Scientific Research Publishing

Engineering, 2017, 9, 524-539
http://www.scirp.org/journal/eng
ISSN Online: 1947-394X
ISSN Print: 1947-3931

An Approach to Fault Diagnosis of Rotating Machinery Using the Second-Order Statistical Features of Thermal Images and Simplified Fuzzy ARTMAP

Faisal Al Thobiani[1], Van Tung Tran[2], Tiedo Tinga[2]

[1]Faculty of Marine Science, King Abdulaziz University, Jeddah, KSA
[2]Faculty of Engineering Technology, University of Twente, Enschede, The Netherlands
Email: vtung.tran@gmail.com, v.tran@utwente.nl

How to cite this paper: Thobiani, F.A., Tran, V.T. and Tinga, T. (2017) An Approach to Fault Diagnosis of Rotating Machinery Using the Second-Order Statistical Features of Thermal Images and Simplified Fuzzy ARTMAP. *Engineering*, **9**, 524-539.
https://doi.org/10.4236/eng.2017.96033

Received: May 28, 2017
Accepted: June 18, 2017
Published: June 21, 2017

Abstract

Thermal image, or thermogram, becomes a new type of signal for machine condition monitoring and fault diagnosis due to the capability to display real-time temperature distribution and possibility to indicate the machine's operating condition through its temperature. In this paper, an investigation of using the second-order statistical features of thermogram in association with minimum redundancy maximum relevance (mRMR) feature selection and simplified fuzzy ARTMAP (SFAM) classification is conducted for rotating machinery fault diagnosis. The thermograms of different machine conditions are firstly preprocessed for improving the image contrast, removing noise, and cropping to obtain the regions of interest (ROIs). Then, an enhanced algorithm based on bi-dimensional empirical mode decomposition is implemented to further increase the quality of ROIs before the second-order statistical features are extracted from their gray-level co-occurrence matrix (GLCM). The highly relevant features to the machine condition are selected from the total feature set by mRMR and are fed into SFAM to accomplish the fault diagnosis. In order to verify this investigation, the thermograms acquired from different conditions of a fault simulator including normal, misalignment, faulty bearing, and mass unbalance are used. This investigation also provides a comparative study of SFAM and other traditional methods such as back-propagation and probabilistic neural networks. The results show that the second-order statistical features used in this framework can provide a plausible accuracy in fault diagnosis of rotating machinery.

Keywords

Thermal Images, Second-Order Statistical Features, Gray-Level

Co-Occurrence Matrix, Minimum Redundancy Maximum Relevance, Rotating Machinery Fault Diagnosis, Simplified Fuzzy ARTMAP

1. Introduction

Together with the development of science and technology, modern rotating machinery in industry has been increasingly developing toward large scale, high speed operation, more precision, and high degree of automation. In the meantime, its structure is gradually becoming more complex, increasing the degree of integration where the entire production could be interrupted once a part or a link fails. These demand to improve the capability of condition monitoring and fault diagnostic technologies and use effective signals so that the potential faults of such machine can be early detected and diagnosed. Traditionally, acoustic and vibration signals are widely used for machine condition monitoring and fault diagnosis due to their easy-to-measure characteristics and their useful information of machine state containing in the features for analysis. Some outstanding works referred in [1]-[6] have been successfully used these signals for machine fault detection and fault diagnosis.

However, it is challenging to extract useful features for acoustics and vibration due to noise contaminating in the acquired signals. Indeed, the most obvious technique to obtain a vibration signal is by direct measurement using vibration transducer rigidly mounted on machine. This not only requires a high-performance vibration transducer which is capable of withstanding harsh environmental condition, but also demands a costly investment for measuring instrument where a large number of measuring points are concerned [7]. A main disadvantage is that the vibration signals contain strong noise which needs an effective signal processing tool to get useful information. Similarly, the acoustic signal is easily contaminated in a normal industrial environment due to the fact that airborne sound from machine is noisy and complex. That is a reason why the acoustic signal has been received slight attention for machinery condition monitoring and fault diagnosis [8]. It could state that alternative signals being more accurate are necessary.

Infrared thermography (IRT) is a non-contact and non-destructive technique, which can detect infrared energy emitted from object, convert it to temperature, and display image of temperature distribution. The digital output image of IRT is called thermogram or thermal image where each pixel of this image has a specific temperature value and the image's contrast is derived from the differences in surface temperature. Therefore, thermogram has recently been considered as a new signal used for condition monitoring and fault diagnosis of rotating machinery. However, in this field, research in image processing incorporated with intelligent system is still at the early stages. In condition monitoring scope, Bagavathiappan *et al.* [9] used infrared thermography to monitor the operating conditions of blower bearings, shaft at the impeller end, and motors in ventila-

tion systems of nuclear plants. The result of this study showed that thermography could assist in detecting the abnormal operation of various components at an early stage of impending failure. Leemans *et al.* [10] evaluated the possibility of IRT to online monitor the element temperatures of an industrial blower, which included a 500 kW electric motor, a drive motor bearing, and two bearings supporting the blower, to detect wear or other defects. In fault diagnosis, thermogram was used by Younus *et al.* [11]. In their study, thermograms of rotating machinery conditions were decomposed by two-dimensional discrete wavelet transform. For each level obtained from the decomposition process, the first-order statistical features were extracted and selected by Mahalanobis distance and relief algorithm to choose salient features. Subsequently, support vector machine (SVM) and linear discriminant analysis were applied as classifiers for each level. Other studies related to the use of IRT for fault diagnosis could be found in the references [12] [13] [14].

Generally, in order to deal with rotating machinery fault diagnosis based on intelligent techniques, the features presenting the characteristics of signal are first extracted. It is similar to the approaches of using thermograms where their extracted features maybe roughly divided into three categories: structural, spectral, and statistical. In structural approaches, texture primitive, the basic element of texture, is used to form a more complex texture pattern by grammar rules that specify the generation of texture pattern [15]. The advantage of the structural approach is that it provides a good symbolic description of the image; however, this feature is more useful for synthesis than analysis tasks [16]. In spectral approaches, the texture image is transformed into frequency domain, and then the extraction of texture features can be carried out by analyzing the power spectrum. The spectral approaches are limited in practice due to lack of either spatial localization or filter resolution at which one can localize a spatial structure in natural textures. Finally, statistical approaches do not attempt to understand explicitly the hierarchical structure of the texture. Instead, they represent the texture indirectly by the non-deterministic properties that govern the distributions and relationships between the grey-levels of an image. This is the reason why statistical texture features are commonly used in machine fault diagnosis.

So far, most of studies of fault diagnosis using thermogram have only focused on the histogram features, which are the first-order statistical texture features, due to their simplicity. The histogram features only provide information related to the grey-level distribution and ignore the spatial interaction among image pixels. They are not able to measure if all low-value grey-levels are positioned together, or they are interchanged with the high-value grey-levels [17]. It was early argued that they were insufficient for adequate texture description and second-order statistical features were required, as efficiently reflected in features computed from the co-occurrence matrix [18]. Furthermore, approaches based on the second-order statistics have achieved higher discrimination rates than the spectral and structural approaches have [19]. Consequently, the second-order statistical features are considered for fault diagnosis in this paper and extracted

from the gray-level co-occurrence matrix (GLCM), which was firstly introduced by Haralick *et al.* [18]. In addition, other features including cluster shade, cluster prominence, and maximum probability proposed in [20] [21] [22] are also investigated.

Based on the features, the machine conditions could be precisely identified through classification models. These classification models have a wide range of approaches which are varied from model-based to artificial intelligence-based. Among these, artificial intelligence (AI) is regularly used owing to their accuracy and flexibility. Such AI models require "minimum configuration intelligence" since no detailed analysis of the fault mechanism is necessary, nor is any modeling of the system required. Once an AI model is used, fault classification can be accomplished without a highly trained and skilled personnel required. A review of techniques including AI for machinery fault diagnosis could be found in the study of Jardine *et al.* [23]. Recently, SVM [24] belonging to statistical approaches has been considered as a remarkable model in classification and attracted much attention by researchers in fault diagnosis. However, in our previous work [25], the comparative performance of SVM and simplified fuzzy ARTMAP (SFAM) [26] was carried out and the result shows that SFAM is superior to SVM in aspect of the accuracy and computational cost. Accordingly, SFAM is used as the classification to diagnose the conditions of rotating machinery in this study. Furthermore, its classification results and those of other traditional AIs such as back-propagation NN and probabilistic NN are carried out to appraise the advantages of the proposed framework.

2. Theoretical Background

2.1. GLCM and Second-Order Statistical Texture Features

GLCM is a matrix of the relative frequencies P_{ij} of two neighboring pixels having grey-level i and j. This matrix is a function of two parameters: relative distance measured in pixel numbers d and their relative orientation θ being quantized to 45˚ intervals. Thus, for different values of θ and d, different GLCMs are generated. Due to the intensive nature of computations involved, only the distance d = 1 or 2 pixels with angles θ = 0˚, 45˚, 90˚, and 135˚ are normally considered [18]. Suppose an image to be analyzed in rectangle and has N_x columns and N_y rows. Suppose that the grey-level appearing at each pixel is quantified to N_g levels. Let $L_x = \{1,2,\cdots,N_x\}$ be the columns, $L_y = \{1,2,\cdots,N_y\}$ be the rows, and $G_x = \{0,1,\cdots,N_g-1\}$ be the set of N_g quantized grey-levels. The set $L_y \times L_x$ is the set of pixels of the image ordered by their row-column designations. The image I can be represented as a function that assigns some grey-level in G to each pixel or pair of coordinates in $L_y \times L_x$; $I: L_y \times L_x \to G$. The unnormalized frequencies can be defined by

$$P(i,j,d,0^\circ) = \#\{((k,l),(m,n)) \in (L_y \times L_x) \times (L_y \times L_x): k-m=0, |l-n| = d, I(k,l) = i, I(m,n) = j\} \quad (1)$$

$$P\left(i,j,d,45^\circ\right)=\#\left\{\left((k,l),(m,n)\right)\in\left(L_y\times L_x\right)\times\left(L_y\times L_x\right):\left(k-m=d,l-n=-d\right)\vee\left(k-m=-d,l-n=d\right),I\left(k,l\right)=i,I\left(m,n\right)=j\right\} \tag{2}$$

$$P\left(i,j,d,90^\circ\right)=\#\left\{\left((k,l),(m,n)\right)\in\left(L_y\times L_x\right)\times\left(L_y\times L_x\right):\left|k-m\right|=d,l-n=0,I\left(k,l\right)=i,I\left(m,n\right)=j\right\} \tag{3}$$

$$P\left(i,j,d,135^\circ\right)=\#\left\{\left((k,l),(m,n)\right)\in\left(L_y\times L_x\right)\times\left(L_y\times L_x\right):\left(k-m=d,l-n=d\right)\vee\left(k-m=-d,l-n=-d\right),I\left(k,l\right)=i,I\left(m,n\right)=j\right\} \tag{4}$$

where # denotes the number of elements in the set.

Using the co-occurrence matrix above, the second-order statistical features are given in Table 1.

2.2. The mRMR Based Feature Selection

Mutual information (MI) is a quantity that measures the level of similarity between features and the level of correlation between feature and class. Accordingly, the MI of features should be minimized to decrease the redundancy among them and MI of feature and class should be maximized to retain the high relevance. mRMR [27] is a MI based feature selection method simultaneously considering both the relevance and the redundancy in a framework. In term of MI, the relevance of a feature set S for the class c is defined by the mean value of all MI values between the individual feature f_i and the class c. The criterion of maximum relevance is given as follow:

$$\max D\left(S,c\right);\ D=\frac{1}{\left|S\right|}\sum_{f_i\in S}I\left(f_i,c\right) \tag{5}$$

The redundancy of all features in the set S is the mean value of all MI values between the feature f_i and f_j. The minimum redundancy criterion is defined as:

$$\min R\left(S\right);\ R=\frac{1}{\left|S\right|^2}\sum_{f_i,f_j\in S}I\left(f_i,f_j\right) \tag{6}$$

The mRMR feature set is obtained by optimizing the conditions described in Equations (5) and (6) simultaneously. In order to optimize these conditions, it is necessary to combine them into a single criterion function. According to [28], the two simplest combinations of these conditions are mutual information difference (MID) and mutual information quotient (MIQ) forms:

$$\max\left(D-R\right) \tag{7}$$

$$\max\left(D/R\right) \tag{8}$$

mRMR uses the following algorithm to solve this optimization problem. The first feature is selected according to Equation (5), *i.e.* the feature with the highest $I\left(f_i,c\right)$. The remaining features are selected in an incremental way: earlier selected features remain in the feature set. Suppose that m features are already selected

Table 1. The second-order statistical features.

Features	Expression
Energy	$f_1 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\{p(i,j)\}^2$
Contrast	$f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j) \middle\| i-j\| = n\right\}$
Correlation	$f_3 = \dfrac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(ij)\,p(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y}$
Variance	$f_4 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i-\mu)^2\,p(i,j)$
Homogeneity	$f_5 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\dfrac{1}{1+(i-j)^2}p(i,j)$
Sum average	$f_6 = \sum_{i=2}^{2N_g} i p_{x+y}(i)$
Sum variance	$f_7 = \sum_{i=2}^{2N_g}(i-f_8)^2\,p_{x+y}(i)$
Sum entropy	$f_8 = -\sum_{i=2}^{2N_g} p_{x+y}(i)\log\left(p_{x+y}(i)\right)$
Entropy	$f_9 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\log\left(p(i,j)\right)$
Difference variance	$f_{10} = \text{variance of } p_{x-y}$
Difference entropy	$f_{11} = -\sum_{i=0}^{N_g-1} p_{x-y}(i)\log\left(p_{x-y}(i)\right)$
Information measures of correlation	$f_{12} = \dfrac{f_9 - HXY1}{\max(HX,HY)}$ $f_{13} = \left(1-\exp\left(-2(HXY2-f_9)\right)\right)^{1/2}$
Maximal correlation coefficient	$f_{14} = (\text{second largest eigenvalue of } Q)^{1/2}$
Auto-correlation	$f_{15} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(ij)\,p(i,j)$
Dissimilarity	$f_{16} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\lvert i-j\rvert p(i,j)$
Cluster shade	$f_{18} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i+j-\mu_x-\mu_y)^4\,p(i,j)$
Cluster prominence	$f_{18} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i+j-\mu_x-\mu_y)^4\,p(i,j)$
Maximum probability	$f_{19} = \max\left(p(i,j)\right)$

- N_g is the number of distinct grey-level in the quantized image.
- $p(i,j)$ is the (i,j)th entry in a normalized GLCM: $p(i,j) = P(i,j)/R$; R is a normalizing constant ($R = \sum_{i\times j} P(i,j)$ in this study.
- μ_x, μ_y, σ_x, and σ_y are the means and standard deviations of p_x and p_y; μ is the mean of μ_x and μ_y; $p_x(i)$ and $p_y(j)$ are the marginal-probabilities:

$$p_x(i) = \sum_{j=1}^{N_g} p(i,j),\ \ p_y(j) = \sum_{i=1}^{N_g} p(i,j)$$

$$\mu_x = \sum_{i=1}^{N_g} i p_x(i),\ \ \mu_y = \sum_{j=1}^{N_g} j p_y(j)$$

$$\sigma_x = \sqrt{\sum_{i=1}^{N_g} P_x(i)(i-\mu_x)^2}$$

$$\sigma_y = \sqrt{\sum_{j=1}^{N_g} P_y(j)(j-\mu_y)^2}$$

$$p_{x+y}(k) = \sum_{\substack{i=1\\ i+j=k}}^{N_g}\sum_{j=1}^{N_g} p(i,j),\ k = 2,3,\cdots,2N_g$$

$$p_{x-y}(k) = \sum_{\substack{i=1\\ |i-j|=k}}^{N_g}\sum_{j=1}^{N_g} p(i,j),\ k = 0,1,\cdots,N_g-1$$

- HX and HY are the entropies of p_x and p_y, respectively:

$$HX = -\sum_{i=1}^{N_g} p_x(i)\log\left(p_x(i)\right),$$

$$HY = -\sum_{j=1}^{N_g} p_y(j)\log\left(p_y(j)\right)$$

$$HXY1 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\log\left(p_x(i)\,p_y(j)\right)$$

$$HXY2 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p_x(i)\,p_y(j)\log\left(p_x(i)\,p_y(j)\right)$$

$$Q(i,j) = \sum_{k=1}^{N_g}\frac{p(i,k)\,p(j,k)}{p_x(i)\,p_y(j)}$$

for the set S, we want to select additional features from the set $\Omega_S = \Omega - S$. We optimize the MI between both features and class label based on the following two conditions:

$$\max_{f_i\in\Omega_S} I(f_i, c) \tag{9}$$

$$\min_{f_i\in\Omega_S}\frac{1}{|S|}\sum_{f_j\in S} I(f_i, f_j) \tag{10}$$

The condition in Equation (9) is equivalent to the maximum relevance condition in Equation (5), while Equation (10) is an approximation of the minimum redundancy condition of Equation (6). The two ways to combine relevance and redundancy described in Equations (7) and (8) lead to the selection criteria of a new feature:

$$\max_{f_i \in \Omega_S} \left\{ I(f_i, c) - \frac{1}{|S|} \sum_{f_j \in S} I(f_i, f_j) \right\} \tag{11}$$

$$\max_{f_i \in \Omega_S} \left\{ I(f_i, c) \Big/ \frac{1}{|S|} \sum_{f_j \in S} I(f_i, f_j) \right\} \tag{12}$$

2.3. Simplified Fuzzy ARTMAP Network (SFAM)

SFAM is a simplified version of fuzzy ARTMAP [29] by reducing the complicated and redundancy architectures that is the main drawback of the original model for classification task. As a result, SFAM is faster than fuzzy ARTMAP and easier to understand. The details of this network could be found in [26].

3. The Proposed Framework

The proposed framework for machinery fault diagnosis is shown in **Figure 1**. This framework is initiated by the capture of thermal images of different machine conditions; then, these images are preprocessed for cropping the regions of interest (ROIs), removing the noise, and enhancing the contrast in ROI using the histogram equalization (HE) algorithm. For further improving the image information to increase the diagnosis accuracy, these images are enhanced by using a combined method of bi-dimensional empirical mode decomposition and PCA fusion (BEMD-PCAF) proposed in our previous study. Further details of this method could be found in [30]. After being enhanced, second-order statistical features are extracted from the GLCMs. Nevertheless, the feature set contains

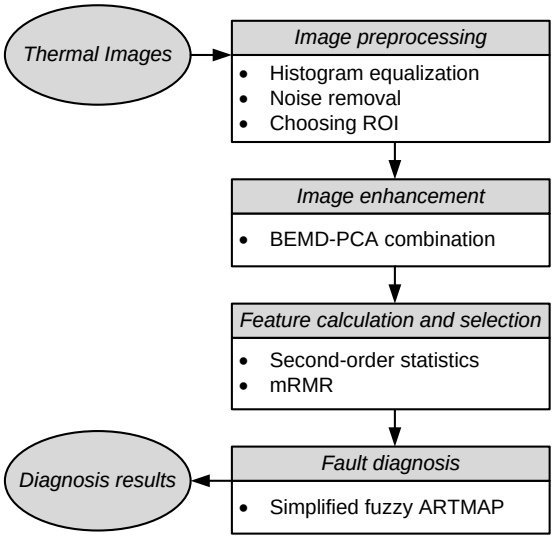


Figure 1. Proposed framework for machinery fault diagnosis.

many redundant as well as relevant features leading to the necessity of feature selection to reduce the computation cost, select prominent features, and eliminate the irrelevant features for avoiding the issue of dimensionality curse. Generally speaking, feature selection methods can broadly fall into three families: filter-based, wrapper-based and embedded methods [31]. Among these, filter-based method is widely used due to its computational efficiency and is employed in this study via mRMR. Finally, the selected feature set is partitioned into training set and test set to build the classifier and validate it, respectively.

4. Experiment

To validate the proposed framework, a series of experiments were carried out by using a fault simulator which consists of driving motor, shaft, disk, PC for saving data, and thermal camera as shown in **Figure 2**. The short shaft, which is of 30 mm diameter and is supported by two ball bearings at the ends, was attached to the shaft of the driving motor through a flexible coupling. This coupling is also used to adjust the misalignment condition on the fault simulator. In order to create the unbalance condition, a disk with many available tapped holes to add extra mass was attached on the shaft. The variable speed DC motor (0.5 HP) with speed up to 3450 rpm was used as the driving motor.

The thermal camera, which is the key device for image acquisition, used for experiments was a long-wave infrared camera from FLIR with the thermal sensitivity of 0.08˚C at 30˚C. Some its parameters require to be set consisting of emissivity (0.9), relative humidity (50%), and distance between the focal length of camera and object (2 m). All of these parameters are chosen according to experimental condition and they were maintained constant in the experiments. The main specifications of the thermal camera and fault simulator are shown in **Table 2**. The experimental procedure for each condition was carried out as following: the speed of the motor was increased gradually up to 900 rpm and was held for five minutes to enable the machine to reach its stable condition. The image acquisition processes for normal, misalignment, mass unbalance, and bearing

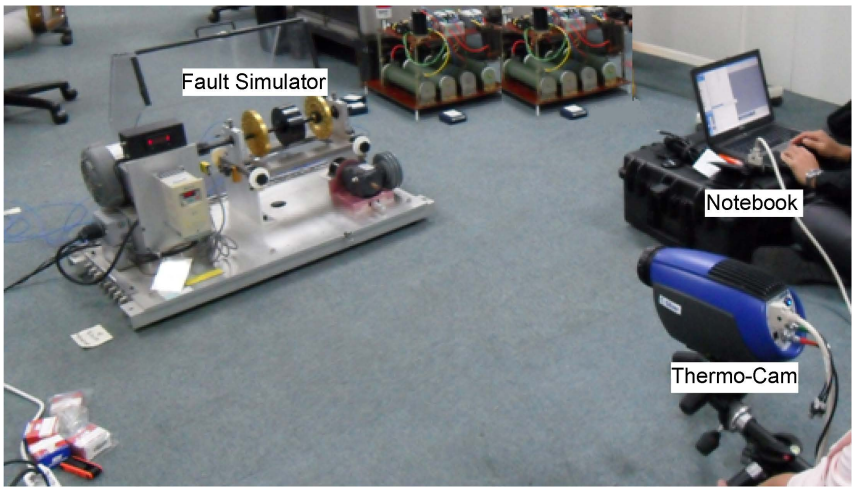


Figure 2. Experimental setup.

fault conditions were then conducted. These faults were created by adjusting the dial screws on the left and right ends of the base plate of the simulator (misalignment), adding a screw 0.02 kg in one of the tapped holes in the rotor disk (mass unbalance), and using outer-race faulty bearing with the defect size 0.3556 mm. For each condition, twenty samples (20) were taken and saved directly to the PC.

5. Results and Discussion

5.1. Image Preprocessing and Feature Calculation

The experimental images collected from different conditions of rotating machinery contain many regions of the fault simulator and background. To focus on fault diagnosis of rotating machinery and avoid unnecessary computation for other regions, a rectangle ROI with the size of 150 × 20 pixels is designated as shown in Figure 3. Then, HE technique is employed for the ROI to augment the

Table 2. Main specification of thermal camera and fault simulator.

Devices	Specification
Thermal camera (FLIR-A 40 series)	• Detector type: focal plane array uncooled microbolometer • Spectral range: 7.5 to 13 μm • Storage temperature range: −40˚C to +70˚C • Solid object materials and surface treatments exhibit emissivity ranging from approximately 0.1 to 0.95 • For short distance, humidity is default value of 50% • Thermal sensitivity: 0.08˚C at 30˚C • Accuracy: ±2˚C • Encapsulation: IP 40 (Determined by connector type)
Fault simulator	• Shaft diameter: 30 mm • Bearing: two ball bearings • Bearing housings: two bearing housings, aluminum horizontally split bracket for simple and easy changes, tapped to accept transducer mount • Bearing housing base: completely movable using jack bolts for easy misalignment in all three planes • Rotors: two rotors, 152.5 mm diameter with two rows of tapped holes at every 20˚ (with lip for introducing unbalance force)

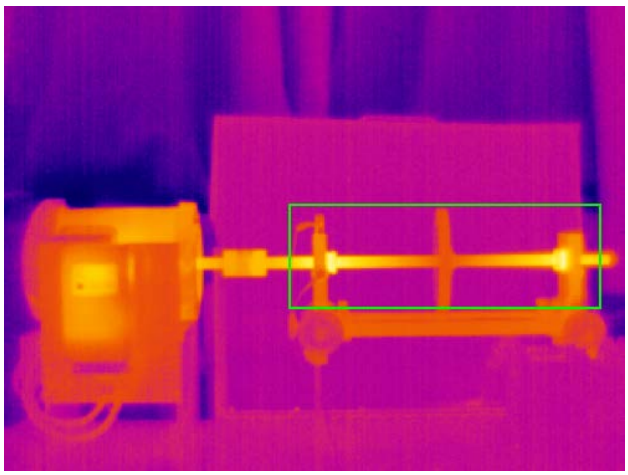


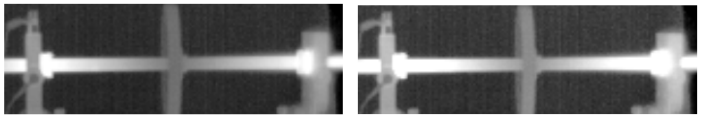
Figure 3. Thermal image and ROI.

contrast. For further improving the image information to increase the diagnosis accuracy, these images are enhanced by BEMD-PCAF. The result of these pre-processing showed in Figure 4 indicates that the visibility of the original image has been improved.

Next, the process of feature calculation is carried out to extract the second-order statistical features introduced in section 2.1. As mentioned, these features are computed from each of the GLCMs obtained by using different values of the relative distance d and the relative orientation θ. The distance d parameter is of importance in the computation of GLCM. As reported in the studies [18] [22], the classification result was best when using features from matrices of $d = 1$ or 2. Hence, the relative distance d as 1 with the orientation θ of 0˚, 45˚, 90˚, and 135˚ is implemented for this study, then averaging these values. In addition, six different values of grey-levels N_g = 8, 16, 32, 64, 128, 256 are also investigated to appraise which value can provide the highest accuracy. Theoretically, 38 features consisting of 19 features mentioned in section 2.1 and their ranges are computed from the image for each grey-level value. However, the feature f_{14}, namely maximal correlation coefficient, is not used in this study due to the fact that some values of $p_x(i)$ or $p_y(j)$ are equal of zeros leading to computational instability. The visualization of the feature distribution in the feature space is shown in Figure 5, where the features of maximum probability, cluster shade, and cluster prominence are presented. It can be seen that the features of image after using BEMD-PCAF are better in cluster of the features being in same condition and superior to separation between the features of different condition than the ones enhancing by HE. This helps the classification more easily attaining the high accuracy without necessity of using any methods to map the features into another space.

5.2. Feature Selection and Classification

The number of features obtained from the previous stage is high dimensionality. Too many features may unnecessarily increase the complexity of the training and classification task; conversely, insufficient selection of features may have a detrimental effect on the classification results. In feature-based techniques, there are two tactics to reduce the high dimensionality as well as select the salient features which are high correlation with the target class label: feature selection and feature extraction. Feature extraction is a technique that transforms the existing features into a lower dimensional space, while feature selection selects a subset of the existing features that optimizes one or more criteria. Notice that the transformation of feature extraction may provide a better discrimination but cannot



(a) (b)

Figure 4. (a) Original image, (b) Enhanced image.

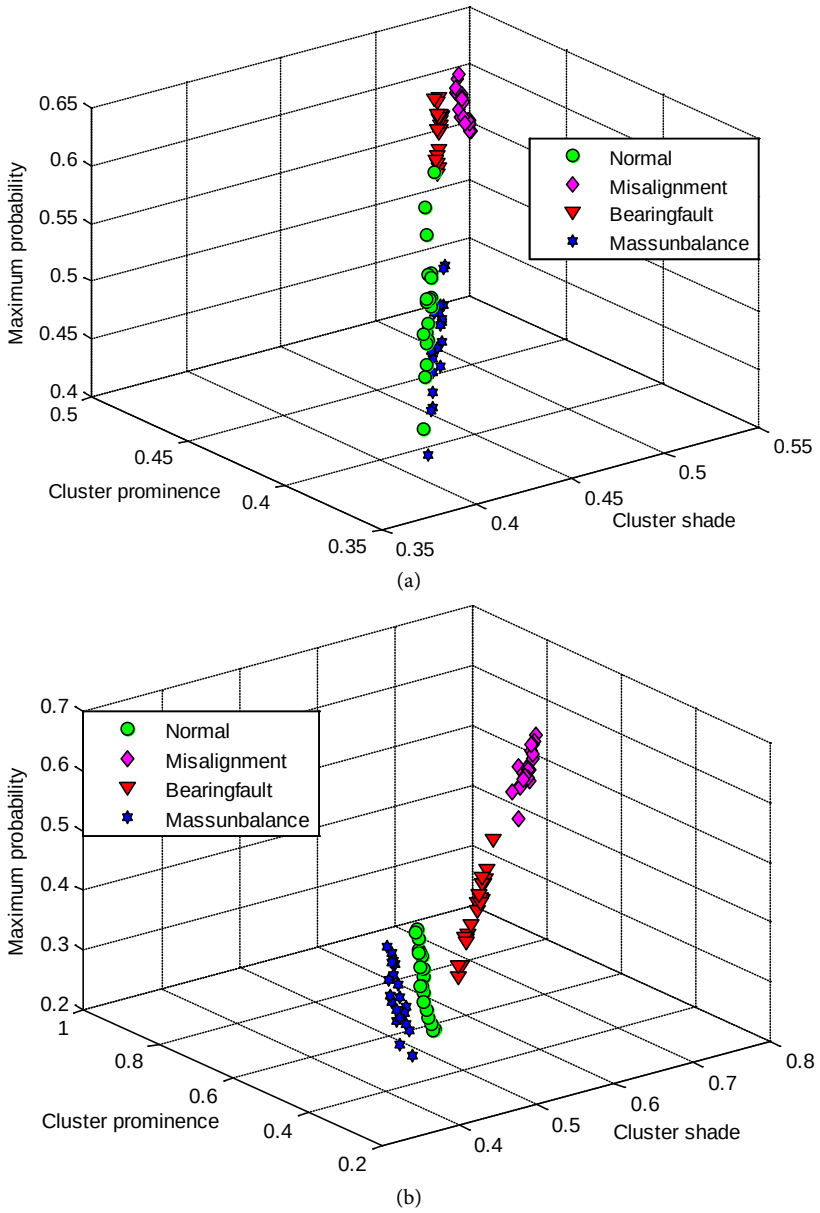


Figure 5. Second-order features of image with grey-level = 8: (a) Enhanced by HE, (b) Enhanced by BEMD-PCAF.

retain the original physical interpretation as feature selection.

In this study, the feature selection method based on mRMR algorithm using MID criterion is used for the purpose of the dimensionality reduction due to its stability in producing feature subsets even MID and MIQ can provide a similar accuracy [32]. Since mRMR is a filter-based method, its subset features necessitate combining with classifier to evaluate the diagnosis accuracy. For each grey-level value, the total features which have 80 samples for 4 machine conditions are randomly partitioned by holdout validation method into 50% for training set to generate the SFAM diagnostic model and the rest for test set to evaluate the model's accuracy. In the training mode, SFAM is trained by basic network setting, *i.e.* fast learning $\beta = 1$ and conservative mode $\alpha = 0.001$. The

value of vigilance parameter (VP) varies from 0 to 0.9 with an increment step of 0.1 to investigate the performance. Due to the randomized selection of samples for the training set and test set, the process of partitioning total features and classes, selecting feature, training and evaluating diagnostic model is repeated 10 times and then average the classification results. In the feature selection process, the number of the selected features is gradually increased from 1 to 36 to determine the number of features sufficing for classifier.

The classification results of SFAM in the testing mode with different values of VPs, grey-levels, and the number of selected features obtained from mRMR are shown in Figure 6. It can be seen that most of the classification accuracy achieve 100% with only one feature selected when the grey-level of 32. This indicates that the selected feature, namely cluster shade in this case, gives a highest relevance with the target labels. However, this accuracy is reduced when increasing the number of selected features over 4. The second grey-level where SFAM classifier can provide a high accuracy and stable for all VP values is of 128; however, the classifier only reaches to 100% accuracy after 3 features selected which are respectively mean of sum average, mean of variance, and cluster prominence. For the other grey-levels, SFAM either achieves lower accuracy, for instance grey-level of 8 and 256, or uses a large number of features to attain a significant accuracy such as grey-level of 64. With the grey-level of 8, SFAM only achieves 99.64% once the selected features are 26 for the value of VP as 0, or achieves 100% when the number of features of 27 for VP of 0.8, which are high computational cost; for other values of VP, SFAM provides a lower accuracy. It is similar for the grey-level of 256. In case of grey-level of 64 where SFAM provides better results, the classification accuracy can attain 100% when the selected features is of 4; however, this only happens in some of VP values such as 0, 0.4, 0.6, 0.7, and 0.8. Thus, the grey-levels of 32 and 128 can give good performance of classification and are chosen for the comparative study in the next section.

5.3. Comparative Performance of SFAM, BPNN, and PNN

As observed in Figure 6, with the VP value of 0.4, SFAM can give a high and stable performance for both grey-levels of 32 and 128. Four selected features used for SFAM in these cases are also applied for back-propagation neural networks (BPNNs) and probabilistic neural networks (PNNs) to evaluate the three-classifier performance. In case of grey-level of 32, the selected features obtained from mRMR are cluster shade, mean of contrast, dissimilarity, and mean of difference variance. In case of grey-level of 128, mean of sum average, mean of variance, cluster prominence, and mean of maximum probability are selected. The networks are trained with ten hidden nodes and Levenberg-Marquardt algorithm. The classification results of the three classifiers are shown in Table 3. It can be seen that SFAM and PNN are superior to BPNN in case of the grey-level of 32. In case of grey-level of 128, BPNN accuracy is higher than ones of SFAM and PNN when one number of selected features is used, and vice versa. Table 3 also shows the training time of the classifiers when one feature is used

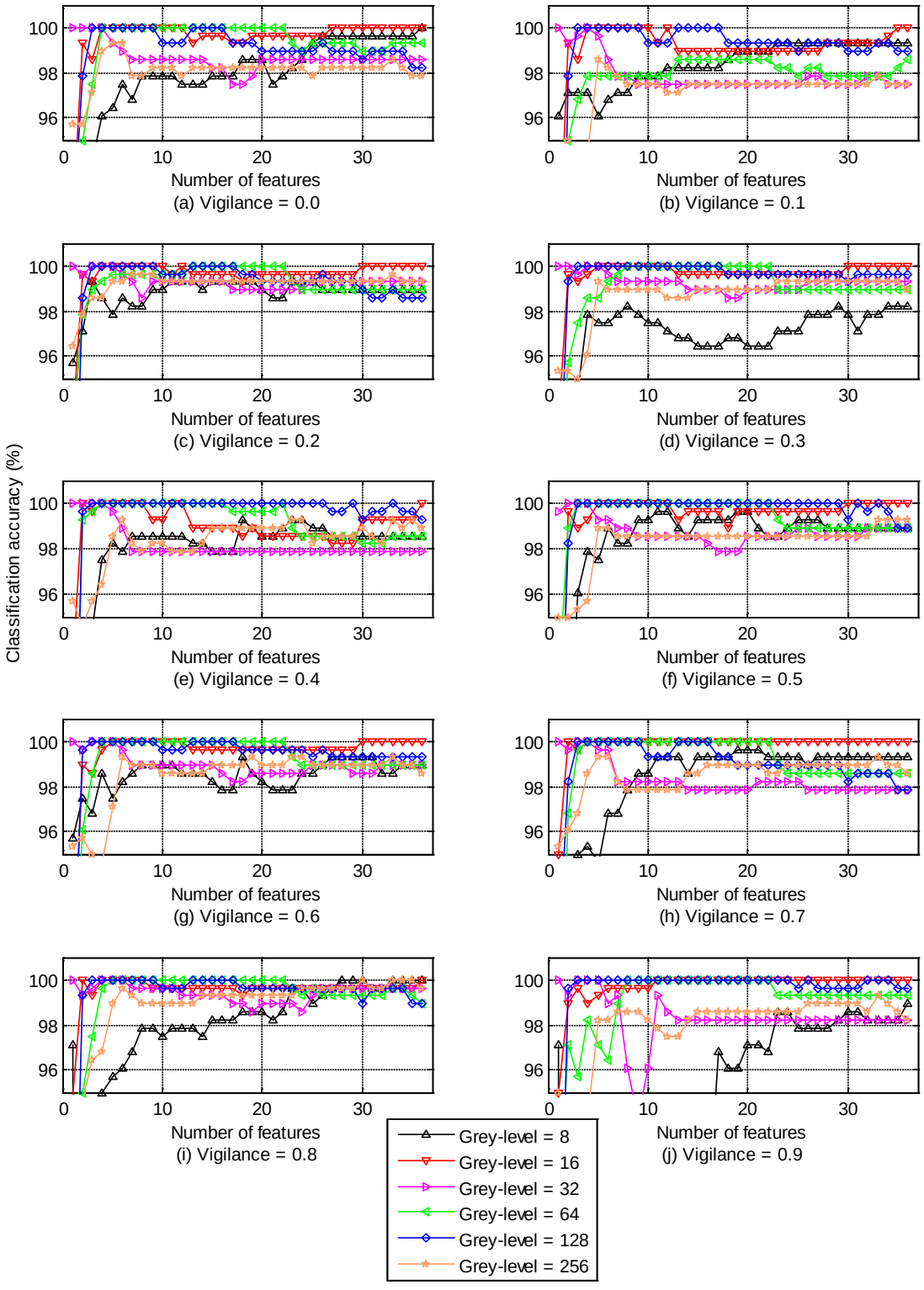


Figure 6. Classification results of SFAM.

for classification. The training time of SFAM is significantly low in comparison with ones of BPNN and PNN. This indicates that SFAM can give better performance with low computational cost that is very useful for real application where huge data could be used.

Table 3. Classification accuracy of SFAM, BPNN, and PNN.

Grey-level	Classifier	Training time (s)	Classification accuracy (%)			
			Number of features			
			1	2	3	4
32	SFAM	0.0018	100	100	100	100
	BPNN	0.4704	99.59	99.88	99.32	99.71
	PNN	0.3533	100	100	100	100
128	SFAM		85	99.64	100	100
	BPNN		91.83	97.37	99.49	99.11
	PNN		81.79	97.5	100	100

6. Conclusion

This paper has presented a new approach of the second-order statistical features of thermal image for fault diagnosis by introducing them into the framework including mRMR and SFAM. The experimental thermal images of a simulator with different conditions such as normal, misalignment, faulty bearing, and mass unbalance are used for this investigation. The second-order statistical features are extracted from these images with the grey-levels of 8, 16, 32, 64, 128, and 256. Then, mRMR based on MID is employed to select the features which have high relevance to the machine condition to input the SFAM classifier. As a result, the classification accuracy of SFAM achieves 100% with only cluster shade feature selected from whole the feature set when the grey-level of 32 is used. In another grey-level of 128, SFAM can reach to 100% accuracy until 3 features are selected. Additionally, a comparative study of the performance of SFAM and other traditional networks BPNN and PNN has been carried out. The results show that SFAM not only provides a better performance but also has insignificantly computational cost. This indicated that SFAM is eminently suitable to use for real fault diagnosis applications.

References

[1] Toutountzakis, T., Tan, C.K. and Mba, D. (2005) Application of Acoustic Emission to Seeded Gear Fault Detection. *NDT & E International*, **38**, 27-36. https://doi.org/10.1016/j.ndteint.2004.06.008

[2] Wu, J.D. and Chuang, C.Q. (2005) Fault Diagnosis of Internal Combustion Engines Using Visual Dot Patterns of Acoustic and Vibration Signals.*NDT & E International*, **38**, 605-614. https://doi.org/10.1016/j.ndteint.2005.02.007

[3] Wang, J. and Hu, H. (2006) Vibration-Based Fault Diagnosis of Pump Using Fuzzy Technique. *Measurement*, **39**, 176-185. https://doi.org/10.1016/j.measurement.2005.07.015

[4] Yang, B.-S., Han, T. and An, J.L. (2004) ART-Kohonenneural Network for Fault Diagnosis of Rotating Machinery. *Mechanical Systems and Signal Processing*, **18**, 645-657. https://doi.org/10.1016/S0888-3270(03)00073-6

[5] Lee, S.K. and White, P.R. (1997) Higher-Order Time-Frequency Analysis and Its Application to Fault Detection in Rotating Machinery. *Mechanical Systems and Signal Processing*, **11**, 637-650. https://doi.org/10.1006/mssp.1997.0098

[6] Cheng, J., Yang, Y. and Yu, D. (2010) The Envelope Order Spectrum Based on Ge-

neralized Demodulation Time-Frequency Analysis and Its Application to Gear Fault Diagnosis. *Mechanical Systems and Signal Processing*, **24**, 508-521. https://doi.org/10.1016/j.ymssp.2009.07.003

[7] Rmda, S. and Alhussein, A. (2013) Petroleum Pumps' Current and Vibration Signatures Analysis Using Wavelet Coherence Technique. *Advances in Acoustics and Vibration*, **2013**, Article ID: 659650, 6 p.

[8] Baydar, N. and Ball, A. (2003) Detection of Gear Failures via Vibration and Acoustic Signals Using Wavelet Transform. *Mechanical Systems and Signal Processing*, **17**, 787-804. https://doi.org/10.1006/mssp.2001.1435

[9] Bagavathiappan, S., Saravanan, T., George, N.P., Philip, J., Jayakumar, T. and Raj, B. (2007) Condition Monitoring of Exhaust System Blowers Using Infrared Thermography. *Insight*, **50**, 512-515. https://doi.org/10.1784/insi.2008.50.9.512

[10] Leemans, V., Destain, M., Kilundu, B. and Dehombreux, P. (2011) Evaluation of the Performance of Infrared Thermography for On-Line Condition Monitoring of Rotating Machines. *Engineering*, **3**, 1030-1039. https://doi.org/10.4236/eng.2011.310128

[11] Younus, A.M.D. and Yang, B.-S. (2012) Intelligent Fault Diagnosis of Rotating Machinery Using Infrared Thermal Image. *Expert Systems with Applications*, **39**, 2082-2091. https://doi.org/10.1016/j.eswa.2011.08.004

[12] Younus, A.M.D., Widodo, A. and Yang, B.-S. (2010) Evaluation of Thermography Image Data for Machine Fault Diagnosis. *Nondestructive Test Evaluation*, **25**, 231-247. https://doi.org/10.1080/10589750903473617

[13] Mazioud, A., Ibos, L., Khlaifi, A. and Durastanti, J.F. (2008) Detection of Rolling Bearing Degradation Using Infrared Thermography. *International Conference on Quantitative InfraRed Thermography*, Krakow, 2-5 July 2008. https://doi.org/10.21611/qirt.2008.02_11_14

[14] Bagavathiappan, S., Lahiri, B.B., Saravanan, T., Philip, J. and Jayakumar, T. (2013) Infrared Thermography for Condition Monitoring—A Review. *Infrared Physics & Technology*, **60**, 35-55. https://doi.org/10.1016/j.infrared.2013.03.006

[15] Sheshadri, H. and Kandaswamy, A. (2007) Experimental Investigation on Breast Tissue Classification Based on Statistical Feature Extraction of Mammograms. *Computerized Medical Imaging and Graphics*, **31**, 46-48. https://doi.org/10.1016/j.compmedimag.2006.09.015

[16] Materka, A. and Strzelecki, M. (1998) Texture analysis Methods—A Review. Institute of Electronics, Tech University of Lodz, Lodz.

[17] Aggarwal, N. and Agrawal, R.K. (2012) First and Second Order Statistics Features for Classification of Magnetic Resonance Brain Images. *Journal of Signal and Information Processing*, **3**, 146-153. https://doi.org/10.4236/jsip.2012.32019

[18] Haralick, R.M., Shanmugam, K. and Dinstein, I. (1973) Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, **3**, 610-621. https://doi.org/10.1109/TSMC.1973.4309314

[19] Weszka, J., Deya, C. and Rosenfeld, A. (1976) A Comparative Study of Texture Measures for Terrain Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, **6**, 269-285. https://doi.org/10.1109/TSMC.1976.5408777

[20] Conners, R.W., Trivedi, M.M. and Harlow, C.A. (1984) Segmentation of a High-Resolution Urban Scene Using Texture Operators. *Computer Vision, Graphics, and Image Processing*, **25**, 273-310. https://doi.org/10.1016/0734-189X(84)90197-X

[21] Haralick, R.M. (1979) Statistical and Structural Approaches to Texture. *Proceeding of the IEEE*, **67**, 786-804. https://doi.org/10.1109/PROC.1979.11328

[22] Soh, L.K. and Tsatsoulis, C. (1999) Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 780-795. https://doi.org/10.1109/36.752194

[23] Jardine, A.K.S., Lin, D. and Banjevic, D. (2006) A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance. *Mechanical Systems and Signal Processing*, **20**, 1483-1510. https://doi.org/10.1016/j.ymssp.2005.09.012

[24] Vapnik, V.N. (1999) The Nature of Statistical Learning Theory. Springer, Berlin.

[25] Tran, V.T., Thobiani, F.A. and Ball, A. (2013) An Application to Transient Current Signal Based Induction Motor Fault Diagnosis of Fourier-Bessel Expansion and Simplified Fuzzy ARTMAP. *Expert Systems with Applications*, **40**, 5372-5384. https://doi.org/10.1016/j.eswa.2013.03.040

[26] Kasuba, T. (1993) Simplified Fuzzy ARTMAP. *AI Expert*, **8**, 19-25.

[27] Peng, H., Long, F. and Ding, C. (2005) Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance and Min-Redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**, 1226-1238. https://doi.org/10.1109/TPAMI.2005.159

[28] Ding, C. and Peng, H.C. (2003) Minimum Redundancy Feature Selection from Microarray Gene Expression Data. *Proceedings of the Computational Systems Bioinformatics*, 2003, 523-528.

[29] Carpenter, C.A., Grossberg, S., Markuzon, N., Reynolds, J.H. and Rosen, D.B. (1992) Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps. *IEEE Transactions on Neural Networks*, **3**, 698-713. https://doi.org/10.1109/72.159059

[30] Tran, V.T., Yang, B.S., Gu, F. and Ball, A. (2013) Thermal Image Enhancement Using Bi-Dimensional Empirical Mode Decomposition in Combination with Relevance Vector Machine for Rotating Machinery Fault Diagnosis. *Mechanical Systems and Signal Processing*, **38**, 601-614. https://doi.org/10.1016/j.ymssp.2013.02.001

[31] Guyon, I. and Elisseeff, A. (2003) An Introduction to Variable and Feature Selection. *The Journal of Machine Learning Research*, **3**, 1157-1182.

[32] Gulgezen, G., Cataltepe, Z. and Yu, L. (2009) Stable and Accurate Feature Selection. *Lecture Notes in Computer Science*, **5781**, 455-468. https://doi.org/10.1007/978-3-642-04180-8_47