

Research Article

Fault Diagnosis for Rolling Bearing under Variable Conditions Based on Image Recognition

Bo Zhou^{1,2} and Yujie Cheng^{1,2}

¹*School of Reliability and Systems Engineering, Beihang University, No. 37, Xueyuan Road, Haidian District, Beijing 100191, China*

²*Science & Technology on Reliability and Environmental Engineering Laboratory, Beijing 100191, China*

Correspondence should be addressed to Yujie Cheng; chengyujie@buaa.edu.cn

Received 25 May 2016; Accepted 14 July 2016

Academic Editor: Minvydas Ragulskis

Copyright © 2016 B. Zhou and Y. Cheng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rolling bearing faults often lead to electromechanical system failure due to its high speed and complex working conditions. Recently, a large amount of fault diagnosis studies for rolling bearing based on vibration data has been reported. However, few studies have focused on fault diagnosis for rolling bearings under variable conditions. This paper proposes a fault diagnosis method based on image recognition for rolling bearings to realize fault classification under variable working conditions. The proposed method includes the following steps. First, the vibration signal data are transformed into a two-dimensional image based on recurrence plot (RP) technique. Next, a popular feature extraction method which has been widely used in the image field, scale invariant feature transform (SIFT), is employed to extract fault features from the two-dimensional RP and subsequently generate a 128-dimensional feature vector. Third, due to the redundancy of the high-dimensional feature, kernel principal component analysis is utilized to reduce the feature dimensionality. Finally, a neural network classifier trained by probabilistic neural network is used to perform fault diagnosis. Verification experiment results demonstrate the effectiveness of the proposed fault diagnosis method for rolling bearings under variable conditions, thereby providing a promising approach to fault diagnosis for rolling bearings.

1. Introduction

Rolling bearings are considered to be a critical mechanical component in industrial applications. The bearings' defects usually lead to a considerable decline in plant productivity and may even cause huge economic losses [1, 2]. Thus, it is important to diagnose rolling bearing fault to keep the bearings in good technical state.

Vibration-based methods have garnered particular attention due to their noninvasive nature and their high reactivity to incipient fault. Therefore, vibration signal analysis is vital for rolling bearings fault diagnosis due to its connection to fault feature extraction accuracy [3]. Aiming to extract the fault features, many feature extraction methods, including Wigner-Ville distribution (WVD) [4], wavelet packet decomposition (WPT) [5, 6], and empirical mode decomposition (EMD) [7–9], have been proposed and have been demonstrated to be powerful. Additionally, many fault diagnosis methods also have been proposed, such as fast spectral

kurtosis based on genetic algorithms [10], multiscale entropy and adaptive neurofuzzy inference system [11], and time varying singular value decomposition [12]. However, most of these methods are proposed based on the assumption that the rolling bearings operate under fixed conditions when performing fault diagnosis. Moreover, the application of these methods is limited because of the tough, complex, and particularly variable working environment of rolling bearings [13, 14]. Therefore, it is important to investigate the fault diagnosis method suitable for varying conditions.

Many studies have researched rolling bearings fault diagnosis. However, thus far, few researchers have studied fault diagnosis under variable conditions. In 1990, Potter [15] proposed a constant angular sampling method (i.e., order tracking) that utilized the electronic impulse angular encoder and solved the frequency smearing phenomenon of the spectrum caused by fluctuating rotating speeds and realized the fault diagnosis for rotating machines. Considering the special analog hardware whose function is sampling data increases

the cost of equipment; Fyfe and Munck [16] developed the computed order tracking (COT) technique based on order tracking to realize fault diagnosis for rotating machines. However, the COT may make the carrier frequencies of the transient responses, which are caused by the faults at various speeds, expand to a wider scope because the natural characteristic of the bearing system hardly changes, which is not beneficial for extracting the fault characteristic. In addition, [13] has proposed a new method for rolling bearing fault diagnosis under variable conditions. This new method utilizes LMD-SVD to extract features, but LMD also has the problem of iterative calculation capacity, frequency aliasing, end effect and other issues. Because of problems associated with the above methods, we need to research a new method for bearing vibration signal feature extraction, a method based on the nonstationary and nonlinear bearing vibration signals, thereby achieving fault diagnosis under variable conditions.

Scale invariant feature transform (SIFT), an image invariant feature extraction method, can recognize the same image when it is rotated, scaled, translational, and affine transformed. By extracting the 128-dimensional feature containing scale, orientation, and location information, SIFT can perform image recognition and matching under translation, rotation, scaling, and brightness changes [17]. Many studies have used SIFT to recognize images. For example, Montazer and Giveki [18] have utilized SIFT to extract image features and match them to a database (i.e., a content-based image retrieval system). Li et al. [19] have employed Robust SIFT to match remote sensing images, and a number of studies have also applied SIFT to such methods as facial expression recognition [20], ear recognition for a new biometric technology [21], and wheat grain classification [22]. Inspired by SIFT, the vibration signals of rolling bearings are considered to be transformed into images. The recurrence plot (RP) is a kind of method to describe the recursive behavior of dynamic orbit in the phase-space reconstruction; it is an important method to analyze the instability of time series. RPs of rolling bearing vibration signals under different conditions reveal translation and scaling characteristics, so RP is employed to transform the vibration signals under different conditions into images and SIFT is utilized to extract the features of transformed RPs, which is without interference of working conditions.

After the 128-dimensional invariant features are extracted, to reduce the data redundancy between the extracted features and the occupation of computer resources, a dimensionality reduction method is utilized to identify the low-dimensional structure hidden in high-dimensional data. Principal component analysis (PCA) is a widely utilized dimension reduction technique performed by linearly transforming a high-dimensional input space onto a lower dimensional one in which the components are uncorrelated. However, PCA will not perform well when the process exhibits nonlinearity. Hence, kernel principal component analysis (KPCA) was developed to overcome the limitations of PCA in dealing with the nonlinear system [23].

This paper is structured as follows. Section 2 first introduces the image transformation method, which generates images for the following recognition. Then, SIFT, the core

of this paper, is described, which is utilized to extract the stable fault features under variable working conditions. Subsequently, KPCA is introduced for the dimensionality reduction. At last, probabilistic neural network (PNN) is described for the final fault classification. Section 3 describes the entire fault diagnosis method for rolling bearing under variable conditions, including descriptions of the experimental data, image transformation, feature extraction, and fault classification. Section 4 includes the results and discussion, and the conclusions are presented in Section 5.

2. Related Theories

2.1. Recurrence Plot. To achieve fault diagnosis under variable conditions, image transformation for SIFT is important to ensure success. Therefore, choosing a good image transformation method is particularly important. On account of the nonlinear and nonstationary characteristics of rolling bearing signals, detecting dynamical changes in complex systems is one of the most difficult problems. Recursiveness is one of the basic characteristics of a dynamic system, and the recurrence plot (RP) based on this characteristic is a good dynamic mainstream shape-description method. Through the black and white dots in the two-dimensional space, the recursive state can be visualized in the phase space [24]. This approach can uncover hidden periodicities in a signal in the recurrence domain. These periodicities are not easily noticeable, and it is an important method that analyzes the periodic, chaotic, and nonstationary of time series. The following theories are related.

The RP analysis is based on the phase-space reconstruction theory, which is described as follows.

- (1) For a time series, u_k ($k = 1, 2, \dots, N$), whose sample interval is Δt , we chose the mutual information method and CAO algorithm to calculate the suitable embedding dimension m and delay time τ , which could reconstitute the time series. The reconstructed time series is

$$x_i = (u_i, u_{i+\tau}, \dots, u_{i+(m-1)\tau}), \quad i = 1, 2, \dots, N - (m-1)\tau. \quad (1)$$

- (2) Calculate the Euclidean norm of x_i and x_j in the reconstructed phase space [25]:

$$S_{ij} = \|x_i - x_j\|, \quad i = 1, 2, \dots, N - (m-1)\tau; \quad j = 1, 2, \dots, N - (m-1)\tau. \quad (2)$$

- (3) Calculate the recurrence value [26]:

$$R(i, j) = H(\varepsilon_i - S_{ij}), \quad i = 1, 2, \dots, N; \quad j = 1, 2, \dots, N, \quad (3)$$

where ε is the threshold value and $H(r)$ is the Heaviside function:

$$H(r) = \begin{cases} 1 & r \geq 0 \\ 0 & r < 0. \end{cases} \quad (4)$$

- (4) Utilize a coordinate graph whose abscissa is i and whose ordinate is j to draw $R(i, j)$, where i and j are the time series labels and the image is RP.

2.1.1. Mutual Information Method. The mutual information method estimating the delay time has been proposed by Fraser and Swinney, based on the Shannon information theory, which is widely used in phase-space reconstruction [27].

The Shannon theory shows that we can obtain the information content of a_i from the event b_j :

$$I_{AB}(a_i, b_j) = \log_2 \left[\frac{P_{AB}(a_i, b_j)}{P_A(a_i) P(b_j)} \right]. \quad (5)$$

The relationship between a_i and b_j could be expressed with comentropy I_{AB} :

$$I_{AB} = \sum_{ij} P_{AB}(a_i, b_j) \log_2 \left[\frac{P_{AB}(a_i, b_j)}{P_A(a_i) P(b_j)} \right]. \quad (6)$$

Apply the theory of the mutual information, and set A is

$$\{A : a_i = x_i = x(t_0 + i\tau_x)\} \quad (7)$$

and set B is

$$\{B : b_i = x_i = x(t_0 + i\tau_x + \tau)\}. \quad (8)$$

Then, (6) translates into

$$I_{AB}(\tau) = \sum_i P[x(t_0 + i\tau_x), x(t_0 + i\tau_x + \tau)] \cdot \log_2 \left\{ \frac{P[x(t_0 + i\tau_x), x(t_0 + i\tau_x + \tau)]}{P[x(t_0 + i\tau_x)] P[x(t_0 + i\tau_x + \tau)]} \right\}. \quad (9)$$

Usually at the beginning, $I_{AB}(\tau)$ is large; therefore, we can obtain an infinite amount of information in $x(t) = x(t + 0)$. $x(t_0 + i\tau_x)$ and $x(t_0 + i\tau_x + \tau)$ are completely independent for chaotic signals when τ is large; when $\tau \rightarrow \infty$, $I(\tau) \rightarrow 0$. Generally the first minimum of $I_{AB}(\tau)$ is selected as the delay time.

2.1.2. CAO Algorithm. The CAO algorithm was proposed by CAO in 1997, and it has excellent properties to clearly distinguish real signal and noise, as well as high computational efficiency [28]. First, we calculated the distance of the points under the embedded dimensionality:

$$a(i, m) = \frac{\|u_i(m+1) - u_{n(i,m)}(m+1)\|}{\|u_i(m) - u_{n(i,m)}(m)\|}, \quad (10)$$

$$i = 1, 2, \dots, N - m,$$

where $\|\cdot\|$ is ∞ the norm of the vector; $u_i(m+1)$ is i th vector after phase-space rebuilding, and the embedded dimension is $m+1$; $u_{n(i,m)}(m+1)$ is the nearest vector from $u_i(m+1)$.

Next, we calculated the average value of the distance change under the same dimension:

$$E(m) = \frac{1}{N-m} \sum_{i=1}^{N-m} a(i, m), \quad (11)$$

where $E(m)$ is the average value of all $a(i, m)$.

Finally, according to the discriminant equation

$$E_1(m) = \frac{E(m+1)}{E(m)} \quad (12)$$

when $m > m_0$, $E_1(m)$ stops changing or changes slowly, and $m_0 + 1$ is the minimum embedding dimension.

2.2. SIFT Theory. Recognizing the images that are rotating, scaling, and translating refers to finding the stable points of the images. These points, such as the corners, blobs, T-junctions, and light spots in dark regions, do not disappear with the rotation, scale, translation, and brightness changes. SIFT was developed by Nurhaida et al. to extract distinctive invariant features from images that can be used to perform reliable matching between different views of an object or scene [29]. SIFT has five basic steps: constructing scale space, extreme points detection, precise location of key points, orientation assignment, and descriptor calculation [30].

2.2.1. Gaussian Blur. SIFT finds key points in the different scale spaces, and the acquisition of scale space needs to be realized using Gaussian blur. Lindeberg has proved that Gaussian convolution kernel is the only kernel to achieve scale transformation, and it is the only linear kernel [31].

Gaussian blur is an image filter that utilizes normal distribution to calculate the fuzzy template, and the template is used to perform convolution operations with the original image to achieve the transition of fuzzy images.

The normal distribution equation of N dimensional space is

$$G(r) = \frac{1}{\sqrt{2\pi\sigma^2}^N} e^{-r^2/(2\sigma^2)}, \quad (13)$$

where σ is the standard deviation of the normal distribution; the larger σ is, the fuzzier image is. r is the fuzzy radius that refers to the distance between the template element and the center of the template. If the two-dimensional template size is $m \times n$, then (x, y) on the template corresponding to the Gauss equation is

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-((x-m/2)^2 + (y-n/2)^2)/2\sigma^2}. \quad (14)$$

According to the value of σ , the size of the Gaussian blur template matrix is $(6\sigma + 1) \times (6\sigma + 1)$. Equation (14) is used to calculate the value of the Gaussian template matrix; convolution is calculated with the original image, and the Gaussian blur image of the original image is obtained.

2.2.2. Scale Space Construction

(1) *Scale Space Theory*. Scale space theory was first proposed by Iijima in 1962, and it was widely used in the field of computer vision after being promoted by Duits et al. [32].

The basic concept of scale space is as follows. A scale parameter is introduced in the image model, and the scale space sequence at multiple scales is obtained through the continuous change of scale parameter. The principal contours are extracted from the scale space of these sequences, and the principal contours are utilized as a feature vector to realize edge detection, corner detection, and feature extraction at different resolutions.

(2) *Representation of Scale Space*. The scale space $L(x, y, \sigma)$ of an image is defined as the convolution calculation between the Gauss function $G(x, y, \sigma)$ and the original image $I(x, y)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad (15)$$

where $*$ represents convolution:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-((x-m/2)^2 + (y-n/2)^2)/2\sigma^2}, \quad (16)$$

where m and n are the dimensionality of the Gaussian template determined by $(6\sigma + 1) \times (6\sigma + 1)$. (x, y) is the pixel location in the image. σ is the scale space factor, the smaller value of which is the least amount of the smoothed image, and the corresponding scale is smaller.

(3) *Gaussian Pyramid*. The pyramid model of an image is as follows: the original image is constantly downsampling, and it generates a series of different sizes of images, ranging from large to small and from the bottom to the top, thereby constructing a tower-shaped model. The original image is the first stratum of the pyramid, and the new image obtained through downsampling is a stratum of the pyramid. The number of strata in the pyramid is jointly decided through the size of the original and top images. The equation is as follows:

$$n = \log_2 \{ \min(M, N) \} - t, \quad (17)$$

$$t \in [0, \log_2 \{ \min(M, N) \}],$$

where M and N are the sizes of the original image and t is logarithm of the minimum dimensionality of the top image.

To reflect the continuity of scale, the Gaussian pyramid introduces the Gaussian filter on the simple downsampling, as shown in Figure 1. The image in each stratum calculates the Gaussian blur using different parameters; thus, each stratum of the pyramid contains multiple Gaussian blur images. The images in each stratum are named octaves. The initial image (bottom image) of an octave in the Gaussian Pyramid is obtained by sampling from the last third image of the previous octave of images.

(4) *DOG Pyramid*. In 2002, Mikolajczyk found that the scale normalization of the Laplacian Gaussian function can produce the most stable image features compared to other feature extraction functions. The difference of Gaussian (DOG)

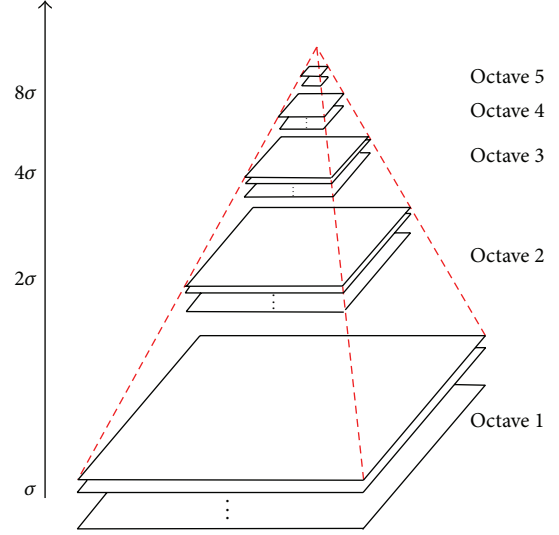


FIGURE 1: Gaussian pyramid.

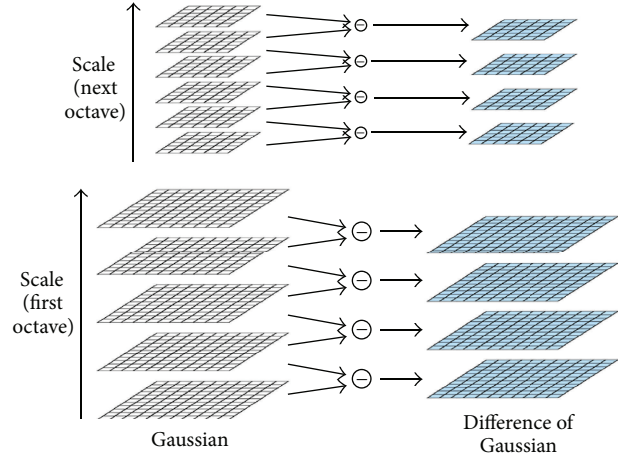


FIGURE 2: Gaussian pyramid generation.

function is similar to the scale normalization of the Laplacian Gaussian function [17]. Therefore, the DOG filter is applied to the input image. The image is gradually downsampled, and the filtering is performed at several scales. Figure 2 demonstrates the creation process for the DOG filters at different scales.

2.2.3. *Extreme Point Detection*. The key points are the local extreme points (in the DOG space) whose initial exploration is accomplished by comparing the two adjacent images of each DOG in the same group. To determine the key points, a $3 \times 3 \times 3$ neighborhood comparison is used, as shown in Figure 3. Each pixel processed by the DOG pyramid is compared with 26 points of its 3-dimensional neighborhood to obtain the maximum or minimum, as the preliminary feature points.

2.2.4. Precise Location of Key Points

(1) *Location of Interpolation*. The extreme points detected by the above methods are the extreme points of the discrete

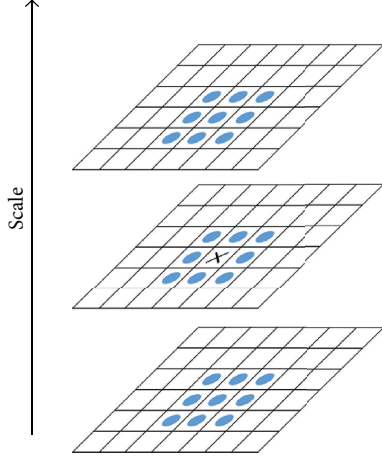


FIGURE 3: Key point localization.

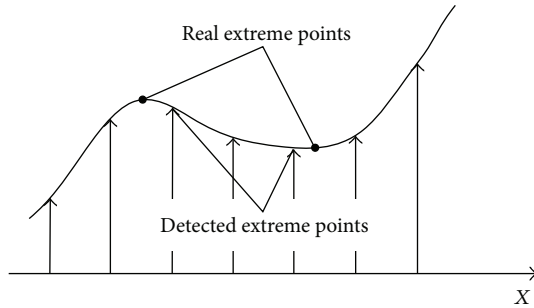


FIGURE 4: Difference between discrete space and continuous space.

space, which are not the real extreme points. Figure 4 shows the difference of the extreme point of a two-dimensional function in discrete and continuous space. SIFT utilizes the linear interpolation method to obtain accurate key points.

(2) *Remove Edge Response*. The detected key points are further examined to choose the “best” candidates. The stability of the resulting set of key points is determined. Locations with low contrast and unstable locations along edges are discarded by calculating the ratio of the square of the matrix trace and determining the Hessian matrix.

2.2.5. Orientation Assignment. To determine the descriptor with rotation invariance, the local features of images are needed to assign a reference orientation for each key point. By calculating the gradient orientations of the neighborhood pixels of key points, the orientation parameter is specified for each feature point. The gradient values and orientations in (x, y) are

$$m(x, y) = \sqrt{[L(x+1, y) - L(x-1, y)]^2 + [L(x, y+1) - L(x, y-1)]^2} \quad (18)$$

$$\theta(x, y) = \tan^{-1} \left\{ \frac{[L(x, y+1) - L(x, y-1)]}{[L(x+1, y) - L(x-1, y)]} \right\}.$$

The gradient histogram statistical method is employed to further ascertain the orientation of the key point. The gradient values of the key points in the neighborhood window are calculated with the key point as the center and 1.5σ as the radius. The 360° of a circle are divided into 36 bins by drawing the gradient histogram, the contribution of each neighborhood point to the orientation of the gradient decreases with the increase of the distance between the neighborhood and the key point. The peak of the histogram is the main orientation of the key points.

After selecting the main orientation, there may also be one or more peaks whose values are more than 80% of the main peak. To enhance the robustness of the match, some key points whose locations and scales are the same as the original key point will be employed.

2.2.6. Descriptor Calculation. After performing the above steps, the key points have the location, scale, and orientation. The next step is to create a descriptor for the key points. First, the coordinate axes are rotated as the key points to ensure the rotation invariance. Then, a 16×16 window is taken; the center of this window has the key points shown in Figure 5(a). Each grid represents a pixel in the scale space of the neighborhood of the key point; the orientation of the arrow represents the gradient orientation of the pixel, and the length of the arrow represents the gradient mode. In the figure, the circle represents the Gauss range weighted. Next, the gradient orientation histogram of 8 orientations is calculated in every 4×4 image, and the seed point is formed by drawing the cumulative value of each gradient orientation, as shown in Figure 5(b). In this figure, a key point is composed of 16 (4×4) seed points, each of which has eight orientation vectors. The key point can generate 128 ($4 \times 4 \times 8$) data sets and then form a 128-dimensional feature vector. The concept of the neighborhood orientation information alliance enhances the antinoise ability and also provides good fault tolerance for the feature matching with the localization error.

2.3. Kernel Principal Component Analysis. KPCA projects the m dimensional observed data matrix ($\mathbf{X} \in \mathbf{R}^m$, input space) onto a high-dimensional feature space \mathbf{F} , which can be expressed as

$$\Phi : \mathbf{R}^m \longrightarrow \mathbf{F}. \quad (19)$$

Similar to PCA, KPCA aims to project a high-dimensional feature space onto a lower space, in which the principal components are linear, uncorrelated combinations of the feature space [33]. The covariance matrix in the feature space can be formulated as

$$\mathbf{C} = \frac{1}{N} \sum_{i=1}^N \Phi(x_i) \Phi^T(x_i). \quad (20)$$

The characteristic equation is

$$\mathbf{C}\mathbf{V} = \lambda\mathbf{V}, \quad (21)$$

where $\Phi(x_i)$ is i th sample in the feature space with zero mean, N denotes the sample size, and T is the transpose operation.

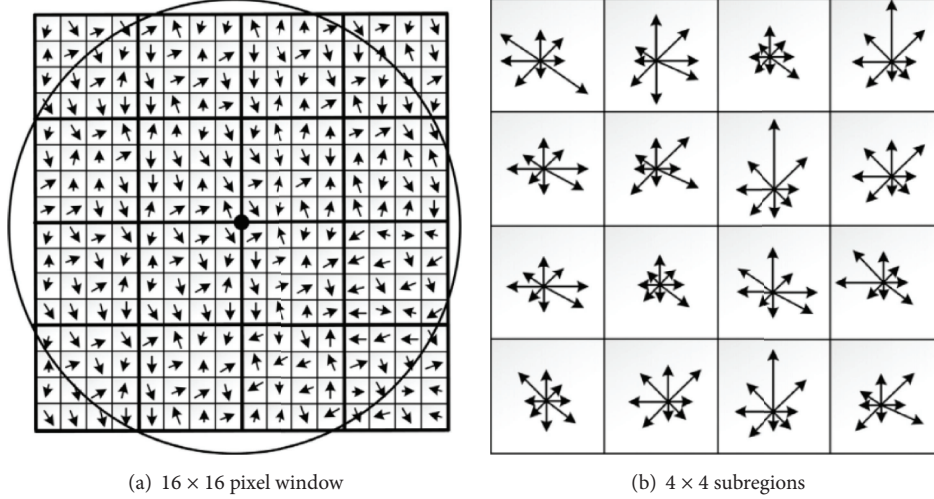


FIGURE 5: Image gradient and key point descriptor.

Let $\theta = [\Phi(x_1), \dots, \Phi(x_N)]$ be the data matrix in the feature space. Hence, C can be expressed as $C = \theta\theta^T/N$. Due to the difficulty of obtaining Φ , a Gram kernel matrix K is determined as follows to directly avoid eigen-decomposition C :

$$K(x_i, x_j) = \Phi^T(x_i), \Phi(x_j), \quad (22)$$

where $K = \theta^T\theta$; therefore, the inner product in the feature space (see (20)) can be obtained by introducing a kernel function to the input space.

Let

$$V = \sum_{i=1}^N \alpha_i \Phi(x_i). \quad (23)$$

Simultaneously, (20), (21), and (22) obtain the following equation:

$$\sum_{i=1}^N \sum_{j=1}^N \Phi(x_i) \alpha_j K_{ij} = N\lambda \sum_{i=1}^N \alpha_i \Phi(x_i). \quad (24)$$

The vector of (23) is

$$K\alpha = N\lambda\alpha = \lambda'\alpha. \quad (25)$$

To extract the principal components, the projection of feature vector V_j in the feature space is calculated:

$$(V_j)^T \Phi(x) = \sum_{i=1}^N \alpha_i \Phi^T(x_i) \Phi(x) = \sum_{i=1}^N \alpha_i K(x_i, x). \quad (26)$$

2.4. Probabilistic Neural Network. The probabilistic neural network was proposed by Specht in 1990 [34]. It is a feed-forward neural network that was developed from the radial basis function, and its theoretical basis is the Bayes minimum

risk rule (Bayes decision theory). As one of the radial basis networks, the PNN is suited to pattern classification, it places the Bayes decision analysis (with the Parzen window function) into the framework of a neural network, and the Bayes classification is produced by combining the Bayes decision and nonparametric estimation of probability density function. It can be described in the following manner. Assuming that there are two fault modes (θ_A and θ_B) for a fault feature sample $X = (x_1, x_2, x_3, \dots, x_n)$,

$$\begin{aligned} \text{if } h_A l_A f_A(X) > h_B l_B f_B(X), \quad \text{then } X \in \theta_A, \\ \text{if } h_A l_A f_A(X) < h_B l_B f_A(X), \quad \text{then } X \in \theta_B, \end{aligned} \quad (27)$$

where h_A and h_B denote the prior probability of fault modes θ_A and θ_B , generally, $h_A = N_A/N$, $h_B = N_B/N$, and N_A and N_B are the number of training samples of θ_A and θ_B , respectively, and N is the total number of training samples. l_A is the cost factor used to classify the feature X sample belonging to θ_A into mode θ_B (falsely), l_B is the cost factor used to classify the feature sample X belonging to θ_B into mode θ_A (falsely). f_A and f_B are the probability density functions of the fault modes θ_A and θ_B , respectively.

Figure 6 shows the PNN structure that demonstrates that the input mode X is divided into 2 types. As shown in Figure 6, the PNN is a feed-forward neural network with a 4-layer structure: the input layer, pattern layer, summation layer, and output layer. The input layer transmits input samples to each node of the pattern layer. The node of the pattern layer calculates the weighted sum of the data passed by the input node, following the operation of a nonlinear operator, which transmits the results to the summation layer. The nonlinear operator is

$$g(z_j) = \exp \left[\frac{(z_j - 1)}{\sigma^2} \right]. \quad (28)$$

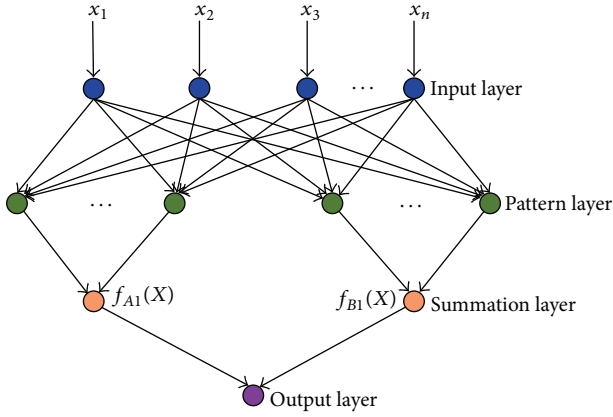


FIGURE 6: The basic PNN structure.

Assuming that X and W are standardized into unit lengths, (1) is equivalent to

$$g(z_j) = \exp \left[-\frac{(W_j - X)^t (W_j - X)}{2\sigma^2} \right], \quad (29)$$

where W is the weight vector.

The summation layer sums up the input from the pattern layer and then obtains the estimated probability densities. The classification result selected by the output layer is the maximum output of the summation layer.

The PNN is equal to the Bayes pattern classification method, which utilizes the Gauss kernel multivariate probability density function. The density function can be estimated as follows:

$$F_A(X) = \frac{1}{(2\pi)^{p/2}} \frac{1}{\delta^p m} \cdot \sum \exp \left[-\frac{(X - X_{Aj})^T (X - X_{Aj})}{2\delta^2} \right], \quad (30)$$

where X is input sample vector, s is the number of sample vectors' variables, X_{Aj} (the weight in the PNN) is j th training vector of fault mode A , and m is the number of training samples belonging to mode A ; δ is the smoothing parameter.

3. Method for Fault Diagnosis of Rolling Bearing under Variable Working Conditions

Inspired by SIFT, this study proposes a novel fault diagnosis for rolling bearings under variable conditions. The diagnostic procedure is shown in Figure 7.

The diagnostic process generally consists of four steps. First, the vibration signals in different fault modes under different conditions are transformed into RPs that are regarded as the objects of SIFT. To more accurately reconstruct the phase space, the delay time and embedded dimensionality are calculated using the mutual information algorithm and CAO algorithm is used to calculate the RPs. Second, the SIFT

extracts the invariable features of the RPs (as described in Section 2), and the constructing scale space, extreme point detection, precise locations of the key points, orientation assignment, and descriptor calculation are determined to achieve the salient invariable features. Third, due to the high-dimensional vector of the extraction features, KPCA is employed by introducing a kernel function to reduce the dimensionality. Finally, the PNN is used as a classifier to diagnose the fault classification using data from one of the conditions to train the neural network and using data from the other conditions to test the proposed method.

4. Results and Discussion

In this section, vibration data of rolling bearings collected from the Case Western Reserve University Bearing Data Center under different working conditions and fault modes were utilized to validate the effectiveness of the proposed method.

4.1. Description of the Experimental Data. The experimental data used to test and verify the proposed method were obtained from the Bearing Data Center of Case Western Reserve University, Cleveland, OH, USA. The experimental setup used a Reliance Electric 2HP motor connected to a dynamometer, which was used as the prime mover to drive a shaft coupled with a bearing housing. Faults (i.e., size 7 mils, 14 mils, 21 mils, and 28 mils) were introduced into the drive-end bearing (6205-2RS JEM SKF) and fan-end (NTN equivalent bearing) of a motor using the electric discharge machining (EDM) method, with the motor speed varied at 1730, 1750, 1772, and 1797 rpm, respectively. These faults were introduced separately at the inner raceway, rolling element (ball), and outer raceway [35]. To quantify the stationary effect of the outer raceway faults, experiments were conducted for the FE and DE bearings, with outer raceway faults located at 3 o'clock, 6 o'clock, and 12 o'clock. An impulsive force was applied to the motor shaft, and the resulting vibration was measured using two accelerometers, one mounted on the motor housing and the other placed at the 12 o'clock position of the outer race of the drive-end bearing. Digital data were collected at 12,000 samples per second, and data were also collected at 48,000 samples per second for the drive-end bearing faults.

In this study, the DE bearing data for the normal, inner race fault, outer race fault, and rolling element fault with the speed varied between the 4 conditions were acquired for the fault pattern classification, and the fault diameters were 21 mils. The fault information (21 mils and outer race fault at 6 o'clock with four speeds), in terms of the test bearings, is listed in Table 1.

4.2. Image Transformation of the Vibration Signals under Different Conditions. In this section, the vibration signals under different conditions are transformed into 2-dimensional images, which facilitate the extraction of invariable features for the fault classification. As previously mentioned, RP can uncover the hidden periodicities in a signal in the recurrence domain, and it is important that the method analyzes the periodic, chaotic, and nonstationary elements of the time

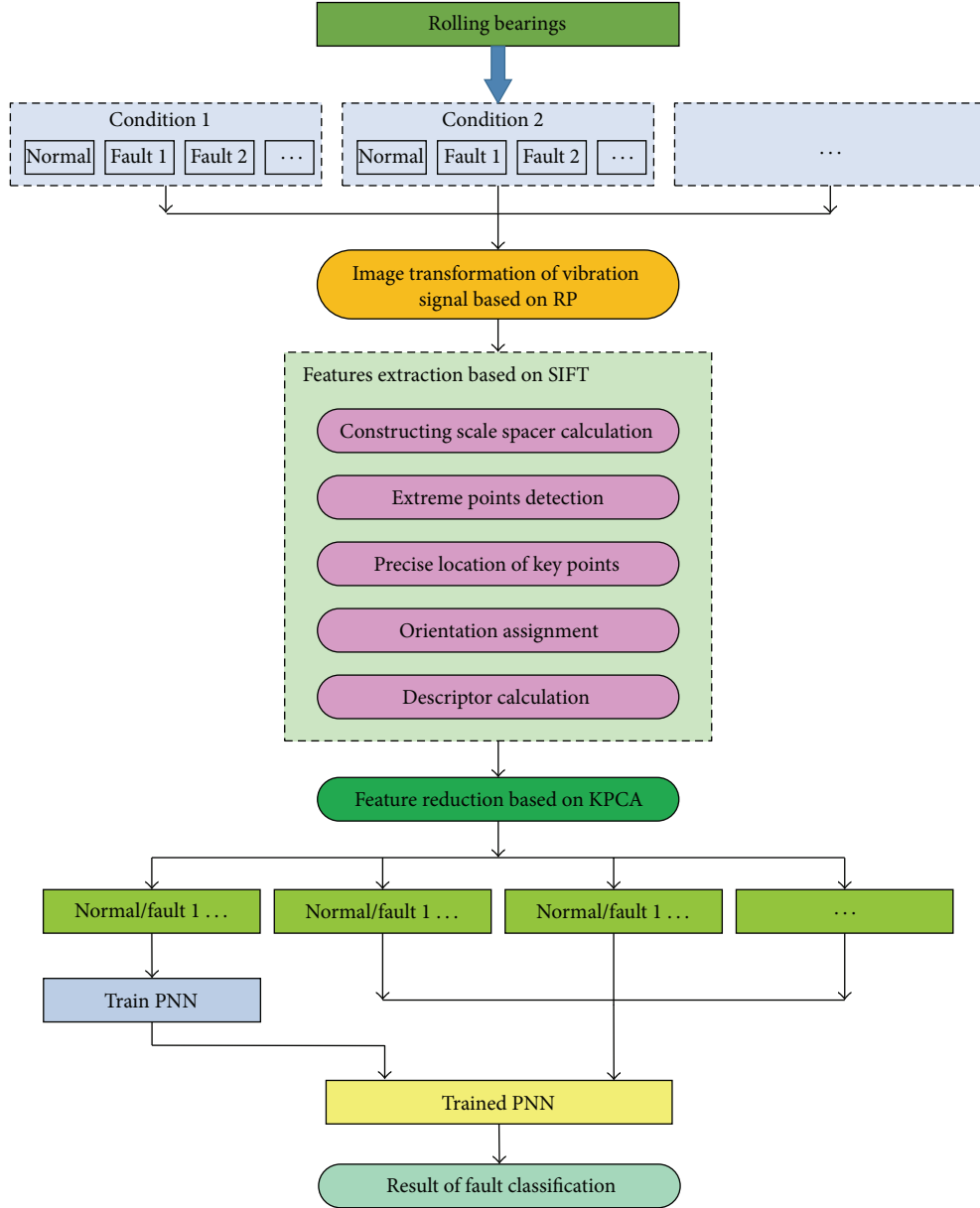


FIGURE 7: A flowchart of the proposed method.

TABLE 1: Data declaration.

Fault diameter (mils)	Motor speed (rpm)	Inner race	Ball	3 o'clock	Outer race 6 o'clock	12 o'clock
21	1797	213.mat	226.mat		238.mat	
	1772	214.mat	227.mat		239.mat	
	1750	215.mat	228.mat		240.mat	
	1730	217.mat	229.mat		241.mat	

series. Thus, RPs are particularly suitable for the image transformation of vibration signals without loss of signal information.

To accurately confirm the suitable embedding dimension, m and delay time τ of each signal for the phase-space

reconstruction, mutual information algorithm, and CAO algorithm were used. The parameters m and τ for each condition are shown in Table 2. Unfortunately, to guarantee the calculation speed, for the data segments of each vibration signal, only 1,000 points were chosen to transform into an RP

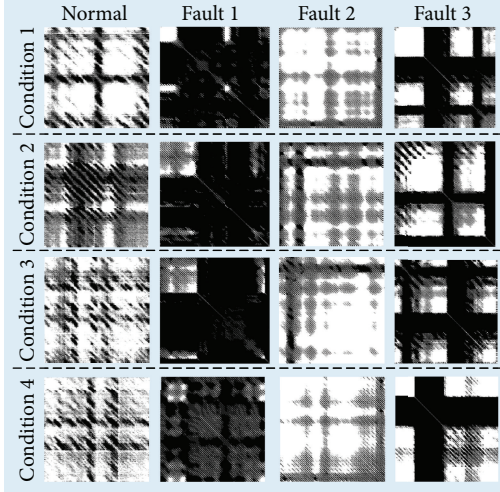


FIGURE 8: RPs transformed by fault mode vibration data under different conditions.

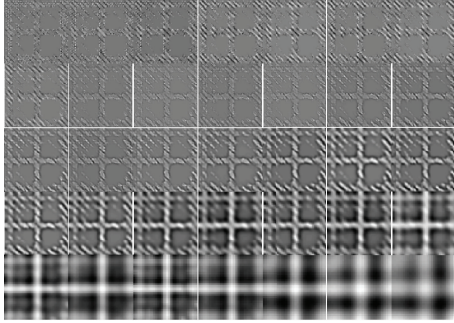


FIGURE 9: The normal DOG scale space under condition 1.

(with dimensionalities of $N \times N$) when reconstructing the phase-space due to calculating the ergodic Euclidean norms of x_i and x_j . This experiment selected a 20-set data segment for each signal. Figure 8 shows the RPs in each fault mode under 4 different conditions that were randomly selected in a 20-set data segment. From the results, we see that the size of the RPs in the different modes under different conditions reveals slight differences.

In Figure 8, the row represents the condition changes, and the column represents the different fault modes (i.e., the first column shows the RPs in the normal state, the second column shows the RPs in the inner fault state, the third column shows the RPs in the element fault state, and the fourth column shows the RPs in the outer fault state). In Figure 8, we can see that the RPs of the different fault modes under different conditions have different structural characteristics, while the same fault modes under different conditions are notably similar. Affected by the condition changes, the RPs under the different conditions show the translation variation, scale variation, and combination of these changes.

4.3. Feature Extraction Based SIFT and the Dimensionality Reduction. In this section, the invariable features in each fault mode under the different conditions are extracted from

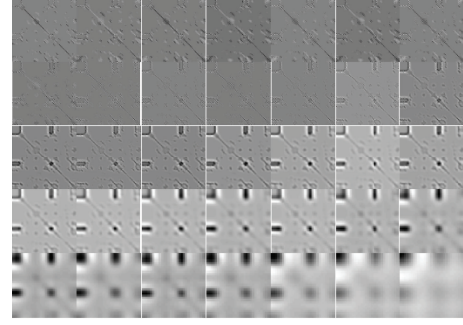


FIGURE 10: The DOG scale space of the inner race fault under condition 1.

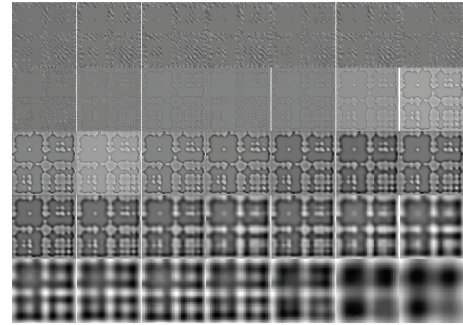


FIGURE 11: The DOG scale space of the element fault under condition 1.

the transformed RPs, based on SIFT. Using SIFT, the scales, orientation, and locations of the key points are calculated. The scale information is obtained by establishing the difference of the Gaussian pyramid, which has 7 octaves. Each octave has 5 strata, and the scale factor σ is utilized to make the image fuzzy between the different strata. Because of the length of the paper, the DOG scale space of four different fault modes (only under condition 1) are shown in Figures 9–12; the locations of the key points are calculated by locating the extreme value points and through further interpolation to determine the exact extreme points on a continuous space. The detected key points of the RPs are shown in Figure 13, and the orientations of the key points are obtained by calculating the gradient orientation of the neighborhood pixels of the key points. The orientation parameters are specified for each key point through the gradient histogram statistics. After performing the above steps, the descriptor of the key point is established through a 128-dimensional vector.

Due to the essential features of RPs hidden in the high-dimensional space, which makes the calculation difficult, the aforementioned KPCA method was used to reduce dimensionality. First, the input space is mapped onto a high-dimensional feature space using the kernel function, and then the PCA is used to reduce the dimensionality. However, the features of the low-dimensional space are also too large and complex to be taken as feature vectors. To solve this problem and to improve the robustness of the feature vectors, singular value decomposition (SVD) was utilized in this

TABLE 2: The experiment parameters of each fault mode under different conditions.

Conditions	Parameters	Normal	Inner race fault	Element fault	Outer race fault
Condition 1	m	15	12	12	12
	τ	4	5	5	5
Condition 2	m	15	12	13	12
	τ	4	5	5	5
Condition 3	m	15	20	13	12
	τ	4	2	5	5
Condition 4	m	16	11	13	12
	τ	4	5	5	5

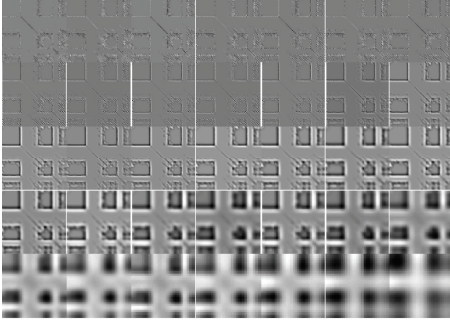


FIGURE 12: The DOG scale space of the outer race fault under condition 1.

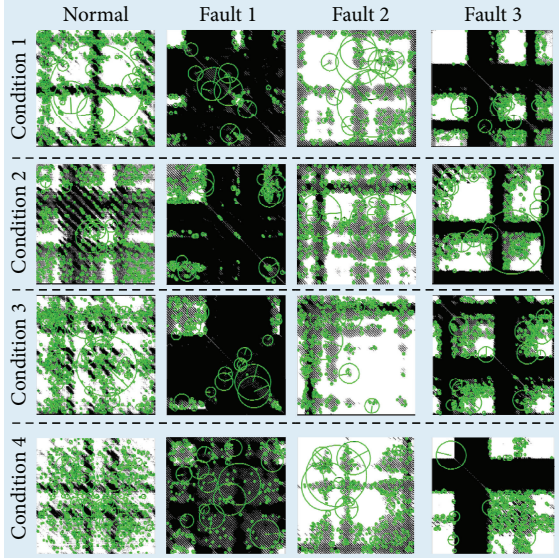


FIGURE 13: The detected interest points in RPs.

paper to compress the scale of the fault feature vectors and to obtain more stable feature vectors [13]. Figure 14 shows the 3-dimensional visual feature points reduced by KPCA and SVD.

4.4. Fault Classification Based on PNN. In this paper, PNN is employed as the classifier to classify the extracted features from vibration signals under different conditions, which are processed by SIFT and KPCA. To verify that the training

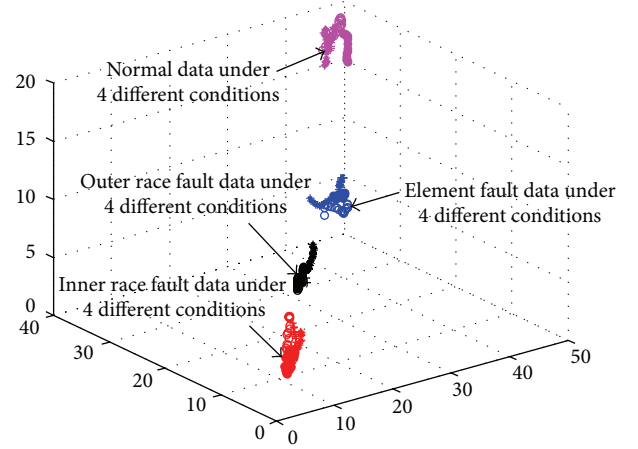


FIGURE 14: The feature scatter diagram in the 3-dimensional space.

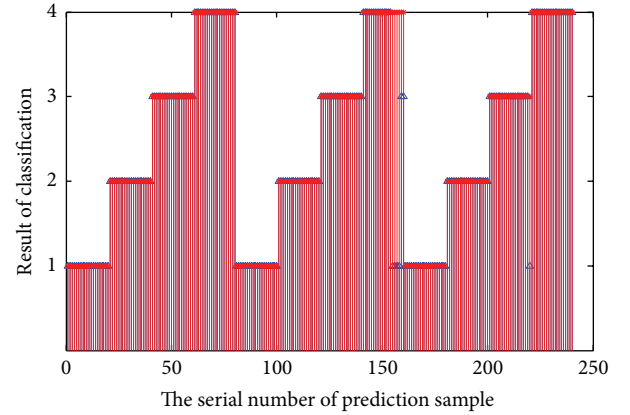
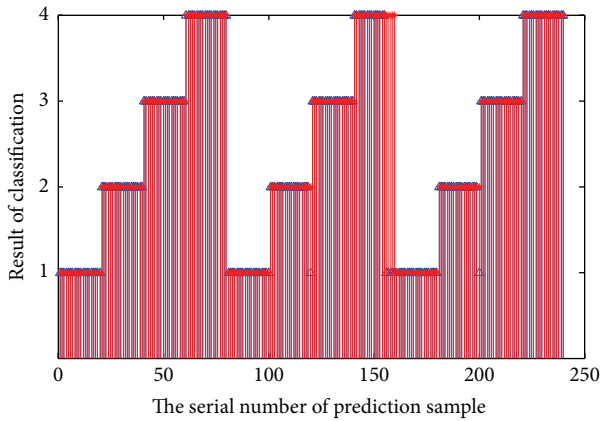


FIGURE 15: The classification result of the first cross-validation group.

data for the different conditions are effective, cross-validation is also necessary; the vibration data collected under each condition are orderly selected as training data, and the data collected under the other three conditions are used as test data, as shown in Table 3.

In each cross-validation, the training data and test data are composed as follows.



The figure is a bar chart titled "Figure 6". The horizontal axis (x-axis) is labeled "The serial number of prediction sample" and has major tick marks at 0, 50, 100, 150, 200, and 250. The vertical axis (y-axis) is labeled "Result of classification" and has major tick marks at 1, 2, 3, and 4. There are approximately 250 red bars in total, representing individual prediction samples. The height of each bar indicates its classification result. The results vary across the samples, with many reaching a value of 4, some reaching 3, 2, or 1. Notably, there are several groups of samples where the classification result is consistently 4, such as around sample numbers 70-80, 140-150, and 220-230.

The results of the PNN classification are shown in Figures 15–18, where Figure 15 is the result of the first group of cross-validation, Figure 16 is the result of the second group of cross-validation, Figure 17 is the result of the third group of cross-validation, and Figure 18 is the result of the fourth group of cross-validation. The red spots are the actual fault mode, and the blue triangles are the result of the PNN classifier. In the vertical axis, 1, 2, 3, and 4 represent the normal, inner race fault, element fault, and outer race fault, respectively. At last, the detailed error samples statistics and classification accuracy of the cross-validation are shown in Tables 4 and 5. From Table 5, it can be seen that the classification accuracy values of the four groups of cross-validation are all higher than 97%, indicating that the proposed method is of high effectiveness.

5. Conclusions

Note: 1, 2, 3, and 4 in the training data conditions and the test data denote 4 different speed conditions, 1797 rpm, 1772 rpm, 1750 rpm, and 1730 rpm, respectively.

Test Data. 20 data groups for each fault mode under the other 3 conditions, a total of 240 groups of data: 1–80 groups of data comprise the first condition, 81–160 groups of data comprise the second condition, and 161–240 groups of data comprise the third condition.

A novel rolling bearing fault diagnosis method under variable conditions, which was originally introduced for image recognition, is described in this paper. First, this method transforms the vibration signals into images, which are expressed in RPs. Then, as mentioned above, through the use of SIFT, the scales, orientation, and locations of the key points are calculated to identify the angular points, peripheral points, and bright points in the dark space of the RPs to extract the invariant features under variable conditions. After creating the key point descriptors, KPCA was employed to reduce the dimensionality of the high-dimensional feature vectors using kernel function to map the feature vector onto a higher-dimensional feature space and then using PCA to reduce the dimensionality. Finally, PNN was utilized as a classifier to execute the fault classification.

Future plans include conducting more experimental and object tests to further test the applicability of the proposed fault diagnosis method. In addition, because of the calculation speed limitation of phase-space reconstruction, new image transformation methods are also needed. To increase the feature extraction speed, the improved SIFT algorithm can be tested.

TABLE 4: The detailed error samples statistics of cross-validation.

The work conditions of training data	The work conditions of test data		Normal	Inner race fault	Element fault	Outer race fault	Mean value/sum
1	2	Number of data	20	20	20	20	80
		Number of wrong samples	0	0	0	0	0
		Error rate	0	0	0	0	0
	3	Number of data	20	20	20	20	80
		Number of wrong samples	0	0	0	6	6
		Error rate	0	0	0	0.3	0.075
	4	Number of data	20	20	20	20	20
		Number of wrong samples	0	0	1	0	1
		Error rate	0	0	0.05	0	0.0125
	1	Number of data	20	20	20	20	80
		Number of wrong samples	0	0	0	0	0
		Error rate	0	0	0	0	0
2	3	Number of data	20	20	20	20	80
		Number of wrong samples	0	1	0	5	6
		Error rate	0	0.05	0	0.25	0.075
	4	Number of data	20	20	20	20	80
		Number of wrong samples	0	1	0	0	1
		Error rate	0	0.05	0	0	0.125
	1	Number of data	20	20	20	20	80
		Number of wrong samples	0	0	1	5	6
		Error rate	0	0	0.05	0.25	0.075
	2	Number of data	20	20	20	20	80
		Number of wrong samples	0	0	0	0	0
		Error rate	0	0	0	0	0
3	4	Number of data	20	20	20	20	80
		Number of wrong samples	0	0	0	0	0
		Error rate	0	0	0	0	0
	1	Number of data	20	20	20	20	80
		Number of wrong samples	0	0	6	0	6
		Error rate	0	0	0.3	0	0.075
	2	Number of data	20	20	20	20	80
		Number of wrong samples	0	0	0	0	0
		Error rate	0	0	0	0	0
	3	Number of data	20	20	20	20	80
		Number of wrong samples	0	0	0	0	0
		Error rate	0	0	0	0	0
4	2	Number of data	20	20	20	20	80
		Number of wrong samples	0	0	0	0	0
		Error rate	0	0	0	0	0
	3	Number of data	20	20	20	20	80
		Number of wrong samples	0	0	0	0	0
		Error rate	0	0	0	0	0

TABLE 5: The result of fault diagnosis for bearing under variable conditions.

Groups of cross validation	1	2	3	4	Mean value/sum
Number of wrong samples	7	7	6	6	26
Number of test data	240	240	240	240	960
Classification accuracy	97.08%	97.08%	97.5%	97.5%	97.29%

Competing Interests

The authors declare that there is not any potential conflict of interests in the research.

Acknowledgments

This study was supported by the Fundamental Research Funds for the Central Universities (Grant no. YWF-16-BJ-J-18) and the National Natural Science Foundation of China (Grant no. 51575021), and the Technology Foundation Program of National Defense (Grant no. Z132013B002).

References

- [1] G. Wang, X. Feng, and C. Liu, "Bearing fault classification based on conditional random field," *Shock and Vibration*, vol. 20, no. 4, pp. 591–600, 2013.
- [2] H. Liu, X. Wang, and C. Lu, "Rolling bearing fault diagnosis based on LCD-TEO and multifractal detrended fluctuation analysis," *Mechanical Systems and Signal Processing*, vol. 60–61, pp. 273–288, 2015.
- [3] Y. Lv, R. Yuan, and G. Song, "Multivariate empirical mode decomposition and its application to fault diagnosis of rolling bearing," *Mechanical Systems and Signal Processing*, vol. 81, pp. 219–234, 2016.
- [4] G. Chen, J. Chen, and G. M. Dong, "Chirplet Wigner-Ville distribution for time-frequency representation and its application," *Mechanical Systems and Signal Processing*, vol. 41, no. 1–2, pp. 1–13, 2013.
- [5] Z. Zhang, Y. Wang, and K. Wang, "Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network," *Journal of Intelligent Manufacturing*, vol. 24, no. 6, pp. 1213–1227, 2013.
- [6] L.-S. Law, J. H. Kim, W. Y. H. Liew, and S.-K. Lee, "An approach based on wavelet packet decomposition and Hilbert-Huang transform (WPD-HHT) for spindle bearings condition monitoring," *Mechanical Systems and Signal Processing*, vol. 33, pp. 197–211, 2012.
- [7] F. Wu and L. Qu, "Diagnosis of subharmonic faults of large rotating machinery based on EMD," *Mechanical Systems and Signal Processing*, vol. 23, no. 2, pp. 467–475, 2009.
- [8] T. Han, D. Jiang, and N. Wang, "The fault feature extraction of rolling bearing based on EMD and difference spectrum of singular value," *Shock and Vibration*, vol. 2016, Article ID 5957179, 14 pages, 2016.
- [9] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 35, no. 1–2, pp. 108–126, 2013.
- [10] Y. Zhang and R. B. Randall, "Rolling element bearing fault diagnosis based on the combination of genetic algorithms and fast kurtogram," *Mechanical Systems and Signal Processing*, vol. 23, no. 5, pp. 1509–1517, 2009.
- [11] L. Zhang, G. Xiong, H. Liu, H. Zou, and W. Guo, "Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference," *Expert Systems with Applications*, vol. 37, no. 8, pp. 6077–6085, 2010.
- [12] S. Zhang, S. Lu, Q. He, and F. Kong, "Time-varying singular value decomposition for periodic transient identification in bearing fault diagnosis," *Journal of Sound and Vibration*, vol. 379, pp. 213–231, 2016.
- [13] Y. Tian, J. Ma, C. Lu, and Z. Wang, "Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine," *Mechanism and Machine Theory*, vol. 90, pp. 175–186, 2015.
- [14] A. B. Ming, W. Zhang, Z. Qin, and F. Chu, "Fault feature extraction and enhancement of rolling element bearing in varying speed condition," *Mechanical Systems and Signal Processing*, vol. 76–77, pp. 367–379, 2016.
- [15] M. C. W. Potter, "Tracking and resampling method and apparatus for monitoring the performance of rotating machines," patents, 1990.
- [16] K. R. Fyfe and E. D. S. Munck, "Analysis of computed order tracking," *Mechanical Systems and Signal Processing*, vol. 11, no. 2, pp. 187–202, 1997.
- [17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [18] G. A. Montazer and D. Giveki, "Content based image retrieval system using clustered scale invariant feature transforms," *Optik*, vol. 126, no. 18, pp. 1695–1699, 2015.
- [19] Q. Li, G. Wang, J. Liu, and S. Chen, "Robust scale-invariant feature matching for remote sensing image registration," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 2, pp. 287–291, 2009.
- [20] H. Soyel and H. Demirel, "Localized discriminative scale invariant feature transform based facial expression recognition," *Computers and Electrical Engineering*, vol. 38, no. 5, pp. 1299–1309, 2012.
- [21] L. Ghoulami, A. Draa, and S. Chikhi, "An ear biometric system based on artificial bees and the scale invariant feature transform," *Expert Systems with Applications*, vol. 57, pp. 49–61, 2016.
- [22] M. Olgun, A. O. Onarcan, K. Özkan et al., "Wheat grain classification by using dense SIFT features with SVM classifier," *Computers and Electronics in Agriculture*, vol. 122, pp. 185–190, 2016.
- [23] C. K. Yoo and I.-B. Lee, "Nonlinear multivariate filtering and bioprocess monitoring for supervising nonlinear biological processes," *Process Biochemistry*, vol. 41, no. 8, pp. 1854–1863, 2006.
- [24] J. Yan, Y. Wang, G. Ouyang, T. Yu, and X. Li, "Using max entropy ratio of recurrence plot to measure electrocorticogram changes in epilepsy patients," *Physica A: Statistical Mechanics and Its Applications*, vol. 443, pp. 109–116, 2016.

- [25] N. Marwan, J. Kurths, and P. Saparin, "Generalised recurrence plot analysis for spatial data," *Physics Letters A*, vol. 360, no. 4-5, pp. 545–551, 2007.
- [26] D. Yang, W.-X. Ren, Y.-D. Hu, and D. Li, "Selection of optimal threshold to construct recurrence plot for structural operational vibration measurements," *Journal of Sound and Vibration*, vol. 349, pp. 361–374, 2015.
- [27] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, no. 2, pp. 1134–1140, 1986.
- [28] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena*, vol. 110, no. 1-2, pp. 43–50, 1997.
- [29] I. Nurhaida, A. Noviyanto, R. Manurung, and A. M. Arymurthy, "Automatic Indonesian's batik pattern recognition using SIFT approach," *Procedia Computer Science*, vol. 59, pp. 567–576, 2015.
- [30] L. Lenc and P. Král, "Automatic face recognition system based on the SIFT features," *Computers and Electrical Engineering*, vol. 46, pp. 256–272, 2013.
- [31] T. Lindeberg, "Scale-space for discrete signals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 3, pp. 234–254, 1990.
- [32] R. Duits, L. Florack, J. de Graaf, and B. ter Haar Romeny, "On the axioms of scale space theory," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 3, pp. 267–298, 2004.
- [33] C.-Y. Cheng, C.-C. Hsu, and M.-C. Chen, "Adaptive kernel principal component analysis (KPCA) for monitoring small disturbances of nonlinear processes," *Industrial & Engineering Chemistry Research*, vol. 49, no. 5, pp. 2254–2262, 2010.
- [34] D. F. Specht, "Applications of probabilistic neural networks," *Neural Networks*, vol. 3, no. 1, pp. 109–118, 1990.
- [35] J. Yu, "Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models," *Mechanical Systems and Signal Processing*, vol. 25, no. 7, pp. 2573–2588, 2011.

