Chapter 9
Local Discriminant Bases for Signal Classification

The goal of analyzing signals from manufacturing machines is to extract relevant *features* from the waveforms to effectively characterize the working conditions of the machines (e.g., tool breakage and gear degradation). As we have shown in Chap. 5, the wavelet packet transform can lead to redundant signal decomposition within certain time–frequency subspaces. When performing wavelet packet transform, the time–frequency subspaces are collectively called the wavelet packet library. Each of the subspaces is denoted as a wavelet packet node. Such a way of signal decomposition provides the possibility of selecting a particularly suited set of wavelet packet nodes out of the wavelet packet library for a specific signal analysis task, such as data compression, regression, or classification (Saito 1994). However, the choice of wavelet packet nodes is dependent on the specific task. For example, the optimal wavelet packet transform technique introduced in Chap. 5 is geared toward signal compression (Coifman and Wicherhauser 1992), in which the wavelet packet nodes are selected based on minimizing an information cost function (e.g., Shannon entropy). This chapter introduces how to choose a good set of wavelet packet nodes from a wavelet packet library, for purpose of signal classification. Such a technique has shown to be effective for monitoring and diagnosis of rotating machines.

9.1 Dissimilarity Measures

To classify signals obtained from a machine under different working status, the *features* extracted from the signals should clearly differentiate different working status of the machine, where each status is considered as a distinct *class*. For example, signals measured on a new gearbox are denoted as one class, while signals measured on a gearbox with broken-teeth are denoted as another class. Such type of features is referred to as "discriminant" features of the signal. The main objective of signal classification by using the wavelet packet transform is to find an optimal set of wavelet packet nodes (each node representing a wavelet packet basis) that are capable of discriminating different *classes* as effectively as possible. This can be achieved by decomposing the signal of interest into different classes, using the local

R.X. Gao and R. Yan, *Wavelets: Theory and Applications for Manufacturing*,
DOI 10.1007/978-1-4419-1545-0_9,

discriminant bases (LDB) algorithm (Saito 1994; Saito and Coifman 1995). The optimal choice of LDBs depends on the nature of the signals and the dissimilarity measures used to distinguish classes. In general, the dissimilarity measure is aimed at evaluating the "statistical distances" of each wavelet packet node among different classes. Numerous dissimilarity measures have been developed (Basseville 1989; Saito 1994; Saito et al. 2002; Umapathy and Krishnan 2006; Umapathy et al. 2007), among which the following four measures have been typically associated with the application of the LDB algorithm.

9.1.1 Relative Entropy

Relative entropy is one of the first dissimilarity measures used for identifying the LDBs (Saito 1994). On the basis of the definition of relative entropy, this dissimilarity measure is defined as:

$$D_1(p^1, p^2) = \sum_{i=1}^{n} p_i^1 \log \frac{p_i^1}{p_i^2} \tag{9.1}$$

where $\sum_{i=1}^{n} p_i^1 = 1$ and $\sum_{i=1}^{n} p_i^2 = 1$. The symbols p^1 and p^2 denote nonnegative sequences, respectively. It is assumed that $\log 0 = -\infty$, $\log(x/0) = +\infty$ for $x > 0$, and $0 \times (\pm\infty) = 0$. The discriminant information $D_1(p^1, p^2)$ between these two sequences measures how differently p^1 and p^2 are distributed. From the definition, it is seen that the nonnegative sequences p^1 and p^2 can be considered as probability density function. Since the normalized energy of wavelet coefficients (i.e., representation of energy distribution) within each wavelet packet basis is actually an expression of the probability density function associated with that wavelet packet node, it can be used as replacement in (9.1). Consequently, the relative entropy can be used as a dissimilarity measure in the LDB algorithm. Furthermore, (9.1) indicates that the relative entropy measure D_1 is nonnegative and will be zero if the two sequences of p^1 and p^2 are the same. The more separate from each other the two sequences are, the higher the relative entropy measure D_1 will be. However, it should be noted that the relative entropy measure shown in (9.1) is only applicable to a two-class problem. For multiple-class problems (e.g., gearbox under four different conditions: such as (a) faultless, (b) slight-worn, (c) medium-worn, and (d) broken-teeth), the dissimilarity measure based on relative entropy is modified as:

$$D_1(\{p^m\}_{m=1}^{L}) = \sum_{a=1}^{L-1} \sum_{b=a+1}^{L} D_1(p^a, p^b) \tag{9.2}$$

where L is the number of classes. Equation (9.2) indicates that the dissimilarity measure of multiple-class problems is the summation of relative entropy for each pair of two-classes among all the classes.

9.1.2 Energy Difference

From the decomposition results of a signal's wavelet packet transform, the normalized energy associated with the wavelet packet node (j, k) is calculated as:

$$E_{j,k} = \frac{\sum_{l=1}^{M} |x_{j,k,l}|^2}{E_{x(t)}} \tag{9.3}$$

In (9.3), the symbols j and k represent the wavelet packet decomposition level and subfrequency band, respectively. These two symbols, collectively, represent a wavelet packet node (j, k). The symbol $x_{j,k,l}$ denotes the lth wavelet packet coefficient within the node (j, k), and the symbol M denotes the total number of coefficients within that node. $E_{x(t)}$ is the total energy contained in the signal.

The difference in the normalized energy associated with wavelet packet node (j, k) between the signals from two classes (denoted as class 1 and class 2) can be defined as a dissimilarity measure, given by:

$$D_2(E^1, E^2) = E^1_{j,k} - E^2_{j,k} \tag{9.4}$$

The symbols $E^1_{j,k}$ and $E^2_{j,k}$ represent the normalized energy associated with wavelet packet node (j, k) from class 1 and class 2 signals, respectively. Since each wavelet packet node corresponds to a time–frequency subspace, the normalized energy computed at a node provides the energy distribution of the signal in a particular subfrequency band. The greater the difference at a particular node in the energy distribution of the two classes, the more significant the node for discriminating the classes will be. Similar to the definition expressed by (9.1), (9.4) represents the energy difference measure used for a two-class problem. For multiple-class problems, the dissimilarity measure based on the energy difference is expressed as:

$$D_2(\{E^m\}_{m=1}^{L}) = \sum_{a=1}^{L-1} \sum_{b=a+1}^{L} D_2(E^a, E^b) \tag{9.5}$$

where L is the number of classes. Equation (9.5) indicates the dissimilarity measure of multiple-class problems is the summation of energy difference for each pair of two-classes among all the classes.

9.1.3 Correlation Index

The dissimilarity measure can also be defined from the correlation between the wavelet packet node (j, k) from class 1 and class 2. This measure can be used to identify those nodes that can detect the difference in the temporal characteristics of the signals between class 1 and class 2. The dissimilarity measure based on the correlation index, which is used in a two-class problem, is formulated as:

$$D_3(x^1, x^2) = \langle x^1_{j,k,l}, x^2_{j,k,l} \rangle \tag{9.6}$$

where the symbols j, k, and l represent decomposition level, subfrequency band, and time position, respectively, and $x^1_{j,k,l}$ and $x^2_{j,k,l}$ are the coefficients of the corresponding wavelet packet node (j, k) of class 1 and class 2. An average low correlation index at a particular node indicates high dissimilarity between the classes. Similarly, for a multiple class problem, the dissimilarity measure based on correlation index is expressed as:

$$D_3(\{x^m\}^L_{m=1}) = \sum_{a=1}^{L-1} \sum_{b=a+1}^{L} D_3(x^a, x^b) \tag{9.7}$$

where L is the number of classes. Equation (9.7) indicates that the dissimilarity measure of multiple-class problems is the summation of correlation index for each pair of two-classes among all the classes.

9.1.4 Nonstationarity

Nonstationarity of the wavelet packet coefficients may also be used to measure the dissimilarity. It is computed as the set of variances along the segments of the wavelet packet coefficients at a given node (j, k). The ratio of this variance between class 1 and class 2 indicates the amount of deviation in the nonstationarity between the two classes. Consequently, the dissimilarity measure based on nonstationarity, which is used in a two-class problem, can be defined as:

$$D_4(v^1, v^2) = \frac{\mathrm{var}[v^1_{j,k}]}{\mathrm{var}[v^2_{j,k}]} \tag{9.8}$$

where the symbols j and k represent the decomposition level and subfrequency band of the wavelet coefficients, respectively. The symbols v^1 and v^2 are variance vectors. Each of them contains L variances, obtained by equally segmenting the wavelet packet coefficients at node (j, k) for class 1 and class 2 signals, respectively. For example, given a signal in class 1 with 4,096 data points, there will be 1,024 wavelet packet coefficients at node (2, 1), since $4{,}096/2^2 = 1{,}024$. If these wavelet packet coefficients are equally partitioned into eight segments (i.e., $L = 8$), then there will be eight elements in variance vector v^1. Each of the elements is calculated from 128 wavelet packet coefficients ($1{,}024/8 = 128$). Variance vector v^2 can be obtained in the same way.

Similarly, for multiple class problems, the dissimilarity measure based on nonstationarity is expressed as:

$$D_4(\{v^m\}_{m=1}^L) = \sum_{a=1}^{L-1} \sum_{b=a+1}^{L} D_4(v^a, v^b) \tag{9.9}$$

where L is the number of classes. Equation (9.9) indicates that the dissimilarity measure of multiple class problems is the summation of nonstationarity for each pair of two-classes among all the classes.

9.2 Local Discriminant Bases

Utilizing one of the dissimilarity measures introduced earlier (e.g., relative entropy), the LDB algorithm can identify wavelet packet nodes that exhibit high discrimination, as indicated by a large statistical distance among the classes.

Let us assume that $\Omega_{0,0}$ denotes the wavelet packet node 0 of the parent tree (i.e., the signal itself). Then at each level, the wavelet packet node $\Omega_{j,k}$ is split into two mutually orthogonal subspaces (i.e., nodes $\Omega_{j+1,2k}$ and $\Omega_{j+1,2k+1}$), given by

$$\Omega_{j,k} = \Omega_{j+1,2k} \oplus \Omega_{j+1,2k+1} \tag{9.10}$$

where j indicates the level of the tree, and k represents the node index in level j, given by $k = 0, \ldots, 2^j - 1$. This process is repeated until level J, giving rise to 2^J mutually orthogonal subspaces. The goal is to select a set of best subspaces that provide maximum dissimilarity information among different classes of the signals. This can be realized by a *pruning* approach, where the wavelet packet tree is pruned in such a way that, starting from the bottom decomposition level, 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 if the children nodes have better discriminative power than that of the parent node. Such a process is executed until it reaches the top level of the decomposition. As a result, the process will end with a subset of wavelet packet nodes that contribute to maximizing the statistical distance among different classes. As an example, Fig. 9.1 shows a wavelet packet tree for a two-level signal decomposition. The LDB algorithm first compares the discriminant

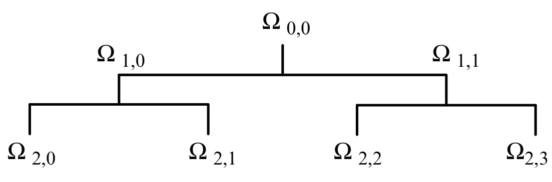

Fig. 9.1 Illustration of all nodes in two-level wavelet packet decomposition

power associated with the coefficients of training signals in different classes at the $\Omega_{1,1}$ node with that of the $\Omega_{2,2}$ and $\Omega_{2,3}$ nodes, respectively. If the relative entropy of $\Omega_{1,1}$ is larger than that of $\Omega_{2,2}$ and $\Omega_{2,3}$, it keeps the bases belonging to node $\Omega_{1,1}$ and omits the other two nodes ($\Omega_{2,2}$ and $\Omega_{2,3}$). Otherwise it keeps the two nodes ($\Omega_{2,2}$ and $\Omega_{2,3}$) and disregards the basis of node $\Omega_{1,1}$. This process is applied to all the nodes in a sequential manner, up to the scale $j = 0$. As a result, a set of complete orthogonal wavelet packet bases having the highest discriminant power are obtained, which can be sorted out further for classification, according to a decreasing order.

Suppose that $A_{j,k}$ represents the desired local discriminant base restricted to the span of $B_{j,k}$, which is a set of wavelet packet coefficients at (j, k) node, and $\Delta_{j,k}$ is the array containing the discriminant measure of the same node, then the LDB algorithm for selecting the optimal wavelet packet base can be summarized as follows (Tafreshi et al. 2005):

LDB Algorithm Given a training dataset that consists of L class of signals $\{\{x_i^{(l)}\}_{i=1}^{N_l}\}_{l=1}^{L}$ with N_l being the total number of training signals in class l,

Step 0: Choose a time–frequency analysis method, such as the wavelet packet transform, to decompose the signals in the training dataset.

Step 1: Select a dissimilarity measure (e.g., relative entropy $D_1(\{p^m\}_{m=1}^{L})$) to apply on the wavelet packet coefficients to the corresponding nodes (j, k) of the wavelet packet trees.

Step 2: Set $A_{J,k} = B_{J,k}$ where $B_{J,k}$ is the basis set spanning subspace of $\Omega_{j,k}$ node (J, k), and then evaluate $\Delta_{J,k}$ for $k = 0, \ldots, 2^J - 1$.

Step 3: Determine the best subspace $A_{j,k}$ for $j = J - 1, \ldots, 0$, $k = 0, \ldots, 2^j - 1$ by the following rule:

Set $\Delta_{j,k}$ as the dissimilarity measure, e.g., $\Delta_{j,k} = D_1(\{p^m\}_{m=1}^{L})$
If $\Delta_{j,k} \geq \Delta_{j+1,2k} + \Delta_{j+1,2k+1}$, i.e., if the discriminant power of a parent node in wavelet packet tree is greater than those of children nodes,
Then
$A_{j,k} = B_{j,k}$
Else
$A_{j,k} = \Delta_{j+1,2k} \bigoplus \Delta_{j+1,2k+1}$ and set $\Delta_{j,k} = \Delta_{j+1,2k} + \Delta_{j+1,2k+1}$.

Step 4: Order sort the chosen basis functions by their power of discrimination in a decreasing order.

Step 5: Select the first $k(\leq l)$ highest discriminant base functions.

After step 3 is performed, a complete orthogonal basis is constructed. Orthogonality of the bases ensures that wavelet coefficients used as features during classification process are uncorrelated as much as possible. Subsequently, one can simply choose the first k highest discriminant bases in step 5 and use the corresponding coefficients as features in a classifier, or employ a statistical method, such as Fisher's criteria, to reduce the dimensionality of the problem first and then apply them into a classifier.

9.3 Case Study

To evaluate the effectiveness of the wavelet packet bases constructed using the LDB algorithm, three classes of signals are synthetically formed:

$$\begin{cases} x^{(1)}(t) = \mathrm{Sine}(t) + n_1(t) & \text{for class 1} \\ x^{(2)}(t) = \mathrm{Gauspuls}(t) + n_2(t) & \text{for class 2} \\ x^{(3)}(t) = Tripuls(t) + n_3(t) & \text{for class 3} \end{cases} \tag{9.11}$$

In (9.11), Sine(t), Gauspuls(t), and Tripuls(t) represent the sinusoidal, Gaussian-modulated sinusoidal pulse, and triangle wave signals, respectively. The terms $n_1(t)$, $n_2(t)$, and $n_3(t)$ represent white noise. For each class, 100 training signals and 1,000 test signals were constructed, and the white noise was regenerated each time. Figure 9.2 shows one sample signal with 64 sampling points from each class. Each sample signal can be decomposed up to the six level (i.e., $2^6 = 64$), and the total number of nodes contained in the wavelet packet library for the signal is 127 (i.e., 1 for the 0 level, 2 for the first level, . . ., 64 for the sixth level).

The LDB algorithm is first applied to the training signals to select a subset of wavelet packet nodes from a wavelet packet library that best discriminate the three classes. Figure 9.3 shows the selected wavelet packet nodes. It can be seen that the selected wavelet packet nodes (highlighted in black color) are distributed across different decomposition levels. Altogether, they form complete orthogonal bases. The first six selected LDB bases are shown in Fig. 9.4, with each containing 64 coefficients. It should be noted that these bases are sorted according to their discriminant power. A complete discrimiant power for all the 64 LDB bases is

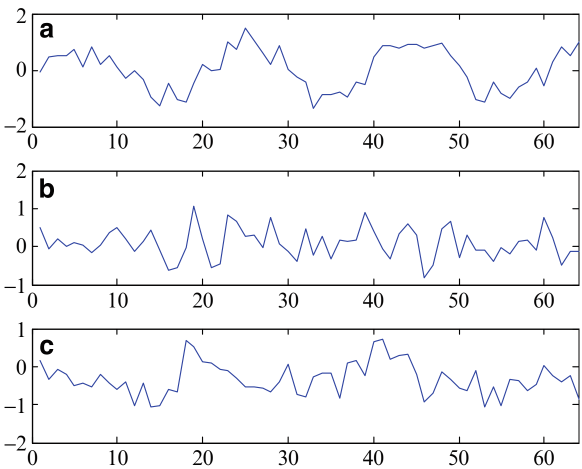

Fig. 9.2 Sample waveform from (**a**) class 1, (**b**) class 2, and (**c**) class 3

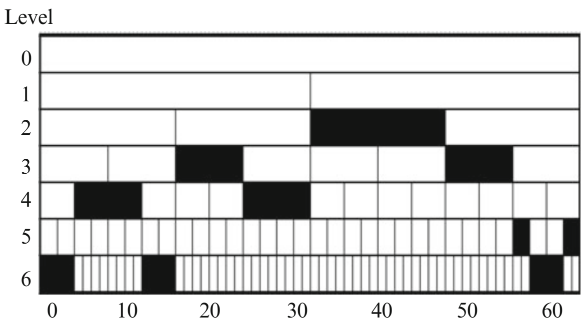

Fig. 9.3 The wavelet packet nodes selected by the LDB algorithm

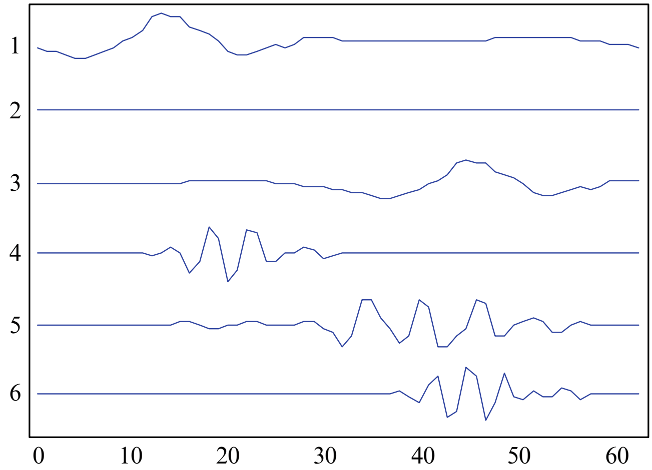

Fig. 9.4 The first six LDB bases selected from the signals

shown in Fig. 9.5 with a decreasing order. A rapid decrease of the discriminant power relative to the LDB bases is seen after the first few bases. Therefore, only the first few bases (e.g., the first six bases) with large discriminative power are considered for purpose of classification.

Wavelet coefficients constructed by projecting the signals onto the selected bases are then used to form *feature* variables for classification. For the training data set, two features (i.e., two wavelet coefficients) generated by the top two LDB bases have produced the clustering result as shown in Fig. 9.6.

To classify the three synthetic signals introduced above, these two features are used as input to a classifier. Various classifiers, such as linear discriminant analysis (LDA), neural network (NN), and support vector machine (SVM) can be considered for this purpose. In this example, the LDA classifier is selected due to its simplicity (Duda et al. 2000). Taking the two features as inputs to the LDA classifier, the

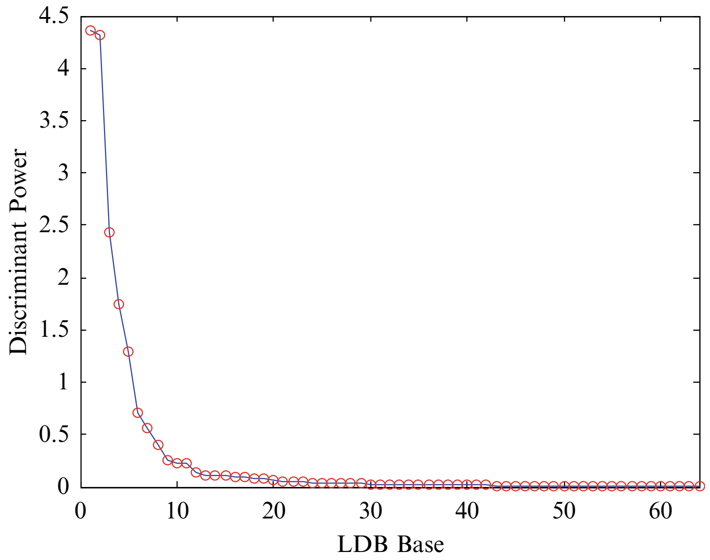

Fig. 9.5 Discriminant power of all 64 LDB bases

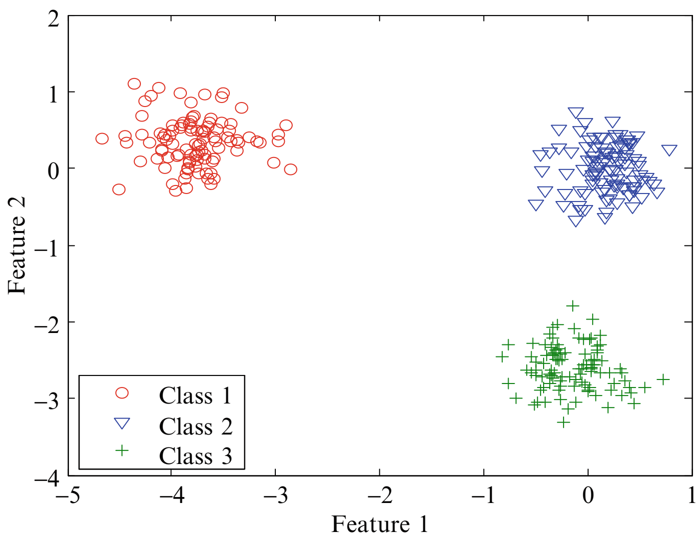

Fig. 9.6 Training signals represented by selected top two LDB features

training dataset are classified. The test signals are subsequently projected onto the top two LDB vectors to produce coefficients. Figure 9.7 indicates the scatter plot of the testing dataset by the two features. It is seen that, using the LDA classifier, all the testing data set are classified successfully.

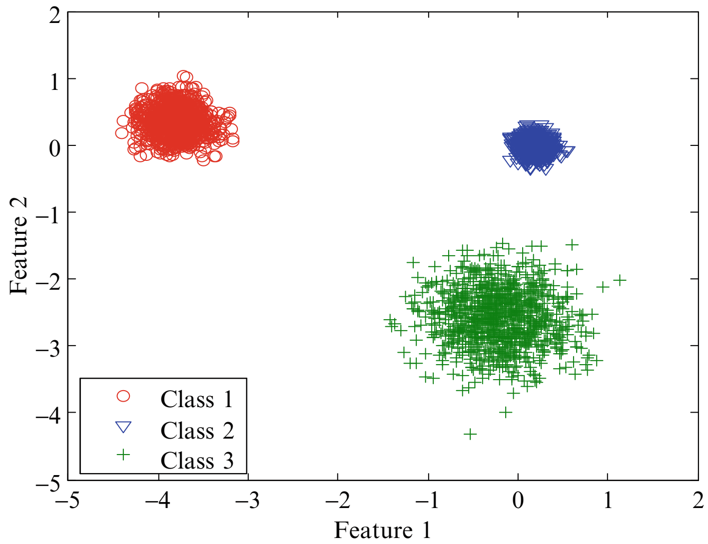

Fig. 9.7 Testing signals represented by selected top two LDB features

The above case study illustrates that the LDB-based wavelet packet base selection method is effective in producing features that enables effective discrimination of different classes.

9.4 Application to Gearbox Defect Classification

Application of the LDB algorithm to real-world classification problems has been reported in various areas, such as geophysical acoustic waveform classification (Saito and Coifman 1997), radar signal classification (Guglielmi 1997), automatic target recognition (Spooner 2001), ultrasonic echoes classification (Christian 2002), audio signal classification (Umapathy et al. 2007), biomedical signal analysis (Englehart et al. 2001; Umapathy and Krishnan 2006), and fault classification of mechanical systems (Tafreshi et al. 2005; Yen and Leong 2006). In this chapter, the LDB algorithm is applied to classifying the severity of defects in gearbox. Figure 9.8 illustrates the experimental set-up (Rafiee et al. 2007) where the vibration from a four-speed motorcycle gearbox is measured. An electrical motor drives the gearbox at a constant nominal rotational speed of 1,420 rpm. A tachometer measures the actual rotational speed to account for fluctuations caused by the load variations. Vibration signals are measured by a triaxial accelerometer mounted on the outer surface of the gearbox's housing, close to the input shaft of the gearbox. Four different working conditions of the test gear, including faultless, slight-worn, medium-worn, and with broken-teeth, are examined by analyzing the measured vibration data. The signals are sampled at 16,384 Hz. The sampled signals under the four different working conditions are shown in Fig. 9.9.

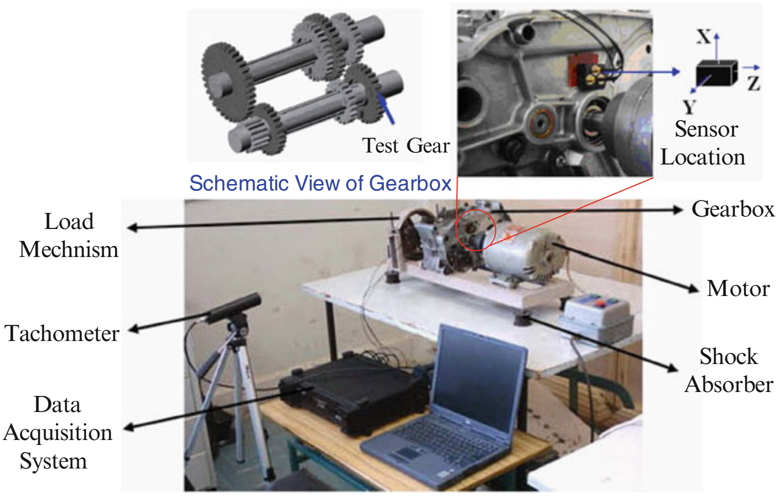

Fig. 9.8 Experimental setup to test a four-speed motorcycle gearbox

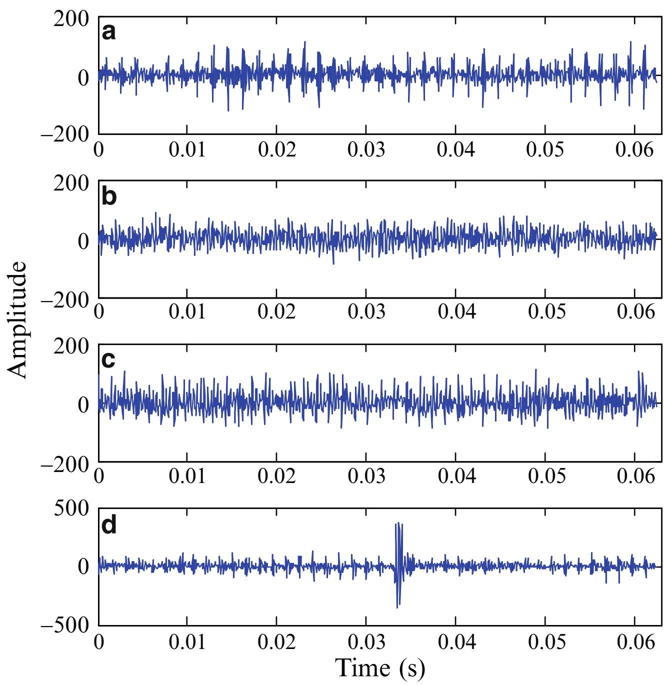

Fig. 9.9 Raw vibration signals of four gearbox conditions: (**a**) faultless, (**b**) slight-worn, (**c**) medium-worn, and (**d**) broken-teeth

To classify the gearbox defect under different working conditions, 60 training signals and 80 testing signals, each containing 1,024 data points, are segmented from the raw signal corresponding to each defect condition. The LDB algorithm is then applied to the training data, where the relative entropy was chosen as the discriminant measure. Figure 9.10 shows the wavelet packet nodes selected from a four-level signal decomposition, and the first 6 LDB bases are shown in Fig. 9.11

To evaluate the effectiveness of the LDB algorithm, the performance of classification by using the testing data are compared between the LDB-selected nodes and all the 16 nodes at the 4th decomposition level. The energy values from the selected nodes are calculated and then used as inputs to a LDA classifier for characterizing the severity of the defect. Figures 9.12 and 9.13 illustrate the distribution of two energy features from node (3, 4) and node (4, 15) for the training and test data, respectively. The results of classification of gearbox defect severity are listed in Table 9.1. It is seen that, for the training data, although the misclassification rate for features with and without basis selection is the same, the LDB-selected features have resulted in a lower misclassification rate (1.56%) than those without (2.19%), for the testing data. This indicates the merit of the LDB in signal classification.

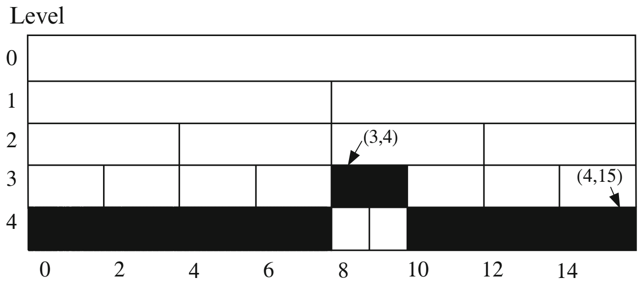

Fig. 9.10 Selected wavelet packet nodes for the gearbox data by the LDB algorithm

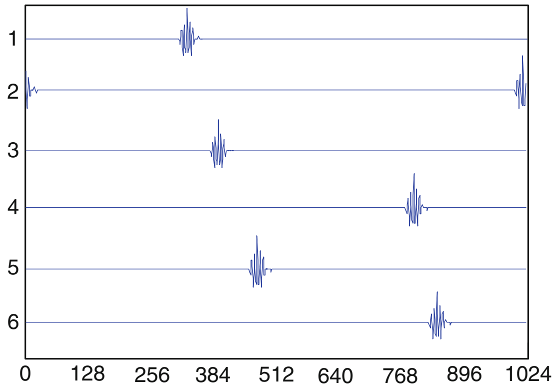

Fig. 9.11 The first six LDB bases selected from the gearbox vibration signals

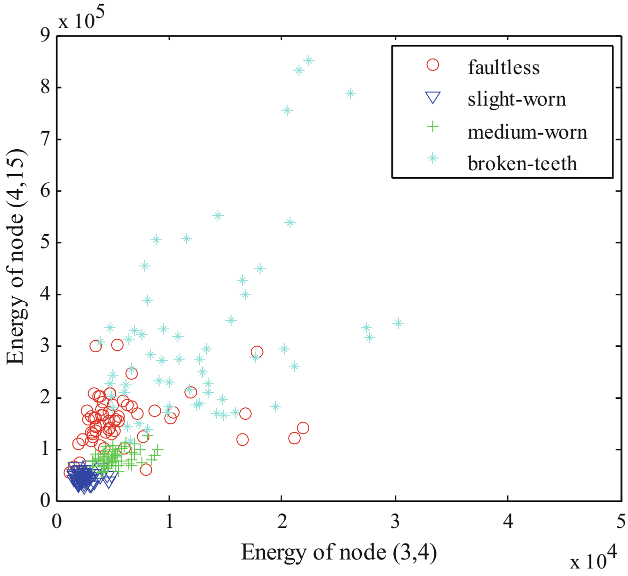

Fig. 9.12 Illustration of training samples by two features

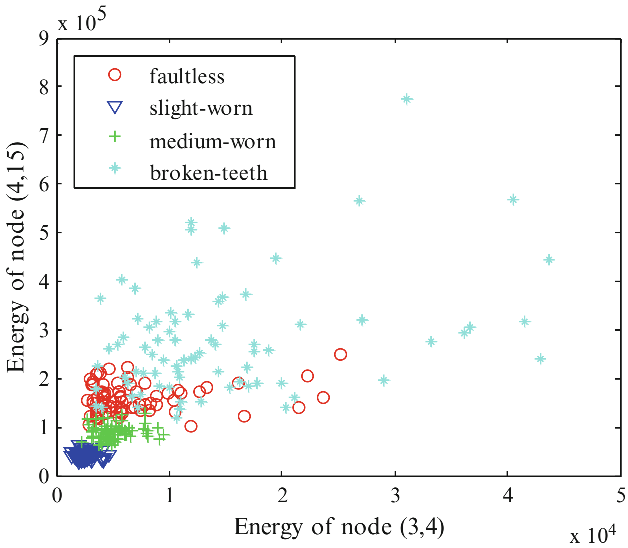

Fig. 9.13 Illustration of testing samples by two features

Table 9.1 Effect of wavelet packet node selection on gearbox defect severity classification

Features	Misclassification rate	
	Training signals (%)	Testing signals (%)
Wavelet packet nodes w/o LDB	0.83	2.19
Wavelet packet nodes w/LDB	0.83	1.56

9.5 Summary

The LDB provides an effective platform for the wavelet packet transform to decompose and classify signals. In this chapter, we have demonstrated that, using this approach, working conditions of a gearbox can be successfully classified by analyzing the measured vibration signals. Research on the theory of LDB has been continued in recent years. For example, the probability density of each class is estimated from the wavelet packet nodes to select the discriminant bases (Saito et al. 2002), and the features derived from this approach has been shown to be more sensitive to phase shifts than those from the original LDBs. Combing the LDB algorithm with signal-adapted filter banks (Strauss et al. 2003), a shape-adapted LDB approach has been developed for bio-signal processing. It can be expected that more powerful algorithms and computational tools are yet to come to better serve the need for signal classification in manufacturing.

9.6 References

Basseville M (1989) Distance measures for signal processing and pattern recognition. Signal Processing 8(4):349–369

Christian B (2002) Local discriminant bases and optimized wavelet to classify ultrasonic echoes: application to indoor mobile robotics. Proc IEEE Sens 1654–1659

Coifman RR, Wicherhauser MV (1992) Entopy-based algorithms for best basis selection. IEEE Trans Inf Theory 38(2):713–718

Duda RO, Hart PE, Stork DG (2000) Pattern classification, 2nd edn.Wiley-Interscience, New York

Englehart K, Hudgins B, Parker PA (2001) A wavelet-based continuous classification scheme for multifunction myoelectric control. IEEE Trans Biomed Eng 48(3):302–311

Guglielmi RJM (1997) Wavelet feature definition and extraction for classification and image processing. Ph.D. Dissertation, Yale University

Rafiee J, Arvani F, Harifi A, Sadeghi MH (2007) Intelligent condition monitoring of a gearbox using artificial neural network. Mech Syst Signal Process 21:1746–1754

Saito N (1994) Local feature extraction and its applications using a library of bases, Ph.D. Dissertaion, Yale University

Saito N, Coifman RR (1995) Local discriminant bases and their applications. J Math Imaging Vis 5(4) 337–358

Satio N, Coifman RR (1997) Extraction of geological information from acoustic well-logging waveforms using time-frequency wavelets. Geophysics 62(6):337–358

Saito N, Coifman RR, Geshwind FB, Warner F (2002) Discriminant feature extraction using empirical probability density estimation and a local basis library. Pattern Recognit 35:2841–2852

Spooner CM (2001) Application of local discriminant bases to HRR-based ATR. In: Proceeding of thirty-fifth asilomar conference on signals, systems and computers, pp 1067–1073

Strauss DJ, Steidl G, Delb W (2003) Feature extraction by shape-adapted local discriminant bases. Signal Processing 83:359–376

Tafreshi R, Sassani F, Ahmadi H, Dumontb G (2005) Local discriminant bases in machine fault diagnosis using vibration signals. Integr Comput Aided Eng, 12:147–158

Umapathy K, Krishnan S (2006) Modified local discriminant bases algorithm and its application in analysis of human knee joint vibration signals. IEEE Trans Biomed Eng 53(3):517–523

Umapathy K, Krishnan S, Rao RK (2007) Audio signal feature extraction and classification using local discriminant bases. IEEE Trans Audio Speech Lang Processing 15(4):1236–1246

Yen GG, Leong WF (2006) Fault classification on vibration data with wavelet based feature selection scheme. ISA Trans 45(2):141–151