

Chapter 6

Gaussian Mixture Models and Hidden Markov Models for Condition Monitoring

6.1 Introduction

Rotating machines are widely used in industry for system operation and process automation. Research shows that the failures of these machines are often linked with bearing failures (Lou et al. 2004). Bearing faults induce high bearing vibrations which generate noise that may even cause the entire rotating machine, such as the electric motor, to function incorrectly. Thus, it is important to include bearing vibration fault detection and diagnosis in industrial motor rotational fault diagnosis systems (Lou et al. 2004). As a result, there is a high demand for cost effective automatic monitoring of bearing vibrations in industrial motor systems.

A variety of fault bearing vibration feature detection techniques exist. These can be classified into three domains, namely: frequency domain analysis, time-frequency domain analysis, and time domain analysis (Ericsson et al. 2004). The frequency domain methods often involve frequency analysis of the vibration signals and look at the periodicity of high frequency transients. This procedure is complicated by the fact that this periodicity may be suppressed (Ericsson et al. 2004). The most commonly used frequency analysis technique for detection and diagnosis of bearing fault is the *envelope analysis*. More details on this technique are found in McFadden and Smith (1984). The main disadvantage of the frequency domain analysis is that it tends to average out transient vibrations and therefore becomes more sensitive to background noise. To overcome this problem, the time-frequency domain analysis is used, which shows how the frequency contents of the signal changes with time. Examples of such analyses are: Short Time Fourier Transform (STFT), the Wigner-Ville Distribution (WVD) and, most importantly, the Wavelet Transform (WT). These techniques are studied in detail in the work of Li et al. (2000).

The last category of the feature detection is the time domain analysis. There are a number of time domain methods that give reasonable results. These methods include the time-series averaging method, the signal enveloping method, the Kurtosis method, and others (Li et al. 2000). Research shows that, unlike the frequency

domain analysis, this technique is less sensitive to suppressions of the impact of periodicity (Ericsson et al. 2004; Li et al. 2000). This chapter introduces a new time domain analysis method, known as *fractal dimension analysis*, which was originally used in image processing and has been recently used in speech recognition (Maragos and Sun 1993; Maragos and Potamianos 1999; Wang et al. 2000). This method is expected to give enormous improvement to the performance of the bearing fault detection and diagnosis because it extracts the non-linear vibration features of each bearing fault. The fractal dimension analysis is based on the Multi-scale Fractal Dimensions (MFD) of short-time bearing vibration segments, derived from non-linear theory (Wang et al. 2000).

Once the bearing vibration features are extracted using one of the three domains mentioned above, then these features can be used for automatic motor bearing fault detection and diagnosis by applying them to a non-linear pattern classifier. The most popular classifier used in bearing fault detection is a Neural Network (NN). Nevertheless, other non-linear classifiers like Gaussian Mixture Model (GMM) and Hidden Markov Model (HMM) have been shown to outperform NN in a number of classification problems, in general, and in speech related problems in particular. Only recently, have researchers such as Purushothama et al. (2005) applied speech pattern classifiers, such as HMM, to the fault detection of mechanical systems because of their success in speech recognition.

This chapter presents a comparative study of HMM and GMM, and introduces time-domain analysis based techniques using fractals to extract the features. Furthermore, the ability of MFD to detect bearing faults was evaluated using both HMM and GMM non-linear pattern classifiers.

The rest of the chapter is arranged as follows: the next section presents the different bearing faults studied in this chapter, followed by the mathematical background to fractal dimensions, HMM, and GMM. Thereafter, the time domain bearing detection and diagnosis framework is presented.

6.2 Background

This section presents, in detail, the different bearing faults studied in this chapter, followed by the mathematical background to fractal dimensions, HMM, and GMM.

6.2.1 The Gaussian Mixture Model (GMM)

A GMM is a weighted sum of M component Gaussian densities, $p(\mathbf{x}|\lambda)$ as given by the equation (Reynolds 1992; Dempster et al. 1977):

$$p(\mathbf{x}|\lambda) = \sum_{i=1}^M w_i p_i(\mathbf{x}) \quad (6.1)$$

with

$$p_i(\mathbf{x}_t) = \frac{1}{(2\pi)^{D/2} \sqrt{\Sigma_i}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_t - \boldsymbol{\mu}_i)' (\Sigma_i)^{-1} (\mathbf{x}_t - \boldsymbol{\mu}_i) \right\} \quad (6.2)$$

Here, \mathbf{x} is a D-dimensional, continuous-valued data vector representing measurements from features, w_i , $i = 1, \dots, M$, are the mixture weights, with mean vector $\boldsymbol{\mu}_i$ and covariance matrix Σ_i . The mixture weights, w_i , satisfy the constraint $\sum_{i=1}^M w_i = 1$.

The entire GMM is parameterized by the mean vectors, covariance matrices, and mixture weights from all component densities and these parameters are together represented by the notation (Reynolds and Rose 1995; Dempster et al. 1977):

$$\lambda = \{\mathbf{w}, \boldsymbol{\mu}, \Sigma\} \quad (6.3)$$

Here, λ is the model, \mathbf{w} , $\boldsymbol{\mu}$, Σ are, respectively, the weights, means, and covariance of the features. The covariance matrices can be full rank or constrained to be diagonal but, in this chapter assumes that it is diagonal. The choice of model architecture, which are the number of components, diagonal covariance matrices and parameter is usually determined by the amount of data available for approximating the GMM parameters and how the GMM is applied in a specific fault identification problem. GMM has the advantage of being able to represent a large class of sample distributions and to form smooth estimates to indiscriminately shaped probability densities.

Given a collection of training vectors, the parameters of this model are estimated by a number of algorithms such as the *Expectation-Maximization (EM)* algorithm and K-means algorithm (Dempster et al. 1977; Reynolds et al. 2000). The EM algorithm was used in this study because it has reasonably fast computational time when compared to other algorithms. The EM algorithm finds the optimum model parameters by iteratively refining GMM parameters to increase the likelihood of the estimated model for the given bearing fault feature vector. More details on the EM algorithm for training a GMM are in the work of Wang and Kootsookos (1998).

Bordes et al. (2007) applied the EM algorithm to image reconstruction. They found that the results were within 10% of the experimental data. Dempster et al. (1977) applied the EM algorithm to missing data, while Ingrassia and Rocci (2007) generalized the EM algorithm to semi-parametric mixture models that, when tested on real data, showed that their method was easy to implement and computationally efficient. Kauermann et al. (2007) used the EM algorithm to recognize polymorphism in pharmacokinetic/pharmacodynamic (PK/PD) phenotypes, while Wang and Hu (2007) improved the EM algorithm's computational load and successfully applied this to brain tissue segmentation. Another successful implementation of the EM algorithm includes binary text classification (Park et al. 2007). Other improvements of the EM algorithm include accelerating the computational speed by Patel et al. (2007). Further information on the implementation of the EM algorithm can be found in Wang et al. (2007), as well as McLachlan and Krishnan (1997).

The aim of maximum likelihood estimation is to identify the model parameters which maximize the likelihood of the GMM, given the training data. For a series of T training vectors $\mathbf{X} = \{x_1, \dots, x_T\}$, the GMM likelihood, assuming independence between the vectors, can be expressed as (Reynolds 1992):

$$p(\mathbf{X}, \lambda) = \prod_{t=1}^T p(\mathbf{x}_t, \lambda) \quad (6.4)$$

For the EM algorithms, the re-estimations are calculated until convergence; and the mixture of weights, means, and variances can, respectively, be written as (Reynolds 1992):

$$\bar{w}_i = \frac{1}{T} \sum_{t=1}^T P(i | \mathbf{x}_t, \lambda) \quad (6.5)$$

$$\bar{\mu}_i = \frac{\sum_{t=1}^T P(i | \mathbf{x}_t, \lambda) \mathbf{x}_t}{\sum_{t=1}^T P(i | \mathbf{x}_t, \lambda)} \quad (6.6)$$

$$\bar{\sigma}_i^2 = \frac{\sum_{t=1}^T P(i | \mathbf{x}_t, \lambda) \mathbf{x}_t^2}{\sum_{t=1}^T P(i | \mathbf{x}_t, \lambda)} - \bar{\mu}_i^2 \quad (6.7)$$

The posterior probability can thus be written as (Reynolds 1992):

$$P(i | \mathbf{x}_t, \lambda) = \frac{w_i p(\mathbf{x}_t | \boldsymbol{\mu}_i, \Sigma_i)}{\sum_{k=1}^M w_k p(\mathbf{x}_t | \boldsymbol{\mu}_k, \Sigma_k)} \quad (6.8)$$

The bearing fault detection or diagnosis using this classifier is then achieved by computing the likelihood of the unknown vibration segment of the different fault models. This likelihood is given by (Dempster et al. 1977):

$$\hat{s} = \arg \max_{1 \leq f \leq F} \sum_{k=1}^K \log p(\mathbf{x}_k | \lambda_f) \quad (6.9)$$

Here F represents the number of faults to be diagonalized, $\mathbf{X} = \{x_1, x_2, \dots, x_K\}$ is the unknown D-dimension bearing fault-vibration segment.

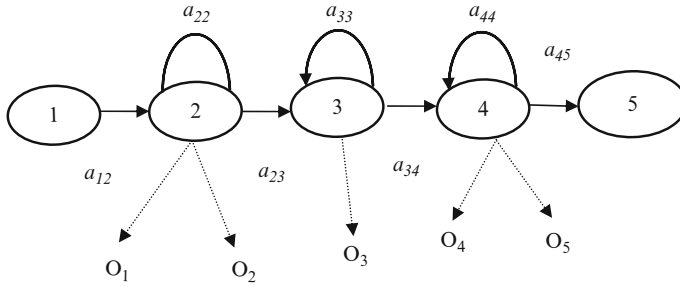


Fig. 6.1 Markov chain with five states with selected state transitions. Here O is the observation and a is the transition probability

6.2.2 The Hidden Markov Model (HMM)

The HMM is a statistical Markov model in which the system being modeled is assumed to be a Markov process with states that are hidden and therefore cannot be observed. In a conventional Markov model, the state is observable, and consequently, the transition probabilities are the only parameters to be estimated while naturally, the output is visible.

Essentially, HMM is a stochastic signal model. HMMs are referred to as Markov sources or probabilistic functions of Markov chains (Rabiner 1989). This model has been applied mostly to speech recognition systems and only recently it has been applied to bearing fault detection. In HMM, the observation is a probabilistic function of the state and this means the resulting model is a doubly emended stochastic process with an underlining stochastic process that is not observable (Rabiner 1989). Nevertheless, this process can only be observed through another stochastic process that produces the sequence. There are a number of possible Markov models, but the left-to-right model is typically applied in speech recognition. The structure of this model is shown in Fig. 6.1 with five states (Rabiner 1989).

Marwala et al. (2006) used bearing vibration signals features which were extracted using a time-domain fractal-based feature extraction technique as well as the HMM and GMM for bearing fault detection. The fractal technique was the Multi-Scale Fractal Dimension and was estimated using the Box-Counting Dimension. The extracted features were then applied to classify faults using the GMM and HMM. The results showed that the HMM outperformed the GMM and that the HMM was computationally more expensive than the GMM.

Boutros and Liang (2011) applied the discrete HMM for the detection and diagnosis of bearing and cutting tool faults. Their method was tested and validated using two situations, tool fracture, and bearing faults. In the first situation, the model correctly detected the state of the tool and, in the second case; the model

classified the severity of the fault seeded into two different engine bearings. The result obtained for fault severity classification was above 95%. In addition to the fault severity, a location index was developed to determine the fault location and gave an average success rate of 96%.

Wong and Lee (2010) successfully applied HMM for fault detection in the shell-and-tube heat exchanger. This method was viewed as a generalization of the mixture-of-Gaussians method and was demonstrated through a problem.

Lee et al. (2010) applied HMM for online degradation assessment and adaptive fault detection of multiple failure modes. Their method, together with statistical process control was used to detect the incidence of faults. This technique permitted the hidden Markov state to be updated with the identification of new states. The results for a turning process showed that the tool wear processes can be successfully detected, and the tool wear processes can be identified.

Calefati et al. (2006) successfully applied HMM for machine faults detection and forecasting in gearboxes. Elsewhere, Zhou and Wang (2005) applied HMM and a principal component analysis to the on-line fault detection and diagnosis in industrial processes, and applied these to case studies from the Tennessee Eastman process.

Menon et al. (2003) applied HMM for incipient fault detection and diagnosis in turbine engines and the effectiveness of the HMM method was compared to a neural network method and a hybrid of principal component analysis and a neural network approach. Their HMM method was found to be more effective than the other methods.

Smyth (1994) applied HMM to fault detection in dynamic systems. It was demonstrated that a pattern recognition system combined with a finite-state HMM was good at modeling temporal characteristics. The model was validated using a real-world fault diagnosis problem and was demonstrated to offer substantial practical advantages.

The complete parameter set needed to define the HMM can be written as (Rabiner 1989; Caelli et al. 2001; Koski 2001):

$$\lambda = \{A, B, \pi\} \quad (6.10)$$

where λ is the model, $A = \{a_{ij}\}$, $B = \{b_{ij}(k)\}$ and $\pi = \{\pi_i\}$ are the transition probability distribution, the observation probability distribution, and initial state distribution, respectively. For example, if we assume that the distribution can be represented by the Gaussian mixture model shown in Eq. 6.2, the equation can be written as:

$$\lambda = \{A, \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \pi\} \quad (6.11)$$

These parameters of a given state, S_i , are defined as (Rabiner 1989; Ching et al. 2003; Purushothama et al. 2005; Ching and Ng 2006):

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i), 1 \leq i, j \leq N \quad (6.12)$$

$$b_{ij}(k) = P(O_k | q_t = S_i), 1 \leq j \leq N, 1 \leq k \leq M \quad (6.13)$$

and

$$\pi_i = P(q_1 = S_i), 1 \leq i \leq N \quad (6.14)$$

Here, q_t is the state at time t and N denotes the number of states. Additionally, O_k is the k^{th} observation and M is the number of distinct observation.

The HMM can be used to simulate the observed state as follows (Rabiner 1989):

1. Let $t = 1$.
2. Create $O_t = v_k \in V$ in accordance with the probability $b_i(k)$.
3. Create a transition of hidden state from $q_t = S_i$ to $q_{t+1} = S_j$ in accordance with the transition probability a_{ij} .
4. Let $t = t + 1$ and go to Step 2 if $t < T$ or else terminate the algorithm.

There are three fundamental issues to be solved for this model to be applied in practice. Firstly, we ought to identify the probability of the observation sequence $\mathbf{O} = O_1, O_2, \dots, O_T$ of visible states generated by the model λ . Secondly, we need a decoding process which identifies a state sequence that maximizes the probability of an observation sequence and this can be realized through the so-called Viterbi algorithm (Rabiner 1989). Thirdly, we need a training process which adjusts model parameters to maximize the probability of the observed sequence.

The next step is to calculate the likelihood of the observed sequence as follows (Rabiner 1989; Ching et al. 2004):

$$P(\mathbf{O}) = \sum_{\text{all possible}} \pi_{q_1} b_{q_1}(O_1) \times \pi_{q_2} b_{q_2}(O_2) \times \dots \times \pi_{q_n} b_{q_n}(O_n) \quad (6.15)$$

To speed up the computation of this, the backward and the forward methods can be applied (Baum 1972). To do this, we define the following (Baum 1972):

$$\alpha_T(i) = P(O_1 O_2 \dots O_T, q_T = S_i) \quad (6.16)$$

The forward technique can be written as follows (Rabiner 1989; Tai et al. 2009):

1. Initialize as follows:

$$\alpha_1(i) = \pi_i b_i(O_1) \text{ for } 1 \leq i \leq N$$

2. Apply the recursion step as follows:

$$\alpha_t(j) = b_j(O_t) \sum_{i=1}^N \alpha_{t-1}(i) a_{ij} \text{ for } 2 \leq t \leq T \text{ and } 1 \leq j \leq N$$

3. Terminate as follows:

$$P(\mathbf{O}) = \sum_{i=1}^N \alpha_T(i)$$

The backward technique can be written as follows by letting (Rabiner 1989; Tai et al. 2009):

$$\beta_t(i) = P(O_{t+1}O_{t+2}\dots O_T | q_t = S_i) \quad (6.17)$$

1. Initialize as follows:

$$\beta_T(i) = 1 \text{ for } 1 \leq i \leq N$$

2. Apply the recursion step as follows:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \text{ for } 1 \leq t \leq T - 1 \text{ and } 1 \leq j \leq N$$

3. Terminate as follows:

$$P(\mathbf{O}) = \sum_{i=1}^N \beta_1(i) \pi_i b_i(O_1)$$

The Baum-Welch estimation procedures with the Maximum Likelihood technique can be used to approximate the model parameters (Rabiner 1989). To explain the use of this procedure it is important to state the following definition (Rabiner 1989; Ching et al. 2004):

$$\xi_t(i, j) = P(q_t = S_i, q_{t+1} = S_j | O, A, B, \pi) \quad (6.18)$$

This is the probability of being in state S_i at time t and having a transition to state S_j at time $t + 1$ given the observed sequence and the model. This can be expanded

as follows (Rabiner 1989; Ching et al. 2004):

$$\xi_t(i, j) = \frac{\alpha_t(i)\alpha_{ij}\beta_{t+1}(j)b_j(O_{t+1})}{\sum_i \sum_j \alpha_t(i)\alpha_{ij}\beta_{t+1}(j)b_j(O_{t+1})} \quad (6.19)$$

We can also define that (Rabiner 1989; Ching et al. 2004):

$$\gamma_t(i) = P(q_t = S_i | O, A, B, \pi) \quad (6.20)$$

This indicates the probability of being in state S_i at time t given the observed sequence and the model. Therefore, we now have (Rabiner 1989; Ching et al. 2004):

$$\gamma_t = \sum_j \xi_t(i, j) \quad (6.21)$$

This procedure can be written as follows (Baum 1972; Rabiner 1989):

1. Select a set of initial parameters $\lambda = \{A, B, \pi\}$ randomly
2. Estimate the parameters using the following equations (Tai et al. 2009)

$$\bar{\pi}_i = \gamma_1(i) \text{ for } 1 \leq i \leq N$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \text{ for } 1 \leq i \leq N, 1 \leq j \leq N$$

$$\bar{b}_j(k) = \frac{\sum_{t=1}^T \gamma_t(j) I_{O_t=k}}{\sum_{t=1}^T \gamma_t(j)} \text{ for } 1 \leq j \leq N, 1 \leq k \leq M$$

$$\text{Here } I_{O_t=k} = \begin{cases} 1 & \text{if } O_t = k \\ 0 & \text{otherwise} \end{cases}$$

3. Set $\bar{A} = \{\bar{a}_{ij}\}_{ij}$, $\bar{B} = \{\bar{b}_j(k)\}_{jk}$ and $\bar{\pi} = \{\bar{\pi}_i\}$
4. Set $\bar{\lambda} = \{\bar{A}, \bar{B}, \bar{\pi}\}$
5. If $\lambda = \bar{\lambda}$, end otherwise let $\lambda = \bar{\lambda}$ and go to Step 2

A more detailed explanation of HMM training using the Baum-Welch re-estimation along with other features of HMM is presented by Rabiner (1989).

The estimation of the hidden state can be conducted using the Viterbi algorithm (Viterbi 1967) to calculate the probability of the hidden states given the HMM parameters and an observed sequence. To do this we can define the following (Rabiner 1989; Tai et al. 2009) which is the maximum probability within a single path:

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P(q_1, q_2, \dots, q_t, O_1 O_2 \dots O_t; q_t = S_i) \quad (6.22)$$

and define (Tai et al. 2009):

$$\delta_t(j) = b_j(O_t) \times \max_i \{\delta_{t-1}(i)a_{ij}\} \quad (6.23)$$

We can therefore solve this problem using dynamic programming as follows (Rabiner 1989; Tai et al. 2009):

1. Initialize $\delta_1(i) = \pi_i b_i(O_1)$ and $\theta_1(i) = 0$ for $1 \leq i \leq N$
2. Solve the following recursion step

$$\delta_t(j) = \max_{1 \leq i \leq N} \delta_{t-1}(i)a_{ij} b_j(O_t) \text{ for } 2 \leq t \leq T \text{ and } 1 \leq j \leq N$$

and

$$\theta_t(j) = \arg \max_{1 \leq i \leq N} \{\delta_{t-1}(i)a_{ij}\} \text{ for } 2 \leq t \leq T \text{ and } 1 \leq j \leq N$$

3. Terminate

$$P^* = \max_{1 \leq i \leq N} \delta_T(i) \text{ and } q_T^* = \arg \max_{1 \leq i \leq N} \delta_T(i)$$

Here P^* is the most likely likelihood and q^* is the most likely state at time T .

4. Backtrack:

$$q_t^* = \theta_{t+1}(q_{t+1}^*) \text{ for } t = T - 1, T - 2, \dots, 2, 1$$

6.2.3 Fractals

For this chapter, fractals were used to analyse the bearing data. A *fractal* is defined as a rough geometric shape that can be divided into various segments, each of which is roughly a reduced-size copy of the whole. This characteristic is known as self-similarity (Mandelbrot 1982). The theory of fractals was described in detail, in Chap. 2. The basis of the idea of fractals extends back to the seventeenth century. There are numerous classes of fractals, characterized as displaying exact self-similarity, quasi self-similarity or statistical self-similarity (Briggs 1992). Even though fractals are a mathematical concept, they are seen in nature, and this has led to their use in the arts they are useful in biomedical sciences, engineering sciences, and speech recognition. A fractal usually has the following characteristics (Falconer 2003):

- It contains a fine structure at randomly small scales.
- It is too irregular to be described using Euclidean geometry.
- It is approximately self-similar.

- It has a Hausdorff dimension (this explained in Chap. 2) which is more than its topological dimension (Pickover 2009).
- It has a basic and recursive description.

Fractals are frequently viewed to be infinitely complex because they look similar at all levels of magnification (Batty 1985; Russ 1994). Examples of natural objects that can be approximated by fractals include clouds, mountain ranges, lightning bolts, and coastlines (Sornette 2004).

Zhang et al. (2010) successfully applied a combined wavelet and fractal method for the fault detection of the opening fault of power electronic circuits based on the singularity of the fault signal from the power electronic equipment. Voltage wave signals were analyzed by applying the wavelet transform and correlative dimensions of the wavelet transform were estimated using fractals.

Yang et al. (2011) applied a fractal correlation dimension for the fault detection in the supply air temperature sensors of air handling unit systems and the results obtained demonstrated that it was more efficient in detecting a relatively small bias fault under noise conditions.

Ikizoglu et al. (2010) applied a Hurst parameter and fractal dimension for fault detection of the bearings in electric motors. The vibration signals were obtained, analyzed in the frequency domain.

Ma (2009) successfully applied fractal analysis for fault detection in the welding process while Shanlin et al. (2007) successfully applied wavelet fractal network for fault detection in a power system generator.

Other successful applications of the wavelet transform in fault detection include its application for distributed power system short-circuit problems (Song et al. 2007), robotics (Yan et al. 2007), short-circuit faults in low-voltage systems (Kang et al. 2006), and Direct Current system grounding (Li et al. 2005).

6.3 Motor Bearing Faults

Vibration measurement is important in advanced conditioning monitoring of mechanical systems. Most bearing vibrations are periodical movements. In general, rolling bearing contains two concentric rings, which are called the inner and outer raceway and these were shown in Chap. 2 (Li et al. 2000). Furthermore, the bearing contains a set of rolling elements that run in the tracts of these raceways. There is a number of standard shapes for the rolling elements such as a ball, cylindrical roller, tapered roller, needle roller, symmetrical and unsymmetrical barrel roller and many more as described by Ocak and Loparo (2004). In this chapter, a ball rolling element is used as was done by Ocak and Loparo (2004).

Three faults are studied in this chapter. These are an inner raceway, an outer raceway, and a rolling element fault. A bearing fault increases the rotational friction of the rotor and, therefore, each fault gives vibration spectra with unique frequency components (Ericsson et al. 2004). It should be taken into account that these

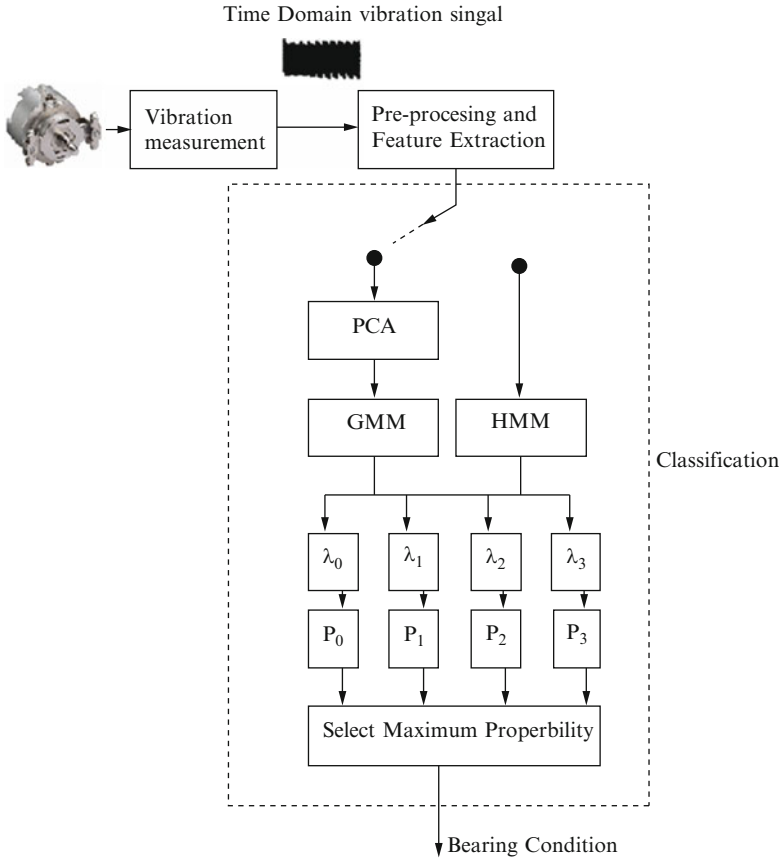


Fig. 6.2 Motor bearing fault detection and diagnosis system

frequency components are a linear function of the running speed and that the two raceway frequencies are also linear functions of the number of balls. The motor bearing condition monitoring systems was implemented by analyzing the vibration signal of all the bearing faults. The vibration signal was produced by Ocak and Loparo (2004) using the impact pulse generated when a ball roller knocks a defect in the raceways or when the defect in the ball knocks the raceways (Li et al. 2000).

The studied motor bearing fault detection and diagnosis system is displayed in Fig. 6.2 (Marwala et al. 2006). The system consists of two major stages after the vibration signal measurement and these are the pre-processing which includes both the feature extraction phase and classification phase.

The initial phase of an automatic fault detection and diagnosis system, as indicated in Fig. 6.3, is signal preprocessing and feature extraction (Marwala et al. 2006). Faults cause a change in the machinery vibration levels and, consequently, the information regarding the health status of the monitored machine is largely contained in the vibration time signal (McClintic et al. 2000). Figure 6.4 shows

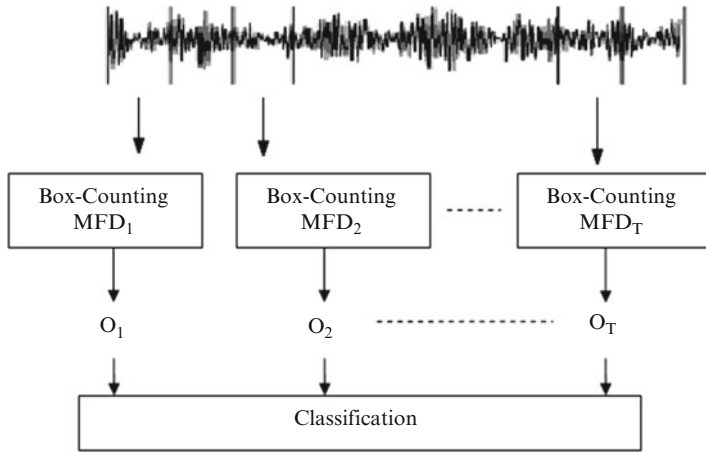


Fig. 6.3 Vibration signal pre-processing and feature extraction

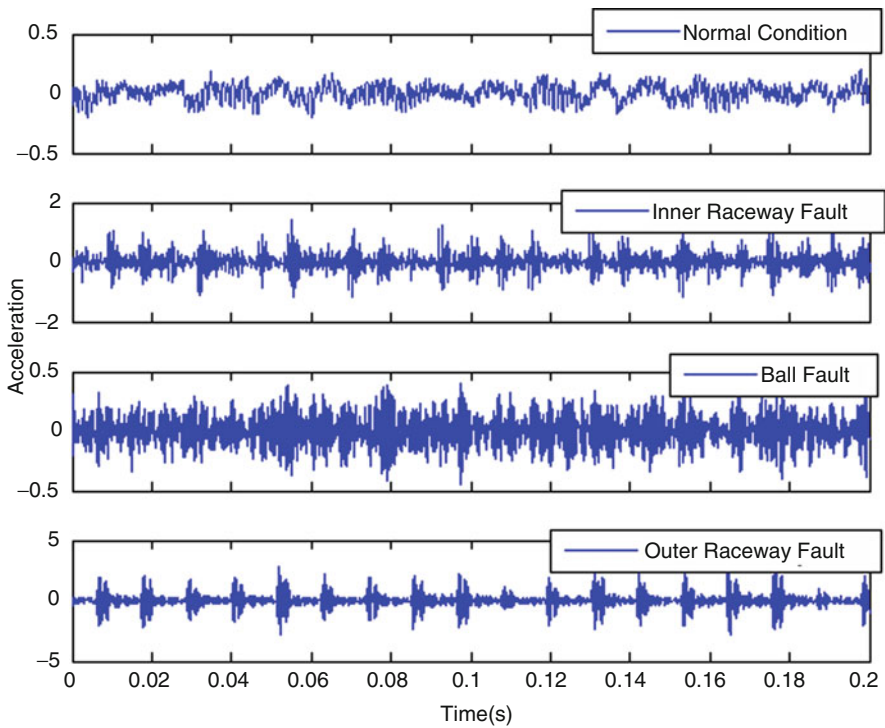


Fig. 6.4 The first 2 s of the vibration signal of the normal, inner raceway fault, ball fault, and outer raceway fault

that the signal is preprocessed by dividing the vibration signals into T windows of equal lengths (Marwala et al. 2006). For this technique to be effective, it should be noted that the width of the window must be more than one revolution of the bearing to ensure that the uniqueness of each vibration fault signal is captured. The preprocessing is followed by extraction of features of each window using the Box-Counting MFD, which forms the observation sequence to be used by the GMM or the HMM classifier. The time domain analysis extracts the non-linear turbulence information of the vibration signal and is expected to give enormous improvement on the performance of the bearing fault detection and diagnosis process.

Due to the large variations of the vibration signal, direct comparison of the signals is difficult. Hence, non-linear pattern classification methods are used to classify different bearing fault conditions. The features extracted were used as inputs to the classification phase of the framework. This chapter compares the performance of the GMM and the HMM classifiers. For the GMM classifier, the principal component analysis (PCA), which was described in Chap. 2, was applied to the feature vector before training to reduce the dimensionality and remove redundant information (Jolliffe 1986). The principal concept behind PCA is to identify the features that explain as much of the total variation in the data as possible with as few of these features as possible. The calculation of the PCA data transformation matrix is based on the eigenvalue decomposition.

The computation of the principal components was conducted as described below (Jolliffe 1986):

- Calculate the covariance matrix of the input data.
- Compute the eigenvalues and eigenvectors of the covariance matrix.
- Preserve the largest eigenvalues and their respective eigenvectors which contains at least 90% of the data.
- Transform the original data into the reduced eigenvectors and, therefore, decrease the number of dimensions of the data.

For more information on the PCA used here to reduce the dimensionality of the feature space, the reader is referred to the work of Jolliffe (1986). In Fig. 6.3, the diagnosis of the motor bearing fault was achieved by calculating the probability of the feature vector, given the entire previously constructed fault model and then the GMM or HMM with maximum probability determined the bearing condition.

This section discusses the experimental database used to evaluate the efficiency of the proposed approach. The performance measure adopted during experimentation is also briefly discussed. The database used to validate the new bearing fault diagnosis discussed in the last section was developed at Rockwell Science Centre by Loparo in 2005. In this data set, single point faults of diameters of 7 mils, 14 mils, and 21 mils (1 mil = 0.001 in.) were introduced using electro-discharge machining. These faults were introduced separately at the inner raceway, rolling element and outer raceway. A more detailed explanation of this data set is presented in (Loparo 2006). The experiments were performed for each fault diameter and this was repeated for two load conditions, which were 1 and 2 hp. The experiment was performed for vibration signals sampled at 12,000 samples per second for the drive

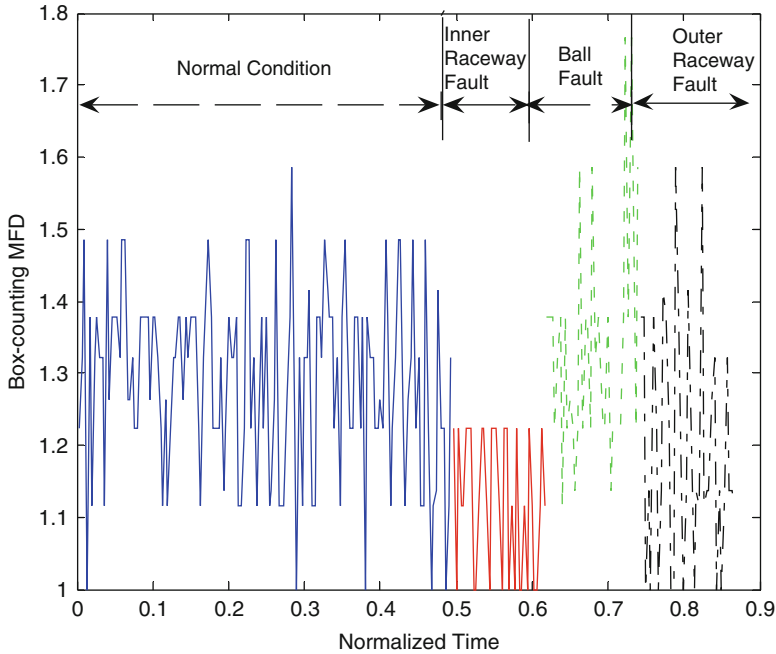


Fig. 6.5 MFD feature extraction comparison for the normal, inner, outer and ball fault for the 1 s vibration

end bearing faults. The vibration signals from this database were divided into equal windows of four revolutions. Half of the resulting sub-signals are used for training and the other half were used for testing.

The main concern was to measure the ability of the system to classify the bearing faults. The performance of the system was measured using the Classification Rate (CR) which is the proportion of fault cases correctly classified.

The optimum HMM architecture, used in the experiment was a 2 state model with a diagonal covariance matrix that contained 10 Gaussian mixtures. The GMM architecture also used a diagonal covariance matrix with three centers. The main advantage of using the diagonal covariance matrix in both cases was that this de-correlated the feature vectors. This was necessary because fractal dimensions of adjacent scales were highly correlated (Maragos and Potamianos 1999).

The first set of experiments measured the effectiveness of the time-domain fractal dimension based feature-extraction using vibration signal of the faults as shown in Fig. 6.5 (Marwala et al. 2006).

Figure 6.5 shows the first 2 s of the vibration signals used. It can be clearly seen that there is fault specific information which must be extracted. Figure 6.6 shows the MFD feature vector which extracts the bearing's fault specific information (Marwala et al. 2006). It should be noted that these features are only for the first second of the vibration signal. Figure 6.6 clearly shows that the presented feature extraction

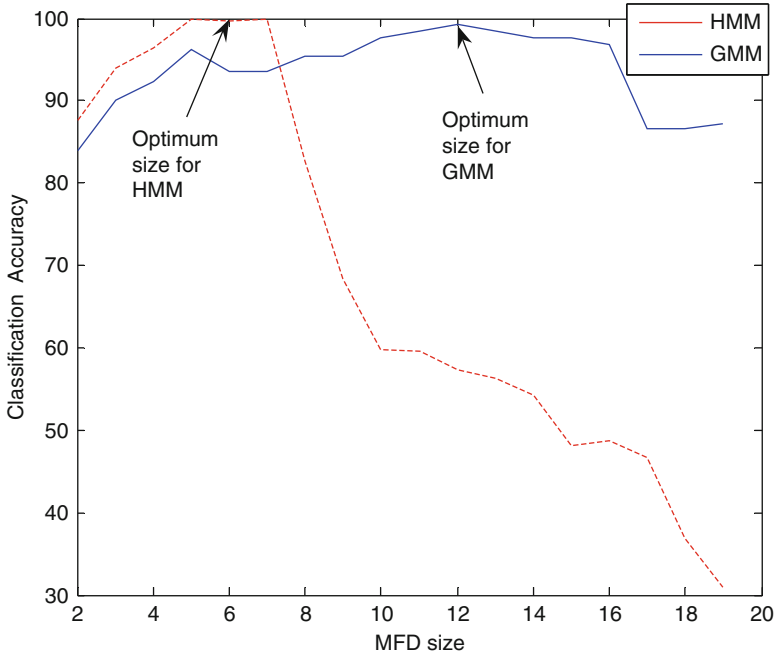


Fig. 6.6 The graph of the change classification rate with change in MFD size

Table 6.1 The classification rate for different loads and fault diameters for the GMM and HMM classifier

Load	7 mils	14 mils	21 mils	Load	7 mils	14 mils
	HMM	GMM	HMM	GMM	HMM	GMM
1	100%	99.2%	100%	98.7%	100%	99%
2	100%	99%	100%	99.1%	100%	99%

method does indeed extract the fault specific features which are used to classify different bearing faults (Marwala et al. 2006). For this reason, the presented MFD feature extraction is expected to give enormous improvement to the performance of the bearing fault detection and diagnosis. Nevertheless, the optimum size of the MFD must be initially found. Figure 6.6 shows the graph of change of the system accuracy with the change of the MFD size. The figure shows that the GMM generally has a large optimum MFD size of 12 compared to 6 for HMM.

Having used the optimum HMM and GMM architecture discussed previously, the classification accuracy that was found for different bearing loads and different bearing fault diameters appears in Table 6.1 for the GMM and the HMM classifier.

Table 6.1 shows that the HMM outperforms the GMM classifier for all cases, with a 100% and 99.2% classification rate for HMM and GMM, respectively. Table 6.1 also shows that changing the bearing load or diameter does not significantly change the classification rate.

Using a Pentium IV with 2.4 GHz processor speed, further experimenting showed that the average training time of HMM was 19.5 s. This was more than 20 times higher than the GMM training time, which was found to be 0.83 s. In summary, even though HMM gave higher classification rate when compared to GMM it was time consuming to train the models when compared to GMM. It is probably worth mentioning that, it was observed that using the PCA dimension reduction technique does not affect the classification rate. Nevertheless, this reduced the dimension from 84 to 11, which makes GMM training even more computationally efficient when compared to training the HMM.

This chapter presented the obtained using MFD short time feature extraction. The results demonstrated that this technique does extract fault specific features. Furthermore, the results showed that for the GMM classifier using PCA, the classification rate was not affected; it simply reduced the dimensionality of the input feature vector which makes the GMM models less complex than the HMM models. Further experimentation revealed that there was an optimum MFD size which gave the optimum classification rate. From the results obtained, it was found that the GMM generally had larger optimum MFD size than the HMM.

The second set of tests that were performed compared the performance of GMM and HMM in classifying the different bearing faults. The test revealed that the HMM outperformed the GMM classifier with a classification rate of 100%. Further testing of these classifiers revealed that, the major disadvantage of the HMM classifier was that it took longer to train than the GMM classifier, even though GMM had larger MFD size than HMM. So, it is recommended that one use the GMM classifier when time is the major issue in that particular application. It was further observed that changing the bearing load or diameter does not significantly affect the classification rate of the presented framework.

6.4 Conclusions

A framework that used a time-domain fractal-based feature extraction method to extract the non-linear turbulent information of the vibration signal has been presented. Using these features together with HMM and GMM classifiers, the results showed that the HMM classifier outperformed the GMM classifier with the HMM giving 100% and the GMM 99.2% classification rate. Nevertheless, the major drawback of the HMM classifier was that it was computationally expensive, taking 20 times longer than the GMM classifier to train.

References

- Batty M (1985) Fractals – geometry between dimensions. *New Sci* 105:31–35
- Baum L (1972) An inequality and associated maximization techniques in statistical estimation for probabilistic function of Markov processes. *Inequality* 3:1–8

- Bordes L, Chauveau D, Vandekerkhove (2007) A stochastic EM algorithm for a semiparametric mixture model. *J Comput Stat Data Anal* 51:5429–5443
- Boutros T, Liang M (2011) Detection and diagnosis of bearing and cutting tool faults using hidden Markov models. *Mech Syst Signal Process* 25:2102–2124
- Briggs J (1992) *Fractals: the patterns of chaos*. Thames and Hudson, London
- Caelli T, McCabe A, Briscoe G (2001) Shape tracking and production using hidden Markov models. In: Bunke H, Caelli T (eds) *Hidden Markov models: applications in computer vision*. World Scientific, Singapore
- Calefati P, Amico B, Lacasella A, Muraca E, Zuo MJ (2006) Machinery faults detection and forecasting using hidden Markov models. In: *Proceedings of the 8th biennial ASME conference on engineering systems design and analysis*, Torino, Italy, pp 895–901
- Ching WK, Ng M (2006) *Markov chains: models, algorithms and applications*, International series on operations research and management science. Springer, New York
- Ching WK, Ng M, Fung E (2003) Higher-order hidden Markov models with applications to DNA sequences. In: Liu J, Cheung Y, Yin H (eds) *Intelligent data engineering and automated learning*. Springer, New York
- Ching WK, Ng M, Wong K (2004) Hidden Markov model and its applications in customer relationship management. *IMA J Manag Math* 15:13–24
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc B* 39:1–38
- Ericsson S, Grip N, Johansson E, Persson LE, Sjöberg R, Strömberg JO (2004) Towards automatic detection of local bearing defects in rotating machines. *Mech Syst Signal Process* 19:509–535
- Falconer K (2003) *Fractal geometry: mathematical foundations and applications*. Wiley, New Jersey
- Ikizoglu S, Caglar R, Seker S (2010) Hurst parameter and fractal dimension concept for fault detection in electric motors. *Int Rev Electric Eng* 5:980–984
- Ingrassia S, Rocci R (2007) A stochastic EM algorithm for a semiparametric mixture model. *Comput Stat Data Anal* 51:5429–5443
- Jolliffe IT (1986) *Principal component analysis*. Springer, New York
- Kang S, Wang B, Kang Y (2006) Early detection for short-circuit fault in low-voltage systems based on fractal exponent wavelet analysis. In: *Proceedings of the SPIE – the international society for optical engineering*: San Diego, California, art. no. 63574Z
- Kauermann G, Xu R, Vaida F (2007) Nonlinear random effects mixture models: maximum likelihood estimation via the EM algorithm. *Comput Stat Data Anal* 51:6614–6623
- Koski T (2001) *Hidden Markov models for bioinformatics*. Kluwer Academic, Dordrecht
- Lee S, Li L, Ni J (2010) Online degradation assessment and adaptive fault detection using modified hidden Markov model. *J Manufacturing Sci Eng Trans ASME* 132:0210101–02101011
- Li B, Chow MY, Tipsuwan Y, Hung JC (2000) Neural-network-based motor rolling bearing fault diagnosis. *IEEE Trans Ind Electron* 47:1060–1068
- Li DH, Wang JF, Shi LT (2005) Application of fractal theory in DC system grounding fault detection. *Automation Electric Power Syst* 29:53–56+84
- Loparo KA (2006) Bearing data center seeded fault test data. <http://www.eecs.case.edu/laboratory/bearing/download.htm>. Last accessed 01 June 2006
- Lou X, Loparo KA, Discenzo FM, Yoo J, Twarowski A (2004) A model-based technique for rolling element bearing fault detection. *Mech Syst Signal Process* 18:1077–1095
- Ma J (2009) The application of fractal analysis to fault detection and diagnoses in course of welded. In: *Proceedings of the 2nd international conference on model and sim*, Manchester, UK, pp 263–266
- Mandelbrot BB (1982) *The fractal geometry of nature*. W.H. Freeman and Co, San Francisco
- Maragos P, Potamianos A (1999) Fractal dimensions of speech sounds: computation and application to automatic speech recognition. *J Acoust Soc Am* 105:1925–1932
- Maragos P, Sun FK (1993) Measuring the fractal dimension of signals: morphological covers and iterative optimization. *IEEE Trans Signal Process* 41:108–121

- Marwala T, Mahola U, Nelwamondo FV (2006) Hidden Markov models and Gaussian mixture models for bearing fault detection using fractals. In: Proceedings of the IEEE international conference on neural networks, Vancouver, Canada, pp 3237–3242
- McClintic K, Lebold M, Maynard K, Byington C, Campbell R (2000) Residual and difference feature analysis with transitional gearbox data. In: Proceedings of the 54th meeting of the society for machinery failure prevention technology, Virginia Beach, pp 635–645
- McFadden PD, Smith JD (1984) Vibration monitoring of rolling element bearings by high frequency resonance technique – a review. *Tribol Int* 77:3–10
- McLachlan G, Krishnan T (1997) The EM algorithm and extensions, Wiley series in probability and statistics. Wiley, New Jersey
- Menon S, Kim K, Uluoyol O, Nwadiogbu EO (2003) Incipient fault detection and diagnosis in turbine engines using hidden Markov models. *Am Soc Mech Eng Int Gas Turbine Inst, Turbo Expo (Publication) IGTI*:493–500
- Ocak H, Loparo KA (2004) Estimation of the running speed and bearing defect frequencies of an induction motor from vibration data. *Mech Syst Signal Process* 18:515–533
- Park J, Qian GQ, Jun Y (2007) Using the revised EM algorithm to remove noisy data for improving the one-against-the-rest method in binary text classification. *Info Process Manag* 43:1281–1293
- Patel AK, Patwardhan AW, Thorat BN (2007) Acceleration schemes with application to the EM algorithm. *Comput Stat Data Anal* 51:3689–3702
- Pickover CA (2009) The math book: from Pythagoras to the 57th dimension, 250 milestones in the history of mathematics. Sterling Publishing Company, New York
- Purushothama V, Narayanan S, Suryana-rayana AN, Prasad B (2005) Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition. *NDT&E Int* 38:654–664
- Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE* 77:257–286
- Reynolds DA (1992) A Gaussian mixture modeling approach to text-independent speaker identification. PhD thesis, Georgia Institute of Technology
- Reynolds DA, Rose RC (1995) Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Trans Acoust Speech Signal Process* 3:72–83
- Reynolds DA, Quatieri TF, Dunn RB (2000) Speaker verification using adapted Gaussian mixture models. *Digit Signal Process* 10:19–41
- Russ JC (1994) Fractal surfaces. Springer, New York
- Shanlin K, Baoshe L, Feng F, Songhua S (2007) Vibration fault detection and diagnosis method of power system generator based on wavelet fractal network. In: Proceedings of the 26th Chinese control conference, Zhangjiajie, China, pp 520–524
- Smyth P (1994) Hidden Markov models for fault detection in dynamic systems. *Pattern Recognit* 27:149–164
- Song Y, Wang G, Chen X (2007) Fault detection and analysis of distributed power system short-circuit using wavelet fractal network. In: Proceedings of the 8th international conference on electron measurement and instruments, Xi'an, China, pp 3422–3425
- Sornette D (2004) Critical phenomena in natural sciences: chaos, fractals, selforganization, and disorder: concepts and tools. Springer, New York
- Tai AH, Ching WK, Chan LY (2009) Detection of machine failure: hidden Markov model approach. *Comput Ind Eng* 57:608–619
- Viterbi A (1967) Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans Info Theory* 13:260–269
- Wang F, Zheng F, Wu W (2000) A C/V segmentation method for Mandarin speech based on multi-scale fractal dimension. In: Proceedings of the international conference on spoken language process, Beijing, China, pp 648–651
- Wang H, Hu Z (2007) Speeding up HMRF.EM algorithms for fast unsupervised image segmentation by bootstrap resampling: application to the brain tissue segmentation. *Signal Process* 87:2544–2559

- Wang YF, Kootsookos PJ (1998) Modeling of low shaft speed bearing faults for condition monitoring. *Mech Syst Signal Process* 12:415–426
- Wang X, Schumitzky A, D'Argenio DZ (2007) Constrained monotone EM algorithms for finite mixture of multivariate Gaussians. *Comput Stat Data Anal* 51:5339–5351
- Wong WC, Lee JH (2010) Fault detection and diagnosis using hidden Markov disturbance models. *Ind Eng Chem Res* 49:7901–7908
- Yan H, Zhang XC, Li G, Yin J, Cheng W (2007) Fault detection for wall-climbing robot using complex wavelet packets transform and fractal theory. *Acta Photonica Sinica* 36:322–325
- Yang XB, Jin XQ, Du ZM, Zhu YH (2011) A novel model-based fault detection method for temperature sensor using fractal correlation dimension. *Build Environ* 46:970–979
- Zhang HT, An Q, Hu ZK, Chen ZW (2010) Fault detection wavelet fractal method of circuit of three-phase bridge rectifier. In: *Proceedings of the international conference on intelligent systems design and engineering applications*, Cairo, Egypt, pp 725–729
- Zhou SY, Wang SQ (2005) On-line fault detection and diagnosis in industrial processes using hidden Markov model. *Dev Chem Eng Miner Process* 13:397–406