University of Huddersfield Repository

Feng, Guojin, Tian, Xiange, Gu, James Xi, Yang, Dingxin, Gu, Fengshou and Ball, Andrew

An adaptive envelope analysis in a wireless sensor network for bearing fault diagnosis using fast kurtogram algorithm

Original Citation


This version is available at http://eprints.hud.ac.uk/21924/

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

http://eprints.hud.ac.uk/
An adaptive envelope analysis in a wireless sensor network for bearing fault diagnosis using fast kurtogram algorithm

Guojin Feng\textsuperscript{a}, Xiangge Tian\textsuperscript{a}, James Xi Gu\textsuperscript{b}, Dingxin Yang\textsuperscript{c}, Fengshou Gu\textsuperscript{a}\textsuperscript{*} and Andrew D. Ball\textsuperscript{a}

\textsuperscript{a}Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield, HD1 3DH, UK
\textsuperscript{b}School of Engineering, Manchester Metropolitan University, Manchester, M15 6BH, UK
\textsuperscript{c}College of Mechatronics and Automation, National University of Defense Technology, Changsha, 410073, PR China

\textbf{ABSTRACT}

This paper proposes a scheme to improve the performance of applying envelope analysis in a wireless sensor network for bearing fault diagnosis. The fast kurtogram is realized on the host computer for determining an optimum band-pass filter for the envelope analysis that is implemented on the wireless sensor node to extract the low frequency fault information. Therefore, the vibration signal can be monitored over the bandwidth limited wireless sensor network with both intelligence and real-time performance. Test results have proved that the diagnostic information for different bearing faults can be successfully extracted using the optimum band-pass filter.

Keywords: Wireless sensor network; envelope analysis; fast kurtogram; bearing fault diagnosis

\textsuperscript{*} Corresponding author. Tel.: +44 (0)1484-473548; fax: +44 (0) 1484 473075; e-mail: f.gu@hud.ac.uk
1. Introduction

Currently, wired online condition monitoring (CM) systems have been successfully employed for the asset analysis, but have so far been mainly restricted to large and critical industrial machines due to their high cost. As the machines become increasingly complicated, the complexity and cost of their maintenance have also grown quickly. One major cost for the wired CM systems is the investment in high quality cables and their installation costs in harsh industrial environment.[1]

With the rapid development of electronics and wireless technology, wireless sensor network (WSN) is gaining increasing popularity in CM field. It brings several inherent advantages, such as low cost, convenience of installation and easy to replace and upgrade.[1] therefore, it can significantly reduce the cost on the asset analysis and make the maintenance management more prevalent and ubiquitous. Currently, the monitoring of the static signals like temperature, pressure, humidity via the WSN has become popular.[2,3] However, the application of dynamic signals over the WSN is still limited, which is restricted by the conflicts between large data amount of dynamic signals and the low band width limitations of the WSN.

Recently, the idea of local [4] (on-sensor [1] / edge [5]) processing has been proposed to solve this conflict, in which way, the acquired data are firstly processed by the embedded algorithms at the sensor node and then only the analyzed results are transmitted over the wireless network. Therefore, the data throughput requirement can be significantly reduced while the useful information for fault diagnosis is still maintained in the transmitted small data set for further analysis. For this purpose, several methods have been implemented by researchers on the wireless sensor node, such as down-sampling [5], fast Fourier transformation (FFT) [1], lossless compression [8,9] and autoregressive (AR) coefficients [6], etc.

As an extensively used diagnostic algorithm for bearing fault diagnosis in the online CM system [7,8], envelope analysis is also a good candidate algorithm for the processing on the wireless sensor node. It transfers the high frequency problem to a low frequency one by extracting out the periodic excitation of the resonance signals [9] and only data in the extracted low frequency band are transmitted through the wireless network. Therefore, the transmission data throughput can be significantly reduced and the real time monitoring of the vibration dynamic signals over the WSN can be achieved. Its implementation on the wireless sensor node has been given out in [10] and its effectiveness for bearing fault diagnosis and transmission data throughput reduction have also been proved through experiments.

However, the performance of envelope analysis is highly affected by the selected band-pass filter. Due to the wide varieties of the bearing resonances, a fixed frequency band may or may not encompass the structural resonances excited by the bearing damage [9]. Thus, there has been an active study on how to choose the most suitable band to perform envelope analysis [9,11]. Traditionally, there are two methods to choose the suitable band-pass filter. One method is searching for a peak at high frequency in response spectra, on the assumption that it would be excited by bearing faults [11]. In this method, a number of trials may be needed before the correct band is selected because of the noise interferences, which obviously is not suitable for automatic applications. The other method is using a hammer impact test to find bearing housing resonances before installing the sensor [11] and a band-pass filter encompass the system resonance is thereafter designed. However, this professional test usually needs an expert to carry out. So it is not convenient for practical factory cases where there are plenty bearing systems. Besides, the system responses may change with bearing conditions, which may cause the fixed filters not suitable anymore.

This problem has now largely been solved by the use of spectral kurtosis (SK) and the kurtogram to find the most impulsive band [11]. In this method, an optimum band-pass filter can be precisely selected automatically. This means that the filter can be updated in an adaptive way and maintains optimum even if the system responses have changed with the bearing condition. However, the computation of the kurtogram is quite costly and not convenient for online industrial purposes, thus, in 2007, a fast kurtogram algorithm is proposed by Antoni to fasten the computation [12].

This paper proposes a scheme that brings the fast kurtogram to the wireless CM system for automatic bearing fault diagnosis. The fast kurtogram is implemented on the host computer for selecting a best frequency band where to filter and demodulate the signal while the envelope analysis algorithm is executed on the wireless sensor node for calculating the envelope spectrum and reducing the data amount for transmission. With this scheme, the implementation of envelope analysis on the wireless sensor node can be made more meaningful and practical.

The following contents are organized as follows. The spectral kurtosis, kurtogram and fast kurtogram are explained in section 2. Section 3 explains the wireless CM system and its working procedure. A bearing test rig and the results of the seeded outer race and inner race fault are given in section 4. Finally, section 5 concludes the paper.

2. Theoretical background

Rolling element bearing is a major source of failure in electromechanical systems. As investigated in [13], bearing faults account for more than 40% of the induction motor’s failure. Therefore, there has been a long history on the health monitoring of bearing and a lot of work has been carried out [9,8,11]. Among them, the early detection of the bearing fault is of major concern because it can effectively prevent catastrophic failures and reduce the economic loss.

Vibration technique is extensively employed for detecting the bearing faults because it is sensitive to the bearing fault and can be monitored in-situ [14]. A vibration signal collected from the bearing is a mixture of the fault signal and the noise signal and the objective of bearing fault detection is to test if the vibration sensor signal contains the faulty bearing signal.

2.1. Spectral kurtosis and kurtogram

The vibration signal from a healthy bearing is considered to follow the Gaussian distribution. If fault exists in the bearing, impulse series will be added into the measured vibration signal which thus doesn’t follow Gaussian distribution anymore. The Gaussian distribution can usually be checked using the global kurtosis. If the signal follows a Gaussian distribution, its kurtosis is almost equal to 0, otherwise, the value will be higher than 0. However, the impulses from the early stage faults are likely to be very weak and can be easily buried by the large background noises, making the fault difficult to be detected using the global kurtosis.

Therefore, spectral kurtosis (SK) was raised to solve this problem, which is a function of frequency that can indicate the impulsiveness of a signal distributing in the frequency domain [11]. Its usage can trace back to 1980s when it was firstly employed for detecting impulsive events in sonar signals [15].
The SK of a signal $x(t)$ can be computed using Eq. (1) based on Short-time Fourier transform (STFT) $X(t,f)$, which is local Fourier transform at time $t$ obtained by moving a window along the signal.

$$K(f) = \frac{\langle |X(t,f)|^4 \rangle}{\langle |X(t,f)|^2 \rangle^2} - 2$$  \hspace{1cm} (1)

where $\langle \cdot \rangle$ denotes the time-averaging operator. The subtraction of 2 is used to enforce $K(f) = 0$ since $X(t,f)$ here is complex Gaussian (instead of 3 for real signals) [16].

According to the definition of SK, in the frequency band dominated by the bearing signal will have a high value of SK, otherwise, the value of SK will be small. Therefore, the SK can be used as a filter function to choose the frequency band that has the highest level of impulsiveness, as shown in Eq. (2).

$$K_{\gamma}(f) = \frac{K_{\delta}(f)}{[1 + \rho(f)]^2}$$  \hspace{1cm} (2)

where $\rho(f)$ is the noise-to-signal ratio function of frequency.

Because the computation of SK is based on STFT, its value is also critically related to the choice of the STFT window length. Therefore, the two-dimensional kurtogram is raised to improve the filter function, which is a function of both centre frequency $f$ and the frequency resolution $\Delta f$. A combination of $f$ and $\Delta f$ determines a value of SK and is named a dyad $(f; \Delta f)$. The dyad $(f; \Delta f)$ with the maximum value is considered as the optimum band-pass filter parameters.

### 2.2. Fast kurtogram

As explained above, the kurtogram contains a host of dyads $(f; \Delta f)$ and a thorough computation of all possible dyads is quite costly and not practical for on-line industrial applications. Therefore, Antoni proposed a fast kurtogram algorithm to solve this problem [12]. Using this algorithm, an equivalent dyad can be computed but with much better speed performance. As an example in [12], on the same computer, for the computation of the same signal with 1.37 s, the fast kurtogram took a fraction of a second while the full kurtogram took more than 10 minutes.

The principle of fast kurtogram is based on an arborescent multi-rate filter-bank structure using a quasi-analytic filter [12]. Let $x(n)$ be a non-stationary process and $h(n)$ a low-pass prototype filter used to construct two quasi-analytic low-pass and high-pass analysis filters $h_{1}(n)$ and $h_{2}(n)$, in the frequency bands $[0; 1/4]$ and $[1/4; 1/2]$, respectively.

$$h_{2}(n) = h(n)e^{j\pi n/4}$$

$$h_{1}(n) = h(n)e^{j3\pi n/4}(j^2 = -1)$$  \hspace{1cm} (3)

Then, these two filters are used to perform the elementary low-pass/high-pass decomposition illustrated in Figure 1. As it shows, the sequence number is doubled after the decomposition, but their respective length is also halved so that the total amount of data remains the same. By using this decomposition, a tree of filter banks can be built, as shown in Figure 2 (a). In the filter-bank tree, there are $2^k$ bands in the corresponding level and the coefficients $c^j_k(n)$ can be interpreted as the complex envelope of signal $x(n)$.

![Figure 1](image1.png)  \hspace{1cm} Figure 1 Basic decomposition

Then, the kurtogram $K^j_k$ is estimated using Eq. (4) by computing the kurtosis of all sequences $c^j_k(n)$, $j = 0, ..., 2^k - 1$, $k = 0, ..., K - 1$.

$$K^j_k = \frac{\langle |c^j_k(n)|^4 \rangle}{\langle |c^j_k(n)|^2 \rangle^2} - 2$$  \hspace{1cm} (4)

To achieve a finer sampling of the dyad plane, an 1/3-binary tree of filter-banks is extended on the binary tree structure shown in Figure 2 (a). Three additional quasi-analytic band-pass filters $g_j(n)$, $j = 0, 1, 2$ are defined with pass-bands $[0; 1/6], [1/6; 1/3]$ and $[1/3; 1/2]$, respectively. These filters are then used to further decompose each sequence $c^j_k(n)$ into three sub-sequences $c^{3+j}_{j+k}(n)$, $j = 0, 1, 2$.

![Figure 2](image2.png)  \hspace{1cm} Figure 2 (a) Fast calculation of the kurtogram by using an arborescent filter-bank structure (b) The complete paving of the (frequency/frequency resolution) plane.

Finally, the kurtogram of the corresponding 1/3-binary tree are also calculated using Eq. (4) and the complete fast kurtogram paving of the $(f, (\Delta f)_2)$ plane is illustrated in Figure 2 (b). For each dyad $(f, (\Delta f)_2)$, a kurtosis value is calculated. The optimum band-pass filter is expected to have the maximum kurtosis value.
3. Implementation of fast kurtogram in the WSN

3.1. Wireless CM system structure

A prototype wireless CM system based on Zigbee is given in Figure 3. It consists of one sink node and several sensor nodes. The Zigbee network is established using the off-the-shelf Xbee Pro @ZB wireless module from Digi international [17]. This popular commercial module is effective, reliable and practical to establish the wireless network [18].

The sink node is the heart of the Zigbee network. It has an Xbee Pro module for wireless communication and a FT232 board for connecting with the host computer through a USB port. All the data from the sensor nodes are transmitted to the host computer through the sink node and the instructions from the host computer are routed to the sensor nodes through the sink node as well.

The sensor node is used to collect the vibration signal, apply local processing algorithms and transmit the processing results to the sink node for further analysis through the wireless network. The vibration signal is measured using a piezoelectric (PE) accelerometer and conditioned by a charge amplifier, voltage amplifier and an anti-aliasing filter. Then, the signal is sampled by the on-chip ADC of the core processor and processed by the embedded algorithms. In this system, a low power state-of-the-art Cortex-M4F microcontroller TM4C1233H6PM is employed as the core processor of the sensor node, which owns both the feature of low power consumption and powerful computing capability. This microcontroller has 256kB FLASH, 32kB SRAM, two 12-bit ADCs with 1 MSPS sampling rate and many peripherals while it costs less than £3.00. Most importantly, a digital signal processing (DSP) unit and a floating point unit (FPU) are integrated inside the processor, allowing intensive computations such as FFT to be implemented [19].

The host computer is used to read data from the sink nodes, implement the fast kurtogram for obtaining the optimum band-pass filter, send instructions to the sensor nodes and display envelope spectrum from the sensor nodes.

In order to accomplish the two tasks of band-pass filter automatic selection and localized envelope analysis in the wireless CM system described above, the system works at two modes which are named as configuration mode and monitoring mode respectively, as shown in Figure 4. The configuration mode is used for updating the optimum band-pass filter coefficients. It operates at the installation of the system and during a period when a significant change appears in the monitoring mode or at a given long time interval. Most of the time, the system works at the monitoring mode, in which the wireless sensor node uses the optimum band-pass filter calculated in the configuration mode to analyze the vibration signal and transmit the envelope spectrum to the host computer for fault diagnosis. Thus, this system can automatically finish the optimum band-pass filter selection and carry out effective condition monitoring.

![Figure 3 Structure of the wireless CM system](image)

3.2. Configuration mode

In this mode, a dataset containing 8192 points of data is firstly collected on the sensor node and transmitted to host computer via the wireless network. Then, the fast kurtogram is computed on host computer to obtain the optimum band-pass filter. As shown in Section 2.2, the filters are designed as a complex type, which is not easy to be implemented on the memory limited microcontroller. Therefore, an equivalent FIR filter with only real part coefficients is generated based on the parameters of the calculated optimum band-pass filter. At last, the coefficients of the FIR filter are sent back to the sensor node where the older coefficients will be renewed for the next coming computation cycle.

![Figure 4 System working procedure](image)
In the test, the sampling rate of the ADC is set at 32 kHz and 8192 points of data means the signal length is about a 0.256 second, which is fairly enough for carrying out the fast kurtogram. As will be discussed in Section 4.1, the frequency range of interest in the bearing system is less than 500 Hz. According to Figure 2 (b), the band width at level 5 has reached 500 Hz with the sampling rate at 32 kHz, therefore, 5 levels’ analysis is sufficient for this system.

3.3. Monitoring mode

Most of the time, the system works in the monitoring mode, in which the envelope analysis is carried out at the sensor node and the calculated envelope spectrum is transmitted wirelessly to the host computer for display and further analysis. Due to limited computing memory size of the sensor node, only small frames of data can be calculated at one time. In order to get a high frequency resolution, a kind of overlap, down-sampling and cascading scheme is employed for real time data processing, in which a large frame of data is divided into several smaller frames and processed in different time division.

The diagram for the data processing scheme is shown in Figure 5. As it shows, the collected data are processed frame by frame with the size of 512 points. The frame first passes through a band-pass filter and combines with the last filtered frame to produce a new frame of 1024 data points. Then the envelope of the newly produced frame is computed using Hilbert transform, as described in [10]. Because distortions exist at the start and end edge of the envelope using Hilbert transform, only middle half of the computed envelope data are reserved and filtered using an 80-order FIR low-pass filter with the cutoff frequency at 500 Hz. Then, the data are down-sampled by a ratio of 1/32 to save computing memory and reduce the calculation amount. After the process, the original 512 points of data shrink to only 16 points while the envelope information of the raw data is still reserved.

As shown in Figure 5, from the second frame, every frame of data will create 16 points of envelope data and after the 129th frame, an envelope signal with 2048 points can be obtained. The consistency of envelope signal is guaranteed by a 50% overlap processing when calculating the envelope.

In the above process, there are two kinds of sampling rate: 32 kHz and 1 kHz. Data with high sampling rate share the same computing space and are computed in different time division. The high sampling rate of the raw data allows the high frequency components can be analyzed and the low sampling rate of the envelope signal allows longer analysis time length and thus higher frequency resolution can be achieved.

Figure 6 gives the flow chart of the one cycle computation on the sensor node. The collected data are firstly processed by the overlap, down-sampling and cascading scheme as explained above. After this, an envelope signal with 2048 points of data is obtained and spectrum is calculated through a forward FFT and complex type amplitude computation. In order to suppress random noises and obtain a reliable envelope spectrum result, the calculated envelope spectrum is averaged by four times. After averaging, the result data are converted to 16-bit unsigned integer format and transmitted to the host computer over the wireless network. When one computation cycle finishes, the computation buffer will be cleared and another cycle will starts for continuous condition monitoring.

For the down-sampled envelope signal, its equivalent sampling rate becomes 1 kHz, therefore, the resolution of the envelope spectrum is about 0.49 Hz. This means the frequency band of interest has about 1024 points of data, which are the data need to be transmitted to the host computer.

The process in grey block in Figure 6 needs real-time performance, which means these calculations should be finished before the next frame arrives. For the sampling rate of 32 kHz and frame size of 512 points, the computation time should be less than 16 milli-seconds (ms). In the test, this process costs about 1053500 cycles, which is measured using the clock cycles in code composer studio (CCS) [20]. As the microcontroller runs at 80 MHz, the computation time for the process is about 13.2 ms, which is smaller than 16 ms requirement. Therefore, the real time requirement can be satisfied.
4. Experimental setup and results

4.1. Bearing test rig

In order to evaluate the performance of the fast kurtogram combining with the embedded envelope analysis algorithm of the wireless CM system, a bearing test rig was setup. As shown in Figure 7 (a), it consists of five main parts: an electrical induction motor, shaft couplings, a DC generator, bearings and motion shaft. Two cases were tested on the same type of bearing but with different fault locations, as shown in Figure 7 (b-d). One bearing is seeded with outer race fault while the other one with inner race fault. A PE accelerometer was mounted on the bearing house horizontally to collect the vibration signal.

The bearing type is N406 cylindrical roller bearing and during the test, the shaft ran at a full speed of 1460 rpm, which is 24.3 Hz. The four localized fault frequencies for the bearing are calculated and listed in Table 1. Among the four fault frequencies, the highest one is the inner race fault frequency at 135.5 Hz, whose 3rd harmonic frequency (406.5 Hz) is within 500 Hz. This means fault frequencies of interests are all within 500 Hz. Therefore, the cut-off frequency of the FIR low-pass filter described in Section 3.3 is set at 500Hz.

![Bearing test rig diagram](image)

Figure 7. (a) Bearing test rig, (b) Roller bearing, (c) Defect on outer race, (d) Defect on inner race

<table>
<thead>
<tr>
<th>Table 1. Fault frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defect location</td>
</tr>
<tr>
<td>Inner race (BFFI)</td>
</tr>
<tr>
<td>Outer race (BPFO)</td>
</tr>
<tr>
<td>Ball (BSF)</td>
</tr>
<tr>
<td>Cage (FTF)</td>
</tr>
</tbody>
</table>

4.2. Outer race fault detection results

A vibration signal (8192 points, approximately 0.25s) measured on the bearing with outer race fault is shown in Figure 8 (a) and its spectrum in Figure 8 (b). This data set are received on the host computer in the configuration mode and calibrated on the host computer before displaying. As shown in Figure 8 (a), periodical spikes are observable which are caused by the defect on the outer race. From its spectrum, it can be seen that the signal has a wide frequency range which makes it difficult to identify the fault types. Several peaks exist in the spectrum and it’s not easy to tell which one is suitable for implementing the envelope analysis.

![Outer race raw data set](image)

Figure 8 Outer race raw dataset: (a) time domain (b) frequency domain

The corresponding fast kurtogram of this data set is displayed in Figure 9, which shows a high non-stationary activity around the carrier frequency $f = 9$ kHz. The dyad which maximises the kurtosis is $\left\{ f_x, \frac{\Delta f}{2} \right\} = \{9 \text{kHz}, 2 \text{kHz}\}$, that is the frequency band [8 kHz; 10 kHz]. According to Figure 8 (b), high peaks are observable in this frequency band. Thus, an 80-order FIR band-pass filter with pass band from 8 kHz to 10 kHz is designed and its coefficients are transmitted to the sensor node for implementing envelope analysis.

The averaged envelope spectrum computed on the sensor node is given in Figure 10. It shows clearly the spectrum lines at the characteristic frequency of outer race fault. Without double, it can be based on to diagnose the outer race fault on the bearing.

![Fast kurtogram](image)

Figure 9 Fast kurtogram of the outer race raw data set, maximum kurtosis is 2.6 at level 3. The optimum band-pass filter is centered at 9 kHz with 2 kHz band width.

![Averaged envelope spectrum](image)

Figure 10 Averaged envelope spectrum for bearing with outer race fault
4.3. Inner race fault detection results

Similarly, a dataset of vibration signal was collected on the bearing with inner race fault, as shown in Figure 11 (a), and its spectrum in Figure 11 (b). The periodical spikes are observable but not as obvious as those in Figure 8 (a). Its spectrum also spread over a wide frequency range and the peaks are a bit different from those in Figure 8 (b), which indicates the resonances of the bearing system have changed slightly with the change of the installed bearing.

![Figure 11 Inner race raw dataset: (a) time domain (b) frequency domain](image)

As shown in Table 2, the data throughput requirement for the raw data is as high as 512 kbps which is even higher than the maximum transmission speed of Zigbee whose bandwidth is 250 kbps. However, after the processing on the sensor node, the data throughput of envelope and its spectrum are significantly reduced to about 10 kbps. If a four time average is carried out on the envelope spectrum, the data throughput requirement can be further decreased to less than 2 kbps. Therefore, multiple sensor nodes would be allowed to coexist in the wireless network for bearing fault diagnosis.

Noteworthy, the performance of the envelope analysis in the wireless CM system is critically related to the selection of the band-pass filter. Therefore, the implementation of fast kurtogram in the system is of significant importance.

5. Conclusions

An adaptive envelope analysis is implemented in a wireless CM system by employing the fast kurtogram algorithm for the automatic selection of the optimum band-pass filter. The fast kurtogram is fulfilled on the host computer due to its complexity and envelope analysis is embedded on the wireless sensor for extracting the envelope spectrum and thus reducing the data throughput requirement. By using this scheme, the vibration signal can be monitored effectively over the bandwidth limited wireless sensor network for bearing fault diagnosis.
References


