Two-level fusion big data compression and reconstruction framework combining second-generation wavelet and lossless compression

Zhang Chuanchao^{1,2}

Received: 10 March 2020 / Accepted: 14 May 2020 / Published online: 6 June 2020
© The Author(s) 2020

Abstract

In view of the characteristics of big data, fuzziness, and real time of data acquisition and transmission in the fuzzy information system faced by aircraft health management, to reduce the load of airborne data processing and transmission system under the condition of limited airborne computing resources and strong time constraints, the data collected by the airborne system are first compressed, and the amount of data are reduced before transmission and reconstructed after transmission. In view of the situation that the compression ratio of primary data compression is too small and the compression time is too long for large-scale fuzzy systems to meet the transmission requirements of the system, this paper combines the advantages of lossy compression method which consumes less time and lossless compression method which has higher compression ratio, and innovatively proposes a two-level data compression and reconstruction framework combining lossy compression and lossless compression. The optimization analysis is carried out. Taking a real aero-engine health sample as an example, the validity, scientificity, and robustness of the proposed framework are verified by comparing with data compression and reconstruction algorithm based on redundant sparse representation and compressed sensing.

Keywords Large-scale fuzzy information system · Data compression and reconstruction · Second-generation wavelet · Lossless compression and reconstruction · Two-level fusion framework · Aircraft health management

Introduction

Aircraft health management parameters and data size are multiplying. In the 1980s, the aircraft health monitoring parameters were only 3000, while up to now, the most advanced civil aircraft has 400,000 monitoring parameters. The historical data of airlines operation and maintenance are increasing day by day. The annual operation data of the major airlines have reached tens TBs. These big data monitoring and acquisition have brought a great burden to the limited air-ground data links, but also increased the workload of maintenance engineers and technicians, and reduced the work efficiency. Therefore, in the era of big data of aircraft, it is imperative to study the compression and

reconstruction algorithm of large data to enable maintenance personnel to determine the failure cause and maintenance measures in time, and make joint efforts to arrange maintenance time to provide technical support, so as to achieve preventive maintenance, improve aircraft dispatch rate, and reduce maintenance costs. However, aircraft health management data are a typical large-scale fuzzy information system with large amount of data and fuzziness.

In recent years, wavelet analysis has become one of the fast signal processing methods to solve the problem of big data. From the view of basis function, wavelet transform absorbs the characteristics of time-shift window function of triangular basis function and in short-time Fourier transform, and forms the basis function of oscillation and attenuation. As long as a suitable wavelet basis function is selected, the signal can be separated reasonably at different times and frequencies, and the result has the ability to describe the local characteristics of the signal in both time and frequency domains. Wavelet and wavelet packet transform provide efficient and powerful tools for non-stationary description, separation of fault characteristic frequencies of mechanical equipment, and extraction of weak information for dynamic signals.

✉ Zhang Chuanchao
574799770@qq.com

¹ School of Information Engineering, Wuhan University of Technology, Wuhan 430070, Hubei, People's Republic of China

² Aviation Industry Corporation of China, Beijing 100028, People's Republic of China

In 1997, Sweldens [1] had used lifting method to construct wavelet function, and proposed the concept of second-generation wavelet transform [2]. In 1998, Daubechies and Sweldens had cooperated to prove that any wavelet transform with a finite impulse response filter could be implemented by a finite-step lifting method. This conclusion establishes the relationship between the classical discrete wavelet transform (also known as the first-generation wavelet transform) and the second-generation wavelet transform, so that more and more wavelet basis functions can be constructed [3]. The second-generation wavelet transform theory is a significant breakthrough in the field of wavelet analysis in recent years. It does not rely on Fourier transform to construct wavelet. It can not only improve the characteristics of existing wavelet by lifting method, but also construct new wavelet by interpolation subdivision method. All operations are carried out in time domain, and the results are the same as those of the first generation of wavelet transform. The second-generation wavelet transform provides a novel method for constructing wavelet, and it has become a hotspot in the research and application of discrete wavelet transform in recent years.

The first-generation wavelet transform realizes multi-scale analysis of signals through scaling and translation operations, so it can effectively extract the time–frequency domain information of non-stationary signals. It provides a variety of basis functions, including triangular basis functions, and is an effective tool for non-stationary signal analysis. Therefore, it is widely used in the field of fault diagnosis and has achieved exciting results. In reference [4], the application of the first-generation wavelet analysis theory in condition monitoring and fault diagnosis is reviewed. Wavelet transform has been applied in time–frequency analysis of fault signal [5], fault feature extraction [6], extreme point detection [7], noise elimination and extraction of weak signal [8], vibration signal compression [9], and system and parameter identification [10]. Although the first-generation wavelet has been widely used in all aspects of fault diagnosis, there are also some limitations in practical application, such as the difficulty of choosing wavelet basis functions, lacking variety, and self-adaptability.

In recent years, many scholars have paid attention to the research of second-generation wavelet in the field of equipment fault diagnosis. Among them, there are many research results about the application of second-generation wavelet in signal compression, signal denoising, fault detection, and fault mode classification. In references [11, 12], the authors implemented a scheme of recording data compression by lifting algorithm and the selection of wavelet bases for fault recording data was studied, the formula for calculating the optimal decomposition level of wavelet was given, the relationship between compression ratio and sampling frequency was studied, and a better compression result was obtained. In reference [13], a vibration signal compression method based

on lifting wavelet is proposed and verified through compression experiments on actual vibration signals. It is proved that this method can obtain a higher compression ratio and retain the waveform characteristics of the original signal.

For signal compression, reference [14] proposed a method of data compression for power system fault recording data based on the second-generation wavelet transform in 2004. In the method, the signal is decomposed by integer wavelet transform firstly, and then, the transform coefficients are coded by Huffman coding method; thus, the lossless compression of recording data is realized. The reference [15] applies the second-generation wavelet transform and Huffman coding technology to data compression of real-time monitoring system for large-scale hydraulic turbines, and the compression rate is more than 10% higher than that of WinZip software. Reference [11] proposed a data compression scheme combining lifting format wavelet transform with embedded zerotree coding. First, lifting format wavelet transform is applied to fault recording data, then threshold quantization is applied to transform coefficients. Second, the quantized coefficients are coded into bits flow by embedded zerotree coding. The compression scheme can realize variable bit rate, reduce computational complexity, and memory occupation, and achieve a better signal-to-noise ratio and compression ratio [11]. Reference [12] compressed the fault recording data by combining integer second-generation wavelet transform with Huffman coding, and only retained and coded the scale coefficients. Although this method improved the compression ratio, there was a problem of transient characteristic distortion of the signal [16].

The second-generation wavelet is a method of constructing wavelet based on interpolation subdivision principle in time or space domain, and not depending on Fourier transform. The second generation of wavelet provides the possibility to construct adaptive and non-linear wavelet basis functions while guaranteeing that it can always be reconstructed completely. In addition, the principle of the second-generation wavelet is clearer, the efficiency is higher, the memory space consumed by standard operation is less than others, and integer wavelet transform can be realized quickly. In view of the situation that the compression ratio of primary data compression is too small and the compression time is too long for large-scale fuzzy information systems to meet the transmission requirements of the system, this paper combines the less time-consuming advantages of the second-generation wavelet transform and the higher compression ratio advantages of lossless compression and reconstruction method, and studies the compression and reconstruction framework combining lossy compression and lossless compression methods for big data.

Principle of wavelet data compression and reconstruction

Data compression and reconstruction of first-generation wavelet

1. Scale coefficient and wavelet coefficient

According to the principle of multiresolution analysis, if the signal $f(t) \in V_{j-1}$ is arbitrary, its expansion in space V_{j-1} is as follows:

$$f(t) = \sum_k C_{j-1,k} 2^{(-j+1)/2} \phi(2^{-j+1}t - k), \quad (1)$$

$f(t)$ will be decomposed once, i.e., it is projected into spaces V_j and U_j , then there is:

$$f(t) = \sum_k C_{j,k} 2^{-j/2} \phi(2^{-j}t - k) + \sum_k d_{j,k} 2^{-j/2} \phi(2^{-j}t - k) \quad (2)$$

where $C_{j,k}$ is the discrete approximation of $f(t)$ in space V_j and is called scale coefficient. $d_{j,k}$ is the discrete value of $f(t)$ in space U_j and is called the wavelet coefficients, which is the discrete wavelet transform $W_{\phi,f}(2^j, k)$ or $DW_{m,n}$, i.e., $DW_{m,n} = d_{j,k}$.

It is knowable that:

$$C_{j,k} = \langle f(t), \phi_{j,k}(t) \rangle = \int_{-\infty}^{\infty} f(t) 2^{-j/2} \overline{\phi(2^{-j}t - k)} dt, \quad (3)$$

$$d_{j,k} = \langle f(t), \varphi_{j,k}(t) \rangle = \int_{-\infty}^{\infty} f(t) 2^{-j/2} \overline{\varphi(2^{-j}t - k)} dt, \quad (4)$$

where $\phi(t)$ is a scale function. $\varphi(t)$ is a wavelet function.

2. The properties of $\phi(t)$, $\varphi(t)$ and mallat pyramid decomposition algorithm

Let to be assumed that $\varphi(t)$ and $\psi(t)$ are the normal orthogonal basis functions, respectively, in scale space V_0 and wavelet space U_0 , then because $V_0 \subset V_{-1}$, $U_0 \subset V_{-1}$, and $\phi(t)$ and $\varphi(t)$ must belong to V_{-1} . That is to say, $\phi(t)$ and $\varphi(t)$ can be linearly expanded by the orthogonal baseline $\phi_{-1,n}(t)$ of space V_{-1} as follows:

$$\phi(t) = \sum_n h(n) \phi_{-1,n}(t) = \sqrt{2} \sum_n h(n) \phi(2t - n), \quad (5a)$$

$$\varphi(t) = \sum_n g(n) \phi_{-1,n}(t) = \sqrt{2} \sum_n g(n) \phi(2t - n), \quad (5b)$$

where $h(n)$ and $g(n)$ are expansion coefficients, that is:

$$h(n) = \langle \phi, \phi_{-1,n} \rangle, \quad g(n) = \langle \varphi, \phi_{-1,n} \rangle. \quad (6)$$

Equations (5a) and (5b) describe the relationship between the two-scale spatial basis functions, which are called two-scale difference engineering. Expansion and translation of the two formulas for time are taken, and then:

$$\phi(2^{-j}t - k) = \sum_n h(n) \sqrt{2} \phi(2^{-j+1}t - 2k - n) \quad (7a)$$

$$\varphi(2^{-j}t - k) = \sum_n g(n) \sqrt{2} \phi(2^{-j+1}t - 2k - n). \quad (7b)$$

If $m = 2k + n$, then Eqs. (7a) and (7b) are changed as in:

$$\phi(2^{-j}t - k) = \sum_n h(m - 2k) \sqrt{2} \phi(2^{-j+1}t - m), \quad (8a)$$

$$\varphi(2^{-j}t - k) = \sum_n g(m - 2k) \sqrt{2} \phi(2^{-j+1}t - m). \quad (8b)$$

Equations (8a) and (8b) are substituted into Eqs. (3) and (4), and then, it is obtained that:

$$C_{j,k} = \sum_m h(m - 2k) \int_{-\infty}^{\infty} f(t) 2^{(-j+1)/2} \overline{\phi(2^{-j+1}t - m)} dt, \quad (9a)$$

$$d_{j,k} = \sum_m g(m - 2k) \int_{-\infty}^{\infty} f(t) 2^{(-j+1)/2} \overline{\varphi(2^{-j+1}t - m)} dt. \quad (9b)$$

And because $C_{j-1,k} = \int_{-\infty}^{\infty} f(t) 2^{(-j+1)/2} \overline{\phi(2^{-j+1}t - m)} dt$, so Eqs. (9a) and (9b) are changed as:

$$C_{j,k} = \sum_m h(m - 2k) C_{j-1,m}, \quad (10a)$$

$$d_{j,k} = \sum_m g(m - 2k) C_{j-1,m}. \quad (10b)$$

It can be seen from Eqs. (10a) and (10b) that the spatial scale coefficients of the space V_j and the spatial wavelet coefficients of the space U_j can be obtained by weighted summation of scale coefficients of the V_{j-1} spatial through filter coefficients $h(n)$ and $g(n)$. If the scale coefficient $C_{j,k}$ of space V_j is decomposed according to Eqs. (10a) and (10b), then the scale coefficients $C_{j+1,k}$ and wavelet coefficients $d_{j+1,k}$ of spaces V_{j+1} and U_{j+1} can be obtained, respectively. Equations (10a) and (10b) are the famous Mallat pyramid decomposition algorithm, which is the fast algorithm of discrete wavelet transform.

3. Principles of decomposition and reconstruction

Fig. 1 Structural schematic diagram of decomposition and reconstruction for one-dimensional wavelet

Fig. 2 Structural schematic diagram of decomposition and reconstruction for two-dimensional wavelet decomposition and reconstruction

According to the previous analysis, the structural schematic diagrams of one-dimensional and two-dimensional wavelet decomposition (transform) and reconstruction (inverse transform) can be given as shown in Figs. 1 and 2. In these figures, $\bar{H}$ and $\bar{L}$ represent the high-pass and low-pass filters used in decomposition, and H and L represent the high-pass and low-pass filters used in reconstruction respectively. $\bar{H}$ and H are conjugated with each other, and $\bar{L}$ and L are conjugated with each other. $\uparrow 2$ represents up sampling according to factor 2. $\downarrow 2$ represents down sampling according to factor 2.

Compression and reconstruction of second-generation wavelet

1. Signal compression principle based on second-generation wavelet

The signal compression method based on wavelet transform first decomposes the being compressed signal, and then, a threshold is set to reserve the wavelet transform coefficients with larger modulus, usually reserving approximate coefficients completely, and the wavelet coefficients with smaller modulus are discarded which are corresponding to noise and redundant information in the signal. Its basic principle is that a threshold at a certain scale l is set as in:

$$\eta_l = \lambda \times \max\{|d_l(n)|\} \quad n = 1, 2, \dots, N/2^l, \quad (11)$$

where N is the length of the being compressed signal. $d_l(n)$ is the wavelet coefficient in the scale l . λ ($0 \leq \lambda \leq 1$) is the proportional coefficients. $\lambda = 0.1$ indicates that the threshold is set to be 0.1 times of wavelet detail coefficient with the largest modulus.

After the signal is compressed in the l level, the wavelet coefficients in this scale are as follows,

$$\hat{d}_l(n) = \begin{cases} d_l(n) & |d_l(n)| \geq \eta_l \\ 0 & |d_l(n)| < \eta_l \end{cases} \quad (12)$$

Therefore, only less data need to be stored and transmitted, i.e., approximate coefficients and less detail coefficients after wavelet decomposition. Therefore, the original signal can be compressed. When the original signal is needed, the stored wavelet transform coefficients can be reconstructed to restore the signal.

The principle of signal compression based on the second-generation wavelet transform is the same as that based on the classical wavelet transform. However, the former has the advantages of faster operation speed, completely standard calculation, integer transformation, no occupancy of auxiliary memory space, and the inverse transformation only needing to take the inverse operation of the forward transformation, and so on. Therefore, it is more suitable for signal compression.

Wavelet transform based on lifting algorithm is called second-generation wavelet transform. It enables people to interpret the basic theory of wavelet in a simple way. The first generation of wavelet transform can find an equivalent lifting scheme. The lifting scheme divides the first generation of wavelet transform into three stages: decompose, predict and update.

- a) To decompose. The input signal $s(i)$ is divided into two smaller subsets $s(i-1)$ and $d(i-1)$. $d(i-1)$ is also known as wavelet subsets. The simplest decomposition method is to divide the input signal $s(i)$ into two groups according to parity. The wavelet generated by this decomposition is called lazy wavelet. The decomposition process can be expressed as $F(s(i)) = (s(i-1), d(i-1))$. Where, $F(s(i))$ is a decomposition process.
- b) To predict. Based on the relativity of original data, the predicted value $P(s(i-1))$ of even sequence $s(i-1)$ is used to predict (or interpolate) odd sequence $d(i-1)$. That is to say, the result after the filter acts on the even signal is taken as the predicted value of the odd signal. The residual signal is the difference between the actual value of the odd signal and the predicted value. In practice, although the subset $d(i-1)$ cannot be accurately predicted from the subset $s(i-1)$, but $P(s(i-1))$ may be very close to $d(i-1)$. Therefore, the difference between $P(s(i-1))$ and $d(i-1)$ can be used to replace the original $d(i-1)$. The new $d(i-1)$ includes less information than the original $d(i-1)$. Therefore, it will get, that is to say $d(i-1) = d(i-1) - \dots$. In this case, the smaller subset $s(i-1)$ and

subset $d(i-1)$ can be used to replace the original signal set $s(i)$. Repeating the decomposition and prediction process, after n step, the original signal set can be expressed by $\{s(n), d(n), d(n-1), d(n-2), \dots, d(1)\}$.

- c) To update. To keep some global characteristics of the original signal set in its subset $s(i-1)$, and keeps certain scalar property $Q(x)$ (such as unchanged mean, vanishing moment, etc.) of the original data, i.e., $Q(s(i-1)) = Q(s(i))$. It is possible to update $s(i-1)$ using the computed small subset $d(i-1)$, so that the former retains its characteristic $Q(x)$, i.e., to construct an U operator to update $s(i-1)$. The definition is as follows: $s(i-1) = s(i-1) + U(d(i-1))$.

From the above analysis, it can see that the lifting method can realize in-situ operation, i.e., the method does not need data other than the output of the previous lifting step. In this way, new data streams can be used to replace old data streams at each point. When the in-situ lifting filter groups are reused, the interleaved wavelet transform coefficients are obtained.

2. Integer second-generation wavelet transform

Generally, the transform coefficients of classical wavelet transform are real values. Even if the original signal sequence is an integer sequence, the wavelet transform coefficients of classical wavelet transform are real values. This restricts the application and data compression of wavelet transform. First, the transform coefficient matrix needs floating-point storage, and generally needs larger storage space. Second, it is difficult to produce a good lossless encoder. Therefore, integer wavelet coefficients are expected to be applied in many applications. The integer wavelet transform refers to the mapping of integer sequences into integer wavelet coefficients, which allow floating-point operations in the process of transformation.

The second-generation wavelet transform based on interpolation subdivision can realize integer transformation by integer operation in the operation process. Taking (2,2), (4,2), and (4,4) wavelets as examples, the formulas corresponding integer transformation are, respectively:

- a) (2,2) integer wavelet transform is as in:

$$\begin{aligned} d_{l+1}(n) &= a_l(2n+1) - \left\lfloor \frac{1}{2}[a_l(2n) + a_l(2n+2)] + \frac{1}{2} \right\rfloor \\ a_{l+1}(n) &= a_l(2n) + \left\lfloor \frac{1}{4}[d_{l+1}(n-1) + d_{l+1}(n)] + \frac{1}{2} \right\rfloor, \end{aligned} \quad (13)$$

- b) (4,2) integer wavelet transform is as in:

$$\begin{aligned} d_{l+1}(n) &= a_l(2n+1) - \left\lfloor \frac{9}{16}[a_l(2n) + a_l(2n+2)] \right. \\ &\quad \left. - \frac{1}{16}[a_l(2n-2) + a_l(2n+4)] + \frac{1}{2} \right\rfloor \\ a_{l+1}(n) &= a_l(2n) + \left\lfloor \frac{1}{4}[d_{l+1}(n-1) + d_{l+1}(n)] + \frac{1}{2} \right\rfloor, \end{aligned} \quad (14)$$

- c) (4,4) integer wavelet transform is as in:

$$\begin{aligned} d_{l+1}(n) &= a_l(2n+1) - \left\lfloor \frac{9}{16}[a_l(2n) + a_l(2n+2)] \right. \\ &\quad \left. - \frac{1}{16}[a_l(2n-2) + a_l(2n+4)] + \frac{1}{2} \right\rfloor \\ a_{l+1}(n) &= a_l(2n+1) + \left\lfloor \frac{9}{32}[d_l(n-1) + d_l(n)] \right. \\ &\quad \left. - \frac{1}{32}[d_l(n-2) + d_l(n+1)] + \frac{1}{2} \right\rfloor. \end{aligned} \quad (15)$$

Where $\lfloor \cdot \rfloor$ means downward truncation and integer.

3. Wavelet decomposition and reconstruction principle of multiresolution analysis

Based on the multiresolution analysis principle, it can be concluded that each subspace has the characteristics of step-by-step inclusion and step-by-step replacement in the frequency band space partition process. If the total bandwidth $(0-\pi)$ occupied by the original signal $f(t)$ is defined as V_0 , then after first-level decomposition V_0 is decomposed into two subspaces: low-frequency subspace V_1 whose frequency band range is $0-\pi/2$, and high-frequency subspace U_1 whose frequency band range is $\pi/2-\pi$. After the second-level decomposition, V_1 is divided into two subspaces: the low frequency subspace V_2 whose frequency band range is $0-\pi/4$ and the high-frequency subspace U_2 whose frequency band range is $\pi/4-\pi/2$. This decomposition continues. The decomposition process of this subspace can be described as in:

$$V_0 = V_1 + U_1; V_1 = V_2 + U_2; \dots; V_{i-1} = V_i + U_i, \quad (16)$$

where $V_i (i = 1, 2, \dots)$ is the low frequency subspace reflecting signal' profile. $U_i (i = 1, 2, \dots)$ is the high-frequency subspace reflecting signal' details.

It can be seen that subspaces have the following characteristics:

- a) Including characteristics step by step, as in:

$$V_0 \supset V_1 \supset V_2 \supset \dots \supset V_i. \quad (17)$$

b) Substitution characteristics step by step, as in:

$$\begin{aligned} V_0 &= V_1 + U_1 = V_2 + U_1 + U_2 = V_3 + U_3 + U_2 + U_1 \\ &= V_i + U_i + U_{i-1} + U_{i-2} + \cdots + U_1. \end{aligned} \quad (18)$$

Therefore, the data compression process based on wavelet is to decompose the spatial function step by step, and then, the structure of each subspace of further analysis and decomposition is carried out on the basis of spatial decomposition, while the data reconstruction process based on wavelet is its inverse transform.

Lossless data compression and reconstruction algorithms

The lossless data compression and reconstruction method tries to change the probability distribution of the source, so that the probability distribution of the signal is as non-uniform as possible. Then, the optimal code method is used to redistribute codewords to each signal, so that the average code length approximates the source. The pressure efficiency of these methods is limited up to their information entropy. Typical lossless data compression and reconstruction methods include Huffman coding [17], run-length coding [18], arithmetic coding [19], dictionary coding [20, 21], and so on. The principles of them are referred to relevant references [17–21], which are not discussed here.

Framework of two-level fusion large data compression and reconstruction based on second-generation wavelet

Signal compression is one of the most successful applications of wavelet transform. Wavelet transform uses compactly supported basis functions to decompose signals, which not only makes the transform more energy-intensive, but also makes the amplitude of most wavelet coefficients very small. Therefore, the wavelet coefficients with smaller amplitude can be discarded by threshold quantization to achieve data compression. Compared with other transform-based signal compression methods, the wavelet-based method can often obtain high compression ratio while retaining more signal singularity, which is very important for fault diagnosis.

Wavelet-based compression is a lossy compression method, but the compression ratio of all compression algorithms is limited. In large data environment, data transmission bandwidth is limited, and it is difficult to meet the requirements of one-time compression. Therefore, to obtain a larger compression ratio while the reconstructed data cannot be too distorted, it can be considered that the two-level signal

Table 1 Steps of the framework and algorithm

Input: Signal V_0 to be compressed

Output: Reconstructed signal V'_0

- 1: the data V_0 are decomposed into low-frequency subspace V_i and high-frequency U_i subspace, i.e. $V_0 = V_i + U_i$, using second-generation wavelet decomposition to compress the data
- 2: the low-frequency subspace V_i is coded and compressed into a low-frequency subspace V_{ii} and a high-frequency subspace U_{ii} , i.e. $V_i = V_{ii} + U_{ii}$, by using step-by-step substitution characteristics and lossless compression and reconstruction methods such as Huffman coding, run-length coding, and arithmetic coding
- 3: to take data transmission after data compression
- 4: to decode the received data step by step, $k = n \dots 3, 2$, and reconstruct the data before lossless compression, i.e., $V_{ii} \oplus U_{ii} = V_i$, using the lossless compression–reconstruction methods such as Huffman coding and run-length coding
- 5: the second-generation wavelet is used to reconstruct the data and the pre-compression data $V_i \oplus U_i = V'_0$ is gained

compression method combines lossless compression method with lossless compression method, i.e., the method may be based on second-generation wavelet and lossless compression reconstruction, such as Huffman coding, run-length coding, arithmetic coding, etc..

Algorithm description

The steps of the framework and algorithm are described in Table 1.

Framework and process

The step which describes two-level data compression and reconstruction algorithm based on second-generation wavelet is shown in Fig. 3. Data compression based on wavelet transform achieves its goal by losing data accuracy, which belongs to distorted coding. Since there is still redundancy in the data after distortion coding, the compression ratio can be further improved by lossless compression coding. Lossless compression methods mainly include Huffman coding, run-length coding, arithmetic coding, dictionary coding, etc. The optimal matching between the probability and the length of codeword among them is mainly achieved using the distribution characteristics of the probability of data or data sequence.

Wavelet decomposition coefficients include scale coefficients and wavelet coefficients. Scale coefficients embody the main characteristics of data and need to be preserved completely. Otherwise, large errors will occur in the reconstruction process. Scale coefficients are further coded and compressed. After threshold quantization, the coefficient matrix of wavelet coefficients is a sparse matrix and achieves

Fig. 3 Two-level data compression and reconstruction method combining one-dimensional second-generation wavelet

compression after doing with a threshold. The reconstruction process is the inverse process of the above process. First, the compressed data are decoded. Second, the decoded data are reconstructed using the second-generation wavelet method to obtain the reconstructed data of two-level fusion compression and reconstruction.

Experimental analysis and results

Experimental data and simulation environment

The real data of an aero-engine on a flight were selected to verify the framework. The flight took place for nearly 13 h with 46,595 sampling points. There are 205 aero-engine monitoring parameters selected. The sampling data occupy 116 MB (121,864,192 bytes) space.

A series of simulation and contrast experiments are carried out using this framework and algorithm. The hardware test platform of the experiment is a computer with Intel Core i7 CPU, 4.0 GHz main frequency, and 16 GB memory. The simulation software platform is with Windows 7 64 bit operating system and Matlab 2014a.

Evaluating metrics

Generally, the compression and reconstruction capability of a data system can be expressed from different aspects by those metrics, such as compression ratio, compression efficiency, compression speed, reconstruction speed, and reconstruction accuracy. At the same time, the complexity of the system is taken into account, including computational complexity of the algorithm, required storage, signal quality, and delay. To evaluate the performance of signal compression and reconstruction algorithms objectively and efficiently, the following metrics are recommended [22]:

1. Compression ratio (abbreviated as CR)

CR is defined as in:

$$CR = N/N_c \times 100\%, \quad (19)$$

where N_c is the length of the wavelet coefficients which is reserved for the final. N is the length of the original signal. CR is the most commonly used compression metrics, which can accurately reflect the changes of data volume before and after compression. However, CR is affected by the sampling rate and quantization accuracy of the original signal data.

2. Mean square error (abbreviated as MSE)

MSE is defined as in:

$$MSE = \sqrt{\|\hat{f}(i) - f(i)\|^2 / \|f(i)\|^2}, \quad (20)$$

where $f(i)$ and $\hat{f}(i)$ are the original and decompressed signals, respectively.

3. Signal-to-noise ratio (abbreviated as SNR)

SNR is defined as in:

$$SNR = 10 \lg \left(\frac{\sum_{i=1}^N |f(i)|^2}{\sum_{i=1}^N |f(i) - \hat{f}(i)|^2} \right), \quad (21)$$

where $f(i)$ and $\hat{f}(i)$ are the original and decompressed signals, respectively.

4. Compression time (abbreviated as CT) and reconstruction time (abbreviated as RT)

Compression time (abbreviated as CT) and reconstruction time (abbreviated as RT) are defined as the running time spent by the algorithm to complete data compression or reconfiguration tasks. CS and RS are of CT and RT respectively. The running time is inversely proportional to the speed. The longer the running time is, the slower the running speed is.

Comparison between the first and second-generation wavelet

To verify the validity of the second-generation wavelet compression reconstruction framework proposed in this paper, the vibration parameters Eng1VibBBA of an aero-engine are taken as an example, and the test sets of different data

Table 2 Performance comparison between the first- and second-generation wavelets

Method	Data size	CR	MSE	SNR	CT (s)	RT (s)
First-generation wavelets	4096	1.03	0.0002	41.44	0.26	0.08
	8192	1.05	0.0014	41.16	0.10	0.10
	16,384	1.05	0.0010	48.35	0.11	0.13
	24,576	1.06	0.0007	51.60	0.12	0.17
	32,768	1.06	0.0006	53.21	0.24	0.20
Second-generation wavelets	4096	1.15	0.032	140.75	0.18	0.030
	8192	1.55	0.040	116.79	0.20	0.038
	16,384	1.87	0.037	117.38	0.68	0.086
	24,576	1.94	0.033	118.35	0.76	0.10
	32,768	1.98	0.032	118.82	0.72	0.11
Mallat algorithm	4096	1.18	0.0039	119.38	0.62	0.012
	8192	1.39	0.060	115.28	0.65	0.011
	16,384	1.18	0.031	121.06	0.68	0.013
	24,576	1.13	0.021	124.02	0.65	0.015
	32,768	1.10	0.016	126.30	0.65	0.014
Wavelet packs	4096	1.08	0.0011	24.99	0.52	0.015
	8192	1.22	0.026	14.93	0.55	0.017
	16,384	1.11	0.014	21.75	0.59	0.018
	24,576	1.08	0.009	25.03	0.57	0.019
	32,768	1.064	0.0068	27.52	0.58	0.021

scales are selected to test. The results show that the second-generation wavelet compression reconstruction method is more effective than the first-generation wavelet compression reconstruction method. The results are shown in Table 2.

The results show that the method proposed in this paper is better than that based on the first-generation wavelet in compression ratio and signal-to-noise ratio, while the compression speed and reconstruction speed are basically the same. It pays the price of reconstructing accuracy. However, it can still meet the requirements of aircraft health management.

Comparison among lossless compression and reconstruction methods

To verify the effectiveness of the proposed fusion compression reconstruction framework based on second-generation wavelet and lossless compression reconstruction method, taking Eng1VibBBA as an example, the vibration parameters of an aero-engine are tested with test sets of different data scales to verify the performance of lossless Huffman coding, arithmetic coding, run-length coding, and dictionary coding. Since Huffman coding, arithmetic coding, run-length coding, and dictionary coding belong to lossless compression reconstruction methods, only three metrics that are CR, CT, and RT need to be verified. The comparison result is shown in Table 3.

In practical systems, the occurrence probability of output symbols varies greatly among different signal source, and it is difficult to know the occurrence probability of each symbol among the source output data in advance. Therefore, it is difficult to ensure that Huffman coding is carried out accurately according to the occurrence probability of each symbol. However, in the aircraft health management system, the data to be compressed have been measured, and the probability of occurrence of all the symbols output from the source can be counted. Then, the probability of character set obtained by the statistical law of the symbols can be used for Huffman coding, which satisfies the preconditions and basic principles of Huffman coding. In addition, because the aircraft is cruising for a large amount of time in flight, the measured values of the monitoring parameters in its health management are not very different or even the same, and the redundancy is large. Therefore, Huffman coding has excellent performance. Run-length encoding is similar to Huffman encoding in principle. The difference is that Huffman encoding counts the probability of the source outputting all symbols, while run-length encoding counts the length of the source output. Based on the redundancy of monitoring parameters in aircraft health management, the data compression performance of run-length coding is second only to Huffman coding.

Different from run-length encoding and Huffman encoding, the key of arithmetic encoding lies in the construction of encoding interval. Because of the large amount data

Table 3 Comparison with lossless compression and reconstruction methods

Method	Data size	CR	CT (s)	RT (s)
Huffman coding	4096	28.64	0.21	0.027
	8192	36.09	0.54	0.019
	16,384	99.9	1.16	0.018
	24,576	144.56	1.58	0.020
	32,768	217.01	2.07	0.019
Arithmetic coding	4096	0.15	0.47	0.34
	8192	0.20	0.92	0.57
	16,384	0.20	1.57	0.98
	24,576	0.17	2.47	1.62
	32,768	0.14	3.73	2.54
Run-length coding	4096	1.39	0.11	0.014
	8192	9.33	0.24	0.018
	16,384	7.87	0.39	0.037
	24,576	10.01	0.48	0.028
	32,768	11.12	0.50	0.036
Dictionary coding				
LZ77	4096	1.33	62.76	0.33
	8192	1.56	176.72	0.78
	16,384	1.51	612.93	3.02
	24,576	1.58	1662.27	4.65
	32,768	1.04	4750.65	15.35
LZ78	4096	3.70	37.48	0.25
	8192	4.06	118.95	0.37
	16,384	3.53	422.79	1.54
	24,576	2.52	1206.33	2.16
	32,768	2.07	2862.16	3.16

to be compressed and the long length of codewords, the construction of encoding interval consumes more time and storage space, and it does not have advantages. Therefore, its compression ratio and compression speed have great disadvantages. Arithmetic coding is more suitable for the compression of complex data structures such as images. This is well verified in Table 3.

The similarity between dictionary-based data compression technology and statistics-based data compression technology is that the rules of statistical characters are also needed in the process of dictionary construction. However, the difference lies in avoiding the precise statistical model, but introducing a pointer access window. The size of the window determines the performance and efficiency of the algorithm. The limitation of LZ77 encoding lies in how far the pointer can go back and the maximum length of the string that can be accessed. LZ77 encoding needs to find such a window in the previous text, which must be the largest matching window. Natural linear lookup is time-consuming and can be accelerated using appropriate data structures to index previous texts.

In contrast to the LZ77 method, where pointers can point to any substring in previous text windows, the LZ78 method limits the referential substring, but it does not have a window that limits the retrospective distance of indexable substrings. By restricting indexable string sets, LZ78 avoids the inefficiency of having more than one encoding representation for the same string. On the contrary, this situation often occurs in the LZ77 encoding method due to the frequent inclusion of many duplicate substrings in the window. The validation data in Table 2 show that, although the compression ratio is guaranteed, the disadvantage of compression time is highlighted because LZ77 and LZ78 face the reciprocating search of matrix windows.

Comparison between two-level fusion frameworks

Based on the above analysis, among the lossy compression and reconstruction methods, the method based on the second-generation wavelet has significant performance advantages and the compression ratio has the most advantages, while the signal-to-noise ratio and compression time are basically the same as those of Mallat algorithm, and it takes more time than other wavelet methods. While the method has the most advantages in reconstruction, and it is more time-consuming than Mallat algorithm, but smaller than other methods. Compression speed and reconstruction speed of the method are very fast, and have good adaptability and robustness to large data, which can effectively meet the needs of aircraft health management. Compression and reconstruction methods based on compressed sensing have some advantages over wavelet-based compression and reconstruction methods in compression ratio and performance. They have similar performance in signal-to-noise ratio, but are at disadvantages in compression speed and reconstruction time, and not adaptable to large data sets. They are not good enough for health management of aircraft. The real-time requirement is obviously insufficient.

Among the lossless compression and reconstruction methods, Huffman coding has the best compression ratio performance, followed by run-length coding and dictionary coding, and arithmetic coding has the worst performance. From the results, it can be concluded that arithmetic coding is not suitable for one-dimensional data compression and reconstruction applications. Huffman coding, run-length coding, and dictionary coding are ranked from best to worst in terms of compression speed and reconstruction speed performance, while the compression time of dictionary coding increases exponentially with the increase of data volume, which fails to meet the real-time requirements of aircraft health management. Based on this, in the compression and reconstruction of aircraft health management data processing, this paper chooses the excellent–excellent combination of second-generation wavelet, Huffman coding, and run-

Table 4 Comparison of two-level fusion data compression and reconstruction framework

Framework proposed in this paper	Data size	CR	MSE	SNR	CT (s)	RT (s)
Combination of second-generation wavelet and Huffman coding	4096	32.85	0.032	140.75	0.39	0.058
	8192	56.05	0.040	116.79	0.73	0.057
	16,384	187.11	0.037	117.38	1.84	0.11
	24,576	280.45	0.033	118.35	2.33	0.12
	32,768	430.12	0.032	118.82	2.79	0.13
Combination of second-generation wavelet and run-length coding	4096	13.59	0.032	140.75	0.29	0.05
	8192	14.49	0.040	116.79	0.44	0.06
	16,384	14.74	0.037	117.38	1.08	0.12
	24,576	19.42	0.033	118.35	1.24	0.13
	32,768	22.04	0.032	118.82	1.22	0.14

Table 5 Performance of the proposed algorithm under the data size 2048×1

Framework proposed in this paper	Data size	CR	MSE	SNR	CT (s)	RT (s)
Combination of second-generation wavelet and Huffman coding	2048	21.04	0.24	124.93	1.02	0.21
Combination of second-generation wavelet and run-length coding	2048	12.98	0.24	124.92	0.61	0.14

length coding. Taking Eng1VibBBA which is a vibration parameter of one aero-engine as an example, the performance of the two-stage fusion data compression and reconstruction framework proposed in this paper is verified, as shown in Table 4.

The calculation results from Table 4 show that the first- and second-level fusion data compression and reconstruction framework combining second-generation wavelet and Huffman coding achieves a very good compression ratio, and the compression speed and reconstruction speed are very agile, and the reconstruction accuracy is very high. The fusion two-level fusion data compression and reconstruction framework combining second-generation wavelet and run-length coding also has a good compression ratio. At the same time, the compression speed, reconstruction speed, and reconstruction accuracy are basically the same as those of the two-level fusion data compression and reconstruction framework combining second-generation wavelet and Huffman coding. These two frameworks can meet the requirements high compression ratio, high reconfiguration accuracy, and real-time requirements for aircraft health manage.

Comparison with compressed sensing theory

Compression and reconstruction method based on compressed sensing is the hotspot and focus in the field of data compression and reconstruction in recent years [23, 24]. It has been well applied in data denoising [25], and data compression and reconstruction [26]. To verify the framework, the performance of this framework is compared with that of compressed sensing. The performance of the framework

under the data scale of 2048×1 is simulated and calculated complementally, as shown in Table 5.

To verify the validity of the framework proposed in this paper, taking the vibration parameter Eng1VibBBA of an aero-engine as an example, selecting test sets of different data scales are tested, and the similar compression reconstruction methods based on compression sensing are verified and compared with the framework. The most typical mature and stable Base-tracking matching pursuit algorithm (BMP) [27], orthogonal matching pursuit algorithm (OMP) [28], and weak orthogonal matching pursuit algorithm (WMP) [29] reconstruction algorithm are selected to use as analysis methods. The validation results are shown in Table 6.

According to the verification results of Tables 4, 5, 6 and Fig. 4, the performance of the framework in compression ratio, reconstruction accuracy, and reconstruction speed is much better than that of compressed sensing. The performances in signal-to-noise ratio are basically the same, while the compression speed is far inferior to compressed sensing. However, the compression speed of framework is also very fast, which can still meet the real-time needs of aircraft health management.

In summary, under the multiple constraints of compression ratio, reconstruction accuracy, compression time, and reconstruction time for aircraft health management, the performance advantages of the framework proposed in this paper are significant.

Robustness analysis for big data

To verify the robustness of the two-stage fusion data compression and reconstruction framework, two other different

Fig. 4 Comparison analysis with compressed sensing methods

characteristic parameters, the second aero-engine vibration parameter Eng2VibBBA and the TPR instruction temperature Eng2TPRCmd, are selected to validate robustness of the framework for different data parameters. The validation results are shown in Table 7 and Fig. 5.

Table 6 Performance comparison with compressed sensing

Method	Data size	CR	MSE	SNR	CT (s)	RT (s)
BMP	2048	2.05	0.21	126.28	0.0052	2.46
	4096	1.42	0.0043	165.35	0.0039	7.72
	8192	1.45	0.063	138.51	0.0030	66.20
OMP	2048	2.05	0.21	126.29	0.0030	7.76
	4096	1.43	0.0043	165.36	0.0092	27.77
	8192	1.81	0.064	138.36	0.0065	147.52
WMP	2048	2.05	0.21	126.29	0.0043	7.79
	4096	1.76	0.12	132.33	0.0036	29.11
	8192	1.44	0.065	138.52	0.0039	167.81

Table 7 Robustness analysis of the two-level fusion framework combining second-generation wavelet and run-length coding

Parameter	Data size	CR	MSE	SNR	CT (s)	RT (s)
Eng2VibBBA	4096	7.47	0.046	137.88	0.42	0.053
	8192	13.33	0.085	135.49	0.53	0.075
	16,384	14.16	0.044	142.13	1.16	0.12
	24,576	16.57	0.030	146.06	0.87	0.091
	32,768	17.83	0.022	148.79	0.96	0.11
Eng2TPRCmd	4096	5.10	0.036	144.18	0.35	0.054
	8192	11.20	0.33	122.04	0.71	0.072
	16,384	21.60	0.92	111.61	1.52	0.12
	24,576	23.72	0.65	115.07	2.38	0.17
	32,768	25.79	0.51	117.54	3.15	0.19

The results show that the performance and robustness of the two-level fusion data compression and reconstruction framework combining second-generation wavelet and run-length coding has little change in compression ratio, reconstruction accuracy, signal-to-noise ratio, compression speed, and reconstruction speed, and has relatively strong robustness. Nevertheless, the two-level fusion data compression and reconstruction framework combining second-generation wavelet and Huffman coding and the two-level fusion data compression and reconstruction framework combining second-generation wavelet and run-length coding can meet the requirements of high compression ratio, high reconstruction accuracy, and real-time performance of aircraft health management.

Conclusion

In view of the characteristics of massive data, fuzziness, and real-time of data acquisition and transmission in the fuzzy information system faced by aircraft health management, to reduce the load of airborne data processing and transmission system under the condition of limited airborne computing resources and strong time constraints, the data collected by

Fig. 5 Robustness analysis of the two-level fusion data compression and reconstruction framework combining second-generation wavelet and run-length coding

the airborne system are first compressed, and the amount of data are reduced before transmission and reconstructed after transmission. In view of the situation that the compression ratio of primary data compression is too small and the compression time is too long for large-scale fuzzy systems to meet the transmission requirements of the system, this paper combines the advantages of lossy compression method which consumes less time and lossless compression method which has higher compression ratio; proposes a two-level data compression and reconstruction framework combining lossy compression and lossless compression; and studies a

two-level data compression and reconstruction framework combining second-generation wavelet and Huffman coding. The optimization analysis is carried out. Taking a real aero-engine health sample as an example, the simulation calculation is carried out, and the following conclusions are obtained:

1. Wavelet transform data compression and reconstruction represented by the second generation of wavelet transform not only has high compression ratio, but also has fast compression speed and reconstruction speed performance, which can meet the real-time requirements of large data processing in aircraft health management. Lossless compression and reconstruction represented by Huffman coding and run-length coding have better compression ratio performance than arithmetic coding and reconstruction speed performance is better than dictionary coding. Dictionary coding represented by LZ77 and LZ78 has a better compression ratio, but its reconstruction time increases exponentially with the increase of data scale, and its adaptability to big data is poor, which fails to meet the requirements of high efficiency big data processing in aircraft health management.
2. The two-level fusion data compression and reconstruction framework and algorithm combining second-generation wavelet and Huffman coding or run-length coding is an optimal-optimal combination, which is verified by an actual aero-engine health samples and compared with the data compression and reconstruction algorithm based on redundant sparse representation and compressed sensing. Effectiveness, scientificity, and robustness can meet the requirements of fast compression and efficient reconstruction of large data processing in aircraft health management.

However, for aircraft health management, the monitoring objects include rotating machinery such as aero-engine and propeller, whose vibration signals are often composed of harmonics, singularities, and noise, so the overall waveform have quasi-periodic characteristics. The redundancy of quasi-periodic signals can be divided into two categories: the one is the redundancy between sampling points contained in each cycle, i.e., the redundancy within a cycle. The other is the redundancy of sampling points between cycles, i.e., the redundancy between the cycles.

For this kind of data, if the original data are compressed only in one-dimensional space, the redundancy between sampling points in a cycle can be effectively eliminated, but it is not very effective for eliminating the redundancy between sampling points in the original signal cycles. A method that can effectively describe the redundancy of quasi-periodic signals is to transform the signals from one-dimensional space to two-dimensional space. First, the original signal sequence is

divided into a series of continuous and separate subsequences according to the cycle period of the cycle. Second, each subsequence is treated as each row of the two-dimensional matrix in sequence to generate a two-dimensional matrix. For the matrix, it still describes the redundancy in an original sequence cycle in a row direction, but each sample in the column direction corresponds to the sample whose interval is integer times of the cycle period in the original sequence.

Therefore, the two-dimensional description method of quasi-periodic signals provides a more compact signal characterization method, which can effectively characterize the redundancy within and between the cycles of the original quasi-periodic sequence, and thus can more conducive to signal compression processing. Therefore, the two-level data compression and reconstruction framework combining the second-generation wavelet fusion lossy compression and lossless compression method proposed in this paper still has a big expansion space, i.e., its performance optimization analysis can be studied in depth under the condition of two-dimensional data. In addition, the framework can be extended to three-level or multi-level data compression and reconstruction framework to further achieve higher compression ratio performance.

Funding The article was funded by China Aviation Science Foundation (Grant no. 20150267001).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Sweldens W (1996) The lifting scheme a custom-design construction of biorthogonal wavelets. *Appl Comput Harmon Anal* 3(2):186–200
- Sweldens W (1998) The lifting scheme: a construction of second-generation wavelets. *SIAM J Math Anal* 29(2):511–546
- Daubechies I, Swelens W (1998) Factoring wavelet transform into lifting steps. *J Fourier Anal Appl* 4(3):245–267
- Peng ZK, Chu FL (2004) Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. *Mech Syst Signal Process* 18(2):199–221
- Peng Z, He Y, Lu Q et al (2003) Feature extraction of the rub-impact rotor system by means of wavelet. *J Sound Vib* 259(4):1000–1010
- Li X, Qu L, Wen G et al (2003) Application of wavelet packet analysis for fault detection in electro-mechanical systems based on torsional vibration measurement. *Mech Syst Signal Process* 17(6):1219–1235
- Sun Q, Tang Y (2002) Singularity analysis using continuous wavelet transform for bearing fault diagnosis. *Mech Syst Signal Process* 16(6):1025–1041
- Altmann J, Mathew J (2001) Multiple band-pass autoregressive demodulation for rolling-element bearing fault diagnostics. *Mech Syst Signal Process* 15(5):963–977
- Staszewski WJ (1998) Wavelet based compression and feature selection for vibration analysis. *J Sound Vib* 211(5):735–760
- Staszewski WJ (1998) Identification of non-linear systems using multi-scale ridges and skeletons of the wavelet transform. *J Sound Vib* 214(4):639–658
- Mingming R, Yanbin Z, Lixin J (2006) Compression method based on lifting wavelet transform and embedded Zerotree coding. *J Xi'an Jiaotong Univ* 40(4):490–496
- Changyou Y, Qixun Y, Wanshun L (2005) A real-time data compression and reconstruction method based on lifting scheme. *Proc CSEE* 25(9):6–10
- Zhang YF, Li J (2006) Wavelet-based vibration sensor data compression technique for civil infrastructure condition monitoring. *J Comput Civ Eng* 20(4):390–399
- Jie Z (2017) Research on fault diagnosis of steam turbine units based on second-generation wavelet analysis. Ph.D dissertation, Tianjin University of Technology, Tianjin, PR. China
- Cheng L (2005) Data processing and process monitoring research based on lifting wavelet. Ph.D dissertation, Zhejiang University, Hangzhou, Zhejiang, PR. China
- Wang H, Zheng L (2001) A study on wavelet data compression of a real-time monitoring system for large hydraulic machines. *J Comput Sci Technol* 16(3):293–296
- Dongnan WU (2012) Data compression, 3rd edn. Electronic Industry Press, Beijing
- Li S, Shang J, Duan Z et al (2018) Fast detection method of quick response code based on run-length coding. *IET Image Proc* 12(4):546–551
- Stefanovich AI, Sushko DV (2017) Reversible data compression by universal arithmetic coding. *Inst Inform Probl* 3:20–45
- Han Y, Wang B, Yang X et al (2019) Efficient regular expression matching on LZ77 compressed strings using negative factors. *World Wide Web*. <https://doi.org/10.1007/s11280-019-00667-z>
- Arroyuelo D, Cánovas R, Navarro G et al (2017) LZ78 compression in low main memory space. In: Presented at the string processing and information retrieval of SPIRE 2017, Palermo, Italy, Sep 26–29
- Rui Z (2009) Research on mechanical fault signal processing method using second generation wavelet. Ph.D dissertation, Harbin Institute of Technology, Harbin, Heilongjiang, PR. China
- Junfeng G (2018) Method of compressed sensing for mechanical vibration signals based on double sparse dictionary model. *J Mech Eng*. <https://doi.org/10.3901/JME.2018.06.118>
- Jung A, Hulsebos M (2018) The network nullspace property for compressed sensing of big data over network. *Front Appl Math Stat*. <https://doi.org/10.3389/fams.2018.00009>
- Tang G, Ma JW, Yang HZ et al (2012) Seismic data denoising based on learning-type overcomplete dictionaries. *Appl Geophys* 9(1):27–32

26. Wang G, Zhao YN (2018) Design of compressed sensing algorithm for coal mine IoT moving measurement data based on a multi-hop network and total variation. *Sensors* 18:1732–1749
27. Wang X, Li G, Wan Q et al (2017) Look-ahead hybrid matching pursuit for multipolarization through-wall radar imaging. *IEEE Trans Geosci Remote Sens* 20:1–10
28. Feng X, Zhang X, Liu C et al (2017) Single-channel and multi-channel orthogonal matching pursuit for seismic trace decomposition. *J Geophys Eng* 14:90–99
29. Shi MM, Li L, Xu JM (2018) Backtracking stagewise weak orthogonal matching pursuit algorithm based on fuzzy threshold. *Video Eng* 42(2):5–9

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.