Fault diagnosis is playing today a crucial role in industrial systems. To improve reliability, safety and efficiency advanced monitoring methods have become increasingly important for many systems. The vibration analysis method is essential in improving condition monitoring and fault diagnosis of rotating machinery. Effective utilization of vibration signals depends upon effectiveness of applied signal processing techniques. In this paper, fault diagnosis is performed using a combination between Wavelet Transform (WT) and Principal Component Analysis (PCA). The WT is employed to decompose the vibration signal of measurements data in different frequency bands. The obtained decomposition levels are used as the input to the PCA method for fault identification using, respectively, the Q-statistic, also called Squared Prediction Error (SPE) and the Q-contribution. Clearly, useful information about the fault can be contained in some levels of wavelet decomposition. For this purpose, the Q-contribution is used as an evaluation criterion to select the optimal level, which contains the maximum information. Associated to spectral analysis and envelope analysis, it allows clear visualization of fault frequencies. The objective of this method is to obtain the information contained in the measured data. The monitoring results using real sensor measurements from a pilot scale are presented and discussed.

Keywords: vibration, fault diagnosis, wavelet analysis, principal component analysis, squared prediction error

1. Introduction

Growing demand for higher performance, safety and reliability of industrial systems has increased the need for condition monitoring and fault diagnosis. During the two past decades, various monitoring methods have been developed, including dynamics, vibration, tribology and non-destructive techniques (Altmann, 1999; Yang, 2004).

The vibration analysis is one of the most important methods used for condition monitoring and fault diagnosis, because it always carries the dynamic information of a system. Effective utilization of vibration signals depends upon the effectiveness of applied signal processing techniques. The analysis of stationary vibration signals has largely been based on well-known spectral techniques such as: Fourier Transform (FT) and Short Time Fourier Transform (STFT) (Seker and Ayaz, 2002; Shibata et al., 2000). Unfortunately, these methods are not suitable for non-stationary signal analysis (Wu and Liu, 2008). In order to solve this problem, Wavelet Transform (WT) has been developed. WT is used to extract approximations and detail coefficients of measurements data with different frequency bands by using successive low-pass and high-pass filtering. This makes the application of WT for non-stationary signal processing an area of active research over the past decade. An overview of the WT used in vibration signal analysis was provided by Litak and Sawicki (2009), Al-Badour et al. (2011) and Yan et al. (2014).
The original signal using WT can be decomposed into approximations and details versions with different resolutions. The decomposed levels will not change their information in the time domain (Gaing, 2004). However, useful information can be contained in some sub-bands. So, the fault can be detected from a given level of resolution. This is based on the choice of an indicator to determine the optimal level where failure can occur. The selection of the most reliable indicator has been studied by several authors. A large number of applications have been reviewed, e.g. Prabhakar et al. (2002) selected periodic impulses of bearing faults in the time domain based on the low and high frequency nature of decomposed levels. Similar analyses were carried out by Purushotham et al. (2005) in order to extract periodic impulses from time signals using discrete wavelet transform at Mel-frequency scales. Chinmaya and Mohanty (2006) used sidebands of gear meshing frequencies as an evaluation criterion for gear faults diagnosis. Djebala et al. (2008) analyzed vibration of faults inducing periodical impulsive forces by selecting the kurtosis as indicator. In another study, Gavrovksa et al. (2009) described the optimal selection of decomposition levels in the wavelet transform used for both high-frequency and low-frequency filtering of the ECG signal. Yaqub et al. (2011) estimated the bandwidth of the resonant frequency band of vibration data by adaptive selection of wavelet decomposition levels. The adaptive criterion was based on saturated dissemination of the signal energy over the wavelet decomposition nodes.

In this work, we propose to use a combination between WT and PCA for improving the vibration monitoring. PCA is a multivariate analysis technique, also a dimension reduction technique. It finds the directions of significant variability in the data by forming linear combinations of variables. Its application for vibration analysis is suggested in several papers (De Moura et al., 2011; Shao et al., 2014).

The aim of the proposed combined method is to provide a solution to the fault diagnosis problem. WT is used to extract approximations and details vectors in order to obtain multiple data series at different resolutions. Clearly, useful information is contained in some decomposition levels. The obtained levels are used as the input to the PCA method for identifying abnormal situations in different frequency bands using the Q-statistic or Squared Prediction Error (SPE) and Q-contributions. In order to extract useful information, the Q-contribution is used as the principal criterion to select the optimal level of resolution. Associated to spectral analysis and envelope analysis, it allows clear visualization of the frequencies faults. The proposed method is evaluated using the experimental measurements data in the cases of mass unbalance and gear fault.

The remainder of this paper is structured as follows. Section 2 presents the fault diagnosis method using WT and PCA along with its formulations. The experimental setup is discussed in Section 3. The monitoring results and discussion are presented in Section 4. Finally, Section 5 concludes our work and contributions.

2. Fault diagnosis method

2.1. Wavelet transform

Wavelet Transform (WT) is a relatively new signal processing tool. Due to its strong capability in the time and frequency domain analysis, it is applied by many researchers in diverse applications (see, for example, Litak et al., 2009; Sen et al., 2010). WT describes a signal by using the correlation with translation and dilatation of a function called mother wavelet; it includes Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT). Let $s(t)$ be the signal, then the CWT of $s(t)$ is defined as

$$CWT(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} s(t) \psi^*\left(\frac{t-b}{a}\right) dt$$ (2.1)
where $\psi^*(t)$ is the conjugate function of the mother wavelet $\psi(t)$, $a$ and $b$ are the dilation (scaling) and translation (shift) parameters, respectively.

DWT is derived from discretization of CWT. The most common discretization is dyadic. DWT is found to yield fast computation of the WT. It is given by

$$DWT(j, k) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} s(t)\psi^*\left(\frac{t - 2^j k}{2^j}\right) \, dt \quad (2.2)$$

where $a$ and $b$ are replaced by $2^j$ and $2^j k$, $j$ is an integer.

A very useful implementation of DWT, called multiresolution analysis (Mallat, 1989), is demonstrated in Fig. 1. The DWT analyzes the signal at different frequency bands with different resolutions by decomposing the signal into several levels: approximations (A1, A2, A3,...) and details (D1, D2, D3,...). The signal is decomposed at the expected level. DWT employs two sets of functions, called scaling function and wavelet function (Mallat, 1989), which are associated with Low-pass (L) and High-pass (H) filters, respectively. The approximations are the high-scale, low-frequency components and the details are the low-scale, high-frequency components of the signal.

2.2. Principal Component Analysis

Principal Component Analysis (PCA) (Chiang et al., 2001; Jolliffe, 2002) is a projection statistical method used for dimensionality reduction. It produces a lower-dimensional representation in a way that preserves the correlation structure between the variables. Given a set of $n$ observations or samples and $m$ variables stacked into a matrix $X$, whose variance-covariance matrix has eigenvalue-eigenvector pairs

$$(\lambda_1, p_1), (\lambda_2, p_2), \ldots, (\lambda_m, p_m) \quad (2.3)$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_m \geq 0$.

The principal decomposition component of $X$ can be represented as

$$X = TP^T + E = \sum_{i=1}^{l} t_i p_i^T + E \quad l < m \quad (2.4)$$

where $T = [t_1, t_2, \ldots, t_l]$ is defined to be the matrix of principal component scores, $P = [p_1, p_2, \ldots, p_l]$ is the matrix of principal component loadings, $E$ is the residual matrix in the sense of the minimum Euclidean norm and $l$ is the index of the Principal Components (PCs).
The identification of the PCA model thus consists in estimating its parameters by an eigenvalue-eigenvector decomposition and determining the number of PCs \( l \) to retain. Many procedures have been proposed for selecting the number of PCs to be retained (Kano and Hasebe, 2001). In this study, we use the experiential method (Nomikos and MacGregor, 1995) which judges that the cumulative sum contribution of the anterior \( l \) PCs is higher than 0.85, as follows

\[
100 \cdot \frac{\sum_{i=1}^{l} \lambda_i}{\sum_{i=1}^{m} \lambda_i} > 85\%
\] (2.5)

The basic idea of the process of fault detection using PCA is to collect normal observation data to establish the PCA model. When a new observation data is subject to fault, these new data can be compared to the PCA model and its threshold. In order to detect the abnormal changes of the new data, the Q-statistic or SPE is used as follows

\[
Q\text{-statistic} = SPE = e^T e = \|x(I - PP^T)\|^2
\] (2.6)

The process is considered normal if

\[
Q\text{-statistic} \leq \delta_Q^2
\] (2.7)

where \( \delta_Q^2 \) denotes the confidence limit or threshold for the Q-statistic. It can be calculated from its approximate distribution (Jackson and Mudholkar, 1979)

\[
\delta_Q^2 = \theta_1 \left[ C_\alpha \sqrt{\frac{2\theta_2 h_0^2}{\theta_1}} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]
\] (2.8)

where \( \theta_i = \sum_{j=k+1}^{m} \lambda_j^i, \ i = 1, 2, 3 \) and \( h_0 = 1 - 2\theta_1 \theta_3/(3\delta_Q^2) \), and \( \lambda_j \) is the eigenvalue, \( C_\alpha \) is the critical value of the normal distribution and \( m \) is the number of all PCs.

The threshold is used to determine whether the data is within range of the model. To compare the new data to the PCA model, a confidence limit of \( \alpha = 95\% \) is used. Any point below the confidence limit line is considered to have a normal variance from the selected number of PCs, and any point above this line is considered to have an abnormally high level of variance.

After the fault is detected, i.e. the new observation data exceed the threshold line of the PCA model but can not be assure in what place the fault appears in the process. An assignable cause is determined by the contribution plot. The contribution plots are bar graphs of the Q-residual contribution of each variable calculated as in equation (2.9) (Xue et al., 2008). The variable having the largest residual produces the worst compliance to the PCA model, and indicates the source of the fault

\[
Q\text{-contribution} = cont_i = \frac{||e_i||^2}{Q\text{-statistic}}
\] (2.9)

where \( e_i \) presents the \( i \)-th element of the residual vector \( e \) and \( cont_i \) is the contribution of the \( i \)-th variable to the total sum of variations in the residual space.

By using the WT, the time domain information will not be lost when the signal is decomposed. In order to extract useful information in different decomposition levels, the contribution plots are used as an evaluation criterion for selecting the optimal level which contains the maximum information about the fault. A flowchart of the fault detection and diagnosis method based on WT and PCA is illustrated in Fig. 2.
3. Experimental setup

The problem of diagnosing the degradation of working conditions of rotating machinery is extremely important in industries to reduce the productivity loss. The measurement of vibration applied to the condition monitoring and fault diagnosis requires different types and levels of equipment and techniques. In the present study, an experimental system is used and the vibration response for mass unbalance and gear fault are obtained.

3.1. System description

The experimental system consists of a test rig built by S’tell Diagnostic (France), a data acquisition system (OROS OR25, 4 channels), piezoelectric accelerometers (PCB Piezotronics 353B34) and a PC (see Fig. 3). The system is driven by a 0.18 kW induction motor giving an output of 0-1500 rpm, controlled by a variable speed drive. To confirm the feasibility of the proposed method, we collect real vibration signals using the experimental test rig illustrated in Fig. 3, where 1 and 2 are gears with 60 and 48 teeth, respectively, H1, H2, H3 and H4 are the bearing housings.

The vibration signals are taken on bearing housing H1 by means of two piezoelectric accelerometers measuring radial vibration, i.e. in the Vertical Direction (VD) and Horizontal Direction (HD). These measurements are repeated for different states at different rotation speeds. The data acquisition is performed using the OR25 software. The vibration signals measured have
length of acquisition of 400 milliseconds. The sampling frequency is 5120 Hz and each signal has 2048 samples.

3.2. Faults description

The experiment described in this paper, performs the condition monitoring of rotating machinery to predict some anomalies that may occur under different measurement conditions such as: mass unbalance and gear fault.

3.2.1. Mass unbalance

Mass unbalance is one of the most common causes of vibration. It is a condition when the center of mass does not coincide with the center of rotation, due to unequal distribution of mass about the center of rotation. It is simulated in this study by an additional weight on the disk. The mass unbalance creates a vibration frequency exactly equal to the rotational speed, with an amplitude proportional to the amount of unbalance (Tandon and Parey, 2006).

Figure 4 represents the measured signals of the mass unbalance at a speed of 900 rpm in VD and HD.

3.2.2. Gear fault

Vibrations of a gear are mainly produced by the shock between teeth of the two wheels. Gear fault is simulated with a gap between teeth. The vibration monitored on a faulty gear generally exhibits a significant level of vibration at the tooth mesh frequency (i.e. the number of teeth on the gear multiplied by its rotational speed) and its harmonics (Tandon and Parey, 2006).

Figure 5 represents the measured signals of the gear fault at a speed of 900 rpm in VD and HD.
4. Results and discussion

The structure of proposed fault diagnosis technique involves two parts: the first one is development and training of the PCA model; the second is testing the process fault based on the trained model (see Fig. 2). The measurements used in training phase represent the normal operating conditions of the system, i.e. system without fault.

The vibration signals used in this work have been gained through the practical measurement, including a normal state, mass unbalance and gear fault. The data collection has been carried out according to the following routine: vibration signals used in the training are taken in VD and HD at seven different rotating speeds between 300 and 1425 rpm. In the testing phase, each fault of the process has been measured in the radial direction at five different rotating speeds: 675, 900, 1125, 1275 and 1425 rpm which, respectively, correspond to 11.25, 15, 18.75, 21.25 and 23.75 Hz.

The multiresolution analysis is applied by using the Daubechies wavelet of the order 4 (db4) and level 4. It may be noted that the same wavelet with the same level of decomposition is applied to each signal for the training and testing phases. The results of db4 decomposition of the vibration signals of the mass unbalance and gear fault collected at 900 rpm in the radial direction are given respectively in Figs. 6 and 7.

After decomposition with db4, the approximations and details vectors of each signal measured at the same rotation speed in the radial direction are collected in a matrix. Thus the input matrices of the PCA method are formed. During the training phase, seven matrices are collected for identifying the PCA model. Through PCA method, the anterior 8 PCs accumulation sum contribution rate is 92.46%, as shown in Table 1. These first 8 PCs are selected for the fault identification.

The detection threshold of Q-statistic is calculated according to equation (2.8), which is 2.2414. To evaluate the fault detection method, the detection ratio is used. It is defined as the number of samples whose Q-statistic values go beyond the threshold to the total samples. If the detection ratio is less than 20%, the fault is not successfully detected (Xu et al., 2008). In normal operating conditions, 2.83% of the Q-statistic samples are above the threshold value. It implies that the model has captured the major correlation and variance among the system variables.

In the testing phase, five matrices of each fault are collected for validating the PCA method. The new data are compared to the PCA model and its threshold. The evaluation results of fault detection rate are summarized in Tables 2 and 3. We show clearly that the majority samples of the Q-statistic at different rotation speeds are above the threshold value. By comparing these values to 2.83% obtained in the normal state, it is also clear that an abnormal situation has occurred in the process.

The objective of the proposed method is to demonstrate the effectiveness of the Q-contribution as a principal criterion for selecting the optimal level of wavelet decomposition.
Fig. 6. Decomposition results of the measured signals of the mass unbalance at 900 rpm in VD (left) and HD (right).

Fig. 7. Decomposition results of the measured signals of the gear fault at 900 rpm in VD (left) and HD (right).
Table 1. PCs values

<table>
<thead>
<tr>
<th>PCs</th>
<th>Eigenvalues</th>
<th>Variance [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.6459</td>
<td>16.46</td>
</tr>
<tr>
<td>2</td>
<td>1.6159</td>
<td>16.16</td>
</tr>
<tr>
<td>3</td>
<td>1.2381</td>
<td>12.38</td>
</tr>
<tr>
<td>4</td>
<td>1.1741</td>
<td>11.74</td>
</tr>
<tr>
<td>5</td>
<td>1.0642</td>
<td>10.64</td>
</tr>
<tr>
<td>6</td>
<td>0.9358</td>
<td>9.36</td>
</tr>
<tr>
<td>7</td>
<td>0.8259</td>
<td>8.26</td>
</tr>
<tr>
<td>8</td>
<td>0.7458</td>
<td>7.46</td>
</tr>
<tr>
<td>9</td>
<td>0.3934</td>
<td>4.13</td>
</tr>
<tr>
<td>10</td>
<td>0.3611</td>
<td>3.54</td>
</tr>
</tbody>
</table>

Table 2. Evaluation results of the mass unbalance detection rate

<table>
<thead>
<tr>
<th>rpm</th>
<th>675</th>
<th>900</th>
<th>1125</th>
<th>1275</th>
<th>1425</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPE [%]</td>
<td>35.59</td>
<td>67.91</td>
<td>79.00</td>
<td>83.30</td>
<td>85.40</td>
</tr>
</tbody>
</table>

Table 3. Evaluation results of the gear fault detection rate

<table>
<thead>
<tr>
<th>rpm</th>
<th>675</th>
<th>900</th>
<th>1125</th>
<th>1275</th>
<th>1425</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPE [%]</td>
<td>70.94</td>
<td>81.59</td>
<td>76.36</td>
<td>80.17</td>
<td>87.50</td>
</tr>
</tbody>
</table>

Fig. 8. Contribution plots of the mass unbalance (left) and the gear fault (right)
The contribution plots are bar graphs of the Q-residual contribution of each variable or decomposed vector (Fig. 8). The level having the largest value produces the worst compliance to the PCA model and indicates the desired level of decomposition. As shown in Fig. 8, decomposition levels 1 to 4 and 6 to 9 present respectively the detail vectors for VD and HD, and levels 5 and 10 stand respectively for the approximation vectors. It shows an obvious difference between the levels. From Fig. 8 (left), it can be seen that levels 5 and 10 have the largest values. We show also in Fig. 8 (right), that the contribution plot using db4 occurs in the third level. So our choice is attached to approximation A4 for the mass unbalance and detail D3 for the gear fault.

In order to diagnose the mass unbalance and the gear fault from the selected level we use, respectively, spectral analysis and envelope analysis. Figure 9 (left) shows the spectra of fourth approximations (A4). Frequency peaks at 11.25, 15, 18.75, 21.25 and 23.75 Hz are present, which could be related to a mass unbalance fault. Figure 9 (right) illustrates the FT of the envelopes of the selected details. It can be seen from this figure that the peaks at the rotation frequencies of the shaft (11.25, 15, 18.75, 21.25 and 23.75 Hz) and their multiples (×2, ×3,...) are present in the frequency spectra. This clearly indicates a gear fault.

![Fig. 9. Spectra of selected levels of the mass unbalance (left) and the gear fault (right)](image)

5. Conclusion

This paper presents a combined approach based on the WT and PCA method to improve the condition monitoring and fault diagnosis of rotating machinery. It is adapted to obtain multiple data series at different levels using wavelet decomposition and reduce the number of variables needed to monitor the process through PCA. The detection method is based on a confidence limit estimated from normal conditions. The aim is to select the optimal level of resolution using the
residual contribution of each decomposed vector, for a possible diagnosis. The proposed method has been tested on real measurement signals collected from a vibration system containing mass unbalance and gear fault. Better experimental results have been obtained by identifying the type of fault. Hence the WT combined with the PCA method is a successful approach to vibration monitoring.

References

6. DAUBECHIES I., 1988, Orthonormal bases of compactly supported wavelets, Communication on Pure and Applied Mathematics, 41, 909-996
10. GAVROVSKA A.M., JEVTIC D.R., RELJIN B.D., 2009, Selection of wavelet decomposition levels in ECG filtering, Proceedings of the International Conference on Telecommunication in Modern Satellite, Cable and Broadcasting Services TELSIKS, Niš-Serbia, DOI: 10.1109/TELSIKS.2009.5339423, 221-224
11. JACKSON J.E., MUDHOLKAR G.S., 1979, Control procedures for residuals associated with principal components analysis, Technometrics, 21, 341-349
17. NOMIKOS P., MACGREGOR J.F., 1995, Multivariate SPC charts for monitoring batch processes, Technometrics, 37, 1, 41-59


*Manuscript received August 13, 2014; accepted for print December 10, 2015*