounded integral control generation," Automatica, vol. 143, p. 110430, 2022.

[14] A. R. Teel, "Global stabilization and restricted tracking for multiple integrators with bounded controls," Systems and Control Letters, vol. 18, no. 3, pp. 165–171, 1992.

Algorithm for Localization and Direction Determination of Threat Impact Using Vision Language Models in Autonomous Transport Systems

A. R. Brazhenko, A. O. Komarov

(1.St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg;

2.NNSTU n.a. R.E. Alekseev, Nizhny Novgorod)

Abstract: An algorithm is proposed for the localization and analysis of threat directions in autonomous transport systems using Vision-Language Models (VLM). The application of VLM enables more accurate identification of the behavior and trajectories of objects in complex scenarios, including dynamic urban conditions and unpredictable environmental changes.

1　Introduction

Modern autonomous transport technologies face challenges related to safe navigation in real-world conditions, where the risk of hazardous events significantly increases. Existing systems[1], [2] often employ classical image analysis algorithms, which do not always effectively solve problems in complex scenarios. In recent years, methods utilizing deep learning for threat detection have been actively developed, including transformer-based approaches[3] that take contextual aspects into account. However, these methods are limited under real-world conditions, such as the presence of overlapping objects or complex interactions between objects. The proposed algorithm, based on Vision-Language Models (VLM), enhances localization efficiency[4] through the integration of visual and textual information, allowing for more accurate consideration of the dynamics and environmental changes.

2　Statement of the problem

This work addresses the task of creating an algorithm capable of localizing objects that pose a threat to the movement of unmanned ground transport and determining their direction based on images obtained from cameras. The algorithm must take into account not only the visual characteristics of the objects but also their behavioral models, which will help more accurately predict their movement. The key aspects of the task include:

The use of Vision-Language Models for generating textual descriptions of a scene based on images, which allows for obtaining contextual information about what is happening.

Localization and tracking of threatening objects in the frame based on the description obtained using VLM, including both static and dynamic elements.

Analysis of the behavior of objects and their trajectories to assess potential threats to the unmanned vehicle.

3　Algorithm for localization and direction determination of threat impact

The algorithm consists of the following steps:

Transmission of the image to the Vision-Language Model (VLM) [5].

Generation of a textual description of the scene, bounding boxes, and threat probabilities.

Extraction of keypoints of the object.

Determination of direction and visualization of object movement.

The algorithm receives an image I from the vehicle's camera, which is passed to the Vision-Language Model based on the attention mechanism[6] and the Kosmos-2 architecture[7], fine-tuned on specialized contextual tasks using the LoRA adapter[8]. A linear layer with a sigmoid activation function was added to the model's architecture for binary classification, predicting the probability of an object belonging to the classes "threat" or "non-threat."

The LoRA adapter adds learnable low-dimensional matrices to the model architecture for adaptive fine-tuning. The input image I is processed by the base model f and then fed into LoRA using the following operations:

$$W_{LoRA} = W + \alpha \cdot A \cdot B \tag{1}$$

where W— the original weights of the base model, A и B — low-dimensional matrices (LoRA adapters) initialized with random values and optimized during training, and α — a scaling factor that regulates the contribution of the adapter. The scaling factor $\alpha = 8$ was empirically chosen for an optimal balance between the performance of the base model and the improvement in adaptation provided by LoRA.

For the binary threat classification task, a loss function based on binary cross-entropy is used, which is optimized to minimize classification error. An additional regularization term is added, considering the specificity of the data and aimed at reducing false positives for the "no threat" class. The loss function is given by the following formula:

$$L = -\frac{1}{N} \Sigma_{i=1}^{N} \left[y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i) + \lambda \cdot R(W_{LoRA}) \right], \tag{2}$$

where N— the number of samples, y_i— the true class label for the i-th sample ($y_i = 1$ for threat, $y_i = 0$ for "no threat"), $\hat{y}_i$ — the predicted probability of belonging to the threat class for the i-th sample, λ — he regularization coefficient, $R(W_{LoRA})$ — the regularization term that depends on the LoRA weights. The regularization coefficient λ, which varied from 0.01 to 1.0 in the experiments, was chosen to be 0.1, providing a balance between reducing false positives and maintaining threat classification accuracy.

The model generates a textual description of the scene, identifying objects and determining their coordinates as bounding boxes. Using the contextual task P, the model produces the following output:

$$VLM : \{I, P\} \rightarrow \{D, B, C\} \tag{3}$$

where D — the textual description of the scene, a B — the set of bounding boxes that define the positions of identified objects in the image, $C = \{c_1, c_2, \dots c_n\}$ — represents the probabilities of objects belonging to the "threat" class. From the generated textual description, entities, their coordinates, and threat classification are extracted, providing structured information about the scene.

To determine the direction of movement of objects (in this research, pedestrians), a vector is

calculated based on the coordinates of key body points obtained using the MediaPipe Pose model [9]:

If previous frames are available, the coordinates of the key points are compared frame by frame:

$$\Delta x = x_t - x_{t-1}$$
$$\Delta y = y_t - y_{t-1}$$

(4)

where (x_t, y_t) — the coordinates of the key points in the current frame, a (x_{t-1}, y_{t-1}) — the coordinates of the key points in the previous frame. Based on these changes, the direction vector is calculated as

$$Direction = a\tan 2(\Delta y, \Delta x)$$

(5)

If previous frames are not available, the direction is determined solely by the coordinates of the feet and the head. The following formula is used:

$$\text{Direction} = \begin{cases} right, & if\ head\_right \wedge left\_foot\_right \wedge right\_foot\_right \\ left, & if\ head\_right \wedge \neg foot\_right \wedge \neg right\_foot\_right \\ forward, & otherwise \end{cases}, \quad (6)$$

where head_right is determined by the position of the head relative to the center of the shoulders, left_foot_right and right_foot_right are determined by the position of the feet relative to the heels. This allows the system to accurately classify the direction of movement of objects even in the absence of information from previous frames.

After calculating the direction vector for each object, the movement direction is visualized by displaying a vector on the image from the geometric center of the object to the endpoint calculated based on the direction. The direction vector V is defined as:

$$V = (x_{end} - x_{center}, y_{end} - y_{center}),$$

(7)

where (x_{center}, y_{center}) — the coordinates of the geometric center of the object, (x_{end}, y_{end}) — the coordinates of the endpoint calculated based on the object's direction of movement.

4   Example of algorithm performance

The input to the algorithm is an image where an object (pedestrian) is crossing the street.

Output from the Vision-Language Model (VLM):

Scene Context: An image of a woman crossing the street while talking on her phone.

Objects:

('a woman', (12, 19), [(0.578125, 0.296875, 0.890625, 0.984375)], {'threat': 0.9, 'no_threat': 0.1}),

('the street', (29, 39), [(0.015625, 0.203125, 0.984375, 0.984375)], {'threat': 0.05, 'no_threat': 0.95}),

('her phone', (57, 66), [(0.640625, 0.421875, 0.671875, 0.484375)], {'threat': 0.3, 'no_threat': 0.7}).

For classifying objects as "threats," a probability threshold of 0.5 is used: if the probability

exceeds 0.5, the object is classified as a threat. In this case, "a woman" with a probability of 0.9 is classified as a "threat," while "the street" and "her phone" (with probabilities of 0.05 and 0.3, respectively) are classified as "non-threats."

For threats, a movement direction vector is calculated. In the output image (Figure 1), the keypoints of the object (pedestrian) are highlighted, and an arrow indicates the direction of movement.

Fig.1   Example of algorithm performance

The algorithm was trained using 1000 images, and the final F1-score for threat classification was 0.89, which confirms the high accuracy of the algorithm. The metric was calculated using the following formula:

$$F1 = 2 \cdot \frac{\mathrm{Pr}\,ecision \cdot \mathrm{Re}\,call}{\mathrm{Pr}\,ecision + \mathrm{Re}\,call},\tag{8}$$

where *Precision* — the proportion of correctly classified objects among all detected objects, Recall — the proportion of correctly classified objects among all true positive cases.

5   Conclusion

The developed algorithm for localization and direction determination of threats in unmanned ground transport demonstrates high efficiency, achieving an F1 score of 0.89 on a collected dataset containing 1000 images. The novelty of the approach lies in the integration of Vision-Language Models (VLM) with fine-tuning on contextual queries, allowing the use of not only visual data but also textual information for a deeper analysis of object behavior and their interaction with the environment. This approach provides more accurate and adaptive threat recognition compared to classical detectors.

In the future, the algorithm will be improved with a focus on adapting VLM for more precise analysis of complex scenes, including scenarios with high object density and changing lighting conditions.

References:

[1]   Umang Goenka, Aaryan Jagetia, Param Patil, Akshay Singh, Taresh Sharma, Poonam Saini: Threat Detection In Self-Driving Vehicles Using Computer Vision, Network Security and Data Sciences MIND-2021, 2021

[2]   Ahmed M. Ibrahim, Rania M. Hassan: Real-Time Collision Warning System BasedonComputer Vision Using

Mono Camera, 2nd Novel Intelligent and Leading Emerging Sci. Conf.(NILES), 2020

[3] https://habr.com/ru/companies/yandex/articles/763348/

[4] Ziang Guo, Zakhar Yagudin, Artem Lykov, Mikhail Konenkov, Dzmitry Tsetserukou: VLM-Auto: VLM-based Autonomous Driving Assistant with Human-like Behavior and Understanding for Complex Road Scenes, IEEE, 2024

[5] Naoki Wake, Atsushi Kanehira, Kazuhiro Sasabuchi1, Jun Takamatsu, and Katsushi Ikeuchi: Open-Vocabulary Action Localization with Iterative Visual Prompting, arXiv preprint arXiv:2408.17422v4, 2024

[6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhini: Attention Is All You Need, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017

[7] Zhiliang Peng, Wenhui Wang, Li Dong, Yaru Hao, Shaohan Huang, Shuming Ma, Furu Wei: Kosmos-2: Grounding Multimodal Large Language Models to the World, arXiv preprint arXiv:2306.14824, 2023

[8] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen: LoRA: Low-Rank Adaptation of Large Language Models, arXiv preprint arXiv:arXiv:2106.09685v2, 2021

[8] Jong-Wook Kim , Jin-Young Choi, Eun-Ju Ha and Jae-Ho Choi: Human Pose Estimation Using MediaPipe Pose and Optimization Method Based on a Humanoid Model, Applied Sciences 13, 2700., 2023

An Optimal Multimodel Trajectory Generation Method

Yumeng Huang[1], Guangyu Liu[1*], Jian Wang[1,2], Ling Zhu[3], Weijie Yu[1]

(1.School of Automation, Hangzhou Dianzi University, Hangzhou, Zhejiang, 310018, China；2. Department of Control Systems and Informatics, Saint-Petersburg State University of Information Technologies Mechanics and Optics, Kronverkski, 49, SaintPetersburg, 197101, RUSSIA; 3. School of Information Technology and Artificial Intelligence, Zhejiang University of Finance and Economics, Hangzhou, 310012, China)

Abstract: Existing trajectory planning methods have some drawbacks such that a given condition of each method generates only a single solution and these methods are difficult to deal with new situations. However, a planner optimization problem is challenging due to non–convex objective function and multiple solutions. In this study, we propose a new multiple-trajectories generation method in corresponding to different modes of non-convex cost function. To simplify the problem, it is reduced to another problem of estimating the density induced by a distribution based on the cost function. The proposed framework helps select a preferable solution from multiple candidate trajectories, thereby making it easier to be learned by a data-driven representation and generation. We evaluated the proposed method with specific tasks in simulated scenarios.

 Keywords: Optimization, trajectory generation, movement primitives.

1   Introduction

For complex industrial environment, effective motion planning ensures that the robot operated safely, optimally, and accurately, adhering the both physical constraints and task-specific requirements. Optimization-based methods[1-2] are popular approaches for motion planning. The idea behind motion optimization is to formulate motion generation as an optimization problem and often focus on improving existing paths.   Robot manipulator systems are nonlinear systems with high degrees of freedom, making the planner optimization problem complex. Existing methods usually only find a single locally optimal solutions and lack generalization to new situations. In practice, the objective function used in planner optimization is often non-convex and solutions needs to be reproduced by users. We hope to develop a method that employ a small number of parameters to model the demonstrated trajectory, enabling quick reproduction of the demonstration. Our methods estimate the multiple modes of the objective function and easier to learned by users in new situations.

2   Problem Formulation

Planner Optimization Problem is a 4-tuple $\{\mathcal{C}, \mathcal{X}, f, \mathcal{D}\}$, where $\mathcal{C}$ is the problem instance space, $\mathcal{X}$ is the planning parameter space, f is the planning objective, and $\mathcal{D}$ is the problem instance

---

①Corresponding author,Yumeng Huang:myh77954@sina.com

distribution. A planner takes as input a problem instance $c \in \mathcal{C}$ and planning parameters $x \in \mathcal{X}$, and performs some computation on c using x. It returns a numeric value indicating the quality of the computation, which inherently follows the distribution $f(R(\xi))$, the planning objective. Given the start configuration q_0 and the goal configuration q_T, the planning task is to plan a smooth and collision-free trajectory $\xi = [q_{0,...,}q_T] \in R^{D \times T}$ which is given by a sequence of robot configurations. This problem can be formulated as an optimization problem

$$\xi^* = \arg \min f(R(\xi)) \tag{1}$$

In actual production environment, we are more interested in problems where there exits multiple solutions and can be learned quickly by imitation. The goal of this study is to learn the solution manifold when optimizing such objective functions and enable our robots have the ability of imitation learning. Instead of finding a single solution, we aim to train a model that represents the distribution of optimal point $p_\theta(\xi \mid z)$ parameterized with a vector θ given by

$$p_\theta(\xi) = \int p_\theta(\xi \mid z) p(z) dz \tag{2}$$

where z is the latent variable. We train the model $p_\theta(\xi \mid z)$ by maximizing the objective function

$$J(\theta) = \mathbb{E}_{\xi \approx p_\theta(\xi)}[f(R(\xi))] \tag{3}$$

where $f(\cdot)$ is monotonically increasing and $f(x) > 0$ for and $x \in R$.

3 Preliminary

In order to realize the mapping from joint space q_t to task space $x_u(q_t)$, a mathematical kinematics modelling of robot is established to compute a viable joint configuration for the robot. Considering the representation of general rigid body motion, involving translation and rotation, S is the object coordinate frame attached to the body relative to an inertial frame T. The homogeneous representation of a rigid motion can be expressed in the form of Lie group

$$SE(3) = \{e^{\hat{\xi}\theta} = \begin{bmatrix} R & p \\ 0 & 1 \end{bmatrix} \mid R \in \mathbb{R}^3, p \in \mathbb{R}^3, RR^T = I, det(R) = 1\} \tag{4}$$

where SE(3) is the special Euclidean group of rigid body transformation in 3D space, R is the orientation of the origin frame S from the origin S and p is the position vector of frame S, relative to frame S.

Chasles-Mozzi theorem states that any physical rigid body motion from one configuration to another can be realized by a combination of a rotation around a certain axis and a translation along the same axis. This combination of rotation and translation can be called screw motion. In the inertial coordinate system, the screw motion of a rigid body is represented by a twist

$$\hat{\xi} = \begin{bmatrix} \hat{\omega} & v \\ 0 & 0 \end{bmatrix} \in se(3) \tag{5}$$

where $\hat{\xi}$ is the coordinate representation of the twist, v is the translational linear velocity, $\hat{\omega}$

is the rotational angular velocity, $\hat{\omega} = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} \in so(3)$. Define the matrix exponent

and the matrix logarithm as follows:

$$exp : \hat{\xi}\theta \in se(3) \to T \in SE(3)$$
$$log : T \in SE(3) \to \hat{\xi}\theta \in se(3)$$

(6)

The motion of a rigid body can be given by deriving the analytical expression for $e^{\hat{\xi}\theta}$

$$e^{\hat{\xi}\theta} = I + \hat{\xi}\theta + \hat{\xi}^2 \frac{\theta^2}{2!} + \hat{\xi}^3 \frac{\theta^3}{3!} + \dots$$

$$= \begin{bmatrix} e^{\hat{\omega}\theta} & G(\theta)v \\ 0 & 1 \end{bmatrix}, G(0) = I\theta + \hat{\omega}\frac{\theta^2}{2!} + \hat{\omega}^3 \frac{\theta^3}{3!} + \dots$$

(7)

Substituting the definition into $\hat{\omega}^3 = -\omega$ into the above equation and the screw motion of a rigid

body can be expressed in the form of exponential coordinates of the twist $\hat{\xi} = (\hat{\omega}, v)$, that is, any

rigid body motion has

$$e^{\hat{\xi}\theta} = \begin{cases} e^{\hat{\xi}\theta} = \begin{bmatrix} e^{\hat{\omega}\theta} & (1-cos\theta)\hat{\omega} + (\theta - sin\theta)\hat{\omega}^2 v \\ 0 & 1 \end{bmatrix}, \|\omega\| = 1 \\ \begin{bmatrix} I & v\theta \\ 0 & 1 \end{bmatrix}, \omega = 0, \|v\| = 1 \end{cases}$$

(8)

According to the screw theory, the motion of each joint of the robot is generated by the motion twist located on the axis of each joint, so the combined description of kinematics can be obtained. If $g_{st}(\theta)$ is used to represent the initial configuration of the object coordinate framem with respect to the inertial coordinate frame T, and $g_{st}(\theta)$ is used to represent the final configuration. The kinematics of an open-chain robot with n degrees of freedom is expressed by an exponential formula as

$$g_{st}(\theta) = e^{\hat{\xi}_1\theta_1} e^{\hat{\xi}_2\theta_2} \dots e^{\hat{\xi}_n\theta_n} g_{st}(0)$$

(9)

The EC66 robot arm is shown in Figure 1. Each joint is directly driven by a brushless motor combined with a 100:1 zero-gap harmonic drive to drive the gears, which greatly increases the torque available at the joint while reducing its maximum speed. The configuration of the 6-DOF serial robots is expressed by the forward kinematics as formula (9) with $n = 6$, where $e^{\hat{\xi}_1\theta_1} \dots e^{\hat{\xi}_6\theta_6}$ represents Rotation matrix of each joint.
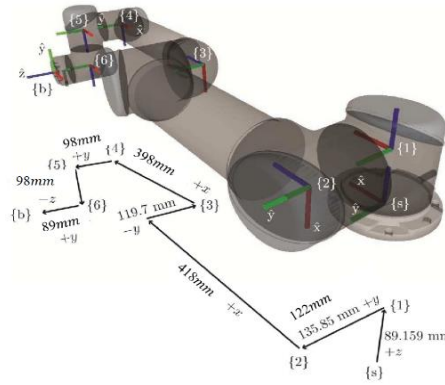
Fig.1 EC66 robot structure.

4 Trajectory learning from multi-solution trajectory Optimization

Consider a distribution over trajectories in the following form:

$$p^{\mathcal{C}}(\xi) = \frac{f(\mathcal{C}(\xi))}{Z} = \frac{e^{-\mathcal{C}(\xi)}}{Z} \tag{10}$$

To address this density estimation challenge, we utilize the importance sampling method. The importance weight is defined as follows:

$$W(\xi) = \frac{p^{\mathcal{C}}(\xi)}{\beta(\xi)} = \frac{f(\mathcal{C}(\xi))}{Z\beta(\xi)} \tag{11}$$

To solve problem (10), it is transformed [3] as minimizaing the KL divergence, which is

$$\theta^* = \arg\min_{\theta} D_{KL}(p^{\mathcal{C}}(\xi) \| p_{\theta}(\xi)) \tag{12}$$

A widely used method for maximizing log likelihood is the variational Bayes expectation-maximization algorithm[4]. One key advantage of the VBEM algorithm is it automatically discards unnecessary clusters and concentrates on distinct modes. So that it facilitates automatic estimation of the number of solutions within our optimal multimodal trajectory generation method. Fig.2 shows the behavior of the cost-weighted density estimation with VBEM.
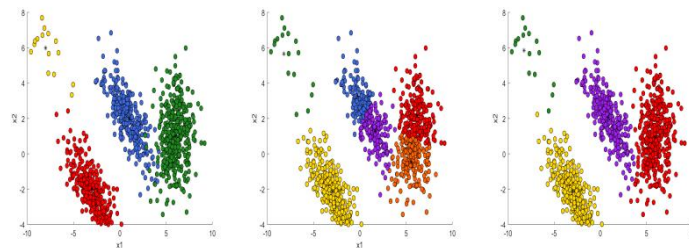
Fig.2 Example of estimating modes of the cost function with importance sampling. (a) shows samples drawn from uniform distributions among 4 types of data intervals. (b) represent 6 clusters found by VBEM. In (c) represent the clusters found by VBEM that is consistent with the original data.

Algorithm 1 Motion planning by learning the multi-solution

Input: Start configuration q_T ;

Output: Planned trajectory ξ ;

1: Initialize the trajectory, e.g., linear interpolation between q_0 and q_T ;

2: Generate N synthetic sample $\xi_{i\,i=1}^{N}$ from $\beta_{traj}(\xi)$

3: Evaluate the objective function $R(\xi)$ and compute the weight $W(\xi)$ for $i = 1,...,N$

4: Convert the trajectory ξ into the trajectory parameter ω using DMPs;

5: Use LWR to determine the parameters ω_i of the weights;

6: Perform dimensionality reduction of the sampled Trajectories parameters, e.g. Laplacian eigenmaps

7: Perform the density estimation using VBEM with importance sampling in (12)

8: Update the trajectories with gradient-descent

9: Project the trajectories onto the constraint solution space

10: if the trajectory ξ is not collision-free then Fine-tune ξ using CHOMP (14)

Return: ξ;

The proposed optimal multimodel trajectory generation method is summarized in Algorithm 1. Trajectaries samples are satisfied as a proposal distribution[5], which is adapted from $\beta_{traj}(\xi) \sim \mathcal{N}(\xi^0, a\mathbf{R})$,where a is a constant, ξ^0 is an initial trajectory, and the covariance matrix $\mathbf{R}$ is given by the Moore-Penrose pesudo inverse of the matrix A defined by the inverse of $A^T A$ .To find multimodel solution of the cost function, we reduce the problem of trajectory optimization to that of the problem of estimatintg the density of the trajectory. Firstly, we apply an importance sampling approach show as line 1-3 in Algorithm 1. In the second step, incorporating the desired property of the trajectory into the trajectory parameterization is essential to reducing the computational cost. For this purpose, we parameterize a trajectory using DMPs method[6] shown as Line 4-5 in Algorithm 1. In our implementation, the density estimation is performed using variational Bayes expectation maximization (VBEM) shown as line 6-7. In the fourth step we locally optimize each trajectory obtained in the first step using a gradient-based method. While the cost weighted density estimation in the first step does not directly minimize the cost function, the second step of our trajectory update directly minimizes the cost function and projects the trajectories onto the space that satisfies the constraints shown as lines 8-9 in Algorithm 1. As the density estimation in the first step of our method is gradient-free, one can interpret that our method alternates be between gradient-free and gradient=based trajectory updates. In trajectory optimization problem in equation (3), $f(R(\xi))$ is the objective function used in previous studies on trajectory optimization[7] given by

$$\mathcal{C}(\xi) = c_{obs}(\xi) + c_{smoothness}(\xi) \tag{13}$$

The first term in equation (13) $c_{obs}(\xi)$, is the penalty for collision with obstacles, the second term is penalty for trajectory smoothness.

The goal of task in Fig.3 is to plan a collision-free trajectory to reach the grasp position of cylinder. Two trajectories are obtained from the proposed optimal multi trajectory generation method. The manipulator goes over the blue collision box in one of the solutions, and behind the box other solution exists. In actual production scenarios, users can choose several solutions to solve

with different situations.
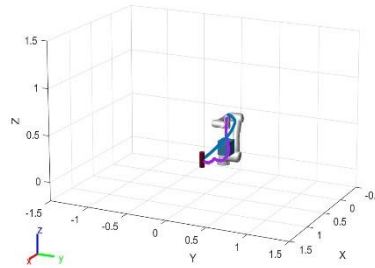
Fig.3   Trajectories in task space with robot EC66.

References:

[1] J. Schulman, Y. Duan, J. Ho, A. Lee, I. Awwal, H. Bradlow, J. Pan, S. Patil, K. Goldberg, and P. Abbeel, "Motion planning with sequential convex optimization and convex collision checking," The International Journal of Robotics Research, vol. 33, no. 9, pp. 1251–1270, 2020/03/01 2014.

[2] M. Mukadam, J. Dong, X. Yan, F. Dellaert, and B. Boots, "Continuoustime gaussian process motion planning via probabilistic inference," The International Journal of Robotics Research, vol. 37, no. 11, pp. 1319– 1340, 2018

[3] Kober J and Peters J (2011). Policy search for motor primitives in robotics.   Machine Learning 84: 171–203.

[4] J. Sung, Z. Ghahramani and S. -Y. Bang, Latent-Space Variational Bayes, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 12, pp. 2236-2242, Dec. 2008.

[5] Kalakrishnan M, Chitta S, Theodorou E, Pastor P and Schaal S (2011). STOMP: Stochastic trajectory optimization for motion planning.   In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). pp. 4569– 4574.

[6] L. Koutras and Z. Doulgeri, Dynamic Movement Primitives for moving goals with temporal scaling adaptation, 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 2020, pp. 144-150, doi: 10.1109/ICRA40945.2020.9196765.

[7] Zucker M, Ratliff N, Dragan A D, et al. CHOMP: Covariant Hamiltonian optimization for motion planning. The International journal of robotics research, 2013, 32(9-10): 1164-1193.

A Scene Building Algorithm for Visualizing Obstacles Using Two Cameras

A. O. Komarov, A. R. Brazhenko

(1.NNSTU n.a. R.E. Alekseev, Nizhny Novgorod; 2.St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg)

Abstract: A building of 2D Bird's-Eye View(BEV) map for visualizing obstacles using information from two cameras is proposed. Based on the depth map, a 2D BEV scene is built, which displays the objects necessary for the functioning of the route planning system. This makes it possible to provide an exact visualization of the environment for subsequent automatic parking, automatic driving or flying in real conditions.

1   Introduction

Currently, the development of autonomous transportation is becoming a key area that will define the future of transportation infrastructure. Autonomous transportation systems have the potential to improve road safety, reduce the number of accidents, and optimize resource usage. One of the most important aspects is the construction of a BEV (Bird's Eye View) map, which provides the autopilot with an accurate representation of the surrounding environment. Modern systems [1], [2], [3] focus on object detection in urban environments, such as traffic signs, pedestrians, and others. However, autonomous vehicles can also be used outside of cities, where a more universal object detection method is required.

This paper is dedicated to an algorithm for constructing a BEV map without the use of LiDAR or radar, unlike the works [4], [5], which employ a multimodal approach. The algorithm demonstrates flexibility under various lighting and weather conditions, in contrast to traditional algorithms such as IPM (Inverse Perspective Mapping), which are sensitive to changes in the environment. The paper presents an approach that uses a neural network to estimate distance, while objects are detected using classical computer vision methods.

2   Problem Statement and Solution Method

The system should automatically process the image, extract depth information of objects, identify obstacles, and display them on a plane corresponding to a BEV (Bird's Eye View) projection. The input consists of an image I_{RGB}, obtained from a camera capturing the surrounding environment in perspective projection. The image I_{RGB} is passed through a trained neural network F_θ (MiDaS[6]) , which generates a depth map $D(x, y)$:

$$D(x, y) = F_\theta(I_{RGB}), \qquad (1)$$

where x, y – pixel coordinates in the image, and $D(x, y)$ - is the distance to the object in the scene for each point.

To obtain objects from the depth map, we will use classical computer vision methods.:

We apply adaptive histogram equalization to obtain a contrast-enhanced image.

We filter the image using Gaussian blur with a kernel 5×5.

We use binarization and morphological filtering to eliminate noise

We find the contours of the obtained objects and create bounding rectangles.

Take camera positioning into account using the function G which transforms the object's bounding rectangle into coordinates on the BEV-map:

$$G(B_i) = \begin{cases} T_{front}\left(\dfrac{\left(x_i+\frac{w}{2}, y_i+\frac{h}{2}, D(x_i,y_i)\right)}{D(x_i,y_i)}\right), & \text{if the front camera is used} \\[2em] T_{rear}\left(\dfrac{\left(x_i+\frac{w}{2}, y_i+\frac{h}{2}, D(x_i,y_i)\right)}{D(x_i,y_i)}\right), & \text{if the rear camera is used} \end{cases} , \qquad (2)$$

where T_{front} — the coordinate transformation matrix from the camera to BEV for the front camera, and T_{rear} — the coordinate transformation matrix for the rear camera, which accounts for the coordinate shift relative to the vehicle (e.g., inversion of direction along the Y-axis).
The center of the object and its width are projected into the BEV coordinate system:

$$P_{cam}(x_{center}, y_{center}, D(x_{center}, y_{center})) = K^{-1} \cdot \begin{bmatrix} x_{center} \\ y_{center} \\ 1 \end{bmatrix} \cdot D(x_{center}, y_{center}), \qquad (3)$$

The final function $F(D(x, y), rear|front)$ - which takes the depth map as input and accounts for the camera position, will include the transformation of bounding boxes to the BEV map, considering the camera position:

$$\{P_{BEV}(X, Y)\} = F(D(x, y), rear|front) = \{G(B_i, rear|front)\} | B_i \in \text{BoundingBoxes}(D(x, y)),$$
$$(4)$$

Thus, for each bounding rectangle B_i we calculate the object's coordinates on the BEV map, considering whether it is the front camera or the rear camera.

3 An example of the algorithm in action

Shows two images from the camera with objects (1 from the front camera and 2 from the rear camera). On the constructed 2D BEV map, three objects are visible: one in front of our vehicle model and two behind it. The white lines represent the camera's field of view.
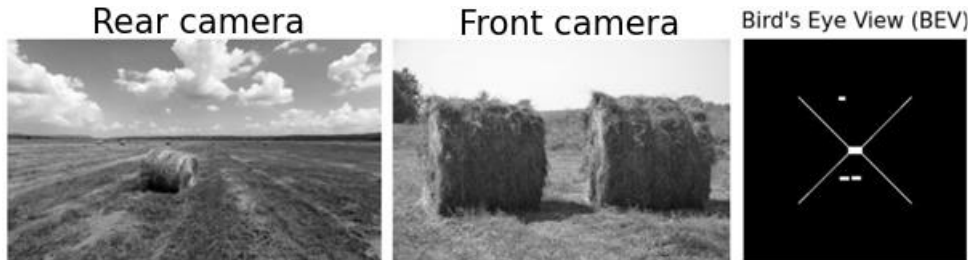
Fig.1 Example of construction 2d BEV

To performance index the quality of the algorithm, we use a combined metric that simultaneously considers both the accuracy of object localization and the ability to determine the distance to objects:

$$\text{LPS} = \alpha \cdot \text{AP} - \beta \cdot \text{DE}_{norm}, \qquad (6)$$

where LPS (Localization Precision Score) – is the quality metric, α and β – are weight coefficients that determine the relative importance of detection accuracy and distance, AP (Average Precision) – is the average precision across various confidence threshold levels for all detected objects, and DE_{norm} – is the distance error.

The algorithm's accuracy based on the LPS metric is 0,57 on 100 images with predominantly agricultural objects, which is a good result for an algorithm that identifies obstacles it has not encountered before.

4 Conclusion

A method for constructing a scene in the Bird's Eye View (BEV) projection for visualizing obstacles based on data from two cameras is proposed. The algorithm includes several key steps: depth map creation, object segmentation in the image, determining their boundaries (bounding boxes), and projecting objects into the BEV coordinate system, considering the camera positions. A metric, Localization Precision Score (LPS), is introduced, which effectively evaluates the algorithm's performance. Testing results on real data show that the algorithm can accurately identify obstacles, providing a high degree of reliability when building a 2D BEV scene.

One of the key advantages of the algorithm is its independence from complex and labor-intensive camera calibration. Moreover, the algorithm does not require the use of expensive sensors such as LiDAR or radar. This makes the proposed approach not only more accessible but also more flexible, ensuring stable performance under various lighting and weather conditions. All of this highlights its potential for use in autonomous driving systems and other fields where efficient environmental visualization is required.

References:

[1] L. Reiher, B. Lampe and L. Eckstein: A Sim2Real Deep Learning Approach for the Transformation of Images from Multiple Vehicle-Mounted Cameras to a Semantically Segmented Image in Bird's Eye View. IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), 2020

[2] Khawlah Bajbaa, Muhammad Usman, Saeed Anwar, Ibrahim Radwan, and Abdul Bais: Bird's-Eye View to Street-View: A Survey. arXiv preprint arXiv: 2405.08961, 2024

[3] Jiayu Yang, Enze Xie, Miaomiao Liu and Jose M. Alvarez: ParametricBEV: Parametric Depth Based Feature Representation Learning for Object Detection and Segmentation in Bird's-Eye View, ICCV, 2023

[4] Wending Zhou, Xu Yan, Yinghong Liao, Yuankai Lin, Jin Huang, Gangming Zhao, Shuguang Cui, Zhen Li: BEV@DC: Bird's-Eye View Assisted Training for Depth Completion, 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 9233-9242

[5] Haotian Hu, Fanyi Wang, Jingwen Su, Yaonong Wang, Laifeng Hu, Weiye Fang, Jingwei Xu, Zhiwang Zhang: EA-LSS: Edge-aware Lift-splat-shot Framework for 3D BEV Object Detection, arXiv preprint arXiv:2303.17895, 2023

[6] Reiner Birkl and Diana Wofk MiDaS v3.1 -- A Model Zoo for Robust Monocular Relative Depth Estimation, arXiv preprint arXiv:2307.14460, 2023

Indoor Pedestrian Location Method by Kalman Filter Based on sEMG-Velocity Mapping Model

Xu Yun, Ma Jiangshan, Zhu Tianyu

(School of Information Science and Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: Aiming at the problem that GPS cannot be used in the indoor environment, an indoor pedestrian location method by Kalman filter based on sEMG-velocity mapping model is proposed. First, sEMG-velocity mapping model is established according to the features extracted from sEMG sensors. Second, pedestrian location estimation model is constructed by Kalman Filter, where the estimated sEMG-velocity is used as the observations. Finally, pedestrian location experiments are carried out. Indoor environment test shows that the maximum location error is about 1.84 m and the location error rate for the total trajectory is about 1.71%. Outdoor environment test shows that the average location error is 2.39 m and the location error rate for the total trajectory is about 0.60%, which is close to the location result based on GPS observations. The proposed pedestrian location method can provide a technical support for the application of indoor environment.

 Key words: pedestrian location; sEMG sensors; indoor environment; inertial measurement unit; feature extraction; Kalman filter

1 Introduction

With the rapid development of artificial intelligence technology and multi-sensor fusion technology, traditional location technology by single sensor cannot satisfy the requirement of indoor location[1]. Therefore, based on the Inertial Measurement Unit (IMU), researchers have focused on the data fusion by other sensors for pedestrian location in the indoor environment, such as barometer, WiFi, magnetic fields, acoustics, Bluetooth, Ultra-Wide Band (UWB) and so on[2-4].

Wang et al.[5] designed a PDR system assisted by acoustic source location and pedestrian movement behavior recognition. A support vector machine algorithm was used to recognize the pedestrian movement behavior of each gait based on data collected by a smartphone. The acoustic location algorithm based on Hamming distance algorithm was proposed to decrease the accumulated error of pedestrian location. Experiment results showed that the recognition accuracy was more than 96%, and location error was about 0.3 m. Huang et al.[6] proposed a Vari-Fi location method based on IMU and WiFi received signal strength. Signal maps was constructed to provide a conditional received signal strength distribution at any given location. The variational inference method was incorporated to optimize the estimated position by optimizing the parameters of the designed inference network by gradient descent-based learning. Location estimation results showed that the

proposed method outperforms the mainstream approaches in both location accuracy and robustness. Liu et al.[7] proposed an indoor pedestrian location scheme. A robust algorithm for UWB location and a heading angle computation strategy for PDR were designed. Then the extended Kalman filter algorithm was designed for information fusion of UWB, PDR, and floor map data. Experiment results indicated that the indoor location accuracy could reach the decimeter level.

Pedestrian location by information fusion of multi sensors has overcome the shortcomings of traditional methods using a single sensor. However, the installation of multi sensors complicates deployment, and the location calculation complexity has also increased. Micro IMU (MIMU) has the advantages of low cost, small size and so on[8], thus it is widely used for pedestrian location by data fusion[9-10]. Surface Electromyography (sEMG) sensors can be used to measure the human muscle activity directly, and they have the advantages of high sensitivity and real-time performance. Therefore, they are widely used for the movement recognition of the pedestrian based on machine learning algorithms[11-12].

2 Pedestrian location based on the fusion of sEMG and MIMU

The proposed indoor pedestrian location scheme based on sEMG and MIMU is shown in Fig.1.
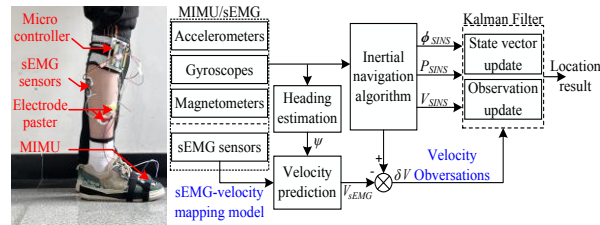
Fig.3 Scheme for indoor pedestrian location

According to the outputs of MIMU worn on the lower limb, inertial navigation algorithm is carried out to calculate the parameters for the pedestrian which contains the velocity, position and the attitude. Then, sEMG-velocity mapping model is constructed based on the output features of sEMG sensors. Additionally, a complementary filtering algorithm is employed to estimate the heading angle of pedestrian by the outputs of magnetometer and accelerometer. And decompose the estimated sEMG-velocity into north and east components. Finally, Kalman filter is applied to realize the estimation of pedestrian location based on the above velocity observations.

2.1 Inertial Navigation Algorithm based on MIMU

Collect the outputs of foot-mounted MIMU and the navigation parameters will be calculated according to the principle of inertial navigation algorithm[13]. In our research, North-East-Up (NEU) navigation coordinate system is used. We only take the ground motion into consideration, and simplify the calculation of the vertical direction.

(1) Pedestrian Velocity Update

In NEU navigation coordinate system, the velocity differential equation for pedestrian is as follows:

$$\dot{V}^n = C_b^n f^b - \left(2\omega_{ie}^n + \omega_{en}^n\right) \times V^n + g^n \tag{1}$$

Where f^b represents the specific force measured by accelerometers, $\omega_{en}^n \times V^n$ denotes the centripetal acceleration generated when pedestrian moves on the earth's surface, $2\omega_{ie}^n \times V^n$ is the

Coriolis acceleration caused by the motion and rotation of earth, g^n is the acceleration of gravity of earth. Decomposing Equation (1) along each axis by:

$$\begin{cases} \dot{V}_N = f_N - (2\omega_{ie}\sin L + \dfrac{V_E}{R_N}\tan L)V_E \\ \dot{V}_E = f_E + (2\omega_{ie}\sin L + \dfrac{V_E}{R_N}\tan L)V_N \end{cases} \tag{2}$$

Where VN and VE are the velocity components in north and east directions; fN and fE are the components of the specific force f in north and east directions; L is the latitude, and λ is the longitude, g is the acceleration of gravity; ωie is the earth's rotation rate, and RM and RN correspond to the radii of curvature of the meridian and prime vertical, respectively. Solve the velocity differential equation in Equation (2), the velocity VSINS of pedestrian motion will be calculated.

(2) Pedestrian Location Update

The location differential equation for pedestrian is as follows:

$$\begin{cases} \dot{L} = \dfrac{V_N}{R_M} \\ \dot{\lambda} = \dfrac{V_E}{R_N\cos L} \end{cases} \tag{3}$$

Solve the location differential equation in Equation (3), the position of pedestrian will be calculated.

As for time progresses, accumulation errors for inertial navigation algorithm will occur due to the drift and bias of inertial sensors[14], which will has a great influence on pedestrian location. To solve this problem, Kalman filter is applied, where velocity observation are used by sEMG sensors.

2.2　sEMG-Velocity Mapping Model

Studies has shown a correlation between the outputs of sEMG sensor and pedestrian walking velocity. Specifically, the collected signal changes will be more obvious for tibialis anterior and gastrocnemius muscles in lower limb. In our research, two sEMG sensors are worn on tibialis anterior and gastrocnemius muscles to find sEMG-velocity mapping model. There are two steps: (1) Gait cycle segmentation; (2) Velocity mapping model establishment.

(1) Gait Cycle Segmentation

Pedestrian walking gait can be segmented into two intervals: zero-velocity interval and non-zero velocity interval. According to the outputs of the accelerometer and gyroscope, we used zero-velocity detection method[15] for gait cycle segmentation, and Fig.2 illustrates a group of segmentation result.
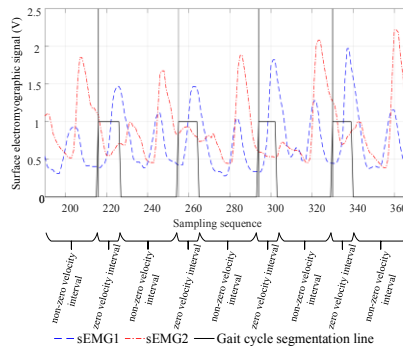
Fig.4　Outputs of sEMG sensors based on zero-velocity detection

In Fig.2, zero-velocity detection method can effectively achieve gait cycle segmentation. And the outputs of two sEMG sensors has the periodicity in each gait cycle.

(2) Velocity Mapping Model Establishment

According to gait cycle segmentation result, the features of sEMG sensors will be extracted for each gait cycle, which contains the mean absolute value (MAV), root mean square (RMS), mean frequency (MF), center frequency (FC), frequency standard deviation (RVF), and root mean square frequency (RMSF). Then, velocity mapping model will be established by BP neural network, as it shows in Fig.3.
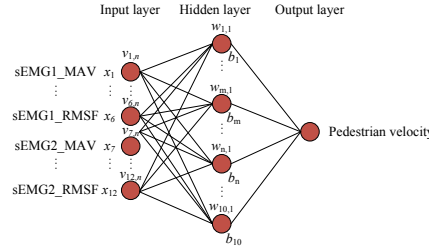
Fig.5 Velocity mapping model by BP neural network

In BP neural network, the input layer contains 12 neurons, which are the 6 output features extracted from the 2-channel sEMG sensors in each gait cycle. The hidden layer contains 10 neurons, where vi,n and wi,n are the weight coefficients of input layer to hidden layer and hidden layer to output layer, respectively. The output layer of BP neural network is the actual pedestrian velocity. During training, the average velocity is used to be the actual pedestrian velocity. Then, a velocity mapping model will be established.

Additionally, according to the outputs of MIMU and magnetometer sensors, the optimized complementary filtering algorithm in[16] is used to estimate the heading angle $\hat{\psi}$. Then, the estimated sEMG-velocity will be decomposed by:

$$\begin{cases} V_{N-sEMG} = V_{sEMG} \cdot \cos\hat{\psi} \\ V_{E-sEMG} = V_{sEMG} \cdot (-\sin\hat{\psi}) \end{cases} \tag{4}$$

Where $\hat{\psi}$ is the heading angle estimated by the optimized complementary filtering algorithm, and VsEMG denotes the estimated sEMG-velocity by velocity mapping model.

2.3 Kalman Filter Based on sEMG-velocity Observations

Define the state vector of Kalman filter as $X=[\delta V_N, \delta V_E, \delta\phi_N, \delta\phi_U, \delta\phi_E, \delta L, \delta\lambda, \nabla_x, \nabla_y, \nabla_z, \varepsilon_x, \varepsilon_y, \varepsilon_z]^T$, where δV_N and δV_E represent the north and east velocity errors. $\delta\phi_N, \delta\phi_U$, and $\delta\phi_E$ are the attitude errors, respectively, δL and $\delta\lambda$ are the latitude error and longitude error. ∇_x, ∇_y, and ∇_z are the bias errors of accelerometers. $\varepsilon_x, \varepsilon_y$, and ε_z are the drift errors of gyroscopes, respectively. Then, the state equation for Kalman filter can be written by:

$$\dot{X} = FX + W \tag{5}$$

The state transition matrix F can be obtained based on the velocity error equation, attitude error equation, and position error equation[17]. Additionally, W represents the process noise matrix.

The observation equation for Kalman filter is given by:

$$Z = HX + V \tag{6}$$

Where V represents the observation noise, and Z is the observation. We use the velocity errors between VsEMG and VSINS as the observation:

$$Z = \begin{bmatrix} \delta V_N \\ \delta V_E \end{bmatrix} = \begin{bmatrix} V_{N-SINS} - V_{N-sEMG} \\ V_{E-SINS} - V_{E-sEMG} \end{bmatrix} \tag{7}$$

Where VN-SINS, VE-SINS are the calculated velocity in north and east direction by inertial navigation algorithm.

The observation matrix H is:

$$H_{2\times13} = \begin{bmatrix} 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \end{bmatrix} \tag{8}$$

According to the procedure arrangement of Kalman filter, pedestrian location will be estimated.

3 Experiment Test and Analysis

We use N200 module and MyoWare sEMG sensor module for the experiment, they are worn on the lower limb of the human body, as it shows in Fig.1. N200 module is a 9-axis IMU, which includes 3-axis accelerometer, 3-axis gyroscope and 3-axis magnetometer. The zero-offset stability of the accelerometer is less than 0.4 mg, and the zero-offset stability of heading gyroscope is 2º/hr, and the zero-offset stability of other two gyroscopes is 5º/hr. Collect the outputs of each sensor by micro controller for the following data process. Then, sEMG-velocity estimation experiment and pedestrian location experiment have been carried out.

3.1 sEMG-Velocity Estimation Experiment

A total of 1,108 data sets have been collected, which include 368 sets of sEMG sensors outputs for normal walking, 239 sets for running, 168 sets for upping the slope, and 172 sets for downing a slope. Divided the dataset into 958 training sets and 150 testing sets randomly for the following estimation.

In the training process, extract the features of sEMG sensors for each gait cycle. The initial training parameters for BP neural network model is set as follows: the input layer size is 12, the hidden layer size is 10, the output layer size is 1, the learning rate is 0.01, and the number of iteration is 1000. Then, the velocity mapping model will be obtained according to BP neural network.

In the testing process, the testing set will be used to evaluate the efficient of sEMG-velocity mapping model. Fig.4 shows the estimated velocities for walking, running, upping the slope and downing a slope.

(a) Velocity estimation and estimation error for walking

(b) Velocity estimation and estimation error for running

(c) Velocity estimation and estimation error for upping the slope

(d) Velocity estimation and estimation error for downing the slope
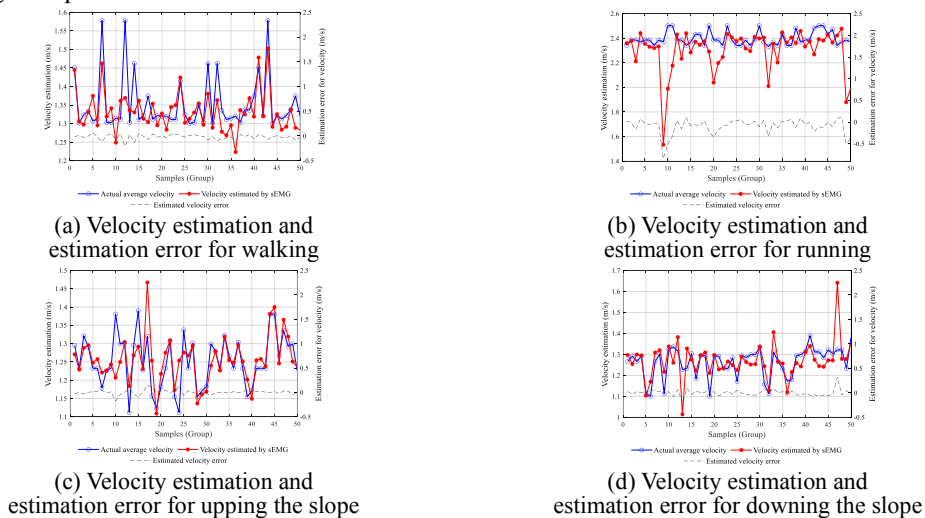
Fig.4 Velocity estimation and estimation error by sEMG

Fig.4 shows that sEMG-velocity mapping model can realize the velocity estimation by the outputs of sEMG sensors by various gaits. The estimated velocity errors for all motions are approximate to ±0.2 m/s, which indicates the efficient of the proposed velocity mapping model. As for running, the estimated velocity error is a litter larger than other motions.

3.2 Pedestrian Location Experiment

We carried out indoor closed loop pedestrian location experiment and outdoor comparison pedestrian location experiment to demonstrate the effectiveness of the proposed indoor pedestrian location method.

(1) Indoor Closed Loop Pedestrian Location Experiment

The indoor closed path is shown in Fig.5. The total length of it is 107.4 m.
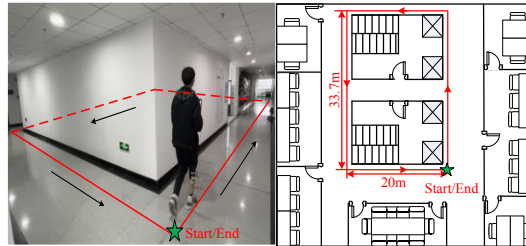
Fig.5   Pedestrian location trajectory for indoor environment

Walk counterclockwise along the indoor closed path, collect the outputs of MIMU and sEMG sensors from the start point to the end point. Use the sEMG-velocity mapping model to estimate the velocity by the outputs of sEMG sensors and decompose it into north and east components, which are shown in Fig.6(a). Pedestrian location result by Kalman filter based on the above observations is presented in Fig.6(b), and the location error is shown in Fig.6(c).
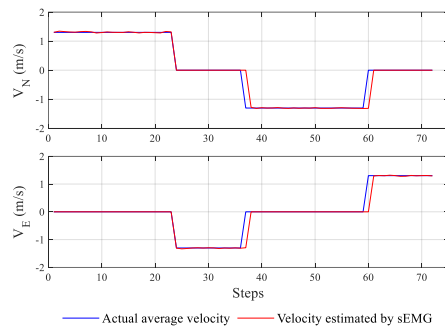
Fig.6(a) Velocity estimated by sEMG sensors
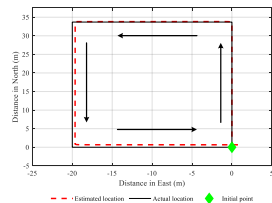
Fig.6(b) Indoor pedestrian location result
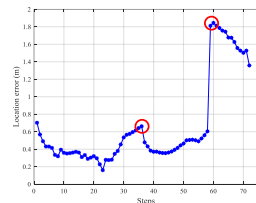
Fig.6(c) Indoor pedestrian location error

Fig.6   Indoor pedestrian location results

In Fig.6(a), the velocity can be accurately estimated according to the outputs of sEMG sensors. However, the velocity estimation errors are obvious when pedestrian is at the turning points. The reason is that our velocity mapping model was trained under the straight-line motion, and there is a

delay for velocity estimation.

As the indoor pedestrian location result shown in Fig.6(b) and Fig.6(c), the proposed indoor location method can achieve a good location result for indoor environment. And the maximum location error is 1.84 m with a total length of 107.4 m, thus the maximum relative location error rate is 1.71% for the total path. It should be emphasized that the red circles in Fig.6(c) mean the moments where the positioning error is relatively large. Comparing with Fig.6(a), the estimated velocity is large at these moments. After these moments, the velocity prediction error decreases, then it will result in a decrease location error.

(2) Outdoor Comparison Pedestrian Location Experiment

Outdoor comparison pedestrian location experiment is carried out on the playground to show the effectiveness of method, and the path is shown in Fig.9.
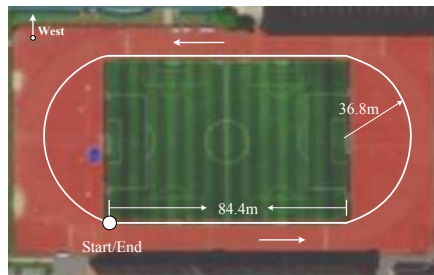
Fig.7   Pedestrian location path for outdoor environment

As for the outdoor experiment, the real-time velocity data from GPS are collected together, and then it will be used for the comparison of location.

According to the velocity mapping model, the pedestrian velocity will be estimated. Decompose it into V_N and V_E, which are shown in Fig.8(a). As for Kalman filter, the estimated velocity by sEMG-velocity mapping model and pedestrian velocity collected by GPS will be used as the observations for the pedestrian location, respectively. Fig.8(b) shows the location results obtained by these two kinds of observations. And the location errors are shown in Fig.8(c).
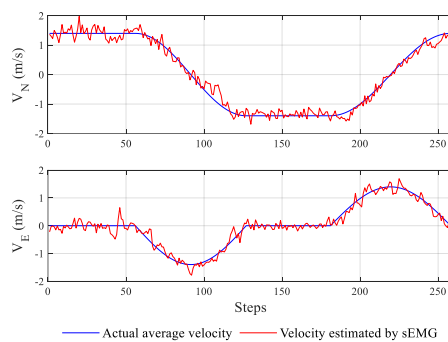
Fig.8(a) Pedestrian velocity estimation for outdoor trajectory
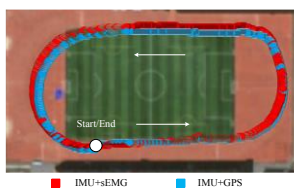
Fig.8(b) Location results for comparison   Fig.8(c) Location error for comparison
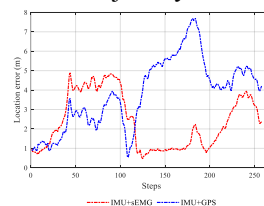
Fig.8   Comparison results for pedestrian location

291

Fig.8(b) and Fig.8(c) show that the location result by two kinds of observations can both achieve a good location result. The average location error based on GPS measured velocity is approximate to 3.82 m, and the location error rate is 0.96% for the total path. Meanwhile, the average location error based on sEMG-velocity mapping model is about 2.39 m, and the average location error rate is 0.60% for the total path. The location results for two kinds of observations are similar, which demonstrates the effectiveness of the proposed pedestrian location method.

However, it should emphasize that there are large location errors for the motion of turning. The reason of it is that sEMG-velocity mapping model is trained under the straight motion, sometimes it is not be suitable for the motion of turning. In the future, we will improve sEMG-velocity mapping model.

4 Conclusion

Indoor pedestrian location method by Kalman filter is studied in this paper. Features are extracted from the outputs of sEMG sensors worn on the tibialis anterior and gastrocnemius muscles of lower limb. And sEMG-velocity mapping model is established according to gait cycle segmentation and BP neural network. Then, pedestrian location estimation model by Kalman filter is established by using the estimated sEMG-velocity as the observations. Indoor closed loop path experiment shows that the maximum location error is 1.84 m, with the maximum relative location error rate is 1.71% for the total path. And the outdoor comparison experiment shows that the average location error of the proposed method is about 2.39 m, with the average location error rate is 0.60% for the total path, which is equivalent to the location estimation result by Kalman Filter using GPS measured velocity as the observations. The proposed indoor pedestrian location method by Kalman filter based on sEMG-Velocity mapping model is effective, and it has a wide application prospects for indoor location.

References

[1] Li Z, Tian ZS, Wang ZC, et al. (2020) Multipath-assisted indoor localization using a single receiver. IEEE Sensors Journal 21(1): 692-705.

[2] Leitch SG, Ahmed QZ, Abbas WB, et al. (2023). On indoor localization using WiFi, BLE, UWB, and IMU technologies. Sensors 23(20): 8598.

[3] Sun C, Zhou J, Jang K, et al. (2023) Indoor localization based on integration of Wi-Fi with geomagnetic and light sensors on an Android device using a DFF network. Electronics 12(24): 5032.

[4] Szyc K, Nikodem M and Zdunek M (2023) Bluetooth low energy indoor localization for large industrial areas and limited infrastructure. Ad Hoc Networks 139: 103024.

[5] Wang M, Duan N, Zhou Z, et al. (2021) Indoor PDR positioning assisted by acoustic source localization, and pedestrian movement behavior recognition, using a dual-microphone smartphone. Wireless Communications and Mobile Computing 2021: 9981802.

[6] Huang H, Yang JF, Fang X, et al. (2023) VariFi: variational inference for indoor pedestrian localization and tracking using IMU and WiFi RSS. IEEE Internet of Things Journal 10(10): 9049-9061.

[7] Liu F, Wang J, Zhang J, et al. (2019) An indoor localization method for pedestrians base on combined UWB/PDR/Floor map. Sensors 19(11): 2578.

[8] Wang JY, Liu JH, Xu XB, et al. (2023) A yaw correction method for pedestrian positioning using two low-cost MIMUs. Measurement 217: 112992.

[9] He GY, Liu MY, Li XY, et al. (2023) Improving the position accuracy of the ZUPT-Aided pedestrian inertial navigation by using a differential layout MIMU array. IEEE Sensors Journal 23(19): 23420-23430.

[10] Jiang P, Chen YP, Zhao BL, et al. (2025) Heading estimation algorithm for pedestrian navigation using a dual-foot-mounted IMU with a millimetre wave radar. Measurement 239: 115497.

[11] Hye NM, Hang U, Chakravarty S, et al. (2023) Artificial intelligence for sEMG-based muscular movement recognition for hand prosthesis. IEEE Access 11: 38850-38863.

[12] Tokas P, Semwal VB and Jain S (2024) Deep ensemble learning approach for lower limb movement recognition from multichannel sEMG signals. Neural Computing and Applications 36(13): 7373-7388.

[13] Lyu P, Wang BQ, Lai JZ, et al. (2023) A factor graph optimization method for high-precision IMU based navigation system. IEEE Transactions on Instrumentation and Measurement 72:9509712.

[14] Hongyang H, Bao L, Feng Z, et al. (2023) A fast intelligent algorithm-based positioning error suppression method for strap-down inertial navigation system. IEEE Sensors Journal 23(5): 5278-5287.

[15] Li QH, Li K and Liang WW (2023) A zero-velocity update method based on neural network and kalman filter for vehicle-mounted inertial navigation system. Measurement Science and Technology 34(4): 045110.

[16] Xu L, Xu Y and Wang FN (2024) Improved pedestrian location method for the indoor environment based on MIMU and sEMG sensors. Journal of Sensors 2024: 2205513.

[17] Kuang J, Xia DZ, Liu T, et al. (2023) Shin-INS: a shin-mounted IMU-based inertial navigation system for pedestrian. IEEE Sensors Journal 23(21): 25760-25769.

Adaptive Tracking Control of Underwater Submarines with State Constraints Under Input Saturation

Wang Di, Liu Mengmeng, Hunag Haoqian, Wang Bing, Wang Junwei

(1. School of Artificial Intelligence and Automation, Hohai University, Changzhou 213000, China)

Abstract: This study proposes an adaptive fixed-time sliding-mode control method with a predefined performance function to ensure that the steady-state response of the closed-loop system is always within a predefined range. A new fixed-time disturbance observer is constructed for real-time estimation and compensation of unknown disturbances. The control algorithm is designed by combining the preset performance control, the fixed-time method, and the non-singular fast terminal sliding mode so that the trajectory tracking error converges to a predefined range. and the auxiliary dynamic equations are introduced for the controller input saturation problem caused by the fixed-time convergence. Finally, the theoretical analysis proves the stability of the closed-loop system.

 Key words: autonomous underwater vehicle (AUV); trajectory tracking; fixed-time control; prescribed performance；

1　Introduction

Underwater vehicles have the ability to perform many complex tasks including marine scientific investigation and underwater target search. They face unknown and complex environments and are also limited by their own state quantities, such as speed and angular velocity, which can be limited by physical limit values.

Underwater vehicle motion control system is an active research field. Many seminal works and papers have been published in conferences and journals. Literature [4] proposes a sliding mode control method for 3D trajectory tracking with exact feedback linearization based on differential geometry, which reduces the system error caused by nonlinear navigators. Literature [5] studies the horizontal plane tracking problem and proposes to introduce the current velocity into the mathematical model of the underwater vehicle, on the basis of which the error equation is established. Literature [6] also addresses 3D trajectory tracking by designing non-singular terminal sliding mode controllers to ensure the error converges to zero in finite time. Compared to conventional USV tracking control approaches, this study introduces the following advancements:

(1) Considering the complex disturbance situations encountered by underwater submersibles underwater, designs a non-singular terminal sliding mode control method based on a fixed-time disturbance observer. The designed fixed-time perturbation observer can accurately observe the external perturbations, and realizes the tracking error to be calibrated to zero in a fixed time.

(2) An adaptive sliding mode control method based on the consideration of input saturation with state constraints for the tracking control of the horizontal plane trajectory of an underwater submarine vehicle. A saturated state auxiliary system is introduced to make the control input continuously smooth.

2　Preparatory knowledge and problem formulation

2.1 Underwater submarine model

In this paper, we only consider the motion in three directions: transverse oscillation, longitudinal inclination and bow direction. establishing the kinematic equations and dynamics model:

$$\dot{\eta} = J(\psi)v \leftrightarrow v = J^{-1}(\psi)\dot{\eta}$$
$$\dot{v} = M^{-1}(-C(v)v - D(v)v + \tau + d) \tag{1}$$

where $\eta = [x, y, \psi]^T$, $v = [u, v, r]^T$ describes its linear and angular velocities, $J(\psi)$ is a rotation matrix, $M(v)$, $C(v)$, $D(v)$ and d are the inertia matrix, the Coriolis and centripetal matrix, the damping matrix, and the unmodeled dynamics.

2.2 Nonlinear model transformations

In order to facilitate the subsequent design of the preset performance function, the tracking error of the trajectory tracking system of the AUV is specified as:

$$e = \eta - \eta_d = \begin{bmatrix} e_x & e_y & e_\psi \end{bmatrix}^T \tag{2}$$

Define the error transformation function as

$$\rho_i(t) = \begin{cases} (\rho_{i,0} - \rho_{i,\infty})\exp\left(1 - \dfrac{T}{T-t}\right) + \rho_{i,\infty}, & t \in [0, T) \\ \rho_{i,\infty}, & t \in [T, +\infty) \end{cases} \tag{3}$$

where $T \in \mathfrak{R}^+$ denotes the scheduled time; $\rho_{i,0}$ and $\rho_{i,\infty}$ is a customized function.

Define the error transformation function as:

$$\varepsilon(t) = \frac{1}{2}\ln\left(\frac{\dfrac{e_i(t)}{\rho(t)} + 1}{1 - \dfrac{e_i(t)}{\rho(t)}}\right) \tag{4}$$

3 Proposed methodology

Lemma 1. Assume that the origin serves as the equilibrium point of the system,

$$x = f(x(t)), x(0) = x_0, f(0, t) = 0 \tag{5}$$

where $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}^n$ if a positive definite, continuously differentiable Lyapunov function $v(x)$ exists such that

$$\dot{v}(x) \leqslant -p_1 v(x)^{q_1} - p_2 v(x)^{q_2} \tag{6}$$

where $0 < q_1 < 1, q_2 > 1$. In such a case, the origin of system (7) is referred to as fixed-time stable, with the convergence time t satisfying

$$t \leqslant \frac{1}{p_1(1-q_1)} + \frac{1}{p_2(1-q_1)} \tag{7}$$

Lemma 2. If $\omega_1, \omega_2, \cdots, \omega_n \geq 0$, then

$$\begin{cases} \displaystyle\sum_{i=1}^{n} \omega_i^k \geq \left(\sum_{i=1}^{n} \omega_i\right)^k, & 0 < k < 1 \\ \displaystyle\sum_{i=1}^{n} \omega_i^k \geq n^{1-k}\left(\sum_{i=1}^{n} \omega_i\right)^k, & 0 < k < \infty \end{cases} \tag{8}$$

Lemma 3. Consider the following system:

$$\dot{x} = -l_1 \operatorname{sig}^{k_1} x - l_2 \operatorname{sig}^{k_2} x, \quad x(0) = x_0 \tag{9}$$

Where $\operatorname{sig}^k(x) = |x|^k \cdot \operatorname{sign}(x), l_1 > 0, l_2 > 0$, $k_1 = (m_1 + 1)/2 + \operatorname{sign}(x-1)(m_1 - 1)/2$, $k_2 = (m_2 + 1)/2 + \operatorname{sign}(x-1)(m_2 - 1)/2$, and $m_1 > 0, 1/2 < m_2 < 1$. This system is a fixed-time stable system. The convergence time is independent of the initial state of the system and satisfies the following:

$$T \leq \frac{1}{l_1(m_1 - 1)} \ln\left(\frac{l_1 + l_2}{l_1}\right) + \frac{1}{l_2(1 - m_2)} \ln\left(\frac{l_1 + l_2}{l_2}\right) \tag{10}$$

4   Controller Design

4.1   Fixed-time perturbation observer design

The auxiliary variable $\varpi(t)$ is introduced as the state estimate of AUV, and the state error variable is defined as: $\Xi(t) = v(t) - \varpi(t) = [\Xi_u, \Xi_v, \Xi_r]^T$

Construct the integral slip mode surface as follows:

$$\begin{cases} \dot{\Theta} = b_1 \dot{\Xi}^{[\alpha]} + b_2 \dot{\Xi}^{[\beta]} \\ s = m\dot{\Xi} + a\Theta = m(\dot{v} - \dot{\varpi}) + a\Theta \end{cases} \tag{11}$$

where a、b_1、b_2、α、β re all greater than 0 and satisfy $0 < \alpha < 1, \beta > 1$.

Derivatives are taken with respect to the sliding mold surface:

$$\dot{s} = \dot{d} - \dot{\hat{d}} + a\dot{\Theta} \tag{12}$$

The estimate for $\hat{d}$ is:

$$\dot{\hat{\zeta}} = a\dot{\Theta} + k_0 \operatorname{sign}(s) + k_1 s^{[\alpha]} + k_2 s^{[\beta]} \tag{13}$$

where k_0, k_1, k_2 are constants greater than 0 and $k_0 \geq |\dot{\zeta}|$ are satisfied.

4.2   Designing an adaptive sliding mode tracking controller

To mitigate the input saturation problem, a following adaptive assist system is designed:

$$\begin{cases} \dot{x}_1 = -c_1 x_1 + x_2 \\ \dot{x}_2 = -c_2 x_2 - JM^{-1}\Delta\tau \end{cases} \tag{14}$$

where $x_1 = [x_{1u}, x_{1v}, x_{1r}]^T$ and $x_2 = [x_{2u}, x_{2v}, x_{2r}]^T$ are auxiliary variables, c_1, c_2 are the appropriate parameter matrices. $\Delta\tau = \tau_c - \tau$ is the difference between the control input and the actual input.

For the variables δ_1, δ_2 a fixed time non-singular terminal sliding mode is constructed:

$$s = \delta_2 + \mu_1 |\delta_1|^\gamma \operatorname{sign}(\delta_1) + \mu_2 |\delta_1|^\lambda \operatorname{sign}(\delta_1) \tag{15}$$

where $\delta_1 = z_1 - x_1 = [\delta_{1u}, \delta_{1v}, \delta_{1r}]^T$, $\delta_2 = z_2 - \dot{x}_1 = [\delta_{2u}, \delta_{2v}, \delta_{2r}]^T$, $\mu_1, \mu_2 > 0, \gamma > \lambda, 1 < \lambda < 2$.

Derivation of Eq. 18 yields:

$$\dot{s} = \dot{\delta}_2 + \mu_1 \gamma |\delta_1|^{\gamma-1} \dot{\delta}_1 + \mu_2 \lambda |\delta_1|^{\lambda-1} \dot{\delta}_1 \tag{16}$$

Choose $\tanh(s/\sigma_1)$ as the sliding mode convergence law

$$\dot{s} = -\rho_1 |s|^{\frac{1}{2}} \tanh(s/\sigma_1) - ks \tag{17}$$

Combining Eq. (17) to Eq. (20), the sliding mode control law can be obtained as

$$\tau = \begin{bmatrix} -\rho_1 |s|^{\frac{1}{2}} \tanh(s/\sigma_1) - ks - \mu_1\gamma |z_1|^{\gamma-1} \dot{e}_1 \\ -\mu_2\lambda |z_1|^{\lambda-1} \dot{z}_1 + \dot{\beta}e + \beta\dot{e} - \dot{\alpha}\dot{e} \end{bmatrix} \frac{M}{\alpha} + M\ddot{\eta}_d + C(v,\eta)\dot{\eta} + D(v,\eta)\dot{\eta} - \hat{d} \tag{18}$$

where $\hat{d}$ is the estimated value of the perturbation observer.

In order to demonstrate that the invention is stabilized at a fixed time, the steps are as follows:

Step 1 Choose the following Lyapunov equation:

$$V_2 = \frac{1}{2} s^T s \tag{19}$$

Derivation of the above equation gives:

$$\dot{V} = s^T(\dot{\alpha}\dot{e} + \alpha\ddot{e} - \dot{\beta}e - \beta\dot{e} + JM^{-1}\Delta\tau + \mu_1\gamma |\delta_1|^{\gamma-1}\dot{\delta}_1 + \mu_2\lambda |\delta_1|^{\lambda-1}\dot{z}_1) \tag{20}$$

Substituting the control law into the above equation gives:

$$\dot{V} = -s^T(\rho_1 |s|^{\frac{1}{2}} \tanh(s/\sigma_1) + ks) = -s^T \begin{bmatrix} \rho_1\left(s/|s|^{\frac{1}{2}}\right) + \text{sign}(s) \cdot \\ \tanh(s/\sigma_1) + ks \end{bmatrix} \leqslant -(u+k)\|s\|^2 \tag{21}$$

where $u = \rho_1 \left[\text{sign}(s)\tanh(s/\sigma_1)\right]_{\min} \left(\sqrt{|s|}\right)_{\max} \geqslant 0$.

Step 2 When the sliding mode reaches the sliding mode surface, there are

$$s \equiv 0, \dot{s} \equiv 0 \tag{22}$$

The association can be shown $\delta_2 + \mu_1 |\delta_1|^{\gamma} \text{sign}(\delta_1) + \mu_2 |\delta_1|^{\lambda} \text{sign}(\delta_1) = 0$. that

$$\dot{\delta}_1 = -\mu_1 \text{sig}^{\gamma}(\delta_1) - \mu_2 \text{sig}^{\lambda}(\delta_1) \tag{23}$$

The above equation satisfies the fixed-time priming.

5 Numerical Simulation Verification

In this part, a series of simulation experiments are proposed to evaluate the superiority of the proposed control scheme, the hydrodynamic parameters are selected as shown in the following table.

Tab1. Hydrodynamic parameters of the Cybership II.

Symbol	Values	Symbol	Values
m	23.800	Y_v	-0.8612
I_z	1.7600	$Y_{|v|v}$	-36.282
x_g	0.0460	Y_r	0.1079
X_u	-0.7230	N_v	0.1052
$X_{|u|u}$	-1.3270	$N_{|v|v}$	5.0437

The total simulation time is 200s, The initial state of the underwater vehicle is adopted respectively as $\eta_0 = [1.5\ 1.4\ 0.5]^T$, $\mu_1 = 0.5, \mu_2 = 0.4, \lambda = 2.5, \gamma = 0.5$, $\rho_1 = [200\ 200\ 50]^T$, $\sigma_1 = [0.01\ 0.01\ 0.9]^T$, $k = [10\ 10\ 5]^T$。

Let the AUV reference trajectory be: $x_d = 5\sin(0.05t), y_d = 6\sin(0.05t), \psi_d = 0.3$

The wind, wave, and current disturbances acting on the USV are assumed to be as follows:

$$d_1 = 21.5\cos(0.1\pi t) + 3.5\sin(1.2\pi t)$$
$$d_2 = 21.5\sin(0.5\pi t) + 4.5\sin(0.1\pi t) \qquad (24)$$
$$d_3 = 11\cos(0.2\pi t) + 12\sin(0.1\pi t)$$

The simulation results demonstrating the control effect of the AUV fixed time controller are shown in Fig. 1 to Fig. 3.
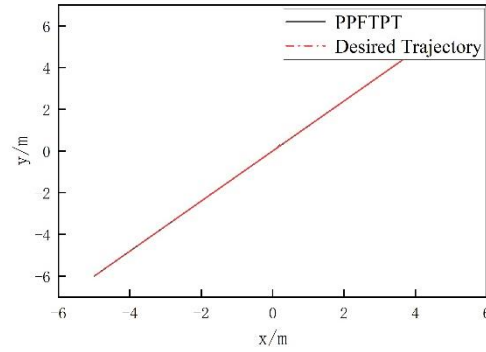
Fig.1　Trajectory tracking with PPFTPT.
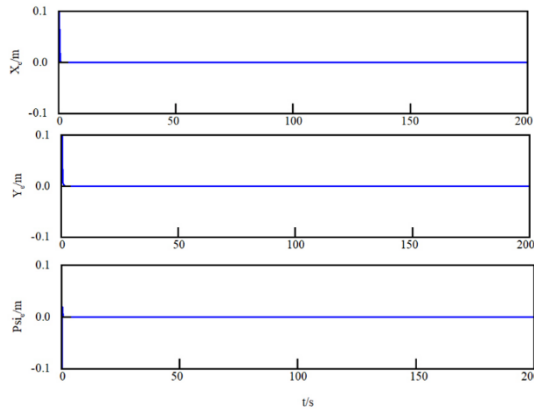
Fig.2　Position error tracking curves　　　　**Fig.3**　Control input curves
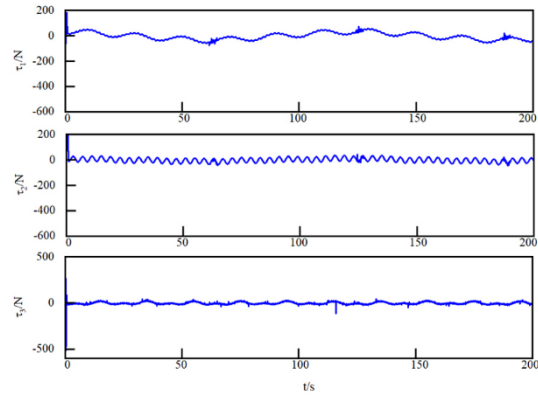
Fig.1 shows the horizontal plane trajectory tracking control curve, the solid line indicates the actual tracking curve obtained under the method of the present invention, and the dashed line indicates the desired trajectory, through which it can be seen that the method proposed in the present invention can realize accurate tracking of the trajectory. Figs. 2 show the longitudinal, transverse and heading angle control input curves, through which it can be seen that the method proposed in the present invention can obtain smooth control input curves, which can effectively reduce jitter. Figs.3 show the tracking error curves for transverse displacement, longitudinal displacement and heading angle, through which it can be seen that the proposed method of the present invention can calm the error to zero in a fixed time.

References:

[1]　XU F, ZHANG L, ZHONG J. Three-Dimensional Path Tracking of Over-Actuated AUVs Based on MPC and Variable Universe S-Plane Algorithms[J]. Journal of Marine Science and Engineering, 2024, 12(3): 418.

[2]　HONG M, GU X, LIU L, et al. Finite time extended state observer based nonsingular fast terminal sliding mode control of flexible-joint manipulators with unknown disturbance[J/OL]. Journal of the Franklin institute, 2023.

[3]　GONZALEZ-GARCIA J, GOMEZ-ESPINOSA A, GARCIA-VALDOVINOS L G, et al. Experimental Validation of a Model-Free High-Order Sliding Mode Controller with Finite-Time Convergence for Trajectory Tracking of Autonomous Underwater Vehicles[J]. SENSORS, 2022, 22(2): 488.

[4] HU Y, SONG Z, ZHANG H. Adaptive sliding mode control with pre-specified performance settings for AUV's trajectory tracking[J]. Ocean Engineering, 2023, 287: 115882.

[5] LI J, DU J, SUN Y, et al. Robust adaptive trajectory tracking control of underactuated autonomous underwater vehicles with prescribed performance[J/OL]. International Journal of Robust and Nonlinear Control, 2019, 29(14)[2024-06-01].

[6] NA J, WANG S, LIU Y J, et al. Finite-Time Convergence Adaptive Neural Network Control for Nonlinear Servo Systems[J]. IEEE Transactions on Cybernetics, 2020, 50(6): 2568-2579.

Quantitative System Analysis of Angular Acceleration Feedback for Attitude Control Synthesis

Hu Yuhui[1], Lin Yue[1]

(1. Beijing Institute of Control Engineering, Beijing 100190, China)

Abstract: Inertial measurement unit (IMU) constituted of the combination of linear accelerometers and rate gyroscopes have been dominated in the control system synthesis of aircrafts for various purposes. However, feedback of high-order angular motion variable, such as angular acceleration, which can enhance the agility of flight vehicles with great demand on maneuverability, has been merely studied. In this paper, a two-loop acceleration Autopilot with Feedback of Angular Acceleration (AFAA) instead of Autopilot with Feedback of Angular Rate (AFAR) is firstly proposed. A new quantitative system analysis approach, namely Degree of Output Controllability (DOC), is derived from the perspective of minimum energy consumption. Meanwhile, null controllable region together with DOO are implemented to numerically reveal the controllability and the robustness of angular acceleration feedback control. Results validate the superiority of AFAA over AFAR and demonstrate the tracking performance enhancement by AFAA.

Key words: Degree of Output Controllability; null controllable region; angular acceleration feedback; quantitative system analysis.

1 Introduction

Nowadays angular acceleration perception technologies have been breaking through some technical bottlenecks and are largely advanced [1, 2]. One of the application platforms for such sensors can be high dynamical flight vehicles, including agile aircrafts and reentry flight vehicles [3, 4]. Generally, acceleration autopilots with multiple feedback loops have been well developed and embedded into the attitude control system. However, compared to the angular-rate feedback loop enclosed in traditional acceleration autopilots, the argumentation of angular-acceleration feedback has been merely studied for flight control systems. From the perspective of system analysis, feedback of angular acceleration can transmit high-order dynamic variation information and thus enhance the control performance, response speed, and the disturbance rejection ability of control systems. Although the estimation and perception of angular acceleration have been successfully employed in a sequence of advanced flight control algorithms [5,6], the characteristics of controllability, stability and robustness of the feedback of angular acceleration has been merely systematically revealed.

In this manuscript, above-mentioned system properties are quantitatively analyzed and compared via quantitative system analysis approach, i.e. degree of output controllability, controllability region and frequency-domain analysis. Generally, the nature of controllability was revealed as "the possibility of realizing arbitrary pole placement via state feedback" [7]. Quantitative controllability analysis, aiming for depicting the inherent properties of the system,

carves out a distinctive approach to the control system analysis, synthesis, and optimization [8, 9]. So far, the state of the art on controllability measures, or namely the degree of controllability, which mainly consist of Gramian matrix-based approaches [10] and PBH modal methods [11] mainly focuses on the comparative analysis and optimization of the system structural parameters such as optimal actuator placement, actuator patterns selection, trajectory optimization. Nevertheless, studies on controllability measures have been merely applied to the analysis of different system feedback loops thus far. Besides, since acceleration autopilots with angular accelerometer is related to output feedback tracking problems, it is more proper to investigate the physical characteristics of such acceleration autopilots in the framework of output controllability.

In this paper, a novel two-loop acceleration autopilot with feedback of angular acceleration (AFAA) for attitude control is proposed. Degree of output controllability (DOC) criterion is derived for the system with direct feedthrough term from input to output introduced by the acceleration measurement. Further, the controllability region of AFAA is investigated and compared relative to the traditional two-loop autopilot with feedback of angular rate (AFAR). The effectiveness of AFAA over AFAR with different static stabilities is validated quantitatively via output controllability analysis and tracking performance comparison.

2 Problem Formulation

2.1 Dynamic equations

Consider the general model for acceleration autopilot derived from nonlinear motion dynamics via coefficient freezing method and small perturbation linearization [12]

$$\begin{cases} \dot{\alpha} = \dot{\vartheta} - b_\alpha \alpha - b_\delta \delta \\ \ddot{\vartheta} = -a_\omega \dot{\vartheta} - a_\alpha \alpha - a_\delta \delta \end{cases} \tag{1}$$

where α and ϑ are the angle of attack and pitching angle, respectively. $a_\alpha, a_\delta, a_\omega, b_\alpha, b_\delta$ are related aerodynamical coefficients.

To realize the control system analysis and synthesis, Eq. (1) in rewritten state-space form

$$\begin{bmatrix} \dot{\alpha} \\ \ddot{\vartheta} \end{bmatrix} = \begin{bmatrix} -b_\alpha & 1 \\ -a_\alpha & -a_\omega \end{bmatrix} \begin{bmatrix} \alpha \\ \dot{\vartheta} \end{bmatrix} + \begin{bmatrix} -b_\delta \\ -\alpha_\delta \end{bmatrix} \delta \tag{2}$$

For AFAA, the angular acceleration is enclosed as the assistance feedback signal in the loop. As for AFAR, the angular rate is feedbacked to enhance the equivalent damping coefficient of vehicles. Output models have the following forms

$$\begin{bmatrix} a_y \\ \ddot{\vartheta} \end{bmatrix} = \begin{bmatrix} Vb_\alpha & 0 \\ -a_\alpha & -a_\omega \end{bmatrix} \begin{bmatrix} \alpha \\ \dot{\vartheta} \end{bmatrix} + \begin{bmatrix} Vb_\delta \\ -a_\delta \end{bmatrix} \delta_e \tag{3}$$

$$\begin{bmatrix} a_y \\ \dot{\vartheta} \end{bmatrix} = \begin{bmatrix} Vb_\alpha & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \dot{\vartheta} \end{bmatrix} + \begin{bmatrix} Vb_{\delta_e} \\ 0 \end{bmatrix} \delta_e \tag{4}$$

The block diagrams of these two autopilots are presented in Fig. 2. K_{DC} denotes the feedthrough command correction gain, and K_A, K_g and K_{aa} are the tunable feedback gains. Linear quadratic output feedback control algorithm proposed in [13] is adopted for parameters

tuning under the identical constraints of crossover frequency.
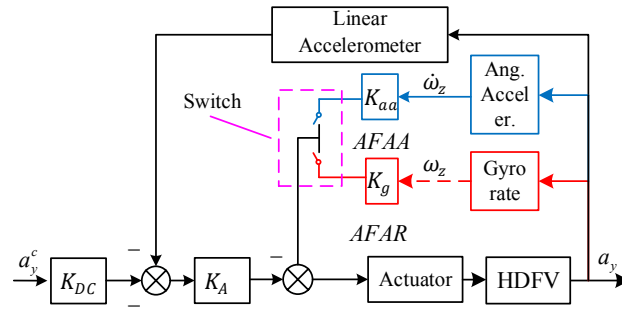
Fig.1   Diagram of two-loop acceleration autopilots AFAA/AFAR

Besides, a prototype based on force balance angular accelerometer SR-207RFR with stable bias and large functioning bandwidth is involved for the system analysis and synthesis [14]

$$G_{aa}(s) = \frac{\dot{\omega}_{meas}}{\dot{\omega}_{act}} = \frac{\left(3.983\times10^6\, s + 1.01\times10^5\right)/K_v}{s^5 + 400.1 s^4 + 1.043\times10^5 s^3 + 1.499\times10^7 s^2 + 9.946\times10^8 s + 1.346\times10^7} \quad (5)$$

where $\dot{\omega}_{meas}$ and $\dot{\omega}_{act}$ refer to measured and actual angular acceleration.

3   Quantitative System Analysis

In this section, a novel controllability criterion, named the degree of output controllability (DOC) is proposed. Then DOC together with controllability region are implemented to quantitively reveal the difference between different output feedback loops.

3.1   Degree of Output Controllability

According to the longitudinal dynamics of AFAA and AFAR derived above, let's consider the following LTV system Σ_o with direct transmission from the input to the output introduced by acceleration measurement as

$$\begin{aligned}
\dot{\mathbf{x}}(t) &= \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \\
\mathbf{y}(t) &= \mathbf{C}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t)
\end{aligned} \quad (6)$$

Then output transfer of the system Σ_o within time intervals $[t_0, t]$ can be described as

$$\mathbf{y}(t) = \mathbf{C}(t)\mathbf{\Phi}(t, t_0)\mathbf{x}_0(t) + \int_{t_0}^{t} \mathbf{\Pi}(t, \tau)\mathbf{u}(\tau)d\tau + \mathbf{D}(t)\mathbf{u}(t) \quad (7)$$

where $\mathbf{\Phi}(t, t_0)$ is the state transition matrix, and the matrix $\mathbf{\Pi}(t, \tau) = \mathbf{C}(t)\mathbf{\Phi}(t, \tau)\mathbf{B}(t)$

For system Σ_0 with direct transmission $\mathbf{Du}$, degree of output controllability can be measured by the minimum energy consumption on transferring the output from $\mathbf{x}_0 = \mathbf{x}(t_0)$ to $\mathbf{y}(t_f) = \mathbf{y}_1$ in the interval $[t_0, t_f]$ as

$$\begin{aligned}
J^* &= \min_{u(t)} \frac{1}{2}\int_{t_0}^{t_f} \mathbf{u}^T(\tau)\mathbf{u}(\tau)d\tau + \frac{1}{2}\xi^2 \\
\text{s.t. } \mathbf{y}_1 &= \mathbf{C}(t_f)\mathbf{\Phi}(t_f, t_0)\mathbf{x}_0 + \int_{t_0}^{t_f} \mathbf{\Pi}(t, \tau)\mathbf{u}(\tau)d\tau + \mathbf{D}(t_f)\xi
\end{aligned} \quad (8)$$

where ξ corresponds to the control input $\mathbf{u}(t_f)$ at the final moment $t = t_f$.

To this strictly convex minimization problem with constraint, the Lagrange multiplier λ is introduced to transform the original formulation to the following conditional extremum problem

$$L(\mathbf{x}, \mathbf{u}, \mathbf{\Lambda}) = \frac{1}{2} \int_{t_0}^{t_f} \mathbf{u}^T(\tau) \mathbf{u}(\tau) d\tau + \frac{1}{2} \xi^2 + \lambda \left[\mathbf{y}_1 - \mathbf{C}(t_f) \mathbf{\Phi}(t_f, t_0) \mathbf{x}_0 - \int_{t_0}^{t_f} \mathbf{\Pi}(t, \tau) \mathbf{u}(\tau) d\tau - \mathbf{D}(t_f) \xi \right] \quad (10)$$

To achieve the extremum of Lagrangian, the following optimality conditions must be satisfied

$$\frac{\partial L}{\partial \mathbf{u}} = \mathbf{u} - \left[\int_{t_0}^{t_f} \mathbf{\Pi}(t, \tau) + \mathbf{D}(t_f) \right]^T \lambda = 0 \quad (9)$$

where λ is obtained by solving $\mathbf{W}_{tf} \lambda = \mathbf{y}_1 - \mathbf{C}(t_f) \mathbf{\Phi}(t_f, t_0) \mathbf{x}_0$.

When the system is output controllable, i.e., the output controllability Gramian $\mathbf{W}_{tf}$ is invertible [15], the piecewise-continuous optimal control input function has the form

$$\mathbf{u}^*(t) = \begin{cases} \mathbf{\Pi}^T(t, \tau) \mathbf{W}_{tf}^{-1} \delta \mathbf{y}, & t \in [t_0, t_f), \\ \mathbf{D}^T(t_f) \mathbf{W}_{tf}^{-1} \delta \mathbf{y}, & t = t_f. \end{cases} \quad (10)$$

where $\delta \mathbf{y} = \mathbf{y}_1 - \mathbf{C}(t_f) \mathbf{\Phi}(t_f, t_0) \mathbf{x}_0$ is defined as the output residual vector

Further, we can get the energy-optimal performance index as follows

$$J^* = \int_{t_0}^{t_f} \left(\mathbf{\Pi}_{tf}^T \mathbf{W}_{tf}^{-1} \delta \mathbf{y} \right)^T \mathbf{\Pi}_{tf}^T \mathbf{W}_{tf}^{-1} \delta \mathbf{y} d\tau + \mathbf{D}^T(t_f) \mathbf{W}_{tf}^{-1} \delta \mathbf{y} \left(\mathbf{D}^T(t_f) \mathbf{W}_{tf}^{-1} \delta \mathbf{y} \right)^T = \delta \mathbf{y}^T \mathbf{W}_{tf}^{-1} \delta \mathbf{y} \quad (11)$$

It can be understood intuitively, that the less minimum energy is required for the output transfer, the higher output controllability that the system owns. The larger the minimum eigenvalue of $\mathbf{W}_{tf}$ is, and then the higher controllability the system possesses. Then, DOC for LTV systems can be defined as

$$D^\dagger = \sigma_{min}(\mathbf{W}_{tf}) \quad (12)$$

where σ_{min} is the smallest eigenvalue of $\mathbf{W}_{tf}$.

With the comparison of Eq. (3) and (4), the difference between AFAA and AFAR lies in the output matrix and the control input formed based on the sensed angular motions. Thus, the proposed DOC criterion can serve as a performance indicator for control system synthesis and optimization.

3.2 Controllability region analysis

Controllability region characterizes the robustness of the control system disturbed away the equilibrium state. A system state is null-controlled if it can be steered to the origin in a finite time with an admissible control. The set of all null-controlled states is called the null controllable region of the system. For high-dynamic flight vehicles with unstable poles, null controllable region $\mathcal{C}$ is a bounded convex open set containing the origin.

Let's consider a general dynamical system with bounded controls

$$\dot{x}(t) = Ax(t) + B\sigma(u(t)) \quad (13)$$

where $\sigma(\cdot)$ denotes the saturating function.

We will characterize the null controllable region $\mathcal{C}$ of the original system by studying the reachable region $\mathcal{R}$ of its time-reversed system as

$$\dot{z}(t) = -Az(t) - B\sigma(v(t)) \quad (14)$$

where $z(t) = x(t_1 - t)$, $z(0) = x_1$, $z(t_1) = x_0$ and $v(t) = u(t_1 - t)$.

According to [16], suppose that A is anti-stable, denote that

$$\varepsilon = \left\{ v(t) = sign(c^t e^{At} B) : t \in \mathrm{R}, c \neq 0 \right\} \quad (15)$$

For a bounded control v, $|v(t)| < 1$ for all $t \in \mathrm{R}$, the reachable region is strictly convex

$$\mathcal{R} = \left\{ z = \int_{-\infty}^{0} e^{A\tau} Bv(\tau) d\tau : v \in \varepsilon \right\} \quad (16)$$

Denote the trajectory of (14) under the control of v as

$$\partial \mathcal{R} = \left\{ \Phi(t,v) = \int_{-\infty}^{t} e^{-A(t-\tau)} B v(\tau) d\tau : t \in \mathrm{R}, v \in \varepsilon \right\} \tag{17}$$

Further, the boundary of the null controllable region is

$$\partial \mathcal{C} = \left\{ \Phi(t,v) : t \in \mathrm{R}, v \in \varepsilon \right\} \tag{18}$$

Particularly, for a second-order anti-stable system, where A has two real eigenvalues

$$\partial \mathcal{C} = \left\{ \pm \left[e^{-At} z_e^- - \int_{-\infty}^{t} e^{-A(t-\tau)} B d\tau \right] : t \in [0,\infty] \right\} = \left\{ \pm \left(-2e^{-At} + I \right) A^{-1} B : t \in [0,\infty] \right\} \tag{19}$$

where $z_e^- = A^{-1} B$ denotes equilibrium point under constant control $u = -1$. If A has a pair of complex eigenvalues $\alpha \pm j\beta, \alpha, \beta > 0$, then

$$\partial \mathcal{C} = \left\{ \pm \left[e^{-At} z_s^- - \int_{-\infty}^{t} e^{-A(t-\tau)} B d\tau \right] : t \in [0, T_p] \right\} = \left\{ \pm \left[e^{-At} z_s^- - \left(I - e^{-At} \right) A^{-1} B \right] : t \in [0, T_p] \right\} \tag{20}$$

where $T_p = \pi / \beta$ and $z_s^- = \left(I + e^{-AT_p} \right)^{-1} \left(I - e^{-AT_p} \right) A^{-1} B$.

4 Simulation and Results

In this section, the controllability and robustness analysis are carried out numerical to compare the performance of attitude control system with different feedback loops. A trim flight condition with a set of nominal aerodynamic coefficients $a_\omega = 1.5$, $b_\alpha = 1.6$, $b_\delta = 0.23$, $a_\delta = 280$, $a_\delta = 250$ is selected for the control system synthesis and controllability analysis. Besides, HDFVs with different stability properties, i.e., static stability $a_\delta = 250$, neutral stability $a_\delta = 0$, and static antistability $a_\delta = -250$, are included for the comparison.

As is displayed in Fig. 2, AFAA with the relatively poor sensor accuracy compared to AFAR, still demonstrates its superiority in faster response speed and smaller steady error during different reference commands. As presented in Fig. 3 a), DOC increases with the degeneration of the static stability properties. Besides, the controllability of AFAA is enhanced compared with the traditional AFAR, which can be validated by the tracking response in Fig. 2. As presented in Fig. 3 b), null controllable region $\partial \mathcal{C}$ expands with a_α closing to 0, and more easily the system stabilizes after the perturbation.
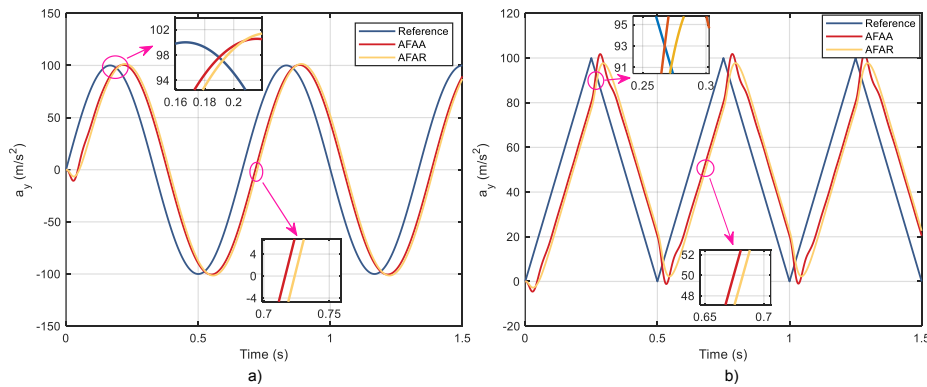
Fig.2 Diagram of two-loop acceleration autopilots AFAA/AFAR
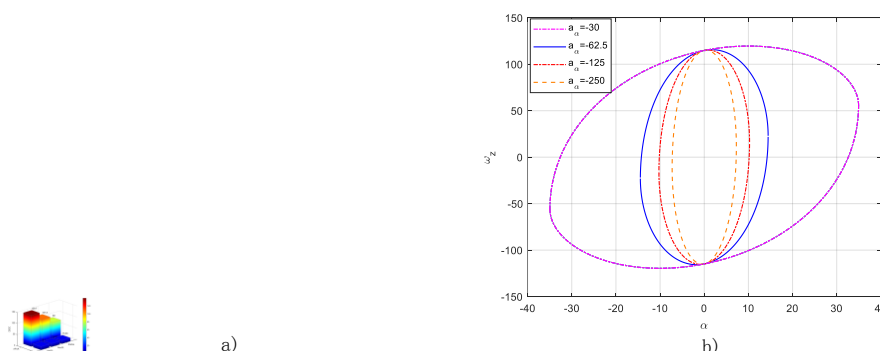
Fig.3　Controllability Analysis of AFAA/AFAR with different stability properties
a)　DOC; b)Controllability region

5　Conclusions

In this study, the general characteristics of the angular acceleration feedback loop are analyzed from the controllability, robustness and stability. A novel criterion for measuring output controllability is proposed based on the minimum energy-control problem. The numerical simulations have revealed that the AFAA owns the higher controllability compared to AFAR, and this deduction is verified by the superiority of AFAA in high maneuvers with faster response and high control accuracy according to simulation results. Besides, the robustness of angular acceleration feedback control is reduced, which recommends the implementation of auxiliary filter during the attitude control synthesis.

References:

[1]　Cheng S, Fu M, Wang M, et al. Modeling for fluid transients in liquid-circular angular accelerometer[J]. IEEE Sensors Journal, 2016, 17(2): 267-273.

[2]　Liu B, Wang J, Chen D, et al. A MEMS-based electrochemical angular accelerometer with integrated plane elec-trodes for seismic motion monitoring[J]. IEEE Sensors Journal, 2020, 20(18): 10469-10475.

[3]　Xu Q, Chang S, Wang Z. Acceleration autopilot design for gliding guided projectiles with less measurement infor-mation[J]. Aerospace Science and Technology, 2018, 77: 256-264

[4]　Li Y, Sun L, Qu X, et al. Acceleration measurement-based incremental nonlinear flight control for air-breathing hypersonic vehicles[J]. Aerospace Science and Technology, 2016, 58: 235-247

[5]　Sieberling S, Chu Q P, Mulder J A. Robust flight control using incremental nonlinear dynamic inversion and angular acceleration prediction[J]. Journal of guidance, control, and dynamics, 2010, 33(6): 1732-1742.

[6]　Cakiroglu C, Van Kampen E J, Chu Q P. Robust incremental nonlinear dynamic inversion control using angular accelerometer feedback[C]//2018 AIAA Guidance, Navigation, and Control Conference. 2018: 1128. Wonham W. On pole assignment in multi-input controllable linear systems[J]. IEEE transactions on automatic control, 1967, 12(6): 660-665

[7]　Wonham W. On pole assignment in multi-input controllable linear systems[J]. IEEE transactions on automatic control, 1967, 12(6): 660-665

[8]　Hu Y, Shen K, Neusypin K A, et al. Hierarchic controllability analysis in high-dynamic guidance for Autonomous Vehicle Landing[J]. IEEE Transactions on Aerospace and Electronic Systems, 2021, 58(3): 1545-1557.

[9]　Cunis T, Kolmanovsky I, Cesnik C E S. Integrating nonlinear controllability into a multidisciplinary design pro-cess[J]. Journal of Guidance, Control, and Dynamics, 2023, 46(6): 1026-1037.

[10] Müller P C, Weber H I. Analysis and optimization of certain qualities of controllability and observability for linear dynamical systems[J]. Automatica, 1972, 8(3): 237-246.

[11] Hamdan A M A, Nayfeh A H. Measures of modal controllability and observability for first-and second-order linear systems[J]. Journal of guidance, control, and dynamics, 1989, 12(3): 421-428.

[12] Mracek C, Ridgely D. Missile longitudinal autopilots: connections between optimal control and classical topolo-gies[C]//AIAA guidance, navigation, and control conference and exhibit. 2005: 6381.

[13] Lewis F L, Vrabie D, Syrmos V L. Optimal control[M]. John Wiley & Sons, 2012.

[14] Jatiningrum D, de Visser C C, van Paassen M, et al. Modelling an angular accelerometer using frequency-response measurements[C]//AIAA Guidance, Navigation, and Control Conference. 2016: 1139.

[15] Danhane B, Lohéac J, Jungers M. Contributions to output controllability for linear time varying systems[J]. IEEE Control Systems Letters, 2021, 6: 1064-1069.

[16] Hu T, Lin Z, Qiu L. An explicit description of null controllable regions of linear systems with saturating actuators[J]. Systems & control letters, 2002, 47(1): 65-78.

Optimized Multi-Robot Area Search for Clustered Target Detection in Limited Sensor Range Scenarios[15]

Jiaqing Ma[1], Yiwen Huo[1], Yang Feng[1], Jun Chen[1,2], Fei Xie[1,2], Qi Mao[1,2]

(1. School of Electrical & Automation Engineering, Nanjing Normal university, Nanjing 210023, China; 2. Jiangsu Key Laboratory of 3D Printing Equipment and Manufacturing, No.2 Xuelin Road, Nanjing, 210023, Jiangsu, China)

Abstract: Inspired by coral reef data gathering using multiple UUVs, this paper tackles distributed multi-robot area examination and target search. Robots navigate through known areas, slowing for close examination to gather precise information and detect targets. Key challenges include limited sensor range and the need to search a large area without prior target distribution knowledge, making optimal search strategy design computationally complex. We compare two target search strategies: a greedy algorithm that exploits detected clusters thoroughly, and a Thompson Sampling-based algorithm balancing exploration and exploitation. Simulations validate the efficiency of these methods, especially in clustered target scenarios, against a baseline sweeping approach.

 Key words: multi-robot systems; sensor-based planning; active information gathering;

1 Introduction

Coral reefs worldwide are threatened by destructive fishing, overfishing, coastal development, and water pollution [1]. Robots have recently been developed for coral reef monitoring, health assessment, and restoration to aid in marine ecosystem recovery [2, 3]. An essential task is for robots to explore underwater, locate damaged coral reefs by gathering high-precision images, and support recovery efforts like transplantation, coral gardening, and substrate enhancement [1]. This paper aims to develop efficient multi-robot target search and examination algorithms, applicable to scenarios like desert trash collection, building and machinery surface inspection, and vegetation sampling in nature reserves.

The task challenges are: (1) targets are static, clustered, and sparsely distributed across an unknown area; (2) the search space is much larger than the robots' field of view, necessitating extensive navigation; (3) locating targets requires thorough, time-intensive investigation. Existing research addresses aspects of target search and exploration but doesn't tackle all these challenges together or focus on optimizing search efficiency for these scenarios.

This paper introduces two multi-robot search algorithms: a brute-force greedy algorithm for locating target clusters through random search and local exploitation, and a Thompson Sampling (TS)-based algorithm that learns target distribution over time. The TS-based approach, using modified Bernoulli TS, directs robots to high-probability areas while balancing exploration and

exploitation, refining searches at cluster edges with a posterior filter. Simulations reveal that the greedy algorithm excels with highly clustered targets, whereas the TS-based method effectively balances search efforts across diverse target distributions.

2 Related Problem

This section explores two problem classes—coverage path planning and exploration—and shows that, despite their similar formulations, no existing method efficiently solves our search and examination problem.

2.1 Coverage Path Planning

Coverage path planning (CPP) methods, surveyed in [6], find collision-free paths that ensure one or multiple robots to pass over all points in a *known* environment. CPP algorithms often generate zigzag [7,8] or spiral [7,8] paths that cover the entire search space. Various task allocation approaches are proposed for a team of robots to complete CPP tasks cooperatively [11]. However, without prior information on the target distribution, robots must thoroughly investigate all locations along a coverage path, leading to inefficiency, especially when targets occupy a small area.

2.2 Exploration

Exploration involves planning paths for mobile robots to investigate all locations in an unknown environment to build a map or detect targets of interest. Authors propose planning approaches to find trajectories along which one [12] or multiple [13] robots visit frontiers, or local minima of a potential information field [14] to reduce the uncertainty of the exploration space. While these approaches suit indoor target search with limited space, some studies focus on efficiently spreading a robot team in a large search area online to minimize target search time cooperatively [15]. All the mentioned approaches focus on selecting the next goal to sweep the unknown environment while balancing task allocation among agents.

Unlike the exploration problem, our search and examination problem assumes robots operate in a known environment, where pre-assigned paths allow for more efficient cooperative coverage with minimal movement. This contrasts with exploration algorithms, which eventually achieve target search but are less efficient.

3 Problem Formulation

A number of static targets are located in a known task space $E \in R^2$. The set of all target locations, i.e., the areas of interest (AoI) in E is denoted $AoI \subset E$. We denote p the position of a point in E. A homogeneous robot team of n agents is tasked to detect all points $p \in AoI$ using perfect isotropic sensors with finite detection radius $f\,ov$. Our approach can be extended to handle heterogeneous and anisotropic sensors by applying the similar strategy from our recent work [4], where we map each sensor to an isotropic field of view with equivalent detection capability. Robots are able to localize their position $Q = q_1, \ldots, q_n$ as well as locations of targets. During traveling, a robot moves at its maximum velocity v_{max} which is too fast for examination tasks, e.g., to gather desired information or interact with the targets. Thus, the robot moves at a slower velocity v_{det} for precise examination once a target is found.

To formulate the task as searching and examining a finite number of locations cover the entire

task space, we tile E by a grid map with K identical grids. To make use of robot's sensing range, K is chosen in a way that a robot can only detect all targets in one grid when it is placed at the center of it. With a slight abuse of notations, k denotes both a grid and the index of it, and K denotes both the entire grid set and the number of total grids. By choosing an $action\ k \in K$, a robot travels to the center of grid k to conduct examination.

4   Greedy Algorithm

The brute force greedy algorithm explores the task space randomly and exploits the surrounding area of a target once found. Each robot r_i starts with an initial position q_0 and ID i, and uses a grid set K for exploration. A set V keeps track of visited grids, shared across the team. Initially, the robot's action list is empty. Upon finding a target at its current grid, the robot adds all unexamined neighboring grids to its action list (Lines 4-5). If all surrounding grids are examined, the robot adds grids from neighboring robots' action lists, starting from the nearest (Lines 6-9). After examining the current grid, the robot selects its next goal from the action list or picks a random unexamined grid if no assistance is needed (Lines 10-16).

Algorithm 2 GreedySearch (q_0, i, K, V)
1:        Initialize an action list $Si \leftarrow \emptyset$
2:        Initialize $q_i \leftarrow q_0$
3:        **while** true **do**
4:        **if** targets are found at q_i **then**
5:                  Push un-examined adjacent grids to Si
6:                      **if** Si $= \emptyset$ **then**
7:        Push Sj to Si for all $j \in Nj$ without duplication in order from nearest to farthest
8:              **end if**
9:        **end if**
10:      **if** $Si = \emptyset$ **then**
11:      Draw $k \in K$ randomly s.t. g_k & V
12:      $goal_i \leftarrow g_k$
13:            **else**
14:      $goal_i \leftarrow Si\,(1)$
15:      $Si \leftarrow Si \setminus Si\,(1)$
16:      **end if**
17:      $V \leftarrow V \cup goal_i$
18:      *Broadcast Si and V to neighbors*
19:      *Travel to and examine $goal_i$*
20:      **end while**

This greedy approach ensures that robots thoroughly explore a target cluster before moving on, assuming targets are clustered together. However, when multiple clusters exist, this method can be inefficient. To address this, a probabilistic TS-based algorithm is introduced in the next section, aiming to balance efficient search for both tightly and loosely clustered targets.

5   Thompson Sampling Algorithm

5.1 Bernoulli Bandit Modeling

The outcome of an action k is denoted as a binary value based on the detection Z received at that location,

$$win_k = \begin{cases} 0, & Z=0 \\ 1, & else, \end{cases} \tag{1}$$

i.e., the outcome is 0 if a robot detects nothing in that grid and 1 otherwise. For each grid $k \in K$, a beta function $beta(\alpha_k, \beta_k)$ is updated by a reward r_k obtained through a reward policy introduced in Section 5.3. Initially, we set $\alpha_k = \beta_k = 1$, meaning that targets are considered with identical probability to appear at any locations in E. Note that outcomes are not immediately received after selecting an action. Instead, there is a time delay between selecting an action and receiving an outcome due to the travel and detecting time it takes.

5.2 Goal Selection

At each step, the robot selects the next grid based on target distribution beliefs and travel time cost to optimize search efficiency. It is rewarded for targeting high-probability grids and minimizing travel time when targets are nearby, promoting local search while deferring distant areas. If no targets are found, the robot relies more on distribution predictions, potentially incurring longer travel times to explore new areas.

The robot is assumed to have repeatedly found targets in the last $c_f \in N$ selected grids, or have repeatedly missed targets in the last $c_m \in N$ selected grids. Once a goal is decided, we define the "travel-reducing" reward obtained by reducing travel time from its current location q to any $p \in E$ by a non-negative function:

$$H(q,p) = \begin{cases} (1-norm) \cdot c_f, & c_f > 0 \\ (1-norm) / c_m, & else, \end{cases} \tag{2}$$

$$norm = \frac{\|q-p\|}{d_{max}} \tag{3}$$

where d_{max} is the distance between two furthest points in E, and $\| \cdot \|$ denotes the Euclidean norm. Note that the Euclidean distance between goal location and robot location can be replaced with length of trajectories if E is not a convex open space and motion planning algorithms are applied. We normalize the distance between q and p by E's dimension in norm. The robot can always be better rewarded by selecting a closer grid to detect, as a smaller norm always leads to a greater H. Then we select an action $k*$ such that

$$u = c \cdot \theta_k + (1-c) \cdot H(q, p_k), \tag{4}$$

$$k* \leftarrow \arg_k \max u, \tag{5}$$

where $\theta_k \sim beta(\alpha_k, \beta_k)$, and $c \in [0,1]$ is a tuning weight. If the robot repeatedly achieves positive outcomes, it gains increasing rewards for reducing travel distances, encouraging exploitation of nearby areas despite the target distribution. Conversely, if the robot fails to find

targets, the reward for minimizing travel time decreases, and the reward for selecting grids with higher target density dominates.

Algorithm 3 GoalSelection (K, α, β, S)

1:      **for** $k = 1, \ldots, K$ **do**

2:          **if** $k \in S$ **then**

3:          Sample $\theta_k \sim \boldsymbol{beta}(\alpha_k, \beta_k)$

4:          **end if**

5:      **end for**

6:      Select $k*$ using Equation 5

7:      $g \leftarrow p_k$

8:      $S \leftarrow S \setminus k*$

9:      Return g, S

5.3   Recursive Target Search

The target search algorithm, outlined in Algorithm 4, allows a robot to search targets online recursively overtime. The algorithm begins with initializing a grid map and beta functions for each grid, outlined in lines 1-2. The temporary goal g is set to the robot's initial position q_0, and the set of all unvisited grids S is initialized with K, as shown in line 3. Lines 4-15 demonstrate the main search process, including a posterior update step and a posterior filtering step, and the goal selection process. Robot terminates when desired task is completed, such as having searched all grids or running out of time.

Algorithm 4 TargetSearch (q_0, K)

1:      Draw K grids over E

2:      Initialize α(1, 1), β(1, 1) for all $\alpha_k \in$ α, $\beta_k \in$ β

3:      Initialize $g \leftarrow q_0$, $S \leftarrow 1, 2, ..., K$

4:      **while** true **do**

5:          $r \leftarrow 0$

6:          Move to g and search

7:          Set r_k and all members in $N(r_k)$ to r_{const}

8:          **if** $win_k = 1$ **then**

9:          α $\leftarrow$ α + r

10:         **else**

11:         β $\leftarrow$ β + r

12:         **end if**

13:         Implement posterior filtering and update β

14:         GoalSelection(K, α, β, S)

15:      **end while**

5.3.1   Posterior Update

To recursively select optimized goals online, a robot must update the target posterior after

getting an outcome at g. In standard MAB problems, a resource's posterior is updated to guide its selection frequency. However, in target search, the same grid isn't explored repeatedly. To leverage search outcomes, the posteriors of other correlated grids must also be updated, in addition to the sampled grid. Thus, a K-dimensional reward set $r = r_1, r_2, ..., r_k$ is created and initialized with zeros at each time step, after which we choose to set not only k's reward, but also the rewards of k's eight neighboring grids $N(r_k)$ to r_{const}, as outlined in lines 5-7 in Algorithm 4. The reward set r is then utilized to update the beta functions following the reward policy outlined in lines 8-12. The reward policy augments α_s of a grid and its neighbors with r_{const} if the outcome is positive, and updates β_s similarly for a zero outcome. In Section 6, we show that this strategy outperforms the baseline sweeping algorithm when targets are clustered, as the neighborhood's target density resembles that of a centered grid, drawing robots to search around detected targets.

5.3.2 Posterior Filtering

While a robot benefits from exploring areas near known targets, it may misestimate target density at cluster edges, avoiding these areas if no targets are detected, which slows target discovery at the boundaries as the search space expands. An example is shown in Fig. 1. After searching for targets distributed as Fig.1 a for a period of time in a 20×20 grids squared search space, while grids with targets inside are with higher α values (Fig.1 b), not all of them are with lower β values (Fig.1 c). The marginal areas of target clusters, marked by translucent green in Fig.1 c, have relative high values, resisting being selected to search by the robot.

| Fig.1 d | Fig.1 d | Fig.1 d | Fig.1 d |
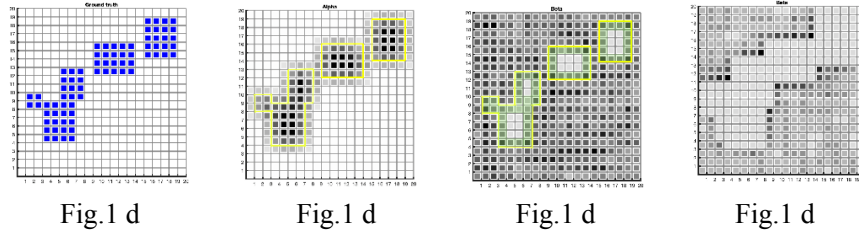
Fig. 1 The figures illustrate the posterior filtering process on a 20×20 grid map. Fig.1 d shows the true target distribution, with targets in blue grids. Fig.1 d displays the α value per grid, where darker grids indicate higher α values; yellow boundaries outline true target clusters. Fig.1 d and Fig.1 d show α values before and after filtering, respectively, with darker grids indicating higher β values. In Fig.1 d, yellow boundaries denote target clusters, and green translucent areas highlight cluster edge grids.

Therefore, before selecting goals using Algorithm 3, we increment the posterior update with a simple filtering process to improve robot's performance in searching cluster edges, as outlined in line 13 in Algorithm 4. We extract a set $A \subset K$ of all grid k_{in} which $\alpha_k > 1$. In other words, A is the set of all grids that targets are found inside or close to them so far. Then we reset $\{\forall \beta_k = 1 | k \in A\}$, indicating that they will not be disregarded even the robot has detected nothing around it. The β map after filtering Fig.1 c is shown as Fig.1 d.

6 Simulation Results

We conduct MATLAB simulations to validate our algorithms through batch tests, comparing them quantitatively with the baseline method. An ablation study demonstrates the effectiveness of posterior update and filtering for the TS-based method. For simplicity, robot displacement follows

first-order dynamics, and rotation is time-free. We choose $r_{const} = 10$ and c = 0.5. Search space is square with 1 m × 1 m grid size based on robot's field of view.

6.1 Quantitative Comparisons

To test whether a team of robot can efficiently examine an area and find targets, eight holonomic robots equipped with an isotropic sensor are tasked to search a 40 m × 40 m square space with striped obstacles. The robot moves at $v_{det} = 0.03$ m/s in order to effectively examine the grids. We test four different target aggregation levels from 1 to 4 depending on the ratio of sensor field of view to the average area of clusters, where level 1 is equivalent to uniform distribution. An example is shown as Fig. 2. For each aggregation level, target distribution is generated randomly in each trial via arbitrarily placing identical numbers and sizes of clusters inside the search space, with possible overlaps.

| level 4 | level 4 | level 4 | level 4 |
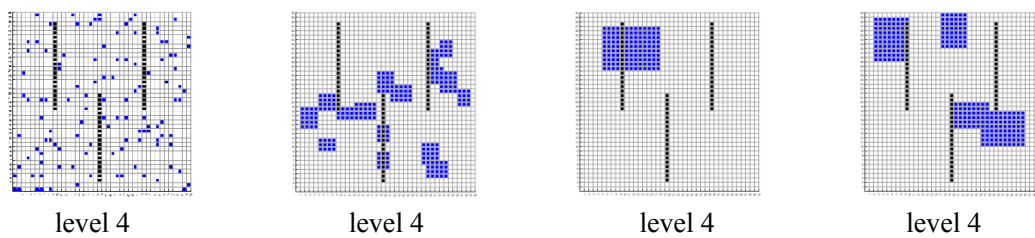
Fig. 2 The figures display four target aggregation levels in a 40 m × 40 m test environment, from level 1 to 4. Black grids represent obstacles that robots cannot cross, while targets are located in blue grids.

We compare our greedy and TS-based algorithms with a baseline sweeping algorithm, which divides the environment into eight rectangular areas and directs robots to sweep each in a zigzag pattern, ensuring efficient coverage of all grids. We first run four batches of tests of four aggregation levels with $v_{max} = 0.3$ m/s respectively, each with 100 trials. For each trial, robots depart from the lower-left corner of the environment and the task is considered finished when all targets are found and examined. As shown in Fig.3 a-Fig.3 d, the sweeping algorithm performs consistently, completing the task in about 3000s for all target distributions, and outperforms when targets are uniformly distributed, as robots must sequentially examine most grids. In contrast, the greedy algorithm performs poorly as aggregation levels change, only outperforming when targets are highly aggregated. The TS-based algorithm, however, shows significant efficiency gains over the sweeping algorithm for targets ranging from slightly to highly clustered, due to its ability to balance exploration and exploitation using past examination data.

Rather than following a shortest coverage path, both our algorithms actively select the next grid based on past outcomes, resulting in longer trajectories. As a result, their efficiency in searching clustered targets decreases when the travel time between grids exceeds the time spent examining them. Fig.3 c, Fig.3 e and Fig.3 f demonstrate this feature and indicate that the more v_{max} greater than v_{det}, the more significant of both greedy and TS-based algorithms.

| Fig.3 f | Fig.3 f | Fig.3 f | Fig.3 f | Fig.3 f | Fig.3 f |
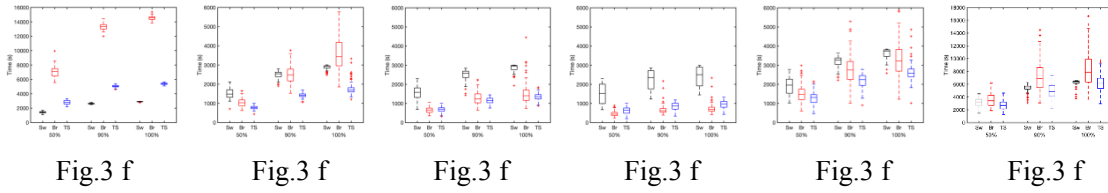
Fig. 3 The boxplots present simulation results for 100 trials of the sweeping (Sw), brute force greedy (Br), and TS-based (TS) algorithms across varying target aggregation levels and robot speeds v_{max}. Fig.3 f–Fig.3 f show the time required to locate and examine 50%, 90%, and 100% of targets at different aggregation levels with $v_{max} = 0.3\ m/s$. Fig.3 f and Fig.3 f display the time to examine all targets at aggregation level 3 with $v_{max} = 0.1\ m/s$ and $0.03\ m/s$, respectively.

6.2 Ablation Study for TS-Based Method

We further validate the efficacy of both posterior update and posterior filtering using TS-based method through a series of tests. In the following tests, a robot is tasked to search a 50 m×50 m square open space with an isotropic sensor. The robot moves at $v_{max} = 1\ m/s$ and $v_{det} = 0.2$ m/s. To randomize target distribution, we select *n* target clusters from 7 different sizes and shapes shown in Fig. 4 and randomly place them inside each tested search space, with possible overlaps.

Cluster	Width	Height
1	1/4	1/3
2	1/3	1/4
3	1/5	1/4
4	1/4	1/5
5	1/7	1/6
6	1/6	1/7
7	1/10	1/10
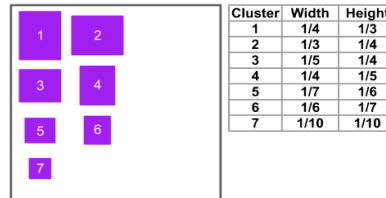
Fig. 4 The figure shows seven blue-numbered target clusters of varying sizes, randomizing target distributions within a black square search space. Cluster dimensions, listed in the legend, are based on the search space's unit side length, rounded to the nearest grid number.

6.2.1 Wasserstein-1 Distance

At each time step, we use Wasserstain-1 distance (W1D) to quantify the error between the distribution of searched locations and the distribution of true target locations. Wasserstein-1 distance is also named the earth-mover distance due to its connection to the optimal transport problem [5]. We denote $P(M)$ the space of probability measures defined on a metric set M. For two probability measures μ, $v \in P(M)$, we define the W1D:

$$W(\mu,\ v) = \inf{}_{\gamma \in \Gamma(\mu, v)} \int \|x - y\| d\gamma(x, y), \tag{6}$$

where $\Gamma(\mu, v)$ is the set of all joint distributions $\gamma(x, y)$ whose marginals are μ and v, respectively. W1D calculates the minimum distance for an earth mover to transport one pile of a unit amount of earth into another, and $\gamma(x, y)$ quantifies the "mass" of earth to be transported from *x* to *y* in order to transform the distributions μ into v.

W1D between the searched and true target sets reflects the robot's effectiveness in goal selection. A lower W1D indicates the robot focuses on areas with higher target density, avoiding wasteful searches in low-density regions. A faster decrease in W1D shows stronger optimization in goal selection.

6.2.2 Results

We run 30 trials, for each of which we generate a target distribution using cluster 1-6, one for each cluster. Three methods are compared, namely the sweeping algorithm, the TS method without posterior filtering, and the TS method with posterior filtering, with the same 30 target distributions generated randomly. In Fig. 5, we plot the W1Ds for 30 trial using each method along with the median search time used in 30 trials to detect 50%, 90%, and 100% of *AoI*. The results show that TS-based methods outperform the sweeping strategy, with or without posterior filtering, as they reduce the time to detect the *AoI*. In Fig.5 c, the zigzag pattern from the sweeping algorithm creates uniformly descending wavy curves. TS-based methods, however, lead to faster W1D drops, highlighting their ability to optimize goal selection online.
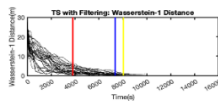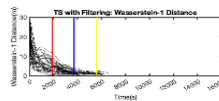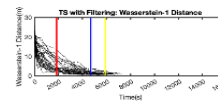
Fig.5 c	Fig.5 c	Fig.5 c

Fig. 5 The results of a batch of trials compare three methods: the sweeping algorithm (Fig.5 c), TS without posterior filtering (Fig.5 c), and TS with posterior filtering (Fig.5 c). Black curves represent 30 W1Ds over time for each method, while red, blue, and yellow lines indicate the median time of 30 trials to reach 50%, 90%, and 100% AoI detection, respectively.

Moreover, the results show that applying posterior filtering significantly reduces search time in locating the final parts of the area of interest (*AoI*), though initial search speed is slower. Early on, limited sampling hinders precise cluster edge detection, but as samples accumulate, filtering quickly aids in finding edges. In contrast, without filtering, W1D exhibits a prolonged, inefficient search for the last targets after an initial rapid descent.

7 Conclusions

This paper addresses multi-robot target search in large areas, where robots locate targets likely distributed in random clusters without prior knowledge. Unlike the greedy approach, the TS-based algorithm iteratively selects search locations using target posteriors and travel time penalties, enabling robots to optimize goal selection based on learned patterns. Simulations show that both algorithms outperform the sweeping method. Future work will extend these methods to 3D environments and distributed systems, with applications in underwater reef monitoring and building facade inspection.

References:

[1] E. Bayraktarov, P. J. Stewart-Sinclair, S. Brisbane, L. BostrÖmEinarsson, M. I. Saunders, C. E. Lovelock, H. P. Possingham, P. J. Mumby, K. A. Wilson, Motivations, success, and cost of coral reef restoration, Restoration Ecology 27 (5) (2019) 981– 991.

[2] T. Manderson, J. Li, N. Dudek, D. Meger, G. Dudek, Robotic coral reef health assessment using automated image analysis, Journal of Field Robotics 34 (1) (2017) 170–187.

[3] Galceran, E., & Carreras, M., "A survey on coverage path planning for robotics," *Robotics and Autonomous Systems*, 61(12), 2013.

[4] J. Chen, P. Dames, Distributed multi-target tracking for heterogeneous mobile sensing networks with limited field of views, in:2021 ieee international conference on robotics and automation (icra),IEEE, 2021, pp. 9058–9064.

[5] F. Santambrogio, Optimal transport for applied mathematicians, Birk user, NY 55 (58-63) (2015) 94.

[6] E. Galceran, M. Carreras, A survey on coverage path planning for robotics, Robotics and Autonomous systems 61 (12) (2013) 1258–1276.

[7] R. B hnemann, N. Lawrance, J. J. Chung, M. Pantic, R. Siegwart,J. Nieto, Revisiting boustrophedon coverage path planning as a generalized traveling salesman problem, in: Field and ser- vice robotics, Springer, 2021, pp. 277–290.

[8] S. Bochkarev, S. L. Smith, On minimizing turns in robot coverage path planning, in: 2016 IEEE international conference on automation science and engineering (CASE), IEEE, 2016, pp. 1237–1242.

[9] T. M. Cabreira, C. Di Franco, P. R. Ferreira, G. C. Buttazzo, Energy-aware spiral coverage path planning for uav photogram- metric applications, IEEE Robotics and automation letters 3 (4) (2018) 3662–3668.

[10] C. Wu, C. Dai, X. Gong, Y.-J. Liu, J. Wang, X. D. Gu, C. C.Wang, Energy-efficient coverage path planning for general terrain surfaces, IEEE Robotics and Automation Letters 4 (3)(2019) 2584–2591.

[11] J. Tang, C. Sun, X. Zhang, Mstc*: Multi-robot coverage path planning under physical constrain, in: 2021 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2021, pp. 2518–2524.

[12] A. Sipahioglu, G. Kirlik, O. Parlaktuna, A. Yazici, Energy con- strained multi-robot   sensor-based coverage path planning us- ing capacitated arc routing approach, Robotics and Autonomous Systems 58 (5) (2010) 529–538.

[13] K. M. Wurm,   C.   Stachniss, W. Burgard,   Coordinated   multi- robot exploration using a segmentation of the environment, in:2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2008, pp. 1160–1165.

[14] J. Vallv , J. Andrade-Cetto, Potential information fields for mo- bile robot exploration, Robotics and Autonomous Systems 69 (2015) 68–79.

[15] L. Heng,   A. Gotovos,   A. Krause, M. Pollefeys,   Efficient vi- sual exploration and coverage with a micro aerial vehicle in un- known environments, in: 2015 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2015, pp.   1071–1078.

A Tightly-Coupled Inertial-LiDAR-Barometer SLAM for Indoor Degraded Scenes

Wu Jialiang[1], Lyu Pin[1], Lai Jizhou[1], Fang Wei[1], Fu Lin[1]

(1. College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: Large indoor scenes are usually similar in structure in the elevation direction, which leads to degradation of the lidar-scanned point cloud in the elevation direction. The traditional UAV localization method of lidar SLAM is prone to mismatching of elevation features. To address this problem, this paper proposes a tightly-coupled inertial-LiDAR-barometer SLAM: the barometer and inertial sensor are introduced into the point cloud initialization process to improve the initial pose matching accuracy; the point cloud registration algorithm based on multivariate normal distribution is decoupled in the horizontal and altitude channels, and the direction of the point cloud registration is constrained to improve the positioning accuracy in environments with degraded elevation. A cabin simulation scene was constructed using the Gazebo simulation platform, and field flight tests were conducted in an indoor environment to verify the proposed method. The results showed that the method in this paper can improve the positioning accuracy of lidar SLAM under degraded elevation features and effectively improve computational efficiency.

Key words: LiDAR SLAM; UAV; degraded scenes; integrated navigation

1 Introduction

In recent years, small UAVs have been widely used in many fields such as inspection, detection, search and rescue by virtue of their low cost and high flexibility. With the development of UAV-related technologies, its application has gradually expanded from outdoor scenarios to indoor scenarios. Since there is no satellite signal indoors, its necessary to use active sensors such as LiDAR for localization [1]. LiDAR SLAM determines the body position by finding the relative position relationship between two frames of point cloud, and the saliency of the surrounding scene features will have a direct impact on the positioning accuracy. In large indoor scenes, the surrounding area is usually composed of vertical walls that have no obvious structural features, which places higher demands on SLAM technology.

LiDAR SLAM algorithms in feature degradation environments are a current research hotspot and usually require the introduction of other sensors for information assistance [2-4]. LiDAR-based combined navigation schemes can be categorized into loosely-coupled algorithms and tightly-coupled algorithms [5-8]. Loosely-coupled algorithm performs navigation computation separately for LiDAR and other sensors, and then fuses their individual navigation results. However, in Loosely-coupled algorithm, each sensor is relatively independent, and although the final navigation result can realize the complementary advantages of the sensors, the low fusion degree of the position information between the sensors does not actually improves the independent navigation

computation accuracy of LiDAR. In comparison with this, tightly-coupled algorithms utilize the redundancy of sensor information to realize the correction of the original point cloud data of LiDAR, which can effectively improve the navigation performance of LiDAR.

Currently, there is literature on the autonomous localization of robots in corridors, pipelines, and other longitudinal feature degradation environments, but there is no effective solution to solve the elevation feature degradation problem in large-scale indoor scenarios.

In response to this, a tightly-coupled inertial-LiDAR-barometer SLAM for indoor degraded scenes is proposed in this paper. Height compensation is used to provide a better initial pose for 3D LiDAR point cloud registration. The problem of point cloud over-matching is solved, and the positioning accuracy and computational efficiency of LiDAR SLAM in the degraded environment in the vertical direction are improved.

2   Design of an improved algorithm architecture for 3D LiDAR-based SLAM for UAVs

In this paper, we propose a tightly-coupled inertial-LiDAR-barometer SLAM. The framework is shown in Fig.1. First, for the problem of insufficient compensation of the initial pose of the point cloud in the sparse feature environment of traditional algorithms, the tight combination of multi-sensor information assisted by inertial/barometer information is realized to improve the perception ability of the body to the environment and fully compensate for the initial pose of the point cloud. Then, considering the problem of point cloud over-matching, we propose to decouple the NDT pose solution for the pre-processed point cloud and restrict the direction of the pose solution after the point cloud registration, so that the algorithm only registers according to the actual overlapping part of the two frames of point cloud. Finally, a graph optimization-based [9] SLAM framework is used to optimize the UAV trajectory in real time and build a map.
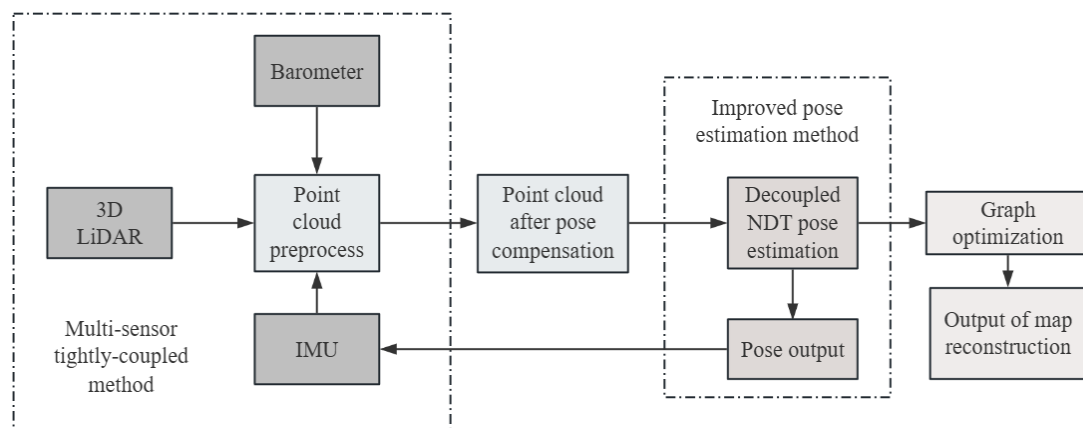
Fig.1 Framework of the improved LiDAR-based localization method

2.1   An improved SLAM algorithm based on decoupled point cloud registration

In this paper, an optimization method for point cloud registration based on pose decoupling is proposed. This method reduces the six-degree-of-freedom pose solution of traditional laser SLAM to three degrees of freedom, separates the height channel from the horizontal position channel, limits the solution space of point cloud registration, and uses height assistance to achieve effective

registration of the actual overlapping parts of the two frames of point clouds. The algorithm architecture is shown in Fig.2.
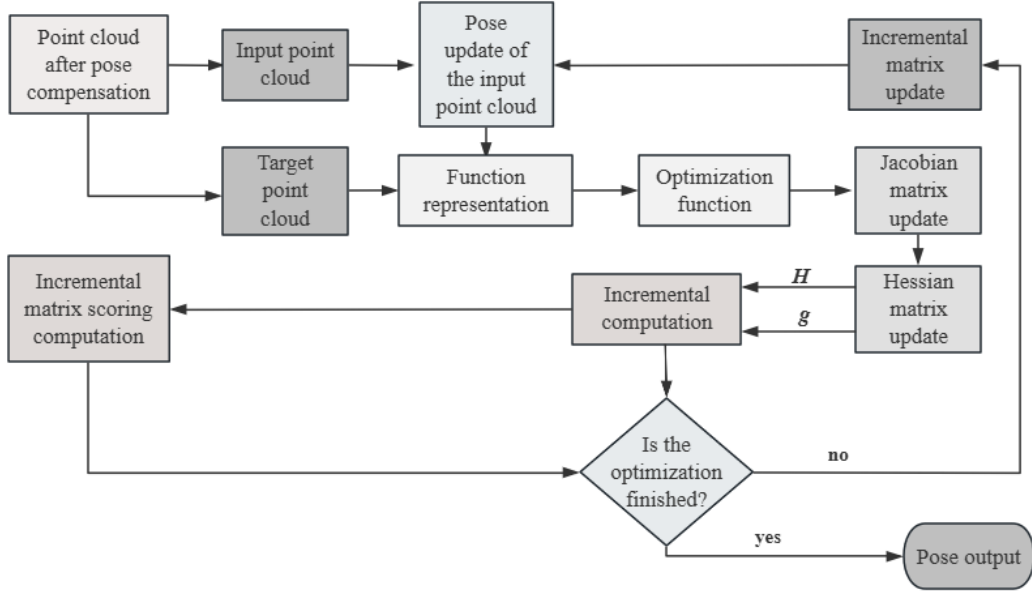
Fig.2 Process flow diagram of the point cloud iterative registration algorithm

2.1.1 An optimization algorithm for point cloud registration based on pose decoupling

The space is discretized into a grid. Based on the point cloud distribution and coordinate data within the spatial grid, the keyframe point cloud is represented using a multivariate normal distribution. Therefore, the spatial point cloud of the keyframe can be represented as:

$$\tilde{S}_c\left(p_{in,i}^b\right) =$$
$$d_1 \exp\left(-\frac{d_2}{2}\left(p_{in,i}^b - \mu_k\right)^T \Sigma_k^{-1}\left(p_{in,i}^b - \mu_k\right)\right) \tag{1}$$

where μ_k is the mean of the point cloud within the cube; Σ_k is the covariance matrix of the point cloud within the cube; $p_{in,i}^b$ is the i-th point in the input point cloud; $\tilde{S}_c\left(p_{in,i}^b\right)$ is the probability value of this point in its corresponding multivariate normal distribution function; d_1 and d_2 are constants.

The optimization function can be constructed:

$$s\left(\Delta P_{k-1,k}^F\right) = \sum_{k=1}^{n} \tilde{S}_c\left(T\left(0, \Delta P_{k-1,k}^F\right) p_{in,k}^{b,init}\right) \tag{2}$$

where $s\left(\Delta P_{k-1,k}^F\right)$ is the total probability value of the input point cloud calculated based on the transformation matrix $T\left(0, \Delta P_{k-1,k}^F\right)$. $P_{k-1,k}^F$ is the change in the UAV's internal pose from frame *k-1* to *k*:

$$\Delta P_{k-1,k}^F = P_k^F - P_{k-1}^F =$$
$$\left[\Delta t x_{b,k-1}^w \quad \Delta t y_{b,k-1}^w \quad \Delta \gamma_{b,k-1}^w\right] \tag{3}$$

The goal is to solve for the change in the internal posture of the UAV corresponding to the maximum value obtained in Eq. (2). The solution method uses the Gauss-Newton iteration method.

The simplified solution formula is as follows:

$$H \Delta P_{k-1,k}^{\mathrm{F}} = -g \tag{4}$$

where H and g are the Hessian matrix and Jacobian matrix derived from the optimization function.

Let:

$$T_{\mathrm{E}} = T\left(\mathbf{0}, \Delta P_{k-1,k,i}^{\mathrm{F,cal}}\right) p_{\mathrm{in},k}^{\mathrm{b,init}} = \begin{bmatrix} x_{k,i}^{\mathrm{b}} & y_{k,i}^{\mathrm{b}} & z_{k,i}^{\mathrm{b}} \end{bmatrix} \tag{5}$$

where T_{E} represents the LiDAR point cloud coordinates of the robot after the i-th iteration of the point cloud at frame k during the pose solution process; $\Delta P_{k-1,k,i}^{\mathrm{F,cal}}$ represents the change in the pose vector of the robot body after the i-th iteration relative to the pose vector of the robot body corresponding to the $k-1$ frame point cloud. Define $\hat{p}_i$ as the i-th element of $\Delta P_{k-1,k}^{\mathrm{F}}$, then:

$$g_i = \frac{\delta s}{\delta \hat{p}_i} = \sum_{k=1}^{n} \frac{\delta \tilde{S}_{\mathrm{c}}(T_{\mathrm{E}})}{\delta T_{\mathrm{E}}} \frac{\delta T_{\mathrm{E}}}{\delta \hat{p}_i} \tag{6}$$

$$H_{ij} = \frac{\delta^2 s}{\delta \hat{p}_i \delta \hat{p}_j} =$$
$$\sum_{k=1}^{n} \left(\frac{\delta T_{\mathrm{E}}^{\mathrm{T}}}{\delta \hat{p}_j} \frac{\delta^2 \tilde{S}_{\mathrm{c}}(T_{\mathrm{E}})}{\delta T_E \delta T_{\mathrm{E}}^{\mathrm{T}}} \frac{\delta T_{\mathrm{E}}}{\delta \hat{p}_i} + \frac{\delta \tilde{S}_{\mathrm{c}}(T_{\mathrm{E}})}{\delta T_{\mathrm{E}}} \frac{\delta^2 T_{\mathrm{E}}}{\delta \hat{p}_i \delta \hat{p}_j} \right) \tag{7}$$

where the Jacobian and Hessian matrices of T_{E} are:

$$J_{\mathrm{E}} = \begin{bmatrix} \dfrac{\delta T_{\mathrm{E}}}{\delta \hat{p}_1} & \dfrac{\delta T_{\mathrm{E}}}{\delta \hat{p}_2} & \dfrac{\delta T_{\mathrm{E}}}{\delta \hat{p}_3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & 0 \end{bmatrix} \tag{8}$$

$$H_{\mathrm{E}} = \frac{\delta^2 T_{\mathrm{E}}}{\delta \hat{p}_i \delta \hat{p}_j} = \begin{cases} \mathbf{0}_{3\times 1} & \text{otherwise} \\ c_{3\times 1} & i = j = 3 \end{cases} \tag{9}$$

where:

$$a = -x_k^{\mathrm{b}} \sin \hat{p}_3 - y_k^{\mathrm{b}} \cos \hat{p}_3 \tag{10}$$
$$b = x_k^{\mathrm{b}} \cos \hat{p}_3 - y_k^{\mathrm{b}} \sin \hat{p}_3 \tag{11}$$

$$c = \begin{bmatrix} -x_k^{\mathrm{b}} \cos \hat{p}_3 + y_k^{\mathrm{b}} \sin \hat{p}_3 \\ -x_k^{\mathrm{b}} \sin \hat{p}_3 - y_k^{\mathrm{b}} \cos \hat{p}_3 \\ 0 \end{bmatrix} \tag{12}$$

2.2 Back-end optimization and graph construction

The algorithm back-end optimizes the keyframe poses in the lidar odometry using a graph optimization algorithm. Finally, the 3D LiDAR keyframe point clouds in the body frame are projected into the global coordinate system to construct the map using the following equation:

$$p_{i,k}^{\mathrm{n}} = T\left(P_i^{\mathrm{key}}\right) \cdot p_{i,k}^{\mathrm{b}} \tag{13}$$

where $p_{i,k}^{\mathrm{b}}$ is the coordinate of the k-th 3D LiDAR point cloud in the machine system in the i-th key frame. P_i^{key} is the pose of the i-th key frame. $p_{i,k}^{\mathrm{n}}$ is the 3D LiDAR point cloud

coordinate of $p_{i,k}^{b}$ after being projected into the global coordinate system using the transformation matrix $T\left(P_i^{key}\right)$.

3 Indoor scene UAV experiments and analysis

To verify the effectiveness of the algorithm in this paper, we conducted experiments in an underground garage scenario with actual UAV flight tests (Fig.3). We choose a tanker hold as the test scenario, which is shown in Fig.4. During flight, the drone uses the algorithm in this paper to perform real-time positioning and mapping, and transmit navigation information to the flight control.
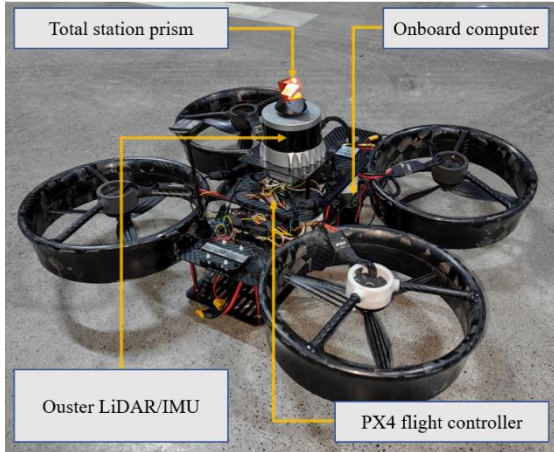
Fig.3 UAV LiDAR SLAM flight test platform

Fig.4 UAV flight test scenario inside the tanker hold.

The comparison and analysis of the test results are shown in Fig.5, Fig.6, Fig.7 and Tab.1.
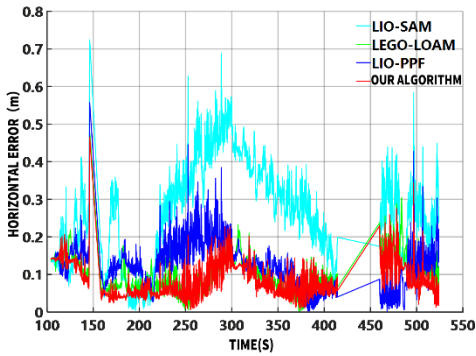
Fig.5 Comparison of horizontal errors in different algorithms
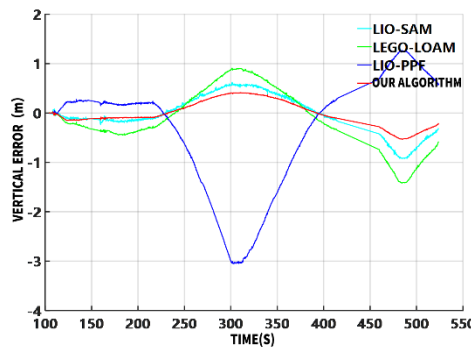
Fig.6 Comparison of vertical errors in different algorithms
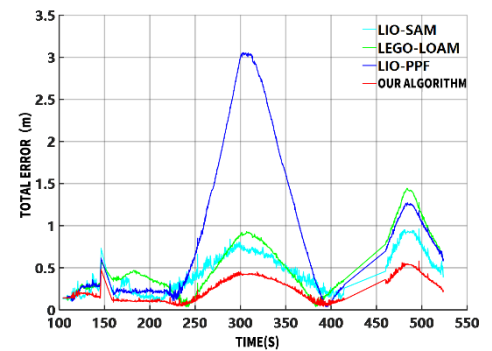
Fig.7 Comparison of total errors in different algorithms

Tab.1 Parameters Comparison of localization accuracy in different algorithms

RMSE	LIO-SAM	LEGO-LOAM	LIO-PPF	OUR ALGORITHM
Horizontal error(m)	0.2988	0.1122	0.1385	**0.0956**
Vertical error (m)	0.3946	0.6167	1.2891	**0.2599**
Total error (m)	0.4949	0.6269	1.2965	**0.2769**

Compared with the LIO-SAM, LEGO-LOAM, and LIO-PPF algorithms, the horizontal localization accuracy of the algorithm in this paper is improved by 68.0%, 14.8%, and 31.0%, respectively. The vertical localization accuracy is improved by 34.1%, 57.9%, and 79.8%, respectively. The total localization accuracy is improved by 44.0%, 55.8%, and 78.6%, respectively, which indicate that the algorithm in this paper has a higher localization accuracy.

4　Conclusion

The paper addresses the issue of autonomous navigation for small-scale flying vehicles in degraded environmental conditions. A decoupled three-dimensional LiDAR SLAM method assisted by multi-source information is proposed. In this method, the calculation and update of the initial point cloud compensation matrix and the point cloud iterative registration algorithm are optimized and improved. Through theoretical analysis and experimental verification, the following conclusions can be drawn:

1)By introducing external altitude and attitude information into the 3D LiDAR SLAM solving process, the decoupled matching of SLAM algorithm in horizontal position and heading channel is realized, which solves the problem of decreased positioning accuracy caused by sparse elevation features.

2) Compared with the traditional SLAM method, the proposed decoupling algorithm reduces the matching dimension from six dimensions to three dimensions, reduces the computational complexity of point cloud matching, and improves the computational efficiency.

References:

[1]　Khan M U, Zaidi S A A, Ishtiaq A, et al. A comparative survey of lidar-slam and lidar based sensor technologies[C]//2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC). IEEE, 2021: 1-8.7.

[2]　KAISER, S.; LANG, C. Integrating Moving Platforms in a SLAM Algorithm for Pedestrian Navigation[J]. Sensors, 2018, 18(12), 4367.

[3]　Gyagenda N, Hatilima J V, Roth H, et al. A review of GNSS-independent UAV navigation techniques[J]. Robotics and Autonomous Systems, 2022, 152: 104069.

[4]　Tiemann J, Ramsey A, Wietfeld C. Enhanced UAV indoor navigation through SLAM-augmented UWB localization[C]//2018 IEEE international conference on communications workshops (ICC workshops). IEEE, 2018: 1-6.

[5]　Xu X, Zhang L, Yang J, et al. A review of multi-sensor fusion slam systems based on 3D LIDAR[J]. Remote Sensing, 2022, 14(12): 2835.

[6]　Shan T, Englot B, Meyers D, et al. Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping[C]//2020 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, 2020: 5135-5142.

[7]　Xu W, Zhang F. Fast-lio: A fast, robust lidar-inertial odometry package by tightly-coupled iterated kalman filter[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 3317-3324.

[8]　Shan T, Englot B, Ratti C, et al. Lvi-sam: Tightly-coupled lidar-visual-inertial odometry via smoothing and mapping[C]//2021 IEEE international conference on robotics and automation (ICRA). IEEE, 2021: 5692-5698.

[9]　Chen S, Zhou B, Jiang C, et al. A lidar/visual slam backend with loop closure detection and graph optimization[J]. Remote Sensing, 2021, 13(14): 2720.

Management of Patient Rehabilitation in Medical Cyberphysical Systems

D.G. Arseniev[1,2], A.E.Misnik[1,2], M.A.Shalukhova [1,2]

(1. Peter the Great St.Petersburg Polytechnic University, St. Petersburg; 2. Inter-state Educational Institution of Higher Education "Belarusian-Russian University", Mogilev)

Abstract: The method of patient rehabilitation management in medical cyberphysical systems based on the use of artificial intelligence is considered. To create a digital model of the musculoskeletal system, computer vision technologies and trajectory analysis of the obtained points are used. By individualising recommendations, the method contributes to improving the effectiveness of patient treatment.

1 Introduction

Medical cyberphysical systems represent a distinct category of cyberphysical systems that integrate physical components and computing technologies with the objective of enhancing the quality of medical care. The effective management of processes in modern medical cyberphysical systems, including patient rehabilitation, necessitates the integration of advanced technologies and methods, as well as an individualised approach to each patient. Among the key challenges associated with the creation and implementation of classical control systems are the complexity of integration into existing workflows, the necessity for subsequent technical support and knowledge updates, and the limited applicability of systems based on statistical data processing [1-3].

The objective of this paper is to present a method for controlling the process of patient rehabilitation in medical cyber-physical systems. The method involves creating a digital model of the musculoskeletal system, which will facilitate more accurate predictions regarding the course of rehabilitation. The proposed approach enhances the availability and efficiency of treatment while reducing the burden on medical personnel.

2 The management of the rehabilitation process in medical cyberphysical systems

In collaboration with the Federal State Budgetary Institution 'Federal Centre for Traumatology, Orthopaedics and Endoprosthesis' of the Ministry of Health of the Russian Federation (Smolensk), which specialises in the provision of high-tech medical care to patients with diseases of the musculoskeletal system, a cyberphysical system for the rehabilitation of patients after joint endoprosthesis has been developed. The ontological approach to system development was implemented using the ontological engineering module of the software and tool environment, which is based on a meta-associative graph including process components in the form of methods and events. The developed system assumes the use of mobile devices equipped with a camera (tablet or mobile phone) to obtain video footage and their interaction with the server module of the system responsible for analysing the obtained data.

The proposed approach to rehabilitation management is based on the use of technical and software tools for the creation and subsequent analysis of a digital model of the human musculoskeletal system. The application of complex assessment, combining qualitative and quantitative methods, allows to comprehensively study the process of changing the state of the musculoskeletal system and to ensure the correct functioning of the developed system. The applied method of functional diagnostics includes the consideration of kinematic parameters, which allow us to determine the characteristics and form of movements, to track and record data on the compliance of the position of the body's support points during the performance of restorative exercises [3-7].

Obtaining the coordinates of the initial points of the joint position in space is carried out with the help of frame-by-frame analysis of the video sequence of the test movement: to determine the angle, the method of functional diagnostics allocates reference points, then, depending on the tested node, the connection vectors are determined. As a result, the obtained measurement information can be considered as the position of the reference points of the human body, the angles of bending of its joints, obtained after processing the video sequence. Each extracted frame is passed through a pre-trained EfficientNet convolutional neural network that extracts features for each frame, which are numerical representations of important aspects of the image characterising physical features of objects. The features are aggregated across all frames to form a digital model of the patient's musculoskeletal system, which is stored in the database of the intelligent decision support system.

The data obtained on the patient's musculoskeletal condition are combined with the patient's profile information for further processing and identification. This is achieved by employing a recurrent LSTM model, which is used to discern patterns between physical features, compliance with recommendations and the effectiveness of individual rehabilitation plans. The clustering of data based on the functional state of the musculoskeletal system enables a more accurate prediction of recovery for patients with similar features. Based on the clusters obtained, a rehabilitation plan is formulated and updated in accordance with the individualised recovery trajectory [8].

In order to evaluate the extent to which the individual rehabilitation plan derived from numerical modelling correlates with the actual outcomes, we examined two groups of patients. The first group (n=46) underwent endoscopic knee joint endoprosthesis and completed their rehabilitation using the prototype of the decision support system. The second group (n=34) did not utilise the decision support system during their rehabilitation. The developed system demonstrated a high degree of predictive capability, with 89.1% of cases (64.1% in the group rehabilitated in accordance with general recommendations) aligning with or exceeding the trajectory constructed by the system and the actual course of rehabilitation. In order to provide an objective evaluation, data on the progress of rehabilitation is still being collected. However, at this stage, rehabilitation therapists have already noted a positive effect of the prototype decision support system due to increased awareness of the functional state of the musculoskeletal system and direct involvement of the patient in the rehabilitation process.

The ability of systems developed based on artificial intelligence to manage and control recommendations in quasi-real time, adapting and modifying them according to the current state of

the patient is an important advantage, allowing the creation of a personalised recovery trajectory and an effective plan for further treatment.

3 Conclusion

The results obtained confirm the significant potential for application of the developed system in clinical practice. As the data accumulates, the system will continue to be improved, providing more accurate and effective interventions in process management, which will open new opportunities for further development and integration of additional functionalities. Further development and adaptation of the system will not only improve the quality of the rehabilitation process, but also significantly reduce the economic costs associated with the long-term treatment and rehabilitation of patients. Further work is being carried out in several directions: extending the system's functionality by integrating new virtual and augmented reality modules to create a visual image of the patient's digital model, and improving machine learning algorithms to analyse large amounts of data and improve the accuracy of the resulting predictions of the most effective rehabilitation strategy and duration.

References:

[1] Burdukovsky, V. N., Chudesova G. P., Actual issues of the use of medical cyberphysical systems / V. N. Burdukovsky, G. P. Chudesova // Strategies and tools of economic management: sectoral and regional aspect : Proceedings of the IX International Scientific and Practical Conference, St. Petersburg, 15 May 2020 / Under the general editorship of V. L. Vasilyonok. - St. Petersburg: Saint-Petersburg National Research University of Information Technologies, Mechanics and Optics, LTD " Scientific and Production Association of Fire Safety Automated Systems ", 2021. – P. 5-9. – EDN QDCDLB.

[2] Levonevskiy D, Motienko A. Using Mobile Applications for Patient Monitoring in Smart Medical Wards[C]//2024 International Russian Automation Conference (RusAutoCon). IEEE, 2024: 833-837.

[3] He, S., Meng, D., Wei, M., [et al.] Proposal and validation of a new approach in tele-rehabilitation with 3D human posture estimation: a randomized controlled trial in older individuals with sarcopenia. BMC Geriatr 24, 586 (2024). https://doi.org/10.1186/s12877-024-05188-7

[4] Misnik A.E. Metagraphs for ontological engineering of complex systems. Prikladnaya informatika= Journal of Applied Informatics, 2022, vol.17, no.2, pp.120-132 (in Russian). DOI: 10.37791/2687-0649-2022-17-2120-132.

[5] Kustova O.V., Khoziainova S.S., Abuseva G.R., Makhotkina N.N., Ponomarenko G.N., Reabilitatsionnye tekhnologii u patsientov posle total'nogo endoprotezirovaniya krupnykh sustavov nizhnikh konechnostei: naukometricheskii analiz [Rehabilitative technologies in patients after total endoprosthesis of lower limbs' major joints: scientometric analysis]. Vopr Kurortol Fizioter Lech Fiz Kult. 2024;101(1):54-61. Russian. doi: 10.17116/kurort202410101154. PMID: 38372738.

[6] Namazov, A. K., Namazov, K. A., Physical rehabilitation as the most important component in the system of medical rehabilitation / A. K. Namazov, K. A. Namazov // Health is the foundation of human potential: problems and ways to solve them. – 2019. – vol. 14, № 2. – P. 903-907. – EDN CEHQUH.

[7] Wang, J., Qiu, K., Peng, H., Fu, J., & Zhu, J. (2019). AI coach: Deep human pose estimation and analysis for personalized athletic training assistance. In Proceedings of the 27th ACM international conference on multimedia, P. 374-382. https://doi.org/10.1145/3343031.3350910

[8] Arseniev D. G., Borisov V. V., Misnik A. E., Gomzhina E. A., Shalukhova M. A., "Intelligent Synthesis of Individual Rehabilitation Plans for Patients after Joint Endoprosthesis Based on Computer Vision Technologies," 2024 7th International Conference on Information Technologies in Engineering Education (Inforino), Moscow, Russian Federation, 2024, pp. 1-5, doi: 10.1109/Inforino60363.2024.10552027.

An End-to-End 4D Millimeter-Wave Radar Camera Extrinsic Calibration Algorithm

Zhuang Long[1], Yao Yiqing[16]

(1. School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China)

Abstract: The online calibration of 4D millimeter-wave radars and cameras plays a significant role in enabling perception and SLAM technologies in complex environments. It bypasses manual labeling, providing real-time capabilities and ease of use. Nevertheless, the sparse nature of 4D radar point clouds poses challenges in establishing correspondences with camera images. To address this, we introduce an online 4D radar-camera calibration approach (4DRC-OC) that employs unified depth map representations for auxiliary training, ensuring feature alignment and modal unification between the two sensors. Given the limited useful information within sparse depth maps, we adopt dynamic convolution to adaptively capture rich feature details. Moreover, we design a Channel-wise Correlation Feature Matching Module (CMCF) that computes correlations between error depth maps and RGB-derived depth maps, enhancing features to aid extrinsic parameter regression. Experimental results on the Dual-Radar dataset demonstrate the superiority of our approach in extrinsic calibration.

 Key words: 4D radar; camera; extrinsic calibration; end-to-end; auxiliary training；

1 Introduction

As an emerging autonomous driving sensor [1], 4D millimeter-wave radar is gradually becoming a solution for environmental perception and SLAM technology in complex environments [2]. Due to the sparsity of its point clouds [3], 4D radar struggles to distinguish semantic information of objects, while camera sensors provide rich texture and semantic information of scenes. Thus, 4D radar and camera fusion is considered a robust combination scheme. For multi-sensor fusion [4], real-time and stable calibration is a prerequisite for efficient fusion.

In autonomous driving systems, mechanical vibrations during vehicle operation can affect the extrinsic parameters of multi-sensors. The proposed online calibration approach aims to address this issue. Currently, offline 4D radar-camera calibration methods rely on manual annotation [5], making real-time extrinsic matrix correction infeasible, and online calibration methods for 4D radar and cameras remain unexplored. This paper draws on LiDAR-camera online calibration algorithms. Previous LiDAR-camera online calibration algorithms [6] used deep learning techniques to extract and match features from input RGB images and LiDAR error depth maps, then regressed the extrinsic matrix (rotation matrix R and translation matrix T). However, for 4D millimeter-wave radar, sparse point clouds hinder feature matching with RGB images. Furthermore, experiments in this paper demonstrate that using traditional LiDAR-camera calibration algorithms like CMRNet [7] makes it difficult for end-to-end 4D radar-camera calibration models to converge. Another issue is

that 4D radar depth maps contain minimal information (most areas are blank), and using fixed convolution kernels leads to redundant feature extraction. Efficiently extracting features from 4D sparse radar depth maps is thus a key focus of this paper.

This paper proposes an end-to-end 4D radar-camera calibration method assisted by depth maps, named 4DRC-OC. Unlike traditional end-to-end calibration algorithms, this method primarily matches the correlation between 4D radar depth maps and error depth maps. This unified modality matching is easier compared to cross-modality matching. Considering the computational burden of monocular depth estimation, 4DRC-OC ingeniously uses monocular depth estimation as an auxiliary training branch. This branch is removed during the inference stage, ensuring depth map feature extraction while reducing network complexity. Another issue is that, the proportion of useful information in 4D radar depth maps is minimal, and using traditional fixed convolution kernels results in low feature extraction efficiency. Therefore, this paper introduces dynamic convolution kernels in the backbone feature extraction process, enabling the network to focus more on useful information in 4D radar depth maps and achieve adaptive feature extraction. Finally, to improve correlation matching computation, this paper designs a Channel-wise Correlation Matching Module (CMCF) to calculate the correlation between error depth maps and RGB-generated depth maps from a channel perspective and enhance features. Subsequent comparative and ablation experiments verify the effectiveness of the proposed method.
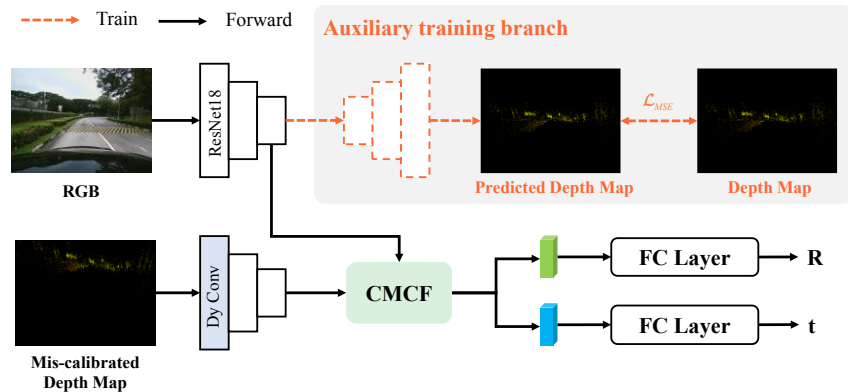
Fig. 1 Overall flowchart of the 4DRC-OC model

2   Method

2.1   Feature Extraction

Fig. 1 shows the overall flowchart of the 4DRC-OC model. The network input consists of the RGB image I_{RGB} and the error depth map I_{MD}, with RGB feature extraction primarily performed using the lightweight ResNet18 network. For the error depth map, considering the sparsity of its projected depth map, inherent convolution is not efficient in feature extraction. Therefore, this paper introduces dynamic convolution kernels [8] for adaptive feature extraction. Depending on the input, the dynamic convolution kernel will adaptively change the convolution kernel parameters, which is particularly beneficial for the sparse point cloud projection depth map of the 4D radar. Therefore, before extracting features from the 4D radar error depth map, this paper introduces a set of dynamic convolution kernels, ensuring that subsequent feature extraction focuses on the useful information of the 4D radar.

2.2 Auxiliary Training Branch

In this paper, we introduce an auxiliary training branch that cleverly removes the up-sampling portion of the depth estimation network and the predictive depth map feature extraction network, greatly reducing the number of parameters while ensuring modal consistency within the network. Specifically, this part performs upsampling on the RGB features f_{RGB}, where the upsampling operation is implemented by a transpose convolution $ConvT(\cdot)$ with a kernel size of 2×2 and an upsampling rate of [2, 2, 2, 2]. Depth prediction is then achieved through a 2D convolution $Conv1\times1(\cdot)$ with a kernel size of 1×1. The specific computation is shown in the following equation:

$$
\begin{aligned}
f_1 &= ConvT(f_{RGB}^4) \\
f_2 &= ConvT(f_1) \\
f_3 &= ConvT(f_2) \\
f_4 &= ConvT(f_3) \\
f_{Depth} &= Conv1\times1(f_4)
\end{aligned}
\tag{1}
$$

where, f_{RGB}^4 represents the fourth layer features of the RGB feature map, i.e., the last layer features; $f_{Depth} \in \mathbb{R}^{B\times1\times W'\times H'}$ represents the predicted depth map; W' and H' represent the width and height of the depth map, respectively. In this paper, $W' = 256$ and $H' = 384$.

2.3 Correlation Matching Module Based on Channel Fusion

After obtaining the depth map features and the error depth map features, feature matching and fusion need to be performed. This paper uses a correlation layer to calculate the matching degree $c(x_1, x_2)$ between the depth map and error depth map, which is essentially a special convolution operation (similar to an inner product), where depth map features convolve with error depth map features. Moreover, this convolution operation does not require training parameters, simplifying the training process. The specific calculation can be expressed by the following formula:

$$
\begin{aligned}
c(x_1, x_2) &= \sum \langle L_1(x_1 + o), L_2(x_2 + o) \rangle \\
O &\in [-k, k] \times [-k, k]
\end{aligned}
\tag{2}
$$

where L_1 and L_2 represent the predicted depth map and the 4D radar error depth map, respectively; x_1 and x_2 represent the pixels on L_1 and L_2, respectively; O represents the expansion region of the correlation calculation, where the expansion region is based on points x_1 and x_2, extending k pixels upwards and downwards, hence the length is $2k+1$.

The matching feature $f_c \in \mathbb{R}^{B\times C'\times W''\times H''}$ can be obtained through the above correlation layer, but this is not sufficient. This paper also fuses the matching feature and error depth map features because the error depth map features are the premise that causes the matching errors. The error depth map feature f_{MD} is first dimension-reduced using a 1×1 convolution kernel to obtain $f_{MD}^{(1)} \in \mathbb{R}^{B\times C'\times W''\times H''}$, where C' represents the channel dimension, and in this paper, it is 81. To achieve fusion, this paper recombines half of the channels of f_c and $f_{MD}^{(1)}$ to obtain new feature sets $U_1 \in \mathbb{R}^{B\times C\times W''\times H''}$ and $U_2 \in \mathbb{R}^{B\times C\times W''\times H''}$, which respectively contain half of the channel features of the aforementioned two and are used for regressing the rotation matrix and the translation matrix.

Finally, the self-attention mechanism is used to enhance the features of U_1 and U_2, which allows the network to focus on useful feature information for the result, facilitating the subsequent regression operations. Specifically, this paper first uses a 1×1 convolution kernel to perform feature dimension reduction and reshaping on U_1 and U_2, then applies the self-attention mechanism to enhance their features. Ultimately, $M_R \in \mathbb{R}^{B \times N \times D}$ and $M_t \in \mathbb{R}^{B \times N \times D}$ are obtained, which are used for regressing the rotation matrix and the translation matrix, where N represents the number of channels, and D represents the hidden layer dimension. The specific calculation is shown as follows:

$$
\begin{aligned}
Z_1 &= \text{Reshape}[\text{Conv}1 \times 1(U_1)] \\
M_R &= \text{Attention}(Z_1) \\
Z_2 &= \text{Reshape}[\text{Conv}1 \times 1(U_2)] \\
M_t &= \text{Attention}(Z_2)
\end{aligned}
\tag{3}
$$

2.4 Regression

The enhanced rotation and translation matrix features need to be reshaped to obtain $M'_R \in \mathbb{R}^{B \times M}$ and $M'_t \in \mathbb{R}^{B \times M}$, where $M = N \times D$. Then, downsampling is performed using a linear layer with a sampling rate of [2, 2]. The final layer directly regresses the rotation and translation matrices using a linear layer, where the rotation matrix is also normalized, resulting in the current 4D radar-camera extrinsic matrix T_0.

3 Experiment

In this paper, experiments were conducted on the latest Dual-Radar dataset [9]. This section mainly introduces the two datasets, evaluation metrics, implementation details, comparison with advanced methods, and ablation experiments.

Tab.1 Comparison results of different models on the dual-radar dataset

Method		Translation (cm)				Rotation (°)			
		E_t	X	Y	Z	E_R	Roll	Pitch	Yaw
CMRNet	Mean	10.5108	8.5220	4.0005	**1.7908**	0.6813	0.3513	0.8525	0.3755
	Median	7.6839	7.5047	3.6206	**1.7466**	0.5582	0.2486	0.1576	0.4661
	Std.	9.8474	6.9811	2.6805	**0.5237**	0.5929	0.3759	1.2416	0.3332
LCCNet	Mean	7.2577	3.0632	2.8168	1.9388	0.6664	0.2026	**0.4626**	**0.2779**
	Median	5.9850	2.3110	1.1605	1.9969	0.5614	0.0972	0.1769	0.3362
	Std.	5.4423	1.6420	3.0447	0.2017	0.4392	0.2133	0.6455	**0.1483**
Our	Mean	**4.7603**	**1.0491**	**0.6326**	2.4042	**0.3865**	**0.0848**	0.0462	0.4033
	Median	**3.7492**	**1.0493**	**0.6423**	1.9683	**0.2387**	**0.0959**	**0.0294**	**0.1577**
	Std.	**4.5384**	**0.0306**	**0.6229**	1.3464	**0.3533**	**0.0332**	**0.0452**	0.5083

3.1 Dataset

The Dual-Radar dataset [9] includes test data from two types of 4D millimeter-wave radars, primarily used for 3D object detection and tracking tasks. The dataset consists of 151 continuous sequences, most of which last for 20 seconds and contain 10,007 synchronized and annotated frames. About two-thirds of the data were collected in normal weather conditions, one-third in adverse weather, and additional data were collected during dawn and night. The dataset uses two types of

4D millimeter-wave radars, Arbe Phoenix and ARS548 RDI, where the Arbe radar contains approximately 6K to 14K points per frame, and the ARS548 contains about 400 to 650 points per frame. The 4D radar-camera extrinsic matrix for this dataset was obtained through offline calibration methods and is used as the ground truth in this paper.

3.2 Comparison to state-of-the-art

There is no mature 4D radar-camera online calibration algorithm yet, and previous methods [11] did not use correlation matching, so their comparability is limited. This paper selects two excellent LiDAR-camera end-to-end calibration algorithms, CMRNet [7] and LCCNet [10], as baseline comparisons, both of which use RGB images and point cloud projected error depth maps as inputs. All experimental results are obtained under the same testing conditions. This paper follows LCCNet's testing conditions, using five sets of dataset input disturbances: $(\pm 20^\circ, \pm 1.5m)$, $(\pm 10^\circ, \pm 1.0m)$, $(\pm 5^\circ, \pm 0.5m)$, $(\pm 2^\circ, \pm 0.2m)$, $(\pm 1^\circ, \pm 0.1m)$. Table 1 presents the online calibration results of three methods on the Dual-Radar dataset, with only the results after the fifth iteration refinement shown for simplicity. It is noted that the method proposed in this paper achieves the smallest translation matrix error and rotation matrix error, with average values of 4.76 cm and 0.39°, respectively. However, the average value for CMRNet is 10.51 cm, which even exceeds the input disturbance of the fifth iteration (10 cm), indicating that the sparsity of the 4D radar point cloud significantly affects the model's convergence. This also proves the convergence difficulty when traditional LiDAR-camera calibration algorithms are applied to 4D radar-camera systems, as mentioned earlier. LCCNet is currently the most classic LiDAR-camera online calibration algorithm, but the errors it achieved are still higher than those of the method proposed in this paper, mainly due to the modality mismatch issue.

Overall, the proposed 4DRC-OC achieves the best calibration results on the latest 4D radar dataset, thanks to the designed dynamic convolution and auxiliary training branches, which help mitigate the sparsity of 4D radar and modality mismatch issues.

4 Conclusion

The 4D millimeter-wave radar-camera online calibration algorithm is constrained by the sparsity of 4D millimeter-wave radar point clouds, which makes it challenging to establish correspondences between sparse point clouds and camera images. This paper introduces a 4D millimeter-wave radar-camera online calibration paradigm, 4DRC-OC, which incorporates an auxiliary training branch for depth estimation to address the modality mismatch issue. Furthermore, to enhance model performance, dynamic convolution is introduced for adaptive feature extraction, and the CMCF correlation matching module is proposed to fuse features from the channel perspective, realizing more thorough correlation matching. When compared with outstanding LiDAR-camera online calibration algorithms on the Dual-Radar dataset, the proposed method achieved the best results, with translation and rotation errors of 4.7603 cm and 0.3865°, respectively, under input disturbances. In the future, we aim to leverage additional underlying information from 4D millimeter-wave radar, such as distance, angle, Doppler, and intensity data, to develop a multi-information fusion calibration algorithm.

References:

[1] Y. Zhuang and M. Li, "4D iRIOM: 4D Imaging Radar Inertial Odometry and Mapping," IEEE ROBOTICS AND AUTOMATION LETTERS, vol. 8, no. 6, 2023.

[2] S. Lu et al., "Efficient Deep-Learning 4D Automotive Radar Odometry Method," IEEE Trans. Intell. Veh., vol. 9, no. 1, pp. 879–892, Jan. 2024.

[3] M. Choi et al., "MSC-RAD4R: ROS-Based Automotive Dataset With 4D Radar," IEEE Robot. Autom. Lett., vol. 8, no. 11, pp. 7194–7201, Nov. 2023.

[4] F. Ding, A. Palffy, D. M. Gavrila, and C. X. Lu, "Hidden Gems: 4D Radar Scene Flow Learning Using Cross-Modal Supervision," in 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada: IEEE, Jun. 2023, pp. 9340–9349.

[5] N. Kumar et al., "An Efficient Approach for Calibration of Automotive Radar–Camera With Real-Time Projection of Multimodal Data," in IEEE Transactions on Radar Systems, vol. 2, pp. 573-582, 2024.

[6] N. Schneider, F. Piewak, C. Stiller and U. Franke, "RegNet: Multimodal sensor registration using deep neural networks," 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 2017, pp. 1803-1810.

[7] D. Cattaneo, M. Vaghi, A. L. Ballardini, S. Fontana, D. G. Sorrenti and W. Burgard, "CMRNet: Camera to LiDAR-Map Registration," 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 2019, pp. 1283-1289.

[8] Y. Chen, X. Dai, M. Liu, D. Chen, L. Yuan and Z. Liu, "Dynamic Convolution: Attention Over Convolution Kernels," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 11027-11036.

[9] X. Zhang et al., "Dual Radar: A Multi-modal Dataset with Dual 4D Radar for Autonomous Driving," 2023, arXiv: 2310.07602.

[10] X. Lv, B. Wang, Z. Dou, D. Ye and S. Wang, "LCCNet: LiDAR and Camera Self-Calibration using Cost Volume Network," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 2021, pp. 2888-2895.

[11] C. Schöller et al., "Targetless Rotational Auto-Calibration of Radar and Camera for Intelligent Transportation Systems," 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 2019, pp. 3934-3941.

Preliminary Exploration of AUV Inertial Self-aided Navigation Method in Non-cooperative Environments

Liu Xixiang[1,2], SHAO Qiantong[1,2], Cheng Xiangzhi[1,2], Yao Yiqing[1,2], Liu Xianjun[3]

(1. School of Instrument Science & Engineering, Southeast University, Nanjing 210096, China;

2. Key Laboratory of Micro-inertial Instrument and Advanced Navigation Technology, Ministry of Education, Southeast University, Nanjing 210096, China;

3. Wuhan Second Ship Design and Research Institute, Wuhan 430205, China)

Abstract: To address the principle deficiency that the autonomous underwater vehicle (AUV) navigation system can only rely on strapdown inertial navigation systems (SINS) for pure inertial navigation calculation, and SINS has theoretical insufficiency of error accumulation, an inertial self-aided method based solely on SINS information is proposed. This method corrects the cumulative error of SINS and reduces the speed of error accumulation subsequently by adding a self-aided segment during the AUV's long voyage. The self-aided navigation method applies a high-pass digital filter to separate the high-frequency periodic motion information, and further regards the separate high-frequency motion as an auxiliary information source for inertial-based integrated navigation. Taking circular motion as an example, the theoretical analysis, mathematical simulation, and AUV navigation test have preliminary verified the effectiveness of the self-aided method. Based on theoretical analysis and experimental results, the key techniques and capability boundaries of this method were further explored[17].

 Key words: AUV; SINS; inertial self-aided; non-cooperative environment

1　INTRODUCTION

With the increasing scarcity of land resources, the ocean which accounts for 71% of the earth's surface area has become the focus of human attention [1]. The support of vehicles is indispensable for the strategy of the ocean, and autonomous underwater vehicles (AUVs), as a class of important platforms, play an important role in various fields [2]. The special characteristics of the water medium make the navigation problem one of the three key/difficult technologies (energy, communication, and navigation/control) for AUVs [3].

At this stage, the navigation and positioning of long-voyage AUVs mainly rely on the integration of SINS and DVL, which utilizes the acoustic Doppler effect to measure the velocity of the vehicle concerning the reflector (e.g., the seafloor or a certain layer of water current) [4][5]. However, the propagation of acoustic signals introduces information leakage and risks being susceptible to detection/capture by non-cooperative objects. AUV navigation systems in underwater non-cooperative environments can only rely on pure inertial navigation calculation by SINS, which suffers from the principal deficiency of error accumulation. Neither the vehicle nor the navigation system in a non-cooperative environment interacts with the external environment or system for

information.

In 2020, the article [6] noticed the maneuvering characteristics of the HOV spiral dive in the context of a large-depth Human Occupied Vehicle (HOV) and accordingly designed an inertial self-aided navigation method, along with mathematical simulation and vehicle test verification. However, there are problems with idealized test conditions and unclear boundary analysis of the method capability. In this article, based on the article [6], we further sort out the basic principles of the method and carry out AUV navigation tests to explore the key technologies and capability boundaries of the method.

2 BASIC PRINCIPLE OF INERTIAL SELF-AIDED METHOD

2.1 Basic principle of inertial self-aided method

The error propagation rule under static base conditions is relatively well studied in the industry，whereas it is difficult to study the SINS error propagation rule under vehicle maneuvering conditions. The feedback coupling relationship is extremely complex, and it involves various issues such as the vehicle maneuvering trajectory, which makes it difficult to get its analytical solution. However, it is possible to study the frequency composition of SINS error under a specific trajectory condition by simulation. Article[13] studied the SINS error frequency components under the condition of circular motion by taking the HOV spiral motion as an example, and under the condition of circular motion, the projection of the vehicle trajectory velocity in the horizontal plane in the east and north directions is the periodic sinusoidal information, as shown in Fig. 1.
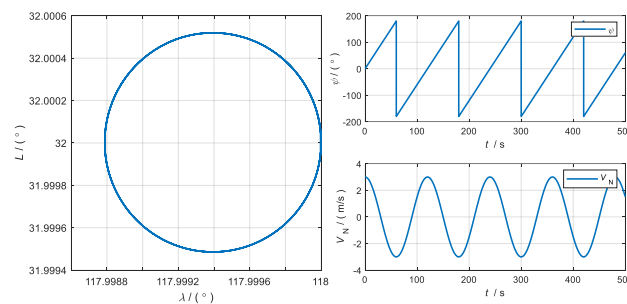
Fig. 1　Circular motion trajectory and motion information
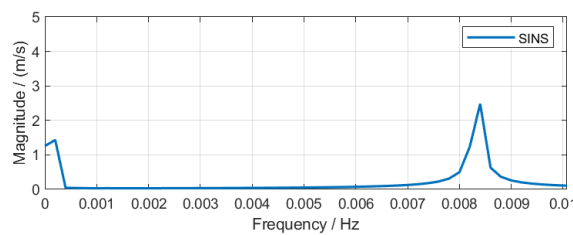
Fig. 2　Frequency components of SINS calculations under circular motion simulation conditions

The simulation test results show that under the excitation of such periodic motions as circular motion, the SINS horizontal velocity behaves as a combination of three low-frequency error information and the periodic motion information, and the periodic circular information has a better differentiation from the Schuler cycle (84.4 min) in the frequency spectrum, as shown in Fig. 2.

2.2 Separation of auxiliary information based on frequency characteristics

Since the periodic motion information has good frequency discrimination from low-frequency

errors in the SINS navigation calculation results，this information can be separated with the help of digital filtering techniques. Digital filters are a common tool in digital signal processing to process mixed signals with multiple frequency components, removing unwanted signal components and retaining valid frequency components. To retain the high-frequency motion information in the SINS calculation results under periodic motion excitation, the high-pass digital filtering technique is used.
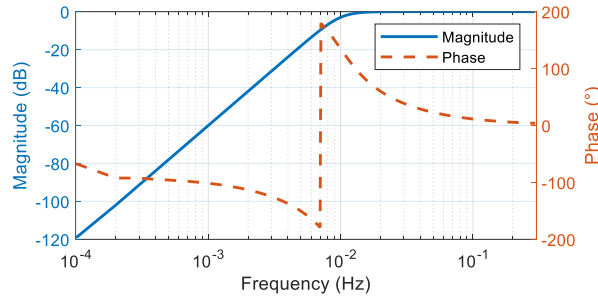
Fig. 3　Amplitude/Phase frequency characteristics of the high-pass digital filter

Fig. 3 shows the amplitude/phase frequency characteristics of the high-pass digital filter. The curves in the figure show that the actual obtained high-pass digital filters always have some amplitude attenuation and phase overrun, and the phase delay will introduce a large separation error for periodic signals.

Noting that for digital filters, a definite frequency corresponds to a definite amplitude attenuation and phase delay, here we introduce an amplitude-phase compensation method with Band-limited Multiple Fourier Linear Combiner (BMFLC), as shown in Fig. 4. Since each frequency component is determined in advance, the amplitude and phase change parameters brought about by the digital filter can be determined in advance, so they can be compensated in the above fitting process, thus obtaining high-precision filtering results.
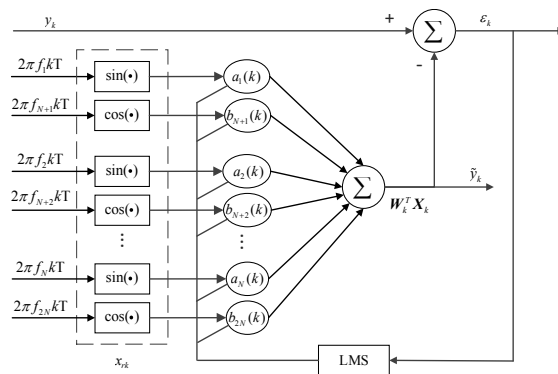
Fig. 4　BMFLC Schematic

2.3　Integrated navigation system based on self-aided velocity matching

After obtaining the separated vehicle high-frequency motion information, the high-frequency information can be directly used as the auxiliary navigation information, which together with SINS constitutes the integrated navigation system based on velocity matching, as shown in Fig. 5. In the integration process, the IMU measurement information is input into the integrated navigation module and the self-aided information extraction module respectively. The self-aided information

extraction module first carries out SINS pure inertial navigation, and inputs the pure inertial navigation calculation results into digital filtering module to extract the auxiliary information; the integrated navigation module carries out SINS navigation calculation and information fusion, and the information fusion unit receives the navigation information from the SINS navigation-solving unit and the auxiliary information from the auxiliary information extraction module for information fusion, and returns the fusion results to the SINS navigation solving unit for various types of correction.
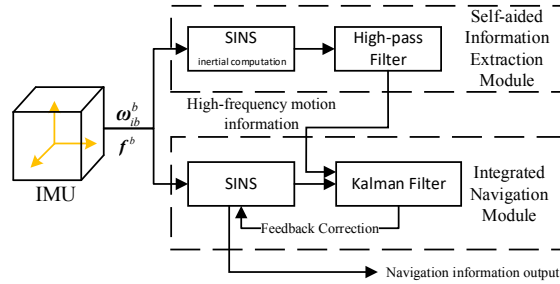
Fig. 5 Integrated Navigation System Based on Self-aided velocity Matching

The state equations and measurement equations in the integrated navigation system based on self-aided velocity matching are not significantly different from the general velocity-matched integrated navigation system, such as the SINS/GNSS integration, and are not listed here. In this article, the misalignment angle, velocity error, position error with gyroscope zero bias and accelerometer zero bias are subsequently chosen as the system state vectors.

3 Experimental validation of an inertial self-aided method

The mathematical simulation and vehicle simulation validation were carried out by literature[6], and the results of the AUV navigation test are further given in this article.

The length of the test was 9500 seconds, and the AUV navigation trajectory acquired with the help of GNSS is shown in Fig. 6. The AUV navigation trajectory generally shows the regularity of circular motion; however, due to the reason of cascading currents in the lake, the AUV circular motion trajectory is superimposed with irregular other motions.
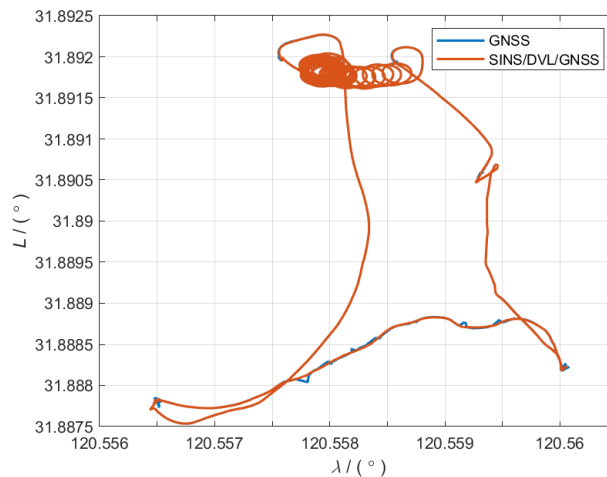
Fig. 6 AUV trajectory

When the test is carried out for 2000 seconds, the AUV performs a circular motion, and the auxiliary velocity is obtained using the BMFLC method in Section 1.2 as shown in Fig. 7. The

statistical results of the accuracy of the separated self-aided velocity compared to the integrated SINS/GNSS/DVL velocity are shown in Tab1, and this error includes the ocean current error.
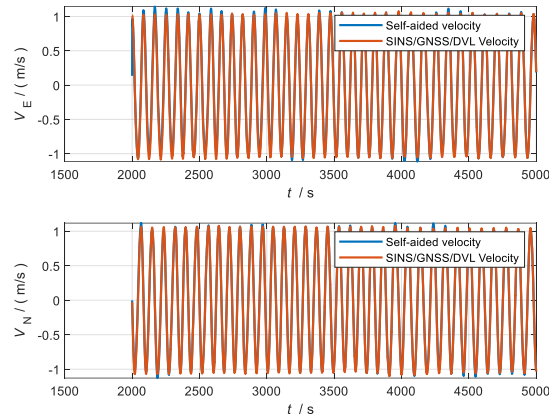
Fig. 7　Velocity Comparison Curve

Tab.1 Self-aided velocity accuracy statistics results

	$\delta V_E / (m/s)$	$\delta V_N / (m/s)$
Root Mean Square Error (RMSE)	0.042	0.027
Mean	0.012	0.003
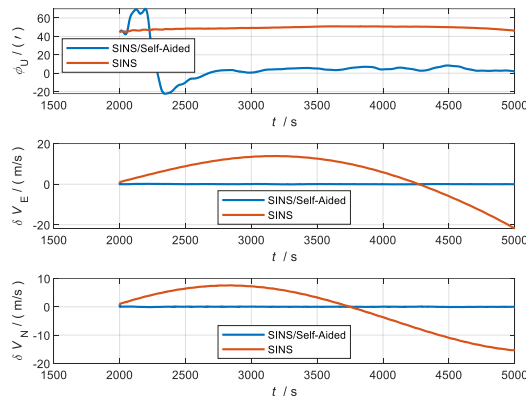
Fig. 8　Comparison curves of heading misalignment angle and horizontal velocity error

　　To simulate the cumulative error in the early part of the long navigation process, the attitude and velocity errors are artificially added to both SINS-calculation sessions in Section 2.3 before entering the self-aided session, which are $\begin{bmatrix} 0.1° & -0.1° & 0.7° \end{bmatrix}$ and $\begin{bmatrix} 1m/s & 1m/s & 0 \end{bmatrix}$ respectively. Since the self-aided velocity cannot introduce position correction information, the position error is not considered here, and only the accumulation process of position error is considered after the integration. The integration results are shown in Fig. 8, compared with the SINS results, the misalignment angle and velocity error of the SINS/self-aided velocity integration results are effectively corrected, while Fig.9 gives the integration position comparison results with SINS/GNSS/DVL.
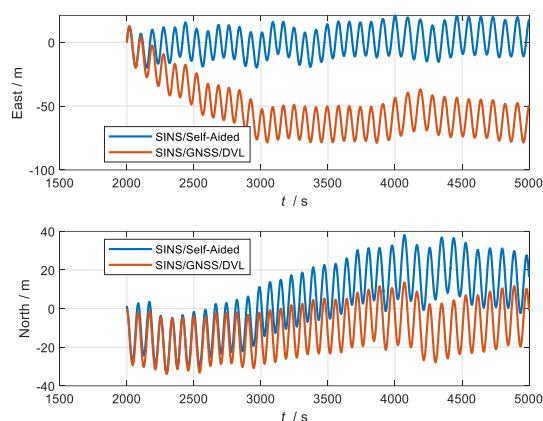
Fig. 9   Navigation Position Comparison Curve

4   Conclusions

This article provides a preliminary discussion on inertial self-aided technology. The AUV navigation system under non-cooperative conditions urgently needs to tap new sources of auxiliary information. Starting from the SINS error frequency characteristics, the self-aided technique utilizes the periodic high-frequency motion of the vehicle to introduce high-frequency information into the SINS navigation information and utilizes digital filters to separate the effective vehicle high-frequency motion information from the invalid three types of low-frequency errors. The separated high-frequency motion information is further utilized as the SINS auxiliary information to correct some of the accumulated errors, provide more accurate initial conditions for the subsequent navigation, and slow down the accumulation of subsequent errors. The navigation test based on circular motion verifies the effectiveness of the inertial self-aided method.

References:

[1]   Feng X S, Li Y P, Xu H X, et al. Development of deep-sea autonomous underwater vehicle and its application in resource survey[J]. Chin. J. Nonferrous Met, 2021, 31: 2746-2756.

[2]   SONG B W, PAN G, ZHANG L C, et al. Development trend and key technologies of autonomous underwater vehicles[J]. Chinese Journal of Ship Research, 2022, 17(5): 27–44.

[3]   Teeneti C R, Truscott T T, Beal D N, et al. Review of wireless charging systems for autonomous underwater vehicles[J]. IEEE Journal of Oceanic Engineering, 2019, 46(1): 68-87.

[4]   LUO H, LIU X X, LIU J W, et al. A method for DVL calibration based on parameter estimation between two point sets[J]. Chinese Journal of Ship Research, 2024, 19(3): 344–352.

[5]   Wang D, Wang B, Huang H, et al. SINS/DVL integrated navigation method based on adaptive particle swarm optimization in complex underwater environment[J]. Journal of Chinese Inertial Technology, 2023, 31(10): 1023-1029+1036.

[6]   Liu X. Research on key technologies of integrated navigation system for deep human occupied vehicle[D]. Southeast University, 2021.

[7]   Feizi N, Patel R V, Kermani M R, et al. Adaptive wave reconstruction through regulated-BMFLC for transparency-enhanced telerobotics over delayed networks[J]. IEEE Transactions on Robotics, 2022, 38(5): 2928-2942.

The Features of Federated Filtering Methods in Nonrecursive Measurement Processing

Yu. A. Litvinenko, O. A. Stepanov, A. M. Isaev

(Concern CSRI Elektropribor, ITMO University, St. Petersburg, Russia)

Abstract: The paper discusses the features of decentralized processing of measurement information, which is aimed to obtain the desired estimate of the state vector of a dynamic system by weighing the local estimates generated in distributed measuring modules. The recursive and non-recursive schemes of measurements processing in federated filters are compared. Conditions for recursive and non-recursive schemes are considered that ensure the coincidence of the estimates and calculated covariance matrices generated in federated filters and the optimal centralized Kalman filter. A methodical example of the Wiener sequence estimation is given to illustrate the methods under consideration.

It is known that the problem of designing algorithms for estimation of a dynamic system state with measurement redundancy can be solved on the basis of both centralized processing, when all available measurements are fed into a single centralized filter, and different decentralized schemes, or so-called federated filtering (FF) algorithms [1-4].

The main advantage of a centralized scheme is the possibility of obtaining an RMS-optimal estimate when solving the estimation problem for a linear system using linear measurements and the Gaussian nature of the system noise and measurement errors. This is achieved through centralized processing of all measurements by means of an optimal Kalman filter that recursively processes the entire set of measurements. However, the disadvantages of centralized schemes in comparison with FF methods are their lower reliability and higher computational load.

The essence of FF methods is that primary processing of information is carried out in so-called local filters (LF), in which local estimates of the dynamic system state are generated based on measurements obtained in local measuring modules. A general estimate is calculated by averaging the local estimates in the so-called master filter. The main advantages of FF algorithms are lower computational load compared to centralized Kalman filters and their immunity against false measurements obtained in one of the measuring modules [1]. FF algorithms have recently been widely used to solve estimation problems based on data from distributed sensors, in particular, in relation to problems of collaborative navigation of unmanned vehicles [7-10]. The features of recursive FF algorithms and their application to navigation information processing are discussed in detail in [1]. Further development of such algorithms is discussed in [2,4-6], which demonstrate the possibility of ensuring the so-called guaranteeing properties of the FF, the essence of which is that the actual error of the general estimate does not exceed the calculated error obtained in the master filter. The main disadvantage of the FF is the fact that in the general case, these algorithms do not provide optimal estimates, and the accuracy characteristic calculated in the master filter is not the

actual covariance matrix of the generated estimate error. Although for some special cases the conditions for the so-called matched adjustment of local filters under which it is possible to derive an optimal estimate in the master filter are formulated [2,4], the authors are not aware of any general answer to the question about the conditions ensuring the coincidence of the estimates and calculated covariance matrices generated in the FF with the results derived in the optimal centralized Kalman filter.

It should be emphasized that the above algorithms are recursive with respect to the incoming measurements. At the same time, in the last decade, there has been an increased interest in nonrecursive processing methods, including algorithms based on factor-graph optimization [11-12]. A number of works show the advantage of nonrecursive schemes over traditional recursive algorithms, for example, the Kalman filter. However, to date, FF methods have only been studied for traditional recursive schemes for processing incoming measurements, and possible advantages of nonrecursive FFs have not been considered in the literature. In this regard, the aim of the proposed work is to study the features of FF methods in the implementation of nonrecursive schemes for processing measurements.

The paper formulates the problem of estimating an unknown random vector x when processing measurements coming from m sensors. The solution of the filtering problem is given for a centralized filter that provides an optimal estimate and an FF that includes m local filters. We consider the recursive and nonrecursive schemes of solution for the both variants of the filter construction. With regard to the traditional recursive FF, the conditions for the matched adjustment of local filters are formulated, while it is noted that in most estimation problems, they do not provide an optimal solution. Also given are the conditions for matched adjustment for an FF designed using a nonrecursive scheme, but it is proved that they ensure the coincidence of the estimates and covariance matrices of the FF estimates and the centralized (optimal) filter in the general case.

As an illustration, an example of estimating the Wiener sequence using the measurements in two measuring modules is given. For the example considered, the conditions for FF adjustment for recursive and nonrecursive measurement processing schemes are compared. It is shown that in order to ensure an optimal FF estimation for recursive processing, it is necessary to comply with the FF adjustment conditions that require additional consideration of measurement noise variances, whereas for a nonrecursive scheme, this is not required in the FF adjustment conditions.

Thus, the paper studies the features of a nonrecursive FF in comparison with a traditional recursive processing scheme. An important difference is shown for a nonrecursive FF, consisting in the possibility of obtaining an optimal state vector estimate, whereas for a recursive processing scheme, FF methods provide optimal estimates only for special cases.

.

References:

[1] Carlson, N.A., Federated filter for fault-tolerant integrated navigation systems, AGARDograph 331, Aerospace Navigation Systems, June 1995, pp. 265-280.

[2] Tupysev, V.A., Methods for designing filtering algorithms with guaranteed quality of parameter estimation in problems of navigation information processing, Dr. Sci. Dissertation, St.Petersburg, 2011.

[3] Stepanov, O.A., Osnovy teorii otsenivaniya s prilozheniyami k zadacham obrabotki navigatsionnoi informatsii (Fundamentals of the Estimation Theory with Applications to the Problems of Navigation Information

Processing), Part 2, Vvedenie v teoriyu otsenivaniya (Introduction to the Estimation Theory), St. Petersburg: TsNII Elektropribor, 2012.

[4] Tupysev, V.A. and Litvinenko, Yu.A., The Effect of the local filter adjustment on the accuracy of federated filters", Proc. MICNON-2015, pp. 349-354.

[5] Zhang, H., Lennox, B., Goulding, P.R., and Wang, Y., Adaptive information sharing factors in federated Kalman filtering, The 15th IFAC World Congress, Barcelona, Spain, 2002, July 21–26, pp. 664–664.

[6] Selezneva, M.S., Proletarsky, A.V., Neusypin, K.A., and Zhang Lifei, Modification of the federated Kalman filter using the observability degree criterion of state variables, 26th St. Petersburg Int. Conf. on Integrated Navigation Systems, St. Petersburg: Elektropribor, 2019.

[7] Maier, A., Kiesel, S., and Trommer, G.F., Performance analysis of federated filter for SAR/TRN/GPS/INS integration, Gyroscopy and Navigation, 2011, vol. 2, no. 4, pp. 293-300.

[8] Weiwei Lyu, Xianghong Cheng, and Jinling Wang, Adaptive federated IMM filter for AUV integrated navigation systems. Sensors 2020, 20, doi:10.3390/s20236806

[9] Pengfei Zhang, Zhenhua Ma, Yin He, Yawen Li, and Wenzheng Cheng, Cooperative positioning method of a multi-UAV based on an adaptive fault-tolerant federated filter. Sensors 2023, 23, https://doi.org/10.3390/s23218823

[10] Tarık Ayabakan, Feza Kerestecioglu. Fault tolerant indoor positioning based on federated Kalman filter. Journal of Signal Processing Systems, 2024, pp. 273–285

[11] Weisong Wen, Tim Pfeifer, Xiwei Bai, and Li-Ta Hsu, Factor graph optimization for GNSS/INS integration: A comparison with the extended Kalman filter, NAVIGATION, 2021; 68:315–331.

[12] Stepanov, O.A, Isaev, A.M., Motorin, A.V., Litvinenko, Yu.A., and Zolotarevich, V.P., Estimation algorithms based on factor-graph optimization vs. Bayesian estimation algorithms for navigational data processing: Differences and interrelation , 31st St. Petersburg Int. Conf. on Integrated Navigation Systems, St. Petersburg: Elektropribor, 2024, pp. 336-344.

Determination of The Spacecraft Orientation Using Instantaneous Measurements

M. A. Ponomareva, A. V. Kramlikh

(Samara National Research University, Samara, Russia)

Abstract: The paper presents a comparison of the most popular algorithms for determining the orientation of a spacecraft based on instantaneous measurements of various physical nature according to the following terms: the speed of algorithms, the accuracy and the type of output information.

1 Introduction

Information about the angular position of the spacecraft is used to link measurements carried out by the on-board equipment and in the spacecraft angular motion control system. There are many algorithms for determining the orientation of the spacecraft. In this paper, algorithms using instantaneous measurements are reviewed. They can determine orientation from two or more vector measurements obtained instantaneously (or with a negligible delay between measurements). Such algorithms are simple, do not require a priori information about the spacecraft position, have high speed, but low accuracy. The proposed report is devoted to the research and comparison of a number of the most commonly used instantaneous algorithms.

2 Mathematical formulation of the problem

The problem of determining the orientation of the spacecraft is to determine the angular position of the coordinate system (CS) associated with the spacecraft relative to some reference CS, for example, the orbital CS, based on available measurements in the associated CS and models of these measurements in the orbital CS. The angular position of the associated CS relative to the orbital CS is described by a rotation matrix $\mathbf{M}$, which can be parameterized by Euler angles or a quaternion [1].

The formula for the relationship of measurements and models of these measurements:

$$\mathbf{b}_i = \mathbf{M} \cdot \mathbf{r}_i , \tag{1}$$

where: $\mathbf{b}_i$ – measurement vectors ($i = \overline{1, N}$, N - the number of unit measurement vectors in the associated CS); $\mathbf{r}_i$ – measurement models vectors in the orbital CS; $\mathbf{M}$ – rotation matrix.

Therefore, the problem is came down to finding the rotation matrix $\mathbf{M}$.

In this paper, the following algorithms for instantaneous measurements are researched: TRIAD [2-4], Optimized TRIAD [5], q-method [4,6,7], QUEST [3,4,7], ESOQ [4,7], ESOQ2 [4,7], SVD [4,7].

3 Results

To compare the algorithms, mathematical modeling was done in Matlab under the following conditions.

The orientation angles were set randomly and the true rotation matrix was calculated.

The elements of the vectors in the orbital CS were generated according to an equally probable law in the range of [-1000;1000]. Then the vectors were normalized, because the directions of these vectors and the angle between them are important, not their length.

The vectors in the associated CS are calculated by formula (1) with the addition of Gaussian noise (the average value is zero, the value of the root mean squre (RMS) was taken as 1% and 10% of the vector modulus). Then the vectors were normalized.

The measure of the error in determining the orientation of the spacecraft is found by the formula

$$\Phi = \arccos\left(\frac{1}{2}\left(tr\left(\mathbf{B}_{ou}\right)-1\right)\right), \tag{2}$$

where $\mathbf{B}_{ou} = \mathbf{M}\cdot\mathbf{M}_1^{\mathrm{T}}$, $\mathbf{M}$ – true rotation matrix, $\mathbf{M}_1$ – found rotation matrix.

For each algorithm, the measure of the orientation error and the speed of operation for 10000 iterations were calculated. And the number of floating point operations was determined for one iteration of the algorithm.
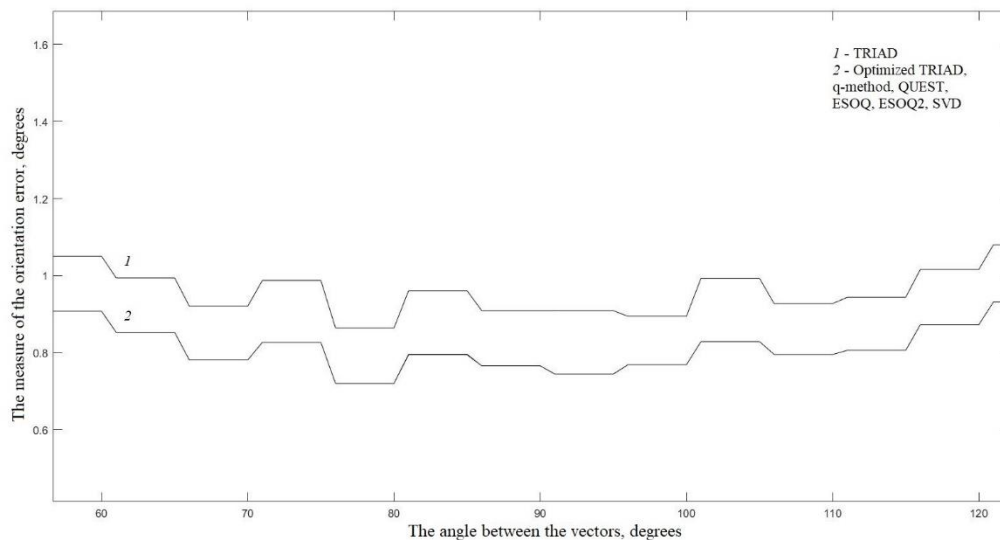
Fig.1 The measure of the orientation error for RMS 1%

Tab.1 Results

Algorithm	Speed*, s	Number of floating point operations **	Type of output information
TRIAD	0,053	11	Rotation matrix
Optimized TRIAD	0,169	42	Rotation matrix
q-method	0,241	59(900)***	Quaternion
QUEST	0,253	227	Quaternion
ESOQ	0,356	125	Quaternion

| ESOQ2 | 0,152 | 79 | Quaternion |
| SVD | 0,175 | 53(650)*** | Rotation matrix |

* The running time of the algorithms is shown for computing 10,000 iterations.

** A function implemented in Matlab was used to calculate floating-point operations [8].

*** The q-Method and SVD algorithms use complex built-in functions for which it is impossible to determine the number of floating-point operations, so the values obtained in other papers are shown in table [7].

4 Conclusion

From results, it can be seen that the least accurate algorithm is TRIAD, other algorithms work with approximately the same accuracy. The fastest algorithms are TRIAD and ESOQ2, the slowest are ESOQ and QUEST. The algorithms with the fewest floating point operations are TRIAD and ESOQ2. According to the results of the comparison, TRIAD and ESOQ2 algorithms performed the best.

References:

[1] Бранец В.Н., Шмыглевский И.П. Применение кватернионов в задачах ориентации твердого тела. М.: Наука, 1973. 320 с.

[2] Black H.D. A Passive System for Determining the Attitude of a Satellite. AIAA Journal. 1964. Vol. 2. №7. P. 1350-1351.

[3] Shuster M.D., Oh S.D. Three-Axis Attitude Determination from Vector Observations. Journal of Guidance and Control. 1981. Vol. 4. №1, Jan.-Feb. P. 70-77.

[4] Markley F.L., Crassidis J.L. Static Attitude Determination Methods. Fundamentals of Spacecraft Attitude Determination and Control. 2014. Chap.5. P. 183-233.

[5] Bar-Itzhack Y.I., Harman R.R. Optimized TRIAD algorithm for attitude determination. Collection of Technical papers (A96-34712 09-12). 1996. P. 422-427.

[6] Davenport P. A vector approach to the algebra of rotations with applications. NASA TN D-4696. 1968.

[7] Markley F.L., D. Mortari. How to estimate attitude from vector observations. AAS 99-427, presented at the AAS/AIAA Astrodynamics Specialist Conference, Girdwood, Alaska. – 1999.

[8] Qian, H. Counting the Floating Point Operations (FLOPS) [Электронный ресурс]. URL: https://www.mathworks.com/matlabcentral/fileexchange/50608-counting-the-floating-point-operations-flops (дата обращения 22.05.2024).

Study on Absolute Accuracy of Cooperative Cluster Navigation in Satellite Denial Environment

Shanwen Xu[1,2], Qinghua Zeng[1], Cong Yu[2], Fangdong Li[1], Yufen Qiu[2]

(1. Navigation Research Center, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China;2. Jiangxi Hongdu Aviation Industry Group, Nanchang 330001, China)

Abstract: With the development of navigation technology and communication technology, the modern battlefield requires more and more collaboration between clusters. On the one hand, compared with a single aircraft, an aircraft cluster can obtain more battlefield resources to accomplish more complex tasks; on the other hand, compared with a manned aircraft cluster, an unmanned aircraft cluster has the advantages of low cost, high sensitivity, and small size, and unmanned clusters have gradually become the main research direction of modern battlefield cluster coordination. This paper takes the same cluster configuration as the research basis, carries out the cluster arrangement through the combination of inertia/scape matching and inertia/terrain matching, and carries out the grouping test by changing the number of nodes carrying inertia/terrain matching within the cluster, and analyzes the three-dimensional positional accuracy of each node's latitude/longitude height among the clusters through software simulation. It is finally concluded that the combined navigation method of inertia/scape matching and inertia/terrain matching can realize the improvement of absolute position accuracy of clusters under the satellite denial environment, which provides a research idea for the intelligent collaboration of future vehicle clusters.

Keywords: cooperative navigation, satellite denial, unmanned clusters, absolute positional accuracy

1   Introduction

Cooperative navigation is a navigation method to enhance the navigation and positioning accuracy of multiple systems through real-time data interaction, and for the cluster, it is to solve and correct the absolute position accuracy of each node through the information interaction of each node, and each node ensures its own positioning accuracy through the constraints of multiple sets of data to ensure the stability of the whole cluster configuration and the efficiency of task completion in the combat[1-3]. For the environment of normal reception of satellite signals, each node only needs to consider its own absolute positional accuracy to realize the stability of the entire cluster formation[4].once the satellite signal is interfered with, the aircraft relying solely on satellite signals will lose its due function, and the entire cluster will be paralyzed under the environment of wide-scale satellite interference, and the fusion of multiple sensors is needed to guarantee the absolute accuracy of the cluster[5-7]. In summary, the research of cooperative technology has a greater impact on the overall positioning accuracy of the cluster.

In response to the above technical requirements, scholars at home and abroad have conducted a large number of studies. Literature [8] analyzes the sensor module in sub-modules, and measures

the device pose through point data and image data. The sensor management method proposed in literature [9] can effectively avoid resource waste of sensors and realize effective tracking of real targets. Article [10] establishes a model for the intelligent planning of collaborative monitoring tasks for marine multi-platform sensors with wide adaptability in response to maritime events involving multiple platforms. For the detection problem, article [11] proposes a heuristic algorithm that can simultaneously determine the detection arc and the start time of detection based on multiple sensors, which improves the completion rate of the detection task. Article [12] designs a collaborative detection mode for multiple sensors under combat processes for different combat distances. In terms of the scheduling of sensor resources, article [13] achieves the optimal sensor scheduling scheme by predicting the scheduling benefits of sensors; article [14] proposes a multi-sensor collaborative detection resource scheduling algorithm based on a combination of proximal policy optimization and a fully connected neural network. Article [15] proposes a redundant information adaptive federated filtering tracking algorithm with outlier monitoring, which achieves high-precision information tracking by making full use of the redundant information from multiple sensors.

The requirements for clustering in the current project are mainly reflected in three aspects: absolute positioning accuracy, relative navigation accuracy and sensor data fusion. Absolute positioning accuracy can be provided according to the satellite and environmental conditions; relative navigation can be constrained by transmitting the navigation information of each node through the data chain; sensor data fusion mainly takes into account the dimensional difference of multiple sensors' information, which can be ultimately localized in the transmission and fusion of position information. In this paper, the absolute positioning accuracy is improved from the combination of different sensors carried by clusters.

2   Navigation program

Under different mission environments, clusters need to be flexibly equipped with sensors according to different situations , and ensure the absolute positioning accuracy of clusters by fully utilizing the battlefield environment resources. This paper takes the special background with obvious undulating terrain and rich interlaced landform features as the background of the combat environment. By selecting and combining multiple sensors and fully utilizing the environment's own environmental information, multi-sensor fusion can effectively prevent interference with satellite signals. The whole navigation process is that the sensors first explore the task environment information, and then the fusion of multiple sensor data corrects the navigation error through the filtering algorithm to realize the improvement of the absolute navigation accuracy of the cluster.

2.1   Navigation strategies

The research background of this paper is a navigation scheme for studying the absolute navigation accuracy of clusters in areas with obvious topographic features in a satellite-denied environment., the cluster uses the combination of multiple nodes carrying inertia/scape matching and inertia/terrain matching respectively to form a grouping, and ensures the absolute navigation and positioning accuracy of the cluster as a whole through the fusion of the two kinds of navigation methods when crossing the land and sea junctions.

Long-endurance aircraft are mainly equipped with computers, satellite navigation receivers, data links, barometric altimeters, radio altimeters, etc., according to background requirements, and for the high-precision long endurance vehicle, it is also equipped with visual sensors. Satellite receivers can provide high-precision absolute position information for the system, the data chain is used to transmit the position information of each node between clusters, the combination of barometric altimeter and radio altimeter, terrain matching can realize high-precision height measurement, and the scene matching in vision technology can independently complete high-precision absolute navigation. Cluster combat has higher requirements for node information processing than single combat, emphasizing the importance of each node and sensor synergy, and its navigation structure is shown in Figure 1.
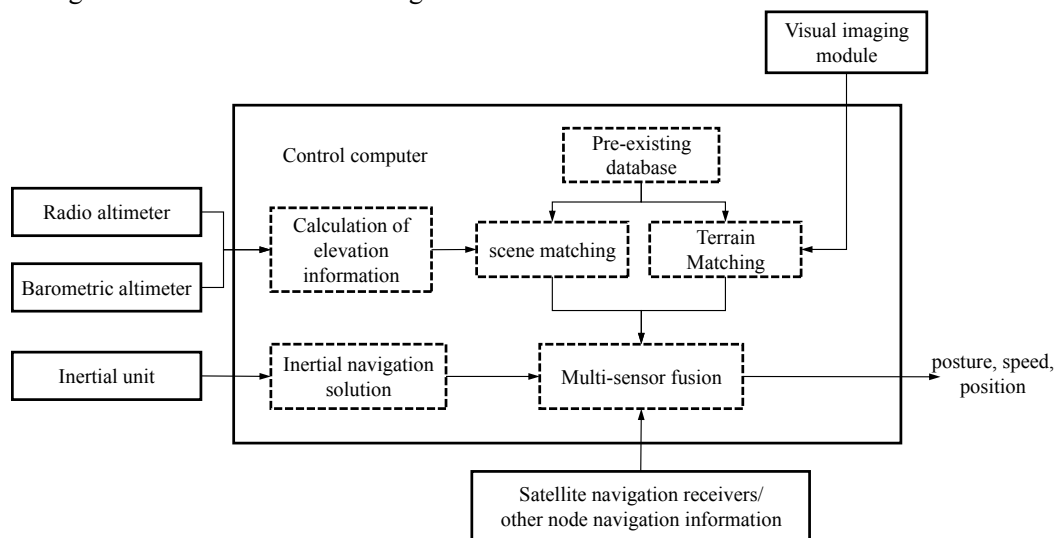
Fig. 1　Block diagram of cluster co-navigation structure

For environments with complex and interwoven terrain and landforms, image matching and terrain matching methods can only be used in this area. When satellite signals are interfered with, a method combining cooperative navigation above the sea surface, cooperative image matching above the target area, and terrain matching is designed. Combining inertial navigation errors and the number of cooperative vehicles, the airspace coverage area for cooperative flight of vehicles is designed, so that the vehicles cooperating in navigation above the sea surface can partially fly reliably to above the target area. The aircraft that fly to the target area first use image matching and terrain matching to locate, and assist other aircraft that have not yet reached the target area to correct their absolute positioning accuracy.

In summary, the priority of vehicle navigation is satellite navigation, image matching, terrain matching and cooperative navigation. Among them, satellite navigation has the highest priority, and the cluster absolutely trusts the navigation information from satellites without interference, and its accuracy can reach the meter level; followed by the visual image matching technology, and with the support of the database, the image matching can provide a positioning accuracy of 50-100 meters; and again, the terrain matching technology, due to the fact that it requires the assistance of the barometric altimeter and the radio altimeter to accomplish it together, and the As it requires the assistance of barometric altimeter and radio altimeter to complete, and has high requirements on heading, its accuracy can reach 200-300 meters; finally, the main function of the cooperative

technology is to measure the distance between nodes and maintain the baseline configuration of the cluster.

2.2 Co-navigation algorithms

In this paper, the error model of the inertial guidance system is used to construct the state equation. Then the state variables are $X = [\ \boldsymbol{\phi}^l \quad \delta \boldsymbol{v}^l \quad \delta \boldsymbol{r}^l \quad \boldsymbol{\varepsilon}^b \quad \nabla^b\ ]^T$, The equation of state is:

$$\dot{X} = FX + GW$$

The measurement equation is:

$$Z(t) = H(t)X(t) + V(t)$$

In the case of a 3-machine cluster, for example, the measurement matrix is:

$$H = [\mathbf{0}_{1\times 3} \quad \mathbf{0}_{1\times 3} \quad H_{1\times 3}^{\rho 12} \quad \mathbf{0}_{1\times 3} \quad \mathbf{0}_{1\times 3}]$$

V is the measurement noise associated with the white noise of the data chain ranging and the 2-node's own localization accuracy. The matrix $H_{1\times 3}^{\rho 12}$ is:

$$H_{1\times 3}^{\rho 12} = \left[\frac{(P_{x1}^e - P_{x2}^e)}{r_{12}^e} \quad \frac{(P_{y1}^e - P_{y2}^e)}{r_{12}^e} \quad \frac{(P_{z1}^e - P_{z2}^e)}{r_{12}^e}\right] T_1$$

Similarly, when 2 each other node collaborates with 1 node through the data chain, the measurement equation is:

$$Z = HX + V$$

where the measurement matrix is:

$$H = \begin{bmatrix} \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & H_{1\times 3}^{\rho 12} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} \\ \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & H_{1\times 3}^{\rho 13} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} \end{bmatrix}$$

V is the measurement noise. The matrices $H_{1\times 3}^{\rho 12}$ and $H_{1\times 3}^{\rho 12}$ are:

$$H_{1\times 3}^{\rho 12} = \left[\frac{(P_{x1}^e - P_{x2}^e)}{r_{12}^e} \quad \frac{(P_{y1}^e - P_{y2}^e)}{r_{12}^e} \quad \frac{(P_{z1}^e - P_{z2}^e)}{r_{12}^e}\right] T_1$$

$$H_{1\times 3}^{\rho 13} = \left[\frac{(P_{x1}^e - P_{x3}^e)}{r_{13}^e} \quad \frac{(P_{y1}^e - P_{y3}^e)}{r_{13}^e} \quad \frac{(P_{z1}^e - P_{z3}^e)}{r_{13}^e}\right] T_1$$

where the variables r_{12}^e and r_{13}^e are:

$$r_{12}^e = \sqrt{(P_{x1}^e - P_{x2}^e)^2 + (P_{y1}^e - P_{y2}^e)^2 + (P_{z1}^e - P_{z2}^e)^2}$$

$$r_{13}^e = \sqrt{(P_{x1}^e - P_{x3}^e)^2 + (P_{y1}^e - P_{y3}^e)^2 + (P_{z1}^e - P_{z3}^e)^2}$$

In the above equation, $P_{x1}^e, P_{y1}^e, P_{z1}^e$ is the triaxial position of node 1 in the Earth system, $P_{x2}^e, P_{y2}^e, P_{z2}^e$ is the triaxial position of node 2 in the Earth system, and $P_{x3}^e, P_{y3}^e, P_{z3}^e$ is the triaxial position of node 3 in the Earth system.

Considering that the traditional KALMAN filtering computation process will involve huge amount of arithmetic matrix operations and inverse operations of high-dimensional matrices. In order to reduce the amount of filtering computation, the sequential filtering algorithm is used for the filtering measurement update, which decomposes the measurement update process of N-dimensional observations into N sub-measurement updates, and inverts the matrix into the inverse of scalar operations. On the other hand, when the number of observation dimensions decreases

during the filtering process, the sequential filtering can avoid the effect of filtering instability due to matrix dimension switching.

The flowchart of the overall algorithm for cluster co-navigation is shown in Fig. 2:
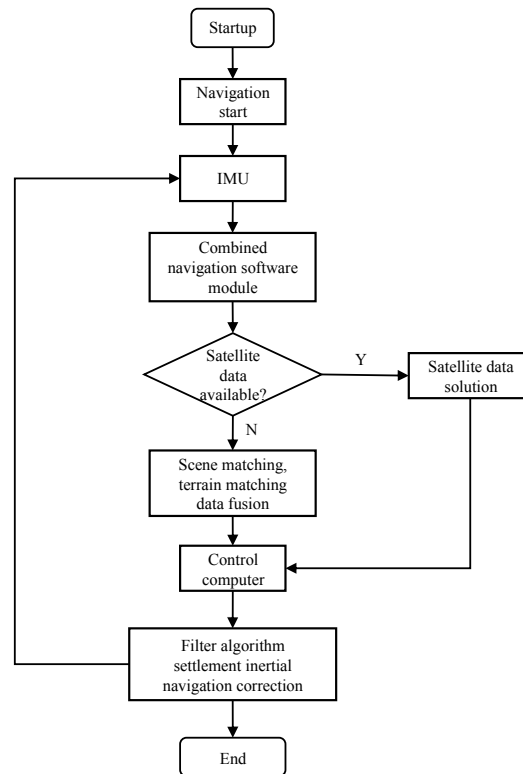
Fig. 2 Flowchart of the overall algorithm for collaborative navigation

3   Experimental program design and result analysis

3.1   Simulation program design

A combination of two/three leader planes and one slave plane is designed for simulation, and the leader plane carries a view-matching sensor and the slave plane carries a terrain-matching sensor. Among them, the zero bias of the inertial gyro is 5°/h (3σ) and the stability of zero bias is 3°/h (1σ); the zero bias of the accelerometer is 1mg (3σ) and the stability of zero bias is 1mg (1σ). The positioning accuracy of the leader is 50m (1σ) in horizontal single axis and vertical direction. The accuracy of data chain ranging is 30m (1σ).

3.2   Simulation results

(1) Based on a single simulation with 2 leaders and 1 slave, the absolute positioning error of the slave through cooperative navigation is shown in Fig. 3:
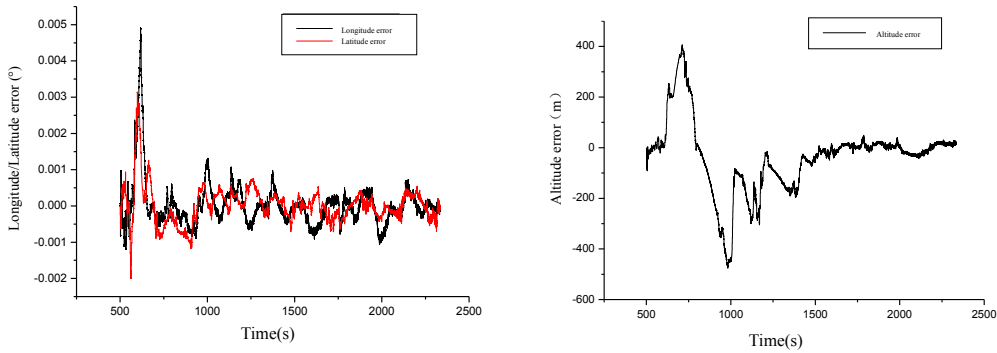
Fig. 3   Absolute positioning error of the first set of cluster slaves

Horizontal position uniaxial error is concentrated within 200m, with a maximum value of 450m. Height error is concentrated within 300m, with a maximum value of 450m.

(2) Based on a single simulation with 3 leaders and one slave, the absolute positioning error of the slave through cooperative navigation is shown in Fig. 4:
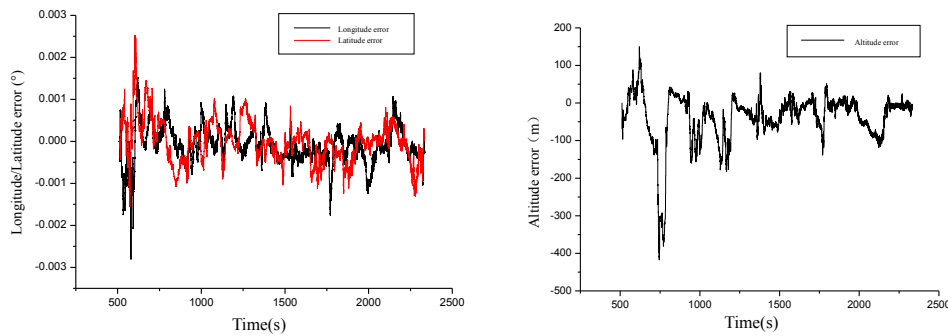
Fig. 4   Absolute positioning error of the second group of cluster slaves

Horizontal position uniaxial error is concentrated within 150m, with a maximum value of 300m. Height error is concentrated within 200m, with a maximum value of 400m.

3.3   Analysis of simulation test results

Simulation scenario	Absolute positioning accuracy	
	Horizontal	Vertical
2 leaders + 1 slave	Concentrated distribution within 200m, maximum 450m	Concentrated distribution within 300m, maximum 450m
3 leaders + 1 slave	Concentrated distribution within 150m, maximum 300m	Concentrated distribution within 200m, maximum 400m

By matching different navigation sensors, the experimental conclusion is that increasing the number of carrier view-matched collars in the cluster can better improve the absolute positioning accuracy of the cluster.

4   Conclusion

This paper describes three main issues of cooperative cluster navigation: absolute positioning accuracy, relative positioning accuracy and multi-sensor data fusion. From a practical point of view, the cluster cooperative navigation scheme for operating in complex terrain and difficult terrain in a

satellite-denied environment is analyzed. and it is finally determined to use a combination of inertia/scenery matching and inertia/terrain matching. In this paper, we choose to analyze the implementability of the scheme from the perspective of improving the absolute navigation and positioning accuracy of clusters, and design a navigation scheme that matches a variety of sensor combination methods within a single cluster, and carry out performance demonstration through software simulation. The experimental results show that the navigation scheme can provide high absolute navigation and positioning accuracy for the cluster. The shortcomings of this paper are the small number of nodes in the cluster and the lack of demonstration of the navigation scheme for large clusters. In the future, we consider adding a decision model to the cluster formation to test whether a larger volume of clusters can realize the improvement of the overall absolute positioning accuracy of the cluster by this scheme.

References:

[1] WEI Shuai-ying, DU Yu-long, HU Bo, et al. Develop Status and Analysis of UAV Swarm Cooperative Navigation Technology in GNSS-denied Environment[J]. NAVIGATION AND CONTROL. 2023,22(04):5-16.

[2] XU Xiao-wei, LAI Ji-zhou, LV Pin. et al. A Literature Review on the Research Status and Progress of Cooperative Navigation Technology for Multiple UAVs[J]. Navigation Position and Timing, 2017,4(04):1-9.

[3] Liu Shuangxi, Xu Xiaoping, Huang Wei, et al. Development and Key Technologies of Foreign Multi-Missile Collaborative Projects[J]. Aero Weaponry, 2024,31(X).

[4] LUAN Houbin, PANG Zhichao, FU Jinlin. Radio ranging collaborative positioning method with UKF in unmanned cluster[J]. SHIP SCIENCE AND TECHNOLOGY, 2024,46(16):186-189.

[5] LEI Ming-bing, LIU Wei-peng, SONG Zhen-hua, et al. Overview and Prospect in Inertial Technology of Air Defense Missile[J]. NAVIGAION AND CONTROL, 2020,19(Z1):88-95.

[6] WANG Wenlong, WANG Xiaofang, LIU Hai, et al. Multi-missiles collaborative navigation and formation optimization design in design in denial environment[J]. FLIGHT DYNAMICS, 2022,40(06):61-71.

[7] WANG Shuai. Design og3D measurement system with multisensory coordination[D]. Zhejiang University, 2018.

[8] WANG Binglu, JIN Yang, ZHANG Lei, et al. Collaborative Perception Method Based on Multisensor Fusion[J]. Journal of Radars, 2024,13(01):87-96.

[9] PANG C, SHAN G L, DUAN X S. Management method for multiple sensors' recognizing and tracking multiple targets cooperatively[J].Journal of Beijing University of Aeronautics, 2019,45(08):1674-1680.

[10] HOU Rui, CHENG Yuting, LI Hui, et al. The AI Planning Technology of Marine Multi-Platform and Multi-Sensor Cooperative Monitoring Task[J]. MARINE INFORMATION, 2020,35(03):11-19.

[11] LONG Xi, CAI Weiwei, YANG Leping. Multi-sensor cooperative planning of space objective detection[J]. JOURNAL OF NATIONAL UNIVERSITY OF DEFENSE TECHNOLOGY, 2024,46(04):37-44.

[12] LI Taotao, GAO Weiliang, WANG Yongkun, et al. Design and simulation of carrier-borne fighter multi-sensor co-detection model[J]. Command Control and Simulation, 2023,45(05):36-42.

[13] ZHANG Yunpu, SHAN Ganlin, FU Qiang. Scheduling method for multi-sensor cooperative ground target tracking[J]. JOURNAL OF XIDIAN UNIVERCITY, 2022,49(05):100-108.

[14] WANG Meng-Qian, LIANG Hao-Xing, GUO Mao-Yun, et al. Resource Scheduling Method of Multi-sensor Cooperative Detective for Flying Targets[J]. ACTA AUTOMATICA SINICA, 2023,49(06):1242-1255.

[15] LIU Jinming, ZHANG Biling, ZHANG Yuyan. Adaptive Federated Filtering Tracking Algorithm Based on Multi-Sensor Redundant Data Cooperative[J]. Journal of Beijing University of Posts and Telecommunications, 2023,46(04):21-26.

Solving The Problem of Relative Navigation Using Deep Neural Networks

V.G. Karaulov[1], A.M. Gruzlikov[2], M. D. Korableva[2]

(1. ITMO University, Concern CSRI Elektropribor, JSC, Saint Petersburg; 2. Concern CSRI Elektropribor, JSC, Saint Petersburg)

Abstract: The problem of relative navigation of an underwater vehicle (UV) for bringing it to a bottom station (BS) is considered. Within the deterministic approach, the statement of the corresponding problem of estimating the UV relative coordinates and orientation angles is presented. The stated problem is solved using the least squares method. To reduce the a priori knowledge uncertainty of the coordinates and orientation angles, it is proposed to use the architecture of deep neural networks. Based on the performed simulation, the root-mean-square errors (RMSE) of estimating the UV relative coordinates and orientation angles are compared using various methods of training fully connected neural networks.

1   Introduction

The paper deals with the problem of determining the coordinates and orientation angles of an underwater vehicle (UV) for solving the problem of bringing it to the bottom station (BS).

The hydroacoustic system of the UV (Fig. 1) includes receiving antennas located under the vehicle keel in the stern and bow parts, and emitters installed at the edges of the base and performing non-directional emission of a signal. It should be noted that the emitted signals are tone signals of the same frequency, the emission of which is spaced in time so that the source of the emission could be uniquely identified by the signal detection period.

The proposed paper continues the authors' works [3-5]. The purpose of the paper is to rectify the drawback associated with the great expenditure of computer memory resources. This is due to the fact that it is necessary to store a large amount of data in the form of a kd-tree used in the K-nearest neighbors method. The present paper is aimed at correcting this problem by using a deep neural network.
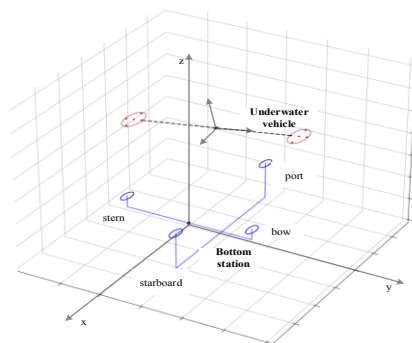
Fig. 1   Layout of UV and BS hydroacoustic system

2   Statement of deterministic problem of estimation

Onboard the UV, there is a group of n hydrophones operating on a single time scale and located in the bow and stern parts. Besides, the BS has m sonar beacons located at the edges (Fig. 1). The beacons emit a tone signal in accordance with the known emission cyclogram, and it is worth noting that the time scales of the UV and the BS do not match.

We will follow the article [2] and record the measurements of pseudoranges between beacons and hydrophones, which are generated by the UV:

$$y_j^i = \hat{c}_0 T_j^i = d_j^i + \delta + v_j^i + \Delta c T_j^i, \, j = \overline{1,n}, i = \overline{1,m}, \tag{1}$$

where i is the beacon number, j is the hydrophone number, d_j^i are the range true values, v_j^i are the interference, δ is the pseudorange measurement error due to the time scale discrepancy (a pseudorandom variable), T_j^i are the measured values of the time of signal arrival from the beacons to the hydrophones, $\hat{c}_0$ is the a priori estimate of the sound speed in the water, Δc is the error of the speed of sound propagation in the water. The interference v_j^i are mutually uncorrelated centered normally distributed random variables with standard deviations σ_j^i. We know the coordinates $x^i, y^i, z^i, i = \overline{1,m}$ of the beacons in the BS-fixed rectangular axes x, y, z, as well as the coordinates $\tilde{x}_j, \tilde{y}_j, \tilde{z}_j, j = \overline{1,n}$ of the hydrophones in the UV-fixed rectangular axes $\tilde{x}, \tilde{y}, \tilde{z}$. Orientation of the UV relative to the BS-fixed coordinate frame will be determined using the heading angles K, trim ψ and roll θ. The center of the UV-fixed coordinate frame is its control point, i.e. the coordinates of the control point in this coordinate frame are $\tilde{x}_0 = \tilde{y}_0 = \tilde{z}_0 = 0$, and in BS-fixed coordinate frame – x_0, y_0, z_0. The a priori information about $x_0, y_0, z_0, K, \psi, \theta, \delta$ is not available.

Let us introduce the designations: $s^i = (x^i, y^i, z^i), i = \overline{1,m}$ – for beacon coordinate vectors in the BS-fixed coordinate frame; $\tilde{p}_j = (\tilde{x}_j, \tilde{y}_j, \tilde{z}_j), j = \overline{1,n}$ – for hydrophone coordinate vectors in the UV-fixed coordinate frame; $p_j = (x_j, y_j, z_j), j = \overline{0,n}$ – for coordinate vectors of the control point and hydrophones in the BS-fixed coordinate frame; $A(K, \psi, \theta)$ – for the matrix of transition [1] from axes $\tilde{x}, \tilde{y}, \tilde{z}$ to axes x, y, z. Note that $p_0 = (x_0, y_0, z_0)$.

Following the article [2], we can write down the formulas for ranges being measured:

$$d_j^i = \left| p_j - s^i \right| = \left| p_0 + A(K, \psi, \theta) \tilde{p}_j - s^i \right|. \tag{2}$$

Using measurements (1) and taking into account (2), it is necessary to estimate the vector p_0 of the UV control point coordinates in the BS-fixed coordinate frame, and the angles K, ψ, θ of the UV orientation relative to this coordinate system, as well as the error in measuring pseudoranges δ and the error of the speed of sound propagation in the water Δc.

3   Fully connected neural network

It is proposed to use a fully connected architecture of a deep neural network to reduce the area of expected knowledge uncertainty of relative coordinates and orientation angles for the first approximation when solving the estimation problem. The feedforward network will be trained on a

data set that is formed on the basis of a geometric model – the preset position of the beacons of an underwater base station and the coordinates of UV receiving elements on a specified coordinate grid and orientation angles table. The training sample is formed according to the following rule:

Let the training sample be formed according to the following rule:

- $(t_a^j - t_b^j)_{j \in 1...4}$ – the difference in the time of signal reception between the pairs of maximally spaced receivers (a, b) with the fixed beacon;

- $(t_i^a - t_i^b - T^{a,b})_{j \in 1...10}$ – the difference in the time of signal reception between the pairs of beacons (a, b) by the fixed receivers and the shift in the time of signal emission by the beacons (with the fixed value $T^{a,b}$).

Thus, the input layer will consist of 100 neurons, the output layer – of 6 neurons. After each hidden layer, the ReLU (Rectified Linear Unit) activation function is applied, and after the last one the activation function is not available.

4 Simulation modeling and obtained results

Simulation modeling of the UV motion down to the BS from the point (0, 0, 50) m was carried out with clockwise rotation about the vertical axis of the UV and with zero angles of roll and pitch. Based on the modeling results, estimates of the root-mean-square errors (RMSE) of the UV coordinates and heading angle were obtained with the help of fully connected neural network architectures using various methods of their training (see Table 1).

Tab.1 Comparison of RMSE of coordinates and heading angle estimates for 5 models with different training methods

		RMSE X, m	RMSE Y, m	RMSE Z, m	RMSE K, °
Model 1 (5x100)	Without BN, L2 = 0	1.21	1.36	4.61	4.7
	With BN, L2 = 0	1.39	1.67	5.45	4.78
	With BN, L2 = 0.001	0.85	1.48	6.84	5.2
	With BN, L2 = 0.01	0.8	1.27	6.2	5.09
	With BN, L2 = 0.1	1.07	1.02	20.5	5.45
Model 2 (10x100)	Without BN, L2 = 0	1.61	2.09	4.78	4.55
	With BN, L2 = 0	1.97	1.14	5.18	5.05
	With BN, L2 = 0.001	1.88	0.91	4.94	4.39
	With BN, L2 = 0.01	1.82	1.8	5.78	5.07
	With BN, L2 = 0.1	0.7	0.61	14.43	5.85
Model 3 (15x100)	Without BN, L2 = 0	1.4	1.84	3.65	4.67
	With BN, L2 = 0	1.59	1.34	5.98	4.99
	With BN, L2 = 0.001	3.34	1.64	4.52	4.89
	With BN, L2 = 0.01	3.99	3.54	5.69	4.85
	With BN, L2 = 0.1	-	-	-	-

Tab.1   (continued)

		RMSE X, m	RMSE Y, m	RMSE Z, m	RMSE K, °
Model 4 (100-500-250-50)	Without BN, L2 = 0	1.75	0.99	4.54	3.89
	With BN, L2 = 0	1.68	2.21	5.21	4.55
	With BN, L2 = 0.001	1.35	2.3	6.41	4.57
	With BN, L2 = 0.01	2.76	2.5	7.93	5.27
	With BN, L2 = 0.1	0.52	0.7	18.85	5.62
Model 5 (500-1000-3000-6000-4000-1000-500-250-100)	Without BN, L2 = 0	1.66	1.73	4.36	4.9
	With BN, L2 = 0	1.35	1.03	4.98	4.62
	With BN, L2 = 0.001	0.46	2.56	5.9	4.59
	With BN, L2 = 0.01	1.07	0.82	5.09	5.5
	With BN, L2 = 0.1	-	-	-	-

The following configurations of fully connected neural networks were analyzed: with batch normalization (BN), with L2 regularization (the table shows the values of the regularization parameter), and 5 network structures, where the first three models have all hidden layers with 100 neurons, and the 4th and 5th models have the structure given in brackets (the number of neurons in each hidden layer is written with a hyphen).

The results show that the use of regularization worsens the prediction accuracy, while the 4th and 5th models with a regularization parameter of 0.1 could not learn at all. The use of BN also did not improve the model. As a result, the best accuracy was shown by Model 4, without L2 regularization and BN. Estimates for the roll and pitch angles are not presented, because each model has estimated them with a RMSE of no more than 0.01 degrees.

5   Conclusion

The paper formulates the problem of estimating the relative coordinates and determining the orientation angles of the UV based on signals from hydroacoustic beacons within the deterministic approach. The authors show the possibility of using a deep neural network in solving this problem at the first stage of narrowing the area of a priori knowledge uncertainty of the relative coordinates and orientation angles.

References:

[1] Emel'yantsev, G.I. and Stepanov, A.P., Integrirovannye inertsial'no-sputnikovye sistemy orientatsii i navigatsii (Integrated INS/GNSS Systems of Orientation and Navigation), St. Petersburg: Concern CSRI Elektropribor, 2016.

[2] Koshaev, D.A., AUV relative position and attitude determination using acoustic beacons, Gyroscopy and Navigation, 2022, vol. 13, pp. 262–275.

[3] Gruzlikov, A.M., Short and ultra-short baseline navigation of the AUV for bringing it to the bottom docking device, Proceedings of the 29th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS), 2022, pp. 1–3.

[4] Gruzlikov, A.M., Karaulov, V.G., Mukhin, D.A., and Shalaev, N.A., Results of testing the positioning algorithm and determining the orientation of the underwater vehicle based on data from hydro-acoustic beacons, Izvestiya YuFu. Tekhnicheskie nauki, 2023, no. 1, pp. 265-275.

[5] Karaulov, V.G., Gruzlikov, A.M., and Zolotarevich, V.P. Recursive algorithm for AUV positioning and

determining its orientation, Proceedings of the 30th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS), 2023, pp. 1-4.

Recursive Iterative Batch Linearized Smoother for Navigation Estimation Problems

O . A . Stepanov, Yu. A . Litvinenko, A . M . Isaev

(Concern CSRI Elektropribor, ITMO University, St. Petersburg, Russia)

Abstract: A recursive iterative batch linearized smoother is described for the general case of solving nonlinear problems with nonlinearities both in the equations for the shaping filter and measurement equations. The efficiency of its application is analyzed by the example of solving practical navigation estimation problems.

Algorithms intended to solve nonlinear problems of navigation information processing are in most cases designed within the framework of the Bayesian stochastic approach [1-9], which is due to the need to obtain not only an estimate of the vector of unknown parameters, but also the corresponding calculated accuracy characteristic. In fact, in this case, the solution of the filtering problem is aimed to find a certain set of parameters describing the type of the posterior probability density function (PDF). Usually, the mathematical expectation calculated with the help of this set is used as an estimate at the algorithm output, and the root-mean-square error of the estimate, as an accuracy characteristic. If the calculated and actual values are consistent, the algorithm is called consistent [10-12].

One of the main requirements for the algorithms intended to solve problems in real time is low computational complexity, which makes it possible not to impose significant requirements when they are implemented in an onboard computer. In this regard, filtering algorithms are often designed using a recursive scheme, which implies sequential processing of measurements, one after another, and it is important that after the measurement was processed at the current step, it is no longer used. In so doing, it is necessary to keep the set of parameters, describing the type of the posterior PDF, from step to step.

Recursive Kalman-type algorithms (KTA), in which the posterior PDF is replaced at each step by its Gaussian approximation, are the most widespread in solving practical problems. Such replacement provides easy implementation of a computationally simple recursive algorithm since, from step to step, it is sufficient to keep only two parameters – the mathematical expectation and the covariance matrix of estimation errors, which fully describe the Gaussian approximation of the PDF. In many cases, when the form of the posterior PDF is close to the Gaussian one, KTAs prove to be consistent and allow obtaining an accuracy estimate close to the optimal one in the RMS sense [3-8]. However, when the form of the posterior PDF is substantially different from the normal one, the errors caused by its approximation will be significant. Due to the recursive nature of the algorithm, these errors will accumulate, which may eventually lead to its inoperability.

To solve such problems, the type of the posterior PDF in the filtering algorithm being

---

designed can be described with the use of a significant set of parameters that retains its complex form, and the recursive nature of the algorithm can be organized based on the known recursive relation [6, 13-18]. A weighted set of delta functions can serve as a set of parameters. However, the algorithms designed in this way mostly turn out to be computationally complicated and difficult to implement in onboard computers, which makes them unsuitable for real-time problems.

There are some problems in which the posterior density, being multiextremal at initial moments of time, takes a single-extremal form close to the Gaussian one in the process of its evolution. Obviously, recursive KTAs will be inoperative for their solution. However, as was shown in [6,19-23], nonrecursive KTAs can provide the accuracy of optimal estimation and be consistent, starting from the moment when the posterior PDF becomes close to the Gaussian one. The input of such algorithms is a batch of all measurements accumulated by the current moment of time, which is why they are usually called batch algorithms [21-23]. For all their advantages, their computational load remains rather high, which is primarily due to the need to invert high-dimensional matrices.

To overcome this drawback, in [23], the authors propose an algorithm called the recursive iterative batch linearized smoother, hereinafter referred to as "Smoother". It is essentially a nonrecursive algorithm, but at the same time, it has the advantage of the algorithms constructed with the use of the recursive scheme, namely, low computational load. This is made possible due to the recursive procedure, which is used, instead of inverting a high-dimensional matrix at each iteration, to find the necessary estimates by solving a smoothing problem. Paper [23] gives only a brief description of such an algorithm, without its formula dependencies and without going into the details of its implementation.

The aim of the paper is to describe in detail the proposed smoother for the general case of the problem whose statement is formulated within the framework of the Bayesian stochastic approach and nonlinearities are contained both in the dynamic equations and equations of measurements. The advantages of the smoother are demonstrated by solving practical problems of navigation information processing.

References:

[1] Kalman, R.E., A new approach to linear filtering and prediction problems, Trans. ASME. Series D, J. Basic Engineering, 82, 1960, pp. 35-45.

[2] Stratonovich, R.L., Conditional Markov processes, in Teoriya veroyatnostei i eyo prilozheniya (Probability Theory and its Applications), 1960, vol. 5, no. 2, pp. 172-195.

[3] Jazwinski, A.H., Stochastic Process and Filtering Theory, New York: Academic Press, 1970.

[4] Gelb, A., Applied Optimal Estimation, M.I.T. Press, Cambridge, 1974.

[5] Dmitriev, S.P., Vysokotochnaya morskaya navigatsiya (High-accuracy Marine Navigation). St.Peterburg: Sudostroyeniye, 1991.

[6] Stepanov, O.A ., Primenenie teorii nelineinoi fil'tratsii v zadachakh obrabotki navigatsionnoi informatsii (Application of Nonlinear Filtering Theory for Navigation Data Processing), 3rd edn., St. Petersburg: Elektropribor, 2003.

[7] Simon D. Optimal State Estimation: Kalman H∞ and Nonlinear Approaches, New Jersey, NJ: John Wiley & Sons, Inc., 2006.

[8] Särkkä S., Bayesian Filtering and Smoothing, Cambridge University Press, 2013.

[9] Stepanov, O .A ., Osnovy teorii otsenivaniya s prilozheniyami k zadacham obrabotki navigatsionnoi informatsii (Fundamentals of theEstimation Theory with Applications to the Problems of Navigation Information Processing), 3rd edn., revised and supplemented, Part 1, Vvedenie v teoriyu fil'tratsii (Introduction to the Filtering Theory), St. Petersburg: TsNII Elektropribor, 2017.

[10] Bar–ShalomY., Li, X., and Kirubarajan, T., Estimation with Applications to Tracking and Navigation, New York, Wiley–Interscience, 2001.

[11] Bolotin, Yu.V., Bragin, A.V., and Gulevskii, D.V., Studying the consistency of the extended Kalman filter in the problem of pedestrian navigation with an SINS fixed on his feet, Giroskopiya i Navigatsiya, 2021, vol. 29, no. 2 (113), pp. 59-77.

[12] Stepanov, O.A. and Isaev, A.M., Methodology of comparative analysis of recursive nonlinear filtering algorithms in problems of navigation information processing based on predictive simulation, Giroskopiya iNavigatsiya, 2023, vol. 31, no. 3 (122), pp. 48-65. EDN: MVWKGC

[13] Busy, R.S., Nonlinear filtering theory, IEEE Trans on Automatic Control, 1965

[14] Doucet, A., Freitas, N. De., and Gordon, N., Sequential Monte Carlo Methods in Practice, New York, NY, USA: Springer-Verlag, 2001.

[15] Chen, Z., Bayesian Filtering: From Kalman Filters to Particle Filters, and Beyond. Adaptive Systems Laboratory Technical Report, McMaster University, 2003, Hamilton.

[16] Sarkka, S., Vehtari, A., and Lampinen, J. Rao-Blackwellized particle filter for multiple target tracking, Information Fusion, 2007, vol. 8 (1), pp. 2-15

[17] Stepanov O.A. and Toropov A.B., Application of sequential Monte Carlo methods using analytical integration procedures in processing navigation information, XII Vserossiiskaya soveschanie po problemam upravleniya VSPU-2014 (XII All-Russian Conferenceon Control Problems VSPU-2014, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia, 2014, pp. 3324-3337.

[18] Stepanov O.A. and Toropov A.B. Primeneniye posledovatel'nykh metodov Monte-Karlo s ispol'zovaniyem protsedur analiticheskogo integrirovaniya pri obrabotke navigatsionnoy informatsii. XII Vserossiyskoye soveshchaniye po problemam upravleniya VSPU-2014. Institut problem upravleniya im. V.A. Trapeznikova ran. 2014. S. 3324-3337.

[19] Dunik, J., Straka, O., Matousek, J., and Brandner, M., Accurate Density-Weighted Convolution for Point-Mass Filter and Predictor. Accepted for IEEE Transactions on Aerospace and Electronic Systems, 2021

[20] Vasiliev V.A., Litvinenko Y.A., Stepanov O.A., Toropov A.B. Sravneniye rekurrentnykh i nerekurrentnykh algoritmov fil'tratsii pri reshenii zadachi navigatsii po geofizicheskim polyam // V sbornike: XIV Vserossiyskaya mul'tikonferentsiya po problemam upravleniya MKPU-2021. materialy XIV mul'tikonferentsii: v 4 t. Rostov-na-Donu, 2021. P. 125-127.

[21] O.A. Stepanov, A.M. Isaev, Y.A. Litvinenko Sravneniye rekurrentnykh i nerekurrentnykh skhem postroyeniya algoritmov,osnovannykh na linearizatsii, pri reshenii odnogo klassa zadach nelineynogo otsenivaniya // Trudy XIV Vserossiyskogo soveshchaniya po problemam upravleniya. M.: Institut problem upravleniya imeni V.A. Trapeznikova RAN, 2024

[22] O.A. Stepanov, Y.A. Litvinenko, A.M. Isaev Iterative Batch Linearized Smoother for the AUV Collaborative Navigation. Materialy XXXI Mezhdunarodnoy nauchno-tekhnicheskoy konferentsii «Sovremennyye tekhnologii v zadachakh upravleniya, avtomatiki i obrabotki informatsii» P. 117, 2024.

[23] Stepanov, O.A., Litvinenko, Y.A., and Isaev, A.M., Comparison of recursive and nonrecursive linearization-based algorithms for one class of nonlinear estimation problems. 10th International Conference on Control, Decision and Information Technologies (CoDIT), 2024 (in press).

[24] Stepanov, O.A., Zolotarevich V.P., Litvinenko Y.A., Isaev A.M., Motorin A.V. Recursive and Nonrecursive Algorithms Applied to Navigation Data Processing: Differences and Interrelation with Factor Graph Optimization Algorithms // 31st Saint Petersburg International conference on Integrated Navigation Systems, 2024.

[25] Stepanov O. A., Litvinenko Yu. A., Isaev A. M. Comparative Analysis of Quasi-Linear Kalman- Type Algorithms in Estimating a Markov Sequence with Nonlinearities in the System and Measurement Equations, Mekhatronika, Avtomatizatsiya, Upravlenie, 2024, vol. 25, no. 11, pp. 585—595 (in Russian).

Comparison of Bayesian Estimation and Factor-Graph Optimization Methods

O . A . Stepanov, Yu. A . Litvinenko, A.V. Motorin, A . M . Isaev

(Concern CSRI Elektropribor, ITMO University, St. Petersburg, Russia)

Abstract: Possibilities and features of designing estimation algorithms based on factor-graph optimization in problems of navigation information processing//navigation processing problems are discussed. The relationship and differences between the Bayesian approach and factor-graph optimization algorithms are considered. It is shown that the main feature of factor-graph optimization algorithms is their non-recursive nature due to batch measurement processing, which is also practiced in the design of algorithms within the Bayesian approach. The differences between the algorithms designed with the use of these two approaches may lie in different ways of factorization of the posterior probability density function (PDF). Examples of recursive and non-recursive estimation algorithms designed with the use of factor- graph optimization and the Bayesian approach are given, including those for the problem of simultaneous localization and mapping.

1 Introduction

Algorithms based on the Bayesian estimation theory are widely used to solve problems of navigation information processing. Significant progress has been made in the field of designing such algorithms, especially with respect to linear estimation problems that are solved by means of recursive Kalman-type algorithms, which are essentially universal in nature. As for nonlinear estimation problems, there is no such a universal algorithm, which is why designing effective estimation algorithms, including those applicable to navigation applications, is still relevant. In this regard, algorithms based factor-graph optimization (FGO) have recently received considerable attention [1 – 15]. Currently, FGO algorithms are actively implemented in the field of navigation information processing, especially in robotic systems [1,3,5,7 – 15], in particular to solve problems of simultaneous localization and mapping (SLAM) [2,4,5,10 – 12], including the cases involving information on geophysical fields. Among the merits of FGO algorithms are their higher accuracy [1,3,7,9] compared to the extended Kalman filter (EKF) and low computational complexity compared to traditional extremum search methods [2,7]. However, the reasons behind the mentioned merits have not been given proper consideration. At the same time, it is desirable to identify these reasons and compare FGO algorithms not only with EKF, but also with other Bayesian algorithms. Such an analysis is possible due to the fact that FGO algorithms rely on factorization of the conditional (posterior) PDF of the estimated parameters, which is the basis for designing optimal Bayesian estimation algorithms. This analysis will make it possible to identify the range of navigation information processing problems for which the use of FGO algorithms is most effective, as well as problems where no significant gains can be expected as compared to Bayesian algorithms.

Thus, this work is devoted to the comparison of algorithms designed on the basis of FGO and

Bayesian methods in problems of navigation information processing problems. The first part of this study deals with linear and nonlinear problems in a general formulation, while the second part considers a SLAM problem.

2 Comparison of methods

Many navigation estimation problems are reduced to a linear Gaussian formulation, that is, estimation of the Gaussian state vector given by a linear shaping filter based on its linear measurements with white-noise Gaussian error. In this case, from the standpoint of the Bayesian approach, the Kalman filter is an optimal solution to the filtering problem. It is also possible to design various smoothing algorithms in the Bayesian approach, in particular the Rauch-Tung-Stribel (RTS) smoothing algorithm. It can be shown that solution of the same problem in terms of FGO methods leads to the algorithms that replicate RTS smoothing [16]. This is a consequence of the factor graph triviality in such a problem, as well as the fact that in this case, the maximum of posterior density coincides with its mathematical expectation. Note that FGO-based algorithms are precisely smoothing algorithms, and it is inappropriate to compare them with filtering algorithms such as the Kalman filter.

Another broad class of problems is represented by nonlinear estimation problems. From the Bayesian approach standpoint, these problems can be solved with the use of linearization, for example, by expanding nonlinear functions in a Taylor series. This leads to suboptimal algorithms, such as EKF and smoothing EKF. When algorithms are designed on the basis of FGO, the same linearization methods as in EKF may be often used, while the factor graph of such a problem remains trivial. However, the EKF implements a recursive processing scheme with linearization for each measurement, while in FGO algorithms, linearization is carried out based on the processing results of the whole accumulated batch of measurements. The latter case is preferential, since it reduces linearization errors due to the fact that the posterior PDF is normalized with the accumulation of measurements, thus ensuring its better approximation. However, the Bayesian approach also allows for nonrecursive algorithms to process a batch of measurements in which the linearization point is adjusted similarly to the FGO based on the results of batch measurement processing, and, moreover, the algorithms can also be designed using other, more accurate, methods for representing the posterior density [15, 17, 18].

The SLAM problem stands apart among nonlinear navigation processing problems. Direct application of standard Bayesian algorithms to solve such a problem leads to an EKF with a state vector of rather high dimension, since the state vector must include all constants describing the map. The covariance matrix of such a vector is filled, which makes the EKF in the SLAM problem computationally complicated. When approaching SLAM from the FGO standpoint, it turns out that it is possible to design a nontrivial factor graph for each of these problems, which corresponds to a special factorization of the posterior density which is different from the traditional one. As a result, this makes it possible to design an algorithm that solves the problem in stages, first, the localization problem, and then, the mapping problem. In this case, there is no need to calculate the full/whole??? covariance matrix for the entire state vector, which significantly reduces the computational complexity of the algorithm [19].

3 Conclusion

It is shown that FGO-based algorithms are smoothing algorithms that maximize the posterior PDF, which in a linear problem is generally equivalent to Bayesian estimation algorithms. Thus, it is impossible to conclude that the accuracy of FGO algorithms in such problems can be improved and computational complexity reduced relative to Bayesian algorithms.

In nonlinear problems, the gain in accuracy of FGO algorithms as compared to recursive Bayesian algorithms is due to the use of a batch of measurements. This, in turn, with the accumulation of measurements, leads to normalization of the posterior distribution density function and, as a consequence, to its better approximation. It is noted that within the framework of the Bayesian approach, it is easy to obtain algorithms similar to the FGO algorithms. Moreover, based on the Bayesian approach, it is possible to design algorithms using other, more accurate, methods for representing the posterior density.

It is shown that in a number of problems containing many constant parameters, such as the SLAM problem, FGO methods allow designing estimation algorithms that are computationally significantly simpler than the algorithms designed by traditional methods from the standpoint of the Bayesian approach without taking into account the features of these problems.

References:

[1] Q. Zeng, W. Chen, J. Liu, and H. Wang, "An Improved Multi-Sensor Fusion Navigation Algorithm Based on the Factor Graph," Sensors, vol. 17, p. 641, Mar. 2017.

[2] E. Olson, AXLE: Computationally-efficient trajectory smoothing using factor graph chains, 2021. IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 2021, pp. 7443-7448, doi: 10.1109/ICRA48506.2021.9561823.

[3] Weisong Wen, Tim Pfeifer, Xiwei Bai Li-Ta Hsu1. Factor graph optimization for GNSS/INS integration: A comparison with the extended Kalman filter. NAVIGATION 2021; 68:315–331.

[4] A.Yu., Scalable algorithms for simultaneous mapping and localization of a swarm of mobile robots, Cand. Sci. (Tech.), St Petersburg, 2021

[5] Xiwei WU, Bing XIAO, Cihang WU, Yiming GUO, Lingwei LI., Factor graph-based navigation and positioning for control system design: A review. Chinese Journal of Aeronautics. Volume 35, Issue 5, May 2022, pp. 25-39

[6] Yanwu Zhai and Songyuan Zhang. A novel LiDAR–IMU–Odometer coupling framework for two-wheeled inverted pendulum (TWIP) robot localization and mapping with nonholonomic constraint factors. Sensors, 2022, 22, 4778. https://doi.org/10.3390/s22134778.

[7] Jun Dai, Songlin Liu, Xiangyang Hao, Zongbin Ren and Xiao Yang. UAV Localization Algorithm Based on Factor Graph Optimization in Complex Scenes. Sensors, 2022, 22, 5862.

[8] Mingxing Chen, Zhi Xiong, Jun Xiong,d Rong Wang. A hybrid cooperative navigation method for UAV swarm based on factor graph and Kalman filter, International Journal of Distributed Sensor Networks, 2022, Vol. 18(1).

[9] Shiji Xin, Xiaoming Wang, Jinglei Zhang, Kai Zhou, and Yufei Chen, A Comparative Study of Factor Graph Optimization-Based and Extended Kalman Filter-Based PPP-B2b/INS Integrated Navigation. Remote Sens. 2023, 15, 5144. https://doi.org/10.3390/rs15215144.

[10] A.A. Penkovsky, J. Mahmoud, M. Mohrat, and S.A. Kolyubin, Robust visual-inertial odometry for ground robots in dynamic environments, 30th St. Petersburg International Conference on Integrated Navigation Systems, 2023, pp. 49–54.

[11] D.A. Cherginets and A.A. Vedyakov, Design of visual-inertial odometry algorithm for a four-legged walking

robot with a stereo camera, 31st Saint Petersburg International conference on Integrated Navigation Systems, 2024, pp. 453–457.

[12] A.R. Abdrashitov, Relative navigation of a group of unmanned aerial vehicles using Bayesian estimates, Trudy XIX Vserossiiskoi nauchno-prakticheskoi konferentsii "Perspektivnye sistemy i problemy upravleniya" (Proc. of the XIX All- Russian scientific and practical conference "Advanced systems and control problems"), 2024.

[13] S.H. Quan, S.H. Chen, Y.L. Zhou S. Zhao, H.Z. Hu, Q. Zhu, A Robust IMU/GNSS/ODO integrated navigation system based on factor graph, 31th Saint Petersburg International Conference on Integrated Navigation Systems, ICINS 2024, Proceedings. 27. 2024

[14] V.G. Karaulov, A.M. Gruzlikov, O.A. Stepanov, and Yu.A. Litvinenko, Solution of the problem of AUV Positioning Relative to a Stationary Docking Station Using a Factor Graph Optimization Algorithm, 31th St. Petersburg International Conference on Integrated Navigation Systems, Conference Proceedings. St. Petersburg, 2024.

[15] O.A. Stepanov, A.M. Isaev, A.V. Motorin, Yu.A. Litvinenko, and V.P. Zolotarevich, Recursive and nonrecursive algorithms applied to navigation data processing: Differences and interrelation with factor graph optimization algorithms, Proc. 31st St. Petersburg International Conference on Integrated Navigation Systems, 2024, pp. 424–432.

[16] A. Aravkin, B. Bell, J. Burke, and G. Pillonetto, Kalman smoothing and block tridiagonal systems: new connections andnumerical stability results, arXiv, 2013, https://doi.org/10.48550/arXiv.1303.5237.

[17] Stepanov, O.A., Primenenie teorii nelineinoifil'tratsii v zadachakh obrabotki navigatsionnoi informatsii (Application of Nonlinear Filtering Theory for Navigation Data Processing), 3rd edn., St. Petersburg: Elektropribor, 2003.

[18] V.A. Vasiliev, Yu.A. Litvinenko, O.A. Stepanov, and A.B. Toropov, Comparison of recursive and nonrecursive filtering

algorithms for map-aided navigation problems, XIV Vserossiiskaya mnogoprofil'naya konferentsiya po problemam upravleniya, MKPU-2021 (XIV All-Russian multi-conference on control problems), Proceedings: in 4 volumes. Rostov- on-Don, 2021, pp. 125-127.

[19] F. Dellaert and GTSAM Contributors, "Georgia Tech Smoothing and Mapping Llibrary", https://github.com/borglab/gtsam, 2022.

Study of The Influence of The Moon on The Inspection Motion of Spacecrafts in High Orbits

V.A. Bolotina, D.P Avariaskin

(Samara National Research University, Samara)

Abstract: In this work, initial conditions are used to ensure inspection motion, providing closed relative trajectories with unperturbed motion. Taking into account the influence of perturbation in the form of the gravitational field of the Moon leads to the evolution of the relative trajectory, and it ceases to be closed. A study of the evolution of the inspection trajectory under the influence of the gravitational field of the Moon is conducted.

1 Introduction

To solve scientific and applied problems in space, the technology of group flight is used, one of the types of which is the inspection motion of spacecraft (SC). Inspection motion is used for the inspection object (hereinafter referred to as the reference SC) to fly around the inspector in order to obtain the necessary information, inspection and diagnostics without direct contact [1]. The inspector can also approach space debris for its inspection or removal from orbit.

The main issue in the inspection problem is the choice of the initial conditions of the group flight. In the case of unperturbed motion, the initial conditions that ensure a closed relative trajectory are found from the condition of equality of the orbital energies of the reference SC and the inspector. However, under the influence of various disturbances, the trajectory ceases to be closed, which leads to the destruction of the inspection motion.

This work is devoted to the study of the influence of the gravitational field of the Moon on the formation of the inspection motion of spacecraft in high orbits. That is, a study was conducted of various initial conditions of motion of the inspector and the reference spacecraft, under which the deviation of the relative trajectory from the nominal (undisturbed) under the influence of the gravitational field of the Moon would be minimal.

2 Mathematical modeling of relative motion

In this paper, we will consider the influence of the Moon's gravitational field on inspection motion in the case when the reference SC moves along a geostationary orbit and the Molniya orbit. Inspection motion is provided in the plane of the orbits under consideration. Closed relative trajectories used to provide inspection of one SC by another can be obtained by selecting the initial conditions of the inspector's motion. In this case, the initial conditions must be selected so that the condition for the formation of a closed relative trajectory is met: the periods of revolution of each SC along its orbit around the attracting center must be equal:

$$T_{ref} = T_i \qquad (1)$$

where: T_{ref}, T_i – the periods of revolution of the reference spacecraft and the inspector, respectively, in their orbits.

Let us introduce two coordinate systems: the absolute geocentric coordinate system (ACS) and the orbital barycentric coordinate system (OCS). Let us consider the disturbed motion of two spacecraft – the reference and the inspector. The center of mass of the reference spacecraft is the origin of the OCS. The model of the motion of the reference spacecraft and the inspector in the ACS, taking into account the influence of the gravitational field of the Moon, are written, respectively, in the form

$$\ddot{\vec{r}}_{\text{оп}} + \frac{\mu}{r_{\text{оп}}^3}\vec{r}_{\text{оп}} = \vec{f}_{\text{Л}}, \ \ \ddot{\vec{r}}_{\text{и}} + \frac{\mu}{r_{\text{и}}^3}\vec{r}_{\text{и}} = \vec{f}_{\text{Л}} \tag{2}$$

where $\ddot{\vec{r}}_{\text{оп}}$ – acceleration of the reference spacecraft; $\mu = 398602$ km³/s² – gravitational parameter of the Earth; $\vec{r}_{\text{оп}}$ – radius vector of the reference spacecraft; $\vec{f}_{\text{Л}}$ – disturbing acceleration caused by the attraction of the Moon; $\ddot{\vec{r}}_{\text{и}}$ – acceleration of the inspector; $\vec{r}_{\text{оп}}$ – inspector radius vector.

The initial conditions of the inspector's motion are determined from the expression for the relationship between the ACS and the OCS:

$$V_{\text{и}}^2 = (v_{x0} + V_{n\text{оп}} + \omega_{\text{оп}}y_0)^2 + \left(v_{y0} + V_{r\text{оп}} - \omega_{\text{оп}}x_0\right)^2 \tag{3}$$

where $\omega_{\text{оп}}$ – angular orbital velocity of the reference spacecraft; $V_{n\text{оп}}, V_{r\text{оп}}$ – normal and radial projections of the reference spacecraft velocity; $v_{y0} = \frac{\omega_{\text{оп}}x_0}{2}$ – inspector initial velocity component [2], $V_{\text{и}} = \sqrt{\frac{2\mu}{r_{\text{и}}} - \frac{\mu}{a_{\text{оп}}}}$ – initial speed of the inspector in the OCS, $r_{\text{и}} = \sqrt{X_{\text{и}}^2 + Y_{\text{и}}^2 + Z_{\text{и}}^2}$ – the value of the radius vector of the inspector. In this case, the fulfillment of (1) is ensured by choosing such a velocity of the inspector that the major semi-axes of the orbits of the inspector and the reference spacecraft are equal. Then the relative trajectory will be closed. This is true only for the case of unperturbed motion. Due to the fact that within the framework of this work the influence of the gravitational field of the Moon is taken into account, the real trajectory will not be closed. Therefore, a parameter is introduced that will be used to estimate the influence of this perturbation. The maximum deviation of the inspection perturbed trajectory from the unperturbed relative trajectory Δr_{max} under the same initial conditions of motion is taken as the evaluation parameter:

$$\Delta r_{max} = r_{\max \text{В}} - r_{\max \text{Н}} \tag{4}$$

where $r_{\max \text{В}}$ – maximum distance between the reference spacecraft and the inspector on the disturbed trajectory, m; $r_{\max \text{Н}}$ – maximum distance between the reference spacecraft and the inspector on an undisturbed trajectory, m.

Using (4), the maximum deviations of the inspection perturbed from the unperturbed relative trajectory were found. The influence of the initial position of the Moon and the reference spacecraft on the inspection motion of the spacecraft was studied. The maximum influence of the Moon occurs at the arguments of the Moon's latitude from 75 to 125 degrees and from 260 to 320 degrees and at the arguments of the reference spacecraft's latitude from -40 to 20 degrees in the Molniya-type orbit.

A study was also conducted on the influence of the initial position of the inspector on the

formation of the inspection motion, taking into account the influence of the Moon. The initial position was set by the angle α, measured from the Ox axis of the OSC to the current position of the inspector. The study showed that the influence of the Moon on the inspection motion is maximum in the case when the reference spacecraft and the inspector are on the same radius vector. Based on the simulation results, a graph was constructed reflecting the dependence of the maximum influence of the Moon on the initial positions of the inspector (Fig. 1).
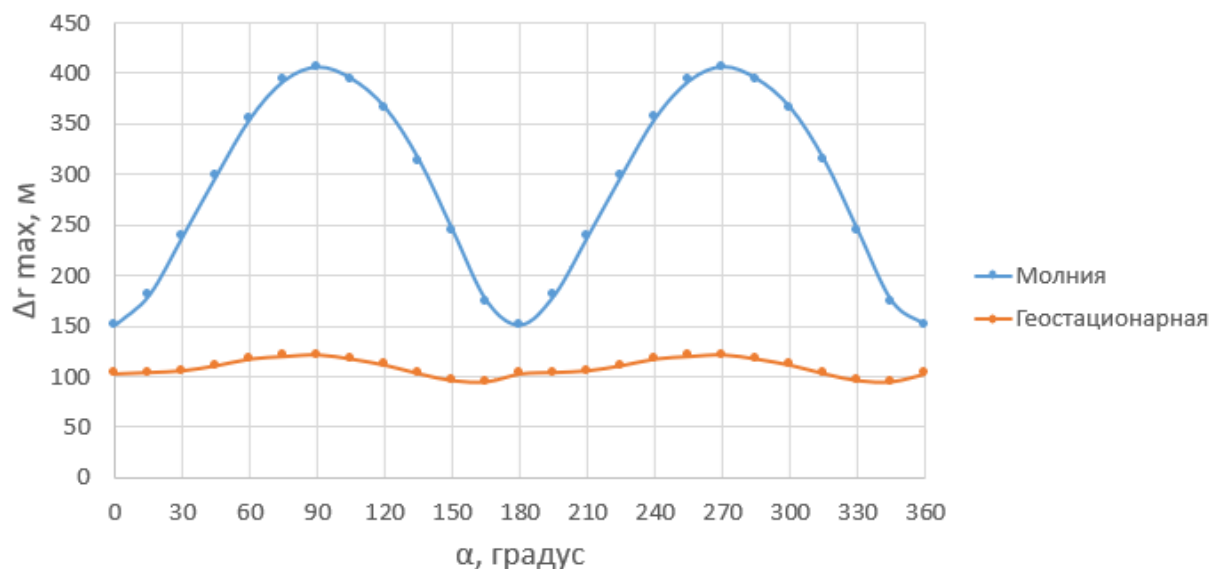
Fig.1 Maximum influence of the Moon depending on the initial position of the inspector

3 Conclusion

A study of the influence of the Moon's gravitational field on the formation of inspection motion of spacecraft in high orbits was conducted. Ranges of arguments of latitude of the Moon and the reference spacecraft were obtained for which the influence of the Moon is maximum. A study of the influence of the Moon's gravitational field on the formation of inspection motion depending on the position of the inspector relative to the reference spacecraft showed that it is maximum when both spacecraft are on the same radius vector. There are no pronounced maxima on the inspection trajectory when moving along a geostationary orbit.

References:

[1] Scherbakov M. S., Medvedev S. A. Study of the possibility of using osculating elements of relative motion in the inspection task // Mathematical modeling, computer and natural experiment in natural sciences. 2022. No. 4

[2] Scherbakov M. S., Avariaskin D. P. Studying problems on choosing stable orbits of nanosatellites to provide passive and periodic relative trajectories // Journal of Physics: Conference Series. – 2020. – Vol. 1536, №1. – 8 p.

Gyroscope Biases Estimation Using Zero-Velocity Updates in Pedestrian Navigation

A. V. Bragin

(Lomonosov Moscow State University, Moscow)

Abstract: Pedestrian navigation system based on foot-mounted inertial measurement units is considered. The problem of gyroscope biases estimation with the help of zero-velocity updates in case of simplified motion equations usage is studied. Analytical observability analysis done with several assumptions shows that gyroscope biases are observable. However numerical analysis done for simulated data indicates that "vertical" gyroscope bias estimates convergence is slow which is unacceptable for practical applications. Further studies should uncover the reason for such behavior of formally observable variables.

1 Introduction

Autonomous pedestrian navigation is on a high demand in different cases: military personnel in a GNSS-denied area, emergency units or firefighters in a large buildings, etc. In the paper pedestrian navigation system based on foot-mounted micromechanical inertial sensors is considered. Foot-mounted configuration allows to use information about foot zero velocity in the stance phase of the step. Zero velocity updates (ZUPT's) make navigation with low-cost and low-accurate micromechanical inertial sensors possible, but cannot prevent the growth of heading error. Heading error is a result of gyroscopes biases which should be estimated and compensated.

A common approach in inertial navigation is to estimate sensors biases during navigation with the help of different external information such as GNSS. In pedestrian navigation the main correction method is ZUPT, but in the literature there is no consensus on the possibility of gyroscopes biases estimation via ZUPT. In [1] authors state that modeling errors in the system affects the navigation accuracy harder than the biases and thus their estimation is not recommended. In [2] analysis of multiple algorithms tests on real data shows that gyros biases estimation shows positive effect in some cases. In [3] authors state that biases are "weakly observable", based on simulated data analysis.

It is shown in the paper that gyroscopes biases are observable during ZUPT in case of simplified motion equations usage. Using simulated data it is shown that despite the observability estimates of "vertical" gyro bias doesn't converge to true values in a reasonable time.

2 Gyroscopes biases estimation

Here are the motion equations and corresponding error equations:

$$
\begin{array}{lllll}
\dot{p}'_n &=& v'_n & \delta p_n &=& \delta v_n + \hat{p}'_n C'^{T}_{sn}(\Delta v_s^0 + v_s^s) & f'_s = f_s + \Delta f_s \\
\ddot{v}'_n &=& C'^{T}_{sn} f'_s + g_n, & \delta v_n &=& \hat{g}_n \beta_n + \Delta f_n + \hat{v}'_n C'^{T}_{sn}(\Delta v_s^0 + v_s^s), & \omega'_s = \omega_s - v_s \\
\dot{C}'_{sn} &=& (\hat{\omega}'_s + \hat{v}_s^0)C'_{sn} & \dot{\beta}_n &=& C'^{T}_{sn}(\Delta v_s^0 + v_s^s) & v_s = v_s^0 + \Delta v_s + v_s^s \\
\dot{v}_s^0 &=& 0 & \dot{\Delta v}_s^0 &=& 0
\end{array}
$$

$$\text{(1)}$$

Here p'_n, v'_n denote coordinates and velocities in navigation frame, C'_{sn} denotes orientation matrix describing transition from navigation frame to sensor frame, $\delta p_n, \delta v_n$ denotes the so-called dynamic coordinates and velocity errors [4], β_n denotes orientation error. f'_s, ω'_s denotes accelerometers and gyroscopes measurements respectively, $\Delta f_s, v_s$ – measurements errors, v_s^0, v_s^s stands for constant and stochastic components of gyroscope biases and Δv_s^0 stands for bias estimation error. Navigation errors are written for flat Earth model [5], these simplified model is widely used in pedestrian navigation.

System (1) is aided with zero velocity. Measurements are formed when the stance phase is detected. SHOE detector was used for this purpose [6].

$$
\begin{array}{ll}
Z' = v'_n & z = Z' - Z = \delta v_n + r_n = Hx + r_n \\
Z = v_n - r_n, & H = [0_3 \quad I_3 \quad 0_3 \quad 0_3]
\end{array}
\quad, \quad
x = \left(\delta p_n^T \quad \delta v_n^T \quad \beta_n^T \quad \Delta v_s^{0T} \right)^T
$$

$$\text{(2)}$$

Here Z, Z' denotes measurements in terms of state vector and z denotes measurements in terms of error vector, r_n denotes measurement errors, H denotes measurement matrix, corresponding to the EKF state vector x. Conventional EKF is used.

3 Observability analysis

Observability analysis was done with a few simplifications. First, stance phase of the step is instantaneous. Second, all steps of the pedestrian are exactly the same. Observability matrix is given

for the state vector $x = \left(\delta v_n^T \quad \beta_n^T \quad \Delta v_s^{0T} \right)^T$, since position error is obviously unobservable:

$$
\mathcal{O}_k =
\begin{bmatrix}
HF^k \\
\cdots \\
HF \\
H
\end{bmatrix}
=
\begin{bmatrix}
I_3 & k\hat{g}_n^{sn}T & k(g_n C_2 + C_3) + \dfrac{k(k-1)}{2}\hat{g}_n T C_1 \\
I_3 & \hat{g}_n^{sn}T & g_n C_2 + C_3 \\
I_3 & 0_3 & 0_3
\end{bmatrix}
\qquad
\begin{aligned}
C_1 &= \int_{\tau=t_k}^{t_{k+1}} C'_{ns}(\tau)d\tau \\
C_2 &= \int_{\tau=t_k}^{t_{k+1}} d\tau \int_{s=t_k}^{\tau} C'_{ns}(s)ds \\
C_3 &= \int_{\tau=t_k}^{t_{k+1}} \hat{v}'_n(\tau)C'_{ns}(\tau)d\tau
\end{aligned}
$$

$$\text{(3)}$$

It can be shown that the unobservable subspace corresponding to the matrix $\mathcal{O}_k$ contains

only the vector $v = \left(0_{1\times3} \quad g_n^T \quad 0_{1\times3} \right)^T$, i.e. only β_3 is unobservable.

4 Gyroscopes biases estimability for the simulated data

Algorithm tests were performed with the help of the simulator, similar to the one, presented in [3]. Coordinates and orientation angles in the swing phase are modeled as harmonic functions. Velocity in the stance phase equals zero. Pith and roll angles were in the stance phase were different in different experiments modeling different sensor position on the foot.
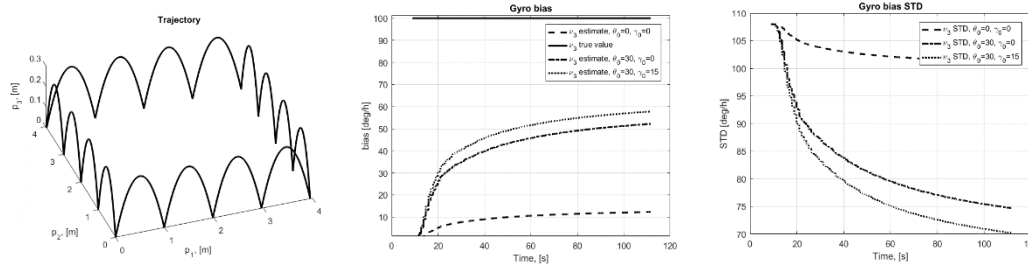
Fig.1 Left – example of the foot trajectory, center – "vertical" gyro bias estimation result for different sensor mounting positions (different pitch/roll angles in the stance phase), right – corresponding covariances.

5 Conclusions

It is shown that for simplified motion equations gyroscopes biases are observable, but the estimates don't converge in a reasonable time. This conclusion is based on a modeling results. Real data differs from the simulated one in the sense that conditions for biases estimation become worse, so further research should be conducted to uncover the influence of the factors not modeled in the presented study.

References:

[1] Nilsson J. O., Skog I., Handel P. A note on the limitations of ZUPTs and the implications on sensor error modeling //2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 13-15th November 2012. – 2012.

[2] Wagner J. F., Kohl M., Gyorfi B. Reevaluation of algorithmic basics for ZUPT-based pedestrian navigation //IEEE access. – 2022. – Vol. 10. – P. 118419-118437

[3] Zhu M., Wu Y., Luo S. A pedestrian navigation system by low-cost dual foot-mounted IMUs and inter-foot ranging //2020 DGON Inertial Sensors and Systems (ISS). – IEEE, 2020. – P. 1-20.

[4] Bolotin Y., Bragin A., Gartseev I. Covariance error analysis for pedestrian dead reckoning with foot mounted imu // CEUR Workshop Proceedings. — 2019. — Vol.2498. — P. 243–250

[5] Saab S. S., Gunnarsson K. T. Automatic alignment and calibration of an inertial navigation system //Proceedings of 1994 IEEE Position, Location and Navigation Symposium-PLANS'94. – IEEE, 1994. – P. 845-852

[6] Skog I. et al. Zero-velocity detection—An algorithm evaluation //IEEE transactions on biomedical engineering. – 2010. – Vol. 57. – №. 11. – P. 2657-2666

Simulation of Trajectory Parameters of Object Motion in Pole Areas Using Quasi-Geodetic Coordinates

S.P. Ablyasova

(Lomonosov Moscow State University, Moscow)

Abstract: The task of simulating the trajectory parameters of an object - coordinates, linear velocity components, attitude angles, and "ideal" readings from inertial sensors such as angular rate sensors and accelerometers when the object is moving above or below the pole - is needed for testing onboard navigation algorithms of inertial navigation systems, such as strapdown inertial navigation system (INS). The unique aspect of the simulation problem mentioned is that traditional parameters such as longitude and heading cannot be used in relevant models, as these parameters are not defined for points lying on the axis of Earth's rotation. It has been demonstrated that this simulation can be performed using the quasi-Greenwich coordinate system, quasi-geodetic (or quasi-geocentric) coordinates, and quasi-angles of orientation, instead of the traditional ones. The zero test of inertial dead reckoning algorithms when the trajectory of an object crosses the axis of Earth's rotation confirms the proposed solution to the problem.

1 Introduction

The task of simulation of an object's trajectory parameters and corresponding inertial sensors readings in case when trajectory meets the axis of Earth rotation, in particular the north pole, is considered. Due to the fact that traditional models of trajectory simulators typically use a geodetic reference frame, these models are not suitable for use in areas near the poles. For this reason, the use of a quasi-Greenwich reference system and an appropriate quasi-geodetic navigation frame, as well as quasi-attitude angles, instead of traditional ones, is relevant. The report presents an approach to simulating the motion trajectory and its corresponding "ideal", i.e., error-free, measurements from INS inertial sensors. Based on this simulation, we compute a set of trajectory parameters to test onboard algorithms (dead reckoning) in polar regions.

2 Quasi-Greenwich and quasi-geodetic coordinate systems

In trajectory simulation procedure it is assumed that the object crosses the pole in the meridian plane with a known value of geodetic longitude λ_0, and that the initial value of geodetic latitude φ_0 is assumed to be close enough to $90°$. In the context of transition to quasi-Greenwich reference system and quasi-geodetic navigation frame, the mentioned meridian is defined as a plane of the quasi-equator. Consequently, the quasi-latitude at the pole equals zero.

The quasi-Greenwich $O\eta^q$ coordinate system is introduced by two consecutive rotations of the Greenwich refernce system $O\eta$ around the axis $O\eta_2$ by the angle $-\frac{\pi}{2}$ and around the new position of the first axis by the angle $-\frac{\pi}{2} + \lambda_0$. The quasi-equator plane coincides with the plane of the meridian with geodetic longitude λ_0.

The quasi-geodetic Ox^{0q} coordinate system is introduced by two consecutive rotations of the Greenwich coordinate system $O\eta^q$ around the axis $O\eta^q{}_3$ by the angle $\lambda^q + \frac{\pi}{2}$ and around the new position of the first axis by the angle $\frac{\pi}{2} - \varphi^{0q}$.

3 Trajectory modelling

In order to model the trajectory in quasi-coordinates the use of kinematic equations motion of IMU accelerometer's proof mass M is proposed:

$$\frac{d}{dt}\varphi^{0q} = \frac{V_N{}^q}{R},$$

$$\frac{d}{dt}\lambda^q = \frac{V_E{}^q}{R \cos\varphi^{0q}}, \tag{1}$$

$$\frac{d}{dt}R = V_{UP}{}^q,$$

where $V_E{}^q$, $V_N{}^q$, $V_{UP}{}^q$ are components of the M velocity vector relative to the axes of the quasi-geodetic coordinate system Ox^{0q}, R represents the modulus of the radius vector OM.

The initial conditions for equations (1) are as follows:

$$\varphi^{0q}(t_0) = 0,$$

$$\lambda^q(t_0) = \frac{\pi}{2} - \varphi^0(t_0),$$

$$R(t_0) = a + h(t_0) - \frac{ae^2}{2}\sin^2\varphi(t_0)(\cos^2\varphi(t_0) - \frac{1}{4}\sin^2\varphi(t_0)) + O(0,1\text{м}), \tag{2}$$

$$\psi^q(t_0) = -\frac{\pi}{2},$$

where $\varphi^0(t_0)$ and $\varphi(t_0)$ are the values of geocentric and geodetic latitudes, respectively, at the initial time t_0, $h(t_0)$ geodetic altitude at time t_0, ψ^q – quasi-heading angle, i.e. the angle between the axis $Mx_2{}^{0q}$ (the direction towards quasi-north) and the projection of the front-back body axis on the plane $Mx_1{}^{0q}x_2{}^{0q}$. This angle is counted in a clockwise direction from the axis $Mx_2{}^{0q}$. Throughout the entire trajectory modelling interval, the quasi-heading angle $\psi^q(t_0) = -\frac{\pi}{2}$, which corresponds to the motion of the object along the quasi-equator towards quasi-longitude decrease.

The kinematic equations (1) are integrated with the initial conditions (2). The trajectory obtained in quasi-geodetic coordinates is used in the modelling of inertial sensor readings.

4 Modelling of inertial sensor readings

Prior to modelling accelerometer readings, it is necessary to determine the absolute angular velocity vector of the body frame Ms fixed to the object. This step is no different from the established methodology for modelling ARS's readings in geodetic coordinates [2].

Accelerometer readings are calculated based on the point M motion equations in the axes Mx^{0q} (see [1], [2]):

$$\frac{d}{dt} A_{x^{0q}\eta^q} = \widehat{\Omega}_{x^{0q}} A_{x^{0q}\eta^q} \;,$$

$$\frac{d}{dt} V_{x^{0q}} = (\widehat{\Omega}_{x^{0q}} + 2\,\widehat{u}_{x^{0q}}) V_{x^{0q}} + f_{x^{0q}} + g_{x^{0q}} \;, \tag{3}$$

where $A_{x^{0q}\eta^q}$ is rotation matrix of the navigation frame Mx^{0q} to the quasi-Greenwich coordinate system $O\eta^q$ respectively, the term $f_{x^{0q}}$ represents the vector of specific force at the point M in the navigation frame (ideal accelerometers measurement), $\widehat{u}_{x^{0q}}$ is a skew-symmetric matrix that corresponds to the vector of angular velocity of the Earth's rotation in the frame Mx^{0q}, $g_{x^{0q}}$ represents the normal gravity vector and $\Omega_{x^{0q}}$ is an angular velocity vector of the quasi-geodetic frame Mx^{0q}.

5 Zero test

Based on the simulated inertial sensor readings, the algorithm of the SINS autonomous inertial calculus was tested. For this purpose the wander azimuth reference frame was used as a navigation frame. This type of navigation frame is traditionally used in aviation. The results of the dead reckoning test demonstrated a practical coincidence between the calculated trajectory parameters and the original trajectory parameters that were the basis of the simulation.

6 Conclusions

The results demonstrate that quasi-coordinates and quasi-orientation angles are the appropriate variables for modelling the trajectory parameters of an object's motion crossing points above and below the pole. This approach allows for the further full-scale dead reckoning algorithms testing on trajectories of this nature.

References:

[1] Vavilova, N.B., Golovan, A.A., and Parusnikov, N.A., Matematicheskie osnovy inertsial'nykh navigatsionnykh sistem. (Mathematical Foundations of Inertial Navigation Systems), Moscow: Moscow State University, 2020.

[2] Bogdanov O.N. Metodika soglasovannogo modelirovaniya izmerenij inercial'nyh datchikov, traektornyh parametrov ob"ekta s prilozheniem k zadacham inercial'noj i sputnikovoj navigacii. (Methodology of agreed modelling of measurements of inertial sensors, trajectory parameters of object with application to problems of inertial and satellite navigation), dissertation for the degree of Candidate of Physical and Mathematical Sciences, 2015

[3] Lukyanov D.P., Mochalov A.V., Odintsov A.A., Vaisgant I.B. Inercial'nye navigacionnye sistemy morskih ob"ektov. (Inertial navigation systems of marine objects), Leningrad : Sudostroenie

An Algorithm for Raw Data Fusion of A Strapdown Inertial Navigation System And Odometer

D.A. Safin

(Lomonosov Moscow State University)

Abstract: The paper describes an algorithm for data fusion of a strapdown inertial navigation system and odometer. The data fusion algorithm corrects the parameters of the inertial and odometry dead reckoning, provides estimates of the instrumental errors of the inertial sensors and the lever arm parameters in the inertial navigation system (INS) and odometer system, as well as the instrumental errors of the odometer, using the Kalman filter. The results of the experimental data processing for different types of land vehicles, including cars and tractors, have been presented. These findings demonstrate the effectiveness of the algorithm that was developed.

1 Introduction

Data fusion, provided by a strapdown inertial navigation system (INS) and an odometer, is widely used in various applications, including road surveying, self-driving vehicles, and agricultural machinery.

The source of initial data used to solve ground vehicle navigation problems is data from inertial sensors, such as accelerometers and gyroscopes, which are part of an inertial navigation system (INS). Additionally, data from an odometer, which records the absolute distance traveled by an object, is used. The given research examines the task of integrating INS (Inertial Navigation System) and odometer data in order to identify potential precision issues with this integration. However, the use of other available navigation information that is recorded, for example, by a global navigation satellite system (GNSS) receiver, is also considered.

Algorithms for data fusion form the basis of land vehicle navigation. They are used both in real-time and in post-processing modes. The Kalman filter is used as an assessment algorithm.

During the development of the algorithm, the following specific tasks were completed:

1. Initial alignment of the INS in the presence of uncontrolled angular motion of the vehicle's body frame.

2. Detecting the movement of an object in reverse, if the odometer measures an unsigned increment of the path.

3. Recalibration of the strapdown INS-odometer system according to results of data processing in previous tracks. Adjusting parameters are: two installation angles of INS instrument frame with respect to odometer measuring axis of the odometer, scale factor error of the odometer.

4. The size effect is due to the mutual displacement of the center of the INS and the odometer measuring point.

5. Application of Kalman-type smoothing algorithm in INS-odometer post-processing mode in case of known coordinates of the final stopping point.

6. Testing of Zero Velocity Update Technology (ZUPT-correction) when the vehicle stops.

2 Specifics of the development of data fusion algorithm

1. A three-dimensional autonomous INS dead reckoning system is applied. Unstable INS altitude channel is being damped.

2. Three-dimensional odometric dead reckoning based on odometer reading and INS orientation parameters is applied.

3. The positional data from odometric dead reckoning are used for INS positional aiding with dampening of its vertical channel.

3 The main mathematical models of the INS integration tasks of the basis of odometer readings:

1. 3D INS error equations

2. 3D odometer dead reckoning error equations

3. INS position aiding based on odometer-derived coordinates.

The following data are included into the state vector of estimated parameters:

1. inertial and odometric dead reckoning errors

2. parameters of inertial sensors instrumental error models

3. scale factor error of the odometer

4. parameters of interrelated geometry of odometer and INS:

- interrelated misalignment of INS–odometer system in yaw and pitch channels
- lever arm parameters in the INS-odometer system

4 Test data, Moscow city

Three diagnostic test runs were carried out using the car which had navigational-grade strapdown INS installed in it.

The results of odometric dead reckoning was used only for INS positioning aiding.
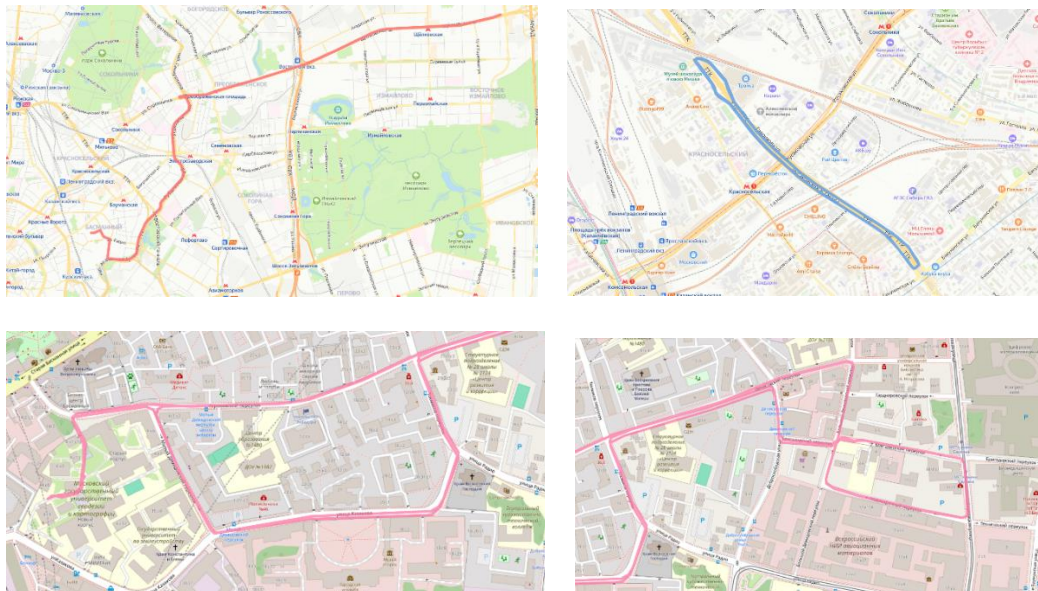
Images 1, 2, 3, 4. Adjusted trajectory on the map (real time algorithm)

In post-processing mode and in all the test rides we have obtained nearly the same estimation of misalignment parameters . In order to improve the precision of navigation we have verified the

effectiveness of misalignment parameters compensation (finishing calibration of strapdown INS-odometer system).

5 Test track using microelectromechanical system-based sensors (MEMS-based sensors)

Positioning data from odometer was used for the aiding purposes, i.e. the result of the odometric dead reckoning and positioning data from global navigation satellite system (GNSS). Estimation algorithm was used in the feedback mode

Image 5.Adjusted trajectory in real time

Image 6. Tractor with MEMS and odometer and GNSS receiver on the map

In post-processing we have received estimation of misaligment parameters and vector of MEMS and odometer bedding point angle that demonstrated a high precision of anticipatorily measurements.

6 Conclusion

The results of test data processing have shown performance efficiency of strapdown INS-odometer integration algorithm, as well as acceptable accuracy in solving navigation tasks using sensors of various accuracy grades. The results demonstrate the effectiveness of finishing calibrations of INS-odometer system.

References:

[1] Vavilova N.B., Golovan A.A., Parusnikov N.A. "Mathematical principles of navigation systems" Publishing house of Moscow University, 2020.

[2] Golovan A.A., Nikitin I.V. "Integration problems of strapdown INS and odometer from the point of view of

the mechanics of adjusting inertial navigation systems" Part 1, Part 2 "Moscow University Journal. Mathematics. Mechanics." №2, pages 69-72, №4, pages 68-72, Moscow, 2015.

[3] Golovan A.A. "Integration solution "Strapdown INS-odometer": positional (place-value) version (variant) // Gyroscopy and navigation. 2021, vol. 29. № 2. pages 110-125. DOI 10.17285/0869-7035.0066.

Research on Consensus and Fixed-Time Control Methods for Multi-Agent Systems

D. Mingazov, M. Kozin

(Saint Petersburg State University, Saint Petersburg)

Abstract: This work examines consensus and fixed-time control algorithms for multi-component systems. The study explores mathematical models using iterative consensus and control methods, and proposes a fixed-time approach based on nonlinear functions. Numerical experiment results are provided to demonstrate the effectiveness of the proposed methods.

1 Introduction

Modern multi-agent systems (MAS) are a critical area of research, particularly in their applications within robotics, autonomous vehicle control, and communication networks. Coordinating agents, each following local behavioral rules and interacting with neighbors, presents a significant challenge. A key issue is achieving a consistent state (consensus) across all agents in the system, despite limited computational resources and time constraints.

Traditional consensus methods, such as linear information exchange models between agents, were introduced in the work of Olfati-Saber et al.[3].These methods enable asymptotic convergence to a common state, making them useful for various practical applications. However,a major drawback of these algorithms is the requirement of unlimited time to reach consensus. This limitation reduces their effectiveness in systems that require convergence within a finite time.

To address this issue, fixed-time control (Fixed-Time Convergence, FTC) methods have been proposed, ensuring system convergence within a fixed time interval regardless of the agents' initial conditions. This approach is based on nonlinear control functions and was thoroughly discussed in the works of Polyakov and Parsegov [4]. These methods guarantee rapid state alignment among agents, allowing consensus to be achieved in fixed time, which makes them appealing for tasks where response speed is critical.

This study investigates combined consensus and FTC methods for multi-agent systems. Specifically, an algorithm is considered that employs nonlinear control functions to accelerate convergence. The control expression is written as follows:

$$U_t = K_1 \left(\sqrt[3]{|x_i - x_j|} \right) + K_2 \left(x_i - x_j \right)^3 \tag{1}$$

where: x_i and x_j are the states of agents, and K_1 and K_2 are constants that regulate the rate of convergence. Using such control functions enables consensus to be reached faster and within a

finite time compared to classical approaches.

This report is focused on examining the effectiveness of combined consensus and fixed-time control methods for multi-agent systems. Numerical experiments are presented to demonstrate the advantages of the proposed approach in managing systems with limited time resources.

2 Problem Statement

We consider a system with N agents, each with a state $x_i(t)$ at time t, where i = 1, 2,..., N. The goal is to develop an algorithm that ensures consensus among all agents within a fixed time. This implies that there exists a time T after which the states of all agents coincide:

$$x_i(T) = x_j(T), \forall i,j \in \{1,2,\ldots,N\} \tag{2}$$

The task is to develop a control algorithm that ensures the alignment of all agents' states within a finite time T, regardless of their initial conditions.

3 Problem Conditions

To solve the problem, it is assumed that agents can interact with a limited number of neighbors according to a communication graph. The interactions between agents are described by an adjacency matrix $A = [a_{ij}]$, where $a_{ij} \neq 0$ indicates a connection between agents i and j, and $a_{ij} = 0$ indicates its absence. The dynamics of the agents' state changes are described as follows:

$$x_i^{(t+1)} = x_i^{(t)} + \sum_{j \neq i} a_{ij} [K_1(\sqrt[3]{\left| x_j^{(t)} - x_i^{(t)} \right|}) + K_2(x_j^{(t)} - x_i^{(t)})^3] \tag{3}$$

where: K_1 and K_2 are positive constants that regulate the rate of convergence. This nonlinear control function includes two components: the first component, involving the cubic root, is responsible for the "soft" alignment of agent states, while the second accelerates the consensus process through the cubic difference of states.

4 Stability Analysis

To analyze the stability of the system, let us consider the Lyapunov function V(t), defined as:

$$V(t) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} (x_j(t) - x_i(t))^2 \tag{4}$$

This function characterizes the divergence of agent states and must decrease over time to ensure convergence. The time derivative of the Lyapunov function is calculated as:

$$\hat{V}(t) = -K_1 \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} (\sqrt[3]{\left| x_i - x_j \right|})^2 - K_2 \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} (x_j(t) - x_i(t))^6 \tag{5}$$

Both terms are decreasing, which confirms the stability of the proposed algorithm and guarantees that all agents converge to a single state within a finite time T.

5 Conclusion

This study examined various dynamic models for controlling multi-agent systems, specifically the consensus protocol, the Round-Robin algorithm, and a modified protocol incorporating Fixed-Time Convergence (FTC) elements. The findings indicate that FTC

significantly enhance the convergence rate of the system due to its adjustable parameters, K_1 and K_2, which allow for precise control over the consensus achievement time.

Optimizing parameters K_1 and K_2 further demonstrated that FTC surpasses traditional methods in convergence speed. Future research could focus on fine-tuning these parameters through adaptive techniques or machine learning approaches. It is also crucial to investigate the application of FTC in the context of partially connected agent networks and under resource constraints.

In summary, this work illustrates the efficacy of FTC algorithms for achieving rapid consensus, and continued research will likely extend their applicability to more complex systems.
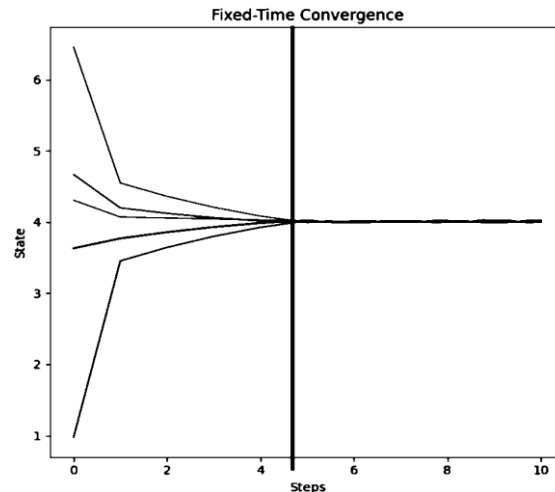
Fig.1   Increase in Convergence Speed for 5 Agents
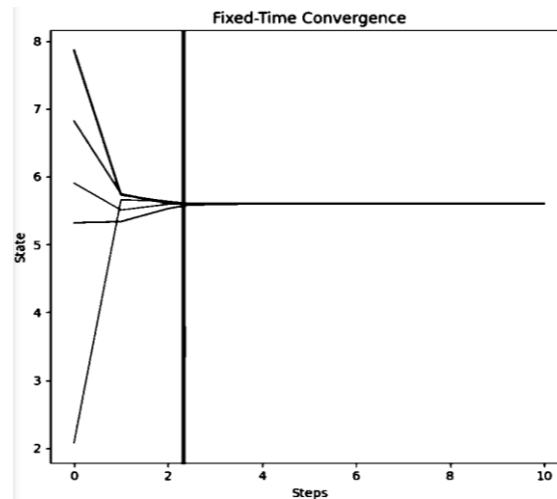
Fig.2   Increase in Convergence Speed for 5 Agents

References:

[1] Amelina N.O., Granichin O.N., Fradkov A.L. Method of Averaged Models for Discrete Adaptive Systems. Moscow: Automation and Telemechanics, 2019,No. 10, pp. 3–36.

[2] Amelina N.O., Fradkov A.L. Approximate Consensus in a Stochastic Dynamic Network with Incomplete Information and Measurement Delays. Moscow: Automation and Telemechanics, 2012,No. 11, pp. 6–29

[3] Olfati-Saber R., Fax A., Murray R.M., Consensus and Cooperation in Networked Multi-Agent Systems,

Proceedings of the IEEE, 2007, Vol. 95, No. 1.

[4] Parsegov S.E., Polyakov A.E., Shcherbakov P.S., Fixed-Time Consensus Algorithm for Multi-Agent Systems with Integrator Dynamics, IFAC Proceedings Volumes, 2013, Vol. 46, No. 27, pp. 110-115.

[5] Chen Y., Wang F., Wang B., Fixed-Time Convergence in Continuous-Time Optimization: A Fractional Approach, IEEE Control Systems Letters, 2022, Vol. 7, P. 631-636

[6] Ge P., Li P., Chen B., Teng F., Fixed-Time Convergent Distributed Observer Design of Linear Systems: A Kernel-Based Approach., IEEE Transactions on Automatic Control, 2023, Vol. 68, P. 4932-4939

Study of The Influence of The Moon on The Inspection Motion of Spacecrafts in High Orbits

V.A. Bolotina, D.P Avariaskin

(Samara National Research University, Samara)

Abstract: In this work, initial conditions are used to ensure inspection motion, providing closed relative trajectories with unperturbed motion. Taking into account the influence of perturbation in the form of the gravitational field of the Moon leads to the evolution of the relative trajectory, and it ceases to be closed. A study of the evolution of the inspection trajectory under the influence of the gravitational field of the Moon is conducted.

1 Introduction

To solve scientific and applied problems in space, the technology of group flight is used, one of the types of which is the inspection motion of spacecraft (SC). Inspection motion is used for the inspection object (hereinafter referred to as the reference SC) to fly around the inspector in order to obtain the necessary information, inspection and diagnostics without direct contact [1]. The inspector can also approach space debris for its inspection or removal from orbit.

The main issue in the inspection problem is the choice of the initial conditions of the group flight. In the case of unperturbed motion, the initial conditions that ensure a closed relative trajectory are found from the condition of equality of the orbital energies of the reference SC and the inspector. However, under the influence of various disturbances, the trajectory ceases to be closed, which leads to the destruction of the inspection motion.

This work is devoted to the study of the influence of the gravitational field of the Moon on the formation of the inspection motion of spacecraft in high orbits. That is, a study was conducted of various initial conditions of motion of the inspector and the reference spacecraft, under which the deviation of the relative trajectory from the nominal (undisturbed) under the influence of the gravitational field of the Moon would be minimal.

2 Mathematical modeling of relative motion

In this paper, we will consider the influence of the Moon's gravitational field on inspection motion in the case when the reference SC moves along a geostationary orbit and the Molniya orbit. Inspection motion is provided in the plane of the orbits under consideration. Closed relative trajectories used to provide inspection of one SC by another can be obtained by selecting the initial conditions of the inspector's motion. In this case, the initial conditions must be selected so that the condition for the formation of a closed relative trajectory is met: the periods of revolution of each SC along its orbit around the attracting center must be equal:

$$T_{ref} = T_i \qquad (1)$$

where: T_{ref}, T_i – the periods of revolution of the reference spacecraft and the inspector, respectively, in their orbits.

Let us introduce two coordinate systems: the absolute geocentric coordinate system (ACS) and the orbital barycentric coordinate system (OCS). Let us consider the disturbed motion of two

381

spacecraft – the reference and the inspector. The center of mass of the reference spacecraft is the origin of the OCS. The model of the motion of the reference spacecraft and the inspector in the ACS, taking into account the influence of the gravitational field of the Moon, are written, respectively, in the form

$$\ddot{\vec{r}}_{\text{оп}} + \frac{\mu}{r_{\text{оп}}^3}\vec{r}_{\text{оп}} = \vec{f}_Л, \ \ \ddot{\vec{r}}_и + \frac{\mu}{r_и^3}\vec{r}_и = \vec{f}_Л \tag{2}$$

where $\ddot{\vec{r}}_{\text{оп}}$ – acceleration of the reference spacecraft; $\mu = 398602$ km³/s² – gravitational parameter of the Earth; $\vec{r}_{\text{оп}}$ – radius vector of the reference spacecraft; $\vec{f}_Л$ – disturbing acceleration caused by the attraction of the Moon; $\ddot{\vec{r}}_и$ – acceleration of the inspector; $\vec{r}_{\text{оп}}$ – inspector radius vector.

The initial conditions of the inspector's motion are determined from the expression for the relationship between the ACS and the OCS:

$$V_и^2 = (v_{x0} + V_{n\text{оп}} + \omega_{\text{оп}}y_0)^2 + \left(v_{y0} + V_{r\text{оп}} - \omega_{\text{оп}}x_0\right)^2 \tag{3}$$

where $\omega_{\text{оп}}$ – angular orbital velocity of the reference spacecraft; $V_{n\text{оп}}, V_{r\text{оп}}$ – normal and radial projections of the reference spacecraft velocity; $v_{y0} = \frac{\omega_{\text{оп}}x_0}{2}$ – inspector initial velocity component

[2], $V_и = \sqrt{\frac{2\mu}{r_и} - \frac{\mu}{a_{\text{оп}}}}$ – initial speed of the inspector in the OCS, $r_и = \sqrt{X_и^2 + Y_и^2 + Z_и^2}$ – the value of the radius vector of the inspector. In this case, the fulfillment of (1) is ensured by choosing such a velocity of the inspector that the major semi-axes of the orbits of the inspector and the reference spacecraft are equal. Then the relative trajectory will be closed. This is true only for the case of unperturbed motion. Due to the fact that within the framework of this work the influence of the gravitational field of the Moon is taken into account, the real trajectory will not be closed. Therefore, a parameter is introduced that will be used to estimate the influence of this perturbation. The maximum deviation of the inspection perturbed trajectory from the unperturbed relative trajectory Δr_{max} under the same initial conditions of motion is taken as the evaluation parameter:

$$\Delta r_{max} = r_{\max В} - r_{\max Н} \tag{4}$$

where $r_{\max В}$ – maximum distance between the reference spacecraft and the inspector on the disturbed trajectory, m; $r_{\max Н}$ – maximum distance between the reference spacecraft and the inspector on an undisturbed trajectory, m.

Using (4), the maximum deviations of the inspection perturbed from the unperturbed relative trajectory were found. The influence of the initial position of the Moon and the reference spacecraft on the inspection motion of the spacecraft was studied. The maximum influence of the Moon occurs at the arguments of the Moon's latitude from 75 to 125 degrees and from 260 to 320 degrees and at the arguments of the reference spacecraft's latitude from -40 to 20 degrees in the Molniya-type orbit.

Fig.1 Maximum influence of the Moon depending on the initial position of the inspector

A study was also conducted on the influence of the initial position of the inspector on the formation of the inspection motion, taking into account the influence of the Moon. The initial position was set by the angle α, measured from the Ox axis of the OSC to the current position of the inspector. The study showed that the influence of the Moon on the inspection motion is maximum in the case when the reference spacecraft and the inspector are on the same radius vector. Based on the simulation results, a graph was constructed reflecting the dependence of the maximum influence of the Moon on the initial positions of the inspector (Fig. 1).

3 Conclusion

A study of the influence of the Moon's gravitational field on the formation of inspection motion of spacecraft in high orbits was conducted. Ranges of arguments of latitude of the Moon and the reference spacecraft were obtained for which the influence of the Moon is maximum. A study of the influence of the Moon's gravitational field on the formation of inspection motion depending on the position of the inspector relative to the reference spacecraft showed that it is maximum when both spacecraft are on the same radius vector. There are no pronounced maxima on the inspection trajectory when moving along a geostationary orbit.

References:

[1] Scherbakov M. S., Medvedev S. A. Study of the possibility of using osculating elements of relative motion in the inspection task // Mathematical modeling, computer and natural experiment in natural sciences. 2022. No. 4

[2] Scherbakov M. S., Avariaskin D. P. Studying problems on choosing stable orbits of nanosatellites to provide passive and periodic relative trajectories // Journal of Physics: Conference Series. – 2020. – Vol. 1536, №1. – 8 p.

2024

Chinese-Russian
"Navigation and Motion Control"
Youth Forum

Topic 3: Global Navigation Satellite System

Resilient Factor Graph Optimization Navigation Method Cooperated with Filtering Theory

Qian Meng[1,2], Zuliang Shen[1,2], Fanchen Meng[3], Huishuo Feng[3] and Zhiqiang Gao[3]

(1. School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China;

2. Key Laboratory of Micro-Inertial Instruments and Advanced Navigation Technology, Ministry of Education, Southeast University, Nanjing 210096, China

3. Beijing Institute of Aerospace Control Devices, Beijing 100039, China)

Abstract: In order to restrain the divergence of integrated navigation and positioning caused by Global Navigation Satellite System (GNSS) outliers or faults in complex scenarios, and to realize accurate navigation and positioning in pedestrian navigation, a resilient factor graph pedestrian dead reckoning (PDR)/GNSS integrated navigation algorithm cooperated with filtering theory is proposed. Firstly, the method uses the fault-free and deviation segments of IGG-III function to generate equivalent weight factors to adjust the observation noise matrix of Extended Kalman filter (EKF) and factor graph optimization (FGO) in real time to reduce the influence of observation gross error on navigation and positioning. Then, to solve the problem that loosely coupled PDR/GNSS has no observation redundancy and is easily affected by long-term multi-epoch anomaly observation, the prediction innovation is used to reconfigure the effective observations, and the reconfigured observation is put into FGO to realize navigation and positioning. The experimental results show that the resilient factor graph algorithm based on filtering cooperation has higher positioning accuracy and robustness than classical FGO.

 Key words: factor graph optimization; Kalman filter; Resilient navigation; Reconfigure;

1 Introduction

Pedestrian navigation based on smart terminals is experiencing rapid development due to the devices multi-sensor integration and efficient on-chip data processing capabilities. Pedestrian dead reckoning/Global Navigation Satellite System (PDR/GNSS) integration is the classic solution for pedestrian navigation in outdoor scenarios. Smart terminals typically include sensors such as gyroscopes, accelerometers, and magnetometers. In PDR, pedestrian gait is detected using sensor information from accelerometers, allowing for the estimation of stride length. This is then combined with data from magnetometers to calculate heading, enabling the estimation of pedestrian displacement and position based on stride length and heading. However, due to limitations related to cost, size, and power consumption, smart terminals often utilize low-precision Micro-Electro-Mechanical Systems (MEMS) gyroscopes and accelerometers, resulting in rapid divergence of PDR accuracy with increasing walking distance. The combination with GNSS can effectively overcome these drawbacks. When GNSS is unavailable, PDR can provide short-term high-precision location

information; when GNSS is functioning well, it can correct the positioning errors of PDR. The integration of GNSS and PDR forms a more robust combined navigation system. However, in complex environments such as urban canyons and high-rise residential buildings, it faces challenges due to GNSS vulnerability and PDR's turning errors. While loosely coupled PDR/GNSS based on Extended Kalman filter (EKF) is widely used, it lacks observation redundancy and is susceptible to GNSS gross errors[1]. Robust filtering methods like RKF-based, and optimized particle filters have been proposed to address these issues [2][3]. However, these methods rely on the Markov property assumption, which is restrictive in complex environments[4]. Factor graph optimization (FGO) offers a promising approach for addressing non-linearity and non-Gaussian noise[5]. However, FGO suffers from challenges in recognizing gross errors, initial state selection, and lacks an error covariance matrix for posterior state evaluation [6][7]. For the multi-sensor integrated navigation, the current literature focuses on the filtering-based integration and optimization-based integration independently in most cases. However, filtering and optimization are not mutually exclusive, but rather, their cooperation and complementarity represent a significant future direction. This paper proposes a resilient factor graph optimization cooperated with filter theory and evaluates it with PDR/GNSS integrated navigation. The algorithm utilizes IGG-III function-based weight factors to mitigate gross errors in EKF and FGO synchronously. Then, EKF-based solutions are used as initial values for FGO iteration. At the same time, prediction innovation is employed to reconfigure faulty observations and enhance the continuity. This research aims to improve the resilience and reconfigurability of FGO-based multi-sensor integration, paving the way for more accurate and resilient PDR/GNSS navigation in complex environments.

2 PDR/GNSS integrated navigation algorithm

2.1 PDR Mechanism

The PDR algorithm utilizes sensor information from smart terminals, such as smartphones, employing key steps (gait detection, stride length estimation, and heading estimation). This allows it to obtain a pedestrian's stride length and heading angle per unit time. Based on the current target's position and heading, the algorithm predicts the target's position and heading for the next epoch, enabling real-time estimation of the target's location [8].

Assume that the position, step length, and heading Angle of the previous epoch user are (x_{k-1}, y_{k-1})、SL_{k-1}、θ_{k-1}, respectively, the position of the user at the current epoch time can be solved by formula (1)

$$\begin{cases} x_k = x_{k-1} + SL_{k-1} \cdot \cos(\theta_{k-1}) \\ y_k = y_{k-1} + SL_{k-1} \cdot \sin(\theta_{k-1}) \end{cases} \tag{1}$$

Where x_{k-1}、y_{k-1} is the east and north position coordinates of the epoch $k-1$; SL_{k-1} and θ_{k-1} is the step length estimate and heading estimate from epoch $k-1$ to k.

2.2 FGO-based PDR/GNSS Integration

As mentioned earlier, PDR achieves continuous navigation and positioning by step detection, step length estimation, and heading estimation to obtain position increments between consecutive

epochs. Considering the PDR positioning principle and the trajectory recursion equation corresponding to formula (1), the PDR factor based on the position increment constraint can be expressed as:

$$\| err_{k+1}^{PDR} \|^2_{\Omega_{k+1}^{PDR}} = \| p_{k+1} - f(p_k) \|^2_{\Omega_{k+1}^{PDR}}$$

$$= \left\| p_{k+1} - \left(p_k + \begin{bmatrix} SL_k \cdot \cos(\theta_k) \\ SL_k \cdot \sin(\theta_k) \end{bmatrix} \right) \right\|^2_{\Omega_{k+1}^{PDR}} \tag{2}$$

Where err_{k+1}^{PDR} and Ω_{k+1}^{PDR} respectively represents the error function and covariance matrix of PDR factor under epoch $k+1$; p_k and p_{k+1} representing the positions of the pedestrian at epoches k and $k+1$. Similarly, the cost function corresponding to the GNSS factor is:

$$\| err_{k+1}^{GNSS} \|^2_{\Omega_{k+1}^{GNSS}} = \| p_{k+1} - p_{k+1}^{GNSS} \|^2_{\Omega_{k+1}^{GNSS}} \tag{3}$$

Where err_{k+1}^{GNSS} and Ω_{k+1}^{GNSS} respectively represents the error function and covariance matrix of GNSS factor under epoch $k+1$.

The state is related to the propagation model. Therefore, it is more feasible to estimate a set of states using all previous measurements. The PDR state transition equations can be regarded as constraints and connections between states, and the corresponding GNSS information also serves as a constraint on the state. Therefore, in a GNSS/PDR integrated navigation system, the estimation of the optimal state can be transformed into minimizing the cost function of all constraints. Suppose that the navigation system state set x contains the state from the first epoch to the last epoch, i.e., $x = (x_1, x_2, \cdots, x_{k+1})$ Considering all PDR and GNSS factors, the optimal estimator of the state set is:

$$x = arg\min \left(\sum_{i=1}^{k} \left(\| err_{i+1}^{PDR} \|^2_{\Omega_{i+1}^{PDR}} + \| err_{i+1}^{GNSS} \|^2_{\Omega_{i+1}^{GNSS}} \right) \right) \tag{4}$$

Where $arg\min(\cdot)$ represents the minimization of the cost function, and the optimal estimation of the state set can be obtained by minimizing the cost function, typically using the Levenberg-Marquardt (LM) algorithm.

3 Resilient factor graph optimization cooperated with robust filtering

It is obvious that GNSS factor covariance matrix Ω_{k+1}^{GNSS} is very important to the accuracy of FGO, but the classical FGO can not accurately estimate it in complex scenarios. Therefore, the robust filtering based on IGG-III function as shown in formula (5) is proposed to cooperate with resilient FGO.

$$a_i = \begin{cases} 1 & |\gamma_i| \leqslant k_0 \\ \dfrac{k_0}{|\gamma_i|} \left(\dfrac{k_1 - |\gamma_i|}{k_1 - k_0} \right)^2 & k_0 < |\gamma_i| < k_1 \\ 0 & |\gamma_i| > k_1 \end{cases} \tag{5}$$

Where γ_i is the standardized innovation of the epoch i; k_0 and k_1 is the robust coefficient. Generally, the value k_0 ranges from 1.0 to 1.5 and k_1 ranges from 2.5 to 3.0.

The classical EKF algorithm is also unable to detect and isolate the observed gross error

effectively, but the REKF algorithm based on IGG-III function calculates the applicable weight matrix for each iteration to ensure the robustness of the estimates The difference between REKF algorithm and EKF algorithm lies in the construction of Kalman gain:

$$K_k^* = \hat{P}_{k/k-1} H^T \left(H \hat{P}_{k/k-1} H^T + R_k a_k^{-1} \right)^{-1} \tag{6}$$

Where K_k^* is the variance expansion Kalman gain matrix; a_k is the expansion matrix and the diagonal elements of the expansion matrix based on IGG-III function.

A test statistic is derived by standardizing the innovation vector, which is calculated as the difference between the current observation and the EKF's one-step prediction, using the corresponding covariance matrix. When the test statistic is less than k_0, the GNSS observation is judged to be trouble-free, no processing is carried out. When the test statistic is greater than k_0 and less than k_1, the GNSS observation is judged to be deviated, the integrated navigation system will use the weight factor calculated from IGG-III function to expand the observed noise matrix. When the test statistic is greater than k_1, the GNSS observation is judged to be a fault, the IGG-III function will zero weight the observation. However, loosely coupled PDR/GNSS lacks observation redundancy, the zero-weight processing of observation faults for a long time can only delay the divergence of navigation system. Therefore, this paper will reconfigure abnormal GNSS observation faults instead of zero-weight processing of IGG-III function, and the specific steps are as follows:

1) Innovation prediction based on classical exponential smoothing algorithm

Suppose that from epoch 1 to k, the corresponding innovation sequence is γ_1, γ_2, γ_3, … γ_k, then the first exponential smoothing predicted value $S_k^{(1)}$ of the epoch is:

$$S_k^{(1)} = S_{k-1}^{(1)} + \alpha \left(\gamma_k - S_{k-1}^{(1)} \right) = \alpha \gamma_k + (1-\alpha) S_{k-1}^{(1)} \tag{7}$$

Where $S_{k-1}^{(1)}$ is the exponential smoothing algorithm predicted value of the epoch $k-1$; α is the smoothness coefficient and the general choice $\alpha = 0.3$. To enhance the adaptability and robustness of the prediction, a sliding window approach is employed with the exponential smoothing algorithm, considering the dynamic nature of the innovation sequence data. A window size of 10 is chosen to optimize efficiency and accuracy, predicting the innovation sequence value for epoch using this sliding window method, weighted coefficient follows an exponential decay pattern and can be used for smoothing the innovation sequence.

$$S_k^{(1)} = \alpha \gamma_k + \alpha(1-\alpha) \gamma_{k-1} + \alpha(1-\alpha)^2 \gamma_{k-2} + \cdots + \alpha(1-\alpha)^9 \gamma_{k-9} \tag{8}$$

2) Fault amplitude and observation reconfigure

Using exponential smoothing algorithm to predict the innovation value and the abnormal innovation vector, the amplitude of the observed fault can be obtained and used to reconfigure the abnormal observation:

$$A_k = \left| \tilde{\gamma}_k - \gamma_k \right| \tag{9}$$

$$\tilde{z}_k = z_k \pm A_k \tag{10}$$

Where A_k represents the anomaly observation amplitude of the epoch k; $\tilde{\gamma}_k$ represents

exponential smoothing forecast innovation vector; γ_k represents innovation vector with anomalous observations. $\widetilde{z}_k$ represents the reconfigured observation values of the epoch k; z_k represents the abnormal observation values of the epoch k. After obtaining the reconfigured observation value, the reconfigured observations are applied to the factor graph and a new GNSS factor is obtained to realize the global optimization of state estimation. The flow of resilient factor graph with EKF collaboration is shown in **Fig 1**.
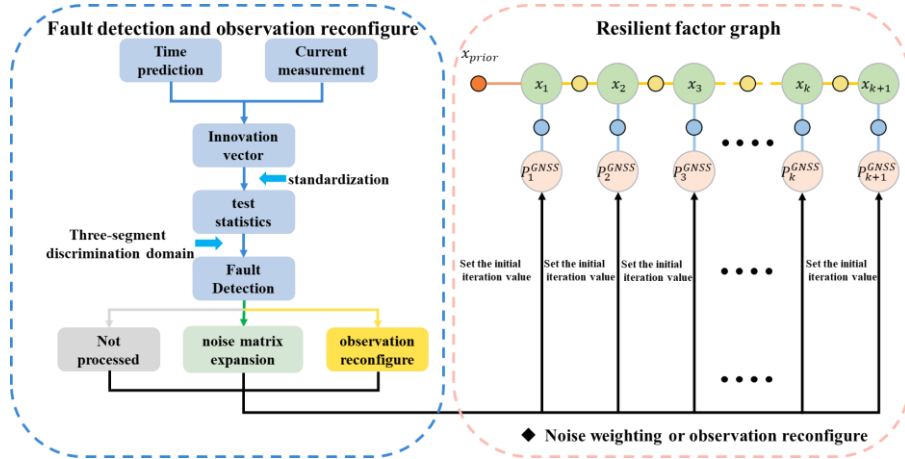
Fig 1   Flow chart of resilient factor graph Cooperated with Filtering

4 Evaluation and analysis of actual experiments

This section will evaluate the superiority of the proposed algorithm in PDR/GNSS navigation. The experiment utilized a Redmi K60 smartphone and an Android app to collect internal sensor data. The experimental scenario data was collected near high-rise residential areas at Southeast University, and the experimental route included the occlusion conditions required for complex environments, as shown in **Fig 2**(a). Simultaneously, the positioning results computed by the NovAtel SPAN-CPT system as shown in **Fig 2**(b) was used as a reference to calculate horizontal positioning errors.

(a) Experimental scenario and route   (b) Datasets collecting equipment

Fig 2   Data acquisition scenario and integrated navigation reference device

The following is a comparison of GNSS positioning, EKF algorithm, FGO algorithm, resilient FGO cooperated with filtering theory (R-FGO) and Resilient FGO enhanced by observation reconfiguration (OR-FGO) proposed in this paper. The trajectory and horizontal position error

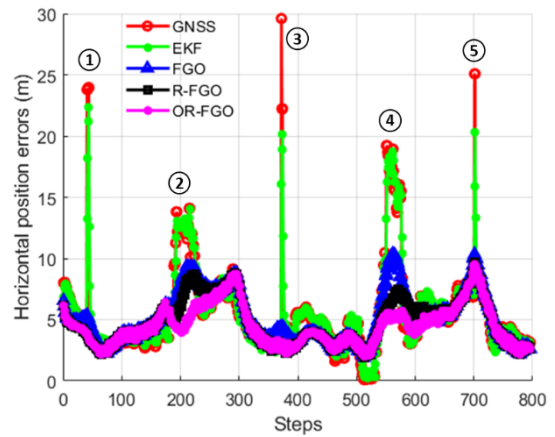results of five different algorithms are shown in **Fig 3** and **Fig 4** respectively. According to the GNSS horizontal positioning error corresponding to the red line in **Fig 4**, there are 5 abnormal GNSS observation mutations. Among them, the first, third and fifth segments are single epoch faults; The second and fourth segments are multi-epoch faults. It can be seen that algorithms except GNSS and EKF can detect and isolate single-epoch faults well. However, for multi-epoch faults, FGO algorithm and R-FGO algorithm can constrain multi-epoch faults but cannot completely eliminate their effects. The OR-FGO algorithm can not only play a good anti-error role for single epoch faults, but also play a good role in eliminating multi-epoch faults. **Tab.1** shows the statistical results of the position errors of the five positioning algorithms. It can be seen that OR-FGO algorithm is superior to all other methods, and the mean horizontal position error is reduced by 19.64%,12.28% and 6.25% comparing with EKF, FGO and R-FGO algorithm, respectively. The maximum horizontal position error is reduced by 57.2%,7.90% and 0.93% comparing with EKF, FGO and R-FGO algorithm, respectively.
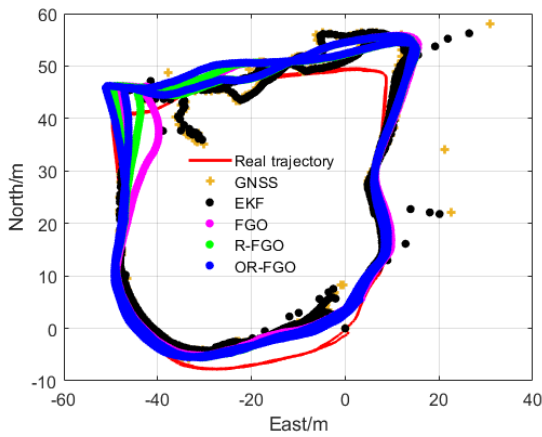
Fig.3 Schematic diagram of trajectories for different schemes   **Fig.4** Horizontal position error in complex scenes

Tab.1 Performance of the various positioning methods

method Horizontal error	GNSS	EKF	FGO	R-FGO	OR-FGO
MEAN(m)	5.55	5.60	5.13	4.80	4.50
RMS(m)	6.70	6.58	5.54	5.14	4.78
STD(m)	3.76	3.45	2.08	1.86	1.61
MAX(m)	29.5978	22.34	10.38	9.65	9.56

4　Conclusion

A resilient FGO algorithm cooperated with filtering theory is proposed to solve the problem of adaptive identification of observed faults and lack of covariance matrix to evaluate the posterior state for classical FGO. The algorithm addresses the inaccurate estimation of the measurement covariance matrix in classical graph optimization navigation algorithms by introducing a resilient factor graph based on IGG-III function coordination. Simultaneously, to further enhance convergence speed, avoid local minima, and improve robustness, the solution from the EKF is used as the iteration initial values for the FGO. Additionally, the algorithm tackles the issue of loosely

coupled PDR/GNSS location dispersion under multi-epoch observation faults by utilizing historical innovation fitting fault innovation to reconfigure GNSS observations and applying these reconfigured observations to FGO for optimal estimation. Experimental results demonstrate the effectiveness of the proposed method in utilizing observational data and processing it accordingly, ultimately enhancing the availability and continuity of positioning information for pedestrian navigation.

References:

[1]   Zhu N, Marais J, Betaille D, et al. GNSS position integrity in urban environments: A review of literature[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(9): 2762-2778.

[2]   Jiang C, Chen Y, Liu Z, et al. A Probabilistic Method-Based Smartphone GNSS Fault Detection and Exclusion System Utilizing PDR Step Length[J]. Remote Sensing, 2023, 15(20): 4993.

[3]   Guoliang X U, Zhen L I, Jun T A O, et al. RTK+ PDR fusion positioning for smart phone in complex environment[J]. Bulletin of Surveying and Mapping, 2021 (12): 44.

[4]   Wen W, Pfeifer T, Bai X, et al. Factor graph optimization for GNSS/INS integration: A comparison with the extended Kalman filter[J]. NAVIGATION: Journal of the Institute of Navigation, 2021, 68(2): 315-331.

[5]   Jiang C, Chen Y, Chen C, et al. Smartphone PDR/GNSS integration via factor graph optimization for pedestrian navigation[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-12.

[6]   Watson R M, Gross J N. Robust navigation in GNSS degraded environment using graph optimization[C]//Proceedings of the 30th international technical meeting of the satellite division of the institute of navigation (ION GNSS+ 2017). 2017: 2906-2918.

[7]   Meng Q, Song Y, Li S, et al. Resilient tightly coupled INS/UWB integration method for indoor UAV navigation under challenging scenarios[J]. Defence Technology, 2023, 22: 185-196.

[8]   Xu L, Xiong Z, Liu J, et al. A novel pedestrian dead reckoning algorithm for multi-mode recognition based on smartphones[J]. Remote Sensing, 2019, 11(3): 294.

Modern Precision of Attitude Determination Using Multi-Antenna Satellite Navigation System

F.S. Kapralov

(Lomonosov Moscow State University, Moscow, Russia)

Abstract: The report focuses on the analysis of the precision of attitude determination by range satellite measurements from a multi-antenna navigation system. Results are presented from the processing of a real experiment with five satellite antennas within one meter of each other. A reference orientation is obtained with a precision multiple times higher than the current typical precision of attitude determination using a multi-antenna satellite system. This allows for a comprehensive examination of attitude angle errors.

1 Introduction

The problem of attitude determination with sub-tenths of a degree accuracy occurs frequently in many applications. It is well-known that the problem can be solved with the accuracy by a high-grade strapdown inertial navigation system (INS). The main drawbacks that make the use of INS in some applications unreasonable are its typical mass, size, power supply, and price. To overcome these strapdown INS limitations, multi-antenna Global Navigation Satellite Systems (GNSS) with more than two antennas can be used to determine attitude angles to within 0.1°. The modern accuracy of attitude determination using dual- or triple-antenna systems is about the first tenth of a degree, adjusted to a 1-meter baseline length [1]. A baseline is defined as a vector between the phase centers of GNSS antennas.

The most common approach to solving the problem is as follows. For each baseline, the coordinates of which are known a priori in the body reference frame, the estimate of the baseline coordinates in the navigation, e.g., geodetic, reference frame must be obtained. So, for each baseline we have a linear mixed integer least squares problem. The problem arises from the integer ambiguity of the phase measurement difference and, as a rule of thumb, is solved numerically using LAMBDA [2]. We use its numerical modification, so called MLAMBDA [3]. After resolving the integer ambiguity and estimating the baseline coordinates in the navigation and body reference frames, it is possible to determine the attitude of the body reference frame relative to the navigation frame by solving the optimization problem for the corresponding attitude matrix [4].

2 Real experiment

In August 2023, a multi-antenna GNSS experiment was performed in Moscow under relatively favorable urban conditions. GNSS antennas were mounted on a wooden board, which is the body of the navigation system carrier object. The board is positioned approximately horizontally. During the record of the satellite measurements, 7 static positions were realized using three 90° rotations around the vertical in a clockwise direction and then three similar rotations in the opposite direction. The duration of each static position is 20 minutes, and the total duration of the experiment is approximately 2.5 hours. The main features of the experiment are as follows:

- multiple antennas: 2 Antcom G8Ant GNSS antennas are connected to the NovAtel PwrPak7D GNSS receiver, 1 NovAtel GNSS-804 GNSS antenna is connected to the NovAtel PwrPak7 GNSS receiver, 2 Javad AirAnt GNSS antennas are connected to the 2 Javad Prego GNSS receivers;

- multiple GNSS: all GNSS receivers process GPS and GLONASS signals, and NovAtel receivers also process Galileo and Beidou signals;

- short baselines: the maximum distance between any two GNSS antennas is less than 1 meter.

3 Attitude angles reference

The main objective of the experiment is to estimate the accuracy of attitude determination using multi-antenna GNSS. The experiment is designed to provide reference attitude angles with an accuracy several times better than the typical accuracy of the multi-antenna GNSS solution. In this work, the attitude of the carrier object relative to a geodetic reference frame is described by the three angles of roll, pitch, and yaw (heading).

4 Reference yaw angle

The carrier object was in the initial position after the end of the main experiment. One of the five antennas was then repositioned so that the long baseline (more than 4 meters) could be formed with all the other antennas on the wooden board. Satellite measurements were recorded separately for the five antennas in this static configuration. The antenna positions relative to the body reference frame were measured, and an estimate of its accuracy was obtained. Exploiting the fact that the yaw angle error is inversely proportional to the baseline length, a reference yaw estimate can be obtained for the short baselines in the main experiment.

5 Reference pitch and roll angles

To estimate the reference roll and pitch angles in all static positions of the experiment, accelerometer measurements from a pre-calibrated microelectromechanical INS were used. The INS was mounted on top of the wooden board together with GNSS antennas. We reduce the problem of estimating pitch and roll angles to a least squares problem. In the measurement model, we consider a variability of the object body's rotational axis, and angular misalignments of the INS instrumental axes relative to the carrier body reference frame. An upper bound on the accuracy of the roll and pitch angles for each baseline is also estimated.

6 Attitude angles analysis

The results of analyzing the attitude angle errors in the experiment for different baselines are presented. The attitude errors of the baselines are decomposed into two components: low-frequency systematic and high-frequency stochastic. For each component, its partial influence on the total attitude error is analyzed.

References:

[1] Raskaliyev, A., Patel, S.H., Sobh, T.M., Ibrayev, A. GNSS-based attitude determination techniques—a comprehensive literature survey. IEEE Access, 8, 2020, pp. 24873-24886. DOI: 10.1109/ACCESS.2020.2970083.

[2] Teunissen P.J. Least-squares estimation of the integer GPS ambiguities. Invited lecture, section IV theory and methodology, IAG general meeting, Beijing, China, 1993, pp. 1-16.

[3] Chang, X.W., Yang, X., Zhou, T. MLAMBDA: A modified LAMBDA method for integer least-squares estimation. Journal of geodesy, 79, 2005, pp. 552-565. DOI: 10.1007/s00190-005-0004-x.

[4] Wahba G. A least squares estimate of satellite attitude. SIAM review, 7(3), 1965, pp. 409-409. DOI: 10.1137/1007077.

2024

Chinese-Russian
"Navigation and Motion Control"
Youth Forum

Topic 4: Gravity-Assisted Navigation

The Current State of Strapdown Airborne Gravimetry. Methodology of Gravimeter Data Postprocessing

A.A. Golovan, V.S. Vyazmin

(Lomonosov Moscow State University, Moscow, Russia)

Abstract: Airborne gravimetry is a branch of geophysics that studies the methods of measuring Earth's gravity on board an aircraft. The paper discusses the current state of airborne gravimetry and the methodology of data postprocessing in the case of using a strapdown gravity meter based on a strapdown inertial navigation system and GNSS receivers.

1 Introduction

Airborne gravimetry is a method of airborne geophysics that deals with determining the vertical component of the gravity disturbance vector (gravity disturbance) from airborne gravimeter measurements on board an aircraft (a fixed-wing aircraft or helicopter). Inertial gravimeters based on a gyro-stabilized platform are traditionally used in airborne gravimetry, e.g., L&R, Chekan-AM, GT-2A and others instruments [1]. In the paper, we discuss the use of a strapdown airborne gravimeter, which is alternative to the traditional inertial gravimeters. A strapdown airborne gravimeter is based on a strapdown inertial navigation system (INS) (or an inertial measurement unit) and global navigation satellite system (GNSS) receivers [2-4]. The gravity disturbance determination from strapdown gravimeter measurements is carried out in the posprocessing mode and includes several stages. The paper describes the key stages of data postprocessing (the determination of GNSS solutions, INS initial alignment, and INS-GNSS integration) and the corresponding approaches (algorithms). A brief description of the developed methodology for raw data postprocessing is presented.

2 Methodology of strapdown gravimeter data postprocessing

We formulate the gravimeter data postprocessing problem as determining the gravity disturbance $\Delta g = g - g_0$ (where g and g_0 are the vertical components of the actual and normal gravity vectors, respectively) on the aircraft flight path from the raw measurements of the gravimeter's INS inertial sensors (accelerometers and gyroscopes) and GNSS onboard and ground-based receivers. The developed data postprocessing methodology includes the following stages:

1.Raw data quality control (checking for possible data loss and instruments failures, analyzing the temperature variations inside the gravimeter, etc.).

2.The initial alignment of the gravimeter's INS and calibration of the INS accelerometers (determining the in-run biases and scale factor errors).

3.Determining the GNSS antenna position and velocity from the raw GNSS measurements (using the carrier-phase differential mode).

4. The INS-GNSS integration.

5.Determining the gravimetric solution (gravity disturbance estimation on the flight path).

Now we briefly describe the key aspects of the developed data postprocessing merthodology. At the INS initial alignment (stage 2), the attitude angles of gravimeter's INS at the aircraft standstill are determined and the INS accelerometers are calibrated (the accelerometer biases and scale factor errors are estimated) [2]. The difficulty of the problem is the uncontrolled angular motion of the gravimeter due to the wind at the aerodrome, turning on and off the aircraft engines, the work of the crew, etc.

At the INS-GNSS integration (stage 4), the parameters of the aircraft flight path and the INS inertial sensor instrumental errors are estimated. The mathematical formulation of the problem is based on the INS error propagation equations and observation equations derived for a GNSS aided INS (with the GNSS positions and velocities used as the observations) [1, 5]. The state vector of the constructed spate-space model includes the following parameters:

- the latitude, longitude and height of the accelerometers proof mass above the reference ellipsoid;
- the INS dynamic velocity errors (the east and north components);
- the INS attitude errors (the misalignments of the vertical and azimuthal error);
- the accelerometers and gyroscopes in-run biases;
- the GNSS antenna offsets with respect to the INS (lever arm);
- the time-synchronization errors.

It is well known that the INS-GNSS integration problem given the measurement error stochastic models is reduced to the standard linear optimal estimation problem. The solution to the problem is provided by Kalman filtering and smoothing [6].

The formulation of the gravimetric problem (stage 5) is based on the fundamental equation of airborne gravimetry [1], which is the equation of the INS accelerometers proof mass motion. We use the form of the equation in the projection onto the geodetic frame (East, North, Up). However, other forms can also be used [1]. After replacing the unknown terms in the equation by the inertial sensor measurements and GNSS data, the gravimetric problem can be reduced to a standard linear optimal estimation. The state vector includes the following unknown parameters: the gravity disturbance, residual angular errors (misalignments of the vertical), vertical accelerometer scale factor error, lever-arm errors, and time-synchronization errors. The estimates of the parameters are provided by the Kalman filter and optimal smoother.

3   Numerical example of gravimeter data postprocessing

In Fig. 1, the numerical results from the airborne gravimeter data postprocessing using the developed methodology are presented. The survey flight was carried out in 2021 with the use of a strapdown gravimeter and the traditional gravimeter based on a stabilized platform (GT-2A) [1]. A fixed-wing aircraft was used in this survey flight. In Fig. 1 (on the left), the position errors of the strapdown gravimeter's INS dead-reckoning (the autonomous mode) are presented. The GNSS positions were used as a reference solution. The maximum error of the gravimeter's INS position solution is 3 km, which is the accuracy level of the gravimeter's sensors after the calibration.

In Fig. 1 (on the right), the gravity disturbance estimates based on the raw data from the strapdown and traditional airborne gravimeters are presented. The standard deviation of the difference between two gravity disturbance estimates is 1.1 mGal, which shows a good agreement of the results. Also shown are the gravity data derived from an Earth's global gravity field model of high degree and order (XGM2019) [7]. Fig. 1 shows that the global gravity data have much lower spatial resolution comparing to the arirborne gravimetry gravity data.
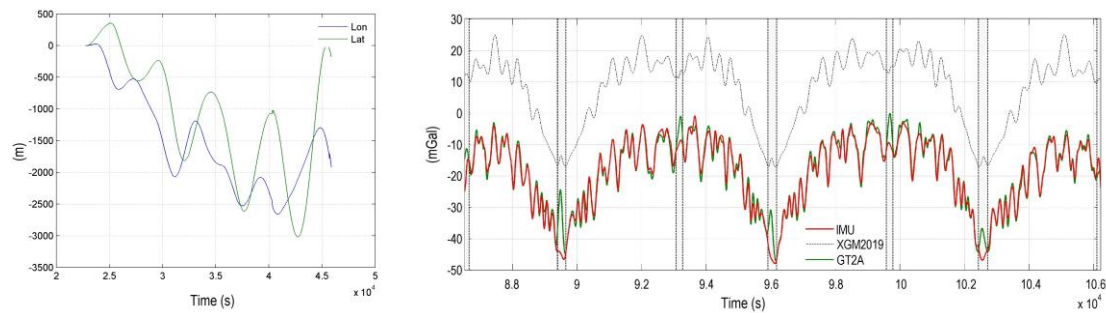
Fig. 1 On the left: the strapdown gravimeter's INS dead-reckoning errors, km. On the right: the gravity disturbance estimates from the strapdown airborne gravimeter data (red) and stabilized-platform gravimeter data (green); the global gravity field model data (black), mGal.

4 Conclusions

The current trend in airborne gravimetry is to use strapdown gravity meters rather than traditional ones, which are based on stabilized platforms. In the paper, the key aspects of postprocessing raw data from a strapdown gravimeter are briefly described. The numerical example from a survey flight with a state-of-the-art strapdown gravimeter are also presented. The results of the data postprocessing (the gravity disturbance estimates) are in good agreement with the results from a classical airborne gravimeter.

References:

[1] Methods and technologies for measuring the earth's gravity field parameters. Cham, Switzerland, Springer. 2022. 388 p.

[2] Golovan A.A., Vyazmin V.S. Methodology of Airborne Gravimetry Surveying and Strapdown Gravimeter Data Processing, Gyroscopy and Navigation, 2023, vol. 14, no. 1, pp. 36-47.

[3] Cai S., Zhang K., Wu M., Huang Y. The method on acceleration extraction from strapdown airborne scalar gravimeter based on flight dynamics. Adv. Mater. Res. 2012. P. 2152–2156.

[4] Senobari M.S. New results in airborne vector gravimetry using strapdown INS/DGPS. J. Geod. 2010. V. 84: P. 277–291.

[5] Jekeli C. Inertial navigation systems with geodetic applications. Walter de Gruyter, New York. 2001.

[6] Kailath T., Sayed A.H., Hassibi B. Linear Estimation. Prentice Hall, Upper Saddle River, NJ, USA, 2000. 854p.

[7] Zingerle P., Pail R., Gruber T., Oikonomidou X. The combined global gravity field model XGM2019e. J Geod. 2020. V. 94. P. 1-12

The Development and Application of Gravity Compensation Inertial Navigation Technology

Hu Qiqin[1], Luo Kaixin[1], Xiong Zhiming[1], Yu Ruihang[1], Cao Juliang[1]

(1. College of Intelligent Science and Technology, National University of Defense Technology, Changsha, Hunan 410073, China)

Abstract: In response to the complex marine environment, this paper primarily explores methods to enhance the accuracy of Inertial Navigation Systems(INS) using gravity compensation techniques. Starting with the current demands for gravity disturbance compensation, the paper begins by defining models of gravity disturbances and analyzes the propagation mechanism of INS errors caused by these disturbances. It introduces various compensation methods, including spherical harmonic model compensation, actual data compensation, filtering estimation compensation, real-time measurement compensation, multi-source information fusion compensation, and spectral analysis compensation, and evaluates the strengths and weaknesses of each approach. Finally, the paper summarizes the development prospects of gravity compensation technology in underwater navigation applications and proposes several suggestions for improvement. Despite the numerous challenges faced by gravity compensation technology, it is of great significance for improving underwater navigation accuracy. The research presented in this paper provides a reference for realizing "dock-to-dock" passive autonomous navigation research.

 Key words: Marine; Inertial Navigation System; Gravity Disturbance; Gravity Disturbance Compensation;

1 Introduction

In the complex and variable marine environment, ships and vessels face numerous challenges, among which the most pressing is achieving high-precision underwater navigation and positioning. The difficulty of underwater navigation lies not only in the positioning itself but also in the complex and variable underwater environment rendering many traditional navigation methods (such as satellite navigation) ineffective. In contrast, INS relies on their inertial measurement units for measurement, without relying on external data input, thus maintaining high autonomy and passivity, making it the significant means of underwater navigation. However, the accuracy of positioning in an INS is significantly affected by the manufacturing processes and algorithms used in the device. Over time, errors can accumulate, leading to decreased accuracy as navigation time increases. Unless effective error correction measures are implemented, the INS may fail to achieve the necessary level of navigation precision. Therefore, the development of integrated navigation

systems with INS as the core and supplemented by various means has become the key direction for the development of underwater navigation and positioning systems.

To address the challenges of underwater navigation, the gravity field has emerged as a vital area of research, particularly when compared to terrain matching and geomagnetic matching. As a stable characteristic of the Earth that is passive and non-intersecting, the gravity field is closely tied to the Earth's intrinsic properties and exhibits a slow rate of change. Its applications are diverse, extending to fields such as surveying, geological exploration, and astrophysics. However, a limitation arises from Einstein's equivalence principle, which states that accelerometers cannot differentiate between acceleration caused by the carrier's motion and gravitational acceleration. Consequently, gravity field information is typically derived from standard gravity field models. This limitation has become a significant factor hindering the advancement of high-precision INS. Gelb noted that the ultimate accuracy of inertial navigation heavily relies on the precision of gravity field information.[1]. Additionally, correcting the systematic errors caused by vertical deflection is essential for achieving INS with an accuracy of 0.001°/h or better.[2]. Although international high-precision inertial navigation indicators have already surpassed 1n mile/15 days[3, 4], there is still a long way to go to achieve the "dock-to-dock (Port-to-Port)" goal of underwater passive autonomous navigation[5]. Therefore, accurately extracting gravity disturbance information and effectively integrating it into the inertial navigation system for compensation is crucial for advancing the development of gravity-aided inertial navigation technology and forms the basis for future research and technological innovation[6].

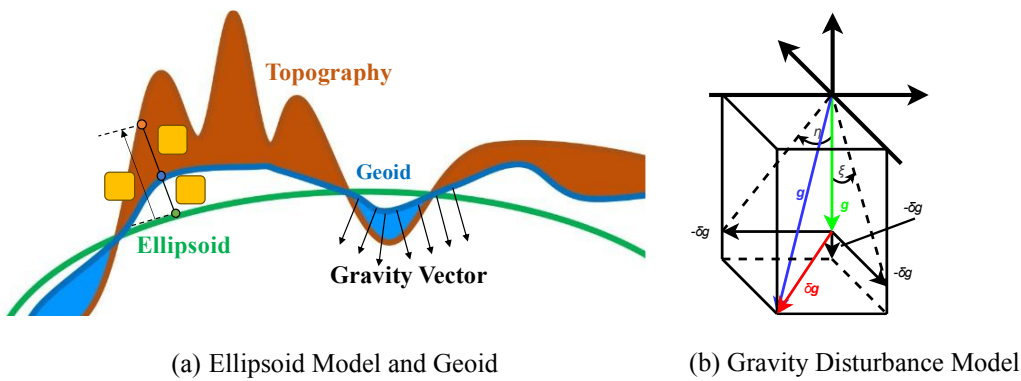## 2 Gravity disturbance definition

The Earth can be approximated as an ellipsoid that rotates uniformly around its minor axis, with its surface forming an equipotential surface. Under this assumption, the calculation of normal gravity can be realized through the Taylor series expansion of the Somigliana formula[7]. However, in reality, the Earth is not an ideal rotating ellipsoid, hence there is a discrepancy between the actual gravity and the normal gravity, as shown in Fig.1, where H represents the geoid height, N represents the height of the geoid relative to the ellipsoid surface, and $h_t = H + N$ represents the topographic height. The difference between the actual gravity and the normal gravity $\delta g = g - g_0$ is commonly defined as the gravity disturbance. In the local east-north-up coordinate system, the projection of the gravity disturbance in the direction of normal gravity is the gravity anomaly, $\Delta g = -\delta g_U$. The angle between the actual gravity and the normal gravity in the horizontal direction is called the deflection of the vertical(DOV), with the component in the meridian direction denoted as η and the component in the prime vertical direction denoted as ξ. When the deflection of the vertical satisfies the small angle condition, the approximate relationship as follow

$$
\begin{aligned}
\eta &\approx -\delta g_E / \gamma \\
\xi &\approx -\delta g_N / \gamma
\end{aligned}
\tag{1}
$$

Based on the geometric relationships, the connection between the true gravity, normal gravity, deflection of the vertical, and gravity anomaly is as follows:

$$
\begin{aligned}
\boldsymbol{g}^n &= \boldsymbol{\gamma}^n + \boldsymbol{\delta g} \\
&= \begin{bmatrix} -\gamma\eta & -\gamma\xi & -\gamma - \Delta g \end{bmatrix}
\end{aligned}
\tag{2}
$$

Specifically, a deflection of the vertical of 1 arcsec can result in approximately 5 milliGal (mGal) of horizontal gravity error[8]. Studies have also found that in the Earth's oceanic regions[2], 80% of the areas have an average deflection of the vertical of 5 arcseconds (arcsec), 15% of the areas have 15 arcsec, and 5% of the areas can reach up to 1 arcminute (arcmin). Furthermore, under static conditions, in long-duration navigation exceeding 20 hours, an increase of 10 mGal in the horizontal gravity disturbance component can lead to an increase of about 130 meters in the position error of high-precision shipborne INS[9]. Given that the largest global gravity disturbance value can reach 500 mGal[2], it is particularly important for INS pursuing high precision to compensate for gravity disturbances.

<div align="center">

(a) Ellipsoid Model and Geoid   (b) Gravity Disturbance Model

Fig.1   Associated models and definitions

</div>

3   Impact of gravity disturbance on INS error

Huang *et al.* suggested dividing the development of gravity-aided inertial navigation technology into two phases: gravity compensation and gravity correction is particularly important for INS pursuing high precision to compensate for gravity disturbances[6]. During the gravity compensation phase, the integration of gravity disturbance information into the solution loop of the inertial navigation system aims to optimize the system's mechanical arrangement, thereby suppressing the proliferation of errors. However, error propagation in INS is a dynamic and complex process involving various error sources, including instrument errors, initial condition errors, algorithmic errors, and navigation environment errors. These error sources are coupled with each other, making it difficult to analyze any single variable in isolation. Therefore, this paper will delve into how gravity disturbances affect the solution process of the inertial navigation system and analyze the mechanisms of their impact.

3.1   Gravity disturbance model and error

In inertial navigation, the propagation of error sources over time is described by constructing differential equations for position, velocity, and attitude errors. Based on the gravity disturbance model described in the previous section, within the navigation frame, there is a velocity differential equation in the inertial navigation solution as follows

$$\boldsymbol{\delta g}^n = \dot{\boldsymbol{v}}^n - \boldsymbol{C}_b^n \boldsymbol{f}^b + (2\boldsymbol{\omega}_{ie}^n + \boldsymbol{\omega}_{en}^n) \times \boldsymbol{v}^n - \boldsymbol{\gamma}^n \tag{3}$$

where $\dot{\boldsymbol{v}}^n$ is velocity differential in the n-frame, $\boldsymbol{C}_b^n$ is the attitude transfer matrix from b-

frame to n-frame, f^b, ω_{ie}^n, ω_{en}^n and γ^n are the vector about specific force, the Earth angular velocity projected on n-frame, the angular velocity of n-frame relative to e-frame and the normal gravity value. If the calculation errors of normal gravity are not considered a priori, the observed gravity δg^n can be treated as the actual gravity disturbance. When the velocity differential equation takes into account the gravity disturbance and separates it from the normal gravity, the error model of the gravity disturbance can be further derived from equation (3)

$$
\begin{aligned}
\mathrm{d}\boldsymbol{\delta g}^n = \boldsymbol{\delta \dot{v}}^n + \left[\phi^n \times\right] \boldsymbol{C}_b^n \boldsymbol{f}^b - \boldsymbol{C}_b^n \boldsymbol{\delta f}^b \\
+ \left(2\delta\omega_{ie}^n + \delta\omega_{en}^n\right) \times \boldsymbol{v}^n + \left(2\omega_{ie}^n + \omega_{en}^n\right) \times \boldsymbol{\delta v}^n - \boldsymbol{\delta\gamma}^n
\end{aligned}
\tag{4}
$$

where $\mathrm{d}\boldsymbol{\delta g}^n$ is the error of gravity disturbance, $\left[\phi^n \times\right]$ is a skew-symmetric matrix of ϕ^n, $\boldsymbol{\delta\gamma}^n$ is the error of the normal gravity model. When considering the synchronization time difference $\mathrm{d}T$ between INS and GNSS, equation (4) can indeed incorporate error terms $\ddot{v}^n\mathrm{d}T$. Therefore the measurement errors of gravity disturbances are mainly caused by dynamic acceleration errors $\delta\dot{v}^n$, attitude errors $\left[\phi^n \times\right]\boldsymbol{C}_b^n \boldsymbol{f}^n$, and specific force errors $\boldsymbol{C}_b^n \boldsymbol{\delta f}^b$.

3.2 Impact on velocity

When the gravity disturbance is not considered, the differential equation for velocity error can be obtained as follows,

$$
\begin{aligned}
\boldsymbol{\delta \dot{v}}^n = \boldsymbol{C}_b^n \boldsymbol{\delta f}^b - \left[\phi^n \times\right] \boldsymbol{C}_b^n \boldsymbol{f}^b - \left(2\delta\omega_{ie}^n + \delta\omega_{en}^n\right) \times \boldsymbol{v}^n \\
+ \boldsymbol{\delta v}^n \times \left(2\omega_{ie}^n + \omega_{en}^n\right) + \boldsymbol{\delta g}^n
\end{aligned}
\tag{5}
$$

According to equation(5), the impact of gravity disturbance on the accelerometer measurement of velocity is the most significant, especially in the horizontal channels. This influence further affects the solution of the carrier's position after the horizontal velocity is integrated. Reference[10] observes through simulation experiments with different magnitudes of gravity disturbance that when the gravity disturbance exceeds 100 mGal, the maximum position error of a purely inertial navigation system within a day may exceed 4n miles. However, this study lacks a quantitative analysis of the error propagation characteristics. In contrast, reference [11] models gravity disturbance using a first-order Markov model and quantitatively determines the order of magnitude of position error caused by gravity disturbance through correlation analysis. In addition, since the velocity error is simultaneously affected by the accuracy of specific force measurement and gravity disturbance, the horizontal component of gravity disturbance plays a role in the propagation path of velocity error that is equivalent to the zero bias of the accelerometer. Therefore, gravity disturbance can be regarded as an equivalent item to the accelerometer zero bias. Based on this understanding, Chang et al. conducted in-depth research on the relationship between accelerometer zero bias and gravity disturbance in the horizontal direction[12], and developed corresponding compensation strategies based on the difference in magnitude between the two.

3.3 Impact on attitude

In integrated Strapdown Inertial Navigation Systems (SINS) with GNSS, both theoretical and experimental evidence has shown that the introduction of vertical deflection can lead to tilt errors[13, 14], hence the need for compensation of vertical deflection to improve attitude accuracy.

$$\dot{\boldsymbol{\phi}} = -\boldsymbol{\omega}_{in}^{n} \times \boldsymbol{\phi} + \delta\boldsymbol{\omega}_{in}^{n} - \delta\boldsymbol{\omega}_{ib}^{n} \tag{6}$$

Although the differential equation for attitude error does not explicitly include the term for gravity disturbance, its impact on attitude error can be indirectly understood by analyzing how gravity disturbance affects velocity error. Gravity disturbance first affects velocity measurement, then affects position calculation, and finally, the changes in both of these affect attitude error. Fig.2 shows the error propagation path of gravity disturbance. Initially, gravity disturbance couples with the device error of the accelerometer, producing velocity error after the first time integration. This velocity error leads to position error after further integration. At the same time, velocity error and position error combine with the device error of the gyroscope, forming angular acceleration error, which is transformed into attitude angle error after integration. The attitude angle error causes a rotation bias between the carrier coordinate system and the navigation coordinate system, ultimately affecting the output of the accelerometer and reducing navigation accuracy.
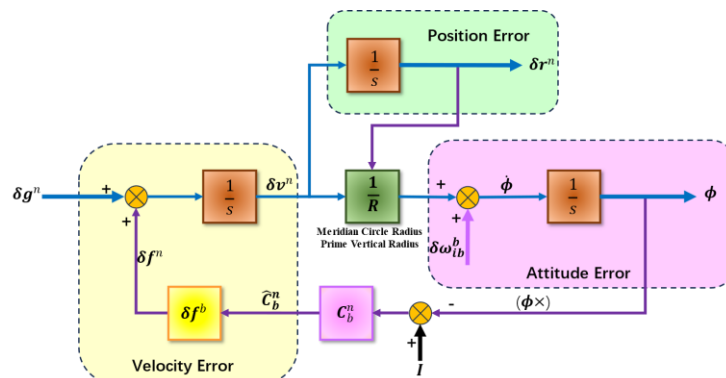
Fig.2 The error propagation of gravity disturbance

Further research indicates that during the motion of the carrier[8, 15-17], when its velocity is high, the attitude error is mainly affected by the mid-low frequency components (wavelength $\lambda_0 > 30km$) of the horizontal gravity disturbance; as the speed decreases, the influence of high-frequency components becomes more significant. Carriers moving at low speeds are more susceptible to the detailed characteristics of the gravity field, which play a more critical role in attitude error. Therefore, accurately capturing and compensating for the high-frequency signals of horizontal gravity disturbance is essential for improving the accuracy of gravity compensation.

3.4 Impact on initial alignment

During the initial alignment phase of an inertial navigation system (INS), if the effect of gravity disturbance is neglected[18, 19], and it is considered as an error in the gravity model and incorporated into the differential equation of velocity error, it will lead to attitude angle errors in the north, east, and vertical directions:

$$\begin{cases} \phi_N = \dfrac{\delta \dot{v}_E - \delta f_E - \delta g_E}{\gamma} = \eta - \dfrac{\delta f_E}{\gamma} \\[2mm] \phi_E = \dfrac{-\delta \dot{v}_E + \delta g_N + \delta f_N}{\gamma} = -\xi + \dfrac{\delta f_N}{\gamma} \\[2mm] \phi_D = \left(\dfrac{\delta f_E}{\gamma} - \eta \right) \tan L + \dfrac{\delta \omega_{ib,E}^n}{\omega_e \cos L} \end{cases} \tag{7}$$

Specifically, the attitude angle errors in the north and vertical directions will be associated with the eastward vertical deflection, while the attitude angle error in the east direction will be related to the northward vertical deflection. This finding emphasizes the importance of considering gravity disturbances during the initial alignment process to ensure an accurate estimation of attitude angle errors.

Researchers have conducted an in-depth analysis of the impact of gravity disturbances during the initial alignment process of INS. Zhu *et al.* utilized marine gravity field information and a Laser Gyro Attitude and Heading Measurement System (AHMS) to analyze and simulate the impact of Different Operational Verticals (DOV) on the initial alignment accuracy under various working modes[15]. The results indicated that the alignment accuracy of a single-axis rotation system decreases with an increase in DOV, and as the alignment time extends, the attitude error gradually stabilizes to the theoretically calculated value. Similarly, reference [20] fine-tuned the attitude error during the initial alignment process of a single-axis rotation inertial navigation system using a rotation modulation method and established a theoretical limit error equation considering the coupling between the Inertial Measurement Unit (IMU) and DOV. However, some research found that after using the EGM2008 model for gravity disturbance compensation during the initial alignment process[21], the accuracy of the inertial navigation system actually decreased. This may be due to errors inherent in the model, resulting in inaccurate estimation of the attitude transfer matrix after compensation, thus introducing additional errors. Deviating from traditional alignment methods, reference [22] proposed an innovative dual-alignment method that improves the precision and efficiency of vertical deflection estimation by dual-aligning and tracking the attitude of the Inertial Navigation System (INS). Despite this, the method mainly focuses on the estimation of DOV and may not fully consider the impact of other error sources in the system. Reference [23] addressed the misalignment between the alignment coordinate system and the navigation coordinate system caused by gravity disturbances by proposing two compensation methods based on velocity and attitude, and experiments showed that these methods could improve the positioning accuracy of ship navigation by about 10%. In addition, reference [24] found that in the absence of external observations, Zero Velocity Update (ZUPT) has a certain constraining effect on the divergence of errors in the inertial navigation system, especially when facing rapidly changing horizontal gravity anomalies, effective reduction of horizontal position errors can be achieved through gravity compensation[25].

4 Compensation method

Current gravity disturbance compensation strategies encompass a variety of methods, including compensation based on spherical harmonic models, pre-stored data interpolation, filtering

estimation, real-time measurement, multi-source data fusion, and frequency band information compensation. Each method has its unique advantages and limitations. The specific choice of method should be combined with the specific application scenario, data quality, and hardware requirements. By carefully designing strategies or effectively combining various methods, better compensation effects can be achieved.

4.1 Spherical harmonic model compensation

Spherical harmonic mode compensation methods rely on publicly available global gravity field models. These methods calculate the horizontal gravity disturbance values at specific coordinates by setting model coefficients of different orders, processing them through a series of steps, and applying them to inertial navigation equations for compensation. Since the 1970s, with the increasing use of satellite data for gravity field reconstruction, models such as GEM (Goddard Earth Model) and TEG (Texas Earth Gravity models) have made significant progress in combined gravity modeling[26, 27]. However, the construction of marine gravity field models still faces challenges, mainly due to the insufficient simulation of marine terrain, especially the height difference between the sea surface and the geoid. Altimeter-based marine geoid and gravity field models are important data sources for calculating composite gravity field models[28, 29]. Currently, the two main publicly available spherical harmonic function models that provide global gravity field information are the EGM2008 model released by the National Geospatial-Intelligence Agency (NGA) and the EIGEN-6C4 model released by the German Research Centre for Geosciences (GFZ Potsdam). Fig.3 shows the relevant computational results of the EIGEN-6C4 model and the EGM2008 model.

**Tab.1   Comparison of gravity disturbance compensation methods**

Method	Advantage	Disadvantage
Spherical harmonic models	1. Well-established theoretical models; 2. Access to public data and models.	1. High computational complexity and demand for computing resources; 2. Poor accuracy in local areas; 3. Low resolution.
Pre-stored data interpolation	1. Reflects actual gravity field; 2. High precision; 3. Enables detailed compensation.	1. Limited coverage; 2. Lower precision in complex terrain; 3. Requires extensive measurement work.
Filtering estimation	1. Real-time compensation update; 2. Adapts to dynamic environments; 3. Enhances navigation robustness.	1. Difficult to decouple gravity measurement and inertial compensation; 2. Requires accurate models and noise characteristics; 3. Limited adaptability to complex environments.
Real-time measurement	1. Suitable for various measurement environments; 2. Directly reflects gravity distribution details.	1. High experimental costs; 2. Complex data processing algorithms; 3. Compensation delay issues

| Multi-source data fusion | 1. Complements strengths and weaknesses of different data sources; 2. More comprehensive gravity field information. | 1. Complex data fusion algorithms; 2. High requirements for data synchronization and consistency. |
| Frequency band information | 1. Strongly targeted compensation; 2. Compensates specific frequency disturbances. | 1. Not suitable for non-stationary disturbances; 2. Limited effectiveness on non-linear disturbances. |

In a spherical coordinate system, the gravity disturbance potential can be expressed as:

$$T = \frac{GM}{r} \sum_{n=2}^{N} \left(\frac{R_e}{r}\right)^n \sum_{m=0}^{n} (\bar{C}_{nm}^* \cos m\lambda + \bar{S}_{nm}^* \sin m\lambda) \bar{P}_{nm}(\cos\theta) \quad (8)$$

where G represents the Newtonian gravitational constant, M represents the mass of the Earth, θ is the polar angle, which can be calculated from $\theta = \pi/2 - L$, λ is the geographical longitude, R_e is the semi-major axis of the reference ellipsoid, r is the vector length from the center of the ellipsoid to the computation point, $\bar{C}_{nm}^*$ and $\bar{S}_{nm}^*$ are the coefficients of the spherical harmonic model of degree n and order m, $\bar{P}_{nm}(\cos\theta)$ represent the normalized associated Legendre functions.
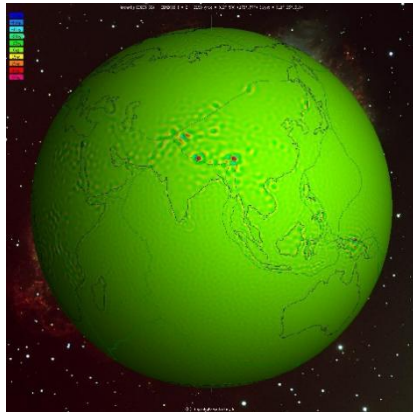
(a) EIGEN-6C4 model gravity anomaly differences with EGM2008 model
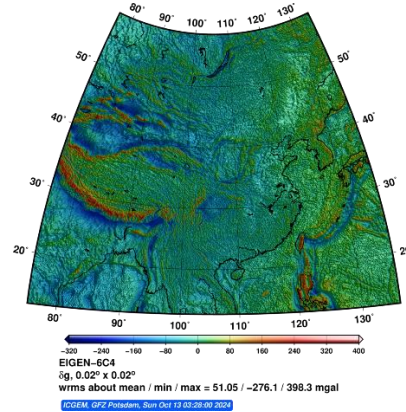
(b) EIGEN-6C4 gravity disturbance grid calculation results

Fig.3 Extracting gravity disturbance values using spherical harmonic models[30]

Based on the geometric relationship and definition between gravity and gravitational potential, taking the derivative of the gravity disturbance potential allows us to calculate the deflection of the vertical, and subsequently, the horizontal gravity disturbance can be determined. Once the coefficients of the spherical harmonic model are determined theoretically, the deflection of the vertical and gravity disturbances can be calculated. In practical applications, the uneven distribution of global gravity data and the large number of required coefficients often limit the model's ability to reflect the long-wavelength characteristics of the global gravity field, typically resorting to truncated low-order expansions. Additionally, incomplete gravity data in local areas further lead to a lack of high-frequency information in spherical harmonic models, affecting the model's accuracy and resolution. To address these issues, a DOV calculation method based on along-track gravity anomaly compensation was proposed in [22]. This method improves computational efficiency and reduces

dependence on satellite data by discretizing the Vening-Meinesz formula[31, 32].

Reference [33] has confirmed that using high-resolution spherical harmonic gravity models can more accurately describe the Earth's gravity field than traditional models. Based on this finding, Reference [34] developed two new gravity compensation methods that directly apply spherical harmonic models to calculate the gravity vector through data interpolation from an offline database, effectively reducing the influence of Schuler oscillations. Furthermore, Reference [35] explored the application of spherical harmonic models and gravity databases for gravity disturbance compensation in astronomical/inertial combined systems. To improve the real-time performance of gravity compensation, Reference [10] proposed a gravity prediction model based on the back-propagation neural network (BPNN), which uses gravity disturbance data within the planned flight area for training to achieve real-time compensation for the solution of a pure inertial navigation system, with effects comparable to traditional spherical harmonic model compensation methods. However, this method mainly focuses on real-time acquisition of gravity disturbances without considering the specific implementation of the compensation scheme, and its application is limited to predefined areas. In addition, selecting spherical harmonic models of different orders based on different terrain distributions may produce different compensation effects[36, 37], making the determination of the optimal order selection a topic worthy of further research.

4.2 Pre-stored data interpolation compensation

Gravity disturbance compensation strategy based on pre-stored data interpolation leverages existing gravity databases or actual gravimeter data and applies interpolation techniques to predict gravity disturbances along the navigation path[38, 39]. Common interpolation methods include inverse distance weighting[40], bilinear interpolation[41], Kriging interpolation, and improved quadratic surface Shepard interpolation methods[42, 43]. Furthermore, a gravity disturbance prediction model based on Extreme Learning Machine (ELM) integrates predicted gravity disturbance values into the inertial navigation system's error equations[44], effectively suppressing the propagation of errors in the inertial navigation system. A neural network model combining mind evolutionary computation (MEC), back propagation (BP), and AdaBoost algorithms processes actual data and inputs INS position coordinates into a well-trained neural network to obtain estimated values of gravity disturbances for compensation in the INS error equations[45]. Compared to traditional interpolation methods such as least-square collocation (LSC), this model can address non-linear issues between input and output data, reducing the maximum position error by 28%. A wavelet neural network predicts gravity disturbance data within a specific area and finds that[46], compared to inverse distance weighting interpolation, this method can reduce velocity errors by about 0.2 m/s and position errors by about 3000 meters after gravity disturbance compensation. Compared to global models, integrating field measurement data and fine-tuning the spherical harmonic function model for specific areas significantly improves model accuracy[25, 47, 48], but this method is only applicable to specific areas with prior data.

Tab.2   Comparison of the interpolation methods

Method	Advantages	Disadvantages
Inverse distance weighting	1. Computationally simple.	1. Low accuracy; 2. Sensitive to outliers.

Bilinear interpolation	1.High computational efficiency; 2. Smooth transition effects.	1. Does not account for non-linear relationships; 2. Prone to detail loss.
Kriging interpolation	1. High accuracy; 2. Smooth and continuous predictions.	1. High computational complexity; 2. Sensitive to variogram selection and parameter estimation.
Improved quadratic surface Shepard interpolation methods	1. Computationally efficient; 2. Suitable for large-scale scattered data; 3. Suitable for large-scale scattered data	1. Limited by data point distribution; 2. Accuracy may be restricted with sparse or unevenly distributed points; 3. Accuracy may be restricted with sparse or unevenly distributed points

4.3　Filtering estimation compensation

Filtering estimation compensation involves combining INS with other navigation sensors that are not affected by gravity disturbances to form an integrated navigation system for optimal estimation and compensation of gravity disturbance components. Reference [49] simulated the estimation and compensation of gravity disturbances based on the principles of Kalman filtering, but the gravity disturbance model constructed was relatively simple, the parameter settings were overly idealized, and its dependence on GPS signals limited its applicability in certain application scenarios. In contrast, Reference [50] revealed the inherent coupling relationships between gravity disturbances, their rates, and gradients, establishing a state-space model that accurately reflects the time-varying characteristics of gravity disturbances and proposing an autonomous high-precision real-time estimation and compensation scheme for rotating INS. An 80km vehicle test showed its horizontal positioning accuracy to be better than 50m.

Literature[51] uses a decaying memory Kalman filter to estimate the compensated attitude angle and differentiates it from the attitude angle of the strapdown inertial navigation system to obtain the deflection of the vertical along the trajectory. However, this method does not consider the attitude error during the initial alignment process, leading to the accumulation of errors during navigation. Literature [52] employs a ship heave model-based Rao-Blackwellised filter to identify genetic algorithm parameters, thereby obtaining more accurate gravity estimates. To reduce the impact of filter latency on the estimation of deflection of the vertical, literature [15] proposes two real-time compensation schemes: model-based selective dynamic feedback (SDF) and Q-tuned adaptive observation method (Q-AOM). After establishing an inertial navigation system error model that includes gravity disturbances, Zhong *et al.* and others concluded that the vertical gravity disturbance can be neglected for system attitude[53]. Although they attempted to compensate using a deflection of the vertical database, the effect was not significant. In contrast, by introducing a noise-driven term in the accelerometer and using Kalman filtering technology for real-time estimation of residual errors, the compensation accuracy was significantly improved, which is consistent with the conclusion in literature [12] regarding the coupling of accelerometer bias and

gravity disturbance. In addition to compensating for inertial navigation system parameters, literature [54] further models gravity anomalies as a first-order Gaussian-Markov process, integrating it into the Kalman filter of inertial/satellite combined navigation, effectively reducing the impact of gravity anomalies on height estimation, but the final estimation accuracy may be affected due to other unconsidered error sources. Literature [55] combines the direct-difference-modeling method (DD-M) of gravity disturbance measurement, which integrates differential method and time series analysis, into the Kalman filter of the position and orientation system (POS) to estimate and compensate for gravity disturbances, thereby significantly improving the accuracy of horizontal attitude.

4.4　Real-time measurement compensation

Real-time measurement compensation is a strategy that optimizes the mechanical arrangement by feeding back the real-time output of a gravimeter into the inertial navigation system's solution process. Despite the high application cost and susceptibility to limitations in gravity measurement accuracy and speed, it holds great application prospects. The gravity research team at the National University of Defense Technology, in response to the application needs of gravity-aided navigation in the maritime environment[56, 57], proposed a gravity disturbance compensation method based on the principles of Newtonian mechanics[18, 58]. The advantage of this method is that it can determine the relative relationship between the inertial navigation system and the gravimeter in real-time measurements and compensate for velocity and position in the navigation frame, thus avoiding the impact of attitude transfer matrix errors. However, the low-pass filtering used in the data processing may introduce delays, leading to poor compensation effects during rapid changes in gravity disturbances. Additionally, due to the presence of attitude errors, this method cannot completely eliminate the part of the accelerometer output affected by gravity disturbances.

4.5　Multi-source data fusion compensation

A single spherical harmonic model or real-time measurement technique often fails to fully describe the complexity of the actual gravity field. Therefore, by integrating different data sources, the details of the gravity field can be more finely supplemented, thereby improving the overall reliability of the data[59]. Multi-source data fusion is mainly achieved through two major categories of methods: statistical methods and analytical methods[60]. The statistical method usually employs the least squares collocation method, while the analytical method uses spherical harmonic functions to integrate gravity anomalies[61, 62]. This process involves using the residuals of local gravity data to adjust the spherical harmonic function coefficients, gradually integrating the actual measured data into the model. To reduce the noise amplification problem caused by small singular values in the covariance matrix, Tikhonov regularization method was introduced to optimize the least squares collocation, thereby enhancing the precision and stability of the collocation solution[63]. The multi-source gravity data fusion algorithm proposed in [18] eliminates errors caused by bias through correlation processing and further integrates prediction models to enhance the real-time performance of the algorithm. Radial basis functions, due to their good localization characteristics in both spatial and frequency domains, are widely used to fuse multi-source gravity data, and their solution efficiency is superior to the traditional least squares collocation method[64]. Many scholars

have successfully established global or local gravity field models using this method[65-72].

4.6 Frequency band information compensation

For slow-moving carriers, high-frequency signals in horizontal gravity disturbances significantly impact the positioning accuracy of INS. Therefore, a deep analysis of the relationship between gravity disturbance frequency characteristics and INS error characteristics is crucial for effective gravity disturbance compensation. Jekeli proposed a method to estimate short-wave gravity vectors without external attitude updates, based on covariance analysis in both time and frequency domains, under the stable condition of uncompensated gyro drift[8]. Reference[36]further points out that the horizontal channel acts as a low-pass filter for high-frequency signals of gravity disturbances and believes that compensation within a certain cutoff frequency range will not sacrifice accuracy. Based on the spherical harmonic model, references [73] and [74] proposed a high-precision gravity disturbance compensation method that combines spherical harmonic model compensation with state estimation compensation. This method uses spherical harmonic model compensation for long waves and utilizes a Markov model for state estimation of medium and short-wave gravity disturbance components, resulting in better overall compensation effects compared to single compensation methods. Reference [75] uses the Residual Terrain Model (RTM) technology to restore high-frequency gravity field signals to make up for the shortcomings of high-order gravity field models in short-wave signals. Reference [76] also points out that the smoothing effect of terrain correction is only effective for the high-frequency part of gravity signals. In addition, reference [77] studied the relationship between horizontal gravity disturbances and INS horizontal attitude errors in pure inertial calculation and combined navigation modes under uniform linear motion. By analyzing the amplitude-frequency curves of gravity disturbance signals of different wavelengths, it was concluded that attitude errors in combined navigation modes require compensation with high-frequency gravity disturbance signals. At the same time, the study also pointed out that overly detailed gravity disturbance data does not necessarily improve the accuracy of attitude errors, making the choice of compensation resolution a matter of concern.

5 Summary

Oceans play a crucial role in humanity's exploration of nature. Inertial navigation systems (INS), with their autonomous and passive features, show promise for underwater navigation. However, the unique marine environment and the limitations of current devices hinder the effectiveness of passive auxiliary inertial navigation. Leveraging the Earth's gravitational field can enhance underwater navigation. Strapdown gravimeters, known for their compactness and ease of integration, have been successfully used in aviation, land, and marine scenarios. As the precision of inertial devices increases, compensating for gravitational disturbances can effectively reduce INS errors. However, many gravity compensation methods rely on spherical harmonic models and databases that often lack the necessary precision and resolution, necessitating interpolation. While real-time measurements from gravimeters and filtering estimation algorithms are viable, their accuracy and complexity can limit their application. This article summarizes existing compensation methods to guide future research in gravity compensation algorithms.

References:

[1] Gelb A. Geodetic and geophysical uncertainties-Fundamental limitations on terrestrial inertial navigation; proceedings of the Control and Flight Dynamics Conference, F, 1968 [C].

[2] Peshekhonov VG. Problem of the vertical deflection in high-precision inertial navigation [J]. Gyroscopy and Navigation, 2020, 11: 255-62.

[3] Zhao Kun, Hu Xiaomao, Liu Bohan. Long-endurance fiber optic gyroscope INS for ships and its development trends [J]. Journal of Chinese Inertial Technology 2022, 30(03): 281-7.

[4] Yang Jinjing, Cai Shanjun, Yuan Run, et al. Adual-axis inertial navigation system rotation modulation method stabilised in the inertial coordinate system [J]. Navigation Positioning and Timing, 2023, 10(02): 57-64.

[5] Moryl J. Advanced submarine navigation systems [J]. Oceanographic Literature Review, 1997, 5(44): 523.

[6] Huang Motao, Ouyang Yongzhong, Bian Shaofeng, et al. Analysis and Reflections on the Development of Underwater Gravity-Aided Inertial Navigation Technology in the United States and Russia [J]. Geomatics and Information Science of Wuhan University, 2024.

[7] Somigliana Carlo. Teoria generale del campo gravitazionale dell'ellissoide di rotazione [J]. Memorie della Società Astronomia Italiana, Vol 4, p 425, 1929, 4: 425.

[8] Jekeli Christopher. Airborne vector gravimetry using precise, position-aided inertial measurement units [J]. Department of Geodetic Science and Surveying, The Ohio State University, Columbus, USA, 1994, Vol.69(No.1): 1-11.

[9] Lu Xin, Lian Junxiang, Wu Meiping. Research of Gravity Error Compensation in Marine Inertial Navigation System [J]. Navigation and Control, 2010, 9(04): 15-21.

[10] Gao D. Y., Hu B. Q., Qin F. J., et al. A Real-Time Gravity Compensation Method for INS Based on BPNN [J]. IEEE Sens J, 2021, 21(12): 13584-93.

[11] Wang Jing, Yang Gongliu, Li Xiangyun, Zhou Xiao. Error indicator analysis for gravity disturbing vector's influence on inertial navigation system [J]. Journal of Chinese Inertial Technology, 2016, 24(3): 285-90.

[12] Chang L. B., Qin F. J., Wu M. P. Gravity Disturbance Compensation for Inertial Navigation System [J]. IEEE Trans Instrum Meas, 2019, 68(10): 3751-65.

[13] Grejner-Brzezinska DOROTA A, Wang Jin. Gravity Modeling for High-Accuracy GPS/INS Integration [J]. Navigation, 1998, 45(3): 209-20.

[14] Dai Dongkai, Wang Xingshu, Zhan Dejun, et al. Dynamic measurement of high-frequency deflections of the vertical based on the observation of INS/GNSS integration attitude error [J]. J Appl Geophys, 2015, 119: 89-98.

[15] Zhu Jing. Study on Gravity aided Technologies inShipborne Attitude Measurement with Ring Laser Gyro [D]; National University of Defense Technology, 2018.

[16] Harriman David W, Harrison J Chris. Gravity-induced errors in airborne inertial navigation [J]. Journal of Guidance, Control, and Dynamics, 1986, 9(4): 419-26.

[17] Zhang P. P., Wu L., Bao L. F., et al. Gravity disturbance compensation for dual-axis rotary modulation inertial navigation system [J]. Front Mar Sci, 2023, 10: 14.

[18] Luo Kaixin. Research on Key Technologies of Ocean Gravity-aided Inertial Navigation [D]; National University of Defense Technology, 2023.

[19] Soler Tomás, Han Jen-Yu, Weston Neil D. On deflection of the vertical components and their transformations [J]. Journal of surveying engineering, 2014, 140(2): 04014005.

[20] Hao S. W., Zhang Z. L., Zhou Z. F., et al. Analysis of DOV estimation in initial alignment based on Single-axis rotating SINS [J]. Measurement, 2022, 204: 14.

[21] Tie Junbo, Wu Meiping, Cao Juliang, et al. The impact of initial alignment on compensation for deflection of vertical in inertial navigation; proceedings of the 2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM), F, 2017 [C]. IEEE.

[22] Hao Shiwen, Zhang Zhili, Zhou Zhaofa, et al. Relative estimation method of DOV based on double-alignment of SINS [J]. IEEE Geosci Remote Sens Lett, 2022, 19: 1-5.

[23] Tie J. B., Cao J. L., Wu M. P., et al. Compensation of Horizontal Gravity Disturbances for High Precision

Inertial Navigation [J]. Sensors, 2018, 18(3): 21.

[24] Wu Mingqiang, Liao Shikang. Analysis of the Influence of Horizontal Gravity Anomaly on the Positioning Accuracy of Inertial Positioning and Orientation System [J]. OPTICS & OPTOELECTRONIC TECHNOLOGY, 2023, 21(03): 74-8.

[25] Weng J., Liu J. N., Jiao M. X., Kou K. Analysis and on-line compensation of gravity disturbance in a high-precision inertial navigation system [J]. GPS Solut, 2020, 24(3): 8.

[26] Lerch FJ, Wagner CA, Smith DE, et al. Gravitational field models for the Earth (GEM 1 and 2) [R], 1972.

[27] Tapley BD, Shum CK, Ries JC, et al. The TEG-3 geopotential model; proceedings of the Gravity, Geoid and Marine Geodesy: International Symposium No 117 Tokyo, Japan, September 30–October 5, 1996, F, 1997 [C]. Springer.

[28] Pavlis Nikolaos K, Holmes Simon A, Kenyon Steve C, Factor John K. The development and evaluation of the Earth Gravitational Model 2008 (EGM2008) [J]. Journal of geophysical research: solid earth, 2012, 117(B4).

[29] Flechtner Frank, Reigber Christoph, Rummel Reiner, Balmino Georges. Satellite gravimetry: a review of its realization [J]. Surv Geophys, 2021, 42(5): 1029-74.

[30] Ince E. S., Barthelmes F., Reissland S., et al. ICGEM-15 years of successful collection and distribution of global gravitational models, associated services, and future plans [J]. Earth Syst Sci Data, 2019, 11(2): 647-74.

[31] Ning JS, Chao DB, Li JC. The spherical convolution form of Vening-Meinesz formula [J]. Acta Geodaetica et Cartographica Sinica, 1994, 23(3): 161-6.

[32] Chen JY. Methods for computing deflections of the vertical by modifying Vening-Meinesz'function [J]. Department of Geodetic Science and Surveying, The Ohio State University, Columbus, USA, 1982, 56: 9-26.

[33] Wang J., Yang G. L., Li X. Y., Zhou X. Application of the spherical harmonic gravity model in high precision inertial navigation systems [J]. Meas Sci Technol, 2016, 27(9): 10.

[34] Wu Ruonan, Wu Qiuping, Han Fengtian, et al. Gravity compensation using EGM2008 for high-precision long-term inertial navigation systems [J]. Sensors, 2016, 16(12): 2177.

[35] Leng Yue, Zhong Sheng. Compensation method for gravity disturbance in celestial/inertial integrated system [J]. School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, 430074, China;National Key Laboratory of Multispectral Information Intellige, 2024, Vol.46(No.4): 1357-63.

[36] Zhu Hongbao, Tian Ye, Yan Gongmin, et al. Research on real-time gravity compensation system based on gravity model order adaptation [J]. Navigation Positioning and Timing, 2024, 11(04): 55-64.

[37] Zhou Xiao, Yang Gongliu, Wang Jing, Wen Zeyang. A combined gravity compensation method for INS using the simplified gravity model and gravity database [J]. Sensors, 2018, 18(5): 1552.

[38] Kwon Jay Hyoun. Gravity compensation methods for precision INS; proceedings of the Proceedings of the 60th Annual Meeting of The Institute of Navigation (2004), F, 2004 [C].

[39] Arora Nitin, Russell Ryan P. Fast efficient and adaptive interpolation of the geopotential [J]. J Guid Control Dyn, 2014, 38: 1345-55.

[40] Hofmann-Wellenhof Bernhard, Moritz Helmut. Physical geodesy [M]. Springer Science & Business Media, 2006.

[41] Sarzeaud O, Lequentrec-Lalancette M-F, Rouxel D. Optimal interpolation of gravity maps using a modified neural network [J]. Math Geosci, 2009, 41(4): 379-95.

[42] Bishop Garner C. Gravitational field maps and navigational errors [unmanned underwater vehicles] [J]. IEEE J Ocean Eng, 2002, 27(3): 726-37.

[43] Wu Taiqi, Huang Motao, Lu Xiuping. Gravity Map Creating Technology in Gravity Matching Navigation [J]. J Chin Inert Technol, 2007, 15: 438-41.

[44] Zhou X., Yang G. L., Cai Q. Z., Wang J. A Novel Gravity Compensation Method for High Precision Free-INS Based on "Extreme Learning Machine" [J]. Sensors, 2016, 16(12): 14.

[45] Zhou X., Yang G. L., Wang J., Li J. An improved gravity compensation method for high-precision free-INS based on MEC-BP-AdaBoost [J]. Meas Sci Technol, 2016, 27(12): 10.

[46] Zhou Xiao, Yang Gongliu, Cai Qingzhong. Compensation on gravity disturbance for high precision INS based on wavelet neural network [J]. Journal of Chinese Inertial Technology, 2014, 24(5): 571-6.

[47] Zhu Z. S., Zhao B., Guo Y. Y., Zhou X. Y. Research on gravity vertical deflection on attitude of position and orientation system and compensation method [J]. Aerosp Sci Technol, 2019, 85: 495-504.

[48] Yansheng WU, Zongqiang WANG, Zhang Bing. Comprehensive compensation method for the influence of disturbing gravity field on long-range rocket guidance computing [J]. Chinese Journal of Aeronautics, 2022, 35(3): 408-18.

[49] Zhou Xiao, Yang Gongliu, Wang Jing, et al. Estimation and compensation for gravity disturbance based on Kalman filtering in inertial navigation [J]. Chinese Inertial Technology, 2015, 23(06): 721-6.

[50] Yang J., Wang X. L., Wang B., Hu X. D. An autonomous and high-accuracy gravity disturbance compensation scheme for rotary inertial navigation system [J]. Meas Sci Technol, 2024, 35(8): 14.

[51] An W., Xu J. N., He H. Y., Jiang P. F. A Method of Deflection of the Vertical Measurement Based on Attitude Difference Compensation [J]. IEEE Sens J, 2021, 21(12): 13125-36.

[52] Stepanov Oleg A., Koshaev Dmitry A., Motorin Andrei V., et al. Algorithms for Integrated Processing of Marine Gravimeter Data and GNSS Measurements [J]. IFAC-PapersOnLine, 2020, Vol.53(No.2): 500-5.

[53] Zhong Mingfei, Zhan Dejun, Xu Rong, Zhong Lin. Research on The Effects of Gravity Disturbances onHigh-Precision RLG Attitude Measurement System; proceedings of the The 6th China Satellite Navigation Academic Conference, Xi 'an, Shaanxi, China, F, 2015 [C].

[54] Xiong Hao, Zhao Yingwei, Wang Xingshu, et al. An analysis of the effect of gravity anomaly to height estimation in high-precision INS/GNSS integrated navigation systems [J]. IEEE Sens J, 2018, 19(7): 2713-21.

[55] Fang Jiancheng, Chen Linzhouting, Yao Jifeng. An accurate gravity compensation method for high-precision airborne POS [J]. IEEE Trans Geosci Remote Sensing, 2013, 52(8): 4564-73.

[56] Wu Meiping, Cai Shaokun, Yu Ruihang, Cao Juliang. Research Progress on Strapdown Gravimetry Technology [J]. Navigation and Control, 2020, 19(Z1): 161-9.

[57] Xiong Zhiming. Research on Key Technologies for Underwater Gravimetry Based On Multi-Sensor Fusion [D]; National University of Defense Technology, 2021.

[58] Luo Kaixin and Yu, Ruihang and Wu, Meiping and Cao, Juliang and Huang, Yulong. A Novel Gravity Disturbance Compensation Inertial Navigation Method based on Newtonian Mechanics [J]. IEEE Trans Instrum Meas, 2024: 1-.

[59] Wang Hubiao, Wang Yong, Lu Yang, Zan J. Inversion of marine gravity anomalies by combinating multi-altimeter data and shipborne gravimetric data [J]. Journal of Geodesy and Geodynamics (in Chinese), 2005, 25(1): 81-5.

[60] Li Fangming. Research on Underwater Gravity AidedPositioning and Gravity Map Navigability Analysis [D]; Harbin Engineering University, 2021.

[61] Kern Michael. An analysis of the combination and downward continuation of satellite, airborne and terrestrial gravity data [M]. Citeseer, 2003.

[62] Cheng Yi. A Study on Heterogeneous Marine Gravity Data Combination Technology [D], 2008.

[63] Huang MT, Ouyang YZ, Zhai GJ, et al. Tikhonov regularization collocation for multi-source gravity data fusion processing [J]. Hydrographic Surveying and Charting, 2013: 6-12.

[64] Wittwer TB. Regional gravity field modeling with radial basis functions [J]. 2009.

[65] Bentel Katrin, Schmidt Michael, Gerlach Christian. Different radial basis functions and their applicability for regional gravity field representation on the sphere [J]. GEM-International Journal on Geomathematics, 2013, 4: 67-96.

[66] Bentel Katrin, Schmidt Michael. Combining different types of gravity observations in regional gravity modeling in spherical radial basis functions; proceedings of the VIII Hotine-Marussi Symposium on Mathematical Geodesy: Proceedings of the Symposium in Rome, 17-21 June, 2013, F, 2016 [C]. Springer.

[67] Pham TD, Tran T. Solutions to pseudo differential equations using spherical radial basis functions [J]. Bulletin of the Australian Mathematical Society, 2009, 79(3): 473-85.

[68] Klees R, Tenzer Robert, Prutkin I, Wittwer T. A data-driven approach to local gravity field modelling using

spherical radial basis functions [J]. J Geodesy, 2008, 82: 457-71.

[69] Zhiwei MA, Yang LU, Yi TU, et al. Regional gravity field modeling with Abel-Poisson radial basis functions [J]. Acta Geodaetica et Cartographica Sinica, 2016, 45(9): 1019.

[70] Yang Fan, Xu Hou-Ze, Zhong Min, et al. GRACE global temporal gravity recovery through the radial basis function approach [J]. Chinese Journal of Geophysics, 2017, 60(4): 1332-46.

[71] Mahbuby Hany, Safari Abdolreza, Foroughi Ismael. Local gravity field modeling using spherical radial basis functions and a genetic algorithm [J]. Comptes Rendus Géoscience, 2017, 349(3): 106-13.

[72] Tenzer Robert, Klees R. The choice of the spherical radial basis functions in local gravity field modeling [J]. Studia Geophysica et Geodaetica, 2008, 52: 287-304.

[73] Liu Yuxin, Wang Xinlong, Wang Xun, et al. An Accurate Gravity Disturbance Compensation MethodBased on Spherical Harmonic Model and Multi-Sensor Fusion [J]. AERO WEAPONRY, 2023, 30(1): 104-13.

[74] Dai Dongkai, Wang Xingshu, Zhan Dejun, et al. An improved method for gravity disturbances compensation in INS/GPS integrated navigation; proceedings of the 2014 12th International Conference on Signal Processing (ICSP), F, 2014 [C]. IEEE.

[75] Panpan ZHANG, Lin WU, Lifeng BAO, et al. Research on gravity compensation of inertial navigation system based on multispectral gravity disturbance [J]. Acta Geodaetica et Cartographica Sinica, 52(8): 1255.

[76] Tenzer Robert, Novák P. Effect of the Long-Wavelength Topographical Correction on the Low-Degree Earth's Gravity Field; proceedings of the Gravity, Geoid and Earth Observation: IAG Commission 2: Gravity Field, Chania, Crete, Greece, 23-27 June 2008, F, 2010 [C]. Springer.

[77] An Wen, Xu Jiangning, Wu Miao, Li Feng. Influence of gravity disturbance for SINS horizontal attitude [J]. Journal of National University of Defense Technology, 2021, 43(03): 135-41.

An Improved ICCP Matching Algorithm Based on HHM

Quan Sihang[1], Wang Zhanghai[2], Huhuizhu[1]

（1. The College of Optical Science And Engineering, Zhejiang University, Hangzhou, China;

2. Center for Novel Computing Sensing and Intelligent Processing, Zhejiang Lab, Hangzhou, China.)

Abstract: The iterative closest contour point (ICCP) algorithm is widely used in matching navigation. However, the ICCP algorithm is vulnerable to divergence in cases such as large initial errors. This paper proposes an improved ICCP matching algorithm based on the Hidden Markov Model (HMM), which establishes the correlation mapping between the INS sequence and the set of trusted points. Then the optimal sequence of the closest contour point can be obtained by using the Viterbi algorithm. Simulation results show that the HMM-ICCP algorithm improves the accuracy by 89.7%, 51.2%, 56.4%, and 38.5% compared to the ICCP algorithm in four typical cases.

 Keywords: geomagnetic matching algorithm; iterative closest contour point; Hidden Markov Model;

1   Introduction

A geomagnetic-aided navigation system consists of four parts: an INS, a geomagnetic database map, a magnetic measurement sensor, and a matching algorithm. Among them, the geomagnetic matching algorithm, as the fundamental technology of geomagnetic navigation, has sparked numerous research studies [1-3].

ICCP algorithm uses a rigid translation model to iteratively approach the nearest contour sequence, which can effectively eliminate position and heading errors. However, when the initial position error is large, ICCP tends to fall into local optimization. Thus, Wang et al. [4] proposed a multipath parallel ICCP algorithm, which effectively solves the convergence problem when the initial error is large. Xu et al. [5] proposed a PSO-ICCP matching method that introduces the trajectory correlation and improves the multi-matching mechanism of ICCP. To address the sensitivity of the ICCP algorithm to noise, Shao et al. [6] combined a probabilistic data correlation method with the ICCP algorithm to enhance its robustness.

In the matching algorithms mentioned above, the sample points are independent. Hence, it is easy for mismatching to occur when multiple contours are close to the true trajectory and either parallel or perpendicular to it, leading to decreased matching accuracy or divergent results. Consequently, the position correlation of INS sample points is introduced to the scalar matching algorithm in this paper. The Hidden Markov Model (HMM) is used to encode the relationship between every point of the INS sequence and every to-be-matched contour point on the corresponding contour line, which are called trusted points. Meanwhile, the selection of the nearest contour point sequence depends on the principle of maximum probability. The primary contributions

---

can be summarized as follows:

1. A selection criterion for the set of trusted points is established.

2. The HMM model is employed to encode INS sequences and trusted point sets. This fully considers the sequential characteristics of the INS data. With the aid of the complex state space and transition probability, it can better reflect the mapping relationship between the INS data and the trusted point sets.

3. The Viterbi algorithm is utilized in the selection of the closest contour points. It takes into account the probability of a series of observations in different states rather than solely relying on spatial distance.

2 Method

The flowchart of the proposed HMM-ICCP geomagnetic matching algorithm is shown in Fig.1. The primary differences from the traditional ICCP algorithm are in the trusted points selection, the closest contour points selection, and the sequence update, which are yellow highlighted in the flowchart. Similarly, while calculating the rigid transformation, the closest contour points are adjusted appropriately.
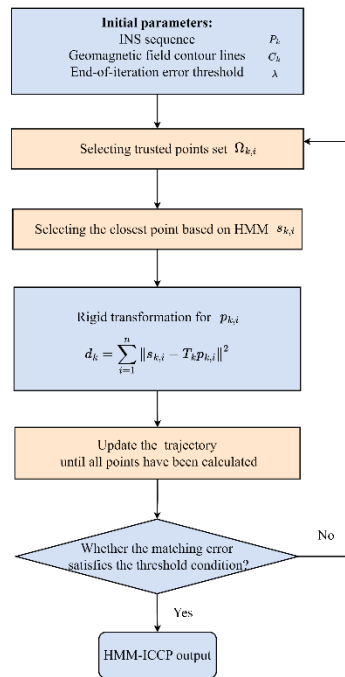
Fig. 1 Flowchart of HMM-ICCP algorithm for geomagnetic matching Navigation.

2.1 Selection of Trusted Points

Based on the assumption that the true trajectory is close to the INS trajectory, the area for the contour search is limited according to the initial error to reduce the amount of computation. The following criteria are met:

$$U = \left\{ (a_k, b_k) \mid\mid a_k - a_k^p \mid < 2\rho_x, \mid b_k - b_k^p \mid < 2\rho_y \right\} \tag{53}$$

where (a_k, b_k) is the coordinate value of the points in the search region, corresponding to latitude and longitude, respectively. (a_k^p, b_k^p) is the coordinate value corresponding to the INS trajectory,

and ρ is the initial position error. After that, the contour extraction is carried out in the specified region, and all the points on the extracted contour are referred to as the trust points.

2.2 HMM-ICCP matching algorithm

This paper uses HMM to model the process of selecting and evaluating those potential close-by points (the set of trusted points), considering that some of these state transfers are more probable than others. Assuming that the position obtained from INS P_k serves as the observation variable of the HMM, and the and the trusted points from S_k are the hidden state variables of the HMM. Thus, the measurements of each position are matched with the corresponding trusted point sets, and the connection among all trusted point sets is determined by the order of the INS sequence.

Formally speaking, suppose that the discrete state of the HMM is a set of trusted points corresponding to each point in INS observations. It is desirable to find the nearest contour point for each INS point. As indicated in Fig .2, the INS sequence contains 5 samples. At the time $t = 1$, the corresponding trusted point set consists of 5 elements. At the time $t = 2$, the corresponding trusted point set has 3 elements. In reality, during $t = 1$ and $t = 2$, even though several transfer routes can be accessed, there is only one likely route. By analogy, we will get an optimal path.
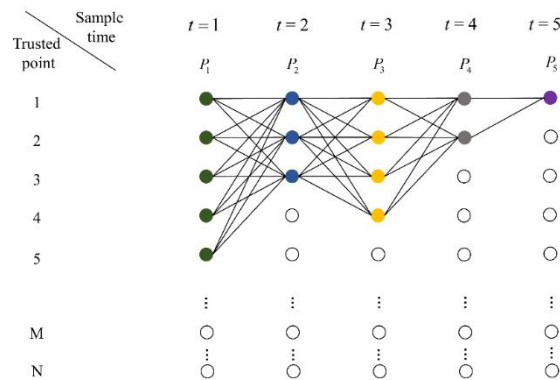

**Fig. 2** Demonstration of state transitions in HMM.

3 Simulation

Based on the EMAG2_V3 global magnetic anomaly model [7], we choose part of the South China Sea at sea level, with longitude ranging from 112° to 115° and latitude ranging from 17° to 19°. We then plot the grid data as the geomagnetic base map, with a grid resolution of 2 arc-minutes. Fig. 3 displays the plotted grid plane and surface. Moreover, the carrier is set to have an initial position error of 0.13° in latitude and 0.08° in longitude. The simulation is based on four cases, respectively: 1) the real track is perpendicular to the contour lines; 2) the real track is parallel to the contour lines; 3) there are multiple contour lines of the same value near the real track; and 4) the real track is within the obvious region of geomagnetic features.

(a)   (b)
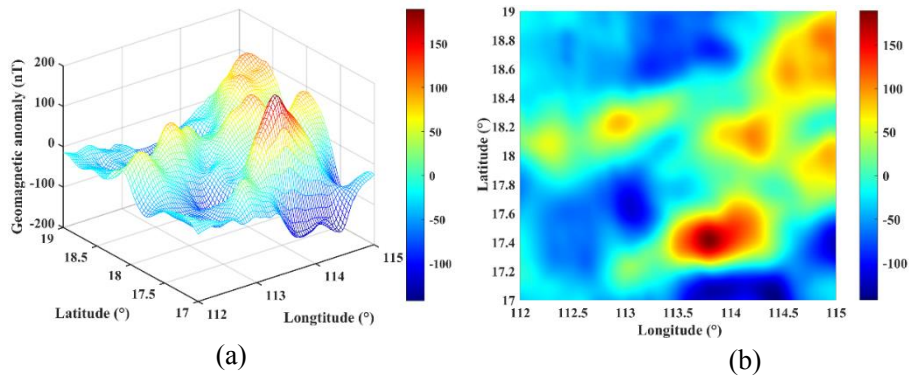
Fig. 3 Geomagnetic anomaly data. (a) Surface. (b) Plane.

To verify the validation of the proposed algorithm, we compare it with the MAGCOM and ICCP, which are typical geomagnetic matching algorithms. The matching results under the specific four scenarios in the simulation are shown in Fig. 4.
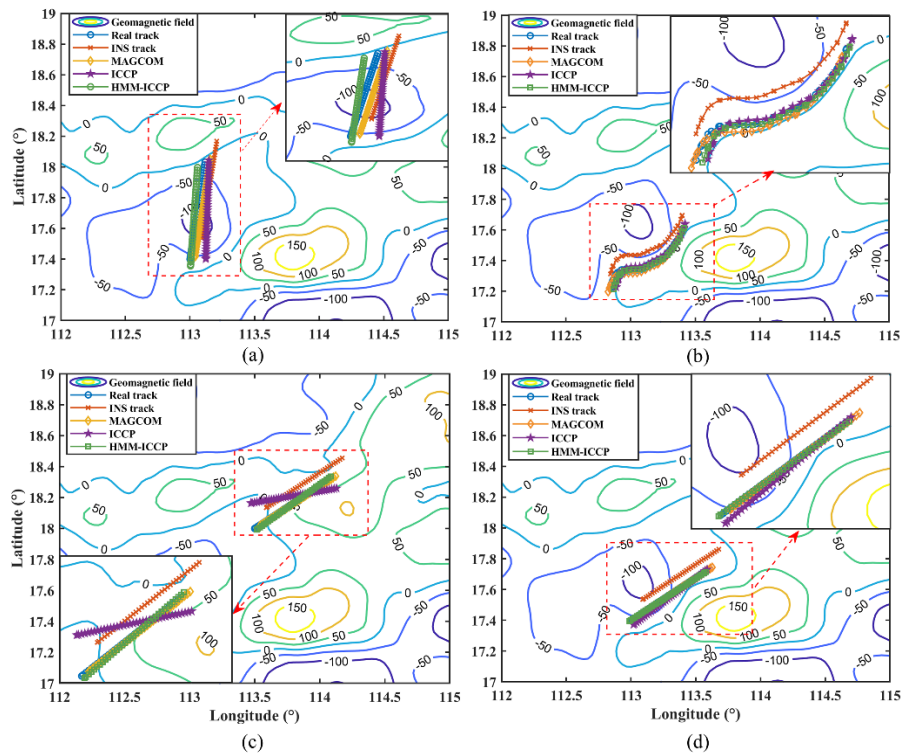
Fig. 4 Matching results of different algorithms in typical conditions. (a) Vertical relationship between contour lines and real track. (b) The parallel relationship between contour lines and real track. (c) Existence of multiple contour lines with the same value. (d) Existence of non-obvious geomagnetic features.

Fig. 4 illustrates the correction results of different algorithms. It can be seen that the ICCP algorithm has obvious mismatches and a large error with the real track in the first three cases, especially when there are multiple contours with the same value close to the real track. Although the MAGCOM algorithm has a small error in the parallel situation, it still has a significant error overall, and the error will increase further in the presence of initial angle errors. However, in the

processing of HMM-ICCP matching, there is a quite significant reduction in error in all typical conditions.

The mean matching errors of the three algorithms in four typical scenarios are shown in Fig. 5 and Tab. 1. All three methods exhibit large errors in the vertical situation, with the ICCP and MAGCOM having errors of more than three grids and up to a dozen kilometers. The HMM-ICCP has a 38.5% improvement in matching accuracy compared to the ICCP. However, the deviations are still close to two grids. In the case of multiple contour lines of the same value near the real track, the HMM-ICCP matching significantly reduces the error by 89.7% compared to the ICCP algorithm. This is because the HMM-ICCP algorithm fully utilizes the correlation between the INS sequence and the trusted points. Thus, it can select the closest contour points more accurately to obtain the optimal trajectory, whereas the traditional ICCP algorithm, which is based on Euclidean distance, incorrectly selects the closest contour points, leading to the occurrence of obvious mismatches. Similarly, in the parallel and non-obvious cases, the error of HMM-ICCP is reduced by 51.2% and 56.4% compared to the ICCP method, respectively. Moreover, in the processing of the HMM-ICCP algorithm, corrected tracks have a small matching error within one grid in the latter three cases, which demonstrates the great superiority of better applicability.
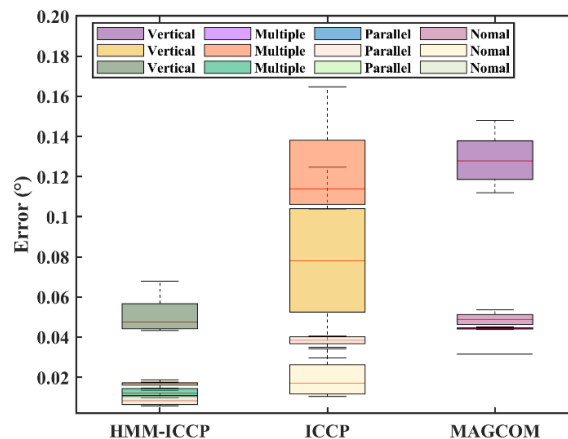

Fig.6 Matching errors of different algorithms under four typical cases.

Tab.1 The RMSE values for three different algorithmic errors in four typical cases

	MAGCOM (RMSE)	ICCP (RMSE)	HMM-ICCP (RMSE)
Vertical relation	0.1299	0.0836	0.0514
Parallel relation	0.0315	0.0408	0.0199
The existence of multiple contour lines with the same value	0.0444	0.124	0.0128
The existence of non-obvious geomagnetic features	0.0489	0.0209	0.0091

4   Conclusion

This study proposes an improved ICCP algorithm based on the Hidden Markov Model (HMM-

ICCP) to tackle the problem that the traditional ICCP algorithm is vulnerable when the trajectory is perpendicular or parallel to the contours; when there are multiple contours near the trajectory and when the existence of non-obvious geomagnetic features. Due to the advantages of the HMM in reflecting the mapping patterns of the data and the ability to model the probability of the sequence, the proposed HMM-ICCP method performs outstandingly with 38.5%, 51.2%, 56.4%, and 89.7% enhancement in accuracy respectively, compared to the ICCP algorithm.

References:

[1] El-Sheimy, Naser and Ahmed A. Youssef. Inertial sensors technologies for navigation applications: state of the art and future trends[J]. Satellite Navigation, 2020, 1: 1-21.

[2] Wu Z, Hu X, Wu M, et al. An experimental evaluation of autonomous underwater vehicle localization on geomagnetic map[J]. Applied Physics Letters, 2013, 103(10):9-18.

[3] B. Wang, L. Yu, Z. Deng, et al. A Particle Filter-Based Matching Algorithm With Gravity Sample Vector for Underwater Gravity Aided Navigation[J]. IEEE/ASME Transactions on Mechatronics, 2016, 21(3): 1399-1408.

[4] H. Wang, X. Xu, and T. Zhang. Multipath Parallel ICCP Underwater Terrain Matching Algorithm Based on Multibeam Bathymetric Data[J]. IEEE Access, 2018, 6:48708–48715.

[5] N. Xu, L. Wang, T. Wu, et al. An innovative PSO-ICCP matching algorithm for geomagnetic navigation[J]. Measurement, 2022, 193:110958.

[6] J. Xiao, X. Duan, and X. Qi. Iterated closest contour point algorithm for geomagnetic matching based on probability data association[J]. Zhongguo Guanxing Jishu XuebaoJournal Chin. Inert. Technol., 2018, 26: 202–208.

[7] N. C. for E. Information, "EMAG2v3: Earth Magnetic Anomaly Grid (2-arc-minute resolution)." Available: https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ngdc.mgg.geophysical_models:EMAG2_V3

An Approach to Estimating the Gravity Vector Disturbance Based on Repeated Airborne Observations

I. A. Akimov

(Lomonosov Moscow State University, Moscow, Russia)

Abstract: An approach to estimating the gravity vector disturbance from airborne observations at repeated lines is presented. The use of a strapdown gravimeter is assumed. The approach is based on the error equations of gravimeter's inertial measurement unit and GNSS. A model of the gravity vector is constructed using the cubic splines. The unknown coefficients are estimated by the Kalman filter. The numerical results from simulated data processing are presented.

1   Introduction

This paper is devoted to the airborne vector gravimetry problem with using a strapdown airborne gravimeter that includes an Inertial Measurement Unit (IMU) and Global Navigation Satellite System (GNSS) receivers. Usually, only the vertical component of the gravity disturbance vector (GDV) is determined [1, 2]. The determination of the horizontal components of the gravity disturbance vector is more complicated because they are observed in a combination with the IMU systematic errors [3]. Some approaches use an a priori information about the GDV, while the others use additional measurements [3-5]. Our approach proposed in this paper uses both.

Several methods based on an a priori model of the GDV as a time-dependent function were proposed previously [1, 3], but the achieved accuracy of the GDV estimates is not sufficient to use them in practice. We propose a new approach that assumes repeated flight lines and uses a new a priori model of the GDV horizontal components: they are assumed to depend on the distance along the flight lines. Mathematically, the model of the GDV components is introduced as a linear combination of the cubic B-splines at a line and the problem of airborne gravimetry is reduced to the IMU-GNSS integration and Kalman filtering. We show that the new approach allows to increase an accuracy of the GDV estimation.

2   Problem statement and estimation algorithm

Let us introduce the geodetic frame as ME,N,Up, where M is the origin of the frame and coincides with in the IMU accelerometers proof mass. Then introduce the gravity disturbance vector $\Delta g = g - g_0$ as the difference between the actual gravity and normal gravity (e.g., computed using Helmert's formula) vectors in the geodetic frame. There are two parts of airborne vector gravimetry problem: estimation of the vertical component and estimation of the horizontal components. Solution to the first problem is well established, see, for example, [1, 2]. Our aim is to solve the second problem.

3   Estimation of the GDV at the repeated lines

The basics of the problem is the error equations of an IMU aided by GNSS [6]. The state vector of the

error equations includes the following variables (the indices E,N,Up mean the projections to the geodetic frame, indices z1,2,3 mean the projections to the IMU body frame):

ΔgE, ΔgN are the east and north components of GDV;

$\Delta fz1$, $\Delta fz2$ are the biases of the IMU accelerometers;

$vz1$, $vz2$, $vz3$ are the drifts of the IMU gyroscopes;

$\beta1$, $\beta2$, $\beta3$ are the attitude errors;

δVE, δVN are the dynamic velocity errors.

The a priori GDV model is introduced as a linear combination of cubic B-splines [7]:

$$\Delta gE(s(t)) = \Sigma cEiBi(s(t)), \Delta gN(s(t)) = \Sigma cNiBi(s(t)), \qquad (1)$$

where s=s(t) is the distance along the line, Bi(s) is the B-spline, cEi,, cNi are the unknown constants, i=1,…, N, i is the knot number, and N is the total number of knots at the flight line.

Let us include cEi,, cNi in the state vector of IMU error equations and set the optimal stochastic estimation problem [8]. The estimate of the state vector (that consists of 2N+10 components) is determined through Kalman filtering [8]. The GDV component estimates are computed using (1) from the estimates of cEi,, cNi obtained at the last iteration of the Kalman filter, i=1,…, N.

4   Data processing results

Below are the results of the algorithm testing based on the simulated data processing. The ideal IMU and GNSS data were simulated at two repeated lines. The systematic errors and noises were added to the ideal data according to the typical errors of a modern strapdown gravimeter and included the accelerometer biases and gyroscopes drifts. The values of the systematic errors are $\Delta fz1$=30 mGal, $\Delta fz2$ =–40 mGal, vz=0.003 deg/h.

The results of the GDV estimation are shown in Fig. 1. The reference gravity data is given by the global gravity model EGM2008. The standard deviation (STD) of the estimation error is 2 mGal, the mean value is 3-4 mGal. Note that a standard approach that uses the model of GDV as a time-dependent function shows much lower accuracy: the STD is 3-5 mGal and the mean value is 8-12 mGal The results are shown in Fig. 1 and summarized in Table 1.
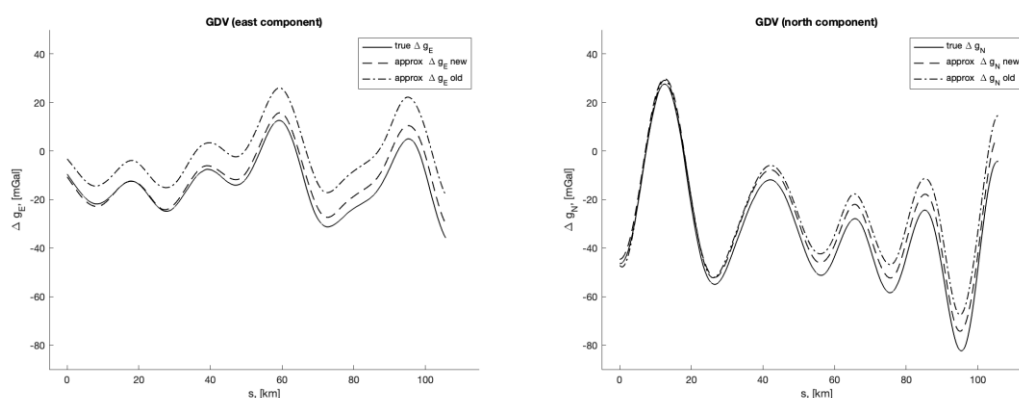
Fig.1 The GDV estimates in the new approach (dashed), in the standard (time-dependent) approach (dashed with dots) and the reference data (solid) at a line. Left: the east component, right: the north component, mGal.

Tab.2 The statistics from GDV estimation at two repeated lines

Statistics	New approach	Standard approach
Mean of Δg_E estimate error, mGal	2,74	12,24
Mean of Δg_N estimate error, mGal	4,72	7,82
STD of Δg_E estimate error, mGal	2,40	3,44
STD of Δg_N estimate error, mGal	1,78	5,19

5 Conclusions

The problem of the GDV estimation from strapdown airborne gravimeter measurements at repeated lines was solved. The a priori model of GDV was designed using the cubic B-splines parametrized by the distance along the line. The airborne vector gravimetry problem was reduced to optimal stochastic and Kalman filtering. The efficiency of the estimation algorithm was approved in simulated data processing, which showed good results even with two repeated lines (low STD and mean values were achieved).

References:

[1] Methods and technologies for measuring the earth's gravity field parameters. Cham, Switzerland, Springer. 2022. 388 p.

[2] Golovan A.A., Vyazmin V.S. Methodology of airborne gravimetry surveying and strapdown gravimeter data processing. Gyroscopy Navig. 2023. V. 14. P. 36–47.

[3] Kwon J.H., Jekeli C. The effect of stochastic gravity models in airborne vector gravimetry. J. Geoph., 2002. P. 770–776.

[4] Cai W., Zhang K., Wu M. Improving airborne strapdown vector gravimetry using stabilized horizontal components. J. Appl. Geophys., 2013. V. 98. P. 79–89.

[5] Fang K., Cai T. An algorithm for strapdown airborne gravity disturbance vector measurement based on high-precision navigation and EGM2008. Sensors. 2024. V. 24. P. 5899.

[6] Farrell J. A. Aided navigation systems: GPS and high rate sensors. 2008. New York, NY, McGraw-Hill.

[7] Cheney E.W., Kincaid D.R. Numerical Mathematics and Computing. Brooks/Cole Publishing Co., 2007. 784p.

[8] Kailath T., Sayed A.H., Hassibi B. Linear Estimation. Prentice Hall, Upper Saddle River, NJ, USA, 2000. 854p.

Assessments of The Accuracy of Modern Global Models of The Earth's Gravity Field and Experience of Their Use in Performing Marine Gravimetric Surveys

P. S. Mikhailov, V. N. Koneshov 22, V. N. Solovyev

(Schmidt Institute of Physics of the Earth of the Russian Academy of Sciences, Moscow)

Abstract: Presents estimates of the accuracy of modern global models of the Earth's gravity field, obtained from comparisons with area surveys. As a result of estimates made at different latitudes and over different structures, patterns of error distribution in models and features of their change were revealed. Examples of practical application of models in marine surveys are considered.

1   Introduction

The global satellite model of the Earth's gravitational field contains data about anomalies in free air at the geoid level. The possibilities for practical use of models depend on such characteristics as resolution and accuracy of anomaly representation. Thanks to the satellite altimetry method, modern models have better detail and accuracy specifically in the World Ocean. The maximum resolution of a modern model based on altimetry data has now reached 1 arc minute. This resolution allows the creation of anomaly maps at a scale of up to 1:200000. At the same time, satellite data contain larger errors than instrumental measurements, therefore, using data from area marine surveys as more highly accurate, we can judge the reliability of global models. In addition, global satellite models of the Earth's gravity field are currently already finding effective application in the process of performing marine gravimetric surveys. Below are some examples of how the authors have used global models to clarification the systematic drift of a marine gravimeter and to perform a marine survey without initial and final still readings at ports calls. [1].

2   Estimates of Global Models

The most effective and simple way to estimate a model field is to compare the model data with high-precision direct area surveys, for example with marine surveys. At the Schmidt Institute of Physics of the Earth, global comparisons of modern global gravity models with high-precision marine surveys at different latitudes and over various geomorphological structures of the World Ocean were carried out. The objective of these studies was to determine the actual accuracy of the models, identify the main trends in their development and patterns in the distribution of errors.

The assessments showed that in the area of   abyssal plains, all the studied models have close accuracy relative to marine gravity surveys. The error of the model data in terms of the standard deviation in such regions is estimated to be around 1 mGal. However, with the increase in the magnitude of the horizontal gradient of the change in the anomalous field, the error of the models increases significantly, as do the discrepancies in accuracy between the early and new models. For example, in the areas of mid-ocean ridges, the error of the EGM2008 model according to the

standard deviation is estimated at 5 mGal, the Chinese SGG-UGM-2 model [2] at 4 mGal, and the latest Sandwell and Smith [3] and DTU [4] models are already at 2.5 – 2.7 mGal. It was also revealed that the errors of modern global models in almost all areas of the World Ocean depend to a greater extent on the magnitude of the full gradient of the anomalous field itself and do not depend on latitude [5].

Using estimates obtained in regions containing different values of the horizontal gradient of the anomalous field structure, it is possible to determine the general functional dependence of the error of the global model on the value of the horizontal gradient and, thus, obtain estimated information about the expected error of this model in any region of interest in the World Ocean.

3   Using global models in marine surveys

The increasing accuracy of global models allows them to be effectively used in the practice of performing marine gravimetric surveys. Using model data on the Earth's anomalous gravity field, such problems as clarification the systematic drift of a marine gravimeter and quality control of the devices operations can be solved.

As a rule, a marine gravimeter has a continuous zero drift, individual for each device over a certain range. The drift velocity is usually determined during initial still readings and is refined during final still readings. Another well-known way to refine the drift rate of a gravimeter, if it is linear, is to evaluate the instrument readings at the intersection points of the survey lines. However, in modern practice, gravity surveys are often carried out together with seismic surveys, and modern seismic vessels can operate practically without port calling. Moreover, 3D seismic surveys often do not involve intersecting (tie) lines. In such a situation, the only way to determine the operating parameters of the gravimeter is to use model data on gravity field anomalies.

Since the global model does not contain systematic error components, comparison of gravimeter measurements with the model field allows us to identify systematic drift in the instrument readings. Moreover, with the increasing accuracy and detail of modern models, it becomes possible to determine not only the speed, but also the nature of the drift and take this into account when processing the survey. Figure 1 shows an example of the gravimeter performance evaluation at the testing site, after it was switched on from a "cold" state directly into the sea. The figure also shows the change in drift velocity caused by the transient process.
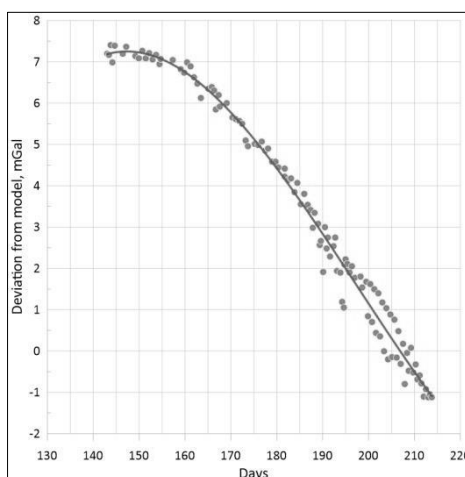
Fig.1 Determination of the speed and nature of the gravimeter drift in the sea by comparing measurements on lines with data from the global model.

The second task in which the global model finds its applicability is quality control of the correct operation of devices. At the stage of quality control of marine measurements, it is necessary to promptly evaluate the gravimeter readings. If there are several gravimeters on board, this is done by comparing their readings, but differences in the readings of two gravimeters also lead to ambiguity. Using a model field allows us to identify deviations in the operation of one or several gravimeters on board and take measures to eliminate possible problems.

The third task is to reduce the survey to the level of a global model field. In the practice of modern marine gravimetry, reliable IGSN71 stations are not always available in all ports of departure of the vessel. And in the case of turning on the gravimeter directly at the sea, there is no tie to the absolute value of gravity in principle. The global model field is represented at sea at the geoid level, so it can be used to calculate the absolute level of the marine survey, which also allows for joint processing of surveys performed from different coastal GSN71 stations or in their absence.

4 Conclusion

Actual estimates of modern gravity field models in open areas of the World Ocean are shown. In terms of accuracy and detail of anomaly representation, satellite models cannot yet replace instrumental marine and airborne gravimetric surveys. However, they can be used to plan surveys and during their production in order to improve the quality and reliability of the data obtained.

References:

[1] Peshekhonov V. G., Stepanov O. A. (eds.). Methods and Technologies for Measuring the Earth's Gravity Field Parameters. Springer Nature (Switzerland). 2022. P. 396. DOI: 10.1007/978-3-031-11158-7.

[2] Wei Liang, Jiancheng Li, Xinyu Xu, Shengjun Zhang, Yongqi Zhao A High-Resolution Earth's Gravity Field Model SGG-UGM-2 from GOCE, GRACE, Satellite Altimetry, and EGM2008 // Research Geodesy and Survey Engineering. 2020. Vol. 6. Issue 8. Pp. 860 – 878. DOI: 10.1016/j.eng.2020.05.008

[3] Sandwell, D. T., Müller R. D., Smith W. H. F., Garcia E., Francis R. New global marine gravity model from CryoSat-2 and Jason-1 reveals buried tectonic structure. Science. 2014. Vol. 346. No. 6205. pp. 65-67. DOI: 10.1126/science.1258213.

[4] Andersen O.B., Knudsen P. The DTU17 global marine gravity field: First validation results, in International Association of Geodesy Symposia. Cham. Springer. 2020. pp. 83-87. DOI: 10.1007/1345_2019_65

[5] Mikhailov, P.S., Koneshov, V.N., Solovyev, V.N., Zheleznyak L. K.   New Results of Estimation of Modern

Global Ultrahigh-Degree Models of the Earth's Gravity Field in the World Ocean. Gyroscopy Navig. 2022. 13, pp. 210－221 DOI: 10.1134/S2075108722040095

Results of The Development of A High-Precision Relative Gravimeter and Test Stands for Its Research

R.A. Davlatov, S.B. Akpanbetov, D.S. Bobrov

(FSUE "VNIIFTRI", Mendeleevo settlement, Moscow region, Russia)

Abstract: The paper presents the results of the development of three prototype high-precision relative gravimeters called "Peshikhod." The sensitive system of the gravimeter is made of metal, which differentiates it from foreign quartz gravimeters produced by Scintrex Ltd. The work includes the results of tests conducted on the prototypes using specialized test stands.

1 Introduction

Currently, high-precision relative gravimeters are used in geological exploration work, topographic and geodetic support, and also for the formation of navigation gravimetric maps for a promising navigation system based on the Earth's gravitational field [1-3].

Today, almost 100% of the world's fleet of relative gravimeters are represented by products from Scintrex Ltd (Canada) - CG-5 Autograv and CG-6 Autograv. At the same time, the use of foreign portable gravimeters in the territory of the Russian Federation is limited. Thus, it became necessary to develop a domestic high-precision relative gravimeter, which is not inferior in accuracy to foreign analogues.

2 Operating principle

The operating principle of the developed small-sized portable high-precision relative gravimeter is based on the use of precision spring scales [4]. The change in gravity is determined by measuring the value of the extension of the metal spring 1 (Fig. 1) according to the following formula:

$$\Delta g = \frac{k}{m} \Delta l \, , \tag{1}$$

where k - th $_e$ proportionality coefficient taking into account the spring stiffness;

Δl - spring extension;

m - is the mass of the weight.

To assess the change in the length of spring 1 (Fig. 1), a capacitive displacement reading system based on a three-plate capacitor is used. The capacitor has three plates, two stationary and a third movable one between them (it is also the sensitive element 3). The flat sensitive element 3 has a special horizontal suspension on a thread 4 stretched between the arms of a specially heat-treated "bow" 5.
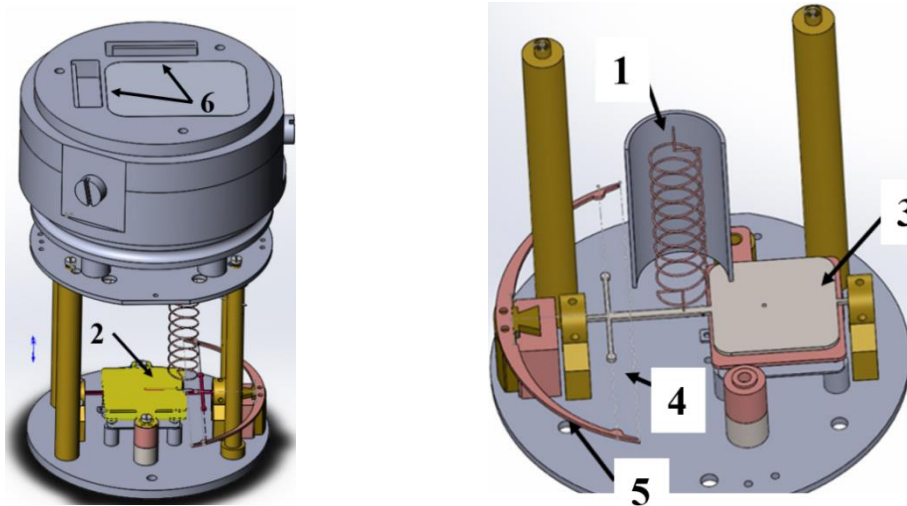
Fig. 1 Model of the gravimeter sensitive system

3 Stands for research

For the experimental evaluation of the sensitivity of the capacitive pickup system from the gravimeter sensor, a stand was developed consisting of stationary plates placed on the platform and a movable plate installed on a precision platform for movement. The sensitivity assessment was performed by comparing the measurement result of the capacitive pickup system with the nominal value of the specified micro-displacement of the "moving" plate. In this case, several movements are performed with a step of 5 nm. As a result, the accuracy of the developed capacitive system was 0.60 nm.

To evaluate the sensitivity of the prototypes, a tilt-and-turn stand was developed, which allows changing the angles between the directions of gravity and the sensitive axis of the gravimeter. This results in an "apparent" change in gravity at the gravimeter location, which can be calculated. The deviation of the measured value from the calculated one will allow us to estimate the sensitivity.

To evaluate the error of prototypes, a vertical displacement stand was developed, which allows gravimeters to be raised up to 1.5 meters. First, measurements of the increment of the acceleration of gravity during vertical movement are performed using another gravimeter whose parameters are known (for example, the CG-6 Autograv gravimeter [5]). Then, using experimental samples, the same series of measurements of the increase in gravity force is carried out. The difference between the values obtained by the experimental samples and the "standard" gravimeter will allow us to estimate the error of the created gravimeters.

4 Conclusion

The paper presents the results of the development of three prototypes of a high-precision relative gravimeter "Peshekhod" with a metal sensitive system. Special stands have been developed for testing prototypes: for assessing the sensitivity of the capacitive system, the sensitivity of prototypes using the tilt method, and for assessing accuracy using the vertical displacement method.

References:

[1]   Denisenko O.V., Pustovoit V.I., Silvestrov I.S., Fateev V.F. Problems of development of seamless assistive navigation technology in GNSS GLONASS based on measurements of geophysical field parameters // Almanac of modern metrology. 2020. No. 4 (24). P. 127-160.

[2]   V. G. Peshekhonov and O. A. Stepanov (eds.) «Methods and Technologies for Measuring the Earth's Gravity Field Parameters» // Earth Systems Data and Models 5, https://doi.org/10.1007/978-3-031-11158-7.

[3]   H. Liu et al., «Comprehensive Features Matching Algorithm for Gravity Aided Navigation» // IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 1505305, doi: 10.1109/LGRS.2022.3192408.

[4]   Akpanbetov S. B., Bobrov D. S., Davlatov R. A., Denisenko O. V., Donchenko S. I., Kravchenko A. I., Fateev V. F. Patent application No. 2024106567/28 (014535) dated 13.03.2024. "Spring sensitive element of a gravimeter".

[5]   Shan-liang Liu, Jin-yang Feng, Qi-yu Wang, Duo-wu Su, Chun-jian Li, and Shu-qing Wu "Investigation on the dynamic characteristics of CG-6 relative gravimeter for the micro-gravity network", Proc. SPIE 11053, Tenth International Symposium on Precision Engineering Measurements and Instrumentation, 110531V (7 March 2019);https://doi.org/10.1117/12.2511622.

Field Geophysical Complex for Creating Navigation Gravity-Magnetic Maps for Autonomous Navigation Systems

M.M. Murzabekov, D.S. Bobrov, V.P. Lopatin, E.S. Skvortsov

(FSUE "VNIIFTRI", Mendeleevo settlement, Moscow region, Russia)

Abstract: The paper presents a field geophysical complex, including a plumb line deviation meter and a high-precision relative gravimeter for creating maps of the Earth's gravity field (EGF) parameters and an onboard quantum magnetometer and a magnetovariation station for creating maps of the Earth's magnetic field (EMF) parameters. These maps can become the basis for autonomous navigation systems by EMF and EGF. The maps measured using the proposed geophysical complex and directions for further research are presented.

1 Introduction

The components of integrated systems of autonomous navigation using the Earth's magnetic and gravitational fields (EMF and EGF) are the user navigation equipment (UNE) of the GNSS, the inertial navigation system (INS), the onboard measuring device and the reference navigation map of the parameters of these fields [1-6]. In such a navigation system, when external conditions do not affect the operation of the GNSS NAP, the "INS+NAP" integration mode is used. In cases where the NAP is not capable of solving the navigation problem, the correction of the deviation of the INS readings is carried out based on measurements of the parameters of the EMF and EGF. In this case, the reference map of the EMF and EGF is used to determine the current location based on the results of comparison with the measurement data of the onboard measuring device.

It should be noted separately, according to the works [7, 8], that the EGF maps are also needed as a source of corrective information for the INS. These works highlight that for promising INS, the error in determining the parameters of the GPS will make a significant contribution to the accuracy of the navigation solution.

The paper presents a field geophysical complex that allows creating maps of GPZ and MPZ parameters with previously unattainable accuracy and discreteness, which can be used in the creation and testing of integrated autonomous navigation systems.

2 Composition of the field geophysical complex

The geophysical complex includes: 1) EGF parameter meters – deflection of the vertical [9] and high-precision relative gravimeter of the Scintrex CG-5 type (Fig. 1a); 2) EMF parameter meters – an onboard quantum magnetometer placed on a quadcopter and a magnetovariation station (Fig. 1b).

a)   b)
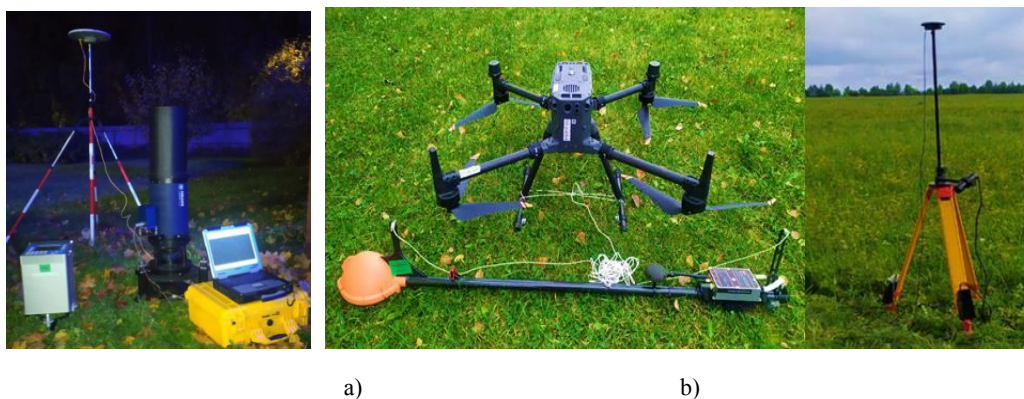
Fig. 1   Geophysical complex

3   Map creation results.

3.1   EGF maps

Joint measurements of the acceleration of gravity g (AGG) and the components of the deflection of the vertical (DoV) ξ, η using a gravimeter and an astrometer were carried out on the territory of the Moscow region at a test site with linear dimensions of approximately 20x20 km. The polygon includes 32 points with an average distance of approximately 4 km from each other; the measurement time at each point does not exceed 1 hour.

To create DoV and AGG maps, a uniform grid with a step of 100 m was created, including measurement points. The interpolation of the DoV and AGG values into points of a uniform grid was performed using the well-known "removal-recovery" method, which allows taking into account the influence of topographic masses and increases the accuracy of map creation [10].

Additionally, the components of the AGG g_x, g_y and the gravitational gradients T_{ij} were calculated using these maps [11]. Fig. 2 shows examples of maps of some parameters of the EGF.
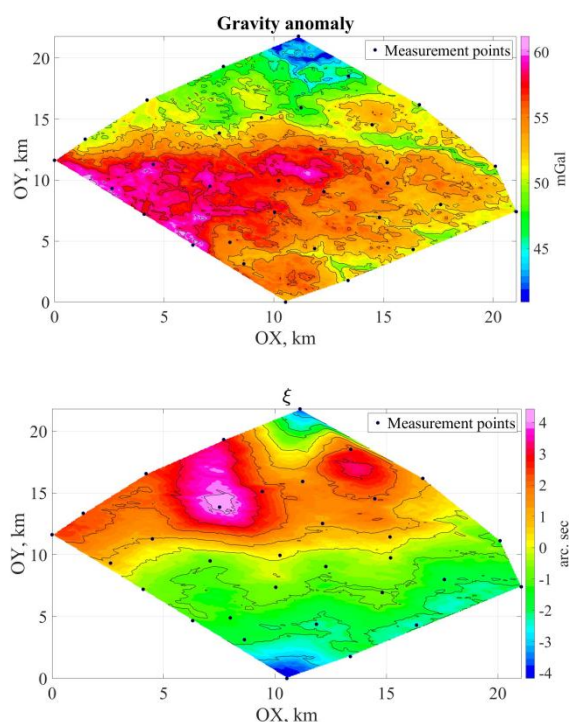

Fig. 2   Examples of created EGF maps

3.2 EMF maps

To create maps of the anomalous component of the magnetic field induction, measurements were carried out in tacks at a flight altitude of 150 m with a distance between tacks of 200 m. A magnetovariation station was located in the measurement area. The map of the anomalous component of EMF induction, created based on the results of these measurements, is shown in Fig. 3a. The created map of the EMF is planned to be used in testing the model of the navigation system, which is being created at the Federal State Unitary Enterprise VNIIFTRI. The model includes the above-mentioned magnetometers and quadcopter, as well as a INS on MEMS sensors (Fig. 3b).

a)   b)
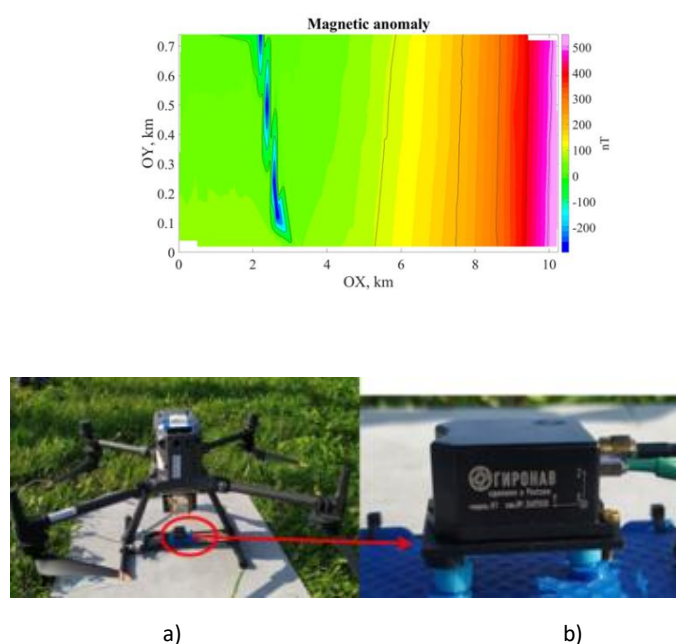
Fig.3 Anomalous EMF map

4 Conclusion

The paper presents a geophysical complex including EGF and EMF parameter meters. This complex allows creating maps of AGG, DoV, components of AGG, gravity gradients, as well as a map of the anomalous component of the EMF. These maps can be used, in particular, in the creation and testing of integrated navigation systems for the EGF and EMF.

The EMF map created using the complex will become a reference when testing the developed layout of the navigation system.

References:

[1] Denisenko O.V., Pustovoit V.I., Silvestrov I.S., Fateev V.F. Problems of development of seamless assistive navigation technology in GNSS GLONASS based on measurements of geophysical field parameters // Almanac of modern metrology. 2020. No. 4 (24). P. 127-160.

[2] Sazonova T.V.   Experimental studies of the accuracy characteristics of correlation-extreme navigation systems based on the Earth's magnetic field // Almanac of modern metrology. 2020. No. 4 (24). P. 86-96.

[3] Stepanov O.A., Nosov A.S. Algorithm for correcting a navigation system based on map and measuring device data that does not require preliminary estimation of the field values along the traveled trajectory // Gyroscopy and navigation. 2020. Vol. 28, No. 2 (109). P. 70-90. DOI 10.17285/0869-7035.0029.

[4] Canciani, A.J., Raquet, J.F. Airborne Magnetic Anomaly Navigation // IEEE Transactions on Aerospace and Electronic Systems. 2017. Vol. 53, no. 1. P. 67–80. DOI: 10.1109/TAES.2017.2649238.
V. G. Peshekhonov and O. A. Stepanov (eds.) «Methods and Technologies for Measuring the Earth's Gravity Field Parameters» // Earth Systems Data and Models 5, https://doi.org/10.1007/978-3-031-11158-7.

[5] H. Liu et al., «Comprehensive Features Matching Algorithm for Gravity Aided Navigation» // IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 1505305, doi: 10.1109/LGRS.2022.3192408.

[6] V. G. Peshekhonov and O. A. Stepanov (eds.) «Methods and Technologies for Measuring the Earth's Gravity Field Parameters» // Earth Systems Data and Models 5, https://doi.org/10.1007/978-3-031-11158-7.

[7] Popadyev V.V., Soroka A.I., Polubekhin A.I., et al. Correction capabilities of inertial navigation systems based on gravimetric maps of the earth // Scientific Bulletin of Moscow State Technical University of Civil Aviation. 2015. No. 222. Pp. 90-97.

[8] Peshekhonov V.G. The problem of plumb line deviation in high-precision inertial navigation // Gyroscopy and Navigation. 2020. Vol. 28, No. 4 (111). P. 3-15. DOI 10.17285/0869-7035.0046.

[9] Murzabekov M.M., Fateev V.F., Yuzefovich P.A.   Measurements of plumb line deviations at a well-known Moscow attraction using a digital astrometer // Astronomical Journal. 2020. Vol. 97, No. 10. Pp. 873-880. DOI: 10.31857/S0004629920100059.

[10] Meng Y., Hirt C., Pail R. TGF: A New MATLAB-based Software for Terrain-related Gravity Field Calculations // Remote Sensing. 2020. 12. 1063. DOI: 10.3390/rs12071063.

[11] Ogorodova L.V. Normal field and definition of anomalous potential. Moscow: MIIGAiK. 2010. P. 106.

2024

Chinese-Russian
"Navigation and Motion Control"
Youth Forum

Topic 5: Unmanned Aerial Vehicle

An Outdoor Robust Visual/Inertial Localization Method Based on Temperature Compensation for UAVs

Liu Fengyu[1,2], Cheng Xianghong[1,2]

(1. Key Laboratory of Micro-inertial Instrument and Advanced Navigation Technology, Ministry of Education, Southeast University, Nanjing 210096, China;

2. School of Instrument Science & Engineering, Southeast University, Nanjing 210096, China)

Abstract: In GNSS-denied environments, outdoor UAV localization accuracy is often reduced due to lighting changes and sparse texture features affecting visual sensors, and temperature variations and high-frequency vibrations impacting MEMS-IMU, causing significant bias errors. To address these issues, a robust Visual/Inertial localization method based on temperature compensation is proposed. An LSTM-based IMU bias online compensation network, with a Teacher Forcing mechanism to enhance learning capabilities, is designed. An IMU error model is constructed using calibrated IMU measurements, temperature data, and true bias values from gyroscopes and accelerometers to train the network and develop a UAV bias model. During operation, the model provides real-time online compensation for gyroscope and accelerometer bias. A joint optimization function is used, incorporating IMU measurements with compensated bias and visual residuals to obtain localization results, with a visual adaptive factor mitigating visual degradation from outdoor conditions. Experiments show a 54.1% reduction in average absolute positioning error compared to VINS-Mono, and ablation studies confirm enhanced robustness.

 Key words: Temperature compensation; UAV; Visual/Inertial; Localization

1　Introduction

With the rapid development of unmanned technology, the demand for UAVs in fields such as disaster monitoring, and battlefield search is steadily increasing [1,2]. Microelectromechanical system inertial measurement units (MEMS-IMU) are widely used in UAVs due to their small size and low power consumption. However, they suffer from error accumulation over time. Recently, with advancements in computer vision [3], leveraging the complementary characteristics of MEMS-IMU and visual sensors has attracted increasing research interest, leading to the development of vision/inertial navigation fusion algorithms [4]. In 2017, Shao et al. [5] released VINS-Mono, a complete vision-inertial simultaneous localization and mapping (SLAM) system based on tightly-coupled graph optimization. In 2021, Carlos et al. [6] proposed ORB-SLAM3, which incorporates IMU data based on ORB-SLAM2 [7], utilizing a maximum a posteriori estimation algorithm for a more effective Visual/Inertial initialization process.

However, MEMS-IMU bias errors are prone to nonlinear changes due to high-frequency UAV vibrations and temperature fluctuations in outdoor environments. Additionally, in areas with complex lighting conditions or low texture features, the visual sensor may degrade, producing outliers and failing to provide reliable visual features for joint optimization, which reduces

localization accuracy and prevents accurate estimation of IMU bias. To address this, Wang et al. [8] proposed an online IMU self-calibration method that adaptively adjusts the IMU intrinsic parameter optimization weight based on movement intensity. Huang et al. [9] improved the DCN network [10] by using convolutional layers to estimate IMU system error matrices, compensating for time-varying gyroscope errors through the DCN network. Wu et al. [11] introduced an online bias calibration method based on an improved long short-term memory (LSTM) network to learn IMU bias errors and compensate for IMU bias under visual degradation. To address unreliable visual features, He et al. [12] proposed PLE-SLAM, which extracts both point and line features in a SLAM framework integrated with IMU and leverages semantic information and geometric constraints in the map to eliminate dynamic features.

Current methods do not simultaneously consider temperature effects on IMU bias and the impact of outliers due to visual degradation on joint optimization. Therefore, this paper proposes an outdoor robust Visual/Inertial localization method based on temperature compensation for UAVs. The main contributions are as follow:

- Design an LSTM-based online temperature compensation network to provide real-time compensation for gyroscope and accelerometer bias errors during UAV operation.
- Introduce a visual adaptive factor to mitigate the impact of outliers from degraded visual scenes in outdoor environments, enhancing the reliability of visual features in optimization.
- Demonstrate the localization accuracy and robustness of our method on a self-built UAV.

2 Methodology
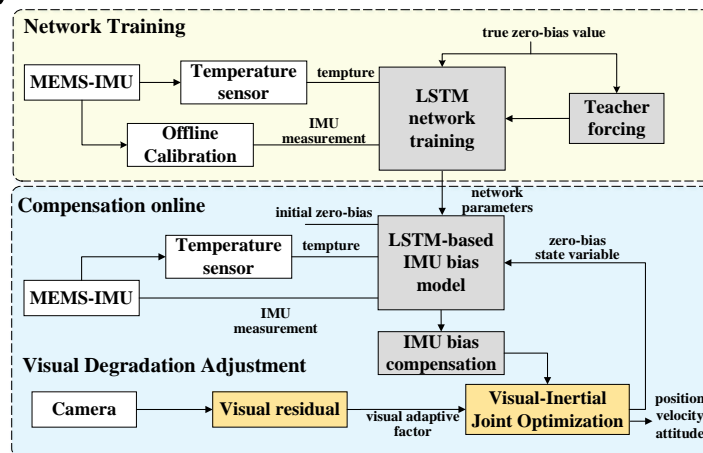
Fig.1　The schematic of the proposed method

The scheme is shown in Fig. 1 and is divided into three main parts. In Network Training section, an LSTM-based online bias compensation network is first designed, integrating the Teacher Forcing mechanism [13] to eliminate accumulated errors in the network and enhance its learning capability. Next, an IMU measurement model is constructed, where calibration parameters obtained from offline calibration are used to compensate for installation errors in the IMU measurements. Then, the LSTM network is trained using the compensated IMU measurement data, environmental temperature information, and the true bias values of the gyroscope and accelerometer to construct the UAV bias model based on LSTM. In Online Compensation section, the trained UAV bias model is applied during UAV operation to perform real-time compensation for the gyroscope and

accelerometer biases at each IMU measurement. In Visual Degradation Adjustment section, a visual adaptive factor is designed to adapt to visual degradation caused by outdoor scenes, enhancing the reliability of visual features in the optimization process. Finally, joint optimization is performed using the bias-compensated IMU measurements and visual residuals, yielding refined localization information.

2.1 IMU measurement model

At time k, the IMU measurement model can be rewritten as:

$$
\begin{aligned}
\boldsymbol{a}^s(t) &= (\boldsymbol{K}_a + \boldsymbol{T}_a)\boldsymbol{a}^b(t) + \boldsymbol{b}_{a0}(t) + \boldsymbol{b}_{Ta0}(t) + \boldsymbol{R}_w^b(t)\boldsymbol{g}^w + \boldsymbol{n}_a(t) \\
\boldsymbol{\omega}^s(t) &= (\boldsymbol{K}_g + \boldsymbol{T}_g)\boldsymbol{\omega}^b(t) + \boldsymbol{b}_{g0}(t) + \boldsymbol{b}_{Tg0}(t) + \boldsymbol{n}_g(t)
\end{aligned} \tag{1}
$$

where $\boldsymbol{a}^s(t)$, $\boldsymbol{\omega}^s(t)$ represent the accelerometer and gyroscope measurements, while $\boldsymbol{a}^b(t)$, $\boldsymbol{\omega}^b(t)$ denote their true values. K_a, K_g and T_a, T_g are the scale factors and temperature coefficients of the accelerometer and gyroscope, respectively. $\boldsymbol{b}_{a0}(t)$, $\boldsymbol{b}_{g0}(t)$ and $\boldsymbol{b}_{Ta0}(t)$, $\boldsymbol{b}_{Tg0}(t)$ respectively refer to the zero-bias and zero-bias temperature error, of the accelerometer and gyroscope. In online operational, installation errors of the MEMS-IMU remain relatively stable. Therefore, installation error obtained through of online calibration becomes:

$$
\begin{aligned}
\boldsymbol{a}^s(t) &= \hat{\boldsymbol{a}}^b(t) + \boldsymbol{b}_a(t) + \boldsymbol{R}_w^b(t)\boldsymbol{g}^w + \boldsymbol{n}_a(t) \\
\boldsymbol{\omega}^s(t) &= \hat{\boldsymbol{\omega}}^b(t) + \boldsymbol{b}_g(t) + \boldsymbol{n}_g(t)
\end{aligned} \tag{2}
$$

where

$$
\begin{aligned}
\boldsymbol{b}_a(t) &= \boldsymbol{b}_{a0}(t) + \boldsymbol{b}_{Ta0}(t) + \delta\boldsymbol{K}_a\boldsymbol{a}(t) \\
\boldsymbol{b}_g(t) &= \boldsymbol{b}_{g0}(t) + \boldsymbol{b}_{Tg0}(t) + \delta\boldsymbol{K}_g\boldsymbol{\omega}(t)
\end{aligned} \tag{3}
$$

where $\hat{\boldsymbol{a}}^b(t)$, $\hat{\boldsymbol{\omega}}^b(t)$ represent the calibrated IMU measurements, while $\boldsymbol{b}_a(t)$, $\boldsymbol{b}_g(t)$ denote the accelerometer and gyroscope bias errors, respectively. These biases include zero-bias error, zero-bias temperature error, and scale factor errors.

2.2 LSTM-based online bias compensation network

The bias of IMU changes over time, making it a time series with temporal dependencies. To predict future bias values, historical IMU measurement data and temperature data are used as inputs in a neural network model. An LSTM neural network function F is defined, along with its associated parameters θ, for this prediction task:

$$
F(\theta) : (\boldsymbol{x}_{b_k}^{b_{k+1}}, T_k, \boldsymbol{b}_{a,k}, \boldsymbol{b}_{g,k}) \mapsto (\hat{\boldsymbol{b}}_{a,k+1}, \hat{\boldsymbol{b}}_{g,k+1}) \tag{4}
$$

On the left side of the above equation, the network input includes IMU measurements $\boldsymbol{x}_{b_k}^{b_{k+1}}$ $=[\hat{\boldsymbol{a}}^b, \hat{\boldsymbol{\omega}}^b]$ from time k to $k+1$, the temperature T_k recorded by the temperature sensor at time k, and the accelerometer and gyroscope biases $\boldsymbol{b}_{a,k}$ and $\boldsymbol{b}_{g,k}$. The right side of the equation represents the network output, which is the estimated accelerometer and gyroscope biases, $\hat{\boldsymbol{b}}_{a,k+1}$ and $\hat{\boldsymbol{b}}_{g,k+1}$. The predicted bias is used to compensate the IMU measurement, and the corresponding IMU state $\hat{\boldsymbol{x}}_{k+1}$ quantity is solved.

When using the LSTM neural network to predict IMU bias, the network incorporates the predicted result from the previous time step as input for the next time step during training. If the

prediction at the previous time step is not accurate, the error accumulates and propagates, leading to poor training performance. To address these issues, a Teacher Forcing-LSTM network structure is designed. This structure introduces the Teacher Forcing mechanism, which effectively reduces the accumulated errors and significantly improves the model's training performance.

2.3 Visual/Inertial Joint Optimization

In outdoor environments, visual sensors are prone to factors such as significant changes in lighting intensity and sparse environmental features, which can reduce the image feature extraction and matching capabilities. The involvement of image outliers in the integrated navigation system can lead to a decline in system localization accuracy. Therefore, an adaptive visual factor, s_{visual} is designed to measure the degradation degree of the visual sensor.

$$s_{visual} = u / w$$
$$\Phi(\omega_{visual}) = s_{visual}(\sqrt{\omega_{visual}^2} - 1)^2 = \frac{u}{w}(\sqrt{\omega_{visual}^2} - 1)^2 \tag{5}$$

where u represents the number of tracked visual features, and w denotes the size of the sliding window. The greater the number of visual features tracked within the sliding window, the more accurate the visual information becomes. ω_{visual} reflects the weight of the visual reprojection residual in the optimization process. $\Phi(\omega_{visual})$ represents the penalty function of the adaptive visual factor. Ultimately, the adaptive joint optimization function is constructed based on this adaptive factor.

$$X = \min_X \left\{ \left\| r_{prior} \right\|^2 + \sum_m^{N_1} \left\| r_m^{imu} \right\|^2 + \sum_n^{N_2} \omega_{visual} \left\| r_n^{visual} \right\|^2 + \Phi(\omega_{visual}) \right\} \tag{6}$$

where X represents the state variables to be optimized, r_{prior} denotes the marginalized prior information, r_m^{imu} and r_n^{camera} respectively represents the residual of the m-th IMU measurement n-th visual reprojection within the sliding window.

3 Experiments

The gyroscope bias prediction network and the accelerometerbias prediction network are trained separately. The training uses the Adam optimizer, and to prevent overfitting, the gyroscope bias network has a batch size of 128, 200 training epochs, a learning rate of 1×10^{-5}, and a weight decay coefficient of 1×10^{-5}. The accelerometer bias network has a batch size of 64, 150 training epochs, a learning rate of 5×10^{-5}, and a weight decay coefficient of 1×10^{-5}. The training data is split into 70% training, 15% validation, and 15% testing. Supervised learning is used to train the networks, with mean squared error as the evaluation metric. The loss function is set as follows:

$$loss = \frac{1}{n}\sum_{k=1}^{n} \left\| b_k - \hat{b}_k \right\|^2 \tag{7}$$

3.1 Offline Temperature Compensation Verification

The MEMS-IMU is placed in a temperature-controlled box, where the temperature variation inside the box is controlled and data from the gyroscope and accelerometer are collected. The sensor output at this stage represents the corresponding bias. The temperature range inside the box is set from -10°C to 50°C based on the actual operating temperature range of the UAV. To fully consider

the impact of temperature change rate on bias variation, the temperature difference between IMU and the internal temperature sensor, as well as the discrepancy in bias variation between the heating and cooling processes, zero-bias temperature error calibration experiment is designed based on the ramp method as follows:

Training data collection: The temperature change rate is set to 0.5°C/min to decrease from 50°C to 30°C, then 1°C/min from 30°C to 10°C, and finally 1.5°C/min down to -10°C. The temperature is then increased in the same manner, with a 10-minute hold at each 10°C interval to stabilize the internal temperature of the instrument.

Testing data collection: Data is collected in five groups with the following conditions: two groups each for heating and cooling at rates of 0.5°C/min and 1°C/min, and one group for cooling and then heating at 1.5°C/min. Each data set lasts 500 seconds, with temperature changes ranging from -10°C to 50°C.

Tab.1   The results of the temperature error compensation on the test set

| Dataset		Polynomial fitting			LSTM network		
		x	y	z	x	y	z
Set1	Acce(m/s^2)	0.0113	0.0177	0.0101	0.0043	0.0039	0.0086
	gyro(rad/s)	0.0006403	0.0054	0.0003847	0.0006158	0.0007874	0.0003695
Set2	Acce(m/s^2)	0.0041	0.0069	0.0229	0.0033	0.0045	0.0089
	Gyro(rad/s)	0.0025	0.0043	0.0008334	0.0011	0.0008105	0.0009298
Set3	Acce(m/s^2)	0.0075	0.0109	0.0265	0.0037	0.0033	0.0127
	Gyro(rad/s)	0.0006504	0.0066	0.0004431	0.0009683	0.0006130	0.0003747
Set4	Acce(m/s^2)	0.0040	0.0057	0.0247	0.0037	0.0034	0.0219
	Gyro(rad/s)	0.0023	0.0065	0.0008215	0.0014	0.0009415	0.0006069
Set5	Acce(m/s^2)	0.0051	0.0105	0.0110	0.0049	0.0077	0.0074
	Gyro(rad/s)	0.0030	0.0054	0.0009123	0.0018	0.0008999	0.00050902
Average	Acce(m/s^2)	0.0064	0.0103	0.0190	0.0040	0.0046	0.0119
	Gyro(rad/s)	0.0018	0.0056	0.0010	0.0012	0.0008105	0.0005580

To compare the calibration performance of zero-bias temperature errors, polynomial fitting with wavelet threshold denoising and LSTM neural network methods are applied. The root mean square error (RMSE) of the IMU output after compensation is calculated to compare the effectiveness of the two methods. As shown in Table 1, the average accuracy of the three-axis accelerometer output using the BP neural network fitting method increased by 42.8% compared to polynomial fitting, while the accuracy of the three-axis gyroscope output improved by 69.4%. Overall, the LSTM neural network fitting method outperforms polynomial fitting, as it can more comprehensively account for the inter-coupling relationships among temperature, temperature change rate, and temperature change markers, resulting in a more accurate IMU zero-bias temperature error model and improved measurement precision after compensation.

3.2 UAV Integrated Navigation Experiment

To validate the effectiveness of the proposed method, a UAV experimental platform is built. The platform is based on the PX4 autopilot, with a 50A 4-in-1 ESC and a 2810-900KV brushless motor for power. The flight controller is the CUAV V5 nano, and the power supply is a 6s 5500mAh 75C model aircraft battery. The sensors include an Intel D435i camera, Xsens Mti-7 MEMS-IMU, Livox mid360 LiDAR, and MTF-01 optical flow module. Additionally, an Intel onboard computer is mounted. The completed UAV platform is shown in Fig. 2 left.

To simulate a degraded scenario, a courtyard on campus was selected to mimic a GNSS-denied environment, with nighttime conditions to simulate visual degradation. The semi-open courtyard also experienced temperature variations. The true IMU bias required for training were obtained by LiDAR/Inertial integrated navigation method, FAST-LIO [14], a highly efficient and robust odometry framework, which uses a tightly-coupled iterative extended Kalman filter to fuse IMU measurements with LiDAR point clouds, compensating for motion distortion through a backpropagation process. During the information fusion, the IMU bias state is continuously updated and output, which is aligned with the IMU measurement timestamps and used as the true IMU bias values for network training. The total training time for the gyroscope and accelerometer bias networks was approximately 40 minutes.
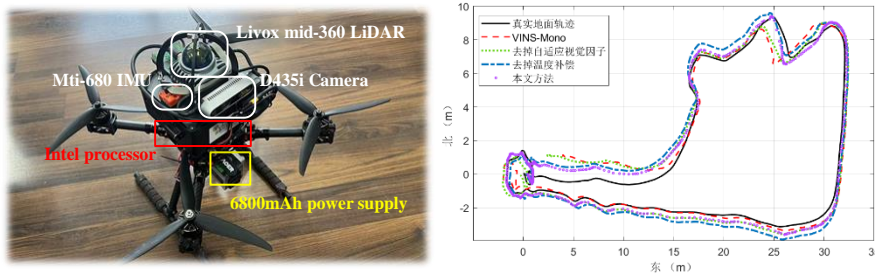
Fig.2 The UAV platform and Visual/Inertial localization error

The proposed method is compared with VINS-Mono, as well as with methods where the temperature compensation and visual adaptive factor modules are separately removed. The resulting localization performance is shown in Fig. 2 right. It is clear that the proposed method significantly enhances the localization accuracy of the UAV Visual/Inertial system. A more detailed analysis reveals that, in this scenario, the adaptive visual factor has a substantial impact on the robustness of the localization. When the adaptive visual factor is removed, the localization error increases notably. This is due to the poor lighting conditions in this scenario, leading to insufficient visual feature extraction. Without the inclusion of the visual adaptive factor, the influence of visual outliers on the system reliability becomes more pronounced. Additionally, removing temperature compensation also degrades the localization accuracy, as the scenario involves temperature fluctuations. If the impact of temperature on bias is not accounted for, the localization precision will decrease. The specific absolute position errors of the navigation trajectory, along with the percentage improvement in localization accuracy relative to the VINS-Mono method, are presented in Table 2. The quantitative results further demonstrate that the proposed method achieves a significant improvement in localization accuracy.

Tab.2 The distance of start-to-end (Unit: m)

Method	VINS-Mono	Ours without s_{visual}	Ours without LSTM	Ours
Errors	1.33	0.93	0.73	0.61
Improvement	-	30.1%	45.1%	54.1%

4 Conclusion

This paper proposes a temperature-compensated robust localization method for UAV vision/inertial navigation in field environments. The method effectively addresses issues of localization degradation due to illumination changes, sparse texture features, and IMU performance

drops under temperature fluctuations and high-frequency vibrations in GNSS-denied environments. By designing an LSTM-based online bias compensation network and introducing the Teacher Forcing mechanism, the method enhances the learning capability and compensation accuracy. The constructed IMU error model enables online bias compensation during actual flight. Experimental results demonstrate a significant improvement in localization accuracy over VINS-Mono on UAVs, reducing the average absolute position error by 54.1%. Additionally, the effectiveness of the designed temperature compensation and visual adaptive factor in enhancing the robustness of the integrated navigation system is validated.

References:

[1] Liu Feng-yu, Cao Yi, Cheng Xiang-hong, et al. Transformer-Based Local-to-Global LiDAR-Camera Targetless Calibration with Multiple Constraints[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73:1-13.

[2] Zhou R, Almustafa M K, Nehdi M L, et al. Automated localization of dike leakage outlets using UAV-borne thermography and YOLO-based object detectors[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2024, 218: 551-573.

[3] Yang X, Bist R B, Paneru B, et al. Computer Vision-Based cybernetics systems for promoting modern poultry Farming: A critical review[J]. Computers and Electronics in Agriculture, 2024, 225: 109339.

[4] Liu Feng-yu, Cao Yi, Cheng Xiang-hong, et al. A visual SLAM method assisted by IMU and deep learning in indoor dynamic blurred scenes[J]. Measurement Science and Technology, 2023, 35(2): 025105.

[5] Qin Tong, Li Pei-liang, Shen Shao-jie. Vins-mono: A robust and versatile monocular visual-inertial state estimator[J]. IEEE transactions on robotics, 2018, 34(4): 1004-1020.

[6] Campos C, Elvira R, Rodríguez J J G, et al. Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.

[7] Mur-Artal R, Tardós J D. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras[J]. IEEE transactions on robotics, 2017, 33(5): 1255-1262.

[8] Wang Zi-hui, Cheng Xiang-hong. Adaptive optimization online IMU self-calibration method for visual-inertial navigation systems[J]. Measurement, 2021, 180: 109478.

[9] Huang Feng-rong, Yi Bo-heng, Wang Xu, et al. Vehicle inertial navigation method based on deep learning and motion constraints[J]. Journal of Chinese Inertial Technology, 2022, 30(05): 569-575.

[10] Brossard M, Bonnabel S, Barrau A. Denoising IMU gyroscopes with deep learning for open-loop attitude estimation[J]. IEEE Robotics and Automation Letters, 2020, 5(3): 4796-4803.

[11] Cheng Xiang-hong, Wu Xin-yi, Liu Feng-yu. An improved LSTM neural network online calibration method of MEMS-IMU bias for UAV[J], Journal of Chinese Inertial Technology, 2024,32(03):213-218.

[12] He Jia-ming, Li Ming-rui, Wang Yang-yang, et al. PLE-SLAM: A Visual-Inertial SLAM Based on Point-Line Features and Efficient IMU Initialization[J]. arXiv preprint arXiv:2401.01081, 2024.

[13] Li Shi-yang, Wang Tong-tong, Li Guo-yuan, et al. Short-term ship roll motion prediction using the encoder–decoder Bi-LSTM with teacher forcing[J]. Ocean Engineering, 2024, 295: 116917.

[14] Xu Wei, Cai Yi-xi, He Dong-jiao, et al. Fast-lio2: Fast direct lidar-inertial odometry[J]. IEEE Transactions on Robotics, 2022, 38(4): 2053-2073.

Research on A Vehicle-Mounted Composite Vertical Wing UAV Take-Off and Landing Platform Device

Wang Yizhe[1], Chen Zhihao[1] , Peng Xiaojun[1] , Li Xiangjie[1] ,Yang Zhijun[1]

(1.No.16 Institute of No.9 Academe of China Aerospace Technology Corporation, Xi'an 710100,China)

Abstract: Aiming at the research hotspot of the combined application of vertical composite wing UAVs and vehicle-mounted chassis, a vehicle-mounted take-off and landing platformfor vertical composite wing UAVs is designed, which can realise the functions of wind direction detection, elevation and attitude adjustment, rotation locking, landing and recovery, etc. The use of this device has certain reference significance for the subsequent application of vertical lift composite wing UAVs combined with vehicle-mounted platforms.

 Key words: Vertical composite wing UAV; Vehicle-mounted ; Dropping platform.

1 Summary

In recent years, with the rapid development of UAV (Unmanned Aerial Vehicle) technology, UAVs have been widely used in military, civil and other fields. On the other hand, due to its unique characteristics, the vertical composite wing UAV has attracted many research institutions and commercial enterprises at home and abroad to conduct in-depth research[1-4]. Vertical composite wing UAV not only takes into account the characteristics of rotary wing UAV vertical take-off and landing, rotary stop, but also retains the advantages of fixed wing UAV long range and long endurance; the combination of UAV and vehicle chassis can be used, and can flexibly choose different operation modes according to the environmental situation, and further carry out all-weather, all-day, multi-target, multi-mission reconnaissance and surveillance, battlefield intelligence gathering and radiation source detection and other combat tasks[5]; Therefore, it has gradually become a research hotspot.

The vehicle-mounted take-off and landing platform for vertical composite wing UAV device designed in this paper is mainly designed for vertical composite wing UAVs with three-point landing gear, and it can realise the automatic lifting and locking of the UAV on the vehicle-mounted chassis in the process of transportation; Before take-off, it can detect the real-time change of wind speed and wind direction in the environment of the UAV, and adjust the direction of the nose of the UAV to ensure that the UAV takes off into the wind; when the UAV lands, it can unfold the recovery net to expand the landing area of the UAV, to ensure that the UAV lands safely on the landing platform and withdraws safely with the vehicle-mounted chassis.

2 Structural design and parameters

The structure of the vehicle-mounted take-off and landing platform for vertical composite wing UAV described herein consists mainly of a lift assembly, a rotary lock assembly and a recovery net

---

assembly. Among them, the rotary locking assembly includes the main platform, the wind speed and direction sensor, and the rotary locking mechanism. The specific composition is shown in Figure 1, the 3D model is shown in Figure 2, and the main performance indices and technical parameters are shown in Table 1.
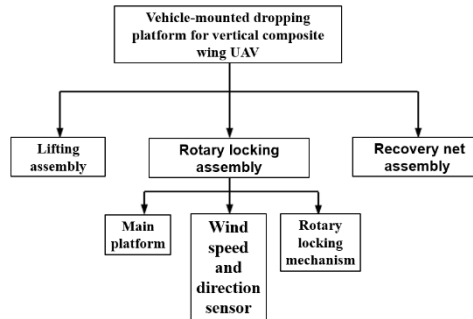
Fig.1 Structural composition of vehicle-mounted take-off and landing platform
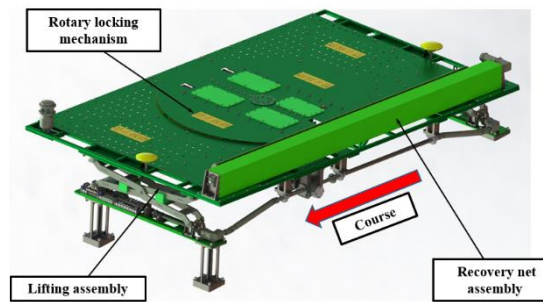
**Fig.2** Vehicle-mounted take-off and landing platform for vertical composite wing UAV 3D Model

Tab1 Key performance indicators and technical parameters

No.	Performance Indicators	Technical Parameters
1	Pendant mass/kg	16
2	Lifting stroke/mm	≮300
3	Pivoting angle/°	±180
4	Aircraft clamping force in locked condition/kg	≮20
5	Recovery net size/mm	Length 2300×Width 2500
6	Detectable uniform speed /m·s-1	≮15
7	Lifting time/s	30
8	Rotation time/s	18
9	Recovery net opening time/s	20
10	Aircraft load mass/kg	≯20
11	Power supply and consumption	DC28V、≯800W
12	Total weight of equipment /kg	≤260

2.1 Lifting assembly

The lift assembly is symmetrically arranged on both sides of the main platform to complete the lifting and lowering of the UAV placed on the surface of the main platform, which mainly consists of lift assembly A, lift power parts, lift assembly B, inclined iron and pad block, and the 3D model is shown in Figure 3. The inclined iron is modelled on the platform design of the vehicle body and is connected to the pad block, which is used in conjunction with the vehicle body to adjust the lift assemblies A and B on both sides to the same horizontal position.

When receiving the corresponding lifting instructions, the control system drives the lift motor to rotate, using the cross reducer to amplify the torque, output to both sides of the universal joint shaft end, the universal joint drives the lift A and lift B lower right angle reducer action, driving the platform for lifting or lowering action. The lifting motor is equipped with brake structure, can be self-locking stagnation at any position in the lifting process; emergency rescue operation, you can increase the temporary unlocking power supply, in the manual input of the cross reducer into the manual crank, to achieve the lifting components of the emergency operation.
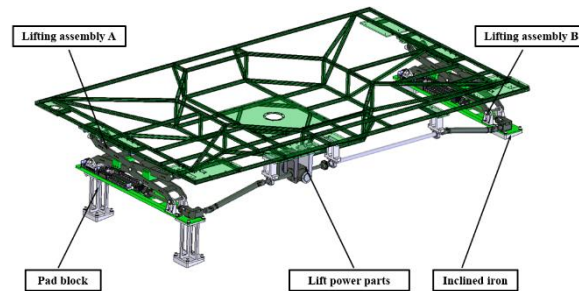
Fig.3 Schematic 3D model of lift assembly

2.2 Rotary locking assembly

The rotary locking assembly is installed above the lifting assembly and connected to the lifting assembly through the guide slide support. It mainly consists of the main platform, rotary locking mechanism, wind speed and direction sensor, and Beidou differential system. The 3D model is shown in Figure 4.
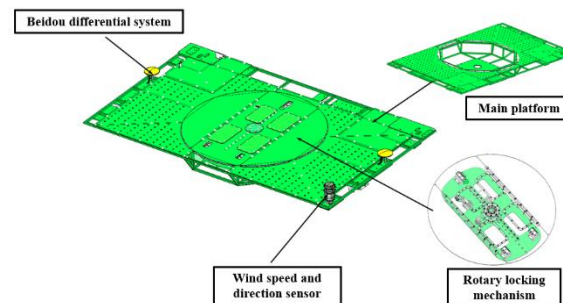

Fig.4 Schematic diagram of the structure of the rotary locking assembly

The main platform body is made of 6063 aluminium alloy pipe welding process, which is mainly used for the installation of the rotary locking mechanism, the recycling net assembly, the wind speed and direction sensor, and the Beidou differential system.

The wind speed and direction sensor is installed at the end of the main platform, which can sense the wind speed and direction of the surrounding environment in real time, and measure the real-time wind speed and direction values during the preparation of the UAV flight. The object is shown in Figure 5.

The upper end surface of the main platform of the landing platform is installed with a rotary locking mechanism, and the 3D model is shown in Figure 5. It mainly includes the rotation mechanism and the locking mechanism. According to the feedback parameters from the wind direction sensor, the rotation mechanism rotates and adjusts the nose direction of the UAV before take-off and after landing, so as to realise the rotational attitude adjustment function of the UAV; the locking mechanism acts to realise the function of locking and releasing the UAV.
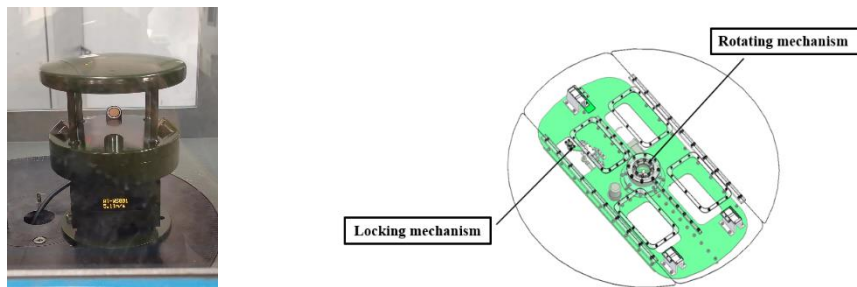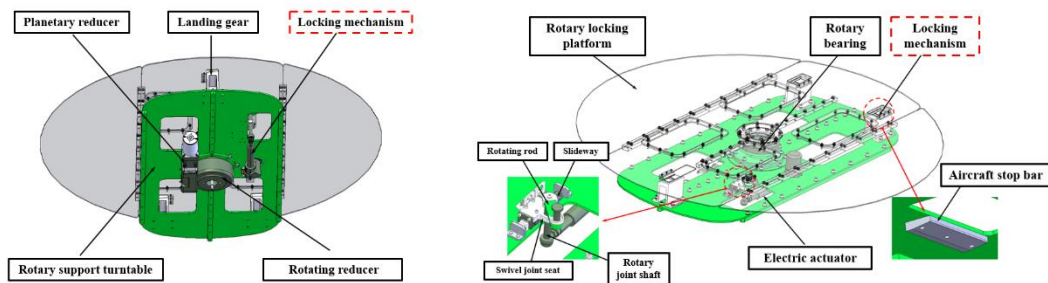

a)   wind speed and direction sensor   b)   Rotary locking mechanism

Fig.5 Schematic diagram of the wind speed and direction sensor and the rotary locking mechanism

The 3D model of the rotating mechanism is shown in Figure 6. The rotating mechanism is mainly driven by the motor to rotate the rotating reducer to rotate the rotating support turntable, the rotation stroke is 360°, which can meet the requirement of the UAV's nose direction swing before take-off. The rotary support turntable also serves as the mounting base of the locking mechanism, which drives the locking mechanism to rotate synchronously with the turntable. In addition, according to the position of the UAV's three-point landing gear, the upper part of the turntable is installed with three landing pads for placing the UAV's landing gear, and the landing pads cooperate with the landing gear to achieve the positioning of the UAV, and cooperate with the locking mechanism to achieve the locking of the UAV.

a)   Rotating mechanism   b)   Locking mechanism

Fig.6 Schematic diagram of the composition of the rotating mechanism and the locking mechanism

The 3D model of the locking mechanism is shown in Figure 6. The working principle is to convert the linear motion of the electric actuator into a micro-motion in the circumferential direction

of the rotary locking platform, thus achieving the locking of the UAV.

If the locking mechanism is not in operation, the landing pads will leak completely. When the UAV is manually placed on the landing pad, the locking mechanism works; the electric actuator moves, driving the rotating linkage to rotate the upper rotating locking platform in the circumferential micro-rotation, and presses the bottom of the UAV's landing gear 'barbs' (a type of part installed on the bottom of the UAV's landing gear, used for the UAV and the landing platform to achieve locking and recovery) on one side and pushes it down.) is pressed on one side to achieve positioning and locking of the UAV's landing gear pivot point. The locking mechanism and the underside of the landing gear are shown in Figure 7.
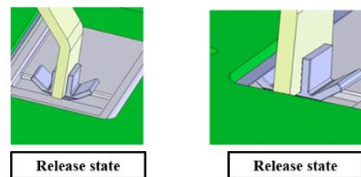

Fig.7 Schematic of the locking mechanism with the bottom of the landing gear

2.3 Recovery net assembly

The recovery net assembly is installed at the top of the main landing platform. When not in use, the recovery net assembly stores the recovery net, which has no effect on the UAV's rotation and launch. Upon receiving a command to land the UAV, the recovery net assembly extends the recovery net over the entire platform with a maximum area of 2500mm×2300mm (the size of the outer contour of the recovery net), which is used to complete the landing and recovery of the UAV.

The schematic diagram of the operating principle of the recovery net and the schematic diagram of the recovery state and deployment state of the recovery net is shown in Figure 8. When the recycling net assembly is in operation, the motor drives the reducer to move, and the movement of the reducer directly drives the rotation of the recycling net reel through the worm gear, on the one hand, to realise the action of unfolding and recycling the recycling net; on the other hand, the worm gear is coordinated with the worm wheel to drive the curved arm portion (the curved arm portion is internally mounted with a spring and can be unfolded automatically when there is no external force) to rotate along the centre of the curved arm mounting seat and to cooperate with the movement of the recycling net reel to realise the movement of the recycling net assembly. When the recycling net is unfolded, the crank arm can be self-locked by the worm gear to stop at any position.
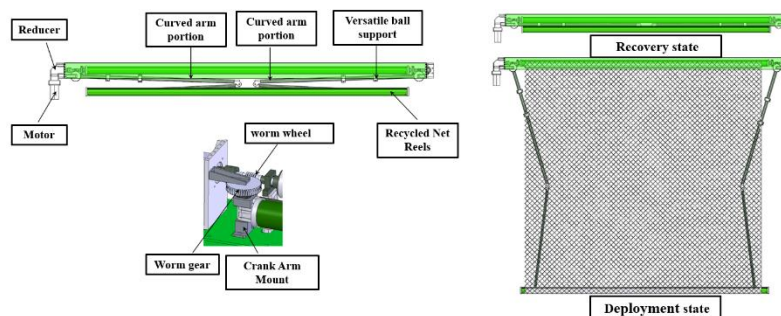
Fig.8 Schematic diagram of the composition and spreading of the recycling network components

3   Control strategy design

The movement of the landing platform device is operated by the vehicle operator in a single step, and the control process is mainly divided into the movement planning of the take-off process, the landing process and other processes. The control strategy of the take-off process and landing operation is shown in Figure 9 below.

When the take-off task is performed, the wind speed and direction sensor detects the wind speed and feeds back to the upper computer, and the operator judges whether to perform the take-off task according to the real-time wind speed value; if the take-off task is performed, the operator sends the command of "lift the platform" through the flight control system, and the lift component A and the lift component B are synchronously lifted under the drive of the lift power component. When the top slider reaches the upper limit of the proximity switch position, the lifting operation stops and is transmitted back to the host computer; the operator sends the "UAV Wind Direction Command", the wind speed and direction sensor detects the real-time wind direction and takes the average value of the wind speed and direction after 10 calculations and then transmits it back to the host computer, which then controls the rotation mechanism to complete the rotation of the UAV against the wind; after the wind direction is completed, the operator sends the "Wind Direction Command" to the host computer. Upon completion of the wind pairing, the operator gives the 'unlock aircraft' command, the electrical push rod of the locking mechanism moves to the position of the unlock proximity switch to complete the unlocking process and release the UAV; upon completion of the release of the UAV, the operator gives the 'reset' command and the launch and recovery platform returns to its initial state to complete the launch process. After the release of the UAV, the operator gives the 'reset' command and the recovery platform returns to its initial state to complete the launch process.

When performing the landing task, the operator sends the 'Raise Lift Platform' command through the flight control system, and when the slider on the top of the lift component reaches the upper limit of the proximity switch position, the lifting action stops and is transmitted back to the top unit; the operator sends the 'deploy recovery net' command, the recovery net deploys The operator sends the 'deploy recovery net' command, the recovery net component deploys the recovery net, the curved arm portion rotates to the limit proximity switch position on the recovery net, the recovery net deployment action is completed and transmitted back to the master computer; the operator lands the UAV on the recycling net and then sends the command 'Lifting Platform Return', the lifting component drives the UAV down to the position of the lower limit proximity switch when the action stops, and it is transmitted back to the upper computer and completes the landing process The operator lands the UAV on the recycling net and sends the command 'Return to Position'.
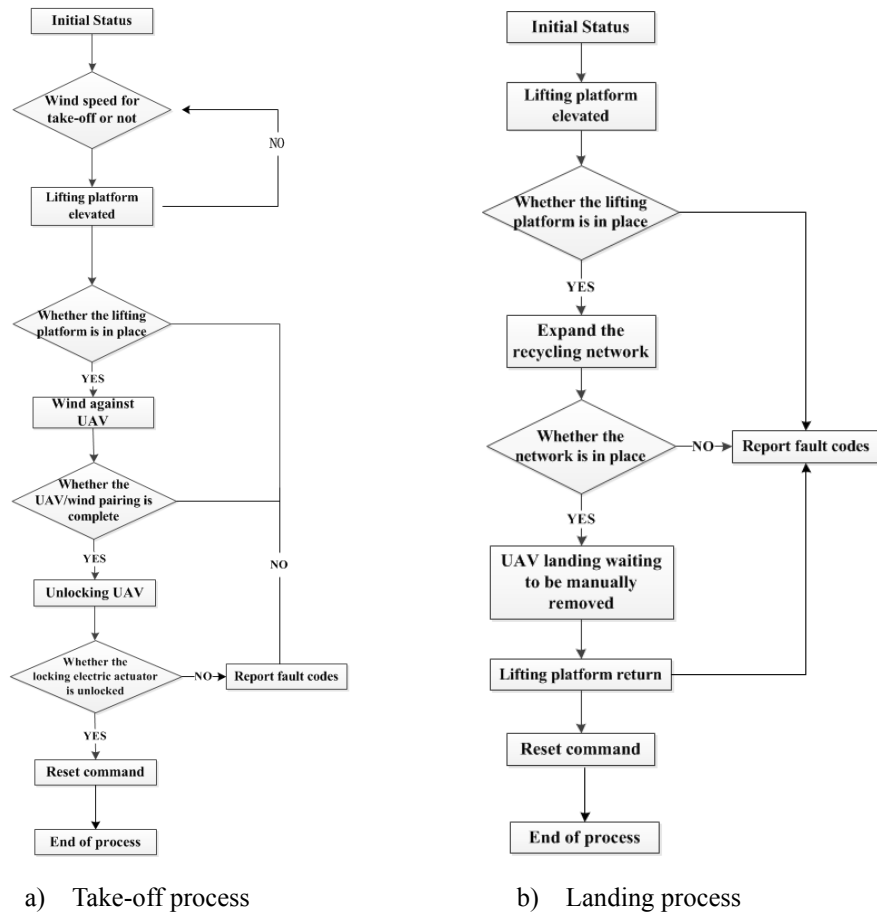
a) Take-off process   b) Landing process

Fig.9 Take-off and Landing process control strategy for landing platforms

4 Conclusion

This paper develops a vehicle-mounted take-off and landing platform for vertical composite wing UAV which mainly comprises a lift assembly, a rotary lock assembly and a recovery net assembly, and which can perform the functions of lifting and descending a UAV placed on the surface of the main platform, rotating the UAV into the wind and locking and releasing the UAV after detecting the wind speed and wind direction, and landing and recovering the UAV after completing the mission, which is of certain reference and significance in the application field of UAV vertical take-off and landing platform. The field of application of the vertical take-off and landing platform for UAVs has certain reference significance.

References:

[1] Liu K, Ye Fuchen. Development dynamics and trend analysis of vertical take-off and landing vehicles [J] Progress in Aeronautical Engineering, 2015, 6(2):127-138.

[2] Yu Jinyong, Wang Chao. Vertical take-off and landing UAV technology development status and outlook [J],Flying Missile, 2017(5):37-42.

[3] HE Xiaojiu, LI Yanbin, ZHU Feng, et al. Development status and design and manufacturing key technology of foreign vertical take-off and landing UAV [J]. Flying Missile, 2016(6):22-27.

[4] Wang Bo. Research on Hybrid Vertical Take-off and Landing Vehicle Technology [D]. Changsha: National University of Defence Science and Technology, 2016.

[5] Fang Y, Chen WH. Method of making a take-off and landing platform for a tethered UAV: CN209814318U [P].2019-12-20.

On The Navigation Problem of Unmanned Wheeled Agricultural Machinery Using Inertial, Satellites Information and Odometry

D. I. Smolianov

(Lomonosov Moscow State University, Moscow)

Abstract: The work describes the development, testing, and implementation of navigation algorithms for unmanned wheeled agricultural vehicles, vehicles when the navigation equipment consists a low-grade inertial navigation system, a receiver of global navigation satellite systems and odometry data — odometer-derived velocity and a steering sensor.

1 Introduction

The navigation problem is one of the most important in the operation of unmanned agricultural machinery. The positioning accuracy of the corresponding equipment needs to be a few centimetres for the treatment of most agricultural crops. High accuracy is also important to reduce fuel costs and ensure safe operation of agricultural machinery for humans. The reliability of the navigation solution depends on its stability to outliers and short-term losses of GNSS data.

Navigation equipment includes a low-grade inertial navigation system (INS) with microelectromechanical sensors (MEMS), a single antenna GNSS receiver (subsequently referred to as GNSS) and odometry — an odometer, and a steering sensor.

The paper considers two approaches to GNSS/INS/Odometry data fusion: loose and tight integrations. There are approaches to loose and tight integrations with odometer data [1, 2]. This work expands on these approaches by adding a new source of information – a steering sensor. We compare mentioned approaches.

Also, the work is devoted to a method of INS velocity fusion based on kinematic model properties of navigation objects – so-called component-wise ZUPT (Zero-Velocity Update Technology). Mentioned fusion does not require new sources of navigation information, but is based on the property of the movement of the ground vehicle. The relevant covariance analysis shows the observability of accelerometer biases in fusion proposed.

2 Kinematic models

The work examines the two most common configurations of agricultural vehicles: classic and articulated. Most tractors and combine harvesters have the classic configuration, i.e., a four-wheeled vehicle, which is steered by the front or the rear axle. Most tractors are steered by the front axle, while the rear axle steers most combines. The bicycle kinematic model [3, 4] is a sufficiently good approximation of classic vehicles without lateral slip. The articulated configuration implies two frames connected by a joint, each of which has two wheels fixed to the corresponding frame. In this configuration, one can make a turn by changing the angle between the frames. The articulated kinematic model [5, 6] is a sufficiently good approximation of articulated vehicles without lateral slip.

The kinematic equations of these models are used to construct the odometric dead reckoning in the loose integration with odometry and to construct the aiding measurements in the tight integration, which we discuss further.

3 Estimation problem

The traditional approach to solving this problem proposes the usage of geodetic reference frame as navigation frame in INS dead reckoning. We use GNSS data aiding if GNSS measurements are available. The fusion is implemented through feedbacks based on the estimates of the corresponding Kalman filter (KF). Methodically, the fusion comes down to solving the estimation problem (for illustration in continuous time)

$$\dot{y} = Ay + q, z = Hy + r, \tag{1}$$

where y is the state vector, A is the matrix that corresponds to the INS linear error equations used and to the model of the odometer and steering instrumental errors, q, r are vector-valued zero-mean white-noise processes, z is the measurement vector, generated by GNSS and odometer data, H is the corresponding observation matrix.

4 Loose INS/Odometry integration

The loose INS/Odometry integration implies two simultaneously working reckonings — inertial and odometric, as well as two integrated estimation problems. The inertial estimation problem was described in [1] with the state vector

$$y_1 = \left(\Delta x^T, \delta V^T, \alpha_1, \alpha_2, \beta_3, k^v, \varkappa_1, \varkappa_3, v_0^T, \Delta f_0^T \right)^T, \tag{2}$$

where Δx is INS position errors, δV is the vector of INS dynamic velocity errors, α_1, α_2 indicate the deflections of virtual horizon, β_3 is the azimuth attitude error, k^v is the odometer scale factor error, $\varkappa_1, \varkappa_3$ are the misalignment angles between the «measuring» axis of the odometer wheel and the INS instrument axes, v_0 is the vector of the angular rate sensors (ARS) drifts, Δf_0 is the biases vector of the accelerometers.

The state vector of the odometric estimation problem is

$$y_2 = \left(\Delta x_1{}^*, \Delta x_2{}^*, \Delta \psi, k^{v*}, k^\delta{}_0, k^\delta{}_1 \right)^T, \tag{3}$$

where $k^\delta{}_0, k^\delta{}_1$ are a null bias and a scale factor of steering sensor.

The two state vectors shown have common components, but the values of the estimates of these components may differ; the subscript (*) is used to demonstrate this fact.

The dynamic equations for y_1 are based on traditional INS error equations [7]. Dynamic equations for y_2 are based on linearised odometric reckoning error equations.

This approach can be called traditional, combining two traditional approaches to ground-based navigation — inertial and odometric. The advantage of this approach is the independent execution of two parallel estimation tasks, which increases the reliability in the event of failure of one of the components.

5 Tight INS/Odometry integration

The tight INS/Odometry integration implies one reckoning — inertial dead reckoning. In this case, the estimation problem is also single, but with an extended state vector

$$y = \left(\Delta x^T, \delta V^T, \alpha_1, \alpha_2, \beta_3, k^v, \varkappa_1, \varkappa_3, k^\delta{}_0, k^\delta{}_1, v_0{}^T, \Delta f_0{}^T\right)^T. \tag{4}$$

Odometry data is used as aiding measurements in this approach.

The advantage of this approach is that it eliminates the additional computational costs associated with the redundancy of estimation problems that arise from the loose integration approach.

6 Component-wise ZUPT

Here, component-wise ZUPT (Zero-Velocity Update Technology) means aiding by zero components of the heading point velocity vector. This method allows aiding even in the absence of odometer measurements, and, moreover, does not require stopping, unlike the classic ZUPT method. The heading point is a point whose velocity vector has direction along the longitudinal axis of the vehicle without any lateral slip. This property can be described by a non-holonomic constraint of the following form

$$\dot{x}_1 \cos\psi - \dot{x}_2 \sin\psi = 0, \tag{5}$$

where $x_{1,}x_2$ are the horizontal coordinates, ψ is the heading angle.

7 Conclusion

We present two approaches for integrating INS and odometry data: loose and tight integration. We consider two kinematic configurations: classic and articulated. We demonstrate that these two methods are equivalent in terms of their behaviour with respect to accelerometer biases and odometry coefficient estimates. Furthermore, we present the so-called component-wise ZUPT approach for INS aiding, which is based on kinematic models of the movement for ground wheeled vehicles. This approach does not require any additional navigation sensors. Based on covariance analysis, we demonstrate the observability of accelerometer biases in periodic curve trajectories. All the presented approaches have been implemented and tested on different models of agricultural machinery under real operating conditions.

References:

[1] Golovan A.A, Nikitin I.V. Combined use of strapdown inertial navigation systems and odometers from the standpoint of mechanics of inertial navigation systems. Part 1. Moscow University Mechanics Bulletin. Volume 70, pages 46–49. 2015

[2] Golovan A.A. INS/Odometer Integration: Positional Approach. Gyroscopy and Navigation. v. 12. #2, 2021. DOI: 10.1134/S2075108721020048

[3] Polack, Philip & Altché, Florent & Novel, Brigitte & de La Fortelle, Arnaud. The kinematic bicycle model: A consistent model for planning feasible trajectories for autonomous vehicles. 2017 IEEE Intelligent Vehicles Symposium (IV). PP. 812-818, 2017. doi:10.1109/IVS.2017.7995816.

[4] Min, Haigen & Wu, & Cheng, & Zhao. Kinematic and Dynamic Vehicle Model-Assisted Global Positioning Method for Autonomous Vehicles with Low-Cost GPS/Camera/In-Vehicle Sensors. Sensors 19. 5430., 2019. DOI:10.3390/s19245430.

[5] Nayl T. Modeling, Control and Path Planning for an Articulated Vehicle. Luleå tekniska universitet, Luleå, Sweden, 2013.

[6] Delrobaei M. Design and Steering Control of a Center-Articulated Mobile Robot Module. Hindawi Publishing Corporation. Journal of Robotics. Volume 2011, Article ID 621879. DOI:10.1155/2011/621879

[7] Vavilova N.B., Golovan A.A., Parusnikov N.A. Mathematical foundations of inertial navigation systems. Moscow University Publishing House. 2020

Identification of Spatiotemporal Clusters in Multi-Agent Systems Using an Endemic Model

V. Smetanina

(Saint Petersburg State University, Saint Petersburg)

Abstract: The paper presents an approach to cluster detection in distributed multi-agent systems using an endemic model. By modelling the "activation" of an agent as an infectious process, the study demonstrates that patterns of joint activation indicate spatial proximity, without requiring explicit information about the distance between agents.

1 Introduction

Modern systems are characterized by increasing complexity and dimensionality, which makes centralized control strategies resource-intensive in terms of computing and communication costs. This has led to an increase in interest in distributed multi-agent systems that are used in various fields such as unmanned vehicles, distributed sensor networks, group robots and aircraft.

The complexity of such systems creates serious problems for traditional control methods. These methods are often unable to effectively control each individual agent, and they also do not take into ac- count the properties of the system at a macro level [1]. By investigating how individual agents sponta- neously form consistent structures at meso-levels, we can obtain important information about how to control multidimensional systems [2], [3]. In addition, many of these systems also have self-organizing properties, which dynamically form clusters. In this regard, control methods need to adapt to these dy- namic structures. The paper [4] describes the formation of mesoscopic structures using the example of an aircraft with a large number of "feathers" distributed over the surface, i.e. elements with pressure sensors and rotary devices. It was shown that in conditions of turbulence, when the reaction of the system is nonlinear, the collective movement of the agents of the "feathers" can lead to their self-organization and a change in the flow regime of the aircraft body, thereby forming clusters of agents on the surface of the plane.

Cluster control strategies show better results compared to control approaches for macro and micro levels. Thus, in [5] an adaptive cluster control strategy with feedback was developed. Obtaining the clustering structure of the entire system was based on the compressed sensing method for a compact representation of the aggregated state of the agent, i.e. based on compressed measurements. Compared to micro-scale and macroscale approaches, cluster control outperformed these strategies in terms of convergence time, efficiency, and accuracy.

The problem of efficient cluster identification remains a matter of great urgency. The proposed work considers the formation of clusters based on periodic activation of agents and subsequent analysis of the dynamics of their spread using the endemic SIR (Susceptible, Infected Recovered) model. Just as the spread of infection in the SIR model demonstrates possible contacts in society

[6], it is assumed that groups of agents, often activated together, are in spatial proximity to each other. By accurately identify- ing clusters that represent cohesive groups of interacting agents, we gain a deeper understanding of the internal organization of the system and can develop more effective control strategies. Unlike static ap- proaches, the SIR model reflects the temporal evolution of agent interactions through the process of "activation", which allows us to identify clusters that arise, disintegrate and transform overtime.

The proposed work is devoted to the identification of clusters in a distributed multi-agent system through the periodic activation of individual agents.

2 Problem statement

Consider a system with N agents (in a set of N = {1,...,N}), which are placed into abounded region and are tasked to reach a certain goal point x^*. Each agent, possesses a state $x_i[t] \in R^d$, which may act as its location on a coordinate plane (therefore d = 2) or in a higher-dimensional coordinate space.

The aim is to analyse the spatiotemporal patterns of active agents overtime and draw a conclusion about the underlying spatial distribution. In this way, identify groups of agents that are often activated together, which will indicate their close proximity.

3 Conditions for solving the problem

To solve the problem, it is assumed that agents can inter- act in accordance with the communication graph. The interactions of agents are described through the adjacency matrix $A=$ $[a_{ij}]$, where $a_{ij} \neq 0$ means the presence of a connection between agents i and j, and $a_{ij} = 0$ means its absence. The dynamics of changes in the states of agents is described as follows: with discrete time intervals $t = 0, 1, 2, \ldots$ a subset of agents is randomly selected for "calling". Let $C_t \subseteq \{1, 2, \ldots, N\}$ be a set of agents called at time t. The selection of C_t is carried out by uniform sampling with a fixed probability. When agent i is called at time t (i.e. $i \in C_t$), its "nearest neighbors" become "active". The function $N(i, r)$, which returns a set of agents within a radius r of agent i, is described:

$$N(i,r) = \left\{ j \quad \in \{1,2,...,n\} : \| p_i - p_j \| \leq r \right\} \tag{1}$$

Where p is the coordinates of the location.

Activation propagation is described by the SIR dynamic model (simplified view):

$$\dot{G} = F(G) = \begin{cases} \dot{S} = S(A - S) - \beta_\gamma(t)IS \\ i = \beta_\gamma IS - \dfrac{rI}{a + I} \end{cases}, \tag{2}$$

where $\beta_\gamma = \beta_0(1 + \Phi(\omega t))$, $G(t) = (S(t), I(t))$, A is the tolerance coefficient for susceptible agents in the absence of activation, Y is the amplitude of periodic oscillations, $\omega > 0$ frequency, β_0 is the rate of activation transmission in the absence of periodicity, a is the delay coefficient during recovery.
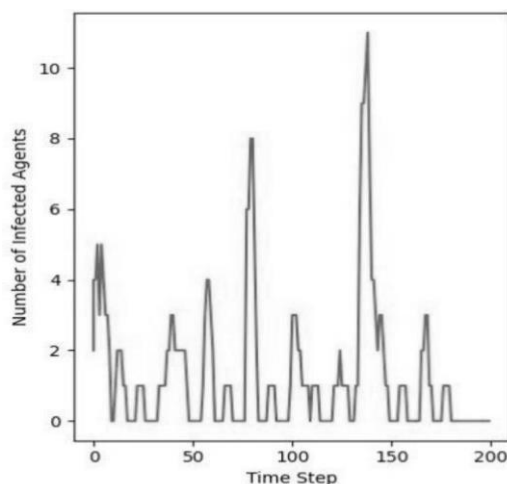
Fig.1 The number of "active" agents overtime

4 Conclusion

As a result of the research, a model- ling environment was developed to visualize emerging spatial structures. Analysis of the patterns of pathogen activation over time reveals distinct temporal spikes as shown in Fig. 1. These spikes correspond exactly to moments when the spread of infection in SIR models is simulated during the "activation" process, agents are geo- graphically grouped and isolated. Further research could be aimed at developing predictive models to predict activation patterns and system behavior which would allow more accurate identification of clusters and consequently better control.

References:

[1]   Li H, Li Y, Zhao H V. Modeling decision process in multi-agent systems: A graphical markov game based approach[C]//2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE, 2020: 197-204.

[2]   Amelin K, Granichin O, Leonova A, et al. A NEW METHOD OF ADAPTIVE MESOSCALE CONTROL IN COMPLEX MULTIAGENT NETWORKED DYNAMICAL SYSTEMS[J]. CYBERNETICS AND PHYSICS Учредители: Институт проблем машиноведения РАН, 2022, 11(4): 175-189.

[3]   Cristín J, Fernández-López P, Lloret-Cabot R, et al. Spatiotemporal organization of ant foraging from a complex systems perspective[J]. Scientific Reports, 2024, 14(1): 12801.

[4]   Granichin O, Khantuleva T, Amelina N. Adaptation of aircraft's wings elements in turbulent flows by local voting protocol[J]. IFAC-PapersOnLine, 2017, 50(1): 1904-1909.

[5]   Erofeeva V, Ershov V, Granichin O, et al. Adaptive Distributed Cluster Flow Control for a Group of Autonomous Robots[J]. IFAC-PapersOnLine, 2023, 56(2): 8690-8695.

[6]   de Carvalho J P S M, Rodrigues A A. Strange attractors in a dynamical system inspired by a seasonally forced SIR model[J]. Physica D: Nonlinear Phenomena, 2022, 434: 133268.

Others

Management of Industrial Distributed Systems Based on Open Information Technologies

V.V. Potekhin, E.V. Kuklin, P.K. Smirnov

(Scientific head professor, vice-rector for international activities, chief scientific officer of the laboratory "Intelligent Control Systems" Arseniev, Dmitry, Germanovich.)

Abstract: This paper discusses the Open Process Automation Standard (OPAS), which is designed for industrial automation systems. The distinction between Information Technology (IT) and Operational Technology (OT) is analyzed, as well as the main challenges currently facing the OT sector. Existing OPAS prototypes and testbeds are discussed, demonstrating the feasibility of open standards in industry. The paper emphasizes the importance of moving towards open architectures that foster collaborative development and innovation in automation.

1 Introduction

With the advent of the concept of open source software and open architectures, the business structures and ways of working of organizations are undergoing significant changes. While the information technology field has seen a rapid growth of open solutions, the operational technology (OT) field is still dominated by proprietary systems. Limitations due to closed interfaces and lack of compatibility between devices from different vendors are hampering the sector's growth. In this context, the Open Process Automation Forum (OPAF) is developing the OPAS standard aimed at creating open solutions for process automation.

2 Rehabilitation management in medical cyber-physical systems

Information technology (IT) encompasses the systems and tools used to process and manage data, including software, servers and networks. Operational Technology (OT), in turn, is concerned with the management of physical processes and systems such as production lines, power grids, and transportation systems. The main difference between the two is that IT is more data and information oriented, whereas OT focuses on physical interaction with the world around us. Despite the differences, IT and OT are beginning to overlap, especially in the context of the Internet of Things (IoT) and Industrial Internet of Things (IIoT) concepts, requiring organizations to rethink their strategies and approaches to systems development.

The major challenges facing the OT sector include:

1. Closed systems: Proprietary solutions limit the ability to integrate and collaborate across devices and platforms.

2. Long hardware lifespan: Legacy technologies can be difficult to upgrade, preventing the adoption of new solutions.

3. security: The increasing number of connected devices creates new vulnerabilities and security risks for production systems.

Moving to open standards in OT provides the following benefits:

- Increased interoperability: Open interfaces allow devices from different manufacturers to

work on the same system.

- Reduced costs: Using open solutions reduces dependence on specific vendors and reduces licensing costs.

- Accelerated adoption of new technologies: Open standards enable developers to adapt and deploy new solutions faster.

3 Conclusion

The OPAS standard represents an important step towards modernization and openness of operational technologies. The move to open architectures can address many of the current challenges of the OT sector, including integration, security, and innovation. Adopting such standards will allow organizations to not only improve their operational processes, but also create more competitive and adaptive business models.

References:

[1] Потехин В.В., Алексеев А.П., Куклин Е.В., Мисник А.Е., Хитрова А.Д., PROGRAMMING OF OPEN DISTRIBUTED INDUSTRIAL SYSTEMS BASED ON THE INTERNATIONAL STANDARD IEC 61499, COMPUTING, TELECOMMUNICATIONS AND CONTROL, 2024, pp. 12-13

[2] Потехин В.В., Алексеев А.П., Куклин Е.В., Хитрова А.Д., Кожубаев Ю.Н., CLOUD DISTRIBUTED CONTROL SYSTEM BASED ON OPEN PROCESS AUTOMATION PLATFORM, COMPUTING, TELECOMMUNICATIONS AND CONTROL, 2023, pp. 19-21

[3] Zoitl, T., Strasser, C., Sunder, Baier T. Is IEC 61499 in harmony with IEC 61131-3? IEEE Ind. Electron. Mag. 2009, Vol. 3, pp. 49–55.

[4] Church P., Mueller H., Ryan C., Gogouvitis S.V., Goscinski A., Tari Z. Migration of a SCADA system to IaaS clouds – a case study. Church et al. Journal of Cloud Computing: Advances, Systems and Applications. 2017, pp. 6-12.

[5] O-PAS Standard, Version 2.1: Part 1 – Technical Architecture Overview, pp. 6-10.

[6] O-PAS Standard, Version 2.1: Part 1 – Technical Architecture Overview, pp. 12-17.

Digitalization of Industrial Production with Tecnomatix Plant Simulation Software in Exeter Digital Enterprise Systems Laboratory (ExDES)

Mi Kerong

(Common Technology Research Department, Beijing Aerospace Times Laser Inertial Technology Company. Ltd. Beijing 100019, China;)

Abstract: University of Exeter is a world-class, public, top-tier research institution located in Devon, South West England. The Digital Enterprise Systems Laboratory (ExDES) at the University of Exeter's College of Engineering employs a number of cutting-edge technologies to realise the call of Industry 4.0. One of these advanced technologies is three-dimension simulation within the realm of digital twin. The simulation process is based on Siemens Tecnomatix Plant Simulation Software. This paper is based on "a new fast-drilling unit" in the Cyber-Physical lab (CP lab) of ExDES. This paper examines the application of simulation software in evaluating the viability of new investments in plant planning and business consulting.

Key words: Industry 4.0; digital twin; digital manufacturing; digital technology, Siemens Plant Simulation Software; qualitative and quantitative analyses; plant planning; economic benefits

1 The introduction of Tecnomatix Plant Simulation software (TPS software)

TPS software is a Siemens software for 3D simulation environments, simulation and analysis, experiment management, optimization tools, data integration and reporting, collaboration and sharing. This software enables the simulation of production line operations, the optimisation of material flows, the analysis of resource utilisation, the evaluation and optimisation of energy analysis in the production process, the automatic detection of bottlenecks in the production line, the development of production strategies and the production of Gantt chart, which helps to visualise the production plan and schedule. In the following article, TPS software is used in a business analysis for a digital enterprise group project at ENGM027 Engineering Management Science module at the University of Exeter, UK.

1.1 Case study

Tecnomatix Plant Simulation Software is used in a mobile phone case production workshop in terms of plant planning/business consultancy/risk assessment to assess whether a £1million investment is worthwhile. Through this new fast drilling unit, the original drilling time can be reduced from 15 seconds by 5 seconds to 10 seconds now, thus improving the productivity and profit. Given the detailed understanding of the production costs, selling prices, the various steps involved, and the time allocated for each step, the analysis can now proceed. By building a 3D model, running the simulation software, collecting analyzing data, and calculating the return of investment (ROI) or payback period and bottleneck, it can be determined whether the investment in

the plant is worthwhile or not.

The production process of a mobile phone case is as follows:

Overview of the Manufacturing Process

Currently, the mobile phone cases are produced using the Cyber-Physical Lab with the plant layout similar to the one described in Figure 1. The Cyber-Physical Lab has nine stations, namely:

1. Case top cover supply station
2. Transportation bridge (Branch 3)
3. Robotino
4. Transportation bridge (Branch 2)
5. Measuring station
6. Case back cover supply station
7. Case press station
8. Drilling unit
9. Final product sorting station

Fig. 6 Overview of the Manufacturing Process (University of Exeter, 2021)
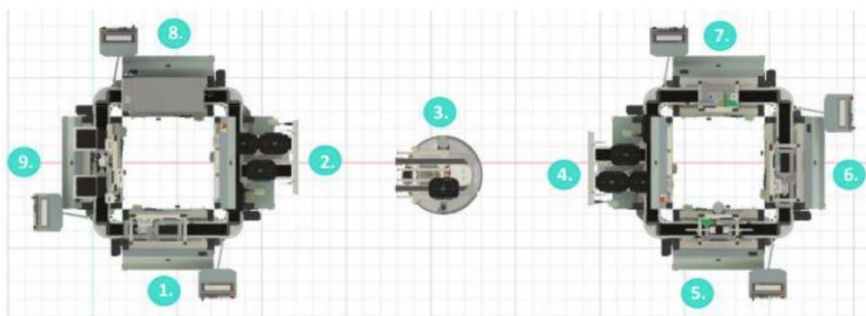
Fig. 7 The Cyber-Physical Lab station layout (University of Exeter, 2021)

Time taken to produce each part of the phone case in the CP lab is as follows:

Tab. 3 Process times for the stations in the Cyber-Physical Lab

Station	Time (seconds)
Case top cover supply station (Magazine Front Station)	5
Measuring Station	10
Case back cover supply station (Magazine Back Station)	5
Transportation bridge (Branch 2)	5
Transportation between islands	35
Transportation bridge (Branch 3)	5
Case Press Station	10
Drilling Station	15
Final product sorting station (Workpiece Output Station)	5

Tab. 2 The function corresponding to each station is as follows

Station	Functions
Case top cover supply station (Magazine Front Station)	To distribute the top cover of the phone cases according to the order information
Measuring Station	analogue differential measurement of workpieces at different measuring points
Case back cover supply station (Magazine Back Station)	To distribute the back cover of the phone cases according to the order information
Transportation bridge (Branch 2)	To transport workpiece carrier to the next working station
Transportation between islands (Robotino)	To transport carriers between two islands equipped with a laser scanner
Transportation bridge (Branch 3)	To transport workpiece carrier to the next working station
Case Press Station	To press the workpiece
Drilling Station	Drilling holes in the workpiece
Final product sorting station (Workpiece Output Station)	To output a cubic workpiece on two roller slides. The left side is dedicated to good quality products The right side is for faulty products
The carrier and the pallet	To carry the front cover and the back cover of the phone case

1.1.1 The solution process is as follows:

The question is to carry out an analysis of whether this value of £100 million is worth it from the point of view of commerciality. All we need to do is carry out the actions below:

- Payback Period: Calculate the payback period. If the cost of £100 million can be recouped in a short period of time and generate a large profit, then the investment is worthwhile.
- Bottleneck analysis: Use Tecnomatix Plant Simulation Software to simulate the production process and then analyse whether the bottleneck is the drilling unit. If the drilling unit does not constitute the production bottleneck, then reducing its operational time will not lead to increased overall output and efficiency, as it does not reside on the critical path. Activities on the critical path determine the minimum completion time of the project. Therefore, the economic efficiency will not be improved either.

## 2 The specific modelling process is described below:

2.1 Original Model

First of all, according to the process and time requirements of the question above, construct an original model of the original mobile phone case production line process.
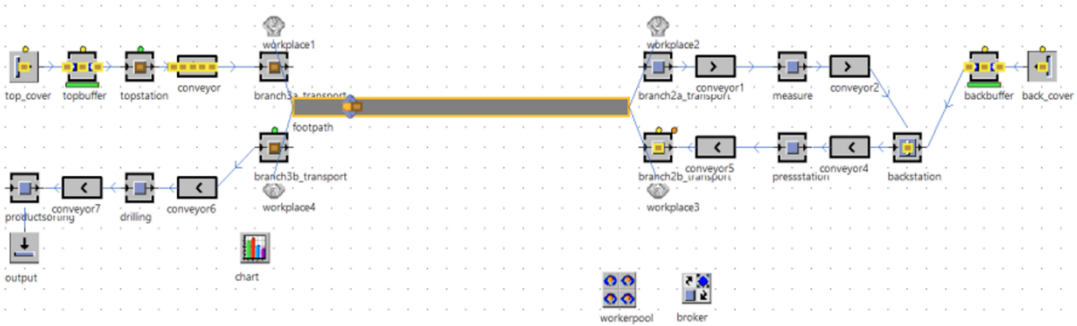
2.2 Improved Model

On the basis of the original model, only the time of the drilling unit is shortened, and the improved model is constructed.

Calculate the payback period and bottle neck by running the above two models.

Here is my drawing based on the requirements shown above of the flow chart of the original model:
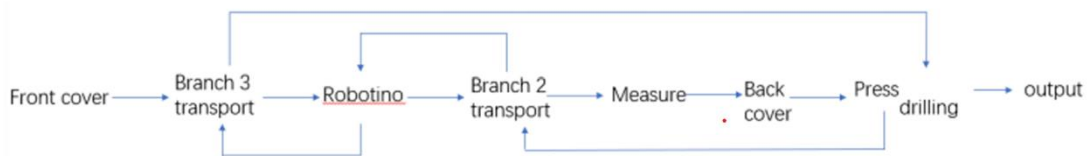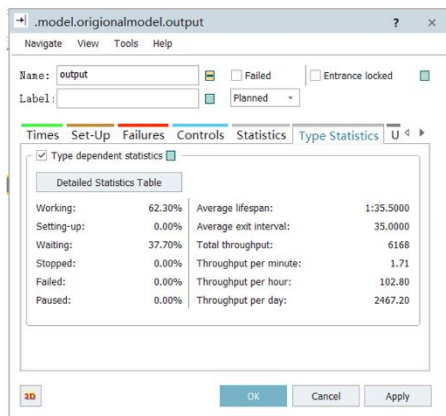

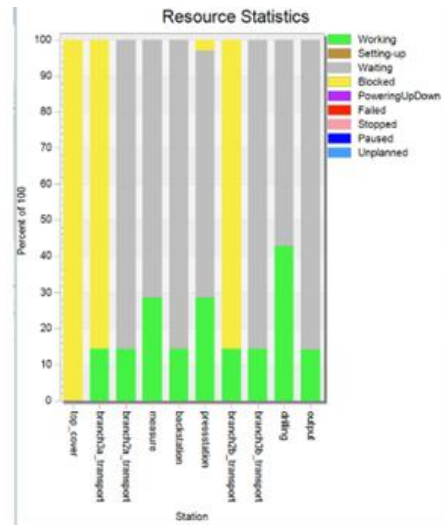**Fig. 8** The flowchart of the original model

Fig. 9 The simulation of the original model in 2D version

By running the original model, it can be seen that the bottle neck is not the drilling time in the production process but the robotino.

Fig. 10 The hourly output of the original model after running for 60 hours

Fig. 11 The resource statistics of the original model

In Figure 6, the yellow bars indicate a blocked status. The longer the yellow bar, the lower the efficiency, suggesting the bottleneck lies in preceding steps. Although the top cover has the longest yellow bar, indicating significant blocked time, it depends on the upstream supply from the case top cover supply station and thus is not a bottleneck. Instead, significant blocks are observed in Figure 4 at branch 2b transport and branch 3a transport, connected by the Robotino conveyor. Conversely, Figure 6 shows the drilling unit's working time in green is quite long, indicating high efficiency. Therefore, whether analyzing from the perspective of bottlenecks or working time, the drilling unit is not the bottleneck in this smartphone case production line, making new investment in it unwarranted.
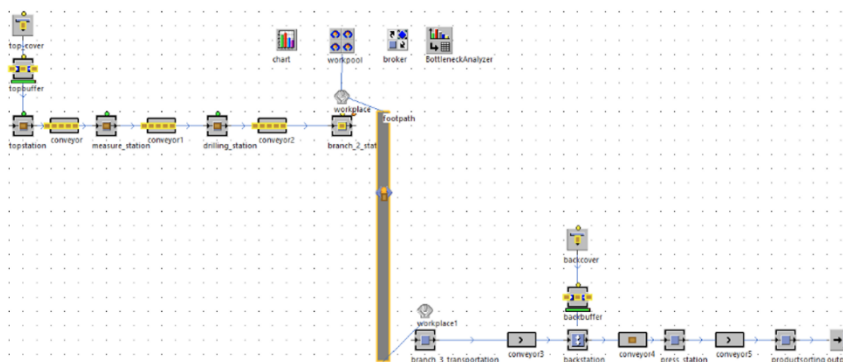
Fig. 12 The simulation of the improved model in 2D version (based on CP lab, drilling time=10s)
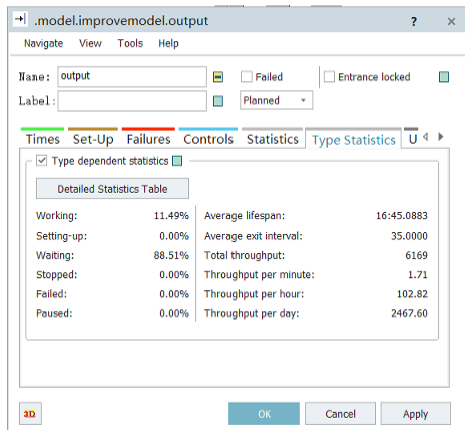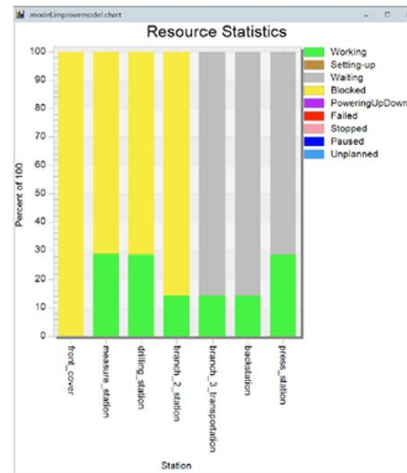
Fig.8

Fig. 9

Fig. 13 The hourly output of the improved model is 102.82 after running for 60 hours

Fig.9 The resource statistics of the improved model after running for 60 hours

Tab. 4 The payback period of the investment of the improved model after applied the "New fast drilling unit"

Different models	Different hourly output	Time taken to recover costs (weeks)
the initial model	102.80	12742.0999
The improved model	102.82	

The calculation process goes as the follows:

Increased hourly output=The improved model hourly output-the initial model hourly output =102. 82-102.80=0.02 cases/hour

The profits of each phone case=the sales prices/each phone- the cost price/each phone=£6.54

Weekly increased profits=the profits/each phone case* Increased hourly output*working hours/ week= 6.54*0.02*60=£7.848

Return of the investment (weeks)= the amount of the new investment/weekly increased profits= £100000/£7.848=12,742.0999(weeks)

2.3 Conclusion

This investment is not worthwhile because the payback period for this $100 million investment is too long and it does not solve the problem of the bottleneck in the production of mobile phone cases.

3 To conclude

3D simulation software provides a new approach to factory planning and factory business consulting. This simulation software realises the method of interconnection between the virtual world of product development and production planning with the physical world of production

system and product performance. The feasibility and profitability of investments are analysed from a quantitative and qualitative point of view. There are other cases where Tecnomatix Plant Simulation Software has been used in real production:

- Automotive Production Line: This study simulates various aspects of an automotive production line, such as sheet metal work, welding, and assembly, with the aim of optimizing production processes and enhancing overall efficiency.
- Component production: This study simulates the production processes of small components, including automotive parts and electronic components, to ensure production continuity and timely delivery.
- Logistics and Warehousing: Simulate logistics and warehousing systems to optimise inventory management, transport routes and loading/unloading processes to reduce costs and improve operational efficiency.
- Production Strategy Optimisation: Find the best production solution by simulating different production strategies, such as adjusting production speed and deploying human resources.
- Energy analysis: Use energy analysis tools to assess and optimise energy consumption in the production process to reduce production costs and environmental impact.

References:

[1] University of Exeter, Exeter Digital Enterprise Systems Laboratory (ExDES) , 2021

CHINESE–RUSSIAN
"NAVIGATION AND MOTION CONTROL"
YOUTH FORUM