Fault Diagnosis for Rotating Machinery: A Method based on Image Processing

Chen Lu[1,2], Yang Wang[1,2], Minvydas Ragulskis[3], Yujie Cheng[1,2]*

1 School of Reliability and Systems Engineering, Beihang University, Xueyuan Road No.37, Haidian District, Beijing, China, **2** Science & Technology on Reliability and Environmental Engineering Laboratory, Xueyuan Road No.37, Haidian District, Beijing, China, **3** Research Group for Mathematical and Numerical Analysis of Dynamical Systems, Kaunas University of Technology, Studentu 50-146, Kaunas LT-51368, Lithuania

* chengyujie@buaa.edu.cn

Abstract

Rotating machinery is one of the most typical types of mechanical equipment and plays a significant role in industrial applications. Condition monitoring and fault diagnosis of rotating machinery has gained wide attention for its significance in preventing catastrophic accident and guaranteeing sufficient maintenance. With the development of science and technology, fault diagnosis methods based on multi-disciplines are becoming the focus in the field of fault diagnosis of rotating machinery. This paper presents a multi-discipline method based on image-processing for fault diagnosis of rotating machinery. Different from traditional analysis method in one-dimensional space, this study employs computing method in the field of image processing to realize automatic feature extraction and fault diagnosis in a two-dimensional space. The proposed method mainly includes the following steps. First, the vibration signal is transformed into a bi-spectrum contour map utilizing bi-spectrum technology, which provides a basis for the following image-based feature extraction. Then, an emerging approach in the field of image processing for feature extraction, speeded-up robust features, is employed to automatically exact fault features from the transformed bi-spectrum contour map and finally form a high-dimensional feature vector. To reduce the dimensionality of the feature vector, thus highlighting main fault features and reducing subsequent computing resources, t-Distributed Stochastic Neighbor Embedding is adopt to reduce the dimensionality of the feature vector. At last, probabilistic neural network is introduced for fault identification. Two typical rotating machinery, axial piston hydraulic pump and self-priming centrifugal pumps, are selected to demonstrate the effectiveness of the proposed method. Results show that the proposed method based on image-processing achieves a high accuracy, thus providing a highly effective means to fault diagnosis for rotating machinery.

Introduction

Rotating machinery, an important and common component in most critical systems, is vital to the reliable operation of the entire system. An unexpected failure of a rotating machinery may cause sudden breakdown of the whole machinery equipment, bringing about enormous financial losses or even personnel casualties [1, 2]. Therefore, condition monitoring and fault diagnosis of rotating machinery is of utmost significance for security and reliability in industrial manufacturing [3, 4].

Rotating machinery under an abnormal state is accompanied with changes in vibration [5]. Thus, the vibration signal analysis method has been widely applied to fault diagnostics of rotating machinery. Feature extraction is a vital stage that determines diagnosis accuracy [6]. Considerable research has been performed on feature extraction based on signal decomposition[7], among which self-adaptive decomposition methods have been widely spread. Empirical mode decomposition (EMD) [7–10] and local mean decomposition (LMD) [11, 12] are two typical self-adaptive decomposition methods. However, problems such as envelope error, mode mixing and end effect exist in EMD and LMD, which directly influence the diagnostic accuracy [12, 13]. Furthermore, EMD and LMD are usually used combined with time domain feature extraction, such as complexity measures [14], skewness, and kurtosis [15]. The selection of the time domain features is usually performed manually by field experts, whose knowledge and skill make difference in the diagnosis accuracy. The above disadvantages limit the application of these methods in practice. Therefore, it is very necessary to put forward a fault diagnosis method which can realize automatic feature extraction with a high accuracy and strong robustness.

Nowadays, with the rapid development of science and technology, fault diagnosis methods based on multi-disciplines have attracted more and more attention of researchers in the field of fault diagnosis. Many interdisciplinary methods have been proposed for fault diagnosis. In reference [16], genetic algorithm (natural evolution theory) combined with support vector machine (machine learning) is applied to bearings for fault diagnosis. In reference [17], ant colony algorithm (bionics) is employed to realize fault diagnosis for rotating machinery. However, few studies have been reported to realize fault diagnosis for rotating machinery employing calculation methods in the field of image processing. Essentially, fault diagnosis for rotating machinery is a process of fault mode recognition, which is extremely similar to the process of image recognition. Both fault classification and image recognition belong to the category of pattern recognition. Therefore, it is of high feasibility to introduce the calculation methods in the field of image processing to fault diagnosis for rotating machinery.

In this paper, a novel approach based on image processing is presented to realize automatic feature extraction for rotating machinery with a high accuracy and strong robustness and finally realize fault classification.

To introduce the computing method of image processing to the field of fault diagnosis, first it is necessary to convert the vibration signal into an image. The higher order spectrum method is a typical signal processing method [18–21]; it has been developed rapidly over the past few years and offers outstanding performance in processing non-Gaussian, non-linear, non-minimum phase, and non-stationary signals, colored Gaussian noise, and blind signals [22]. High order statistics (HOS) have been proved to provide abundant diagnostic information [23–26]. Bi-spectrum is a subset of higher order spectrum and is defined in terms of the third-order cumulate. It retains phase information of the signal with a high resistance to noise [9, 27]. Therefore, the bi-spectrum contour map is employed to realize the transformation from vibration signals to images.

Owing to the development of the image automatic feature extraction technique in recent decades, the scale invariant feature transform (SIFT) method is recognized as the most

appealing descriptor for practical uses and for matching features with good robustness and high accuracy [28–30]. However, some obvious shortcomings associate with SIFT, such as high resource consumption, high time complexity, and large computational time requirements [31]. To address these disadvantages, researchers have implemented various improvements to SIFT [32–36]. In 2008, Bay et al. proposed the modified SIFT method, speeded up robust features (SURF) [36], which provides faster and better performance in the feature point detection and description scheme. SURF is combined with integral images and Haar wavelets with the characteristics of rotation and scale invariance. As a novel feature extraction algorithm, SURF not only brings an obvious advantage in the area of computational speed, but also performs excellent in repeatability, distinctiveness and robustness equal or close to previous methods. Rotating machinery is susceptible to interference of working conditions and noise, the invariance of SURF can extract more stability fault features. Therefore, SURF is employed to extract fault features from the transformed bi-spectrum contour map.

The SURF algorithm generates a 64-dimensional feature vector for each detected feature point. Considering the high dimensionality may result in feature redundancy and a waste of resources for subsequent calculation, it is necessary to reduce the dimensionality of the SURF feature vector. As an emerging dimensionality reduction technique, t-distributed stochastic neighbor embedding (t-SNE) can maintain the consistency of neighborhood probability distribution between high-dimensional and low-dimensional space, thus avoiding information loss as much as possible [37]. Therefore, t-SNE is used to reduce the dimensionality of the SURF feature vectors. Finally, a classical neural network, probabilistic neural network (PNN) is introduced for identification of fault modes.

Two typical types of rotating machinery, axial piston hydraulic pump and self-priming centrifugal pump, are selected to verify the proposed fault diagnosis method based on image processing. The experimental results demonstrate that the proposed method reaches a high accuracy.

Methodology

The proposed method contains four major step: (1) Image transformation of vibration signal based on bi-spectrum. (2) Feature extraction based on SURF (3) Dimension reduction based on t-SNE (4) Fault diagnosis based on PNN. Fig 1 shows the proposed fault diagnosis scheme.

Bi-spectrum counter map generation

A vibration signal of rotating machinery is a complex signal with non-liner, non-stationary signal with working noise, bi-spectrum is an effective tool to process such signal. Therefore, bi-spectrum technology is applied to transform vibration signal into image (bi-spectrum contour map), which prepare for the application of image processing technology.

Higher-order spectrum are defined by higher-order cumulant. So does bi-spectrum, which is defined by third-order cumulant. For a random signal $x(t)$, the third-order cumulant can be expressed as:

$$C_{3x}(\tau_1, \ \tau_2) = E\{x(t)x(t+\tau_1)x(t+\tau_2)\} \tag{1}$$

By assuming that the third-order cumulant $C_{3x}(\tau_1, \tau_2)$ is convergent, the *third-order* cumulant spectrum is defined as the 2-dimensional Fourier transform of the *third-order* cumulant:

$$B_x(w_1, w_2) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} C_{3x}(\tau_1, \tau_2) e^{-j(w_1\tau_1 + w_2\tau_2)} \tag{2}$$
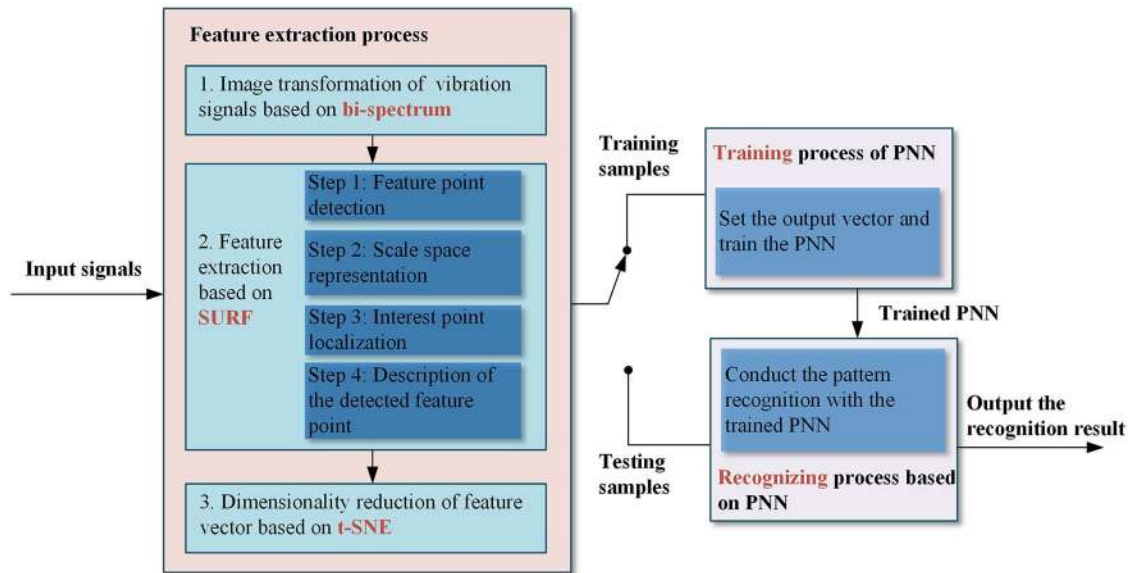
Fig 1. The proposed fault diagnosis scheme based on image processing.

The bi-spectrum of a signal should be estimated under a certain "optimum" standard. In practice, the bi-spectrum can be estimated only according to limited observation data. The estimated method of the bi-spectrum method is the same as the power spectrum estimation method, including the parametric model and non-parametric model. Compared with the indirect method, less computation is required by the direct method [38, 39]. Therefore, the direct estimation method is adopted in this paper. The flow of the algorithm is described as follows:

1. By assuming that the observation data are of finite length, the sampling frequency is f_s. In the bi-spectrum domain, the number of the points are ω_1 and ω_2, and thus the frequency sampling interval is $\Delta_0 = f_s/N_0$. Divide $x(t)$ into k segments, with each segment containing M points; (i.e., $N = KM$), then subtract the mean of each sample.

2. The discrete flourier transform (DFT) is applied for the jth data segment; that is:

$$X^j(\lambda) = \frac{1}{M} \sum_{i=1}^{M} x^j(i) \exp\left(-j\frac{2\pi}{M} i\lambda\right) \tag{3}$$

3. According to the coefficient of DFT, calculate the bi-spectrum estimation of each segment:

$$B_x^j(\lambda_1, \lambda_2) = \frac{1}{\Delta_0^2} \sum_{k_1=-L_1}^{L} \sum_{k_2=-L_1}^{L} X^j(\lambda_1 + k_1) X^j(\lambda_2 + k_2) X^j(\lambda_1 + k_1 + \lambda_2 + k_2) \tag{4}$$

4. According to the result of the bi-spectrum estimation $B_{xx}^j(\lambda_1, \lambda_2)$ of each segment, calculate the mean of $B_{xx}^j(\lambda_1, \lambda_2)$ and then obtain the bi-spectrum estimate of the observation

data $x(i)$:

$$B_x(\omega_1, \omega_2) = \frac{1}{K} \sum_{j=1}^{k} B_x^j(\omega_1, \omega_2) \tag{5}$$

Where $\omega_1 = [2\pi f_s/N_0]\lambda_1$, $\omega_2 = [2\pi f_s/N_0]\lambda_2$.

Feature extraction based on SURF

After the completion of the image conversion, the next step is to use the technology of image processing to extract the fault information contained in the bi-spectrum contour map. As a classical method in the field of image processing, SURF process the advantages of high precision and high speed. Its application is has been found in the field of visual tracking, underwater motion estimation, and other related fields for its novelty. Herein, SURF is employed to extract and describe feature points of the bi-spectrum contour map. The basic procedure is as follows:

Feature point detection. Feature point detection in the SURF algorithm is based on Hessian matrices; the feature point is located according to the local maximum of the Hessian matrix. Consider a certain point (x, y) in row x, column y and image I. The Hessian matrix $H(x, \sigma)$ at x with scale σ Gaussian filter is defined as:

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xx}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \tag{6}$$

where $L_{xx}(x, \sigma)$, $L_{xy}(x, \sigma)$, $L_{yy}(x, \sigma)$ are the second-order partial derivative and two-dimension convolution of point (x, y) of the image I, respectively.

The Gaussian function is applicable in scale-space analysis, but in practical application process es, it needs to be discretized and cropped, as shown in the left column of Fig 2. This results in a loss in repeatability during the process of image rotation around odd multiples of $\pi/4$. However, it is inevitable that it be discretized and cropped. Given Lowe's success in approximating LoG with DoG, Bay improved the approximation for the second-order Gaussian derivative even further with box filters, as shown in right column of Fig 2. There is a clear benefit for processing integral images with a convolution template after approximation. The template consists of a simple rectangle whose computation is independent of the size of the template. Hence, the computational efficiency is greatly accelerated.

In [36], experimental data indicates that the approximation does not degrade the performance of the convolution template. The convolution template in Fig 2 are approximations of a Gaussian function with $\sigma = 1.2$, which is the lowest scale for calculating the blob response maps. We denote the approximation template and convolution by D_{xx}, D_{yy} and D_{xy}, which are used instead of $L_{xx}(x, \sigma)$, $L_{xy}(x, \sigma)$, and $L_{yy}(x, \sigma)$ to obtain the approximate Hessian matrix H. That is:

$$|H| = D_{xy}(X)D_{xy}(X) - [0.9D_{xy}(X)]^2 \tag{7}$$

where w is set as 0.9. According to (1), we calculate and record the response of each point to obtain the response map in scale σ.

Scale space representation. To search the feature points in the image with reducing-enlarging relation, the detection operator should have the ability to search the feature point in the same physical location at different scale. Scale space is usually donated by an image pyramid. The images are repeatedly smoothed with a Gaussian filter and then sub-sampled for sake
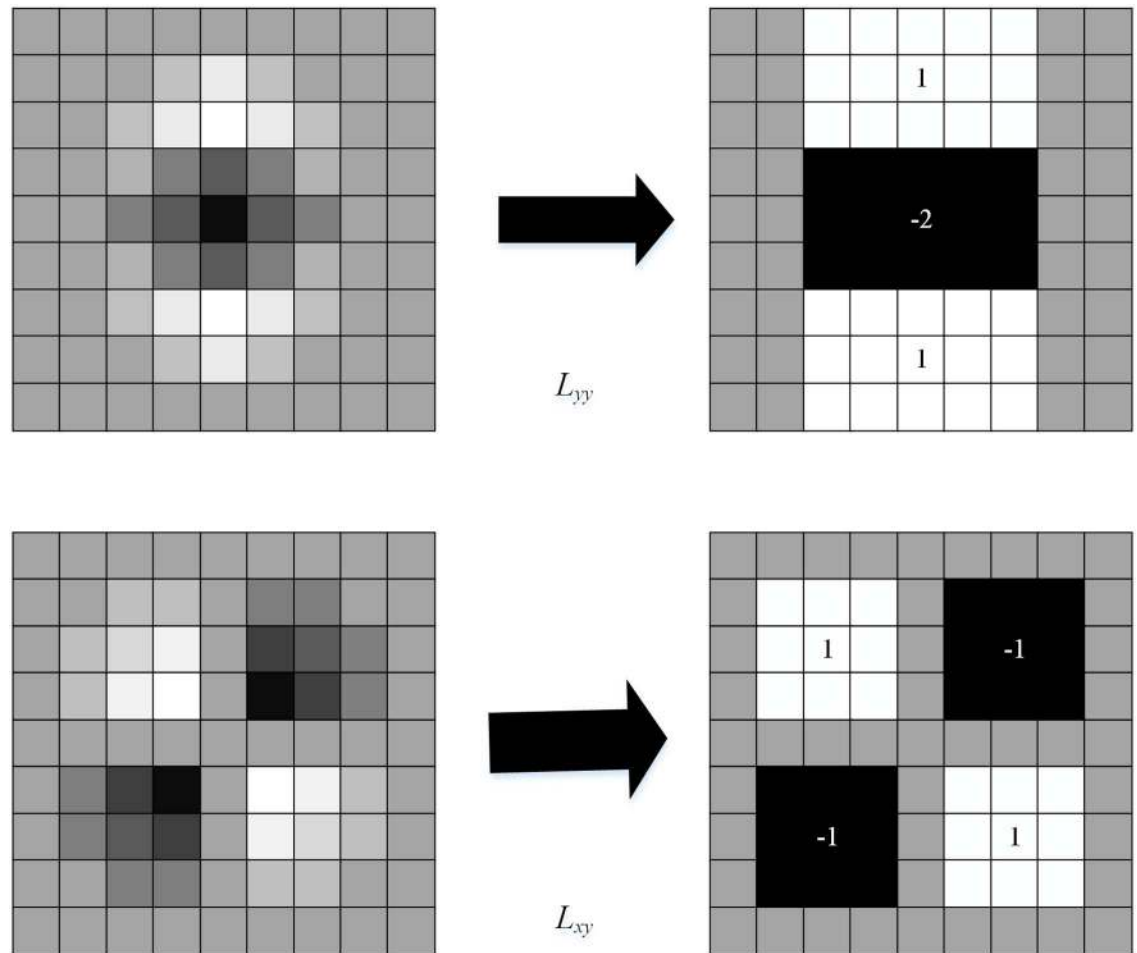
Fig 2. Approximated second order derivatives with box filters.

of achieving a higher level of the pyramid. Lowe subtracts these pyramid layers to obtain the DoG (Difference of Gaussians) images where edges and blobs can be found.

In contrast to this method, SURF processes the original image with box filters of different size. Owning to the application of an integral image, the computation speed between box filters of different size are the same. In the left column of Fig 3, two 9*9 approximate templates are considered to be an initial scale template (approximating Gaussian derivatives with $\sigma = 1.2$). S refers to the scale of approximate template, where $s = \sigma = 1.2$. The initial layer is noted as the convolution between approximate templates of the initial scale of the image. The following layers are obtained by the convolution between gradually increasing s and the initial image. To guarantee the existence of cardinality and its central pixel, adjacent templates always differ by an even number.

Each octave consists of 4 templates. In the first octave, adjacent templates differ by 6 pixels; in the second octave, adjacent templates differ by 8 pixel. In the third octave, adjacent templates differ by 24 pixels, and so on. If the size of a template is N*N, the scale of the template is $S = 1.2*9/N$. We calculate the response at each point and record the response; the response map can be obtained at a different scale. Thus, a 3-dimension scale space can be constructed.

Interest point localization. We seek the extremum of scale image in (x, y, σ) according to the Fast-Hessian matrix. First, non-maxima suppression is performed at the extreme points of
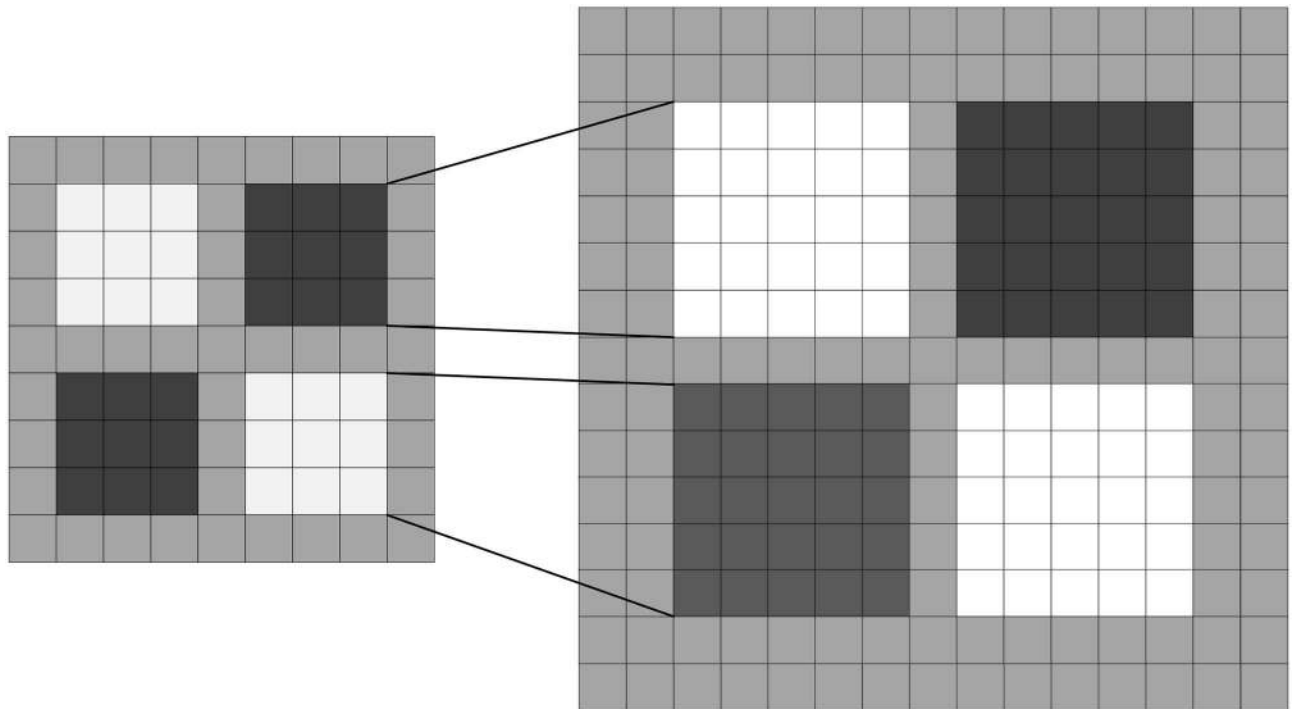
Fig 3. Filters D_{xy} for two successive scale levels (9×9 and 15×15).

the 3×3×3 circular neighborhood. Only extreme points greater or less than 26 compared to the neighborhood value of the scale adjacent up and down and the scale itself can be selected as feature points. To locate candidate feature points with sub-pixel resolution, we interpolate between scale space and image space to obtain a stable feature point and its scale value in its location.

Description of the detected feature point. To enable the descriptor to exhibit rotation invariance, the direction of feature point should be obtained first. Construct a wavelet response whose center is a feature point and whose radius is 6 s (s is the scale of the feature point). Process the image with Haar wavelet of size 4 s; the wavelet response in the direction of the x-axis and y-axis can be obtained. Haar wavelet templates are shown in Fig 4. The template on the left is used to calculate the response in the x-axis direction, whereas the one on the right is used to calculate the response in the y-axis direction. The black parts are denoted by -1, and the white parts are denoted by +1.

The neighborhood of the wavelet response is in the directions of the x-axis and y-axis; assign different Gaussian weight coefficients to these responses. A local direction vector is formed by summing the Harr wavelet in the direction of the x-axis and y-axis within 60°. Traverse the entire circle, then select the direction of the longest vector as the main direction of this feature point. The main direction determination of the feature point is illustrated in Fig 5.

Construct SURF feature vector. Construct a window region whose center is a feature point and whose side length is 20σ. Next, divide the window region into 4×4 sub-regions. A 5×5 sampling point is obtained from the sub-region. Seek the wavelet response in the directions of the x-axis and y-axis of each sampling point, donated as dx and dy, respectively. A Gaussian filter is applied to dx and dy in each sub-region. The filter center is taken as the feature point, and a four-dimensional eigenvector (Σdx, Σdy, $\Sigma|dx|$, $\Sigma|dy|$) is formed by summing dx, dy, $|dx|$, $|dy|$ in each sub-region. The four-dimensional eigenvector makes up 4 dimensions of the

Fig 4. Haar wavelet filters.

descriptor. Each descriptor consists of 4 dimensions. Thus, 4×4×4 = 64 dimensions are obtained and constitute the SURF descriptor.

Dimensionality reduction based on t-SNE

The feature points extracted by SURF are described in terms of a 64-dimensional descriptor. The high dimensionality will result in feature redundancy and a waste of resources for subsequent calculation. Therefore, it is necessary to reduce the dimensionality of the SURF feature vector. Manifold learning methods are widely used dimensionality reduction methods, which can be divided into two types, linear methods and nonlinear methods. Linear manifold learning methods include principal component analysis (PCA), multidimensional scaling (MDS), etc.; nonlinear manifold learning methods include Isometric Feature Mapping (Isomap), locally-linear bedding (LLE), etc[40].

Stochastic neighbor embedding (SNE) is one of the best performed nonlinear manifold learning algorithm whose core idea is to maintain the consistency of neighborhood probability distribution between high-dimensional and low-dimensional space. SNE transfers traditional Euclidean distance-based similarity measurement to conditional probability-based similarity measurement: in high dimensional observation space, the Gaussian distribution is adopted to simulate the similarity relationship between observation samples. Similarity between x_i and x_j
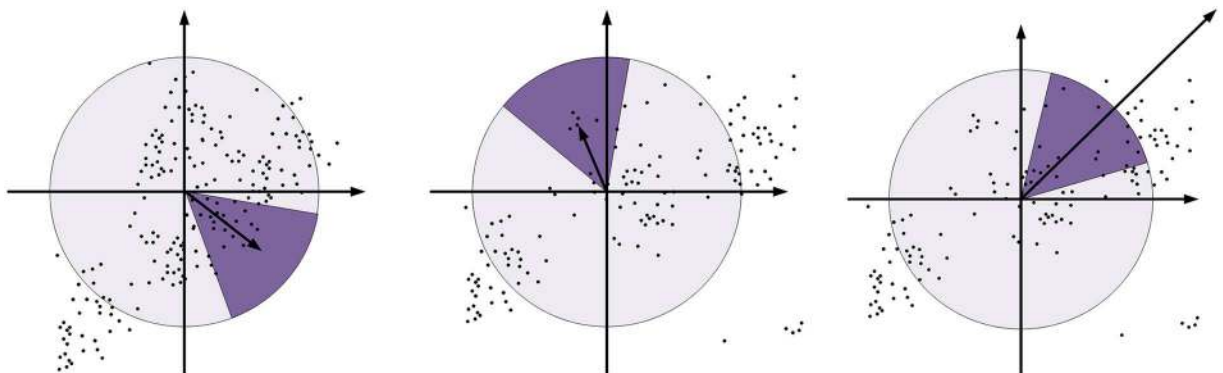

Fig 5. Determining the main direction of the feature point.

$p_{j|i}$ is denoted as follows:

$$p_{j|i} = \frac{\exp(-\|x_j - x_i\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp - \|x_j - x_i\|^2 / 2\sigma_i^2} \tag{8}$$

Where σ_i is the bandwidth of the Gaussian kernel function in the observation sample x_i. $p_{j|i}$ is the probability of x_j chosen by x_i as its neighbor sample. The parameter $p_{j|i}$ obeys a Gaussian distribution in which the variance is σ_i^2 and the mean value is x_i. The probability of x_i and x_j being adjacent to each other is

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n} \tag{9}$$

In low-dimensional space, SNE continues to adopt the Gaussian distribution to measure the similarity between low-distinction samples. However, two obvious shortcomings exists with SNE: (1) the objective function is too complex to optimize, and the gradient is not as concise as desired; (2) the so-called "crowding problem"; that is, when the data are far apart from each other in high-dimensional space, they must be gathered in the process of mapping to low-dimensional space. In response, t-distributed stochastic neighbor embedding (t-SNE) is introduced to alleviate these problems[37].

To solve the first problem, the characteristic of symmetry is adopted to simplify the objective function and optimize the gradient form, which is referred to as symmetric SNE. According to probability theory, the SNE objective function minimizes the sum of distances of the conditional probability distribution $p_{j|i}$ (high-dimension) and $p_{i|j}$ (low-dimension) for corresponding points. It equals the two following joint probability distributions of P (high-dimension) and Q (low-dimension):

$$C = KL(P\|Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{10}$$

After adopting a joint probability distribution instead of a conditional probability distribution, the formula is more concise and understandable.

As for the second question, the t-distribution function is introduce to alleviate the "crowding problem". That is, the t-distribution function is used to measure the similarity of points in low-dimensional space. The joint probability distribution function is as follows:

$$q_{ij} = \frac{(1 + \|y_i - y_j\|_2^2)^{-1}}{\sum_{k \neq 1} (1 + \|y_i - y_j\|_2^2)^{-1}} \tag{11}$$

Here, the t-distribution function (DOF is 1) is applied because of its special advantageous characteristic: $(1 + \|y_i - y_j\|_2^2)^{-1}$ is the reciprocal of the distance of points far from each other in low-dimensional space to $\|y_i - y_k\|^2$. That means that in low-dimensional space, the presentation of the joint probability distribution is insensitive to the distance of points. In addition, in theory, the t-distribution function offers the same performance as the Gaussian function because the t-distribution function can be express as the infinity Gaussian function. Thus, the gradient of t-SNE is:

$$\frac{\delta C}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j)(1 + \|y_i - y_j\|_2^2)^{-1} \tag{12}$$

In conclusion, t-SNE focuses on:

1. The characteristic of non-similarity is associated with points far from each other. The characteristic of similarity is associated with points close to each other. The t-distribution function is introduced to exert an "exclusion" to process the non-similar points;

2. The application of the t-distribution function make it easier to optimize.

The t-SNE approach applies a probability density distribution function to measure the similarity and distribution characteristics. Then, t-SNE accomplishes property preservation by minimizing the K-L distance probability density in high-dimensional and low-dimensional space. The t-distribution function is employed to measure the similarity of points in low-dimensional space, which simplifies the gradient form and improves computational speed. Most important is to better process the non-similarity points and thus alleviate the "crowding problem".

In most research reports, t-SNE exhibits better performance than Sammon mapping, isomap and locally linear embedding.

Fault diagnosis based on PNN

PNN is a feedforward neural network developed from radial basis function; its theoretical basis is the Bayes minimum risk rule (Bayes decision theory). As one of the radial basis function networks, it can be used for pattern recognition[41].

Fault diagnosis based on PNN is a generally acceptable decision-making method in probability statistics. By assuming faults mode A and B, for fault feature sample X to be recognized, according to Bayes minimum risk rule, if $H_A L_A F_A(X) > H_B L_B F_B(X)$, $X \in A$; if $H_A L_A F_A(X) < H_B L_B F_B(X)$, $X \in B$. Here, H_A and H_B denote the prior probability of fault mode A and B. L_A is the cost factor of classifying feature X sample (belonging to A) into mode B falsely. L_B is the cost factor of classifying feature sample X (belonging to B) into mode A falsely. F_A and F_B are the density functions of fault mode A, B. Generally, they can only be obtained by the statistic of the fault feature sample, which cannot be obtained precisely. The density function can be estimated as follows:

$$F_A(X) = \frac{1}{(2\pi)^{p/2} \delta^p} \frac{1}{m_t} \sum \exp\left[-\frac{(X - X_{Ai})^T (X - X_{Ai})}{2\delta^2} \right] \tag{13}$$

where X_{Ai} is the i th training vector of fault mode A; m_t is the total number of training samples of fault mode A; and δ is a smoothing parameter whose value determines the width of the mitriform curve centered on samples.

PNN is a feedforward neural network with a parallel 4-year structure: input layer, pattern layer, summation layer and output layer, as shown in Fig 6. The input layer passes input samples to each node of the pattern layer; the pattern layer weighted summation input vector is passed by input nodes, which then pass it to the summation layer after calculation by the nonlinear operator. The purpose of the summation layer is to sum up the input from the pattern layer and obtain the estimated probability densities. The classification result selected by the output layer is the maximum of the output of the summation layer in terms of the corresponding classification results.

Experiment Results

In order to verify the effectiveness of the proposed method, two typical rotating machinery, self-priming centrifugal pump and axial piston hydraulic pump are selected for case verification.
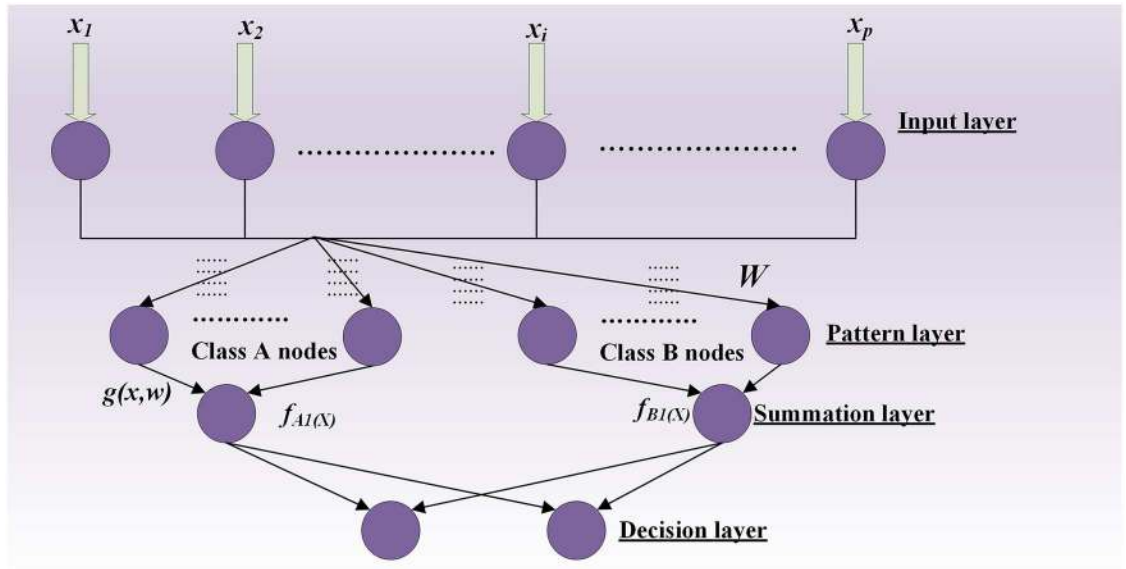
Fig 6. Basic structure of PNN.

Case 1. Fault diagnosis for self-priming centrifugal pump based on SURF and t-SNE

(1) Data preparation. The data of self-priming centrifugal pump are collected on a self-priming centrifugal pump data acquisition system, as shown in Fig 7. The acceleration sensor is installed above the motor housing, and the sensor is fixed on a specific pedestal.

According to the requirement of fault diagnosis for centrifugal pump, a data acquisition experimental scheme is created for the fault insertion test. The test covers primarily fault modes. The experiment items are listed in Table 1.
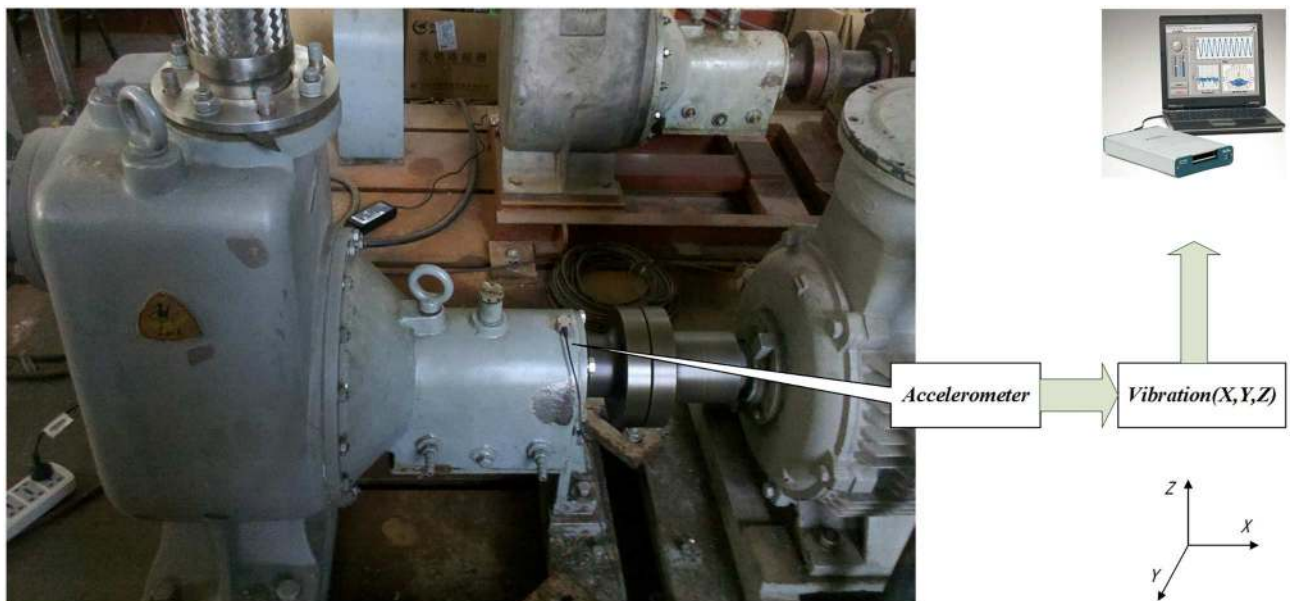
Fig 7. Self-priming centrifugal pump data acquisition system.

Table 1. Experiment items of centrifugal pumps fault injection.

Test object	Failure test	Normal test
rolling bearings	bearing inner race wearing test	Bearing normal operation test
	bearing outer race wearing test	
	bearing rollers wearing test	
Impeller	impeller wearing test	

doi:10.1371/journal.pone.0164111.t001

In the experiment, the rotation speed is 2,900/min. An acceleration sensor is employed when sampling. The sample frequency is 10239Hz.

Vibration data are collected under normal conditions and fault conditions, including bearing roller wearing, inner race wearing, and outer race wearing fault conditions, as well as impeller wearing fault condition. The sampling time is 2s for each set, and one set is collected every 5 seconds.

(2) Feature extraction based on SURF and t-SNE. Figs 8–12 shows the bi-spectrum counter maps of bearing roller wearing, inner race wearing, outer race wearing, and normal conditions, and centrifugal pump impeller wearing fault condition, respectively. For each condition, three groups of data sets are selected to generate bi-spectrum counter maps for comparison.
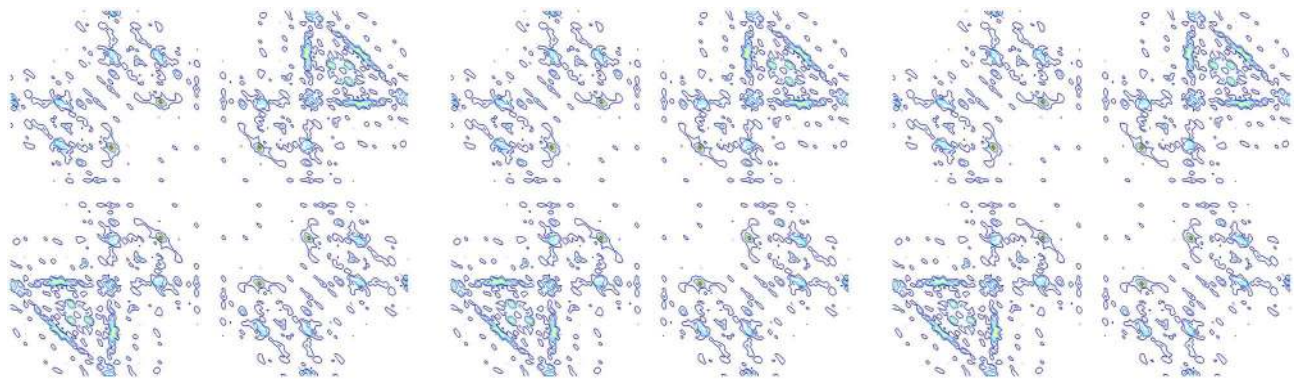

Fig 8. Bi-spectrum counter maps under bearing roller wearing fault condition.
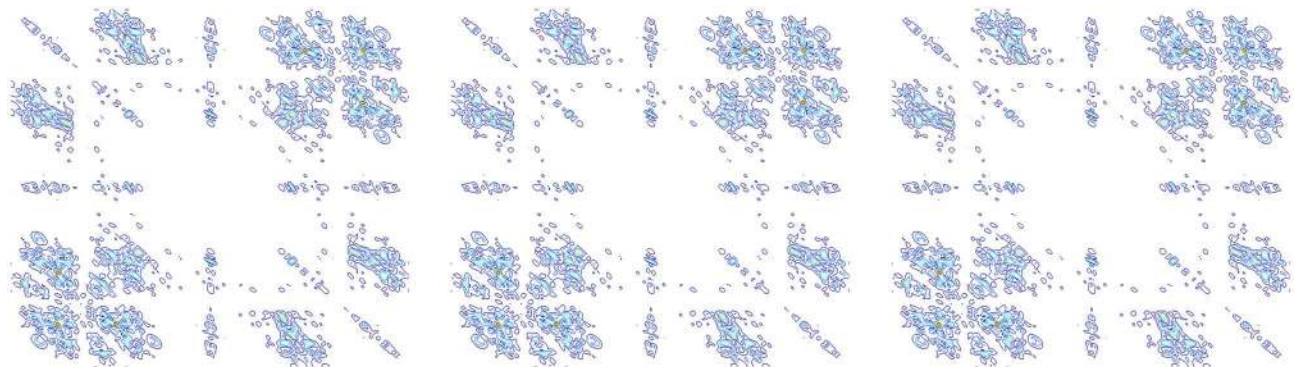
doi:10.1371/journal.pone.0164111.g008

Fig 9. Bi-spectrum counter maps under bearing inner race wearing fault condition.
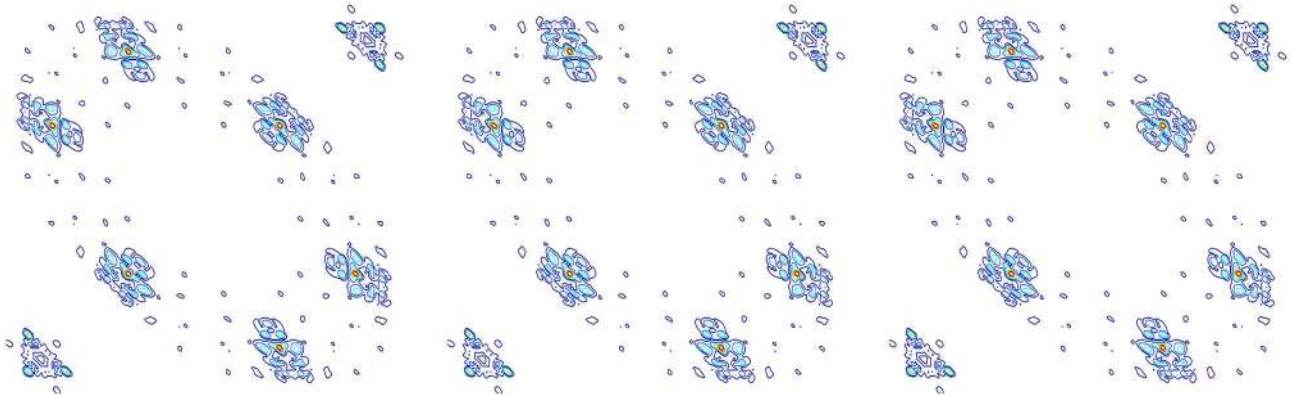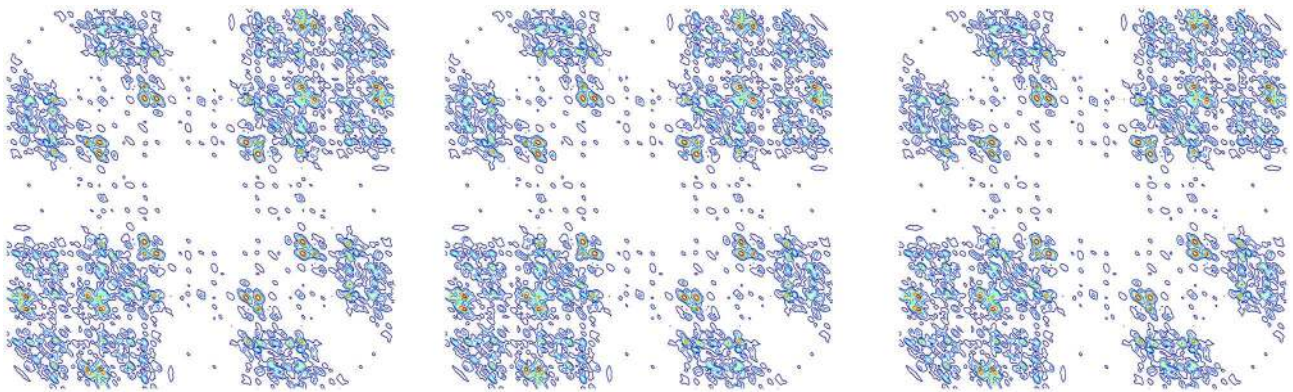
doi:10.1371/journal.pone.0164111.g009

Fig 10. Bi-spectrum counter maps under bearing outer race wearing fault condition.

Fig 11. Bi-spectrum counter maps under normal condition.
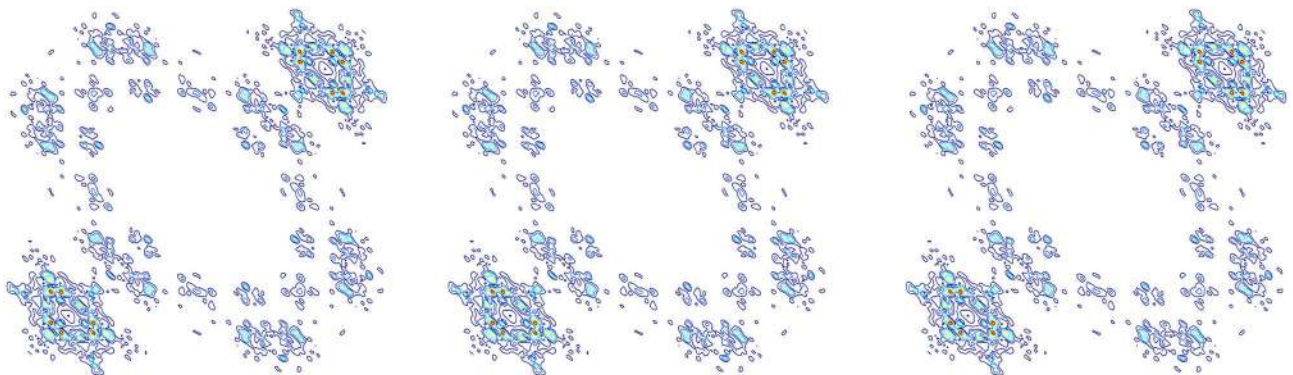
Fig 12. Bi-spectrum counter maps under impeller wearing fault condition.

In the experiment, the SURF descriptor extracts the feature point of the bi-spectrum with 64-dimension form. By using t-SNE, the origin features of the datasets are reduced automatically to 20 dimensions. Fig 13 shows the distribution of the first three features extracted using t-SNE and without using t-SNE, respectively. From the first three features, we can conclude
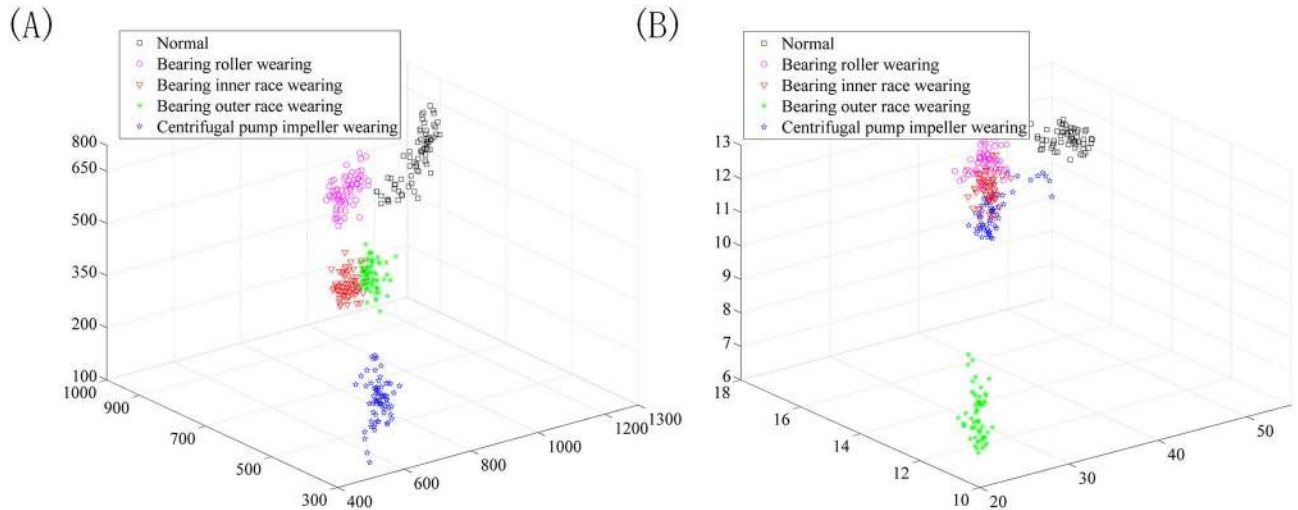
Fig 13. The first three features extracted using t-SNE (a) and without using t-SNE (b).

that the feature information is more concentrated and exhibits a strong ability for separability after dimension reduction by t-SNE, which provides a desirable input for classification.

(3) Fault diagnosis result based on PNN. The input eigenvector of PNN is extracted by SURF and t-SNE. To display the result visually, the classification result is contrasted with the actual result. The accuracy rate is defined as the odds ratio of the correct results and the total results.

Four set cross validation is adopted to verify the accuracy of the proposed method. For each fault mode, 60 sets of data are collected. The length of the data is 1024. Divide the 60 sets into 4 groups; each group is selected as training data in turn, whereas the others are selected as test data. The composition of the data is shown as Table 2.

The fault diagnosis of PNN is shown as follows. Fig 14 show the results of 4 set cross validation. The red circle is the actual fault category, and the blue triangle is the fault diagnosis result. Annotations 1~5 in vertical axis represent the bearing roller wearing, the bearing inner race wearing, the normal condition, the centrifugal pump impeller wearing and the bearing outer race wearing. Table 3 presents a summary of the cross-validation results.

From the diagnosis of PNN we can conclude that all of the accuracy rates exceed 96%. The cross- validation results of the first, second, third, and fourth sets are 99.11%, 100.00%, 96%, and 98.22%, respectively. Average classification accuracy is as high as 98.33%, which verifies the effectiveness of the proposed method.

Table 2. The data composition of the self-priming centrifugal pump for cross validation.

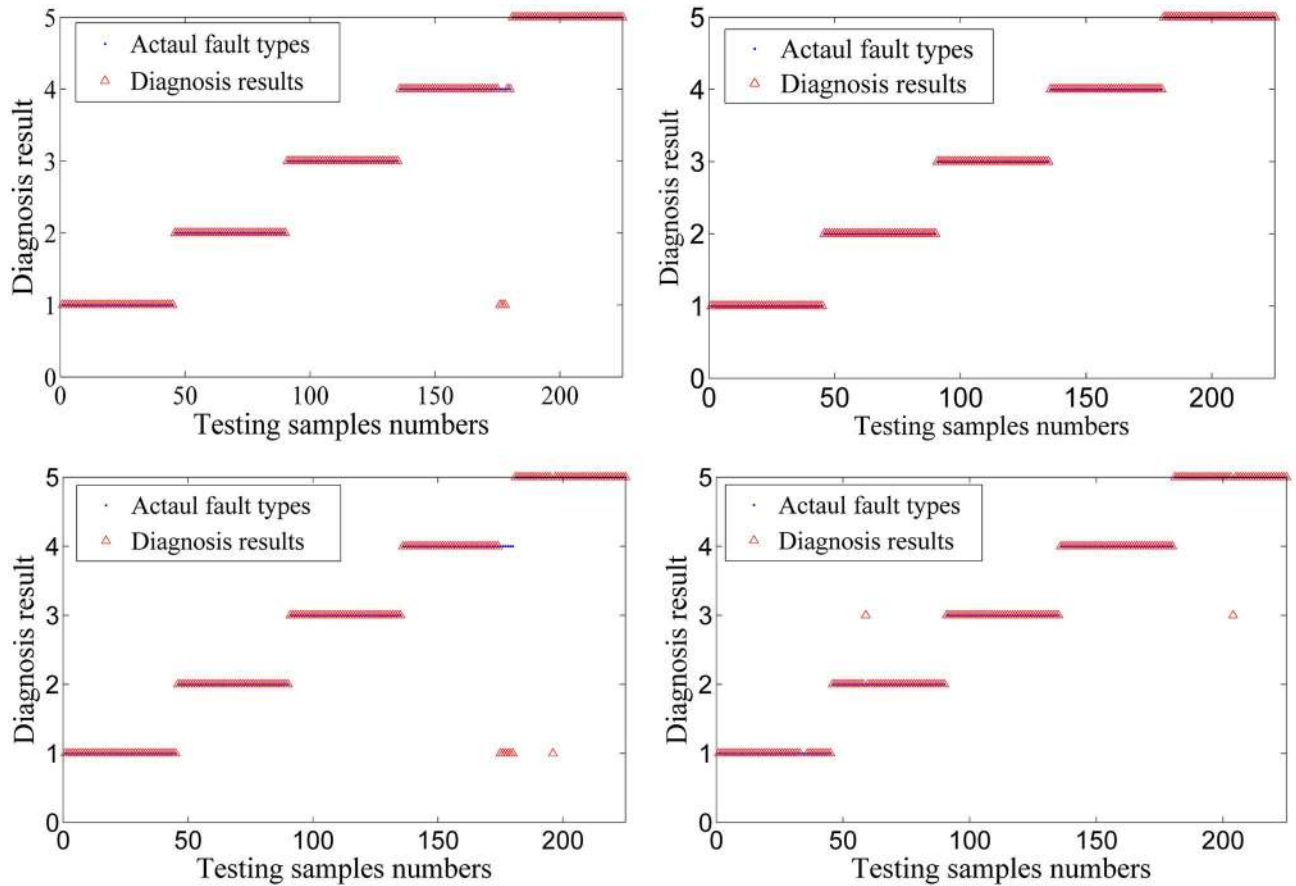|	bearing roller wearing	bearing inner race wearing	normal	centrifugal pump impeller wearing	the bearing outer race wearing	total amount
number of data points	60	60	60	60	60	300
amount of training data	15	15	15	15	15	75
amount of test data	45	45	45	45	45	225

Fig 14. Results of 4 groups of cross validation.

doi:10.1371/journal.pone.0164111.g014

Table 3. The error rate of 4 groups of cross-validation.

Cross-validation No.		bearing roller wearing	bearing inner race wearing	normal	centrifugal pump impeller wearing	bearing outer race wearing	Mean value/total
1	samples in test set	45	45	45	45	45	225
	error samples	0	0	0	2	0	2
	accuracy	1	1	1	0.9556	1	0.9911
2	samples in test set	45	45	45	45	45	135
	error samples	0	0	0	0	0	0
	accuracy	0	0	0	0	0	1
3	samples in test set	45	45	45	45	45	225
	error samples	0	0	0	6	1	9
	accuracy	1	1	1	0.8667	0.9778	0.9600
4	samples in test set	45	45	45	45	45	225
	error samples	2	1	0	0	1	4
	accuracy	0.9556	0.9778	1	1	0.9778	0.9822

doi:10.1371/journal.pone.0164111.t003

Fig 15. Axial piston hydraulic pump system.

doi:10.1371/journal.pone.0164111.g015

Case 2. Fault diagnosis for axial piston hydraulic pump based on SURF and t-SNE

(1) Data preparation. The testing equipment for the axial piston hydraulic pump is shown in Fig 15. In the experiment, the rotation speed is set as 5280r/min and the corresponding spindle frequency is 88Hz. An accelerograph is installed at the end face of the pump. The sample frequency is 1k Hz. The data collected contains fault modes of normal, piston shoes and swashplate wearing, and valve plate wearing. A data set of 500 points for each team is selected for analysis.

(2) Feature extraction based on SURF and t-SNE. Bi-spectrum estimation using the direct method for an axial piston hydraulic pump system is conducted. We segment each sample into M records of 512 points each. The spectrum contour map of normal, piston shoes and swashplate wearing, and valve plate wearing are shown in Figs 16–18, respectively. The bi-spectrum of three data sets are selected for comparison.

From the spectrum contour it can be observed that the same types of fault mode are similar; although there are obvious differences between different fault modes, there are invariant features between the same fault modes. The differences between different fault modes can be used to distinguish fault types.

In the experiment, the SURF descriptor converts the bi-spectrum into a 64-dimension vector. Using t-SNE, the origin features of the datasets are reduced automatically to 20 dimensions. Fig 19 shows the distribution of the first three features extracted using t-SNE and without using t-SNE. From the first three features, we can conclude that the feature information is more concentrated and exhibits a strong ability of separability after dimension reduction by t-SNE, which provides a desirable input for classification.
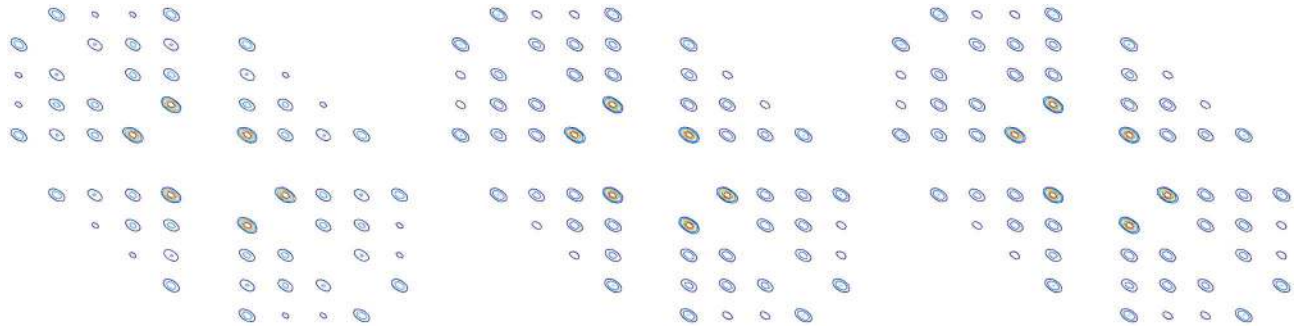
Fig 16. Bi-spectrum counter map of normal.
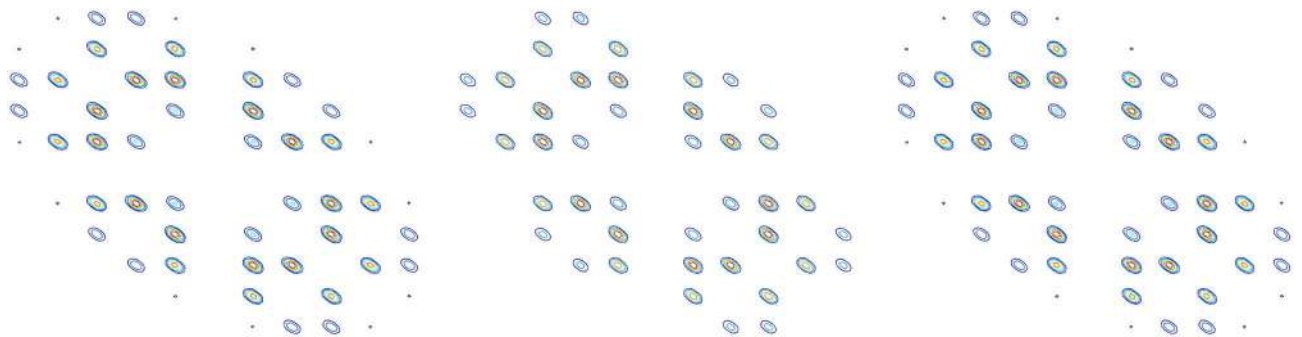
doi:10.1371/journal.pone.0164111.g016

Fig 17. Bi-spectrum counter map of valve plate wearing.

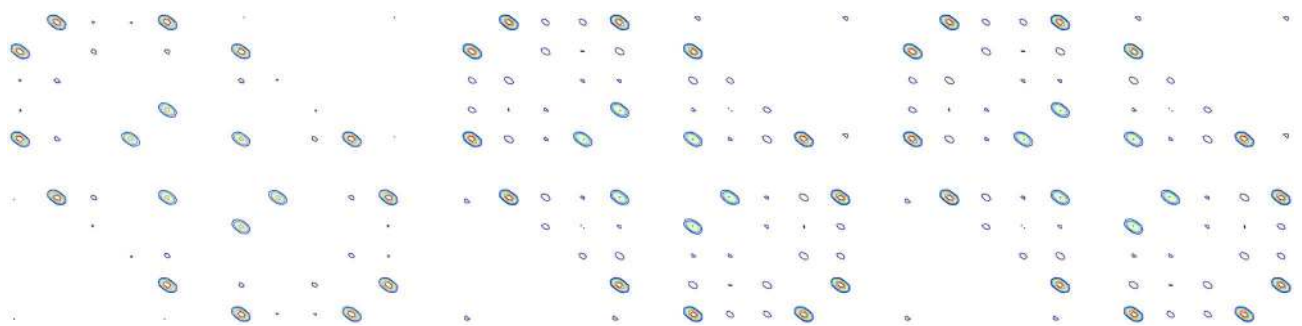doi:10.1371/journal.pone.0164111.g017

Fig 18. Bi-spectrum counter map of piston shoes and swashplate wearing.

doi:10.1371/journal.pone.0164111.g018

(3) Fault diagnosis result based on PNN. The input eigenvector of PNN is extracted by SURF and t-SNE. To display the result visually, the classification result is contrasted with the actual result. The accuracy rate is defined as the odds ratio of the correct results and the total results.

Four set cross validation is adopted to verify the accuracy of the proposed method. For each fault mode, 60 sets of data are collected. The data set length is 500. We divide the 60 sets into 4 groups; each group is selected as training data in turn, whereas the others are selected as test data. The composition of the data is shown as Table 4.

The fault diagnosis of PNN is shown as follows. Fig 20 show the results of 4 sets cross validation. The red circle is the actual fault category, and the blue triangle is the fault diagnosis result.
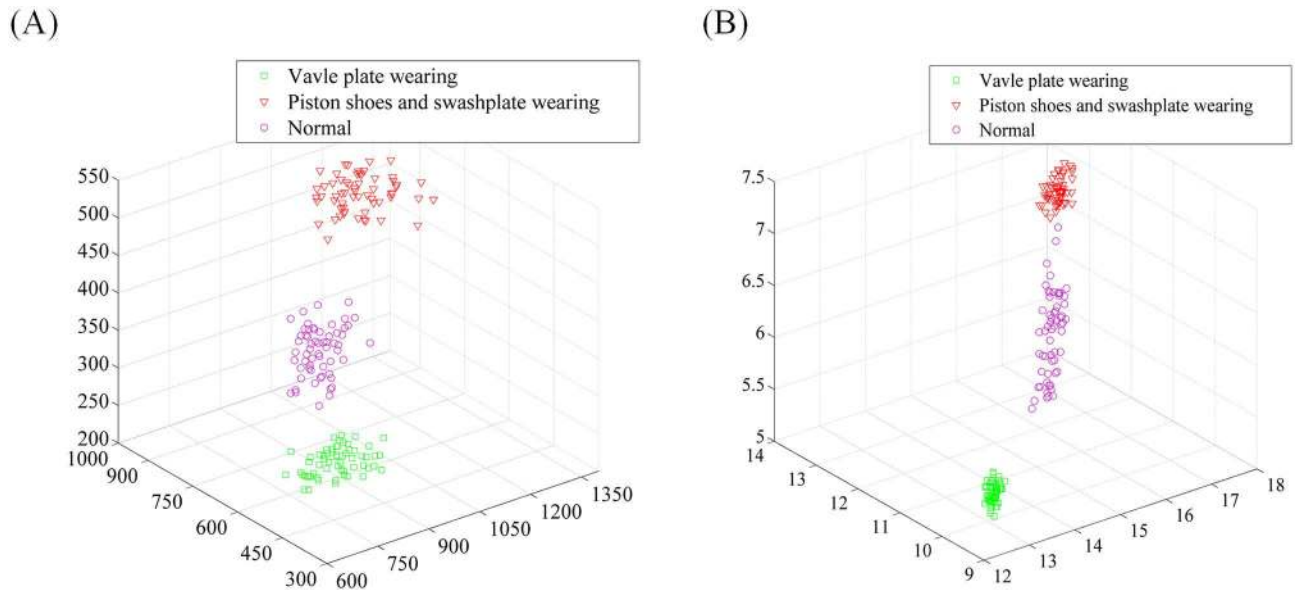
Fig 19. The first three features extracted using t-SNE (a) and without using t-SNE (b).

The numbers 1~3 in the vertical axis represent normal, piston shoes and swashplate wearing and valve plate wearing. Table 5 presents a summary of the cross validation results.

From the results diagnosed by PNN, we can conclude that the accuracy rate of all four set cross validation tests has exceeded to 97%. The cross- validation results of the first, second, third, and fourth sets are 97.04%, 100.00%, 100.00%, and 97.78%, respectively. Average classification accuracy is as high as 98.71%, which verifies the effectiveness of the proposed method.

Conclusions

In this paper, we present a novel rotating machinery diagnosis method based on image recognition that contains four major steps: First, the bi-spectrum is employed to transform the initial vibration signal into an image (bi-spectrum counter map). Next, SURF is first introduced to extract feature points of the image automatically. To reduce the dimension while describing the feature points accurately as much as possible, the manifold dimension reduction method t-SNE is used to map the high-dimensional features to low-dimensional space. Based on the feature vectors extracted by SURF and t-SNE, PNN is applied to enable fault mode recognition.

The proposed image-recognition-based fault diagnostic method for rotating machine first introduces the image interest point extraction method to provide a fault diagnosis and achieve feature extraction of the bi-spectrum automatically. Thus, the method avoids the limitations of relying on a diagnostician for feature extraction. Favorable results for two types of typical rotating machinery demonstrate that our method can improve robustness and generalization ability while maintaining accuracy of classification. Our subsequent work will be focused as follows:

Table 4. The data composition of axial piston hydraulic pump for cross validation.

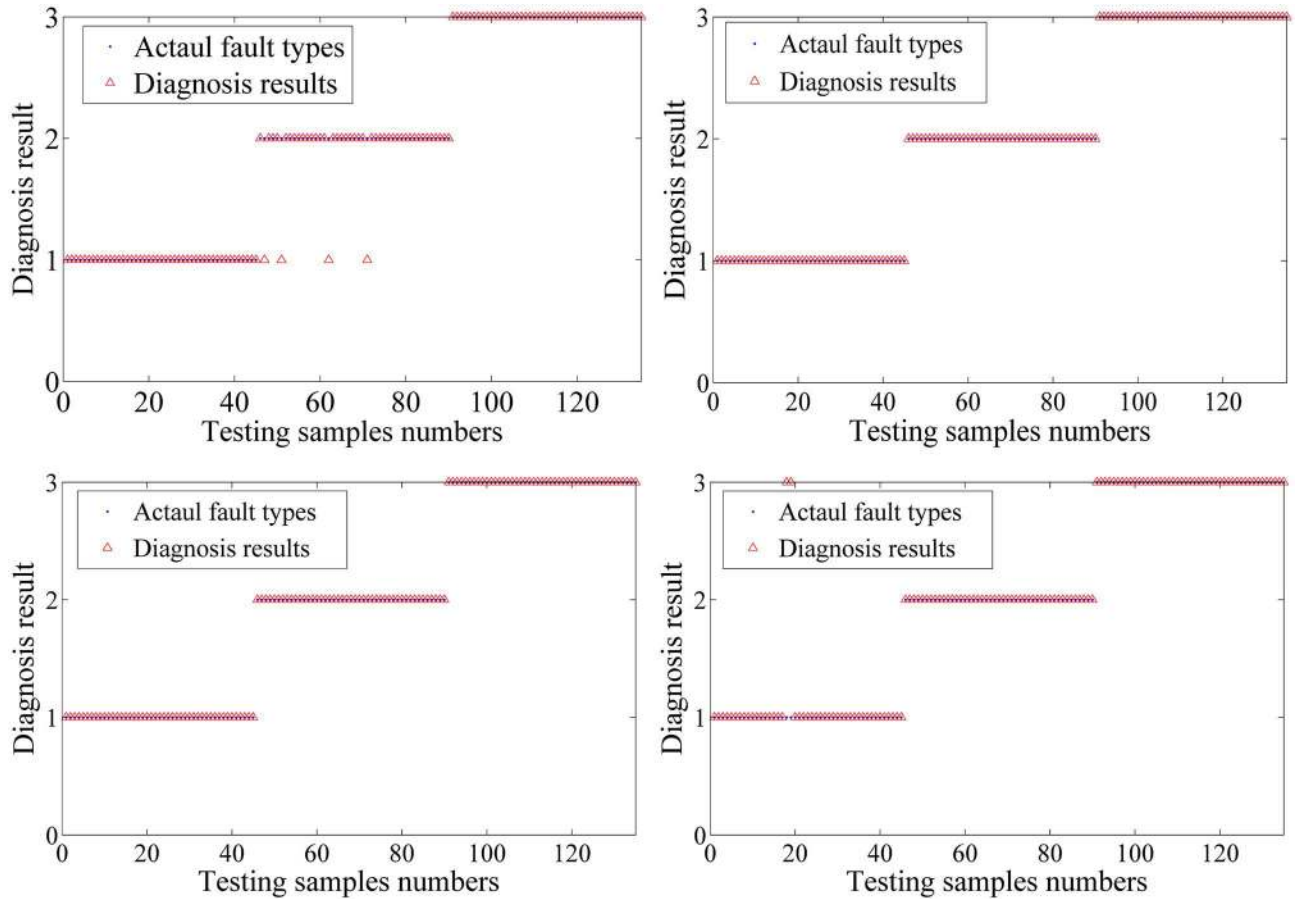|	normal	piston shoes and swashplate wearing	valve plate wearing	Total
number of data points	60	60	60	180
amount of training data	15	15	15	45
amount of testing data	45	45	45	135

Fig 20. Result of 4 set cross validation.

doi:10.1371/journal.pone.0164111.g020

1. Apply this method to more types of machinery.

2. Improve computing speed while maintaining diagnostic accuracy.

3. Conduct further research about the fault diagnosis technologies with fluctuated working conditions.

Table 5. The error rate of 4 groups of cross-validation.

No. of cross validation		Normal	piston shoes and swashplate wearing	valve plate wearing	mean value/total
1	samples in test set	45	45	45	135
	error samples	0	4	0	4
	accuracy	100%	91.11%	100%	97.04%
2	samples in test set	45	45	45	135
	error samples	0	0	0	0
	accuracy	100%	100%	100%	100%
3	samples in test set	45	45	45	135
	error samples	0	0	0	0
	accuracy	100%	100%	100%	100%
4	samples in test set	45	45	45	135
	error samples	2	0	0	2
	accuracy	95.56%	100%	100%	97.78%

doi:10.1371/journal.pone.0164111.t005

Supporting Information

S1 File. The raw data of self-priming centrifugal pump.
(RAR)

S2 File. The raw data of hydraulic pump.
(RAR)

Acknowledgments

Author Contributions

Conceptualization: YC.

Data curation: YW.

Formal analysis: YC.

Funding acquisition: CL.

Investigation: YC.

Methodology: YC YW.

Project administration: CL.

Resources: CL.

Software: YW YC.

Supervision: MR CL.

Validation: YW.

Visualization: YW.

Writing – original draft: YW.

Writing – review & editing: YC YW.

References

1. Chen J, Li Z, Pan J, Chen G, Zi Y, Yuan J, et al. Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review. Mechanical Systems and Signal Processing. 2016; 70:1–35. doi: 10.1016/j.ymssp.2015.08.023

2. Liu Z, He Z, Guo W, Tang Z. A hybrid fault diagnosis method based on second generation wavelet denoising and local mean decomposition for rotating machinery. ISA transactions. 2016. doi: 10.1016/j.isatra.2015.12.009 PMID: 26753616

3. Zhang Y, Tang B, Xiao X. Time–frequency interpretation of multi-frequency signal from rotating machinery using an improved Hilbert–Huang transform. Measurement. 2016; 82:221–39.

4. Lin J, Dou C. The diagnostic line: A novel criterion for condition monitoring of rotating machinery. ISA transactions. 2015; 59:232–42. doi: 10.1016/j.isatra.2015.10.003 PMID: 26603941

5. Cempel C, Tabaszewski M. Multidimensional condition monitoring of machines in non-stationary operation. Mechanical Systems and Signal Processing. 2007; 21(3):1233–41. doi: 10.1016/j.ymssp.2006.04.001

6. Li W, Zhu Z, Jiang F, Zhou G, Chen G. Fault diagnosis of rotating machinery with a novel statistical feature extraction and evaluation method. Mechanical Systems and Signal Processing. 2015; 50:414–26. doi: 10.1016/j.ymssp.2014.05.034

7. Yu Y, Junsheng C. A roller bearing fault diagnosis method based on EMD energy entropy and ANN. Journal of sound and vibration. 2006; 294(1):269–77. doi: 10.1016/j.jsv.2005.11.002

8. Yang W, Tavner P. Empirical mode decomposition, an adaptive approach for interpreting shaft vibratory signals of large rotating machinery. Journal of Sound and Vibration. 2009; 321(3):1144–70. doi: 10.1016/j.jsv.2008.10.012

9. Saidi L, Ali JB, Fnaiech F. Bi-spectrum based-EMD applied to the non-stationary vibration signals for bearing faults diagnosis. ISA transactions. 2014; 53(5):1650–60. doi: 10.1016/j.isatra.2014.06.002 PMID: 24975564

10. Kumar R, Ramaswamy R, Mallick BN. Local properties of vigilance states: EMD analysis of EEG signals during sleep-waking states of freely moving rats. PloS one. 2013; 8(10):e78174. doi: 10.1371/journal.pone.0078174 PMID: 24167606

11. Liu H, Han M. A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings. Mechanism and Machine Theory. 2014; 75:67–78. doi: 10.1016/j.mechmachtheory.2014.01.011

12. Luo S, Cheng J, Ao H. Application of LCD-SVD technique and CRO-SVM method to fault diagnosis for roller bearing. Shock and Vibration. 2015; 2015:847802. doi: 10.1155/2015/847802

13. Wang Z, Han Z, Gu F, Gu JX, Ning S. A novel procedure for diagnosing multiple faults in rotating machinery. ISA transactions. 2015; 55:208–18. doi: 10.1016/j.isatra.2014.09.006 PMID: 25440946

14. Huang J, Hu X, Geng X. An intelligent fault diagnosis method of high voltage circuit breaker based on improved EMD energy entropy and multi-class support vector machine. Electric Power Systems Research. 2011; 81(2):400–7. doi: 10.1016/j.epsr.2010.10.029

15. Samanta B, Al-Balushi K. Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. Mechanical systems and signal processing. 2003; 17(2):317–28. doi: 10.1006/mssp.2001.1462

16. Fei S-w, Zhang X-b. Fault diagnosis of power transformer based on support vector machine with genetic algorithm. Expert Systems with Applications. 2009; 36(8):11352–7. doi: 10.1016/j.eswa.2009.03.022

17. Zhang X, Chen W, Wang B, Chen X. Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization. Neurocomputing. 2015; 167:260–79. doi: 10.1016/j.neucom.2015.04.069

18. Xiaoli X, Zhanglei J, Hongjun W, Guoxin W, Yunbo Z, Peng C, et al. Application of the state deterioration evolution based on bi-spectrum in wind turbine. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science. 2013:0954406213511964. doi: 10.1177/0954406213511964

19. Parker BE, Ware HA, Wipf DP, Tompkins WR, Clark BR, Larson EC, et al. Fault diagnostics using statistical change detection in the bispectral domain. Mechanical systems and signal processing. 2000; 14(4):561–70.

20. Park H, Jang B, Powers EJ, Grady WM, Arapostathis A, editors. Machine condition monitoring utilizing a novel bispectral change detection. Power Engineering Society General Meeting, 2007 IEEE; 2007: IEEE.

21. Saidi L, Ali JB, Fnaiech F. Application of higher order spectral features and support vector machines for bearing faults classification. ISA transactions. 2015; 54:193–206. doi: 10.1016/j.isatra.2014.08.007 PMID: 25282095

22. Zhou Y, Chen J, Dong G, Xiao W, Wang Z. Application of the horizontal slice of cyclic bispectrum in rolling element bearings diagnosis. Mechanical Systems and Signal Processing. 2012; 26:229–43. doi: 10.1016/j.ymssp.2011.07.006

23. Jiang L, Liu Y, Li X, Tang S. Using bispectral distribution as a feature for rotating machinery fault diagnosis. Measurement. 2011; 44(7):1284–92. doi: 10.1016/j.measurement.2011.03.024

24. Liu W, Han J. A fuzzy clustering-based binary threshold bispectrum estimation approach. Neural Computing and Applications. 2012; 21(1):385–92. doi: 10.1007/s00521-012-1050-y

25. Liu W, Han J. Rolling element bearing fault recognition approach based on fuzzy clustering bispectrum estimation. Shock and Vibration. 2013; 20(2):213–25. doi: 10.1155/2013/241937

26. Saidi L, Fnaiech F, Henao H, Capolino G, Cirrincione G. Diagnosis of broken-bars fault in induction machines using higher order spectral analysis. ISA transactions. 2013; 52(1):140–8. doi: 10.1016/j. isatra.2012.08.003 PMID: 22999985

27. Dong G, Chen J, Zhao F. A frequency-shifted bispectrum for rolling element bearing diagnosis. Journal of Sound and Vibration. 2015; 339:396–418. doi: 10.1016/j.jsv.2014.11.015

28. Lindeberg T. Scale invariant feature transform. Scholarpedia. 2012; 7(5):10491. doi: 10.4249/ scholarpedia.10491

29. Yue S, Wang Q, Zhao R. Robust wide baseline feature point matching based on scale invariant feature descriptor. Advanced Intelligent Computing Theories and Applications With Aspects of Theoretical and Methodological Issues: Springer; 2008. p. 329–36. doi: 10.1007/978-3-540-87442-3_42

30. Wang Y, Ban X, Chen J, Hu B, Yang X. License plate recognition based on SIFT feature. Optik-International Journal for Light and Electron Optics. 2015; 126(21):2895–901. doi: 10.1016/j.ijleo.2015.07.040

31. Rublee E, Rabaud V, Konolige K, Bradski G, editors. ORB: an efficient alternative to SIFT or SURF. Computer Vision (ICCV), 2011 IEEE International Conference on; 2011: IEEE.

32. Ke Y, Sukthankar R, editors. PCA-SIFT: A more distinctive representation for local image descriptors. Computer Vision and Pattern Recognition, 2004 CVPR 2004 Proceedings of the 2004 IEEE Computer Society Conference on; 2004: IEEE.

33. Abdel-Hakim AE, Farag AA, editors. CSIFT: A SIFT descriptor with color invariant characteristics. Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on; 2006: IEEE.

34. Morel J-M, Yu G. ASIFT: A new framework for fully affine invariant image comparison. SIAM Journal on Imaging Sciences. 2009; 2(2):438–69. doi: 10.1137/080732730

35. Mikolajczyk K, Schmid C. A performance evaluation of local descriptors. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2005; 27(10):1615–30. doi: 10.1109/TPAMI.2005.188

36. Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features. Computer vision–ECCV 2006: Springer; 2006. p. 404–17.

37. Van der Maaten L, Hinton G. Visualizing data using t-SNE. Journal of Machine Learning Research. 2008; 9(2579–2605):85.

38. Collis W, White P, Hammond J. Higher-order spectra: the bispectrum and trispectrum. Mechanical systems and signal processing. 1998; 12(3):375–94. doi: 10.1006/mssp.1997.0145

39. Carrión MC, Gallego A, Ruiz DP, Morente JA, Giménez G. A block-data recursive-in-order method based on reflection coefficients for bispectrum estimation using AR-modeling. Signal processing. 1995; 47(1):19–32.

40. Bregler C, Omohundro SM, editors. Nonlinear manifold learning for visual speech recognition. Computer Vision, 1995 Proceedings, Fifth International Conference on; 1995: IEEE.

41. Specht DF. Probabilistic neural networks. Neural networks. 1990; 3(1):109–18. doi: 10.1016/0893-6080(90)90049-Q