Research Article

Modified Kernel Marginal Fisher Analysis for Feature Extraction and Its Application to Bearing Fault Diagnosis

Li Jiang and Shunsheng Guo

School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070, China

Correspondence should be addressed to Li Jiang; happyjl0929@163.com

The high-dimensional features of defective bearings usually include redundant and irrelevant information, which will degrade the diagnosis performance. Thus, it is critical to extract the sensitive low-dimensional characteristics for improving diagnosis performance. This paper proposes modified kernel marginal Fisher analysis (MKMFA) for feature extraction with dimensionality reduction. Due to its outstanding performance in enhancing the intraclass compactness and interclass dispersibility, MKMFA is capable of effectively extracting the sensitive low-dimensional manifold characteristics beneficial to subsequent pattern classification even for few training samples. A MKMFA- based fault diagnosis model is presented and applied to identify different bearing faults. It firstly utilizes MKMFA to directly extract the low-dimensional manifold characteristics from the raw time-series signal samples in high-dimensional ambient space. Subsequently, the sensitive low-dimensional characteristics in feature space are inputted into K-nearest neighbor classifier so as to distinguish various fault patterns. The four-fault-type and ten-fault-severity bearing fault diagnosis experiment results show the feasibility and superiority of the proposed scheme in comparison with the other five methods.

1. Introduction

Rolling bearing is frequently used in rotating machinery. The damaged bearings are often the leading cause of the catastrophic machine breakdown and big economic loss [1, 2]. Therefore, it is significant for rolling bearings to implement fault diagnosis so as to prevent fatal malfunction of rotating machinery and even human casualties [3, 4]. Feature extraction is significant for bearing fault diagnosis. There are common feature extraction approaches to bearing fault diagnosis. The time-domain statistical features (e.g., mean, root mean square, etc.) [5] are extracted from the periodic time-series signals. The frequency-domain statistical features (e.g., mean frequency, etc.) [6] are extracted from the frequency spectrums of faulty signals. The classical time–frequency analysis techniques are suitable to nonlinear and nonstationary signals, like empirical mode decomposition (EMD) [7]. The features in the three domains are usually high-dimensional to obtain comprehensive faulty information. A large number of redundant and irrelevant characteristics will degrade the diagnosis performance and increase the computing consumption. Conversely, a few salient features will improve the fault identification accuracy and alleviate the computation burden [8]. Consequently, it is a great challenge for rolling bearings to extract the sensitive low-dimensional characteristics for improving fault diagnosis performance.

The vibration signals of faulty machinery are weak nonlinear or strong nonlinear due to the instantaneous variations in friction, damping, and loads [9]. Kernel Fisher discriminant analysis (KFDA) [10] and Kernel principal component analysis (KPCA) [11] are classical nonlinear feature extraction methods with dimensionality reduction. Liu et al. utilized KFDA to capture low-dimensional characteristics of planetary gearboxes [12]. Shao et al. employed KPCA to capture low-dimensional characteristics from the 16-dimensional wavelet packet energy features of a gear system [13]. However, KPCA neglects the significant discriminant information related to the subsequent pattern classification. Despite KFDA being supervised, it may perform not well due to overlooking the non-Gaussian distribution characteristics of faulty samples. Moreover, both KPCA and KFDA

effectively discover only the global structure in Euclidean space. They may not effectively excavate the underlying manifold structure which is more beneficial to classification assignment compared with the global structure, if the high-dimensional samples locate or keep close to a low-dimensional manifold [14, 15].

Many studies reveal that manifold learning geometrically motivated can well handle the high-dimensional nonlinear samples and exploit the inherent low-dimensional manifold structure [16–18]. Some recent investigations have demonstrated that manifold learning methods can extract the sensitive low-dimensional manifold characteristics beneficial to pattern classification for bearing fault diagnosis [19–22]. As one of the representative manifold learning techniques, marginal Fisher analysis (MFA) [23] algorithm was successfully applied to face recognition [23, 24] and gait recognition [25]. It has also been proved to be an effective methodology for bearing fault diagnosis [5]. MFA is a linear method in essence. Hence, kernel MFA (KMFA) was proposed by applying the kernel trick to MFA [23]. Although KMFA is a prominent approach, it has some flaws for bearing feature extraction. Firstly, most kernel-based algorithms, like KMFA and KFDA, usually take advantage of Gaussian radial basis function (RBF) as the kernel function. Nevertheless, it is an open issue to select the best kernel parameters for improving their feature extraction performances [26]. Secondly, the classical KMFA algorithm (not utilizing PCA preprocessing) may encounter the singular problem if there is no sufficient training (or labeled) faulty sample especially for expensive and critical machine. In such case, the classical KMFA algorithm fails to obtain a stable solution and effectively extract the sensitive low-dimensional manifold features of mechanical equipment. Thirdly, the similarities of the two neighborhood graphs in most extended KMFA algorithm are defined to be either 1 or 0, which do not simultaneously utilize the label information and distance relationship of sample points.

In view of the aforementioned deficiencies, this paper presents modified kernel marginal Fisher analysis (MKMFA) algorithm to make KMFA more robust for feature extraction and pattern classification. MKMFA utilizes the data-dependent kernel function [27] without the selection of the best kernel parameters. Additionally, it introduces a manifold regularization term to solve the singular problem and simultaneously incorporates the label information as well as the distance relationship of sample points into the two similarities to further enhance its classification capability. Subsequently, the MKMFA-based fault diagnosis model is presented and applied to identify various bearing faults. Unlike the traditional signal processing-based fault diagnosis techniques, it is unnecessary for the proposed scheme to extract the high-dimensional multidomain features by signal processing approaches and then reduce the feature dimension before pattern classification. By implementing MKMFA algorithm, it directly extracts the optimal low-dimensional manifold characteristics from the time-series signal samples in high-dimensional ambient space and simultaneously preserves the inherent manifold structure related to fault patterns. Finally,

K-nearest neighbor (KNN) classifier is employed to identify various fault operations of rolling bearings in category space.

The remainder of the paper is organized as follows. The principle of MKMFA algorithm is addressed in Section 2. In Section 3, the MKMFA-based fault diagnosis model is presented and applied to identify various bearing faults. Finally, Section 4 gives the concluding remarks.

2. MKMFA Algorithm

KMFA is designed to capture the low-dimensional manifold characteristics embedded in high-dimensional ambient space based on graph embedding framework. The outline of graph embedding framework and KMFA algorithm are first briefly presented. For more details on them, readers can refer to [23]. Given a D-dimensional sample set of N samples $\mathbf{X} = [x_1, x_2, \ldots, x_N]$, $x_i \in R^D$. Suppose $c(x_i) \in \{1, 2, \ldots, N_c\}$ denotes the class label of the sample x_i and N_c is the number of classes. The low-dimensional representation of the high-dimensional sample set $\mathbf{X}$ is denoted by a vector $\mathbf{Y} = [y_1, y_2, \ldots, y_N]$, $y_i = \mathbf{Z}^T x_i$ ($\mathbf{Z}$ is a transformation matrix, $y_i \in R^d$ and $d \ll D$).

2.1. Outline of Graph Embedding Framework. Assume $\mathbf{G}_c = \{\mathbf{X}, \mathbf{S}^c\}$ to be an intrinsic graph with vertex set $\mathbf{X}$ and similarity matrix $\mathbf{S}^c$. The similarity matrix $\mathbf{S}^c \in R^{N \times N}$ of the intrinsic graph represents the similarities between vertexes. The diagonal matrix $\mathbf{D}^c$ of the intrinsic graph $\mathbf{G}_c$ is defined as

$$\mathbf{D}^c = \left(D_{ii}^c\right)_{N \times N} = \sum_{i \neq j} S_{ij}^c, \quad \forall i. \tag{1}$$

Suppose $\mathbf{G}_p = \{\mathbf{X}, \mathbf{S}^p\}$ to be a penalty graph with vertex set $\mathbf{X}$ and similarity matrix $\mathbf{S}^p$. The similarity matrix $\mathbf{S}^p \in R^{N \times N}$ of the penalty graph reflects the suppressed similarity properties between vertexes. The diagonal matrix $\mathbf{D}^p$ of the penalty graph $\mathbf{G}_p$ is defined as

$$\mathbf{D}^p = \left(D_{ii}^p\right)_{N \times N} = \sum_{i \neq j} S_{ij}^p, \quad \forall i. \tag{2}$$

The graph embedding aims at seeking low-dimensional representations of the vertexes in high-dimensional space and simultaneously preserving their similarities. The objection function of the similarity preserving criterion in graph embedding framework is depicted as

$$\begin{aligned}
\mathbf{Z}^* &= \arg \min_{\mathbf{Y}^T \mathbf{A} \mathbf{Y} = g} \sum_{i \neq j} \left\| y_i - y_j \right\|^2 S_{ij}^c \\
&= \arg \min_{\mathbf{Y}^T \mathbf{A} \mathbf{Y} = g} \sum_{i \neq j} \left\| \mathbf{Z}^T x_i - \mathbf{Z}^T x_j \right\|^2 S_{ij}^c,
\end{aligned} \tag{3}$$

where g is a constant and $\mathbf{A}$ is a constraint matrix, which may be the Laplacian matrix of a penalty graph.

Class 1

x_i

x_j
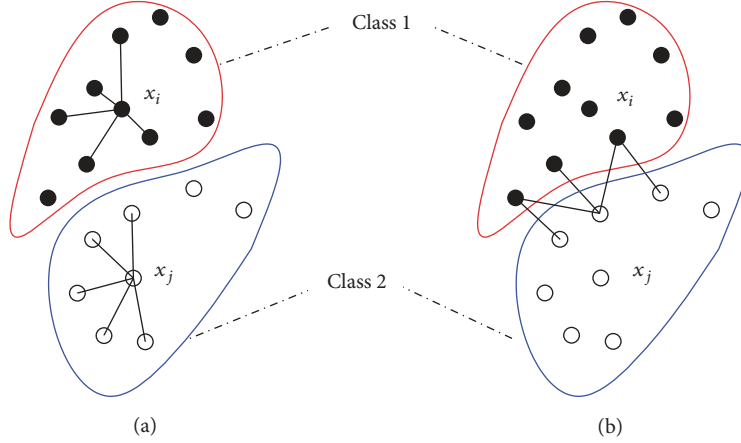
Class 2

x_i

x_j

(a)

(b)

Figure 1: Neighborhood graphs: (a) intrinsic graph and (b) penalty graph.

The linear graph embedding in (3) postulates that the low-dimensional embeddings of the vertexes are linearly projected from the high-dimensional space. Thereby, the kernel trick is applied to the linear graph embedding so as to acquire the nonlinear embedding and fully excavate the inherent geometry structure. Suppose the D-dimensional sample set $\mathbf{X}$ is projected to a reproducing kernel Hilbert space (RKHS) H by a nonlinear mapping ϕ: $\mathbf{X} \subset R^D \rightarrow \phi(\mathbf{X}) \subset H$. The inner product in RKHS is defined as $\langle \phi(s) \cdot \phi(t) \rangle = k(s, t)$. RBF is usually employed as the kernel function $k(s, t)$ and the kernel parameter is the kernel width σ. The elements of the kernel Gram matrix $\mathbf{K}$ is $K_{ij} = k(x_i, x_j)$, and thus the nonlinear graph embedding in RKHS is expressed as

$$\mathbf{Z}_\phi^* = \underset{[\phi(\mathbf{X})]^T \mathbf{Z}_\phi \mathbf{B} \mathbf{Z}_\phi^T \phi(\mathbf{X})=c}{\arg\min} \sum_{i \neq j} \left\| \mathbf{Z}_\phi^T \phi(x_i) - \mathbf{Z}_\phi^T \phi(x_j) \right\|^2 S_{ij}^c. \quad (4)$$

By the reproducing kernel theory, the transformation matrix $\mathbf{Z}_\phi$ is a linear combination of $\phi(x_i)$. Accordingly, there exists a vector $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_N]^T$ satisfying

$$\mathbf{Z}_\phi = \sum_{j=1}^N \alpha_j \phi(x_j) = \phi(\mathbf{X}) \boldsymbol{\alpha}. \quad (5)$$

2.2. Outline of KMFA. On the basis of the label information and local adjacency relationship of samples in high-dimensional space, KMFA defines two neighborhood graphs (both shown in Figure 1) to illustrate the inherent geometrical structure. The vertex pairs are linked in the intrinsic graph if the sample x_i and its k_1 neighbors fall into the same classes. If the data point x_i and its k_2 neighbors belong to different classes, the vertexes in the penalty graph are connected. The intraclass similarity matrix $\mathbf{S}^c \in R^{N \times N}$ of the intrinsic graph is defined as

$$S_{ij}^c = \begin{cases} 1, & \text{if } x_j \text{ is } k_1 \text{ intraclass neighboring point of } x_i \text{ or } x_i \text{ is } k_1 \text{ intraclass neighboring point of } x_j \\ 0, & \text{else.} \end{cases} \quad (6)$$

The interclass similarity matrix $\mathbf{S}^p \in R^{N \times N}$ of the penalty graph is defined as

$$S_{ij}^p = \begin{cases} 1, & \text{if } x_j \text{ is } k_2 \text{ interclass neighboring point of } x_i \text{ or } x_i \text{ is } k_2 \text{ interclass neighboring point of } x_j \\ 0, & \text{else.} \end{cases} \quad (7)$$

The intraclass compactness of the intrinsic graph is depicted by the term

$$\mathbf{S}_c = \sum_{ij} \left\| y_i - y_j \right\|^2 S_{ij}^c = \sum_{ij} \left\| \mathbf{Z}_\phi^\mathrm{T} \phi(x_i) - \mathbf{Z}_\phi^\mathrm{T} \phi(x_j) \right\|^2 S_{ij}^c$$

$$= 2\boldsymbol{\alpha}^\mathrm{T} \mathbf{K} \left(\mathbf{D}^c - \mathbf{S}^c \right) \mathbf{K}\boldsymbol{\alpha}. \tag{8}$$

The interclass dispersibility of the penalty graph is illustrated by the term

$$\mathbf{S}_p = \sum_{ij} \left\| y_i - y_j \right\|^2 S_{ij}^p = \sum_{ij} \left\| \mathbf{Z}_\phi^\mathrm{T} \phi(x_i) - \mathbf{Z}_\phi^\mathrm{T} \phi(x_j) \right\|^2 S_{ij}^p$$

$$= 2\boldsymbol{\alpha}^\mathrm{T} \mathbf{K} \left(\mathbf{D}^p - \mathbf{S}^p \right) \mathbf{K}\boldsymbol{\alpha}. \tag{9}$$

KMFA aims to seek an optimal mapping direction which pulls the intraclass nearest neighbors in the intrinsic graph close and pushes the interclass nearest neighbors in the penalty graph far. Thus, marginal Fisher criterion is maximizing the interclass dispersibility and minimizing the intraclass compactness. The objection function of KMFA algorithm is defined as follows:

$$\boldsymbol{\alpha}^* = \arg\max_{\boldsymbol{\alpha}} \frac{\mathbf{S}_p}{\mathbf{S}_c} = \arg\max_{\boldsymbol{\alpha}} \frac{\boldsymbol{\alpha}^\mathrm{T} \mathbf{K} \left(\mathbf{D}^p - \mathbf{S}^p \right) \mathbf{K}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^\mathrm{T} \mathbf{K} \left(\mathbf{D}^c - \mathbf{S}^c \right) \mathbf{K}\boldsymbol{\alpha}}. \tag{10}$$

By matrix transformation theory, the objective function in (10) can be converted to solve the following generalized maximum eigenvalue decomposition problem:

$$\mathbf{K} \left(\mathbf{D}^p - \mathbf{S}^p \right) \mathbf{K}\boldsymbol{\alpha} = \lambda \mathbf{K} \left(\mathbf{D}^c - \mathbf{S}^c \right) \mathbf{K}\boldsymbol{\alpha}. \tag{11}$$

If the number of features exceeds that of the training samples, the intraclass compactness matrix $\mathbf{S}_c$ may encounter the singular problem. Under the circumstances, the eigenvalue decomposition problem in (11) is ill-posed and could not obtain a stable solution. Consequently, the classical KMFA algorithm (not utilizing PCA preprocessing) firstly embeds the original high-dimensional sample set into a PCA subspace ahead of constructing the two neighborhood graphs. Hence, the final mapping matrix of KMFA algorithm is defined as

$$\mathbf{Z}_{\mathrm{KMFA}} = \mathbf{Z}_{\mathrm{PCA}} \boldsymbol{\alpha}^*, \tag{12}$$

where $\mathbf{Z}_{\mathrm{PCA}}$ is the mapping matrix of the PCA subspace.

2.3. Modified KMFA. In KMFA algorithm, the similarities of the two neighborhood graphs are simultaneously defined to be either 1 or 0, which ignore the label information as well as the distance relationship of samples. Although the intraclass similarity of supervised kernel locality preserving projection (SKLPP) [28] algorithm between vertex pairs is defined to be heat kernel related to the distances of sample points, it involves the kernel width parameter and defines the interclass similarity to be 0. Furthermore, the two kinds of weights make the algorithms prone to be sensitive to the noise and overfit the sample points [29]. So as to enhance the compactness of the samples in the same classes and the dispersibility of the samples from different classes, MKMFA algorithm incorporates the label information and the distance relationship into the similarities of KMFA to guide the construction of the two neighborhood graphs. Motivated by [29], the definitions of the two similarities in MKMFA algorithm are stated below. In the intrinsic graph $\mathbf{R}_c = \{\mathbf{X}, \mathbf{H}^c\}$ of MKMFA algorithm, the entries of the intraclass similarity matrix $\mathbf{H}^c \in R^{N \times N}$ are defined as

$$H_{ij}^c$$

$$= \begin{cases} 1 + \exp\left(-\dfrac{\left\| x_j - x_i \right\|^2}{b}\right), & \text{if } x_j \text{ is } k_1 \text{ intraclass neighboring point of } x_i \text{ or } x_i \text{ is } k_1 \text{ intraclass neighboring point of } x_j \\ 0, & \text{else,} \end{cases} \tag{13}$$

where the parameter b is a regulator, which is equal to the square of the average Euclidean distances between all samples.

In the penalty graph $\mathbf{R}_p = \{\mathbf{X}, \mathbf{H}^p\}$ of MKMFA algorithm, the entries of the interclass similarity matrix $\mathbf{H}^p \in R^{N \times N}$ are illustrated as

$$H_{ij}^p$$

$$= \begin{cases} 1 - \exp\left(-\dfrac{\left\| x_j - x_i \right\|^2}{b}\right), & \text{if } x_j \text{ is } k_2 \text{ interclass neighboring point of } x_i \text{ or } x_i \text{ is } k_2 \text{ interclass neighboring point of } x_j \\ 0, & \text{else.} \end{cases} \tag{14}$$

Because $0 \leq 1 - \exp(-(\|x_j - x_i\|^2/b)) \leq 1$ and $1 \leq 1 + \exp(-(\|x_j - x_i\|^2/b)) \leq 2$, the interclass similarity of MKMFA algorithm is smaller than that of KMFA algorithm and the intraclass similarity of MKMFA algorithm is larger than that of KMFA algorithm. According to marginal Fisher criterion, the higher similarity of the two neighboring samples in the same class will bring about the smaller distance between their corresponding low-dimensional representations. In contrast, the lower similarity of the two neighboring samples from different classes will lead to the larger distance between their corresponding low-dimensional embeddings. Thus, compared to KMFA algorithm, MKMFA algorithm pulls the neighboring sample points in the same classes closer and pushes the neighboring sample points from different classes farther. Additionally, the similarity of the intrinsic graph is larger than that of the penalty graph in MKMFA algorithm. It means two sample points with the same label will have relatively high similarity. On the contrary, two sample points with lower similarity will have more possibility of having different labels. On account of the two similarities that are controlled in a certain range, it results in the goal of the noise suppression. These endow that the two similarities of MKMFA algorithm are helpful for improving the discriminant ability and suppressing the noise.

It is not easy for KMFA to select the best kernel width σ for improving the feature extraction performance. Wang et al. proposed manifold adaptive nonparameter kernel [30], which can well capture the nonlinear property. However, it is not an easy and efficient method to calculate the kernel Gram matrix. Thus, MKMFA algorithm utilizes the data-dependent kernel function [27] constructed by the covariance matrix to reduce the influence of kernel parameter selection on the feature extraction performance. The data-dependent kernel function is described as

$$k\left(x_i, x_j\right) = \exp\left\{-\frac{1}{2}\left(x_i - x_j\right)^{\mathrm{T}} \mathbf{J}^{-1}\left(x_i - x_j\right)\right\}, \quad (15)$$

where $\mathbf{J}$ denotes the covariance matrix of the sample set $\mathbf{X}$.

The intraclass compactness of MKMFA algorithm is depicted as

$$\begin{aligned}
\widetilde{\mathbf{S}}_c &= \sum_{ij} \left\| \left(\mathbf{Z}_\phi\right)^{\mathrm{T}} \phi\left(x_i\right) - \left(\mathbf{Z}_\phi\right)^{\mathrm{T}} \phi\left(x_j\right) \right\|^2 H_{ij}^c \\
&= 2\boldsymbol{\alpha}^{\mathrm{T}} \mathbf{K}\left(\mathbf{O}^c - \mathbf{H}^c\right) \mathbf{K}\boldsymbol{\alpha},
\end{aligned} \quad (16)$$

where the diagonal matrix $\mathbf{O}^c = (O_{ii}^c)_{N \times N} = \sum_{i \neq j} H_{ij}^c$.

The interclass dispersibility of MKMFA algorithm is expressed as

$$\begin{aligned}
\widetilde{\mathbf{S}}_p &= \sum_{ij} \left\| \left(\mathbf{Z}_\phi\right)^{\mathrm{T}} \phi\left(x_i\right) - \left(\mathbf{Z}_\phi\right)^{\mathrm{T}} \phi\left(x_j\right) \right\|^2 H_{ij}^p \\
&= 2\boldsymbol{\alpha}^{\mathrm{T}} \mathbf{K}\left(\mathbf{O}^p - \mathbf{H}^p\right) \mathbf{K}\boldsymbol{\alpha},
\end{aligned} \quad (17)$$

where the diagonal matrix $\mathbf{O}^p = (O_{ii}^p)_{N \times N} = \sum_{i \neq j} H_{ij}^p$.

Therefore, the objection function can be characterized as follows:

$$\boldsymbol{\alpha}^* = \arg\max_{\boldsymbol{\alpha}} \frac{\boldsymbol{\alpha}^{\mathrm{T}} \mathbf{K}\left(\mathbf{O}^p - \mathbf{H}^p\right) \mathbf{K}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^{\mathrm{T}} \mathbf{K}\left(\mathbf{O}^c - \mathbf{H}^c\right) \mathbf{K}\boldsymbol{\alpha}}. \quad (18)$$

Just like KMFA algorithm, (18) also may suffer from singular problem when only a small number of samples are available. In order to obtain good generalization capability and avoid the singular problem, a common approach is mapping the original samples in the high-dimensional space into a PCA subspace ahead of constructing the neighborhood graphs [14, 23]. Although the preprocessing scheme can suppress noise and avoid the singular problem, the unsupervised PCA algorithm does not employ the label information. Thus, the features extracted by KMFA algorithm (utilizing PCA preprocessing) may discard some useful discriminate information in favor of pattern classification. The second method is transforming the ratio form of marginal Fisher criterion into the difference form [30, 31]. The third method is calculating the mapping direction in the null space of the intraclass compactness matrix. Lin et al. proposed Kernel Null Space MFA for face recognition [32]. The above three techniques disregard employing the underlying geometry of samples. Another way is introducing a manifold regularization term so as to deeply exploit the inherent manifold structure. Wei et al. utilized Laplacian penalty function as the regularization term [33]. Regularized KMFA (RKMFA) [34] and semisupervised KMFA (SSKMFA) [35] were proposed to deal with the singular problem and applied to bearing feature extraction. Motivated by [36], marginal Fisher criterion is modified by introducing the underlying manifold structure as the regularization term. It can be described as

$$\begin{aligned}
&\boldsymbol{\alpha}^* \\
&= \arg\max_{\boldsymbol{\alpha}} \frac{\boldsymbol{\alpha}^{\mathrm{T}} \mathbf{K}\left(\mathbf{O}^p - \mathbf{H}^p\right) \mathbf{K}\boldsymbol{\alpha}}{(1 - \beta)\boldsymbol{\alpha}^{\mathrm{T}} \mathbf{K}\left(\mathbf{O}^c - \mathbf{H}^c\right) \mathbf{K}\boldsymbol{\alpha} + \beta\boldsymbol{\alpha}^{\mathrm{T}} \mathbf{K}\left(\mathbf{O}^p - \mathbf{H}^p\right) \mathbf{K}\boldsymbol{\alpha}},
\end{aligned} \quad (19)$$

where $0 \leq \beta \leq 1$ controls the smoothness of the regularization term.

The procedure of MKMFA algorithm is stated below.

Step 1 (constructing two neighborhood graphs). According to the local neighborhood relationship and label information of sample points, MKMFA algorithm constructs the intrinsic graph $\mathbf{R}_c = \{\mathbf{X}, \mathbf{H}^c\}$ and the penalty graph $\mathbf{R}_p = \{\mathbf{X}, \mathbf{H}^p\}$. The intraclass similarity matrix $\mathbf{H}^c \in R^{N \times N}$ of the intrinsic graph is defined in (13) and the interclass similarity matrix $\mathbf{H}^p \in R^{N \times N}$ of the penalty graph is defined in (14).

Step 2 (calculating the kernel Gram matrix). The entries of the kernel Gram matrix $\mathbf{K} \in R^{N \times N}$ are $K_{ij} = k(x_i, x_j)$. Thus, the kernel Gram matrix $\mathbf{K}$ is acquired according to (15).

Step 3 (seeking the optimal projecting direction). The optimal projecting direction $\boldsymbol{\alpha}^*$ in (19) is given by solving the
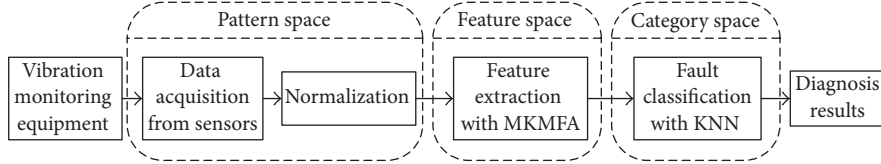
FIGURE 2: The flow chart of the proposed fault diagnosis procedure.

following generalized maximum eigenvalue decomposition problem:

$$\mathbf{K}\left(\mathbf{O}^{p} - \mathbf{H}^{p}\right)\mathbf{K}\boldsymbol{\alpha}$$
$$= \lambda\left((1 - \beta)\mathbf{K}\left(\mathbf{O}^{c} - \mathbf{H}^{c}\right)\mathbf{K} + \beta\mathbf{K}\left(\mathbf{O}^{p} - \mathbf{H}^{p}\right)\mathbf{K}\right)\boldsymbol{\alpha}, \quad (20)$$

where the diagonal matrix $\mathbf{O}^{c} \in R^{N \times N}$ of the intrinsic graph and the diagonal matrix $\mathbf{O}^{p} \in R^{N \times N}$ of the penalty graph are depicted in Section 2.3.

Step 4 (calculating the low-dimensional representations). The low-dimensional embedding of the original high-dimensional signal sample x_i is obtained as

$$y_j = \boldsymbol{\alpha}^{*}\mathbf{K}\left(:, x_i\right), \quad (21)$$

where $\mathbf{K}(:, x_i) = [\mathbf{K}(x_1, x_i), \mathbf{K}(x_2, x_i), \dots, \mathbf{K}(x_N, x_i)]^{\mathrm{T}}$.

There are three parameters to be preset for WKMFA, such as the regularization parameter β, the intraclass neighboring point number k_1, and the interclass neighboring point number k_2. The regularization parameter β is set to be 0.01 by experience. As recommended in [23], the intraclass neighboring point number k_1 is selected as 5. Fivefold cross-validation is employed to select the best interclass neighboring point number k_2, which ranges from 5 to 70 with a step size of 5 [35].

3. Bearing Fault Diagnosis Based on MKMFA

So as to verify the effectiveness of MKMFA algorithm for feature extraction and pattern classification, the MKMFA-based fault diagnosis model is presented and applied to identify various bearing faults.

3.1. The Structure of the Diagnosis System. Rolling bearing fault classification is essentially multiple-manifolds learning problem [37]. From the viewpoint of geometry, the high-dimensional signal samples in the same fault state have the same topology or space distribution, and their low-dimensional embeddings reside on or near a submanifold [38]. On the other hand, the high-dimensional signal samples in different classes have different geometric property, and their low-dimensional representatives are located on different submanifolds. Owing to the fact that MKMFA algorithm simultaneously considers the same submanifold compactness and different submanifolds dispersibility, the MKMFA-based bearing fault diagnosis model was presented. Figure 2 shows the entire process of the proposed scheme.
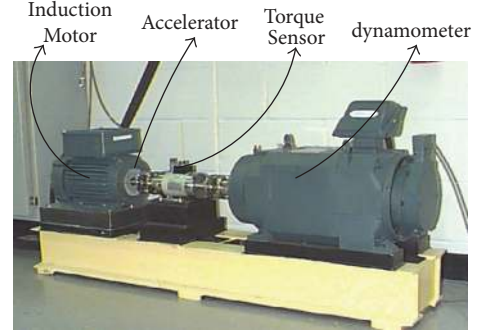

FIGURE 3: The test rig.

The time-series vibration signal series $s(i)$ are collected from the vibration monitoring equipment by sensors. Subsequently, the signal samples $s_i = [s(i), s(i + 1), \dots, s(i + D - 1)]$ are normalized to zero mean and unit variance, where D denotes the feature dimension of signal samples and is equal to the sampling point number of each signal sample. Thus, the signal sample set $\mathbf{X} = [s_1, s_2, \dots, s_N]$ is acquired in high-dimensional pattern space. Via MKMFA algorithm learning the underlying manifold structures of high-dimensional signal samples and excavating the inherent fault information of different submanifolds, the signal samples in high-dimensional pattern space are mapped to a low-dimensional feature space, in which the intraclass nearest neighbors become closer while the interclass nearest neighbors get farther. Thus, the sensitive low-dimensional manifold characteristics related to the fault patterns are extracted from the high-dimensional pattern space and finally inputted into a category space. Thereby, the various fault patterns of rolling bearings are identified by KNN classifier in the category space.

3.2. Vibration Data Acquisition. The experimental data of the rolling bearings are available from the Bearing Data Center [39]. It has been validated in several researches [19, 22, 35, 37, 40] and become a standard data set of bearings. The detailed descriptions of the experimental system are illustrated in [40]. Figure 3 shows the experimental setup, which is composed of a motor, a torque sensor, and a dynamometer controlled to gain different torque load levels. The rotational speed of the motor was ranged from 1730 to 1797 r/min according to different loads (0, 1, 2 and 3 hp). The deep groove ball bearings at the drive end were tested under four kinds of single point faults (normal, ball fault, inner, and outer race fault). Each fault type covers three

TABLE 1: Description of the data sets.

Data set	Training sample size	Testing sample size	Defect size (inches)	Load (hp)	Fault type	Classification label
A	10	90	0.021	1	I	1
	10	90	0.021	2	B	2
	10	90	0.021	3	O	3
	10	90	0	0	N	4
B	Varied from 10 to 90 each class	The remaining samples	0.021	1	I	1
			0.021	2	B	2
			0.021	3	O	3
			0	0	N	4
C	50	50	0.007	1	I	1
	50	50	0.007	1	B	2
	50	50	0.007	1	O	3
	50	50	0.014	2	I	4
	50	50	0.014	2	B	5
	50	50	0.014	2	O	6
	50	50	0.021	3	I	7
	50	50	0.021	3	B	8
	50	50	0.021	3	O	9
	50	50	0	0	N	10

I: inner race fault, B: ball fault, O: outer race fault, and N: normal.

kinds of damage sizes (fault diameter: 0.007, 0.014, and 0.021 inches) with defect depth of 0.011 inches. An accelerometer is installed on the 12 o'clock direction of the motor shell. The data acquisition system contains a high bandwidth amplifier and a 16 channel DAT recorder with low-pass antialiasing filters at the input stage. The vibration signals of the defective bearings were collected at 12000 sampling points per second. One hundred signal samples were acquired for each fault state. Each signal sample consists of 1024 sampling points. So as to evaluate the proposed bearing fault diagnosis scheme, we conducted several experiments over three data subsets (A–C) acquired from the experimental system. Table 1 gives a detailed description of the experimental data sets.

Data set A comprises 400 signal samples covering four kinds of loads (0, 1, 2, and 3 hp) and fault types (normal condition, ball fault, inner, and outer race fault) with the damage size of 0.021 inches. Each operating condition consists of 100 signal samples, which are divided into 10 training samples and 90 testing samples. It is a four-submanifold learning problem corresponding to four kinds of bearing fault categories. The experiments were conducted on data set A to evaluate the feature extraction and fault classification performance of MKMFA algorithm.

Data set B is similar to data set A. Data set B also consists of 400 signal samples, whose operating conditions are identical with those of data set A. But data set B varies the training sample sizes of each class. It increases with a step size of 20 from 10 to 90. Hence, the corresponding remaining samples are used for testing. It is also a four-submanifold learning problem corresponding to four kinds of bearing fault categories. The experiments over data set B were aimed at

assessing the effect of the training sample size on the fault recognition ability of MKMFA algorithm.

Data set C comprises 1000 signal samples involving four kinds of loads (0, 1, 2, and 3 hp) and fault types (normal condition, ball fault, inner, and outer race fault). Each fault type contains three kinds of damage sizes (0.007, 0.014, and 0.021 inches). Each operating condition consists of 100 signal samples, which are split into 50 training samples and 50 testing samples. It is a ten-submanifold learning problem corresponding to ten kinds of bearing fault severities. The purpose of performing the experiment on data set C is to further investigate the fault classification performance of the proposed fault diagnosis scheme under complicated operating conditions.

3.3. Feature Extraction and Pattern Classification. So as to evaluate the effectiveness and exhibit the superiority of MKMFA algorithm for bearing feature extraction and fault identification, we conducted several experiments on the three data sets in Table 1 and made a comparison with KPCA, KFDA, KMFA, SKLPP, RKMFA, and SSKMFA. The feature dimension of each signal sample in the three data sets is 1024, which is larger than their training sample sizes.

3.3.1. Bearing Fault Categories Identification. An investigation was performed on data set A to evaluate the feature extraction performance of MKMFA algorithm. The low-dimensional features are directly extracted from the high-dimensional pattern space by utilizing the six feature extraction methods. The first two mapping results of these methods
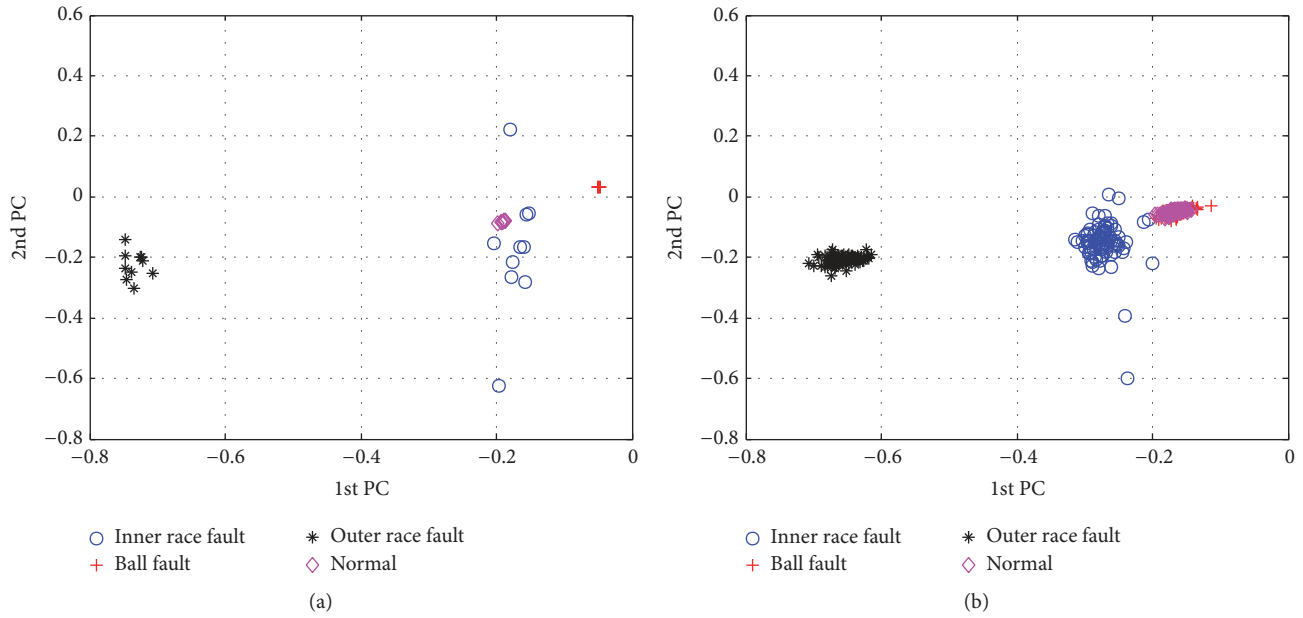
FIGURE 4: Feature extraction with KMFA: (a) training set and (b) testing set.
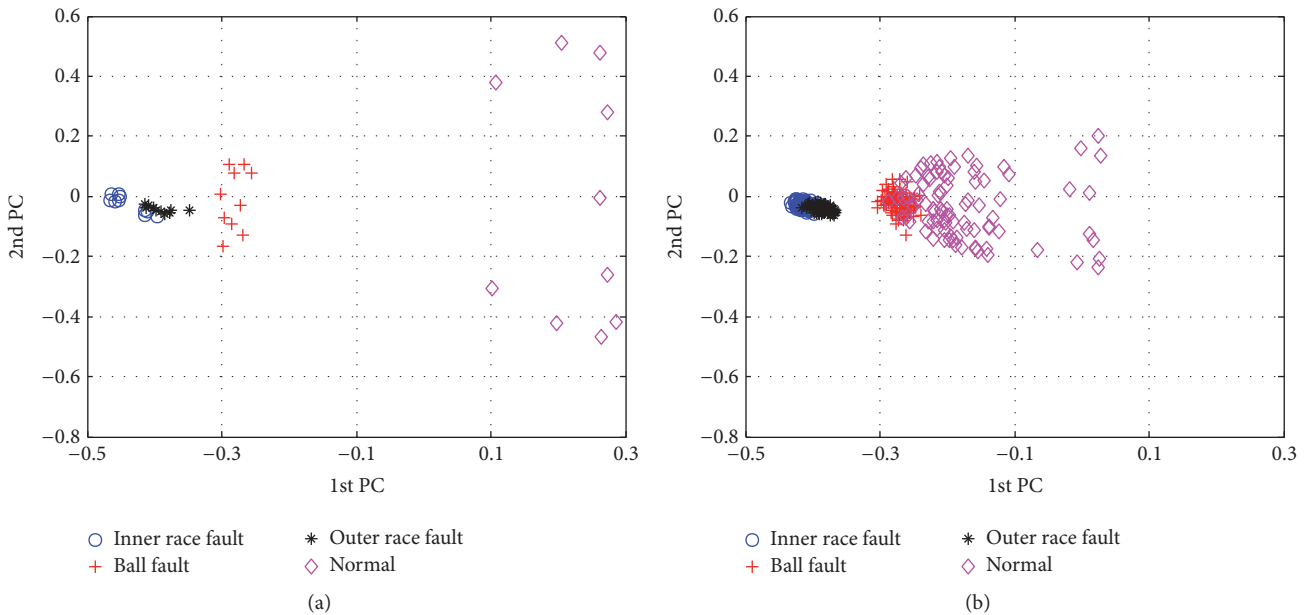

FIGURE 5: Feature extraction with KPCA: (a) training set and (b) testing set.

are plotted for intuitional display. Figure 4 reveals that the clustering result of the training set with KMFA is much dispersed for inner race fault, so that it brings about the crossing area between normal condition and inner race fault. Except outer race fault, KMFA could not clearly separate the other three faults for the testing set. Figure 5 exhibits that the clustering and separate performances of KPCA features are not so good because of the mixed inner and outer race fault. Figures 6 and 7 show that both KFDA and SKLPP can distinguish each type of fault and have a good clustering result for the training set. Nevertheless, KFDA could not completely

recognize two types of faults for the testing set because of the overlap region between normal condition and outer race fault. Additionally, the testing set of normal condition with SKLPP is not well clustered. Figure 8 demonstrates that the classification boundary of normal condition and ball fault is not very clear for the RKMFA features of the testing set. Additionally, the two types of faults overlap together for the RKMFA features of the training set. Figure 9 displays that MKMFA has better feature extraction performance over the above five methods, as it yields better clustering effect and more clear separation boundaries. In comparison with KMFA
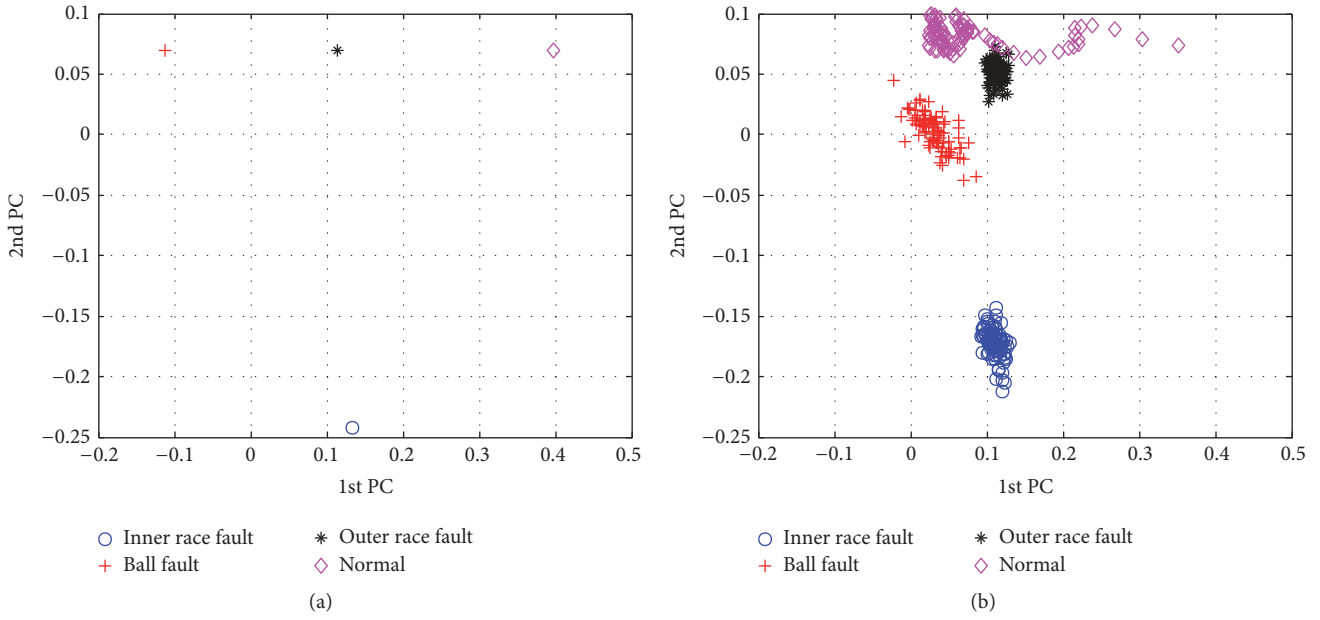
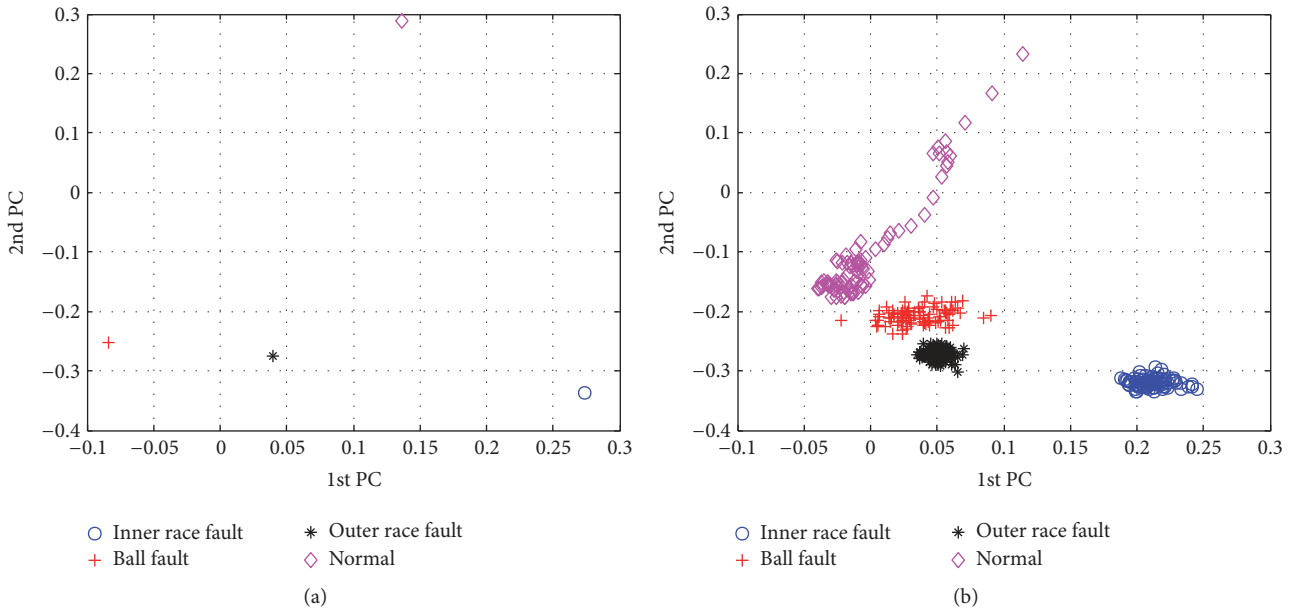FIGURE 6: Feature extraction with KFDA: (a) training set and (b) testing set.

FIGURE 7: Feature extraction with SKLPP: (a) training set and (b) testing set.

and RKMFA algorithm, MKMFA pulls the neighboring sample points in the same classes closer and pushes the neighboring sample points from different classes farther. The reason is that the two similarities of MKMFA algorithm are weighted. On the grounds of the above experimental results, it is demonstrated that MKMFA is able to enhance the intraclass compactness and interclass dispersibility. Compared with the other five feature extraction techniques, MKMFA is more effective to capture the sensitive low-dimensional manifold characteristics related to the nature of different bearing faults.

For objectively assessing the fault classification performance of the six feature extraction approaches, the low-dimensional mapping results of them are fed into KNN classifier as the final evaluation criteria. Table 2 displays their recognition accuracies and the corresponding parameter settings. It can be seen from Table 2 that the feature dimension of MKMFA ($d = 5$) is lower than that of KPCA ($d = 20$). Nevertheless, the classification performances of the former (100% for ball fault, normal condition, and inner race fault) surpass those of the latter (6.67%, 13.34% and 80%, resp.). It results from the fact that KPCA does not take advantage of
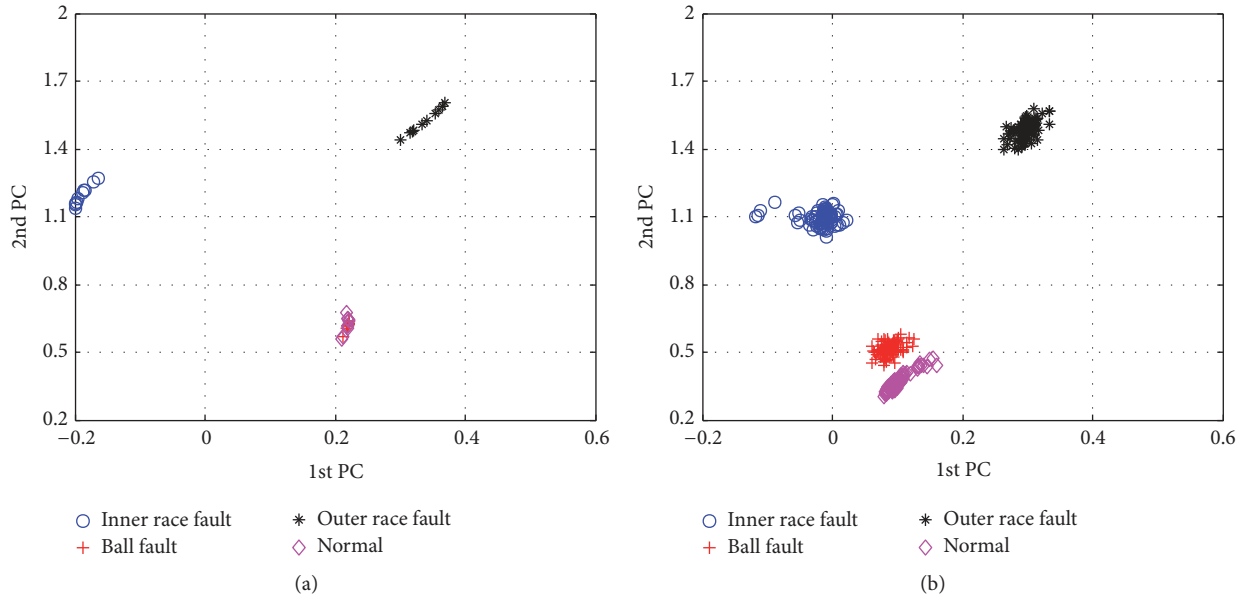
FIGURE 8: Feature extraction with RKMFA: (a) training set and (b) testing set.
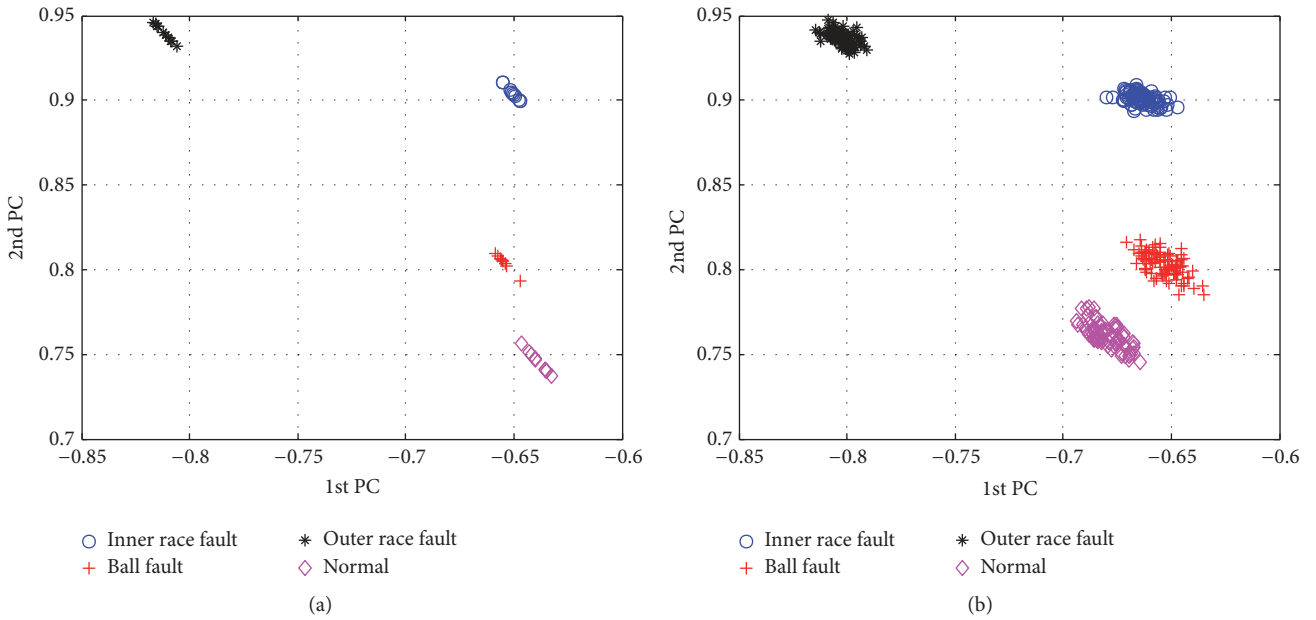

FIGURE 9: Feature extraction with MKMFA: (a) training set and (b) testing set.

TABLE 2: The recognition rates (%) of KNN classifier based on six feature extraction approaches.

| Fault type | Methods | | | | | |
	KMFA $(\sigma = 5, d = 6)$	KPCA $(d = 20, \sigma = 5)$	KFDA $(d = 3, \sigma = 3)$	SKLPP $(d = 5, \sigma = 3)$	RKMFA $(\sigma = 3, d = 10)$	MKMFA $(d = 5)$
Inner race fault	98.89	80.00	100	100	100	100
Ball fault	6.67	6.67	97.78	100	100	100
Outer race fault	100	100	100	100	100	100
Normal	100	13.34	15.56	13.34	88.89	100

TABLE 3: The average recognition rates (%) of KNN for various training sample sizes each class.

Training sample size	KPCA ($d = 20, \sigma = 5$)	KFDA ($d = 3, \sigma = 3$)	SKLPP ($d = 5, \sigma = 3$)	KMFA ($\sigma = 5$) (k_1, k_2, d)	RKMFA (σ, d)	MKMFA (d)
10	56.11	78.33	81.11	63.61 (5, 10, 6)	97.22 (3, 10)	98.93 (5)
30	82.14	79.29	89.17	75 (10, 20, 7)	97.86 (3, 6)	99.29 (4)
50	87.00	81.00	91.50	88.5 (5, 40, 11)	99.64 (8, 5)	100.00 (4)
70	96.67	87.50	100.00	100 (10, 40, 15)	100.00 (5, 6)	100.00 (5)
90	100.00	100.00	100.00	100 (5, 10, 5)	100.00 (5, 6)	100.00 (4)

any class information related to fault classification. Compared with KFDA, the identification accuracies of MKMFA (100% for normal condition and ball fault) increase by 84.44% and 2.22%, respectively. Although KFDA is supervised, it cannot excavate the underlying manifold structure. These results also indicate that the local structure information extracted by MKMFA could be more effective than the global feature information of the Euclidean space extracted by KPCA and KFDA. The classification performance of the proposed approach outperforms those of KMFA-based and SKLPP-based fault diagnosis methods. The reason is that KMFA and SKLPP lose some useful discriminant information by using PCA as preprocessing although they are able to capture the manifold structure. Compared with RKMFA, MKMFA employs fewer features to achieve better diagnosis results. Hence, the proposed approach is effective to extract the most sensitive low-dimensional manifold characteristics beneficial to fault classification. The reason is that the proposed approach effectively makes use of the class information and the underlying geometric structure of faulty samples. On the other hand, the modified intraclass and interclass similarities are helpful for deeply exploiting the underlying manifold structure.

We used data set B to assess the influence of the training sample sizes on the recognition rates of different feature extraction methods. The training samples are selected randomly. Each experiment is conducted by ten trials in the following experiments. Table 3 and Figure 10 display the average recognition rates of the six feature extraction methods with different training sample sizes per class. The neighboring parameters of MKMFA are set to $k_1 = 5$ and $k_2 = 10$; those of RKMFA are set to $k_1 = 5$ and $k_2 = 20$. It is indicated that the recognition performances are improved as the training samples increase. This is because overfitting is less likely to occur when more training samples are available for KPCA, KFDA, and KMFA. The classification accuracies of KPCA, KFDA, KMFA, and SKLPP are much lower than that of MKMFA as the training sample number is equal to 10. By comparison, the classification rates based on RKMFA and MKMFA hold the least fluctuation when varying the training sample number. It results from the fact that RKMFA and MKMFA introduce the regularization term incorporating the intrinsic manifold structure to reduce the effect of the insufficient training samples. In comparison with RKMFA, MKMFA utilizes few features to achieve higher diagnosis accuracies. Therefore, MKMFA has the best classification performance even though the training sample number
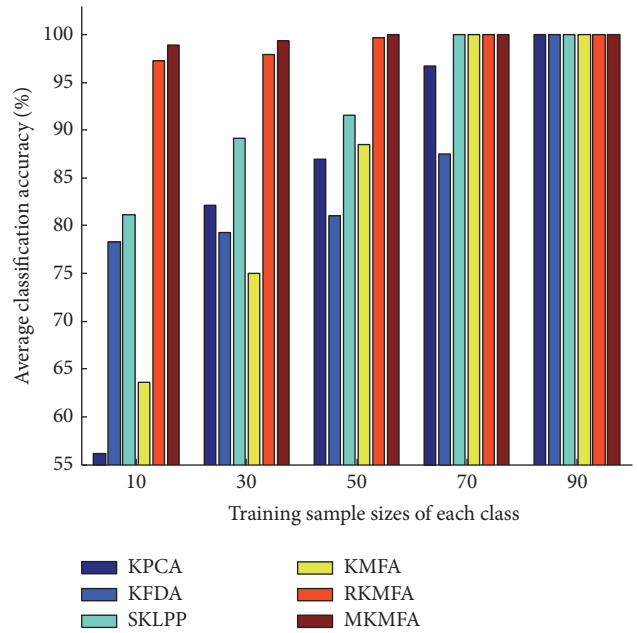

FIGURE 10: The comparison of the average accuracies for various training sample sizes.

is small compared with the other five feature extraction methods.

For KMFA, RKMFA, and MKMFA algorithm, k_1 neighboring points dominate the intraclass compactness and k_2 neighboring points govern the interclass dispersibility. Hence, the two neighboring points number roles are critical for the construction of the two neighborhood graphs and the subsequent diagnosis assignment. Several experiments were implemented by changing the number of the two neighboring points. The training and testing sample sizes per class are both set to be 50. As illustrated in Figure 11(a), the classification accuracies based on KMFA fluctuate below 90% when the values of k_2 are small. Figure 11(b) reveals that the identification accuracies of RKMFA have the relatively small fluctuations. Figure 11(c) shows that the classification rates of MKMFA are stable and maintained at a high level while varying the two neighboring points number. Compared with KMFA and RKMFA, MKMFA-based bearing fault diagnosis approach is robust and convenient in virtue of without making great effort to tune the two nearest neighbors' number.
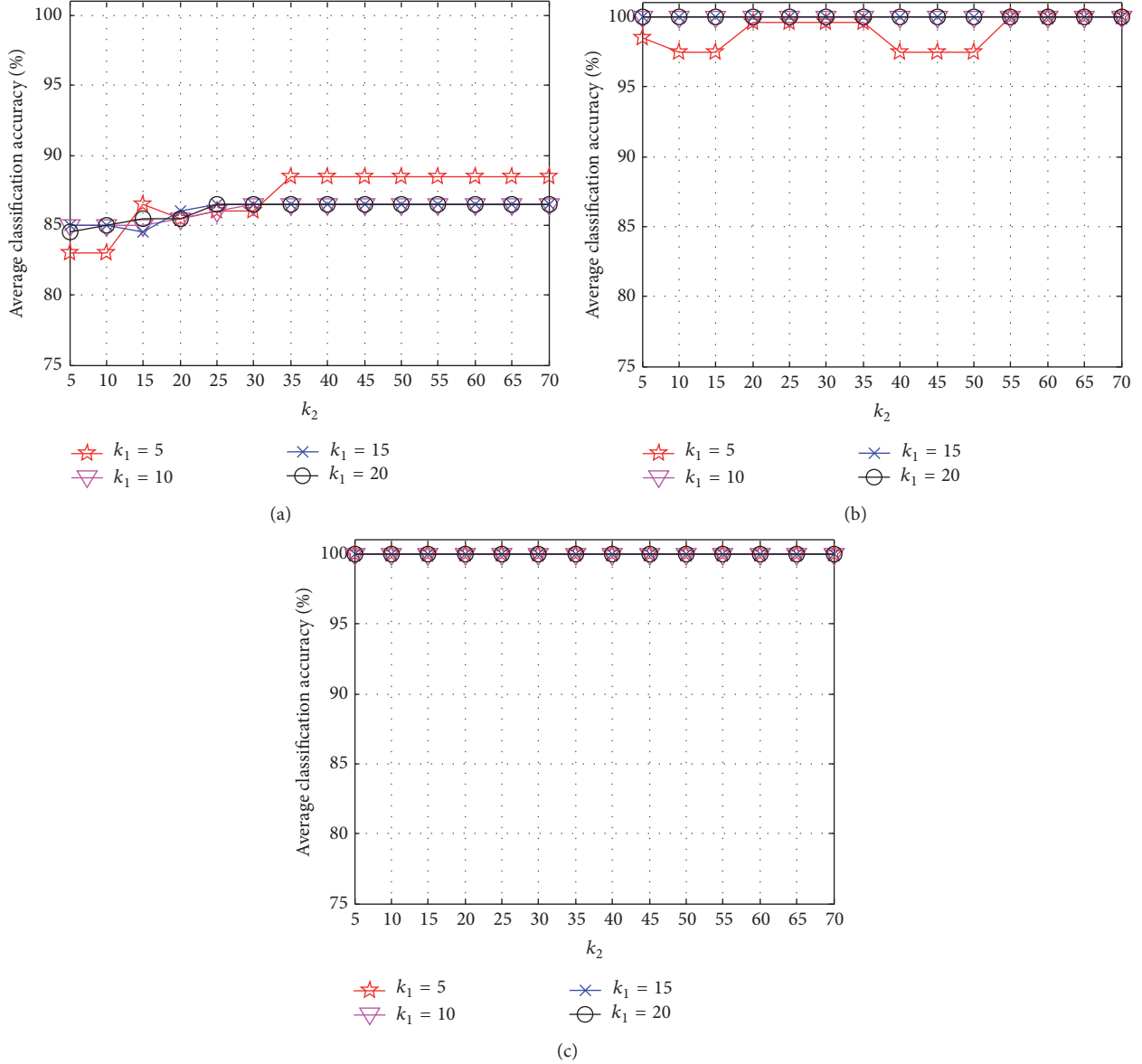
FIGURE 11: The classification accuracies of (a) KMFA, (b) RKMFA, and (c) MKMFA.

The parameter β controls the smoothness of the regularization term in MKMFA algorithm. As the training sample sizes in each class are 10, 50, and 90, the effect of the parameter β on the recognition rates is illustrated in Figure 12. As can be seen, the classification performances of KMFA are superior to those of MKMFA and RKMFA with β as zero. It stems from the fact that under the circumstances, MKMFA and RKMFA are exactly the classical KMFA without mapping the original samples to a PCA subspace beforehand. Except the parameter β which is equal to zero, the recognition rates of MKMFA are higher than those of KMFA for different training sample sizes. Thus, the regularization term of MKMFA can improve the diagnosis performance of KMFA. Compared with RKMFA, the classification accuracies of MKMFA hold the relatively smaller fluctuations for different β. It reveals that it is not very difficult for MKMFA to

select the best parameter β for enhancing its classification capability.

3.3.2. Bearing Fault Severities Identification. The experiments were conducted on data set C to recognize ten types of bearing fault severity conditions. So as to quantitatively describe the superiority of the six feature extraction methods, the definition of within-class scatter and between-class scatter is [9]

$$\mathbf{S}_w = \sum_{l=1}^{N_c} \sum_{f_k \in l} \left(f_k - \mu_f^l \right) \left(f_k - \mu_f^l \right)^{\mathrm{T}}$$

$$\mathbf{S}_b = \sum_{l=1}^{N_c} \left(\mu_f^l - \overline{\mu}_f \right) \left(\mu_f^l - \overline{\mu}_f \right)^{\mathrm{T}}, \tag{22}$$
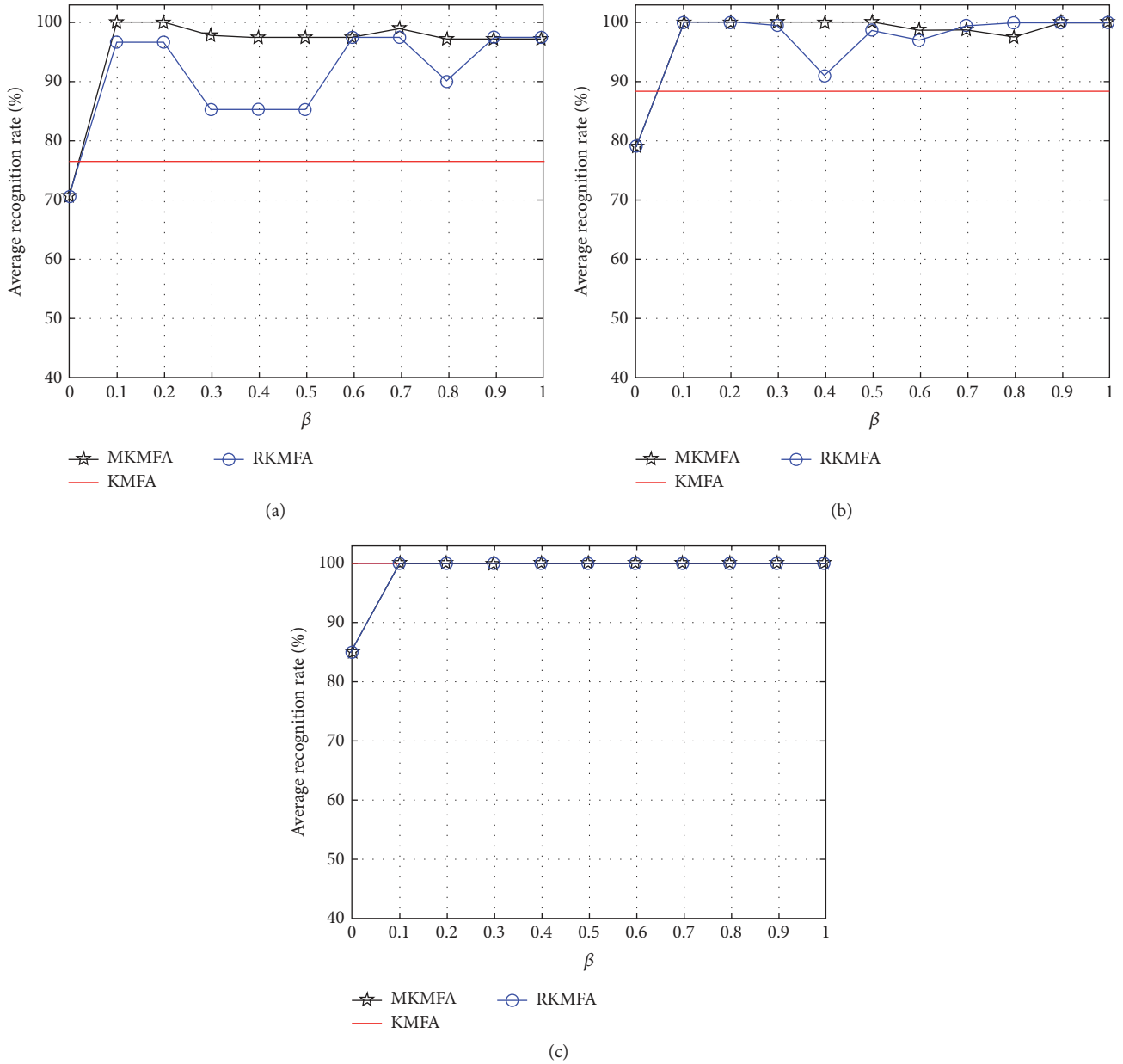
FIGURE 12: The average recognition rate as the training sample sizes are (a) 10, (b) 50, and (c) 90.

where f_k $(k = 1, 2, \ldots, d)$ is the feature vector, μ_f^l is the mean of the feature vectors in the lth class, and $\overline{\mu}_f$ is the mean of all feature vectors.

The within-class scatter $\mathbf{S}_w$ describes the compactness of the samples in the same classes and the between-class scatter $\mathbf{S}_b$ characterizes the dispersibility of the samples from different classes. Thus, it is beneficial to fault classification for smaller $\mathbf{S}_w$ values and bigger $\mathbf{S}_b$ values. Table 4 shows the two parameter values and the classification accuracies of the six feature extraction methods based on KNN classifier. The parameters of KMFA, SSKMFA, and MKMFA are set as $k_1 = 5$ and $k_2 = 10$. As can be seen, MKMFA features have the smallest $\mathbf{S}_w$ value and the biggest $\mathbf{S}_b$ value compared with the

other five feature extraction schemes. It reveals that MKMFA has the best clustering property and classification capability in comparison with KPCA, KFDA, KMFA, SSKMFA, and SKLPP. Compared with the other five supervised feature extraction techniques, KPCA has the lowest recognition rate due to discarding some useful discriminate information in favor of pattern classification. Additionally, MKMFA algorithm based on KNN classifier achieves the highest identification rate in all the six feature extraction techniques. The reason is that MKMFA algorithm employs the class label information and the distances relationship of sample points to guide the construction of local neighborhood graphs. Consequently, the discrimination performance of

TABLE 4: Performance comparisons for various feature extraction approaches.

Methods	S_b	S_w	Average accuracy (%)
KMFA ($d = 12, \sigma = 3$)	1.455	3.11×10^{-3}	85.36
KPCA ($d = 20, \sigma = 5$)	0.4329	0.7747	73.20
KFDA ($d = 9, \sigma = 1$)	2.238	3.50×10^{-3}	89.63
SKLPP ($d = 10, \sigma = 5$)	3.086	1.24×10^{-3}	91.27
MKMFA ($d = 11, \beta = 0.01$)	6.501	8.71×10^{-5}	97.45
SSKMFA ($d = 11, \sigma = 1, \beta = 0.1$)	3.473	4.59×10^{-4}	95.39

TABLE 5: The average recognition rates (%) of the simple features based on KNN.

Case 1: fault categories identification					Case 2: fault severities identification
Training sample size of each class					
10	30	50	70	90	
68.89	94.29	98.33	99.67	99.95	53.25

the obtained low-dimensional manifold features extracted by MKMFA algorithm can be further improved when focusing on classification assignment.

3.4. Discussions. Compared to KPCA, KFDA, KMFA, RKMFA, SSKMFA, and SKLPP, the above experimental results demonstrate that MKMFA algorithm has remarkable superiority. The main reasons are as follows. Firstly, MKMFA algorithm employs the discriminant information and local neighborhood relationship of signal samples to construct the two neighborhood graphs. Secondly, it incorporates the class label information and distance relationship of signal samples into the two similarities and thus further enhances the intraclass compactness and interclass dispersibility. Thirdly, it introduces a manifold regularization term to cope with the singular problem and employs the nonparametric kernel function to reduce the influence of kernel parameter selection on feature extraction performance. As a result, the advanced low-dimensional manifold characteristics extracted by MKMFA algorithm are related to the nature of bearing fault patterns by excavating the inherent manifold structures of different submanifolds.

Some simple feature extraction methods are widely applied to bearing fault diagnosis. Hence, we employed the simple features, including ten time-domain statistical features and six EMD energy entropies as illustrated in [35], to analyze the above two cases for comparison. Table 5 shows the average recognition rates of the 16 simple features based on KNN classifier. In comparison with Figure 10 and Tables 3 and 4, the simple features have lower recognition rates than MKMFA features for the two cases. The reason is that the MKMFA algorithm extracts the sensitive low-dimensional manifold characteristics related to fault patterns by learning the underlying manifold structures of high-dimensional signal samples. However, the simple feature extraction methods only give attention to some specific contents of faulty signals. Thus, it is necessary to explore the advanced feature extraction methods to improve fault classification performance, which is the goal of our study.

4. Conclusions

This paper presents improved kernel marginal Fisher analysis (MKMFA) algorithm for feature extraction with dimensionality reduction, which employs the label information and distances relationship of faulty samples, introduces a manifold regularization term, and utilizes the data-dependent kernel function. MKMFA effectively extracts the optimal low-dimensional manifold characteristics from the time-series signal samples in high-dimensional ambient space. It is efficient to transform the complicated two-stage (feature extraction and dimensionality reduction) procedure into a relatively simple one-step process, which boils down to the generalized maximum eigenvalue decomposition problem. Compared with KPCA, KFDA, KMFA, RKMFA, and SKLPP, the feature extraction experiments on four categories of bearing faults reveal that our proposed feature extraction scheme is more effective to capture the sensitive low-dimensional manifold characteristics beneficial to pattern classification due to its good clustering and separation properties. The feature evaluation experiments on ten types of bearing fault severities show its superiority in comparison with KPCA, KFDA, KMFA, SSKMFA, and SKLPP. Based on MKMFA algorithm, a fault diagnosis model is presented and applied to identify different bearing faults. When varying the training sample sizes in the four-fault-type comparison experiments, it is demonstrated that the classification performances of MKMFA are significantly improved even for insufficient training samples. The ten-fault-severity comparison experiments of rolling bearings exhibit its outstanding fault recognition capability compared with the other five feature extraction methods. It is robust and easily applied to bearing fault classification without great effort to tune the parameters in MKMFA. The proposed diagnosis scheme has confirmed its effectiveness of recognizing bearing faults and can be easily applied to fault diagnosis of other components as well.

Competing Interests

Acknowledgments

References

[1] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 2006.

[2] M. Grasso, S. Chatterton, P. Pennacchi, and B. Colosimo, "A data-driven method to enhance vibration signal decomposition for rolling bearing fault analysis," *Mechanical Systems and Signal Processing*, vol. 81, pp. 126–147, 2016.

[3] R. Liu, B. Yang, X. Zhang, S. Wang, and X. Chen, "Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 75, pp. 345–370, 2016.

[4] H. Al-Bugharbee and I. Trendafilova, "A fault diagnosis methodology for rolling element bearings based on advanced signal pretreatment and autoregressive modelling," *Journal of Sound and Vibration*, vol. 369, pp. 246–265, 2016.

[5] L. Jiang, T. L. Shi, and J. P. Xuan, "Fault diagnosis of rolling bearings based on Marginal Fisher analysis," *Journal of Vibration and Control*, vol. 20, no. 3, pp. 470–480, 2014.

[6] Z. Xu, J. Xuan, T. Shi, B. Wu, and Y. Hu, "A novel fault diagnosis method of bearing based on improved fuzzy ARTMAP and modified distance discriminant technique," *Expert Systems with Applications*, vol. 36, no. 9, pp. 11801–11807, 2009.

[7] Y. Lv, R. Yuan, and G. Song, "Multivariate empirical mode decomposition and its application to fault diagnosis of rolling bearing," *Mechanical Systems and Signal Processing*, vol. 81, pp. 219–234, 2016.

[8] T. Chen, G. Y. Tian, A. Sophian, and P. W. Que, "Feature extraction and selection for defect classification of pulsed eddy current NDT," *NDT & E International*, vol. 41, no. 6, pp. 467–476, 2008.

[9] Q. He, "Time-frequency manifold for nonlinear feature extraction in machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 200–218, 2013.

[10] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Muller, "Fisher discriminant analysis with kernels," in *Proceedings of the 9th IEEE Workshop on Neural Networks for Signal Processing (NNSP '99)*, pp. 41–48, Madison, Wis, USA, August 1999.

[11] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[12] Z. Liu, J. Qu, M. J. Zuo, and H.-B. Xu, "Fault level diagnosis for planetary gearboxes using hybrid kernel feature selection and kernel Fisher discriminant analysis," *International Journal of Advanced Manufacturing Technology*, vol. 67, no. 5-8, pp. 1217–1230, 2013.

[13] R. Shao, W. Hu, Y. Wang, and X. Qi, "The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet transform," *Measurement*, vol. 54, pp. 118–132, 2014.

[14] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.

[15] D. Cai, X. F. He, K. Zhou, J. W. Han, and H. J. Bao, "Locality sensitive discriminant analysis," in *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI '07)*, pp. 714–719, Hyderabad, India, January 2007.

[16] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[17] H. S. Seung and D. D. Lee, "The manifold ways of perception," *Science*, vol. 290, no. 5500, pp. 2268–2269, 2000.

[18] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[19] B. Li and Y. Zhang, "Supervised locally linear embedding projection (SLLEP) for machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 25, no. 8, pp. 3125–3134, 2011.

[20] S. Dong, L. Chen, B. Tang, X. Xu, Z. Gao, and J. Liu, "Rotating machine fault diagnosis based on optimal morphological filter and local tangent space alignment," *Shock and Vibration*, vol. 2015, Article ID 893504, 9 pages, 2015.

[21] Y. Wang, G. Xu, L. Liang, and K. Jiang, "Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 54-55, pp. 259–276, 2015.

[22] W. Li, S. Zhang, and S. Rakheja, "Feature denoising and nearest-farthest distance preserving projection for machine fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 1, pp. 393–404, 2016.

[23] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–51, 2007.

[24] A. Maronidis, A. Tefas, and I. Pitas, "Subclass graph embedding and a marginal fisher analysis paradigm," *Pattern Recognition*, vol. 48, no. 12, pp. 4024–4035, 2015.

[25] D. Xu, S. Yan, D. Tao, S. Lin, and H.-J. Zhang, "Marginal Fisher analysis and its variants for human gait recognition and content-based image retrieval," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2811–2821, 2007.

[26] J. B. Li, Y. H. Wang, S. C. Chu, and J. F. Roddick, "Kernel self-optimization learning for kernel-based feature extraction and recognition," *Information Sciences*, vol. 257, no. 2, pp. 70–80, 2014.

[27] J. Wang, J. T. Kwok, H. C. Shen, and L. Quan, "Data-dependent kernels for high-dimensional data classification," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '05)*, pp. 102–107, Montreal, Canada, August 2005.

[28] J. Cheng, Q. Liu, H. Lu, and Y.-W. Chen, "Supervised kernel locality preserving projections for face recognition," *Neurocomputing*, vol. 67, pp. 443–449, 2005.

[29] W. K. Wong and H. T. Zhao, "Supervised optimal locality preserving projection," *Pattern Recognition*, vol. 45, no. 1, pp. 186–197, 2012.

[30] Z. Wang, X. Sun, L. Sun, and Y. Huang, "Semisupervised kernel marginal fisher analysis for face recognition," *The Scientific World Journal*, vol. 2013, Article ID 981840, 13 pages, 2013.

[31] J. He, L. Ding, L. Jiang, and L. Huang, "A flexible and efficient algorithm for regularized Marginal Fisher analysis," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '14)*, pp. 4198–4205, IEEE, Beijing, China, July 2014.

[32] Y. E. Lin, X. Z. Liang, and H. P. Zhou, "Kernel null space marginal fisher analysis for face recognition," in *Proceedings of the 4th International Conference on Advances in Materials and Manufacturing (ICAMMP '13)*, pp. 1065–1068, Kunming, China, December 2013.

[33] Z. Wei, C. Liu, and N. Li, "Weighted Marginal Fisher analysis with Spatially Smooth for aircraft recognition," *Chinese Journal of Aeronautics*, vol. 27, no. 1, pp. 110–116, 2014.

[34] L. Jiang, T. L. Shi, and J. P. Xuan, "An intelligent fault diagnosis method of rolling bearings based on regularized kernel Marginal Fisher analysis," in *Proceedings of the 25th International Congress on Condition Monitoring and Diagnostic Engineering (COMADEM '12)*, Huddersfield, UK, June 2012.

[35] L. Jiang, J. Xuan, and T. Shi, "Feature extraction based on semi-supervised kernel Marginal Fisher analysis and its application in bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 41, no. 1-2, pp. 113–126, 2013.

[36] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition," *Pattern Recognition Letters*, vol. 26, no. 2, pp. 181–191, 2005.

[37] M. Li, J. Xu, J. Yang, D. Yang, and D. Wang, "Multiple manifolds analysis and its application to fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 23, no. 8, pp. 2500–2509, 2009.

[38] Q. Jiang, M. Jia, J. Hu, and F. Xu, "Machinery fault diagnosis using supervised manifold learning," *Mechanical Systems and Signal Processing*, vol. 23, no. 7, pp. 2301–2311, 2009.

[39] "Case Western Reserve University bearing data center," 2015, http://csegroups.case.edu/bearingdatacenter/home/.

[40] K. A. Loparo and X. S. Lou, "Bearing fault diagnosis based on wavelet transform and fuzzy inference," *Mechanical Systems and Signal Processing*, vol. 18, no. 5, pp. 1077–1095, 2004.