Research Article

Machine Fault Classification Based on Local Discriminant Bases and Locality Preserving Projections

Qingbo He,¹ Xiaoxi Ding,¹ and Yuanyuan Pan²

¹ Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China, Hefei, Anhui 230026, China

² Anhui Vocational College of City Management, Hefei, Anhui 231635, China

Correspondence should be addressed to Qingbo He; qbhe@ustc.edu.cn

Received 27 April 2014; Accepted 1 June 2014; Published 26 June 2014

Academic Editor: Weihua Li

Copyright © 2014 Qingbo He et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine fault classification is an important task for intelligent identification of the health patterns for a mechanical system being monitored. Effective feature extraction of vibration data is very critical to reliable classification of machine faults with different types and severities. In this paper, a new method is proposed to acquire the sensitive features through a combination of local discriminant bases (LDB) and locality preserving projections (LPP). In the method, the LDB is employed to select the optimal wavelet packet (WP) nodes that exhibit high discrimination from a redundant WP library of wavelet packet transform (WPT). Considering that the obtained discriminatory features on these selected nodes characterize the class pattern in different sensitivity, the LPP is then applied to address mining inherent class pattern feature embedded in the raw features. The proposed feature extraction method combines the merits of LDB and LPP and extracts the inherent pattern structure embedded in the discriminatory feature values of samples in different classes. Therefore, the proposed feature not only considers the discriminatory features themselves but also considers the dynamic sensitive class pattern structure. The effectiveness of the proposed feature is verified by case studies on vibration data-based classification of bearing fault types and severities.

1. Introduction

Machine fault classification is an important task for intelligent identification of the condition patterns for the system being monitored. For a mechanical system, vibration monitoring is often employed to evaluate the system dynamics. A specific application considered in this paper is to monitor health condition of a machine or its components, such as bearings, for timely identifying possible faults, which is increasingly significant to reduce machine downtime and ensure high productivity. Once a fault happens in a machine, it makes sense to identify the fault type or the fault severity through vibration data analysis so that time and safety can be guaranteed. There are many reasons leading to machine failures. For instance, poor lubrication, acid corrosion, and plastic deformation could cause the bearing to work in an abnormal condition, respectively [1]. In addition, typical damages of the rolling bearing are located at outer raceway, inner raceway, or rolling element. To effectively monitor and

recognize the machine condition, the major challenge is to extract reliable features from vibration data which are often disturbed by the environment noise. The traditional features, such as the time-domain features and the frequency-domain ones, are often applied to fault diagnosis [2–6]. However, the pattern of vibration signals demonstrates many nonlinear characteristics and the methods mentioned above cannot extract these nonlinear features effectively for classifying fault types and severities. Therefore, this study intends to find a good feature representation of the raw signals that yields higher discriminatory information.

Wavelet transform has the ability to well express the nonstationary signals and represent sensitive features with its multiresolution capability, which has achieved a great success in fault classification [7]. As one of the most widely used wavelet transform methods, the wavelet packet transform (WPT) is well-known for its orthogonal, complete, and local property [8]. WPT leads to a redundant binary tree of a signal with a set of time-frequency subspaces each of which is made

up of a wavelet packet (WP) base vector. The whole subspaces are called a WP library. As we know, different WP bases give rise to different representation of a given signal. Thus, it is important to select optimal WP bases out of the whole WP library for enhanced signal analysis or classification. For classification, the main objective is to find an optimal set of WP nodes that yield high discriminatory information for discriminating different classes as much as possible. This can be realized by the local discriminant bases (LDB) [9]. The algorithm identifies optimal LDBs with high discriminatory information by using a dissimilarity measure on the given dataset. Many related works have been conducted in the last two decades to demonstrate the effectiveness of the LDB to achieve a good classification through selecting the optimal WP bases among various redundant WP subspaces [9–18]. Although the discriminatory features can be obtained by selecting WP nodes via the LDB, different node displays different sensitivity in characterizing class information. In the machine learning-based classification approach, the classification accuracy will mainly depend on the sensitive features. Current methods mainly employ the most sensitive bases for classification. We focus on another approach, which is using the dimensionality reduction techniques to mine more sensitive features in the whole set of discriminatory features by LDB. Therefore, in this study, one challenge is how to extract the most useful and sensitive information hidden in high-dimensional data based on the selected WP nodes.

In the past few decades, many useful dimensionality reduction techniques have been employed for fault diagnosis and classification [2, 3, 19–27]. These techniques can be simply divided into two types: linear and nonlinear approaches. Linear dimensionality reduction aims to find a set of low-dimensional bases from high-dimensional data through the linear transformation. Two of the well-known linear learning methods are principal component analysis (PCA) [19, 20] and linear discriminant analysis (LDA) [21, 22]. The other type, nonlinear dimensionality reduction, searches for nonlinear structure hidden in high-dimensional data. There are two main nonlinear approaches including kernel-based techniques [2, 23, 24] and manifold learning techniques [25–27]. Manifold learning pursuits the goal to embed data that originally lies in a high-dimensional space in a lower dimensional space while preserving local characteristic properties, for example, local geometric property (Isomap [28]), local embedding structure (LLE [29]), local adjacency relations (LE [30]), and local tangent space information (LTSA [31]). Although these nonlinear manifold learning methods have been effectively developed to machine fault classification, they need heavy computation cost and are complex to be extended for fault classification of a new data [25–27, 32]. He and Niyogi [33] proposed a new linear model, locality preserving projections (LPP), which can reveal the nonlinear manifold structure embedded in the dataset with a kernel that maintains the local information. LPP is provided with the remarkable superiority that it can form an explicit map to the manifold learning algorithm, which is linear and easily operational. Some works have indicated that LPP is beneficial to feature extraction in machine fault classification [3, 24]. Hence, the LPP is employed in this study to extract

the sensitive information hidden in the raw feature data from the selected WP nodes.

In this paper, based on energy features of the nodes selected by LDB algorithm from the WP library, a new effective feature is proposed to mine the nonlinear pattern information by LPP in the case of bearing fault classification. The proposed feature intends to overcome the weakness of the discriminatory WP nodes for characterizing the fault pattern in different sensitivity. Specifically, vibration signals from the bearings with different fault types and severities are firstly decomposed into the WP library and the LDB is then applied to identify the optimal WP subspaces that supply maximum dissimilarity information among them. After that, the root energy of the selected nodes constitutes a raw feature set. Due to the redundant property of the features in representing the fault pattern, some important sensitive information may be submerged among them. Therefore, the LPP is employed to extract the nonlinear sensitive pattern information embedded in the dataset. These sensitive features are finally chosen as inputs to a diagnostic classifier for characterizing bearing types and severities.

The rest of this paper is organized as follows. Section 2 describes the theoretical background and major principle of the proposed feature extraction method that combines the LDB and the LPP. In Section 3, experimental results on bearing fault classification are used to verify the effectiveness of the proposed method as compared to other traditional feature extraction methods. Finally, conclusions are provided in Section 4.

2. Theoretical Background

2.1. WPT for Signal Decomposition. The WPT is an excellent signal decomposition tool with well-known properties of being orthogonal, complete, and local [8]. In operation, the WPT utilizes a series of low-pass and high-pass filters to filter a signal being analyzed recursively. Through this way, a signal $x(t)$ can be decomposed into a set of WP nodes with the form of a full binary tree by the WPT. Each node possesses a specific time-frequency subspace. Let $\Omega_{0,0}$ denote a vector space $\mathbb{R}^n$ corresponding to the node 0 of the parent tree. Then at each level the vector space is split into two mutually orthogonal subspaces by a pair of low-pass and high-pass filters. The split process can be given by

$$\Omega_{j,k} = \Omega_{j+1,2k} \oplus \Omega_{j+1,2k+1}, \quad (1)$$

where j indicates the level of the tree and k represents the node index in level j with $k = 0, \dots, 2^j - 1$. This process is repeated until level J , giving rise to 2^J mutually orthogonal subspaces.

Each subspace $\Omega_{j,k}$ is spanned by a set of base vectors $\{w_{j,k,m}\}_{m=0}^{2^{n_0-j}-1}$, where $n_0 = \log_2 N \geq J$ (N is signal length and n_0 is the maximum level of signal decomposition). The vector $w_{j,k,m}$ represents the WP base function indexed by the triplet (j, k, m) representing scale, frequency band (oscillation), and time position, respectively.

The WP coefficients of signal $x(t)$ can be calculated in the inner product of the signal with every WP function as follows:

$$\alpha_{j,k,m} = \langle x, w_{j,k,m} \rangle = \int_{-\infty}^{\infty} x(t) w_{j,k,m}(t) dt, \quad (2)$$

where $\alpha_{j,k,m}$ denotes the k th set of WP coefficients at the j th scale parameter and m is the translation parameter. In other words, the signal $x(t)$ is decomposed into 2^J subspaces with coefficients $\{\alpha_{j,k,m}\}_{m=0}^{2^{n_0-j}-1}$ in each subspace.

The signal $x(t)$ can then be expressed as

$$x(t) = \sum_{j,k,m} [\alpha_{j,k,m}]_i \cdot w_{j,k,m}, \quad (3)$$

where, in the index (j, k, m) , (j, k) corresponds to the terminal (leaf) nodes and $\alpha_{j,k,m}$ are the base vector coefficients at position (m) .

2.2. LDB for WP Selection. The LDB is a pruning algorithm that identifies the subspaces and their bases that exhibit high discrimination between signal classes using a given dissimilarity measure [9]. The optimal selection of LDB subspaces for a given dataset is driven by the nature of the dataset and the dissimilarity measure. Dissimilarity measure is designed to evaluate the “statistical distances” among different classes for each WP node. Numerous dissimilarity measures have been developed so far, such as relative entropy, energy difference, correlation index, and nonstationarity. In this paper, relative entropy is investigated as the dissimilarity measure in identifying optimal WP subspaces.

The LDB algorithm is used to identify the WP nodes that exhibit high discrimination, indicated by large statistical distance between classes. A set of training signals for all L classes are decomposed into full binary WP trees of order J . Let each of the signals in the training set be denoted by x_i^l , where the index i and l correspond to the i th training signal in the l th class. The WP tree is pruned by the LDB algorithm in such a way that a node is split if the cumulative discriminative measure of the children nodes is greater than that of the parent node. In other words, a node is split only if the children nodes have better discriminative power than that of the parent node. As a result, the process will end with a subset of terminal WP nodes that contribute to maximizing the statistical distance between different classes.

Mathematically, the LDB selection process is described as follows. Suppose that $A_{j,k}$ represents the desired local discriminant base restricted to the span of $B_{j,k}$, which is a set of base vectors at (j, k) node, and $\Delta_{j,k}$ is the array containing the discriminant measure of the same node.

LDB Algorithm. A training dataset consisting of L class of signals $\{x_i^{(l)}\}_{i=1}^{N_l}$ with N_l being the total number of training signals in class l is given.

Step 1. Choose a time-frequency decomposition method, such as the WPT, to decompose the signals contained in the training dataset.

Step 2. Construct time-frequency energy maps C_l for $l = 1, \dots, L$ on the WP coefficients. Here, C_l is calculated by accumulating the squares of expansion coefficients of the signals at each position followed by a normalization with respect to the total energy of all the training signals belonging to class l as follows:

$$C_l(j, k, m) \equiv \frac{\sum_{i=1}^{N_l} (w_{j,k,m}^T x_i^{(l)})^2}{\sum_{i=1}^{N_l} \|x_i^{(l)}\|^2}, \quad (4)$$

where $j = 0, 1, \dots, J$, $k = 0, 1, \dots, 2^j - 1$, $m = 0, 1, \dots, 2^{n_0-j} - 1$.

Step 3. Set $A_{j,k} = B_{j,k}$, where $B_{j,k}$ is the base set spanning subspace of $\Omega_{j,k}$ node (j, k) and then evaluate $\Delta_{j,k} = D(\{C_l(j, k, \cdot)\}_{l=1}^L)$ for $k = 0, \dots, 2^j - 1$. Let $p^{(l)} = C_l(j, k, \cdot)$; for multiple class problems, the dissimilarity measure based on relative entropy is expressed as

$$D(\{p^{(l)}\}_{l=1}^L) = \sum_{a=1}^{L-1} \sum_{b=a+1}^L \sum_{i=1}^{N_l} p_i^{(a)} \log \frac{p_i^{(a)}}{p_i^{(b)}}. \quad (5)$$

Step 4. Determine the best subspace $A_{j,k}$ for $j = J-1, \dots, 0$, $k = 0, \dots, 2^j - 1$ by the following rule:

set $\Delta_{j,k} = D(\{C_l(j, k, \cdot)\}_{l=1}^L)$;

if $\Delta_{j,k} \geq \Delta_{j+1,2k} + \Delta_{j+1,2k+1}$, that is, if discriminatory power of a parent node in WP tree is greater than those of children nodes;

then $A_{j,k} = B_{j,k}$;

else $A_{j,k} = A_{j+1,2k} \oplus A_{j+1,2k+1}$ and set $\Delta_{j,k} = \Delta_{j+1,2k} + \Delta_{j+1,2k+1}$.

Step 5. Order the chosen base functions by their power of discrimination.

Step 6. Use k (normally much less than N) most discriminant base functions for constructing classifiers.

After Step 4 is performed, a complete set of orthogonal bases are constructed. Orthogonality of the bases ensures that wavelet coefficients used as features during classification process are uncorrelated as much as possible. Subsequently, one can use all the WP coefficients from each of the terminal nodes of the pruned tree or just use their subset with k highest discriminant bases in Step 6 or employ a statistical method to produce low-dimensional features as the input features of a classifier for discriminating different classes. In this paper, the WP coefficients of the selected optimal WP nodes are taken for calculating the root energy contained in each node. Mathematically, for the WP coefficients $W_{j,k}(i)$, $i = 1, \dots, 2^{n_0-j}$, from each node (j, k) , their root energy is calculated as

$$E_{j,k} = \sqrt{\sum_{i=1}^{2^{n_0-j}} W_{j,k}(i)^2}. \quad (6)$$

Totally, the root energy values of all of the selected nodes are put together to formulate a vector denoted by $\mathbf{E}_S$ (where $S = \{(j, k)\}$ is the subscript set of the selected WP nodes $W_{j,k}$) which is conveniently called the LDB feature. The set of root energy values contained in the WP nodes at the final level J of the WPT is called WPT feature in this paper and denoted by $\mathbf{E}_J$.

2.3. LPP for Feature Pattern Mining. In this section, we briefly describe the LPP algorithm of learning a locality preserving subspace from a high-dimensional data containing the sample values of a feature vector $\mathbf{E}_S$ or $\mathbf{E}_J$. Let $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n] \in \mathbb{R}^{q \times n}$ denote a data matrix, representing a set of q -dimensional samples of size n with zero mean. Now, consider the problem of representing the data matrix $\mathbf{Y}$ by a single vector $\mathbf{x} = [x_1, \dots, x_n]$ such that x_i represents $\mathbf{y}_i$. We will thus find a linear mapping, denoted by a transformation vector $\mathbf{w} \in \mathbb{R}^q$, from the q -dimensional space to a one-dimensional space, so that $\mathbf{w}^T \mathbf{y}_i = x_i$. LPP is a technique that seeks to preserve the intrinsic geometry of the data and local structure. The criterion of the objective function for choosing a map of the LPP is as follows:

$$\min \sum_{i,j} (x_i - x_j)^2 S_{ij}, \quad (7)$$

where x_i is the one-dimensional representation of $\mathbf{y}_i$ and the matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ is a similarity matrix.

A possible way of defining $\mathbf{S}$ is as follows:

$$S_{ij} = \begin{cases} \exp\left(\frac{-\|\mathbf{y}_i - \mathbf{y}_j\|^2}{t}\right), & \|\mathbf{y}_i - \mathbf{y}_j\|^2 < \varepsilon \\ 0 & \text{otherwise,} \end{cases} \quad (8)$$

where parameter $t \in \mathbb{R}$ and ε defines the radius of the local neighborhood and is sufficiently small but bigger than 0. Two samples $\mathbf{y}_i$ and $\mathbf{y}_j$ are viewed within a local ε -neighborhood provided that $\|\mathbf{y}_i - \mathbf{y}_j\|^2 < \varepsilon$.

The objective function in (7) with the choice of symmetric weights S_{ij} ($S_{ij} = S_{ji}$) will be heavily penalized if neighboring points $\mathbf{y}_i$ and $\mathbf{y}_j$ are mapped far apart, that is, if $(x_i - x_j)^2$ is large. Therefore, minimizing the objective function is to ensure that if $\mathbf{y}_i$ and $\mathbf{y}_j$ are close, then x_i and x_j are close as well. Based on this point, the local structure of the input data can be preserved. Following some algebraic steps, we can get

$$\begin{aligned} & \frac{1}{2} \sum_{i,j} (x_i - x_j)^2 S_{ij} \\ &= \frac{1}{2} \sum_{i,j} (\mathbf{w}^T \mathbf{y}_i - \mathbf{w}^T \mathbf{y}_j)^2 S_{ij} \\ &= \sum_{i,j} \mathbf{w}^T \mathbf{y}_i S_{ij} \mathbf{y}_i^T \mathbf{w} - \sum_{i,j} \mathbf{w}^T \mathbf{y}_i S_{ij} \mathbf{y}_j^T \mathbf{w} \\ &= \mathbf{w}^T \mathbf{Y} (\mathbf{D} - \mathbf{S}) \mathbf{Y}^T \mathbf{w} = \mathbf{w}^T \mathbf{Y} \mathbf{L} \mathbf{Y}^T \mathbf{w}, \end{aligned} \quad (9)$$

where $\mathbf{D}$ is a diagonal matrix; its entries are column (or row since $\mathbf{S}$ is symmetric) sums of $\mathbf{S}$, $D_{ii} = \sum_j S_{ij}$. $\mathbf{L} = \mathbf{D} - \mathbf{S}$ is

the Laplacian matrix. The bigger the value D_{ii} (corresponding to x_i) is, the more important x_i is. Therefore, the LPP algorithm imposes a constraint as follows:

$$\mathbf{x}^T \mathbf{D} \mathbf{x} = 1 \implies \mathbf{w}^T \mathbf{Y} \mathbf{D} \mathbf{Y}^T \mathbf{w} = 1. \quad (10)$$

Then, the minimization problem reduces to finding

$$\begin{aligned} & \underset{\mathbf{w}}{\operatorname{argmin}} \quad \mathbf{w}^T \mathbf{Y} \mathbf{L} \mathbf{Y}^T \mathbf{w} \\ & \text{s.t.} \quad \mathbf{w}^T \mathbf{Y} \mathbf{D} \mathbf{Y}^T \mathbf{w} = 1. \end{aligned} \quad (11)$$

The transformation vector $\mathbf{w}$ that minimizes the objective function is finally given by the minimum eigenvalue solution to the generalized eigenvalue problem:

$$\mathbf{w}^T \mathbf{Y} \mathbf{L} \mathbf{Y}^T \mathbf{w} = \lambda \mathbf{w}^T \mathbf{Y} \mathbf{D} \mathbf{Y}^T \mathbf{w}, \quad (12)$$

where the matrices $\mathbf{Y} \mathbf{L} \mathbf{Y}^T$ and $\mathbf{Y} \mathbf{D} \mathbf{Y}^T$ are symmetric and positive semidefinite. The top several projective vectors that minimize the objective function are the optimal linear approximations to the eigenfunctions of the Laplace Beltrami operator on the manifold, so they are capable of discovering the nonlinear manifold structure [33]. In this paper, the top several projective vectors are chosen as the mapping vectors to represent the LDB feature. They characterize the inherent class pattern and thus are hoped to mine the useful sensitive features for classification.

2.4. Proposed Feature Extraction Scheme for Data Classification. In the techniques mentioned above, the LDB and the LPP techniques have specific merits for classification. Specifically, the LDB algorithm focuses on identification of optimal decomposition subspaces for discriminatory feature extraction, while the LPP addresses the nonlinear pattern structure that represents the inherent condition class pattern. In other words, the LDB focuses on extraction of optimal raw features but each feature characterizes the class pattern in different sensitivity or local sensitivity, while LPP mainly addresses mining inherent class pattern feature embedded in the raw features. Therefore, this paper is proposed to combine the merits of these two techniques for a novel feature extraction. Specifically, the novel feature addresses extracting the inherent pattern structure embedded in the optimal WP nodes. Therefore, the proposed feature not only considers the static discriminatory WP node features themselves but also considers the dynamic sensitive class pattern structure embedded in the samples.

The idea of the proposed feature is illustrated in Figure 1. It can be found that although the optimal WP nodes (filled in black in Figure 1) have been selected through the LDB algorithm, they have different sensitivity in characterizing the class pattern. However, after conducting the LPP algorithm on the feature values, a new sensitive feature that clearly represents the class pattern is effectively extracted. In this process, the sensitive feature characterizes a nonlinear class pattern manifold embedded in sample values of the raw features. This indicates that LPP is beneficial to improve the class sensitivity of the selected

FIGURE 1: Principle of the proposed feature extraction scheme based on combination of LDB and LPP.

discriminatory features. Therefore, this new kind of feature can be well used to vibration data-based machine fault classification.

Based on the principle of combination of LDB and LPP, the proposed feature extraction algorithm can be then described as follows.

LDB-LPP Feature Extraction Algorithm. A training dataset consisting of L class of signals and a testing dataset are given.

Step 1. Conduct the WPT to decompose the signals contained in the dataset into the WP library with level J via (2).

If the signal is from the training dataset, go to Step 2.

Else, if the signal is from the testing dataset, then go to Step 3.

Step 2. Conduct the LDB algorithm to identify the optimal WP nodes that supply maximum dissimilarity information among the training dataset.

Step 3. Calculate the root energy of the coefficients of selected WP nodes to constitute a raw feature set E_S via (6).

If the signal is from the training dataset, go to Step 4.

Else, if the signal is from the testing dataset, then go to Step 5.

Step 4. Conduct the LPP algorithm to the raw feature value sets of the training dataset to obtain the mapping matrix through solving (12); then go to Step 5.

Step 5. Use the mapping matrix in Step 4 to calculate the new feature values of the dataset.

The proposed features have the most sensitive discriminatory capability and are thus chosen as inputs to a diagnostic classifier for characterizing data classes. To make it clearer, the flowchart of the proposed algorithm is shown in Figure 2 as well as the scheme of machine fault classification. The machine fault classification scheme includes two parts: the LDB-LPP feature values are firstly extracted for both the training and testing signals and then a diagnostic classifier is trained for classification of the fault signals.

3. Experimental Results and Analysis

In order to evaluate the effectiveness of the feature extraction scheme proposed above for machine fault classification, the bearing data with multiple faults from real bearing experiments are analyzed in this study.

3.1. Experimental Dataset. The experimental data are from Case Western Reserve University Bearing Data Center [34]. The experimental setup consists of four parts which are

FIGURE 2: Flowchart of the proposed LDB-LPP feature extraction algorithm and machine fault classification.

FIGURE 3: A schematic of the experimental system.

TABLE 1: Description of datasets for bearing data classification.

Dataset	Number of samples (training/testing)	Fault type	Fault size (inch)	Label of class
A	50/50	Normal	0	1
		Ball	0.007	2
		Inner race	0.007	3
		Outer race	0.007	4
		Ball	0.014	5
B	25/25	Inner race	0.014	6
		Outer race	0.014	7
		Ball	0.021	8
		Inner race	0.021	9
		Outer race	0.021	10

an induction motor, a dynamometer, a torque transducer, and control electronics. The resulting vibration was measured by an accelerometer being mounted to the motor housing at the drive end of the motor as illustrated in Figure 3. The accelerometer is a vibration sensor with a bandwidth up to 5000 Hz and a 1 V/g output. Single point faults of size 0.007, 0.014, 0.021, and 0.028 inches were set on the drive-end bearings by using the electric discharge machining approach. These faults were set, respectively, on rolling element, inner raceway, and outer raceway in the experiments. The sampling frequency of the data is 12 kHz with the sample length being 2000 and the motor speed was 1748 rev/min.

Datasets A and B to be analyzed consist of ten classes (class labels are marked in Table 1) covering different bearing fault types and severities as listed in Table 1. In the datasets, there are four different fault types including normal, outer-race fault, inner-race fault, and ball fault, and each of the last three fault types includes three different defect sizes of 0.007, 0.014, and 0.021 inches, respectively. In dataset A, the samples are split into 500 training ones (50 in each class) and 500

testing ones (50 in each class), while dataset B contains 250 training samples (25 in each class) and 250 testing samples (25 in each class). This is a complex ten-class problem to identify both the fault type and the fault size for the operating bearing conditions.

3.2. Feature Evaluation. In this study, the decomposition level of the WPT is set to be 6 and the Daubechies 8 wavelet is employed. The selected nodes by the LDB are shown in Figure 4. In the following study, the root energy of a signal decomposed into each selected node is calculated as the raw features in the proposed study, while that in each node at the last layer is used to form the traditional WPT feature for a comparison.

To quantitatively evaluate the capability of LDB feature in pattern classification, three common clustering evaluation metrics are analyzed as follows. The first is a widely used discriminant factor. Suppose that there is a feature vector $\{f_1, f_2, \dots, f_d\}$, where d is the dimension of feature; then the discriminant factor is defined as follows:

$$S = \frac{S_B}{S_W}, \quad (13)$$

where S_B indicates the between-class scatter to describe the scattered level among different classes, while S_W is the within-class scatter which represents the concentrated level in the same classes. These two scatters are, respectively, defined as

$$S_B = \sum_{l=1}^L N_l \|\mu_f^l - \bar{\mu}_f\|, \quad (14)$$

$$S_W = \sum_{l=1}^L \sum_{j=1}^{N_l} \|f_j - \mu_f^l\|,$$

where N_l is the total number and μ_f^l is the average feature vector for samples in the l th class and $\bar{\mu}_f$ is the total average of the feature vectors for all classes. It can be seen that the discriminant factor S is a comprehensive indicator that combines between-class scatter and within-class scatter. A larger discriminant factor is better for classification purpose to characterize the discriminating capability of the given feature.

FIGURE 4: Optimal WP nodes selected by the LDB at level 6: (a) dataset A and (b) dataset B.

The other two clustering evaluation metrics are the cluster accuracy (ACC) and normalized mutual information (NMI) metrics [35], which are defined as follows, respectively.

Cluster Accuracy (ACC). Assuming that s_i and r_i are the acquired new label and the provided real label of the given point μ_i , the ACC is defined as follows:

$$\text{ACC} = \frac{\sum_{i=1}^{Ns} \delta(r_i, \text{Map}(s_i))}{Ns}, \quad (15)$$

$$\delta(p, q) = \begin{cases} 1 & p = q \\ 0 & \text{otherwise,} \end{cases}$$

where Ns is the total number of the samples and $\text{Map}(s_i)$ is the optimal mapping function that ranges each s_i to match the real label and can be found by the Kuhn-Munkres (KM) algorithm. Here, we assume that the relationship of the identified clusters with the predefined classes is known. Thus, it is easy to imagine that a larger ACC value indicates better clustering and generally better classification.

Normalized Mutual Information (NMI). It is a mutual information (MI) metric and defined as

$$\text{NMI} = \frac{\sum_{i=1}^L \sum_{j=1}^L t_{i,j} \log((Ns \cdot t_{i,j}) / t_i \tilde{t}_j)}{\sqrt{\sum_{i=1}^L t_i \log(t_i / Ns)} \sqrt{\sum_{j=1}^L \tilde{t}_j \log(\tilde{t}_j / Ns)}}, \quad (16)$$

where t_i is the number of the acquired samples in class C_i and $\tilde{t}_j$ is the number of the provided samples in the ground truth class C_j . In addition, $t_{i,j}$ is the number of the intersected samples between class C_i and class C_j . A larger NMI reveals better clustering performance, which is beneficial to classification.

For a visible purpose and a fair comparison, the dimensions of the LDB and the WPT features are both reduced to 3 by using dimensionality reduction techniques including the LPP and the traditional PCA. Note that the LDB feature followed by the LPP just generates the proposed feature in this study. We then calculated the mentioned

TABLE 2: Clustering evaluation of different features for dataset A.

Metric	PCA		LPP	
	WPT	LDB	WPT	LDB
S	3.0900	4.0740	7.8135	8.7864
ACC	0.4700	0.7560	0.9740	0.9800
NMI	0.7914	0.9090	0.9546	0.9623

three clustering evaluation metrics. Here, k -means clustering method is applied to obtain the cluster label in the reduced 3-dimensional features before calculating the ACC and NMI. The k number used in the k -means clustering method is set as L which is the number of the class. What is more, to realize efficient and stable convergence, we set the L initial points as the intermediate point of each class in mathematics.

The neighborhood parameter of LPP is taken as 12. As an illustration, the scatter plots of the ten-class dataset A for training data are drawn in Figure 5. It can be seen that the LDB feature shows a better classification capability than the WPT features. On the other hand, it can be also found that the LPP has a much more excellent classification capability than the PCA. Therefore, the LDB-LPP shows the best classification capability in the between-class and within-class scatter performance. The extracted feature patterns are also demonstrated in Figure 6, where it can be clearly seen that the third LPP of LDB feature values characterizes a better difference for each class as compared to the third LPP of WPT feature values. Moreover, the quantitative results as listed in Table 2 also support the above statements. It can be seen that the clustering evaluation metrics S, ACC, and NMI of the LDB feature are higher than those of the WPT feature, and LPP performs much better than PCA. The combination of LDB and LPP shows the most beneficial performance for classification. Moreover, the clustering evaluation of dataset B (with half the number of samples of dataset A) is also computed here as shown in Table 3. These clustering evaluation values show the same tendency and indicate that the LPP can learn a good nonlinear class pattern structure among the discriminatory LDB feature values.

FIGURE 5: Representation of training samples of bearing dataset A: (a) PCA of WPT feature values, (b) PCA of LDB feature values, (c) LPP of WPT feature values, and (d) LPP of LDB feature values.

TABLE 3: Clustering evaluation of different features for dataset B.

Metric	PCA		LPP	
	WPT	LDB	WPT	LDB
S	3.0260	3.7721	9.5638	9.7630
ACC	0.5120	0.7720	0.9280	1
NMI	0.779	0.8149	0.9040	1

3.3. Classification of Fault Types and Severities. To further evaluate the performance of the proposed feature in data classification, the ten-class datasets A and B are employed for fault classification by comparing various features. In this study, the proposed LDB-LPP feature is compared to traditional feature extraction methods including PCA, LDA, LPP, supervised LPP (SLPP) [36], LE, and LLE. Among the six methods, PCA and LPP are unsupervised linear techniques,

LDA and SLPP are supervised linear techniques, and LE and LLE are nonlinear manifold learning techniques.

To emphasize the feature performance, the nearest mean classifier, one of the simplest and the most intuitive statistical classifiers, is applied for classification in this study. This classifier is based on the principle of the closest Euclidean distance and the concept of similarity that similar patterns should be assigned to the same class. In this study, the mean vector of the training data in each class is used to represent each pattern class. Patterns of samples can be distinguished according to the minimum distance criterion which means maximum similarity. Moreover, another advanced classifier, Gaussian mixture model (GMM) classifier, is also applied in this study.

The recognition accuracy of the proposed LDB-LPP feature and the other comparison features (extracted from the WPT feature without node selection) are shown in Table 4. It can be seen that the proposed LDB-LPP feature

FIGURE 6: The first three projections of training samples of bearing dataset A: (a~c) LPPs of WPT feature values and (d~f) LPPs of LDB feature values.

TABLE 4: Recognition accuracy (%) performance for dataset A and dataset B.

Classifier	Feature Extraction	Dataset A		Dataset B	
		Training	Testing	Training	Testing
Nearest Mean	PCA	84.6	84.2	85.6	82
	LPP	98.8	99	98.8	96
	SLPP	98.6	99	99.6	97.6
	LE	97.8	97.2	95.2	90.4
	LLE	97.2	98.2	98	98.4
	LDA	98.6	98.2	100	98.4
GMM	LDB-LPP	99.8	100	99.2	99.6
	PCA	97.8	90	96.8	83.6
	LPP	99.2	99	99.6	97.6
	SLPP	99.2	98.8	100	97.6
	LE	89	88.4	96.4	90.8
	LLE	98.8	99.2	100	99.6
	LDA	99.2	97.8	100	96.8
	LDB-LPP	100	100	100	100

outperforms the traditional features achieved by PCA, LDA, LPP, SLPP, LE, and LLE, which verifies the benefits of the LDB for choosing discriminatory features. Moreover, the GMM

classifier further improves the recognition rate of the nearest mean classifier. Note that the recognition accuracy of dataset A is generally higher than that of dataset B because the number of samples in dataset A is bigger. It can be found that the promotion of the recognition accuracy of the LDB-LPP feature in comparison with the other features becomes more obvious in dataset B than in dataset A for two classifiers. For instance, dataset A shows an average promotion 1% for testing by considering the LDB in the LPP feature extraction, while dataset B displays an average promotion 3% (for two classifiers) for testing. Figures 7 and 8 intuitively display that the proposed LDB-LPP feature performs the best among all the comparison features. In this study, the testing recognition accuracy based on LDB-LPP feature is equal to or very close to 100% for two classifiers. These results imply that the proposed LDB joint LPP feature extraction method could obtain significant achievements in improving classification accuracy.

4. Conclusions

This paper presents a feature extraction method which integrates the LDB and the LPP to explore the useful and

FIGURE 7: Recognition accuracy of ten-class classification by the nearest mean classifier: (a) dataset A and (b) dataset B.

FIGURE 8: Recognition accuracy of ten-class classification by GMM classifier: (a) dataset A and (b) dataset B.

powerful characteristics for vibration data-based machine fault classification. The LDB is used to select the most discriminant WP bases from a library of redundant and orthogonal time-frequency subspaces. The input features are produced by the selected optimal wavelet bases but they possess different sensitivity in characterizing class information. The LPP is then employed to acquire the sensitive feature that characterizes the inherent class pattern feature embedded in the raw features for a much better identification accuracy. The proposed feature extraction method combines the merits of the LDB and the LPP and thus displays valuable benefits for data classification. To verify the effectiveness of

the proposed method, the vibration data representing different bearing fault types and severities are analyzed by comparing with other features extracted from the WPT feature. The experimental results for bearing fault classification indicate that the LDB-LPP feature is more effective than those feature extraction methods based on the WPT feature without base selection. The presented LDB joint LPP feature extraction method is also hoped to be well-suited to other machine fault classification, such as gears, spindles, and cutting tools, due to the excellent feature representation for the class patterns.

Moreover, the technical aspects in the proposed LDB-LPP feature extraction framework can be further improved and

strengthened. First, this paper fairly compares the LDB feature and WPT feature in the same decomposition level, which can validate the benefits of the LDB in data classification. However, how to select the well-suited decomposition level in the LDB is still an open issue in the further study. Second, LPP is a typical and effective feature extraction method which obtains the manifold structure in a linear projection. Although the LPP has been successfully used to overcome the weakness of the LDB in this study, it is meaningful to apply the new well-performed manifold learning methods instead of LPP in the proposed framework to further enhance the performance of data classification. This should also depend on how complex the data to be analyzed is. Other possible applications on complex classification remained to be studied in the future.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 51005221), the Research Fund for the Doctoral Program of Higher Education of China (Grant no. 20103402120017) and the Program for New Century Excellent Talents in University, China (Grant no. NCET-13-0539). The authors would like to thank Case Western Reserve University for offering free download of the bearing data and the anonymous reviewers for their constructive and valuable comments.

References

- [1] H. Qiu, J. Lee, J. Lin, and G. Yu, "Robust performance degradation assessment methods for enhanced rolling element bearing prognostics," *Advanced Engineering Informatics*, vol. 17, no. 3-4, pp. 127-140, 2003.
- [2] Q. He, F. Kong, and R. Yan, "Subspace-based gearbox condition monitoring by kernel principal component analysis," *Mechanical Systems and Signal Processing*, vol. 21, no. 4, pp. 1755-1772, 2007.
- [3] J. Yu, "Bearing performance degradation assessment using locality preserving projections," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7440-7450, 2011.
- [4] Y. Lei, J. Lin, Z. He, and D. Kong, "A method based on multi-sensor data fusion for fault detection of planetary gearboxes," *Sensors*, vol. 12, no. 2, pp. 2005-2017, 2012.
- [5] Z. Liu, X. Chen, Z. He, and Z. Shen, "LMD method and multi-class RWSVM of fault diagnosis for rotating machinery using condition monitoring information," *Sensors*, vol. 13, no. 7, pp. 8679-8694, 2013.
- [6] S. Wang, X. Sun, and C. Li, "Wind turbine gearbox fault diagnosis method based on Riemannian manifold," *Mathematical Problems in Engineering*, vol. 2014, Article ID 153656, 10 pages, 2014.
- [7] R. Yan, R. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: a review with applications," *Signal Processing*, vol. 96, pp. 1-15, 2014.
- [8] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 713-718, 1992.
- [9] N. Saito and R. R. Coifman, "Local discriminant bases and their applications," *Journal of Mathematical Imaging and Vision*, vol. 5, no. 4, pp. 337-358, 1995.
- [10] E. Hulata, R. Segev, Y. Shapira, M. Benveniste, and E. Ben-Jacob, "Detection and sorting of neural spikes using wavelet packets," *Physical Review Letters*, vol. 85, no. 21, pp. 4637-4640, 2000.
- [11] N. Saito, R. R. Coifman, F. B. Geshwind, and F. Warner, "Discriminant feature extraction using empirical probability density estimation and a local basis library," *Pattern Recognition*, vol. 35, no. 12, pp. 2841-2852, 2002.
- [12] D. J. Strauss, G. Steidl, and W. Delb, "Feature extraction by shape-adapted local discriminant bases," *Signal Processing*, vol. 83, no. 2, pp. 359-376, 2003.
- [13] D. Li, W. Pedrycz, and N. J. Pizzi, "Fuzzy wavelet packet based feature extraction method and its application to biomedical signal classification," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 6, pp. 1132-1139, 2005.
- [14] R. Tafreshi, F. Sassani, H. Ahmadi, and G. Dumont, "Local discriminant bases in machine fault diagnosis using vibration signals," *Integrated Computer-Aided Engineering*, vol. 12, no. 2, pp. 147-158, 2005.
- [15] K. Umamathy and S. Krishnan, "Modified local discriminant bases algorithm and its application in analysis of human knee joint vibration signals," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 3, pp. 517-523, 2006.
- [16] K. Umamathy, S. Krishnan, and R. K. Rao, "Audio signal feature extraction and classification using local discriminant bases," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 4, pp. 1236-1246, 2007.
- [17] Q. He, R. Yan, and R. X. Gao, "Wavelet packet base selection for gearbox defect severity classification," in *Proceedings of the Prognostics and System Health Management Conference (PHM '10)*, Macau, China, January 2010.
- [18] D. Kim, J. J. Liu, and C. H. Han, "Determination of steel quality based on discriminating textural feature selection," *Chemical Engineering Science*, vol. 66, no. 23, pp. 6264-6271, 2011.
- [19] Q. He, R. Yan, F. Kong, and R. Du, "Machine condition monitoring using principal component representations," *Mechanical Systems and Signal Processing*, vol. 23, no. 2, pp. 446-466, 2009.
- [20] R. Yan and R. Gao, "Wavelet domain principal feature analysis for spindle health diagnosis," *Structural Health Monitoring*, vol. 10, no. 6, pp. 631-642, 2011.
- [21] L. H. Chiang, E. L. Russell, and R. D. Braatz, "Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 50, no. 2, pp. 243-252, 2000.
- [22] X. Jin, M. Zhao, T. Chow, and M. Pecht, "Motor bearing fault diagnosis using trace ratio linear discriminant analysis," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2441-2451, 2014.
- [23] A. Widodo and B. Yang, "Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors," *Expert Systems with Applications*, vol. 33, no. 1, pp. 241-250, 2007.
- [24] C. Sun, Z. Zhang, Z. He, Z. Shen, B. Chen, and W. Xiao, "Novel method for bearing performance degradation assessment? A

- kernel locality preserving projection based approach,” *Proceedings of the Institution of Mechanical Engineers C: Journal of Mechanical Engineering Science*, vol. 228, pp. 548–560, 2013.
- [25] 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.
 - [26] 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.
 - [27] Z. Su, B. Tang, J. Ma, and L. Deng, “Fault diagnosis method based on incremental enhanced supervised locally linear embedding and adaptive nearest neighbor classifier,” *Measurement*, vol. 48, pp. 136–148, 2014.
 - [28] 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.
 - [29] M. Belkin and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
 - [30] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
 - [31] Z. Zhang and H. Zha, “Principal manifolds and nonlinear dimensionality reduction via tangent space alignment,” *SIAM Journal on Scientific Computing*, vol. 26, no. 1, pp. 313–338, 2005.
 - [32] M. H. C. Law and A. K. Jain, “Incremental nonlinear dimensionality reduction by manifold learning,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 377–391, 2006.
 - [33] X. He and P. Niyogi, “Locality preserving projections,” in *Proceedings of the Conference on Advances in Neural Information Processing Systems (NIPS ’03)*, vol. 16, pp. 234–241, Vancouver, Canada, December 2003.
 - [34] Bearing data center, <http://csegroups.case.edu/bearingdatacenter/home>.
 - [35] M. Zhao, X. Jin, Z. Zhang, and B. Li, “Fault diagnosis of rolling element bearings via discriminative subspace learning: visualization and classification,” *Expert Systems with Applications*, vol. 41, no. 7, pp. 3391–3401, 2014.
 - [36] D. Cai, X. He, and J. Han, “Using graph model for face analysis,” Tech. Rep. no. 2636, 2005.