

Intelligent fault diagnosis of rolling element bearings using sparse wavelet energy based on overcomplete DWT and basis pursuit

Cong Wang · Meng Gan · Chang'an Zhu

Received: 30 September 2014 / Accepted: 12 February 2015 / Published online: 17 February 2015
© Springer Science+Business Media New York 2015

Abstract This paper explores sparse time-frequency distribution (TFD) using overcomplete discrete wavelet transform (DWT) and sparse representation techniques. This distribution is discovered for characterizing the periodic transient information embedded in rolling element bearings and extracting effective features that can discriminate different fault conditions. Based on the sparse TFD, a new sparse wavelet energy (SWE) feature is obtained by three main steps: first, an overcomplete discrete DWT is employed to decompose the fault signal and construct a redundant dictionary; second, the redundant dictionary is optimized by basis pursuit to obtain the sparsest TFD; finally, SWE is calculated from the new TFD to produce a feature vector for each signal. SWE features that combine the merits of overcomplete DWT and sparse representation techniques can precisely reveal fault-induced information, thereby exhibiting valuable properties for automatic fault identification by intelligent classifiers. The effectiveness and advantages of the proposed features are confirmed by simulation and the practical fault pattern recognition of rolling bearings.

Keywords Overcomplete discrete wavelet transform (DWT) · Basis pursuit · Sparse wavelet energy (SWE) · Intelligent fault diagnosis

Introduction

Rapid development in modern industry has generated more complex machines. Condition monitoring and fault diagnosis for modern mechanical equipment is increasingly important to prevent economic loss and numerous researches have been conducted in this field (Zhang et al. 2013; Wells et al. 2013; Yu et al. 2014). Defects in rolling element bearings are major factors in machinery failure and have elicited considerable attention. Bearing fault diagnosis is conducted by data acquisition, feature extraction and intelligent classification. Data acquisition and intelligent classification are relatively easy to implement. Typical method to collect data from a mechanical system is using accelerometers attached to the machines, and several widely used algorithms have been employed to create the intelligent classifier, such as the k -nearest neighbor (Gharavian et al. 2013), fisher discriminant analysis (Jiang et al. 2013), artificial neural network (Wang and Cui 2013; He et al. 2013; Mortada et al. 2014), and the support vector machine (SVM) (Konar and Chattopadhyay 2011). However, feature extraction remains challenging because fault-induced transient impulses existing in vibration signals are usually contaminated by noise.

Researchers have proposed numerous methods to extract features in different situation, for example, Boskoski and Juricic extracted Renyi entropy values from wavelet packet coefficients of vibration signals to detect mechanical faults in rotational drives (Bokoski and Juricic 2012), Li used adaptive filter and robust statistical features to detect faults in wind turbine transmission system (Li and Frogley 2013). These studies found that features can be extracted in three domains including time domain, frequency domain, and time frequency domain. Time domain features are extracted based on time series analysis, such as the statistical features (variance, kurtosis) and the autoregressive model (Li et al.

C. Wang (✉) · M. Gan · C. Zhu
Department of Precision Machinery and Precision Instrumentation,
University of Science and Technology of China (USTC), Hefei 230027,
AnHui, People's Republic of China
e-mail: wcc12345@mail.ustc.edu.cn

M. Gan
e-mail: ganmeng@mail.ustc.edu.cn

2012). Frequency domain features are achieved from spectrum analysis, such as subband energy (Zhao et al. 2005). Features in the time or frequency domains focus only on specific signal content that cannot comprehensively consider fault-related information because defect-induced impulses are non-stationary with time-varying frequencies. Contrarily, time-frequency features can present a synthetic consideration for mechanical fault detection by characterizing varying frequency information at different times. Commonly used time-frequency analysis methods include short-time Fourier transform (Klein et al. 2001), Wigner–Ville distribution (Baydar and Ball 2001), wavelet transform (WT) (Wang et al. 2011) and empirical mode decomposition (Peng et al. 2005; Rai and Mohanty 2007). Among these techniques, WT is outstanding in rotary machine diagnosis because its multi-resolution merit is suitable for analyzing signals with transient impulses. Continuous wavelet transform (CWT) and discrete wavelet transform (DWT) are two categories of WT, each with its own merits as well as deficiencies.

CWT can calculate wavelet coefficients on any scale to reveal signal features completely (Lin and Zuo 2003). However, the high computation cost hinders fast diagnosis. DWT (Mori et al. 1996) can improve decomposition efficiency, but limitations exist in two aspects: (1) lack in shift-invariance causes waveform distortion and (2) fixed frequency domain sampling manner leads to low resolution and severe frequency aliasing. Wavelet packet transform (WPT) (Zhang et al. 2013; Pandya et al. 2014), as an extension of DWT, can decompose signals with higher frequency resolution by analyzing them in both low and high frequency bands. However, it inherits the other problems of DWT. These drawbacks sometimes prevent DWT or WPT from effectively capturing the fault information of rolling bearings.

Recently, overcomplete techniques have become a well recognized tool in signal processing (Kovacevic and Chebira 2007), and numerous overcomplete WT have been designed and utilized for application (Chui and He 2000; Selesnick 2011). Comparing with traditional DWT and WPT, overcomplete DWT have significant advantages. First, overcomplete DWT can achieve higher frequency resolution. Second, overcomplete DWT can be approximately shift-invariant. Third, the redundant basis can help solve the frequency aliasing problems. Besides, with a certain set of basis functions, overcomplete DWT can focus on specific physical properties of the signal. In the current study, Selesnick's tunable Q-factor wavelet transform (TQWT) (Selesnick 2011), as a kind of overcomplete DWT, will be employed because it can reflect the oscillatory behavior of the signal except for the mentioned advantages.

In order to find a better representation for signal, sparsity-based method is proposed and has received several notable achievements in the field of machinery fault diagnosis (Liu et al. 2002; Yang et al. 2005; Feng and Chu 2007). Its basic

principle is to construct a signal as a linear combination of transform basis (atoms) from an overcomplete dictionary (Grbovic et al. 2012). Two well-known methods are utilized to obtain a sparse representation of the signal, which are respectively matching pursuit (Mallat and Zhang 1993) and basis pursuit (Chen et al. 2001). Matching pursuit is suitable for orthogonal dictionary and has fast computational speed, whereas basis pursuit is characterized by super-resolution and better sparsity. The advantages of basis pursuit indicates it has the capacity to extract more informative intrinsic features accurately from vibration signals without frequency aliasing. The process can be conducted by the following two steps: redundant dictionary design and sparse coefficients solving. Redundant dictionary can be designed by K-SVD algorithm (Rusu and Dumitrescu 2012), shift-invariant sparse coding algorithm (Plumbley et al. 2006), or redundant signal transform basis like wavelet packet basis (Yang et al. 2005). Sparse coefficients can be calculated by greedy pursuit algorithms (Bahmani et al. 2013), l_p norm regularization algorithms (Marjanovic and Solo 2012) and iterative shrinkage algorithms (Beygi et al. 2012).

Considering the merits of overcomplete DWT and sparse representation technique, this paper proposes a pioneering sparse wavelet energy (SWE) feature for diagnosing rolling element bearings. SWE features are achieved from a sparse representation of wavelet-based time-frequency distribution obtained by basis pursuit. The redundant dictionary for basis pursuit is designed by TQWT, which can reveal the oscillatory properties of the signal. Then, the dictionary is optimized to achieve the sparse wavelet-based distribution by the split variable augmented Lagrangian shrinkage algorithm (SALSA) (Selesnick 2011), which can efficiently handle various problems. After optimization, SWE features can be obtained from sparse wavelet-based distribution. Its physical meaning is easy to interpret and has the property of sparsity and high resolution. Moreover, SWE feature is suitable for exploring the intrinsic characteristics of bearing fault signals due to its sensitivity to impulses, which greatly benefits machinery diagnosis. The advantages of SWE features are confirmed in both simulation and experiment by comparing with several traditional features.

The remainder of this paper is organized as follows. “Sparsity-related theory” section introduces the basic theory of sparse representation method. “Overcomplete DWT” section presents the design of redundant dictionary by overcomplete DWT. “Automatic fault diagnosis based on SWE” section describes the realization of intelligent diagnosis based on SWE features. “Simulation” section illustrates the procedure and preliminary validation of the proposed method using simulated bearing data. “Engineering validation” section further confirms the advantages of SWE features by experiment. Conclusions are drawn in “Conclusion” section.

Sparsity-related theory

Basis pursuit

Consider signal \mathbf{x} with p points, which can be viewed as a vector in \mathbb{R}^p . A redundant dictionary $\mathbf{A} = \{\mathbf{a}^1, \mathbf{a}^2, \dots, \mathbf{a}^n\}$ consists of n vectors $\mathbf{a}^j \in \mathbb{R}^p$, that span the entire space \mathbb{R}^p with $n > p$. Signal \mathbf{x} can be represented as the superposition of basis functions:

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

where $\mathbf{s} = [s^1, s^2, \dots, s^n]^T$ are the coefficients for basis functions. Among all possible coefficient sets, the sparsest can be achieved by the minimization of l_0 norm (Donoho and Huo 2001), which is defined as the number of nonzero elements. This process can be expressed as Eq. (2).

$$\underset{\mathbf{s}}{\operatorname{argmin}} \|\mathbf{s}\|_0, \quad \text{s.t. } \mathbf{x} = \mathbf{A}\mathbf{s}. \tag{2}$$

Unfortunately, Eq. (2) is a non-convex optimization problem which is difficult to deal with. Therefore, in practice, the sparse solution to Eq. (1) is usually obtained by solving the optimization problem in Eq. (3),

$$\underset{\mathbf{s}}{\operatorname{argmin}} \|\mathbf{s}\|_1, \quad \text{s.t. } \mathbf{x} = \mathbf{A}\mathbf{s}, \tag{3}$$

where $\|\mathbf{s}\|_1$ is the l_1 norm of \mathbf{s} as defined in Eq. (4).

$$\|\mathbf{s}\|_1 := \sum_{n=0}^{N-1} |s(n)|. \tag{4}$$

Equation (3) is known as the basis pursuit (BP) problem (Chen et al. 2001) that can provide the sparsest solution for the l_0 problem for most large scale redundant systems (Donoho 2006).

Signal \mathbf{x} usually contains noise in practice. Thus, solving Eq. (3) exactly is unreasonable. Generally, we can find an approximate solution by changing the optimization function to Eq. (5), which is called the basis pursuit denoising (BPD) problem (Gunn et al. 2002),

$$\underset{\mathbf{s}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{A}\mathbf{s}\|_2^2 + \lambda \|\mathbf{s}\|_1, \tag{5}$$

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

Algorithm to solve the BP and BPD problems

Several effective approaches have been developed for solving the BP and BPD problems, such as primal-dual log

barrier interior point method (Chen et al. 1998) and iterative shrinkage/thresholding algorithm (ISTA) (Michailovich 2011). Recently, a novel SALSA method is proposed by Afonso (Afonso et al. 2010). SALSA will be used in the present study because of its flexibility in handling various problems and its fast convergence in practice. SALSA will change the unconstrained optimization formulation in Eq. (5) to a constrained problem [Eq. (6)] based on a variable splitting technique,

$$\underset{\mathbf{s}, \mathbf{u}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{A}\mathbf{u}\|_2^2 + \lambda \|\mathbf{u}\|_1, \quad \text{s.t. } \mathbf{s} - \mathbf{u} = 0. \tag{6}$$

where \mathbf{u} is the created new variable. This problem can be solved by an augmented Lagrangian method (ALM) (Mateos et al. 2010), more specifically, the alternating direction method of multipliers (ADMM) (Mateos et al. 2010) by the following update equations:

$$\begin{aligned} \mathbf{u}^{(k+1)} &= \underset{\mathbf{u}}{\operatorname{argmin}} \lambda \|\mathbf{u}\|_1 + \frac{\mu}{2} \|\mathbf{u} - \mathbf{s}^{(k)} - \mathbf{d}^{(k)}\|_2^2, \\ \mathbf{s}^{(k+1)} &= \underset{\mathbf{s}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{x} - \mathbf{A}\mathbf{s}\|_2^2 + \frac{\mu}{2} \|\mathbf{u}^{(k+1)} - \mathbf{s} - \mathbf{d}^{(k)}\|_2^2, \\ \mathbf{d}^{(k+1)} &= \mathbf{d}^{(k)} - \left(\mathbf{u}^{(k)} - \mathbf{s}^{(k)} \right), \end{aligned} \tag{7}$$

where k is the iteration index and μ is a user specified penalty parameter.

According to the previously mentioned theory, we need to find a set of basis to construct an underdetermined system that can be optimized by SALSA to achieve a sparse signal representation. In the current study, overcomplete DWT will be used to construct the undetermined system because it can highlight impulses representing fault information of rolling bearings.

Overcomplete DWT

DWT

The WT of signal $x(n)$ ($n = 1, 2, \dots, N$) is shown in Eq. (8),

$$W(a, b) = \langle x(n), \psi_{a,b}(n) \rangle = \sum_{n=1}^N x(n) \frac{1}{\sqrt{a}} \psi^* \left(\frac{n-b}{a} \right), \tag{8}$$

where the asterisk represents complex conjugate, ψ is the mother wavelet function, a and b denote the scale factor and translational factor, respectively. In practice, DWT is usually employed to implement the transform in Eq. (8) for the benefit of computational convenience and easy invertibility. By making $a = 2^j$, $b = k2^j$, DWT can be realized as

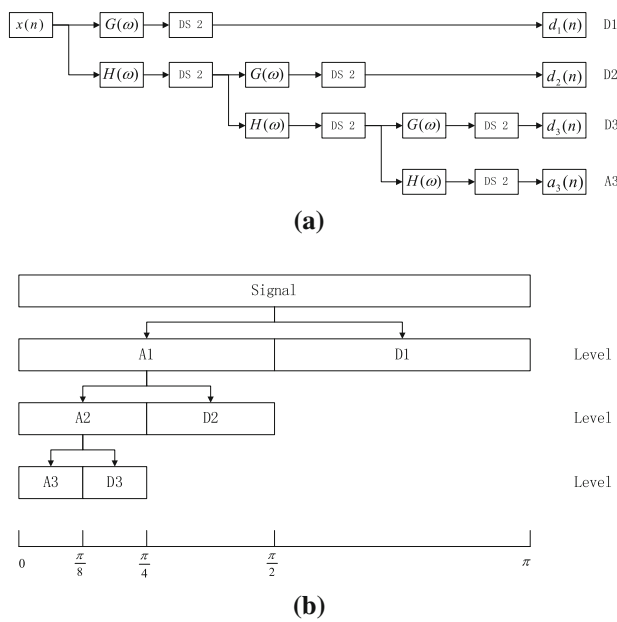


Fig. 1 Demonstration of the DWT process: **a** the decomposition procedure and **b** the corresponding frequency partition manner

$$dwt(j, k) = \frac{1}{\sqrt{2^j}} \sum_{n=1}^N x(n) \psi^* \left(\frac{n - k2^j}{2^j} \right) dt. \quad (9)$$

Practically, DWT can be implemented by Mallat's iterated algorithm (Rajpoot et al. 2008), which recursively convolves the low-pass channel series with low pass filter $h(n)$ and band pass filter $g(n)$ and subsequently downsamples the filtered series in each channel by a factor of 2. The process of a three-level DWT can be illustrated by Fig. 1a, and the frequency partition manner is described in Fig. 1b. DWT has been proven a powerful tool for time-frequency signal analysis. However, the frequency aliasing problem and its poor frequency resolution limits its application in mechanical fault diagnosis because it cannot effectively identify periodic impulses.

Realization of overcomplete DWT

Overcomplete DWT is also implemented by an iterated two-channel filterbank. In the current paper, Selsnick's TQWT (Selesnick 2011) is employed to conduct overcomplete DWT, which is conceptually simple and can be implemented efficiently.

Unlike DWT, the filters, on which TQWT is based are specified directly in the frequency domain. Transform is implemented by iteratively applying the two-channel filter banks on its low-pass channel, followed by the low-pass scaling and high-pass scaling operations. For low-pass scaling with parameter α , the output signal has a sampling rate of αf_s , where f_s is the sampling rate of the input signal. Simi-

larly, for high-pass scaling with parameter β , the output signal has a sampling rate of βf_s . The scaling parameters satisfy $0 < \alpha < 1$, and $0 < \beta \leq 1$ to ensure that the WT will not be overly redundant. Meanwhile, to achieve a oversampled results, we also require $\alpha + \beta > 1$. Generally, the parameters are set based on Eq. (10),

$$\beta = \frac{2}{Q+1}, \quad \alpha = 1 - \frac{\beta}{r}, \quad (10)$$

where Q is the quality factor and r is the redundant factor. The physical interpretation and selection principle of Q and r can be found in (Selesnick 2011). The low-pass filter and high-pass filter are defined as Eqs. (11) and (12),

$$H(\omega) = \begin{cases} 1, & |\omega| \leq (1-\beta)\pi \\ \theta \left(\frac{\omega + (\beta-1)\pi}{\alpha + \beta - 1} \right), & (1-\beta)\pi < |\omega| < \alpha\pi \\ 0, & \alpha\pi \leq |\omega| \leq \pi \end{cases} \quad (11)$$

$$G(\omega) = \begin{cases} 0, & |\omega| \leq (1-\beta)\pi \\ \theta \left(\frac{\alpha\pi - \omega}{\alpha + \beta - 1} \right), & (1-\beta)\pi < |\omega| < \alpha\pi \\ 1, & \alpha\pi \leq |\omega| \leq \pi \end{cases} \quad (12)$$

where $\theta(\omega) = \frac{1}{2}(1 + \cos(\omega))\sqrt{2 - \cos(\omega)}$, $|\omega| \leq \pi$. The mathematical expression of the low-pass scaling and high-pass scaling can be found in Eq. (13).

$$\text{LPS}(\alpha) : Y(\omega) = X(\alpha\omega), \quad |\omega| \leq \pi,$$

$$\text{HPS}(\beta) : Y(\omega) = X \left(\beta\omega + \frac{\omega}{|\omega|} (1-\beta)\omega \right), \quad 0 < |\omega| \leq \pi. \quad (13)$$

The TQWT decomposition process is completely illustrated in Fig. 2a, whereas the corresponding inverse transform is shown in Fig. 2b. With the achieved overcomplete DWT results, it will be possible for us to get more accurate sparse decomposition results by the optimization algorithm.

Automatic fault diagnosis based on SWE

TQWT decomposition level determination for bearing diagnosis

For certain TQWT decomposition parameters, corresponding wavelets represent different frequency subbands and the decomposition result should cover the informative frequencies of the signal. Accordingly, we synthesize the fault-related frequency band and the frequency response of wavelets for TQWT for an effective diagnosis. The informative frequency band can be easily localized in the 2-D time-frequency distribution.

Supposing a signal with period transient impulses has a CWT-based TFD shown in Fig. 3a, the normalized fault-

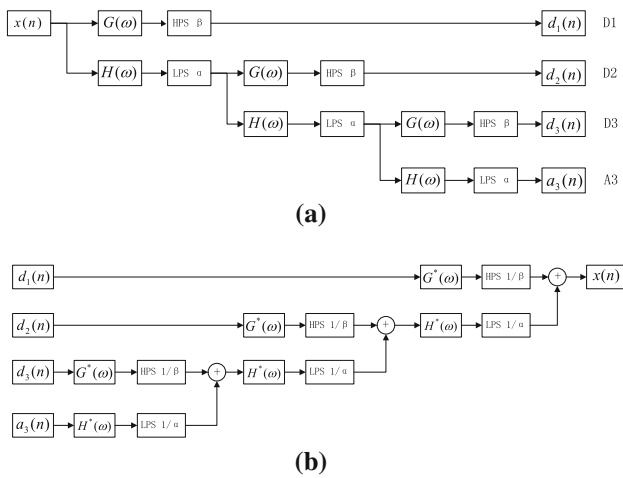


Fig. 2 Demonstration of the TQWT process: **a** the decomposition procedure and **b** the inverse TQWT process

related frequency lies in [0.15, 0.35] as the dash line indicates. The frequency responses of the wavelets for TQWT are shown in Fig. 3b. For fully cover the interested part highlight in Fig. 3b, we should define the lower boundary $J_{low} \leq 2$ and the upper boundary $J_{up} \geq 7$. Generally, the range can be set a little larger than necessary. With the decomposition boundary determined, $J_{up} - J_{low} + 2$ subbands are obtained according to Fig. 2a. In practical fault diagnosis containing multiple fault classes, the boundary should be determined by the union of different cases, which will be further discussed in the engineering section (“Engineering validation” section).

SWE features

A set of oversampled DWT coefficients s is obtained by applying TQWT to a signal x as shown in Eq. (14).

$$s = \text{TQWT}(x). \tag{14}$$

To achieve an ideal set of coefficients, which is the sparsest, we need to solve the optimization problem expressed in Eq. (15),

$$\underset{s}{\text{argmin}} \|x - \text{ITQWT}(s)\|_2^2 + \lambda \|s\|_1, \tag{15}$$

where ITQWT stands for the inverse transform of TQWT. The above equation is the BPD problem described in “Basis pursuit” section, and can be efficiently solved by SALSA. The sparsest result has the most concentrated energy best reflecting the period impulses of the machinery signal. The proportion of subband average energy can be regarded as a significant feature for machinery fault diagnosis, which is the proposed SWE feature. Suppose signal X has m subbands

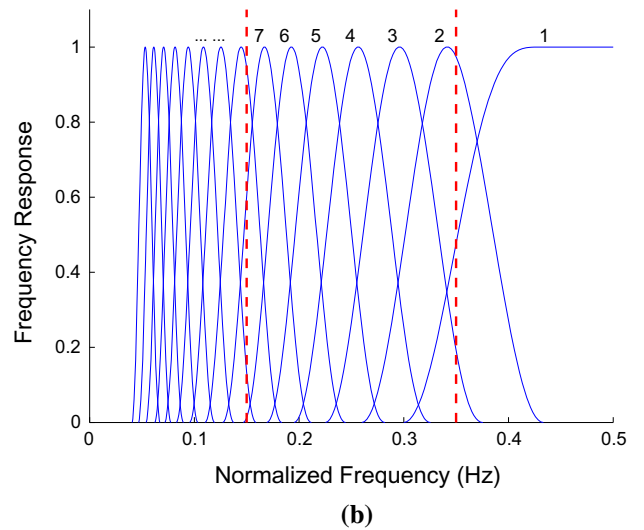
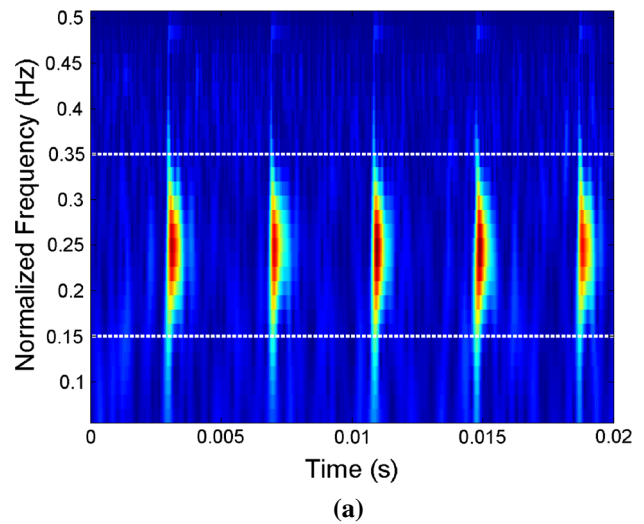


Fig. 3 Determination of the decomposing level: **a** CWT result of a signal and **b** frequency responses of wavelets in TQWT

$\{w^{(1)}, w^{(2)}, \dots, w^{(m)}\}$. The SWE feature of the i th subband with $l(i)$ points can be calculated as shown in Eq. (16).

$$\text{SWE}(i) = \frac{E_i}{E_{\text{total}}} = \frac{\left(\sum_{j=1}^{l(i)} w^{(i)}(j)^2\right) / l(i)}{\sum_{i=1}^m \left(\left(\sum_{j=1}^{l(i)} w^{(i)}(j)^2\right) / l(i)\right)}. \tag{16}$$

Automatic fault diagnosis

The diagnosis process includes the training part which infers a feature space for fault types and the testing part which evaluates the accuracy of the training model. The detailed steps to conduct intelligent fault diagnosis based on SWE features can be summarized as follows:

- Step 1: The dataset is randomly divided into the training and testing sets.
- Step 2: TQWT is applied to obtain the oversampled DWT subbands of each sample.
- Step 3: SALSA is employed to achieve the sparse representation of the oversampled DWT subbands.
- Step 4: The proportion of subband energy is calculated as the SWE features for fault diagnosis.
- Step 5: A classifier is trained by the SWE features of the training samples.
- Step 6: The misclassification rate of testing samples is checked to evaluate the performance of the classifier. If the classification accuracy rate is satisfactory, then the classifier can be finally used to conduct intelligent fault diagnosis.

Simulation

Vibration model construction

To illustrate the procedure and the preliminary validation of the proposed SWE features, a simulated bearing fault signal is constructed. The periodical impulses representing fault information can be described in Eq. (17),

$$I(t) = \sum_{k=1}^t A \exp \left\{ \frac{-\xi}{\sqrt{1-\xi^2}} [2\pi f_0(t - kT)] \right\} \sin [2\pi f_0(t - kT)], \quad (17)$$

where $A = 4$ is the initial magnitude of the simulated vibration signal, $f_0 = 2000$ Hz is the central frequency of the resonance band, $\xi = 0.04$ represents the damping ratio and $T = 0.02$ s denotes the repetition period. The vibration signal of the rolling bearing generally includes periodic impulses, harmonic components, and noise. Therefore, the vibration model is constructed as shown in Eq. (18), where $n(t)$ denotes white noise.

$$h(t) = I(t) + B \sin(2\pi f_1 t) + C \cos(2\pi f_2 t) + D n(t). \quad (18)$$

In this case, $B = 0.2$, $C = 0.15$, $D = 0.4$, $f_1 = 500$, and $f_2 = 900$ Hz. The waveform of the simulated signal is presented in Fig. 4.

SWE features of the simulated signal

The CWT-based TFD of the simulated signal is demonstrated in Fig. 3a. With the analysis in ‘‘SWE features’’ section, the decomposition boundaries of TQWT are set to be $J_{low} = 1$ and $J_{up} = 8$ based on the wavelet frequency responses in Fig. 3b ($Q = 4$, $r = 3$). Period impulses are detected in four subbands (subband 4–7) as shown in Fig. 5a because of the

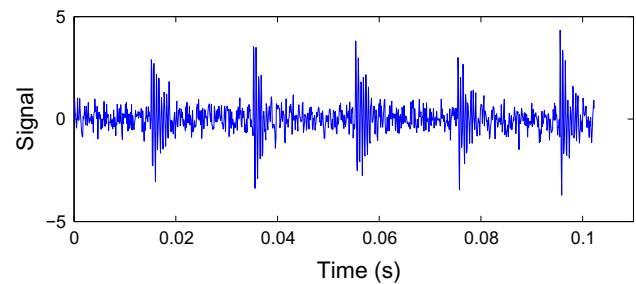


Fig. 4 Waveform of the simulated signal

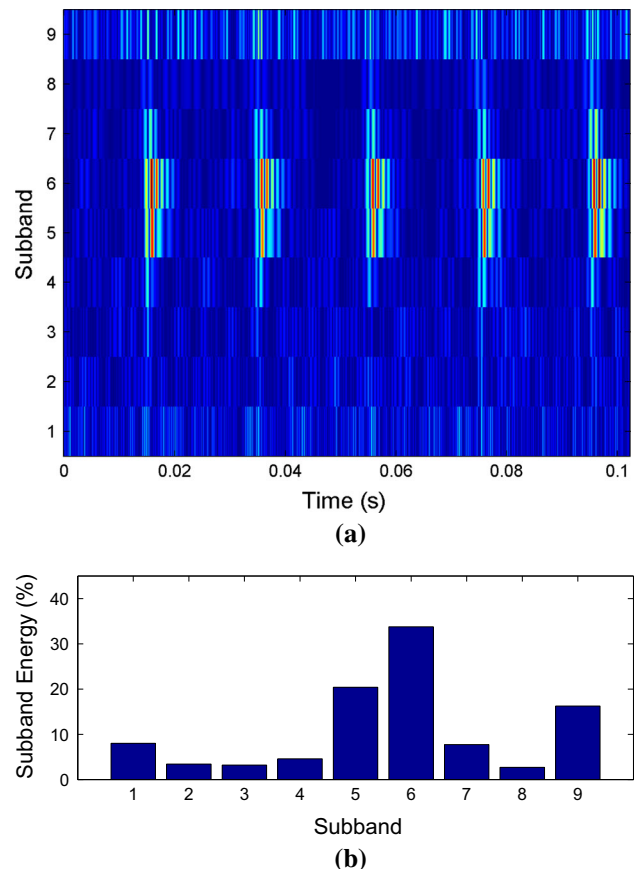


Fig. 5 Illustration of the overcomplete DWT: **a** the distribution of coefficients in each subband and **b** the corresponding subband energy distribution

aliasing of different subbands. This result can be improved through sparse-related techniques. By applying SALSA, a sparse representation of overcomplete DWT coefficients can be obtained and the result is shown in Fig. 6. Obviously, distribution in Fig. 6a is much clearer, with period impulses concentrated in subbands 5 and 6. The sparsity can reduce frequency aliasing and ensure intrinsic energy flow from low to high level with little noise, which can be considered as the intrinsic structure embedded in the signal. It can be inferred that different health states of the bearing signal possess different energy flows, which motivates its application in machin-

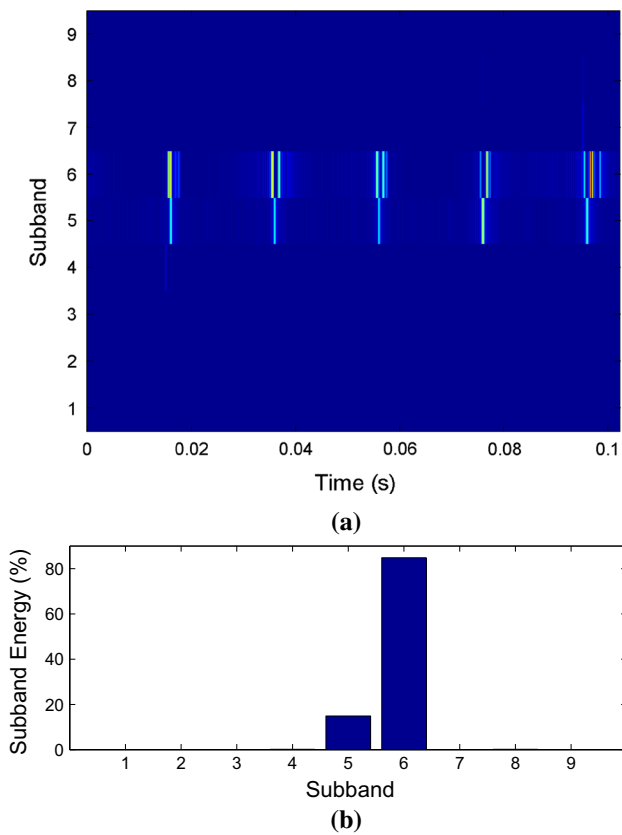


Fig. 6 Illustration of the sparsity-based wavelet transform: **a** the distribution of coefficients in each subband and **b** the corresponding SWE features

ery fault diagnosis. Therefore, we use the average energy proportion of each subband as a signature representing the fault information calculated by Eq. (16). The result is illustrated in Fig. 6b.

Comparison with other methods

The same simulated signal is analyzed by DWT and WPT for comparison, which are widely used in fault diagnosis. According to “SWE features of the simulated signal” section, energy flow is captured from TFD. An 8-level DWT decomposition result and its corresponding subband energy is shown in Fig. 7. Figure 7a with five periodic impulses indicates the effectiveness of DWT decomposition, and energy flow in Fig. 7b reflects the energy distribution in different subbands. However, the distribution in Fig. 7a has a few confusing fuzzy lines and the impulse information can be found in subband six except the energy-concentrated subbands (7 and 8), which is caused by the frequency aliasing of DWT. Contrarily, the SWE-based method proposed in this paper can compensate for the drawback and achieve accurate energy flow. Moreover, the little energy in subbands 1–4 in Fig. 7b

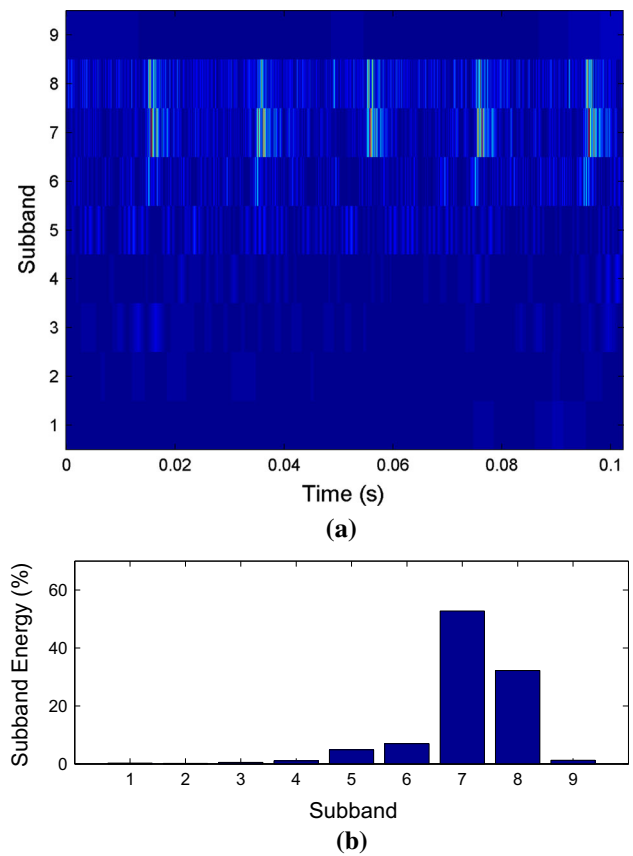


Fig. 7 Illustration of the DWT: **a** the distribution of coefficients in each subband and **b** the corresponding subband energy features

indicates the low frequency resolution of DWT, which may lead to the missing of signal information.

WPT is conducted in three levels for the simulated signal with $2^3 = 8$ subbands obtained. The transform coefficients and corresponding subband energy are presented in Fig. 8a. WPT performs a complete decomposition in each level to achieve both low-frequency and high-frequency components, thereby obtaining a higher frequency resolution than DWT. The WPT coefficients in Fig. 8a demonstrate obvious periodic impulses embedded in the signal in subbands 4–6. However, the distribution is not as clear as Fig. 6a and frequency aliasing still exists, leading to a more dispersed energy along the frequency axis as shown in Fig. 8b.

The comparison verifies that the proposed processing method based on overcomplete DWT and SALSA guarantees high resolution in frequency and accurate reflection of fault information. These merits are beneficial to characterize a more representative intrinsic energy pattern of the fault signal than traditional WT techniques, thereby rendering the new SWE features recognizable for mechanical fault identification.

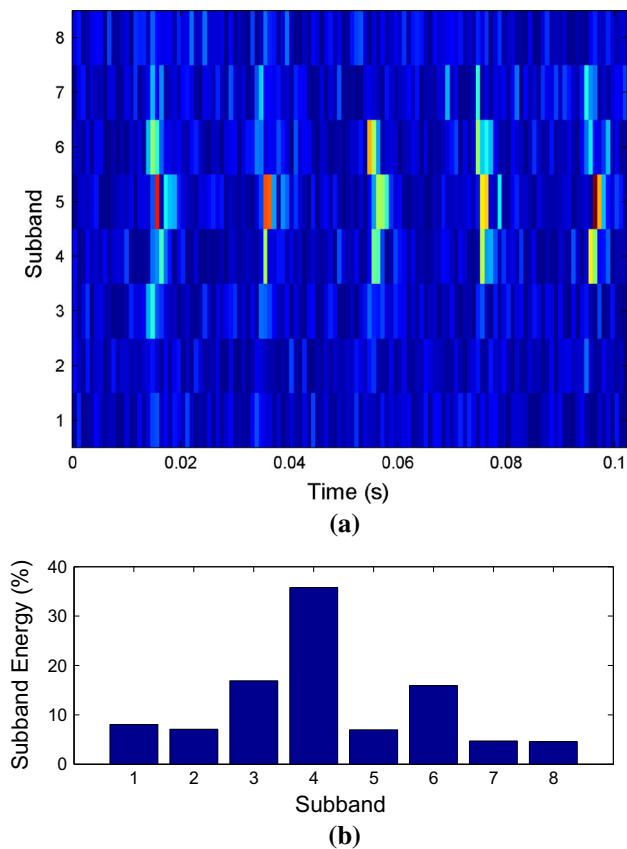


Fig. 8 Illustration of the WPT: **a** the distribution of coefficients in each subband and **b** the corresponding subband energy features

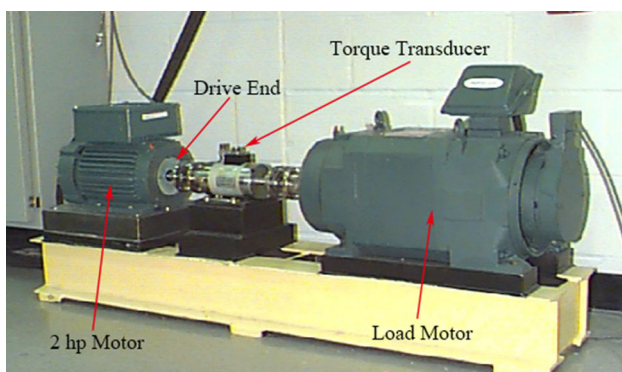


Fig. 9 Experimental setup for acquiring vibration signals of the rolling element bearings

Engineering validation

Instruction of the dataset

The proposed SWE feature is evaluated on a rolling bearing dataset from the Case Western Reserve University Bearing Data Center. The experimental apparatus presented in Fig. 9 consists of the following main parts: a 2 hp motor on the left, a torque transducer and a dynamometer in the middle and a load

motor on the right. The testing groove ball bearing supports the motor shaft at the drive end, on which single-point faults are seeded. Vibration data are collected by accelerometers attached to the housing with magnetic bases at the 12 o'clock position at the drive end. The sampling frequency is set at 12 kHz. The dataset includes four health conditions of the rolling bearing: healthy; rolling element defect; inner-race defect and outer-race defect. Each fault condition has defects with three different sizes: 0.007, 0.014 and 0.021 inches. Detailed parameters are listed in Table 1. Typically, sample signals and their corresponding Fourier spectrum under 10 different states are illustrated in Fig. 10.

TQWT decomposition level determination for bearing signals

In this rolling bearing case, Q is set to 4 and r is set to 4 to obtain the preliminary wavelet in advance referring to (Selesnick 2011). The same bearing fault locations usually share similar fault-related frequency bands in TFD. Thus, we select three typical defective signals in different locations to determine the effective energy band. The obtained CWT results and frequency responses of wavelets used in TQWT are shown in Fig. 11. The fault-related band is [0.15, 0.35] for rolling-element defect; [0.1, 0.37] for inner-raceway defect and [0.1, 0.3] for outer-raceway defect. The final frequency band determined by the union of three conditions is [0.1, 0.37]. According to Fig. 11d, the decomposition boundary should be determined as $J_{low} \leq 2$ and $J_{up} \geq 14$. A larger decomposition level is needed to ensure the whole fault information is embodied. Therefore, J_{low} is set to 1 and J_{up} is set to 15.

SWE feature extraction

After a redundant dictionary is constructed by TQWT, SALSA is utilized to achieve a sparse set of coefficients. The SWE features calculated by Eq. (16) for samples under 10 conditions are shown in Fig. 12a. It can be learned that energy is mainly concentrated in 1 or 2 subbands, each of which presents unique energy distribution that is beneficial for classifiers to identify the condition of the bearing signals. To illustrate the classification capability of the SWE feature intuitively, principle component analysis (PCA) (Humberstone et al. 2012; Shao et al. 2014) is applied to the features as a dimension reduction method. The target dimension is set to 3 to generate a visualization clustering result as shown in Fig. 12b. The SWE features in low-dimension show reasonable distribution for original high-dimensional features. Samples in the same class are gathered to a cluster and separated from the others with a significant distance. The remarkable clustering result is strong proof that the SWE feature is effective in fault identification of the rolling bearings.

Table 1 Parameters of the rolling bearings

Class label	Fault location	Fault size (in.)	Training sample no.	Testing sample no.	Sample length
S1	Healthy	0	50	20	1024
S2	Rolling element	0.007	50	20	1024
S3	Rolling element	0.014	50	20	1024
S4	Rolling element	0.021	50	20	1024
S5	Inner raceway	0.007	50	20	1024
S6	Inner raceway	0.014	50	20	1024
S7	Inner raceway	0.021	50	20	1024
S8	Outer raceway	0.007	50	20	1024
S9	Outer raceway	0.014	50	20	1024
S10	Outer raceway	0.021	50	20	1024

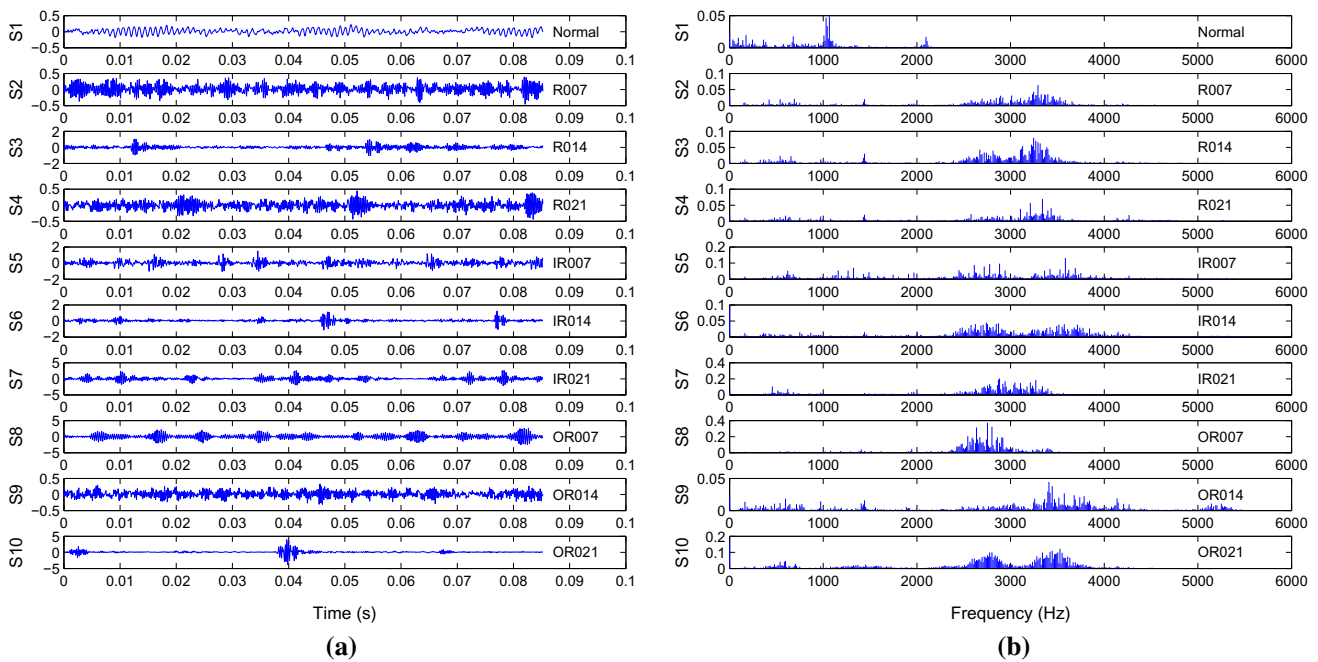


Fig. 10 Demonstration of the bearing signals: **a** the original waveform and **b** the corresponding Fourier spectrum

To highlight the advantages of the SWE features, the DWT and WPT subband energy features are extracted as a comparison. In this case, 15-level DWT is employed and the DWT features extracted from the samples of 10 cases are presented in Fig. 13a. It can be seen that energy concentrated in subbands 2 and 3 express similarities among many conditions, indicating that DWT-based energy flow cannot effectively identify bearing faults. The frequency aliasing and low frequency resolution render DWT ignore some unique signal information as discussed in “Simulation” section. Such a result can also explain the weakness of DWT-based features in bearing fault diagnosis. More intuitive proof can be achieved by PCA clustering result presented in Fig. 13b. DWT-based PCA clustering results are completely indis-

cernible. The only fault type we can identify is the healthy one, which is consistent with the result in Fig. 13a. Thus, DWT subband features cannot express intrinsic fault information in this rolling bearing diagnosis.

WPT has higher resolution than DWT, therefore, WPT subband energy feature has arguably more elaborate information than DWT. The level of WPT is set to 4 to obtain $2^4 = 16$ subbands. WPT subband energy of samples from 10 different conditions are shown in Fig. 14a. WPT subband energy features perform better than the DWT subband energy features with most conditions presenting distinguishable characteristics. However, similarities remain in WPT-based features, such as S2, S3, S4 and S8. These features cannot be inferred to rival SWE features for the following

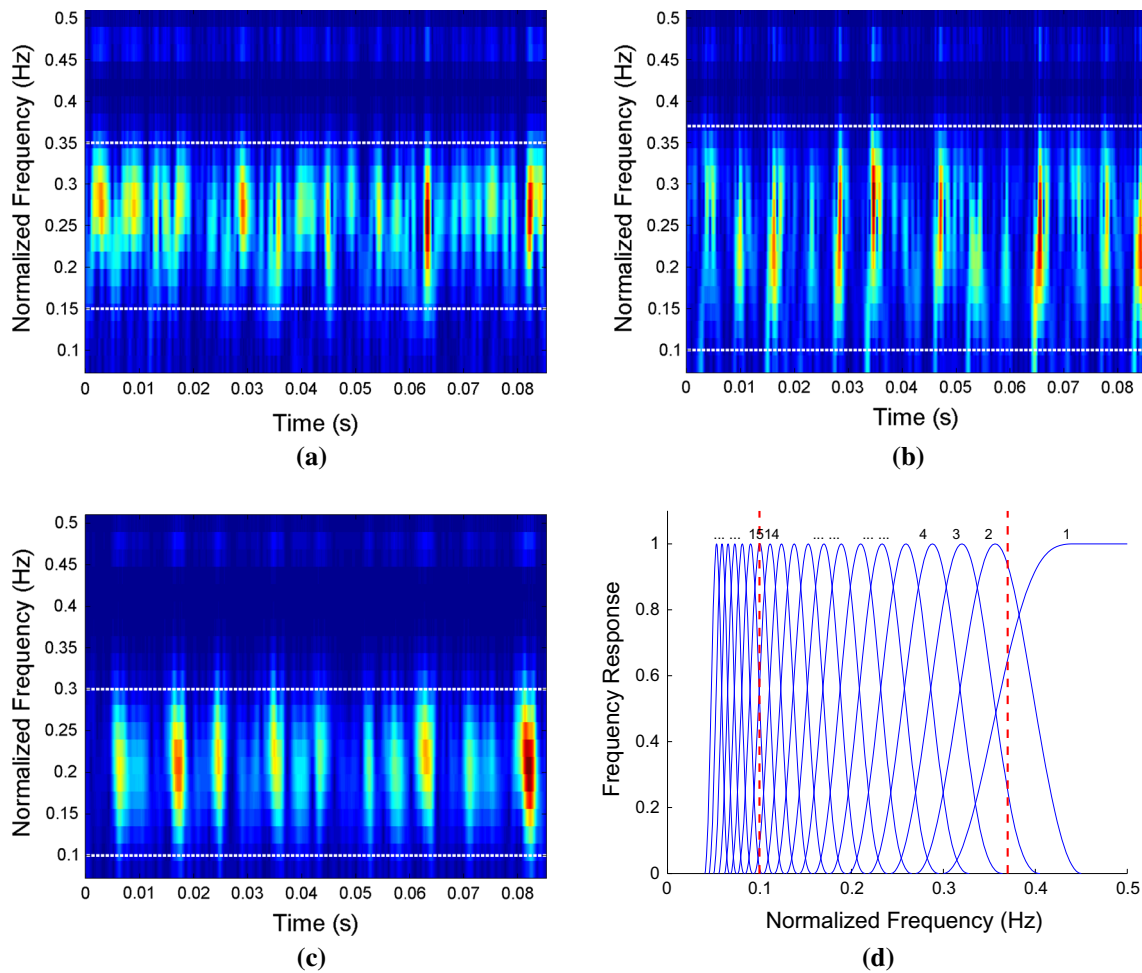


Fig. 11 Determination of the decomposing level: **a** CWT result of the rolling-element fault case, **b** CWT result of the inner-raceway fault case, **c** CWT result of the outer-raceway fault case and **d** frequency response of the wavelets in TQWT

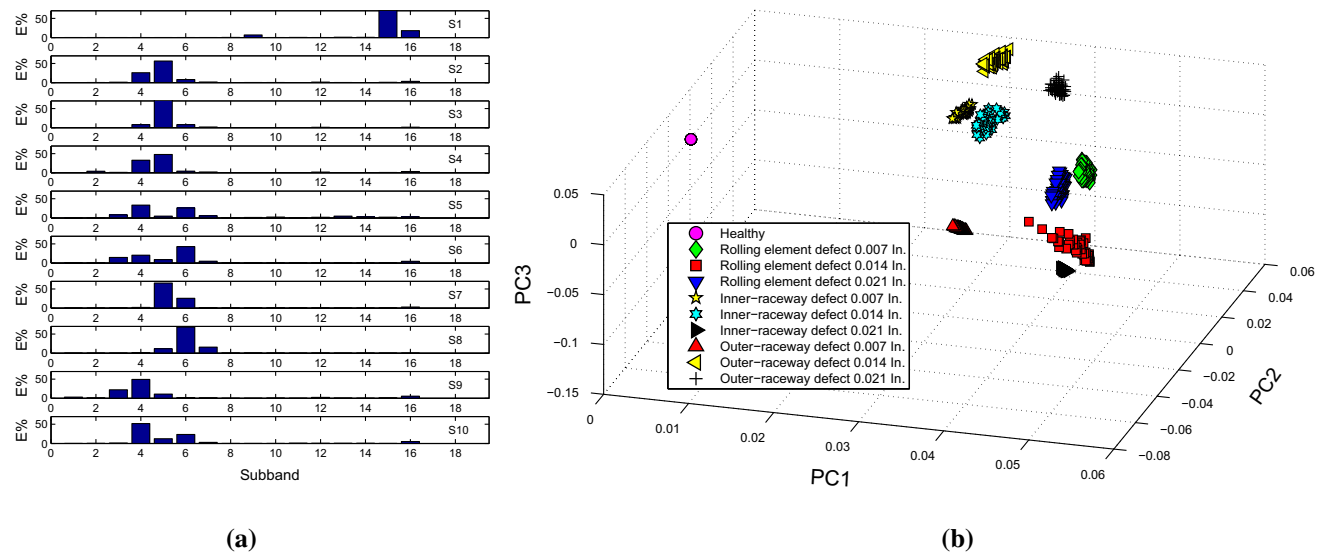


Fig. 12 Illustration of the SWE features of the bearings: **a** the distribution of the SWE features in 10 cases and **b** the intuitive clustering result

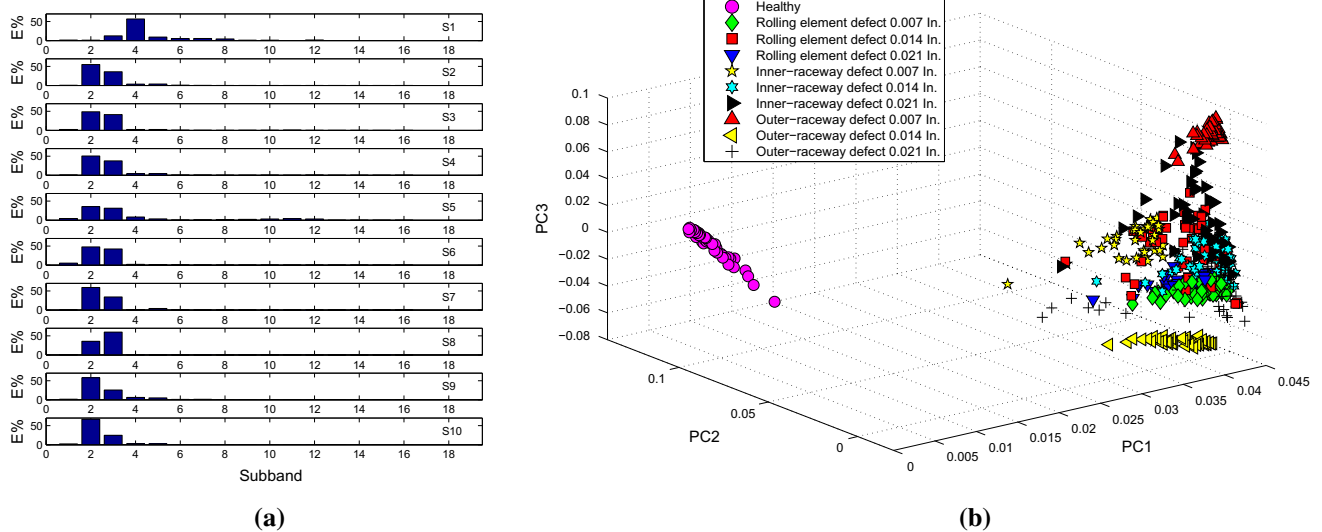


Fig. 13 Illustration of the DWT subband energy features of the bearings: **a** the distribution of the DWT-based features in 10 cases and **b** the intuitive clustering result

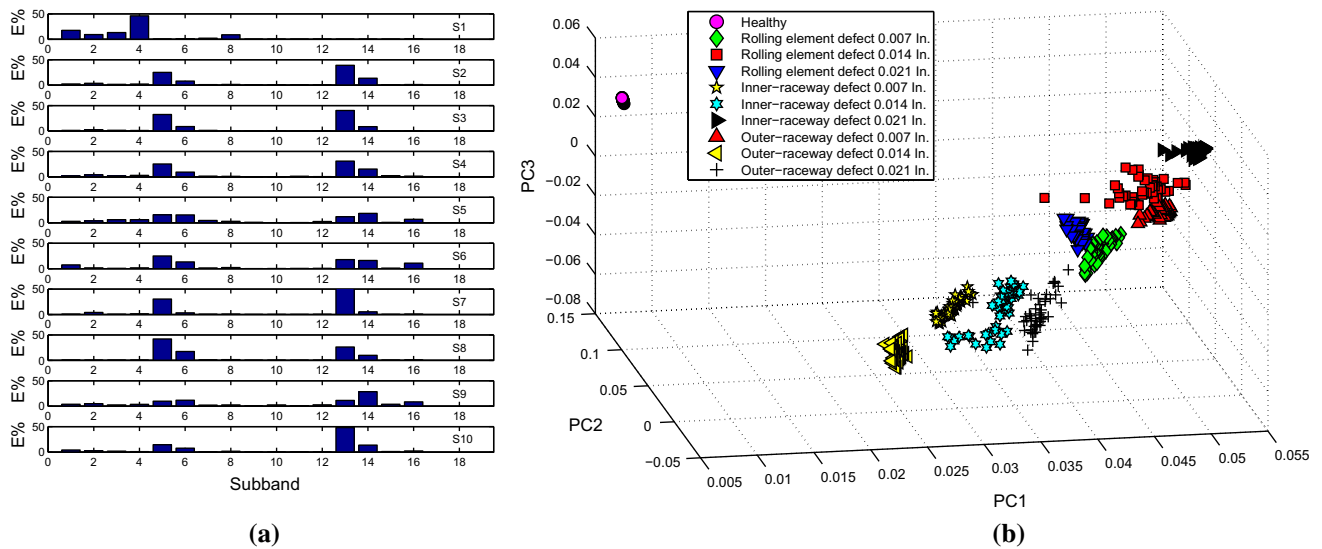


Fig. 14 Illustration of the WPT subband energy features of the bearings: **a** the distribution of the WPT-based features in 10 cases and **b** the intuitive clustering result

reasons: (1) WPT-based features are not as concentrated as the SWE features for frequency aliasing and (2) the extracted WPT-based features cannot reveal fault impulse information as clearly as SWE features according to “Simulation” section. Similar with SWE and DWT-based features, the WPT features are illustrated intuitively by PCA in Fig. 14b. WPT demonstrates better clustering than DWT, however, significant overlapping is found between S2 and S5, as well as S3 and S8. The overlapping is detrimental for effective and accurate fault pattern recognition. Moreover, it puts forward higher requests to the classifier.

The advantages of SWE features can be preliminarily seen by the comparison with DWT and WPT subband energy

features in both feature distribution and PCA-based clustering. To further confirm the effectiveness and practical value of SWE features, two commonly used intelligent classifiers were employed to achieve the final diagnosis result, as provided in “Intelligent fault diagnosis based on SWE features” section.

Intelligent fault diagnosis based on SWE features

The last step of intelligent fault pattern recognition is conducted by intelligent classifiers. In the current study, two commonly used intelligent classifiers are employed: nearest neighbor classifier (Gharavian et al. 2013) and SVM

Table 2 Classification results of the bearing data by the nearest neighbor classifier

Feature	DWT subband energy		WPT subband energy		SWE	
	Training	Testing	Training	Testing	Traning	Testing
Mean recognition rate (%)	100	67.05	100	88.5	100	97.95
Best recognition rate (%)	100	73.5	100	90.5	100	99
Mean standard deviation (10^{-4})	0	7.121	0	4.238	0	0.899
Best standard deviation (10^{-4})	0	6.763	0	3.425	0	0.45

Table 3 Classification results of the bearing data by the SVM classifier

Feature	DWT subband energy		WPT subband energy		SWE	
	Training	Testing	Training	Testing	Traning	Testing
Mean recognition rate (%)	79.7	76.1	99.96	91.85	99.16	98.65
Best recognition rate (%)	81.2	81	100	93.50	99.6	100
Mean standard deviation (10^{-4})	2.384	6.2	0.008	3.331	0.161	0.611
Best standard deviation (10^{-4})	2.219	5.075	0	2.813	0.077	0

(Konar and Chattopadhyay 2011). The nearest neighbor classifier is based on the intuitive concept that data instances of the same class should be closer in the feature space. It is conducted by calculating the distance of a new sample to all samples in the training data, and class is determined by the sample nearest to the new one. The SVM classifier is developed from the optimal separating plane under linearly separable condition. It determines the optimal separating hyperplane that minimizes the generalization error by maximizing the margin between the separating hyperplane and the nearest sample points, thereby realizing the classification. The nearest neighbor classifier is simple and intuitive while the SVM classifier generally has higher accuracy. Moreover, to confirm the merits of SWE features, the subband energy features obtained by DWT and WPT are also used for the same classifiers as a comparison.

The simple nearest neighbor classifier is first employed for the 10-class problem. The classification recognition rate η and standard deviation σ are calculated by Eqs. (19) and (20),

$$\eta = 1 - \sum_{i=1}^M P(\omega_i) \frac{k_i}{N_i}, \quad (19)$$

$$\sigma = \sum_{i=1}^M P^2(\omega_i) \frac{P_i(1 - P_i)}{N_i}, \quad (20)$$

where M is the class number, $P(\omega_i)$ is the occurrence probability of class ω_i with N_i samples, k_i is the misclassified number in ω_i , $P_i = k_i/N_i$. A higher η means a better classification and a smaller σ indicates a more convincing result. It can be inferred from Eq. (20) that the recognition rate estimated by Eq. (19) is asymptotically consistent only as

$N_i \rightarrow \infty$. In this case, the training set has 50 samples and the testing set has 20 samples according to Table 1. In order to get more convincing evaluation, we conduct a cross validation by randomly selecting training and testing samples for 50 times. Diagnosis results of three feature extraction techniques based on the nearest neighbor classifier over 50 runs are listed in Table 2. The standard deviation is around 10^{-4} for all the three features, which indicates the estimation result is acceptable. In comparison, the DWT subband energy feature has the lowest recognition rate, where nearly 1/4 testing samples are misclassified. This poor performance is consistent with feature distribution in Fig. 13a and sample clustering in Fig. 13b. The WPT subband energy feature with higher frequency resolution has a more satisfactory recognition rate, which proves that WPT-based feature is more effective than DWT-based feature. This better classification ability can be inferred from the feature demonstration in Fig. 14. SWE feature has the highest recognition rate because it is the sparse optimal result of a high resolution overcomplete DWT. Stated differently, SWE can reflect more elaborate content of the signal and reveal the intrinsic fault-related information.

Classification is then conducted by SVM classifier. As listed in Table 3, the recognition rate is higher than the nearest neighbor classifier for the testing samples and the standard deviation is smaller, which indicates a better performance of the SVM classifier.

In this case, the WPT-based feature remains more effective than the DWT-based one, and the SWE feature steadily exhibits the highest mean recognition rate. All of these results powerfully prove that SWE has significant advantages over DWT and WPT-based features for bearing fault classification.

Discussions

1. The main contribution of this paper is proposing a new SWE feature based on overcomplete DWT and sparsity theory. Overcomplete DWT brings high frequency resolution to time–frequency decomposition and the sparsity-related technique can effectively eliminate the frequency aliasing to obtain a concentrated energy distribution. Therefore, SWE feature can accurately reveal fault-related information, thereby approaching the intrinsic pattern of fault bearing signals. The above studies have comprehensively shown the benefits of the SWE feature in clustering and classification.
2. PCA is used to generate a visualization clustering result for intuitively interpreting the advantages of SWE features. Final diagnosis process by classifiers is performed on original SWE features rather than dimension-reduced features because PCA as a linear dimension reduction method will sacrifice non-linear information, which may be important to pattern recognition.
3. Two intelligent classifiers were employed in this study: nearest neighbor classifier and SVM classifier. Nearest neighbor classifier is almost the simplest method to emphasize the benefits of the SWE features, and SVM classifier is more advanced and illustrates the practical application. Some parameters need to be determined in practice for SVM, such as kernel function, kernel argument and penalty parameter (Gharavian et al. 2013). In this study, radial basis function kernel with kernel argument set to 1 is employed and the penalty parameter is set to 2500. The parameters are chosen roughly by experiment because the main point of this paper is highlighting the advantages of SWE features. More theoretical and elaborate parameters determination methods can be found in (Hsu and Lin 2002; Lorena and de Carvalho 2008).
4. Fault feature extraction by basis pursuit can also be conducted using wavelet packet dictionary according to (Yang et al. 2005) with experiments exhibiting the effectiveness. However, unlike the overcomplete DWT dictionary employed in this paper, the wavelet packet dictionary has little redundant information in frequency domain, thus limiting the sparse optimization (Chen et al. 2001). Take the simulated signal in Eq. (18) for further discussion, the basis pursuit result using wavelet packet dictionary is shown in Fig. 15. Comparing Figs. 5a and 15a, we can find that both of them indicate obvious period concentration around impulses along the time axis, but the result using wavelet packet dictionary (Fig. 15b) is severely dispersed along frequency axis in contrary to the result using overcomplete DWT dictionary (Fig. 5b). Stated differently, basis pursuit using wavelet packet dictionary cannot solve the frequency aliasing problem.

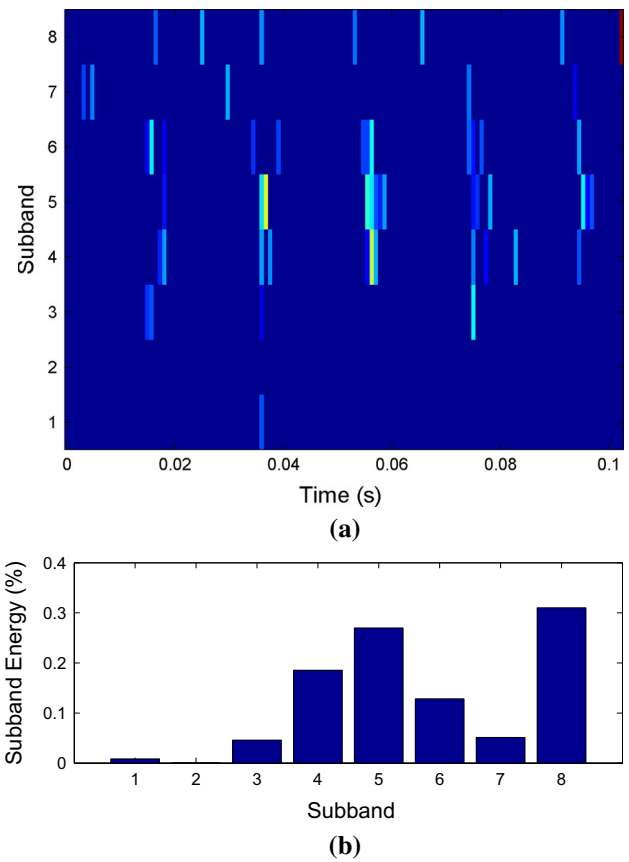


Fig. 15 Illustration of basis pursuit result using wavelet packet dictionary: **a** the distribution of coefficients in each subband and **b** the corresponding subband energy features

Therefore, the proposed SWE feature with redundant information in both time and frequency domains perform better than basis pursuit using wavelet packet dictionary.

Conclusion

In this study, a new energy-related feature, called SWE is proposed based on overcomplete DWT and basis pursuit. SWE feature is the sparse optimization result of the wavelet redundant dictionary, and reveals signal impulse information, which means it can be a meaningful characteristic for fault pattern recognition of rolling bearings. Compared with traditional DWT and WPT subband energy features, SWE has the advantages of high resolution, concentrated energy and clear physical significance. With the parameters set properly, the SWE feature is evaluated to have remarkable clustering result, outperforming the DWT-based feature and the WPT-based feature. The practical application of SWE features in distinguishing bearing fault signals under 10 health conditions further confirms its merits, with the highest recognition rate in both the nearest neighbor classification and the SVM

classification. All evidence indicates that the SWE feature has valuable potential and significant practicability in rolling bearing fault diagnosis.

Acknowledgements This work was supported by the National Key Basic Research Program of China (973 Program) under Grant No. 2014CB049500 and the Key Technologies R&D Program of Anhui Province under Grant No. 1301021005.

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.
- Bahmani, S., Raj, B., & Boufounos, P. T. (2013). Greedy sparsity-constrained optimization. *Journal of Machine Learning Research*, *14*, 807–841.
- Baydar, N., & Ball, A. (2001). A comparative study of acoustic and vibration signals in detection of gear failures using wigner–ville distribution. *Mechanical System and Signal Processing*, *15*(6), 1091–1107.
- Beygi, S., Kafashan, M., Bahrami, H. R., & Mugler, D. H. (2012). The iterative shrinkage method for impulsive noise reduction from images. *Measurement Science and Technology*, *23*(11), 114009.
- 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.
- Chen, S. S. B., Donoho, D. L., & Saunders, M. A. (1998). Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, *20*(1), 33–61.
- Chen, S. S. B., Donoho, D. L., & Saunders, M. A. (2001). Atomic decomposition by basis pursuit. *SIAM Review*, *43*(1), 129–159.
- Chui, C. K., & He, W. J. (2000). Compactly supported tight frames associated with refinable functions. *Applied and Computational Harmonic Analysis*, *8*(3), 293–319.
- Donoho, D. L. (2006). For most large underdetermined systems of linear equations the minimal $l(1)$ -norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, *59*(6), 797–829.
- Donoho, D. L., & Huo, X. M. (2001). Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, *47*(7), 2845–2862.
- Feng, Z. P., & Chu, F. L. (2007). Application of atomic decomposition to gear damage detection. *Journal of Sound and Vibration*, *302*(1–2), 138–151.
- Gharavian, M. H., Ganj, F. A., Ohadi, A. R., & Bafroui, H. H. (2013). Comparison of fda-based and pca-based features in fault diagnosis of automobile gearboxes. *Neurocomputing*, *121*, 150–159.
- Grbovic, M., Li, W. C., Xu, P., Usadi, A. K., Song, L. M., & Vucetic, S. (2012). Decentralized fault detection and diagnosis via sparse pca based decomposition and maximum entropy decision fusion. *Journal of Process Control*, *22*(4), 738–750.
- 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, S. G., He, Z., & Wang, G. A. (2013). Online monitoring and fault identification of mean shifts in bivariate processes using decision tree learning techniques. *Journal of Intelligent Manufacturing*, *24*(1), 25–34.
- Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, *13*(2), 415–425.
- Humberstone, M., Wood, B., & Henkel, J. (2012). Differentiating between expanded and fault conditions using principal component analysis. *Journal of Intelligent Manufacturing*, *23*(2), 179–188.
- Jiang, L., Xuan, J. P., & Shi, T. L. (2013). Feature extraction based on semi-supervised kernel marginal fisher analysis and its application in bearing fault diagnosis. *Mechanical Systems and Signal Processing*, *41*(1–2), 113–126.
- Klein, R., Ingman, D., & Braun, S. (2001). Non-stationary signals: Phase-energy approach theory and simulations. *Mechanical System and Signal Processing*, *15*(6), 1061–1089.
- Konar, P., & Chattopadhyay, P. (2011). Bearing fault detection of induction motor using wavelet and support vector machines (svms). *Applied Soft Computing*, *11*(6), 4203–4211.
- Kovacevic, J., & Chebira, A. (2007). Life beyond bases: The advent of frames (part II). *IEEE Signal Processing Magazine*, *24*(5), 115–125.
- Li, R., & Frogley, M. (2013). On-line fault detection in wind turbine transmission system using adaptive filter and robust statistical features. *International Journal of Prognostics and Healthy Management Special issue on Wind Turbine PHM*, *4*(19), 120–129.
- Li, R. Y., Sopon, P., & He, D. (2012). Fault features extraction for bearing prognostics. *Journal of Intelligent Manufacturing*, *23*(2), 313–321.
- Lin, J., & Zuo, M. J. (2003). Gearbox fault diagnosis using adaptive wavelet filter. *Mechanical Systems and Signal Processing*, *17*(6), 1259–1269.
- Liu, B., Ling, S. F., & Gribonval, R. (2002). Bearing failure detection using matching pursuit. *Ndt and E International*, *35*(4), 255–262.
- Lorena, A. C., & de Carvalho, A. C. P. L. F. (2008). Evolutionary tuning of SVM parameter values in multiclass problems. *Neurocomputing*, *71*, 3326–3334.
- Mallat, S., & Zhang, Z. (1993). Matching pursuit with timefrequency dictionaries. *IEEE Transactions on Signal Processing*, *41*(12), 3397–3415.
- Marjanovic, G., & Solo, V. (2012). On $l(q)$ optimization and matrix completion. *IEEE Transactions on Signal Processing*, *60*(11), 5714–5724.
- Mateos, G., Bazerque, J. A., & Giannakis, G. B. (2010). Distributed sparse linear regression. *IEEE Transactions on Signal Processing*, *58*(10), 5262–5276.
- Michailovich, O. V. (2011). An iterative shrinkage approach to total-variation image restoration. *IEEE Transactions on Image Processing*, *20*(5), 1281–1299.
- Mori, K., Kasashima, N., Yoshioka, T., & Ueno, Y. (1996). Prediction of spalling on a ball bearing by applying the discrete wavelet transform to vibration signals. *Wear*, *195*(1–2), 162–168.
- 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.
- Pandya, D. H., Upadhyay, S. H., & Harsha, S. P. (2014). Fault diagnosis of rolling element bearing by using multinomial logistic regression and wavelet packet transform. *Soft Computing*, *18*(2), 255–266.
- Peng, Z. K., Tse, P. W., & Chu, F. L. (2005). A comparison study of improved hilbert-huang transform and wavelet transform: Application to fault diagnosis for rolling bearing. *Mechanical Systems and Signal Processing*, *19*(5), 974–988.
- Plumbley, M. D., Abdallah, S. A., Blumensath, T., & Davies, M. E. (2006). Sparse representations of polyphonic music. *Signal Processing*, *86*(3), 417–431.
- Rai, V. K., & Mohanty, A. R. (2007). Bearing fault diagnosis using FFT of intrinsic mode functions in hilbert-huang transform. *Mechanical Systems and Signal Processing*, *21*(6), 2607–2615.

- Rajpoot, K., Rajpoot, N., & Noble, J. A. (2008). Discrete wavelet diffusion for image denoising. *Image and Signal Processing*, 5099, 20–28.
- Rusu, C., & Dumitrescu, B. (2012). Stagewise k-svd to design efficient dictionaries for sparse representations. *IEEE Signal Processing Letters*, 19(10), 631–634.
- Selesnick, I. W. (2011). Wavelet transform with tunable q-factor. *IEEE Transactions on Signal Processing*, 59(8), 3560–3575.
- Shao, R. P., Hu, W. T., Wang, Y. Y., & Qi, X. K. (2014). The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet transform. *Measurement*, 54, 118–132.
- 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.
- Wang, S. B., Huang, W. G., & Zhu, Z. K. (2011). Transient modeling and parameter identification based on wavelet and correlation filtering for rotating machine fault diagnosis. *Mechanical Systems and Signal Processing*, 25(4), 1299–1320.
- Wells, L. J., Megahed, F. M., Niziolek, C. B., Camelio, J. A., & Woodall, W. H. (2013). Statistical process monitoring approach for high-density point clouds. *Journal of Intelligent Manufacturing*, 24(6), 1267–1279.
- Yang, H. Y., Mathew, J., & Ma, L. (2005). Fault diagnosis of rolling element bearings using basis pursuit. *Mechanical Systems and Signal Processing*, 19(2), 341–356.
- Yu, H. C., Lin, K. Y., & Chien, C. F. (2014). Hierarchical indices to detect equipment condition changes with high dimensional data for semiconductor manufacturing. *Journal of Intelligent Manufacturing*, 25(5), 933–943.
- Zhang, X. H., Deng, Z. H., Liu, W., & Cao, H. (2013). Combining rough set and case based reasoning for process conditions selection in camshaft grinding. *Journal of Intelligent Manufacturing*, 24(2), 211–224.
- 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.
- Zhao, D. L., Ma, W., & Liang, W. K. (2005). On data fusion fault diagnosis and simulation of hydroelectric units vibration. *Proceedings of the CSEE*, 25(20), 137–42.