IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 18, 2020, accepted March 2, 2020, date of publication March 5, 2020, date of current version March 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2978521

A Novel Algorithm for the Fault Diagnosis of a Redundant Inertial Measurement Unit

XIAOQIANG HU, XIAOLI ZHANG, XIAFU PENG, AND DELONG YANG
Department of Automation, School of Aeronautics and Astronautics, Xiamen University, Xiamen 361005, China

Corresponding author: Xiaoli Zhang (zhxl@xmu.edu.cn)

This work was supported in part by the "Inertial Technology" Key Laboratory Foundation of National Defense Science and Technology of China under Grant 61425060201162506007, and in part by the Industry-University Cooperation Project (IUCP) of Fujian Province under Grant 2017H6101, and in part by the Fundamental Research Funds for the Central Universities under Grant 20720190129.

ABSTRACT Although a number of fault diagnosis algorithms for inertial sensors have been proposed in previous decades, the performance of these algorithms needs to be improved with regard to small faults. In this paper, we introduce a data driven-based algorithm, namely, SaPD, for the anomaly detection and output reconstruction of a redundant inertial measurement unit (RIMU). SaPD implements the fault identification of an inertial apparatus by combining an artificial neural network with the Q contribution plots method in parity space. To improve the performance of the fault detection part, in particular for small faults, we introduce a novel hyperplane that measures the distances between inputs and the primary-neuron set obtained from a self-organizing incremental neural network (SOINN). We also employ the Q contribution plots of sensors in the fault isolation part by analyzing historical data with principal component analysis (PCA). We perform quantitative evaluations in a realistic simulation environment, which demonstrates that the proposed SaPD algorithm outperforms other related algorithms in terms of the fault identification accuracy of tiny faults with an acceptable computational complexity.

INDEX TERMS Inertial navigation, anomaly detection, fault diagnosis, artificial neural networks, principal component analysis.

I. INTRODUCTION

The fundamental tasks of fault detection and isolation (FDI) include the effective detection of faults and the accurate isolation of these faults from healthy components in the shortest time possible. A good diagnostic for navigation systems improves robustness and increases the system availability [1]. Recently, many works have been performed regarding the fault diagnosis task for integrated navigation systems under the assumption that inertial navigation system (INS), as a common reference, is infallible [2], [3]. Thus, research on the FDI of INS is of great significance.

The redundant inertial measurement unit (RIMU) provides a hardware foundation for fault tolerance design in INS [4]. The structures can be divided into orthogonal configurations and nonorthogonal configurations. In addition, the latter scheme has greater advantages than the former in terms of improving reliability and reducing costs [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa Rahimi Azghadi.

There have been multiple prior methods exploring the fault diagnosis of the RIMU. These methods were basically classified into two categories. The first category involved conventional statistical approaches such as the generalized likelihood ratio (GLR) method [6], [9], optimal vector test (OPT) method [7], [10] and singular value decomposition (SVD) method [8], [11]. The second category investigated pattern recognition approaches on the basis of parity analysis. The approaches in the latter class were developed for either the realization of fault diagnosis with minimal redundancy or the improvement of algorithm performance with sufficient redundancy.

Regarding minimal redundancy, the reliability of a four-gyro scheme is 1.75 times that of a nonredundant system without a significant cost increase [12]. However, a gyro-quadruplet can't isolate faults via statistical approaches with only one redundancy. The neural network observer [13], support vector machine (SVM) [14], [15], principal component analysis (PCA) [16] and linear prediction [5] methods were thus used to present diagnosis algorithms auxiliary to the parity analysis.

Hybrid algorithms combining expert knowledge, physical models, or data-driven methods have become one of the main developing methods used in fault diagnosis and health management [17], [18]. In the context of algorithm performance improvement, there have been numerous works introducing intelligent methods into the fault diagnosis of the RIMU. To accelerate the diagnosis of a gradual fault, a residual quality of the parity equation, and the fuzzy evaluation of this quality were presented [19]. An SVM multiclass fault classifier was applied in a nine-gyro unit to identify the abnormal false alarm in the OPT method [20]. An algorithm based on intersection fault-tolerant fusion and an unscented Kalman filter (UKF) was proposed for a configuration of 5 gyros [21], which realized fast detection of faults. The Q contribution plots of gyros were used when a fault was detected in parity space, extending the application of PCA to navigation systems [22]. To handle the influence of high-frequency noise and singular points, a low-pass filter was then added to the PCA-based algorithm, which increased the robustness of the diagnosis algorithm [23]. With the high-frequency noise separated by discrete wavelet transform (DWT), the isolation probability of the extended parity space approach (EPSA) significantly increased [24].

The main drawback to two-step fault diagnosis algorithms, i.e., fault detection first and isolation later, was the need for a fault detection (FD) algorithm to guarantee the reliability of the isolation part. The performance of the whole diagnosis algorithm was limited by the detection part. In addition, the hybrid methods above have achieved good results in tests for faults of large amplitude. However, when it comes to small faults, the performance of these methods regarding fault identification rate and computational complexity needs to be improved. Therefore, further studies on the fault diagnosis algorithm of RIMU, especially on a sensitive FD method for two-step algorithm, are still necessary.

In this paper, we study the FDI algorithm for a RIMU (gyroscopes or accelerometer) with a regular dodecahedron configuration. We propose a novel "SOINN (self-organizing incremental neural network) and PCA-based fault diagnosis algorithm", i.e., SaPD, to enhance the small fault detection and isolation abilities. The neuron training block first analyzes the topological structure of historical data in parity space via fast-SOINN processing. The primary neuron set is then sent to the FD part to compose a hyperplane that is irregular and rugged for anomaly monitoring. We employ the method in [22] to form the fault isolation (FI) part because of the outstanding effectiveness of this method. When a fault is detected, the Q contribution plots of the sensors are estimated by the PCA model. These diagnosis results are then used to rebuild the outputs of the RIMU measurements. Simulation results show that the proposed SaPD algorithm significantly outperforms other methods [6], [7], [20], [22] in terms of the identification accuracy for very small faults in an acceptable computation time.

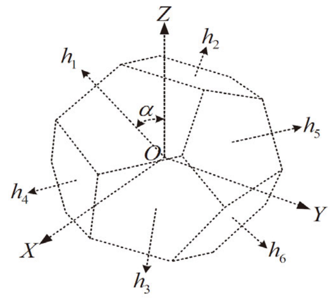

FIGURE 1. Regular dodecahedron configuration of the RIMU.

Our contributions are multifold:

(1) We design a novel fault detection algorithm to enhance the performance of two-step fault diagnosis method based on the topological structure of a historical dataset;

(2) We employ Q-contribution plots and the SOINN-based algorithm to form an outstanding hybrid fault diagnosis algorithm;

(3) We provide a thorough quantitative comparison between the proposed SaPD and other benchmark methods for reference.

II. CONFIGURATION STRUCTURE

The geometric aspect plays an important role in the RIMU design. In this work, a six-gyros structure with dodecahedron configuration, which is an optimization of reliability and cost compared with other redundancies, is used as shown in Fig. 1; the observation matrix related to this scheme is expressed by (1), and the measurement equation is expressed by (2):

$$\boldsymbol{H} = \begin{bmatrix} 0.5257 & 0 & 0.8507 \\ -0.5257 & 0 & 0.8507 \\ 0.8507 & 0.5257 & 0 \\ 0.8507 & -0.5257 & 0 \\ 0 & 0.8507 & 0.5257 \\ 0 & 0.8507 & -0.5257 \end{bmatrix} \tag{1}$$

$$\boldsymbol{M} = \boldsymbol{H}\boldsymbol{\omega} + \boldsymbol{\varepsilon} + \boldsymbol{f} \tag{2}$$

where $\boldsymbol{M}$ is the sensor measurement vector (6×1); $\boldsymbol{H}$ is the observation matrix (6×3) related to the sensor axis system; $\boldsymbol{\omega}$ is the state vector (3×1) in the tri-orthogonal axis system ($\boldsymbol{\omega} = [\omega_x \omega_y \omega_z]^T$); $\boldsymbol{\varepsilon}$ is the Gaussian measurement noise vector (6×1); and $\boldsymbol{f}$ is the fault vector (6×1).

$\boldsymbol{H}$ satisfies $\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H} = 2\boldsymbol{I}_3$, where $\boldsymbol{I}_3$ is the identity matrix (3×3), which meets the optimal navigation criterion of the RIMU [25], i.e., the measurement noise $\boldsymbol{\varepsilon}$ has a minimal impact on the estimated state vector $\hat{\boldsymbol{\omega}}$.

The parity vector $\boldsymbol{R}$ associated with this configuration is given by

$$\boldsymbol{R} = \boldsymbol{V}(\boldsymbol{H}\boldsymbol{\omega}) + \boldsymbol{V}(\boldsymbol{\varepsilon} + \boldsymbol{f}) \tag{3}$$

where the decoupling matrix $\boldsymbol{V}$ is obtained from the null space of $\boldsymbol{H}$, so $\boldsymbol{V}\boldsymbol{H} = 0$. Potter added an orthogonal constraint to the decoupling matrix, that is, $\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{I}_3$. Thus, the decoupling matrix obtained by the Potter algorithm [9] and used in this paper is shown as follows:

$$V = \begin{bmatrix} 0.7071 & -0.3163 & -0.3162 & -0.3162 & -0.3162 & 0.3162 \\ 0 & 0.6324 & 0.1954 & 0.1954 & -0.5117 & 0.5117 \\ 0 & 0 & 0.6015 & -0.6015 & -0.3717 & -0.3717 \end{bmatrix} \quad (4)$$

The residual vector $\boldsymbol{R}$ has the optimal characteristics of fault diagnosis, i.e., the fault detection performance of every sensor is the same because $\boldsymbol{H}$ satisfies the optimal navigation criterion and $\boldsymbol{V}$ is obtained via the Potter algorithm [26].

III. FUNDAMENTAL THEORY

A. SELF-ORGANIZING INCREMENTAL NEURAL NETWORK SCHEME

SOINN is a neural network model with a two-layer structure for incremental learning [27]. This model adaptively approximates the density distribution and realizes the topological representation of data by a group of neurons. However, choosing the right time for turning the first level of learning to the second level is very difficult. Thus, some improvement of SOINN only retain the first layer of the original SOINN in structure, making the whole algorithm more suitable for real-time incremental learning [28].

SOINN consists of edge set $N \subseteq C \times C$ and neuron set $C = \{c_1, c_2, \cdots, c_k\}$. Neuron c_i has three attributes: the codebook vector (also named the neuron distribution) $\boldsymbol{W}_{c_i}$, which is the coordinate position of the neuron; the neuron density M_{c_i}, which is the number of wins in competition for the data representation; and the similarity threshold T_{c_i}, which is the effective range of representation. Edge $n\{c_i, c_j\}$ represents the connections between neurons c_i and c_j. In addition, the parameter age$\{c_i, c_j\}$ indicates the activity of the neuronal connection [29].

SOINN defines two operations, namely, the within-class insertion and between-class insertion, for dynamic node adjustment. The within-class insertion is used to reduce the quantization error of the data, and is in some cases omitted to simplify the training process. The between-class insertion, as a key to realizing incremental learning, is used to adapt SOINN to unknown input data without affecting the previous learning results. For a new input sample, the learning algorithm selects the first two closest neurons that satisfy the insertion conditions expressed by (5) and then measures the similarity between them.

$$\begin{cases} s_1 = \arg\min\limits_{c_i \in C} ||\boldsymbol{\xi} - \boldsymbol{W}_{c_i}|| \\ s_2 = \arg\min\limits_{c_i \in C \backslash s_1} ||\boldsymbol{\xi} - \boldsymbol{W}_{c_i}|| \end{cases} \quad (5)$$

where $\boldsymbol{\xi}$ is the new input vector, s_1 and s_2 are the first two closest neurons.

If the points meet the condition

$$d(\boldsymbol{W}_{s_1}, \boldsymbol{\xi}) > T_{s_1} \quad \text{or } d(\boldsymbol{W}_{s_2}, \boldsymbol{\xi}) > T_{s_2}. \quad (6)$$

where $\boldsymbol{W}_{s_1}, \boldsymbol{W}_{s_2}$ are the codebook vectors of s_1 and s_2, respectively; T_{s_1}, T_{s_2} are the similarity thresholds of s_1 and s_2, respectively; $d(\cdot)$ is the similarity measure function (usually the Euclidean distance), SOINN generates an neuron node to represent the possible new pattern. The similarity threshold of the new neuron can be calculated as follows:

$$T_{c_r} = \left\| \boldsymbol{W}_{c_r} - \boldsymbol{W}_{s_1} \right\|_2 \quad (7)$$

where c_r is the new neuron; $\boldsymbol{W}_{c_r} = \boldsymbol{\xi}$ is the codebook vector of c_r; T_{c_r} is the similarity threshold of c_r.

If the input $\boldsymbol{\xi}$ does not meet condition (6), then the neuron distribution of the winning nodes s_1, s_2 should be adjusted. Neuron distribution adjustment is essentially a vector quantization (VQ) process. The algorithm converges to the local optimal solution by the gradient descent method, shown as follows:

$$\boldsymbol{W}_{s_i}^{(+)} \leftarrow \boldsymbol{W}_{s_i}^{(-)} + \varepsilon_{s_i}(\boldsymbol{\xi} - \boldsymbol{W}_{s_i}^{(-)}), \quad (i = 1, 2) \quad (8)$$

where ε_{s_i} is learning rate of s_i, given by SOINN. To achieve convergence with the learning algorithm, the learning rates should meet the following conditions:

$$\sum_{t=1}^{\infty} \varepsilon_{s_i}(t) = \infty, \sum_{t=1}^{\infty} \varepsilon_{s_i}^2(t) < \infty, \quad (i = 1, 2) \quad (9)$$

SOINN connects s_1 and s_2 with an edge $n\{s_1, s_2\}$ or resets the parameter age$\{\cdot\}$ to 1 if the edge already exists. Each edge records the input of the data through age$\{\cdot\}$, and the edge is deleted if age$\{\cdot\}$ exceeds the threshold agemax. Under the condition that the neurons are dense enough, this method, based on competitive Hebbian learning rules (CHL), can perfectly build a topological connection between neurons [30].

In practical applications, noise or outliers often exist in data, which causes SOINN to generate unnecessary nodes for these data. To alleviate this problem, SOINN performs network denoising after a learning cycle λ_d. The neuron nodes that meet the following condition are found and deleted:

$$M_{c_i} < c_d \bar{M}_C \;\&\; n_{c_i} \leq 1 \quad (10)$$

where c_d is the denoising ratio, $\bar{M}_C$ is the average density of neurons in SOINN, and n_{c_i} is the number of neighbor nodes of neuron c_i.

B. PRINCIPAL COMPONENT ANALYSIS THEORY

Principal component analysis (PCA) is a method based on the decomposition of eigenvalues with a covariance matrix or mathematically with the aid of singular value decomposition (SVD). PCA is commonly used in multivariate statistical analysis, reducing the dimensionality of the dataset space [31].

Let $\boldsymbol{x}$ be an $N \times 1$ vector, and $\boldsymbol{X}$ be an $N \times L$ matrix composed by an L time series of $\boldsymbol{x}$ ($L > N$). To obtain an effective direction of variation, the matrix $\boldsymbol{X}$ is normalized to matrix $\bar{\boldsymbol{X}}$ using its mean and standard deviation, that is,

$$\bar{\boldsymbol{X}}_i = \frac{\boldsymbol{X}_i - \boldsymbol{E}_X}{\boldsymbol{S}_X}, \quad (i = 1, 2, \cdots, L) \tag{11}$$

where $\boldsymbol{X}_i$ is a column of $\boldsymbol{X}$, $\bar{\boldsymbol{X}}_i$ is the normalization of $\boldsymbol{X}_i$, $\boldsymbol{E}_X$ is the mean vector ($N \times 1$), $\boldsymbol{S}_X$ is the standard deviation vector ($N \times 1$). Then, the covariance matrix $\sum_X$ can be calculated as $\sum_X = \bar{X}\bar{X}^T$.

The eigenvalues λ_i and corresponding eigen-vectors $\boldsymbol{p}_i (i = 1, 2, \cdots, N)$ of matrix $\boldsymbol{X}$ are obtained by solving the following equation set:

$$\begin{cases} |\lambda \boldsymbol{I} - \sum_X| = 0 \\ (\lambda \boldsymbol{I} - \sum_X)\boldsymbol{p} = 0 \end{cases} \tag{12}$$

where $\boldsymbol{p}_i$ satisfies:

$$\boldsymbol{p}_i^{\mathrm{T}} \boldsymbol{p}_j = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}, \quad (i, j = 1, 2, \cdots, N) \tag{13}$$

By arranging the eigenvalues in decreasing order, we can select the first n_{m} eigenvalues as principal components considering a properly criterion. Besides, the corresponding eigenvectors form the loading matrix $\boldsymbol{P}_{\mathrm{m}} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_{n_{\mathrm{m}}}]$ of principal components. Then, when a new sample vector $\boldsymbol{x}$ is input, it can be decomposed as follows:

$$\boldsymbol{x} = \hat{\boldsymbol{x}} + \tilde{\boldsymbol{x}} \tag{14}$$

where $\hat{\boldsymbol{x}} = \boldsymbol{P}_{\mathrm{m}} \boldsymbol{V}_{\mathrm{m}(\boldsymbol{x})}$ is the principal component subspace projection, $\boldsymbol{V}_{\mathrm{m}(\boldsymbol{x})} = \boldsymbol{P}_{\mathrm{m}}^{\mathrm{T}} \boldsymbol{x}$ is the PC score vector, and $\tilde{\boldsymbol{x}} = (\boldsymbol{I} - \boldsymbol{P}_{\mathrm{m}} \boldsymbol{P}_{\mathrm{m}}^{\mathrm{T}})\boldsymbol{x}$ is the error subspace projection [32].

According to the above principal component model, statistics T^2 and statistics Q (also as known as the squared prediction error, SPE) can be calculated and compared with their thresholds T_α^2 and Q_α (α is the confidence level of the distribution) [33]. Once $T^2 > T_\alpha^2$ or $Q > Q_\alpha$, a fault is detected in the vector $\boldsymbol{x}$. Then, for the isolation of the detected fault, the contribution plots C^Q of each element of $\boldsymbol{x}$ are calculated as follows:

$$C_i^Q = ||x_i - \hat{x}_i||_2^2 \quad (i = 1, 2, \cdots, N) \tag{15}$$

In addition, the state component with maximum contribution is identified as a fault item and separated from the system.

IV. PROPOSED METHOD

A. FAULT DETECTION ALGORITHM

Historical data $\boldsymbol{M}_{\mathrm{T}}$ is transformed into a parity vector set $\boldsymbol{R}_{\mathrm{T}}$ via the decoupling matrix $\boldsymbol{V}$. The topological structure $\{C, N\}$ of historical data, trained by the SOINN algorithm, is used to represent the normal system. To reduce the computational complexity of the fault detection algorithm, we select some of nodes that meet (16) from the neuron set C and use them as the primary neuron set C'.

$$n_{c_i} > \mathrm{rate_n} \cdot E(n_c) \tag{16}$$

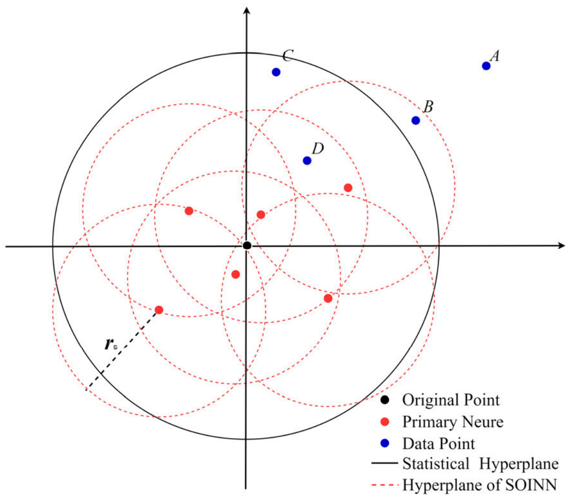

FIGURE 2. Principle of anomaly detection algorithm.

where $E(n_c)$ is the average connection number and $\mathrm{rate_n}$ is the scaling parameter.

Based on the work above, the principle of anomaly monitoring algorithm based on SOINN is shown in Fig. 2 with a 2D example. The training algorithm based on SOINN generates a primary neuron set in the parity space. With each primary neuron as the center and a fixed r_{S} as the radius, a series of separating hyperplanes can be obtained. The inner circle of the hyperplane is an area of normal data, while the area outside this circle is unreliable. Then, the normal data area $A_{C'}$ determined by the whole algorithm is as follows:

$$A_{C'} = A_1 \cup A_2 \cup \cdots \cup A_{n'_C} \tag{17}$$

where n'_{C} is the number of primary neurons, and A_i is the area surrounded by the hyperplane of node c'_i. Therefore, the classification surface of the SOINN-based fault detection algorithm is the envelope surface of $A_{C'}$. Any parity vector of the RIMU located outside this surface is judged as faulty.

When a sensor measurement vector $\boldsymbol{M}_{\mathrm{L}}$ is sent to the fault diagnosis block, this vector is first changed into a parity vector $\boldsymbol{\xi}$ via the decoupling matrix $\boldsymbol{V}$ in (4), as shown at the bottom of the previous page. Then, the minimum distance $r_{\boldsymbol{\xi}}$ between $\boldsymbol{\xi}$ and the primary neuron set C' is calculated as follows:

$$r_{\xi} = \min_{c_i \in C'} \left\| \boldsymbol{\xi} - \boldsymbol{W}_{c_i} \right\|_2 \tag{18}$$

The determination of the operation status, i.e., $\mathrm{FD}(\boldsymbol{M}_{\mathrm{L}})$, is made by comparing the distance r_{ξ} with the threshold as follows:

$$\mathrm{FD}(\boldsymbol{M}_{\mathrm{L}}) = \begin{cases} \text{fault}, & r_\xi > r_{\mathrm{S}} \\ \text{normal}, & r_\xi \leq r_{\mathrm{S}} \end{cases} \tag{19}$$

where r_{S} is the threshold.

In this paper, we use a scaling of the maximum radius of class C' as the fault detection threshold, that is

$$r_{\mathrm{s}} = \mathrm{rate_r} \cdot r_{\mathrm{b}} \tag{20}$$

$$r_b = \max_{c_i \in C'} \left\| \boldsymbol{W}_{c_i} - \boldsymbol{b} \right\|_2 \tag{21}$$

where r_b is the maximum radius of the primary neuron set, $\mathrm{rate_r}$ is the scaling factor, and $\boldsymbol{b}$ is the barycenter of set C' obtained by the k-means algorithm.

Compared with the classification surface of the general likelihood ratio or other statistical methods, the separating hyperplane made by SOINN is irregular and rugged. Considering the following points with different faults, that is, the points A, B, C and D, we discuss the similarities and differences in fault detection between the two classification surfaces, as shown in Fig. 2. The datapoint (point A) of a large fault is far from the origin of the coordinate space, which makes it easy to determine that the fault exists. The datapoint (point D) of a small fault is highly similar to the normal data, which leads to either our algorithm or the statistical method missing detection of the fault. The main difference between these two kinds of classification surfaces is the diagnosis result for a fault that is not too obvious or too small. As shown at point B and point C, the statistical classification surface (the black solid line) judges point B as faulty and point C as normal because it determines the type of datapoints only by the distance from the origin. While the SOINN-based surface (the red dashed line) obtains the opposite results, that is, point B is normal and point C is faulty, because it also considers the distribution of normal data in parity space. these two methods provide different judgments for operation states.

The SOINN-based algorithm trains the classification surface through a large amount of historical data. Compared with a statistical classification surface that is fixed and has the same thresholds in all directions, our fault detection surface is more in line with the characteristics of the system under normal working conditions. Consequently, our surface can provide better performance for fault detection.

B. FAULT DIAGNOSIS AND RECONSTRUCTION

The Q contribution plots show excellent performance in fault isolation. Therefore, when an anomaly is detected by the SOINN-based algorithm, we use PCA method to separate the fault components.

Using the decoupling matrix (4) obtained by the Potter algorithm, the data of the six-gyro module is changed into a parity space ($\mathbb{R}^3$). the first two eigenvalues are chosen as the primary components (the cumulative variance of the primary components is approximately 67%), and the corresponding eigenvectors form the loading matrix $\boldsymbol{P}_\mathrm{m}$.

To eliminate the impact of the data magnitude, the parity vector $\boldsymbol{\xi}$ is normalized by the mean and the standard deviation of the training data, that is,

$$\bar{\boldsymbol{\xi}} = \frac{\boldsymbol{\xi} - \boldsymbol{E}_{\boldsymbol{R}_\mathrm{T}}}{\boldsymbol{S}_{\boldsymbol{R}_\mathrm{T}}} \tag{22}$$

where $\bar{\boldsymbol{\xi}}$ is the normalize vector of $\boldsymbol{\xi}$, $\boldsymbol{E}_{\boldsymbol{R}_\mathrm{T}}$ and $\boldsymbol{S}_{\boldsymbol{R}_\mathrm{T}}$ are the mean vector (3×1) and standard deviation vector (3×1) of the training data $\boldsymbol{R}_\mathrm{T}$, respectively.

The Q contribution plots of the parity vector are calculated by (15) and then converted to the gyros using the least squares estimation method [22]:

$$\boldsymbol{C}_\mathrm{G}^{Q} = (\boldsymbol{V}^\mathrm{T}\boldsymbol{V})^{-1}\boldsymbol{V}^\mathrm{T}\boldsymbol{C}^{Q} \tag{23}$$

where $\boldsymbol{C}_\mathrm{G}^{Q}$ is the Q contribution vector (6×1), representing the contribution of the gyros to the Q statistic. Then, the isolation result of the fault, i.e., $\mathrm{FI}(\boldsymbol{M}_\mathrm{L})$, is made as follows:

$$\mathrm{FI}(\boldsymbol{M}_\mathrm{L}) = \arg_i \max(\boldsymbol{C}_{\mathrm{G},i}^{\mathrm{Q}}) \quad (i = 1, 2, \cdots, 6) \tag{24}$$

The fault information matrix $(6 \times 6)\boldsymbol{W}$ is used to record the system state. When the system is normal, the matrix $\boldsymbol{W}$ is the unit matrix. When a fault is detected and isolated by the SOINN and PCA-based algorithm, the corresponding row of the matrix $\boldsymbol{W}$ is set to all zeros. The reconstruction of the system output $\hat{\boldsymbol{\omega}}$ via the weighted least squares method is as follows:

$$\hat{\boldsymbol{\omega}} = (\boldsymbol{H}^\mathrm{T}\boldsymbol{W}\boldsymbol{H})^{-1}\boldsymbol{H}^\mathrm{T}\boldsymbol{W}\boldsymbol{M} \tag{25}$$

A diagram of the SaPD algorithm is shown in Fig. 3.

V. EXPERIMENTAL RESULTS AND COMPARISONS

A. SIMULATION DESCRIPTION

To verify the effectiveness of the SaPD algorithm, a digital simulation platform is established based on a high-precision strapdown inertial navigation system (SINS) toolbox [34]. The carrier is assumed to be a fixed-wing aircraft, and a 150 s trajectory simulation of the entire navigation procedure is designed as a dynamic trace. The simulated trajectory curve is shown in Fig. 4. Based on the trajectory above, an inverse concept of the SINS algorithm was used to generate the measurements of the tri-orthogonal axis system, as shown in Fig. 5.

The angle increments are further transformed into the six gyros data through the configuration matrix $\boldsymbol{H}$. Then, during the training process and every testing process, a 0.5 °/h$^{1/2}$ random walk σ, which means a 1.4544×10^{-5} rad standard deviation of noise acting on the angle increment data under a 100 Hz frequency, is rejoined to each gyro to obtain the training dataset and the testing datasets.

A fault diagnosis algorithm should have the highest possible accuracy of state recognition, while meeting real-time requirements. To validate the proposed method, we first assess the parameter sensitivity and dynamic fault detection of SaPD, and then compare it with other benchmark methods under the same false alarm rate.

Referring to other related works, we use the following four indexes to assess and compare the performance of the fault diagnosis algorithms: the false alarm rate (AR), which is the probability that a normal sample is judged as faulty; the fault detection rate (DR), which is the probability that a faulty sample is judged as faulty; the fault isolation rate (IR), which is the probability that a faulty sensor is correctly isolated when it is detected; the fault identification accuracy (FIA), which is the probability that a fault is accurately located. The

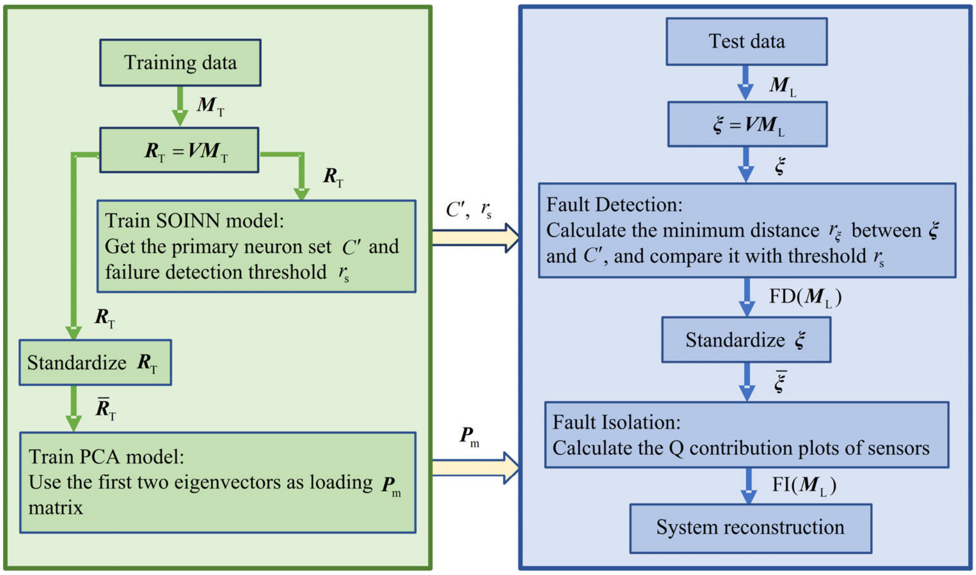

FIGURE 3. Diagram of the SOINN and PCA-based fault diagnosis algorithm (SaPD).

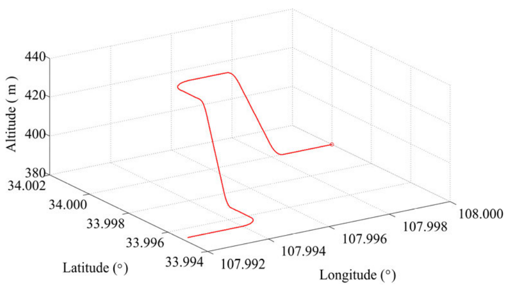

FIGURE 4. Trajectory curve simulation.

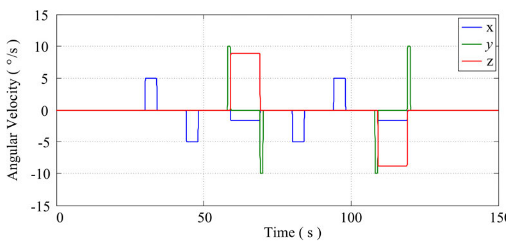

FIGURE 5. The angular velocity curves of three coordinate axes in the body frame.

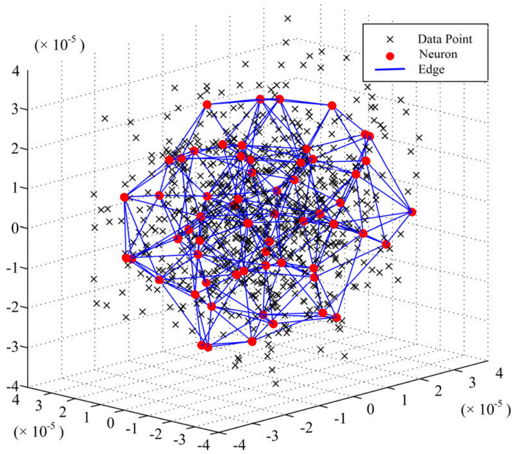

FIGURE 6. Topological structure of regular dodecahedron.

first three indicators are obtained through the statistics of the experimental results, while the FIA is the product of the DR and IR.

B. ACCURACY ASSESSMENT

1) SENSITIVITY ANALYSIS

Fast-SOINN is taken in this paper for the historical data training. The parameters are set as follows: the age threshold $age\max = 50$, the learning cycle $\lambda_d = 100$, and the denoising ratio $c_d = 0.5$. The topology of the RIMU in parity space, trained by the SOINN algorithm, is shown in Fig. 6. In addition, the distribution of the primary neuron set is obtained with $rate_n = 1.2$.

Controlled experiments are established to conduct a sensitivity analysis. Different step faults, that is, 3σ to 6σ, are added on gyro-3 after 50 s, respectively. The influences of the threshold on the AR and FIA are studied by changing the scale parameter $rate_r$ in a specified range, as given in Fig. 7 and Fig. 8. The AR value decreases from 10.07% to 0.19%, and the FIA values from 3σ to 6σ decrease from 58.08% to 9.65%, 79.68% to 25.24%, 92.82% to 50.30% and 98.43% to 75.84%, respectively. Therefore, the AR and FIA are sensitive to the threshold, i.e., the scale parameter $rate_r$.

In addition, we set $rate_r$ in the range of [1.55,1.60] for an approximately 1% false alarm rate, which is used behind.

2) DYNAMIC FAULT DETECTION

A fault of tiny state is always dynamic. To verify the dynamic fault detection capability of the proposed algorithm, we add

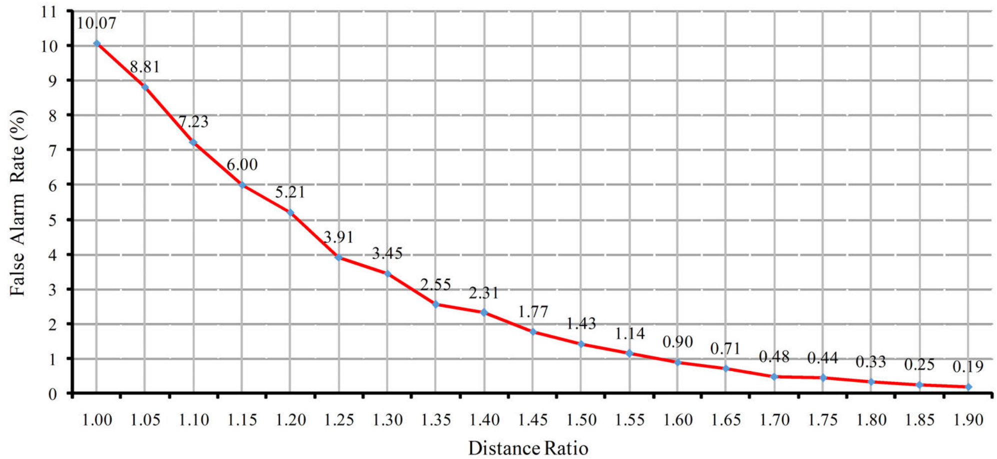

FIGURE 7. False alarm rates of SaPD with different distance ratios.

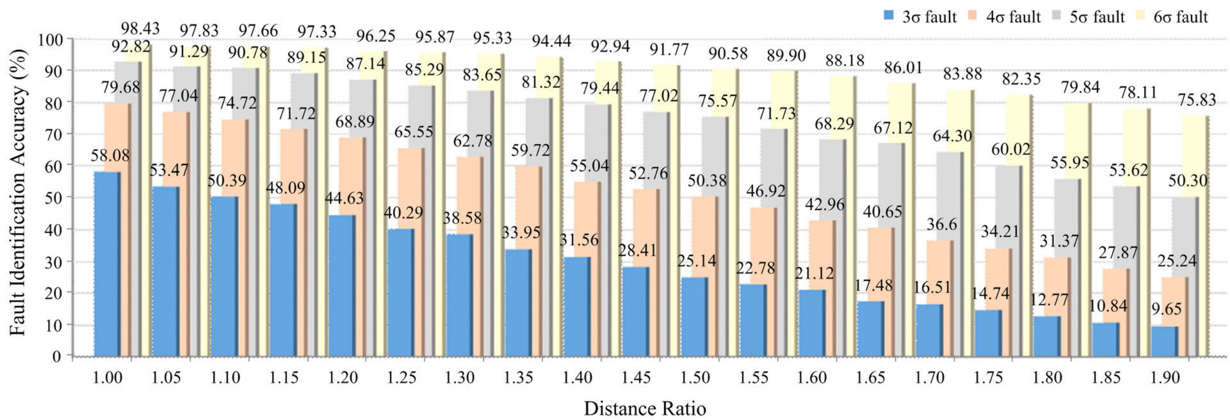

FIGURE 8. False identification rates of SaPD with different distance ratios.

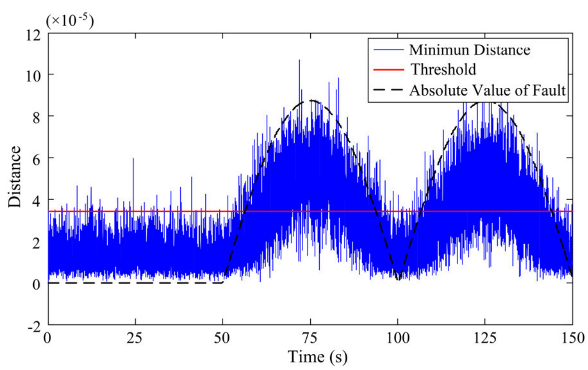

FIGURE 9. Detection result of SaPD for dynamic fault.

a sinusoidal fault to the third gyro after 50 s. The amplitude of the fault is 6σ and the period is 100 s. The diagnosis result is shown in Fig. 9. It is obviously that SaPD can recognize the change of this kind of tiny fault when the fault reaches a certain size.

C. METHOD COMPARISONS

The simulation trajectory takes 150 s, with 15000 samples. A grouping experimental study with different step faults, that is, 1σ to 6σ, is conducted, in which the faults are added on gyro-3 after 50 s. To show both the strengths and weakness of our algorithm, we evaluate the performance of SaPD and compare it with those of other methods. The details of these methods and their parameters, such as the threshold P and confidence level α, can be seen in Table 1.

There are two differences between our experiment and paper [20] that should be declared. 1) The penalty function C_S and RBF (radial basis kernel function) kernel parameters λ_S have little impact on SVM training in our experiment, which can be realized by setting $C_S = 1$ and $\lambda_S = 1$. Therefore, there is no relevant parameter optimization in the simulation. 2) The faults in the training data are very important to SVM training. In our experiment, $f_T = 5\sigma$ is the minimum fault that can make the training converge within the maximum number of iterations, while the training fault in paper [20] is $f_T = 2\sigma$.

TABLE 1. Comparative algorithms with corresponding arguments.

No	Algorithm	Parameters
1	GLR [6]	P_G=11.344 (Critical value of $\chi^2_{0.99}(3)$)
2	OPT [7]	P_O=6.635 (Critical value of $\chi^2_{0.99}(1)$)
3		P_O=9.635
4		P_O=11.344
5	OPT+SVM [20]	C_S=1 , λ_S=1 , $f_T=7\sigma$
6		C_S=1 , λ_S=1 , $f_T=8\sigma$
7		C_S=1 , λ_S=1 , $f_T=8.5\sigma$
8		C_S=1 , λ_S=1 , $f_T=8.8\sigma$
9		C_S=1 , λ_S=1 , $f_T=9\sigma$
10	GLR+PCA [22]	n_m=2 , α_{T^2}=99% , α_Q=99.9%
11		n_m=2 , α_{T^2}=99.5% , α_Q=99.5%
12		n_m=2 , α_{T^2}=99.9% , α_Q=99%
13	SaPD	$rate_r$=1.55
14		$rate_r$=1.60

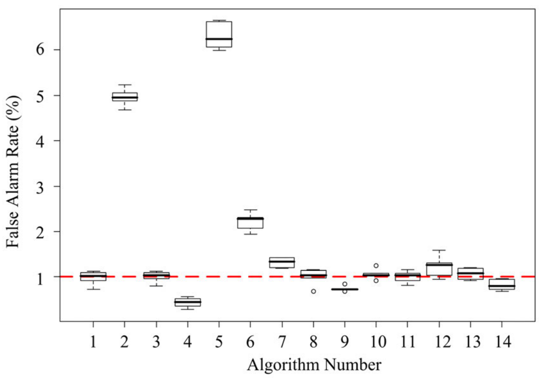

FIGURE 10. Boxplot of false alarm rate.

1) FALSE ALARM RATE ANALYSIS

To express the experimental results concisely, the false alarm rate of each situation is counted, and the results are shown in Fig. 10, in the form of a boxplot.

The false alarm rate decreases with parameter changes in the algorithms: the threshold increases for OPT (as in algorithms 2, 3 and 4), the training faults decrease for OPT + SVM (as in algorithms 5 through 9), and $rate_r$ decreases for SaPD (as in algorithm 13 and 14).

2) FAULT IDENTIFICATION ACCURACY ANALYSIS

For a fault diagnosis algorithm, an AR decrease is always accompanied by a decrease in DR and vice versa. To compare the performances of these methods, we select parameter settings with approximately 1% AR and compare the three other indexes. The DR and IR of the selected algorithms under different faults are shown in Table 2.

To evaluate the algorithm more intuitively, the FIA is obtained by multiplying DR and IR, and the results are shown in Fig. 11. It can be concluded that SaPD and OPT + SVM are obviously superior to the other algorithms. The two algorithms perform similarly for large faults. However, SaPD performs the best for small faults. The specific analysis is as follows.

1) The FIA is improved with increasing fault size for each algorithm;

2) The algorithms can be divided into three echelons in decreasing order: OPT + SVM and SaPD, GLR and OPT, GLR + PCA. In addition, the FIA of the former is better than that of the latter except for two specific cases: the GLR + PCA algorithm performs better than the second echelon for the first two faults 1σ and 2σ; and this algorithm is also better than the OPT + SVM algorithm for fault 1σ;

3) The FIA differences between the echelons decrease with increasing fault. For example, The FIA of the SaPD algorithm is 3.67 times, 2.08 times, 1.53 times, 1.23 times, 1.10 times and 1.05 times of those of OPT; The FIA of the SaPD algorithm is 1.40 times, 1.51 times, 1.64 times, 1.43 times, 1.26 times and 1.15 times of those of GLT + PCA (comparing A13 with A11);

4) The dominance relation of the FIA in the same echelon changes with the faults. For example, the FIA of SaPD is almost that of OPT + SVM for faults 4σ-6σ, while SaPD gains an advantage for faults 1σ-3σ, that is, by 1.92% (1.92 times), 2.16% (1.33 times) and 2.37% (1.11 times), respectively.

3) COMPUTATION COMPLEXITY ANALYSIS

Now, we analyze the computational complexity of these methods. Algorithms 10 and 11 both adopt GLR + PCA, but with different confidence sets, so we only consider algorithm 11 here. The simulation experiment is carried out on a consumer laptop with a 1.60 GHz Intel(R) Core (TM) i5-8265U CPU and 8.00 GB RAM. The running times of the fault diagnosis programs on the whole flight trajectory are shown in Fig. 12. The results indicate that SaPD matches the real time demand of RIMU fault diagnosis, while OPT + SVM

with similar FIA cannot. The specific analysis is as follows.

1) The methods can be divided into three echelons according to the computational complexity: OPT + SVM, SAPD and GLR + PCA, GLR and OPT, ordered from highest to lowest;

2) The computational complexity in different echelons varies greatly, and only the latter two echelons meet the real-time requirements at a 100 Hz output frequency. The running time of OPT + SVM ranges from 161.52 s to 205.23 s, i.e., 10.77 to 13.68 ms for every sample; SaPD in the second echelon ranges from 7.15 s to 48.75 s in total, i.e., 0.48 ms to 3.25 ms for every sample; and the average times for OPT and GLR in the last echelon are approximately 0.23 s and 0.56 s, which means that the diagnostic periods are 0.02 ms and 0.04 ms, respectively.

3) The diagnostic time for OPT + SVM decreases with increasing fault size. OPT + SVM consists of six SVM fault classifiers in series. A sample is sent to the next classifier if it is determined to be normal, and this process continues until the sample is deemed faulty or all the classifiers make

TABLE 2. Fault detection rate and isolation rate of the selected algorithms.

No	Method	1 σ		2 σ		3 σ		4 σ		5 σ		6 σ	
		DR	IR	DR	IR	DR	IR	DR	IR	DR	IR	DR	IR
1	GLR	2.27	41.85	7.12	60.53	19.28	72.46	40.61	84.39	67.24	91.63	87.36	95.59
3	OPT	2.08	36.06	6.62	63.14	19.77	76.73	43.16	86.98	69.82	92.72	88.39	96.10
8	OPT+SVM	2.18	65.60	7.41	88.39	21.95	94.26	47.79	96.97	73.97	98.07	91.19	98.78
10	GLR+PCA	1.70	100.00	4.00	100.00	11.28	100	27.30	100.00	52.17	100.00	77.10	100.00
11	GLR+PCA	1.97	100.00	5.75	100.00	14.08	100	32.28	100.00	56.37	100.00	77.90	100.00
13	SaPD	2.75	100.00	8.71	100.00	23.06	100	46.19	100.00	71.04	100.00	89.28	100.00

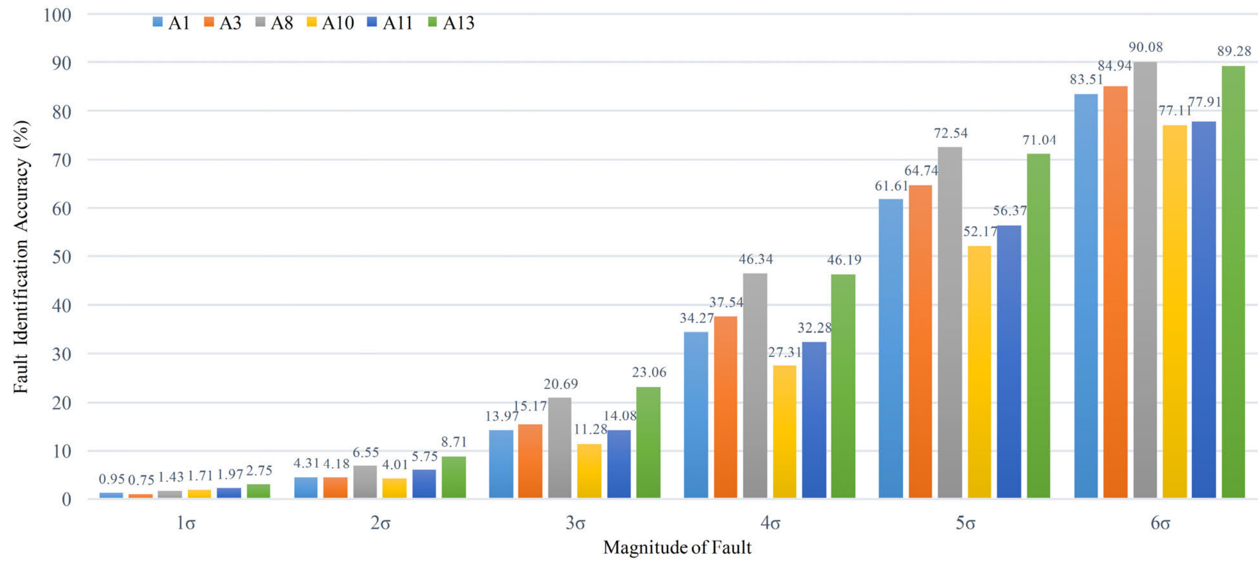

FIGURE 11. Fault identification accuracy of the selected algorithms.

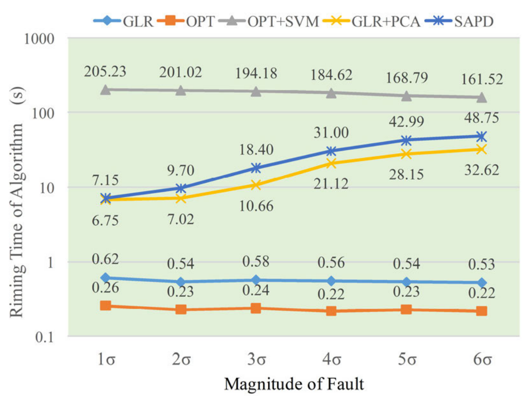

FIGURE 12. Running time of the selected algorithms.

a decision. Thus, for a small fault, most of the samples are determined by six classifiers in turn because a small fault can easily be missed in detection, and the diagnosis processing of a large fault can end without the sample going through all six classifiers.

4) The diagnosis times of the SaPD and GLR + PCA algorithms increase with increasing fault size. These two algorithms call the fault isolation program only when a fault is detected. Therefore, the diagnostic processing of a small fault often skips the isolation component because of missed detection. Furthermore, SaPD needs to calculate multiple distances for threshold comparison, while GLR + PCA only calculates T2 statistics and Q statistics, which makes SaPD take more time.

VI. CONCLUSION

In this paper, we presented a real-time FDI algorithm, namely, SaPD, and applied this algorithm to an RIMU with six-gyro redundancy. The fault detection method, based on parity space topology obtained from the SOINN algorithm, was developed regarding the fault detection part, providing improved results in terms of the correct detection of small faults (approximately 2.37% to 11.78% growth of FIA for a 3 σ fault). Q contribution plots based on PCA were used in the fault isolation part to obtain reliable isolation results, thus avoiding missed separation from the normal sensor when a fault was detected by the SOINN-based algorithm. The simulation results demonstrated that the proposed SaPD algorithm can recognize the change of a dynamic fault when the fault reaches a certain size and outperform other methods with respect to the fault identification accuracy for tiny faults while meeting real-time requirements.

REFERENCES

[1] S. Li, J. Mei, Q. Qu, and K. Sun, "Research on SINS/GPS/CNS fault-tolerant integrated navigation system with air data system assistance," in *Proc. IEEE Chin. Guid., Navigat. Control Conf. (CGNCC)*, Nanjing, China, Aug. 2016, pp. 2366–2370.

[2] Y. Wu, X. Wang, P. Cao, and Z. Yang, "A novel fault detection algorithm for inertial navigation system," *Trans. Beijing Ins. Technol.*, vol. 35, no. 5, pp. 494–499, May 2015.

[3] X. Li, Y. Sun, H. Mo, and C. Yao, "Redundancy configuration of 9-gyro based on MIMU," *J. Harbin Ins. Technol.*, vol. 41, no. 5, pp. 90–94, May 2009.

[4] F. Geng, X. Zhu, and Z. Zhou, "Research on an effective integrated navigation failure-tolerance filtering technology," *J. North-western Polytech. Univ.*, vol. 34, no. 3, pp. 449–455, Jun. 2016.

[5] J. Cheng, X. Sun, H. Mu, and X. Fang, "Multi-fault detection and isolation for redundancy strapdown inertial navigation system," *J. Harbin Engr. Univ.*, vol. 39, no. 2, pp. 358–364, Feb. 2018.

[6] K. C. Daly, E. Gai, and J. V. Harrison, "Generalized likelihood test for FDI in redundant sensor configurations," *J. Guid. Control*, vol. 2, no. 1, pp. 9–17, Jan. 1979.

[7] H. Jin and H. Yue Zhang, "Optimal parity vector sensitive to designated sensor fault," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 4, pp. 1122–1128, Oct. 1999.

[8] D. S. Shim and C. K. Yang, "Geometric FDI based on SVD for redundant inertial sensor systems," in *Proc. 5th ASCC*, Melbourne, Australia, Jul. 2004, pp. 1094–1100.

[9] T. Zhang, W. Fu, Z. Ren, and J. Yan, "New decoupling matrix method for generalized likelihood fault detection and isolation of redundant IMU," *Jour. Solid Rocket Technol.*, vol. 40, no. 4, pp. 532–536, Apr. 2017.

[10] B. Li and Z. Zhang, "Optimal parity vector technology applied to fault detection of redundant inertial measurement unit," *Aerosp. Control*, vol. 34, no. 2, pp. 86–90, Apr. 2016.

[11] Z. Ren, W. Fu, and T. Zhang, "New SVD method in FDI of redundant IMU," *Chin. J. Sci. Instrum.*, vol. 37, no. 2, pp. 412–419, Feb. 2016.

[12] W. Wu, Z. Ren, and T. Zhang, "Improved FDI Method for a Gyro-quadruplet," *Command Control Simul.*, vol. 37, no. 1, pp. 128–131, Feb. 2015.

[13] C. Wang and X. Li, "Fault diagnosis method for gyro in double-redundant strapdown control system," *Aeronaut. Manuf. Technol.*, no. 10, pp. 88–91, Oct. 2010.

[14] J. Jiang, K. Tang, Y. Zhang, and D. Dai, "Fault prediction for gyroscopes based on wavelet transform and SVR," *Farm Mach. Using Maintenance*, no. 7, pp. 1–3, Jul. 2018.

[15] Y. Li, "Fault diagnosis for gyroscope based on incremental fuzzy support vector machine," M.S. thesis, Dept. Automat., NUAA, Nanjing, China, 2015.

[16] É. J. de Oliveira, I. Milagre da Fonseca, and H. Koiti Kuga, "Fault detection and isolation in inertial measurement units based on -CUSUM and wavelet packet," *Math. Problems Eng.*, vol. 2013, Aug. 2013, Art. no. 869293.

[17] B. Cai, H. Liu, and M. Xie, "A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks," *Mech. Syst. Signal Process.*, vol. 80, pp. 31–44, Dec. 2016.

[18] B. Cai, X. Shao, Y. Liu, X. Kong, H. Wang, H. Xu, and W. Ge, "Remaining useful life estimation of structure systems under the influence of multiple causes: Subsea pipelines as a case study," *IEEE Trans. Ind. Electron.*, to be published, doi: 10.1109/TIE.2019.2931491.

[19] Y. Li and X. Wu, "Gradual fault-tolerance design of redundant SINS," *J. Projectiles, Rockets, Missiles Guid.*, vol. 28, no. 2, pp. 57–60, Apr. 2008.

[20] H. Liang, H. Xu, and Z. Lv, "The fault diagnosis method of redundant gyroscopes based on support vector machine," *Aerosp. Control*, vol. 32, no. 5, pp. 77–83, Oct. 2014.

[21] W. Zhou, X. Li, and S. Zhao, "Fault-tolerant algorithm and information management for redundant strapdown inertial navigation system," *J. Nat. Univ. Defense Technol.*, vol. 40, no. 1, pp. 32–36, Feb. 2018.

[22] M. Zheng, L. Wu, L. Li, C. Liu, L. Wang, and S. Sun, "A modified method for fault detection and isolation of redundant inertial measurement Unit in dynamic environment," in *Proc. CCC*, Dalian, China, Jul. 2017, pp. 6998–7002.

[23] F. fen Wang, T. Zhang, and W. X. Fu, "An improved principal component analysis algorithm on FDI of redundant inertial measurement unit," in *Proc. 37th Chin. Control Conf. (CCC)*, Wuhan, China, Jul. 2018, pp. 6082–6086.

[24] W. Lee and C. Park, "Double fault detection of cone-shaped redundant IMUs using wavelet transformation and EPSA," *Sensors*, vol. 14, no. 2, pp. 3428–3444, Feb. 2014.

[25] J. Harrison and E. Gai, "Evaluating sensor orientations for navigation performance and failure detection," *IEEE Trans. Aerosp. Electron. Syst.*, vols. AES–13, no. 6, pp. 631–643, Nov. 1977.

[26] Y. Gao, J. Zhu, and Y. Zhao, "Study on redundancy configuration of strap-down inertial gyros," *Tactical Missile Technol.*, vol. 2, pp. 62–66, Feb. 2008.

[27] S. Furao and O. Hasegawa, "An incremental network for on-line unsupervised classification and topology learning," *Neural Netw.*, vol. 19, no. 1, pp. 90–106, Jan. 2006.

[28] S. Furao, T. Ogura, and O. Hasegawa, "An enhanced self-organizing incremental neural network for online unsupervised learning," *Neural Netw.*, vol. 20, no. 8, pp. 893–903, Oct. 2007.

[29] T. Qiu, F. Shen, and J. Zhao, "Review of self-organizing incremental neural network," *Jour. Softw.*, vol. 27, no. 9, pp. 2230–2247, Sep. 2016.

[30] T. Martinetz and K. Schulten, "Topology representing networks," *Neural Netw.*, vol. 7, no. 3, pp. 507–522, Jan. 1994.

[31] J. Harmouche, C. Delpha, and D. Diallo, "Incipient fault detection and diagnosis based on Kullback-Leibler divergence using principal component analysis: Part I," *Signal Process.*, vol. 94, no. 1, pp. 278–287, Jan. 2014.

[32] C.-L. Wen, "Relative PCA with applications of data compression and fault diagnosis," *Acta Automatica Sinica*, vol. 34, no. 9, pp. 1128–1139, Mar. 2009.

[33] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, "A review of process fault detection and diagnosis. Part III: Process History based Methods," *Comput. Chem. Eng.*, vol. 27, no. 3, pp. 327–346, Mar. 2003.

[34] G. Yan, J. Weng, "Simulation of strapdown inertial navigation and integrated navigation," in *Lecture notes on strapdown inertial navigation algorithm and integrated navigation principle*. Xian, China: Northwestern Polytechnical Univ., 2016, pp. 178–182. [Online]. http://blog.sina.com.cn/s/blog_40edfdc90102xd2t.html

XIAOQIANG HU received the B.S. degree from the School of Information Science and Technology and the M.S. degree from the School of Aeronautics and Astronautics, Xiamen University, Fujian, China, where he is currently pursuing the Ph.D. degree with the Department of Automation. His research interests include robotics, integrated navigation, and fault diagnosis.

XIAOLI ZHANG received the Ph.D. degree in control theory and application from Northeastern University, China. From 2002 to 2003, he was a Postdoctoral Research Fellow with the Department of Automation, Tsinghua University. Since 2004, he has been with the Department of Automation, Xiamen University, China, where he is currently an Associate Professor. His main research interests include switched systems, nonlinear systems, and robust control.

XIAFU PENG received the M.S. degree in control theory and applications from the Harbin Institute of Naval Engineering, Harbin, China, and the Ph.D. degree in control science and engineering from Harbin Engineering University, Harbin, in 2001. Since 2002, he has been with the Department of Automation, Xiamen University, China, where he is currently a Professor. His research interests include inertial technology and intelligent transportation systems.

DELONG YANG received the B.S. degree from the School of Mathematical Science, Huaibei Normal University, Huaibei, China, and the M.S. degree from the School of Science, Jimei University, Xiamen, China. He is currently pursuing the Ph.D. degree with the Department of Automation, Xiamen University, Xiamen. His research interests include robotics, computer vision, deep learning, and image processing.