IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 6, 2016, accepted January 8, 2017, date of publication March 1, 2017, date of current version May 17, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2675940

A Fault Diagnosis Approach for Rolling Element Bearings Based on RSGWPT-LCD Bilayer Screening and Extreme Learning Machine

QINGBIN TONG[1], JUNCI CAO[1], BAOZHU HAN[1], XIAODONG ZHANG[1], ZHENGWEI NIE[2], JIAMIN WANG[2], YUYI LIN[2], AND WEIDONG ZHANG[3]

[1]School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China
[2]Department of Mechanical and Aerospace Engineering, University of Missouri, Columbia, MO 65211 USA
[3]State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing 102206, China

Corresponding author: Q. Tong (tqbin818@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51577007, and in part by the State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources under Grant LAPS15019.

ABSTRACT The faults of rolling element bearings can result in the deterioration of machine operating conditions; how to assess the working condition and identify the fault of the rolling element bearing has become a key issue for ensuring the safe operation of modern rotating machineries. This paper presents a novel hybrid approach that detects bearing faults and monitors the operating status of rolling element bearings in modern rotating machineries. Based on redundant second-generation wavelet packet transform and local characteristic-scale decomposition , this method is implemented to extract the fault features, the vibration signal is adaptively decomposed into a number of desired intrinsic scale components by two-step screening processes based on the energy ratio, and reduce random noises and eliminate the pseudofrequency components. The fault features are then used to implement the identification classification of faults using singular value decomposition and extreme learning machine. The approach is evaluated by simulation and practical bearing vibration signals under different conditions. The experiment results show that the proposed approach is feasible and effective for the fault diagnosis of rolling element bearing.

INDEX TERMS Rolling element bearings, fault diagnosis, redundant second generation wavelet packet transform (RSGWPT), local characteristic-scale decomposition (LCD), energy ratio, singular value decomposition (SVD), extreme learning machine (ELM).

I. INTRODUCTION

Rolling element bearings are regarded as one of the most common components in modern rotating machinery. The failure of rolling element bearings can result in the deterioration of machine operating conditions [1]. Therefore, how to assess the working condition and identify the fault of the rolling element bearing has become a key issue for ensuring the safe operation of modern rotating machineries. [2]. Vibration signal contains abundant fault information that a series of impact impulses will occur when a rolling element bearing possesses a localized fault [3], [4]. Therefore, vibration-based analysis is widely applied in the condition monitoring and fault diagnosis of rolling element bearings [5]–[7]. Usually, original signals are not adequate to identify the existence of a fault; therefore, fault features are extracted from the time domain, frequency domain, or time-frequency domain analysis [8]. However, the vibration signal of rolling element bearings is nonstationary signal. It is very difficult to identify the fault of the rolling element bearing only in the time domain and the frequency domain analysis. Time-frequency domain methods, such as spectral kurtosis (SK) and wavelet transform (WT), are able to process nonlinear and nonstationary signals.

Spectral kurtosis (SK) analysis has been used as an effective method for detecting localized faults of rolling element bearing from the nonstationary signals. SK can identify the frequency band that contains fault bearing information by detecting impulse series. However, the values of SK strongly depend on the choice of the frequency resolution of the estimator when dealing with nonlinear and nonstationary signals. The extreme case where the frequency resolution is infinite leads systematically to a zero SK as a consequence

of the Central Limit Theorem. A coarse frequency resolution may not allow the SK to detect a narrow-band transient buried in additive stationary broad-band noise. Wavelet transform (WT) is a multi-resolution analysis method and used widely in the field of machinery condition monitoring and fault detection, which possesses good local property in both time domain and frequency domain [9], [10]. However, WT cannot split the high frequency band where the modulation information of machine fault always exists. Wavelet package transform (WPT) is an extension of WT, which can split the high frequency band and possesses better time-frequency localization of signals [11]–[13]. Second generation wavelet transform (SGWT) is a new wavelet construction method using lifting scheme [14], [15]. The resolution of SGWT varies with the decomposition levels. It gives good time and poor frequency resolution at high-frequency sub-band, and good frequency and poor time resolution at low-frequency sub-band. In order to obtain a higher resolution in the high-frequency sub-band in which the faulty characteristics always exist, second generation wavelet packet transform (SGWPT) has been constructed and the detail coefficients at each level are further decomposed to obtain their approximation and detail components. However, SGWPT does not have time invariant property in the signal decomposition. This may result in the loss of useful fault information for feature extraction and fault diagnosis. The redundant second generation wavelet package transform (RSGWPT) possesses time invariant property and overcomes the disadvantage of lifting scheme, by getting rid of the split step and zero padding of prediction operator and update operator, which makes the approximation and detail signals at all levels are the same length as the original signal [16]–[19]. RSGWPT not only has more detailed local time-frequency description of the signal, but also inhibit the frequency aliasing components of the analysis result because of the absence of the split and merge step in the decomposition and reconstruction stage. Consequently, the statistical features extracted from the coefficients of RSGWPT have a greater ability to detect the fault signal.

Empirical mode decomposition (EMD) provides a different method to nolinear and nonstationary signal processing in time-frequency domain. It is based on the local characteristic time scales and could self-adaptively decompose the signal into a set of intrinsic mode functions (IMFs) [20]. IMFs represent the local characteristic information of the original signal. Therefore, the EMD method has received more attention for its application in condition identification and fault diagnosis of rolling element bearings [21], [22]. However, when the EMD method is applied to the nolinear and nonstationary signal, the original signal cannot be accurately decomposed because of mode mixing, the end effects and the unexplainable negative frequency. Wu and Huang developed ensemble empirical mode decomposition (EEMD) to improve EMD, by adding noise to the original signal and calculating the means of the IMFs repeatedly [23]. Local mean decomposition (LMD) is a self-adaptive signal processing method [24]. Some researchers have applied LMD to the fault diagnosis of rotating machine and illustrated the LMD method is better than the EMD method in mode mixing, end effects, etc [25]–[27]. Nevertheless, LMD is inherently dedicated for the decomposition of the AM-FM signals; when used in the processing of impact signals, the algorithm usually does not converge. Recently, a new signal decomposition method called local characteristic-scale decomposition (LCD) was proposed, the original signal is decomposed into a set of intrinsic scale components (ISCs) by defining a new base line and iteration-stopping criterion. Furthermore, it is verified that LCD has an advantage over EMD in processing nonlinear and non-stationary signals because LCD can reduce invalid components and mode mixing [28], [29].

Artificial neural network (ANN) has once been used in fault diagnosis and provided satisfying results. However, with insufficient samples, the inherent deficiencies such as underfitting and local optimal solution will become obvious issues. Therefore, the application of ANN is limited. Support vector machine (SVM) is an outstanding method of pattern recognition, which is based on statistical learning theory and especially suitable for small sample cases. Additionally, SVM has the better generalization ability compared to ANN, and it can guarantee the local and global optimal solution are exactly the same. However, SVM still cannot provide a perfect solution due to the low sparsity of the model and the margin trade-off parameter must be estimated. Recently, extreme learning machine (ELM) has been originally developed based on single-hidden layer feed-forward neural networks (SLFNs) with the weights in hidden nodes randomly assigned and the output weights decided by the Moore-Penorose pseudoinverse. It has been shown that ELM has an extremely high learning speed and a good generalization performance than other traditional methods.

Many faults of the rolling element bearing in rotating machines produce series of impacts causing by spalling faults, cracks or excessive clearances, which in turn generate transient vibration signals. Some methods have been developed to detect the fault characteristics and identify the faults of rolling element bearing by means of specific processing. However, a consistent difficulty is that available vibration signals are often severely corrupted by strong levels of background noise. The background noise often embodies strong vibrations from several competing sources (e.g. harmonics of rotating parts, random impulses from friction and contact forces, flow noise, etc.) which span a large frequency range and seriously mask the signal of interest.

In order to more accurately extract the fault characteristics and identify the operating conditions, a fault diagnosis approach for rolling element bearings based on RSGWPT, LCD and ELM is presented in this paper. The RSGWPT with first-stage screening processes based on the energy ratio is taken as the pre-filter process unit. Then, the selected feature packets will be decomposed by LCD with second-stage screening processes. Applying the spectrum analysis on those desired ISCs, the fault characteristics are

easily extracted. Finally, singular value decomposition (SVD) is used to decompose the matrix which consists of desired ISCs to generate feature vectors. Feature vectors are input to ELM to specify the fault type.

II. REDUNDANT SECOND GENERATION PACKET WAVELET TRANSFORM

Second generation wavelet transform (SGWT) is a new wavelet construction method using lifting scheme in the time domain. The decomposition stage of SGWT consists of three steps: split, predict and update. In the split stage, the signal a_j is split into even samples and odd samples at level j.

$$a_{j+1} = a_j(2k), \quad d_{j+1} = a_j(2k+1) \tag{1}$$

Then a designed prediction operator P is applied on a_{j+1} to predict odd samples. The difference between predicted d_{j+1} and odd samples is regarded as the detail coefficients of a_j. The prediction step is described as follows:

$$d_{j+1}(k) = d_{j+1}(k) - \sum_{i=-N/2+1}^{N/2} p_i a_{j+1}(k+i) \tag{2}$$

where p_i are the coefficients of prediction operator P with the length of N.

In the update step, an update operator U is designed and applied on d_{j+1} to update even samples. The sum of the result and even samples a_{j+1} is regarded as the approximate coefficients of a_j.

$$a_{j+1}(k) = a_{j+1}(k) + \sum_{r=-M/2+1}^{M/2} u_r d_{j+1}(k+r-1) \tag{3}$$

where u_r are the coefficients of update operator U with the length of M.

The detail and approximation coefficients at different levels can be obtained by iterating the above three steps on the output approximate signal a.

Reversely, the reconstruction procedure of SGWT includes inverse update stage, inverse prediction stage and merging stage.

$$\begin{cases} a_{j+1}(k) = a_{j+1}(k) - \sum_{r=-M/2+1}^{M/2} u_r d_{j+1}(k+r-1) \\ d_{j+1}(k) = d_{j+1}(k) + \sum_{i=-N/2+1}^{N/2} p_i a_{j+1}(k+i) \\ a_j(2k) = a_{j+1} \\ a_j(2k+1) = d_{j+1} \end{cases} \tag{4}$$

The prediction operator P and the update operator U are essential parameters in SGWT. The operators P and U are built by means of interpolating subdivision method (ISM) [41]. The selection of different P and U is equivalent to choosing different biorthogonal wavelet filters [42]. The decomposition and reconstruction stages of SGWT are depicted in Fig. 1.

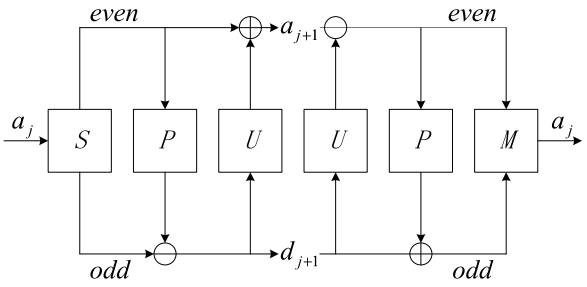

FIGURE 1. Decomposition and reconstruction stages of SGWT.

Second generation wavelet packet transform (SGWPT) is generated based on SGWT which only performs the multi-resolution analysis in low frequency bands. SGWPT is conducted to obtain a higher resolution in the high frequency band, and hence the detail coefficients at each level are further decomposed to obtain their approximation and detail components.

The decomposition and reconstruction of SGWPT are expressed as below. The decomposition of SGWPT includes three steps: split, prediction and update. In the first stage of decomposition, $X_{j,m}$ is split into even samples $X_{j,me}$ and odd samples $X_{j,mo}$,

$$X_{j,me} = X_{j,m}(2k), \quad X_{j,mo} = X_{j,m}(2k+1) \tag{5}$$

where $X_{j,m}$ represents the coefficients of the m-th node at decomposition level j.

Then the prediction step and update step of SGWPT at level $j+1$ are performed according to the following formula:

$$\begin{cases} X_{j+1,2} = X_{j,1o} - P\left(X_{j,1e}\right) \\ X_{j+1,1} = X_{j,1e} + U\left(X_{j+1,2}\right) \\ \quad \cdots \\ X_{j+1,2^{j+1}} = X_{j,2^j o} - P\left(X_{j,2^j e}\right) \\ X_{j+1,2^{j+1}-1} = X_{j,2^j e} + U\left(X_{j+1,2^{j+1}}\right) \end{cases} \tag{6}$$

The reconstruction procedure of SGWPT consists of inverse update step, inverse prediction step and merging step. In the reconstruction procedure, the subband coefficients to be reconstructed are reserved, and then other subband coefficients are set to be zeroes. The reconstruction procedure is expressed as below:

$$\begin{cases} X_{j,2^j e} = X_{j+1,2^{j+1}-1} - U\left(X_{j+1,2^{j+1}}\right) \\ X_{j,2^j o} = X_{j+1,2^{j+1}} + P\left(X_{j,2^j e}\right) \\ X_{j,2^j}(2i) = X_{j,2^j e} \\ X_{j,2^j}(2i+1) = X_{j,2^j o} \\ \quad \cdots \\ X_{j,1e} = X_{j+1,1} - U\left(X_{j+1,2}\right) \\ X_{j,1o} = X_{j+1,2} + P\left(X_{j,1e}\right) \\ X_{j,1}(2i) = X_{j,1e} \\ X_{j,1}(2i+1) = X_{j,1o} \end{cases} \tag{7}$$

Overall, Fig. 2 illustrates the decomposition and the reconstruction procedure of SGWPT.

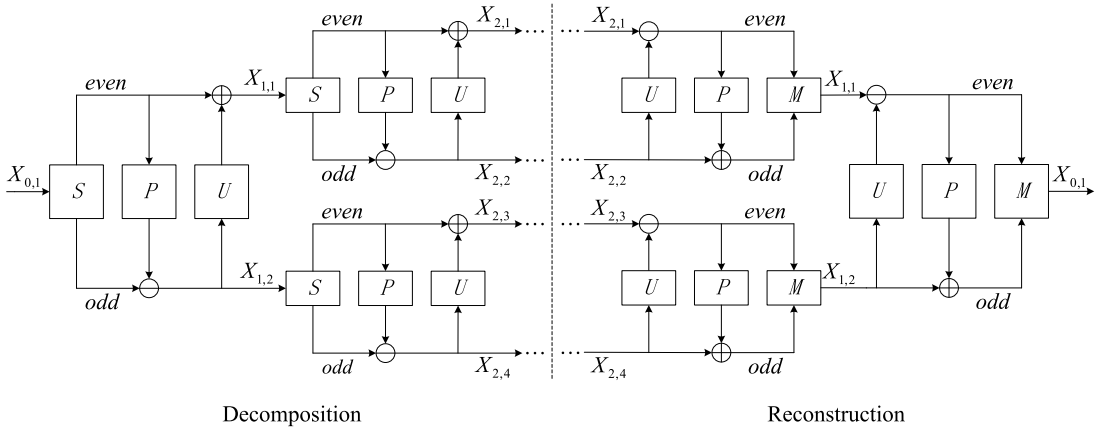

FIGURE 2. Decomposition and reconstruction stages of SGWPT.

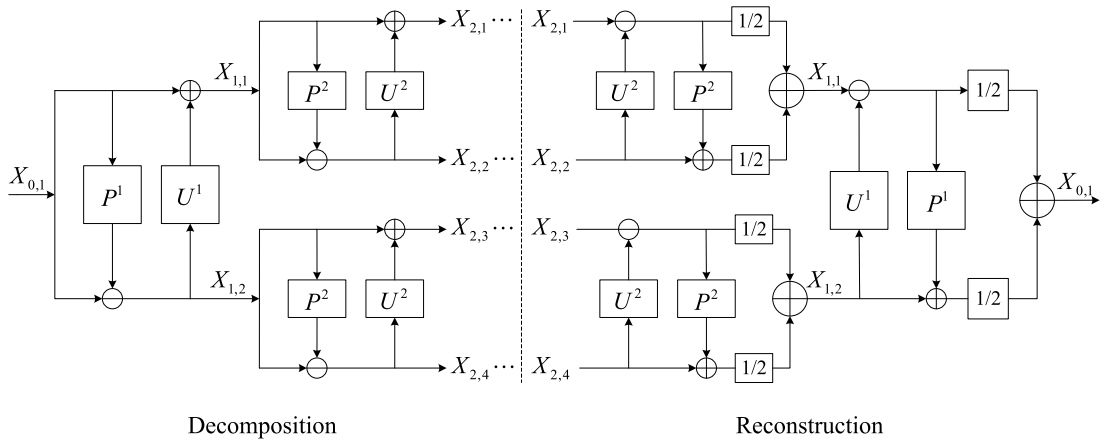

FIGURE 3. Decomposition and reconstruction stages of RSGWPT.

The RSGWPT is constructed by applying the redundant lifting scheme on the SGWPT. In the redundant lifting scheme, P^j and U^j represent the prediction and update operators of the redundant lifting scheme at level j, the coefficients of P^j and U^j are obtained by padding P_i and U_r of initial operator P and U with zeroes [16].

$$p_i^j = p_0^0, \underbrace{0, \ldots, 0}_{2^j-1}, p_1^0, \underbrace{0, \ldots, 0}_{2^j-1}, p_2^0, \ldots, p_{N-2}^0, \underbrace{0, \ldots, 0}_{2^j-1}, p_{N-1}^0 \tag{8}$$

$$u_r^j = u_0^0, \underbrace{0, \ldots, 0}_{2^j-1}, u_1^0, \underbrace{0, \ldots, 0}_{2^j-1}, u_2^0, \ldots, u_{M-2}^0, \underbrace{0, \ldots, 0}_{2^j-1}, u_{M-1}^0 \tag{9}$$

The decomposition process of RSGWPT at level j are performed by using P^j and U^j, which are expressed as follows.

$$\begin{cases} X_{j+1,2} = X_{j,1} - P^{j+1}\left(X_{j,1}\right) \\ X_{j+1,1} = X_{j,1} + U^{j+1}\left(X_{j+1,2}\right) \\ \quad \cdots \\ X_{j+1,2^{j+1}} = X_{j,2^j} - P^{j+1}\left(X_{j,2^j}\right) \\ X_{j+1,2^{j+1}-1} = X_{j,2^j} + U^{j+1}\left(X_{j+1,2^{j+1}}\right) \end{cases} \tag{10}$$

The reconstruction procedure of RSGWPT can be obtained from its decomposition procedure, which consists of inverse update step, inverse prediction step and merging step.

$$\begin{cases} X_{j,2^j}^u = X_{j+1,2^{j+1}-1} - U^{j+1}\left(X_{j+1,2^{j+1}}\right) \\ X_{j,2^j}^p = X_{j+1,2^{j+1}} + P^{j+1}\left(X_{j,2^j}^u\right) \\ X_{j,2^j} = \dfrac{1}{2}\left(X_{j,2^j}^u + X_{j,2^j}^p\right) \\ \quad \cdots \\ X_{j,1}^u = X_{j+1,1} - U^{j+1}\left(X_{j+1,2}\right) \\ X_{j,1}^p = X_{j+1,2} + P^{j+1}\left(X_{j,1}^u\right) \\ X_{j,1} = \dfrac{1}{2}\left(X_{j,1}^u + X_{j,1}^p\right) \end{cases} \tag{11}$$

The forward and inverse transform of RSGWPT are shown in Fig. 3. Owing to the coefficients at all levels have the same length as that of the input signal, the decomposition results of RSGWPT possess time invariant property and keep the information of the raw signal perfectly.

III. LOCAL CHARACTERISTIC-SCALE DECOMPOSITION

Local characteristic-scale decomposition (LCD) method is a self-adaptive time-frequency analysis approach. According to the local characteristic scale, LCD presents a definition

of a mono-component with physical meaning. A complex signal can be decomposed into a sum of intrinsic scale components (ISCs) which are independent of each other and a residue by defining a base line and iteration-stopping criterion. That is:

$$x(t) = \sum_{i=1}^{n} ISC_i(t) + r_n(t) \tag{12}$$

where $ISC_i(t)$ is the i-th ISC and $r_n(t)$ is the residue. As a constant sequence of monotonic sequence, $r_n(t)$ represents the average trend of the signal generally. The ISC component c_i $(i = 1, \ldots, n)$ represents a signal of different frequencies from high to low. Based on the definition of the ISC component, the signal $x(t)$ can be decomposed into numbers of ISCs by using the LCD method in the following way [28]:

(1) Suppose (t_k, x_k) $(k = 1, \ldots, M)$ express all the extrema of the signal $x(t)$, then calculate A_k $(k = 2, \ldots, M-1)$ and corresponding L_k $(k = 2, \ldots, M-1)$ according to Eq. (11), respectively.

$$A_k = x_{k-1} + \frac{t_k - t_{k-1}}{t_{k+1} - t_{k-1}} (x_{k+1} - x_{k-1}) \tag{13}$$

$$L_k = aA_k + (1-a)\, x_k, \quad k = 2, \ldots, M-1 \tag{14}$$

where a = 0.5. When calculating A_1 and A_M, the boundary of the data would be extended [30]. By the extension of two end extrema, (t_0, x_0) and (t_{M+1}, x_{M+1}) can be obtained. We can get L_1 and L_M from Eqs. (11).

(2) All the $L_k (k = 1, \ldots, M)$ are connected by a cubic spline as the baseline, noted as $SL_1(t)$. By contrast, the base line is defined as the mean of the upper envelope and the lower envelope. The difference between the original signal and the base line $SL_1(t)$ is the first component $h_1(t)$. Theoretically, $h_1(t)$ should be the first ISC.

$$h_1(t) = x(t) - SL_1(t) \tag{15}$$

If $h_1(t)$ meets the conditions to be the ISC, take $h_1(t)$ as the first ISC of $x(t)$; otherwise

(3) Take $h_1(t)$ as the original signal and repeat the above process until $h_m(t)$ satisfies the ISC conditions after m iterations of the computation. Then, $h_m(t)$ is expressed as the first ISC $ISC_1(t)$.

(4) Separate the first ISC from $x(t)$ and the residue is recorded as $r_1(t)$:

$$r_1(t) = x(t) - ISC_1(t) \tag{16}$$

(5) Take the $r_1(t)$ as the original signal to be processed and repeat the above steps until the residue $r_n(t)$ satisfies the three threshold criterion. So $x(t)$ is decomposed into n ISCs and a residue as shown in Eq. (16). As an improvement method to SD criterion [31], the three threshold criterion [32] is adopted for the sifting process to guarantee that the ISC components retain enough physical sense of both the amplitude and frequency modulations, namely, meet the definition.

IV. BILAYER FEATURE SCREENING METHOD BASED ON RSGWPT AND LCD

When there is a fault in the rolling element bearings, the useful characteristic frequency components generally contain very little energy and are submerged by background noise. The background noise has a wide frequency band and higher-level macrostructural vibrations, and it is difficult to extract the useful fault characteristic frequency components. To solve this problem, the RSGWPT with first-stage screening processes based on the energy ratio is taken as the pre-filter process unit to reduce random noises, and decomposes the signal into a series of narrow frequency bands and enhances the weak fault characteristic components in the different narrow frequency bands. Then, the selected feature packets will be decomposed by LCD, in which second-stage screening processes are proposed to eliminate the pseudo low-frequency components of ISCs. Applying the spectrum analysis on those desired ISCs, the fault characteristics are easily extracted.

The discrete signal $x(t)$, can be written as

$$x(t) = \left\{ X_{0,1}(k) \,|k = 1, 2, \cdots, K \right\} \tag{17}$$

where K is the length of the signal. By computing j-level RSGWPT, the signal $x(t)$ is decomposed into 2^j sets of sub-band coefficients $C_{j,m}$ with the length of $K/2^j$. The sequence of these sets at the j-th level is $m = 1^0, 2^1, \cdots, 2^j$ and each set of these coefficients can be viewed as a node in a binary wavelet packet tree. The decomposition coefficients $C_{j,m}$ can be given.

$$C_{j,m} = \left\{ X_{j,m}(k) \,\middle|\, k = 1, 2, \cdots K/2^j \right\} \tag{18}$$

Each node of the RSGWPT tree is indexed with a pair of integers (j, m) where j is the corresponding level of decomposition and m is the order of the node position in the specific level. From the decomposition coefficients $X_{j,m}$, a reconstructed signal $p_j^m(t)$ of length N can be obtained by implementing the redundant second generation wavelet packet reconstruction. In reconstruction procedure, the coefficients to be reconstructed are reserved, and the other coefficients in level j are set to be zeroes, and $j_{\max} \leq \log_2 K$.

The redundant ISCs have not only lower frequency than the lowest frequency of the signal $x(t)$ but also less amplitude. This can be explained that the redundant ISCs are generated due to leaking in the decomposing process of RSGWPT. Therefore, if the energy of $v_i(t)$ is almost equal to that of $x(t)$, $h_i(t)$ is viewed as the redundant reconstruction signal. The energy ratio of $v_i(t)$ to $x(t)$ will be constructed to eliminate the redundant ISCs.

It is defined as

$$L_i = \frac{\int v_i^2(t)dt}{\int x^2(t)dt} = \frac{\int [x(t) - h_i(t)]^2\, dt}{\int x^2(t)dt} \tag{19}$$

If L_i is less than T_0 (in general, $0.95 \leq T_0 < 1$), $h_i(t)$ $(1 \leq i \leq m)$ are viewed as the real ISC. The energy ratio criterion is logical to eliminate the redundant ISCs. It is based on the fact that the redundant ISCs has a lower frequency and less amplitude than the lowest frequency of the signal $x(t)$.

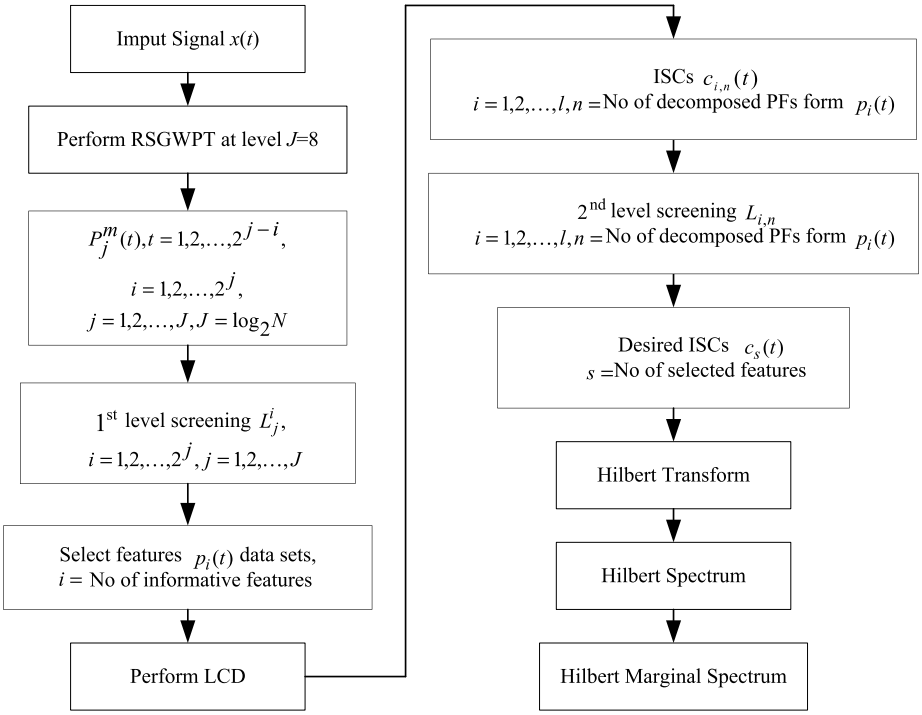

FIGURE 4. Flow chart of proposed RSGWPT-LCD based two stages screening processes.

Next, the significance of feature packets can be assessed quantitatively by introducing the 1^{st} level screening processes which use the energy ratio L_j^m between $v_{j,m}^2(t)$ and signal $x^2(t)$ as given [33]:

$$L_j^m = \frac{\int v_{j,m}^2 dt}{\int x^2(t)dt} = \frac{\int \left[x(t) - p_j^m(t)\right]^2 dt}{\int x^2(t)dt} \tag{20}$$

where $v_{j,m}(t)$ is the difference between the input signal $x(t)$ and $p_j^m(t)$. On the other hand, $p_j^m(t)$ represent time series wavelet packet coefficients with m-th packet on j-th resolution.

When the 1^{st} screening process is finished, the process of LCD will be applied to decompose the selected feature packet $p_i(t)$ data sets. Next, the resultant ISCs from each selected feature packet is stored as preselected ISCs $C_{i,n}(t)$ data sets. Subsequently, the energy ratio $L_{i,n}$ as given in Eq. (21), is used in the 2nd level screening processes. The purpose of 2nd screening processes is to identify and select the most vital ISCs from the list of ISCs $C_{i,n}(t)$ data sets.

$$L_{i,n} = \frac{\int v_{j,m}^2 dt}{\int p^2(t)dt} = \frac{\int \left[p(t) - c_{i,n}(t)\right]^2 dt}{\int p_i^2(t)dt} \tag{21}$$

In order to extract vital signal features, computing energy ratio is an important step as part of this algorithm. For those with energy ratio $L_{i,n}$ lower than 0.99 thresholds, their corresponding $C_{i,n}(t)$ is viewed as a real ISC and subsequently stored in the data sets $C_s(t)$. Those ISCs, which are stored in the data sets $C_s(t)$, are defined as ''desired ISCs''. Otherwise, it is recognized as a redundant ISC and therefore will be eliminated. Finally, using Hilbert transform, it can be obtained the Hilbert spectrum and Hilbert marginal spectrum. The flow chart of the proposed RSGWPT-LCD based two stages screening processes is summarized in Fig. 4.

V. FAULT SINGULAR VALUE EXTRACTION BASED ON SVD

According to the matrix theory, singular values generated by SVD present the inherent feature of matrix and possess the characteristics of favorable stability, scale invariance and rotating invariance. Hence, the singular values of the matrix whose rows are desired ISCs are very feasible to be the feature vector for ELM training and testing.

The matrix consist of a set of desired ISCs is divided to two initial feature matrice A and B, they are described respectively by

$$\begin{aligned} A &= \begin{bmatrix} c_1 & c_2 & \cdots & c_J \end{bmatrix}^T, \\ B &= \begin{bmatrix} c_{J+1} & c_{J+2} & \cdots & c_S \end{bmatrix}^T \end{aligned} \tag{22}$$

If S is even, $J \leq S/2$. Otherwise, $J \leq (S+1)/2$. The substantive features of the disturbing signal are characterized by the matrix A and B.

Let C be a real matrix with N rows and M columns. There exist orthogonal matrices U and V with the size of $N \times N$ and $M \times M$ such that [37]

$$C = U \Lambda V^T \tag{23}$$

where Λ is a $N \times M$ diagonal matrix with nonnegative diagonal elements. These diagonal elements $\sigma_i (i = 1, 2, \cdots p)$,

$p = \min(N, M)$ arranged in descending order are termed the singular values of the matrix C.

Then the initial feature matrices A and B are processed by the SVD respectively, the obtained singular values $\sigma_{A,j}$ and $\sigma_{B,j}$ are expressed as follows:

$$\sigma_{A,j} = \left[\sigma_{A,j}^1, \sigma_{A,j}^2, \cdots, \sigma_{A,j}^J\right] \tag{24}$$

$$\sigma_{B,j} = \left[\sigma_{B,j}^{J+1}, \sigma_{B,j}^{J+2}, \cdots, \sigma_{B,j}^S\right] \tag{25}$$

where $\sigma_{A,j}^1 \geq \sigma_{A,j}^2 \geq \cdots \geq \sigma_{A,j}^J$, $\sigma_{B,j}^{J+1} \geq \sigma_{B,j}^{J+2} \geq \cdots \geq \sigma_{B,j}^S$. The vector $\left[\sigma_{A,j}, \sigma_{B,j}\right]$ is chosen as the feature vector.

VI. EXTREME LEARNING MACHINE

Huang et al. present the ELM method, which is a single step constrained optimization solution for single hidden layer feed forward networks and extended for non-neuron like networks [34], [35]. This method is superior to the traditional iterative gradient decent method in training speed.

Suppose a training dataset with q classes $\{x_i, y_i\}$, $i = 1, \ldots, N$ and $y_i \in R^q$. $x_i \in R^d$ is a d-dimensional data point. If $G(x)$ is an infinitely differentiable activation function in the hidden layer, the output of SLFN is given by

$$o_i = \sum_{j=1}^{L} \beta_j G\left(a_j, b_j, x_i\right) \tag{26}$$

where L is the number of hidden nodes, a_j and b_j are the jth hidden node's learning parameters assigned randomly. $\beta_j \in R^q$ is the output weight vector. The SLFN output of the matrix form can be presented as below

$$O = H\beta \tag{27}$$

where

$$H = \begin{bmatrix} G(a_1, b_1, x_1) & \cdots & G(a_L, b_L, x_1) \\ \vdots & \ddots & \vdots \\ G(a_1, b_1, x_N) & \cdots & G(a_L, b_L, x_N) \end{bmatrix}_{N \times L} \tag{28}$$

H is called the hidden layer output matrix of the network. H_{ij} represents the jth hidden node with respect to the input samples x_i. $\beta = [\beta_1, \beta_2, \cdots, \beta_L]^T$ and $O = [o_1, o_2, \cdots, o_N]^T$. Therefore, the objective function of ELM can be written as follows

$$\beta = H^+ Y = (H^T H)^{-1} H^T Y \tag{29}$$

where $Y = [y_1, y_2, \cdots, y_N]^T$ and H^+ is the Moore-Penrose generalized inverse of matrix H. The above output weight matrix β minimizes the cost function $\|O - Y\|$.

VII. EXPERIMENTAL ANALYSIS AND DISCUSSION

A. SIMULATION ANALYSIS

In order to verify the effectiveness of the proposed method in this paper, a simulation experiment is designed and carried out. When fault exist in the roller bearing, due to the effect of period impulse, the high-frequency shock vibration is then generated and the amplitude is modulated. To simulate this phenomenon, a simulated signal is designed. The simulated signal is sampled at 1024Hz, and it consists of two carrier frequency components (150 and 400Hz) in time intervals [0, 0.5] and [0.5, 1] which are modulated with 20 and 60Hz. In practice, noise existing in the original signal often behaves like white noise; it covers the entire frequency domain. A white noise with signal to noise ratio of -1dB is added into the input signal. The simulated signal $x(t)$ is expressed as:

$$x(t) = \begin{cases} (1 + \cos(2\pi 20 t_1)) \cdot \cos(2\pi 150 t_1) & 0 \leq t_1 < 0.5 \\ (1 + \cos(2\pi 60 t_2)) \cdot \cos(2\pi 400 t_2) & 0.5 \leq t_2 < 1 \\ White\ noise(S/N = -1dB) \end{cases} \tag{30}$$

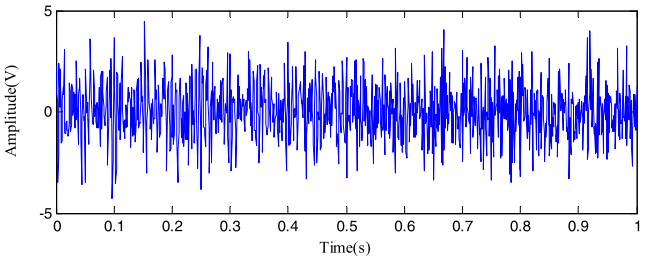

FIGURE 5. Time domain plot of the simulated signal.

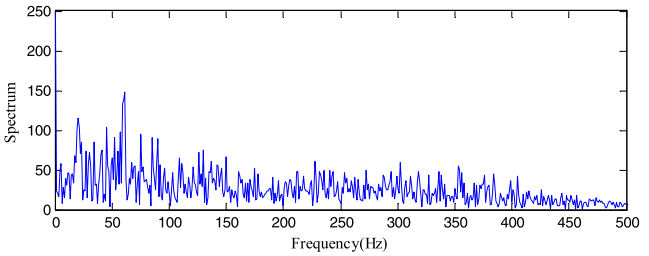

FIGURE 6. Envelope spectrum of simulated signal.

The simulated signal is shown in Fig.5. Obviously the frequency components are masked by the noise, and we can hardly find the features of the modulation in Fig.5. Envelope analysis is one of the most important methods for fault detection in frequency domain. The envelope signal $x_e(t)$ is obtained by using Hilbert transform to demodulate the simulated signal. Theoretically, the main frequency components of the envelope signal should be the modulating frequency (20Hz, 60Hz). The envelope spectrum of the simulated signal is displayed in Fig. 6. Because of a large number of frequency components, it is difficult to clearly detect the modulating frequency of 20Hz and 60Hz from this figure. Therefore, the traditional method in the frequency domain no longer applies to extract such modulation features buried in the heavy background noise.

The simulated signal is processed using RSGWPT, LCD and the proposed RSGWPT-LCD respectively to demonstrate the effectiveness of the energy ratio selection criterion and the proposed method. Fig. 7, 8 and 9 illustrate the processing

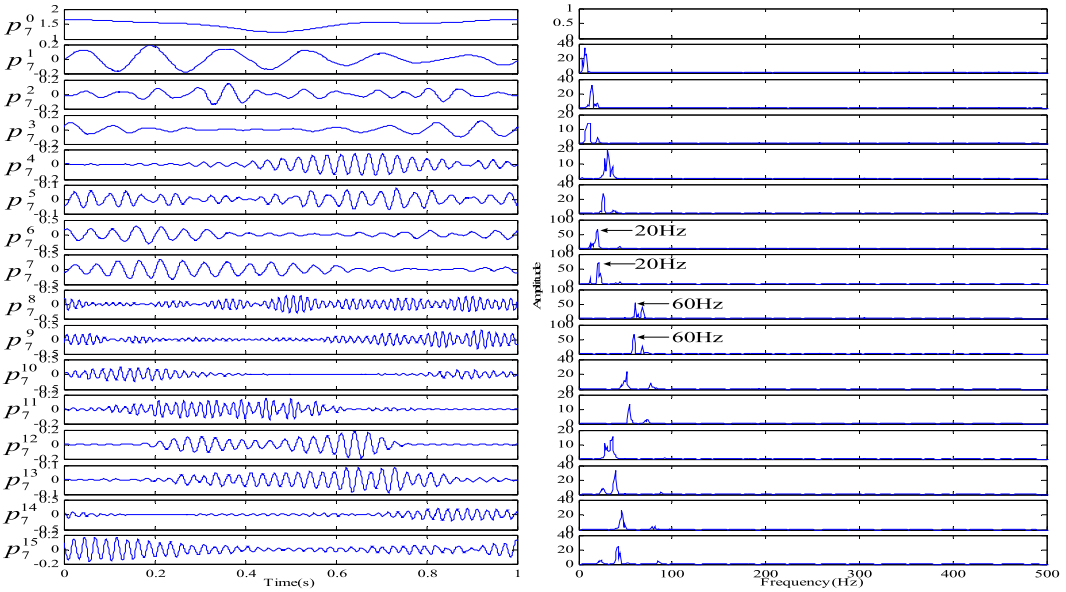

FIGURE 7. Decomposition results with 8-level RSGWPT and corresponding envelope spectrums.

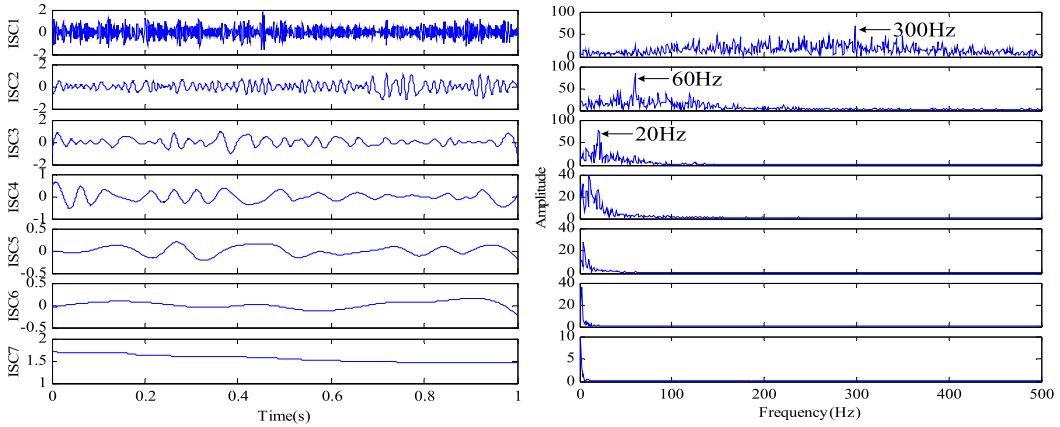

FIGURE 8. Decomposition results with LCD and corresponding envelope spectrums of ISCs.

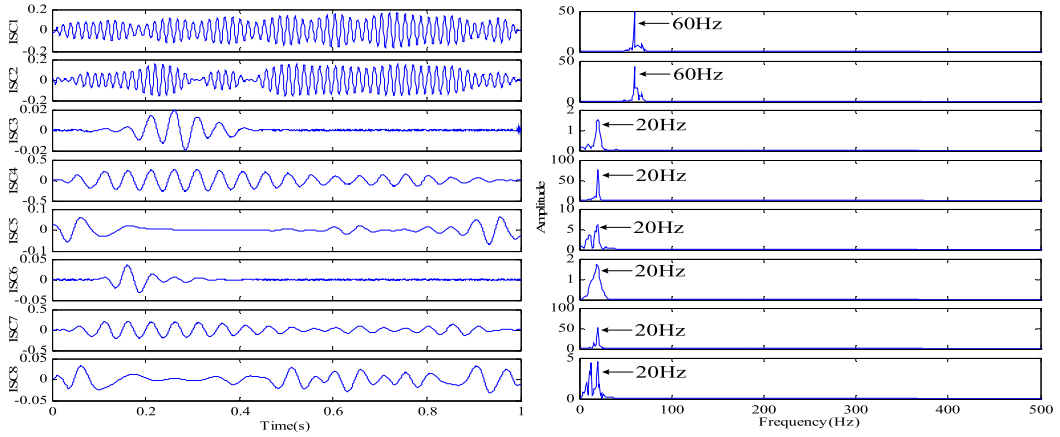

FIGURE 9. Decomposition results with proposed RSGWPT-LCD and corresponding envelope spectrums.

results and corresponding envelope spectrums. The calculation results of the energy ratio are listed in Tables 1 and 2.

By computing the 8-level RSGWPT on the signal $x(t)$, there are 256 sets of sub-band reconstructed signals. To reduce the length of this paper, only first 16 reconstructed signals are shown. From the decomposition results with RSGWPT and corresponding envelope spectrums shown in Fig. 5, it can be seen that only four reconstructed signals

TABLE 1. Energy ratios between envelope signal and each sub-band reconstructed signal.

Reconstructed signal	p_7^0	p_7^1	p_7^2	p_7^3	p_7^4	p_7^5	p_7^6	p_7^7
Energy ratio	0.9989	0.9969	0.9968	0.9984	0.9978	0.9986	0.9858	0.9867
								(continue Table 1)
Reconstructed signal	p_7^8	p_7^9	p_7^{10}	p_7^{11}	p_7^{12}	p_7^{13}	p_7^{14}	p_7^{15}
Energy ratio	0.9884	0.9890	0.9980	0.9996	0.9982	0.9963	0.9983	0.9969

TABLE 2. Energy ratios between filtered signal and each ISC.

Method	Energy Ratio							
	ISC1	ISC 2	ISC 3	ISC 4	ISC 5	ISC 6	ISC 7	ISC 8
LCD	0.8382	0.9057	0.9690	1.0112	1.0066	0.9977	1.0030	/
Proposed method	0.0094	0.0089	0.9540	0.0298	0.9333	0.9303	0.0273	0.9274

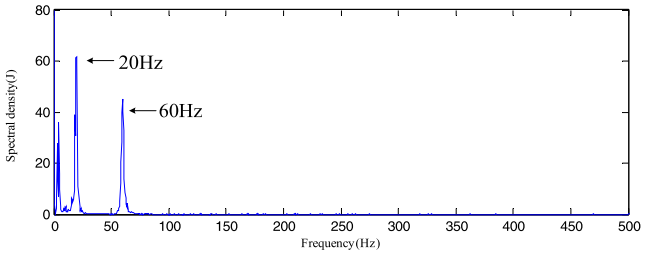

FIGURE 10. Hilbert marginal spectrum of simulated signal using RSGWPT-EMD.

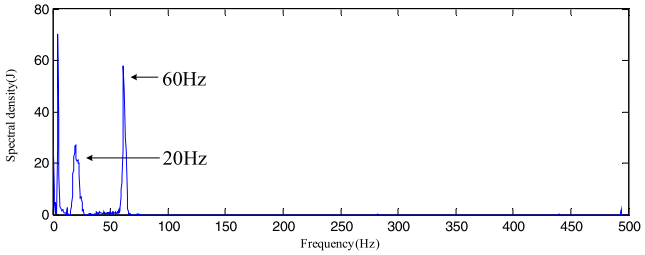

FIGURE 11. Hilbert marginal spectrum of simulated signal using WPT-LCD.

mainly contain the modulating frequency components. For example, the reconstructed signals $p_7^6(t)$ and $p_7^7(t)$ mainly contain the modulating frequency 20 Hz. The modulating frequency 60 Hz exists in the signal $p_7^8(t)$ and $p_7^9(t)$. It means that only four of first 16 reconstructed signals are meaningful signal features. And the rest are redundant reconstructed signals that consist of useless frequency components, which may mislead our identification.

For the calculated energy ratio listed in Table. 1, there are four energy ratio values of sub-band reconstructed signals are lower than 0.99. With energy ratio selection criterion, these reconstructed signals are regarded as the meaningful features, which coincide with the analyzed results of the decomposition results shown in Fig. 7. It indicates that the sub-band reconstructed signals with low energy ratio are more informative than other reconstructed signals.

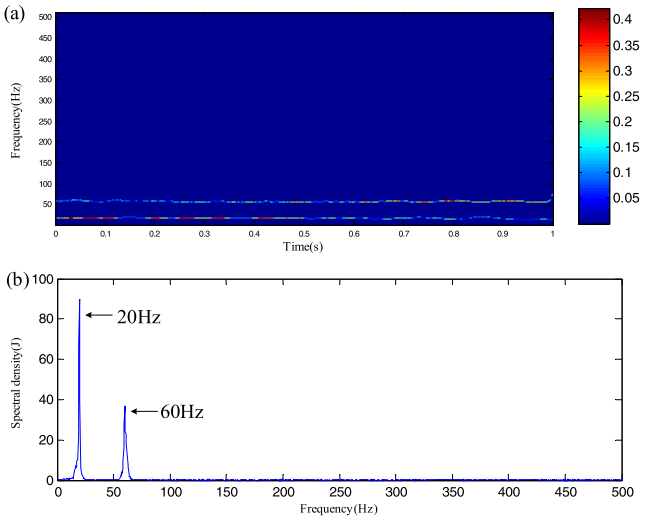

FIGURE 12. (a) Hilbert spectrum of simulated signal using proposed method. (b) Hilbert marginal spectrum of simulated signal using proposed method.

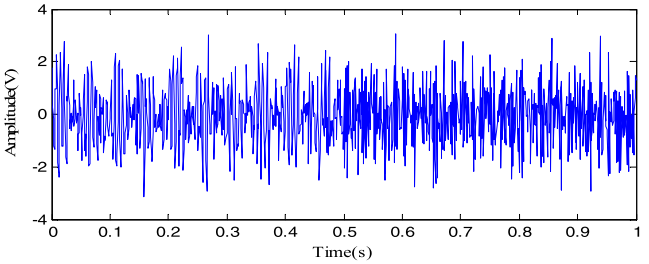

FIGURE 13. Time domain plot of the simulated signal under $S/N = 3$dB.

The LCD decomposition results of the signal and their envelope spectrum is shown in Fig. 8. As illustrated in Fig. 8, ISC1contains the frequency multiplication of modulating frequency while ISC2 and ISC3 contain the modulating frequency 60 Hz and 20 Hz. However, the rest of the PRCs consist of a lot of redundant low-frequency components. So, only ISC1, ISC2 and ISC3 are real components of the

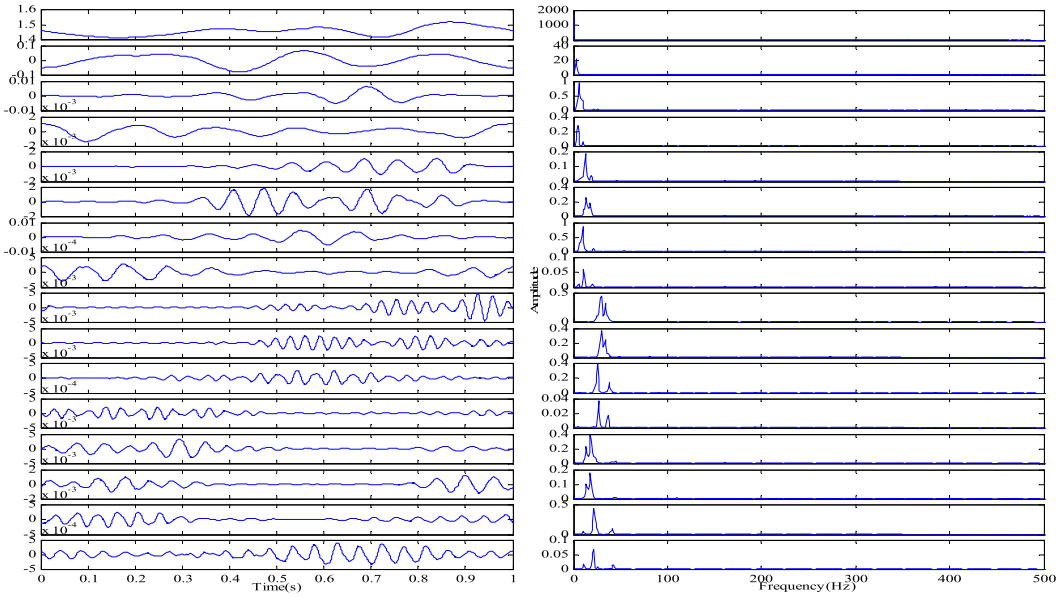

FIGURE 14. Decomposition results with 8-level RSGWPT and corresponding envelope spectrums under $S/N = 3$dB.

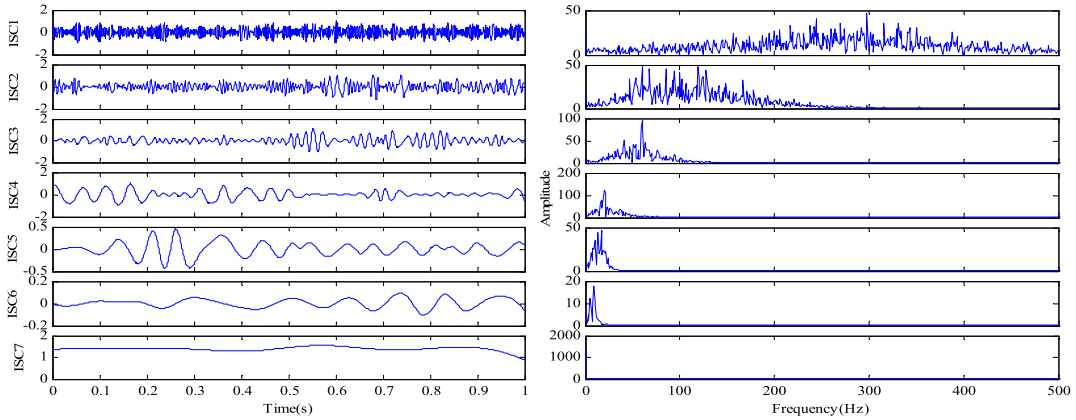

FIGURE 15. Decomposition results with LCD and corresponding envelope spectrums of ISCs under $S/N = 3$dB.

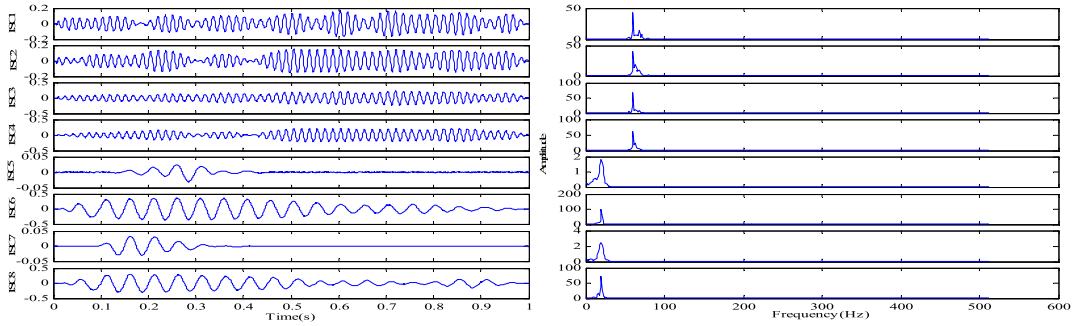

FIGURE 16. Decomposition results with RSGWPT-LCD and corresponding envelope spectrums under $S/N = 3$dB.

signal and the others are the pseudo components. The pseudo components contain meaningless frequency, which may mislead our analysis. Meanwhile, the noise existed in the ISCs may result in ambiguous identification results.

Considering calculated energy ratio for LCD listed in Table 2, it can be seen that the energy ratio values of the first three ISCs are lower than 0.99. With energy ratio selection criterion, the first two ISCs are viewed as the meaningful

ISCs and others are the redundant ISCs, which coincide with the analyzed results of the decomposition results shown in Fig. 7. It can be found that the ISCs with low energy ratio are more informative than other ISCs.

The results obtained by RSGWPT and LCD cannot fully present the real frequency pattern of the signal. Then the proposed method is applied to analyze the envelope signal. The desired ISCs and their frequency spectrums are illustrated in Fig. 9. It can be seen that the proposed method is able to clearly detect the modulating frequency 20 and 60 Hz. Besides, there are no redundant components in the decomposition results of proposed method. Different from the results of RSGWPT and LCD method, in which exist many redundant components and most of the components are highly corrupted with noise, the RSGWPT-LCD technique can effectively remove noise of the signal and extract the frequency feature of the signal. From Table 2, it can be seen that all the desired ISCs generated by the RSGWPT-LCD have low energy ratio. According to energy ratio selection criterion, all the desired ISCs are meaningful ISCs, which have been demonstrated by analyzing the decomposition results shown in Fig. 9. It can be found that the processing results of RSGWPT-LCD are superior to that of RSGWPT and LCD.

To further examine the effectiveness of the proposed method, two stages screening processes based on the wavelet packet transform (WPT) and LCD will be applied for comparison, which is designated as WPT-LCD. Another comparison is with respect to the RSGWPT and empirical mode decomposition (EMD) where two stages screening processes are also used, designated as RSGWPT-EMD. The analyzed results are shown in Fig. 10 and 11. Compared with envelope spectrum, the processing results of RSGWPT-EMD and WPT-LCD are able to present the modulating frequencies of 20Hz and 60Hz. However, they also contain some meaningless information, which may effect the fault detection.

The Hilbert spectrum and Hilbert marginal spectrum using the proposed RSGWPT-LCD are presented in Fig. 12. The results show that the modulating frequencies of 20Hz and 60Hz can easily be identified. Considering the spectrum by other methods, the spectrum based on RGSWPT-LCD presents a much clearer message about the frequency components embedded in the simulated signal. Furthermore, RGSWPT-LCD has successfully removed the noise related frequency components.

In order to further verify the performance of the proposed method under different noise levels, the comparative experiments of the simulated signal are given. The simulated signals under $S/N = 3$dB are shown in Fig.13.

The simulated signal is processed using RSGWPT, LCD and the proposed RSGWPT-LCD respectively. Fig. 14, 15 and 16 illustrate the processing results and corresponding envelope spectrums under $S/N = 3$dB. According to the processing results and corresponding envelope spectrums under different noise levels, it can be seen that the processing results of RSGWPT contain the meaningful and redundant reconstructed signal features in the first 16 reconstructed signals. The LCD decomposition results of the signal contain the modulating frequency 60 Hz and 20 Hz. However, the modulation frequency is not obvious, and submerged in the useless frequency. The proposed RSGWPT-LCD can clearly detect the modulating frequency 20 and 60 Hz. Besides, there are no redundant components in the decomposition results.

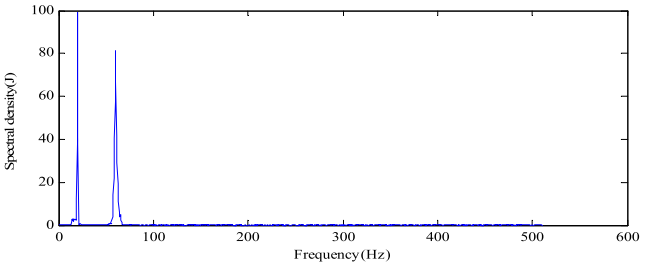

FIGURE 17. Hilbert marginal spectrum of signal simulated using proposed method. under $S/N = 3$dB.

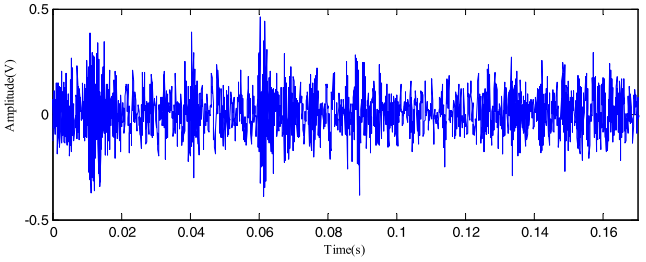

FIGURE 18. Time domain plot of signal with rolling element defect.

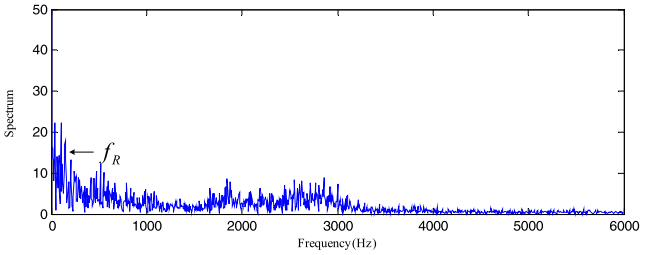

FIGURE 19. Envelope spectrum of signal with rolling element defect.

Different from the results of RSGWPT and LCD method under different noise levels, in which exist many redundant components and most of the components are highly corrupted with noise, the RSGWPT-LCD technique can effectively remove noise of the signal and extract the frequency feature of the signal. The Hilbert marginal spectrum using the proposed RSGWPT-LCD is presented in Fig. 17. The processing results show that the modulating frequencies of 20Hz and 60Hz can easily be identified. The spectrum based on the proposed method presents a much clearer message about the frequency components embedded in the simulated signal. Furthermore, RGSWPT-LCD has successfully removed the noise related frequency components.

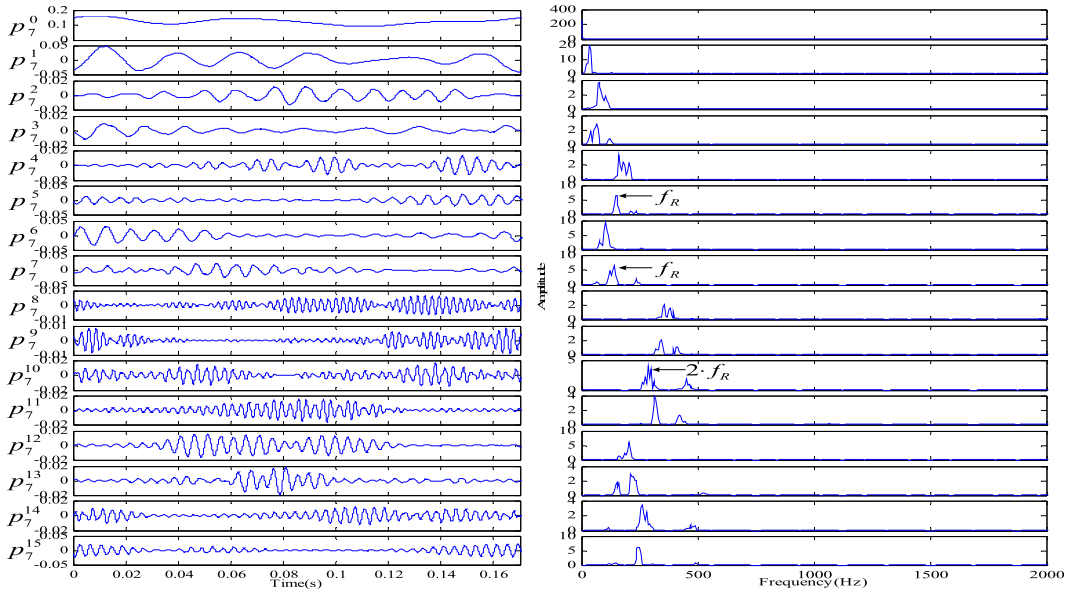

FIGURE 20. Decomposition results with rolling element defect and corresponding envelope spectrum using RSGWPT.

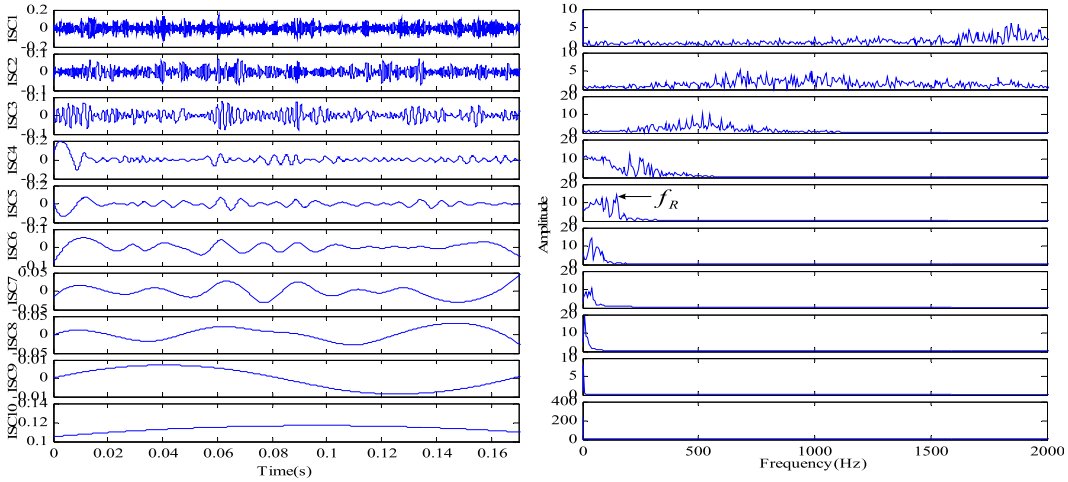

FIGURE 21. Decomposition results with rolling element defect and corresponding envelope spectrum using LCD.

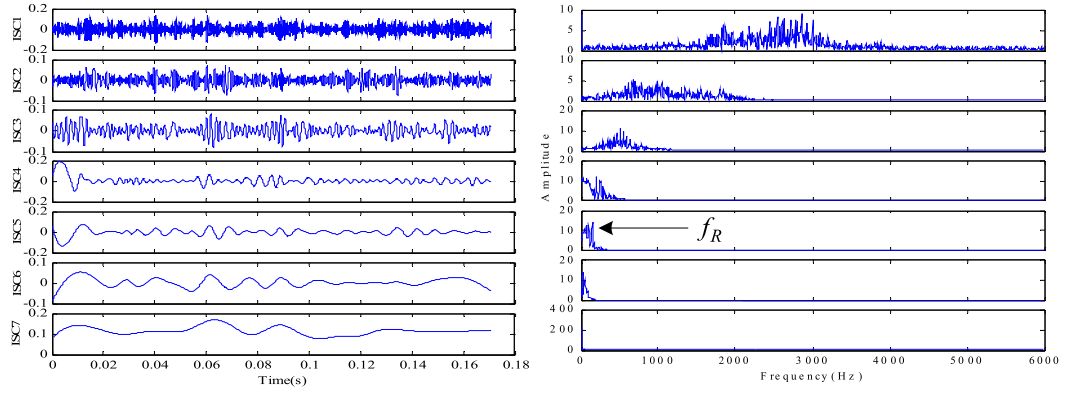

FIGURE 22. Decomposition results with rolling element defect and corresponding envelope spectrum using LMD.

B. FAULT DIAGNOSIS OF ROLLING ELEMENT BEARINGS

In order to further verify the proposed method, the actual experiment on fault identification of rolling bearings is conducted in this section. The vibration data of rolling bearings are provided by Case Western Reserve Lab (CWRU) [36]. The test stand consists of a 2hp,

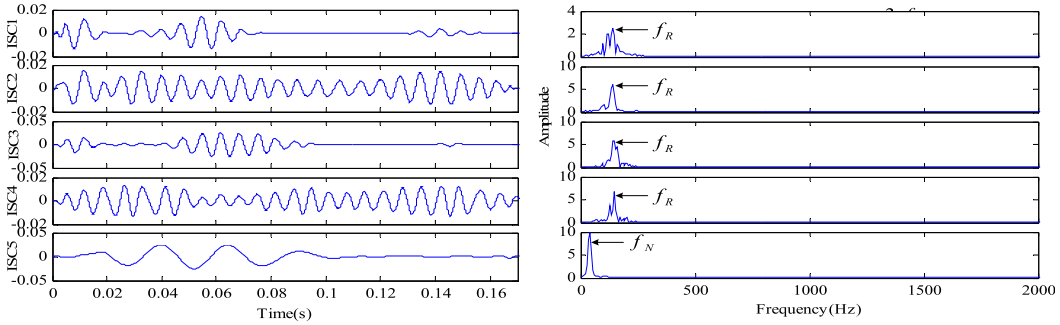

FIGURE 23. Decomposition results with rolling element defect and corresponding envelope spectrum using RSGWPT-LCD.

three-phase induction motor, a torque transducer/encoder and a dynamometer. Fault locations cover inner raceway, outer raceway (at 6:00, 3:00 and 12:00 directions) and rolling element, with the same defect sizes of 0.007 or 0.021 in. Vibration data was collected at the sampling frequency of 12 KHz by using accelerometers which are mounted at the drive end of the motor. The characteristic frequency of the rolling element defect is calculated to be at 141.17 Hz, the inner race defect and the outer race defect is 162.19 Hz and 107.37 Hz respectively based on the geometric parameters.

Fig. 18 illustrates vibration waveform of rolling element defect under 0 hp load with diameters of 7 mills. The diagnosis of fault for rolling elements may be the most challenging task in bearing health condition monitoring due to its non-stationary and nonlinear characteristic. While operating a roller bearing with local fault, the impulse is created, the high-frequency shock vibration is then generated and the vibration amplitude is modulated by the impulse force. In order to obtain the fault characteristic frequency, the vibration signal of roller bearing need to demodulate. Now, the vibration signal is directly demodulated by the Hilbert transform, and Fig. 19 shows the demodulation spectrum. From the demodulation spectrum, it can be seen that the characteristic frequency is submerged in heavy background signals relevant to the rotary speed of rotor and other noise, which may bring errors to the identification.

Then the envelope signal is decomposed by the RSGWPT, LCD and LMD respectively. Fig.20 presents the first 16 sub-band reconstructed signals of the 8-level RSGWPT method and their envelope spectrum. From Fig. 20, it can be seen that although some reconstructed signals (p_7^5, p_7^7 and p_7^{10}) contain the fault characteristic frequencies, there are many meaningless components in which there is no obvious line at the positions corresponding to $i \cdot f_R$, $i = 1, 2, 3, \cdots$. Meanwhile, there still exists a degree of frequency mixing phenomenon in the p_7^7, p_7^{10} and p_7^{13}, etc. With LCD method, the 1th-10th ISC components are obtained as shown in Fig. 21. It reveals that only the fifth ISC contains the fault characteristic frequency information (141 Hz). The first four ISCs contain frequency multiplication, but the useful information is buried in heavy background. The rest ISCs are irrelevant

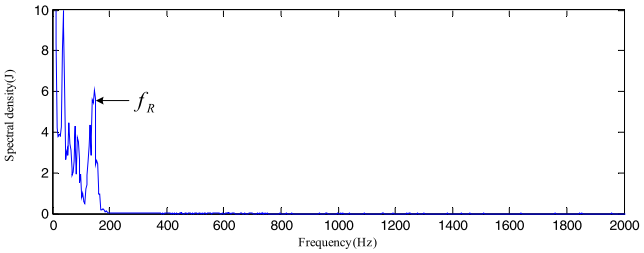

FIGURE 24. Hilbert marginal spectrum of rolling element defect using RSGWPT-EMD.

or noise related. Besides, the noise existed in the ISCs would bring inaccurate diagnosis results. Fig.22 presents the seven sub-band reconstructed signals of LMD decomposition and their envelope spectrums. It can be seen from Fig.22 that only the fifth decomposition signal contains the fault characteristic frequency, there are many meaningless frequency components in which there is no obvious line at the positions corresponding to $i \cdot f_R$, $i = 1, 2, 3, \cdots$. Simultaneously, the useful information is buried in heavy background.

Through RSGWPT-LCD method, 5 desired ISCs was acquired. Fig.23 depicts the decomposition results as well as their corresponding envelope spectrum. From Fig.23 it can be found that there are obvious spectral lines at the position of the rolling element fault characteristic frequency f_R(141Hz). According to the analysis results, it can be determined there are defects on the rolling element without any trouble. Compared to RSGWPT and LCD, the RSGWPT-LCD can not only eliminate the redundant components but also reduce the effect of noise. It is clear that the proposed method outperforms RSGWPT and LCD in extracting the characteristic frequency component.

In addition, the following methods are adopted to demonstrate the effectiveness of the proposed method: RSGWPT-EMD, WPT-LCD and LMD. Fig. 24, 25, 26 and 27 illustrate the processing results of RSGWPT-EMD, WPT-LCD, LMD and proposed method. Compared to the results of RSGWPT-EMD, WPT-LCD and LMD, the defect feature of proposed method is more obvious and clear. The proposed RSGWPT-LCD technique provides the best diagnostic information, due to its efficient noise reduction and adap-

TABLE 3. Faults identification of based on relevant method.

Data set	The number of training samples	The number of testing samples	Defect size (inches)	Operating condition	Label of classification
A	20	20	0	Normal	1
	20×3	20×3	0.007	Three types defect	2,3,4
B	20	20	0	Normal	1
	20×3	20×3	0.021	Three types defect	2,3,4
C	20	20	0	Normal	1
	20×3	20×3	0.007	Three types defect	2,3,4
	20×3	20×3	0.021	Three types defect	5,6,7

TABLE 4. Average classification accuracy of ELM based on relevant method.

Algorithms	Classification accuracy of data set A (%)	Classification accuracy of data set B (%)	Classification accuracy of data set C (%)
LMD+ELM	96.63	96.32	91.64
LCD+ELM	96.81	96.54	91.87
RSGWPT+ELM	96.86	96.59	91.85
WPT-LCD+ELM	97.89	97.81	95.03
RSGWPT-EMD+ELM	97.95	97.95	95.87
DRSGWPT-LCD+ELM	99.65	99.58	99.45

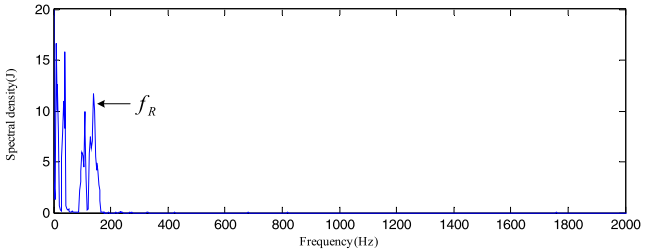

FIGURE 25. Hilbert marginal spectrum of rolling element defect using WPT-LCD.

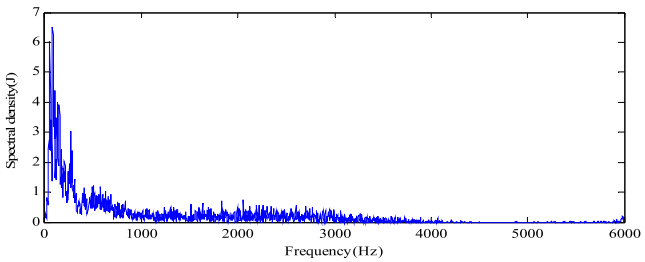

FIGURE 26. Hilbert marginal spectrum of rolling element defect using LMD.

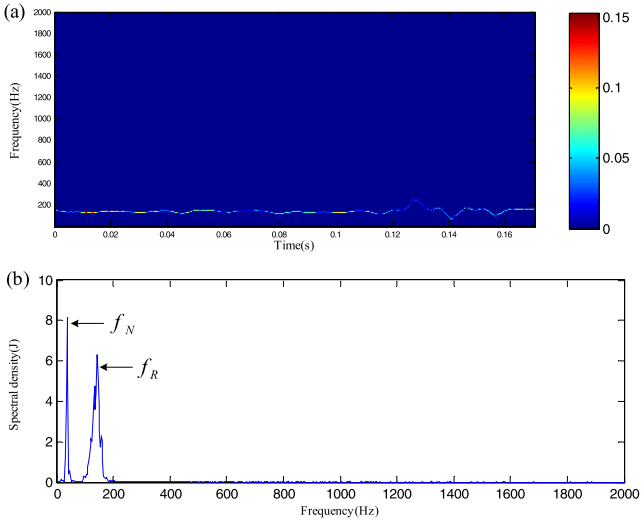

FIGURE 27. (a) Hilbert spectrum of rolling element defect using proposed method (b) Hilbert marginal spectrum of rolling element defect using proposed method.

tive ISC selection using the proposed two-stage screening. Although the RSGWPT-EMD (Fig. 24) and WPT-LCD technique (Fig. 25) can also detect 141Hz, other strong frequency components on the spectral map may result in false diagnostic result.

In order to provide an intuitional distinguish result, ELM is used to diagnose the bearing condition after RSGWPT-LCD. For each vibration signal that represents one working condition, the first 120000 points are divided into 40 sub-signals. As a result, 40 samples are acquired for each working condition, and there are 7×40 samples in total. SVD technique is applied to the feature matrix which consists of desired ISCs to generate the feature vectors. Consequently, for each working condition, a matrix of the feature vector is generated.

The feature vectors are divided into training samples and testing samples. The detailed descriptions of the three data sets are shown in Table 3. To show the superiority of the feature extraction by SVD based on RSGWPT-LCD, some comparison are done. For avoiding the tweak problem, each experiment is repeated 30 times, and the average classification accuracy of three data sets is listed in Table 4.

For all the data sets, it can be noticed that the proposed method provides better classification accuracy (99.65%, 99.58%, 99.45%) compared with RSGWPT (96.86%, 96.59%, 91.85%), LCD (96.81%, 96.54%, 91.87%) and LMD (96.63%, 96.32%, 91.64%). The RSGWPT-LCD

combines advantage of RSGWPT and LCD, and at the same time it eliminates most of redundant features based on two-stage screening processes. Hence, the feature extraction ability of the proposed method is significantly stronger than the previous two methods. Moreover, the classification accuracy of RSGWPT-LCD+ELM (99.65%, 99.58%, 99.45%) is higher than those of the WPT-LCD (97.89%, 97.81%, 95.03%) and RSGWPT-EMD (97.95%, 97.95%, 95.87%). Thus, the features extracted by SVD based on RSGWPT-LCD are more representative than features extracted by SVD based on WPT-LCD, RSGWPT-EMD and LMD.

In the case of data sets C, the classification accuracies of all the method are lower than that of data set A or B, which can be explained by the larger number of classes (seven classes) than data sets A or B (four classes). With the class number increasing the advantage of RSGWPT-LCD+ELM is more prominent. It is clear that the performance of RSGWPT-LCD+ELM is significant better than that of other methods. In data set A or B, 4 working conditions were considered, the test accuracies of proposed method are also higher than the other methods but it is not as significant as in data set C.

VIII. CONCLUSION

The faults of the rolling element bearing in rotating machines produce series of impacts causing by spalling faults, cracks or excessive clearances. Fault vibration signals are often severely corrupted by strong levels of background noise which span a large frequency range and seriously mask the signal of interest. In order to accurately evaluate the working condition and identify the fault of the rolling element bearing, a novel method based on RSGWPT, LCD and ELM for fault diagnosis of rolling element bearings is presented. The RSGWPT with first-stage screening processes based on the energy ratio is taken as the pre-filter process unit to reduce random noises, and decomposes the signal into a series of narrow frequency bands and enhances the weak fault characteristic components in the different narrow frequency bands. The selected feature packets will be decomposed by LCD, in which second-stage screening processes are proposed to eliminate the pseudo low-frequency components of ISCs. Applying the spectrum analysis on those desired ISCs generated by the proposed method, the fault features are easily extracted. Simulation experiment is implemented with a signal containing different frequency components and a practical signal of defective bearings with rolling element fault is also utilized. The processing results are compared with that of RSGWPT-EMD and WPT-LCD. Results show that the proposed method is more accurate and effective than other related methods in extracting the fault feature of vibration signal. Finally, seven classes of fault datasets are used in fault diagnosis, the feature vectors are generated by applying SVD to the matrix whose rows are desired ISCs and identified by ELM. The diagnosis results are compared with related method in testing identification accuracy. The results show that the proposed method performs well for fault diagnosis of rolling element bearing.

REFERENCES

[1] L. Yu, Z. Junhong, B. Fengrong, L. Jiewei, and M. Wenpeng, "A fault diagnosis approach for diesel engine valve train based on improved ITD and SDAG-RVM," *Meas. Sci. Technol.*, vol. 26, no. 2, Dec. 2015, Art. no. 025003.

[2] W. Sui, S. Osman, and W. Wang, "An adaptive envelope spectrum technique for bearing fault detection," *Meas. Sci. Technol.*, vol. 25, no. 9, Jul. 2014, Art. no. 095004.

[3] F. Jiang, Z. Zhu, W. Li, G. Chen, and G. Zhou, "Robust condition monitoring and fault diagnosis of rolling element bearings using improved EEMD and statistical features," *Meas. Sci. Technol.*, vol. 25, no. 2, Dec. 2014, Art. no. 025003.

[4] Y. G. Lei, J. Lin, Z. J. He, and Y. Y. Zi, "Application of an improved kurtogram method for fault diagnosis of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 25, no. 5, pp. 1738–1749, Jan. 2011.

[5] X. Liu, L. Bo, X. He, and M. Veidt, "Application of correlation matching for automatic bearing fault diagnosis," *J. Sound Vib.*, vol. 331, no. 26, pp. 5838–5852, Dec. 2012.

[6] B. Muruganatham, M. A. Sanjith, B. Krishnakumar, and S. A. V. S. Murty, "Roller element bearing fault diagnosis using singular spectrum analysis," *Mech. Syst. Signal Process.*, vol. 35, no. 1, pp. 150–166, Dec. 2013.

[7] W. Wang and H. Lee, "An energy kurtosis demodulation technique for signal denoising and bearing fault detection," *Meas. Sci. Technol.*, vol. 24, no. 2, Jan. 2013, Art. no. 025601.

[8] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, Oct. 2006.

[9] W. J. Wang and P. D. McFadden, "Application of wavelets to gearbox vibration signals for fault detection," *J. Sound Vib.*, vol. 192, no. 5, pp. 927–939, May 1996.

[10] Z. K. Peng and F. L. Chu, "Application of the wavelet transform in machine condition monitoring and fault diagnostics: A review with bibliography," *Mech. Syst. Signal Process.*, vol. 18, no. 2, pp. 199–221, Mar. 2004.

[11] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*, vol. 4. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[12] X. Fan and M. J. Zuo, "Gearbox fault detection using Hilbert and wavelet packet transform," *Mech. Syst. Signal Process.*, vol. 20, no. 4, pp. 966–982, May 2006.

[13] Q. S. Xu and Z. G. Lia, "Recognition of wear mode using multi-variable synthesis approach based on wavelet packet and improved three-line method," *Mech. Syst. Signal Process.*, vol. 21, no. 8, pp. 3146–3166, Nov. 2007.

[14] W. Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets," *Appl. Comput. Harmon. Anal.*, vol. 3, no. 2, pp. 186–200, Apr. 1996.

[15] W. Sweldens, "The lifting scheme: A construction of second generation wavelets," *SIAM J. Math. Anal.*, vol. 29, no. 2, pp. 511–546, Aug. 1998.

[16] N. Li, R. Zhou, Q. Hu, and X. Liu, "Mechanical fault diagnosis based on redundant second generation wavelet packet transform, neighborhood rough set and support vector machine," *Mech. Syst. Signal Process.*, vol. 28, no. 2, pp. 608–621, Apr. 2012.

[17] R. Zhou, W. Bao, N. Li, X. Huang, and D. Yu, "Mechanical equipment fault diagnosis based on redundant second generation wavelet packet transform," *Digit. Signal Process.*, vol. 20, no. 1, pp. 276–288, Jan. 2010.

[18] C. S. Lee, C. K. Lee, and K. Y. Yoo, "New lifting based structure for undecimated wavelet transform," *Electron. Lett.*, vol. 36, no. 22, pp. 1894–1895, Oct. 2000.

[19] Z. Li, Z. He, Y. Zi, and X. Chen, "Bearing condition monitoring based on shock pulse method and improved redundant lifting scheme," *Math. Comput. Simul.*, vol. 7, no. 3, pp. 318–338, Dec. 2008.

[20] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London Ser. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[21] J. Cheng, D. Yu, J. Tang, and Y. Yang, "Application of SVM and SVD technique based on EMD to the fault diagnosis of the rotating machinery," *Shock Vib.*, vol. 16, no. 1, pp. 89–98, 2009.

[22] V. K. Rai and A. R. Mohanty, "Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2607–2615, Aug. 2007.

[23] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, Jan. 2009.
[24] J. S. Smith, "The local mean decomposition and its application to EEG perception data," *J. Roy. Soc. Interface*, vol. 2, no. 5, pp. 443–454, Dec. 2005.
[25] B. Chen *et al.*, "A demodulating approach based on local mean decomposition and its applications in mechanical fault diagnosis," *Meas. Sci. Technol.*, vol. 22, no. 5, Mar. 2011, Art. no. 055704.
[26] C. Park, D. Looney, M. M. Van Hulle, and D. P. Mandic, "The complex local mean decomposition," *Neurocomputing*, vol. 74, no. 6, pp. 867–875, Feb. 2011.
[27] W. Y. Liu, W. H. Zhang, J. G. Han, and G. F. Wang, "A new wind turbine fault diagnosis method based on the local mean decomposition," *Renew. Energy*, vol. 48, no. 12, pp. 411–415, Dec. 2012.
[28] J. Zheng, J. Cheng, and Y. Yang, "A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy," *Mech. Mach. Theory*, vol. 70, pp. 441–453, Dec. 2013.
[29] H. Liu, X. Wang, and C. Lu, "Rolling bearing fault diagnosis based on LCD–TEO and multifractal detrended fluctuation analysis," *Mech. Syst. Signal Process.*, vol. 60, no. 8, pp. 273–288, Aug. 2015.
[30] K. Qi, Z. He, and Y. Zi, "Cosine window-based boundary processing method for EMD and its application in rubbing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 21, no. 7, pp. 2750–2760, Oct. 2007.
[31] N. E. Huang and Z. Wu, "A review on Hilbert–Huang transform: Method and its applications to geophysical studies," *Rev. Geophys.*, vol. 46, no. 2, pp. 1–23, Jun. 2008.
[32] G. Rilling, P. Flandrin, and P. Gonçalves, "On empirical mode decomposition and its algorithms," in *Proc. IEEE-EURASIP Workshop Nonlinear Signal Image Process.*, Grado, Italy, Jun. 2003.
[33] S. Guo, G. Gu, and C. Li, "An algorithm for improving Hilbert–Huang transform," in *Proc. ICCS*, vol. 3. 2007, pp. 137–140.
[34] G.-B. Huang, D. H. Wang, and Y. Lan, "Extreme learning machines: A survey," *Int. J. Mach. Learn. Cybern.*, vol. 2, no. 2, pp. 107–122, Jun. 2011.
[35] B. Mirza, Z. Lin, and N. Liu, "Ensemble of subset online sequential extreme learning machine for class imbalance and concept drift," *Neurocomputing*, vol. 149, pp. 316–329, Feb. 2015.
[36] *Seeded Fault Test Data From Bearing Data Center of Case Western Reserve University*, accessed on Feb. 2016. [Online]. Available: http://csegroups.case.edu/bearingdatacenter
[37] X. Zhang and J. Zhou, "Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines," *Mech. Syst. Signal Process.*, vol. 41, no. 1, pp. 127–140, Aug. 2013.

QINGBIN TONG received the M.S. degree in electrical engineering from the Kunming University of Science and Technology, Kunming, China, in 2005, and the Ph.D. degree in instrument science and technology from the Harbin Institute of Technology, China, in 2008. He was a Post-Doctoral Research Associate with Beijing Jiaotong University and a Visiting Scholar with the University of Missouri, Columbia, MO, USA. He is currently an Associate Professor with the School of Electrical Engineering, Beijing Jiaotong University. His research interests include mechanical vibration, fault diagnosis, damage assessment and life prediction technology, engineering applications of artificial intelligence, rail transit traction drive, and control technology.

JUNCI CAO received the B.S. and M.S. degrees from the Harbin University of Science and Technology, Harbin, China, in 2001 and 2004, respectively, and the Ph.D. degree from the Harbin Institute of Technology, Harbin, in 2008, all in electrical engineering. He is currently an Associate Professor with the School of Electrical Engineering, Beijing Jiaotong University. His research interests include asynchronous traction drive system design, synchronous traction drive system design, simulation and reliability analysis, special motor design and integrated physical field research, motor optimization design, and system energy-saving projects.

BAOZHU HAN received the M.S. degree in electrical engineering from Beijing Jiaotong University, China, in 2016. She is currently an Assistant with the Department of Mechanical and Electrical Engineering, Taiyuan University. Her research interests include mechanical vibration, fault diagnosis, damage assessment and life prediction technology, engineering applications of artificial intelligence, and information processing technology.

XIAODONG ZHANG received the M.S. degree in electrical engineering from Tianjin University, Tianjin, China, in 1991, and the Ph.D. degree in electrical engineering from Beijing Jiaotong University, China, in 2012. He is currently a Professor with the School of Electrical Engineering, Beijing Jiaotong University. His research interests include mechanical vibration, fault diagnosis, damage assessment and life prediction technology, engineering applications of artiiňĄcial intelligence, and electromagnetic compatibility.

ZHENGWEI NIE received the M.S. degree in power engineering and engineering thermophysics from Xi'an Jiaotong University, Xi'an, China, in 2013. He is currently pursuing the Ph.D. degree in mechanical engineering with the University of Missouri, Columbia, MO, USA. His current research interests include CFD-based optimization and simulation, lightweight materials, fabrication of porous materials, and metamaterials.

JIAMIN WANG received the bachelor's degree from the East China University of Science and Technology, Shanghai, in 2015. He is currently pursuing the M.S. degree with the Mechanical and Aerospace Engineering Department, University of Missouri, Columbia, MO, USA. His research focuses on control and robotics. His major researches include dynamic analysis, control and application of automotive vehicles, such as quad-copters and spherical wheeled robots.

YUYI LIN received the M.S. degree from the University of California at Los Angeles, Los Angeles, CA, USA, and the Ph.D. degree from the University of California at Berkeley, Berkeley, CA, USA, in 1984 and 1989, respectively. Since 1990, he has been with the Department of Mechanical and Aerospace Engineering, College of Engineering, University of Missouri, USA, where he is currently an Associate Professor. He is a fellow of ASME. His research interests include dynamics and control of mechanisms, optimization, numerical methods, and artificial intelligence techniques.

WEIDONG ZHANG received the Ph.D. degree in electrical theory and new technology from North China Electric Power University, China, in 2003. He is currently a Professor with the School of Electrical and Electronic Engineering and the School of Electrical Engineering, North China Electric Power University. His research interests mainly include optical fiber sensing, signal analysis and processing, and power system electromagnetic compatibility.