CrossMark

A supervised sparsity-based wavelet feature for bearing fault diagnosis

Cong Wang[1] · Meng Gan[1] · Chang'an Zhu[1]

Abstract This paper proposes a supervised sparsity-based wavelet feature (SSWF) for the detection of bearing fault, which combines wavelet packet transform (WPT) and sparse coding. SSWF is extracted from vibration signals by four main steps: (1) construct a WPT vector using the fault-related WPT coefficients; (2) design a structured dictionary that combines the signal characteristics and class information; (3) use the dictionary to implement the sparse coding of the WPT vectors, which can be solved by basis pursuit (BP) and (4) calculate the SSWF from the sparse coefficients. During the process, WPT can detect the fault occurrence of the bearing signal. Sparse coding based on a structured dictionary can find a robust representation of the signal and at the same time, integrate the class information. Therefore, SSWF is able to stably and discriminatively reflect different fault types, which indicates its potential in bearing fault diagnosis. Experiments on two bearing cases are conducted to verify the advantages of SSWF in the detection of bearing faults.

Keywords Wavelet packet transform (WPT) · Sparse coding · Structured dictionary · Supervised sparsity-based wavelet feature (SSWF) · Machinery fault diagnosis

✉ Cong Wang
wcc12345@mail.ustc.edu.cn

Meng Gan
ganmeng@mail.ustc.edu.cn

Chang'an Zhu
changan@ustc.edu.cn

[1] Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China (USTC), Hefei 230027, Anhui, People's Republic of China

Introduction

Fault diagnosis and prognosis have received considerable attentions because of their importance in preventing unexpected machinery breakdowns (Mortada et al. 2014; Aydin et al. 2015; Shukla et al. 2015). Bearing defects are major factors for a number of machinery failure in modern industry, which have been widely studied in recent years. Vibration signal based data-driven technology is the most commonly used method for bearing fault diagnosis. To effectively monitor and recognize the bearing condition, the challenge lies in extracting reliable features from vibration data that are usually disturbed by environment noise.

Time–frequency transform is a useful tool for extracting bearing fault features because it provides a synthetical analysis of time and frequency information. Applying time–frequency analysis appropriately is able to discover the changing frequency knowledge over time. For rotating machines, this information is also an effective tool to reveal the periodic transient pulses resulting from bearing faults. A number of time–frequency transforms have been developed, including the short-time Fourier transform (Chandra and Sekhar 2016), Hilbert–Huang transform (Siracusano et al. 2016) and wavelet transform (Li et al. 2015a, b). Among these methods, wavelet transform is able to conduct a multi-resolution analysis for signals with transient impulses. As a result, it is more powerful than the others in bearing fault diagnosis (Yan et al. 2014).

Wavelet packet transform (WPT) is a widely used wavelet method, which has some quite attractive characteristics, such as orthogonal, complete, and local properties (Coifman and Wickerhauser 1992). Based on the WPT coefficients, a number of feature extraction methods are proposed to identify different fault types of rolling element bearings, such as the sub-frequency energy (Zarei and Poshtan 2007), the Kurto-

sis value (Li et al. 2008) and the Rényi entropy (Bokoski and Juricic 2012). These features are extracted in a certain pattern to detect signal's energy distribution or transient impulses. Nonetheless, they cannot comprehensively reveal the structure of the signal, thus showing less effectiveness for some modern machinery signals with various interference.

Sparse coding is a popular topic in recent years and has received much attention in signal processing (Pavlidi et al. 2013), image processing (Yeganli et al. 2015) and computer vision (Sui et al. 2015). A number of state-of-art fault diagnosis algorithms are proposed based on this novel signal representation technique (He et al. 2016; Zhang et al. 2016). The basic principle of sparse coding is that signal can be expressed by a few columns, which are also called atoms, from a specific dictionary. In most cases, two main steps are needed to finish the representation, defined as dictionary construction and coefficient solving. A direct way to design a dictionary is using existing orthogonal basis, such as discrete cosine transform basis and wavelets. Another way is to learn a dictionary from the data, which is more complicate but can better meet the requirements of the processing data. Representative dictionary learning algorithms include the locality constrained linear coding (Rahmani et al. 2016) and k-singular value decomposition (Aharon et al. 2006). In this study, we intended to extract fault features in a supervised way, which means the dictionary should capture both data character and class information. An efficient and effective method is to organize the training samples in a class-guided order to construct a structured dictionary as employed in Wright's research (Wright et al. 2009). After the dictionary design, we can achieve the sparse coefficients by any of the two algorithms, including matching pursuit (MP) (Mallat and Zhang 1993) and basis pursuit (BP) (Chen et al. 2001). MP is famous for its fast convergence rate, whereas BP has the benefits of super-resolution and better sparsity. In this study, BP will be adopted to achieve more accurate results.

Inspired by WPT and sparse representation theory, we propose the supervised sparsity-based wavelet feature (SSWF) for the diagnosis of rolling element bearings. The novel feature is extracted mainly by the following steps: (1) construct a WPT vector using the fault-related WPT coefficients; (2) design a structured dictionary that combines the signal characteristics and class information; (3) use the dictionary to implement the sparse coding of the WPT vectors, which can be solved by BP and (4) calculate the SSWF from the sparse coefficients. Details about the proposed feature extraction method will be further described in "Method for extracting SSWFs" section. As seen from the feature extraction process, SSWF benefits from three aspects. WPT can detect the fault occurrence of bearing signals. The structured dictionary integrates the class information with the signal characters, rendering a supervised feature extraction process. Sparse coding represents the signal in a robust way. These advantages make SSWF representative and discriminative, thereby exhibiting valuable properties in bearing fault recognition.

The remainder of this paper is composed of the following parts. "Brief review of WPT" and "Sparse coding" section review the WPT method and sparse representation theory, respectively. "Method for extracting SSWFs" section describes the detail process for extracting the proposed SSWF. "Experiment results" section using two experiments on rolling element bearings to verify the advantages of SSWF. "Conclusions" section provides the conclusions of the whole study.

Brief review of WPT

WPT has a convenient implementation in practice with the help of a low-pass filter $h(k)$ and a high-pass filter $g(k)$. The filters are designed based on $\psi(t)$ and $\phi(t)$ by Eq. (1), which are called the wavelet function and scaling function, respectively (Mallat 1989).

$$\phi(t) = \sqrt{2} \sum_k h(k) \phi(2t - k)$$
$$\psi(t) = \sqrt{2} \sum_k g(k) \phi(2t - k), \tag{1}$$

where $\sum_k h(k) = \sqrt{2}$ and $\sum_k g(k) = 0$. An intuitive demonstration of the WPT results is illustrated in Fig. 1. The mathematical expression is

$$d_{j+1,2n} = \sum_m h(m - 2k) d_{j,n}$$
$$d_{j+1,2n+1} = \sum_m g(m - 2k) d_{j,n}, \tag{2}$$

where j represents the decomposition level, n indicates the node index, and m denotes the coefficient length.

Sparse coding

Signal sparse representation

The definition of sparse representation explains that an input $\mathbf{x}$ can be represented by a dictionary $\mathbf{A}$ as Eq. (3),

$$\mathbf{x} = \mathbf{As} = \sum_{j=1}^n \mathbf{a}^j s^j, \tag{3}$$

where $\mathbf{s} = [s^1, s^2, \ldots, s^n]^T$ are the coefficients for the dictionary. The sparsest coefficient set can be obtained by searching for the solution with the minimum nonzero ele-
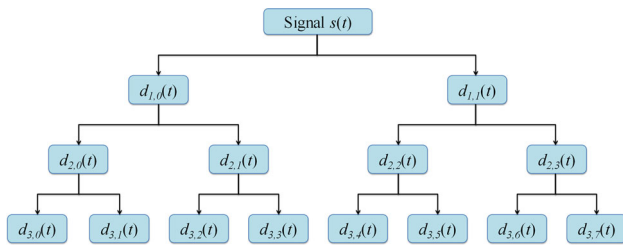
Fig. 1 Demonstration of a 3-level WPT

ments number. Mathematically, the optimization can be written as

$$\arg\min_{\mathbf{s}} \|\mathbf{s}\|_0, \quad s.t. \quad \mathbf{x} = \mathbf{As}, \tag{4}$$

where $\|\cdot\|_0$ is the l_0 norm, representing the number of nonzero elements in a vector.

Coefficients solving

Solving Eq. (4) exactly is difficult because of the NP-hard problem (Amaldi and Kann 1998). In practice, people change the l_0 optimization to Eq. (5) to seek for an approximate solution.

$$\arg\min_{\mathbf{s}} \|\mathbf{s}\|_1, \quad s.t. \quad \mathbf{x} = \mathbf{As}, \tag{5}$$

$\|\mathbf{s}\|_1$ indicates the l_1 norm defined by Eq. (6).

$$\|\mathbf{s}\|_1 := \sum_{n=0}^{N-1} |s(n)|. \tag{6}$$

Equation (5) is called BP (Chen et al. 2001), which is an effective method for solving sparse coefficients. Generally, the vibration signal of mechanical equipment measured from sensors inevitably contains some noise, which makes the solution to the basis pursuit denoising (BPD) problem (Eq. (7)) is more useful in practise (Gunn et al. 2002),

$$\arg\min_{\mathbf{s}} \|\mathbf{x} - \mathbf{As}\|_2^2 + \lambda \|\mathbf{s}\|_1, \tag{7}$$

where $\|\mathbf{x}\|_2^2 := \sum_{n=0}^{N-1} |x(n)|^2$, and λ is the Lagrange multiplier.

The BP and BPD problems can be solved by several algorithms, like the gradient projection sparse reconstruction (GPSR) (Figueiredo et al. 2007), interior-point method (Kim et al. 2007), fast iterative-shrinkage thresholding algorithm (FISTA) (Beck and Teboulle 2009) and split variable augmented Lagrangian shrinkage (SALSA) (Afonso et al. 2010). In this study, the SALSA algorithm is employed because of its efficiency and simplicity. In the SALSA framework, the

unconstrained problem in Eq. (7) is replaced by a constrained optimization [(Eq. (8)] using variable splitting,

$$\arg\min_{\mathbf{s},\mathbf{u}} \|\mathbf{x} - \mathbf{Au}\|_2^2 + \lambda \|\mathbf{u}\|_1, \quad s.t. \quad \mathbf{s} - \mathbf{u} = 0, \tag{8}$$

where $\mathbf{u}$ represents a created new variable. This problem can be solved using the following update procedure (Mateos et al. 2010).

$$\mathbf{u}^{(k+1)} = \arg\min_{\mathbf{u}} \lambda \|\mathbf{u}\|_1 + \frac{\mu}{2}\|\mathbf{u} - \mathbf{s}^{(k)} - \mathbf{d}^{(k)}\|_2^2,$$

$$\mathbf{s}^{(k+1)} = \arg\min_{\mathbf{s}} \frac{1}{2}\|\mathbf{x} - \mathbf{As}\|_2^2 + \frac{\mu}{2}\|\mathbf{u}^{(k+1)} - \mathbf{s} - \mathbf{d}^{(k)}\|_2^2,$$

$$\mathbf{d}^{(k+1)} = \mathbf{d}^{(k)} - (\mathbf{u}^{(k)} - \mathbf{s}^{(k)}). \tag{9}$$

k denotes the iteration index and μ represents the penalty parameter.

Method for extracting SSWFs

Features extracted from WPT coefficients are widely used in machinery fault diagnosis because WPT is able to detect fault-related impulses embedded in bearing signals (Gao and Yan 2006). Nevertheless, these features are lack of class label information, which is quite helpful in fault identification. To improve the recognition accuracy, this study proposes the SSWF obtained from sparse coding with WPT-based structure dictionaries (i.e. the dictionary atoms have correspondence to the class labels), which indicates the new feature can combine the class label and the signal characteristics, thus providing more discriminative information. Our method includes the following parts: WPT-based band selection and SSWF extraction.

WPT frequency band selection

Considering a set of WPT subbands $\{\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \ldots \mathbf{w}^{(m)}\}$, the average energy of $\mathbf{w}^{(i)}$ can be expressed as

$$e(i) = \frac{E_i}{E_{\text{total}}} = \frac{\left(\sum_{j=1}^{l} w^{(i)}(j)^2\right)/l}{\sum_{i=1}^{m}\left(\left(\sum_{j=1}^{l} w^{(i)}(j)^2\right)/l\right)}, \tag{10}$$

where, l is the subband length. The boundary of the selected subbands can be determined when $e(i)$ attenuates from the peak to a threshold for the first time. The threshold is set at 5 % in the current study. Considering the case shown in Fig. 2, the left boundary is 7 and the right boundary is 10, which indicates subbands 7–10 contains the major vibration information of the signal. In practice, the final range should be determined by the union of all fault-related subbands to
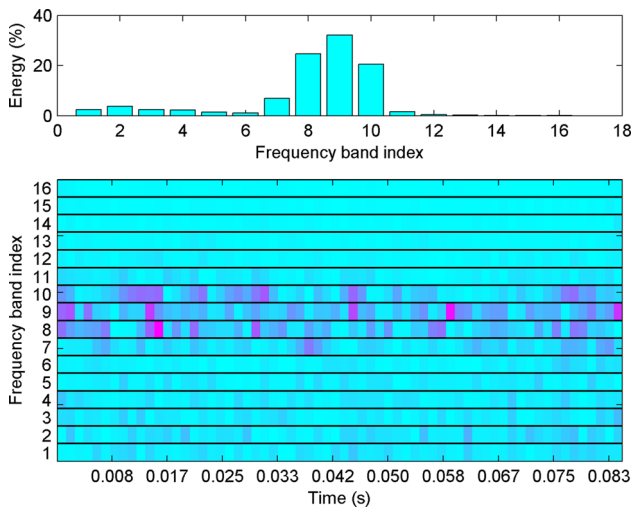
Fig. 2 Energy distribution of a WPT spectrum

comprehensively present useful frequency information. The selected WPT bands of a signal are arranged as a column vector as:

$$\mathbf{x} = [\mathbf{w}_s^1, \mathbf{w}_s^2, \ldots, \mathbf{w}_s^k]^T, \tag{11}$$

where $\mathbf{w}_s^i$ ($i = 1, 2, \ldots, k$) are subbands selected from the signal.

SSWF calculation

To obtain the SSWF, a structured dictionary need to be constructed based on WPT signatures associated with the class labels. Supposing we have c classes of subjects, the dictionary $\mathbf{A}$ is designed as:

$$\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_c], \tag{12}$$

where $\mathbf{A}_i$ is the subset of the training WPT vectors from class i. For a selected WPT vector $\mathbf{x}$ from a signal, the sparse representation $\hat{\mathbf{s}}$ can be solved by Eq. (6). Then, $\mathbf{x}$ can be represented as:

$$\mathbf{x} \approx \mathbf{A}\hat{\mathbf{s}} = \mathbf{A}_1\mathbf{s}_1 + \mathbf{A}_2\mathbf{s}_2 + \cdots + \mathbf{A}_c\mathbf{s}_c \tag{13}$$

More intuitively, the eqnarray can be illustrated in Fig. 3. It can be seen that $\mathbf{x}$ can be approximately described by the linear combination of atoms and coefficients. A larger coefficient indicates the atom plays more important role in representing the original signal. Ideally, most of the coefficients are zero except those associated with a specific class, which the signal belongs to. In that case, the coefficients will have the ability to discard the unrelated data and preserve the class information. Accordingly, we propose a novel feature formulated by the sum of absolute value of the coefficients
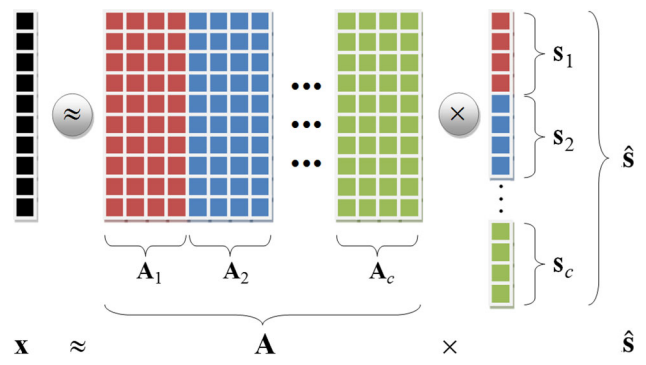

Fig. 3 Sparse representation of a signal by a structured dictionary

corresponding to a certain class, which can be regarded as an indicator for the similarity between the signal and class. Mathematical description is presented in Eq. (14).

$$f_i = \sum_{j=1}^{n_i} |\mathbf{s}_i(j)| = \|\mathbf{s}_i\|_1 \tag{14}$$

Combination of f_i for all the classes is the SSWF proposed in this study.

$$\text{SSWF} = [f_1, f_2, \ldots, f_c] \tag{15}$$

The whole process can be summarized in Fig. 4. SSWF has the following merits in theory: (1) WPT helps find the fault impulses of the signal; (2) the structured dictionary integrates the class information with the signal characters, rendering a supervised feature extraction process and (3) sparse coding represents the signal in a robust way. These advantages make SSWF outperform most traditional WPT features in bearing fault diagnosis, which will be proved by the following engineering case.

Experiment results

We verify the performance of SSWF on two bearing cases. Features such as WPT subband energy, Kurtosis and Rényi entropy are also calculated as a comparison.

Bearing case I

Data instruction

Firstly, we evaluate the SSWF using a benchmark bearing dataset from the Case Western Reserve Lab (CWRU) (Lei et al. 2008; Ding et al. 2015). The experimental setup is shown in Fig. 5. Signals are collected by accelerometers with a 12 kHz sampling frequency from the testing bearing, which is located at the drive end. Bearings with 0.007-in. single point faults at the outer raceway, inner raceway and
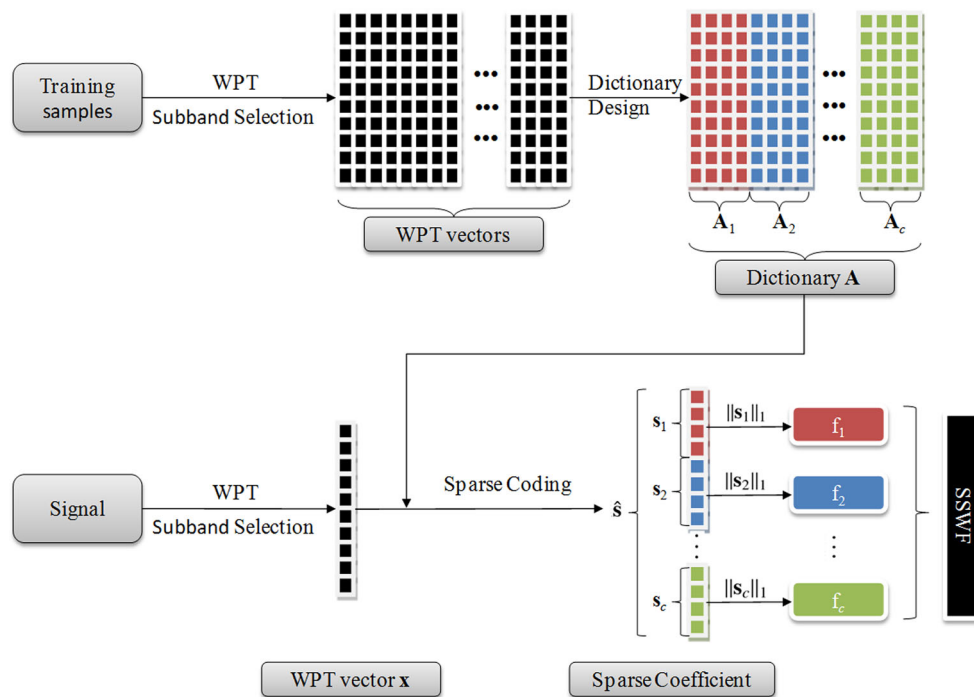
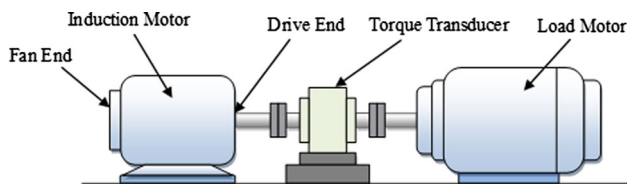**Fig. 4** Procedure of SSWF extraction method

Fig. 5 Experimental apparatus in bearing case I

the rolling-elements are collected under four different motor speeds. In total, three defective states and one healthy state are used for the evaluation of SSWF. Typical waveforms of different fault types are illustrated in Fig. 6. We collected 400 samples for each bearing health condition under different motor speeds. The sample length is set at 1024. These 400 samples are randomly split into training and testing datasets, each containing 200 samples. Considering there are 4 health states, we will finally obtain a training dataset with 800 samples and a testing dataset with the same size.

Bearing case I subband selection

The WPT decomposition level for this case is set at 4 and the corresponding results are demonstrated in Fig. 7. Energy is concentrated in subbands 7–10 for rolling-element defect, 7–11 for inner-raceway and outer-raceway fault. Accordingly ("WPT frequency band selection" section), the final subband range is 7–11 for keeping the fault-related information as much as possible.
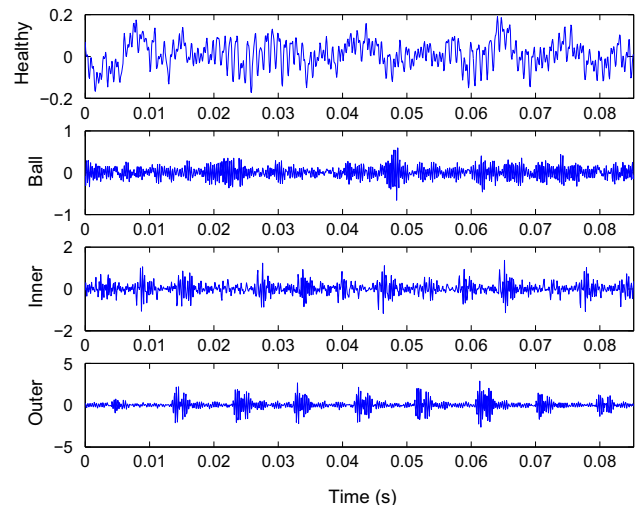

Fig. 6 Waveforms for different fault types in bearing case I

SSWFs for bearing case I

Based on the description in "SSWF calculation" section, we construct the dictionary using the fault-related WPT subbands of the training samples. Considering the 1024-point sample signal, the data length of each subband should be $1024/2^4 = 64$. The 5 selected subbands form a WPT vector with $64 \times 5 = 320$ points [Eq. (11)], thus achieving a 320×800 dictionary **A**. The sparse coefficients can be solved by SALSA [Eq. (9)] with $\lambda = 0.04$, $\mu = 2$. Based on
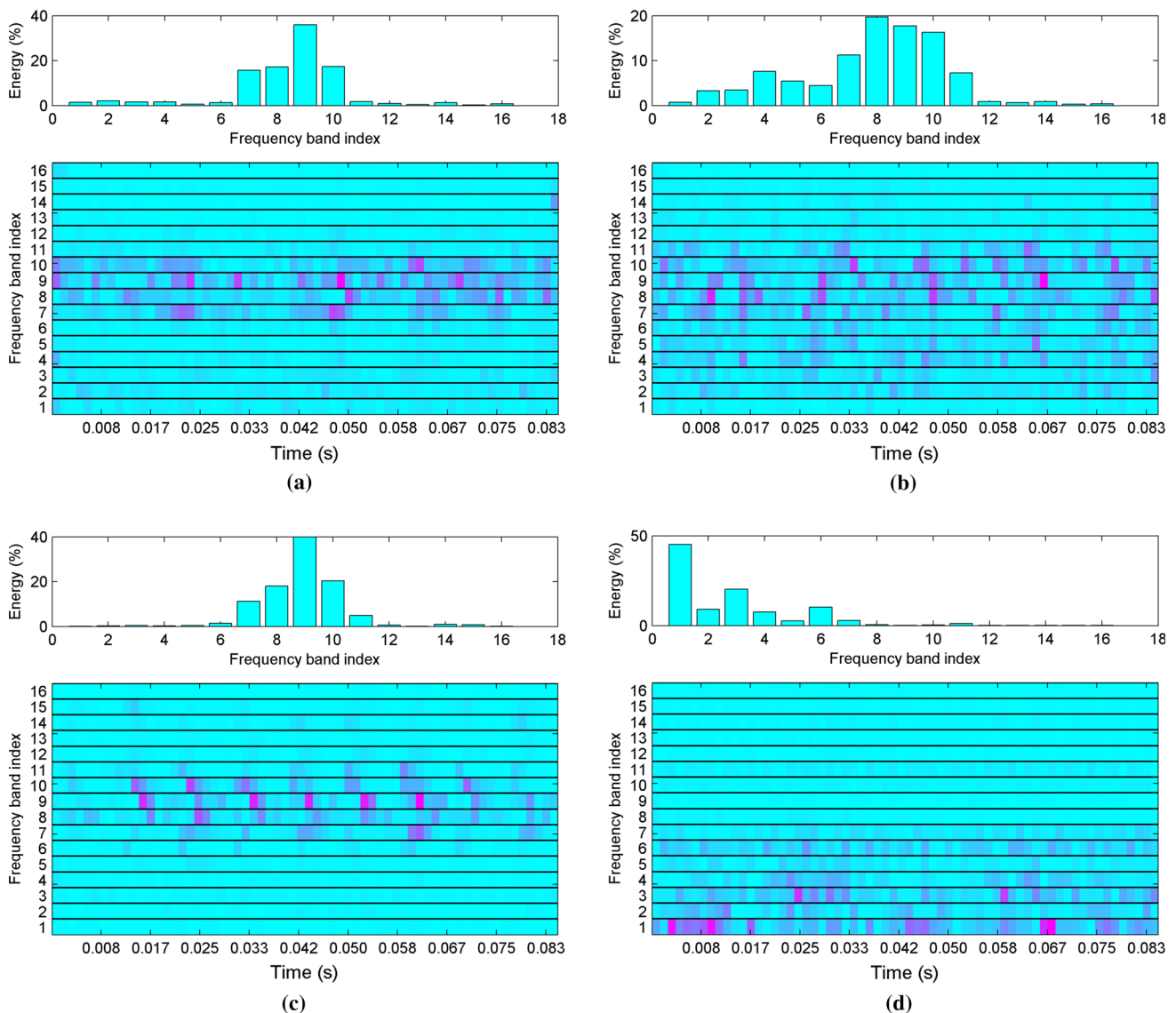
Fig. 7 WPT results of four health conditions in bearing case I: **a** rolling-elements defect; **b** inner-raceway defect; **c** outer-raceway defect and **d** healthy

the coefficients, SSWF is obtained by Eq. (15). The dimension of SSWF is coincident with the class number $c = 4$. Intuitive demonstration of each dimension of SSWFs for the training samples can be found in Fig. 8. It is clearly presented that SSWF has quite regular distribution. Each dimension highlights a certain health state, which indicates the SSWF can be regarded as an indicator of the similarity between the sample signal and a certain class in the structured dictionary. Therefore, with a well construct dictionary and a stable representation (sparse coding), SSWF can definitely reveal the internal relationship of the signal and different fault types, thus achieving a successful fault recognition.

Classification evaluation

The final diagnosis result is presented in this part. For a comprehensive evaluation, three widely used WPT-based features, including subband energy (Zarei and Poshtan 2007), Kurtosis (Li et al. 2008) and Rényi entropy (Bokoski and Juricic 2012), are calculated as a comparison. The classification is conducted by the simple nearest neighbor classifier (Onan 2016), which implement the classification based on sample distance in the feature space.

As discussed before ("Data instruction" section), the training and testing dataset are randomly constructed. To make the result more convincing, the classification is conducted
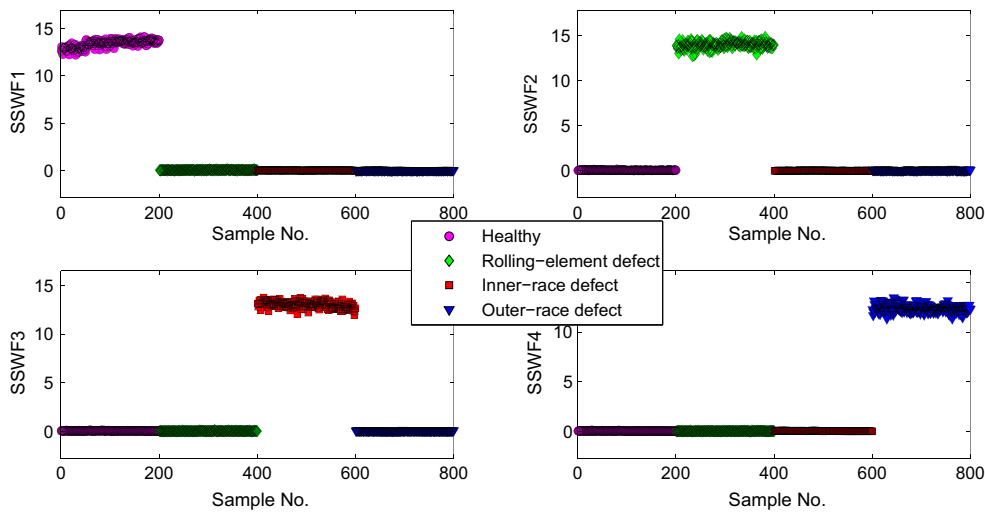
Fig. 8 SSWF features for bearing case I

Table 1 Classification error rates (mean ± standard-deviation %) for different features of bearing case I

Item	SSWF	Subband energy	Kurtosis	Rényi entropy
Healthy	0	0	0	0.06 ± 0.03
Rolling-element	0.06 ± 0.13	0.51 ± 0.22	0.74 ± 0.23	0.36 ± 0.33
Inner-raceway	0.1 ± 0.08	0.21 ± 0.14	1.27 ± 0.38	0.19 ± 0.11
Outer-raceway	0.15 ± 0.12	0.59 ± 0.18	1.35 ± 0.52	0.4 ± 0.19
Overall	0.43 ± 0.33	1.31 ± 0.27	3.36 ± 0.55	1.01 ± 0.44

Table 2 Wilcoxon test for the classification result of bearing case I

Item	SSWF > Subband energy	SSWF > Kurtosis	SSWF > Rényi entropy
Significance level	3.97×10^{-4}	3.97×10^{-4}	0.0043

repeatly for 50 times using different randomly constructed dataset. The recognition error rates are reported in Table 1, which tells that the SSWF is superior to the other three features with the lowest error rate in the recognition of all kinds of fault types. Besides, to conduct a more convincing evaluation, we use the Wilcoxon test (Wilcoxon 1945) to check the significance of the result as listed in Table 2. The significance level of all the comparisons is lower than 0.05 %, which further confirms the proposed SSWF performs better than the subband energy, Kurtosis and Rényi entropy features.

Bearing case II

Data instruction

To highlight the advantages of the SSWF, another set of bearings was experimented using the apparatus in Fig. 9. Different defects are set in the test bearings as illustrated in Fig. 10. During the experiment, the motor speed was set at 900, 1200, 1350 and 1450 rpm for each test. Signals are collected by accelerometers with the sample frequency set at 10 kHz. Figure 11 presents waveforms for different health states. We can see that noise corruption of this case is heavier than bearing case I, especially for the rolling-element defect with weak regularity of fault-induced impulsive information. The training and testing datasets are formed by the approach similar to Bearing case I. Each dataset contains 800 samples with the signal length set at 1024.

Bearing case II subband selection

This set of signals are decomposed by WPT in three levels, resulting in $2^3 = 8$ subbands with 128 points each. The WPT results of signals under different health conditions are demonstrated in Fig. 12, which shows the fault-related subbands are roughly in the range 3–8.

SSWFs for bearing case II

Realign the 6 selected subbands to construct a 768-point WPT vector, thus achieving a 768×800 dictionary **A**. The sparse coefficients are solved by SALSA [Eq. (9)] with

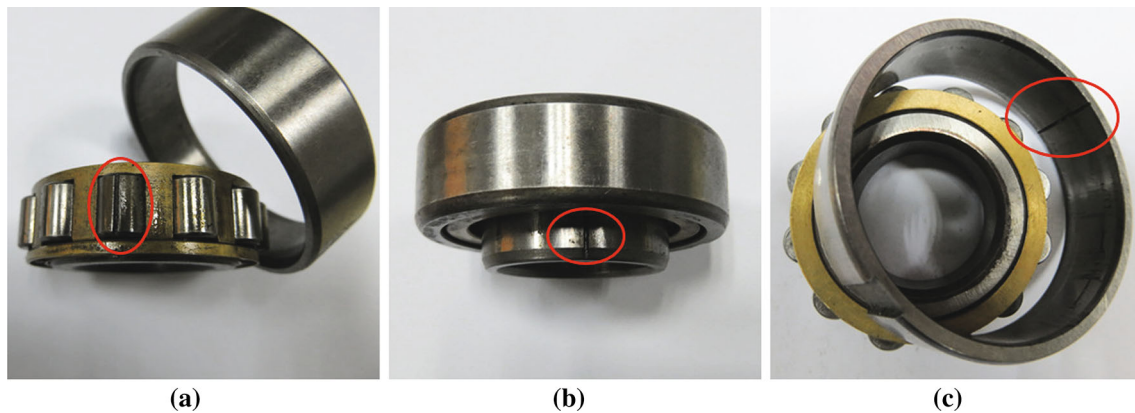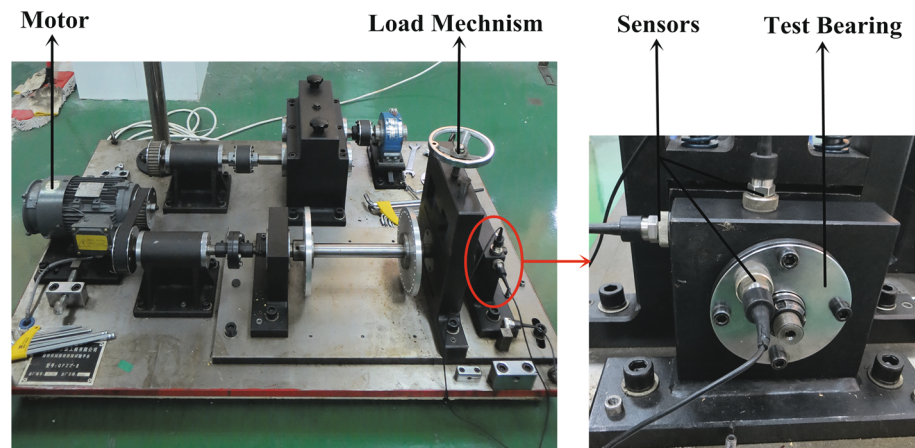**Fig. 9** Experimental apparatus in bearing case II

Fig. 10 Defectives in bearing case II: **a** rolling-element defect; **b** inner-raceway defect and **c** outer-raceway defect
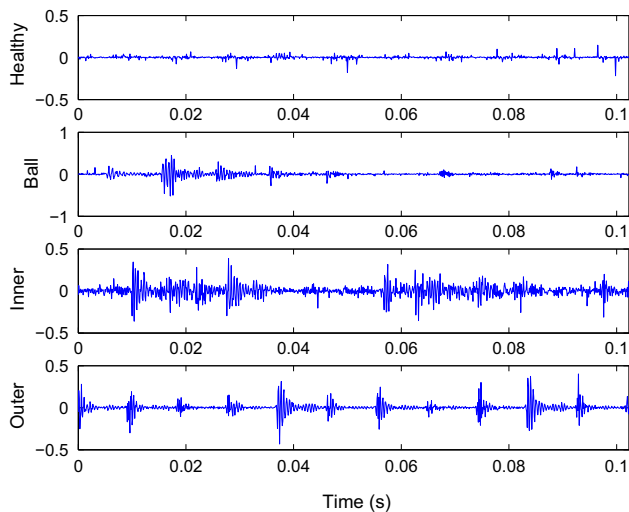

Fig. 11 Waveforms for different fault types in bearing case II

$\lambda = 0.025$, $\mu = 2.5$. Based on the coefficients, SSWF is obtained by Eq. (15). The dimension of SSWF is also 4 in this case and intuitive demonstration of each SSWF dimension for the training samples can be found in Fig. 13. This

distribution is still with strong regularity like bearing case I even under heavier noise corruption. Features from different dimensions separate one health class from the others and the variance within one class is small enough to achieve an excellent fault classification. Again, this feature illustrates the merits of strong stability and discriminability.

Classification evaluation

Again, the SSWF is compared with other traditional features (subband energy, Kurtosis and Rényi entropy) using the nearest neighbor classifier. The training and testing dataset are still randomly constructed for 50 times as described in "Classification evaluation" section. The classification error rates are summarized in Table 3. In comparison, the SSWF is still more effective than the others. which reaches the lowest overall error rates 0.75 %. As for the recognition for a certain fault type, SSWF performs the best except for the diagnosis of outer-raceway defect (the Rényi entropy makes no mistake). The Wilcoxon test result is shown in Table 4 with all the significance level lower than 0.05 %, which again proves the superiority of SSWF in detecting bearing faults.
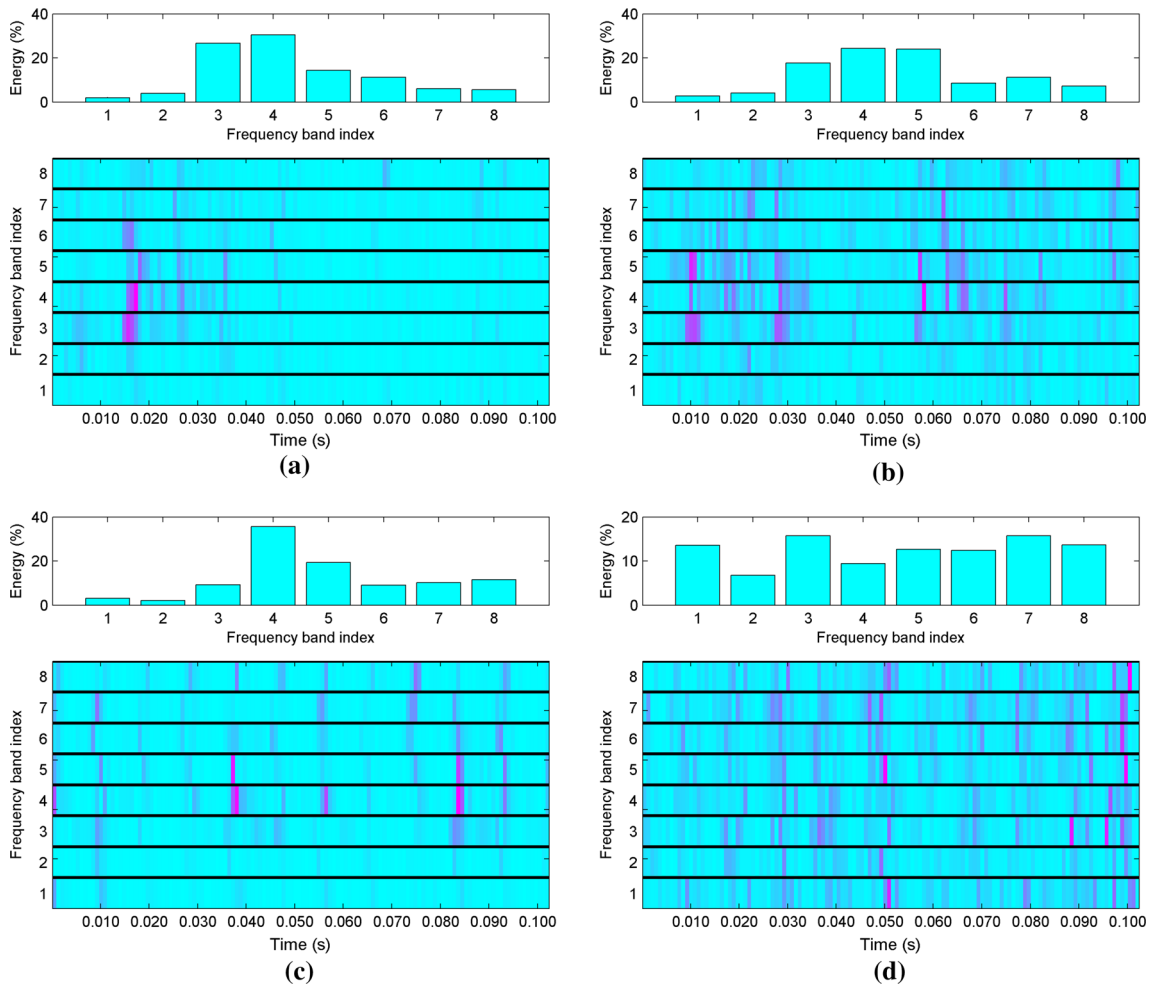
Fig. 12 WPT results of four health conditions in bearing case II: **a** rolling-elements defect; **b** inner-raceway defect; **c** outer-raceway defect and **d** healthy
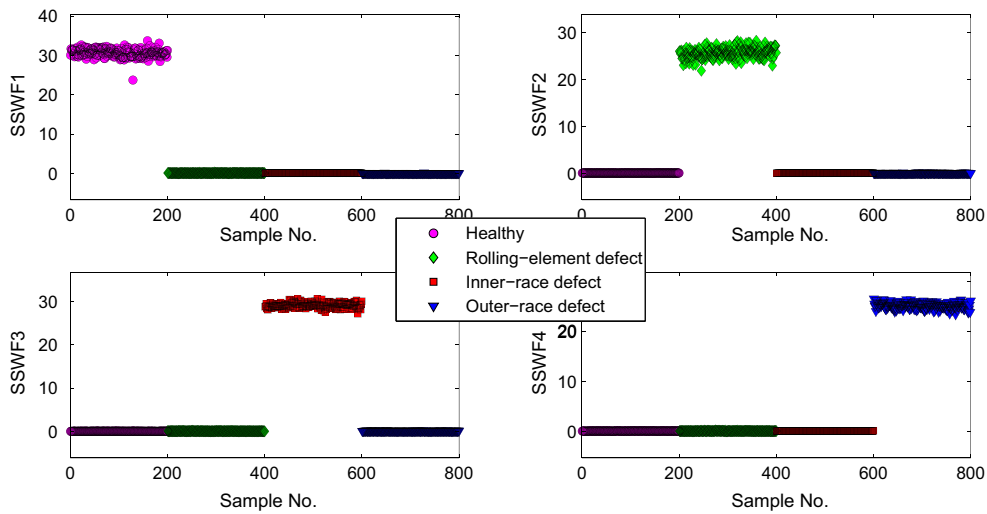

Fig. 13 SSWF features for bearing case II

Table 3 Classification error rates (mean ± standard-deviation %) for different features of bearing case II

Item	SSWF	Subband energy	Kurtosis	Rényi entropy
Healthy	0	0	1.07 ± 0.54	0.51 ± 0.22
Rolling-element	0.36 ± 0.28	0.62 ± 0.28	0.93 ± 0.30	0.45 ± 0.15
Inner-raceway	0.26 ± 0.21	0.31 ± 0.09	0.91 ± 0.39	0.3 ± 0.12
Outer-raceway	0.13 ± 0.1	0.45 ± 0.16	0.15 ± 0.13	0
Overall	0.75 ± 0.3	1.38 ± 0.18	3.06 ± 0.83	1.26 ± 0.22

Table 4 Wilcoxon test for the classification result of bearing case II

Item	SSWF > Subband energy	SSWF > Kurtosis	SSWF > Rényi entropy
Significance level	2.97×10^{-4}	1.77×10^{-4}	0.0012

Discussions

This study proposes the SSWF for mechanical fault diagnosis, which can preserve the intrinsic vibration pattern of the original signal and sparsely represents it in a robust way. The SSWF can accurately reveal signal fault pattern by activating the atoms from a specific class of the dictionary and discard useless ones. The excellent performance in experiments shows the stability and discriminability of SSWF, which can be observed in Figs. 8 and 13 as well as the classification accuracy of two cases.

The feature extraction algorithm can be conducted quite efficiently. Take bearing case I for example, it takes about 0.01 s to extract the SSWFs on a workstation with dual-core 2.70 GHz Athlon processors and 8 GB of memory. This can meet the requirement of real-time diagnosis. Besides, SSWF is more robust to disturbance and has higher recognition rate, which makes it quite competitive in modern machinery diagnosis.

In practical diagnosis of complex modern machines, more advanced classifiers can be utilized for classification instead of the nearest neighbor classifier, which is intended to highlight the advantages of SSWF. Classifiers such as neural network (Zhang et al. 2013; Wang and Cui 2013) and support vector machine (Missoum et al. 2014) are more powerful than the nearest neighbor classifier in most cases, which can help improve the recognition accuracy.

Conclusions

A novel supervised feature (SSWF) was proposed in this paper for bearing fault diagnosis using WPT and sparse coding. SSWF is calculated from the sparse coefficients using a specifically designed structured dictionary, which integrates signal characteristics and class information. For a certain sample, only the atoms related with the corresponding class in the dictionary can be activated. Accordingly, SSWF can reflect the internal relationship between the sample signal and different classes in the structured dictionary and highlights a certain health state. The clear and regular distribution of SSWF illustrates its effectiveness in bearing fault diagnosis. Moreover, the bearing experiments of SSWF in identifying different fault types further proves its superiority, which shows higher recognition rate than the other traditional features, including WPT subband energy, Kurtosis and Rényi entropy. All these advantages make SSWF attractive and competitive for the detection of bearing fault in modern industry.

References

Afonso, M. V., Bioucas-Dias, J. M., & Figueiredo, M. A. T. (2010). Fast image recovery using variable splitting and constrained optimization. *IEEE Transactions on Image Processing*, *19*(9), 2345–2356.

Aharon, M., Elad, M., & Bruckstein, A. (2006). K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, *54*(11), 4311–4322.

Amaldi, E., & Kann, V. (1998). On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. *Theoretical Computer Science*, *209*(1–2), 237–260.

Aydin, I., Karakose, M., & Akin, E. (2015). Combined intelligent methods based on wireless sensor networks for condition monitoring and fault diagnosis. *Journal of Intelligent Manufacturing*, *26*(4), 717–729.

Beck, A., & Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, *2*(1), 183–202.

Bokoski, P., & Juricic, D. (2012). Fault detection of mechanical drives under variable operating conditions based on wavelet packet Renyi entropy signatures. *Mechanical Systems and Signal Processing*, *31*, 369–381.

Chandra, N. H., & Sekhar, A. S. (2016). Fault detection in rotor bearing systems using time frequency techniques. *Mechanical Systems and Signal Processing*, *72–73*, 105–133.

Chen, S. S. B., Donoho, D. L., & Saunders, M. A. (2001). Atomic decomposition by basis pursuit. *Siam Review*, *43*(1), 129–159.

Coifman, R. R., & Wickerhauser, M. V. (1992). Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, *38*(2), 713–718.

Ding, X. X., He, Q. B., & Luo, N. W. (2015). A fusion feature and its improvement based on locality preserving projections for rolling element bearing fault classification. *Journal of Sound and Vibration*, *335*, 367–383.

Figueiredo, M. A. T., Nowak, R. D., & Wright, S. J. (2007). Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing*, *1*(4), 586–597.

Gao, R. X., & Yan, R. (2006). Non-stationary signal processing for bearing health monitoring. *International Journal of Manufacturing Research*, *1*(1), 18–40.

Gunn, R. N., Gunn, S. R., Turkheimer, F. E., Aston, J. A. D., & Cunningham, T. J. (2002). Positron emission tomography compartmental models: A basis pursuit strategy for kinetic modeling. *Journal of Cerebral Blood Flow and Metabolism*, *22*(12), 1425–1439.

He, W. P., Ding, Y., Zi, Y. Y., & Selesnick, I. W. (2016). Sparsity-based algorithm for detecting faults in rotating machines. *Mechanical Systems and Signal Processing*, *72–73*, 46–64.

Kim, S. J., Koh, K., Lustig, M., Boyd, S., & Gorinevsky, D. (2007). An interior-point method for large-scale l(1)-regularized least squares. *IEEE Journal of Selected Topics in Signal Processing*, *1*(4), 606–617.

Lei, Y. G., He, Z. J., & Zi, Y. Y. (2008). A new approach to intelligent fault diagnosis of rotating machinery. *Expert Systems with Applications*, *35*(4), 1593–1600.

Li, F. C., Meng, G., Ye, L., & Chen, P. (2008). Wavelet transform-based higher-order statistics for fault diagnosis in rolling element bearings. *Journal of Vibration and Control*, *14*(11), 1691–1709.

Li, H. K., Lian, X. T., Guo, C., & Zhao, P. S. (2015). Investigation on early fault classification for rolling element bearing based on the optimal frequency band determination. *Journal of Intelligent Manufacturing*, *26*(1), 189–198.

Li, H. K., Wang, Y. H., Zhao, P. S., Zhang, X. W., & Zhou, P. L. (2015). Cutting tool operational reliability prediction based on acoustic emission and logistic regression model. *Journal of Intelligent Manufacturing*, *26*(5), 923–931.

Mallat, S. G., & Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing, 41*, 3397–3415.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition—the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(7), 674–693.

Mateos, G., Bazerque, J. A., & Giannakis, G. B. (2010). Distributed sparse linear regression. *IEEE Transactions on Signal Processing*, *58*(10), 5262–5276.

Missoum, S., Vergez, C., & Doc, J. B. (2014). Explicit mapping of acoustic regimes for wind instruments. *Journal of Sound and Vibration*, *333*(20), 5018–5029.

Mortada, M. A., Yacout, S., & Lakis, A. (2014). Fault diagnosis in power transformers using multi-class logical analysis of data. *Journal of Intelligent Manufacturing*, *25*(6), 1429–1439.

Onan, A. (2016). Classifier and feature set ensembles for web page classification. *Journal of Information Science*, *42*(2), 150–165.

Pavlidi, D., Griffin, A., Puigt, M., & Mouchtaris, A. (2013). Real-time multiple sound source localization and counting using a circular microphone array. *IEEE Transactions on Audio Speech and Language Processing*, *21*(10), 2193–2206.

Rahmani, H., Huynh, D. Q., Mahmood, A., & Mian, A. (2016). Discriminative human action classification using locality-constrained linear coding. *Pattern Recognition Letters*, *72*, 62–71.

Shukla, N., Ceglarek, D., & Tiwari, M. K. (2015). Key characteristics-based sensor distribution in multi-station assembly processes. *Journal of Intelligent Manufacturing*, *26*(1), 43–58.

Siracusano, G., Lamonaca, F., Tomasello, R., Garesci, F., La Corte, A., Carni, D. L., et al. (2016). A framework for the damage evaluation of acoustic emission signals through Hilbert–Huang transform. *Mechanical Systems and Signal Processing*, *75*, 109–122.

Sui, Y., Zhang, S. L., & Zhang, L. (2015). Robust visual tracking via sparsity-induced subspace learning. *IEEE Transactions on Image Processing*, *24*(12), 4686–4700.

Wang, G. F., & Cui, Y. H. (2013). On line tool wear monitoring based on auto associative neural network. *Journal of Intelligent Manufacturing*, *24*(6), 1085–1094.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80–83.

Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(2), 210–227.

Yan, R. Q., Gao, R. X., & Chen, X. F. (2014). Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Processing*, *96*, 1–15.

Yeganli, F., Nazzal, M., & Ozkaramanli, H. (2015). Image super-resolution via sparse representation over multiple learned dictionaries based on edge sharpness and gradient phase angle. *Signal Image and Video Processing*, *9*, 285–293.

Zarei, J., & Poshtan, J. (2007). Bearing fault detection using wavelet packet transform of induction motor stator current. *Tribology International*, *40*(5), 763–769.

Zhang, H., Chen, X. F., Du, Z. H., Li, X., & Yan, R. Q. (2016). Nonlocal sparse model with adaptive structural clustering for feature extraction of aero-engine bearings. *Journal of Sound and Vibration*, *368*, 223–248.

Zhang, Z. Y., Wang, Y., & Wang, K. S. (2013). Fault diagnosis and prognosis using wavelet packet decomposition, fourier transform and artificial neural network. *Journal of Intelligent Manufacturing*, *24*(6), 1213–1227.