OPEN ACCESS

2nd International Electronic Conference on Entropy and Its Applications
15-30 November 2015

www.sciforum.net/conference/ecea-2

Conference Proceedings Paper – Entropy

Early fault detection and diagnosis in bearings based on logarithmic energy entropy and statistical pattern recognition

Aleksandar Brkovic [1,2], Dragoljub Gajic [2,3,4,*], Jovan Gligorijevic [1,2], Ivana Savic-Gajic [5,6], Olga Georgieva [4] and Stefano Di Gennaro [6]

1 Faculty of Engineering, University of Kragujevac, Serbia
2 Tetra Pak, Gornji Milanovac Packaging Materials Plant, Serbia
3 School of Electrical Engineering, University of Belgrade, Serbia
4 Faculty of Mathematics and Informatics, University of Sofia, Bulgaria
5 Faculty of Technology, University of Nis, Serbia
6 Center of Excellence DEWS, University of L'Aquila, Italy

* Author to whom correspondence should be addressed; E-mail address: dragoljubgajic@gmail.com.

Published: 13 November 2015

Abstract: Rolling-element bearings are among the most common components of rotating machinery available in various industries. Mechanical wear and defective bearings can cause machinery to decrease its reliability and efficiency. Therefore it is very important to detect their faults in an early stage in order to assure a safe and efficient operation. We present a new technique for an early fault detection and diagnosis in rolling-element bearings based on vibration signal analysis. After normalization and the wavelet decomposition of vibration signals, the logarithmic energy entropy of obtained wavelet coefficients, as a measure of the degree of order/disorder, is extracted in a few sub-bands of interest. Then the feature space dimension is optimally reduced to two using scatter matrices. In the reduced two-dimensional feature space the fault detection and diagnosis is carried out by quadratic classifiers. Accuracy of the new technique was tested on four classes of the recorded vibrations signals, *i.e.* normal, with the fault of inner race, outer race and balls operation. An overall accuracy of 100% was achieved. The new technique can be used to increase productivity and efficiency in industry by preventing unexpected faulty operation of bearings.

Keywords: Fault diagnosis; rolling bearings; wavelet transform; logarithmic energy entropy; quadratic classifiers

1. Introduction

Predictive maintenance together with an early fault detection and diagnosis (FDD) plays an increasingly important role in order to achieve the efficient and sustainable operation of rotating machinery already installed and running at the present time. The typical lifetime of such machinery is between 30 and 50 years. Rolling-element bearings are among the most common components of rotating machinery available in various industries from agriculture to aerospace. They operate under high loading and severe conditions. As shown in Figure 1 their faults often occur gradually and represent one of the foremost causes of failures in rotating machinery. Defective bearings generate various forces causing high amplitude of vibration and thus decreasing efficiency. For example, in the case of a water pumping station bearing faults can increase vibration level up to 85%, while efficiency decreases 18% [1]. Therefore it is very important to avoid deteriorating condition, degraded efficiency and unexpected failures using a reliable, fast and automated technique for early FDD in bearings.

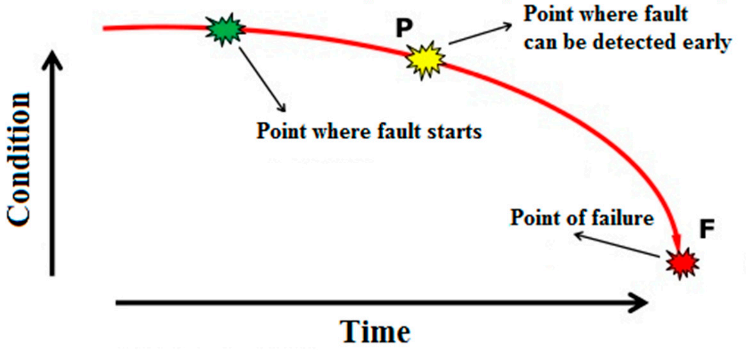

Figure 1. Predictive maintenance Potential to Functional Failure (P-F) curve

Many techniques for FDD in bearings based on vibration signal analysis have emerged in recent years. Generally, an FDD can be decomposed into three steps: data acquisition, feature extraction, and classification. An effective feature extraction as the key step represents a mapping of vibration signals from their original measured space to the feature space which contains more valuable information for FDD. Even though time-domain features such as peak, mean, root mean square, variance, skewness and kurtosis have also been employed as input features to train a bearing FDD classifier the fast Fourier transform (FFT) is one of the most widely used and well-established feature extraction methods [2]. However, the FFT-based techniques are not suitable for analysis of non-stationary signals. Since vibration signals often contain non-stationary components, for a successful FDD it is very important to reveal such information as well. Thus, a supplementary technique for non-stationary signal analysis is necessary. Time-frequency techniques such as the Wigner–Ville distribution (WVD) [3] and the short-time Fourier transform (STFT) [4] also have their own disadvantages. The WVD bilinear characteristic leads to interference terms in the time-frequency domain while the STFT results in a constant resolution for all frequencies since it uses the same window for the analysis of the entire signal. The wavelet transform very accurately resolves all these deficiencies and provides good frequency resolution and low time resolution for low-frequency components as well as low frequency resolution and good time resolution for high-frequency components. Therefore the wavelet transform has been widely applied in the field of vibration signal analysis and feature extraction for bearing FDD [5,6]. A precise classification as the next step directly depends on previously extracted features, *i.e.* there is no a classifier which can make up for the information lost during the feature extraction. As in the case of the feature extraction, we can come across a wide range of classifiers used for FDD in bearings. The classifiers based on artificial neural networks [7–9] and fuzzy logic [10,11] demonstrated a highly reliable classification. However, one of the disadvantages of these classification approaches is that they require the availability of a very large training set and a large number of parameters, which have to be selected or adjusted to obtain good results [12]. Therefore, there is a strong need to make the classification process simpler, faster and accurate using the minimum number of features and parameters.

In this paper a new technique for early FDD is proposed. As shown in Figure 2 it has several steps. The first step is acquisition of vibration signals as well as their preprocessing which includes normalization and segmentation. In the next step and after the wavelet transform of vibration signals 6 representative features, *i.e.* the logarithmic energy entropy of the wavelet coefficients in six sub-bands, are extracted in the time-frequency domain. In the third step the feature space dimension is optimally reduced to two using scatter matrices while in the final step in total three quadratic classifiers are designed [13], the one for detection and the other two for diagnosis of bearing faults. Using this new approach, the overall complexity of FDD is decreased and at the same time a very high accuracy maintained compared with already available techniques which employ more complex training algorithms.

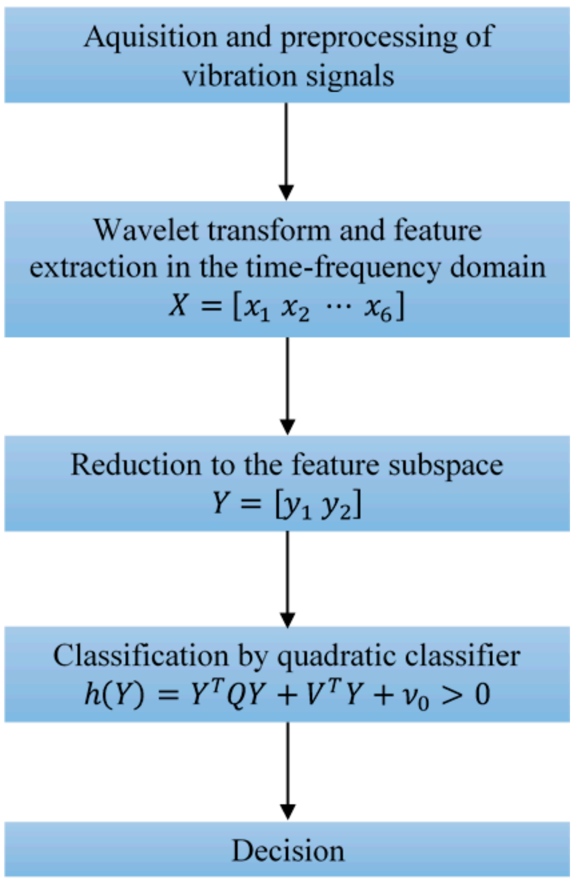

Figure 2. Flowchart of the new technique for early FDD

2. Materials and methods

2.1. Acquisition and preprocessing of the vibration signals

In order to test the capability of the new technique the bearing data obtained from the Case Western Reserve University (CWRU) Bearing Data Center [10] is used since it has become a standard reference in the field of FDD in bearings. A ball bearing as one shown in Figure 3 was installed in a motor driven mechanical system shown in Figure 4. A three-phase induction motor was connected to a dynamometer and a torque sensor by a self-aligning coupling. The dynamometer was controlled so that desired torque load levels can be achieved. An accelerometer with a bandwidth up to 5000 Hz and a 1 V/g output was mounted on the motor housing to acquire the vibration signals from the bearing.

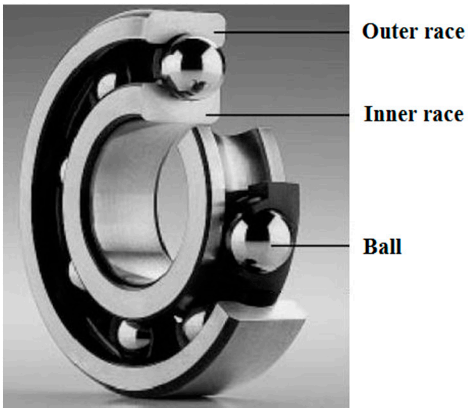

Figure 3. Ball bearing

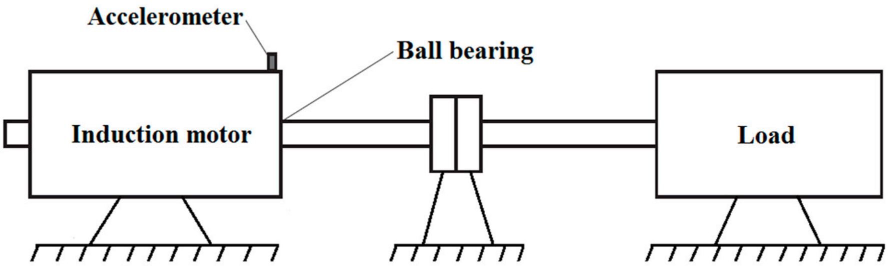

Figure 4. Experimental system

The data collection system consisted of a high bandwidth amplifier particularly designed for vibration signals and a data recorder with a sampling frequency of 12000 Hz. The sampling rate is ample having in mind that the frequency content of interest in the recorded vibration signals did not exceed 5000 Hz. The data recorder was equipped with low-pass filters at the input stage for anti-aliasing. In total four sets of data were obtained from the experimental system: (i) under normal conditions; (ii) with inner race faults; (iii) with ball faults and (iv) with outer race faults. The faults ranging from 0.007 inches to 0.40 inches in diameter and 0.011 inches in depth were introduced separately at the bearing elements using the electrical discharge machining (EDM) method. The ball bearing was tested under four different loads (0, 1, 2 and 3 hp) while the shaft rotating speeds were 1730, 1750, 1772 and 1797 rpm. Only the smallest fault diameter was selected for this study since we were interested in an early FDD. In order to make the entire technique more robust and less dependent on the vibration signal magnitude the recorded vibration signals were normalized to zero mean and unit variance. The vibration signals collected from each of four different conditions are divided into 256 segments of 1024 sample each, as shown in Figure 5. In total 1024 segments were used, 512 for design and 512 for testing of the new technique for early FDD in bearings.

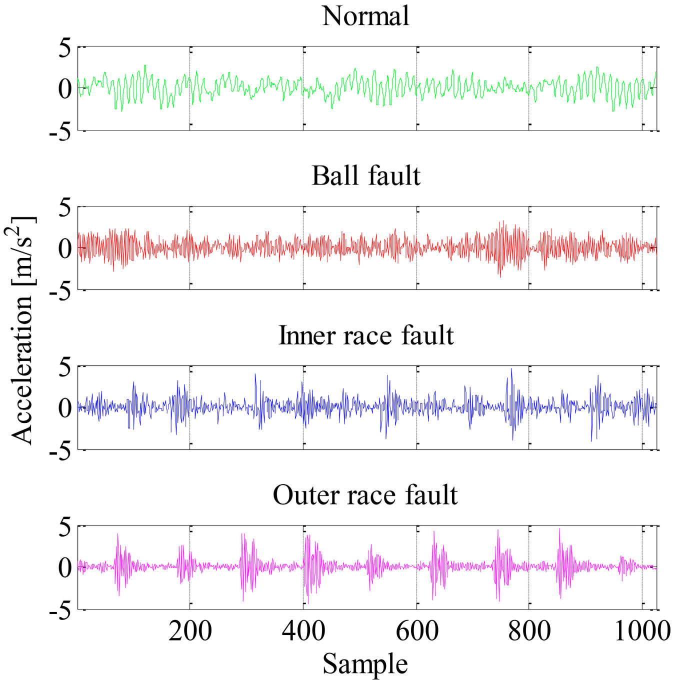

Figure 5. Segments of the vibration signals collected from four different conditions of the ball bearing

2.2. Wavelet decomposition and logarithmic energy entropy

We are already familiar with the fact that a signal can be presented as a linear combination of its basic functions. A unit impulse function whose power is limited and whose mean differs from zero is the basic function of the signal in the time domain, whereas in the frequency domain, this role is assigned to the sinusoidal function that has infinite power, and a zero mean. When using the wavelet transform to transform the signal from the time domain to the time-frequency domain, the basic function is the wavelet. The wavelet is a function of limited power, *i.e.* duration, and a zero mean [14], and for which the following is valid

$$\sum_{n=-\infty}^{\infty} |\psi[n]|^2 < \infty, \sum_{n=-\infty}^{\infty} \psi[n] = 0\,. \quad (1)$$

The wavelet that is moved, or translated, in time for b samples and scaled by the so-called dilation parameter a is given by

$$\psi_{ab}[n] = \frac{1}{\sqrt{a}}\psi\left[\frac{n-b}{a}\right]. \quad (2)$$

By changing the dilation parameter, the basic wavelet ($a = 1$) changes its width, and thus spreads ($a > 1$) or contracts ($0 \leq a < 1$) in the time domain as shown in Figure 5. In the analysis of non-stationary signals, the possibility of changing the width of the wavelet represents a significant advantage, considering the fact that wider wavelets can be used to extract slower changes, *i.e.* low frequency components, and narrower wavelets can be used to extract faster changes, *i.e.* high frequency components. Following the selection of the values of parameters a and b it is possible to transform segments of the signal $x[k]$ of N samples, and calculate the wavelet transform coefficients in the following way

$$w_{ab}[n] = \sum_{\tau=1}^{N} x[\tau]\psi_{ab}[n-\tau]\,, 1 \leq n \leq N \quad (3)$$

What is actually being extracted from the signal are only those frequencies that are within the wavelet frequency band $\psi_{ab}[n]$ *i.e.* the signals are filtrated by the wavelet $\psi_{ab}[n]$. Based on the coefficients obtained in this way, the original signal can also be reconstructed in the time domain using an inverse wavelet transform. Of course, if necessary, it is possible to also independently reconstruct the part of the signal which is filtered, as well as the part that was rejected by the wavelet $\psi_{ab}[n]$ on the basis of the so-called detail and approximation coefficients respectively, which are of course a function of the transformation coefficients $\psi_{ab}[n]$.

Figure 6. Sinusoid and two wavelets with different width

Parameters a and b can continuously change, which is not so practical especially bearing in mind that the signal can be completely and accurately transformed and reconstructed by using a smaller and finite number of wavelets, that is, by using a limited number of discrete values of parameters a and b, which is also known as the discrete wavelet transform (DWT). In this case, parameters a and b are the powers of 2, which gives us the dyadic orthogonal wavelet network with frequency bands which do not overlap each other. The dilation parameter a, as the power of 2, at each subsequent higher level of transformation, doubles in value in comparison to the value from the previous level, which means that the wavelet becomes twice as wide in the time domain, and has a frequency band that is half as narrow and twice as low. This actually decreases the resolution of the transformed signal in the time domain twofold, increasing it twice as much in the frequency domain. Thus, the signal frequency band from the previous level is split into two halves at every next level, into a higher band which contains higher frequencies and describes the finer changes, or details, and a lower band that contains lower frequencies and represents an approximation of the signal from the previous level. This technique is also known as the wavelet decomposition.

Entropy-based wavelet decomposition presented by Coifman and Wickerhauser [15] is used to compute logarithmic energy entropy. Essentially, entropy tells us how much information is carried by a signal, *i.e.* how much randomness is in the signal. The logarithmic energy entropy of the wavelet coefficients $w[n], 1 \leq n \leq N$, as a finite length random discrete variable, with probability distribution function $p(w)$ is defined by

$$H(w) = -\sum_{i=1}^{N} \left(log_2 \left(p_i(w) \right) \right)^2 \quad (4)$$

where i indicates one of the discrete states. This entropy is smaller if each discrete state has about the same probability of occurrence.

2.3. Reduction of the feature space dimension

Let an n-dimensional vector $X = [x_1\ \mathrm{x}_2\ \cdots\ \mathrm{x}_n]^T$ be transformed through the application of a certain linear transformation into an n-dimensional vector $Y = [y_1\ \mathrm{y}_2\ \cdots\ \mathrm{y}_n]^T = A^T X$ where A is an n-dimensional transformation matrix. Mapping of the vector X into the space made up by the eigenvectors $\Phi_1, \Phi_2, \dots, \Phi_n$ of its covariance matrix Σ_X is known as the principal component analysis (PCA) [13]. When reducing the feature space dimension using the PCA [16] the performance of each feature $x_1, \mathrm{x}_2, \dots, \mathrm{x}_n$ is characterized by its eigenvalue $\lambda_1, \lambda_2, \dots, \lambda_n$, respectively. Thus, by rejecting features we should first reject those with the smallest eigenvalue, *i.e.* with the smallest variance in the new feature space. For example, in the case of the dimension reduction from two to one shown in Figure 7 the mapped feature y_2 would be rejected as less informative even though it has better discriminatory potential than y_1.

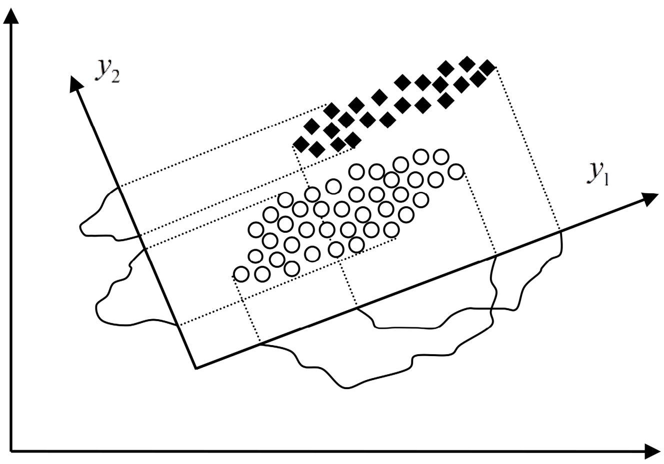

Figure 7. Different approaches to reduction of the feature space dimension

Unlike the PCA, the dimension reduction based on scatter matrices [13] is of special significance in this paper since it takes into consideration the very purpose of the dimension reduction, that is, the classification. Let L be the number of classes which should be classified and M_i and $\Sigma_i, i = 1 \cdots L$ the mean vectors and the covariance matrices of these classes, respectively. Then the within-class scatter matrix can be defined by:

$$S_W = \sum_{i=1}^{L} P_i\, E\{(X - M_i)(X - M_i)^T / \omega_i\} = \sum_{i=1}^{L} P_i\, \Sigma_i \quad (5)$$

and the between-class scatter matrix as:

$$S_B = \sum_{i=1}^{L} P_i \, (M_i - M_0)(M_i - M_0)^T \ (6)$$

where M_0 is the joint vector of mathematical expectation for all the classes together, that is

$$M_0 = E\{X\} = \sum_{i=1}^{L} P_i \, M_i. \ (7)$$

In addition the mixed scatter matrix can be defined by:

$$S_M = E\{(X - M_0)(X - M_0)^T\} = S_W + S_B. \ (8)$$

Then the problem of dimension reduction is reduced to the identification of the $n \times m$ transformation matrix A which maps the random vector X of dimension n onto the random vector $Y = A^T X$ of dimension m and at the same time maximizes the criteria $J = tr(S_W^{-1} S_B)$. This criteria is invariant to non-singular linear transformations and results into transformation matrix that takes the following form:

$$A = [\Psi_1 \ \Psi_2 \ \cdots \ \Psi_m] \ (9)$$

where Ψ_i, $i = 1, \dots, m$ are the eigenvectors of the matrix $S_W^{-1} S_B$ which correspond to its m greatest eigenvalues. The dimension reduction based on scatter matrices applied to the case shown in Figure 7 would result into selection of the mapped feature y_2. Obviously it is much better choice than y_1 selected by the PCA in terms of more accurate classification as the main goal of the dimension reduction.

2.4. Design of quadratic classifiers

Quadratic classifiers are already known to be very good and robust solutions to the problems of classification of vectors whose statistical features are either unknown or change over time [13]. In addition they also allow visual insight into the classification results. A quadratic classifier to be designed in the two-dimensional feature subspace $Y = [y_1 \ y_2]^T$ can be defined by the following equation:

$$h(Y) = Y^T Q Y + V^T Y + v_0 = [y_1 \ y_2] \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + [v_1 \ v_2] \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + v_0. \ (10)$$

The matrix Q, vector V and scalar v_0 are the unknowns to be optimally determined. The quadratic equation (10) can be represented in a linear form as:

$$h(Y) = [q_{11} \ q_{12} \ q_{22} \ v_1 \ v_2] \begin{bmatrix} y_1^2 \\ 2y_1 y_2 \\ y_2^2 \\ y_1 \\ y_2 \end{bmatrix} + v_0 = V_z^T Z + v_0. \ (11)$$

In order to achieve the largest possible between-class and shortest within-class scattering we have selected the following function as the optimization criterion [13]:

$$f = \frac{P_1 {\eta_1}^2 + P_2 {\eta_2}^2}{P_1 {\sigma_1}^2 + P_2 {\sigma_2}^2} \ (12)$$

where P_1 and P_2 are probabilities and

$$\eta_l = E\{h(Z)/\omega_l\} = E\{V_z^T Z + v_0/\omega_l\} = V_z^T M_l + v_0 \ (13)$$

$${\sigma_l}^2 = var\{h(Z)/\omega_l\} = var\{V_z^T Z + v_0/\omega_l\} = V_z^T \Sigma_l V_z. \ (14)$$

M_l and Σ_l are the mean vectors and covariance matrices, respectively, of the vector Z for each of the two classes l to be classified. After optimization of the function f, the optimal vector V_z, and thus the optimal matrix Q and vector V, gets the following form

$$V_z = \begin{bmatrix} q_{11} \\ q_{12} \\ q_{22} \\ v_1 \\ v_2 \end{bmatrix} = [P_1\Sigma_1 + P_2\Sigma_2]^{-1}(M_2 - M_1) \quad (15)$$

while the optimal scalar is

$$v_0 = -V_z^T(P_1M_1 + P_2M_2) \quad (16)$$

which finishes the design of the quadratic classifier.

3. Results and discussion

In order to extract representative features and before the application of the wavelet transform it is necessary to choose the type of the basic wavelet as well as the number of resolution levels into which the vibration signal segments will be decomposed. After analysis of several types of the basic wavelets, the fourth-order Daubechies wavelet was selected since it demonstrated better discriminatory potential and also has good localizing properties both in the time domain and the frequency domain [14]. The five-level wavelet decomposition of the vibration signals recorded with a sampling frequency of 12000 Hz resulted into the following six frequency sub-bands A_5 0-187.5 Hz, D_5 187.5-375 Hz, D_4 375-750 Hz, D_3 750-1500 Hz, D_2 1500-3000 Hz and D_1 3000-6000 Hz.

Following the wavelet decomposition, the logarithmic energy entropy of the obtained wavelet coefficients in each sub-band were extracted as representative features in the time-frequency domain. The entropy is a measure of the degree of order/disorder of the signal. So it can provide useful information about the underlying dynamical process associated with the signal. For example, a very ordered process can be a signal with a narrow band spectrum. Its energy will be almost zero except for the sub-band that includes the representative signal frequency while the entropy will be near zero. A signal generated by a random process will represent a disordered behaviour and will have significant contribution from all frequency sub-bands. Consequently its energy will be almost equal for all sub-bands while the entropy will take maximum value.

In total 6 features were extracted for each of 1024 analyzed segments. Now each original segment recorded in the time domain is represented by its feature vector $X = [x_1 x_2 \cdots x_6]^T$, and thus a point in the feature space with a dimension of 6. The extracted features together with their mean values and standard deviations for all four different classes of interest are presented in Table 1. Obviously in presence of a bearing fault there is a shift in energy in the vibration signals from lower to higher sub-bands while at the same time the disorder in the vibration signals in lower sub-bands decreases and higher sub-bands increases. In Table 1 and Figure 8, it can also be noticed that most of the extracted features have a certain potential for the fault detection but not for the fault diagnosis.

Table 1. Mean and standard deviation of the features extracted from different sub-bands

Index	Feature	Sub-band	Normal		Ball fault		Inner race fault		Outer race fault	
			μ	σ	μ	σ	μ	σ	μ	σ
x_1	Log. En. Entropy	A_5	20.62	16.63	-99.70	17.29	-124.2	16.97	-182.5	14.96
x_2	Log. En. Entropy	D_5	-9.674	13.20	-86.74	14.37	-113.6	13.63	-157.5	14.94
x_3	Log. En. Entropy	D_4	-50.41	17.60	-146.2	22.97	-119.6	13.85	-256.7	21.39
x_4	Log. En. Entropy	D_3	-2.009	39.46	-337.6	50.89	-221.0	26.74	-560.4	58.71
x_5	Log. En. Entropy	D_2	-423.9	37.51	-210.8	33.56	-323.7	33.03	-504.2	46.38
x_6	Log. En. Entropy	D_1	-2145	148.5	-597.1	61.37	-981.6	54.50	-1211	57.77

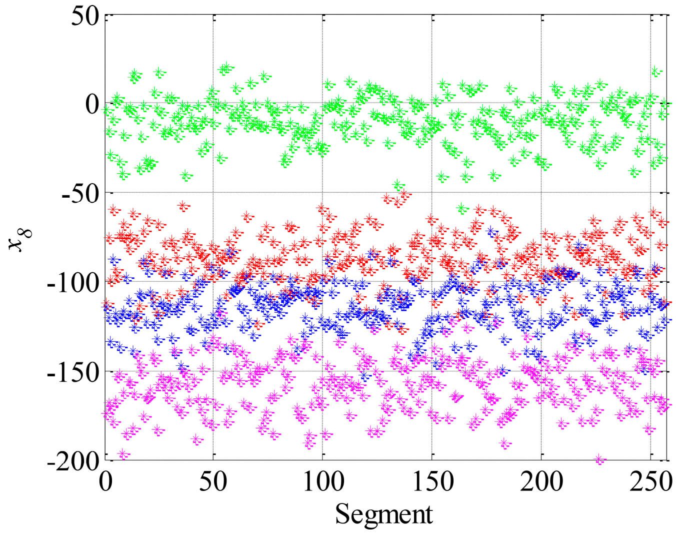

Figure 8. Entropy of the wavelet coefficients in the sub-band D_5, normal (green), ball fault (red), inner race fault (blue) and outer race fault (magenta)

Since none of the extracted features is sufficiently reliable for FDD it is necessary to find their optimal combination in order to achieve a better separability between different classes of bearing condition. That is usually done by a mapping of the existing feature space into a new one whose dimension can be reduced without any significant loss of information that makes the classification process much simpler. Although the PCA is one of the most widely used techniques for reduction of the feature space dimension [16] in this paper we apply the technique based on scatter matrices [13] since it is more suitable for classification problems as described in Section 2.3. At first, we reduce the feature space dimension to enable the fault detection and then repeat the same procedure in order to diagnose the detected fault as shown in Figures 9 and 10. Obviously in this way separabilty between different classes is increased compared with Figure 8. After the dimension reduction we designed suitable quadratic classifiers following the procedure described in Section 2.4, that is also the last step in design of the new technique for early FDD. The first quadratic classifiers shown in Figure 9 separates normal from faulty condition, *i.e.* performs the fault detection, while the other two quadratic classifiers shown in Figure 10 are able to separate all three different bearing faults from each other, and thus perform the fault diagnosis.

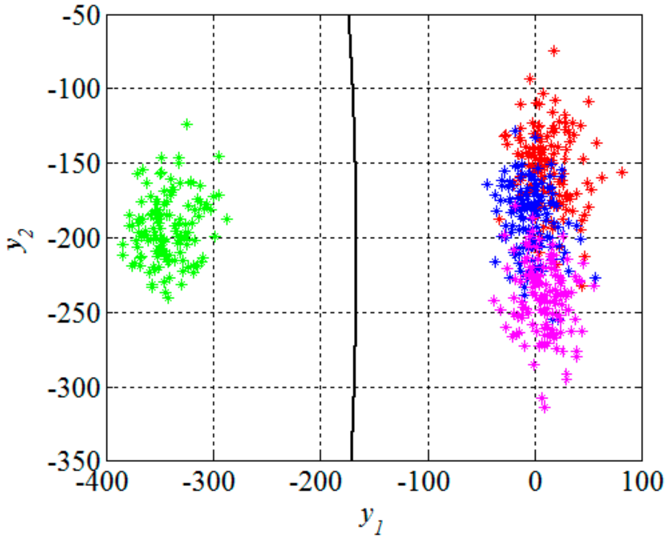

Figure 9. Dimension reduction and classification for the fault detection, normal (green) and faulty (red, blue and magenta) segments of the design set

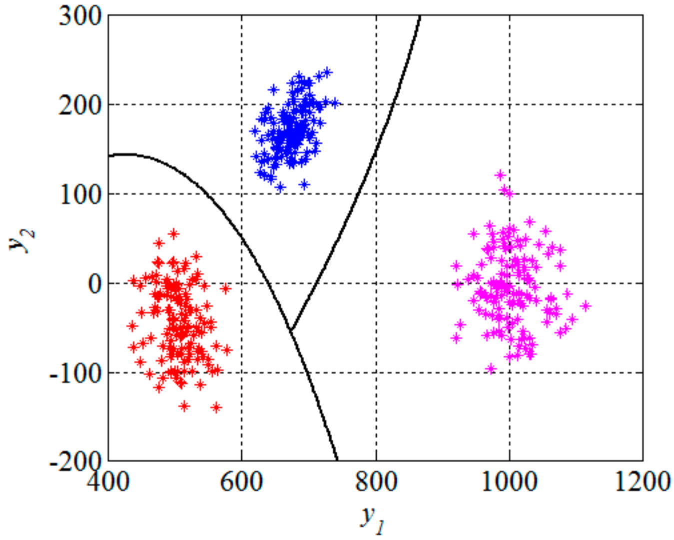

Figure 10. Dimension reduction and classification for the fault diagnosis, segments of the design set with ball fault (red), inner race fault (blue) and outer race fault (magenta)

Unlike in the previous two figures where the design set of 512 segments is shown, Figures 11 and 12 show the remaining set of 512 segments used to test the performance of the designed classifiers and thus the new technique for early FDD as well. Statistical performances such as sensitivity, specificity and accuracy of the new technique are estimated based on the classification results. The sensitivity is defined as a ratio between the number of correctly classified segments and the total number of the segments for each of classes individually. The specificity is also calculated for each of classes individually and it represents the ratio between the number of correctly classified segments of the other classes and the total number of the segments in them. The accuracy is calculated as the ratio between the total number of correctly classified segments and the total number of segments in all classes together. As it can be noticed in Figures 11 and 12 all these three statistical performances of the new technique for early FDD in bearings are equal to 100%.

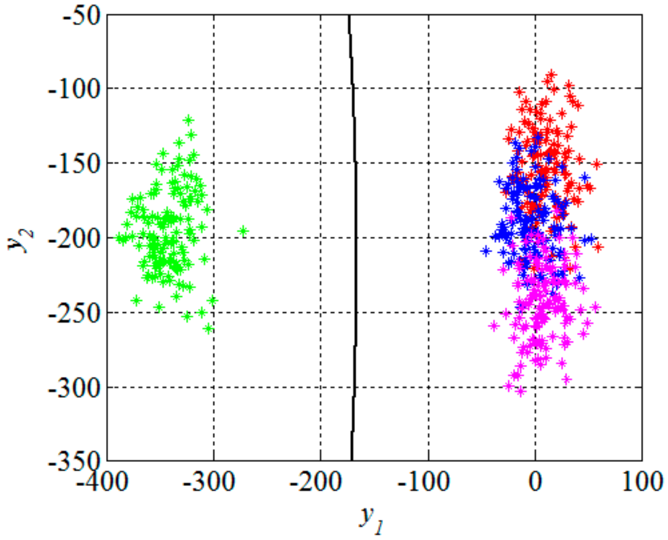

Figure 11. Dimension reduction and classification for the fault detection, normal (green) and faulty (red, blue and magenta) segments of the testing set

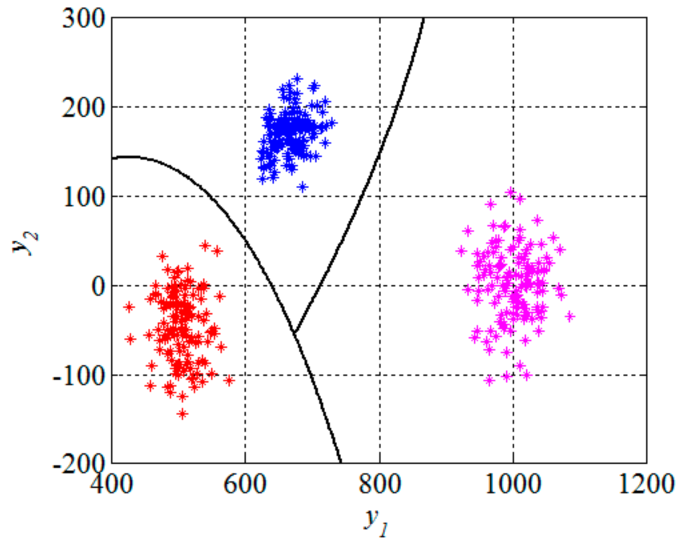

Figure 12. Dimension reduction and classification for the fault diagnosis, segments of the testing set with ball fault (red), inner race fault (blue) and outer race fault (magenta)

Having in mind the results of other techniques available in the literature and tested on the same vibrations signals, e.g. 41 papers published in Mechanical Systems and Signal Processing between 2004 and early 2015 [17], the new technique demonstrated a very good performance which is either better or comparable with other available techniques which usually deploy much more complex algorithms. In addition, in this work only the segments with the smallest fault diameter were used because we were interested in incipient FDD. It should also be emphasized that the vibration signals were normalized before the feature extraction. In that way we managed to overcome one of main disadvantages of other techniques in terms of application in a real production environment since most of them also rely on the amplitude of the vibration signals as one of the key discriminatory features. However, the amplitude has been found as unreliable in real applications since it varies even with healthy bearings, e.g. depending on their load.

4. Conclusion

In order to further increase productivity and energy efficiency of rotating machinery it is necessary to deploy an advanced techniques for an early FDD. In this paper we described such a technique to be used in rotating-element bearings as the most common components of rotating machinery. The new technique based on the logarithmic energy entropy and statistical pattern recognition demonstrated a very high accuracy that is either better or comparable with other available techniques which in most cases require a very large training set and a large number of parameters to be selected and adjusted to obtain good results. A special attention has been paid to robustness of the new technique not only during the feature extraction and the reduction of the feature space dimension but also during the classification process that resulted in the choice of quadratic classifiers known for both their simplicity and a high level of robustness in the applications of this type. Quadratic classifiers have also one more important advantage that is possibility of visualization of the classification results in two-dimensional space. As part of our future work we plan to test the new technique in a real production environment.

Acknowledgments

The support from the Marie Curie FP7-ITN InnHF, Contract No: PITN-GA-2011- 289837 and the Erasmus Mundus Action II EUROWEB+, Contract No: 552125-EM-1-2014-1-SE-ERA MUNDUS-EMA21is gratefully acknowledged.

References

1. Abu-Zeid MA, Abdel-Rahman SM. Bearing problems' effects on the dynamic performance of pumping stations. Alex Eng J **2013**; 52:241–8.
2. Tandon N, Choudhury A. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. Tribol Int **1999**; 32:469–80.
3. Russell PC, Cosgrave J, Tomtsis D, Vourdas A, Stergioulas L, Jones GR. Extraction of information from acoustic vibration signals using Gabor transform type devices. Meas Sci Technol **1998**; 9:1282–90.
4. Koo IS, Kim WW. Development of reactor coolant pump vibration monitoring and a diagnostic system in the nuclear power plant. ISA Trans **2000**; 39:309–16.
5. Peng ZK, Chu FL. Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. Mech Syst Signal Pr **2004**; 18:199–221.
6. Kankar PK, Sharma SC, Harsha SP. Fault diagnosis of ball bearings using continuous wavelet transform. Appl Soft Comput **2011**; 11:2300–12.

7. Lazzerini B, Volpi SL. Classifier ensembles to improve the robustness to noise of bearing fault diagnosis. Pattern Anal Appl **2013**; 16:235:51.
8. Samanta B, Al-Balushi KR. Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. Mech Syst Signal Pr **2003**; 17:317–28.
9. Samanta B, Al-Balushi KR, Al-Araimi SA. Artificial neural networks and genetic algorithm for bearing fault detection. Soft Comput **2006**; 10:264–71.
10. Lou X, Loparo KA. Bearing fault diagnosis based on wavelet transform and fuzzy inference. Mech Syst Signal Pr **2004**; 18:1077–95.
11. Lei Y, He Z, Zi Y, Hu Q. Fault diagnosis of rotating machinery based on a new hybrid clustering algorithm. Int J Adv Manuf Tech **2008**; 35:968–77.
12. Stepanic P, Latinovic IV, Djurovic Z. A new approach to detection of defects in rolling element bearings based on statistical pattern recognition. Int J Adv Manuf Tech **2009**; 45:91–100.
13. Fukunaga K. Introduction to Statistical Pattern Recognition. Boston: Academic Press; 1990.
14. Rao R, Bopardikar A. (1998). Wavelet transforms: Introduction to theory and applications. Addison-Wesley Longman; 1998.
15. Coifman RR, Wickerhauser MV. Entropy-based Algorithms for best basis selection. IEEE Trans Inf Theory **1992**; 38:713–18.
16. Sun W, Chen J, Li J. Decision tree and PCA-based fault diagnosis of rotating machinery. Mech Syst Signal Pr **2007**; 21:1300–17.
17. Smith WA, Randall RB. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. Mech Syst Signal Pr **2015**; 64–65:100–31.