

Article

# Fault Severity Monitoring of Rolling Bearings Based on Texture Feature Extraction of Sparse Time–Frequency Images

Yan Du <sup>1</sup> , Yingpin Chen <sup>2,3</sup>, Guoying Meng <sup>1,\*</sup>, Jun Ding <sup>1,4</sup> and Yajing Xiao <sup>1</sup>

<sup>1</sup> School of Mechanical Electronic and Information Engineering, China University of Mining and Technology, Beijing 100083, China; duyan042188@163.com (Y.D.); dingjun0823@163.com (J.D.); xiaoyajing001@163.com (Y.X.)

<sup>2</sup> School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China; 110500617@163.com

<sup>3</sup> School of Physics and Information Engineering, Minnan Normal University, Zhangzhou 363000, China

<sup>4</sup> School of Mechanical-Electrical Engineering, North China Institute of Science and Technology, Yanjiao 101601, China

\* Correspondence: mgy@cumtb.edu.cn; Tel.: +86-139-1023-5962

Received: 21 July 2018; Accepted: 21 August 2018; Published: 3 September 2018



**Abstract:** Rolling bearings are important components of rotating machines. For their preventive maintenance, it is not enough to know whether there is any fault or the fault type. For an effective maintenance, a fault severity monitoring needs to be conducted. Currently, the bearing fault diagnosis method based on time–frequency image (TFI) recognition is attracting increasing attention. This paper contributes to the ongoing investigation by proposing a new approach for the fault severity monitoring of rolling bearings based on the texture feature extraction of sparse TFIs. The first and main step is to obtain accurate TFIs from the vibration signals of rolling bearings. Traditional time–frequency analysis methods have disadvantages such as low resolution and cross-term interference. Therefore, the TFIs obtained cannot satisfactorily express the time–frequency characteristics of bearing vibration signals. To solve this problem, a sparse time–frequency analysis method based on the first-order primal-dual algorithm (STFA-PD) was developed in this paper. Unlike traditional time–frequency analysis methods, the time–frequency analysis model of the STFA-PD method is based on the theory of sparse representation, and is solved using the first-order primal-dual algorithm. For employing the sparse constraint in the frequency domain, the STFA-PD obtains a higher time–frequency resolution and is free from cross-term interference, as the model is based on a linear time–frequency analysis method. The gray level co-occurrence matrix is then employed to extract texture features from the sparse TFIs as input features for classifiers. Vibration signals of rolling bearings with different fault severity degrees are used to validate the proposed approach. The experimental results show that the developed STFA-PD outperforms traditional time–frequency analysis methods in terms of the accuracy and effectiveness for the fault severity monitoring of rolling bearings.

**Keywords:** rolling bearing; vibration signal; fault severity monitoring; sparse time–frequency analysis; gray level co-occurrence matrix; texture feature

## 1. Introduction

Rolling bearings are important and vulnerable components widely used in rotating machinery. The running state of rolling bearings significantly affects the performance, safety, and reliability of the overall machinery [1–3]. Therefore, the state detection and fault diagnosis for rolling bearings have

important practical significance. As bearing failure is a dynamic process, it is insufficient to know whether there is any fault or the fault type for the preventive maintenance of bearings in engineering applications. Determining the failure evolution process and accurately monitoring the fault severity are paramount for an effective bearing maintenance [4–6]. Consequently, the fault severity monitoring of rolling bearings is becoming increasingly important [7,8].

It would be instructive and practical to describe overall modes of bearing failure first. Bearing failure is not always due to vibration. There are many reasons for bearing failure. In fact, in most cases, the initiating source of failure is lubrication. The lack of a lubricant film between the rolling element and races will result in increased friction and wear. These conditions lead to generated heat, and thus, surface scoring and scuffing and even bearing dimensional instability. Of course, these can lead to vibration; however, by the time this occurs, the bearing is really in a failed state and needs to be replaced. There are three sources of failure for which the timely monitoring of vibrations is effective for prevention. The first one is poor assembly or loss of preload or interference fit, which leads to out-of-balance rotation and gradually excessive cage responses [9]. The second one is the variable compliance vibration which is inherent to ball and rolling element bearings [10,11]. Any amplitude increase at this frequency can lead to premature resonance and occurs with loss of preload or radial interference fit. The last one is increased contact loads, which cause fatigue spalling as the result of sub-surface stresses [12], leading to pits, cracks, etc. on the rolling surfaces. In this study, we focus on monitoring the fault severity of rolling bearings caused by the latter source of failure using vibration signals.

The first step in fault severity monitoring is obtaining the appropriate signals. As discussed before, the vibration signal is suitable for monitoring the fault severity. Moreover, compared to signals of temperature, sound, stress, and strain, a vibration signal carries considerable information regarding bearing health conditions and is easy to collect. Thereafter, the key step is to process the vibration signals. Time-domain and frequency-domain signal analysis methods, such as time-domain feature extraction, fast Fourier transform (FFT), and envelope analysis, were used to process vibration signals in the past [13–16]. However, the fault vibration signals of rolling bearings often exhibit nonlinear and non-stationary characteristics, making it difficult to accurately recognize the fault severity degree of bearings only in the time domain or in the frequency domain. Some non-stationary signal processing methods were developed. For example, the cyclostationarity method is a commonly used method to process non-stationary signals and was applied to some extent in the field of fault diagnosis of rotating machinery in the past few years [17–19]. Compared to the methods mentioned above, the time–frequency analysis methods are more effective when dealing with nonlinear and non-stationary signals, because they can accurately describe the local time–frequency characteristics of non-stationary signals by revealing the constituent frequency components and their time-variation characteristics. In fact, these methods are widely used in the field of mechanical fault diagnosis [20–22]. To date, the traditional time–frequency analysis methods, which are also commonly used methods, can be roughly divided into two classes [20]: linear time–frequency analysis methods, such as short-time Fourier transform (STFT), Gabor transform [23,24], wavelet transform (WT), and the S transform (ST); and bilinear/quadratic time–frequency analysis represented by the Wigner–Ville distribution (WVD) [22,25–29]. In fact, WT is a time-scale method, but there is a relationship between scale and frequency. Thus, WT can be used to conduct a time–frequency analysis for a signal. In addition, WT is also used to decompose a vibration signal into its wavelet components and choose the components that are related to the specific desired frequencies [30]. However, a substantial amount of useful information may be lost.

Governed by the Heisenberg uncertainty principle, the time and frequency resolutions of linear time–frequency analysis methods cannot simultaneously achieve the desired optimum effects. As a result, the time–frequency resolution is often low. On the contrary, bilinear time–frequency analysis methods have a higher time–frequency resolution. However, their time–frequency distributions are

inevitably subjected to cross-term interference, which is not the case with linear time–frequency representations [20].

In recent years, the sparse representation (SR) theory attracted widespread attention and was applied to many fields, particularly to signal and image processing [31–36]. With the application of SR theory in the field of time–frequency analysis, a new time–frequency representation (TFR) model based on SR and STFT emerged [37,38]. The time–frequency analysis problem is essentially converted into a sparse optimization problem based on the L0-norm constraint. The time–frequency distribution with higher concentration and resolution can be obtained by solving the sparse optimization problem. Moreover, as the model is based on the STFT framework, there is no cross-item interference. As a result, this model can better describe the time–frequency characteristics of the signals and was successfully applied to the analysis of non-stationary signals, such as seismic and radar signals [39,40]. In this paper, this model was applied to conduct the time–frequency analysis of bearing vibration signals. However, the optimization problem with respect to this model is an L0-norm minimization problem, i.e., a non-convex optimization problem, which is NP-hard and very difficult to solve. To address this issue, a first-order primal-dual algorithm-based sparse time–frequency analysis method (STFA-PD) was developed. Firstly, the non-convex optimization problem of L0-norm was relaxed to an L1-norm minimization problem using a convex relaxation technique. Secondly, the first-order primal-dual algorithm was employed to solve the problem [41,42].

Several time–frequency images (TFIs) reflecting the bearing condition can be obtained by conducting the time–frequency analysis of vibration signals. Currently, there are two types of bearing condition recognition methods based on TFIs [43,44]. The first one is an artificial recognition method performed by skilled experts. This method relies heavily on individual experience and skill level and has low accuracy and efficiency for bearing condition recognition. The other is an automatic classification using computers, which is attracting increasing attention. Although TFIs can effectively characterize the local time–frequency properties of rolling bearing vibration signals, they cannot be directly used as features for an automatic classification, because the data dimensions of TFIs are too numerous for such a classification. Therefore, image dimensionality reduction and feature extraction need to be conducted before classification. Some algebraic feature extraction methods, such as two-direction two-dimensional principal component analysis (TD-2DPCA) [45], two-direction two-dimensional linear discriminant analysis (TD-2DLDA) [46], and two-dimensional non-negative matrix factorization (2DNMF) [43], were proposed for the dimensionality reduction of TFIs. The low-dimensional features obtained through dimensionality reduction are directly used for automatic classification. However, as the horizontal distribution of the frequency properties in TFIs changes with time, the features will be inevitably affected by the influence of the time-domain truncation. As a result, even the TFIs with the same fault severity may have completely different features.

Bearing vibration signals of different states can be used to express different characteristics in the time–frequency domain, resulting in TFIs with different textures. Additionally, the texture features can be expressed as global statistics of local features, which has little to do with the changes in the horizontal distribution with time. Therefore, texture-feature-based image classification can be used to distinguish the fault severity degrees of bearings. The most widely used texture descriptor for image recognition is the gray level co-occurrence matrix (GLCM), proposed by Haralick et al. [47] in 1973, as it can accurately describe the roughness, repetition direction, and complexity of the texture of an image. GLCM is a square matrix. GLCMs with different parameters can be easily obtained from a single image, while many statistical features can be easily obtained from each GLCM. Compared to GLCM, the statistical features are more convenient for classification. Consequently, GLCM-based texture features, i.e., statistical features, can be extracted from TFIs to obtain fault severity information.

A new approach for the fault severity monitoring of rolling bearings based on the texture feature extraction of sparse TFIs is proposed in this paper. The novelty is that a new time–frequency analysis method, named STFA-PD, was developed for analyzing the bearing vibration signal. A high-quality

TFI without the cross-term interference, which we call sparse TFI, can be obtained from the vibration signal using the STFA-PD method. Thereafter, the GLCM can be used to extract the texture features from sparse TFIs as the input features for classifiers. Finally, the fault severity recognition can be realized. Figure 1 shows a flowchart of the proposed bearing fault severity monitoring approach.

The rest of this paper is organized as follows: in Section 2, the STFA-PD is described in detail, including the establishment of the model and the algorithm solution to the sparse optimization problem. In addition, the superiority of the STFA-PD is verified using a simulated signal. The calculation of GLCM and its statistical features are briefly reviewed in Section 3. In Section 4, the vibration signals of the rolling bearings with different fault severity degrees are used to validate the proposed approach. Finally, the conclusions of this investigation are given in Section 5.

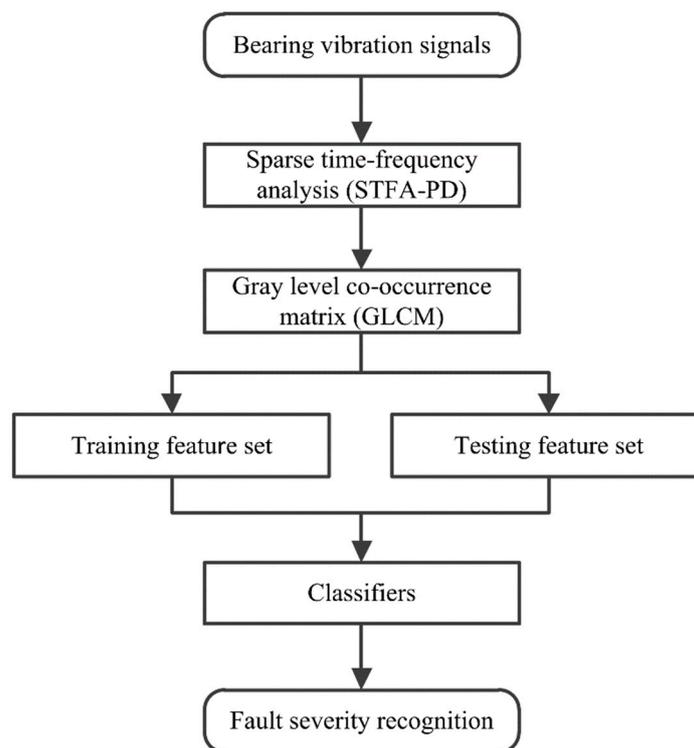


Figure 1. Flowchart of the proposed approach.

## 2. First-Order Primal-Dual Algorithm-Based Sparse Time–Frequency Analysis

### 2.1. Model of Sparse Time–Frequency Analysis

The short-time Fourier transform (STFT) is one of the earliest and most widely used time–frequency analysis methods. The basic idea of the STFT can be summarized as described below. A short time window is added to the desired signal, and the local Fourier spectrum of the windowed signal is calculated. By sliding the window with respect to time, the time–frequency analysis of the signal can be carried out [20,48]. The STFT of the continuous-time signal  $x(t)$  can be defined as follows:

$$X(t, \omega) = \int_{-\infty}^{+\infty} x(\tau)w(\tau - t) \exp(-j\omega\tau)d\tau, \quad (1)$$

where  $w(t)$  is the sliding window function.

Correspondingly, for a discrete signal  $x(n)$  of period  $N$ , where  $n = 0, 1, \dots, N - 1$ , the discrete STFT can be defined as follows:

$$X(m, l) = \sum_{n=0}^{N-1} x(n)w(n - m)W_N^{nl}, \tag{2}$$

where  $l = 0, 1, \dots, N - 1$ , and  $W_N = \exp(-j2\pi/N)$ .

According to Equation (1), the STFT  $X(t, \omega)$  of the continuous-time signal  $x(t)$  can be regarded as a Fourier transform (FT) of the windowed signal  $x(\tau)w(\tau - t)$ . On the contrary, when  $X(t, \omega)$  and window functions  $w(t)$  are given, the original signal can be directly recovered by inverse Fourier transform (IFT), which can be represented as follows:

$$x(\tau)w(\tau - t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(t, \omega) \exp(j\omega\tau) d\omega. \tag{3}$$

Correspondingly, for the windowed signal  $x_m(n) = x(n)w(n - m)$  of  $x(n)$  at time  $m$ , the discrete STFT can be defined as follows:

$$x_m(n) = \frac{1}{N} \sum_{l=0}^{N-1} X(m, l)W_N^{-nl}. \tag{4}$$

The vector form of  $x_m(n)$  can be expressed as follows:

$$\mathbf{x}_m = \mathbf{W}_N \mathbf{X}_m, \tag{5}$$

where  $\mathbf{x}_m = [x_m(0), x_m(1), \dots, x_m(N - 1)]^T$  and  $\mathbf{X}_m = [X_m(m, 0), X_m(m, 1), \dots, X_m(m, N - 1)]^T$  are the windowed signal and Fourier spectrum at time  $m$ , respectively;  $\mathbf{W}_N$  is the inverse Fourier transform matrix; and  $[\mathbf{W}_N]_{k,l} = 1/N \cdot W_N^{-(k-1)(l-1)}$ , where  $k, l = 1, 2, \dots, N$ .

The short-time truncated sub-signal can be given as follows:

$$\mathbf{y}_m = \mathbf{Z}_m \mathbf{x}_m = \mathbf{Z}_m \mathbf{W}_N \mathbf{X}_m, \tag{6}$$

where  $\mathbf{y}_m \in \mathbf{R}^{U \times 1}$  is the short-time truncated sub-signal,  $\mathbf{Z}_m \in \mathbf{R}^{U \times N}$  is the sliding window at time  $m$ , and the sliding window  $\mathbf{Z}_{m+1}$  at time  $m + 1$  is the circular shift of  $\mathbf{Z}_m$  with the column values shifted to right by one.

In Equation (6),  $\mathbf{Z}_m \mathbf{W}_N$  is a partial Fourier matrix, and its columns may be highly correlated, where this cannot guarantee the uniqueness of the solution of  $\mathbf{X}_m$ . The windowed signal  $\mathbf{x}_m$  at time  $m$  is sparse in the frequency domain. Inspired by SR theory, the special spectrum  $\mathbf{X}_m$  can be easily obtained by adding the sparse constraint to fit the sparse prior knowledge of the spectrum. The sparse constraint problem can be modeled as Equation (7).

$$\hat{\mathbf{X}}_m = \underset{\mathbf{X}_m}{\operatorname{argmin}} \|\mathbf{X}_m\|_0 \text{ s.t. } \mathbf{y}_m = \mathbf{Z}_m \mathbf{W}_N \mathbf{X}_m, \tag{7}$$

where  $\|\cdot\|_0$  denotes the L0-norm. By comparing Equations (6) and (7), we find that  $\mathbf{Z}_m \mathbf{W}_N$  plays the role of a dictionary of SR. Considering that the L0-norm minimization problem is NP-hard, the L1-norm is generally approximated to the L0-norm, i.e., the convex relaxation is considered an L1-norm minimization problem, or an interference in practical situations. Therefore, we can estimate the spectrum by solving the sparse constraint problem modeled as Equation (8).

$$\hat{\mathbf{X}}_m = \underset{\mathbf{X}_m}{\operatorname{argmin}} \|\mathbf{X}_m\|_1 \text{ s.t. } \|\mathbf{Z}_m \mathbf{W}_N \mathbf{X}_m - \mathbf{y}_m\|_2 \leq \varepsilon_m, \tag{8}$$

where  $\|\cdot\|_1$  denote the L1-norm,  $\hat{\mathbf{X}}_m$  is the estimated signal spectrum at time  $m$ , and  $\varepsilon_m$  is a user-specified parameter related to noise.

The sparse time–frequency spectrum  $\hat{\mathbf{X}} = [\hat{\mathbf{X}}_0, \hat{\mathbf{X}}_1, \dots, \hat{\mathbf{X}}_{N-1}]$  can be finally obtained by solving a group of optimization problems, Equation (8), for all  $m = 0, 1, \dots, N - 1$ . The group of optimization problems is called the sparse time–frequency model.

### 2.2. Solution of the Model Based on the First-Order Primal-Dual Algorithm

As the model is a set of convex optimization problems with constraints, it can be rewritten as an unconstrained optimization problem, i.e., Equation (9), which is convenient for solving, as follows:

$$\hat{\mathbf{X}}_m = \underset{\mathbf{X}_m}{\operatorname{argmin}} \mu \|\mathbf{X}_m\|_1 + \frac{1}{2} \|\mathbf{Z}_m \mathbf{W}_N \mathbf{X}_m - \mathbf{y}_m\|_2^2, \tag{9}$$

where  $\mu \|\mathbf{X}_m\|_1$  is the sparse term,  $\mu > 0$  is the sparse coefficient, and  $\frac{1}{2} \|\mathbf{Z}_m \mathbf{W}_N \mathbf{X}_m - \mathbf{y}_m\|_2^2$  is the fidelity term.

Assuming  $F(\mu \mathbf{X}_m) = \mu \|\mathbf{X}_m\|_1$  and  $G(\mathbf{X}_m) = \frac{1}{2} \|\mathbf{Z}_m \mathbf{W}_N \mathbf{X}_m - \mathbf{y}_m\|_2^2$ , according to the first-order primal-dual algorithm, the solution can be trivially given by the following sequence of convex problems:

$$\begin{cases} \mathbf{Y}_m^{(n+1)} = (\mathbf{I} + \sigma \partial F^*)^{-1} (\mathbf{Y}_m^{(n)} + \mu \sigma \bar{\mathbf{X}}_m^{(n)}) \\ \mathbf{X}_m^{(n+1)} = (\mathbf{I} + \tau \partial G)^{-1} (\mathbf{X}_m^{(n)} - \mu \tau \mathbf{Y}_m^{(n+1)}) \\ \bar{\mathbf{X}}_m^{(n+1)} = \mathbf{X}_m^{(n+1)} + \theta (\mathbf{X}_m^{(n+1)} - \mathbf{X}_m^{(n)}) \end{cases}, \tag{10}$$

where  $\mathbf{I} \in \mathbf{R}^{N \times N}$  is the unit matrix,  $(\mathbf{I} + \sigma \partial F^*)^{-1}$  and  $(\mathbf{I} + \tau \partial G)^{-1}$  are resolvent operators,  $F^*$  is the convex conjugate of  $F$ ,  $F^*(\mathbf{Y}_m) = \begin{cases} 0 & \text{if } \mathbf{Y}_m \in \mathbf{Y} \\ +\infty & \text{if } \mathbf{Y}_m \notin \mathbf{Y} \end{cases}$  is the convex conjugate function of  $F(\mu \mathbf{X}_m)$ ,  $\mathbf{Y}_m \in \mathbf{Y}$  is the dual variable, and  $\mathbf{Y}$  is the convex set defined as follows:

$$\mathbf{Y} = \{\mathbf{Y}_m \in \mathbf{Y} : \|\mathbf{Y}_m\|_\infty \leq 1\}. \tag{11}$$

The resolvent operator with respect to  $F^*$  can be obtained as follows:

$$\begin{aligned} \mathbf{Y}_m^{(n+1)} &= (\mathbf{I} + \sigma \partial F^*)^{-1} (\mathbf{Y}_m^{(n)} + \mu \sigma \bar{\mathbf{X}}_m^{(n)}) \\ &= \underset{\mathbf{Y}_m^{(n+1)}}{\operatorname{arg min}} \left\{ \frac{1}{2} \|\mathbf{Y}_m^{(n+1)} - (\mathbf{Y}_m^{(n)} + \mu \sigma \bar{\mathbf{X}}_m^{(n)})\|_2^2 + F^*(\mathbf{Y}_m^{(n+1)}) \right\}. \end{aligned} \tag{12}$$

As  $F^*(\mathbf{Y}_m)$  is the indicator function for a convex set, the resolvent operator with respect to  $F^*$  reduces to pointwise Euclidean projectors:

$$\mathbf{Y}_m^{(n+1)} = \frac{\mathbf{Y}_m^{(n)} + \mu \sigma \bar{\mathbf{X}}_m^{(n)}}{\max(1, |\mathbf{Y}_m^{(n)} + \mu \sigma \bar{\mathbf{X}}_m^{(n)}|)}. \tag{13}$$

The resolvent operator with respect to  $G$  can be expressed as follows:

$$\begin{aligned} \mathbf{X}_m^{(n+1)} &= (\mathbf{I} + \tau \partial G)^{-1} (\mathbf{X}_m^{(n)} - \mu \tau \mathbf{Y}_m^{(n+1)}) \\ &= \underset{\mathbf{X}_m^{(n+1)}}{\operatorname{arg min}} \left\{ \frac{1}{2} \|\mathbf{X}_m^{(n+1)} - (\mathbf{X}_m^{(n)} - \mu \tau \mathbf{Y}_m^{(n+1)})\|_2^2 + \frac{\tau}{2} \|\mathbf{Z}_m \mathbf{W}_N \mathbf{X}_m^{(n+1)} - \mathbf{y}_m\|_2^2 \right\}. \end{aligned} \tag{14}$$

The resolvent operator with respect to  $G$  is a simple convex quadratic programming problem, the solution to which is given as follows:

$$\mathbf{X}_m^{(n+1)} = \frac{(\mathbf{X}_m^{(n)} - \mu \tau \mathbf{Y}_m^{(n+1)}) + \tau (\mathbf{Z}_m \mathbf{W}_N)^H \mathbf{y}_m}{\mathbf{I} + \tau (\mathbf{Z}_m \mathbf{W}_N)^H (\mathbf{Z}_m \mathbf{W}_N)}. \tag{15}$$

The sparse time–frequency spectrum  $\hat{X} = [\hat{X}_0, \hat{X}_1, \dots, \hat{X}_{N-1}]$  of the signal can be finally obtained by solving a group of optimization problems, such as Equation (9), using the first-order primal-dual algorithm for  $m = 0, 1, \dots, N - 1$ .

The pseudocode of the STFA-PD is summarized in Algorithm 1.

---

**Algorithm 1.** Sparse time–frequency analysis method based on the first-order primal-dual algorithm (STFA-PD)

---

**Input:**  $y_m, Z_m W_N, \mu, \tau, \sigma > 0, \theta \in [0, 1]$   
**Output:**  $X_m$   
**Initialize:**  $X_m^{(1)} = 0, Y_m^{(1)} = 0, \bar{X}_m^{(1)} = X_m^{(1)}, Maxiter = 100$   
**1: For**  $m = 0 : N - 1$  **do**  
**2: For**  $n = 1 : Maxiter$  **do**  
**3:**  $Y_m^{(n+1)} = \frac{Y_m^{(n)} + \mu\sigma\bar{X}_m^{(n)}}{\max(1, |Y_m^{(n)} + \mu\sigma\bar{X}_m^{(n)}|)}$   
**4:**  $X_m^{(n+1)} = \frac{(X_m^{(n)} - \mu\tau Y_m^{(n+1)}) + \tau(Z_m W_N)^H y_m}{I + \tau(Z_m W_N)^H (Z_m W_N)}$   
**5:**  $\bar{X}_m^{(n+1)} = X_m^{(n+1)} + \theta(X_m^{(n+1)} - X_m^{(n)})$   
**6: End For**  
**7: End for**  
**8: Return**  $X_m^{(n+1)}$  **as**  $X_m$

---

### 2.3. Simulation Analysis

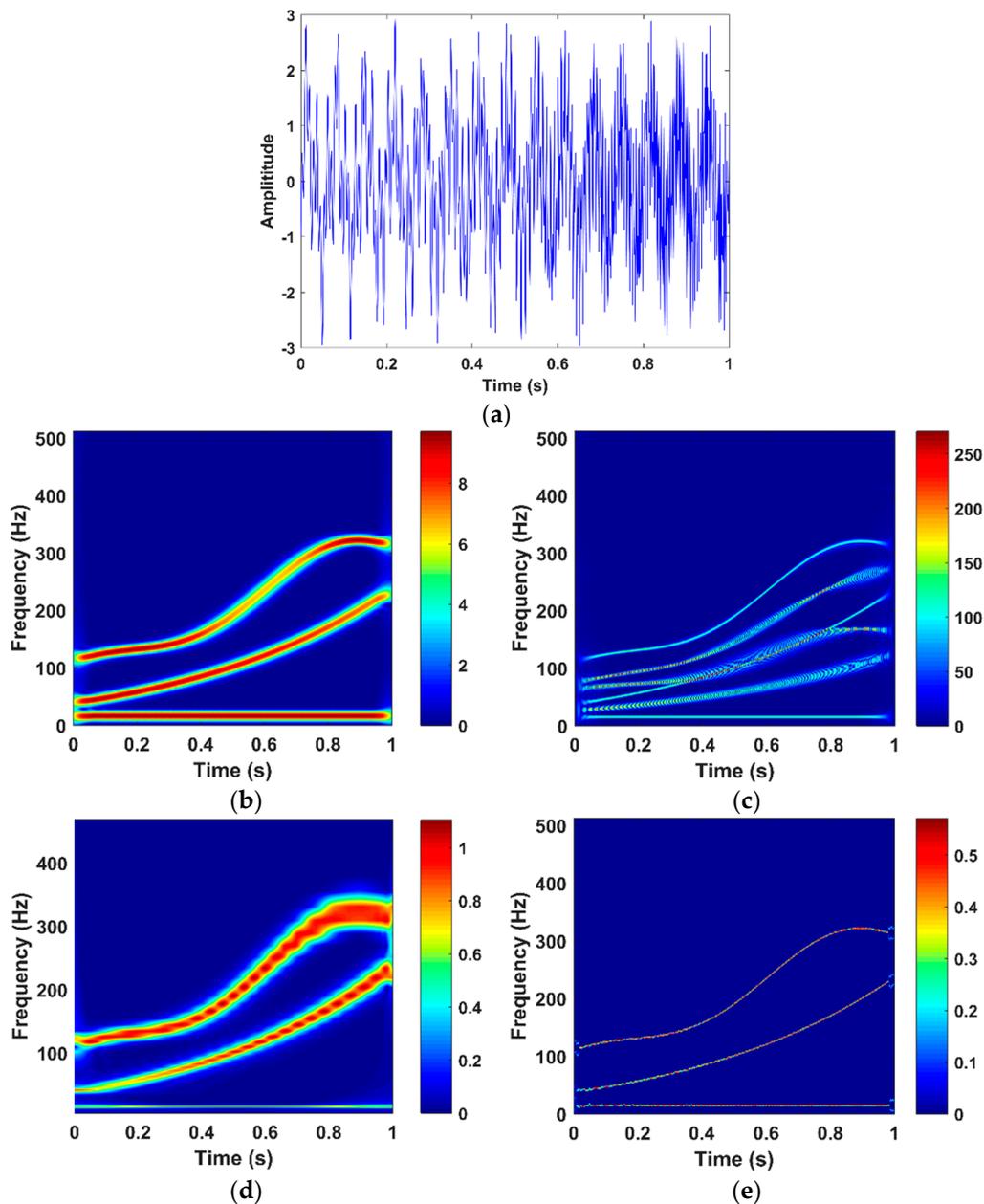
To validate the effectiveness and superiority of the STFA-PD, a multi-component simulated signal, defined by Equation (16), was employed:

$$\begin{cases} s_1(t) = \sin(30\pi t), \\ s_2(t) = \sin(30 + 75\pi t \cdot 3^t) \\ s_3(t) = \sin[200\pi t + 200\pi t^2 + 20\pi t \cos(2\pi t)] \\ s(t) = s_1(t) + s_2(t) + s_3(t) \end{cases} \quad (16)$$

where  $s_1(t)$  was a sinusoidal signal and both  $s_2(t)$  and  $s_3(t)$  were nonlinear frequency-modulated signals. The sampling frequency was set to 1000 Hz, and the sample length was 1 s.

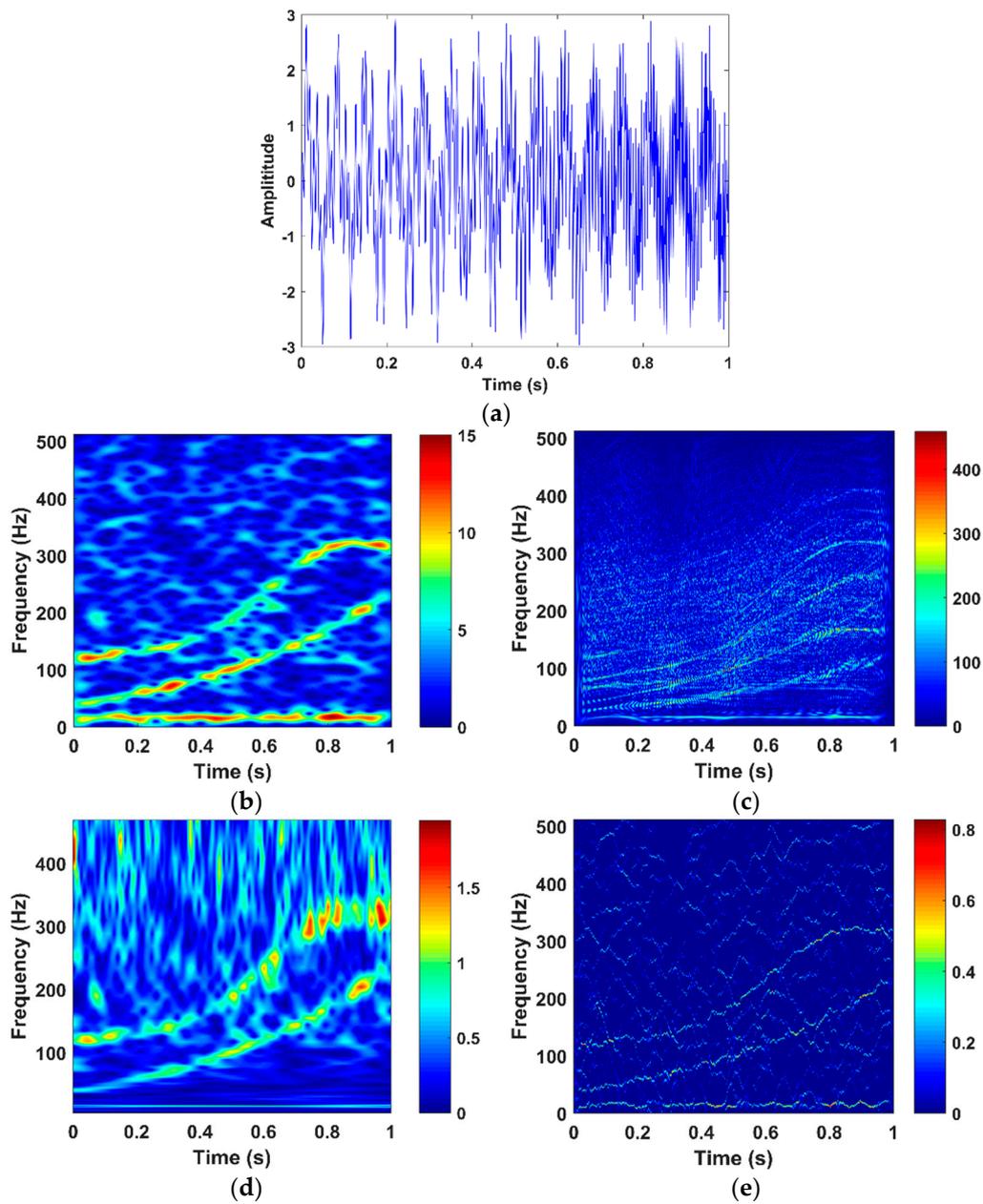
Figure 2a shows the time-domain waveform of the simulated signal  $s(t)$ , which was a non-stationary signal with a frequency that varied with time. Figure 2b,e show the TFIs obtained using the STFT, pseudo WVD (PWVD), WT, and the STFA-PD. In this study, the bump mother wavelet was used in WT of the simulated signal, as well as bearing vibration signals.

Figure 2b shows that the resolution of the traditional STFT was low either in the lower frequency band or in the higher frequency band, and the time–frequency distribution of the signal was not accurately reflected. By contrast, the TFI obtained using the PWVD was good in terms of the time–frequency resolution, but suffered from cross-term interferences, and thus, did not precisely reflect the time–frequency characteristics of the signal, as shown in Figure 2c. As shown in Figure 2d, the TFI obtained using the WT had a good time–frequency resolution in the lower frequency band, but a poor resolution in the higher frequency band. As the WT was essentially a type of time-scale transform and as there was no good correspondence between the scale and the frequency of the signal, the WT did not accurately express the time-based variations in the frequency of each component of the signal. Figure 2e shows the TFI obtained using the STFA-PD method. Comparing Figure 2e with Figure 2b–d, we see that the STFA-PD method had higher time–frequency concentration and resolution than the other methods.



**Figure 2.** (a) Time-domain waveform of the simulated signal  $s(t)$  and its time–frequency images (TFIs) obtained using (b) short-time Fourier transform (STFT), (c) pseudo Wigner–Ville distribution (PWVD), (d) wavelet transform (WT), and (e) the sparse time–frequency analysis method based on the first-order primal–dual algorithm (STFA-PD).

To validate the denoising performance of the STFA-PD method, the Gaussian white noise was artificially added to the simulated signal  $s(t)$ . The noisy signal was denoted as  $z(t)$ . Figure 3a shows the time-domain waveform of the noisy signal  $z(t)$  with the signal-to-noise ratio (SNR) of  $-2$  dB. Figure 3b–e show the TFIs obtained using STFT, PWVD, WT, and STFA-PD. It can be seen there were different degrees of noise interferences in the TFIs. As displayed in Figure 3c, The TFI obtained by PWVD seemed more sensitive to the noise than others. The time–frequency characteristic of the signal  $s(t)$  was almost masked by the noise and cross-term interferences. From Figure 3e, we can find that The TFI obtained by STFA-PD still had a high time–frequency resolution even if there was a high level of noise.



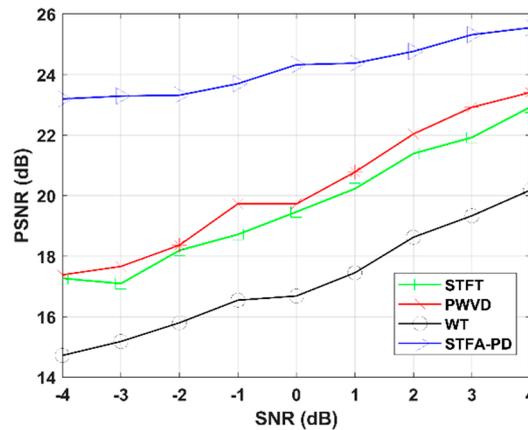
**Figure 3.** (a) Time-domain waveform of the noisy signal  $z(t)$  with the signal-to-noise ratio (SNR) of  $-2$  dB, and its TFIs obtained using (b) STFT, (c) PWVD, (d) WT, and (e) STFA-PD.

To further quantitatively analyze and evaluate the denoising performance of the STFA-PD method, the peak signal-to-noise ratio (PSNR) was adopted as a denoising performance evaluation measurement. The PSNR is defined as follows:

$$\text{PSNR}(J, K) = 10 \log_{10} \frac{\text{MAX}_J^2}{\frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (J_{ij} - K_{ij})^2}, \quad (17)$$

where  $J$  is the TFI of the simulated signal  $s(t)$ , and  $K$  is the TFI of the noisy signal  $z(t)$ ; the bigger the PSNR is, the closer  $J$  and  $K$  are;  $\text{MAX}_J$  is the maximal value of  $J$ . Figure 4 shows the PSNR values obtained by different methods (i.e., STFT, PWVD, WT, and the proposed STFA-PD method) at different

SNR levels. It can be seen the STFA-PD method obtained bigger a PSNR value than the other three methods, which meant the STFA-PD method was better in terms of denoising performance than others. Moreover, it can also be seen that the PSNR value became smaller along with the SNR getting lower. However, the PSNR value obtained by STFA-PD fell more slowly than others. Even at a low SNR of  $-4$  dB, the PSNR value obtained by STFA-PD was still high.



**Figure 4.** Peak SNR (PSNR) values obtained by different methods (i.e., STFT, PWVD, WT, and the proposed STFA-PD method) at different SNR levels.

### 3. GLCM-Based Image Texture Features

#### 3.1. Brief Introduction of GLCM

The texture is formed by the repeated appearance of the gray distribution in spatial positions. Thus, there will be a certain gray relationship between two pixels separated by a certain distance in the image space, which is the spatial correlation characteristic of gray in the image. In image recognition, the GLCM is a commonly used method for image texture analysis [47,49]. By studying the spatial correlation characteristics of gray levels in the image, the GLCM can be used to accurately describe image texture features. The GLCM can be obtained by counting the co-occurrence frequency of gray levels of a pair of pixels with a certain positional relationship in the two-dimensional (2D) space of a gray image.

Figure 5 shows the calculation schematic of the GLCM. Given a pixel pair with gray level  $i$  and  $j$ , the co-occurrence frequency  $P(i, j, d, \theta)$  of the pixel pairs can be defined as follows:

$$P(i, j, d, \theta) = \Gamma \{ [(x, y), (x + \Delta x, y + \Delta y)] | f(x, y) = i, f(x + \Delta x, y + \Delta y) = j \}, \quad (18)$$

where  $\Gamma$  denotes the total number of pixel pairs satisfying the conditions,  $i, j = 0, 1, 2, \dots, N - 1$  is the gray level,  $(x, y)$  are the coordinates of the pixel with gray level  $i$ ,  $(x + \Delta x, y + \Delta y)$  are the coordinates of the pixel with gray level  $j$ ,  $\Delta x$  and  $\Delta y$  are the position offsets in the horizontal and vertical directions, respectively,  $d = \sqrt{\Delta x^2 + \Delta y^2}$  is the generation step, and  $\theta$  is the direction of generation.

The matrix composed of frequency  $P(i, j, d, \theta)$  with specified values of  $d$  and  $\theta$  is the GLCM, an  $N \times N$  square matrix. Four GLCMs with  $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$  are generally used for texture analysis.

The generation step  $d$  is determined by the horizontal and vertical offsets  $(\Delta x, \Delta y)$ . Therefore, the GLCMs of an image can change with respect to  $(\Delta x, \Delta y)$ . The value of  $(\Delta x, \Delta y)$  should be selected according to the characteristics of the periodic distribution of texture. For thinner and more slowly changing texture images, lower values should be taken, and higher values should be chosen for images of textures that change quickly.

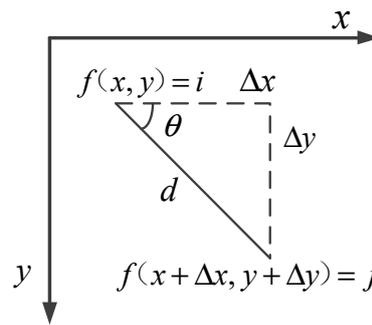


Figure 5. Calculation schematic of gray level co-occurrence matrix (GLCM).

### 3.2. Statistical Features of GLCM

For the texture analysis of an image, the GLCM is often not directly used as the texture features of the image. In fact, the statistical features extracted from the GLCM are used as the texture features.

Before extracting the statistical features from the GLCM, the frequency  $P(i, j, d, \theta)$  should be normalized to co-occurrence probability  $g(i, j)$  as follows:

$$g(i, j) = P(i, j) / \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j). \tag{19}$$

Haraclick et al. [47] suggested many statistical features, such as maximum probability, entropy, contrast, correlation, energy, and inverse differential moment. Three of them were calculated here: entropy, contrast, and energy. This is because the results of an experiment conducted by Baraldi and Parmiggiani [50] showed that these statistical features are suitable for effective classification and image recognition.

Entropy is used to characterize the information capacity of an image and to characterize the complexity or non-uniformity of image texture. The entropy is highest when the image is full of texture and lowest when there is no texture:

$$\text{Entropy} = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} g(i, j) \log(g(i, j)). \tag{20}$$

Contrast is a measure of the degree of texture clarity, and it is expressed as in Equation (20). The deeper the texture groove, the greater the contrast in the image, and vice versa. For example, the contrast of a constant image is zero, as there is only a single non-zero element in the GLCM.

$$\text{Contrast} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i - j)^2 g(i, j). \tag{21}$$

Energy, also known as the angular second moment (ASM), can be defined as follows:

$$\text{Energy} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} g^2(i, j). \tag{22}$$

It characterizes the uniformity of the gray distribution of the image texture, and can reflect its roughness. When image texture is coarser, i.e., when the frequency distribution in a TFI is more concentrated, the energy is greater and vice versa.

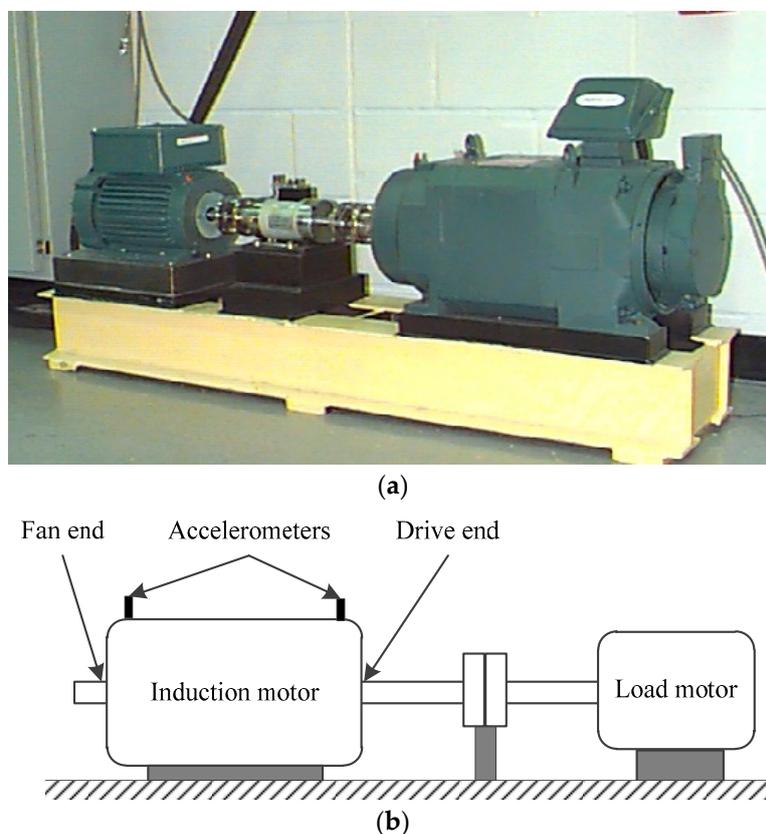
In this study, four GLCMs with  $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$  were employed. We extracted three types of statistical features from the four GLCMs as texture features for characterizing the TFIs of the vibration signals. The 12 features were expressed as T1, T2, ..., T12, where T1, T2, T3, and T4 represent

entropy features; T5, T6, T7, and T8 represent contrast features; and T9, T10, T11, and T12 represent energy features of the four GLCMs.

#### 4. Experimental Validation and Discussion

##### 4.1. Experimental Data and Setup

The bearing data used in this study were obtained from the Bearing Data Center of the Electrical Engineering Laboratory at the Case Western Reserve University (CWRU) [51–53]. Figure 6 shows the bearing test stand and its structural diagram. The test stand included a 2 horsepower motor, a torque transducer/encoder, a dynamometer, and control electronics. The motor bearings were seeded with single-point faults using electro-discharge machining (EDM). Faults with different severity degrees ranging from 0.007 to 0.040 inches in diameter were introduced separately at the inner race, outer race, and ball. The vibration data were collected using accelerometers, which were installed on the housing using magnetic bases and placed at a 12 o'clock position on both the drive end and fan end of the motor housing. The sampling frequency was set to 12 kHz.



**Figure 6.** (a) Bearing test stand, and (b) its structural diagram.

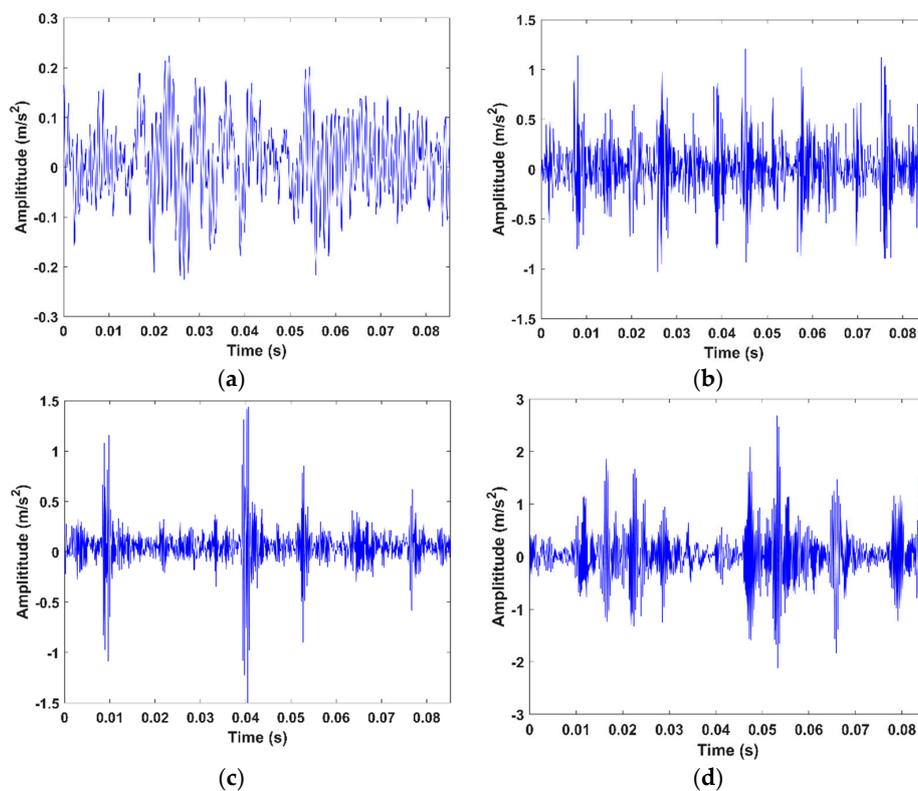
The 6205-2RS JEM SKF deep groove ball bearings were chosen as study object and were seeded with faults of four different severity degrees (0.000, 0.007, 0.014, and 0.021 in) at the inner race, outer race, and ball. The vibration signals were collected under the condition of 0 horsepower (HP) load. Table 1 lists the test arrangement of the bearing fault severity. Twenty samples were chosen for each severity degree, with a total of 80 samples for one test. Each sample contained 1024 data points. In the following section, the results of Test 1 for each step are described in detail. At the end, the results of the other two tests are given, and the results of all the tests are compared.

**Table 1.** Test arrangement of the bearing fault severity.

Test	Fault Type	Severity Degree (inches)
1	Inner race fault	0.000
		0.007
		0.014
		0.021
2	Outer race fault	0.000
		0.007
		0.014
		0.021
3	Ball fault	0.000
		0.007
		0.014
		0.021

4.2. STEFA-PD of Vibration Signals of Rolling Bearings

Figure 6 shows the time-domain waveforms of the vibration signals of bearings with different fault severity degrees at the inner race. Comparing Figure 7a with Figure 7b–d, it is obvious that we can easily determine whether a fault occurs or not, as the signal of the inner race with a fault of 0.000 in was a normal signal distinct from fault signals. From Figure 7b–d, we find that all fault signals have prominent impact components with different amplitudes. However, the differences between the signals of different fault severity degrees were insufficient to accurately distinguish the fault severity degrees of the inner race.

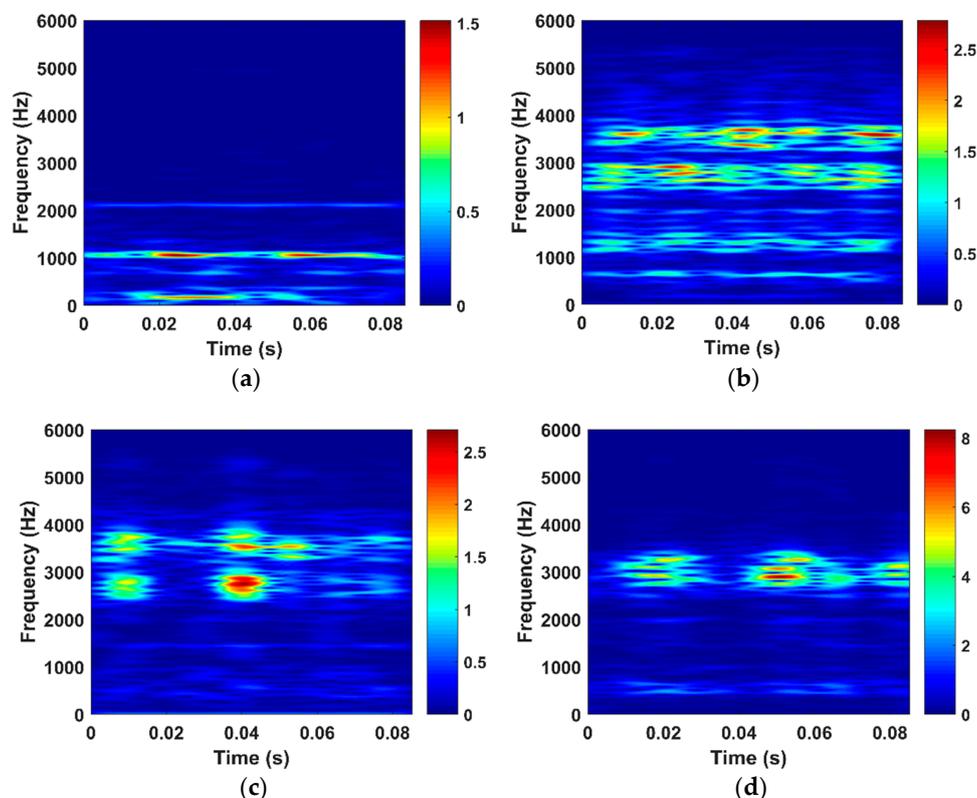


**Figure 7.** Time-domain waveforms of the vibration signals of bearings with four different faults severity degrees at the inner race: (a) 0.000 in, (b) 0.007 in, (c) 0.014 in, and (d) 0.021 in.

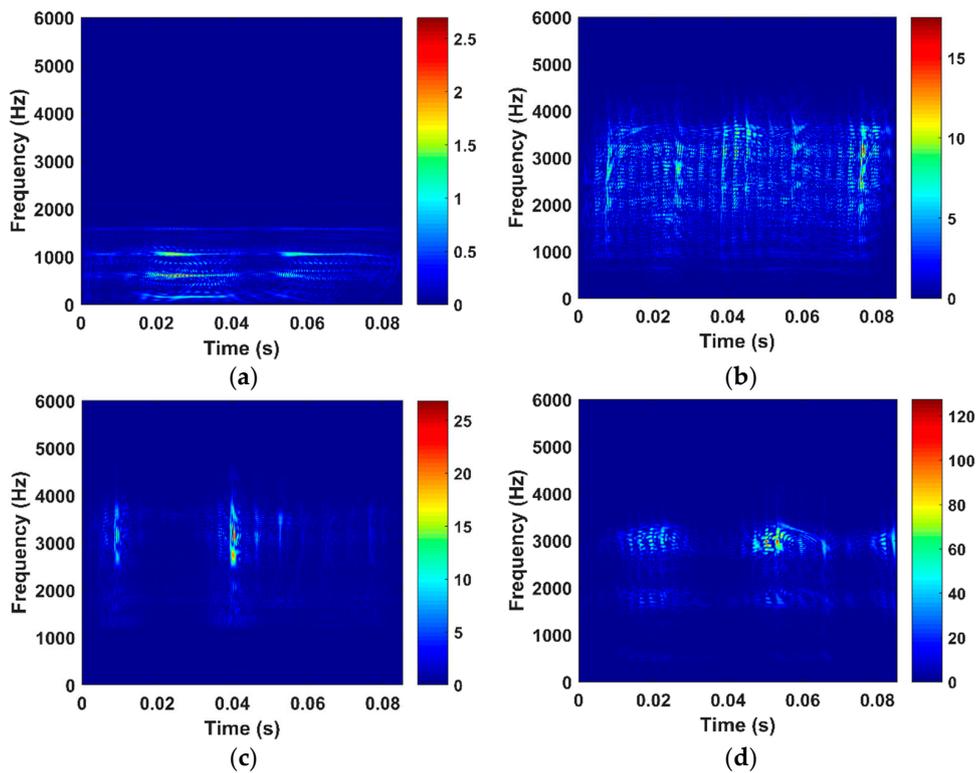
In this study, the STFA-PD method was employed to process the vibration signals. For comparison, the STFT, PWVD, and WT were used to conduct the time–frequency analysis of the signals with different fault severity degrees at the inner race. Figures 8–11 show the TFIs obtained using the STFT, PWVD, WT, and STFA-PD, respectively.

The energy distribution in the frequency domain of the normal signal was found to be narrow and relatively centralized in the low-frequency band. When an inner race fault occurs, the energy distributions became very wide, and the energy was mostly concentrated in the high-frequency band ranging from 2500 Hz to 3500 Hz. All TFIs obtained from the fault signals showed certain proportions of high-frequency impact components corresponding to the time-domain waveforms of the vibration signals. Moreover, the proportion of high-frequency impact components increased with an increase in the fault severity degree.

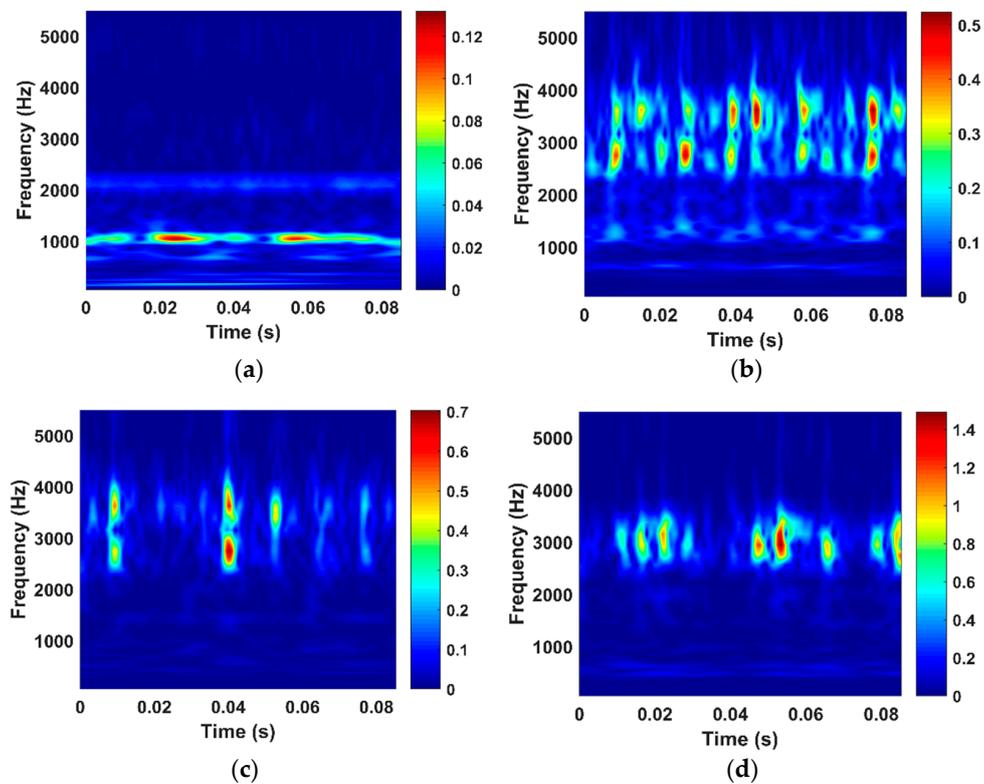
Figure 8 shows that the time–frequency concentration and resolution of the traditional STFT were low and the TFIs could not accurately express the time–frequency characteristics of the vibration signals. In addition, there were obvious noise interferences in the TFIs. On the contrary, the TFIs obtained using the PWVD were good in terms of the time–frequency concentration and resolution, and the TFIs of different fault severity degrees could be accurately distinguished. However, they suffered from cross-term interferences, and thus, could not precisely reflect the time–frequency characteristics of the vibration signals, as shown in Figure 9. As shown in Figure 10, similar to the STFT method, the TFIs obtained using the WT were not subjected to cross-term interferences. Moreover, we found that the WT had a relatively good distinctiveness. However, the frequency resolution obtained using the WT was lower than that obtained using the STFT, although its time resolution was better. Additionally, as the WT is essentially a type of time-scale transform and there is no good correspondence between scale and frequency, the WT could not accurately express the frequency changes in the signals with respect to time.



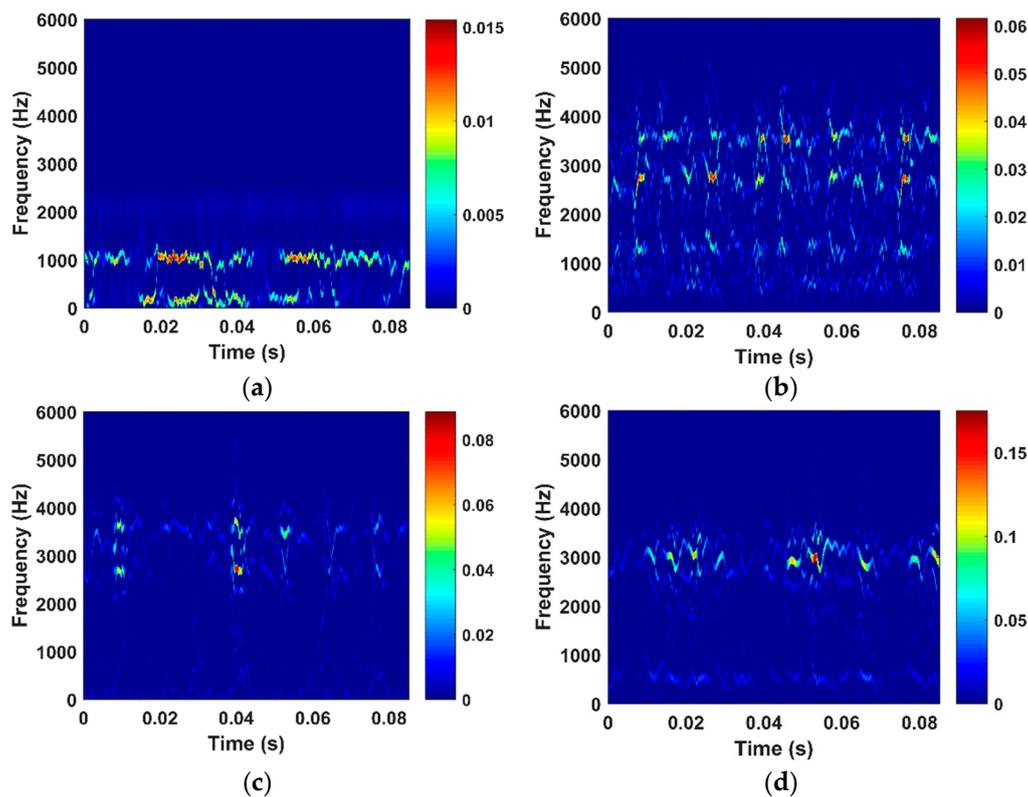
**Figure 8.** TFIs obtained using the STFT from bearing vibration signals of different severity degrees: (a) 0.000 in, (b) 0.007 in, (c) 0.014 in, and (d) 0.021 in.



**Figure 9.** TFIs obtained using the PWVD from bearing vibration signals of different severity degrees: (a) 0.000 in, (b) 0.007 in, (c) 0.014 in, and (d) 0.021 in.



**Figure 10.** TFIs obtained using the WT from bearing vibration signals of different severity degrees: (a) 0.000 in, (b) 0.007 in, (c) 0.014 in, and (d) 0.021 in.



**Figure 11.** TFIs obtained using the STFA-PD from bearing vibration signals of different severity degrees: (a) 0.000 in, (b) 0.007 in, (c) 0.014 in, and (d) 0.021 in.

Compared to Figures 8–10, Figure 11 shows that the time–frequency concentration and resolution of the STFA-PD were higher than the STFT, PWVD, and WT methods, and there was no cross-term interference problem. Moreover, the STFA-PD filtered out a lot of noise, which made the characteristics of the high-frequency impact more obvious. The differences between the TFIs of different fault severity degrees were higher than those obtained using the other three methods. As a result, the STFA-PD can better express the time–frequency characteristics of the signals, and thus, build a good foundation to extract the texture features for an effective characterization of bearings with different fault severity degrees.

#### 4.3. GLCM-Based Texture Features of Sparse Time–Frequency Images

To extract the GLCM-based texture features from sparse TFIs, the colored TFIs shown in Figure 10 should be transformed into gray images. In this study, the weighted average method was used to convert the color images into corresponding gray images, each of which had a pixel size of  $429 \times 543$  and contained 256 gray levels ranging from 0 to 255. Moreover, given the characteristics of the TFIs shown above, the offset values  $(\Delta x, \Delta y)$  were taken as  $(1, 0)$ ,  $(1, 1)$ ,  $(0, 1)$ , and  $(-1, 1)$  corresponding to the four directions  $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ , respectively. We then obtained the GLCMs with  $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ , and thus, extracted the 12 texture features for each gray image.

As mentioned above, 80 samples were chosen for Test 1. By extracting the GLCM-based texture features from each sample, we obtained values of the 12 features in Test 1, as shown in Figure 12. The samples were numbered from 1 to 80. No. 1 to No. 20 denoted normal (0.000 in) samples, No. 21 to No. 40 denoted samples with a mild fault severity degree (0.007 in) at the inner race, No. 41 to No. 60 denoted samples with a moderate fault severity degree (0.014 in) at the inner race, and No. 61 to No. 80 denoted samples with a severe fault severity degree (0.021 in) at the inner race.

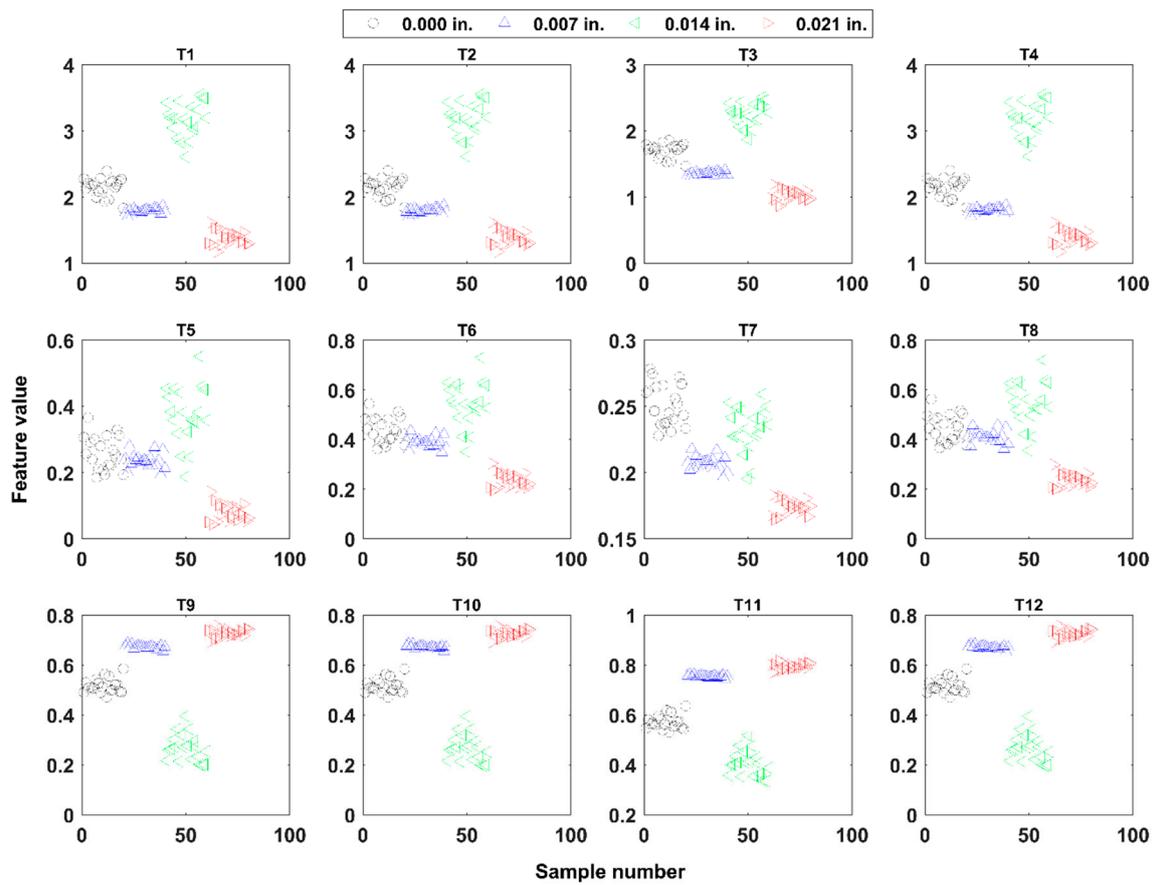


Figure 12. Values of 12 features of samples in Test 1.

The 12 texture features were found to have a relatively low discreteness for the same fault severity degree and good discrimination capability between different fault severity degrees. Therefore, the features could be used to conduct the classification of fault severity degrees of the bearings. From a discrimination viewpoint, among the 12 features, the energy features (T9, T10, T11, and T12) were the best, followed by the entropy features (T1, T2, T3, and T4). The contrast features (T5, T6, T7, and T8) performed the worst. The distribution range of each contrast feature of the samples was wide, and the distributions of the samples with different fault severity degrees overlapped with each other in various degrees.

In addition, we found that, although a single feature had a certain degree of discrimination, it could not completely separate different fault severity degrees, given the observed overlap of sample distributions with different fault severity degrees. Consequently, it was necessary to employ multiple features for pattern recognition.

#### 4.4. Fault Severity Classification

To realize automatic classification of the GLCM-based texture features and evaluate the performance of the sparse TFI texture feature extraction method proposed in this paper, the 1-nearest neighbor classifier (1NNC) [54] and least squares support vector machine (LS-SVM) [55–57] were employed. Both these classifiers are commonly used for pattern recognition, and have stable classification performance and high efficiency. The LS-SVM was performed using LS-SVMlab1.8 [58], and the cross-validation method was used to optimize the parameters. Moreover, the linear kernel function was selected as the kernel function of the LS-SVM.

As mentioned above, 80 samples were chosen for each test. The dataset of each test was divided into two groups: 40 samples (20 samples for each severity degree) were used as the training dataset

and the remaining 40 samples as the testing dataset. The process of random grouping, training, and classification was conducted 10 times to obtain more robust results. The average classification accuracy was used as a performance evaluation index for the proposed method.

Table 2 lists the results of the bearing fault severity recognition. The performance of the STFA-PD was the best of the four time–frequency analysis methods, because it had higher time–frequency concentration and resolution than the others, which made the texture features extracted from the sparse TFIs more accurate for representing differences between the samples with different fault severity degrees.

**Table 2.** Results of bearing fault severity recognition.

Method	Accuracy (%)		
	Test 1	Test 2	Test 3
STFA-PD	99.50	99.25	83.75
STFT	97.00	97.00	77.25
PWVD	85.75	94.00	75.00
WT	91.50	98.75	74.00
STFA-PD	100.00	100.00	96.50
STFT	100.00	100.00	84.25
PWVD	98.25	99.00	84.50
WT	100.00	100.00	93.75

Moreover, the results of the first two tests were better than those of the third one, because the third test was on the ball fault severity recognition. The vibration signals of ball faults were more complex than those of outer race and inner race faults, making the TFIs more complex. Consequently, it was more difficult to use texture features to characterize the signals and obtain better classification accuracy in Test 3 compared with Tests 1 and 2. Nevertheless, the classification accuracy of LS-SVM using the STFA-PD-based texture features was 96.50% in Test 3. The WT-based texture features also yielded satisfactory results; however, they were not always as good as those obtained using the STFA-PD-based features.

## 5. Conclusions

In this paper, an approach for the bearing fault severity monitoring based on the texture feature extraction of sparse TFIs was proposed. The STFA-PD was developed to obtain the sparse TFIs. Then, the GLCM-based texture features of the sparse TFIs were extracted for an automatic classification to realize fault severity monitoring.

The developed STFA-PD was validated using a simulated signal. The result showed that the STFA-PD method can overcome the drawbacks of the traditional time–frequency analysis methods, such as low resolution and cross-term interference. Moreover, it was found that the STFA-PD method is better in terms of denoising performance. Therefore, the STFA-PD method can be used to obtain high-quality TFIs without suffering from cross-term interference. Since the STFA-PD method is a universal time-frequency analysis method, it can be used to analyze other non-stationary signals.

Experiments were conducted on the popular CWRU bearing dataset to validate the effectiveness of the proposed bearing fault severity monitoring approach. It was found that the STFA-PD method obtained the most high-quality TFIs of bearing vibration signals. As a result, the proposed approach obtained higher overall accuracy than those based on STFT, PWVD, and WT. This study clearly showed that the proposed approach has significant potential to be a powerful tool for the fault severity monitoring of rolling bearings and can be applied to other rotating machinery such as gearboxes and rotors.

However, there were some limitations to the proposed approach. One is that a training feature set is needed for this approach. In practice, this set is often difficult to obtain, which restricts the

application of the approach. Future work will focus on monitoring the bearing fault severity without a training feature set. Noting that 1NNC and LS-SVM are supervised learning algorithms, in future work, we will attempt to use some semi-supervised or unsupervised learning algorithms to recognize the bearing fault severity. Another limitation of the proposed approach is that the GLCM-based features are not correlated with physical significance. We cannot identify the bearing fault severity from the feature values. In the future, we intend to search for or construct more powerful features with physical meaning to more effectively capture the pattern of fault severity. The energy (or amplitude) information of signals expressed by the TFIs will be used in future studies to extract features correlated with physical significance.

**Author Contributions:** All authors discussed and agreed upon the idea and made scientific contributions. Y.D. and Y.C. conceived the original idea and designed the experiments; J.D. and Y.X. conducted the experiments; Y.D. wrote the paper. G.M. is the advisor of Y.D., provided the funding, and revised the paper.

**Funding:** This research was funded by the National Key Research and Development Program of China (Grant number 2016YFC0600900), the Yue Qi Distinguished Scholar Project of China University of Mining & Technology(Beijing) (Grant number 800015Z1145), the National Natural Science Foundation of China (Grant number U1361127), the Fundamental Research Funds for the Central Universities (Grant number 00/800015HJ), the Education and Scientific Research Foundation of Education Department of Fujian Province for Middle-aged and Young Teachers (Grant number JAT170352), the Foundation of Department of Education of Guangdong Province (Grant number 2017KCXTD015), and the Open Foundation of Digital Signal and Image Processing Key Laboratory of Guangdong Province (Grant number 2017GDDSIPL-01). The APC was funded by the National Natural Science Foundation of China (Grant number U1361127).

**Acknowledgments:** The authors would like to thank for the supports of all the fundings. The authors would also like to thank Editage [www.editage.cn] for English language editing.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Nomenclature

TFIs	time–frequency image
SR	sparse representation
GLCM	gray level co-occurrence matrix
STFA-PD	sparse time–frequency analysis method based on the first-order primal-dual algorithm
FFT	fast Fourier transform
STFT	short-time Fourier transform
ST	S transform
WVD	Wigner–Ville distribution
PWVD	pseudo Wigner–Ville distribution
WT	wavelet transform
TFR	time–frequency representation
TD-2DPCA	two-direction two-dimensional principal component analysis
TD-2DLDA	two-direction two-dimensional linear discriminant analysis
2DNMF	two-dimensional non-negative matrix factorization
FT	Fourier transform
IFT	inverse Fourier transform
$x(t)$	continuous-time signal
$w(t)$	continuous-time sliding window function of time index $t$
$X(t, \omega)$	STFT coefficient of $x(t)$ at time index $t$ and frequency $\omega$
$x(n)$	discrete signal of period $N$ , where $n = 0, 1, \dots, N - 1$
$w(n)$	discrete sliding window function of time index $n$
$X(m, l)$	discrete STFT coefficient of $x(t)$ at time index $m$ and frequency $l$
$N$	period of discrete signal $x(n)$
$W_N$	operator of Fourier transform coefficient $W_N = \exp(-j2\pi/N)$
$W_N^{ml}$	Fourier transform coefficient
$W_N$	inverse Fourier transform matrix $[W_N]_{k,l} = 1/N \cdot W_N^{-(k-1)(l-1)}$ ; where $k, l = 1, 2, \dots, N$

$x_m$	windowed signal at time $m$ , $x_m = [x_m(0), x_m(1), \dots, x_m(N-1)]^T$
$X_m$	Fourier spectrum of $x_m$ , $X_m = [X_m(m, 0), X_m(m, 1), \dots, X_m(m, N-1)]^T$
$y_m$	short-time truncated sub-signal at time index $m$
$Z_m$	sliding window matrix at time $m$ , $Z_m \in \mathbf{R}^{U \times N}$
$Z_{m+1}$	sliding window matrix at time $m+1$ , $Z_m \in \mathbf{R}^{U \times N}$
$\hat{X}_m$	estimated signal spectrum at time $m$
	sparse time–frequency spectrum, $\hat{X} = [\hat{X}_0, \hat{X}_1, \dots, \hat{X}_{N-1}]$
$\varepsilon_m$	user-specified parameter related to the noise
$\mu$	sparse coefficient, $\mu > 0$
$\mathbf{Y}$	convex set, $\mathbf{Y} = \{\mathbf{Y}_m \in \mathbf{Y} : \ \mathbf{Y}_m\ _\infty \leq 1\}$
$\mathbf{Y}_m$	dual variable, $\mathbf{Y}_m \in \mathbf{Y}$
$F(\mu X_m)$	sparse term $\mu \ X_m\ _1$ of Equation (9)
$F^*(\mathbf{Y}_m)$	convex conjugate function of $F(\mu X_m)$ , $F^*(\mathbf{Y}_m) = \begin{cases} 0 & \text{if } \mathbf{Y}_m \in \mathbf{Y} \\ +\infty & \text{if } \mathbf{Y}_m \notin \mathbf{Y} \end{cases}$
$G(X_m)$	fidelity term $\frac{1}{2} \ Z_m W_N X_m - y_m\ _2^2$ of Equation (9)
$\tau, \sigma$	leaning ratios, $\tau, \sigma > 0$
$\theta$	acceleration factor, $\theta \in [0, 1]$
<i>Maxiter</i>	the maximum number of iterations
$X_m^{(n)}$	value of $X_m$ in the $n$ th iteration
$X_m^{(n+1)}$	value of $X_m$ in the $(n+1)$ th iteration
$Y_m^{(n)}$	value of $Y_m$ in the $n$ th iteration
$Y_m^{(n+1)}$	value of $Y_m$ in the $(n+1)$ th iteration
$\bar{X}_m^{(n)}$	acceleration variable of $X_m$ in the $n$ th iteration
$\bar{X}_m^{(n+1)}$	acceleration variable of $X_m$ in the $(n+1)$ th iteration
$s_1(t)$	a sinusoidal signal, $s_1(t) = \sin(30\pi t)$
$s_2(t)$	a nonlinear frequency-modulated signal, $s_2(t) = \sin(30 + 75\pi t \cdot 3^t)$
$s_3(t)$	a nonlinear frequency-modulated signal, $s_3(t) = \sin[200\pi t + 200\pi t^2 + 20\pi t \cos(2\pi t)]$
$s(t)$	a multi-component simulated signal, $s(t) = s_1(t) + s_2(t) + s_3(t)$
$z(t)$	a noisy signal
$P(i, j, d, \theta)$	co-occurrence frequency of a pixel pairs
$\Gamma\{\cdot\}$	total number satisfying the conditions
$g(i, j)$	co-occurrence probability $g(i, j) = P(i, j) / \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j)$

## References

1. Cui, L.; Chunqing, M.A.; Zhang, F.; Wang, H. Quantitative diagnosis of fault severity trend of rolling element bearings. *Chin. J. Mech. Eng.* **2015**, *28*, 1254–1260. [[CrossRef](#)]
2. Glowacz, A.; Glowacz, W.; Glowacz, Z.; Kozik, J. Early fault diagnosis of bearing and stator faults of the single-phase induction motor using acoustic signals. *Measurement* **2018**, *113*, 1–9. [[CrossRef](#)]
3. Berredjem, T.; Benidir, M. Bearing faults diagnosis using fuzzy expert system relying on an improved range overlaps and similarity method. *Expert Syst. Appl.* **2018**, *108*, 134–142. [[CrossRef](#)]
4. Hong, H.; Liang, M. Fault severity assessment for rolling element bearings using the Lempel–Ziv complexity and continuous wavelet transform. *J. Sound Vib.* **2009**, *320*, 452–468. [[CrossRef](#)]
5. Hong, S.; Zhou, Z.; Zio, E.; Hong, K. Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method. *Dig. Signal Process.* **2014**, *27*, 159–166. [[CrossRef](#)]
6. Duan, Z.; Wu, T.; Guo, S.; Shao, T.; Malekian, R.; Li, Z. Development and trend of condition monitoring and fault diagnosis of multi-sensors information fusion for rolling bearings: A review. *Int. J. Adv. Manuf. Technol.* **2018**, *96*, 803–819. [[CrossRef](#)]
7. Wang, H.; Chen, J. Performance degradation assessment of rolling bearing based on bispectrum and support vector data description. *J. Vib. Control* **2013**, *20*, 2032–2041. [[CrossRef](#)]
8. Cerrada, M.; Sánchez, R.; Li, C.; Pacheco, F.; Cabrera, D.; José, V.D.O.; Vásquez, R.E. A review on data-driven fault severity assessment in rolling bearings. *Mech. Syst. Signal Process.* **2018**, *99*, 169–196. [[CrossRef](#)]

9. Molina, M.A.; Gohar, R. Hydrodynamic lubrication of ball-bearing cage pockets. *ARCHIVE J. Mech. Eng. Sci.* **1978**, *20*, 11–20. [[CrossRef](#)]
10. Rahnejat, H.; Gohar, R. The vibrations of radial ball bearings. *ARCHIVE Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci.* **1985**, *199*, 181–193. [[CrossRef](#)]
11. Wardle, F.P. Vibration forces produced by waviness of the rolling surfaces of thrust loaded ball bearings, part 1: Theory. *Proc. Inst. Mech. Eng.* **1988**, *202*, 305–312. [[CrossRef](#)]
12. Johns-Rahnejat, P.; Gohar, R. Point contact elastohydrodynamic pressure distribution and sub-surface stress field. In Proceedings of the International Tri-Annual Conference on Multi-Body Dynamics: Monitoring and Simulation Techniques, Bradford, UK, 25–27 March 1997; pp. 161–177.
13. Rai, A.; Upadhyay, S.H. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribol. Int.* **2016**, *96*, 289–306. [[CrossRef](#)]
14. Borghesani, P.; Ricci, R.; Chatterton, S.; Pennacchi, P. A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions. *Mech. Syst. Signal Process.* **2013**, *38*, 23–35. [[CrossRef](#)]
15. Randall, R.B.; Antoni, J. Rolling element bearing diagnostics—A tutorial. *Mech. Syst. Signal Process.* **2011**, *25*, 485–520. [[CrossRef](#)]
16. Peeters, C.; Guillaume, P.; Helsen, J. A comparison of cepstral editing methods as signal pre-processing techniques for vibration-based bearing fault detection. *Mech. Syst. Signal Process.* **2017**, *91*, 354–381. [[CrossRef](#)]
17. Bi, G.; Chen, J.; He, J.; Zhou, F.; Zhang, G.C. Application of degree of cyclostationarity in rolling element bearing diagnosis. *Key Eng. Mater.* **2005**, *293–294*, 347–354. [[CrossRef](#)]
18. Liu, X.; Bo, L.; Peng, C. Application of order cyclostationary demodulation to damage detection in a direct-driven wind turbine bearing. *Meas. Sci. Technol.* **2014**, *25*, 25004–25011. [[CrossRef](#)]
19. Acuña, D.Q.; Vicuña, C.M. *Damage Assessment of Rolling Element Bearing Using Cyclostationary Processing of Ae Signals with Electromagnetic Interference*; Springer International Publishing: Basel, Switzerland, 2015; pp. 43–54.
20. Feng, Z.; Liang, M.; Chu, F. Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples. *Mech. Syst. Signal Process.* **2013**, *38*, 165–205. [[CrossRef](#)]
21. Zhao, M.; Tang, B.; Tan, Q. Fault diagnosis of rolling element bearing based on s transform and gray level co-occurrence matrix. *Meas. Sci. Technol.* **2015**, *26*, 085008. [[CrossRef](#)]
22. Peng, Z.K.; Chu, F.L. Application of the wavelet transform in machine condition monitoring and fault diagnostics: A review with bibliography. *Mech. Syst. Signal Process.* **2004**, *18*, 199–221. [[CrossRef](#)]
23. Chen, Y.P.; Peng, Z.M.; He, Z.H.; Lin, T.; Zhang, D.J. The optimal fractional gabor transform based on the adaptive window function and its application. *Appl. Geophys.* **2013**, *10*, 305–313. [[CrossRef](#)]
24. Chen, Y.; Peng, Z.; Cheng, Z.; Tian, L. Seismic signal time-frequency analysis based on multi-directional window using greedy strategy. *J. Appl. Geophys.* **2017**, *143*, 116–128. [[CrossRef](#)]
25. Cohen, L. Time-frequency distribution—A review. *Proc. IEEE* **1989**, *77*, 941–981. [[CrossRef](#)]
26. Hlawatsch, F.; Boudreaux-Bartels, G.F. Linear and quadratic time-frequency signal representations. *IEEE Signal Process. Mag.* **1992**, *9*, 21–67. [[CrossRef](#)]
27. Hess-Nielsen, N.; Wickerhauser, M.V. Wavelets and time-frequency analysis. *Proc. IEEE* **1996**, *84*, 523–540. [[CrossRef](#)]
28. Sejdíć, E.; Djurović, I.; Jiang, J. Time–frequency feature representation using energy concentration: An overview of recent advances. *Digit. Signal Process.* **2009**, *19*, 153–183. [[CrossRef](#)]
29. Stockwell, R.G.; Mansinha, L.; Lowe, R.P. Localization of the complex spectrum: The s transform. *IEEE Trans. Signal Process.* **1996**, *44*, 998–1001. [[CrossRef](#)]
30. Vafaei, S.; Rahnejat, H. Indicated repeatable runout with wavelet decomposition (irr-wd) for effective determination of bearing-induced vibration. *J. Sound Vib.* **2003**, *260*, 67–82. [[CrossRef](#)]
31. Mallat, S.G.; Zhang, Z. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.* **1993**, *41*, 3397–3415. [[CrossRef](#)]
32. Chen, S.S.; Donoho, D.L.; Saunders, M.A. Atomic decomposition by basis pursuit. *Siam Rev.* **2001**, *43*, 129–159. [[CrossRef](#)]

33. Borgnat, P.; Flandrin, P. Time-frequency localization from sparsity constraints. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March 2008–4 April 2008; pp. 3785–3788.
34. Elad, M. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*; Springer Publishing Company, Incorporated: Basel, Switzerland, 2010; pp. 1094–1097.
35. Cheng, H.; Liu, Z.; Yang, L.; Chen, X. Sparse representation and learning in visual recognition: Theory and applications. *Signal Process.* **2013**, *93*, 1408–1425. [[CrossRef](#)]
36. Wang, C.; Gan, M.; Zhu, C.A. Fault feature extraction of rolling element bearings based on wavelet packet transform and sparse representation theory. *J. Intell. Manuf.* **2015**, *11*, 1–15. [[CrossRef](#)]
37. Liu, Z.; Wei, X.; Li, X. Aliasing-free micro-doppler analysis based on short-time compressed sensing. *Signal Process. IET* **2013**, *8*, 176–187. [[CrossRef](#)]
38. Liu, Z.; You, P.; Wei, X.; Liao, D.; Li, X. High resolution time-frequency distribution based on short-time sparse representation. *Circ. Syst. Signal Process.* **2014**, *33*, 3949–3965. [[CrossRef](#)]
39. Gholami, A. Sparse time-frequency decomposition and some applications. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 3598–3604. [[CrossRef](#)]
40. Hu, J.; He, X.; Li, W.; Ai, H.; Li, H.; Xie, J. Parameter estimation of maneuvering targets in othr based on sparse time-frequency representation. *J. Syst. Eng. Electron.* **2016**, *27*, 574–580. [[CrossRef](#)]
41. Chambolle, A.; Pock, T. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **2011**, *40*, 120–145. [[CrossRef](#)]
42. Sidky, E.Y.; Jørgensen, J.H.; Pan, X. Convex optimization problem prototyping for image reconstruction in computed tomography with the chambolle-pock algorithm. *Phys. Med. Biol.* **2012**, *57*, 3065–3091. [[CrossRef](#)] [[PubMed](#)]
43. Li, B.; Zhang, P.L.; Liu, D.S.; Mi, S.S.; Ren, G.Q.; Tian, H. Feature extraction for rolling element bearing fault diagnosis utilizing generalized s transform and two-dimensional non-negative matrix factorization. *J. Sound Vib.* **2011**, *330*, 2388–2399. [[CrossRef](#)]
44. Wang, W.G.; Liu, Z.S. Classification of time-frequency images based on locality-constrained linear coding optimization model for rotating machinery fault diagnosis. *ARCHIVE Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci.* **2015**, *229*, 3350–3360. [[CrossRef](#)]
45. Li, W.; Lin, L.; Shan, W. Bearing fault diagnosis based on generalized s-transform and two-directional 2dpca. *Zhendong Ceshi Yu Zhenduan/J. Vib. Meas. Diagn.* **2015**, *35*, 499–506.
46. Li, B.; Zhang, P.L.; Liu, D.S.; Mi, S.S.; Liu, P.Y. Classification of time-frequency representations based on two-direction 2ddlda for gear fault diagnosis. *Appl. Soft Comput. J.* **2011**, *11*, 5299–5305. [[CrossRef](#)]
47. Haraclick, R.M. Texture features for image classification. *IEEE Trans. Syst. Man. Cybern.* **1973**, *3*, 610–621. [[CrossRef](#)]
48. Nawab, S.H. Short-Time Fourier Transform. *Adv. Top. Signal Process.* **1988**, *32*, 289–337.
49. Pantic, I.; Dimitrijevic, D.; Nestic, D.; Petrovic, D. Gray level co-occurrence matrix algorithm as pattern recognition biosensor for oxidopamine-induced changes in lymphocyte chromatin architecture. *J. Theor. Biol.* **2016**, *406*, 124–128. [[CrossRef](#)] [[PubMed](#)]
50. Baraldi, A.; Parmiggiani, F. Investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 293–304. [[CrossRef](#)]
51. Smith, W.A.; Randall, R.B. Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mech. Syst. Signal Process.* **2015**, *64–65*, 100–131. [[CrossRef](#)]
52. Loparo, K.A. Bearings Vibration Data Set, Case Western Reserve University. Available online: <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file> (accessed on 27 October 2017).
53. Pang, B.; Tang, G.; Tian, T.; Zhou, C. Rolling bearing fault diagnosis based on an improved htt transform. *Sensors* **2018**, *18*, 1203. [[CrossRef](#)] [[PubMed](#)]
54. Grother, P.J.; Candela, G.T.; Blue, J.L. Fast implementations of nearest neighbor classifiers. *Pattern Recognit.* **1997**, *30*, 459–465. [[CrossRef](#)]
55. Ye, J.; Xiong, T. Svm versus least squares svm. *J. Mach. Learn. Res.* **2007**, *2*, 644–651.
56. Zhang, Y.; Zhang, C.; Sun, J.; Jingjing, G. Improved wind speed prediction using empirical mode decomposition. *Adv. Electr. Comput. Eng.* **2018**, *18*, 3–10. [[CrossRef](#)]

57. Zhang, C.; Peng, Z.; Chen, S.; Li, Z.; Wang, J. A gearbox fault diagnosis method based on frequency-modulated empirical mode decomposition and support vector machine. *ARCHIVE Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci.* **2016**, *232*, 369–380. [CrossRef]
58. Ls-svmlab1.8. Available online: <http://www.esat.kuleuven.ac.be/sista/lssvmlab/> (accessed on 14 January 2018).



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).