

# Rolling element bearing fault recognition approach based on fuzzy clustering bispectrum estimation

W.Y. Liu\* and J.G. Han

*School of Mechanical and Electrical Engineering, Jiangsu Normal University, Xuzhou, Jiangsu, China*

Received 26 June 2011

Revised 15 February 2012

**Abstract.** A rolling element bearing fault recognition approach is proposed in this paper. This method combines the basic Higher-order spectrum (HOS) theory and fuzzy clustering method in data mining area. In the first step, all the bispectrum estimation results of the training samples and test samples are turned into binary feature images. Secondly, the binary feature images of the training samples are used to construct object templates including kernel images and domain images. Every fault category has one object templates. At last, by calculating the distances between test samples' binary feature images and the different object templates, the object classification and pattern recognition can be effectively accomplished. Bearing is the most important and much easier to be damaged component in rotating machinery. Furthermore, there exist large amounts of noise jamming and nonlinear coupling components in bearing vibration signals. The Higher Order Cumulants (HOC), which can quantitatively describe the nonlinear characteristic signals with close relationship between the mechanical faults, is introduced in this paper to de-noise the raw bearing vibration signals and obtain the bispectrum estimation pictures. In the experimental part, the rolling bearing fault diagnosis experiment results proved that the classification was completely correct.

Keywords: Fault recognition, fuzzy clustering, bispectrum estimation, bearing fault

## 1. Introduction

Some rotating machinery failure features which are inconspicuous in stable conditions can reflect many special features [1–3]. Bearing is much easier to be damaged in rotating machinery operation process. For the easy changed machine processing state, there exist large amounts of nonlinear coupling components in bearing vibration signals [4–6]. Especially in bad working environment, various faults may occur, such as crack, vortex move, friction and oil film vortex move etc. These faults' occurrence or development often leads to the dynamic signals having nonlinear characteristics, which can characterize the existence of some certain faults [7,8]. But in actual operation condition monitoring, the collected vibration signals have strong noise and low Signals to Noise Ratio (SNR). Therefore it is difficult to identify fault categories accurately from the raw vibration signals [9].

Therefore it needs some better de-noising methods to de-noise the raw vibration signals, and then extract the feature components using feature extraction technologies. In the de-noising and fault diagnosis process, if the de-noising effect is not ideal, it is easy to skip key fault information, which will influence subsequent fault diagnosis process. Especially for the early faults or in strong background noise, even the good effect de-noising method also have difficult to ensure a perfect de-noising effect. Therefore, some researchers began to introduce some analysis

---

\*Corresponding author: W.Y. Liu, School of Mechanical and Electrical Engineering, Jiangsu Normal University, Xuzhou 221116, Jiangsu, China. E-mail: liuwenyi1984@yahoo.com.cn.

methods which had anti-noise ability, and successfully introduced Higher Order Cumulants (HOC) into the nonlinear non-Gaussian signals processing [10–13].

The HOC of signals and noises have different characteristics, so the Gaussian noise interference can be filtered by the HOC analysis. The Higher-order spectrum (HOS), which is one of the most useful common cumulants, can describe random signals from higher probability structure characterization and can quantitatively describe the nonlinear characteristic signals which have close relationship with the mechanical faults. The HOS can also make up the limitations in Fourier analysis, which makes it be a new fault diagnosis method and an effective tool to analyze the nonlinear signals. Signal's HOC is insensitive to additive Gaussian noise and symmetric non-Gaussian noise, so it can detect non-Gaussian characteristics in the output signals and inhibit Gaussian interference, retain signals phase information such as special ability. Meanwhile, the HOC can well testing phase coupling phenomenon, removing phase frequency components without phase coupling, which is more suitable for rotating machinery equipment bearing fault diagnosis and analysis. Generally speaking, Gaussian noise's HOC ( $k > 2$ ) is constantly equal to zero and useful vibration signals components' HOC cannot be interfered by Gaussian noise. Therefore, the Gaussian noise interference can be suppressed by the HOC analysis [14–16].

Based on the above analysis and aimed at the non-Gaussianity of the rotating machinery bearing vibration signals, a fuzzy clustering bispectrum estimation approach is proposed in this paper. The method combines the basic HOS theory and fuzzy clustering technique in data mining, and is applied on the rolling element bearing fault diagnosis.

To address the issues discussed above, this paper is organized as follows. Section 1 gives a brief introduction of the research background to introduce the new method. Section 2 introduces the basic HOS theory. Section 3 describes the threshold processing on the bispectrum estimation results of the analyzed bearing vibration data. Section 4 describes the minimum distance classification to introduce the main steps of the new method. Section 5 gives the experimental analysis to prove the effectiveness of the new method. Finally, the main conclusions of this paper are given in Section 6.

## 2. The basic HOS theory

In order to introduce the definition of HOC easily, the characteristic equation of single random variable, which is defined as below [17–20]: If a random variable  $x$  has probability density  $f(x)$ , its characteristic equation can be defined as

$$\Phi(\omega) = \int_{-\infty}^{+\infty} f(x)e^{j\omega x} dx \quad (1)$$

In other words, the characteristic equation can be expressed as the Fourier transform of probability density  $f(x)$ . But in many situations, formula (1) can be written in Laplace operator form as below

$$\Phi(s) = \int_{-\infty}^{+\infty} f(x)e^{sx} dx = E\{e^{sx}\} \quad (2)$$

Take  $k$  times derivative process to formula (2), and we can get formula (3)

$$\Phi^k(s) = E\{x^k e^{sx}\} \quad (3)$$

Then, we can deduce formula (4)

$$\Phi^k(0) = E\{x^k\} = m_k \quad (4)$$

In other words, a  $k$ -times derivative of equation  $\Phi(s)$  in zero is equal to  $x$ 's  $k$ -th order moment  $m_k$ . In this reason, the  $\Phi(s)$  or  $\Phi(\omega)$  is often called moments generating equation of random variable  $x$ .

The formula (5) below

$$\Psi(\omega) = \ln \Phi(\omega) \quad (5)$$

is called cumulant production function of  $x$ . From formula (5) we can get the definition of cumulant, which can be thought as moments in which the dependence on moments of lower order has been removed. Mathematically they are defined as the coefficients of the Taylor series expansion of the natural logarithm of the characteristic function [21–25].

$K$ -th order cumulant  $c_k$  of random variable  $x$  can be defined as

$$c_k = \left. \frac{d^k \Psi(s)}{ds^k} \right|_{s=0} \tag{6}$$

That is to say, the  $k$ -th order cumulant of  $x$  is its cumulant production function  $\Psi(s)$ 's  $k$  times derivative in origin values.

Define  $x = [x_1, \dots, x^k]^T$  as a random variable and take  $\gamma = \gamma_1 + \gamma_2 + \dots + \gamma_k$  times derivative to its characteristic equation  $\Phi(\omega_1, \dots, \omega_k)$ , then we can get:

$$\frac{\partial^\gamma \Phi(\omega_1, \dots, \omega_k)}{\partial \omega_1^{\gamma_1} \dots \partial \omega_k^{\gamma_k}} = j^\gamma E\{x_1^{\gamma_1} \dots x_k^{\gamma_k} e^{j(\omega_1 x_1 + \dots + \omega_k x_k)}\} \tag{7}$$

Define  $\omega_1 = \dots = \omega_k = 0$ , then:

$$m_{\gamma_1 \dots \gamma_k} = E\{x_1^{\gamma_1} \dots x_k^{\gamma_k}\} = (-j)^\gamma \left. \frac{\partial^\gamma \Phi(\omega_1, \dots, \omega_k)}{\partial \omega_1^{\gamma_1} \dots \partial \omega_k^{\gamma_k}} \right|_{\omega_1 = \dots = \omega_k = 0} \tag{8}$$

This is the definition of random variable  $x$ 's  $\gamma$ -th order moment. In the same manner, we can deduce the definition of  $\gamma$ -th order cumulant of  $x = [x_1, \dots, x^k]$  and its cumulant production function  $\Psi(\omega_1, \dots, \omega_k)$

$$c_{\gamma_1 \dots \gamma_k} = (-j)^\gamma \left. \frac{\partial^\gamma \Psi(\omega_1, \dots, \omega_k)}{\partial \omega_1^{\gamma_1} \dots \partial \omega_k^{\gamma_k}} \right|_{\omega_1 = \dots = \omega_k = 0} \tag{9}$$

When  $\gamma_1 = \dots = \gamma_k = 1$ , we can deduce the  $k$ -th order moment and  $k$ -th order cumulant, which is defined respectively as below:

$$m_k = m_{1, \dots, 1} = \text{mom}(x_1, \dots, x^k) \tag{10}$$

$$c_k = c_{1, \dots, 1} = \text{cum}(x_1, \dots, x^k) \tag{11}$$

If  $\{x(n)\}$  is zero-mean  $k$ -th order steady random process, then the  $k$ -th order moment and  $k$ -th order cumulant of the process can be defined as

$$m_{kx}(\tau_1, \dots, \tau_{k-1}) = \text{mom}\{x(n), x(n + \tau_1), \dots, x(n + \tau_{k-1})\} \tag{12}$$

$$c_{kx}(\tau_1, \dots, \tau_{k-1}) = \text{com}\{x(n), x(n + \tau_1), \dots, x(n + \tau_{k-1})\} \tag{13}$$

At first we study the zero-mean random process  $\{x(n)\}$ , the power spectral density of signal  $x(n)$  is defined as its Fourier transform of correlation function. Similarly, they can also be quoted corresponding to the HOC torque and the definition of the spectrum.

If the high order moment  $m_{kx}(\tau_1, \dots, \tau_{k-1})$  is absolutely summable,

$$\sum_{\tau_1 = -\infty}^{\infty} \dots \sum_{\tau_{k-1} = -\infty}^{\infty} |m_{kx}(\tau_1, \dots, \tau_{k-1})| < \infty \tag{14}$$

Then the  $k$ -th order moment spectrum is defined as the  $k$ -th order moment's  $k - 1$  dimension Fourier transform

$$M_{kx}(\omega_1, \dots, \omega_{k-1}) = \sum_{\tau_1 = -\infty}^{\infty} \dots \sum_{\tau_{k-1} = -\infty}^{\infty} m_{kx}(\tau_1, \dots, \tau_{k-1}) \exp \left[ -j \sum_{i=1}^{k-1} \omega_i \tau_i \right] \tag{15}$$

If the HOC  $c_{kx}(\tau_1, \dots, \tau_{k-1})$  is absolutely summable,

$$\sum_{\tau_1=-\infty}^{\infty} \dots \sum_{\tau_{k-1}=-\infty}^{\infty} |c_{kx}(\tau_1, \dots, \tau_{k-1})| < \infty \tag{16}$$

Then the  $k$ -th order cumulant spectrum is defined as the  $k$ -th order cumulant's  $k - 1$  dimension Fourier transform

$$S_{kx}(\omega_1, \dots, \omega_{k-1}) = \sum_{\tau_1=-\infty}^{\infty} \dots \sum_{\tau_{k-1}=-\infty}^{\infty} c_{kx}(\tau_1, \dots, \tau_{k-1}) \exp \left[ -j \sum_{i=1}^{k-1} \omega_i \tau_i \right] \tag{17}$$

The most commonly used HOS is third order spectrum, which can also be call bispectrum, thus the HOS in this paper is generally referred to bispectrum.

$$B_x(\omega_1, \omega_2) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} c_{3x}(\tau_1, \tau_2) e^{-j(\omega_1 \tau_1 + \omega_2 \tau_2)} \tag{18}$$

In real random process, third order cumulant has 6 kind of symmetrical forms,

$$\begin{aligned} C_{3x}(m, n) &= C_{3x}(n, m) = C_{3x}(-n, m - n) = C_{3x}(n - m, -m) \\ &= C_{3x}(m - n, -n) = C_{3x}(-m, n - m) \end{aligned} \tag{19}$$

According to the definition of bispectrum and symmetry of third order cumulant, we can deduce that the symmetrical forms of bispectrum  $B_x(\omega_1, \omega_2)$  can be expressed as formula (20) [26–28]

$$\begin{aligned} B_x(\omega_1, \omega_2) &= B_x(\omega_2, \omega_1) = B_x^*(-\omega_2, -\omega_1) \\ &= B_x^*(-\omega_1, -\omega_2) = B_x(-\omega_1 - \omega_2, \omega_2) = B_x(-\omega_1 - \omega_2, \omega_1) \\ &= B_x(\omega_1, -\omega_1 - \omega_2) = B_x(\omega_2, -\omega_1 - \omega_2) \end{aligned} \tag{20}$$

Furthermore, the bispectrum has the periodicity in  $2\pi$ , namely,

$$B_x(\omega_1 + 2\pi, \omega_2 + 2\pi) = B_x(\omega_1, \omega_2) \tag{21}$$

### 3. Threshold of bispectrum estimation

#### 3.1. Mapping relationships between bispectrum and faults

In order to analyze the mapping relationships between bispectrum and faults, four different rolling bearing fault vibration signals have been selected to take bispectrum estimation analysis. These four faults are normal state, inner fault, outer fault and ball fault [29]. The bispectrum estimation analysis is using direct method [17]. Firstly the fixed speed (1797 r/min) different fault category bearing data are analyzed by bispectrum estimation, including four fault categories: normal state, inner fault, outer fault and roller fault. The bispectrum features of the four bearing faults are respectively shown in Figs 1(a)–(d).

The X-label  $f_1$  and Y-label  $f_2$  are the two main parameters in bispectrum, they are corresponding to the  $\omega_1$  and  $\omega_2$  in formula (18),  $\omega_1 = 2\pi f_1$  and  $\omega_2 = 2\pi f_2$ . From the above analysis process we can see, the features have similar properties in similar target and obvious differences in different target categories. Therefore the bispectrum can be used to clearly distinguish different fault categories, and better reflect the similarities and differences between different categories. This means that there exist certain mapping relations between the bispectrum estimation and the bearing fault mode, thus the bispectrum estimation can be used as a fault characteristic description.

There is a direct way to classify fault categories using the bispectrum estimation values, which dispose the bispectrum estimation of the unknown target and the known target templates in a special way. If the bispectrum estimation of the unknown target has relationships with the templates, we can judge that they belong to the same category. However, because the bispectrum estimation results are two-dimensional data in this process, both the data and calculation amount are very big in the relevant data processing. In some special areas the bispectrum results are approach to zero and there are many calculation blanks. If all the data are used directly as the characteristic input features, the following classifier design and calculation will also be disturbed. Therefore it is necessary to select the smallest dimension characteristic which is able to distinguish between targets. So we select nonzero bispectrum areas and regard them as characteristics of targets, then design the structure of the target templates.

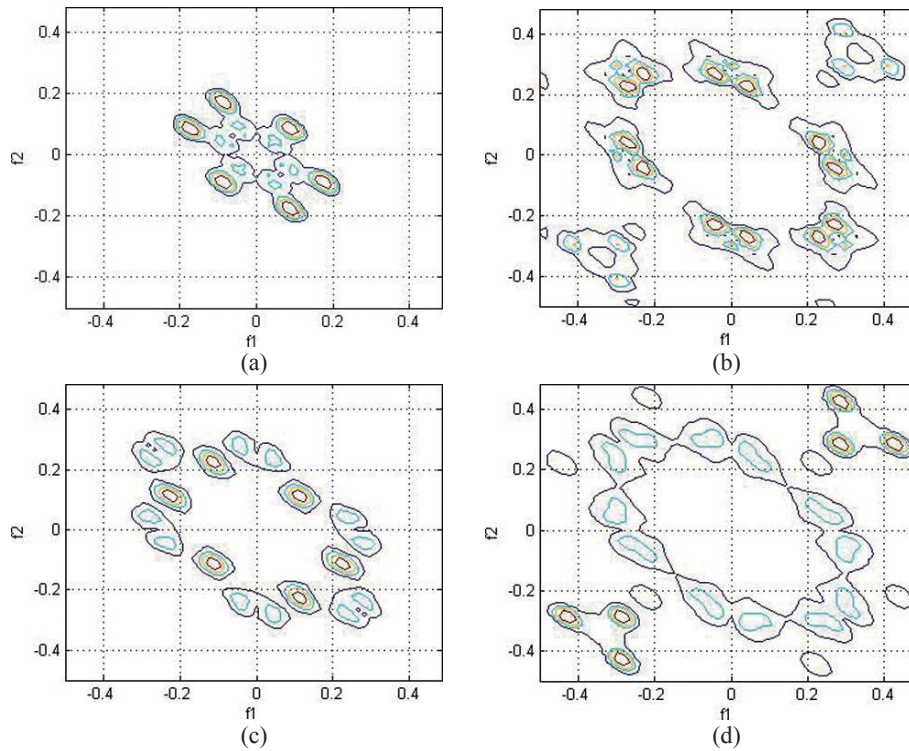


Fig. 1. Bispectrum feature images of different faults (a) Normal state (b) Outer fault (c) Inner fault (d) Roller fault.

### 3.2. Threshold of bispectrum estimation

The bispectrum estimation results are messy but in numerical images they are embodied in black and white color. Therefore, in order to reduce the computation cost and improve operation efficiency, it is necessary to do binarization processing, i.e., to handle the nonzero complex data to 1 or 0, which can greatly reduce the computational cost. Binarization processing usually needs threshold processing, namely to select a suitable threshold in analyzed bispectrum estimation, and to compare the threshold and the other bispectrum estimation. If the bispectrum estimation results are greater than this threshold, the results will be set to 1, otherwise the results will be set to 0. This process is similar to the wavelet de-noising threshold processing. Similarly, bispectrum binarization threshold processing also exist two difficulties, one is threshold way and the other is the threshold selection. Threshold method also can be divided into hard threshold and soft threshold. Theoretically, results obtained by the soft threshold are better than that by hard threshold. However, in this process, it is un-necessary to increase the calculation amount for improving the effect of a small fraction. So finally, we determine to choose the hard threshold way in bispectrum estimation threshold processing, which can be described as:

If we define the bispectrum estimation of the signals  $x(t)$  as  $B_{isx}(\omega_1, \omega_2)$ , take the threshold processing to the bispectrum estimation and set the binary image as the feature image. Define

$$B_{istx}(\omega_1, \omega_2) = \begin{cases} 1 & B_{isx}(\omega_1, \omega_2) \geq T \\ 0 & B_{isx}(\omega_1, \omega_2) < T \end{cases} \quad (22)$$

Where,  $T$  is the threshold.

In the threshold selection aspect, there are various ways, for instance, according to the maximum value of bispectrum to determine the threshold. Define

$$T = \alpha \cdot \max[B_{istx}(\omega_1, \omega_2)] \quad (23)$$

Table 1  
Distances between test sample and each fault template

	Fault 1	Fault 2	Fault 3
Test sample	0	25	51

Where, the parameter  $\alpha$  is a variable value in the interval  $\{0, 1\}$ .  $\alpha = 0$  means there hasn't binarization processing in the bispectrum estimation. In this situation,  $B_{isx}(\omega_1, \omega_2) = B_{istx}(\omega_1, \omega_2)$ . When  $\alpha = 1$ , all the bispectrum estimation data are threshold to zero, only reserving the biggest data. These two situations are both extreme and should be avoided. In normal circumstances, the parameter  $\alpha$  should select between 0.4 and 0.6. This method, determine the threshold value according to the maximum bispectrum value, has a drawback, which is in special circumstances, if suddenly appeared a bigger bispectrum value, much bigger than other bispectrum values, such as when  $\alpha = 0.4$  calculating the threshold value is still greater than other bispectrum estimation value. At this time, it is possible to take a lot of useful information threshold zero, thereby losing useful feature information.

In view of this situation, there is also a determined mean value threshold method, which is determined by the mean value of the bispectrum estimation. Define

$$T = \frac{1}{N \times N} \sum_{\omega_1=0}^{N-1} \sum_{\omega_2=0}^{N-1} B_{isx}(\omega_1, \omega_2) \tag{24}$$

This threshold method can ensure the suitable value in all bispectrum estimation results, in the sudden extremum situation, still can better determine the threshold, and has a good effect. The computation is reduced and the calculation efficiency can be improved after the threshold processing, and the key information of the fault category of distinguishing can be retained. At this time the bispectrum estimation binary image can be directly used for the next step of fault classification, so next we will introduce the fault pattern recognition method.

#### 4. Fuzzy clustering bispectrum estimation fault recognition

##### 4.1. Minimum distance classification

In the cluster analysis, the analyzed target is often called object library and the components in the object library are called objects. Each object represents a category or corresponding to one fault type, which contains many training samples. This is the basic structure of object library analysis. The basic principle of the clustering analysis is to extract objects which have obvious characteristics with the training samples, then to distinguish the objects between each other effectively [30–35]. This paper uses a method based on the objective function of fuzzy clustering, which mainly construct the object templates and nearest classifier, and then deal with the template of different categories to accomplish the template matching and pattern recognition. The core of this method is the minimum distance classification, whose mainly working principle is introduced below through a simple example.

As shown in Figs 2(a)–(c) are three typical fault feature images, (d) is the test sample feature image need to be classified.

As shown in Fig. 2, we can easily judge through macroscopic manner that (d) is similar to (a) and belongs to fault category (a). However, the computer can't give the judge result by "macroscopic manner" at once. Therefore we need to find a suitable way for the computer to judge the fault classification. We put the three faults and one test sample into data gridding processing and turn all the samples into  $10 \times 10$  matrix. Then we define the white area as 0 and the colored area as 1. Three kinds of fault samples and test sample are shown below in Figs 3(a)–(d).

Therefore, we can judge which kind of faults this sample belongs to according to the "distance", which is defined as the total summation of the corresponding two matrix's differences. For two matrix  $A_{m \times n}$  and  $B_{m \times n}$ , the distance  $D_{AB}$  between  $A_{m \times n}$  and  $B_{m \times n}$  can be defined as

$$D_{AB} = \sum_{i=1}^m \sum_{j=1}^n |A_{i,j} - B_{i,j}| \tag{25}$$

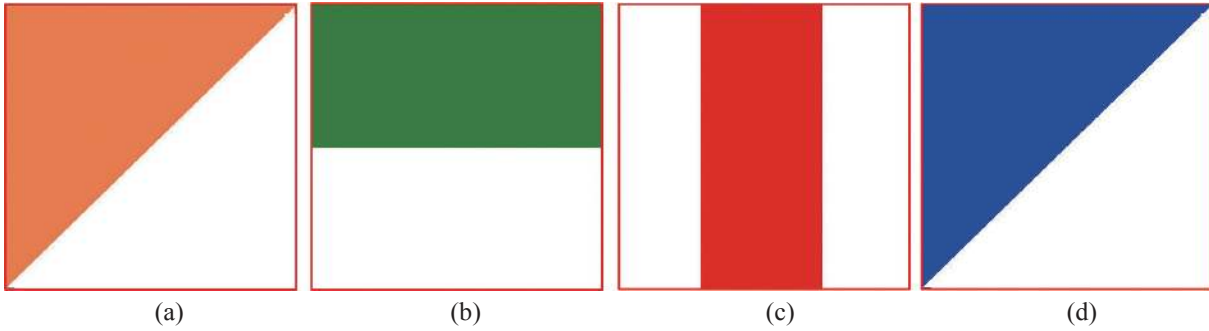


Fig. 2. Fault feature images and test sample feature image (a) Sample 1 (b) Sample 2 (c) Sample 3 (d) Test sample.

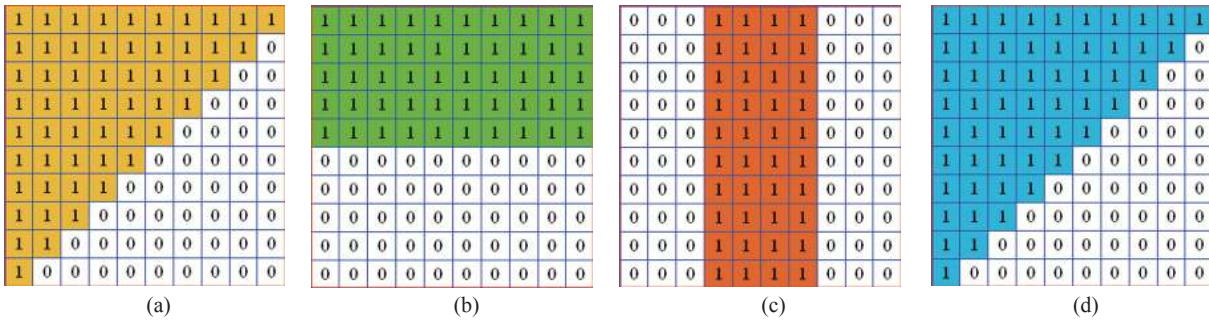


Fig. 3. Data gridding results of the fault template and the test sample (a) Sample 1 (b) Sample 2 (c) Sample 3 (d) Test sample.

The smaller the "distance" between test sample and the fault, the test sample is more closer to the fault category, can be judged that they belong to the same class. The bigger the distance between test sample and the fault category, the test sample is much farther from the fault category, can be judged that they belong to different classes.

We calculate each distance between test sample and three fault templates, the results are shown in Table 1.

The distance between the test sample and fault 1 is the minimum one, so we can judge that test sample belongs to the first category. This is the basic principle of minimum distance method. The principle of fuzzy clustering bispectrum estimation fault diagnosis method is similar to minimum distance method. This method converts the bispectrum value of the signals to a matrix only contains 0 and 1 (this can be called binarization threshold processing), and then construct the different target templates. The test sample which is also converted into 0 and 1 is used to test minimum "distance", in order to judge its fault category.

#### 4.2. Construct the target template

Assuming an analyzed object library have  $c$  class targets  $\{X_1, X_2, \dots, X_c\}$ , each object have  $m$  training samples. Define  $\text{Tmp}_i = \{K_i, D_i\} (i = 1, 2, \dots, c)$  as the object template, which contains two parts. One is  $K_i$  called the  $i$ -class target's kernel, defined as the public section of all samples in that object. The other  $D_i$  is called the  $i$ -class target's domain, defined as the union section of all samples in that object. It also can be said that all of the target within the sample contains kernel, and are contained by the domain. Combined with the analysis objects in this section, for bispectrum binary feature image  $B_{isk}(\omega_1, \omega_2)$ , the kernel and domain can be defined separately as follows:

$$K_i(\omega_1, \omega_2) = \bigcap_{k=1}^m \{B_{isk}(\omega_1, \omega_2)\} \tag{26}$$

$$D_i(\omega_1, \omega_2) = \bigcup_{k=1}^m \{B_{isk}(\omega_1, \omega_2)\} \tag{27}$$

Where,  $B_{isk}(\omega_1, \omega_2)$  is the bispectrum binary feature image with  $k$  training samples in  $i$ -class objects. According to definition functions (26) and (27) we know that  $K_i$  expresses the generality of  $i$ -class objects, in other words is the public section, which constructs the common features of the object.  $D_i$  expresses the distribution of  $i$ -class object bispectrum, that is to say, this domain includes all the bispectrum binary feature images. The templates constructed by the kernel images and the domain images are not only the common bispectrum features between the same object, but also the differences between their distribution.

After getting the definition of kernel and domain, we can classify the objects by the minimum distance method, in which the minimum distance classifier is constructed according to the decision rules.

### 4.3. Minimum distance classifier

In the minimum distance classification method, there are two kind of basic judgment rules, one is  $1 - NN$  rule and the other is  $k - NN$  rule. In the hypothesis in  $c$  class problems with  $X = \{x_1, x_2, \dots, x_n\}$  samples, classification property of each sample is certain. Then in class  $X$  the sample number belongs to which sample category is also certain, which is defined as  $n_1, n_2, \dots, n_c$ .  $x_i^k$  is used to describe the  $k$  sample in category  $w_i$ , if we want to classify a random sample  $x$ , the  $1 - NN$  rule can be described as follows [36,37].

If

$$D(x, x_i^k) = \min_{i=1, \dots, c} \{D(x, x_i^k), k = 1, 2, \dots, n_i\} \tag{28}$$

Then in the function  $x \in w_i$ ,  $D(x, x_i^k)$  is the distance between sample  $c$  and sample  $k$  in category  $w_i$ , this distance is commonly used Euclidean distance to measure. This rule shows that as long as the distance between samples  $x$  and one category sample is the nearest, the sample  $x$  belongs to this category. Namely the sample  $x$  is relegated to the nearest distance subordinate category, so this method can also be called minimum distance decision rule.

The  $k - NN$  rule can be extended from  $1 - NN$  rule. On the above definition conditions, for every sample  $x$  to be classified, we can find  $k$  nearest distances, which belong to  $c$  categories. Define  $k = k_1 + k_2 + \dots + k_c$ , where  $k_i$  is the sample number of the category  $w_i$ . Then the  $k - NN$  rule can be defined as follows.

If

$$k_i = \max_{j=1, \dots, c} \{k_j\} \tag{29}$$

then  $x \in w_i$ .

This definition can also be described as that the sample  $x$  is classified to the category most samples belong to. In the feature space, the distances between different category features are often bigger than the distances between same category features. Through the analysis, the same category features are often gathered together and distribute around the “kernel”. For an input feature vector of random unknown category, by comparing the distance between the vector and the “kernel”, we can judge which category it belongs to.

Generally speaking, if we want to inspect a mode classifier, it needs to build two categories of samples for inspection, one is training samples and the other is test samples. Among them the training sample is known belongs to some certain category to build component classifier, and the test sample is used to validate classification effect. Usually we need firstly using training samples to determine classifier’s parameters and design classifier, then using test samples to classify the samples and test its classification effect.

For the minimum distance analysis method, its realization process can be described as: after constructing the kernel and domain, obtain the target template, and then accomplish the classification by testing the distances between samples  $x(t)$ ’s template feature image  $B_{isx}$  and object template  $Tmp_i$ . Its decision function is

$$d_i(X) = \min_{i=1}^c \{d(B_{isx}, Tmp_i)\} \tag{30}$$

Its category discrimination is

$$g = \arg \min_i \{d_i(X)\} \tag{31}$$



Table 2  
Geometry dimension of the bearing (mm)

Type	Inside diameter	Outside diameter	Thickness	Ball diameter	Pitch diameter
SKF6203	17.0002	39.9999	11.9990	6.7462	28.4988
SKF6205	25.0012	51.9989	0.5906	15.0012	39.0398

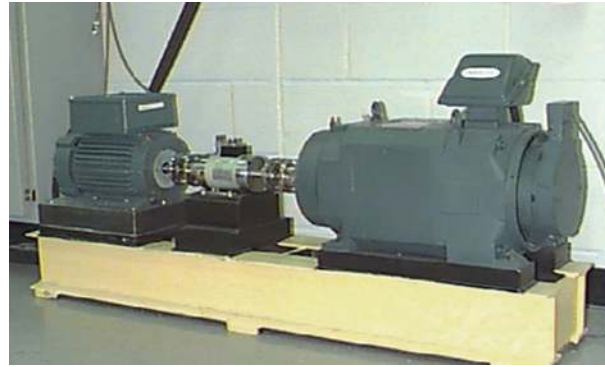


Fig. 4. Sketch map of the bearing test-rig.

Where,  $g$  is the category number,  $d(B_{isx}, Tmp_i)$  is the distance between binary image  $B_{isx}$  and the  $i$ -category template  $Tmp_i$ .  $d(B_{isx}, Tmp_i)$  is defined as follows:

$$\begin{aligned}
 d(B_x, M_i) &= \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \text{sgn}(D_i(m, n) - B_{isx}(m, n)) \\
 &\quad + \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \text{sgn}(B_{isx}(m, n) - K_i(m, n)) \\
 &= d(D - B) + d(B - K)
 \end{aligned} \tag{32}$$

Where,  $\text{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}$ , is the symbols function.  $d(B - K)$  is the one-way distance between the kernel  $K_i$  and the test sample feature image  $B_{isx}$ . This one-way distance is used to measure how far the samples leave the kernel.  $d(D - B)$  is the one-way distance between the domain  $D_i$  and  $B_{isx}$ . This one-way distance is used to measure the degree how far the samples leave this target domain.

In this algorithm processing, the various feature images are carried out a threshold binary processing, the obtained results only embodied in 0 and 1 two kinds. So this method can not only greatly improve the computation efficiency but also simplify the calculation process and increase the computer real-time processing ability.

### 5. Experimental analysis

Experiments were conducted using a 2 hp reliance electric motor, and the acceleration data was collected at locations near to and remote from the motor bearings. Motor bearings were seeded with faults using electro-discharge machining. Faults ranging from 0.007 inches in diameter to 0.040 inches in diameter were introduced separately at the inner raceway, outer raceway and the rolling ball. Faulted bearings were reinstalled into the test motor and vibration data was recorded for motor power of 0 to 3 horsepower (motor speeds of 1797 to 1720 RPM) [29].

As shown in Fig. 4 above, the test stand consists of a 2 hp motor, a torque transducer/encoder, a dynamometer, and control electronics. The test SKF bearings, which support the motor shaft, are used for the 7, 14 and 21 mils diameter faults. Drive end and fan end bearing specifications, including bearing geometry and defect frequencies are listed in Tables 2 and 3.

Table 3  
Multiple of bearing running speed (Hz)

Type	Inner	Outer	Cage train	Ball
SKF 6203	4.9469	3.0530	0.3817	3.9874
SKF 6205	5.4152	3.5848	0.39828	4.7135

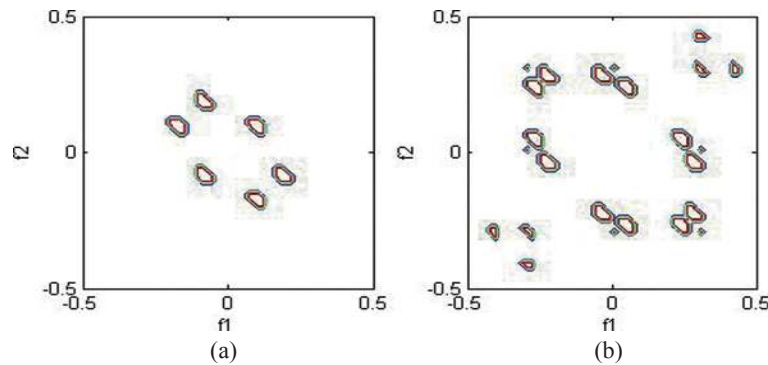


Fig. 5. Two fault feature images (a) inner fault binary feature image (b) ball fault binary feature image.

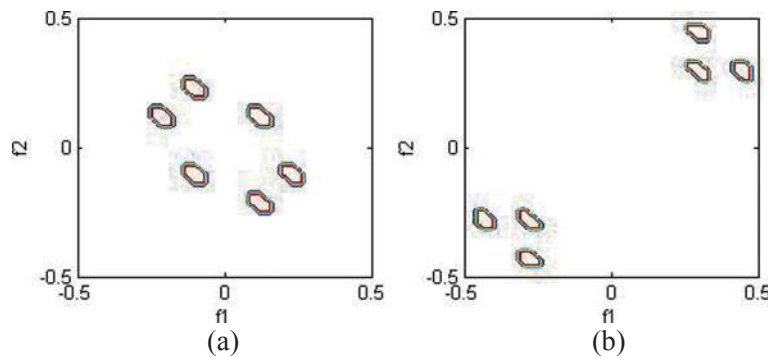


Fig. 6. Kernel images of object template (a) Inner fault kernel image (b) Ball fault kernel image.

Bearing vibration data was collected using accelerometers, which were attached to the housing with magnetic bases. Accelerometers were placed at the 12 o'clock position at both the drive end and fan end of the motor housing. Digital data was collected at 48,000 samples per second for drive end bearing faults. Speed and the horsepower data were collected by the torque transducer/encoder equipment.

We classify the different fault type 6205 rolling bearing vibration signals, using the above bispectrum binary image feature extraction, template structure and template matching classification method. In each category 12 group data are selected in normal state, inner fault, outer fault and ball fault, in which 9 groups are training samples and 3 groups are test samples.

At first, carry on the bispectrum estimation on the each 12 group data in every 4 fault states, then the binarization threshold processing. 12 group results (each group 4 results, total 48 binary results) are obtained in the binarization threshold processing, 2 of them are shown in Figs 5(a), (b) below, which is respectively represent inner fault binary feature image and ball fault binary feature image.

From Fig. 5 we can see, after the binarization processing to bispectrum estimation, all data are translated into 1 and 0, which in the image is embodied in the black and white. In this way, the data are greatly simplified and redundant unnecessary or noneffective data are suppressed.

On the base of the binary feature images, object templates are constructed including the kernel and the domain. In this situation, the fault category is  $c = 4$ ,  $i = 1, 2, 3, 4$ . Construct the object templates using the above method to the 9 group (in total 36 classes samples) training samples, and get 1 group in total 4 classes object kernel templates,

Table 4  
Distances between test samples and object templates

Test samples		$d(B_x, T_1)$	$d(B_x, T_2)$	$d(B_x, T_3)$	$d(B_x, T_4)$	Diagnosis results
Normal	Nor01	<u>230</u>	990	783	690	Correct
	Nor02	<u>196</u>	1123	786	709	Correct
	Nor03	<u>253</u>	984	775	762	Correct
Inner fault	Inner01	764	<u>414</u>	765	630	Correct
	Inner02	734	<u>443</u>	803	635	Correct
	Inner03	726	<u>419</u>	882	683	Correct
Outer fault	Out01	537	924	<u>373</u>	754	Correct
	Out02	590	983	<u>397</u>	703	Correct
	Out03	525	1020	<u>376</u>	694	Correct
Ball fault	Ball01	612	945	867	<u>196</u>	Correct
	Ball02	687	903	895	<u>241</u>	Correct
	Ball03	634	956	887	<u>218</u>	Correct

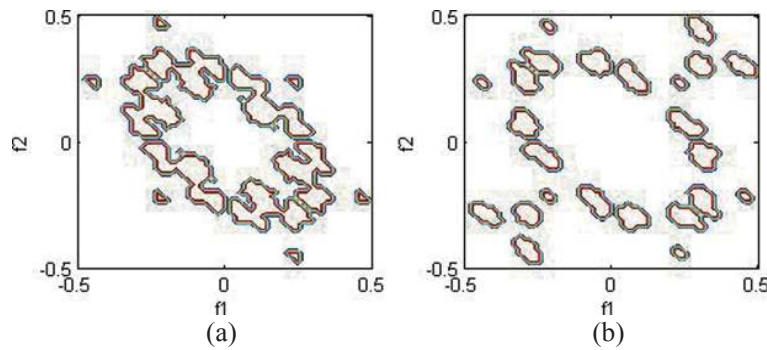


Fig. 7. Domain images of object template (a) Inner fault domain image (b) Ball fault domain image.

named  $K_1, K_2, K_3, K_4$ .  $K_3$  and  $K_4$  are shown in Figs 6(a), (b) respectively.

From Fig. 6 we can see that, the analyzed data are further simplified by the object template kernel image. One kernel image is constructed by 9 binary feature images in the same category. In other words the inner fault kernel image in Fig. 6(a) are constructed by 9 inner fault binary feature images. The kernel images reflect the public characteristics of the training samples, namely the common attributes in the same state.

Then construct the domain images according to the above method, using 9 group 36 classes training samples. One group (including 4 classes) domain templates are obtained and defined as  $D_1, D_2, D_3, D_4$ .  $D_3$  and  $D_4$  are shown in Figs 7(a), (b) respectively.

From Fig. 7 we can see, the domain image of the object template further extend the data, and one domain image is constructed by 9 binary feature images. In other words, the inner fault domain image is constructed by 9 inner fault binary feature images. The domain image reflects the all attributes of the same category state. Therefore we can say that, the kernel images and the corresponding binary feature images are embodied in the corresponding domain images.

The object templates, including kernel images and domain images, are defined as  $T_1, T_2, T_3, T_4$ . Then the other 3 group of test samples are used to validate the classification effect. The calculation results are shown in Table 4, in which  $d(B_x, T_1)$  is normal object template,  $d(B_x, T_2)$  is inner fault object template,  $d(B_x, T_3)$  is outer fault object template and  $d(B_x, T_4)$  is ball fault object template.

As shown in Table 4, there are 3 test samples in each bearing state. By calculating the distances between test samples and target templates, the results are obtained and shown in the same line. Through the calculation result analysis, we can conclude that every fault diagnosis is correct. Such as in normal state, the distances between 3 group testing samples and the normal target template are the minimum. The other three kind of distances, such as to the outer fault templates, inner ring fault templates, roller fault templates, are much larger. Similarly, in other fault categories the test distances acquired are the minimum in the corresponding fault target templates.

## 6. Conclusions

- (1) We used bispectrum estimation to analyze the different categories of rolling bearings vibration signals. The research results showed that the picture shape of the bispectrum estimation in the same fault is similar to each other, while in different categories is different from each other.
- (2) The rotate speed of the bearing has little influence on the bispectrum results in the same fault category.
- (3) The binary feature images, which are used to construct object templates including kernel images and domain images, are obtained by threshold processing on the bispectrum estimation. By calculating the distances between samples to the different object templates, object classification and pattern recognition can be effectively achieved.
- (4) The rolling bearing fault diagnosis example used 9 group training samples to construct object templates and 3 group test samples for effect validation. The results showed that the distance between test sample to the same category fault target template was the minimum in the same comparison group and the classification was completely correct.

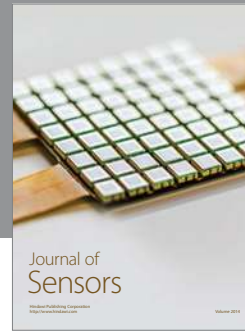
## Acknowledgments

This research was supported by the Scientific research support project for teachers with doctor's degree, Xuzhou normal university, China (Grant No. 11XLR15), the National Natural Science Foundation of China (Grant No. 51075347).

## References

- [1] L. Renaudin, F. Bonnardot, O. Musy, J.B. Doray and D. Rémond, Natural roller bearing fault detection by angular measurement of true instantaneous angular speed, *Mechanical Systems and Signals Processing* **24**(7) (October 2010), 1998–2011.
- [2] S.H. Ghafari, E.M. Abdel-Rahman, F. Golnaraghi and F. Ismail, Vibrations of balanced fault-free ball bearings, *Journal of Sound and Vibration* **329**(9)(26) (April 2010), 1332–1347.
- [3] M. Liang and I.S. Bozchalooi, An energy operator approach to joint application of amplitude and frequency-demodulations for bearing fault detection, *Mechanical Systems and Signals Processing* **24**(5) (July 2010), 1473–1494.
- [4] M. Elforjani and D. Mba, Accelerated natural fault diagnosis in slow speed bearings with acoustic emission, *Fracture Mechanics* **77**(1) (January 2010), 112–127.
- [5] Q. He, D.M. Du and X.H. Wang, Autoregressive model-based vibration fault diagnosis of rolling bearing, *Noise and Vibration Worldwide* **41**(10) (1 November 2010), 22–28.
- [6] S. Geetha, N. Ishwarya and N. Kamaraj, Audio steganalysis with hausdorff distance higher order statistics using a rule based decision tree paradigm, *Expert Systems with Applications* **37**(12) (December 2010), 7469–7482.
- [7] P.K. Kankar, S.C. Sharma and S.P. Harsha, Fault diagnosis of ball bearings using machine learning methods, *Expert Systems with Applications* **38**(3) (March 2011), 1876–1886.
- [8] H. Li, Y.P. Zhang and H.Q. Zheng, Bearing fault detection and diagnosis based on order tracking and teager-huang transform, *Journal of Mechanical Science and Technology* **24**(3) (March 2010), 811–822.
- [9] H. He, Multitaper higher-order spectral analysis of nonlinear multivariate random processes, *Canada degree of Doctor of Philosophy Queen's University Kingston Ontario*, 2008.
- [10] J.M. Smulko and L.B. Kish, Higher-order statistics for fluctuation-enhanced gas-sensing, *Sensors and Materials* **16**(6) (2004), 291–299.
- [11] Y. Birkelund and A. Hanssen, Multitaper estimators for bispectra, *Proc of the IEEE Workshop on Higher-Order Statistics* (1999), 207–211.
- [12] Y. Birkelund, A. Hanssen and E.J. Powers, Multitaper estimators of polyspectra, *Signals Processing* **83** (2003), 545–559.
- [13] L.S. Qu and Z.J. He, *Mechanical failure diagnostics*, Science and Technology Press, Shanghai, 1986.
- [14] Y.B. Nouredine, Extraction of signals buried in noise, Part I: Fundamentals, *Signals Processing* **86** (2006), 2464–2478.
- [15] Y.B. Nouredine, Extraction of signals buried in noise, Part II: Experimental results, *Signals Processing* **86** (2006), 2994–3011.
- [16] Z.N. Lai, Z.T. Wu, Y.Y. He et al., Hidden markov model-based fault diagnostics method in speed-up and speed-down process for rotating machinery, *Mechanical Systems and Signals Processing* **19** (2005), 329–339.
- [17] W.B. Collis, P.R. White and J.K. Hammond, Higher-order spectra: The bispectrum and trispectrum, *Mechanical Systems and Signals Processing* **12**(3) (May 1998), 375–394.
- [18] A.E. Shermeh and R. Ghazalian, Recognition of communication signals types using genetic algorithm and support vector machines based on the higher order statistics, *Digital Signals Processing: A Review Journal* **20**(6) (December 2010), 1748–1757.
- [19] Y. Takahashi, H. Saruwatari, K. Shikano and K. Kondo, Musical-noise analysis in methods of integrating microphone array and spectral subtraction based on higher-order statistics, *Eurasip Journal on Advances in Signals Processing*, 2010.

- [20] J.E. Laier, Spectral analysis of a high-order hermitian algorithm for structural dynamics, *Applied Mathematical Modelling* **35**(2) (February 2011), 965–971.
- [21] A. Napolitano and M. Tesauro, Almost-periodic higher order statistic estimation, *IEEE Transactions on Information Theory* **57**(1) (January 2011), 514–533.
- [22] V. Choqueuse, A. Mansour, G. Burel, L. Collin and K. Yao, Blind channel estimation for STBC systems using higher-order statistics, *IEEE Transactions on Wireless Communications* **10**(2) (February 2011), 495–505.
- [23] N.C. Sagias, F.I. Lazarakis, A.A. Alexandridis, K.P. Dangakis and G.S. Tombras, Higher order capacity statistics of diversity receivers, *Wireless Personal Communications* **56**(4) (February 2011), 649–668.
- [24] G. Birot, L. Albera and P. Chevalier, Sequential high-resolution direction finding from higher order statistics, *IEEE Transactions on Signals Processing* **58**(8) (August 2010), 4144–4155.
- [25] J.L. Tan and A.Z.B. Sha’Ameri, Adaptive optimal kernel smooth-windowed wigner-ville bispectrum for digital communication signals, *Signals Processing* **91**(4) (April 2011), 931–937.
- [26] A. Lanatá, G. Valenza, C. Mancuso and E.P. Scilingo, Robust multiple cardiac arrhythmia detection through bispectrum analysis, *Expert Systems with Applications* **38**(6) (June 2011), 6798–6804.
- [27] C.R.P. Courtney, S.A. Neild, P.D. Wilcox and B.W. Drinkwater, Application of the bispectrum for detection of small nonlinearities excited sinusoidally, *Journal of Sound and Vibration* **329**(20) (2010), 4279–4293.
- [28] C.J. Gu, S.Y. Zhang, K. Liu and H. Huang, Fuzzy kernel k-means clustering method based on immune genetic algorithm, *Journal of Computational Information Systems* **7**(1) (January 2011), 221–231.
- [29] <http://www.eecs.case.edu/laboratory/bearing/download.htm>.
- [30] H. Izakian and A. Abraham, Fuzzy c-means and fuzzy swarm for fuzzy clustering problem, *Expert Systems with Applications* **38**(3) (March 2011), 1835–1838.
- [31] A. Karahoca and D. Karahoca, GSM churn management by using fuzzy c-means clustering and adaptive neuro fuzzy inference system, *Expert Systems with Applications* **38**(3) (March 2011), 1814–1822.
- [32] S.-C. Horng, F.-Y. Yang and S.-S. Lin, Hierarchical fuzzy clustering decision tree for classifying recipes of ion implanter, *Expert Systems with Applications* **38**(1) (January 2011), 933–940.
- [33] Y. Ozbay, R. Ceylan and B. Karlik, Integration of type-2 fuzzy clustering and wavelet transform in a neural network based ECG classifier, *Expert Systems with Applications* **38**(1) (January 2011), 1004–1010.
- [34] F. Naghdy, Fuzzy clustering of human motor motion, *Applied Soft Computing Journal* **11**(1) (January 2011), 927–935.
- [35] Q. Tan, Q. He, W.Z. Zhao, Z.Z. Shi and E.S. Lee, An improved FCMBP fuzzy clustering method based on evolutionary programming, *Computers and Mathematics with Applications* **61**(4) (February 2011), 1129–1144.
- [36] N. Sawalhi and R.B. Randall, Helicopter gearbox bearing blind fault identification using a range of analysis techniques, *Australian Journal of Mechanical Engineering* **5**(2) (2008), 157–168.
- [37] V. Sugumaran and K.I. Ramachandran, Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing, *Mechanical Systems and Signals Processing* **5**(21) (2007), 2237–2247.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

