



University of HUDDERSFIELD

University of Huddersfield Repository

Tran, Van Tung, Yang, Bo-Suk, Gu, Fengshou and Ball, Andrew

Thermal Image Enhancement using Bi-dimensional Empirical Mode Decomposition in Combination with Relevance Vector Machine for Rotating Machinery Fault Diagnosis

Original Citation

Tran, Van Tung, Yang, Bo-Suk, Gu, Fengshou and Ball, Andrew (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 (2). pp. 601-614. ISSN 0888-3270

This version is available at <http://eprints.hud.ac.uk/id/eprint/16904/>

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/>

Thermal Image Enhancement using Bi-dimensional Empirical Mode Decomposition in Combination with Relevance Vector Machine for Rotating Machinery Fault Diagnosis

Van Tung Tran^{a*}, Bo-Suk Yang^b, Fengshou Gu^a, Andrew Ball^a

^a *School of Computing and Engineering, University of Huddersfield,
Queensgate, Huddersfield HD1 3DH, UK*

^b *Department of Mechanical & Automotive Engineering, Pukyong National University
San 100, Yongdang-dong, Namgu, Busan 608-739, South Korea*

Abstract

In this study, a novel fault diagnosis system for rotating machinery using thermal imaging is proposed. This system consists of bi-dimensional empirical mode decomposition (BEMD) for image enhancement, a generalized discriminant analysis (GDA) for feature reduction, and a relevance vector machine (RVM) for fault classification. Firstly, the thermal image obtained from machine conditions is decomposed into intrinsic mode functions (IMFs) by using BEMD. At each decomposed level, the IMF is expanded and fused with the residue by grey-scale transformation and principal component analysis fusion technique, respectively. The enhanced image is then formed by the improved IMFs in reconstruction process. Subsequently, feature extraction is applied for the enhanced images to obtain histogram features which characterize the thermal image and contain useful information for diagnosis. The high dimensionality of the achieved feature set can be reduced by GDA implementation. Moreover, GDA also assists in the increase of the feature cluster separation. Finally, the diagnostic results are performed by RVM. The proposed system is applied and validated with the thermal images of a fault simulator. A comparative study of the classification results obtained from RVM, support vector machines, and adaptive neuro-fuzzy inference system is also performed to appraise the accuracy of these models. The results show that the proposed diagnosis system is capable of improving the classification accuracy and efficiently assisting in rotating machinery fault diagnosis.

Keywords: Thermal image; Bi-dimensional empirical mode decomposition; Rotating machinery fault diagnosis; Generalized discriminant analysis; Relevance vector machines

1. Introduction

Rotating machinery covers a wide range of mechanical equipment and is of importance in industrial applications. Therefore, faults occurring in rotating machinery may severely affect operations in industry and even safety. In order to minimize the number of breakdowns as well as to increase the reliability, rotating machinery condition should be monitored for symptoms and incipient fault detection. By this, the life of machinery could be prolonged and the catastrophic consequences of unplanned failure could be avoided. Consequently, fault diagnosis and condition monitoring of rotating machinery have been the research subject in recent years. Traditionally, to monitor the conditions and diagnose the faults of rotating machinery, acoustics and vibration are commonly used signals due to their easy-to-measure characteristics and analysis. Some outstanding works [1-6] have successfully employed these signals for fault detection and diagnosis areas. However, it is necessary to increase the fault diagnosis capability and implement suitable signals which intensify the fault detection ability for highly automatic rotating machinery.

In recent times, infrared thermal image has been considered as a new signal which can be applied for fault diagnosis because of the possibility to indicate the object's operating condition through its temperature. In the study of Bagavathiappan et al. [7], the operating conditions of blower bearings, shaft at the impeller end, and motors in ventilation systems of nuclear plants were monitored by the temperature obtained from thermal images. 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. Younus et al. [8] used thermal imaging for diagnosing the faults of rotating machinery in which principal component analysis (PCA) and independent component analysis (ICA) were utilized as feature extraction tools whilst support vector machine (SVM) was employed as a classifier. Moreover, the classification accuracy of SVM was also compared with other classifiers such as parzen probabilistic neural networks, fuzzy k-nearest neighbor, and adaptive resonance theory-Kohonen neural network. In another work in [9], thermal images of rotating machinery conditions were decomposed by two-dimensional discrete wavelet transform. For each level obtained from the decomposition process, histogram features was extracted and selected by Mahalanobis distance and relief algorithm to choose salient features. Subsequently, SVM and linear discriminant analysis were applied as classifiers for each level.

Generally, the classification accuracy of those studies was not significant because the thermal images were not enhanced before implementing feature extraction or classification. Image enhancement, which is a function of image processing, aims to augment some information in an image as a specific requirement, weaken or remove some unwanted information, so that it changes the original image into a more suitable form for human

observations or computer analysis. Traditionally, image enhancement technologies can be divided into two categories [10]: image enhancement on spatial domain and image enhancement on frequency domain. Several methods consisting of histogram equalization, adaptive contrast enhancement, smoothing, sharpening, color processing, etc. have been utilized in these categories.

The first commonly used method for image enhancement is histogram equalization (HE). As a consequence, HE flattens the density distribution of the resultant image and enhances the contrast of the image, since it has an effect of stretching dynamic range [11]. However, HE changes the brightness of the image significantly and makes the image saturated with very bright or dark intensity values [12]. The second method is un-sharp masking (USM). The classic linear USM is implemented by passing a low-contrast image through a linear two-dimensional high-pass filter and then adding a fraction of its output to the origin. This method enlarges the image noise, particularly in uniform areas of even slightly noisy images and causes overshoot artifacts in high-contrast regions. HE and USM are only the compromise between de-noising and enhancing image details. Furthermore, they are also less sensitive to noise presented in the input. Another commonly used method is wavelet transformation. The main advantage of this method is that no artificial information is introduced into the enhanced image. This allows some flexibility in different applications. Nevertheless, wavelet transformation has a major drawback where the basis function has to be defined a priori and this choice may influence the final results.

To overcome the shortcomings of the above traditional methods, several image enhancement methods were proposed in which some of them were based on bi-dimensional empirical mode decomposition (BEMD). BEMD is a two-dimensional generalization of the classical empirical mode decomposition (EMD), which was proposed by Huang et al. [13] to extract the frequency components of a signal. Theoretically, any complicated signal can be decomposed into a set of intrinsic mode functions (IMFs) based upon the local characteristic time scale of the signal. EMD is self-adaptive and highly efficient in non-stationary data analysis because IMFs, working as the basis functions, are determined by the signal itself. Accordingly, BEMD overcomes the wavelet transformation drawback and provides a promising image processing technique that can be applied in various areas of image analysis, e.g. image fusion [14], image compression [15-16], texture analysis [17-18], feature extraction [19], etc. Numerous expansions of BEMD are afterwards developed for image enhancement. Qin et al. [20] used BEMD to decompose the medical images and the high frequency information was then expanded to obtain the enhanced images. Çelebi et al. [21] utilized BEMD to separately decompose the color channels of underwater color images into IMFs; then, the weighted IMFs of each channel were combined to reconstruct the enhanced images [22]. Other image enhancement methods based on ensemble EMD [23] and fast and adaptive BEMD [24] were

developed by Bakhtiari et al. in [22] and [25], respectively. In this study, a novel method which is a combination of BEMD and PCA fusion technique [26] is proposed to enhance the thermal images of machine conditions. The thermal images are iteratively decomposed into the IMFs. At each decomposition step, the obtained IMF is expanded by grey-scale transformation and then fused with the residue by using PCA fusion. The enhanced image is obtained from the reconstruction once the decomposition process is terminated.

Next, the enhanced images are the input of the feature extraction process to draw out useful information. According to Umbaugh [27], the image features consist of histogram, spectral, texture, and color. Among these, histogram features which are truly statistical features are a compact representation of image characteristics without requiring knowledge about them. Moreover, histogram features also provide information about the characteristics of the intensity grey level distribution for the image. Hence, they are suitable for fully automatic characterization of images and are used in this study. Histogram features consist of mean, standard deviation, skewness, energy, entropy, and kurtosis. Normally, these features are rarely usable because of the huge dimensionality leading to the increase of computational complexity and the decrease of accuracy in fault diagnosis. Therefore, dimensionality reduction is the necessity for data preprocessing techniques. In fault diagnosis, there have been numerous approaches for feature reduction such as ICA, PCA [28], and genetic algorithms [29], etc. In this study, the generalized discriminant analysis (GDA) [30] based feature reduction is investigated with the aim of improving the classification performance.

Selecting the model is the last stage in this system in order to diagnose the machine conditions. The classification models have a wide range of approaches which are varied from model-based to pattern recognition-based. Amongst these, machine learning and artificial intelligence are regularly used for machine fault diagnosis system due to their accuracy and flexibility. In machine learning techniques, the relevance vector machine (RVM) which was originally introduced by Tipping [31] is one of the remarkable models. However, the original RVM could only be implemented for binary classification which merely encounters in real applications in general and fault diagnosis in particular. Therefore, RVM was developed from the original for multi-class classification by introducing multinomial logistic regression [32] and applied for fault diagnosis of low speed bearing [33]. In this study, multi-class RVM is used for classifying the different machine conditions such as normal, misalignment, mass unbalance, and bearing fault. A comparative study of the classification results obtained from RVM, SVM [34], and adaptive neuro-fuzzy inference system (ANFIS) [35] is also carried out to appraise the accuracy of these models.

2. Background knowledge

2.1. Histogram features

Let's consider an image I , the first-order histogram probability $P(g)$ can be defined as:

$$P(g) = \frac{N(g)}{M} \quad (1)$$

where M is the number of pixels in the image I , $N(g)$ is the number of pixels at grey level g . The histogram features of image based on the first-order histogram probability are mean, standard deviation, skew, energy, entropy, and kurtosis. These features are expressed as followings

2.1.1. Mean: is the average value that gives some information about general brightness of image. Denote L as total number of grey levels for the available range from 0 to 255 for the image. The mean can be defined as:

$$\bar{g} = \sum_{g=0}^{L-1} gP(g) = \sum_r \sum_c \frac{I(r,c)}{M} \quad (2)$$

where $I(r,c)$ is the grey level value of the image I at point (r,c) ; r and c are respectively row and column.

2.1.2. Standard deviation: is the square root of the variance. It provides us something about the contrast and also describes the spread in the data, so a high contrast image will have a high variance, vice versa. Standard deviation is defined as follows:

$$\sigma_g = \sqrt{\sum_{g=0}^{L-1} (g - \bar{g})^2 P(g)} \quad (3)$$

2.1.3. Skewness: measures the asymmetry about the mean in the grey-level distribution. It is defined as:

$$S = \frac{1}{\sigma_g^3} \sum_{g=0}^{L-1} (g - \bar{g})^3 P(g) \quad (4)$$

The skewness could also be measured by using the mean, mode, and the standard deviation where the mode is defined as the peak or highest value. This method is more computationally efficient, especially when the mean and the standard deviation have already been calculated

$$S' = \frac{\bar{g} - \text{mode}}{\sigma_g} \quad (5)$$

2.1.4. Kurtosis: is a measure characterized the peakedness or flatness of the grey-level distribution. It is given by the ratio of the fourth central moment and the square of the variance.

$$K = \frac{\sum_{g=0}^{L-1} (g - \bar{g})^4}{\sigma_g^4} \quad (6)$$

2.1.5. Energy: is a measure that tells us something about how the grey levels are distributed:

$$ENERGY = \sum_{g=0}^{L-1} [P(g)]^2 \quad (7)$$

The energy measure has a maximum value of 1 for an image with a constant value, and it gets increasingly smaller as the pixel values are distributed across more grey-level values.

2.1.6. *Entropy*: is a measure that provides us how many bits we need to code the image data and is given by:

$$ENTROPY = -\sum_{g=0}^{L-1} P(g) \log_2 [P(g)] \quad (8)$$

2.2. *Bi-dimensional empirical mode decomposition (BEMD)*

The central idea of the BEMD is the sifting process to decompose any given image into its frequency components, namely intrinsic mode functions (IMFs), which satisfy two conditions:

- The number of extrema and the number of zero crossings must either be equal or differ at most by one.
- At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima are zero.

Given the digital image $I = f(x,y)$, $x = 1, \dots, K$, $y = 1, \dots, L$, where K and L are the total number of rows and columns in the discrete-domain image, respectively. The decomposition process is summarized as follows [17, 36]:

1. Initialization: $r_0(x, y) = I(x, y)$ (the residual) and $i = 1$ (index number of IMF)

2. Extraction of the i th IMF:

- (i) Initialize $h_0(x, y) = r_{i-1}(x, y)$, $j = 1$
- (ii) Identify all the extrema involving maxima and minima of $h_{j-1}(x, y)$
- (iii) Compute the upper envelope $u_{j-1}(x, y)$ and the lower envelope $l_{j-1}(x, y)$ of $h_{j-1}(x, y)$ by connecting maxima points and minima points using surface interpolation, respectively.
- (iv) Calculate the mean of the upper and the lower envelopes:

$$m_{j-1}(x, y) = [u_{j-1}(x, y) + l_{j-1}(x, y)] / 2 \quad (9)$$

(v) Update

$$h_j(x, y) = h_{j-1}(x, y) - m_{j-1}(x, y) \quad (10)$$

and $j = j + 1$

(vi) Calculate stopping criterion using the normalized standard deviation (SD) which

can be computed from the two consecutive sifting results:

$$SD_{ij} = \sum_{k=1}^K \sum_{l=1}^L \frac{|h_{i(j-1)}(k,l) - h_{ij}(k,l)|^2}{h_{i(j-1)}^2(k,l)} \quad (11)$$

(vii) Repeat steps (ii) to (vi) until $SD_{ij} \leq \xi$, where ξ is an a priori constant, then

$$s_i(x, y) = h_j(x, y) \text{ is } i\text{th IMF}$$

3. Update the residual $r_i(x, y) = r_{i-1}(x, y) - s_i(x, y)$

4. Repeat steps (2)-(3) with $i = i + 1$ until the number of extrema in $r_i(x, y)$ is less than 2

At the end of the decomposition, the original image $I(x,y)$ can be reconstructed using the following equation

$$I(x, y) = \sum_{k=1}^n c_k(x, y) + r_n(x, y) \quad (12)$$

where $r_n(x,y)$ denotes the final residue, n is the number of IMFs, $c_k(x,y)$ is the k th IMF.

2.3. Generalized discriminant analysis for feature reduction

The generalized discriminant analysis (GDA) [30, 37-38] deals with a nonlinear classification using a kernel function Φ which maps the original space \mathbf{X} into new high-dimensional features space \mathbf{Z} . The within-class scatter and between-class scatter matrix of the nonlinearly mapped data is:

$$B^\phi = \sum_{c=1}^C M_c m_c^\phi (m_c^\phi)^T, \quad W^\phi = \sum_{c=1}^C \sum_{x \in X_c} \phi(x) \phi(x)^T \quad (13)$$

where m_c^ϕ is the mean of class X_c and Z , M_c is the number of samples belonging to X_c . The aim of the GDA is to find such projection matrix U^ϕ that maximizes the ratio

$$U_{opt}^\phi = \arg \max_{U^\phi} \frac{|(U^\phi)^T B^\phi U^\phi|}{|(U^\phi)^T W^\phi U^\phi|} = [u_1^\phi, u_2^\phi, \dots, u_N^\phi] \quad (14)$$

The vectors u^ϕ can be found as the solution of the generalized eigenvalue problem i.e., $B^\phi u_i^\phi = \lambda_i W^\phi u_i^\phi$. The training vectors are supposed to be centered (zero mean, unit variance) in the feature space Z . From the theory of reproducing kernels, any solution $u^\phi \in Z$ must lie in the span of all training samples, i.e.

$$u^\phi = \sum_{c=1}^C \sum_{i=1}^{M_c} \alpha_{ci} \phi(x_{ci}) \quad (15)$$

where α_{ci} are some real weights and x_{ci} is the i th sample of the class c . The solution is obtained

by solving

$$\lambda = \frac{\alpha^T KDK\alpha}{\alpha^T KK\alpha} \quad (16)$$

where $\alpha = (\alpha_c)$, $c = 1, \dots, C$ is a vector of weights with $\alpha_c = (\alpha_{ci})$, $i = 1, \dots, M_c$. The kernel matrix $K(M \times M)$ is composed of the dot products of nonlinearly mapped data, i.e.

$$K = (K_{kl}); \quad k = 1, \dots, C; \quad l = 1, \dots, C \quad (17)$$

where $K_{kl} = (k(x_{ki}, x_{lj})); \quad i = 1, \dots, M_k; \quad j = 1, \dots, M_l$. The matrix $D(M \times M)$ is a block diagonal matrix such that

$$D = (D_c); \quad c = 1, \dots, C \quad (18)$$

where the c th on the diagonal has all elements equal to $1/M_c$. Solving the eigenvalue problem yields the coefficient vector α that defines the projection vectors $u^\phi \in Z$. A projection of a testing vector x_{test} is computed as [25]

$$(u^\phi)^T \phi(x_{test}) = \sum_{c=1}^C \sum_{i=1}^{M_c} \alpha_{ci} k(x_{ci}, x_{test}) \quad (19)$$

The procedure of the proposed algorithm could be summarized as follows:

- Compute the matrices K and D by solving the Eq. (17) and Eq. (18).
- Decompose K using eigenvector decomposition
- Compute eigenvectors α and eigenvalues of Eq. (14)
- Compute u^ϕ using α_{ci} from Eq. (15) and normalize
- Compute projections of the test points onto the eigenvectors u^ϕ using Eq. (19)

2.4. Relevance vector machine

The RVM has the probabilistic Bayesian learning framework, which outputs probabilities of class membership. Originally, RVM was derived and experimented on binary classification where it was desired to predict the posterior probability of membership of one of the classes given the input \mathbf{x} . Recently, the RVM was developed for multi-class classification using multinomial logistic regression [32]. In the following, the basic theory of RVM is briefly introduced. Further details of the RVM could be found in [31]. For binary classification, following statistical convention and generalizing the linear model by applying the logistic sigmoid function $\sigma(y) = 1/(1 + e^{-y})$ to $y(\mathbf{x})$ and adopting the Bernoulli distribution for $P(\mathbf{t} | \mathbf{x})$, the likelihood is written as [31]:

$$P(\mathbf{t} | \mathbf{x}) = \prod_{n=1}^N \sigma\{y(\mathbf{x}_n; \mathbf{w})\}^{t_n} [1 - \sigma\{y(\mathbf{x}_n; \mathbf{w})\}]^{1-t_n} \quad (20)$$

where the targets $t_i \in \{0,1\}$. However, the weights \mathbf{w} cannot be analytically integrated, and so are denied the closed-form expression for either the weight posterior $p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha})$ or the marginal likelihood $P(\mathbf{t} | \boldsymbol{\alpha})$, with a hyper-parameter vector $\boldsymbol{\alpha}$. Due to the fact that the weights cannot be analytically obtained, the approximation procedure proposed by MacKay [39], which is based on Laplace's method, is used as the following:

1. For a fixed value of $\boldsymbol{\alpha}$, the most possible maximum posteriori weights (\mathbf{w}_{MP}) are found, giving the location of the mode of the posterior distribution. Since $p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}) \propto P(\mathbf{t} | \mathbf{w})p(\mathbf{w} | \boldsymbol{\alpha})$, this is equivalent to finding the maximum, over \mathbf{w} , of

$$\log\{P(\mathbf{t} | \mathbf{w})p(\mathbf{w} | \boldsymbol{\alpha})\} = \sum_{n=1}^N [t_n \log y_n + (1-t_n) \log(1-y_n)] - \frac{1}{2} \mathbf{w}^T \mathbf{A} \mathbf{w} \quad (21)$$

where $y_n = \sigma\{y(\mathbf{x}_n; \mathbf{w})\}$, $\mathbf{A} = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$ for the current values of $\boldsymbol{\alpha}$. This is a penalized logistic log-likelihood function and necessitates iterative maximization. Thus, the following procedure adopts the iteratively reweighed least-squares algorithm [40] to find the \mathbf{w}_{MP} .

2. Eq. (21) is differentiated twice, then give

$$\nabla_{\mathbf{w}} \nabla_{\mathbf{w}} \log p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}) |_{\mathbf{w}_{MP}} = -(\boldsymbol{\Phi}^T \mathbf{B} \boldsymbol{\Phi} + \mathbf{A}) \quad (22)$$

where $\mathbf{B} = \text{diag}(\beta_1, \beta_2, \dots, \beta_N)$ is a diagonal matrix with $\beta_n = \sigma\{y(\mathbf{x}_n)\}[1 - \sigma\{y(\mathbf{x}_n)\}]$, and $\boldsymbol{\Phi}$ is the $N \times (N+1)$ design matrix with $\boldsymbol{\Phi} = [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_N)]^T$, where in $\phi(\mathbf{x}_n) = [1, K(\mathbf{x}_n, \mathbf{x}_1), K(\mathbf{x}_n, \mathbf{x}_2), \dots, K(\mathbf{x}_n, \mathbf{x}_N)]^T$. This result is then negated and inverted to give the covariance $\boldsymbol{\Sigma}$ for a Gaussian approximation to the posterior over weights centered at \mathbf{w}_{MP}

3. Using the statistics $\boldsymbol{\Sigma}$ and \mathbf{w}_{MP} , the hyper-parameter $\boldsymbol{\alpha}$ is updated by

$$\alpha_i^{new} = \frac{\gamma_i}{\mathbf{w}_{MP}^2} \quad (23)$$

where $\gamma_i = 1 - \alpha_i \sum_{ii}$; \sum_{ii} is the i th diagonal element of the covariance $\boldsymbol{\Sigma} = (\boldsymbol{\Phi}^T \mathbf{B} \boldsymbol{\Phi} + \mathbf{A})^{-1}$ and $\mathbf{w}_{MP} = \boldsymbol{\Sigma} \boldsymbol{\Phi}^T \mathbf{B} \mathbf{t}$

For solving the multi-class problem, the original formulation of RVM essentially treats the K multi-class problem as a series on n one-against-all binary classification problem. This would translate into training n binary classifiers independently. The likelihood in Eq. (20) is generalized to standard multinomial form:

$$P(\mathbf{t} | \mathbf{w}) = \prod_{n=1}^N \prod_{k=1}^K \sigma\{y_k(\mathbf{x}_n; \mathbf{w}_k)\}^{t_{nk}} \quad (24)$$

where t_{nk} is the indicator variable for observation n to be a member of class k ; y_k is the predictor of class k . Here, a true multi-class likelihood can be stated as [41]:

$$P(\mathbf{t} | \mathbf{w}) = \prod_{n=1}^N \prod_{k=1}^K \sigma\{y_k; y_1, y_2, \dots, y_k\}^{t_{nk}} \quad (25)$$

where the predictors of each class y_k is coupled in the multinomial logic function (or soft-max)

$$\sigma(y_k; y_1, y_2, \dots, y_k) = \frac{e^{y_k}}{e^{y_1} + e^{y_2} + \dots + e^{y_k}} \quad (26)$$

3. The proposed system for fault diagnosis

The proposed system for rotating machinery fault diagnosis using thermal imaging is shown in Fig. 1. This system consists of several modules: image capture, image preprocessing, image enhancement, histogram feature extraction, feature reduction, and classification. Thermal images are captured from the machine conditions such as normal condition, misalignment, mass unbalance, and bearing fault. These images are then processed by the preprocessing module for cropping the region of interest (ROI), removing the noise, and enhancing the contrast in ROI using the HE algorithm before they are input to the image enhancement module.

Fig. 1 Proposed system for thermal image based fault diagnosis

A method for ameliorating the image quality in the enhancement module is proposed as follows:

- (i) BEMD decomposes the original images into IMFs which reduce the frequency information gradually. At each step of decomposition, the high-frequency information part being IMF and low-frequency information part being the residue are obtained. The former expresses the image texture whilst the latter expresses the content of the image.
- (ii) Once the IMF is obtained, it is expanded to be clearer and more prominent by multiplying by factor k . This factor is set in $1 < k < 3$. For k too high, the highest frequency is augmented too much hence the borders become too prominent. For k smaller than 1, results in the loss of borders.
- (iii) The expanded IMF is fused with the residue by using PCA.
- (iv) The steps from (i) to (iii) are iteratively carried out until the decomposition process is terminated.
- (v) Finally, the enhanced image is reconstructed from the fused IMFs.

After being enhanced, the images are input to the feature extraction module for computing the histogram features. However, as mentioned before, the features are normally of high dimensions, are not well clustered, and may contain a large amount of redundant features. These will decrease significantly the performance and increase the computational time if the features are directly input into the classifier. Therefore, GDA is employed to transform the exiting features into lower dimension space and allow for the construction of nonlinear mappings that maximize the class separability in the feature set. Finally, these features will be split into training and test data to generate the diagnostic model through the learning process and to validate the model.

4. Experiment

To validate the proposed system, a series of experiments were carried out using a fault simulator which consists of driving motor, shaft, disk, PC for saving data, and thermal camera as shown in Fig. 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 screwed holes to add extra mass were attached on the shaft. The variable speed DC motor (0.5 HP) with speed up to 3,450 rpm was used as the driving motor. Table 1 shows the main specifications of thermal camera and fault simulator. This camera used in the experiments was a long-wave infrared camera from FLIR with a thermal sensitivity of 0.08 °C at 30 °C.

Fig. 2 Experimental setup

Table 1 Main specification of thermal camera and fault simulator

The thermal camera is the key device and some its parameters require to be set due to their importance for data acquisition, especially for thermal image data. The most important parameter is emissivity and the other parameters are relative humidity, scale temperature, focal length of camera, and distance. All of these parameters are chosen according to experimental condition. In this study, all parameters were maintained constant. The experiment for each condition was carried out as followings: the speed of the motor was increased gradually up to 900 rpm. This speed 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 fault conditions were then conducted. Data from the thermo-cam were saved directly to the PC. The detailed descriptions of image data in four machine condition experiments are shown in Table 2.

Table 2 Detailed descriptions of image data

5. Results and discussion

Fig. 3 shows one of the original thermal images of machine conditions. For the purpose of rotating machinery fault diagnosis and reduction of image processing computation, ROI is chosen from original image as a rectangle, size 150×20 pixels. This size is likewise applied for other images. Subsequently, these ROIs are preprocessed by using the HE technique to enhance the contrast. Then, histogram feature extraction is carried out to describe the characteristics of the machine conditions. In total, 480 feature values (80×6) have been extracted from the ROIs. The images after being enhanced by the HE technique and their features are now called original images and original features, respectively.

Fig. 3 Original thermal image and ROI

In order to observe the feature distribution of the original image, three features can be arbitrarily selected from the feature set. This visualization only provides the information to understand how the features distribute in same machine condition and how the clusters of the features separate in the different conditions. From this visualization, the reason of misclassification, if any, can be estimated based on the overlap of the features. Fig. 4 shows the distribution of the three-first features involving skewness, standard deviation (SD), and the mean of the original images. It can be seen that the features of machine conditions are not well clustered and overlapped with each other even though the images have been enhanced by HE. This significantly decreases the accuracy of classification which leads to the misunderstanding of existing machine condition. Obviously, HE algorithm is not adequate to improve the images for achieving good diagnosis.

To increase the separation among the feature clusters and reduce the feature dimensionality for effective computation, GDA is continuously applied to the original feature set. The sufficient amount of necessary features to characterize the machine condition can be gained based on the eigenvalue of the covariance matrix [42]. As shown in Fig. 5, the number of features is reduced from 6 to 3. Also the structure of the features related to the different conditions is reconstructed and presented in Fig. 6. It can be observed that the features of the same condition in a new space are now located close to each other and are separated from other conditions in comparison to the original ones. Thus, GDA not only increases the effective computation but also assists in a better discrimination of different machine conditions. However, there still exists the overlap between normal and bearing fault conditions, which can cause the misclassification in diagnosis process. It also indicates that the assistance of GDA is not enough to achieve the highest diagnosis performance with the original images.

Fig. 4 Original features

Fig. 5 Eigenvalue of covariance matrix for feature reduction

Fig. 6 Original features obtained from GDA

Obviously, it is necessary to apply an enhancement method to ameliorate the image quality. In this study, the enhancement method based on BEMD is employed in which the stopping criterion and the k factor are chosen as 0.1 and 1.35, respectively. The results of the decomposition process and enhancement according to the proposed method are respectively presented in Figs. 7 and 8. As observed in Fig. 7, the fused IMF of each decomposition level is much better than the IMF derived by the conventional technique. This results in an improved visibility of the image obtained from the reconstruction process as shown in Fig. 8. Similarly, histogram features are also extracted from these enhanced images and GDA is then employed to these extracted features. Figs. 9 and 10 are correspondingly depicted the three-first features of enhanced images and the structure of the features in the GDA space. Evidently, after enhancing, the features are well separated into groups which are similar in characteristics and there is no overlap between the machine conditions. This shows that the proposed enhancement method has assisted in improving the image quality.

Fig. 7 a) Original image, b) IMFs, c) Residues

Fig. 8 a) Original image, b) Enhanced image

Fig. 9 Enhanced image features

Fig. 10 Enhanced image features obtained from GDA

The next stage of the proposed system is the diagnosis task or classification where the features attained from GDA become the inputs to the classifiers involving multi-class RVM, SVM, and ANFIS. In case of RVM, some parameters are predefined e.g. the Gaussian kernel is used as the basis function, and the kernel parameter is set to 0.1. In case of SVM, two algorithms involved one-against-one (SVM-OAO) and one-against-all (SVM-OAA) are used. Furthermore, some parameters are predefined for this classifier such as the regularizing parameter C and the kernel parameter γ are set to 100 and 0.1, respectively. Regarding to the ANFIS classifier, three inputs are used according to the number of features after GDA reduction. For each input, a bell shape is chosen for each membership function (MF) and the number of MFs is chosen as 2. The parameters of these MFs, which are premise parameters and consequent parameters, are automatically adjusted through the learning process in order that the outputs of ANFIS model match the actual classes in training data. The convergence of root mean squared error (RMSE) is utilized for evaluating and terminating the learning process. In the case that either the decreasing rate of the RMSE or the performance is not significant, the learning process will be terminated.

Firstly, the training set, which is randomly partitioned from the total samples by holdout validation method, is used to generate the classification models. In case of ANFIS, the RMSE of the output reaches the convergent stage after 200 training epochs. Also, the premise parameters of the membership functions corresponding to the inputs are changed during the training process. As depicted in Fig. 11, the second and the third membership functions have reformed the shapes for the sake of network convergence according to the given classes in the training data. After being trained, the classifiers are tested against the test data to validate the accuracy in diagnosis task. The classification results of all classifiers in the training and testing processes are presented in Table 3.

Fig. 11 Membership functions

In the training process, all the classifiers including RVM, SVM-OAO, SVM-OAA, and ANFIS achieve 100% accuracy without any misclassification out of 40 samples of training data for both original features and enhanced features. This indicates that the classifiers are well trained and can be applied for diagnosing faults. However, in the testing process in which these classifiers are validated against the test data, the accuracy of each classifier is different. For the original features, the classification accuracy of SVM-OAO and RVM are same as 97.5% whilst SVM-OAA and ANFIS are respectively as 95% and 82.5%. These show that the accuracy of ANFIS is the lowest in comparison with others even though its learning result has been completed. Additionally, these results also indicate that SVM-OAO is superior to SVM-OAA. The miss-classifications are due to the overlap of machine condition features. For the features of enhanced images, the accuracy of ANFIS achieves only 97.5%, whilst RVM, SVM-OAO, and SVM-OAA classifiers achieve 100% accuracy. Obviously, the classification accuracy of all classifiers using the enhanced features is higher than that of the original features. This concludes that the proposed enhancement method based on BEMD has significantly improved the quality of images which results in the increase of the classification accuracy. From the classification results, it can be realized that the accuracy of RVM and SVM-OAO is the same and both are higher than that of SVM-OAA. In order to appraise the efficiency of RVM and SVM-OAO, their training time are considered and shown in Table 4. Evidently, the time for training RVM is much smaller in comparison with that of SVM-AOA. Thereby, RVM is very suitable to use as the fault classifier in real applications.

Table 4 Time-consumption in the training process of RVM and SVM-OAO

6. Conclusions

In this study, the novel fault diagnosis system using thermal images consisting of BEMD based image enhancement, GDA for feature reduction, and RVM for classification have been presented. The thermal images captured from machine conditions are firstly preprocessed by the HE algorithm to enhance the image contrast, removing noise, and cropping to obtain the ROI. These images are further improved by the enhancement method based on the combination of BEMD and PCA fusion. Then, the histogram feature extraction and the GDA feature reduction are respectively carried out to extract the features of enhanced images, reduce the high dimension and increase the cluster separability of the feature data. The classification process provides a comparative study of RVM, SVM including one-against-one and one-against-all algorithms, and ANFIS in fault diagnosis.

As a result, the classification accuracy of the enhanced images is much higher than that of the original images. This can conclude that the proposed image enhancement method has significantly assisted in improving the image quality. Furthermore, the performances of RVM and SVM-OAO are considerably superior to SVM-OAA in the case of the original images. It indicates that RVM and SVM-OAO have great potential for fault diagnosis using thermal imaging. In training time comparison, RVM is faster than SVM-OAO, which is necessary and useful in real applications. Accordingly, the proposed system with RVM used as classifier is capable of efficiently assisting in machine fault diagnosis.

Acknowledgments

This study is carried out in memory of Professor Bo-Suk Yang.

References

- [1] T. Toutountzakis, C.K. Tan, D. Mba, Application of acoustic emission to seeded gear fault detection, *NDT & E International* 38 (2005) 27-36.
- [2] J. D. Wu, C.Q. Chuang, Fault diagnosis of internal combustion engines using visual dot patterns of acoustic and vibration signals, *NDT & E International* 38 (2005) 605-614.
- [3] J. Wang, H. Hu, Vibration-based fault diagnosis of pump using fuzzy technique, *Measurement* 39 (2006) 176-185.
- [4] B.S. Yang, T. Han, J.L. An, ART-Kohonen neural network for fault diagnosis of rotating machinery, *Mechanical Systems and Signal Processing* 18 (2004) 645-657.
- [5] S.K. Lee, P.R. White, Higher-order time-frequency analysis and its application to fault detection in rotating machinery, *Mechanical Systems and Signal Processing* 11 (1997) 637-650.
- [6] J. Cheng, Y. Yang, D. Yu, The envelope order spectrum based on generalized demodulation time-frequency analysis and its application to gear fault diagnosis, *Mechanical Systems and Signal Processing* 24 (2010) 508-521.
- [7] S. Bagavathiappan, T. Saravanan, N.P. George, J. Philip, T. Jayakumar, B. Raj, Condition monitoring of exhaust system blowers using infrared thermography, *Insight* 50 (2007)

512-515.

- [8] A. Younus, A. Widodo, B.S. Yang, Evaluation of thermography image data for machine fault diagnosis, *Nondestructive Testing and Evaluation* 25 (2010) 231-247.
- [9] A. Younus, B.S. Yang, Intelligent fault diagnosis of rotating machinery using infrared thermal image, *Expert System with Applications* 39 (2012) 2082-2091.
- [10] R.C. Gonzalez, R.E. Woods, *Digital image processing*, 3rd ed., MA, Addison-Wesley, 1993.
- [11] Y.T. Kim, Contrast enhancement using brightness preserving bi-histogram equalization, *IEEE Transactions on Consumer Electronics* 43 (1997) 1-8.
- [12] D. Menotti, L. Najman, J. Facon, A.A. de Araujo, Multi-histogram equalization methods for contrast enhancement and brightness preserving, *IEEE Transactions on Consumer Electronics* 53 (2007) 1186-1194.
- [13] N.E. Huang et al., The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceeding of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences* 454 (1998) 903-995.
- [14] H. Hariharan, A. Gribok, M. Abidi, A. Koschan, Image fusion and enhancement via empirical mode decomposition, *Pattern Recognition Research* 1 (2006) 16-32.
- [15] A. Linderhed, 2D empirical mode decompositions in the spirit of image compression, *SPIE Proceedings of Wavelet and Independent Component Analysis Applications IX* 4738 (2002) 1-8.
- [16] A. Linderhed, Compression by image empirical mode decomposition, *IEEE International Conference on Image Processing*, Vol. 1, 2005, pp. I553-6.
- [17] J.C. Nunes, Y. Bouaoune, E. Delechelle, O. Niang, Ph. Bunel, Image analysis by bidimensional empirical mode decomposition, *Image and Vision Computing* 21 (2003) 1019-1026
- [18] J.C. Nunes, S. Guyot, E. Deléchelle, Texture analysis based on local analysis of the bidimensional empirical mode decomposition, *Machine Vision and Applications* 16 (3) (2005) 177-188.
- [19] J. Wan, L. Ren, C. Zhao, Image feature extraction based on the two-dimensional empirical mode decomposition, *Congress on Image and Signal Processing* 1 (2008) 627-631.
- [20] X. Qin, S. Liu, W. Zhengqiang, J. Han, Medical image enhancement method based on 2D empirical mode decomposition, *The 2nd International Conference on Bioinformatics and Biomedical Engineering*, 2008, pp. 2533-2536
- [21] A.T. Çelebi, S. Ertürk, Empirical mode decomposition based visual enhancement of underwater images, *The 2nd International on Image Processing Theory, Tools and Applications*, 2010, pp. 221-224
- [22] S. Bakhtiari, S. Aгаian, M. Jamshidi, A novel empirical mode decomposition based system for medical image enhancement, *IEEE International Systems Conference*, 2011, pp. 145-148
- [23] Z. Wu, N.E. Huang, Ensemble empirical mode decomposition: a noise assisted data analysis method, *Advances in Adaptive Data Analysis* 1(2009) 1-41.
- [24] S.M.A. Bhuiyan, R.R. Adhami, J.F. Khan, Fast and adaptive bidimensional empirical mode decomposition using order-statistics filter based envelope estimation, *EURASIP Journal on Advances in Signal Processing*, Vol. 2008 (2008) 1-18.
- [25] S. Bakhtiari, S. Aгаian, M. Jamshidi, An enhanced empirical mode decomposition based method for image enhancement, *IEEE International Conference on Systems, Man, and Cybernetics*, 2011, pp. 2681 - 2686
- [26] V.P.S. Naidu, J.R. Raol, Pixel-level image fusion using wavelets and principal component analysis, *Defence Science Journal* 58(2008) 338-352.
- [27] S.E. Umbaugh, *Digital Image Analysis and Processing*, CRC Press, 2005.
- [28] A. Widodo, B.S. Yang, Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors, *Expert System with Applications* 33

- (2007) 241-250.
- [29] W. Siedlecki, J. Sklansky, A note on genetic algorithm for large scale feature selection, *Pattern Recognition Letter* 10 (1989) 246-259.
 - [30] G. Baudat, F. Anouar, Generalized discriminant analysis using a kernel approach, *Neural Computation* 12 (2000) 2385-2404.
 - [31] M.E.Tipping, Sparse Bayesian learning and relevance vector machine, *Journal of Machine Learning Research* 1 (2001) 211-244.
 - [32] B. Khishnapuram, L. Carin, M.A.T. Figueiredo, A.J. Hartemink, Sparse multinomial logistic regression: Fast algorithm and generalization bounds, *IEEE Transaction on Pattern Analysis and Machine Intelligence* 27 (2005) 957-968.
 - [33] A. Widodo, E.Y. Kim, J.D. Son, B.S. Yang, A.C.C. Tan, D.S. Gu, B.K. Choi, J. Mathew, Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine, *Expert Systems with Applications* 36 (2009) 7252-7261
 - [34] V.N. Vapnik, *The Nature of Statistical Learning Theory*, New York, Springer, 1999.
 - [35] J.S.R. Jang, ANFIS: Adaptive-Network-based Fuzzy Inference System, *IEEE Transactions on System, Man and Cybernetics* 23 (1993) 665-685.
 - [36] J. C. Nunes, O. Niang, Y. Bouaoune, E. Delechelle, Ph. Bunel, Bidimensional empirical mode decomposition modified for texture analysis, *Lecture Notes in Computer Science* 2749 (2003) 171-177.
 - [37] T.-K. Kim, J. Kittler, Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (2005) 318-327
 - [38] S. Singh, S. Silakari, Generalized discriminant analysis algorithm for feature reduction in cyber attack detection system, *International Journal of Computer Science and Information Security* 6 (2009) 173-180
 - [39] D.J.C. MacKay, The evidence framework applied to classification network, *Neural Computation* 4 (1992) 720-736
 - [40] I.T. Nabney, Efficient training of RBF networks for classification, *Proceedings of the 9th International Conference on Artificial Neural Networks*, Edinburgh, UK, 1999, pp. 210-215.
 - [41] H. Zhang, J. Malik, Selecting shape features using multi-class relevance vector machine, *Technical Report No. UCB/EECS-2005-6*, University of California at Berkeley.
 - [42] B.-S. Yang, A. Widodo, *Introduction of intelligent machine fault diagnosis and prognosis*, Nova Science Publishers Inc., New York, 2009

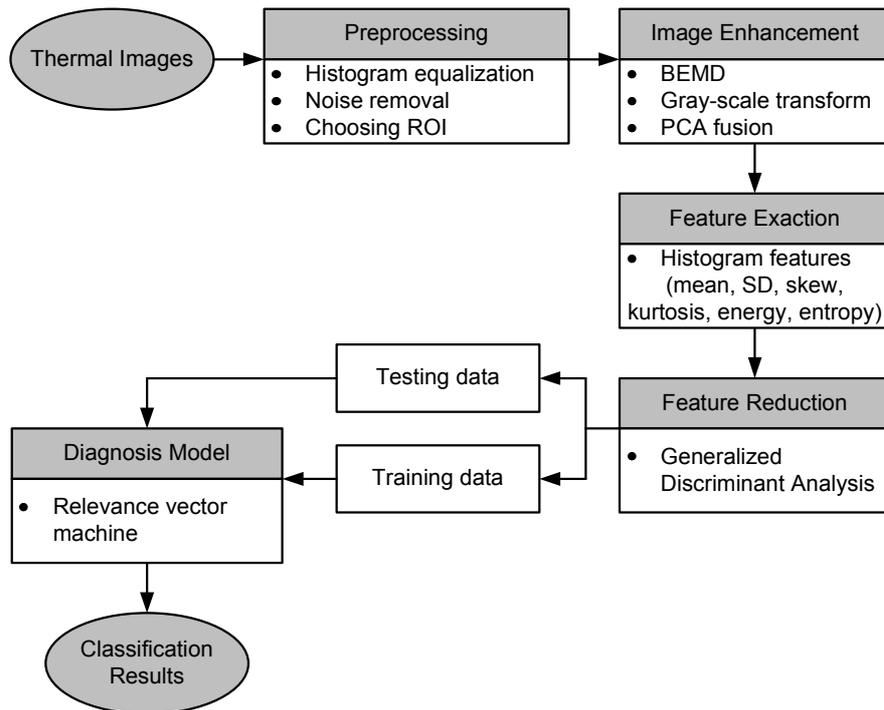


Fig. 1 Proposed system for thermal image based fault diagnosis



Fig. 2 Experimental setup

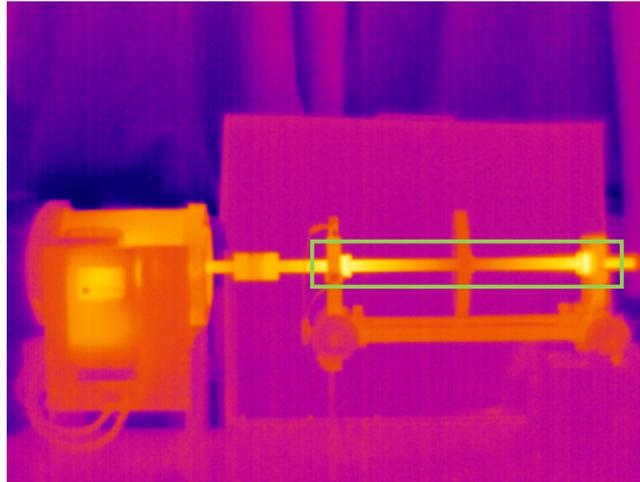


Fig. 3 Original thermal image and ROI

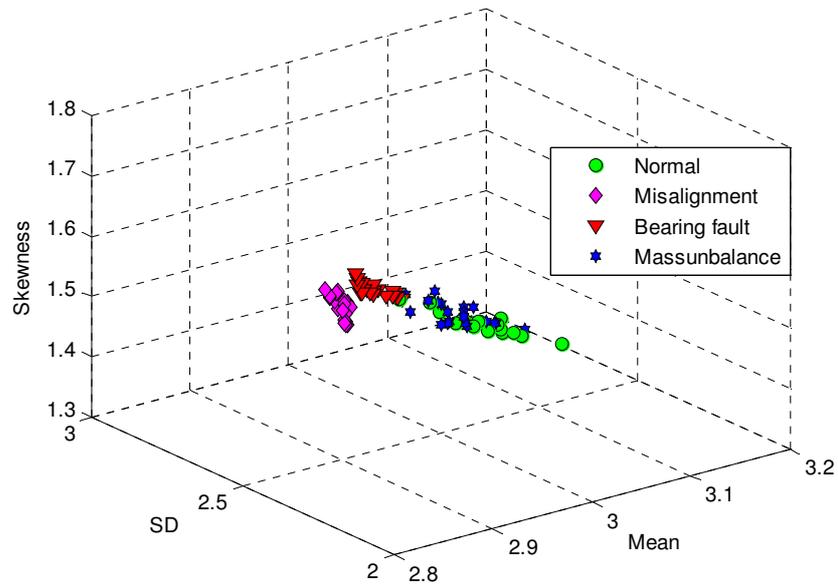


Fig. 4 Original features

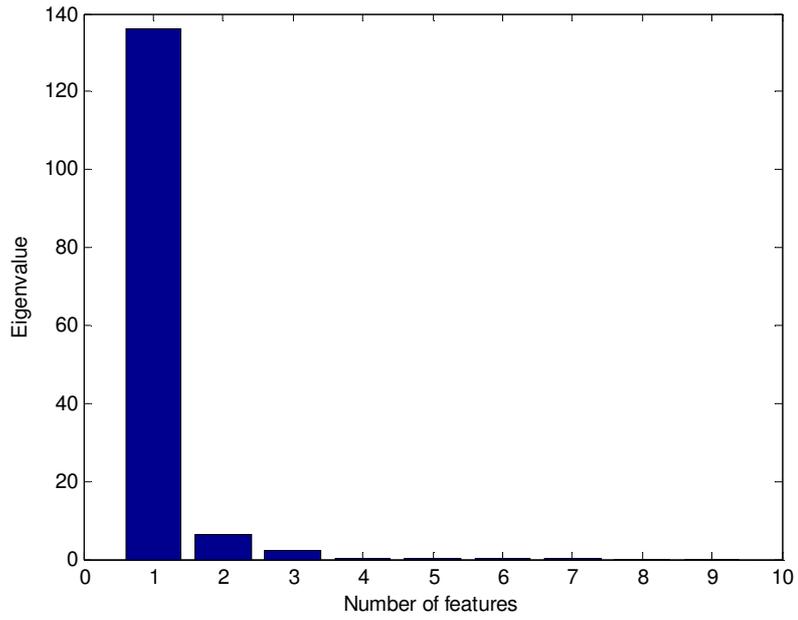


Fig. 5 Eigenvalue of covariance matrix for feature reduction

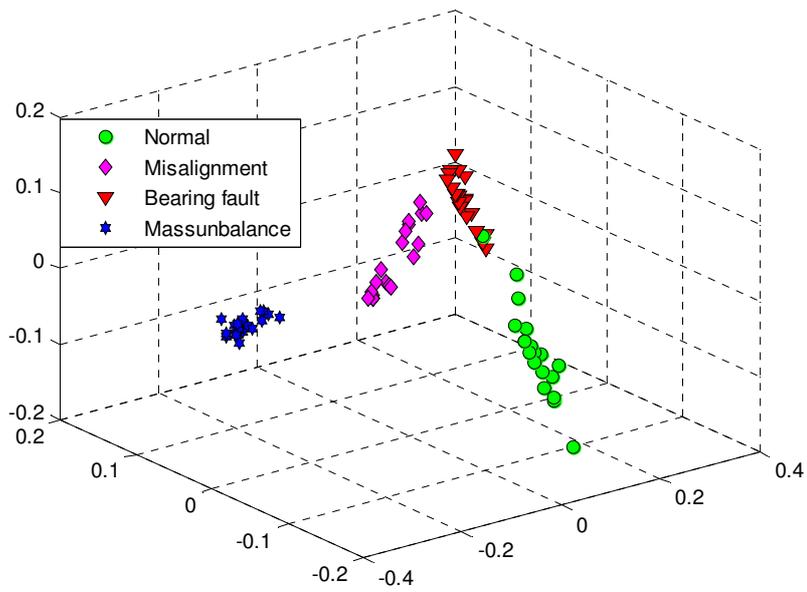


Fig. 6 Original features obtained from GDA

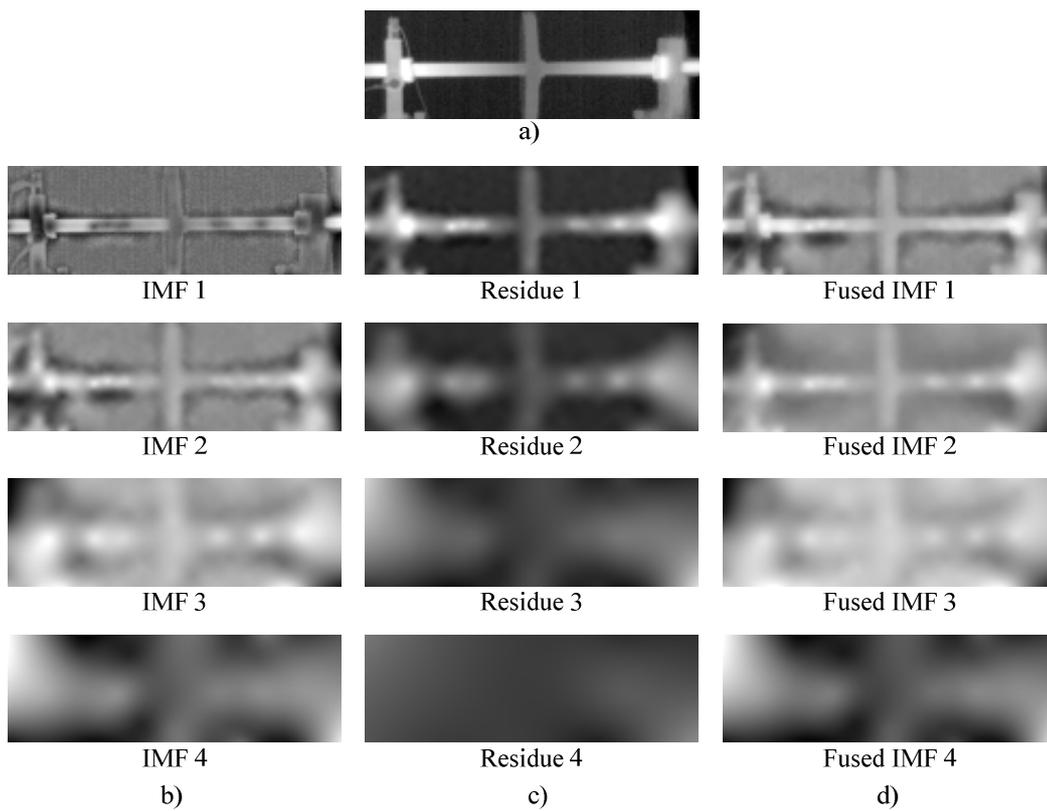


Fig. 7 a) Original image, b) IMFs, c) Residues, d) Fused IMFs

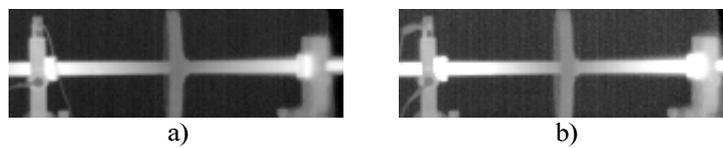


Fig. 8 a) Original image, b) Enhanced image

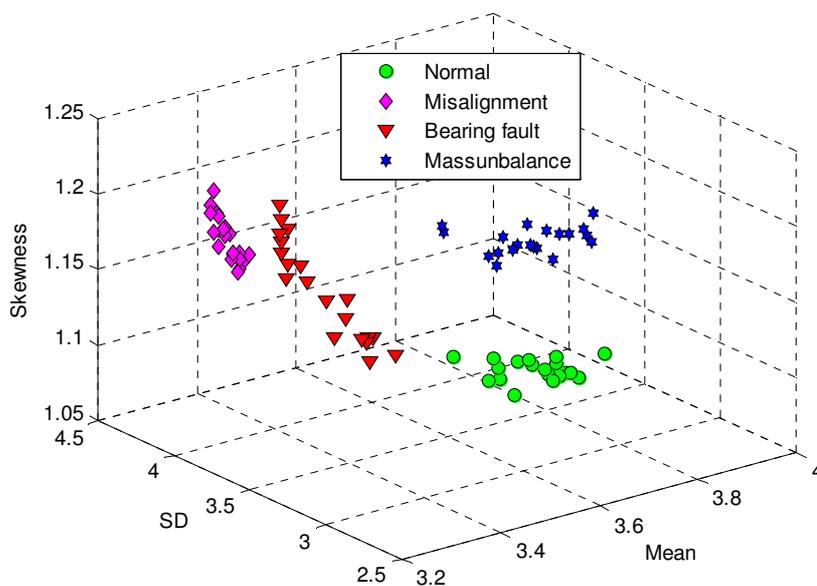


Fig. 9 Enhanced image features

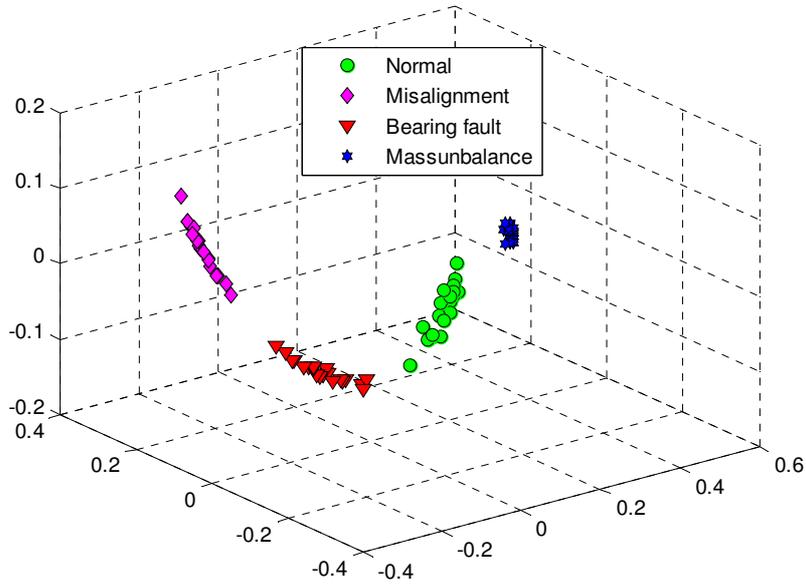
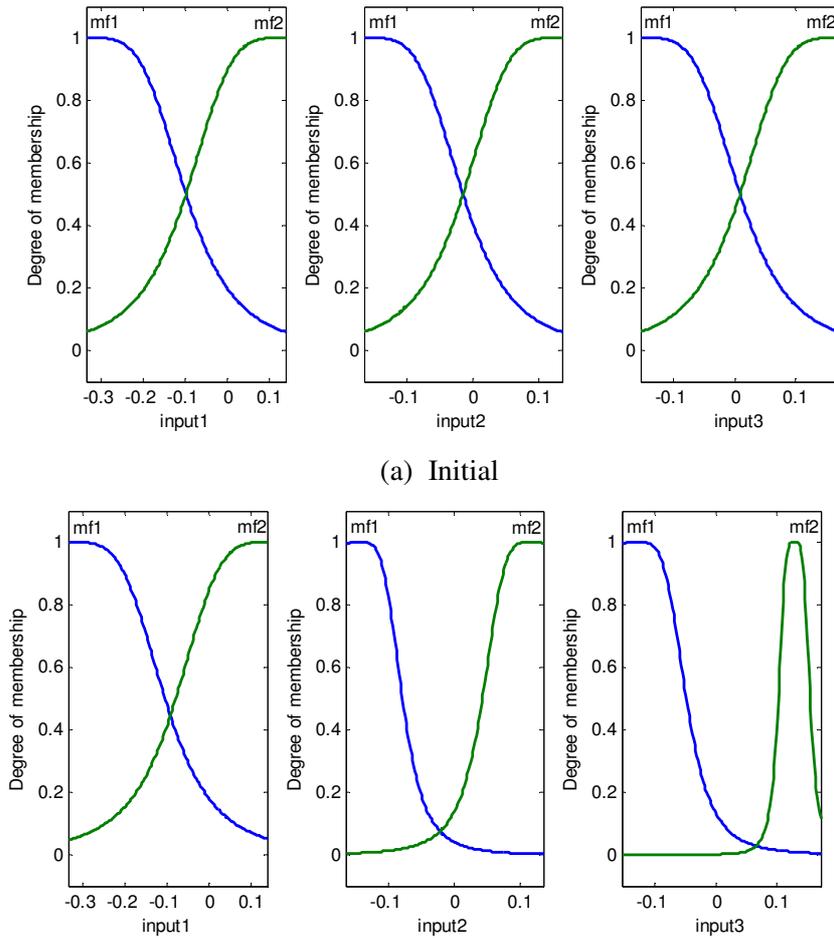
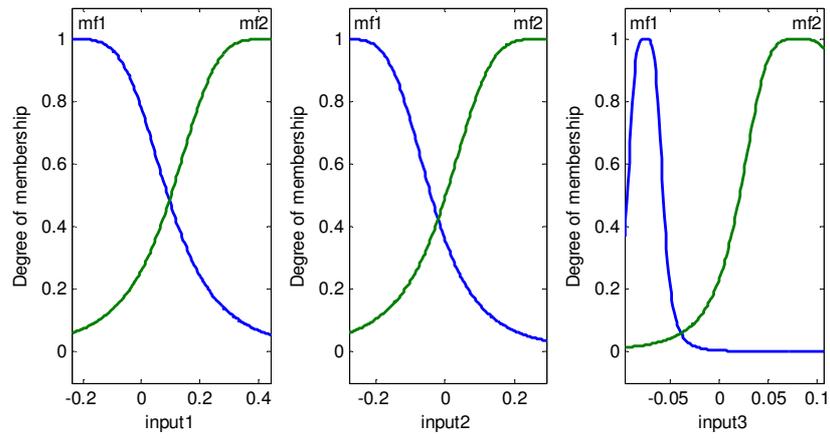


Fig. 10 Enhanced image features obtained from GDA



(b) Final membership functions of original images



(c) Final membership functions of enhanced images

Fig. 11 Membership functions

Table 1 Main specification of thermal camera and fault simulator

Devices	Specification
Thermal camera (FLIR-A 40 series)	<ul style="list-style-type: none">• Detector type: focal plane array uncooled microbolometer• Spectral range: 7.5 to 13 μm• Storage temperature range: $-40\text{ }^{\circ}\text{C}$ to $+70\text{ }^{\circ}\text{C}$• 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: $0.08\text{ }^{\circ}\text{C}$ at $30\text{ }^{\circ}\text{C}$• Accuracy: $\pm 2^{\circ}\text{C}$• Encapsulation: IP 40 (Determined by connector type)
Fault simulator	<ul style="list-style-type: none">• 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, 6" diameter with two rows of tapped holes at every 20° (with lip for introducing unbalance force)

Table 2 Detailed description of image data

Label of Classes	Machine Condition	No. of Samples	No. of Training Samples	No. of Testing Samples
C1	Normal	20	10	10
C2	Misalignment	20	10	10
C3	Bearing fault	20	10	10
C4	Mass unbalance	20	10	10
Total samples		80	40	40

Table 3 Results of classification

Features	Process	Classification Accuracy (%)			
		SVM-OAO	SVM-OAA	ANFIS	RVM
Original	Training	100	100	100	100
	Testing	97.5	95	82.5	97.5
Enhanced	Training	100	100	100	100
	Testing	100	100	97.5	100

Table 4 Time-consumption in the training process of RVM and SVM-OAO

Features	RVM	SVM-OAO
Original	2.15s	6.6s
Enhanced	2.0s	5.15s