IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 22, 2018, accepted January 12, 2019, date of publication January 17, 2019, date of current version February 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2893331

A Rotating Machinery Fault Diagnosis Method Based on Feature Learning of Thermal Images

ZHEN JIA[1], **ZHENBAO LIU**[1], **(Senior Member, IEEE),**
CHI-MAN VONG[2], **(Senior Member, IEEE), AND**
MICHAEL PECHT[3,4], **(Fellow, IEEE)**
[1]School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China
[2]Department of Computer and Information Science, University of Macau, Macau 999078, China
[3]Center for Advanced Life Cycle Engineering, University of Maryland, College Park, MD 20742 USA
[4]Prognostics and Health Management Center, City University of Hong Kong, Hong Kong

Corresponding author: Zhenbao Liu (liuzhenbao@nwpu.edu.cn)

This work was supported in part by the Natural Science Foundation of China under Grant 61672430, in part by the Shaanxi Key Research and Development Program under Grant S2019-YF-ZDCXL-ZDLGY-0227, in part by the Aeronautical Science Fund under Grant BK1829-02-3009, and in part by the NWPU Basic Research Fund under Grant 3102018jcc001.

ABSTRACT The rotating machinery plays a vital role in industrial systems, in which unexpected mechanical faults during operation can lead to severe consequences. For fault prevention, many fault diagnostic methods based on vibration signals are available in the literature. However, the vibration signals are obtained by using different types of sensors, which can cause sensor installation issues and damage the rotating machinery. In addition, this kind of data acquisition through vibration signal induces a large amount of signal noise during machine operation, which will challenge the later fault diagnosis. A recent fault detection method based on infrared thermography (IRT) for rotating machinery avoids these issues. However, the corresponding literature is limited by the fact that the characteristics of the manual design cannot characterize the fault completely so that the diagnostic accuracy cannot exceed the diagnostic method based on the vibration signals. This paper introduces a popular image feature extraction method into the fault diagnosis of rotating machinery based on IRT for the first time. First, capturing the IRT images of the rotating machinery in different states, and then two popular feature extraction methods for IRT images, bag-of-visual-word, and convolutional neural network, are tested in turn. Finally, the extracted features are classified to implement the automatic fault diagnosis. The developed method is applied to analyze the experimental IRT images collected from bearings, and the results demonstrate that the developed method is more effective than the traditional methods based on vibration signals.

INDEX TERMS Fault diagnosis, infrared thermography, convolutional neural network, bag-of-visual-words, feature recognition.

I. INTRODUCTION

Modern industrial applications including automobiles and generators, use rotating machinery, whose failure can cause different levels of damage [1]. In most cases, these failures refer to the discrepancy or residual of a mechanical component that goes above a certain threshold. For example, in rotor systems, the most common malfunction is unbalance whose primary symptom is abnormal vibration. Unbalance leads to fatigue of machine components. Worst of all, the wear on the bearing will further damage the seal, resulting in a decline in the performance of the machine [2].

Material fatigue is the major cause of rotating machinery failure. Other causalities, such as improper operation and installation, abnormally heavy load, and insufficient lubricant, can lead to failure or even damage of rotating machinery. Diagnosis and repair of rotating machinery is time-consuming, and its sudden suspension of operation or damage usually incurs a huge economic loss [3]. Due to its importance, numerous prognostics and health management (PHM) techniques have been developed for rotating machinery [4], such as acoustic emission [5], [6], vibration analysis [7]–[9], and eddy-current [10], [11].

The PHM technology for mechanical fault diagnosis in the existing literature is mostly based on the vibration signal of rotating machinery [12]–[15]. A significant disadvantage of vibration signal-based diagnosis is the noise issue. It is well known that a large amount of noise is induced within the captured vibration signals due to many environmental factors such as temperature and electromagnetic inference, during machine operation. Although many methods [16], [17] have been proposed to denoise the vibration signals, these methods are destructive to the fault information implied in the vibration signal, and hence can result in degradation of the fault diagnosis performance. Noise in images is represented by isolated pixels or blocks that cause strong visual effects. Nowadays, noise on IRT images used to detect the health status of equipment is very small, so it is not necessary to consider the issue of denoising. In order to solve the problem of denoising, infrared thermography (IRT) has been applied to the fault diagnosis of rotating machinery [18], [19]. IRT has been extensively used in areas such as maintenance of electrical installations. Its use in electrical machines has been mainly circumscribed to the detection of faults in static machines, such as power transformers [20]. The principle of IRT detection is to analyze whether the electrical equipment has a fault and determine the fault location by detecting the IRT image photographed by infrared camera during operation. By now, IRT has become a matured and widely accepted condition monitoring tool that can effectively judge the temperature rise and operating state of the equipment, where the temperature is measured in real time in a non-contact manner.

Applying IRT to the fault diagnosis of rotating machinery can not only avoid the noise problem in the traditional vibration signal-based fault diagnosis methods as mentioned above, IRT images have many other significant advantages such as non-contact, easy setup, non-invasiveness, high sensitivity and resolution. Therefore, IRT has gradually appeared in the field of mechanical fault diagnosis in recent years [21]–[25]. In [22], IRT is first used for classification of different machine conditions, where two-dimensional discrete wavelet transform is used to decompose the thermal image, and then Mahalanobis distance and relief algorithm are employed for feature selection, and finally support vector machine (SVM) is used as classifiers. Similar to the diagnostic procedure in [22], the fault detection method based on IRT includes three stages, i) thermal images are acquired; ii) features are extracted; and iii) feature classification is conducted to realize fault diagnosis. In addition, [23] and [24] using thermal imaging-based methods to diagnose rotating machinery failures is also in accordance with this diagnostic process.

The above studies all adopt the traditional pattern recognition method. Their difference lies in the feature determination step. It can be said that the researchers manually designed the features to be extracted. However, the diagnostic accuracy of rotating machinery is highly dependent on feature selection, and improper selection of features cannot guarantee high diagnostic accuracy. Existing applications using traditional feature extraction methods (Mahalanobis distance, relief algorithms, etc.) have proven that IRT can achieve mechanical fault diagnosis, but in comparative experiments, the accuracy of the IRT-based diagnostic method does not exceed the accuracy based on vibration signals, which is limited by the fact that the characteristics of the manual design cannot fully characterize the fault. Different from the feature extraction methods in the literature, in this work, two popular image feature extraction methods, namely, bag-of-visual-word (BoVW) [26] and convolutional neural network (CNN) [27] are tested. Firstly, the IRT image features are extracted by BoVW or CNN, and then the extracted features are classified by SVM for fault diagnosis. Experiments show that the developed method can realize high-precision diagnosis of rotating machinery by infrared diagnostic technology, and the diagnostic accuracy is higher than the traditional method based on vibration signals.

Contributions: Our approach is made possible by the following technical contributions:

1) Unlike the existing fault diagnosis methods for rotating machinery, which are almost based on vibration signals, this paper presents a fault diagnosis method based on IRT images. IRT images have many significant advantages such as non-contact, easy setup, non-invasive, high sensitivity and resolution. In addition, image-based fault diagnosis avoids the important problem of denoising the vibration signal.

2) An image-based fault diagnosis scheme for rotating machinery is developed, which can automatically detect mechanical faults with high accuracy. In this study, an experimental platform is established to obtain IRT images of different fault states, which verifies the effectiveness of this method.

Section II describes the procedures and techniques of the two feature extraction methods used in this paper and briefly introduces the steps of fault detection and isolation. Section III first introduces the experimental platform, and then presents two sets of comparative experiments, and gives performance analysis of the two comparative experiments. Finally, conclusions are drawn in Section V.

II. IRT IMAGE-BASED FAULT DIAGNOSIS

Image-based diagnosis is a kind of data-driven diagnostic method where the data is a set of IRT images other than vibration signals, and the IRT images are taken by infrared camera. Inspired by the mature image recognition techniques, two popular feature extraction methods are employed in this work: BoVW [28] and CNN, which is a kind of deep learning method [29].

Image-based diagnostics have several advantages: i) image identification refers to processing image data and then classifying images. The thermal image-based rotating machinery fault diagnosis method used in this paper also achieves the final fault diagnosis target by directly processing the thermal images. Therefore, no expert knowledge on the application is required, such as material properties, structures, or failure mechanisms, because the knowledge can be learned directly

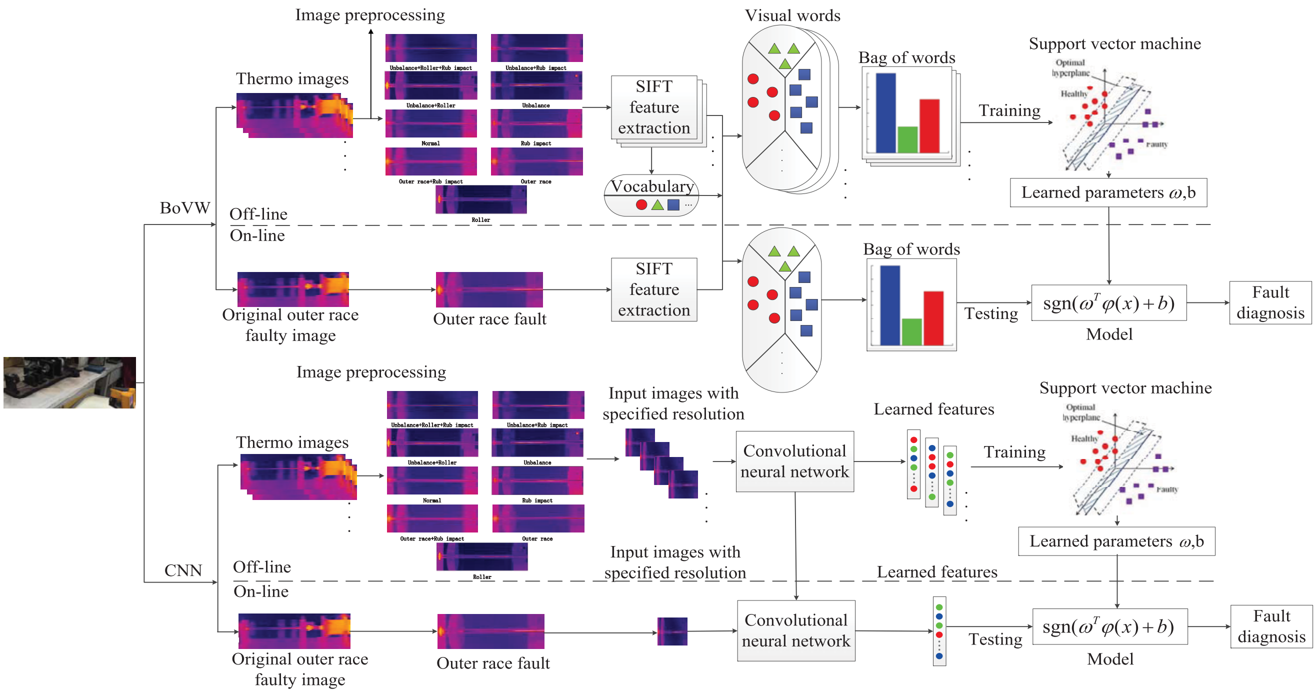


FIGURE 1. Flowchart of the developed IRT image-based fault diagnosis.

from the images; ii) IRT images do not possess troublesome noise induced by environmental factors and machine operation as in vibration signals; iii) IRT images are captured in a non-contact and non-intrusive way that avoids rubbing damage to the machinery or malfunctioning of the contacting sensors. Based on this idea, a framework of image-based fault diagnosis is proposed as shown in Fig. 1. The two diagnostic methods are composed of two parts, namely, offline and online [30]. Among them, the offline part corresponds to the training process, and the online part corresponds to the testing process. The IRT images taken of the entire test rig contains faulty components and other devices (motors, photoelectric counters, etc.). The difference in IRT images under different fault conditions is mainly reflected in the faulty component part. Therefore, in order to make the image processing process more efficient, the method developed in this paper includes a preprocessing step of cutting off the redundant part of the complete IRT image of the test rig as shown in Fig. 1, leaving only the range of interest, that is, the part containing the faulty component. The rules between features and faults are learned after the training process, while the testing process is used to observe the effect of fault diagnosis methods. In this figure, the process of fault diagnosis and isolation is introduced through the example of the outer race fault.

Both BoVW and CNN have been conducted and compared in the experiments. For BoVW, the diagnosis framework is conducted as follows: i) capturing and preprocessing the IRT images of the region of interest, followed by scale-invariant feature transform (SIFT) algorithm [31] to extract visual features from the IRT images; ii) from the extracted visual features, a visual vocabulary is generated by which the BoVW of each image is computed for fault diagnosis; and iii) the extracted BoVW features are used as input of the classification algorithms such as SVM. For CNN, the framework is conducted as follows: i) capturing and preprocessing the IRT images of the region of interest, followed by resizing the image resolution to fit input of the network; ii) constructing the network structure and learning the CNN feature of each image for fault classification; and iii) the extracted CNN features are used for input to the SVM classifier. The output of the SVM classifier indicates the fault state of the rotating machinery including which component is faulty. The following sections describe these steps in details.

A. BAG-OF-VISUAL-WORD

BoW (bag of words) gains great success in text retrieval, leading to its extension for image processing called BoVW. Similar to BoW, BoVW represents an image as a histogram (i.e., frequencies of occurrences) of visual features (i.e., words) in the image, which is an unordered set of non-distinctive discrete words with certain level of invariance to the spatial location of objects in the image. This histogram of visual words can be regarded as the effective features for image classification.

1) SIFT FEATURE EXTRACTION

Practically, fault diagnosis at the component level cannot be achieved by analyzing the whole IRT image but only a few ‘‘interested’’ regions in the IRT image. The reason is obvious that most IRT images are very similar under different fault states. Therefore, the analysis is focused on the characteristics of the "interest points" of these few regions, such as corners and blobs.

In [32], Lindeberg has shown that stable locations of the interest points in scale space can be efficiently detected through scale-space extrema in the difference-of-Gaussian function [33]. By convolving the scale-space extrema with the image $D(p, \sigma)$, where $p = (x, y)$ is a point in the image, the difference of two nearby scales separated by a constant multiplicative factor κ can be computed:

$$\begin{aligned} D(p, \sigma) &= (G(p, \kappa\sigma) - G(p, \sigma)) * I(p) \\ &= L(p, \kappa\sigma) - L(p, \kappa\sigma) \end{aligned} \quad (1)$$

where $*$ is the convolution operation in p. The scale space of an image is defined as follows,

$$L(p, \sigma) = G(p, \sigma) * I(p) \quad (2)$$

which is the convolution of a variable-scale Gaussian, $G(p, \sigma)$ with an input image, $I(p)$.

$$G(p, \sigma) = \frac{1}{2\pi\sigma^2} exp{-}(x^2 + y^2)/2\sigma^2 \quad (3)$$

The local extrema of the function D can be detected through accurate localization of interest points. Following the method in the literature [34], the Taylor expansion of the scale-space function $D(p, \sigma)$ is computed up to quadratic term only:

$$D(p, \sigma) = D + \frac{\partial D^T}{\partial p} p + \frac{1}{2} p^T \frac{\partial^2 D}{\partial p^2} p \quad (4)$$

By taking the partial derative of Eq. (4) with respect to p and setting to zero, the location of the extremum $\hat{p}$ is found. Subsequently, these local extrema, i.e., interest points, constitute an abstraction of the image. At each of these local extrema, a consistent orientation is assigned so that the local information of the image (represented as a vector of gradients) can be obtained as the SIFT descriptor. This process has been implemented in the Open Source Computer Vision Library (OpenCV) environment. Default parameters are employed in the visual feature extraction while the feature vectors are 128-dimensional. After this step, each image is a collection of vectors of the same dimension (128 for SIFT), where the order of different vectors is unimportant. Since the IRT images are acquired under slightly different (up/down/left/right) angles (± 5 degrees) and the image brightness taken from different angles is varying, the SIFT descriptor must be robust to noise and illumination changes of images, which is an expectant property for some compensations on slightly different angles when we acquire the IRT images. Experimentally, the number of SIFT features is adjusted from 20 to 80 due to the different fault states. Finally, the presentation of a fault state (of 100 images) is constituted from about 2000 to 8000 SIFT features.

2) BoVW FEATURE GENERATION

Before feature encoding, a representative of several similar interest points called vocabulary is produced using k-means clustering [35]. First, a set of k clusters is learned by clustering the features extracted from each fault state into a specified vocabulary size. Then, the centers of the learned clusters are defined as the vocabulary. In current application, a set of 128-dimensional vectors of SIFT features $(x_1, x_2, \ldots, x_n)$ is given. Through k-means clustering, the n SIFT vectors are partitioned into k different sets such that $S = \{S_1, S_2, \ldots, S_k\}$ with minimized intra-cluster squared summed error (SSE). Equivalently, it aims to find

$$S = \underset{S}{\text{argmin}} \sum_{i=1}^{k} \sum_{x \in S_i} ||x - \mu_i||^2 \quad (5)$$

where μ_i is the mean vector in S_i. In the experiments, the vocabulary size k is set to 1000 for better discrimination capability. With this constructed vocabulary S, the extracted SIFT features are then quantized as the label of the closest cluster [36]. Finally, an image is abstracted as the frequency counts or histogram of the quantized SIFT features $[f_1, f_2, ..f_i, \ldots, f_k]$ where f_i is the frequency of i^{th} visual word in the image. This encoding from SIFT features to visual words according to the idea of nearest neighbor in the vocabulary can generate an effective BoVW histogram with spatial invariance.

B. CONVOLUTIONAL NEURAL NETWORK

CNN is a deep learning method that has achieved great success in image classification. The application of CNN in fault diagnosis tasks is essentially an image classification problem. CNN provides a powerful framework to learn hierarchical features of images as a feature extractor based on deep learning. Furthermore, an advantage of this network is that it is highly variable for translation, scaling, tilting or other deformations, thanks to the fact that the convolution operation is close to the mechanism of the human eye capturing features. In this paper, CNN is employed to learn IRT image features, and Fig. 2 shows its structure and working principle by taking the network of this paper as an example.

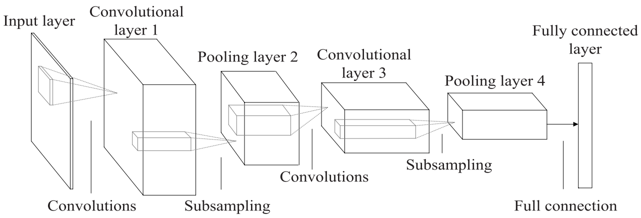


FIGURE 2. The principle of feature learning for the CNN model.

1) CONVOLUTIONAL NEURAL NETWORK

The CNN model employs multiple layers, including convolutional layer, pooling layer, and connection layer, to process the input image. CNN is trained with multiple stages, and the input and output sets of arrays of each stage are called feature maps. In this study, the IRT images were used as input to the network. The output feature map of the current stage is considered the input feature map for the next stage. Each stage consists of three steps: convolutional operation, nonlinearity transformation, and feature pooling. Several such three-part stages form the complete structure of the CNN model, followed by SVM to perform the classification. The details of CNN are described as follows.

The L-layer CNN model can be described as a series of linear transformations, nonlinear symmetric squashing operations (such as *sigmoid* function or *tanh* function) and pooling/subsampling operations. The network treats the input image as three-dimensional arrays, including the number of feature maps, the height and width of the maps. For layer l, F_l is defined as the output of the l-th stage and is given by:

$$F_l = pool(\tanh(W_l * F_{l-1} + b_l)) \tag{6}$$

where $l \in 1, \ldots, L$, W_l is the convolutional kernel, b_l is the bias parameter of the l-th layer. The input image is the initial feature map, i.e., $F_0 = I$. The whole network is made up of all layers stacked one after another.

In the CNN model, the *tanh* function is the point-wise hyperbolic tangent function that is applied to the output of every convolutional and pooling layer. The pool operation is a function considering a neighborhood of activations and generating one activation in each neighborhood. The fully connected layer is the last layer of the network. This layer can be regarded as a feature descriptor, which is the final feature representation of the input image by the CNN network. Its output is fed to the SVM classifier.

As depicted in Fig. 2, the CNN model constructed in this paper contains five layers with weights; the first and third ones are convolutional, the second and fourth ones are pooling, and the last layer is fully connected. All layers are connected to one another in turn. The first convolutional layer filters the 200 × 200 × 3 input image with 40 kernels of size 12 × 12 × 3 with a stride of 4 pixels. The second pooling layer uses a 40 kernels of size 2 × 2 × 40 with a stride of 2 pixels. The third convolutional layer takes as input the output of the second pooling layer and filters it with 60 kernels of size 6 × 6 × 40. The fourth pooling layer uses 60 kernels of size 2 × 2 × 60 with a stride of 2 pixels. The fifth fully connected layers is connected to the fourth pooling layers. Then the output of the fifth layer is used as the input to the SVM classifier.

C. FAULT DETECTION AND ISOLATION

In this paper, the problem of fault detection and isolation is considered as a multi-class classification. Given a set of training data, features are extracted from different fault states for rotating machinery. Under these extracted features with the corresponding fault states, a fault dictionary is produced containing different classes of fault states. Noteworthy, the no-fault state is considered as a class in the fault dictionary. SVM is employed to learn a classifier from the extracted features along with their corresponding fault states. For diagnosis, the extracted features of a new image are diagnosed using the trained SVM classifier. Note that a no-fault state is considered one of the classes in SVM.

There are many SVM classification models [37], e.g., c-support vector classification, ν-support vector classification, and distribution estimation. Each classification model contains a kernel function, i.e., linear, polynomial, RBF, or sigmoid kernels, which are evaluated and formulated as shown in Table 1. Here, γ, r, and d are kernel parameters. In the study, after testing the combination of different kinds of SVM classifiers and kernel functions, it is found that the combination of c-support vector classification and linear kernel achieves the best classification accuracy.

TABLE 1. Commonly used kernel functions.

Kernel	$K(x_i, x_j)$
Linear	$x_i^T x_j$
Polynomial	$(\gamma x_i^T x_j + r)^d, \gamma > 0$
Gaussian RBF	$\exp(-\gamma\|\|x_i - x_j\|\|^2), \gamma > 0$
Sigmoid	$\tanh(\gamma x_i^T x_j + r)$

III. EXPERIMENTAL SETUP

In order to test the developed image-based diagnosis on rotating machinery, the IRT images were chosen as the target signals for analysis, and two popular classes of feature extraction methods were employed in this work: BoVW, and CNN. Finally the SVM classification performed the task of fault detection and isolation.

A fault simulator (WS-ZHT1 type multifunctional rotor test rig) as shown in Fig. 3 was employed to provide different kinds of faults of driving motors, bearings, flywheels, shafts and rotors. A long shaft of 10 mm diameter was supported by two ball bearings. One end of the shaft was attached to the motor spindle through a flexible coupling. In this paper, 9 fault states of the bearing were collected (note that the normal state was also considered a fault state), including various

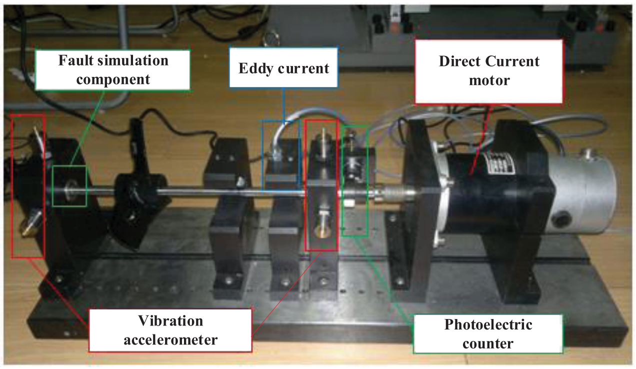


FIGURE 3. Experimental setup for demonstrating the developed approach.

single fault types and compound fault types. Compound faults refer to the simultaneous occurrence of multiple types of faults, that is, multiple faults are coupled together. In addition to roller fault state, outer race fault state and normal state, the unbalance state was simulated by attaching weights on a flywheel with threaded holes. As for the rub impact fault, a copper blade was fixed by the threaded holes on the stand in order to connect the copper blade and adjust the gap between the copper blade and the shaft. When the copper and the shaft come into contact with each other, the rub impact fault happens. The motor was driven by a variable speed direct current (DC) motor with speed up to 2000 rpm. The specifications of the fault simulator and the thermal camera (Fluke Ti32) are shown in Table 2.

TABLE 2. Specifications of IRT camera and fault simulator.

Fluke Ti32	• $8.0\mu m$ to $14\mu m$ (long-wave) infrared spectral band • Environmental temperature is 27°C • -20°C to +600°C temperature measurement range • Solid object materials and surface treatments exhibit emissivity ranging from approximately 0.1 to 0.95 • Focal-Plane Array, uncooled micro bolometer detector, 640×480 pixels • 0.05°C at 30°C thermal sensitivity
WS-ZHT1	• Shaft diameter: 10mm • Bearing: two ball bearings • Bearing housings: two aluminum brackets • Rotors: one rotor, 8cm diameter with two rows of tapped holes at every 20° • DC motor speed from 0 to 10000 rpm

The thermal camera is the key device for data acquisition whose parameters must be carefully set for accurate IRT images. The parameters of the thermal camera are automatically configured by the camera itself. Among all parameters, the most important one is emissivity. The emissivity of a object refers to the ratio of the radiated power of the object to the radiating power of a black body at the same temperature. It is related to the surface state of the object (including the surface temperature of the object, the surface roughness, and the presence of surface oxides, surface impurities, or coatings). The diagnostic object of this paper is the rolling bearing, and the material of the rolling bearing is steel. For metallic materials, surface roughness and surface temperature will have a greater impact on emissivity. In general, the greater the surface roughness of the metal, the greater the emissivity of the metal. The emissivity of the metal increases as the surface temperature increases. As shown in Table 2, the thermal imager used in the experiments herein is suitable for materials with an emissivity of 0.1 to 0.95. The other parameters such as relative humidity, scale temperature, focal length of camera, and distance are configured as in Table 2. All of these parameters were chosen according to experiment condition. In this study, all fault states (such as normal, unbalance, roller, outer race, rub impact, unbalance and rub impact, unbalance and roller, outer race and rub impact, unbalance and roller and rub impact) are measured under the same parameters setting of the Fluke Ti32 for the experiment. Data from the Fluke Ti32 were saved directly to the computer. In order to verify the robustness of our developed image-based diagnosis, the images were acquired under slightly different (up/down/left/right) angles (±5 degrees).

For the machine configuration, the speed of the motor was gradually raised up to 2000 rpm for data acquisition. We collected two sets of data (Group 1 and Group 2), each of which contains 100 images. Group 1 is the set of 100 images captured during the transient state of the machine (from the beginning to 5 minutes of machine operation). The speed of 2000 rpm was held for 15 minutes until the machine reached its steady state. Then, another set of 100 images was captured as Group 2 (from 15 minutes to the end of the experiment). It is well known that the images captured under transient state may significantly vary even though they belong to the same fault. On the other hand, such variation is significantly reduced for those images captured under an operational state.

Fig. 4 shows the IRT image of the machine state, with a black rectangle encircling the interesting part of the experiment, the bearing. The black rectangle is a predefined bounding box whose resolution is 480 × 200 in order to realize the automatic cutting of an IRT image (whose resolution is 640 × 480). The final experimental IRT images consist of 9 machine conditions as listed in Table 3. Actually, segmentation is performed to the portion (specified by the black rectangle) of the original IRT image in order to exclude non-defective parts, such as driving motor and some other parts, and reduce the scope of image processing. In this way, it will be beneficial to the next step of feature extraction by deleting non-defective portions, because the extracted features about the black rectangle contain more fault information and are more representative than the extracted features of the whole image. After segmentation, the IRT images of all faults under transient and steady states are presented in Figs. 5 and 6.

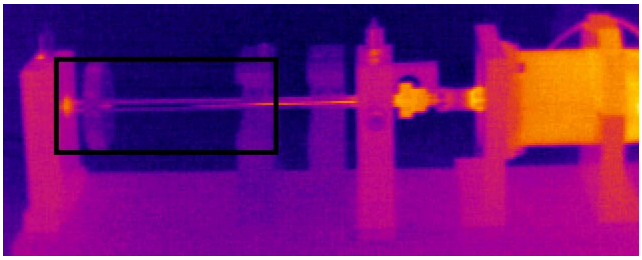


FIGURE 4. An IRT image and the region of interest.

In order to verify the effectiveness of the two feature extraction methods (i.e., BoVW and CNN), we intentionally captured the IRT images under both transient state and steady state. There are 9 fault states corresponding to labels F1 to F9. Each fault state has been captured for 100 samples in the experiment. Each sample is a collected IRT image with 480 × 200 pixels. For every fault state, the 100 captured IRT images are halved for training and testing, respectively.

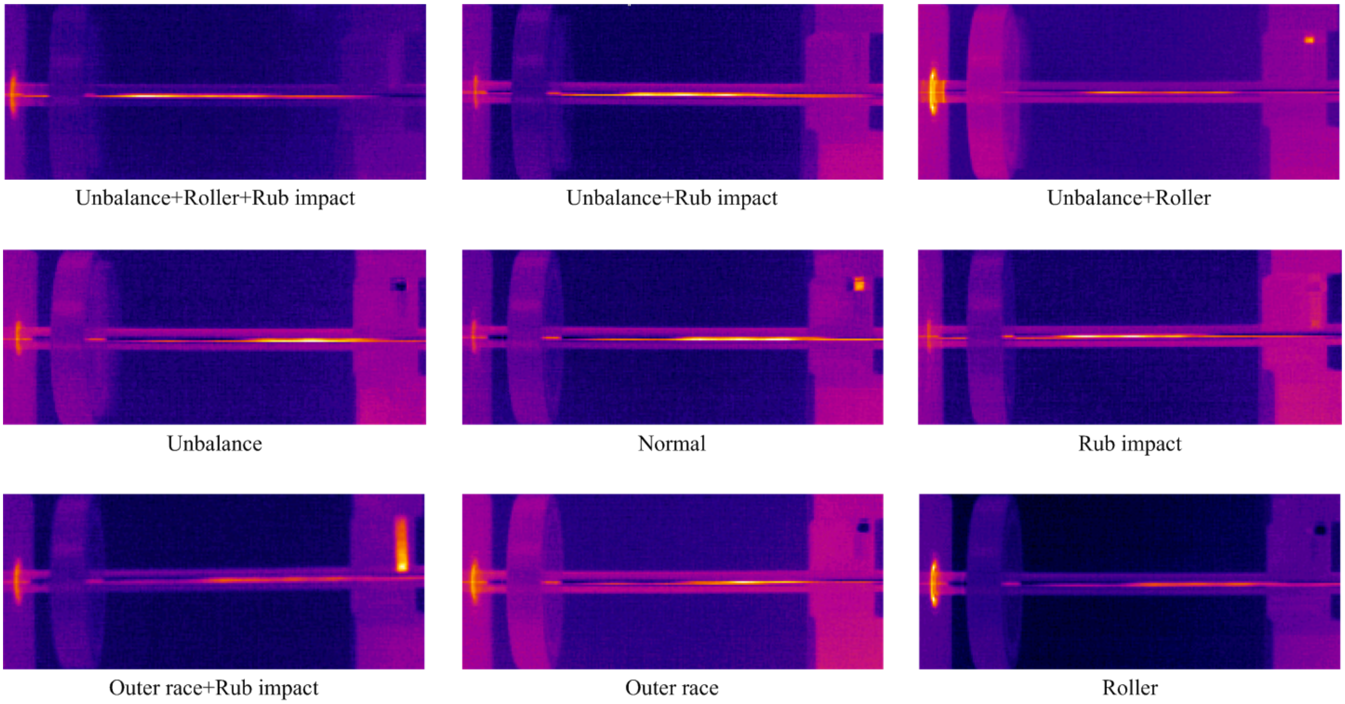

Unbalance+Roller+Rub impact
Unbalance+Rub impact
Unbalance+Roller
Unbalance
Normal
Rub impact
Outer race+Rub impact
Outer race
Roller

FIGURE 5. IRT images under transient state.

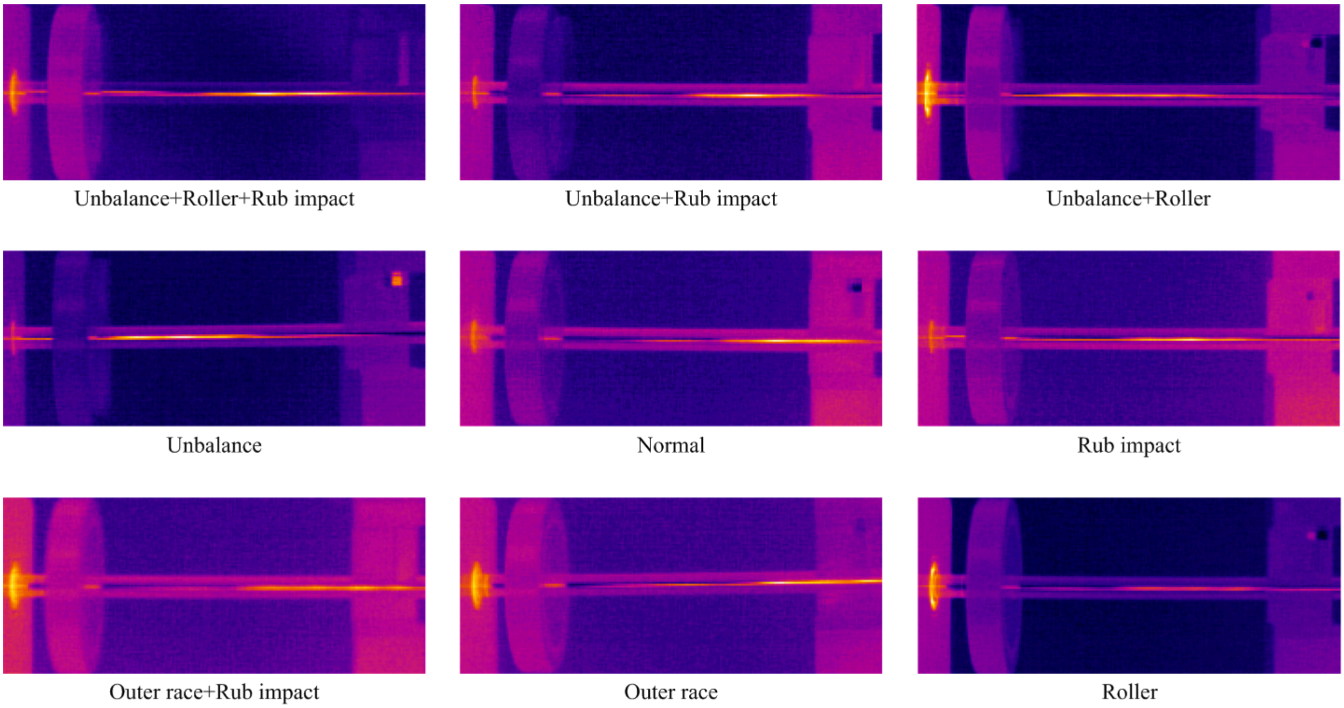

Unbalance+Roller+Rub impact
Unbalance+Rub impact
Unbalance+Roller
Unbalance
Normal
Rub impact
Outer race+Rub impact
Outer race
Roller

FIGURE 6. IRT images under steady state.

This ensures a relatively adequate training set and a fully tested diagnostic algorithm. Therefore, under both transient and steady states, there are 100 no-fault and 800 faulty cases for the purposes of training and testing, respectively. The setup details of the experimental data are shown in Table 3. Therefore, the total number of training samples and test samples is both 450 (each sample is a 480×200 pixel image). It is very hard to effectively distinguish the 9 operation conditions from the collected IRT images according to Figs. 5 and 6. After SVM training, the trained classifier was applied to the

TABLE 3. Details of captured IRT images under transient state.

Machine condition	Size of total IRT images	Training IRT images	Testing IRT images	Label
Unbalance+Roller +Rub impact	100	50	50	F1
Unbalance+Roller	100	50	50	F2
Unbalance	100	50	50	F3
Normal	100	50	50	F4
Rub impact	100	50	50	F5
Unbalance+Rub impact	100	50	50	F6
Outer race+Rub impact	100	50	50	F7
Outer race	100	50	50	F8
Roller	100	50	50	F9

testing cases to evaluate the discriminability of the extracted features. The procedure to handle IRT images under steady state was similarly conducted.

IV. RESULTS AND ANALYSIS

In order to assess the performance of the developed IRT image-based fault diagnosis method, two types of analyses were carried out in the fault feature extraction step. In the first analysis, the classification results are compared based on features, respectively, of the two popular feature learning methods, BoVW and CNN [38]. In the second analysis, the results of fault diagnosis based on IRT maps and traditionally used vibration signals are compared. Finally, the diagnostic performance was verified under different metrics to evaluate the performance of the above several fault diagnosis methods.

A. BOVW VERSUS CNN

Two different feature learning methods based on BoVW and CNN were used to produce diagnostic decisions. The classification accuracy of the SVM classifier was evaluated under features learned by BoVW and CNN, respectively. The average classification accuracy was considered as the evaluation measure throughout the first experiment. The different fault condition settings are shown in Table 3, which contains a total of 9 types of faults. The IRT images were acquired under these fault conditions, and then divided into training sets and test sets, and fault features were collected for classification. Under these fault conditions, IRT images of the test elements were collected and then divided into training and testing sets, on which fault features were captured for classification. Subsequently, the SVM was selected for classification.

During the training phase, the SVM classifier was trained by features and corresponding fault labels. The training and testing datasets consisted of faulty IRT images of the bearings captured under different fault conditions. Then, through the fault feature extraction method, BoVW and CNN, the training and testing features were captured from training and testing datasets, respectively. After training the classifiers, the fault labels for testing features were predicted. The classification results for fault class diagnosis produced by SVM classifier are shown in Tables 4, 5, and 7, respectively.

TABLE 4. Fault diagnostic results based on IRT images with BoVW features under transient state.

	F1	F2	F3	F4	F5	F6	F7	F8	F9
F1	50	-	-	-	-	-	-	-	-
F2	-	50	-	-	-	-	-	-	-
F3	-	-	49	-	-	1	-	-	-
F4	-	-	-	50	-	-	-	-	-
F5	-	-	-	-	50	-	-	-	-
F6	-	-	1	-	-	49	-	-	-
F7	-	-	-	-	-	-	48	-	2
F8	-	-	-	-	-	-	-	49	1
F9	-	-	-	-	-	-	-	-	50

TABLE 5. Fault diagnostic results based on IRT images with BoVW features under steady state.

	F1	F2	F3	F4	F5	F6	F7	F8	F9
F1	50	-	-	-	-	-	-	-	-
F2	-	50	-	-	-	-	-	-	-
F3	-	-	49	1	-	-	-	-	-
F4	-	-	-	50	-	-	-	-	-
F5	-	-	-	-	50	-	-	-	-
F6	-	-	-	-	-	50	-	-	-
F7	-	-	-	-	-	-	50	-	-
F8	-	-	-	-	-	-	-	50	-
F9	-	-	-	-	-	-	-	-	50

The distributions of BoVW features under transient state and steady state are shown in Figs. 7 and 8, respectively (*Remark*: The X-axis and Y-axis in Figs. 7 and 8 are meaningless, which are merely scale representation. The 9 different colors represent the 9 different faults.). It is well known that if the intra-class distance is smaller and the inter-class distance is bigger, the classification result is better. The visualization of BoVW features under transient state in Fig. 7 is more disordered than the BoVW features under steady state in Fig. 8. Compared to Fig. 8, in Fig. 7, different faults are not completely distributed at different regions because some regions are overlapping. For example, the regions of classes F3 and F6 are overlapping (specified by the black rectangle), which is consistent with the diagnostic performance contrast under two states shown in Tables 4 and 5. After the feature extraction is completed, the SVM is used for classification to implement fault diagnosis, that is, fault type classification.

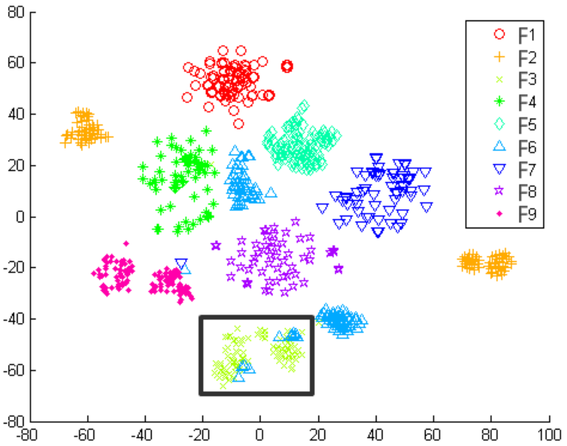


FIGURE 7. Visualization of BoVW features under transient state.

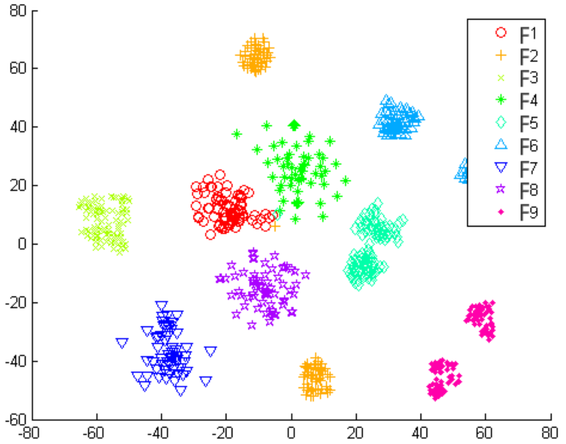


FIGURE 8. Visualization of BoVW features under steady state.

Tables 4 and 5 show the fault diagnosis for the bearing under transient state and steady state using the BoVW feature extraction method, respectively. As can be seen from Table 4, under the transient state, there are four types of faults (F3, F6, F7, and F8) that are incorrectly classified. However, from Table 5, it can be seen that only one misclassification case occurred in the steady state. It is obvious that the fault diagnosis effect under the steady state is better than that under the transient state. In addition, the BoVW feature extraction method based on IRT image has excellent overall diagnostic performance, indicating that BoVW can effectively capture highly discriminative semantic features. The four assessment types are shown in Table 6. It can be seen that the diagnostic performance of both the two states based on BoVW feature learning method achieved a high score, which implies that BoVW works effectively in capturing features for diagnosis of rotating machinery. Furthermore, the SVM classifier can identify and isolate faulty types with a nearly 100% accuracy in both states (see Precision in Table 6 for details), which can be concluded as fault isolation in addition to fault diagnosis.

TABLE 6. Metrics evaluation of fault diagnostic performance based on IRT images with BoVW features.

Performance Metrics	Transient State	Steady State
False positives	0	0
False negatives	0	0
Accuracy	98.9%	99.8%
Precision	100%	99.8%

In order to evaluate the diagnosability of the developed IRT image-based fault diagnosis method, four evaluation measures are employed: 1) false positives: the number of cases that the classifier indicated a fault in the bearing while there is no fault actually; 2) false negatives: the number of cases that the classifier declares a no-fault situation in the circuit while there is a fault actually; 3) accuracy: the proportion of cases whose label is correctly predicted in the total cases; 4) precision: the proportion of cases correctly predicted as faulty in all cases predicted to be faulty.

CNN is another popular feature extraction method for images based on deep learning, which can be an alternative tool to BoVW feature learning. For a fair comparison, CNN uses the same training and testing data as BoVW. The structure of the CNN determined in this experiment, as described in Sec. II B, consists of five layers, first two sets of convolutional and pooling layers, followed by a fully connected layer. As shown in Table 7, the use of CNN features based on IRT images achieves perfect fault diagnosis results in both transient and steady state, and 450 test samples are correctly classified into corresponding fault states, achieving 100% diagnostic accuracy. As feature extraction methods, the features extracted by the CNN and BoVW from the IRT images are classified by the same classifier, and the accuracy based on CNN is higher than that of BoVW. From the results shown in Tables 4, 5, and 7, it may be explained that in CNN,

TABLE 7. Fault diagnostic results based on IRT images with CNN features under both transient and steady state.

	F1	F2	F3	F4	F5	F6	F7	F8	F9
F1	50	-	-	-	-	-	-	-	-
F2	-	50	-	-	-	-	-	-	-
F3	-	-	50	-	-	-	-	-	-
F4	-	-	-	50	-	-	-	-	-
F5	-	-	-	-	50	-	-	-	-
F6	-	-	-	-	-	50	-	-	-
F7	-	-	-	-	-	-	50	-	-
F8	-	-	-	-	-	-	-	50	-
F9	-	-	-	-	-	-	-	-	50

the feature extraction process is performed layer by layer, and the ability to express features becomes stronger as the layer deepens, which enables the final output features to accurately reflect the characteristics of the input information, whereas BoVW first captures the ''interest points'', on which learning and classification are carried out.

B. IRT IMAGE-BASED METHOD VERSUS VIBRATION SIGNAL-BASED METHOD

A comparative experiment was also set up to compare the accuracy of fault diagnosis based on IRT images and traditional vibration signals. In this comparative experiment, the IRT image-based method used CNN for feature extraction (because the performance of CNN is overall better than BOVW), and the vibration signal-based method used shift invariant sparse coding (SISC) [9] for feature extraction. In this paper, an adaptive feature extraction scheme based on sparse coding is developed. The sparse representation of the signal can represent the diagnostic information in an efficient way. The bearing fault diagnosis is carried out in the vibration signal experiment, and the scheme has good diagnostic performance. The experimental results of better diagnostic accuracy prove that the proposed features can effectively represent fault information. Subsequently, the features extracted by the two methods are classified using SVM.

The same fault simulator (WS-ZHT1) was used again as the vibration signal generator and the speed of the motor was still set to 2000 rpm. An accelerometer was installed on the bearing of the drive end side. A 16-channel digital audio tape (DAT) recorder was used to collect vibration signals under a sampling frequency of 12 kHz when it reached steady state. Only three single fault types were diagnosed in [9], and the composite fault state was not considered. Therefore, in this paper, only the diagnostic accuracies of these three fault states are compared. In this comparative experiment, the diagnostic results of three common fault classes including outer race (F8), roller (F9) and normal (F4) were compared. Sparse feature extraction was also performed on the time domain vibration signal. The extracted feature was divided into two equal halves for training and testing. Experimental parameters were as follows, each vibration data set had 100 samples truncating into time-series with a 1024-point window block, and 10 atoms, each with a length of 128-point, which were learned from one state of the bearing data. The fault diagnostic results based on vibration signals under steady state are presented in Table 8. It can be seen from the classification accuracies comparison results of the two diagnostic methods given in Table 9 that the vibration features perform slightly worse diagnosis of the mechanical condition.

From the results, we can consider the reason why the image-based method is better than the signal-based method. On the one hand, the vibration signal is doped with noise from the operating environment of the machine, such as temperature and electromagnetic interference. On the other hand, the signal-based method needs to utilize accelerometers and eddy-current sensors. Because these sensors must be installed inside the machinery, the problem of improper sensor installation can affect the measurement accuracy. However, IRT images are non-contact, non-intrusive, easy to set up, and have high sensitivity and resolution, which avoids the above two problems existing in the vibration signal-based method. Therefore, the developed IRT image-based method using CNN is superior to the traditional vibration signal-based method.

TABLE 8. **Fault diagnostic results based on vibration signals under steady state.**

	F4	F8	F9
F4	50	-	-
F8	-	49	1
F9	-	2	48

TABLE 9. **Classification accuracies comparison based on vibration signals and IRT images.**

Fault State \ Signal	Vibration	Thermal image
Outer race	$49/50 = 98\%$	$50/50 = 100\%$
Roller	$48/50 = 96\%$	$50/50 = 100\%$
Normal	$50/50 = 100\%$	$50/50 = 100\%$

C. ANALYSIS OF THE IMPACT OF ENVIRONMENT ON EXPERIMENTAL PERFORMANCE

For a particular device, the ambient temperature at which it is located is generally stable. The IRT camera used in this paper has an operating temperature range of −10°C to +50°C and a measurement temperature range of −20°C to +600°C. Taking the CNN with outstanding performance in this paper as an example, the image feature extraction method based on CNN includes two parts, namely, the offline training part and the online test part. The high-precision fault diagnosis performance in this paper has proven that CNN has strong feature learning ability. Therefore, the diagnostic accuracy can be guaranteed as long as the IRT camera operates within the allowable operating temperature range of −10°C to +50°C. In addition, as shown in Table 2, the IRT camera used in this paper is suitable for materials with an emissivity of 0.1 to 0.95, which ensures that the emissivity is still within the applicable range when the ambient temperature varies or the lighting intensity changes.

V. CONCLUSION

There are many advantages to fault diagnosis based on infrared thermography (IRT) images, such as non-contact, ease-of-setup, and high sensitivity and resolution, which avoids expensive equipment such as accelerometers and eddy-current sensors those are regarded to be installed inside the machinery. Moreover, image-based fault diagnosis avoids the significant issue of denoising on the vibration signal. However, the existing IRT image-based methods for diagnosing rotational machinery faults involve two steps: feature extraction using signal processing techniques and fault classification using the shallow learning models (e.g., the most frequently used support vector machine (SVM) classifiers), which inevitably leads to the diagnostic performance that relies heavily on feature extraction. This is why the IRT image-based diagnostic performance cannot go beyond the traditional vibration signal-based diagnostic methods. In recent years, the successful application of bag-of-visual-word (BoVW) and convolutional neural network (CNN) in computer vision proved that these two models are more powerful in describing image features. Therefore, these two methods are introduced to analyze IRT images in this paper.

In this study, an IRT image-based diagnosis scheme was developed to classify key rotating machinery conditions. Instead of the existing IRT-based method for diagnosis using traditional pattern recognition, our approach used BoVW and CNN. The approach consists of IRT image data acquisition, feature extraction, and condition identification. The two pattern recognition methods, BoVW and CNN, were both used as the classifier to mine the features of the IRT images, and it was found that CNN can give better results than BoVW. By setting a variety of fault states for the rotating machinery, the traditional vibration signal-based diagnostic methods and the developed IRT image-based method using CNN are used for diagnostic comparison experiments. Experimental results show that the classification accuracy based on IRT images using CNN is superior to that based on vibration signal, which breaks the previous situation based on traditional vibration signals to rule the field of rotating machinery fault diagnosis. The results also show that the diagnostic method based on IRT images for rotating machinery deserves more in-depth research.

REFERENCES

[1] Z. Li, J. Chen, Y. Zi, and J. Pan, "Independence-oriented VMD to identify fault feature for wheel set bearing fault diagnosis of high speed locomotive," *Mech. Syst. Signal Process.*, vol. 85, pp. 512–529, Feb. 2017.

[2] J. Zhou, L. Wee, and Z. Zhong, "A knowledge base system for rotary equipment fault detection and diagnosis," in *Proc. Int. Conf. Control Autom. Robot. Vis.*, Dec. 2010, pp. 1335–1340.

[3] A. Saxena and A. Saad, "Evolving an artificial neural network classifier for condition monitoring of rotating mechanical systems," *Appl. Soft Comput.*, vol. 7, no. 1, pp. 441–454, 2007.

[4] Z. Liu, Z. Jia, C. Vong, J. Han, C. Yan, and M. Pecht, "A patent analysis of prognostics and health management (PHM) innovations for electrical systems," *IEEE Access*, vol. 6, pp. 18088–18107, 2018.

[5] A. Morhain and D. Mba, "Bearing defect diagnosis and acoustic emission," *Proc. Inst. Mech. Eng. J, J. Eng. Tribol.*, vol. 217, no. 4, pp. 257–272, 2003.

[6] C. K. Tan, P. Irving, and D. Mba, "A comparative experimental study on the diagnostic and prognostic capabilities of acoustics emission, vibration and spectrometric oil analysis for spur gears," *Mech. Syst. Signal Process.*, vol. 21, no. 1, pp. 208–233, Jan. 2007.

[7] G. Dong and J. Chen, "Noise resistant time frequency analysis and application in fault diagnosis of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 33, pp. 212–236, Nov. 2012.

[8] Y. Li, G. Li, Y. Yang, X. Liang, and M. Xu, "A fault diagnosis scheme for planetary gearboxes using adaptive multi-scale morphology filter and modified hierarchical permutation entropy," *Mech. Syst. Signal Process.*, vol. 105, pp. 319–337, May 2018.

[9] H. Liu, C. Liu, and Y. Huang, "Adaptive feature extraction using sparse coding for machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 25, no. 2, pp. 558–574, 2011.

[10] M. Dowell, G. Sylvester, R. Krupp, and G. Zipfel, "Progress in turbomachinery prognostics and health management via eddy-current sensing," in *Proc. IEEE Aerosp. Conf.*, vol. 6, Mar. 2000, pp. 133–143.

[11] C. Roeseler, A. von Flotow, and P. Tappert, "Monitoring blade passage in turbomachinery through the engine case (no holes)," in *Proc. IEEE Aerosp. Conf.*, vol. 6, Mar. 2002, pp. 3125–3129.

[12] Y. Li, Y. Yang, G. Li, M. Xu, and W. Huang, "A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mRMR feature selection," *Mech. Syst. Signal Process.*, vol. 91, pp. 295–312, Jul. 2017.

[13] H. Shao, H. Jiang, H. Zhang, and T. Liang, "Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2727–2736, Mar. 2018.

[14] Y. Li, Y. Yang, X. Wang, B. Liu, and X. Liang, "Early fault diagnosis of rolling bearings based on hierarchical symbol dynamic entropy and binary tree support vector machine," *J. Sound Vib.*, vol. 428, pp. 72–86, Aug. 2018.

[15] H. Shao, H. Jiang, H. Zhang, W. Duan, T. Liang, and S. Wu, "Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing," *Mech. Syst. Signal Process.*, vol. 100, pp. 743–765, Feb. 2018.

[16] J. Antoni and R. Randall, "Unsupervised noise cancellation for vibration signals: Part I—Evaluation of adaptive algorithms," *Mech. Syst. Signal Process.*, vol. 18, no. 1, pp. 89–101, 2004.

[17] J. Antoni and R. Randall, "Unsupervised noise cancellation for vibration signals: Part II—A novel frequency-domain algorithm," *Mech. Syst. Signal Process.*, vol. 18, no. 1, pp. 103–117, 2004.

[18] D. Lopez-Perez and J. Antonino-Daviu, "Failure detection in industrial electric motors through the use of infrared-based isothermal representation," in *Proc. 43rd Annu. Conf. IEEE-Ind.-Electron.-Soc.*, Oct./Nov. 2017, pp. 3822–3827.

[19] Z. Hui and H. Fuzhen, "An intelligent fault diagnosis method for electrical equipment using infrared images," in *Proc. 34th Chin. Control Conf.*, Jul. 2015, pp. 6372–6376.

[20] M. S. Jadin and S. Taib, "Recent progress in diagnosing the reliability of electrical equipment by using infrared thermography," *Infr. Phys. Technol.*, vol. 55, no. 4, pp. 236–245, 2012.

[21] L. Duan, M. Yao, J. Wang, T. Bai, and L. Zhang, "Segmented infrared image analysis for rotating machinery fault diagnosis," *Infr. Phys. Technol.*, vol. 77, pp. 267–276, 2016.

[22] A. M. D. Younus and B.-S. Yang, "Intelligent fault diagnosis of rotating machinery using infrared thermal image," *Expert Syst. Appl.*, vol. 39, no. 2, pp. 2082–2091, 2012.

[23] G. Lim, D. Bae, and J. Kim, "Fault diagnosis of rotating machine by thermography method on support vector machine," *J. Mech. Sci. Technol.*, vol. 28, no. 8, pp. 2947–2952, 2014.

[24] V. Tran, B. Yang, F. Gu, and A. Ball, "Thermal image enhancement using bi-dimensional empirical mode decomposition in combination with relevance vector machine for rotating machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 38, no. 2, pp. 601–614, 2013.

[25] M. Eftekhari, M. Moallem, S. Sadri, and M.-F. Hsieh, "A novel indicator of stator winding inter-turn fault in induction motor using infrared thermal imaging," *Infr. Phys. Technol.*, vol. 61, pp. 330–336, Nov. 2013.

[26] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 1–22.

[27] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[28] R. Shekhar and C. V. Jawahar, "Word image retrieval using bag of visual words," in *Proc. 10th IAPR Int. Workshop Document Anal. Syst. (DAS)*, Mar. 2012, pp. 297–301.

[29] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated recognition, localization and detection using convolutional networks," in *Proc. Int. Conf. Learn. Represent.*, 2014. [Online]. Available: https://arxiv.org/abs/1312.6229

[30] Z. Liu, T. Liu, J. Han, S. Bu, X. Tang, and M. Pecht, "Signal model-based fault coding for diagnostics and prognostics of analog electronic circuits," *IEEE Trans. Ind. Electron.*, vol. 64, no. 1, pp. 605–614, Jan. 2017.

[31] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[32] T. Lindeberg, "Scale-space theory: A basic tool for analysing structures at different scales," *J. Appl. Statist.*, vol. 21, nos. 1–2, pp. 224–270, Jun. 1994.

[33] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2. Sep. 1999, pp. 1150–1157.

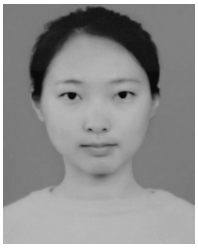
[34] M. Brown and D. Lowe, "Invariant features from interest point groups," in *Proc. Brit. Mach. Vis. Conf.*, vol. 23, 2002, pp. 1–10.

[35] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2007, pp. 1027–1035.

[36] D. Nistér and H. Stewénius, "Scalable recognition with a vocabulary tree," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2006, pp. 2161–2168.

[37] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

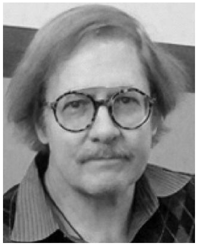
[38] Z. Liu, Z. Jia, C.-M. Vong, S. Bu, and J. Han, "Capturing high-discriminative fault features for electronics-rich analog system via deep learning," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1213–1226, Jun. 2017.



ZHEN JIA was born in Shanxi, China, in 1993. She received the B.S. degree in electrical engineering and automation from the Xi'an University of Architecture and Technology, Xi'an, China, in 2015. She is currently pursuing the Ph.D. degree in means of transport applied engineering with Northwestern Polytechnical University, Xi'an. Her main research interests include prognostics and health management, state monitoring, and deep learning.

ZHENBAO LIU (M'11–SM'18) received the B.S. and M.S. degrees from Northwestern Polytechnical University, Xi'an, China, in 2001 and 2004, respectively, and the Ph.D. degree from the University of Tsukuba, Tsukuba, Japan, in 2009, all in electrical engineering and automation. He was a Visiting Scholar with Simon Fraser University, Canada, in 2012. He is currently a Professor with Northwestern Polytechnical University. His research interests include UAV, prognostics and health management, and aircraft fault diagnosis. He is an Associate Editor of the IEEE ACCESS.

CHI-MAN VONG (M'12–SM'15) received the M.S. and Ph.D. degrees in software engineering from the University of Macau, Macau, in 2000 and 2005, respectively. He is currently an Associate Professor with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau. His research interests include machine learning methods and intelligent systems.



MICHAEL PECHT (M'83–SM'90–F'92) received the M.S. degree in electrical engineering, and the M.S. and Ph.D. degrees in engineering mechanics from the University of Wisconsin Madison, Madison, WI, USA, in 1978, 1979, and 1982, respectively. He is currently the Founder of the Center for Advanced Life Cycle Engineering, University of Maryland at College Park, College Park, MD, USA, which is funded by over 150 of the world's leading electronics companies at over U.S. $6 million/yr. He is also the George E. Dieter Professor of mechanical engineering, and a Professor of applied mathematics with the University of Maryland. He has authored over 20 books on electronic product development, use, and supply chain management, and over 500 technical articles.

• • •