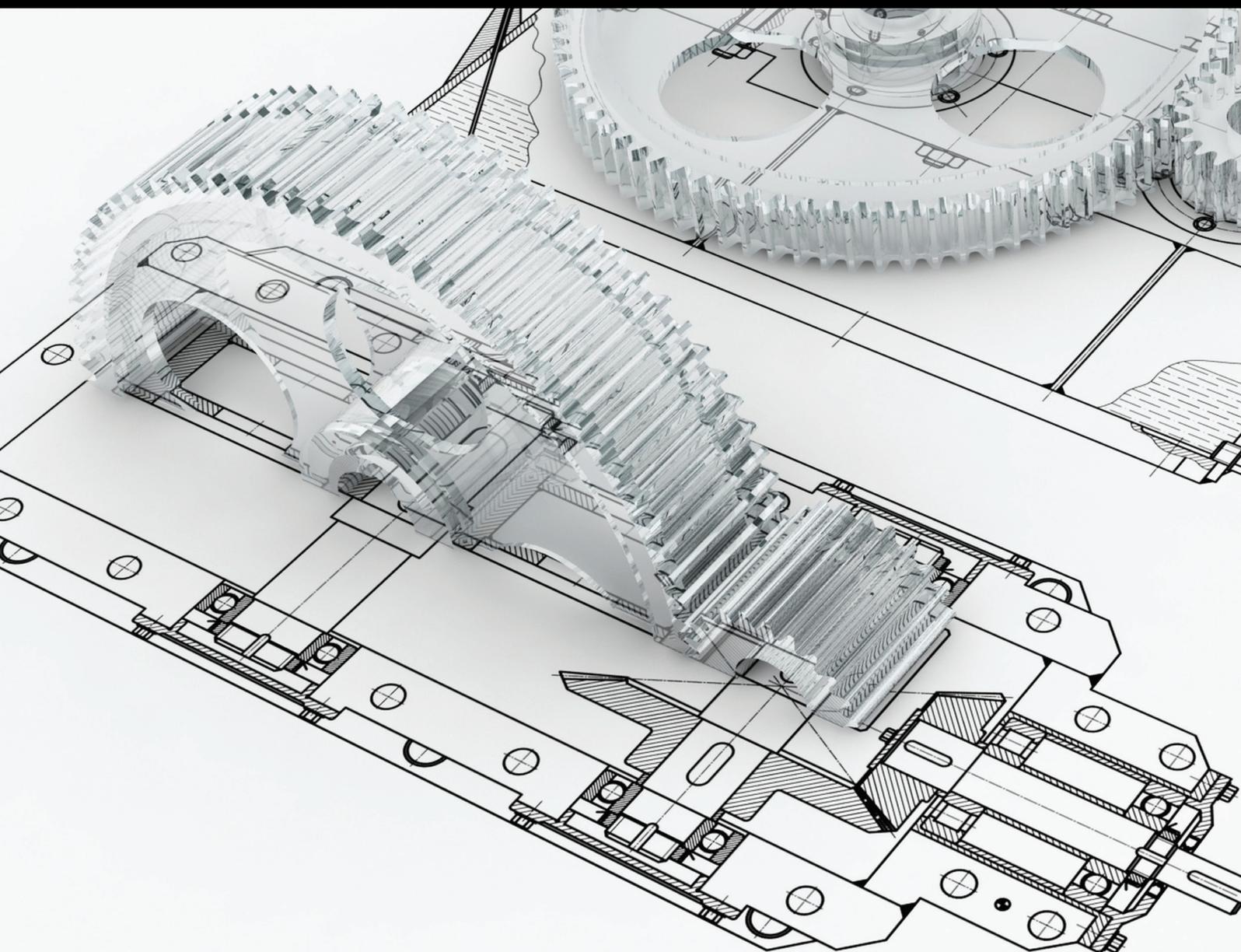


Prognostics and Maintenance for Mechanical Systems in Harsh Environment

Guest Editors: Cheng-liang Liu, Jay Lee, Liang Gong, and Yixiang Huang





Prognostics and Maintenance for Mechanical Systems in Harsh Environment

Advances in Mechanical Engineering

Prognostics and Maintenance for Mechanical Systems in Harsh Environment

Guest Editors: Cheng-liang Liu, Jay Lee, Liang Gong,
and Yixiang Huang



Copyright © 2013 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in "Advances in Mechanical Engineering." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

- Koshi Adachi, Japan
Mehdi Ahmadian, USA
Rehan Ahmed, UK
Muhammad T. Akhtar, Japan
Nacim Alilat, France
M. Affan Badar, USA
Luis Baeza, Spain
R. Balachandran, UK
Claude Bathias, France
Adib Becker, UK
Leonardo Bertini, Italy
L. A. Blunt, UK
Noël Brunetière, France
Marco Ceccarelli, Italy
Fakher Chaari, Tunisia
Chin-Lung Chin, Taiwan
Hyung H. Cho, Republic of Korea
Seung-Bok Choi, Korea
Kangyao Deng, China
Francisco D. Denia, Spain
T.S. Dhanasekaran, USA
Nihad Dukhan, USA
Farzad Ebrahimi, Iran
Ali El Wahed, UK
Bogdan I. Epureanu, USA
M. R. Eslami, Iran
Ali Fatemi, USA
Mario L. Ferrari, Italy
Siegfried Fouvry, France
Ian Frigaard, Canada
Mergen H. Ghayesh, Canada
Lus Godinho, Portugal
- Tian Han, China
Woo-Suck Han, France
F. J. Huera-Huarte, Spain
D. Jalali-Vahid, Iran
Jiin Yuh Jang, Taiwan
Zhongmin Jin, UK
Xiaodong Jing, China
S.-W.Kang, Republic of Korea
Xianwen Kong, UK
Michal Kuciej, Poland
Yaguo Lei, China
Z. Li, The Netherlands
Cheng-Xian Lin, USA
Jaw-Ren Lin, Taiwan
S. Nima Mahmoodi, USA
Oronzio Manca, Italy
Ramiro Martins, Portugal
Aristide F. Massardo, Italy
Francesco Massi, Italy
T.H. New, Singapore
Kim Choon Ng, Singapore
C. T. Nguyen, Canada
Hiroshi Noguchi, Japan
Hakan F. Oztop, Turkey
Duc Truong Pham, UK
Jurij Prezelj, Slovenia
Xiaotun Qiu, USA
Pascal Ray, France
R. L. Reuben, UK
Pedro A.R. Rosa, Portugal
E. de Sá Caetano, Portugal
D.R. Salgado, Spain
- M. Reza Salimpour, Iran
Sunetra Sarkar, India
Pietro Scandura, Italy
A. Seshadri Sekhar, India
A. S. Dalkılıç, Turkey
Liyuan Sheng, China
Xi Shi, China
Seiichi Shiga, Japan
C. S. Shin, Taiwan
Ray W. Snidle, UK
Margaret M. Stack, UK
Neil Stephen, UK
Kumar K. Tamma, USA
Yaya Tan, China
Anand Thite, UK
Cho W. Solomon To, USA
Yoshihiro Tomita, Japan
Shan-Tung Tu, China
S. Velarde-Suárez, Spain
Moran Wang, China
Junwu Wang, China
Jia-Jang Wu, Taiwan
Fengfeng Xi, Canada
Gongnan Xie, China
Hiroshi Yabuno, Japan
Wei Mon Yan, Taiwan
Jianqiao Ye, UK
Byeng D. Youn, USA
Bo Yu, China
Jianbo Yu, China
Zhongrong Zhou, China

Contents

Prognostics and Maintenance for Mechanical Systems in Harsh Environment, Cheng-liang Liu, Jay Lee, Liang Gong, and Yixiang Huang
Volume 2013, Article ID 121340, 2 pages

Study on Immune Relevant Vector Machine Based Intelligent Fault Detection and Diagnosis Algorithm, Zhong-hua Miao, Guang-xing Zhou, Xiao-hua Wang, and Chuang-xin He
Volume 2013, Article ID 548248, 8 pages

Autonomous Navigation of a Surveillance Robot in Harsh Outdoor Road Environments, Youjin Shin, Donghyeon Kim, Hyunsuk Lee, Jooyoung Park, and Woojin Chung
Volume 2013, Article ID 837484, 15 pages

A Computationally Efficient and Adaptive Approach for Online Embedded Machinery Diagnosis in Harsh Environments, Chuan Jiang and Samuel H. Huang
Volume 2013, Article ID 847612, 12 pages

Condition Evaluation of Large Generator Stator Insulation Based on Partial Discharge Measurement, Ruihua Li, Ling Pan, Chen Yan, Hao Li, and Bo Hu
Volume 2013, Article ID 765374, 10 pages

Wavelet Leaders Based Vibration Signals Multifractal Features of Plunger Pump in Truck Crane, Wenliao Du, Chunyan Yang, Ansheng Li, and Liangwen Wang
Volume 2013, Article ID 676404, 11 pages

Performance Assessment for a Fleet of Machines Using a Combined Method of Ant-Based Clustering and CMAC, Lei Zhang, Qixin Cao, and Jay Lee
Volume 2013, Article ID 603071, 12 pages

Improved Visual Hook Capturing and Tracking for Precision Hoisting of Tower Crane, Yanming Li, Shuangyuan Wang, and Bingchu Li
Volume 2013, Article ID 426810, 8 pages

EMD and Wavelet Transform Based Fault Diagnosis for Wind Turbine Gear Box, Qingyu Yang and Dou An
Volume 2013, Article ID 212836, 9 pages

Identification of Magnetizing Inrush and Internal Short-Circuit Fault Current in V/x-Type Traction Transformer, Y. H. Li, X. J. Tian, and X. Q. Li
Volume 2013, Article ID 905202, 8 pages

Thermal Performance Prediction of a Trapezoidal Cavity Absorber for a Linear Fresnel Reflector, Yanhua Lai, Tao Wu, Shuping Che, Zhen Dong, and Mingxin Lyu
Volume 2013, Article ID 615742, 7 pages

Factors That Influence Failure Behaviour and Remaining Useful Life of Mining Equipment Components, Mark Ho and Melinda Hodkiewicz
Volume 2013, Article ID 913048, 9 pages

Editorial

Prognostics and Maintenance for Mechanical Systems in Harsh Environment

Cheng-liang Liu,¹ Jay Lee,² Liang Gong,¹ and Yixiang Huang²

¹ School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

² Department of Mechanical Engineering, University of Cincinnati, Cincinnati, OH 45221, USA

Correspondence should be addressed to Cheng-liang Liu; chlliu@sjtu.edu.cn

Received 26 September 2013; Accepted 26 September 2013

Copyright © 2013 Cheng-liang Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

The idea of preparing this special issue aims at solving emerging problems when condition monitoring is applied to more and more machines running in practical harsh environments. Such machines, including wind turbines, mining and civil engineering machinery, are not easy to be frequently checked and maintained by man power. Most established machine condition monitoring are developed in ideal lab-setting conditions. Considering the real-world highly dynamical load and harsh working conditions, there is a need to develop practical ways to handle various situations and rapidly response to the faults and significant degradation process. Thus, the authors of this special issue were asked to prepare papers on the following topics:

- (i) diagnosis and maintenance of engineering equipment within extreme environment,
- (ii) mechanical structure analysis under extreme environment,
- (iii) decision-making strategies for reliable operations of mechanical systems,
- (iv) information/data mining on mechanical system maintenance,
- (v) robust, self-sustainable remote sensing, and monitoring systems.

2. Prognostics and Maintenance for Mechanical Systems in Harsh Environment: Past, Present, and Future

The safety, reliability, and maintainability of mechanical equipment during their service life are issues of critical importance to be solved by either the manufacturing industry or the service suppliers. In the past, for the machines operating in harsh environment, the reliable and robust design was the only aspect that people could do their best to keep them in the reliable service time as long as possible. The traditional “fail and fix (FAF)” practice was adopted at that time. However, unexpected downtime often may cause great loss because the maintenance could take so much time to repair, or even the replacement of the faulty machine with a new one becomes a must-do due to severe chain damages. Therefore, firstly, people began to assume some certain level of performance degradation for the machinery, without input from the working load or any sensors with itself, and serviced equipment on a routine schedule, whether the service is actually needed or not, which may be called blindly proactive maintenance. With fast development of sensor and tether-free communication technologies, condition-based maintenance (CBM) is found to be more effective approaches in achieving the goals of reliability and maintainability for this mechanical system running in harsh environment, which has become a value-added service, apart from the quality of products in design and manufacturing. The CBM also makes it possible to measure the faults more precisely and based on the evidence on data from running machine, severe

unexpected faults can be avoided by scheduling the precautionary maintenance before they actually occur. Thus, the prognostics shifts the maintenance into the “predict and prevent (PAP)” paradigm. From then on, more and more challenges emerge with wider and wider applications of prognostics, for example, the disturbances in the monitored signals under working process, the problem of feature extraction from long sampling signals, and some inherent disadvantages in the feature-based machine condition recognition. In order to effectively extract the information carried by the monitoring data, redundancy reduction techniques may be taken. The feature-based machine condition recognition also has some inherent disadvantages, such as the “inevitable misdiagnosis” problem and “lacking of historical monitoring data”. For the problem of how to overcome the disturbances under the working process, the blind separation technique based on sparse component analysis could be applied. Other new trends for better applying prognostics may include the methods related to IoT (Internet of Things), cloud computing, big data, and the refactoring database methodology. With data mining technology, the intelligence models should be formed as generative models that can self-detect and self-heal the mechanical systems.

3. Special Issue Overview

In this special issue, the above-mentioned trends are embodied by pioneer research. Classical digital signal processing methodologies, such as Wavelet and EMD, with improved performance and targeted application by W. Du et al. in “*Wavelet leaders based vibration signals multifractal features of plunger pump in truck crane*” and Q. Yang et al. in “*EMD and Wavelet transform based fault diagnosis for wind turbine gear box*” are presented to rapidly and precisely mine the fault data. One of the most challenging issues in prognostics and maintenance fields is to manipulate the incomplete data for evaluating the mechanical systems’ performance. In the paper titled “*Study on immune relevant vector machine based intelligent fault detection and diagnosis algorithm*,” Z. Miao et al. proposes a “self/nonsel” identification principle to effectively tackle this problem with the inspiration of artificial immune system. To efficiently extract features from big data and conduct the dimension-reduction processing, a condition-based monitoring system that is tailored to the embedded deployment is demonstrated in “*A computationally efficient and adaptive approach for online embedded machinery diagnosis in harsh environments*” by C. Jiang and S. H. Huang Boosted by the viewpoint of the whole system performance assurance, in “*Performance assessment for a fleet of machines using a combined method of ant-based clustering and cmac*,” L. Zhang et al. advocate this concept and the relevant analyzing techniques. As spin-off hot topics, some cross-discipline analysis for electrical mechanical system performance is also embraced in this issue, such as “*Condition evaluation of large generator stator insulation based on partial discharge measurement*,”

which highlights the fault detection and quantitative assessment procedures in harsh environment.

Cheng-liang Liu
Jay Lee
Liang Gong
Yixiang Huang

Research Article

Study on Immune Relevant Vector Machine Based Intelligent Fault Detection and Diagnosis Algorithm

Zhong-hua Miao,¹ Guang-xing Zhou,¹ Xiao-hua Wang,¹ and Chuang-xin He²

¹ *Mechatronics Engineering and Automation, Shanghai University, Shanghai 200072, China*

² *School of Mechanical Engineering, Shanghai Jiaotong University, Shanghai 200240, China*

Correspondence should be addressed to Xiao-hua Wang; x.wang@shu.edu.cn

Received 22 April 2013; Revised 5 July 2013; Accepted 7 July 2013

Academic Editor: Cheng-liang Liu

Copyright © 2013 Zhong-hua Miao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An immune relevant vector machine (IRVM) based intelligent classification method is proposed by combining the random real-valued negative selection (RRNS) algorithm and the relevant vector machine (RVM) algorithm. The method proposed is aimed to handle the training problem of missing or incomplete fault sampling data and is inspired by the “self/nonself” recognition principle in the artificial immune systems. The detectors, generated by the RRNS, are treated as the “nonself” training samples and used to train the RVM model together with the “self” training samples. After the training succeeds, the “nonself” detection model, which requires only the “self” training samples, is obtained for the fault detection and diagnosis. It provides a general way solving the problems of this type and can be applied for both fault detection and fault diagnosis. The standard Fisher’s Iris flower dataset is used to experimentally testify the proposed method, and the results are compared with those from the support vector data description (SVDD) method. Experimental results have shown the validity and practicability of the proposed method.

1. Introduction

The system failure of a complex mechatronic machine is usually caused by the failure of its critical components. The nature of the fault diagnosis is to gain the correct feature parameters of the critical parts by monitoring the relevant internal and external signals. Therefore, intelligent fault detection and diagnosis are widely treated as a pattern recognition problems. In the problem, the normal state and the fault state are categorized as different kinds of patterns. A rational model is set up between the state features and different patterns, which is also addressed as the fault detection model and can intelligently recognize the working status of the devices.

The literature on the fault detection methods is very rich. Most of its parts are on the supervised learning methods, such as neural networks [1], support vector machines [2–4], relevance vector machine [5–8], naive Bayes [9–11], and decision tree [12]. Supervised learning methods require a sufficient number of fault data for training; however, fault data are usually insufficient in practice, particularly at the early stages of operation, where probably only normal data samples are

available. Under this circumstance, the traditional intelligent diagnosis methods have their deficiencies: the fault diagnosis model can hardly be trained with only normal samples; moreover, the fault detection model trained with normal data together with only part of the failure data cannot identify unknown faults and always assigns them wrongly to some known categories.

Inspired by the “self/nonself” recognition principle in artificial immune systems, an immune relevant vector machine algorithm is proposed to solve the problems with traditional methods. This method combines the random real-valued negative selection (RRNS) algorithm and the relevant vector machine (RVM) algorithm, detecting the “nonself” mode with only “self” training samples. Based on the proposed algorithm and its utilization with other methods, accurate fault detection and unknown failure detection can be performed with only normal samples or incomplete training samples.

The remainder of this paper is organized as follows. The first section introduces the immune relevant nonself vector machine detection algorithm. The second section describes

the fault detection model based on the immune relevant vector machine. In the third section, standard Fisher's Iris flower data are applied to experimentally verify the proposed algorithm and the model. The results are compared with other anomaly detection algorithms. The fourth section gives the conclusions.

2. Immune Relevant Vector Machine "Nonself" Fault Detection Algorithm

2.1. Definitions

Definition 1 (system state space). System state can be denoted by a feature vector \mathbf{x} , which is normalized and expressed as follows:

$$\mathbf{x} = (x_1, x_2, \dots, x_n) \in [0, 1]^n. \quad (1)$$

System state space is denoted as $\mathbf{G} \subseteq [0, 1]^n$, including all the possible system states.

Definition 2 (self space). The self space is a space of feature vectors with known target status, denoted as \mathbf{S} , where $\mathbf{S} \subset \mathbf{G}$.

Definition 3 (nonself space). The nonself space is the complement of the self space in the system state space, denoted by \mathbf{N}_S , where $\mathbf{S} + \mathbf{N}_S = \mathbf{G}$.

Definition 4 (nonself detection). The nonself detection here includes both the anomaly detection and the multiclass fault diagnosis in the sense of the traditional negative selection algorithm. For the nonself detection problem, a state recognition function is estimated given a self sample set $\mathbf{S}' \subseteq \mathbf{S}$, and then the function is applied to identify whether an observed system state belongs to a self space or a nonself space.

Note that, for brevity, self data/samples denote data/samples in the self space, which refer to normal data or data of a known class, and nonself data/sample denote data/sample in the nonself space, which are abnormal data or data that do not belong to a given class.

2.2. Improved Normalization Method. Each dimension of the feature vector \mathbf{x} must be normalized within the range of $[0, 1]$. When part of the self space is known and the whole system state space is unknown, the literature on how to properly normalize the feature vector is relatively rare. References [13, 14] have proposed a normalization method, where the self sample is normalized into the space $[0, 1]^n$ using the following equation (2):

$$s(i) = \frac{\mathbf{S}(i) - \mathbf{B} \cdot \min[\mathbf{S}(i)]}{\max[\mathbf{S}(i)] - \min[\mathbf{S}(i)]}, \quad i = 1, 2, \dots, n, \quad (2)$$

where $\mathbf{S}(i)$ is the i th column of the original self sample matrix \mathbf{S} , $s(i)$ is the i th column of the normalized self sample matrix, and $\max[\mathbf{S}(i)]$ and $\min[\mathbf{S}(i)]$ are the maximum and the minimum values of $\mathbf{S}(i)$, respectively. $\mathbf{B}' = [1, 1, \dots, 1]_{1 \times n}$ is the dimension of the feature vector space.

The disadvantage of the previous method is as follows.

- (1) Space $[0, 1]^n$ is actually a self space after normalization using (2). Its boundary is either the maximum or

the minimum value of the self samples, which cannot represent the whole system space. Therefore, the detectors generated in this $[0, 1]^n$ space cannot represent the nonself space.

- (2) The original samples must be normalized following the same procedure to guarantee their comparability, so as the unknown samples. However, the dimension of some feature vector of the unknown sample may exceed that of $\max[\mathbf{S}(i)]$ or $\min[\mathbf{S}(i)]$. If all samples are normalized in the same way, the dimension will overflow and the detection will not succeed.

For real-world applications, the range of the feature vector in the nonself space is unknown, and only part of the self space samples can be used to estimate the self space. It is very important to normalize the whole system state space before any other manipulations. An improved normalization method is proposed based on the RRNS and the RVM methods.

Let $\max[\mathbf{G}(i)]$ and $\min[\mathbf{G}(i)]$ be the maximum and the minimum values of the i th feature vector of the system state space \mathbf{G} , respectively, where $i = 1, 2, \dots, n$.

Assumption 5. Assume that the self space is in the middle of the whole system state space. Range of feature vectors in the self space is 1/3 of the whole space. The nonself space occupies the rest larger 1/3 and the smaller 1/3 of the range $[0, 1]$. The following equations exist accordingly:

$$\min[\mathbf{G}(i)] = \min[\mathbf{S}(i)] - (\max[\mathbf{S}(i)] - \min[\mathbf{S}(i)]) \quad (3)$$

$$\max[\mathbf{G}(i)] = \max[\mathbf{S}(i)] + (\max[\mathbf{S}(i)] - \min[\mathbf{S}(i)])$$

$$\max[\mathbf{S}(i)] - \min[\mathbf{S}(i)] = \frac{1}{3} (\max[\mathbf{G}(i)] - \min[\mathbf{G}(i)]). \quad (4)$$

Assumption 6. Assume that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is one original sample with the i th feature vector as X_i , where $i = 1, 2, \dots, n$. If X_i lies out of the range of $\mathbf{G}(i)$, then its value is truncated as follows:

$$\text{if } X_i > \max[\mathbf{G}(i)] \quad \text{then } X_i = \max[\mathbf{G}(i)] \quad (5)$$

$$\text{if } X_i < \min[\mathbf{G}(i)] \quad \text{then } X_i = \min[\mathbf{G}(i)].$$

The following normalization is proposed based on Assumptions 5 and 6:

$$x_i = \frac{X_i - \min[\mathbf{G}(i)]}{\max[\mathbf{G}(i)] - \min[\mathbf{G}(i)]}, \quad i = 1, 2, \dots, n. \quad (6)$$

By substituting (3) into (6), the normalization equation related only to the self space samples is obtained, which is

$$x_i = \frac{X_i - 2 \cdot \min[\mathbf{S}(i)] + \max[\mathbf{S}(i)]}{3 \cdot (\max[\mathbf{S}(i)] - \min[\mathbf{S}(i)])}, \quad i = 1, 2, \dots, n. \quad (7)$$

After normalizing all the n dimensions, the feature vector, $\mathbf{x} = (x_1, x_2, \dots, x_n)$, belongs to $[0, 1]^n$.

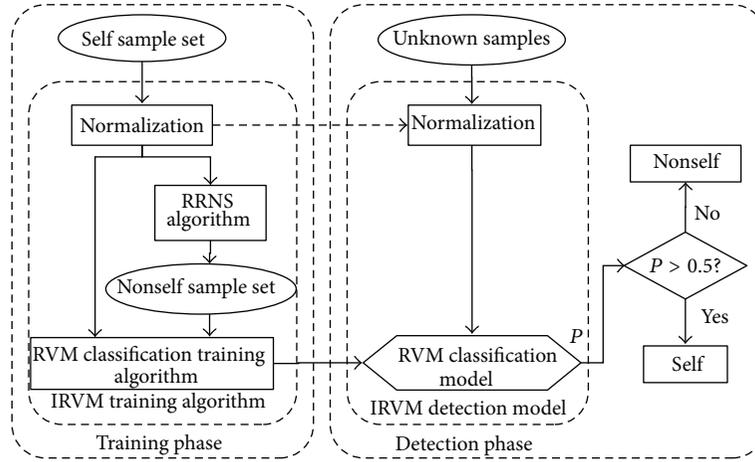


FIGURE 1: Flowchart of the relevance vector machine based nonself-detection algorithm.

Remarks. This algorithm combines RRNS and RVM, where the detectors generated by RRNS are considered as the set of nonself samples, together with the set of self space samples, serving as the two classes of samples for training the RVM classification model. The main purpose here is to construct a hyperplane for the classification between the self space and the nonself space.

- (1) The proposed normalization method maps the original self space into the center of the system state space surrounded by the nonself space. Therefore, it is possible to train and find the hyperplane separating the self space and the nonself space.
- (2) The normalized samples occupy the center 1/3 range. The self space and the nonself space are balanced in either the positive direction or the negative direction, which enables the training of the classifying hyperplane, that is, the smallest irregular hypersphere containing the self space samples.
- (3) After the training succeeds, the truncated samples and the results after truncation are considered as nonself samples. Therefore, the truncation does not affect the classification. The classification model can be applied in the whole \mathbf{R}^n space without dimension overflow.

2.3. Immune Relevant Vector Machine (IRVM) Nonself Detection Algorithm. According to the mechanism of biological immune recognition, the traditional immune recognition algorithm generates a large number of detectors using different negative selection algorithms. The detectors are then compared with unknown state samples to detect anomaly. Although the method is simple and intuitive, its efficiency is usually low because the number of detectors is too large in practice, which inhibits the application of the method.

In the studies here, the detectors are generated as the nonself supervised learning samples because they are evenly distributed in the nonself space, which are then trained using

RVM. Those are the most important ideas of the immune relevant vector machine nonself detection algorithm.

Figure 1 is the flowchart of the IRVM nonself detection algorithm, including the training phase and the detection phase. During the training phase, the training samples are the self space samples instead of the nonself space samples. Three steps are required during the training phase.

Step 1. The system state space normalization: The self sample set is normalized according to (7), denoted as \mathbf{S}' after normalization.

Step 2. Nonself sample generation: A fixed number of detectors are generated in the system state space using RRNS based on the self sample set \mathbf{S}' . The detectors serve as the training sample set in the nonself space, denoted as $\mathbf{N}_{\mathbf{S}'}$.

Step 3. RVM classification model training: Merge the self sample set \mathbf{S}' and the nonself samples set $\mathbf{N}_{\mathbf{S}'}$, input them to the RVM classification training algorithm, and train the RVM classification model using the RBF kernel functions, whose radius is selected as 0.5, which is the center of the normalized range.

The IRVM training algorithm can be considered as a combination of the normalization, RRNS, and the RVM classification algorithm, as shown in Figure 1. IRVM nonself detection algorithm requires only one class of training samples and can obtain the RVM classification model for two classes.

During the detection phase, RVM nonself detection model contains two parts, the normalization process and the RVM classification model training, whose output is the posterior probability. If the posterior probability is greater than 0.5, then the sample is a self sample, otherwise, a nonself sample.

The combination of RRNS algorithm and RVM algorithm in the IRVM nonself detection algorithm exploits the RRNS ability of simulating nonself samples in the nonself space and adopts the RVM advantages of being simple and fast at the same time.

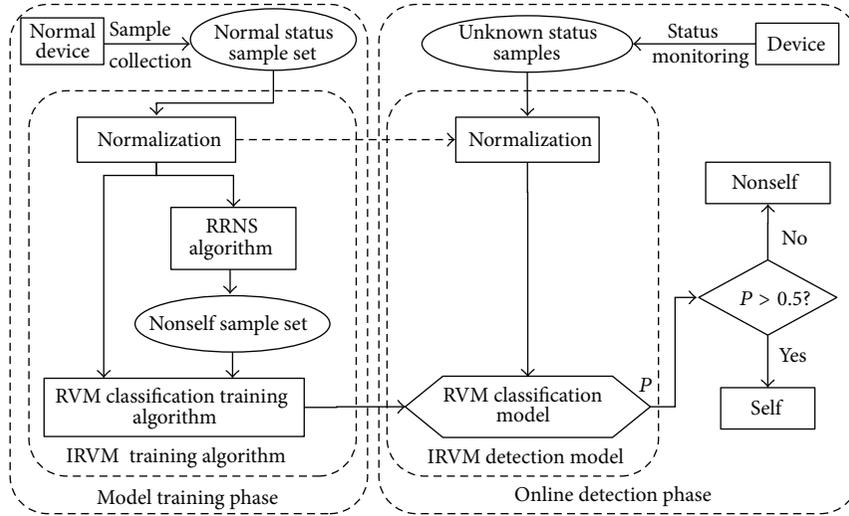


FIGURE 2: Immune relevance vector machine based fault detection system.

3. Fault Detection and Diagnosis Based on Immune Relevant Vector Machine

3.1. Fault Detection Model Based on IRVM. Fault detection is a typical classification problem for two, since only the normal state and the fault state exist. The fault diagnosis model based on IRVM classification needs only the normal samples. The model treats the feature vector space under normal conditions as the self space and the feature vector under fault conditions as the nonself space.

The IRVM fault detection system is shown in Figure 2, including the model training phase and the online detection phase. Compared with Figure 1, the whole system is included in this figure.

Two steps are included in the model training phase.

Step 1. A number of normal state samples are collected, the characteristic features are extracted, the feature vectors are constructed, and the set of normal samples is formulated. According to the real working conditions, different types of data may be collected to validate of the final detection model under different working conditions.

Step 2. The IRVM fault detection model is gained after the training of the normal state samples according to the IRVM training algorithm.

During the online fault detection phase, the state data are collected, the feature vectors are extracted, and the unknown state samples to be tested are then formed and input to the IRVM fault detection model. Whether the state is a fault state or a normal state is determined according to the posterior probability output.

3.2. Fault Diagnosis Model Based on IRVM. Traditional immune detection algorithm determines self and nonself modes and then detects anomaly. In many cases, to detect fault is usually not enough. The class of the faults and the level of faults need detection too, which forms a multiclass

classification problem. Explanation on how to diagnose multiple fault classes using the immune relevant vector machine nonself detection algorithm is illustrated in the following paragraphs.

The framework of IRVM fault diagnosis model for multiple classes is plotted in Figure 3. The model is composed of multiple IRVM single class fault diagnosis models. Each IRVM single class fault diagnosis model is trained in the same way as that shown in Figure 2, while its training sample set is the fault sample set of one specific fault class instead of a normal sample set. During the training phase, one IRVM detection model for one class of failure is trained according to the IRVM training algorithm.

For detecting multiple fault classes, the procedure is as follows: first of all, the unknown sample to be detected is input to every IRVM detection model. Then, a posterior probability output is obtained connecting with one class of failure. Afterwards, the largest posterior probability is selected and the referred class becomes a candidate failure class. If the posterior probability is greater than 0.5, then the referred class of failure is the model output. Otherwise, if no class has the posterior probability greater than 0.5, then this testing sample does not belong to any known fault class and is detected as an unknown class.

The IRVM multi-class fault diagnosis model includes all target fault IRVM detection models together with a combinational detection rules:

IRVM fault diagnosis model

$$= \{ \text{all target fault IRVM detection models} \} \quad (8) \\ + \{ \text{combinational detection rules} \} .$$

The structure of the IRVM multi-class detection model is similar to that of the standard RVM one against All (RVM-OAA) algorithm, but the OAA method does not actually occupy the whole nonself space since it uses the samples other than the target samples as the nonself space sample, which

TABLE 1: Results for the 2-dimensional Iris data.

Experiment results	SVDD			IRVM				
	Tr.Err	Te.FP	Te.FN	Te.Err	Tr.Err	Te.FP	Te.FN	Te.Err
1	0.100	0.1	0.02	0.0273	0	0.0	0.03	0.0273
2	0.100	0.2	0.00	0.0182	0	0.1	0.00	0.0091
3	0.075	0.1	0.01	0.0182	0	0.0	0.02	0.0182
4	0.075	0.2	0.02	0.0364	0	0.1	0.00	0.0091
5	0.075	0.2	0.02	0.0364	0	0.0	0.00	0.0000
Average	0.085	0.16	0.014	0.0273	0	0.04	0.01	0.0127

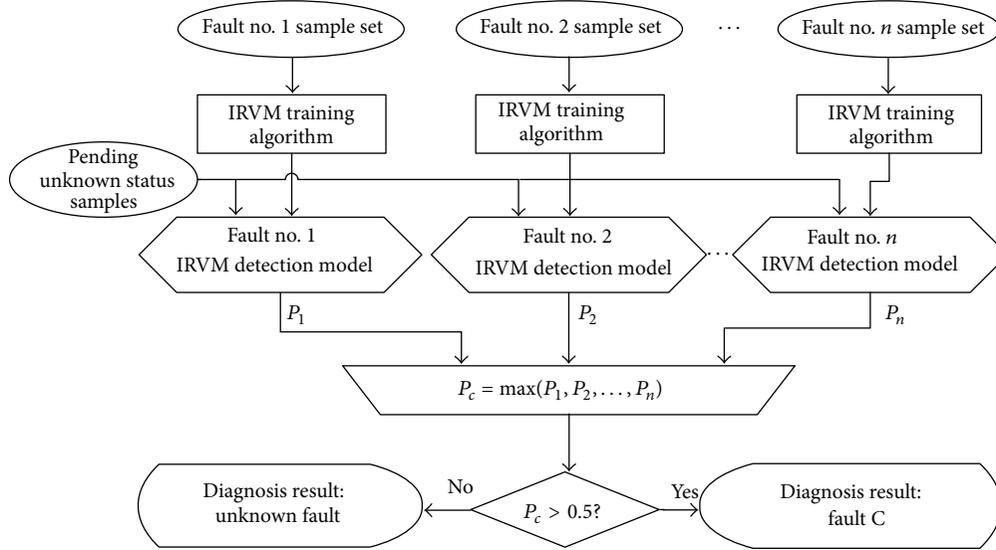


FIGURE 3: Principle of the construction of IRVM based multiclass fault diagnosis model.

may lead to misclassification of unknown classes and wrong diagnosis. Therefore, IRVM has better generalization ability than the RVM-OAA method.

4. Experimental Validation and Analysis

In order to verify the effectiveness of the IRVM nonself detection algorithm, the standard Fisher's Iris dataset is used for experimental verification and analysis. The experiment results are also compared with those of the support vector data description (SVDD) [15] method, another commonly used anomaly detection method in the area of fault detection.

4.1. Experiments on Model Verification with the Fisher's Iris Data. Fisher's Iris is a commonly recognized dataset for algorithm verification in the area of machine learning. The Iris dataset has four features of three types of irises. The three types of irises are Setosa, Virginica, and Versicolor. Each type of irises has 50 samples with a total of 150 sample data. The four features are the petal length, petal width, sepal length, and sepal width.

During the experiments, Setosa is considered as the self type, while Virginica and Versicolor are considered as the nonself types. Forty Setosa samples are randomly chosen as the training samples of the self space and the rest of the samples are all the testing samples. The experiment is repeated

for 5 times to reduce the random error and carried on with 2 features (feature 1 and feature 2) and 4 features separately for visualizing the algorithm effectiveness. The 2-dimensional data distribution is shown in Figure 4. Figure 4(a) is the original data distribution and Figure 4(b) is the data distribution after the normalization.

4.2. Results and Analysis. Tables 1 and 2 list the results for the 2-dimensional and the 4-dimensional Iris data, respectively. Tr.Err, Te.FP, and TE.FN are the training error rate, the testing false positive rate, and training false negative rate, respectively.

From Tables 1 and 2, the following conclusions are drawn:

- (1) IRVM does not have a training error while SVDD does.
- (2) Both IRVM and SVDD have a high accuracy for nonself testing samples; that is, both have low false negative rates. When all the 4 features are considered, the false negative rates are zeros, which indicates that the self set and the nonself set are highly separable.
- (3) IRVM has a better false positive rate than SVDD.
- (4) Generally, IRVM is more accurate and has better generalization ability when the self set and nonself set are highly separable.

TABLE 2: Results for the 4-dimensional Iris data.

Experiment results	SVDD			IRVM				
	Tr.Err	Te.FP	Te.FN	Te.Err	Tr.Err	Te.FP	Te.FN	Te.Err
1	0.125	0.1	0	0.0091	0	0.0	0	0.0000
2	0.100	0.2	0	0.0182	0	0.1	0	0.0091
3	0.100	0.3	0	0.0273	0	0.0	0	0.0000
4	0.075	0.6	0	0.0545	0	0.2	0	0.0182
5	0.125	0.1	0	0.0091	0	0.0	0	0.0000
Average	0.105	0.26	0	0.0236	0	0.06	0	0.0055

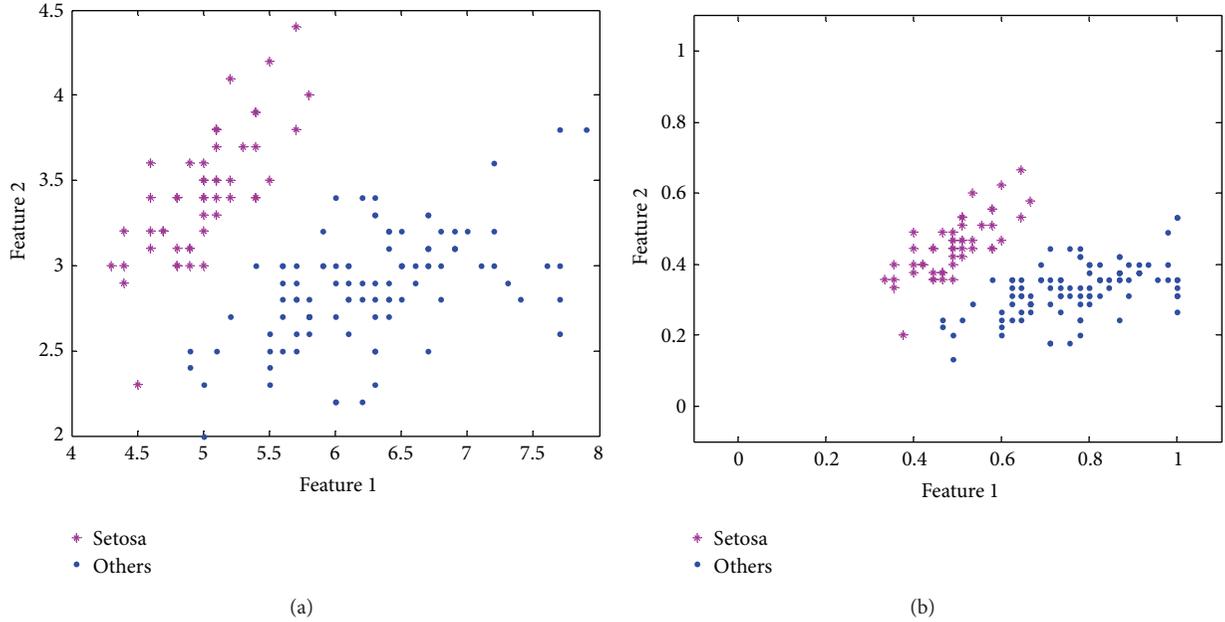


FIGURE 4: Data distribution of the Iris dataset. (a) Original value. (b) Normalized value.

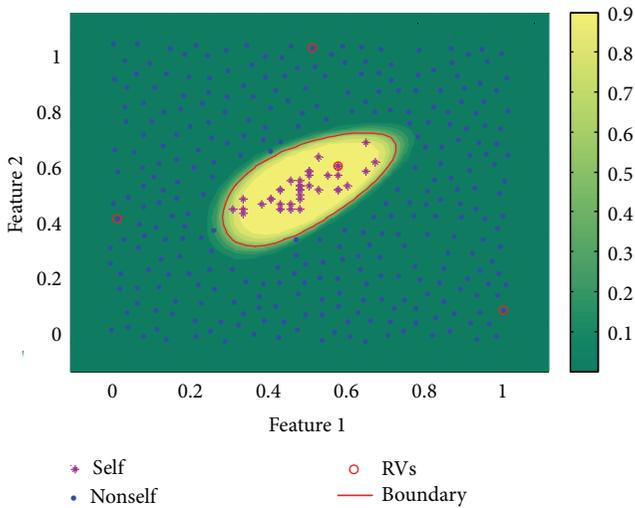


FIGURE 5: Training results of the IRVM experiment of Iris dataset.

The RRNS-RVM training and testing results are depicted in Figures 5 and 6, respectively. Figure 7 is the comparative results using SVDD. Figure 5 shows the distribution of the self

training samples, the distribution of nonself training samples using RRNS, as well as the dividing line of RVM model, the relevant vector points, and the posterior probability distribution. The following is seen:

- (1) RRNS can generate more evenly distributed nonself detectors, that is, nonself samples.
- (2) The dividing line from the classification model trained using IRVM wraps the self training sample set as a closed curve, separating the self space and the nonself space.
- (3) The improved normalization method balances the self space and the nonself space in every direction, which guarantees the training accuracy of the RVM model.

Comparing Figures 5, 6, and 7, by observing the dividing lines of IRVM and SVDD, it is seen that IRVM has zero training error, while SVDD always has some training error. The reason is that IRVM is to find a closure through training, within which is the whole self sample space, while SVDD originates from anomaly detection, that is, to remove all the abnormal samples and minimize the false positive rate [16, 17]. Similarly, IRVM model has a much lower false negative rate than the SVDD model.

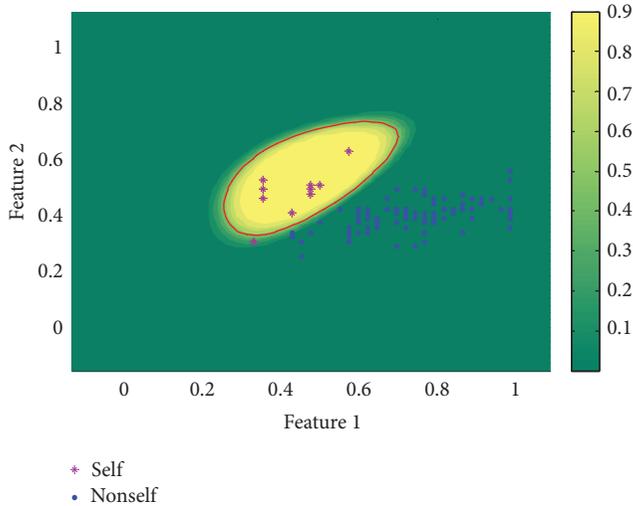


FIGURE 6: Testing results of the IRVM experiment of Iris dataset.

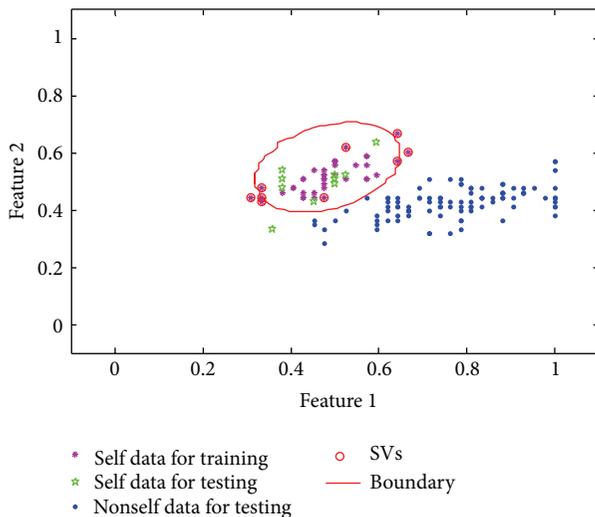


FIGURE 7: Training and testing results of the SVDD experiment of Iris dataset.

Iris dataset is a relatively simple and small dataset. The experiments here are basically for illustrative purposes. Interested readers may use several other datasets [18] to testify the performance of the proposed method.

5. Conclusions

In this paper, an immune relevant vector machine intelligent fault detection and diagnosis method is proposed, inspired by “self/nonself” recognition mechanism in the artificial immune systems. The method is able to handle the efficiency problem due to missing or incomplete fault samples in the traditional intelligent classification algorithm, since it combines the random real-valued negative selection (RRNS) algorithm in the artificial immune system and the relevant vector machine (RVM) algorithm. The detectors generated by RRNS are considered as the nonself training samples,

which are used to train the RVM model together with the self samples. The detection model is then obtained requiring only the self samples. Based on the model and the method, the fault detection can be performed using only the normal samples, and the known and unknown faults can be identified accurately with incomplete training samples. The method proposed adopts the advantage of the RNSS algorithms and outranks the traditional immune algorithms with reduced computation load and improved efficiency.

Acknowledgments

This paper is supported by the National Natural Science Foundation of China under Grant 61175038, Shanghai Education Commission Project (nos. 12YZ010, 12JC1404100, 11CH-05).

References

- [1] J. Chen, *Vibration Monitoring and Fault Detection of Mechanical Equipments*, Shanghai Jiao-Tong University Press, 1999, (Chinese).
- [2] B. Samanta, “Gear fault detection using artificial neural networks and support vector machines with genetic algorithms,” *Mechanical Systems and Signal Processing*, vol. 18, no. 3, pp. 625–644, 2004.
- [3] J. D. Wu and J. J. Chan, “Faulted gear identification of a rotating machinery based on wavelet transform and artificial neural network,” *Expert Systems with Applications*, vol. 36, no. 5, pp. 8862–8875, 2009.
- [4] B. S. Yang, W. W. Hwang, D. J. Kim, and A. C. Tan, “Condition classification of small reciprocating compressor for refrigerators using artificial neural networks and support vector machines,” *Mechanical Systems and Signal Processing*, vol. 19, no. 2, pp. 371–390, 2005.
- [5] B. S. Yang, T. Han, and W. W. Hwang, “Fault diagnosis of rotating machinery based on multi-class support vector machines,” *Journal of Mechanical Science and Technology*, vol. 19, no. 3, pp. 846–859, 2005.
- [6] A. Widodo and B. S. Yang, “Support vector machine in machine condition monitoring and fault diagnosis,” *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2560–2574, 2007.
- [7] E. Ortiz and V. Syrmos, “Support vector machines and wavelet packet analysis for fault detection and identification,” in *Proceedings of IEEE International Conference on Neural Networks (IJCNN '06)*, pp. 3449–3456, Vancouver, Canada, July 2006.
- [8] M. Ge, R. Du, G. Zhang, and Y. Xu, “Fault diagnosis using support vector machine with an application in sheet metal stamping operations,” *Mechanical Systems and Signal Processing*, vol. 18, no. 1, pp. 143–159, 2004.
- [9] L. Y. Lei and Q. Zhang, “Relevance vector machine based bearing fault diagnosis,” in *Proceedings of the International Conference on Machine Learning and Cybernetics*, vol. 1–7, pp. 3492–3496, August 2006.
- [10] C. He, Y. Li, Y. Huang et al., “Relevance vector machine based gear fault detection,” in *Proceedings of the Chinese Conference on Pattern Recognition (CCPR '09)*, pp. 1–5, November 2009.
- [11] A. Widodo, B.-S. Yang, E. Y. Kim, A. C. C. Tan, and J. Mathew, “Fault diagnosis of low speed bearing based on acoustic emission signal and multi-class relevance vector machine,” *Nondestructive Testing and Evaluation*, vol. 24, no. 4, pp. 313–328, 2009.

- [12] Y. L. Zhu, F. Wang, and L. Q. Geng, "Transformer fault diagnosis based on naive bayesian classifier and SVR," in *Proceedings of IEEE Region 10 Conference (TENCON '06)*, pp. 1–4, Hong Kong, China, November 2006.
- [13] F. González, D. Dasgupta, and R. Kozma, "Combining negative selection and classification techniques for anomaly detection," in *Proceedings of the Congress on Evolutionary Computation (CEC '02)*, pp. 705–710, Hawaii, Hawaii, USA, May 2002.
- [14] F. González, D. Dasgupta, and L. Fernando Niño, "A randomized real-valued negative selection algorithm," in *Proceedings of the 2nd International Conference on Artificial Immune System*, pp. 261–272, Edinburgh, UK, September 2003.
- [15] M. E. Tipping and A. C. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *Proceedings of the 9th International Workshop on Artificial Intelligence and Statistics*, 2003.
- [16] J. F. Zhang, Q. H. Yan, Y. L. Zhang, and Z. C. Huang, "Novel fault class detection based on novelty detection methods," in *Intelligent Computing in Signal Processing and Pattern Recognition*, vol. 345 of *Lecture Notes in Control and Information Sciences*, pp. 982–987, 2006.
- [17] D. M. J. Tax, A. Ypma, and R. P. W. Duin, "Pump failure detection using support vector data descriptions," in *Advances in Intelligent Data Analysis*, vol. 1642 of *Lecture Notes in Computer Science*, pp. 415–425, 1999.
- [18] <http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/>.

Research Article

Autonomous Navigation of a Surveillance Robot in Harsh Outdoor Road Environments

**Youjin Shin,¹ Donghyeon Kim,¹ Hyunsuk Lee,¹
Jooyoung Park,² and Woojin Chung¹**

¹ School of Mechanical Engineering, Korea University, Seoul, Republic of Korea

² Department of Control and Instrumentation, Korea University, Seoul, Republic of Korea

Correspondence should be addressed to Woojin Chung; smartrobot@korea.ac.kr

Received 11 April 2013; Revised 18 July 2013; Accepted 18 July 2013

Academic Editor: Yixiang Huang

Copyright © 2013 Youjin Shin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper deals with the autonomous navigation problem of a mobile robot in outdoor road environments. The target application is surveillance in petroleum storage bases. Although there have been remarkable technological achievements recently in the area of outdoor navigation, robotic systems are still expensive due to a large number of high cost sensors. This paper proposes the reliable extraction algorithm of traversable regions using a single onboard Laser Range Finder (LRF) in outdoor road environments. The traversable regions are derived from the classifications of the road surfaces, curbs, and obstacles. The proposed scheme was experimentally tested in success. Since there are many potential applications that require autonomous service robots to move in semistructured road environments, the proposed scheme can be widely used as a low-cost practical solution.

1. Introduction

It is difficult to find reliable solutions for automated diagnosis of mechanical structures in harsh outdoor environment. Figure 1 is a petroleum base that is a typical example of harsh environment. It is common that a petroleum base is close to harbor in order to allow easy access to oil tankers. Maintenance by human operator is difficult because of the risk of explosion and huge workspace. Surveillance for security purpose is also required. Therefore, mobile robotic systems are advantageous. Direct access by a robot with leakage sensors and cameras is essential in practical applications.

The aim of this work is to provide a practical low-cost robotic surveillance solution in outdoor paved road environment. The road in a petroleum base is a restricted area without many dynamic obstacles. Therefore, environmental uncertainty is relatively low. However, a reliable day and night operation is required. The robot should be able to survive in various weather conditions.

Recently, the application area of mobile robots has been expanding [1–4]. Especially, outdoor navigation of mobile

robots has received much attention, fostering a large number of significant technological achievements [5, 6]. Typical examples are the automated robotic vehicles shown in DARPA challenge [7, 8]. Furthermore, some successful technologies can be found, for example, in [9, 10]. These technologies, however, are based on multiple high costs sensing equipment as solutions. Expensive sensing equipment is required to overcome a variety of uncertainties in unstructured outdoor environments. However, the high cost of these technologies limits the practical applications of robotic systems using these technologies.

This paper proposes a reliable extraction algorithm of traversable regions using a single onboard Laser Range Finder (LRF) in outdoor road environments. The traversable regions are derived from the classifications of the road surfaces, curbs, and obstacles.

Detection of road curbs or lanes has been studied. Vision systems are the common approach for road-lane detection. Wang's method [11], Vector Lane Concept [12], and Wu's method [13] are successful vision-based methodologies. However, the major drawback of vision systems is that



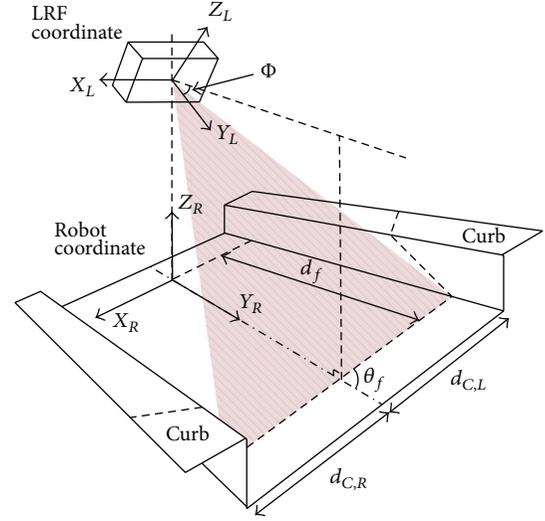
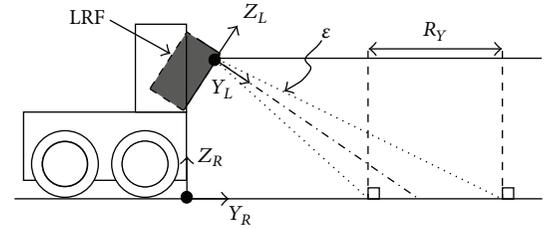
FIGURE 1: Target environment.

TABLE 1: Nomenclature.

Symbols	Meaning
X_L, Y_L, Z_L	LRF local coordinate.
X_R, Y_R, Z_R	Robot local coordinate
Φ	Nominal tilt angle of a LRF
ε	Tolerance of pitch angle
θ_f	Angle between the detected road surface points and Y_R -axis
θ_c	Angle between the detected curb and Y_R -axis
d_f	Horizontal distance from the LRF origin to the road surface points
d_c	Gap between two parallel curbs. Road width.
$d_{C,R}$	Lateral margin from the LRF origin to the right curb
$d_{C,L}$	Lateral margin from the LRF origin to the left curb
X_{pass}	Detected traversable region in x direction
Y_{pass}	Detected traversable region in y direction
R_Y	Expected y range of the road surface points

they have difficulty when confronted with bad weather or illumination change.

LRFs have been widely used. Rho et al. [15, 16] proposed an algorithm for detecting curbs by extracting the lines of a road surface using Hough transforms. Although the algorithm works well in many cases, the curbs and road surfaces have to be sufficiently flat because of the limitation of the Hough transform. In [17], an algorithm is presented for detecting the lines of curbs for some classes of environment using the Extended Kalman Filter (EKF). Kodagoda et al. [18–20] proposed a reliable curb detection algorithm based on the EKF. The algorithm was experimentally verified under a dynamic obstacle environment. LRF data have been obtained in 3D and used to detect the traversable area of the road [21]. However, traversable regions have been commonly detected using vision sensors. Nefian and Bradski [22] proposed a method of image segmentation based on the Bayesian network. An algorithm for detecting a traversable area using a homography segmentation has been presented [23]. Vision sensors, however, have difficulties in coping with bad weather

FIGURE 2: Definition of coordinates and variables. An LRF is mounted at tilt angle Φ . Dotted line represents the nominal road surface. The shaded zone corresponds to the traversable region.FIGURE 3: Expected region of the road surface R_Y .

and illumination change. In [24], a traversable road extraction method was presented for autonomous robot navigation in outdoor cluttered pedestrian walkways. Multiple LRFs are required for implementation.

In this paper, we propose an algorithm for recognition of curbs, the road surfaces, and obstacles using a single LRF in various road conditions. At least one LRF is already installed for obstacle detection, so the proposed algorithm can be applied without additional costs. The traversable region is determined by the locations of curbs and obstacles in the robot's local coordinate. The robot can move safely if traversable region can be detected successfully, even when the localization accuracy is poor. The underlying idea of this study is to utilize the accurate range data of the LRF. Vision systems are excluded to guarantee reliable operation regardless of weather conditions. This paper presents a novel approach to the reliable recognition of road features. Several attributes are defined as geometric features. Experimental attribute data are collected, and the recognition algorithm is inductively established by using the Principal Component Analysis (PCA). The PCA in [25] provides the mathematical framework of combining multiple geometric attributes.

The rest of the paper is organized as follows. In Section 2, the traversable region detection algorithm is presented. Section 3 presents the experimental results of an outdoor

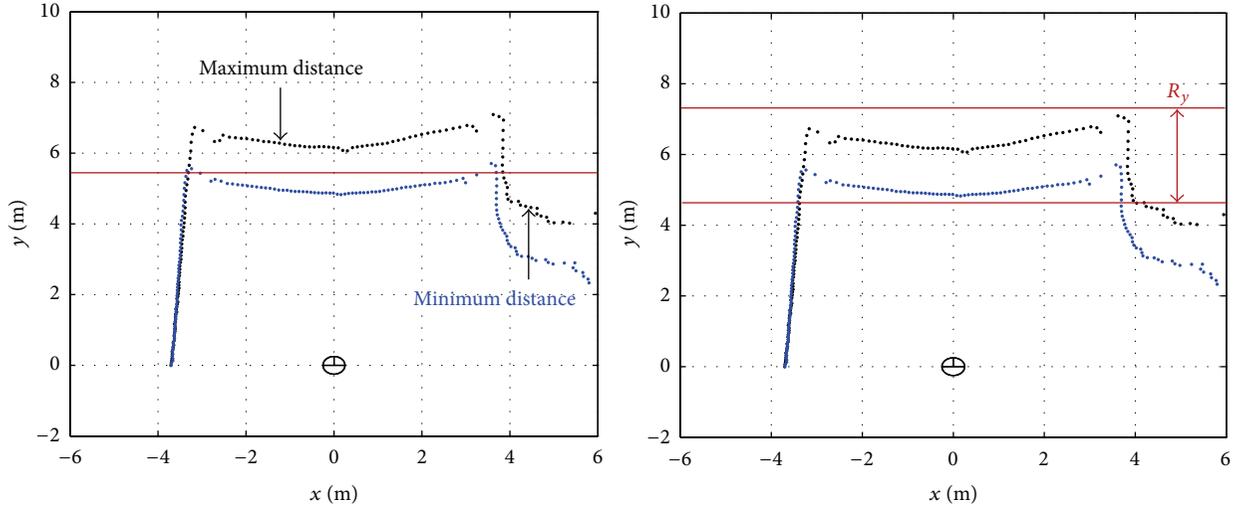


FIGURE 4: LRF data of the road surface and expected region of the road surface R_Y .

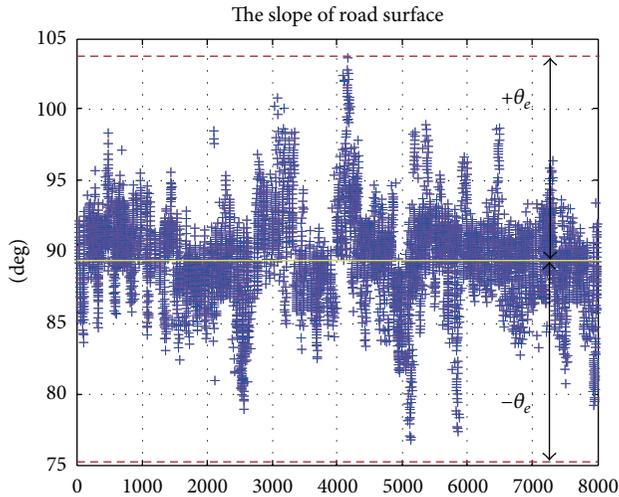


FIGURE 5: Slope (\circ) of the road surface and tolerance (θ_e) of the slope in robot coordinates.

mobile robot in the petroleum storage base environment. Some concluding remarks are presented in Section 4.

2. Traversable Region Detection Algorithm

2.1. System Configuration. Figure 2 presents the coordinate frames and the mounting configuration of the LRF. The fixed LRF looks down on the road at a small tilt angle Φ . The scanning range is assumed to be 180° . Table 1 shows the nomenclature. Variables are described with respect to the robot coordinate. The traversable region corresponds to the road surface between the right and left curbs excluding obstacles.

The road surface points form a straight line if the ground is flat. However, more often, the range images of the road surface show generally curves because most of the road

surfaces are convex. The LRF measurements are affected by the vibration of the robot during motion. Curbs are typically two parallel lines in the range image. However, curb images become distorted for a curved road. There are many different shapes of curbs. In addition, there are no curbs at branches or intersections. Therefore, various uncertainties of LRF measurements should be considered.

The proposed algorithm is composed of three steps. The first step is to determine the pose of the robot with respect to the road surface points (road feature detection). The second step is to extract curbs (curb detection). The third step is to extract the traversable road region with consideration to obstacles (traversable road-region detection).

2.2. Road Feature Detection. The expected region of the road surface is presented in Figure 3. This is derived from the tolerance of the LRF pitch angle. Distribution of LRF data varies with the Y_L -axis as the robot moves along the road. In practice, these variations are caused by the unevenness of the road surface. In order to obtain the tolerance of the pitch angle, an experiment was conducted in the target environment shown in Figure 1. LRF data were accumulated while the robot was traveling. Figure 4 shows the selected LRF data which has minimum and maximum distances to the robot position. By using this information, we can assume the expected road region of the road surface, R_Y , as shown in right Figure 4. The tolerance ε is derived from the expected road region of the road surface, R_Y , as indicated in Figure 3. Therefore, the tolerance ε is determined as 2° .

In order to identify road surface points, the LRF data in the expected region are selected first. Then, consecutive points are clustered. If the robot moves forward on a straight and flat road, the cluster of the road surface points becomes a line. The slope of the road surface line is parallel with the X_R -axis in the ideal case. In practice, the slope of the point cluster may have some variations because of convex road surfaces or robot heading errors. The tolerance errors of heading angle can be determined by experiments in target environment.

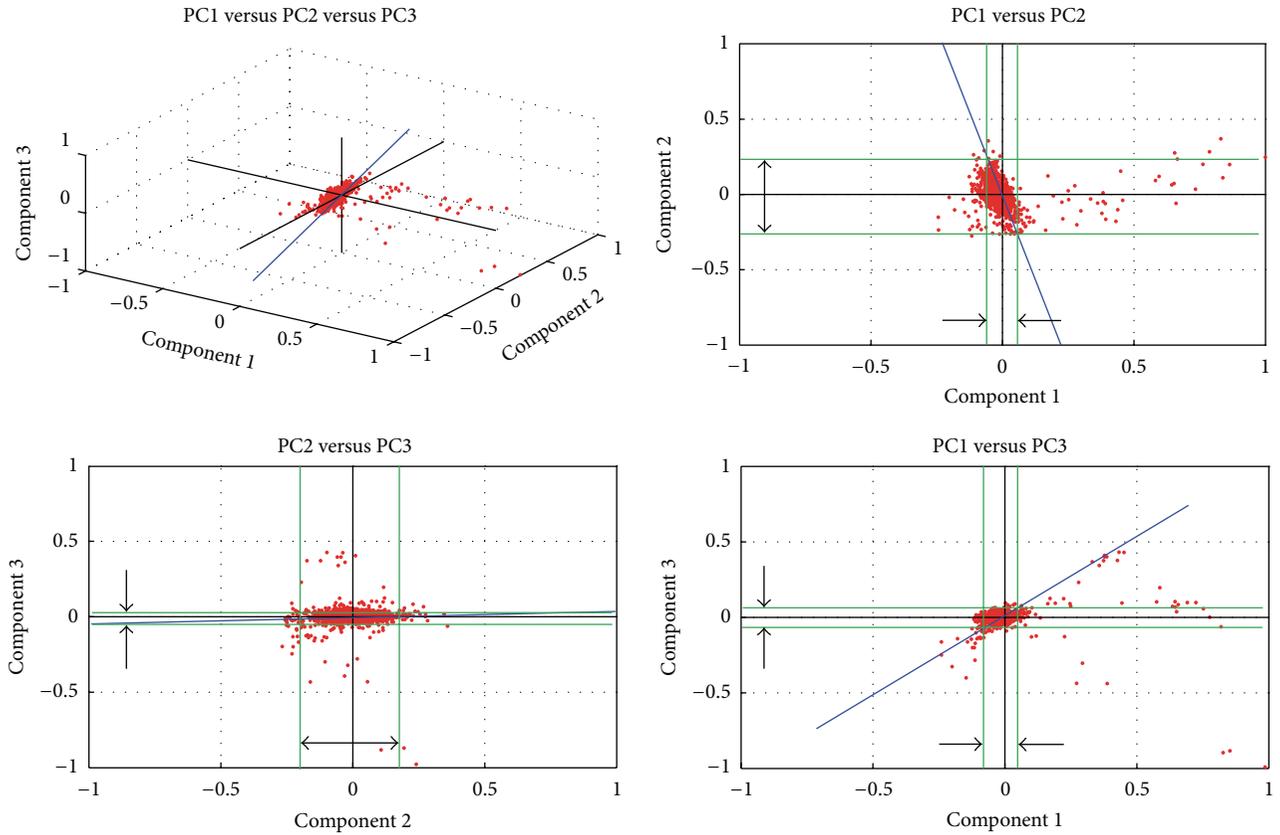


FIGURE 6: Selected principal axis in principal component space.

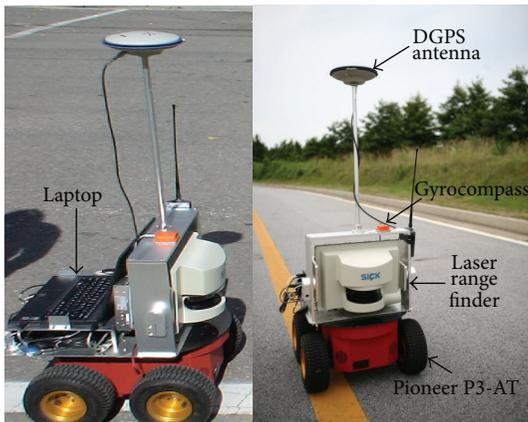


FIGURE 7: Robot platform.

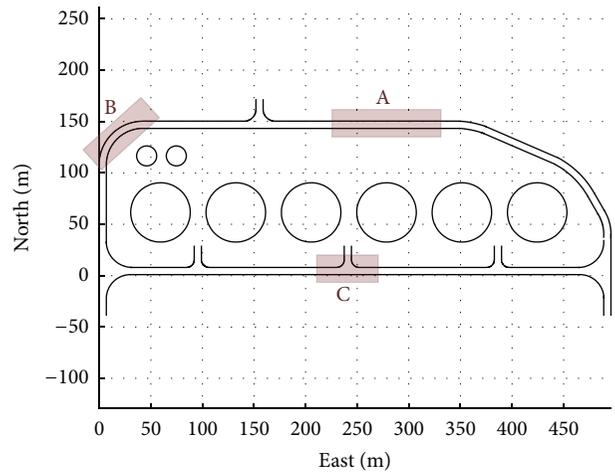


FIGURE 8: Target road environment of various road conditions, and a sector of the given area that is shown in Figure 1.

Figure 5 shows the slope of the road surface with 8000 LRF data.

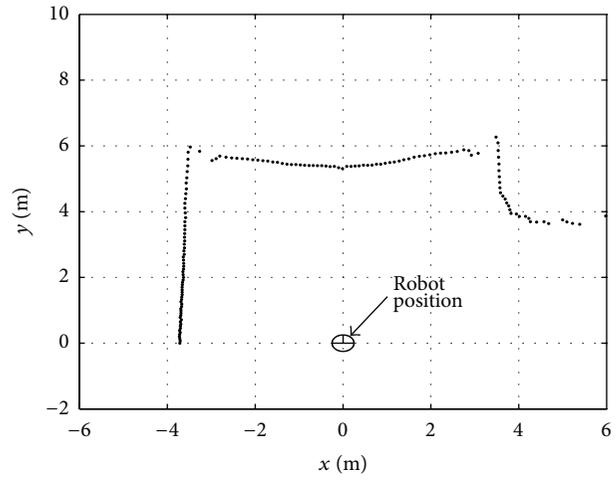
In Figure 5, the yellow line represents the average slope of the road surface, which is 89.6° close to 90° . Therefore, it can be concluded that the slope of the road surface is parallel to the X_R -axis. The red dotted line indicates the upper and lower bounds of the slope of the road surface. Therefore, the tolerance $\pm\theta_e$ is determined as 25° by the difference between the upper and lower bounds. Finally, the extracted clusters

of the road surface points form the road surface. The road feature detection algorithm can be summarized as shown in Algorithm 1.

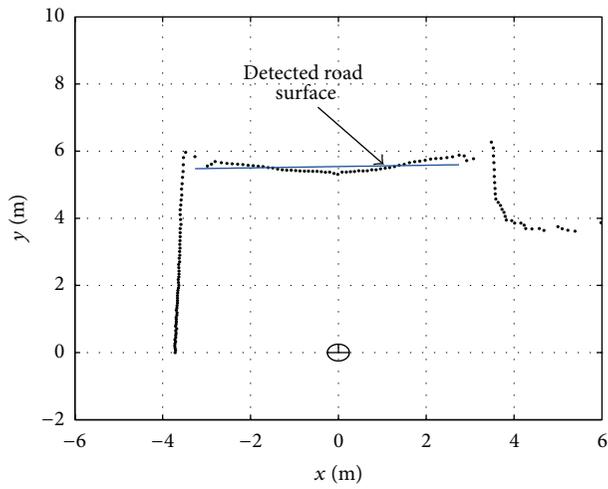
From the computed L_f in Algorithm 1, the road surface can be represented as one line using the conventional least square method. Angle θ_f is the angle difference between the Y_R -axis and the derived line. The horizontal distance from the



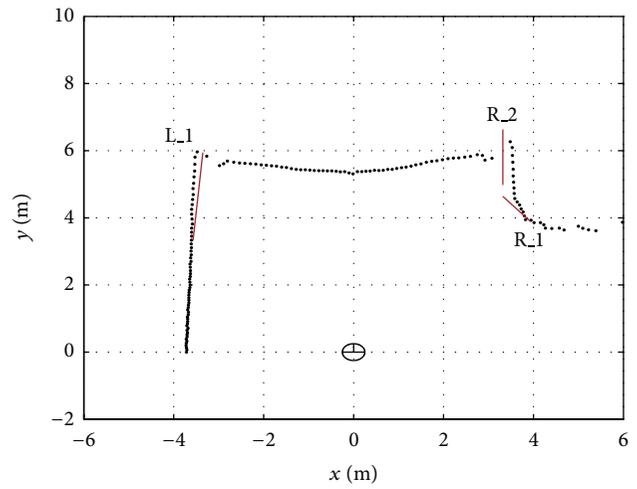
(a)



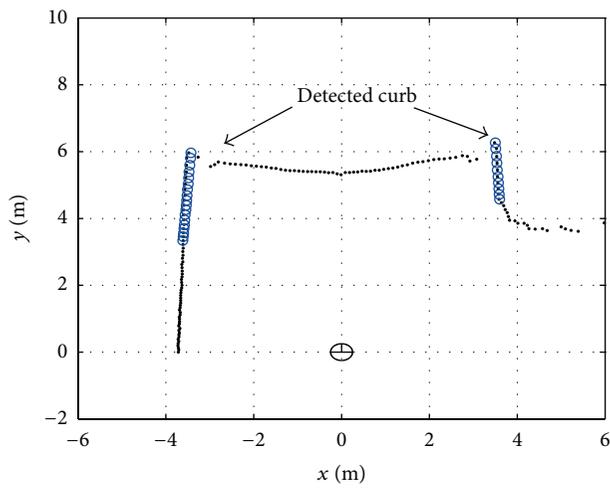
(b)



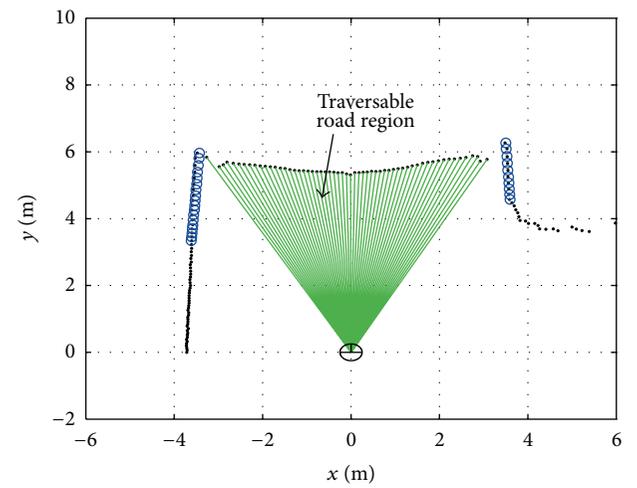
(c)



(d)



(e)



(f)

FIGURE 9: (a) Experimental environment; (b) LRF data; (c) detected road surface; (d) candidate curbs on the left and right; (e) detected curbs; (f) detected traversable road region.

```

(1)  $L_{X,k} \leftarrow x$  coordinate of the LRF measurement
(2)  $L_{Y,k} \leftarrow y$  coordinate of the LRF measurement
(3)  $N \leftarrow$  Count of the LRF measurement
(4) for  $k \leftarrow 1$  to  $N$ 
(5)   if  $L_{Y,k} \in R_Y$ 
(6)     then  $L_r \leftarrow$  Candidate for the road surface
(7)        $\theta_{L,r} \leftarrow$  Slope of candidate for the road surface
(8)        $M \leftarrow$  Count of candidate for the road surface
(9)   for  $l \leftarrow 1$  to  $M$ 
(10)    if  $\|\theta_{L,r}\| \leq \theta_e$ 
(11)     then  $L_f \leftarrow$  save  $L_r$  as the road surface
(12) return ( $L_f$ )

```

ALGORITHM 1: Computing the road feature properties.

```

(1) Select the first  $N$  points, construct a line by the use of a least square method in [14].
(2) Construct a line by employing the next  $N$  points.
(3) Compute Euclidean distance of two lines.
(4) If Euclidean distance is smaller than  $d_m$ , merge the lines and re-compute the line
    parameters, continue (go to 2).
(5) Otherwise, return the line.
(6) Continue with the next  $N$  points, go to 2.

```

ALGORITHM 2: Line segmentation and clustering.

LRF origin to the road surface points is represented by d_f . The average value of y coordinates of the derived line.

2.3. Curb Detection Using the Principal Component Analysis (PCA). At first, by combining consecutive N LRF data, each line segment is constructed. Then, separate line segments are merged into one line when multiple line segments can be considered to be a single line. In order to evaluate the closeness of line segments, Euclidean distance is applied as introduced in [14]. Two line segments $x_i = [\alpha_i \ r_i]^T$ and $x_j = [\alpha_j \ r_j]^T$ in polar coordinate are merged when the Euclidean distance is smaller than threshold d_m . The validation process can be written as the following equation:

$$(x_i - x_j)^T (x_i - x_j) \leq d_m. \quad (1)$$

The procedure of line segmentation and clustering can be summarized as in Algorithm 2.

Extracted line clusters are considered as curb candidates.

We define several attributes from geometric features in order to extract curbs. The defined attributes are shown as follows.

- (Att.1) The angular difference between the curb orientation θ_c and road-surface slope θ_f . This is close to 90° .
- (Att.2) The gap between the road surface horizon distance d_f and the maximum y of the curb points. This is close to 0.
- (Att.3) The angular difference between the left and the right curb orientations. This is close to 0.

(Att.4) The range difference between the road width and the gap of two curbs. This is close to 0. It is assumed that the road width is known.

It is commonly assumed that the robot's heading is parallel with the curb orientation. However, this assumption becomes invalid when there is an inclination of the road surface. (Att.1) extracts the candidate lines of both right and left curbs. The candidate lines are registered as curbs when the candidates satisfy the validation gate, which is a function of (Att.2), (Att.3), and (Att.4). About (Att.2), (Att.3), and (Att.4), we exploit the Principal Component Analysis (PCA) in order to define the validation gate. The PCA provides a mathematical procedure to combine three attributes as a linear combination. Therefore, the range of candidate curbs can be derived from these attributes with different dimensions. At last, prospective curbs can be selected from the candidate curbs within this range.

The validation gate is a scalar, and it is defined as a combination of attributes. There exists a unit vector u that can maximally represent all components. The vector u is defined as the principal axis that reduces the dimensions of the attribute space. In order to reduce the dimensions of attributes to k (where $k < n$), each attribute should be normalized first:

$$\mu = \frac{1}{m} \sum_{i=1}^m x^{(i)} \quad x^{(i)} - \mu \quad (2)$$

$$\sigma_j^2 = \frac{1}{m} \sum_{i=1}^m (x_j^{(i)})^2, \quad x_j^{(i)} = \frac{x_j^{(i)} - \mu}{\sigma_j}. \quad (3)$$

```

(1)  $L_{X,k} \leftarrow x$  coordinate of the LRF measurement
(2)  $L_{Y,k} \leftarrow y$  coordinate of the LRF measurement
(3)  $d_f \leftarrow$  Horizontal distance to road surface
(4)  $N \leftarrow$  Count of the LRF measurement
(5) for  $k \leftarrow 1$  to  $N$ 
(6)   if  $L_{X,k} \leq d_{C,R}$  and  $L_{X,k} \geq -d_{C,L}$ 
(7)     if  $L_{Y,k} \leq d_f$ 
(8)       then  $L_{\text{obstacle}} \leftarrow$  Obstacle data set
(9)     else  $L_{\text{pass}} \leftarrow$  Add  $L_{X,k}$  and  $L_{Y,k}$  to the traversable
        region, data set  $(X_{\text{pass}}, Y_{\text{pass}})$ 
(10) return  $(L_{\text{pass}})$ 

```

ALGORITHM 3: Detect the traversable road region.

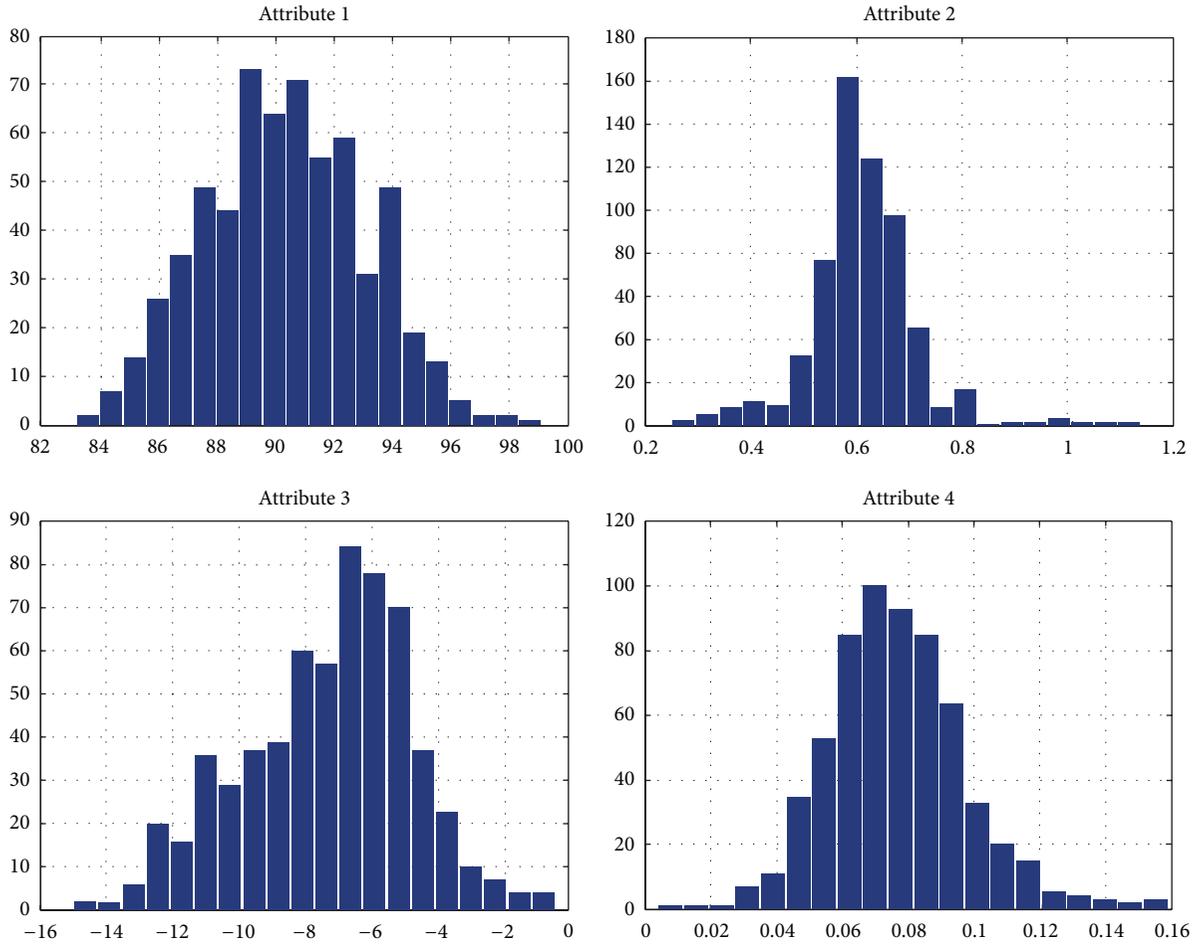


FIGURE 10: PCA attributes in straight road environment.

Means are shifted to 0 by (2). Equation (3) is the rescaling process. The unit vector u can be derived from the following equation:

$$\begin{aligned} \frac{1}{m} \sum_{j=1}^m (x^{(j)T} u)^2 &= \frac{1}{m} \sum_{i=1}^m u^T x^{(i)} x^{(i)T} u \\ &= u^T \left(\frac{1}{m} \sum_{j=1}^m x^{(j)} x^{(j)T} \right) u. \end{aligned} \quad (4)$$

The unit vector u is defined from the eigenvector and the eigenvalue of the covariance matrix $\Sigma = (1/m) \sum_{j=1}^m x^{(j)} x^{(j)T}$ by the following equation:

$$\begin{aligned} u^T \Sigma u &= u^T \lambda u = \lambda u^T u = \lambda \\ \Sigma u &= \lambda u. \end{aligned} \quad (5)$$

In (5), u is the eigenvector of the covariance matrix Σ , and λ is the eigenvalue. The principal axis corresponds to

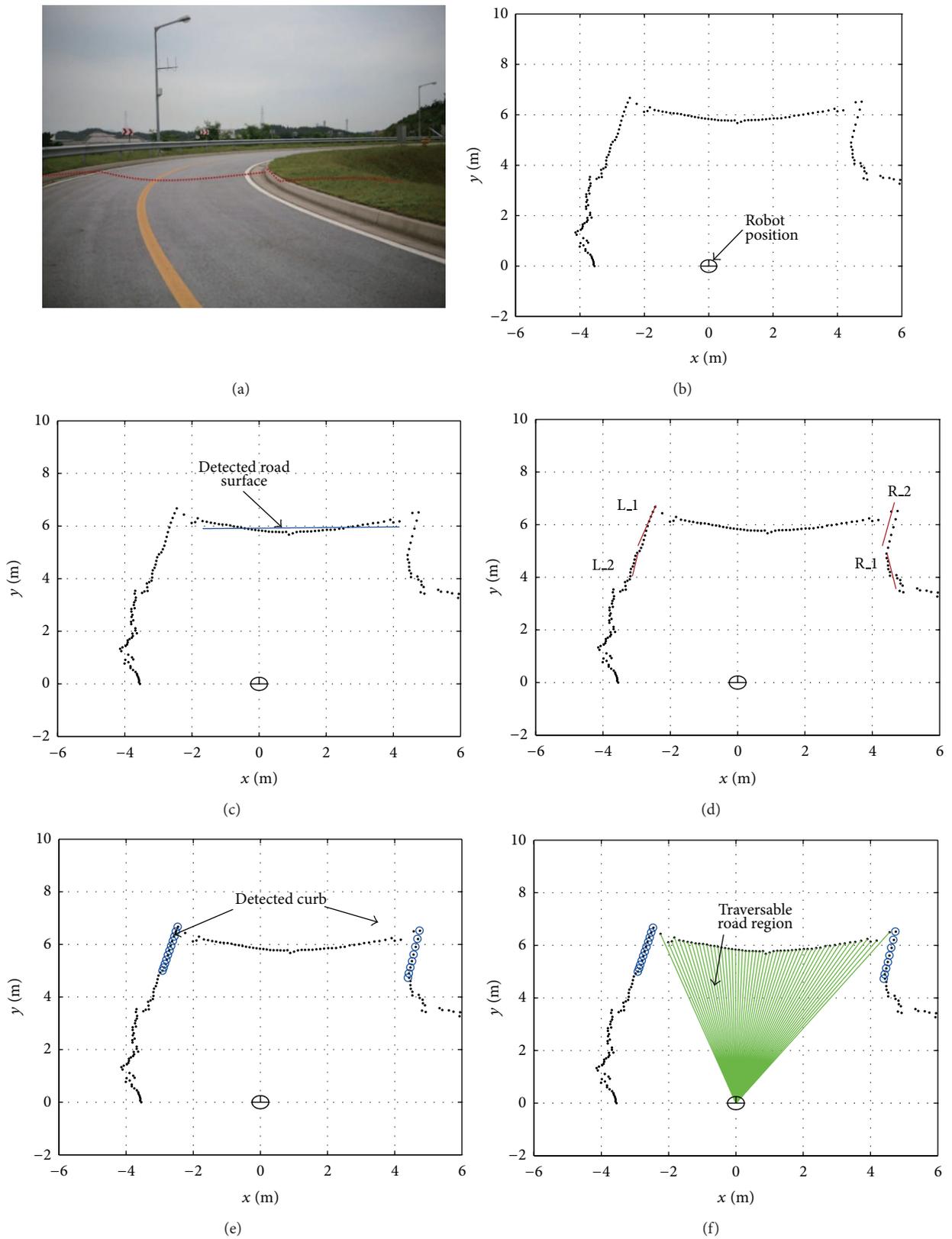


FIGURE 11: (a) Experimental environment; (b) LRF data; (c) detected road surface; (d) candidate curbs on the left and right; (e) detected curbs; (f) detected traversable road region.

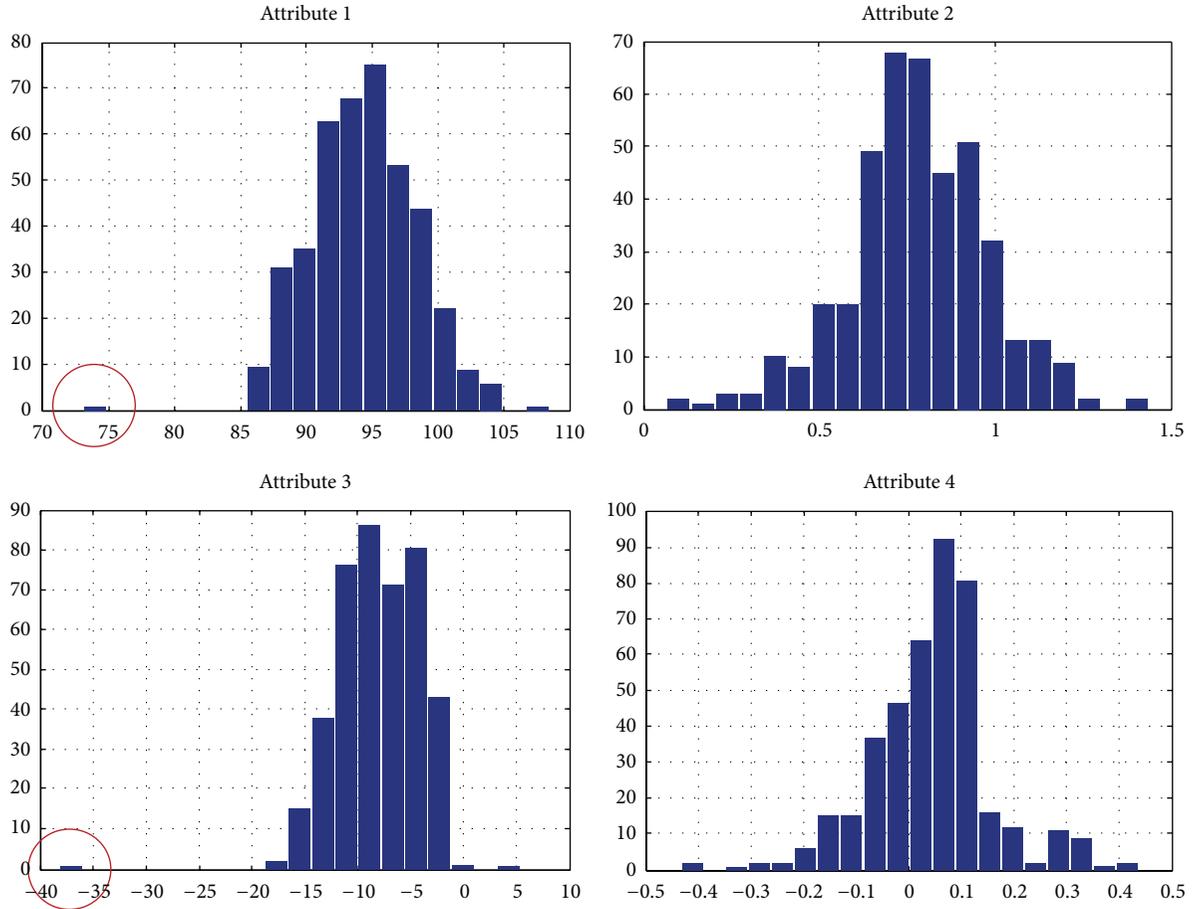


FIGURE 12: PCA attributes in straight road environment.

the eigenvector with the largest eigenvalue. Figure 6 shows the computational result of the PCA algorithm. The blue line represents the selected principal axis, and the green dotted line is the range of the principal axis which can be selected as curbs. The red dots represent the principal component.

The experimental measurements are obtained from a test run in the environment in Figure 1. From LRF measurements, 999 candidate lines were extracted. The principal axis u is shown in the principal component space. The validation gate Vg can be defined by the following equation:

$$Vg = u_1 \cdot X_1 + u_2 \cdot X_2 + u_3 \cdot X_3. \quad (6)$$

A reliable curb recognition algorithm can be obtained by appropriate selection of the range of Vg results. X_1 , X_2 , and X_3 correspond to the principal components of attributes (Att.2), (Att.3), and (Att.4), respectively. The range of the principal component is determined by comparing the collected data of the curbs with the expected values. The derived range is used for obtaining the acceptable ranges of attributes from the experimental results and represented as the green line in Figure 6. Table 2 summarizes the acceptable ranges of attributes from the experimental results.

2.4. Traversable Region and Obstacle Detection. LRF data within the range between the left curb ($d_{C,L}$) and the right

TABLE 2: Acceptable range of curb attributes.

	Att.2 [m]	Att.3 [deg]	Att.4 [m]
Max.	1.5	20	0.5
Min.	0.1	-20	-0.5

TABLE 3: Selection of curbs (values resulting from (6)).

	R.1	R.2
L.1	4.29	0.86

TABLE 4: Selection of curbs (values resulting from (6)).

	R.1	R.2
L.1	1.79	1.75
L.2	1.92	1.88

TABLE 5: Selection of curbs (values resulting from (6)).

	L.1	L.2
R.1	1.19	1.37

curb ($d_{C,R}$) are determined as the expected region. This region is used to detect obstacles and traversable area. If the

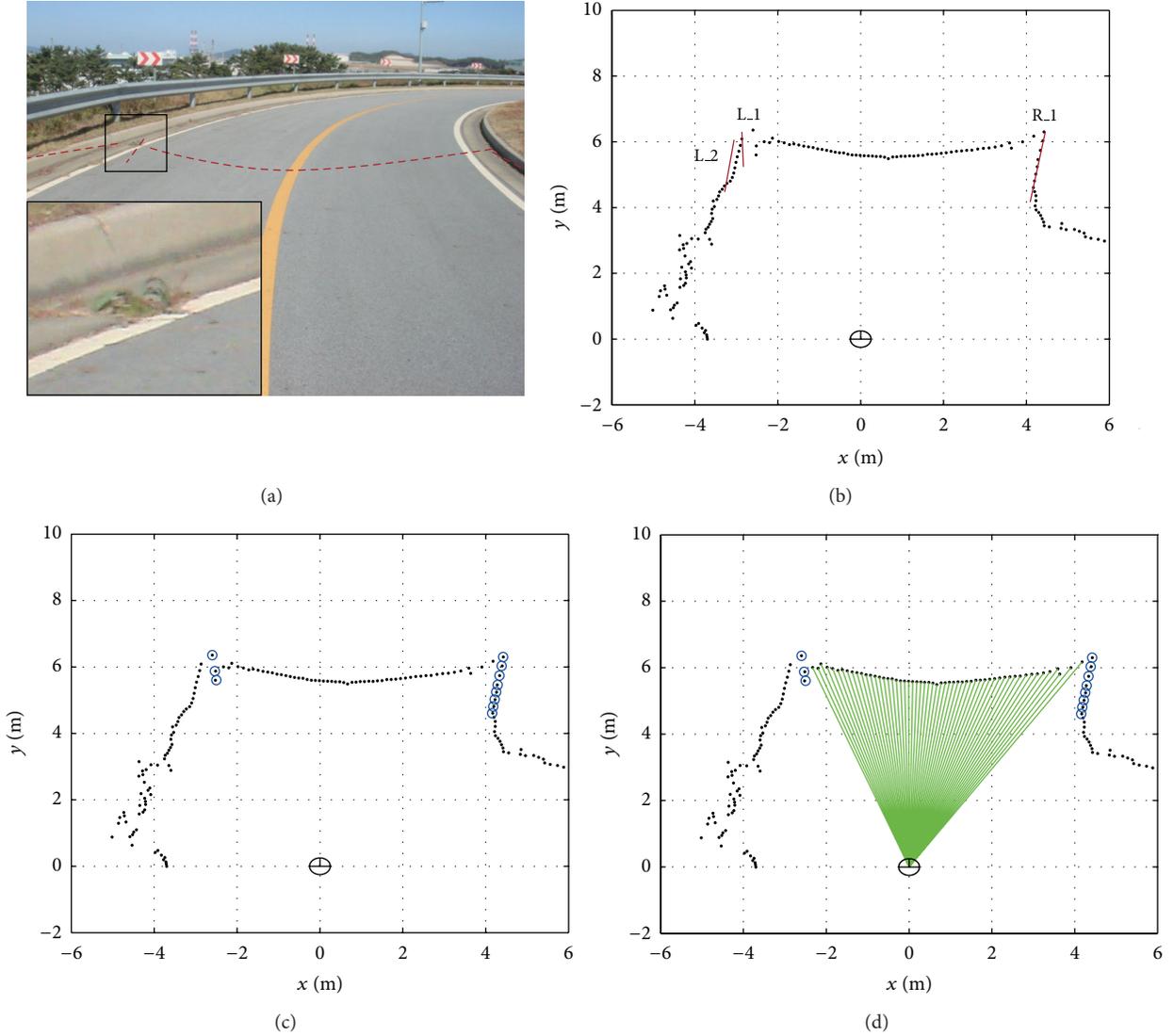


FIGURE 13: (a) Road environment with weeds; (b) candidate curbs on the left and right; (c) detected weeds as the curb; (d) detected traversable road region.

TABLE 6: Expected obstacle region.

	Min. value, m	Max. value, m
X range	-3.58	3.56
Y range	0	3.73

TABLE 7: Expected obstacle region.

	Min. value, m	Max. value, m
X range	-3.41	3.71
Y range	0	3.92

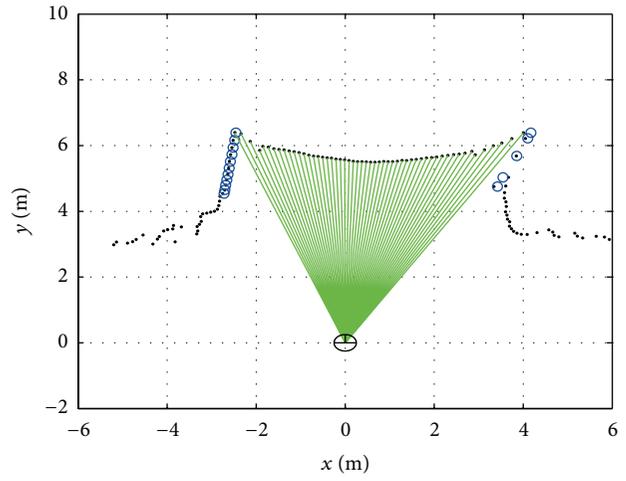
y value of LRF data is lower than d_f , the data represents an obstacle. The expected region without obstacles is considered as a traversable area. The algorithm also enables the robot to calculate the traversable area regardless of curbs by assuming

the locations of the curbs when the value of the road's width is given. The algorithm for the condition in which both left and right curbs exist is as in Algorithm 3.

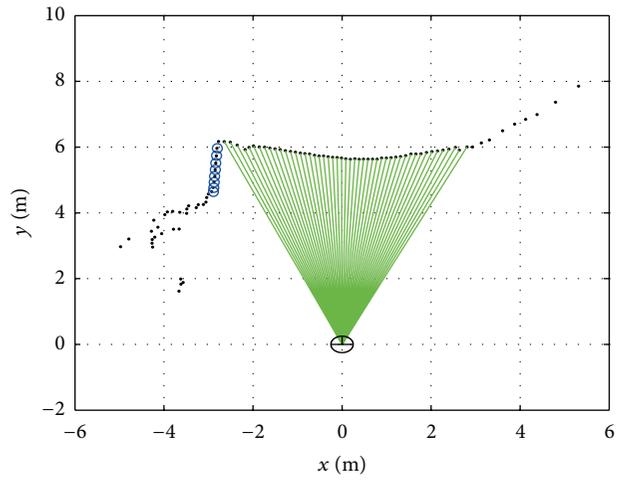
Detected traversable region can guarantee the safety of outdoor navigation regardless of the localization performance since it guides the robot in a safe drive direction in the robot coordinate system. Therefore, this method can also be used for local obstacle avoidance such as VFH [26].

3. Experimental Results

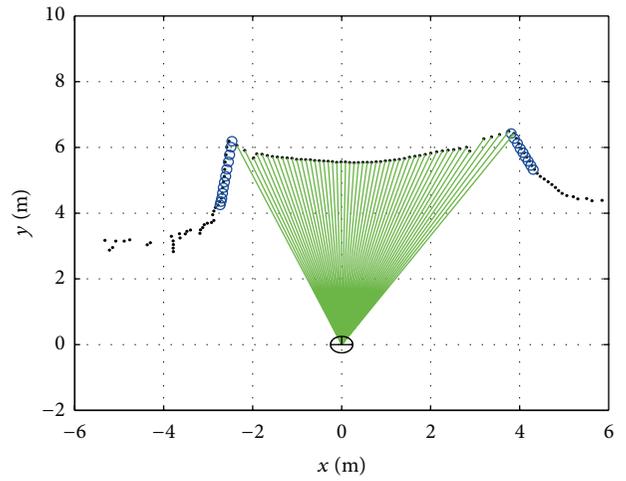
3.1. Experimental Setup. Figure 7 shows the experimental setup. We constructed the navigation system on a P3-AT model of pioneer. It is a four-wheeled robot designed for outdoor use with odometry. A SICK laser sensor (LMS200 in [27]) is mounted on the robot to detect the range around the robot. The maximum range of the LRF is 33 m, and its static



(a)



(b)



(c)

FIGURE 14: Experimental environment and detected curbs and traversable region on a fork in the road in robot coordinates.

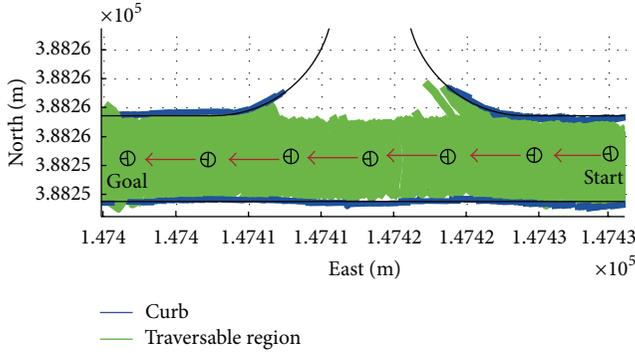


FIGURE 15: Detected curbs and traversable region on a fork in the road in global coordinates.

error is under 0.3 cm. To detect the road surfaces and curbs, the LRF was mounted at a small tilt angle 6.4° . The Novatel DGPS system was attached to localize the robot and the Xsens gyrocompass sensor was used to obtain the direction of the robot. Computations were conducted by a laptop.

The proposed algorithm is applied to the various road environments in petroleum storage bases. Figure 8 presents the setting of the targeted roads in global coordinates. In order to test the robot's performance in various conditions, three different types of roads were selected from the experiment site: a straight road, area A; a curved road, area B; and a forked road, area C. The widths of the roads are between 6 m and 7 m, and the robot traveled in a clockwise direction.

Experiment I was conducted in area A, a straight road with the width of 7 m. While the robot traveled in area A, curbs and traversable area were detected by LRF applying the proposed algorithm. Experiment II was carried out in area B, a curved road with the width of 6 m. Experiment III was conducted on a fork in the road of 6 m in width, area C. In Experiment III, as the robot traveled in a clockwise direction, the curbs on the right side disappeared and thus became undetectable. Experiment IV was carried out in area A with obstacles. The results of robot navigation in these environments are presented as follows.

3.2. Experiment I. As shown in Figure 8 in experiment I, the robot traveled approximately 100 m following the straight center line in area A of the straight road environment. Curbs existed on both sides of the road. While driving, 600 LRF data were acquired. Figure 9 shows the process of the algorithm with the LRF data. In Figure 9(a), the environment setting is presented, and the red dotted line represents the positions of the obtained LRF data. These LRF data are shown in robot coordinate in Figure 9(b). Figure 9(c) shows the result of detected road surface represented as a blue line. Figure 9(d) shows the candidate curbs on both sides of the road. Left and right sides are determined by the heading of the robot, Y_R axis. There is one candidate curb (L_1) on the left and two candidate curbs (R_1, R_2) on the right. With the attributes of each candidate, the value of V_g is derived from the PCA method. The candidate curbs with the lowest value of V_g are selected as curbs. Therefore, L_1 and R_2 are detected as

curbs as shown in Figure 9(e). The calculated values of V_g are indicated in Table 3. Figure 9(f) shows the result of detected traversable road region.

The histograms illustrating the PCA attributes of detected curbs from the results of experiment I are shown in Figure 10. The average value of (Att_1) is 90.07° . This result meets the environmental condition that the angular difference between curb orientation and road-surface slope is almost perpendicular. Other attributes are also in acceptable ranges of curb attributes as shown in Table 2. Distribution of (Att_2) is between 0.2 m and 1.2 m, which is within the acceptable range of (Att_2), between 0 m and 1.5 m. The average value of (Att_3) is -7.5° , and the distribution is between -7.5° and 0° within the acceptable range of (Att_3). The average value of (Att_4) is 0.07 m, and the distribution is between 0 m and 0.16 m within the acceptable range of (Att_4). These results prove that experiment I was conducted successfully for detecting curbs in area A. Also, the processing time of the algorithm between steps was 76 ms on average.

3.3. Experiment II. In experiment II, the robot traveled in area B of the curved road environment. Curbs exist on both sides of the road. While traveling, 418 LRF data were accumulated. Figure 11 presents the process of the algorithm with the LRF data in experiment II. In Figure 11(a), the experimental environment is presented, and red dotted line represents the positions of obtained LRF data. These LRF data are plotted in robot coordinate as shown in Figure 11(b). Figure 11(c) shows the result of detected road surface as a blue line. In Figure 11(d), candidate curbs can be seen on both sides of the road. There are two candidate curbs (L_1, L_2) on the left and two candidate curbs (R_1, R_2) on the right. With the attributes of each candidate, the value of V_g was derived from PCA method. As shown in Figure 11(e), among all candidate curbs, the curbs with the lowest value of V_g , L_1 and R_2, are selected. The calculated values of V_g are indicated in Table 4. Figure 11(f) shows the result of detected traversable road region.

Figure 12 shows the histograms illustrating the PCA attributes of detected curbs from the results of experiment II. All the attributes except for (Att_1) are within the acceptable ranges of curb attributes as shown in Table 2. Because experiment II was conducted on a curve, the difference in angle between curb orientation and road-surface slope is not clearly perpendicular. Thus, the average value of (Att_1) is 93° . Distribution of (Att_2) is between 0 m and 1.5 m, which are within the acceptable range of (Att_2). The average value of (Att_3) is -12° . This result shows that the difference in angle between the left and the right curb orientations is larger on curved roads than on straight road. The average value of (Att_4) is 0.07 m, and the distribution is between 0 m and 0.16 m in the acceptable range of (Att_4). These results show that experiment II was conducted successfully for detecting curbs in area B.

Figure 13 presents the failure case of curb detection that the robot misclassifies the weeds on the road as curbs. This result can be seen as a red circle in Figure 12. The weeds can be found near the left curb as shown in Figure 13(a). Figure 13(b)

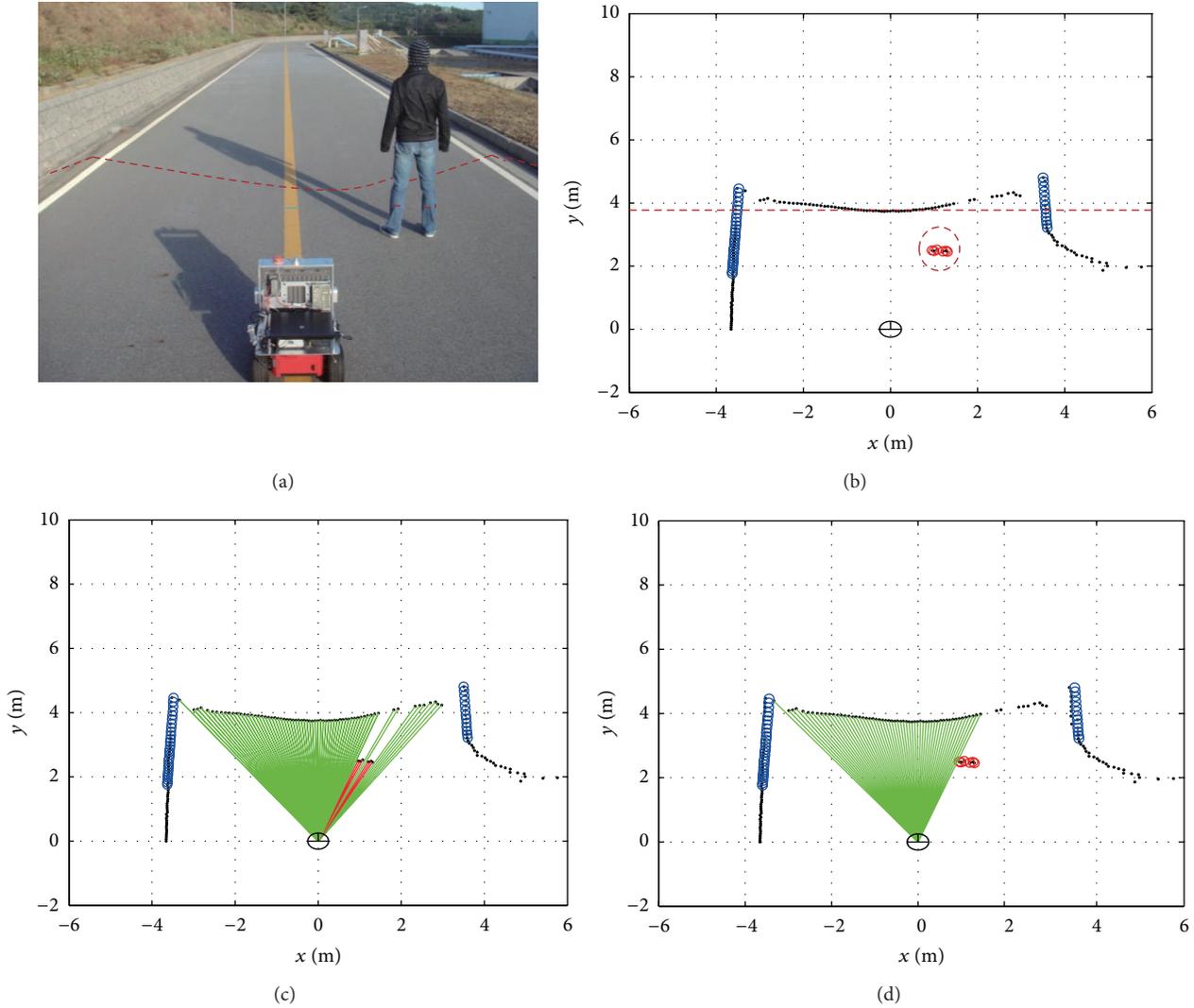


FIGURE 16: (a) Experimental environment; (b) detected curbs and obstacle; (c) candidates of traversable region; (d) detected traversable region.

shows the candidate curbs in this environment. There are two candidate curbs (L₁, L₂) on the left, and there is a candidate curb (R₁) on the right. The calculated values of V_g in (6) are shown in Table 5. Pair (L₁, R₁) is selected as curbs because of the lowest V_g . This is a misclassification because the true curb pair is (L₂, R₁). The number of misclassifications was only one during the movement in area B shown in Figure 8. This result implies the 99.8% success rate (417 success/418 LRF scans). Therefore, partial failure was not a significant concern in practical applications.

3.4. Experiment III. Experiment III was conducted in area C of a fork in the road with the width of 6 m. In this road environment, the robot could not detect curbs on the right side while traveling in a clockwise direction. The robot traveled about 50 m and obtained 351 LRF data. Figure 14(a) presents the LRF data obtained as the robot approached the fork. As shown in Figure 14(b), even when curbs disappeared on the right side of the road while traveling, the robot was able

to detect the traversable region successfully since the width of the road was 6 m, and the information of curbs on the left side was given. Figure 14(c) shows LRF data collected by the robot as it passed by the fork.

Figure 15 shows these results, which were transformed into global coordinates. The blue line in Figure 15 indicates the detected curbs, and the green area presents the detected traversable region. Localization of the robot was estimated by the EKF algorithm using odometry data, DGPS, and gyroscope. In Figure 15, the location of the robot is plotted every 40 steps, and the detected curbs are plotted every 5 steps. These results show that experiment III was conducted successfully for detecting the traversable region and the curbs on the left side. The processing time of the algorithm between steps was 90 ms on average.

3.5. Experiment IV. In order to verify the algorithm in the obstacle environment, Experiment IV was carried out in area A. Figure 16 shows the experimental results of detected

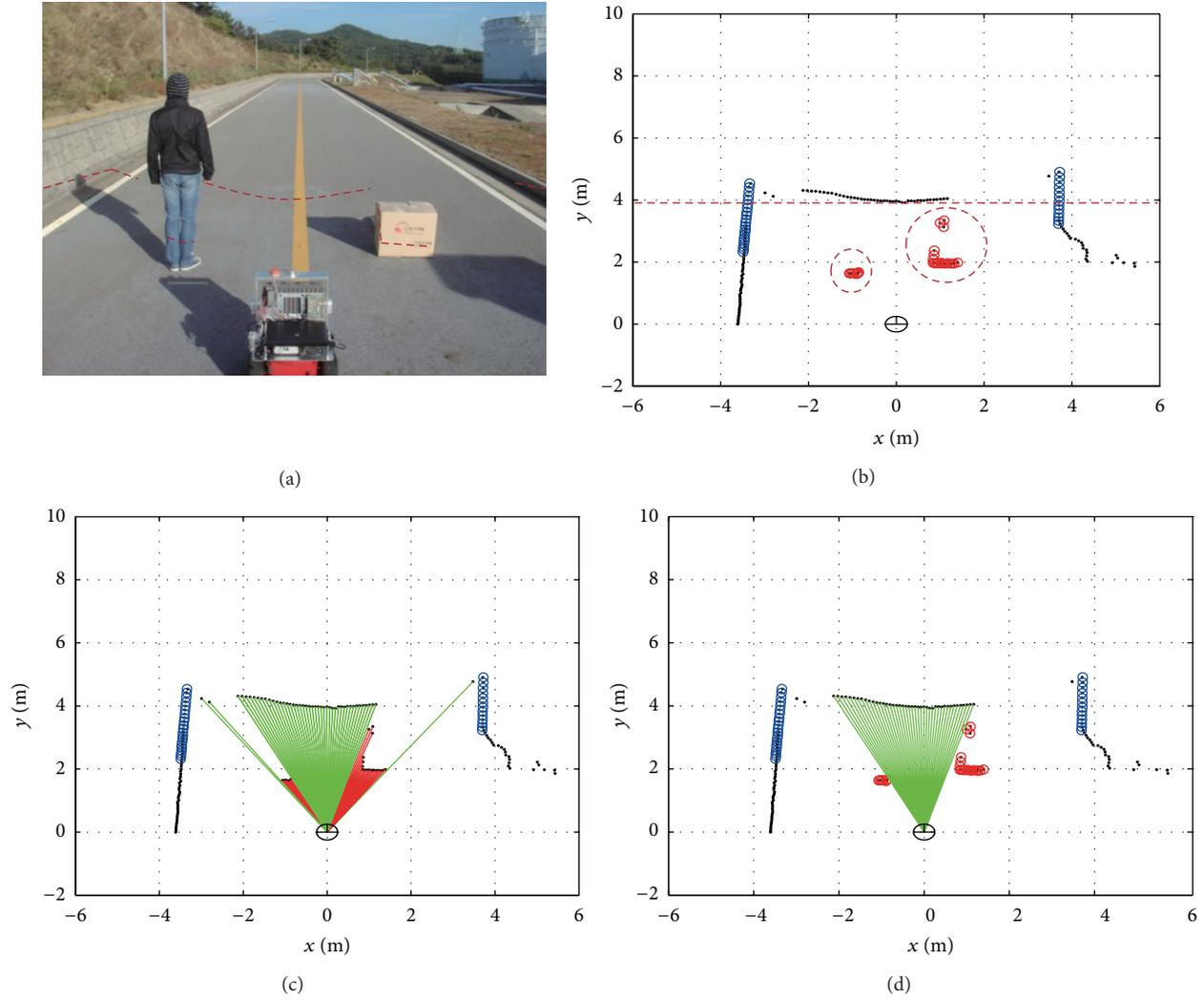


FIGURE 17: (a) Experimental environment; (b) detected curbs and obstacles; (c) candidates of traversable region; (d) detected traversable region.

traversable region when one pedestrian exists. The red dotted line represents the scanning area of LRF. The curbs are represented in blue lines as shown in Figure 16(b). The horizontal distance d_f in Table 1 is shown in the red dashed line. The expected obstacle region is determined by the detected curbs and d_f as indicated in Table 6. The LRF data lying on this region are considered as an obstacle. The red circled data in Figure 16(b) indicate the detected obstacle that corresponds to the pedestrian. From the road surface and the obstacle, the candidates of traversable region are obtained as shown in Figure 16(c). The green regions indicate the candidates of the traversable region, and the red regions show the collision region due to the detected obstacle. Among these candidates of the traversable region, the region with the widest width is finally selected for safety. The finally detected traversable region is shown in Figure 16(d).

Figure 17 shows the experimental results, when one pedestrian and a box exist. The experimental environment is shown in Figure 17(a). The expected obstacle region is

determined by the curbs and the road surface as indicated in Table 7. The LRF data marked by the red circle in Figure 17(b) are considered as obstacles that correspond to a pedestrian and a box, respectively. The candidates of the traversable region are shaded as the green regions. The collision regions due to the obstacles are represented by the red regions. The candidate region that has the widest width is selected as the traversable region.

4. Conclusion

In this paper, we proposed a method that can address the perceptual issues related to outdoor road environments for outdoor navigation. We used a single laser range finder to detect curbs, road surfaces, and obstacles. Hence, this method can be easily adopted for practical applications. By extracting two types of information (on curbs and road surfaces), we increased the reliability of traversable road

detection in relation to the findings of other studies. This method, in combination with the pose estimation method, was verified by experiments for various road environments. This algorithm was computationally efficient and guaranteed the safety of outdoor navigation. Since there are many potential applications of autonomous outdoor mobile robots in semistructured road environments, the proposed method can be widely used as a low-cost practical solution for safe outdoor navigation.

Acknowledgments

This work was supported by the NRF Grant funded by the MEST, Korea (2013-029812). This research was also supported in part by the MKE, Korea, under the Human Resources Development Program supervised by the NIPA (NIPA-2013-H1502-13-1001). The research was also supported in part by the Implementation of Technologies for Identification, Behavior, and Location of Human based on Sensor Network Fusion Program through the MKE (Grant no. 10041629).

References

- [1] B. Kim, J. Park, S. Park, and S. Kang, "Impedance learning for robotic contact tasks using natural actor-critic algorithm," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 40, no. 2, pp. 433–443, 2010.
- [2] S. Kang, K. Komoriya, K. Yokoi, T. Koutoku, B. Kim, and S. Park, "Control of impulsive contact force between mobile manipulator and environment using effective mass and damping controls," *International Journal of Precision Engineering and Manufacturing*, vol. 11, no. 5, pp. 697–704, 2010.
- [3] B. Chu, D. Kim, and D. Hong, "Robotic automation technologies in construction: a review," *International Journal of Precision Engineering and Manufacturing*, vol. 9, no. 3, pp. 85–91, 2008.
- [4] B. Chu, K. Jung, C.-S. Han, and D. Hong, "A survey of climbing robots: locomotion and adhesion," *International Journal of Precision Engineering and Manufacturing*, vol. 11, no. 4, pp. 633–647, 2010.
- [5] Y.-J. Lee and J.-B. Song, "Three-dimensional iterative closest point-based outdoor SLAM using terrain classification," *Intelligent Service Robotics*, vol. 4, no. 2, pp. 147–158, 2011.
- [6] X. Zhang, M. Geimer, P. O. Noack, and L. Grandl, "A semi-autonomous tractor in an intelligent master-slave vehicle system," *Intelligent Service Robotics*, vol. 3, no. 4, pp. 263–269, 2010.
- [7] M. Buehler, K. Iagnemma, and S. Singh, Eds., *The 2005 DARPA Grand Challenge*, Springer, 2007.
- [8] M. Buehler, K. Iagnemma, and S. Singh, Eds., *The DARPA Urban Challenge: Autonomous Vehicles in City Traffic*, Springer, 2010.
- [9] H. Dahlkamp, A. Kaehler, D. Stavens, S. Thrun, and G. Bradski, *Self-Supervised Monocular Road Detection in Desert Terrain*, Robotics Science and Systems, Philadelphia, Pa, USA, 2006.
- [10] S. J. Velat, J. Lee, N. Johnson, and C. D. Crane III, "Vision based vehicle localization for autonomous navigation," in *Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA '07)*, pp. 528–533, June 2007.
- [11] R. Wang, Y. Xu, Libin, and Y. Zhao, "A vision-based road edge detection algorithm," in *Proceedings of the IEEE Intelligent Vehicle Symposium*, vol. 1, pp. 141–147, June 2002.
- [12] Q. B. Truong, B. R. Lee, N. G. Heo, Y. J. Yum, and J. G. Kim, "Lane boundaries detection algorithm using vector lane concept," in *Proceedings of the 10th International Conference on Control, Automation, Robotics and Vision (ICARCV '08)*, pp. 2319–2325, Hanoi, Vietnam, December 2008.
- [13] B.-F. Wu and C.-T. Lin, "Robust lane detection and tracking for driving assistance systems," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC '07)*, pp. 3848–3853, Montréal, Canada, October 2007.
- [14] R. Siegwart and I. R. Nourbakhsh, *Autonomous Mobile Robots*, MIT Press, Cambridge, Mass, USA, 2004.
- [15] C.-W. Roh, S.-H. Kim, M.-J. Kim, S.-C. Kang, and S.-K. Hong, "Development of patrol robot using DGPS and curb detection," *International Journal of Control, Automation and Systems*, vol. 13, pp. 140–146, 2007.
- [16] S.-H. Kim, C.-W. Roh, S.-C. Kang, and M. Park, "Curb detection using laser range finder for reliable outdoor navigation," in *Proceedings of the IEEE International Conference on Advanced Robotics (ICAR '07)*, pp. 930–935, Jeju, Republic of Korea, 2007.
- [17] C. Yu and D. Zhang, "Road curbs detection based on laser radar," in *Proceedings of the 8th International Conference on Signal Processing (ICSP '06)*, Guilin, China, November 2006.
- [18] W. S. Wijesoma, K. R. S. Kodagoda, and A. P. Balasuriya, "Road-boundary detection and tracking using ladar sensing," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 3, pp. 456–464, 2004.
- [19] K. R. S. Kodagoda, W. S. Wijesoma, and A. P. Balasuriya, "Road curb and intersection detection using a 2D LMS," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '02)*, pp. 19–24, Lausanne, Switzerland, October 2002.
- [20] W. S. Wijesoma, K. R. S. Kodagoda, A. P. Balasuriya, and E. K. Teoh, "Road edge and lane boundary detection using laser and vision," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '01)*, pp. 1440–1445, Maui, Hawaii, USA, November 2001.
- [21] Y. Morales, A. Carballo, E. Takeuchi, A. Aburadani, and T. Tsubouchi, "Autonomous robot navigation in outdoor cluttered pedestrian walkways," *Journal of Field Robotics*, vol. 26, no. 8, pp. 609–635, 2009.
- [22] A. V. Nefian and G. R. Bradski, "Detection of drivable corridors for off-road autonomous navigation," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '06)*, pp. 3025–3028, October 2006.
- [23] C. Guo and S. Mita, "Drivable road region detection based on homography estimation with road appearance and driving state models," in *Proceedings of the 4th International Conference on Autonomous Robots and Agents (ICARA '09)*, pp. 204–209, February 2009.
- [24] B. Ma, S. Lakshmanan, and A. O. Hero III, "Simultaneous detection of lane and pavement boundaries using model-based multisensor fusion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, no. 3, pp. 135–147, 2000.
- [25] I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, NY, USA, 2nd edition, 2002.
- [26] J. Borenstein and Y. Koren, "The vector field histogram—fast obstacle avoidance for mobile robots," *IEEE Transactions on Robotics and Automation*, vol. 7, no. 3, pp. 278–288, 1991.
- [27] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*, MIT Press, Cambridge, Mass, USA, 2005.

Research Article

A Computationally Efficient and Adaptive Approach for Online Embedded Machinery Diagnosis in Harsh Environments

Chuan Jiang and Samuel H. Huang

Department of Mechanical Engineering, University of Cincinnati, Cincinnati, OH 45221, USA

Correspondence should be addressed to Chuan Jiang; jiangcn@mail.uc.edu

Received 21 April 2013; Accepted 16 June 2013

Academic Editor: Yixiang Huang

Copyright © 2013 C. Jiang and S. H. Huang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Condition-based monitoring (CBM) has advanced to the stage where industry is now demanding machinery that possesses self-diagnosis ability. This need has spurred the CBM research to be applicable in more expanded areas over the past decades. There are two critical issues in implementing CBM in harsh environments using embedded systems: computational efficiency and adaptability. In this paper, a computationally efficient and adaptive approach including simple principal component analysis (SPCA) for feature dimensionality reduction and K-means clustering for classification is proposed for online embedded machinery diagnosis. Compared with the standard principal component analysis (PCA) and kernel principal component analysis (KPCA), SPCA is adaptive in nature and has lower algorithm complexity when dealing with a large amount of data. The effectiveness of the proposed approach is firstly validated using a standard rolling element bearing test dataset on a personal computer. It is then deployed on an embedded real-time controller and used to monitor a rotating shaft. It was found that the proposed approach scaled well, whereas the standard PCA-based approach broke down when data quantity increased to a certain level. Furthermore, the proposed approach achieved 90% accuracy when diagnosing an induced fault compared to 59% accuracy obtained using the standard PCA-based approach.

1. Introduction

Within the past decades, industrial maintenance has become progressively significant due to the fact that industries are demanding near-zero downtime for operating machines. With the development of sensor and informatics technologies, condition-based maintenance (CBM) approaches have been widely used in various applications [1–3]. One of the challenges in today's CBM research is that mechanical systems in harsh environments are difficult to be maintained due to their big size, faraway location, and limited accessibility to large-scale computing unit. In addition, some unexpected environmental conditions may accelerate the degradation or failure of mechanical components since their design specifications are based on general environments. These factors will result in a failure to process the collected data efficiently and assess the health status of operating machines adaptively. Specifically, efficiency refers to how to quickly analyze the data with limited computational resources and accurately

capture a simple set of features that represents the dynamic characteristics of the operating machine [2]; adaptability means abnormality or faults should be detected in a real-time manner. Thus, an online embedded machinery diagnosis strategy which integrates intelligent feature selection, diagnostic models, and embedded systems needs to be developed to build an on-site knowledge-based platform for machinery fault diagnosis in harsh environments. Although numerous sensor and informatics technologies were applied in fault diagnosis [3–5], the computational capabilities of embedded systems still cannot meet the requirements when processing a high dimensional dataset. The lack of data computation, transfer speed, and data storage space of equipment in harsh environments can also cause the failure to formulate an optimized online CBM strategy, which has already become a bottleneck problem in many practical situations [6]. Therefore, it is inevitable to find ways to reduce the computational burden of processing high dimensional datasets to realize an optimized online embedded CBM application.

This paper proposes a novel computationally efficient approach which integrates a fast and adaptive feature dimensionality reduction algorithm based on simple principal component analysis (SPCA) and an intelligent K-means clustering algorithm for fault diagnosis. Figure 1 depicts the framework of the proposed approach for online machinery diagnosis. Firstly, raw vibration signals collected from a data acquisition system are analyzed by wavelet analysis, and the energy features are obtained and normalized using z-score method. Secondly, the efficient dimensionality reduction tool, SPCA, is used to simplify the feature space and calculate the principal component vectors. Thirdly, the intelligent K-means clustering algorithm is applied to assess machine performance and diagnose if any faults have happened in a real-time manner [6]. Also, the diagnostic accuracy can be calculated by confusion matrix analysis. The key of this approach is the involvement of computationally efficient algorithms for feature dimensionality reduction and adaptive health assessment, which makes it a suitable approach for online embedded machine diagnosis in harsh environments.

The remainder of this paper is organized as follows. In Section 2, the state-of-art is reviewed for the different dimensional reduction algorithms such as PCA, KPCA, SPCA, and the intelligent K-means clustering algorithm. In Section 3, the mathematical background of PCA, KPCA, and SPCA is introduced with a theoretical analysis of the computational complexity. The potential of SPCA in computational speed and adaptability is discussed. Also, the proposed approach is validated using an offline bearing test dataset. In Section 4, a real online embedded case study is given. The proposed approach is applied to analyze vibration signals that are collected from an imbalanced shaft test rig. The diagnostic results are not only validated, but also benchmarked regarding the calculation speed, as well as the adaptability for the scenario of potentially unknown faults in online embedded CBM applications. Section 5 draws the conclusions.

2. State-of-the-Art

Condition monitoring for a mechanical system usually requires installing some necessary sensors, which result in multiple channels for different signals in the dataset. What's more, even for each channel, applying feature extraction methods in time, frequency, and wavelet domain forms a high dimensional feature matrix, which dramatically increases the data dimension used to describe the condition of the operating machines. This problem will not only create a big burden for all the following data processing, transferring, and storage, but also will hinder the expression for the most useful data information and diagnostic decision making.

Numerous efforts have been targeted on developing dimensionality reduction techniques in the past few decades. Principal component analysis (PCA) is one of the most popular dimension reduction techniques to optimize the features without discarding much original information of the feature space. For instance, Malhi and Gao used PCA to identify the most representative features as inputs to a defect classification application with its ability to discriminate

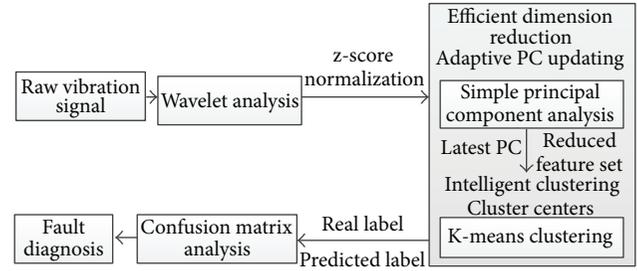


FIGURE 1: The framework of the computationally efficient and adaptive approach.

directions with the largest variance in a dataset [7]. Tumer and Huff extracted the principal “modes” of vibrations from the input data for health monitoring of helicopter gearboxes [8]. Khomfoi and Tolbert applied PCA to reduce the number of neurons before neural network classification, and it helped simplify the problem and achieve good results [9]. For studies on fault diagnosis, PCA can be effectively used for feature reduction from the original features [10]. On the other hand, some extensions of PCA were developed for various applications. For example, kernel principal component analysis (KPCA) was proposed to deal with nonlinear problems by characterizing the nonlinearity and nonstationary of industrial systems [11, 12]. Kim et al. applied KPCA to extract facial features and solved face recognition problems [13]. He et al. also established a subspace of feature vectors for gear condition monitoring through KPCA [14]. Researchers have also used KPCA in condition-based monitoring and fault diagnosis applications [15]. Simple principal component analysis (SPCA), invented by Partridge and Calvo [16], was used for dimensionality reduction of two high-dimensional image databases with a fast convergence rate. However, most of the previous studies focused only on the effectiveness of the methods for data dimension reduction in an offline analysis environment. Little attention has been paid to consider if they are still applicable when used in an online embedded CBM application, one of the future mainstreams of CBM research. Even if offline datasets can be used to simulate an online analysis environment for several online CBM methods [6, 17], some problems, such as efficiency and compatibility, can be overlooked very easily without a real online embedded validation.

In this paper, the proposed approach is particularly designed for online embedded machinery CBM applications. The most significant properties of SPCA, computational speed and adaptability, help to improve the data analysis efficiency and adaptively detect abnormality when working with the intelligent K-means clustering algorithm [18] if the machine degradation or fault happens. Further, this study validates the advantages of the approach in an offline fault diagnosis of rolling element bearing and an online embedded imbalanced shaft diagnostic application, respectively.

3. Online Embedded Machinery Diagnosis Approach

3.1. Dimension Reduction. Basically, the methods to calculate the principal components can be divided into two main categories of calculating the eigenvectors. One is based on matrix method, of which a practical technique is applying the singular value decomposition (SVD) technique to calculate principal component vectors. The most commonly used one is the matrix method. For instance, PCA is a technique that can be used to simplify a dataset. More formally it is defined as a linear transformation that chooses a new coordinate system for the dataset such that the greatest variance by any projection of the dataset comes to lay on the first axis, the second greatest variance on the second axis, and so on. Mathematically, PCA converts feature vectors into lower dimensional random variable with independently-distributed components by finding the eigenvalues and eigenvectors of the covariance matrix to represent the statistical significance and directions of principal components, respectively [19]. The detailed steps can be illustrated as follows.

- (1) Use x_i to represent the i th sample for the original n -dimensional dataset, $i = 1, 2, \dots, m$.
- (2) Subtract off the mean of each measurement type for the original n -dimensional data matrix \mathbf{X} , then the corresponding covariance matrix $\mathbf{R} = \mathbf{X}\mathbf{X}^T$ can be calculated as follows:

$$\mathbf{R} = \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})^T. \quad (1)$$

- (3) Calculate the eigenvalues λ and eigenvectors ν of \mathbf{R} as follows:

$$\lambda \nu = \mathbf{R}\nu. \quad (2)$$

- (4) Sort the eigenvalues and the corresponding eigenvectors, select the first $d \leq n$ eigenvectors, and generate the new data matrix. Here, d is the number of preferred dimensionality.

Another example of the matrix method to compute eigenvalues and eigenvectors is KPCA, which is an extension of the standard PCA using kernel functions to realize a nonlinear mapping. For KPCA, data in the input space is usually mapped to a higher dimensional feature space where its eigenvectors can be calculated. However, the enhancement for the nonlinear feature of KPCA requires much more computational resources, which may lower the efficiency and make it not suitable for online applications. The algorithm steps of KPCA can be shown as follows.

- (1) Use x_i to represent the i th sample for the original n -dimensional dataset, $i = 1, 2, \dots, m$.
- (2) Subtract off the mean of each feature dimension for the original data matrix \mathbf{X} .

- (3) Project the matrix obtained from step 2 to a higher dimension by using a kernel function and normalize the m -by- m kernel matrix \mathbf{K} as follows:

$$\mathbf{K} = k(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j), \quad (3)$$

where x_i and x_j are the sample vectors. k is the kernel function. The common kernel functions proposed by Vapnik et al. [20] mainly include polynomial kernel function, radial basis function, and neural network kernel function in (4)–(6), respectively, as follows:

$$k(x_i, x_j) = (x_i \cdot x_j + 1)^d, \quad (4)$$

$$k(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|}{2\sigma^2}\right), \quad (5)$$

$$k(x_i, x_j) = \tanh((x_i \cdot x_j) + b). \quad (6)$$

- (4) Calculate the eigenvalues λ and the corresponding eigenvectors ν of the normalized kernel matrix \mathbf{K}_n . One has

$$m\lambda\nu = \mathbf{K}_n\nu. \quad (7)$$

- (5) Normalize eigenvectors, select the first $d \leq n$ eigenvectors, and generate the new data matrix.

The second category to determine the principal component vectors is based on the data method. SPCA falls into this category [16]. Compared with the matrix-based methods, the numerical methods can reduce the computational complexity, especially when the dimension of the dataset increases. In SPCA, a data oriented method is adopted to approximate the principal components instead of explicitly calculating the eigenvectors based on the covariance matrix. In addition, SPCA uses a Hebbian-learning-like algorithm to adapt the learning parameters dynamically and quickly find eigenvectors, but avoids the learning parameter tuning problem of the Hebbian learning rule [16]. The procedure of SPCA can be described in Figure 2 and as follows.

- (1) Use x_i to represent the i th sample for the original n -dimensional dataset, $i = 1, 2, \dots, m$.
- (2) Subtract off the mean of each measurement type.
- (3) Initialize an n -dimensional vector e_k^0 , where $k = 1, 2, \dots, d$.
- (4) Calculate the k th principal component vector e_k based on the following equations:

$$y_i = \frac{1}{\|e_k^{j-1}\|} (e_k^{j-1})^T x_i. \quad (8)$$

The purpose of (8) is to map the i th sample using the latest k th principal component vector. So y_i is the projection in the direction of k th principal component vector. It can also be viewed as the “similarity”

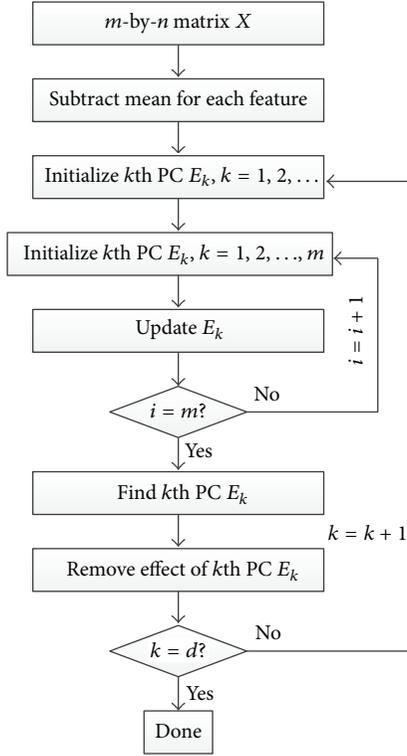


FIGURE 2: SPCA algorithm procedure.

between the i th sample vector and the latest k th principal component vector. One has

$$\phi(y_i, x_i) = y_i x_i. \quad (9)$$

In order to calculate the change of the k th principal component vector, a Hebbian term $\phi(y_i, x_i)$ is defined in (9). It can be found that the larger y_i is, the more influential the i th sample will be in determining the k th principal component vector. Consider

$$e_k^i = \frac{\sum_i \phi(y_i, x_i)}{\|\sum_i \phi(y_i, x_i)\|}. \quad (10)$$

Then in (10), the latest k th principal component vector e_k^i can be obtained by normalization. The $(i + 1)$ th sample will be used to update the k th principal component vector to e_k^{i+1} .

- (5) Up to this point, the k th principal component vector has been obtained after using all m samples in the k th iteration. In order to find the $(k + 1)$ th principal component vector, the following equation has to be used for all data samples to remove the effect of the k th principal component, so as to avoid being found again:

$$x_i' = x_i - (e_k^T x_i) e_k. \quad (11)$$

- (6) Repeat Steps 3–5 until d ($d \leq n$) principal component vectors have been obtained. Figure 3 illustrates the graphical steps to find the first principal component vector in SPCA algorithm.

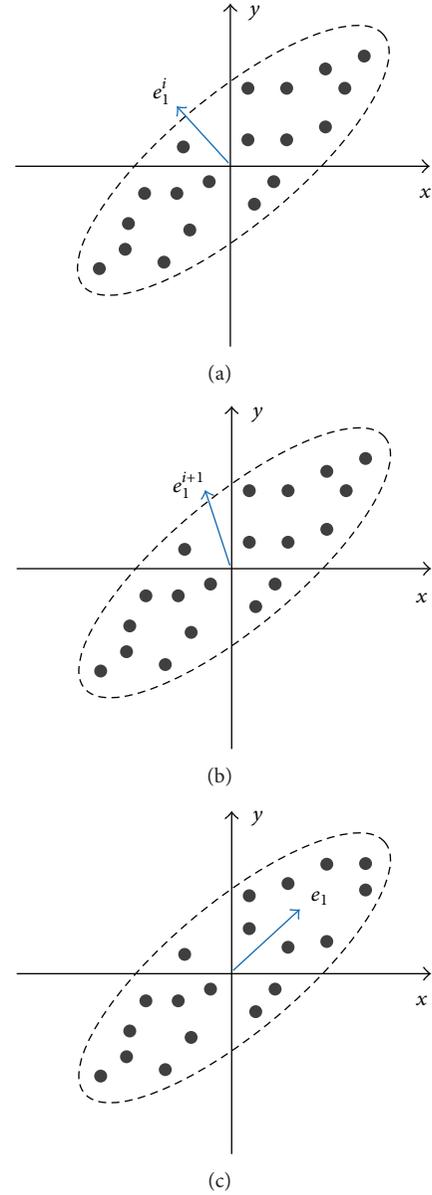


FIGURE 3: SPCA algorithm procedure. (a) Data distribution with an initialized approximate principal component vector; (b) a better approximate principal component vector has been found; and (c) the real first principal component vector of this data distribution is given.

3.2. Dimension Reduction Algorithm Complexity. The computational speed of PCA, SPCA, and KPCA depends on the dimensionality of the feature space. However, it is also affected by different attributes of the input matrix or vectors. For instance, if the objective is to find d ($d \leq n$) principal components in an n -dimensional matrix with m samples using PCA, the time complexity for singular value decomposition (SVD) or Householder-QR technique is $O(mn^2 + n^3)$ and for utilizing Hotelling's power method, it is $O(mn^2 + dn^2)$. Additionally, the computation of KPCA realized by calculating eigenvectors involves a time complexity of $O(m^3)$

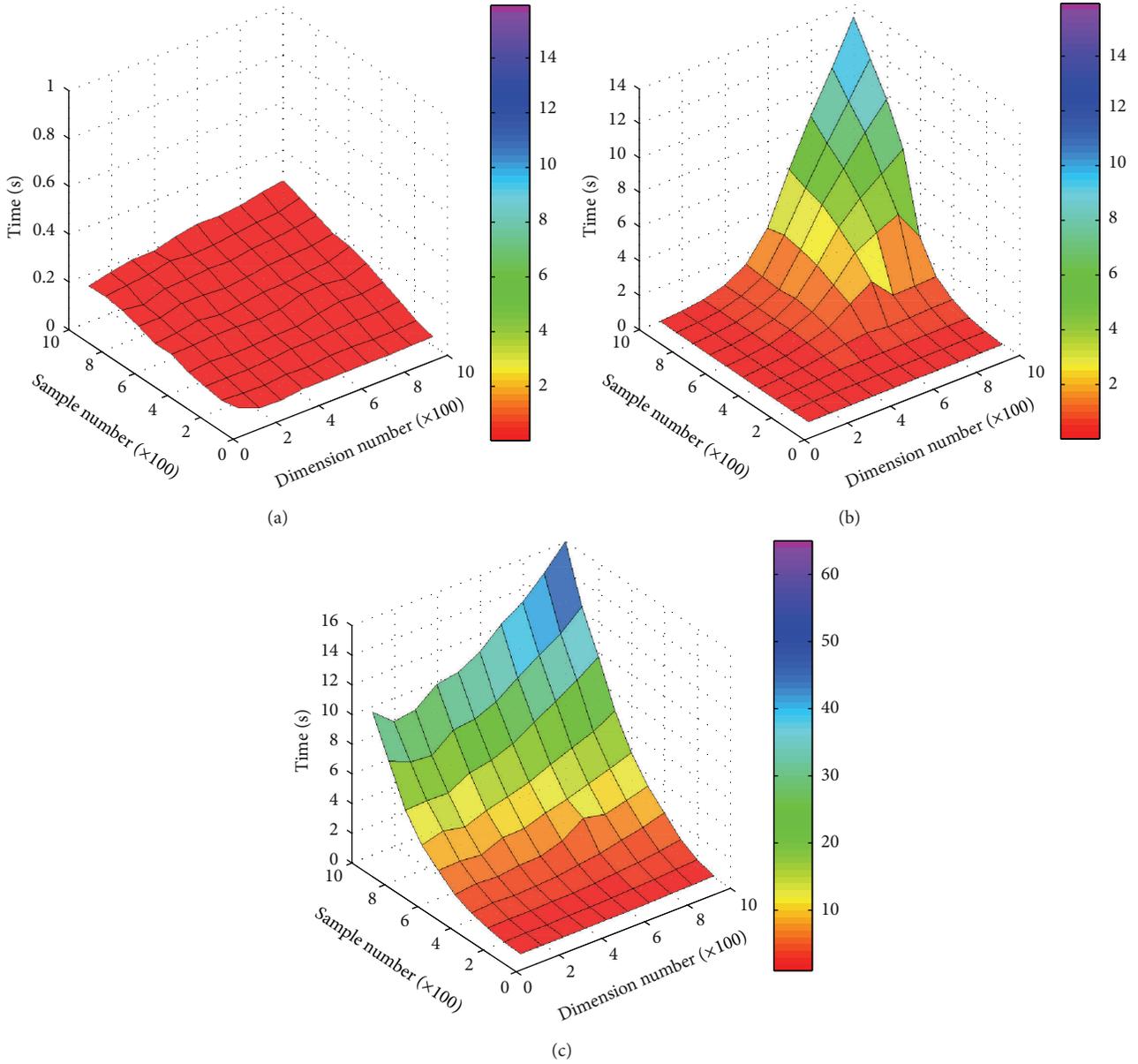


FIGURE 4: Computational time by: (a) SPCA, (b) PCA, and (c) KPCA.

[21], and the time complexity for SPCA is $O(dmn)$ by looking into its computational steps. It can be concluded that the time complexity for PCA increases more dramatically than that of SPCA when the data dimension n gets larger. For KPCA, its calculation speed is slow if the data sample number is large. As a result, implementing PCA or KPCA on large datasets is prohibitively time expensive.

Further, how these variables, that is, target dimensionality, the target number of principal components d , sample number (m), or data matrix dimension (n), influence the time complexity can be studied. PCA, SPCA, and KPCA are applied to a data matrix with randomly generated numbers to study the different computational speeds by changing the row number and the column number from an initial number 100 up to 1000 with an increment of 100 for each step. Figure 4

shows computational speeds for different sizes of the input matrix for PCA, SPCA, and KPCA, respectively.

It can be seen from Figure 4 that the advantage of the computational speed of SPCA gets more apparent when the size of the matrix becomes larger. Its time cost increases much slower than the other two methods do. The computational speed of SPCA is mainly affected by the sample number m , while dimension number n becomes a relatively minor factor, which can be explained by the data-oriented procedure of SPCA. Furthermore, its computational cost is low even when the size of data matrix is very large. Taking the 1000-by-1000 data matrix as an example, the computational time of SPCA is 0.2646 s. It is reduced by 97.8% and 98.4% when compared with PCA (13.110 s) and KPCA (15.957 s), respectively. Similarly, the sample number m is the major factor

for KPCA in terms of computational speed. Nonetheless, the computational consumption becomes prohibitively high with the growing sample number m , and it will exceed the consumption of PCA. On the contrary, the computational speed of PCA is determined by the smaller number between row size and column size. For example, if there is an n -dimensional data matrix with m samples ($m > n$), then n will be the key factor influencing PCA's computational speed. This can be verified by looking into the mathematical steps of PCA, an n -by- n matrix is utilized to calculate the eigenvalues. What's more, the computational time of PCA rises exponentially with the growth of n and will be very slow if n is a large number.

3.3. Online Adaptive Diagnosis Potential of SPCA. For most online monitoring cases, it is desired to find the principal component vectors accurately and quickly with as little computational consumption as possible. However, PCA, as well as KPCA, usually needs all of the data samples to calculate the principal component vectors. Therefore, the computational speed will be very slow when the dataset size is large. This may result in the delay of data processing and significantly weaken the applicability of online analysis.

As previously mentioned, SPCA adopts a data-oriented method to calculate principal component vectors. Thus, it can update the principal component vectors at every iteration, which enables it to utilize the most recent principal component vectors when new data samples come in. This means, it is only necessary to store the latest principal component vectors and discard old data samples instead of saving all data samples to calculate the principal component vectors. As a result, both hardware storage and computational consumption will be saved in embedded systems. Moreover, in some situations where the machine starts to degrade, people may not have the failure data to train the model and use it to detect the degradation or unknown faults quickly, because the principal components vectors are calculated in normal operating conditions. The principal components trained by only normal status data may blur or mislead the clustering results in these cases. To solve it, another way is to train PCA or KPCA frequently to acquire the latest principal component vectors, which is time consuming whenever new data sample comes in and requires a lot of memory to store all the history data. This property of PCA and KPCA will slow down the calculation and may quickly exhaust the computational resources of some embedded systems. Compared with PCA and KPCA, SPCA can always process the new data sample to update the previous-latest principal components. Then, in case new incoming data from some unknown faults is fed into the CBM system, the adaptability of SPCA may help to achieve accurate classification results as well as to identify the new features of the failure status that deviate from the cluster of the normal conditions. It will be shown that this deviation can be viewed as an abnormality or fault indicator by using SPCA instead of PCA and KPCA.

3.4. Offline Approach Validation. Before a real online embedded implementation, the proposed approach is applied in

a fault-seeded bearing test. The vibration dataset used in this work is acquired from a standard rolling element bearing test [22] which has been studied by many researchers [6, 23–25]. The vibration data was acquired by accelerometers at the sampling frequency of 12 KHz. The type of test bearings was the deep groove ball bearing 6205-2RS JEM SKF. During the test, the rotation speed was about 30 Hz (1800 rpm) under different loads (0, 0.75, 1.49 and 2.24 KW).

There are 985 data samples in total from the normal and the fault-seeded bearings. Each data sample has 2048 data points. Among the datasets, there are 276 from the normal conditions, and the rest of 709 samples are of faulty conditions (237 inner race fault samples, 236 outer race fault samples, and 236 ball fault samples). Faults are introduced into either the inner raceways or the outer raceways of the drive-end bearings by the electric discharge machining method. The diameter of the faults is 0.1778 mm with depth of 0.2794 mm.

The second-generation wavelet packet decomposition is used to extract the energy features in the time-frequency domain [6, 26]. Basically, the second generation wavelet transform uses a general construction of the lifting scheme which has three major advantages (1) the fully in-place calculation of wavelet transforms, (2) no requirement for auxiliary memory, and (3) a faster implementation of the wavelet transform [27, 28].

Considering the rotation speed is about 30 Hz and the sampling rate is 12 KHz in this experiment, the decomposition level is chosen to be 7 so that the whole frequency space can be equally divided into 128 intervals. Therefore, the distance between centers of adjacent frequency spans is 46.875 Hz which is close to the normal operating frequency. Thus, the 7-level decomposition of the lifting-based biorthogonal wavelet packet transform is applied to map the raw signals into the wavelet domain, where the packet energy of each decomposition node is calculated and then normalized by the z-score method. Thus, a large matrix of wavelet packet energy for normal condition and three faulty conditions is obtained. The obtained feature space has 128 dimensions.

The datasets are divided into the training sets and the testing sets through a random cross-validation process; 50% of the datasets are used as a training dataset to calculate the principal components. Standard PCA, KPCA, and SPCA are applied to the training dataset, respectively. The result is the product of whole datasets and principal component vectors. For the purpose of visualization, the fault classification results have been plotted (see Figure 5).

It can be seen from Figure 5 that the gaps between each two data clusters are all clear for the applied methods, which indicate that the normal dataset and the three fault datasets can be easily classified. Secondly, the normal datasets scatter much more centralized than the faulty datasets do in the plots of PCA and SPCA, which may reveal an increasing unstableness of the faulty conditions. In addition, as PCA and SPCA share the same objective equations, although with different ways to calculate the principal components, the shapes of the dataset distributions are similar. The distribution shape of the dataset projected by using KPCA is quite different from those done by PCA and SPCA because of its nonlinear property. All the previously mentioned dimension

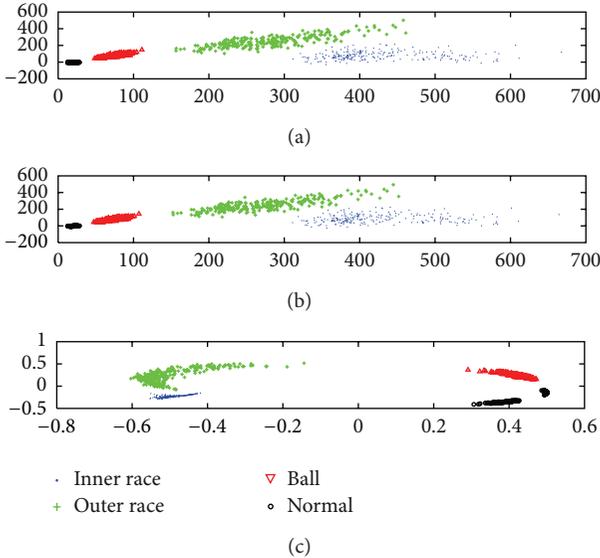


FIGURE 5: Bearing faults classification. (a) Bearing faults classification using PCA, (b) bearing faults classification using SPCA, and (c) bearing faults classification using KPCA.

reduction algorithms are effective for the rolling element bearing diagnosis within a 2-dimensional feature space. In other words, 98.44% of redundant data has been eliminated, that is, from 128 dimensions to only 2 dimensions.

3.4.1. Study on Computational Speed for Online Purpose. For a better implementation in online applications, the computational speed of dimension reduction for different data dimensions regarding the scale or size of the input matrix is studied. The feature dimensionality is adjusted by changing the wavelet decomposition level. The results of processing the training dataset are shown in Table 1. To understand how computational time varies intuitively based on the adjustment of wavelet decomposition level, the comparison among the computational time of PCA, SPCA, and KPCA is shown in Figure 6 as follows.

The observations are:

- (1) As wavelet decomposition level changes incrementally, the dimensionality accordingly increases. The sample number remains constant. The increase in computational time of SPCA when dealing with 32-dimensional and 512-dimensional data is 27.83%; For PCA, the increase in the time consumption is 7600%; and for KPCA, it is 54.23%. It can be seen that the increase in time consumption of SPCA is much smaller than that of PCA and KPCA.
- (2) Although the computational time for PCA is smaller than that of SPCA when the wavelet decomposition levels are 5, 6, and 7, its speed decreases exponentially with the growing feature dimensionality due to the wavelet decomposition levels increasing to 8 and 9, and quickly goes far beyond the time consumption

TABLE 1: The computational speed of different size of data matrix for 493 feature points.

Wavelet decomposition level	5	6	7	8	9
Feature dimensionality	32	64	128	256	512
SPCA time cost (s)	0.097	0.100	0.104	0.111	0.124
PCA time cost (s)	0.013	0.017	0.041	0.219	1.001
KPCA time cost (s)	0.922	0.946	0.985	1.194	1.422

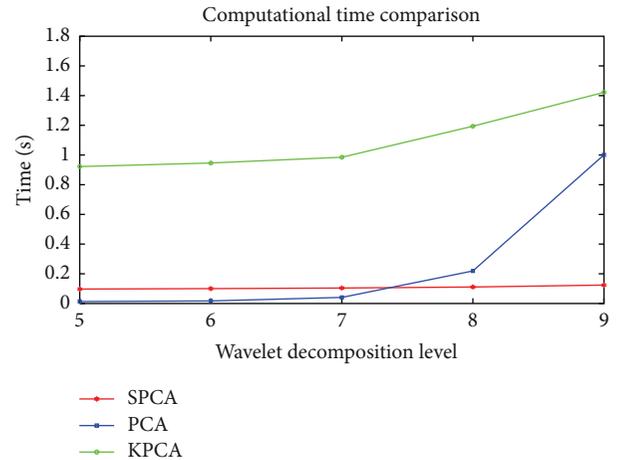


FIGURE 6: The computational time for different wavelet decomposition level (the larger the number wavelet decomposition level is, the higher the dimensionality of the feature space is.)

of SPCA, which corresponds with what has been discussed in Section 3.2.

- (3) The computational time of KPCA is usually much more than those of PCA and SPCA, which indicates that KPCA is not quite suitable for an online application that requires a reasonably quick response time to process the data. On the contrary, SPCA can save computational time, especially for large dimensional feature sets, which may entitle SPCA as a sound option of dimension reduction technique to solve the bottleneck problem due to limited computational resources in online applications.

3.4.2. Study on Adaptability for Online Purpose. In many scenarios of real world condition-monitoring applications, the machine runs normally for a long time after it is installed in a manufacturing line. The critical equipment must be regularly maintained to reliably work in its normal status, which means people often lack the data from the faulty status. For this kind of scenario, it is reasonable to assume that the diagnosis algorithms can be trained by only the dataset from the normal status in many cases. Due to the lack of the faulty information, the adaptability is required for the online applications. Therefore, it is necessary to study the adaptability of the algorithms in order to deal with the unknown faults. In this paper, PCA and SPCA are compared by the following procedure.

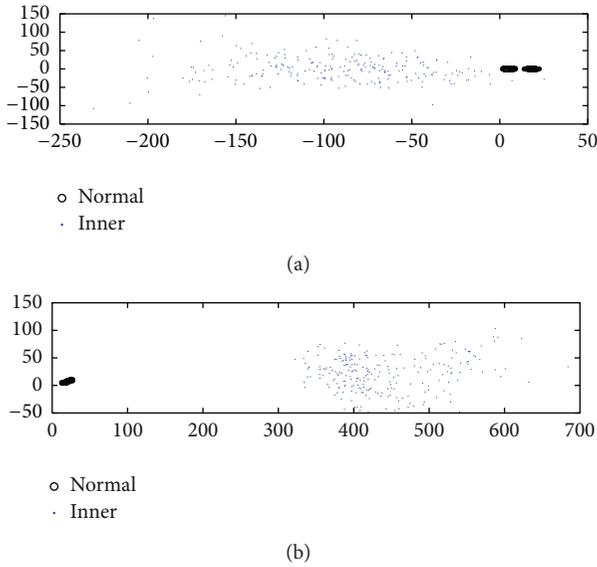


FIGURE 7: (a) Bearing normal and inner race fault classification using PCA. (b) Bearing normal and inner race fault classification using SPCA.

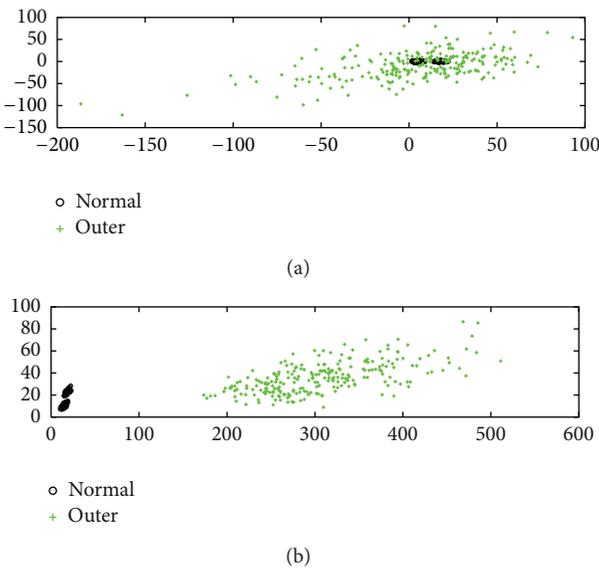


FIGURE 8: (a) Bearing normal and outer race fault classification using PCA. (b) Bearing normal and outer race fault classification using SPCA.

Step 1. Assume that only the data from the normal status of the machine can be obtained in the initiate operation stage. Thus, the PCA and the SPCA are trained by the dataset of the normal status.

Step 2. Assume that the three seeded fault types (inner race fault, outer race fault, and ball fault) are unknown abnormal statuses, which will occur after the normal status, respectively. Thus, the principal component vectors trained in Step 1 will be applied to the unknown abnormal statuses.

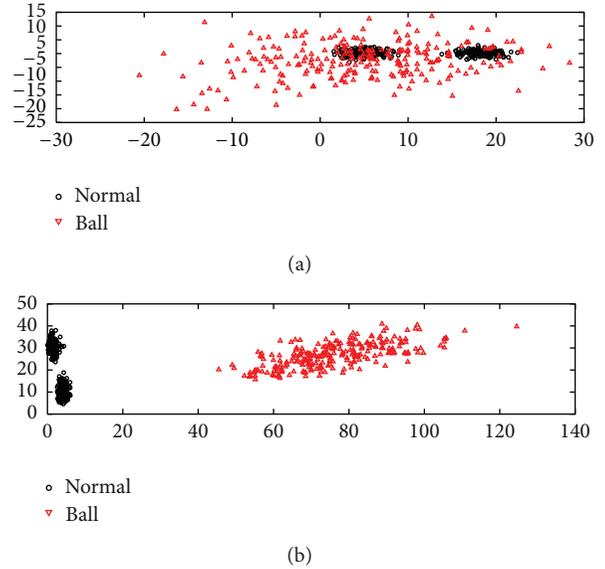


FIGURE 9: (a) Bearing normal and ball fault classification using PCA. (b) Bearing normal and ball fault classification using SPCA.

TABLE 2: Results of the correct classification rate by K-means algorithm.

	PCA	SPCA
Normal and inner race fault	92.215%	100%
Normal and outer race fault	67.383%	100%
Normal and ball fault	57.617%	100%

The KPCA is not included in this comparison because of its slow computational response as mentioned in the previous study. The comparison experiment results of the adaptability study are shown in Figures 7, 8, and 9.

It can be seen from Figures 7, 8, and 9 and Table 2 that when the bearing fault occurs, the normal condition and the faulty condition can hardly be classified by applying PCA, especially for the outer race fault and the ball fault conditions. For the inner race fault, although there are no obvious overlapped clusters, the correct classification rate based on standard PCA is still lower than that based on SPCA. Compared with PCA; the gaps between clusters of feature points from normal and faulty conditions projected by SPCA are much clearer. Therefore, it can be seen that the proposed approach can not only effectively reduce the dimensionality of the feature space, but also adaptively separate the normal condition and the unknown faulty conditions by updating the trained principal component vectors when new data comes in.

4. Online Embedded Case Study for Imbalanced Shaft Diagnosis

4.1. Experiment Setup. In this case study, a test rig consisting of a rotating shaft is used for the purpose of validating the advantages of the proposed approach in online embedded

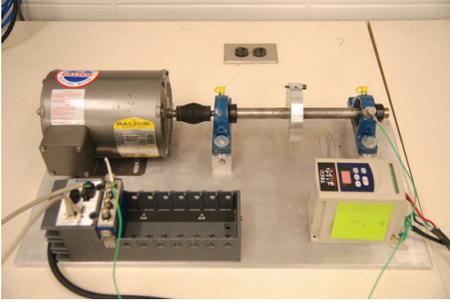


FIGURE 10: Test rig for shaft vibration data acquisition.

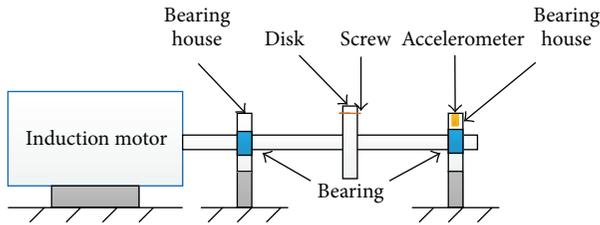


FIGURE 11: The schematic description of the test rig.

CBM applications (see Figures 10 and 11). This test rig consists of a shaft connected to a three-phase industrial motor, a disk attached in the middle of the shaft, two bearings to hold the shaft on each side, and a controller used to set the rotation speed of the shaft. The imbalanced shaft fault can be induced by putting screws into the disk. An accelerometer is mounted on one side of the bearing houses to collect the vibration data.

The condition monitoring algorithms are deployed on an embedded system, which consists of an NI cRIO-9004 real-time controller of 64 MB DRAM, 512 MB compact flash, and an NI cRIO-9104 reconfigurable chassis with an NI 9234 IEPE vibration signal collection module. The algorithms are compiled to an FPGA executable LabVIEW program to analyze the data acquired by the NI 9234 module and the accelerometer. The data sampling rate is set at 5120 Hz, and the rotation speed of the shaft is set at 20 Hz (1200 rpm). The embedded system is connected to a computer for displaying the results of real-time monitoring after data processing is finished.

4.2. Study on Online Embedded Computational Speed. The data processing procedure is the same as the procedure described in the previous case study. The computational speed values of PCA and SPCA are compared by analyzing the matrix of the features of the wavelet packet coefficient energies. Different wavelet packet decomposition levels are tried from 5 to 10. The experiment is implemented on 500 feature points. There are 2048 vibration data samples processed for each feature point. The results are shown in Table 3.

With the increase of wavelet decomposition level, it can be seen that SPCA deployed on the embedded system is still able to reduce the feature dimensionality with limited computational resources when data dimensionality gets

TABLE 3: Computational speed of different feature dimensionality in the embedded system.

Wavelet decomposition level	5	6	7	8	9	10
Feature dimensionality	32	64	128	256	512	1024
SPCA time cost (s)	0.162	0.275	0.477	0.913	1.767	3.636
PCA time cost (s)	0.376	1.249	5.640	32.787	N/A	N/A

higher. Nevertheless, the computational time for PCA goes up exponentially again with increasing dimensionality of the feature vectors when the wavelet packet decomposition level changes from 5 to 8. Out-of-memory errors occur on the embedded system when the wavelet decomposition level further increases to 9 and 10, which makes the results not available (N/A). Further, if Table 3 is compared with Table 1, efficiency of SPCA becomes much more outstanding in a real online embedded implementation, while the usage of PCA can result in a long delay or even breakdown which will weaken the significance of online data processing and timely decision making.

4.3. Study on Adaptability for Online Embedded Application. Besides the scenario of limited online computational resources for the embedded system, another online scenario is lacking the faulty information when the diagnosis algorithm is deployed and applied, which means that the adaptability for an unknown faulty status is required. An online analysis scenario is assumed that people are unaware of the coming faults (to be induced) in the beginning when the algorithms are deployed on the embedded system. Thus, the principal component vectors of standard PCA and SPCA are trained by the normal condition and applied to all the new incoming data, including from the normal and the faulty statuses. The detailed procedure is shown as follows.

- (1) Collect and process the data from normal shaft, and train the principal component vectors using standard PCA and SPCA.
- (2) Induce the imbalanced shaft fault by putting screws into the disk.
- (3) Collect and process the data from the faulty shaft, and reduce its dimensionality using the trained principal component vectors.
- (4) Visualize the projected feature points of both normal and faulty conditions.

Considering the rotation speed is about 20 Hz and the sampling rate is 5.12 KHz in the experiment, the decomposition level is chosen to be 7 so that the whole frequency space can be equally divided into 128 intervals, of which the distance between centers of adjacent frequency spans is 20 Hz which is the same as the normal operating frequency. The test has been run ten times for the purpose of validation. The screenshots of the visualization step during one of the ten tests are shown in Figures 12 and 13. The separation between the normal cluster and the faulty cluster mapped by SPCA is much clearer than that mapped by standard PCA.

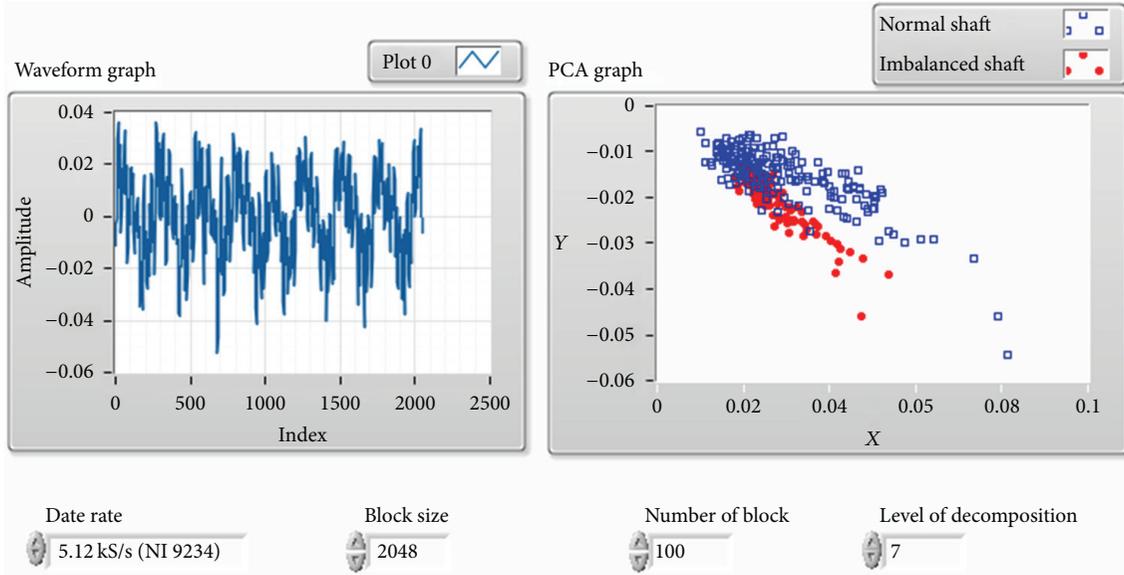


FIGURE 12: Online shaft diagnosis using PCA.

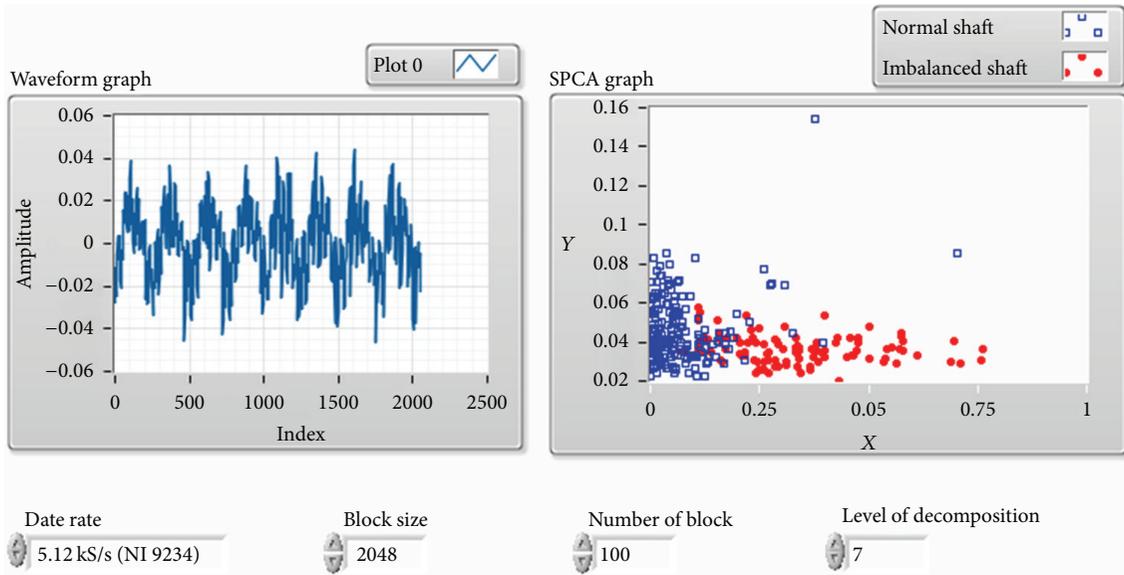


FIGURE 13: Online shaft diagnosis using SPCA.

TABLE 4: The confusion matrix for PCA.

PCA		Predicted class	
		Normal	Imbalanced
Classification accuracy = 59.00%			
Actual class	Normal	139	61
	Imbalanced	62	38

TABLE 5: The confusion matrix for SPCA.

SPCA		Predicted class	
		Normal	Imbalanced
Classification accuracy = 90.33%			
Actual class	Normal	183	19
	Imbalanced	10	88

To find out quantitative classification accuracy for each online analysis scenario, K-means clustering algorithm is used to cluster the feature points from normal and imbalanced faulty statuses, and the diagnosis accuracy in one of the tests is shown in confusion matrices in Tables 4 and 5. The average accuracies for all the tests are 89.80% for SPCA

and 60.80% for standard PCA, respectively. The reason why SPCA can detect the faulty condition with better accuracy than standard PCA in online analysis scenario is because it can utilize the previous principal component vectors and update them adaptively when new data comes in instead of using all data samples to calculate principal component

vectors again. Therefore, the proposed approach could not only save computational resources and data storage space, but also improve the machine fault diagnosis efficiency and accuracy in online embedded CBM applications.

5. Conclusion and Future Work

The paper proposes a computationally efficient and adaptive approach which includes SPCA for feature dimensionality reduction and intelligent K-means clustering for online machine faults diagnosis. The approach is applied to an offline rolling element bearing fault diagnosis test and an online embedded imbalanced shaft fault diagnosis application. While it has been demonstrated that the proposed approach is effective in both applications, it is also validated despite the computational constraints which can be easily neglected in simulated online applications. Results show that the proposed approach runs much faster and is more memory saving for high dimensional datasets and has better adaptability to identify the new unknown abnormal or faulty conditions.

In summary, it can be concluded that the proposed approach is a suitable method for online embedded machinery diagnosis applications in harsh environments. Also, a further study will be conducted on some possible ways to realize machine failure mode identification and remaining useful life prediction in the future.

References

- [1] J. Lee, J. Ni, D. Djurdjanovic, H. Qiu, and H. Liao, "Intelligent prognostics tools and e-maintenance," *Computers in Industry*, vol. 57, no. 6, pp. 476–489, 2006.
- [2] X. Li and S. K. Tso, "Real-time tool condition monitoring using wavelet transforms and fuzzy techniques," *IEEE Transactions on Systems, Man and Cybernetics C*, vol. 30, no. 3, pp. 352–357, 2000.
- [3] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 2006.
- [4] V. T. Tran and B. Yang, "An intelligent condition-based maintenance platform for rotating machinery," *Expert Systems with Applications*, vol. 39, no. 3, pp. 2977–2988, 2012.
- [5] V. T. Tran, H. Thom Pham, B. S. Yang, and T. Tien Nguyen, "Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and support vector machine," *Mechanical Systems and Signal Processing*, vol. 32, pp. 320–330, 2012.
- [6] Y. Huang, C. Liu, X. F. Zha, and Y. Li, "A lean model for performance assessment of machinery using second generation wavelet packet transform and Fisher criterion," *Expert Systems with Applications*, vol. 37, no. 5, pp. 3815–3822, 2010.
- [7] A. Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 6, pp. 1517–1525, 2004.
- [8] I. Y. Tumer and E. M. Huff, "Analysis of triaxial vibration data for health monitoring of helicopter gearboxes," *Journal of Vibration and Acoustics, Transactions of the ASME*, vol. 125, no. 1, pp. 120–128, 2003.
- [9] S. Khomfoi and L. M. Tolbert, "Fault diagnosis and reconfiguration for multilevel inverter drive using AI-based techniques," *IEEE Transactions on Industrial Electronics*, vol. 54, no. 6, pp. 2954–2968, 2007.
- [10] Q. Hu, Z. He, Z. Zhang, and Y. Zi, "Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 688–705, 2007.
- [11] B. Schölkopf, A. Smola, and K. Müller, "Nonlinear component analysis as a Kernel Eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [12] B. Schölkopf, S. Mika, C. J. C. Burges et al., "Input space versus feature space in Kernel-based methods," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1000–1017, 1999.
- [13] K. I. Kim, K. Jung, and H. J. Kim, "Face recognition using Kernel principal component analysis," *IEEE Signal Processing Letters*, vol. 9, no. 2, pp. 40–42, 2002.
- [14] Q. He, F. Kong, and R. Yan, "Subspace-based gearbox condition monitoring by Kernel principal component analysis," *Mechanical Systems and Signal Processing*, vol. 21, no. 4, pp. 1755–1772, 2007.
- [15] R. Sun, F. Tsung, and L. Qu, "Evolving Kernel principal component analysis for fault diagnosis," *Computers and Industrial Engineering*, vol. 53, no. 2, pp. 361–371, 2007.
- [16] M. Partridge and R. A. Calvo, "Fast dimensionality reduction and simple PCA," *Intelligent Data Analysis*, vol. 2, no. 1–4, pp. 203–214, 1998.
- [17] F. Wu, T. Wang, and J. Lee, "An online adaptive condition-based maintenance method for mechanical systems," *Mechanical Systems and Signal Processing*, vol. 24, no. 8, pp. 2985–2995, 2010.
- [18] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. IT-28, no. 2, pp. 129–137, 1982.
- [19] J. Shlens, "A tutorial on principal component analysis," *Measurement*, vol. 51, no. 10003, p. 52, 2005.
- [20] V. Vapnik, S. E. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," *Advances in Neural Information Processing Systems*, pp. 281–287, 1997.
- [21] C. Tat-Jun and D. Suter, "Improving the speed of Kernel PCA on large scale datasets," in *Proceedings of the IEEE International Conference on Video and Signal Based Surveillance (AVSS '06)*, p. 41, November 2006.
- [22] K. A. Loparo, "Bearing Data Center, Case Western Reserve University, United States," <http://www.eecs.case.edu/laboratory/bearing/>, 2011.
- [23] H. Liu, C. Liu, and Y. Huang, "Adaptive feature extraction using sparse coding for machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 558–574, 2011.
- [24] K. C. Gryllias and I. A. Antoniadis, "A Support Vector Machine approach based on physical model training for rolling element bearing fault detection in industrial environments," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 2, pp. 326–344, 2012.
- [25] H. Ocaik and K. A. Loparo, "Estimation of the running speed and bearing defect frequencies of an induction motor from vibration data," *Mechanical Systems and Signal Processing*, vol. 18, no. 3, pp. 515–533, 2004.
- [26] J. Hongkai, H. Zhengjia, D. Chendong, and C. Peng, "Gearbox fault diagnosis using adaptive redundant Lifting Scheme," *Mechanical Systems and Signal Processing*, vol. 20, no. 8, pp. 1992–2006, 2006.

- [27] W. Sweldens, "The lifting scheme: a construction of second generation wavelets," *SIAM Journal on Mathematical Analysis*, vol. 29, no. 2, pp. 511–546, 1998.
- [28] W. Sweldens, "Lifting scheme: a new philosophy in biorthogonal wavelet constructions," in *Proceedings of the Wavelet Applications in Signal and Image Processing III. Part 1 (of 2)*, pp. 68–79, San Diego, Calif, USA, July 1995.

Research Article

Condition Evaluation of Large Generator Stator Insulation Based on Partial Discharge Measurement

Ruihua Li, Ling Pan, Chen Yan, Hao Li, and Bo Hu

School of Electronics and Information Engineering, Tongji University, Shanghai 201804, China

Correspondence should be addressed to Ruihua Li; rhli@tongji.edu.cn

Received 28 May 2013; Accepted 11 July 2013

Academic Editor: Liang Gong

Copyright © 2013 Ruihua Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The large generator faults are mostly from the stator insulation, which is aged by exposure to a combination of thermal, electrical, mechanical, and harsh environment stresses. Condition evaluation of stator insulation is an important measure of ensuring the safe operation and extending the remaining life of large generator. In this paper, a diagnosis method based on the partial discharge (PD) measurement was used for the condition evaluation of the stator insulation. The statistical parameters of partial discharge phase resolved distribution were proposed to assess the aging condition of stator insulation. A partial least square (PLS) approach was used to explore and extract relationships between the statistical parameters of partial discharge distribution with physical properties and performance such as electrical breakdown strength. Results of the PD testing and statistical analysis show that the statistical parameter S_k extent of partial discharge distribution changes much more remarkably with the aging time than do the other parameters. The statistical parameter S_k can be considered as a potential indicator of stator insulation aging. The PLS approach can effectively assess the stator insulation aging condition and can provide a predictive capability for the stator insulation diagnosis.

1. Introduction

Stator winding is a main component of high-voltage (HV) rotating machines (generators and large motors). Stator insulation failure of HV rotating machine can result in catastrophic damages to the equipment, loss or derating of power output, lengthy forced outages, and heavy costs to the utility. A large number of HV rotating machines failed due to stator winding problems, especially insulation deterioration. Electrical, mechanical, thermal, and environmental stresses remain major factors for the insulation degradation and accelerated aging [1]. These stresses impose, individually and in a collective manner, degradation of the stator winding insulation. It has been shown that the stator winding insulation is also one of the most vulnerable components used in an HV rotating machine. Surveys published in [2–4] show that about 30–40% of rotating machine failures are initiated from problems in the stator insulation. Further, a recent survey published in [5] shows that the percentage of stator related failures is about 60–70% for HV rotating machines. The reliability of rotating machines is almost defined by the condition of the stator winding insulation. Therefore, stator

insulation condition assessment is a vital step for the purpose of estimating residual life and reliability of rotating machine, and there has been a lot of effort in identifying the causes of stator insulation degradation and failure and finding methods for assessing the condition of stator insulation systems in the past decades [6].

Historically, several conventional methods and parameters for assessing the high-voltage insulation condition have been proposed in the literature, such as measurements of insulation resistance (IR) and polarization index (PI), recovery voltage and ramped-DC voltage, capacitance (C) and dissipation factor ($\tan \delta$), AC and DC hipot tests, and partial discharge tests, as suggested in some standards [6–10]. All these traditional methods are complementary but incomplete and, on the other hand, the nondestructive parameters used to evaluate the condition of stator winding insulation are still controversial [11]. Therefore, it is necessary to find the sensitive parameters and develop some new assessment methods for the deterioration condition indication and assessment of stator winding insulation.

Generally, there are two kinds of methods to assess the condition of stator winding insulation, which are the direct

method and the indirect method. A direct measure of the integrity of an insulation system may be the breakdown strength under AC stress, as determined through destructive testing, which damages the winding and makes the machine unserviceable. An indirect measure of an insulation system may be nondestructive diagnostic tests, after which the insulation remains serviceable. However, the credibility in estimating the insulation's condition using these nondestructive tests should be well correlated to the results of breakdown tests.

In order to check aging symptoms and correlation factors, such as electrical diagnostic parameters, partial discharge (PD) measurement test was employed to assess the condition of stator insulation in this paper. Several statistical parameters of phase resolved PD distribution, called fingerprint parameters, were proposed. A partial least square (PLS) statistical approach was used to investigate the correlation between the breakdown strength and the PD fingerprint parameters, which is also used to assess the insulation condition assessment of large generator stator winding. The effectiveness of this method was verified by the test results of generator stator bars.

This paper is organized as follows. Section 2 introduces the stator bar experimental specimens and accelerated aging model used in this study. Section 3 describes the parameters test for the condition estimation of large generator stator insulation. The characteristics of PD distributions are analyzed and the results are shown in Section 4. And subsequently, Section 5 discusses some experimental results. Finally, Section 6 provides a conclusion.

2. Stator Bar Specimens and Accelerated Aging Model

The insulation of the model stator bars was made of epoxy/mica resin rich with the thickness of 3.5 mm, and the conductor of model bars was made of aluminum with the length of 1200 mm. The structure of the stator bar is shown as Figure 1.

To study the characteristic parameters and aging regulations of stator bar, the model stator bars were used to perform the accelerating aging test in laboratory. As mentioned above, several stresses were imposed to stator bar insulation in operation. According to IEEE Std.275, thermal, mechanical vibration and environmental stresses were applied in sequence. Considering the aging process of stator bar during the real operation of large generator, a combined multistress test model was proposed, as shown in Figure 2. The electrical, thermal, and mechanical vibration stresses were imposed on stator bars simultaneously.

The acceleration aging tests were conducted on model stator bars at various stress levels for model bar simultaneity. Various stress levels were designed as: (a) electrical stress: aging electrical field intensity is 6.5 kV/mm and aging voltage is 22.75 kV; (b) thermal stress: aging temperature is 180°C; (c) mechanical stress: vibration frequency is 100 Hz and vibration extent is 1000 μm . In order to measure the nondestructive parameters, the insulation surface of ends of

the model bar specimens was coated with silicone carbide paint to suppress corona for application of electrical stress. As shown in Figure 1, the middle three sections of stator bar were used to measure dielectric loss parameters $\tan \delta$ and $\Delta \tan \delta$. The two ends of stator bar were used for the measurement of PD.

3. Parameters Measurement

In this study, PD measurement was performed after every aging cycle. Some PD parameters were employed to assess the condition of stator insulation under multistress aging. Breakdown voltage (VBD) test was performed on all specimens throughout the whole aging period to determine the remaining VBD after obtaining the PD parameters. In this way, the dielectric properties of ground wall insulation degradation with aging time could be obtained from every nondestructive measurement. The relation between the non-destructive parameters and the VBD could be studied based on the final measurement.

3.1. PD Measurement. In order to acquire the PD experimental data, the PD measurement system was proposed, as shown in Figure 3.

All the instruments were placed in a shielding room shown in the dashed frame. The PD analyzer and oscilloscope were connected to the measuring impedance through a high-frequency coaxial cable with characteristic impedance of 50 Ω . The PD measuring system is calibrated for each specimen before measurement by a standard calibrator manufactured by HAEFELY TRENCH.

3.2. AC Destructive Test. AC destructive experiment was performed to detect withstanding voltage ability of the stator bars. The specimens were broken down after each aging cycle in order to acquire the relation between the breakdown voltage (VBD) and the aging time. All breakdown tests were conducted according to the related IEC standard. The number of samples reduced gradually after each aging cycle due to breakdown, and the multistress accelerating aging finished after all samples were broken down. The test results demonstrated that the breakdown site was mainly located at the edge of the stator bar. The electrical, thermal, and mechanical stresses were the most severe at the edge of bar during the multistress accelerating aging, where the electrical field was also the highest during the breakdown test. Therefore, almost all the samples were broken down at their edges.

4. Measurement Analysis and Results

The characteristics of PD and PD detection techniques have gained popularity in the past years, which is now considered as one of the most useful diagnostic tools for generators. Efforts have been made in the course of years to relate PD magnitude, energy, and other parameters to insulation damage. Understanding the mechanisms of PD and degradation

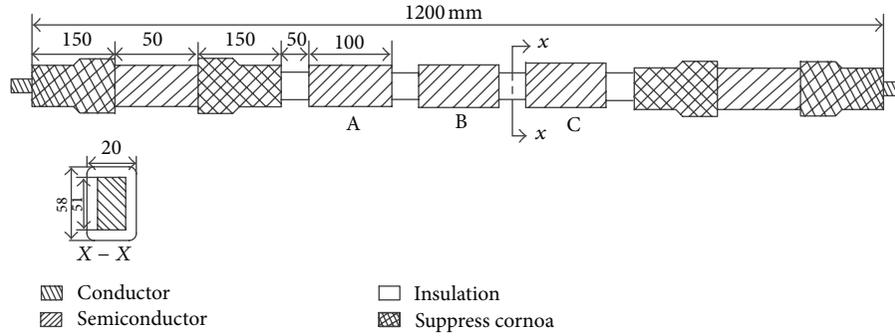


FIGURE 1: Structure of the stator bar.

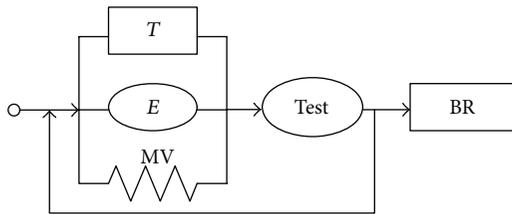


FIGURE 2: The multistress test model, ((T) thermal stress, (E) electrical stress, (MV) mechanical vibration, ($Test$) test parameters, and (BR) breakdown test).

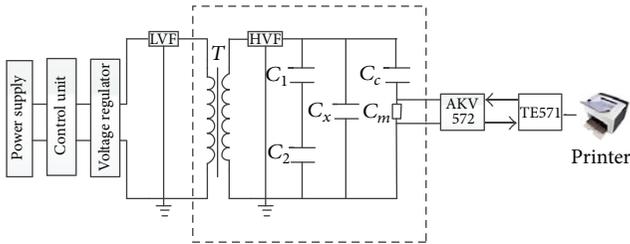


FIGURE 3: Schematic diagram of PD measurement system (LVF: low voltage filter, T: transformer, HVF: high voltage filter, C_x : specimen, C_c : coupling capacitor, C_m : measurement impedance, AKV572: analog amplifier, and TE571: PD analyzer).

of insulating materials is important to improve the reliability of diagnosis.

In recent years, advancement of computer hardware and pattern recognition techniques has provided automation and improvement of the PD interpretation process. As a result of the computer-aided processing [12, 13], massive amounts of PD measurements can be interpreted efficiently and reliably [14, 15]. The well-known pattern recognition methods applied to PD analysis are expert systems, neural networks, fuzzy classifiers, fractal models, statistical methods, and others [16]. Phase resolved plots directly reflect the pulses distribution properties of PD. Phase resolved PD analysis results were obtained and reported in this section. The PD statistical parameters are also noted in this study for PD diagnosis.

4.1. Phase Resolved PD Analysis. PD is a complex physical phenomenon, in which discharges are involved in electrically

weak regions of solid insulation materials. Much work has been done to understand and describe the nature of PD phenomena. In order to obtain meaningful data for pattern recognition, it is necessary to acquire the fingerprints of the PD signals under well-defined conditions for which the cause of PD is known. For many years, PD recognition was performed by visual examination, namely, on an oscilloscope screen. In recent years, the use of computer-aided processing technique has facilitated the automation of the recognition task. Phase resolved partial discharge (PRPD) measurement systems Especially have become very popular [17].

In phase-resolved PD measurement data, PD pulses are grouped by their phase angle with respect to power frequency sine wave. Consequently, the voltage cycles are divided into phase windows representing the phase angle axis (0 to 360°). The test voltage is to be held constant and the voltage phase angle is divided into a suitable number of small bins or windows. The digital PD detector is used to collect all the individual quasi-integrated pulses and quantifying each of these PD pulses by their discharge magnitude (q), the corresponding phase angle, or discharge epoch (φ), at which they occur and their number densities or discharge rates (n) over some chosen interval of time. The analysis software plots these data either as univariate distributions (functions of the phase positions φ) or as a bivariate distribution (such as φ - q - n plots). By taking appropriate averages of these statistical distributions, the observed PD patterns throughout the whole phase angle axis result in two- or three-dimensional patterns [18]. The main distributions, being function of the phase angle φ , are as follows. (a) Maximum pulse height phase resolved distribution $H_{q_{max}}(\varphi)$ and the pulse count distribution $H_n(\varphi)$ that represents the number of the observed discharges in each phase window as a function of the phase angle. (b) The mean pulse height distribution $H_{qn}(\varphi)$ which represents the average amplitude in each phase window as a function of the phase angle. $H_{qn}(\varphi)$ is derived from the total discharge amount in each phase window divided by the number of discharges in the same phase window.

Because of the phase-dependent behavior of PD generated under ac voltages conditions, several quantities as function of the phase angle can be used to describe the PD phenomena. The shapes of $H_{q_{max}}(\varphi)$, $H_{qn}(\varphi)$, $H_n(\varphi)$, and three-dimensional distribution $H_n(\varphi, q)$ changed with aging time. These distributions have characteristic shapes,

which vary with the type of the defects [19]. Experience has shown that the shape belonging to the positive half of the HV sine wave differs much from that of the negative half, so that their characteristic shapes are available for recognition purposes.

In fact, regardless of whether the recognition process is performed by a computer system or experts, standard PD analyses usually apply phase resolved PD pulse height and phase of occurrence distribution to extract information about the characteristics of defects, as well as PD induced aging of the insulating materials [17]. However, it is difficult to describe the changes in deterioration of specimen insulation quantitatively with them. Therefore, authors make an attempt to introduce statistical parameters for analyzing the aging extent of insulation.

4.2. Statistical Parameters Analysis. The profiles of $H_{q_{\max}}(\varphi)$, $H_n(\varphi)$, and $H_{q_n}(\varphi)$ characteristics provide information of defects in insulation, in which partial discharges happen. In theory, the shapes of $H_{q_{\max}}(\varphi)$, $H_n(\varphi)$, and $H_{q_n}(\varphi)$ distributions can be used to classify discharges, either by recognizing them by eye or by observing them with pattern classifying approach. However, a more distinctive method has been based on the application of PD statistical parameters (operators). The term ‘‘operator’’ is used in this paper to denote mathematical descriptions or parameters. These statistical parameters (operators) are nothing more than the mathematical concepts according to the definitions in [18] and there is no relation to the test condition. These statistical parameters (operators) have been proved to be effective for PD analysis [19], especially for detecting the kind of defect or treeing [20]. In this approach, several aspects of the shape of the distributions are described by mathematical functions, as shown in the following.

The first operator is skewness Sk, which describes the asymmetry of a distribution with respect to a normal distribution. $Sk = 0$ means symmetric, $Sk > 0$ means asymmetric with the left side large, and $Sk < 0$ means asymmetric with larger right side. Sk is defined as

$$Sk = \frac{\sum_{i=1}^N (q_i - \mu)^3 p_i}{\sigma^3 \sum_{i=1}^N p_i}, \quad (1)$$

where q_i is the recorded PD value and p_i the probability of frequency of appearance for that value q_i in time window i , μ is the mean $\mu = \sum q_i \cdot p_i / \sum p_i$, and σ is calculated by the variance $\sigma^2 = \sum (q_i - \mu)^2 \cdot p_i / \sum p_i$.

Ku indicates increased sharpness or the amount of concentration of the distribution with respect to the normal distribution. $Ku = 0$ means a normal distribution, $Ku > 0$ means a sharp distribution, and $Ku < 0$ means a flat distribution. Ku is defined as

$$Ku = \frac{\sum_{i=1}^N (q_i - \mu)^4 p_i}{\sigma^4 \sum_{i=1}^N p_i} - 3, \quad (2)$$

C_c is used to evaluate the difference in shape between the distributions both in the positive and negative half cycles. $C_c = 1$ means 100% shape symmetry and, on the other hand, $C_c = 0$ means total asymmetry. The following formula is used to calculate the C_c :

$$C_c = \frac{\sum_{i=1}^N q_i^+ q_i^- - \sum_{i=1}^N q_i^+ \sum_{i=1}^N q_i^- / N}{\sqrt{[\sum_{i=1}^N (q_i^+)^2 - (\sum_{i=1}^N q_i^+)^2 / N][\sum_{i=1}^N (q_i^-)^2 - (\sum_{i=1}^N q_i^-)^2 / N]}}, \quad (3)$$

where q_i^+ and q_i^- are the mean discharge magnitude in i phase window in positive and negative half cycles, respectively, and N is the number of phase window per half cycle.

The asymmetry, Asy of $H_{q_{\max}}(\varphi)$, $H_n(\varphi)$, and $H_{q_n}(\varphi)$ as the quotient of the mean discharge level in the positive and in the negative half of voltage cycle, is defined as

$$Asy = \frac{Q_s^- / N^-}{Q_s^+ / N^+}, \quad (4)$$

where Q_s^+ and Q_s^- are the sum of discharges of the $H_{q_{\max}}(\varphi)$, $H_n(\varphi)$, and $H_{q_n}(\varphi)$ distributions in the positive and the negative half of the voltage cycle; N^+ and N^- are the number of discharges of the $H_{q_{\max}}(\varphi)$, $H_n(\varphi)$ and $H_{q_n}(\varphi)$ distributions in the positive or negative half of the voltage cycle.

As aforementioned, Figures 4 and 5 show phase-position quantities of PD happening in various stator bars insulation, which had been conducted by multistress aging, respectively. An interpretation of the changes of PD statistical operators is listed in Table 1.

The test results indicate that all the specimens show the similar changing trend with little difference, so the test data of only one of the specimens subjected to all aging cycles are given in Figures 4–7.

Figures 4(a) and 4(b) show the positive and negative skewness changes of $H_{q_{\max}}(\varphi)$, $H_{q_n}(\varphi)$, and $H_n(\varphi)$ distributions with aging time. Each point in Figure 4 represents the skewness obtained at rated voltages of 6.6 kV. It is evident that the skewness decreases with the aging time no matter in positive half cycle or in negative half cycle. The changes of skewness of $H_{q_{\max}}(\varphi)$, $H_{q_n}(\varphi)$, and $H_n(\varphi)$ indicate that the change trend of voids is from flat to narrow in ground insulation. The $H_{q_n}(\varphi)$ distributions of dielectric-bounded cavities (square and flat) as well as those of electrode-bounded cavities are characterized by positive values of Sk. Further, $H_{q_n}(\varphi)$ processed for multiple cavities is also represented by positive Sk values [21].

The relations between the kurtosis of $H_{q_{\max}}(\varphi)$, $H_{q_n}(\varphi)$, and $H_n(\varphi)$ distributions and aging time are shown in Figure 5. No linear correlation between the kurtosis of PD distributions and aging time was found. The kurtosis Ku proved to be a good discriminator for multiple cavities [21]. Figure 5 shows that in the case of multiple cavities, the Ku value of the $H_{q_{\max}}(\varphi)$, $H_{q_n}(\varphi)$, and $H_n(\varphi)$ distribution is clearly negative. The result is in accord with the literature [21].

Figure 6 shows the change of cross-correlation factor of PD phase resolved distributions with aging time. It can be seen from Figure 6 that the cross-correlation factor of the $H_{q_{\max}}(\varphi)$, $H_{q_n}(\varphi)$, and $H_n(\varphi)$ distributions decreased

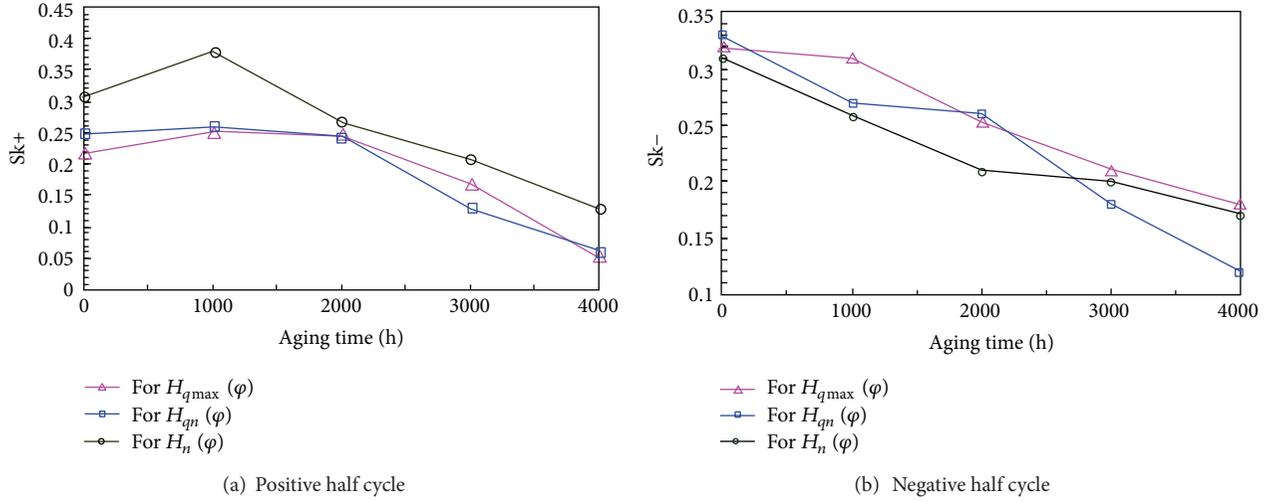


FIGURE 4: Skewness change of phase-resolved distributions of PD.

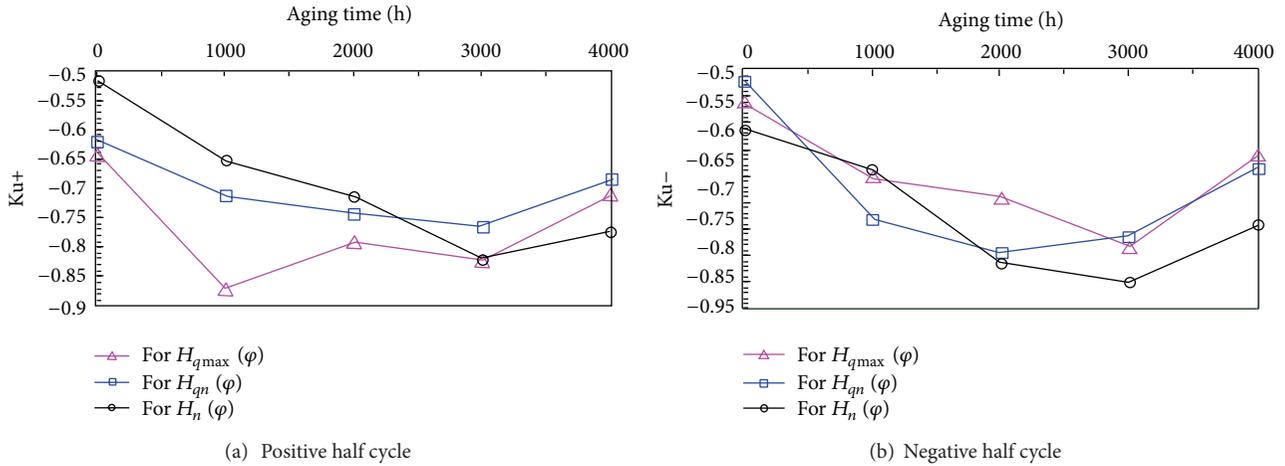


FIGURE 5: Kurtosis change of phase-resolved distributions of PD.

TABLE 1: Changes of PD statistic operators of $H_{q_n}(\varphi)$ distributions during aging.

Aging time	Sk+	Sk-	Ku+	Ku-	Asy	C_c
0 h	0.25	0.33	-0.62	-0.562	-0.29	0.51
1000 h	0.262	0.27	-0.711	-0.7	-0.12	0.46
2000 h	0.244	0.235	-0.743	-0.74	-0.04	0.44
3000 h	0.131	0.167	-0.765	-0.76	0.07	0.33
4000 h	0.061	0.183	-0.684	-0.68	0.01	0.41

nonmonotonously with aging time. Figure 7 indicates the asymmetric degree of $H_{q_{max}}(\varphi)$, $H_{q_n}(\varphi)$, and $H_n(\varphi)$ increase with aging time.

These are better visible during the time that the discharges are stable. The above mentioned characteristics may prove to be useful for the development of a system for the diagnose of defects in insulating constructions.

4.3. Analysis of Relationship between VBD and PD Parameters. Breakdown voltage (VBD) is considered to be the most objective criterion for residual life of large generator. However, breakdown test is destructive. Recent efforts have been directed towards the condition assessment of stator insulation by using nondestructive insulation diagnostic measurement data. A review of extensive research works in this area reveals that the correlation between the nondestructive parameters and residual life (insulation breakdown voltage) is still controversial among researchers. As aforementioned, PD measurement has gained popularity over the years and is now considered as one of the most useful tools for insulation diagnosis and aging assessment of generators. However, the credibility in estimating the insulation's aging condition using PD tests should be well correlated to the results of breakdown tests. So PD data was systematically analyzed using a partial least square (PLS) approach in this study and the relationships between the breakdown voltage VBD and PD data were investigated.

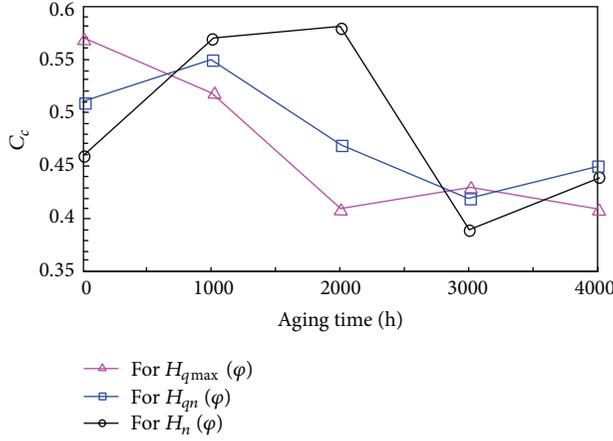


FIGURE 6: C_c change of phase-resolved distributions of PD.

Statistical data analysis has been widely used in establishing models from experimental or historical data. Typical problems in multivariate statistical analysis are high dimensionality and collinearity in a sparse sample data set. The partial least squares (PLS) modeling method is one of the most useful measures for overcoming these problems. PLS is a multivariate statistical data analysis and regression method, which uses projection into latent variables to reduce high-dimensional and strongly correlated data to a much smaller data set that can then be interpreted. Consider a given data base \mathbf{X} of n observations parameterized by p parameters and an output variable \mathbf{Y} , in which all variables are normalized. Only the case of a single output of n by 1 dimension is considered in the present study despite that the PLS regression can also be used with several output variables. In fact, this projection method decomposes variables of high collinearity into one-dimensional variables (input score vector t and output score vector u). The composition of X and Y by score vectors is formulated as follows:

$$X = \sum_{h=1}^m t_h p_h^T + E, \quad (5)$$

$$Y = \sum_{h=1}^m u_h q_h^T + F, \quad (6)$$

where p and q are loading vectors, and E and F are residuals. This relation is known as the PLS outer relation. The relation between score vectors t_h and u_h is known as the inner relation. The problems of sparse sample data set and multicollinearity are therefore solved effectively in regression analysis. For the m choice, a cross validation method [22] can be applied. The threshold of E variance can also be used as stopping criterion.

The PLS outer relation can be further expressed as

$$X = TP^T + E_m = t_1 p_1^T + t_2 p_2^T + \cdots + t_k p_k^T + E_m, \quad (7)$$

$$Y = UQ^T + F_m = u_1 q_1^T + u_2 q_2^T + \cdots + u_k q_k^T + F_m.$$

If a linear relation exists between latent variable matrices \mathbf{U} and \mathbf{T} , then $\mathbf{U} = \mathbf{T}\mathbf{B}$, where \mathbf{B} is a diagonal matrix, termed

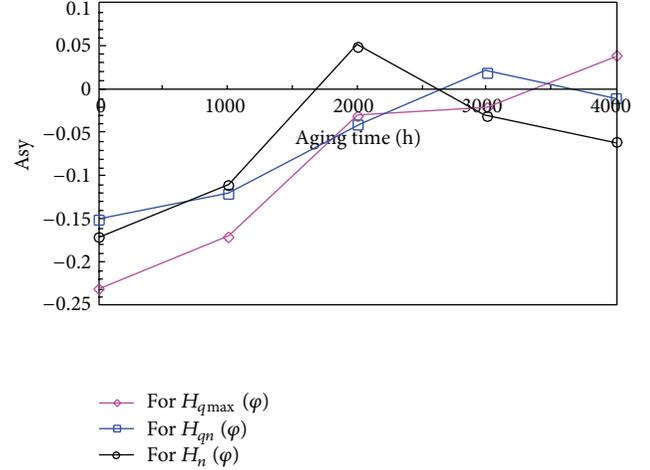


FIGURE 7: Asy change of phase-resolved distributions of PD.

linear PLS model, detailed in [23]. In a previous nonlinear PLS model, [22] indicated $u_h = f_h(t_h) + r_h$, the inner relation between score vectors t and u was defined as

$$U = f(T) + R \quad (8)$$

$$= [f_1(t_1), f_2(t_2), \dots, f_m(t_m)] + [r_1, r_2, \dots, r_m].$$

The function $f(\cdot)$ expresses a linear or nonlinear function, detailed in [22, 24].

Over the multistress aging duration (0, 1000, 2000, 3000, and 4000 hours), PD and destructive tests were performed on 15 aging stator bar specimens after each aging period and a set of data was collected and testing data is shown in Table 2. There are six input variables and one output variable for the analysis model. As in the aforementioned results, the changes of statistical parameters of $H_{q_{max}}(\varphi)$, $H_{q_n}(\varphi)$, and $H_n(\varphi)$ have similar trend curves with aging time. So the statistical parameters of $H_{q_n}(\varphi)$, such as $Sk+$, $Sk-$, $Ku+$, $Ku-$, Asy , and C_c are used as the input variables in this study. The output is the breakdown voltage VBD.

To avoid the mistakes of analysis caused by unreal variations due to different units and/or scalar levels, a centered-compress (standardization) method was adopted in data preprocessing. The data set (including explanatory and response variables) was firstly standardized by

$$x_{x_j}^* = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad (9)$$

$$s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2,$$

where $i = 1, \dots, n$; $j = 1, \dots, p$, n are the sample numbers, p is the explanatory variable numbers and s_j is the standard deviation of explanatory variable x_j . After being standardized, the mean and variance of each variable were equal to 0 and 1, respectively. The standardized preprocessing data is shown in Table 3.

Based on the PLS method, the first principal component t_1 is extracted from data. The t_1 synthesizes information

TABLE 2: Testing data.

No.	Aging duration (hours)	Sk+	Sk-	Ku+	Ku-	Asy	C_c	VBD (kV)
1	0	0.25	0.33	-0.64	-0.52	-0.15	0.51	91.3
2	1000	0.26	0.27	-0.87	-0.77	-0.12	0.55	90.95
3	2000	0.24	0.27	-0.8	-0.81	-0.04	0.47	90.56
4	3000	0.13	0.18	-0.82	-0.78	0.02	0.42	89.75
5	4000	0.06	0.12	-0.7	-0.62	-0.01	0.45	57.1
6	0	0.24	0.36	-0.62	-0.52	-0.19	0.54	92.3
7	1000	0.252	0.259	-0.78	-0.72	-0.17	0.49	91.1
8	2000	0.22	0.2	-0.82	-0.85	-0.11	0.51	90.86
9	3000	0.11	0.191	-0.86	-0.81	-0.05	0.46	89.55
10	4000	0.07	0.144	-0.68	-0.71	0.02	0.44	63.1
11	0	0.251	0.366	-0.62	-0.56	-0.17	0.52	91.5
12	1000	0.26	0.3	-0.7	-0.72	-0.06	0.48	89.97
13	2000	0.21	0.22	-0.74	-0.74	-0.02	0.41	90.96
14	3000	0.13	0.174	-0.76	-0.71	0.01	0.45	90.81
15	4000	0.058	0.131	-0.68	-0.66	0.04	0.42	56.5

TABLE 3: Standardization preprocessing data.

No.	Sk+	Sk-	Ku+	Ku-	Asy	C_c	VBD
1	0.8446	1.1555	1.1819	1.7218	-1.0575	0.7962	0.5173
2	0.9702	0.4206	-1.5547	-0.6696	-0.6768	1.6975	0.4910
3	0.7191	0.6655	-0.7218	-1.0522	0.3384	-0.1052	0.4617
4	-0.6621	-0.6819	-0.9599	-0.7653	1.0998	-1.2318	0.4008
5	-1.5411	-1.4168	0.4680	0.7653	0.7191	-0.5558	-2.0546
6	0.7191	1.5230	1.4199	1.7218	-1.5651	1.4722	0.5925
7	0.8697	0.2858	-0.4839	-0.1913	-1.3113	0.3455	0.5023
8	0.4679	-0.4369	-0.9598	-1.4349	-0.5499	0.7962	0.4842
9	-0.9133	-0.5471	-1.4357	-1.0522	0.2115	-0.3305	0.3857
10	-1.4155	-1.1228	0.7060	-0.0957	1.0998	-0.7811	-1.6033
11	0.8572	1.5965	1.4200	1.3392	-1.3113	1.0215	0.5324
12	0.9702	0.7880	0.4680	-0.1913	0.0846	0.1202	0.4173
13	0.3424	-0.1919	-0.0079	-0.3826	0.5922	-1.4571	0.4918
14	-0.6621	-0.7554	-0.2459	-0.0957	0.9729	-0.5558	0.4805
15	-1.5662	-1.2821	0.7060	0.3826	1.3536	-1.2318	-2.0997

of PD parameters matrix X and characterizes breakdown voltage VBD maximum. Three components (t_1 , t_2 , and t_3) are extracted by these means. Various data messages of the stator winding condition parameters are shown in Figures 8–12.

5. Discussions

The PLS is a relative new data analysis method. Graphical analysis function of PLS is easy to understand various data messages of PD condition parameters. The relation between the PD parameters and breakdown voltage can be presented by t/u plot. The fitting results of each pair of latent variables (t and u) with the PLS inner mapping are shown in Figure 8. In order to analyze the experiments adequately, three components are extracted. The manner of component extraction using the PLS method shows that the information of t_1 extracted from raw data is maximum

(contributions rate is 54.7% by computing); score t_2 takes second place (contributions rate is 33.7%) in comparison with the contribution rate of t_3 , 7.47%. Two scores are determined by the cross-validation method. In other words, the scores t_1 and t_2 (accumulative contributions rate is 88.4%) are enough to relate breakdown voltage VBD of generator stator.

In Figure 8(a), 5, 10, and 15 are data samples of 4000 h aging duration. 4, 9, and 14 are samples with 3000 h aging duration. Figure 8 indicates that relation between the PD parameters and breakdown voltage (VBD) is linear approximately at the initial aging duration. The linear relation is not obvious with increase of aging time (especially after 4000 h). So the relation between the PD parameters and breakdown voltage cannot be characterized with linear relation simply.

The $w_1^*c_1/w_2^*c_2$ plot can express the correlativity between the PD parameters X and breakdown voltage (VBD) Y . The shorter distance stands for the better correlativity between

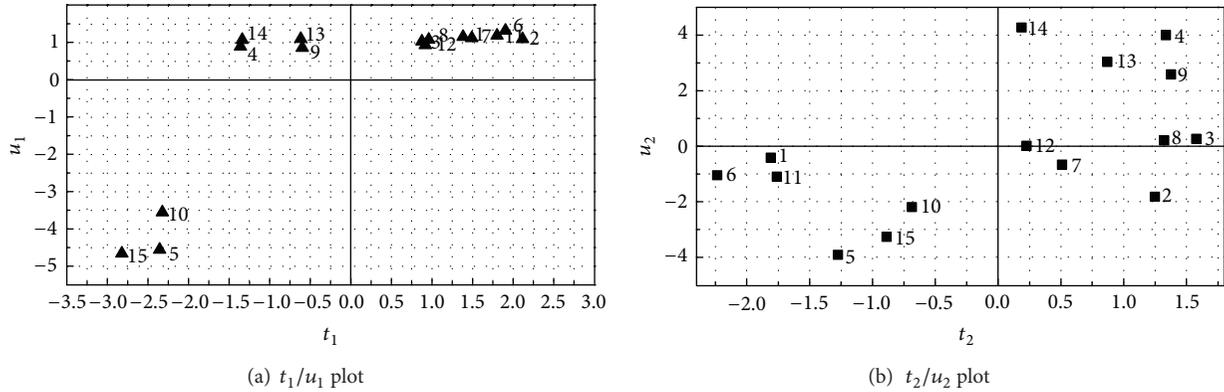


FIGURE 8: t/u plot of PD parameters.

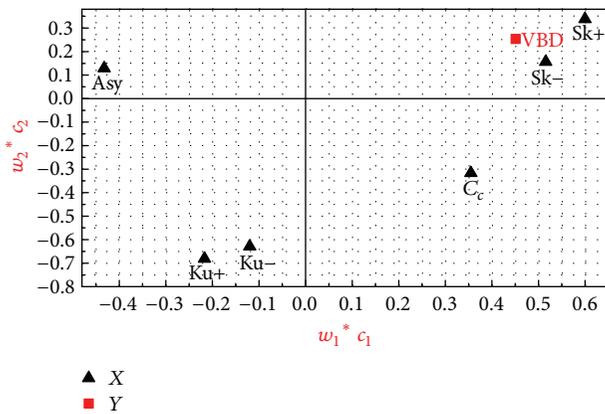


FIGURE 9: $w_1^*c_1/w_2^*c_2$ plot of PD parameters.

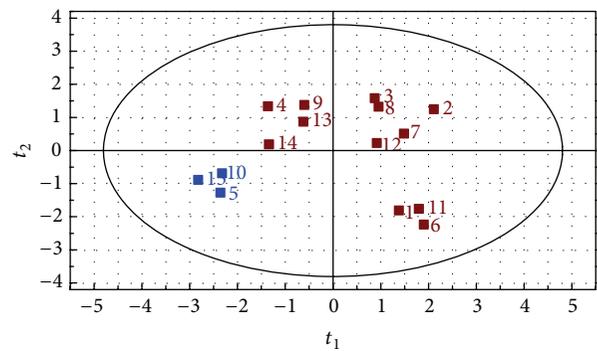


FIGURE 11: t_1/t_2 plot of PD parameters.

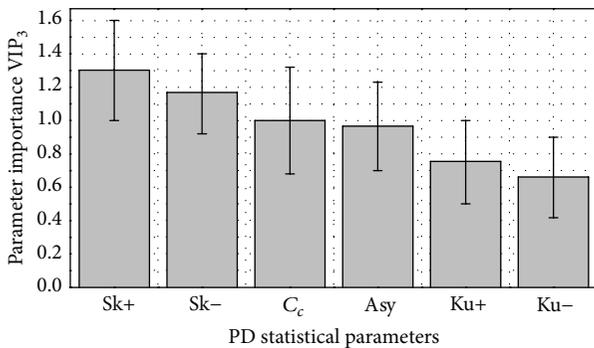


FIGURE 10: VIP plot of PD parameters.

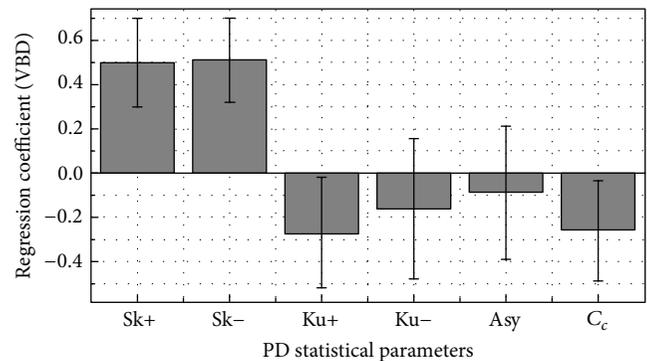


FIGURE 12: Regression coefficients plot of PD parameters.

the X and Y. Figure 9 indicates that there is a positive and stronger relation between skewness Sk and breakdown voltage VBD. This is in conformity with the experimental results of Figure 4, which shows that the changing trend of skewness Sk of PD distribution is sensitive and clear. The parameter C_c has negative stronger relation with VBD. Ku and Asy are hardly related with VBD.

Moreover, PLS method computes the influence on Y (VBD) of every term (PD parameters) in the model, called

VIP (variable importance in the projection). The VIP values indicate the importance of terms in the model and the correlation of the terms to all the responses. Figure 10 shows that the effect of Sk is mostly in regression modeling and secondly in C_c and so on.

Distributing data samples can be observed easily using t_1/t_2 plot in the Hotelling's T^2 ellipse box. Figure 11 indicates

that there are not abnormal data in experimental data. The blue data point is most different with other data points. Clearly, 5, 10, and 15 are data samples of 4000 h aging duration. The breakdown voltage (VBD) of these specimens is the lowest. Moreover, the regression coefficients of a simple PLS fitting model are shown in Figure 12. Although this paper only presents a simple regression model that is not enough to characterize the life model of stator bar, some novel ideas can be directed for further research work. In the near future, we will use improved nonlinear PLS modeling method to characterize life model by increasing aging time.

6. Conclusions

Nondestructive measurement has gained popularity and is now considered as one of the most useful tools for insulation diagnosis and aging assessment of generators. In order to check aging symptoms and correlation factors such as electrical diagnostic parameters, partial discharge (PD) measurement test was employed to assess the condition of stator insulation in this paper. The model stator bars were used to perform the accelerating aging test. The characteristics of the shapes of phase-resolved PD distributions were investigated. The relationships between the breakdown voltage VBD and PD statistical parameters data were systematically analyzed using a partial least square (PLS) approach.

The test results indicate that the phase-resolved PD distributions change much more remarkably with the aging time and can be used to describe the PD phenomena. The profiles of the phase-resolved PD distributions characteristics provide information about defects in insulation in which partial discharges are taking place.

Moreover, the statistical analysis of PD distributions is meaningful for aging-condition assessment of stator bar insulation. It is found clearly that the statistical parameter S_k decreases continuously with the aging time. Comparing with K_u , A_{sy} , and C_c , the statistical parameter S_k of PD distributions is more sensitive to the change of PD activities occurring in the insulation of stator bar during the aging. Therefore, the parameter S_k would be taken as a characteristic parameter for the aging-condition assessment of stator winding insulation.

Because a direct indicator of insulation condition is the breakdown voltage under ac stresses, the breakdown test was also performed after each aging cycle in order to acquire the true aging condition. Correlation between nondestructive test data and destructive data was analyzed by means of PLS statistical model. The analysis results indicate that the PD statistical parameters are correlated to the results of breakdown tests. It is evident that the correlation coefficient of breakdown voltage with the PD statistical parameter S_k was the largest among all the correlation coefficients, which shows that the statistical parameter S_k of PD distributions is more suitable for the aging-condition assessment of the stator-bar insulation than other PD statistical parameters (comparing with K_u , A_{sy} , and C_c).

Consequently, the mentioned above PD characteristics and PLS analysis approach may prove to be useful for the development of a system for the condition assessment of large generator stator insulation.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant no. 51177110) and Shanghai Natural Science Foundation of China (Grant no. 11ZR1439500).

References

- [1] R. Morin and R. Bartnikas, "Multistress aging of stator bars in a three-phase model stator under load cycling conditions," *IEEE Transactions on Energy Conversion*, vol. 27, no. 2, pp. 374–381, 2012.
- [2] P. O'Donnell, "Report of large motor reliability survey of industrial and commercial installations, part I and part II," *IEEE Transactions on Industry Applications*, vol. IA-21, no. 4, pp. 853–872, 1985.
- [3] K. Kimura, T. Iwabuchi, K. Morooka, and Y. Ishikawa, "Fundamentals—a useful index for estimating residual life of motor insulation," *IEEE Electrical Insulation Magazine*, vol. 6, no. 2, pp. 29–35, 1990.
- [4] W. Liu, E. Schaeffer, D. Averty, and L. Loron, "A new approach for form wound machine goundwall insulation diagnosis by means of high frequency model parameter monitoring," in *Proceedings of the IEEE International Symposium on Industrial Electronics (ISIE '07)*, pp. 1270–1275, June 2007.
- [5] C. Sumereder, "Statistical lifetime of hydro generators and failure analysis," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 15, no. 3, pp. 678–685, 2008.
- [6] G. C. Stone, "A perspective on online partial discharge monitoring for assessment of the condition of rotating machine stator winding insulation," *IEEE Electrical Insulation Magazine*, vol. 28, no. 5, pp. 8–13, 2012.
- [7] S. A. Bhumiwat, "Insulation resistance and polarization of rotating machines," in *Proceedings of the 30th Electrical Insulation Conference (EIC '11)*, pp. 249–253, Annapolis, Md, USA, June 2011.
- [8] K. Younsi, P. Neti, M. Shah et al., "On-line capacitance and dissipation factor monitoring of AC stator insulation," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 17, no. 5, pp. 1441–1452, 2010.
- [9] R. Houtepen, L. Chmura, J. J. Smit, B. Quak, P. P. Seitz, and E. Gulski, "Estimation of dielectric loss using damped AC voltages," *IEEE Electrical Insulation Magazine*, vol. 27, no. 3, pp. 20–25, 2011.
- [10] E. David and L. Lamarre, "Progress in DC testing of generator stator windings: theoretical considerations and laboratory tests," *IEEE Transactions on Energy Conversion*, vol. 25, no. 1, pp. 49–58, 2010.
- [11] C. Sumereder, M. Muhr, M. Marketz, C. Rupp, and M. Krüger, "Unconventional diagnostic methods for testing generator stator windings," *IEEE Electrical Insulation Magazine*, vol. 25, no. 5, pp. 18–24, 2009.
- [12] P. Basappa, V. Lakdawala, K. Agarwal, G. A. Gerdin, and G. Shivaraj, "A novel system for generation, acquisition and analyses of partial discharges," in *Proceedings of the 30th Electrical Insulation Conference (EIC '11)*, pp. 166–169, Annapolis, Md, USA, June 2011.
- [13] A. Jambula, V. Nimbole, V. Lakdawala, and P. Basappa, "Development of a digital acquisition system for partial discharge measurement and analyses," in *Proceedings of the IEEE Electrical Insulation Conference (EIC '09)*, pp. 60–63, Montreal, Canada, June 2009.

- [14] Y. Hu, M. Chiampi, G. Crotti, and A. Sardi, "Setting-up of a characterization system for digital PD measuring instruments," in *Proceedings of the Conference on Precision Electromagnetic Measurements (CPEM '10)*, pp. 659–660, June 2010.
- [15] M. Bélec, C. Hudon, and D. N. Nguyen, "Statistical analysis of partial discharge data," in *Proceedings of the IEEE International Symposium on Electrical Insulation (ISEI '06)*, pp. 122–125, June 2006.
- [16] B. Yazici, "Statistical pattern analysis of partial discharge measurements for quality assessment of insulation systems in high-voltage electrical machinery," *IEEE Transactions on Industry Applications*, vol. 40, no. 6, pp. 1579–1594, 2004.
- [17] S. Das and P. Purkait, " $\varphi - q - n$ pattern analysis for understanding partial discharge phenomena in narrow voids," in *Proceedings of the IEEE Power and Energy Society 2008 General Meeting: Conversion and Delivery of Electrical Energy in the 21st Century (PES '08)*, pp. 1–7, July 2008.
- [18] E. Gulski, "Computer-aided measurement of partial discharges in HV equipment," *IEEE Transactions on Electrical Insulation*, vol. 28, no. 6, pp. 969–983, 1993.
- [19] B. Yue, X. Chen, Y. Cheng, J. Song, and H. Xie, "Diagnosis of stator winding insulation of large generator based on partial discharge measurement," *IEEE Transactions on Energy Conversion*, vol. 21, no. 2, pp. 387–395, 2006.
- [20] F. H. Kreuger, E. Gulski, and A. Krivda, "Classification of partial discharges," *IEEE Transactions on Electrical Insulation*, vol. 28, no. 6, pp. 917–931, 1993.
- [21] E. Gulski and F. H. Kreuger, "Computer-aided recognition of discharge sources," *IEEE Transactions on Electrical Insulation*, vol. 27, no. 1, pp. 82–92, 1992.
- [22] S. Wold, J. Trygg, A. Berglund, and H. Antti, "Some recent developments in PLS modeling," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 131–150, 2001.
- [23] K. Helland, H. E. Bernsten, O. S. Borgen, and H. Martens, "Recursive algorithm for partial least squares regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 14, no. 1–3, pp. 129–137, 1992.
- [24] G. Baffi, E. B. Martin, and A. J. Morris, "Non-linear projection to latent structures revisited (the neural network PLS algorithm)," *Computers and Chemical Engineering*, vol. 23, no. 9, pp. 1293–1307, 1999.

Research Article

Wavelet Leaders Based Vibration Signals Multifractal Features of Plunger Pump in Truck Crane

Wenliao Du, Chunyan Yang, Ansheng Li, and Liangwen Wang

School of Mechanical and Electronic Engineering, Zhengzhou University of Light Industry, No. 5 Dongfeng Road, Zhengzhou 450002, China

Correspondence should be addressed to Wenliao Du; dwenliao@zzuli.edu.cn

Received 30 May 2013; Accepted 18 July 2013

Academic Editor: Yixiang Huang

Copyright © 2013 Du Wenliao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Vibration signal of plunger pump in truck crane is a typical nonlinear, nonstationary signal, and the multifractal features are a powerful tool for depicting the geometry features of such signals. The wavelet-leaders based multifractal features extraction method is an attractive tool, which has solid theoretical mathematical support and a simplified calculation procedure. The wavelet-leaders based multifractal features extraction method is compared for the original signal and the denoised signal, and the statistical performance of the obtained features is also introduced based on block bootstrap technology. The effectiveness of the wavelet leaders based method is first verified for a traditional multifractal signal. Then, the application of the proposed method to the pump vibration signals of three working conditions of truck crane: lifting, rotating, and luffing is discussed. The result shows that the geometry features of vibration signals can be obtained with wavelet leaders multifractal analysis method, and the denoising process improves the multifractal features.

1. Introduction

In truck crane, the plunger pump is an essential apartment, and the quality of the pump affects the performance of the hydraulic system, even the whole equipment, directly. So, it is very important to monitor the condition and diagnose the fault of the pump. In order to achieve this goal, a considerable research focused on the vibration signals [1, 2], and extracting the features relevant to the mechanical conditions is of great consequence. Mechanical vibration signal is a typical nonlinear and unstable signal. The traditional feature extraction methods based on linear system do often fail to work effectively. As multifractal features can fully display the distribution of signal's singularities, the geometric characters and the local scaling behaviors [3, 4], wavelet transform based multifractal analysis approaches, are widely used.

Recently, a new multifractal analysis method based on wavelet leaders was proposed by Lashermes, Jaffard, and Wendt [5–7]. Based on discrete wavelet transform (DWT), the new method has solid theoretical mathematical support and complicated calculation is avoided. Furthermore, it successfully describes the characters of spectrum on a full

domain (for the negative orders qs and the positive orders qs). However, in practical applications, the wavelet leaders based vibration analysis also has some unknown characters. First, the mechanical vibration signal always consists of noise. The effect of the noise is worth studying. Second, in the classic reference [6], the block bootstrap based performance access method is employed with a fixed block length. But the dependence of the coefficients on different levels of wavelet is not the same. The selection of the optimal block length of wavelet coefficients vector is another question. The detailed introduction about the previous questions will be researched in this paper, in the following, and comparison studies about wavelet leaders based multifractal features extraction for original and denoised vibration signals of plunger pump in truck crane will be carried out.

This paper is organized as follows. The symptoms of vibration signal of plunger pump in truck crane and wavelet leaders based multifractal features are briefly introduced in Section 1. Section 2 presents the theory of extracting multifractal features with wavelet leaders. In Section 3, a detailed discussion on the scheme of wavelet leaders based vibration signals multifractal features extraction is presented.

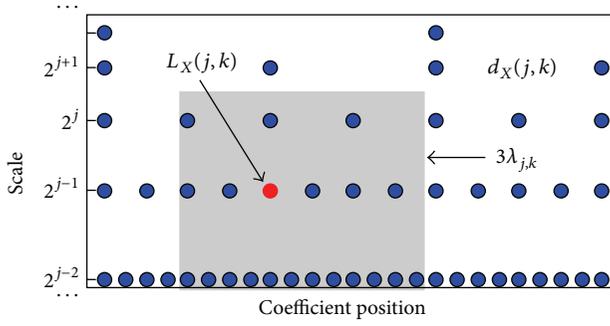


FIGURE 1: Definition of wavelet leaders.

In Section 4, the effectiveness of the proposed multifractal features extraction method is tested by using plunger pump signals, and the results by direct multifractal features extraction are also given for comparison. A conclusion is given in Section 5.

2. Wavelet Leaders Based Multifractal Features

2.1. Introduction of Wavelet Leaders. Suppose a function $\psi_0(t)$ with the vanishing moment N_ψ satisfies the admissible condition then $\psi_0(t)$ can be selected as mother wavelet. Let $\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k), j \in \mathbb{Z}, k \in \mathbb{Z}\}$ denote templates of $\psi_0(t)$ dilated to scales 2^j and translated to time positions $2^j k$, and they form an orthonormal basis of $L^2(\mathbb{R})$. For signal $X = \{x_k, k \in \mathbb{Z}\}$, the discrete wavelet transform is

$$d_X(j, k) = \int_{\mathbb{R}} X(t) 2^{-j} \psi_0(2^{-j}t - k) dt. \quad (1)$$

Further, assume that $\psi_0(t)$ has a compact time support, and define dyadic interval $\lambda = \lambda_{j,k} = [k \cdot 2^j, (k+1) \cdot 2^j]$. Let 3λ denote the union of the interval λ with its two adjacent dyadic intervals: $3\lambda_{j,k} = \lambda_{j,k-1} \cup \lambda_{j,k} \cup \lambda_{j,k+1}$. Then, the wavelet leader of position k in j scale is defined as [5] $L_X(j, k) \equiv L_\lambda = \sup_{\lambda' \subset 3\lambda} |d_{X,\lambda'}|$. Hence, $L_X(j, k)$ consists of the largest wavelet coefficient $d_X(j', k')$, which is calculated at all finer scales $2^{j'} \leq 2^j$ within a narrow time neighborhood $(k-1) \cdot 2^{j'} \leq 2^{j'} \cdot k' < (k+2) \cdot 2^{j'}$, as shown in Figure 1.

2.2. Wavelet Leaders Based Multifractal Features Extraction. Mechanical vibration signal has a typical nonlinear and nonstationary character. When the machine state changes from one work condition to another work condition or fault state, a sudden change always appears in vibration signal. The multifractal analysis can describe such special character of the signal geometric structure [8]. From the scaling exponents function, whether the signal is a single fractal or a multifractal signal can be judged, and the function also reflects the singularity to a certain extent. The longitudinal and the horizontal ordinates of the multifractal spectrum have concise physical meanings, which depict the uniformity of signal measure and reflect the strength of singularity.

For signal X , the wavelet leader of position k in j scale is denoted by $L_X(j, k)$. Let $S_L(q, j)$ be the structure functions,

and the scaling exponents are denoted by $\zeta_L(q)$ [9]; here, q is the order of these multiresolution quantities

$$S_L(q, j) = \frac{1}{n_j} \sum_{k=1}^{n_j} |L_X(j, k)|^q, \quad (2)$$

$$\zeta_L(q) = \liminf_{j \rightarrow 0} \left(\frac{\log_2 S_L(q, j)}{j} \right).$$

The multifractal spectrum $D(h)$ can be obtained by the Legendre transform of the scaling exponents

$$D(h) = \inf_{q \neq 0} (1 + qh - \zeta_L(q)). \quad (3)$$

In addition, with the Chhabra algorithm, the Legendre transform can be avoided, and the empirical expression of multifractal spectrum is obtained [10]

$$\widehat{D}(q) = \sum_{j=j_1}^{j_2} w_j U^L(j, q) \quad (4)$$

$$\widehat{h}(q) = \sum_{j=j_1}^{j_2} w_j V^L(j, q),$$

where $U^L(j, q) = \sum_{k=1}^{n_j} R_X^q(j, k) \log_2 R_X^q(j, k) + \log_2 n_j$, $V^L(j, q) = \sum_{k=1}^{n_j} R_X^q(j, k) \log_2 L_X(j, k)$, and $R_X^q(j, k) = L_X(j, k)^q / \sum_{k=1}^{n_j} L_X(j, k)^q$. The weight w_j satisfies $\sum_{j=j_1}^{j_2} j w_j \equiv 1$ and $\sum_{j=j_1}^{j_2} j w_j \equiv 1$ and can be expressed as $w_j = b_j / (V_0 j - V_1) / (V_0 V_2 - V_1^2)$; here, $V_i = \sum_{j=j_1}^{j_2} j^i b_j$, $i = 0, 1, 2$, and b_j is arbitrary positive number reflecting the confidence levels of $\log_2 S^L(j, q)$.

3. Procedures of Vibration Signals Multifractal Features Extraction

In order to get multifractal features from the vibration signal, the wavelet leaders based multifractal analysis with a wavelet denoising algorithm and bootstrap based performance evaluation are proposed and examined. The procedure for applying the proposed method is given in detail next.

Step 1 (denoising with the wavelet method). In actuality, the vibration signals are often contaminated with noise. It is verified that the noise can affect the value of fractal and further cause the analysis error. Thus, the contaminated signal denoising is one of the important steps in fractal analysis. Due to its good localization and multiresolution features in the time-frequency domain, the wavelet transform has been widely used. In this step, the selection of the wavelet basis and the decomposition level are based on the minimization of the cost functional of entropy [11].

Step 2 (multifractal features extraction). As mentioned in Section 2, the scaling exponents and the multifractal spectrum can reflect the singularity of the vibration signal, so they are selected as the features for condition monitoring and fault diagnosis. In this step, the selection of related parameters must be considered. First, the choice of mother wavelet and

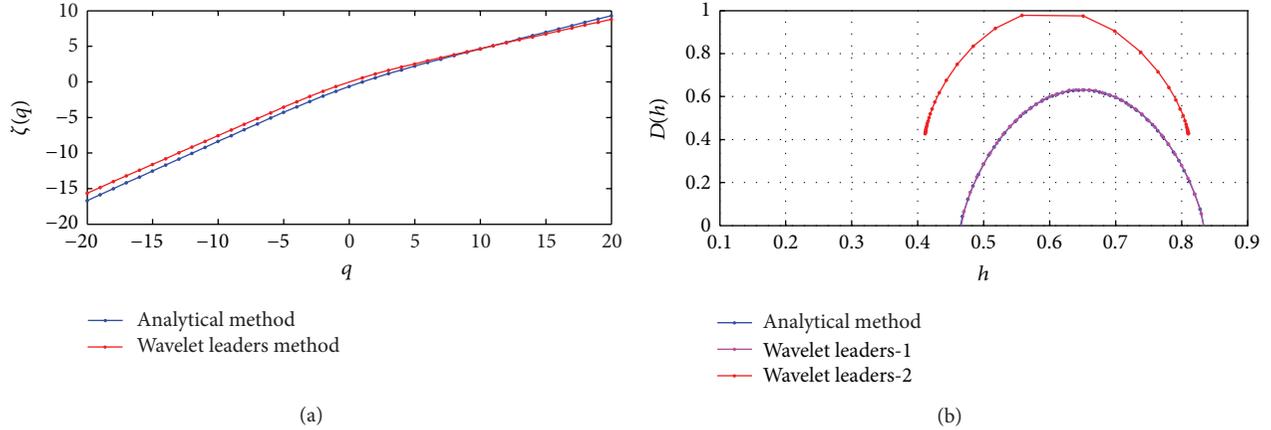


FIGURE 2: Scaling exponents and multifractal spectrum of the “devil staircase.” (a) Scaling exponents obtained by analytical method and wavelet leaders method, respectively. (b) Multifractal spectrum obtained by analytical method, wavelet leaders with Legendre transform method (wavelet leaders-1), and wavelet leaders with Chhabra regression method (wavelet leaders-2), respectively.

its vanishing moment is mainly based on the tradeoff between the relief of frequency aliasing and the preservation of spatial characterization. Second, the orders qs are chosen in a scope that avoids the linearization effect in the estimation procedure of scaling exponents, which is obtained by $\pm\sqrt{2/|c_2|}$ obtained with parabolic approximation [12]; here, c_2 is one of the expansion coefficients of $\zeta_L(q)$ and can be estimated on the vibration data set by bootstrap technology. Third, scaling range of regression in empirical expression of multifractal spectrum is another issue, which can be determined by the plot of $\zeta_L(q)$ versus $\log_2 2^j = j$. When the data actually show a scaling behavior, the corresponding j is appropriate.

Step 3 (performance evaluation of multifractal features). The scaling exponents and the multifractal spectrum belong to a statistical character, and performance analysis of these features is a principal means for exploring the stability and the confidence level. For the practical signal, the confidence level cannot be obtained by analytical method. The nonparameter bootstrap method is used to evaluate the statistical values in this step. Because the wavelet leaders in the same decomposition scale in general cannot satisfy the independent and identically distributed (*i.i.d.*), the block bootstrap technology is adopted for the wavelet leaders, and then the multifractal features with the produced pseudosamples are extracted; finally, the confidence level is obtained by these statistical values.

An optimal block length of wavelet leaders vector is required for the block bootstrap method. Here, based on the flat-top lag windows presented by Politis and White [13], a spectrum estimation approach is adopted to estimate the optimal block length, which guarantees a minimum mean square error of block bootstrap variance estimates.

4. Simulation Study and Plunger Pump Working Conditions Analysis

4.1. Simulation Signal Study. In this section, we select a popular fractal signal “devil staircase,” whose multifractal

spectrum has an analytical solution, to evaluate the performance of the wavelet leaders approach by comparing with the result of the analytical method [14].

For the triadic Cantor set measure μ , define $f : [0, 1] \rightarrow [0, 1]$ as the distribution function of μ ,

$$f(x) = \int_0^x d\mu, \quad \mu \in [0, 1]. \quad (5)$$

As the measure of Cantor set is constructed recursively with the two distribution weights p_i ($i = 1, 2$) and it looks like a staircase, the function is called the “devil staircase” and the graph is self-similar

$$f(x) = \begin{cases} p_1 f(3x) & x \in \left[0, \frac{1}{3}\right] \\ p_1 & x \in \left[\frac{1}{3}, \frac{2}{3}\right] \\ p_1 + p_2 f(3x - 2) & x \in \left[\frac{2}{3}, 1\right]. \end{cases} \quad (6)$$

For this function, the scaling exponents can be calculated analytically by

$$\zeta(q) = \frac{\ln(p_1^q + p_2^q)}{\ln(1/3)}. \quad (7)$$

The multifractal spectrum can be obtained by the following formula:

$$D(h) = \frac{\left(\left(h_{\text{temp}} - 1\right) * \log\left(h_{\text{temp}} - 1\right) - h_{\text{temp}} * \log\left(h_{\text{temp}}\right)\right)}{\left(h_{\text{temp}} * \log(1/3)\right)}. \quad (8)$$

Figure 2(a) shows the scaling exponents of the “devil staircase” function obtained by the analytical method and wavelet leaders method, respectively. It can be seen that the difference is very small, especially for $q > 0$. The multifractal

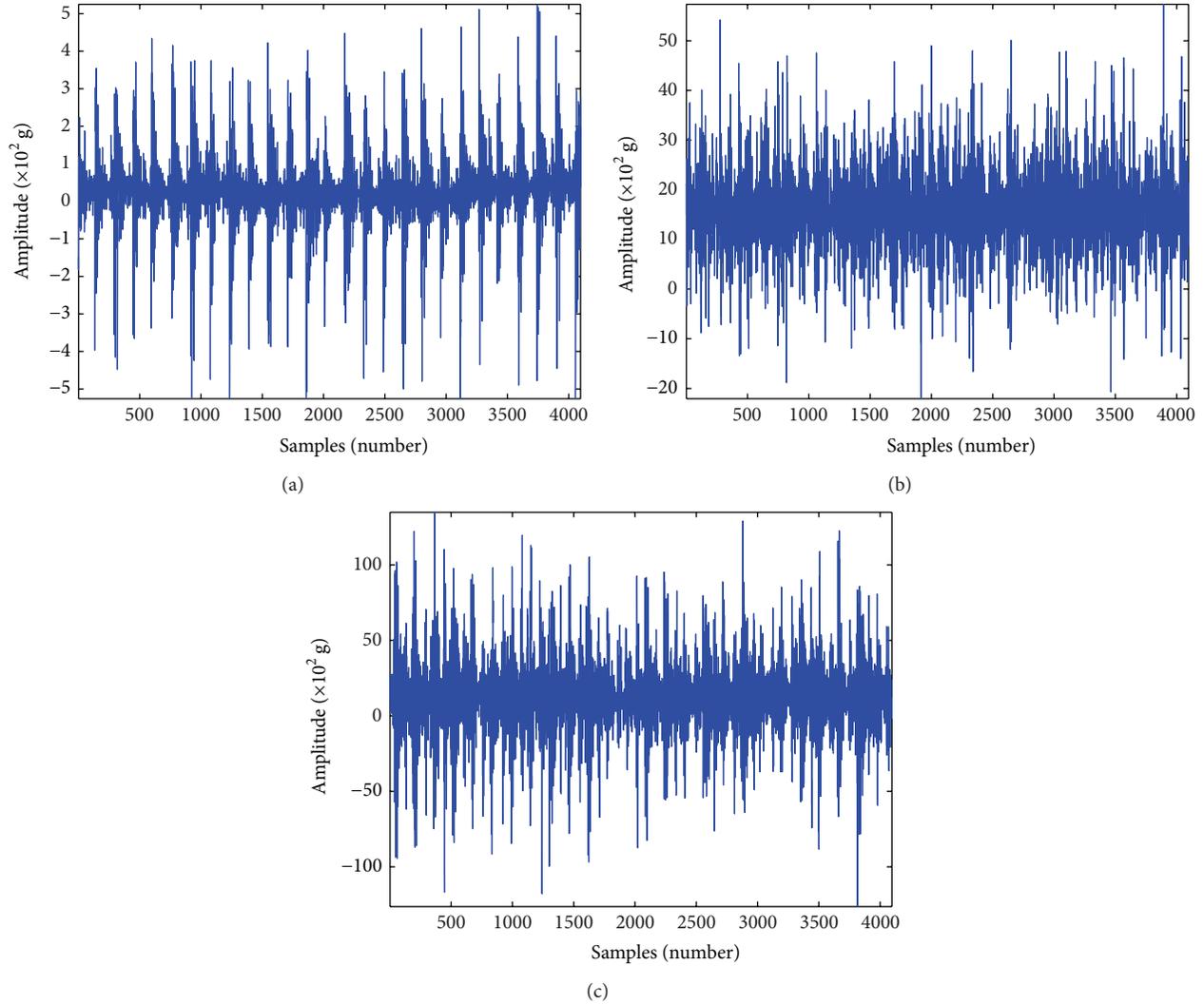


FIGURE 3: Waveform of vibration signals of plunger pump in three working conditions. (a) Hoisting. (b) Rotating. (c) Luffing.

spectra with the previous two approaches are shown in Figure 2(b). The spectrum obtained by the wavelet leaders with Legendre transform method almost coincides with the theoretical spectrum, which validates that the wavelet leaders based scaling exponents are effective. In addition, the multifractal spectrum obtained by the wavelet leaders with Chhabra regression method seems to deviate from the analytical result distinctly. The reason may be the regression with limited scales data and the scale weight distribution. However, the obtained spectrum still reflects the multifractal character of the “devil staircase” function. This is attributed to the two spectra having a similar shape. Furthermore, the three of four parameters used to characterize the spectrum [15], which are the minimum and the maximum of the Holder index corresponding to the two spectra, and the range of the Holder index, are similar; as for the fourth parameter, which is the h^* of the wavelet leaders with Chhabra regression method that corresponds to the Holder index of the maximum Hausdorff dimension, it is about 0.65 and agrees with the analytical method.

4.2. Plunger Pump Working Conditions Analysis with the Proposed Method. In order to realize the condition monitoring or fault diagnosis of plunger pump in truck crane, the working condition is required to judge firstly. In different working conditions, the states of the pump are different. So, the multifractal features of the vibration signal of the pump can be used to determine the working condition of the truck crane.

The plunger pump is seated on the truck crane, and the pump type is Linde HPR130, which has nine plungers. Three accelerometers mounted on the pump surface are used to obtain the vibration signal. The accelerometers are located in the horizontal direction of the bearing (1#), the vertical direction of the bearing (2#), and the horizontal direction of thrust plate (3#), respectively. The experiment data is collected from accelerometer 1#. The rotating speed of the pump is about 850 r/min, and the vibration signals are collected for hoisting, rotating, and luffing conditions at a sampling rate of 20 kHz. The waveforms of the original signals are shown in Figure 3, each consisting of 4096 points.

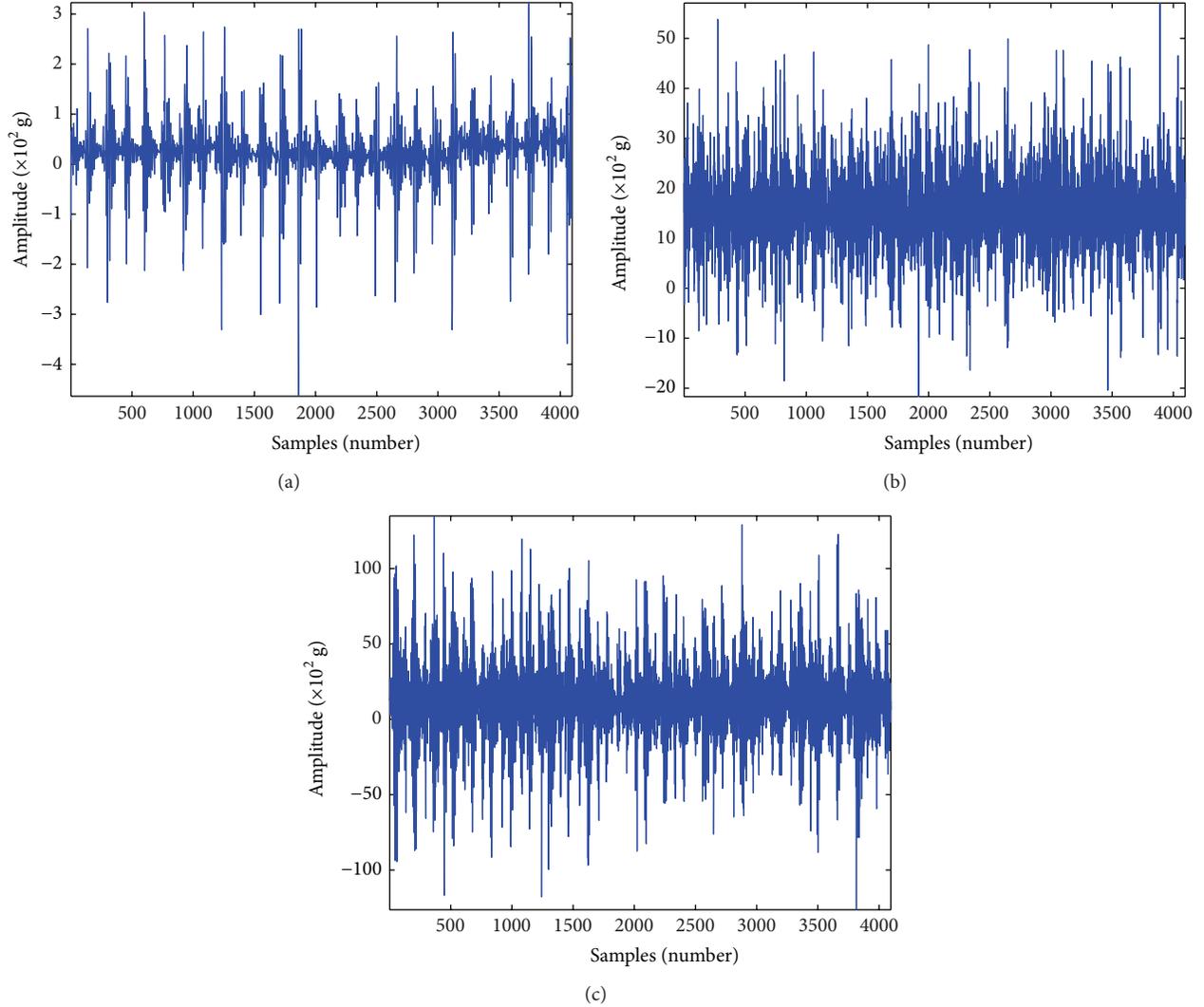


FIGURE 4: Waveform of denoised vibration signals of plunger pump in three working conditions. (a) Hoisting. (b) Rotating. (c) Luffing.

First, the original vibration signal is denoised with wavelet method. The Daubechies' wavelets with different vanishing wavelet moments (from "db1" to "db10") and "sym8" wavelet consist of a wavelet library. The wavelet basis and the decomposition level are selected based on minimizing the cost functional of entropy. For the previous three vibration signals of different working conditions, they are decomposed into 10 levels, and the wavelet basis and the corresponding decomposition level with the minimal Shannon entropy are listed in Table 1, and the denoised signals are shown in Figure 4.

Second, the wavelet leaders based multifractal features extraction method is employed to extract the vibration features. Considering the tradeoff between the relief of frequency aliasing and the preservation of spatial characterization, the db3 Daubechies' wavelet is selected as mother wavelet. For these three denoised vibration signals, c_2 are estimated on the data sets by bootstrap technology. As shown in Figure 5, q are calculated as 2, 6, and 5 for hoisting,

TABLE 1: The wavelet basis and the corresponding decomposition level with the minimal Shannon entropy.

Vibration signal	Wavelet basis	Decomposition level
Hoisting	db 10	2
Rotating	db 10	10
Luffing	db 1	1

rotating, and luffing conditions, respectively. For convenient comparison, the scopes of q are selected as $q_*^\pm = \pm 6$. In order to use the empirical expression of multifractal spectrum, the scaling range of regression should be determined according to the plot of $\zeta_L(q)$ versus $\log_2 2^j = j$. Taking the vibration signal of rotating condition as an example, Figure 6 shows that the scaling range is appropriate in the scope of 2 to 9. According to these parameters, decompose the vibration signals of the mentioned three conditions into nine levels and extract the wavelet leaders of wavelet coefficients in each level, respectively. The multifractal features, including the scaling

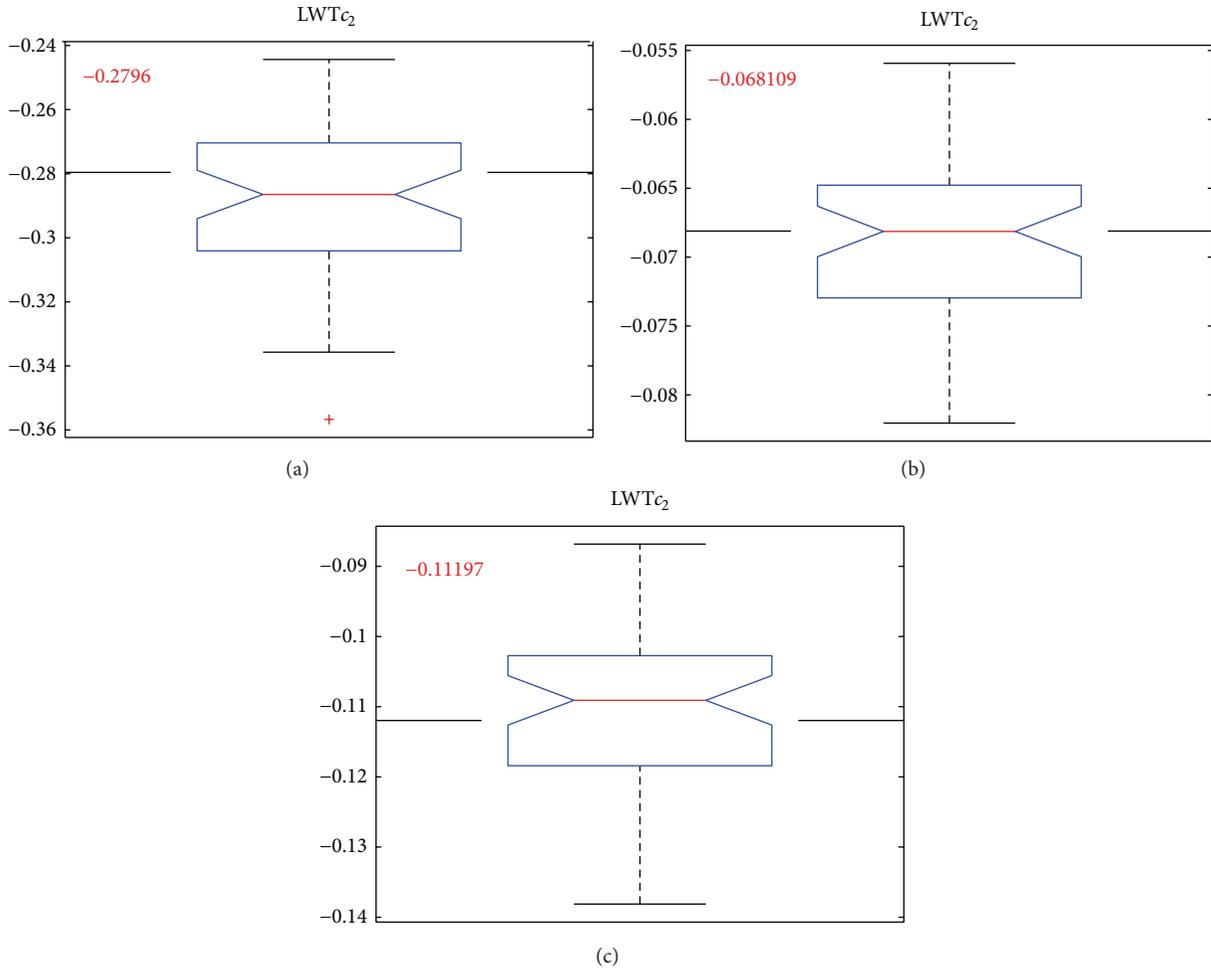


FIGURE 5: Estimation of c_2 of the vibration signals based on bootstrap technology, the number of resample is 49. (a) Hoisting. (b) Rotating. (c) Luffing.

exponents and the multifractal spectrum, are obtained with formulae (2) and (4).

The multifractal features for the three conditions are shown in Figures 7, 8, and 9. The relations of $\zeta(q)$ and q are nonlinear for the three conditions, which indicates that the vibration signals have the multifractal characters. Furthermore, all the multifractal spectra of the three conditions are unimodal functions, which also verifies that the signals have multifractal characters. But the precise singularities of these three signals are different as to the multifractal parameters [15], though all of the maximum of $D_{\max}(h)$ are close to 1. From the curve of multifractal spectrum, for the hoisting condition, the $\Delta h = h_{\max} - h_{\min}$ is about 1.65, which is larger than 0.55 of rotating condition and 0.73 of luffing. For the second parameter $\Delta D = D_{\max} - D_{\min}$, it is 1.1 for hoisting condition and 0.3 and 0.02 for luffing condition and rotating condition, respectively. As for the third parameter $h_0(D_{\max})$, it is 1.0 for hoisting condition and 0.35 and 0.3 for luffing condition and rotating condition. All of the parameters explain that the most singular is the vibration signal of hoisting condition, and the next is that of luffing condition. The vibration signal of rotating condition is the most regular.

TABLE 2: The block length of bootstrap for the vibration signals of three working conditions.

Level	2	3	4	5	6	7	8	9
Hoisting condition	96	33	20	34	20	9	3	1
Rotating condition	14	11	4	11	2	7	3	1
Luffing condition	15	9	9	5	1	1	3	1

In order to evaluate the statistic performance of the obtained multifractal features, the bootstrap technology is utilized to resample the wavelet leaders in each level. Then, the scaling exponents and the multifractal spectrum are calculated with the produced pseudosamples, and the statistic performances of the multifractal features are obtained. The optimal block length of each level is listed in Table 2. It shows that there exists various degrees of correlation among wavelet leaders in the same level. In the small level, the correlation is stronger, but in the 9th level, the leaders are almost independent of each other. For most of the levels, the wavelet leaders of hoisting show a stronger correlation.

Set the sampling times of bootstrap to 49 and calculate the multifractal features with the obtained bootstrap samples. The confidence levels of the multifractal features are shown

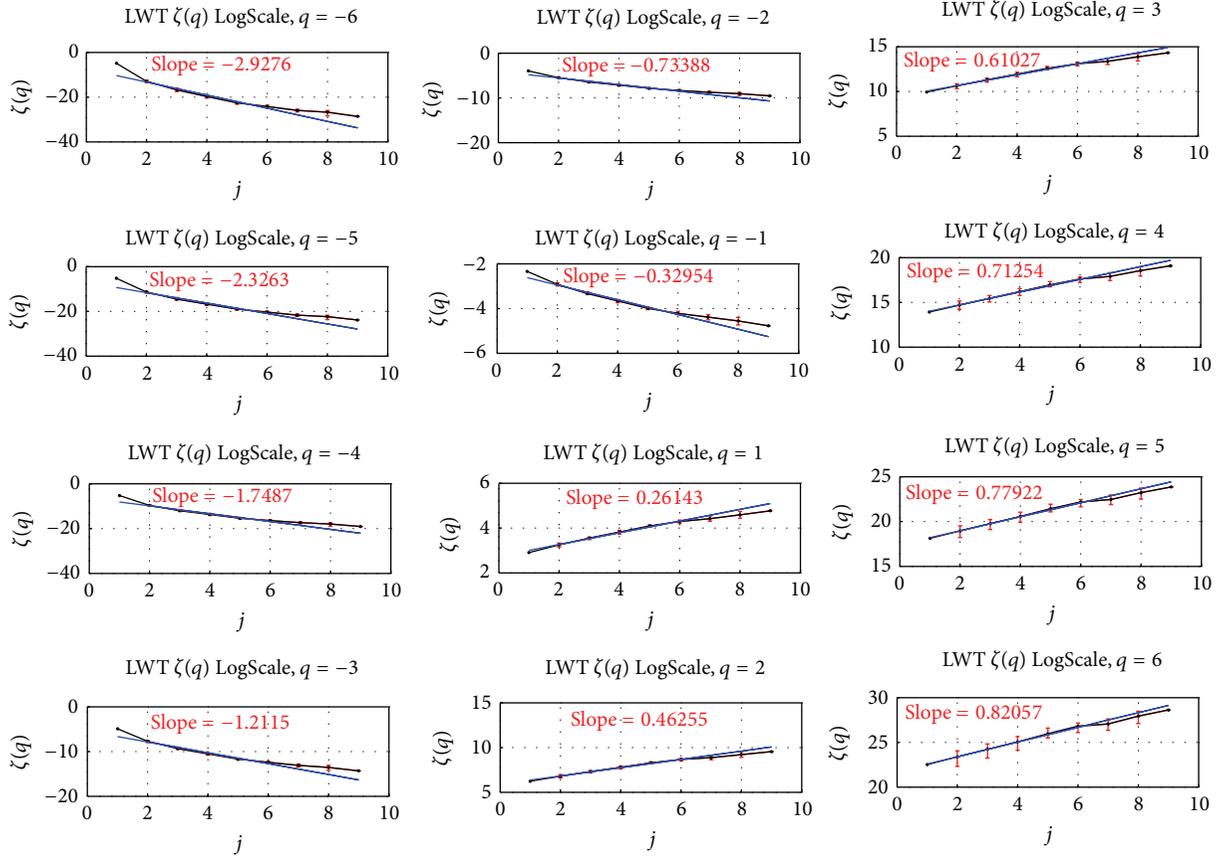


FIGURE 6: Relationship between scaling exponents $\zeta_L(q)$ and scale j for rotating condition.

TABLE 3: The confidence level of scaling exponents of three working conditions.

Order q	Confidence level		
	Hoisting $\zeta_H(q)$	Rotating $\zeta_R(q)$	Luffing $\zeta_L(q)$
-6	[-11.2, -8.10]	[-3.14, -2.66]	[-3.92, -3.33]
-5	[-9.08, -6.58]	[-2.52, -2.09]	[-3.11, -2.64]
-4	[-6.97, -5.09]	[-1.90, -1.56]	[-2.33, -1.98]
-3	[-4.86, -3.67]	[-1.32, -1.08]	[-1.60, -1.36]
-2	[-2.83, -2.32]	[-0.81, -0.66]	[-0.94, -0.81]
-1	[-1.17, -1.04]	[-0.37, -0.30]	[-0.40, -0.34]
1	[0.80, 0.87]	[0.24, 0.30]	[0.23, 0.29]
2	[1.37, 1.56]	[0.41, 0.54]	[0.37, 0.50]
3	[1.83, 2.15]	[0.54, 0.74]	[0.46, 0.64]
4	[2.25, 2.69]	[0.60, 0.91]	[0.50, 0.73]
5	[2.66, 3.19]	[0.62, 1.04]	[0.53, 0.81]
6	[3.06, 3.68]	[0.61, 1.14]	[0.54, 0.88]

in Figure 7 to Figure 9. Table 3 lists the scaling exponents of the previous three conditions, and $\zeta_H(q)$, $\zeta_R(q)$, and $\zeta_L(q)$ correspond to the scaling exponent of hoisting, rotating, and luffing conditions, respectively. From Tables 4, 5, and 6, it further reflects the confidence level of multifractal features

of the vibration signal. In the tables, $D_H(q)$ and $h_H(q)$ are multifractal spectrum value and Holder index of hoisting condition, respectively, and $D_R(q)$ and $h_R(q)$ are multifractal spectrum value and Holder index of rotating condition, respectively, while $D_L(q)$ and $h_L(q)$ correspond to multifractal spectrum value and Holder index of luffing condition, respectively. From Tables 4 to 6, when q is in the scope of $[-2, +2]$, the multifractal features are more concentrative, and the corresponding confidence levels are narrow; out of this scope, the multifractal features are dispersive, and the confidence levels are wide.

The proposed wavelet leaders based multifractal features extraction method is compared with the direct wavelet leaders based multifractal features extraction method. The latter does not include the denoised step. The obtained multifractal features with the direct method are shown in Figures 10, 11, and 12. For the working conditions of rotating and luffing, the shapes of multifractal spectrum change little, and the parameters representing the singularity are $\Delta h = 0.55$, $\Delta D = 0.01$, and $h_0(D_{\max}) = 0.3$ for rotating condition, respectively, and the corresponding parameters are 0.72, 0.3, and 0.35 for luffing condition, respectively. But for the working condition of hoisting condition, the shape of multifractal spectrum changes considerably, and the parameters representing the singularity change to $\Delta h = 0.76$, $\Delta D = 0.42$, and $h_0(D_{\max}) = 0.45$, respectively. This indicates

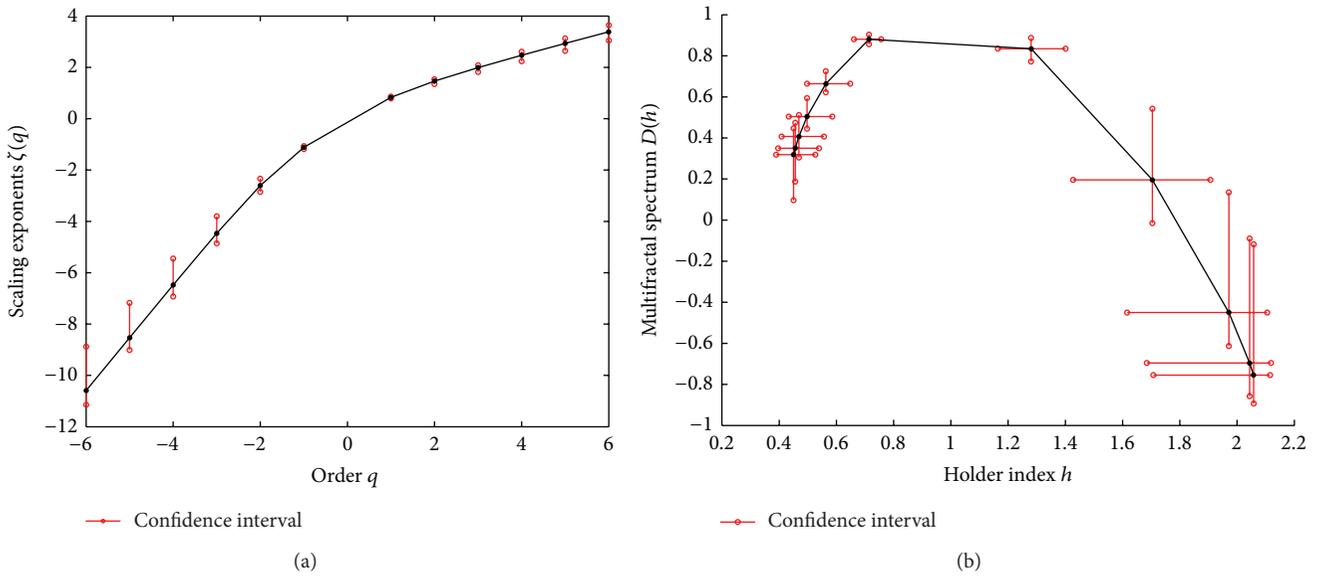


FIGURE 7: Multifractal features and the statistic performance of the vibration signal in hoisting condition.

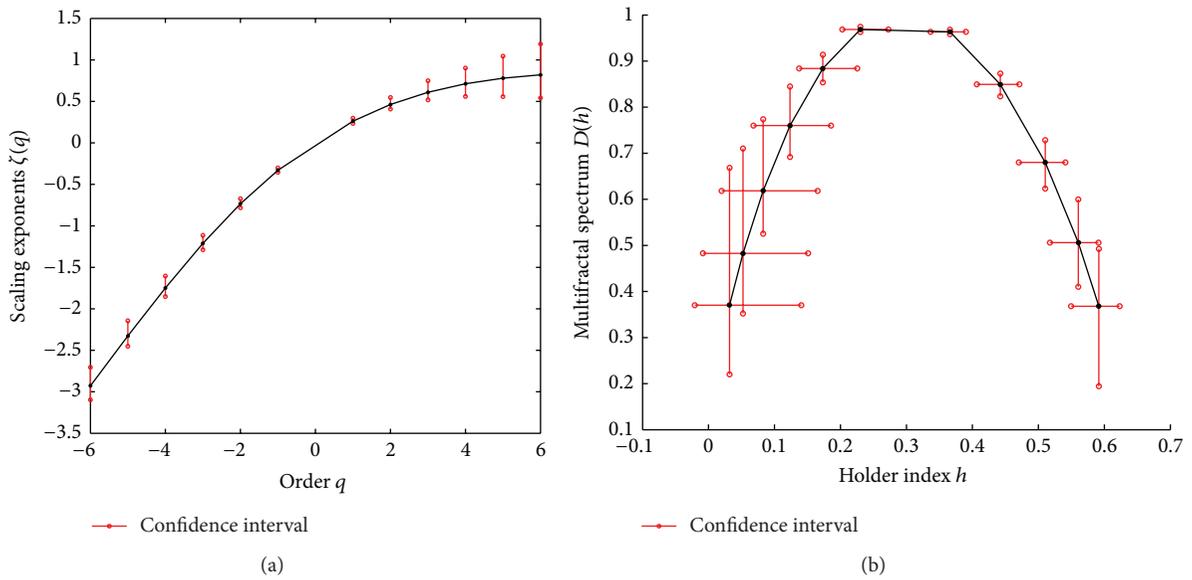


FIGURE 8: Multifractal features and the statistic performance of the vibration signal in rotating condition.

TABLE 4: The confidence level of multifractal spectrum of hoisting working condition.

Confidence level		Confidence level		Confidence level	
Holder index $h_H(q)$	Spectrum value $D_H(q)$	Holder index $h_H(q)$	Spectrum value $D_H(q)$	Holder index $h_H(q)$	Spectrum value $D_H(q)$
[0.40, 0.49]	[0.08, 0.47]	[0.50, 0.63]	[0.62, 0.72]	[1.39, 2.10]	[-0.59, 0.51]
[0.40, 0.51]	[0.17, 0.50]	[0.66, 0.77]	[0.86, 0.91]	[1.46, 2.11]	[-0.85, 0.25]
[0.41, 0.53]	[0.28, 0.52]	[1.16, 1.39]	[0.77, 0.91]	[1.51, 2.11]	[-0.89, 0.05]
[0.44, 0.56]	[0.44, 0.58]	[1.29, 1.92]	[-0.02, 0.73]	[1.54, 2.11]	[-0.88, -0.03]

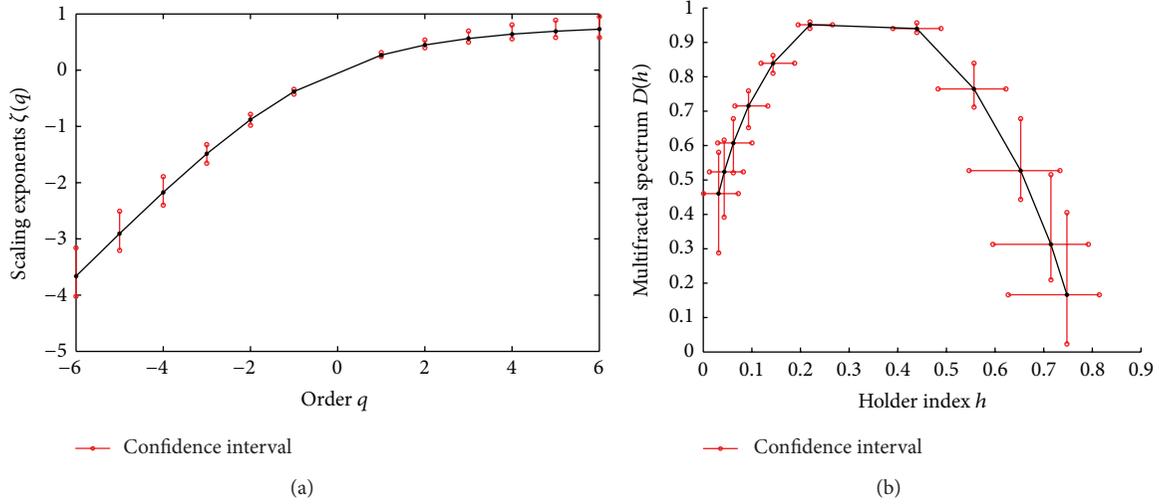


FIGURE 9: Multifractal features and the statistic performance of the vibration signal in luffing condition.

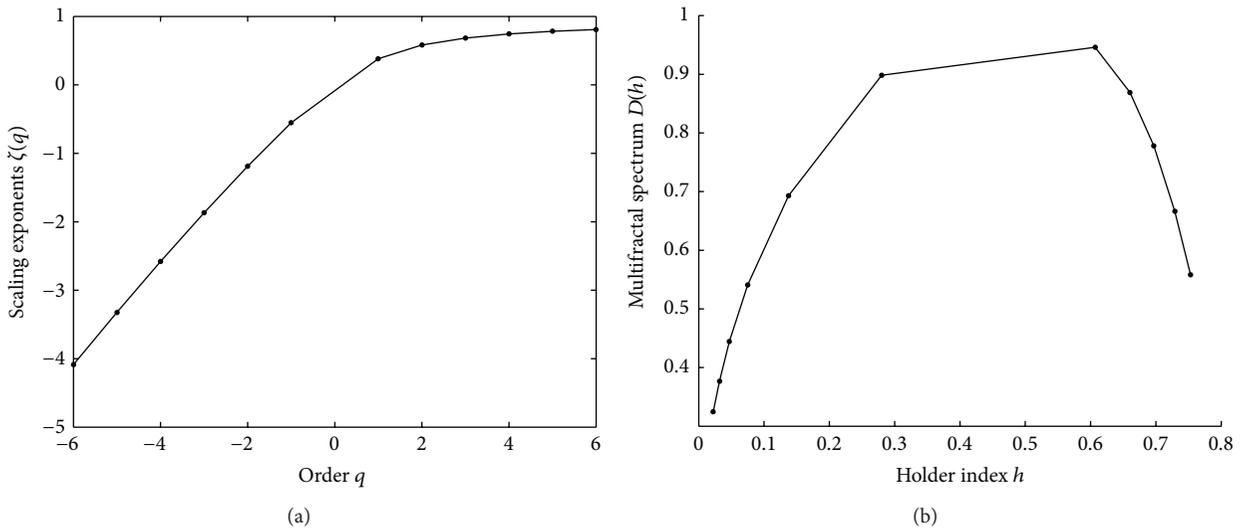


FIGURE 10: Multifractal features and the statistic performance of the vibration signal without denoising process in hoisting condition.

TABLE 5: The confidence level of multifractal spectrum of rotating working condition.

Confidence level		Confidence level		Confidence level	
Holder index	Spectrum value	Holder index	Spectrum value	Holder index	Spectrum value
$h_R(q)$	$D_R(q)$	$h_R(q)$	$D_R(q)$	$h_R(q)$	$D_R(q)$
[-0.02, 0.09]	[0.22, 0.53]	[0.14, 0.22]	[0.85, 0.91]	[0.45, 0.55]	[0.63, 0.76]
[0.00, 0.12]	[0.35, 0.62]	[0.21, 0.27]	[0.96, 0.98]	[0.50, 0.61]	[0.40, 0.62]
[0.03, 0.15]	[0.52, 0.72]	[0.33, 0.40]	[0.96, 0.97]	[0.54, 0.63]	[0.25, 0.52]
[0.08, 0.19]	[0.69, 0.82]	[0.39, 0.49]	[0.82, 0.88]	[0.56, 0.65]	[0.10, 0.47]

TABLE 6: The confidence level of multifractal spectrum of luffing working condition.

Confidence level		Confidence level		Confidence level	
Holder index	Spectrum value	Holder index	Spectrum value	Holder index	Spectrum value
$h_L(q)$	$D_L(q)$	$h_L(q)$	$D_L(q)$	$h_L(q)$	$D_L(q)$
[0.01, 0.06]	[0.38, 0.58]	[0.11, 0.17]	[0.82, 0.86]	[0.58, 0.71]	[0.42, 0.64]
[0.02, 0.07]	[0.45, 0.61]	[0.18, 0.25]	[0.94, 0.96]	[0.63, 0.78]	[0.16, 0.46]
[0.03, 0.09]	[0.56, 0.68]	[0.40, 0.47]	[0.93, 0.95]	[0.67, 0.81]	[0.00, 0.36]
[0.05, 0.12]	[0.68, 0.76]	[0.50, 0.61]	[0.72, 0.82]	[0.68, 0.82]	[-0.12, 0.32]

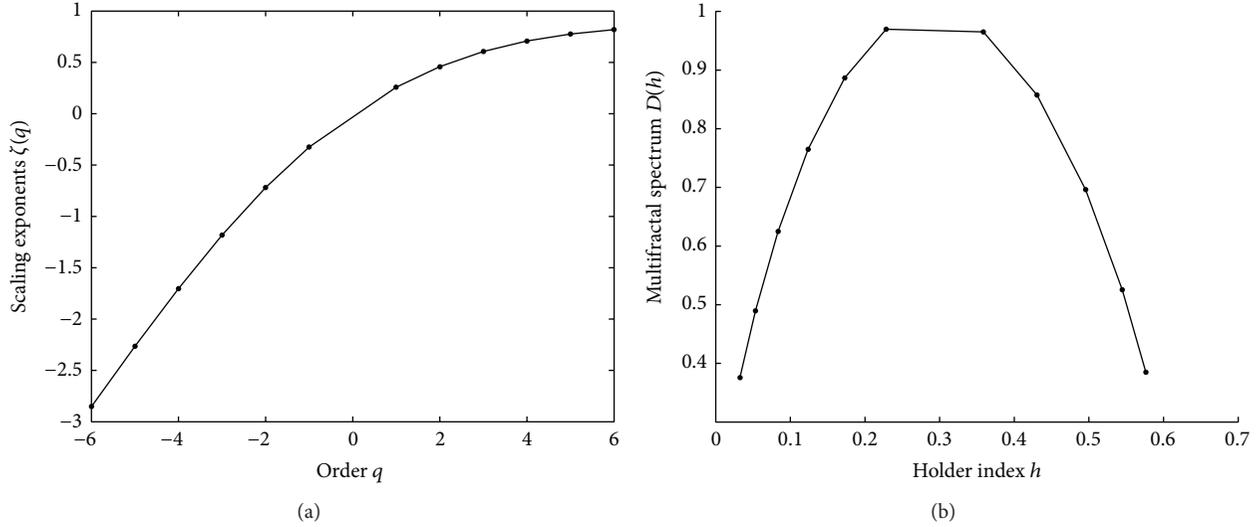


FIGURE 11: Multifractal features and the statistic performance of the vibration signal without denoising process in rotating condition.

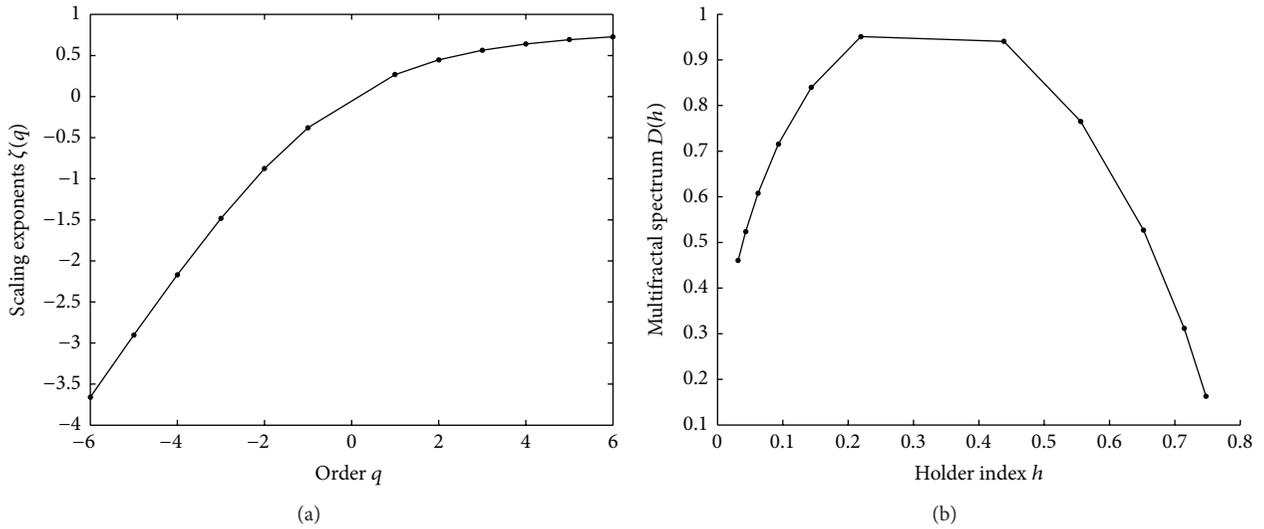


FIGURE 12: Multifractal features and the statistic performance of the vibration signal without denoising process in luffing condition.

that the singularity of the vibration signal is enhanced by the denoised process, and the multifractal characters of the vibration itself are highlighted.

5. Conclusions

In this study, an improved wavelet leaders based multifractal features extraction scheme is proposed for the plunger pump in truck crane. The wavelet transform is used first to denoise the original signal, and then the wavelet leaders based multifractal features are calculated on those obtained denoised signals; finally, to evaluate the performance of the obtained statistic features, the block bootstrap technology is utilized. The case of “devil staircase” verifies that the wavelet leaders based analysis method can extract the multifractal features effectively. The vibration signals of the plunger pump in the truck crane are used to validate the capability of the improved wavelet leaders based multifractal features

for the contaminative signals, and the performance of the proposed method is also compared with that of the direct wavelet leaders based scalogram through this case study. The results show that (1) the proposed method is not only qualified in multifractal features extraction, but also evaluates the performance of the obtained statistic. (2) The proposed method is utilized to analyze the vibration signals of the plunger pump in the truck crane in three working conditions, and the results indicate that the multifractal characters of the vibration are highlighted. (3) The multifractal method with denoised process has better performance, which means that the proposed method is more suitable for contaminative signal analysis in condition monitoring.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant nos. 51205371 and 61175038),

National Basic Research Program of China (“973” Program) (Grant no. 2013CB035400), Innovation Program of Shanghai Committee of Science and Technology (Grant nos. 11dz1121500 and 11JC1405800), and the Research Project of State Key Laboratory of Mechanical System and Vibration (Grant no. MSV-2012-06).

multifractal spectrum of natural series,” *International Journal of Bifurcation and Chaos*, vol. 20, no. 2, pp. 331–339, 2010.

References

- [1] Q. Wang, H. Chen, P. Chua, and L. G. Hian, “Fault detection of water hydraulic motor by demodulated vibration signal analysis with the hilbert transform and genetic algorithm,” *Journal of Testing and Evaluation*, vol. 39, no. 2, pp. 12–15, 2011.
- [2] K. Mollazade, H. Ahmadi, M. Omid, and R. Alimardani, “An intelligent model based on data mining and fuzzy logic for fault diagnosis of external gear hydraulic pumps,” *Insight: Non-Destructive Testing and Condition Monitoring*, vol. 51, no. 11, pp. 594–600, 2009.
- [3] R. Lopes and N. Betrouni, “Fractal and multifractal analysis: a review,” *Medical Image Analysis*, vol. 13, no. 4, pp. 634–649, 2009.
- [4] Z. K. Peng, F. L. Chu, and P. W. Tse, “Singularity analysis of the vibration signals by means of wavelet modulus maximal method,” *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 780–794, 2007.
- [5] B. Lashermes, S. Jaffard, and P. Abry, “Wavelet leader based multifractal analysis,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’05)*, pp. 161–164, March 2005.
- [6] H. Wendt, P. Abry, and S. Jaffard, “Bootstrap for empirical multifractal analysis,” *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 38–48, 2007.
- [7] H. Wendt and P. Abry, “Multifractality tests using bootstrapped wavelet leaders,” *IEEE Transactions on Signal Processing*, vol. 55, no. 10, pp. 4811–4820, 2007.
- [8] Y. Li, W. Du, P. Ye, and C. Liu, “Wavelet leaders multifractal features extraction and performance analysis for vibration signals,” *Journal of Mechanical Engineering*, vol. 49, no. 6, pp. 60–65, 2013.
- [9] E. Serrano and A. Figliola, “Wavelet Leaders: a new method to estimate the multifractal singularity spectra,” *Physica A*, vol. 388, no. 14, pp. 2793–2805, 2009.
- [10] A. B. Chhabra, C. Meneveau, R. V. Jensen, and K. R. Sreenivasan, “Direct determination of the $f(\alpha)$ singularity spectrum and its application to fully developed turbulence,” *Physical Review A*, vol. 40, no. 9, pp. 5284–5294, 1989.
- [11] M. Kupczak, “Ideal choice of an orthogonal basis from a library of bases for denoising,” in *Proceedings of the 28th ASR 2003 Seminar, Instruments and Control*, pp. 181–188, VSB Technical University of Ostrava, 2003.
- [12] P. Abry, H. Wendt, S. Jaffard et al., “Methodology for multifractal analysis of heart rate variability: from LF/HF ratio to wavelet Leaders,” in *Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC ’10)*, pp. 106–109, arg, September 2010.
- [13] D. N. Politis and H. White, “Automatic block-length selection for the dependent bootstrap,” *Econometric Reviews*, vol. 23, no. 1, pp. 53–70, 2004.
- [14] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, Calif, USA, 1999.
- [15] A. Figliola, E. Serrano, G. Paccosi, and M. Rosenblatt, “About the effectiveness of different methods for the estimation of the

Research Article

Performance Assessment for a Fleet of Machines Using a Combined Method of Ant-Based Clustering and CMAC

Lei Zhang,¹ Qixin Cao,¹ and Jay Lee²

¹ Research Institute of Robotics, Shanghai Jiao Tong University, Shanghai 200240, China

² NSF Center for Intelligent Maintenance Systems, University of Cincinnati, Cincinnati, OH 45221, USA

Correspondence should be addressed to Lei Zhang; zhanglei@sjtu.edu.cn

Received 2 April 2013; Revised 15 July 2013; Accepted 17 July 2013

Academic Editor: Liang Gong

Copyright © 2013 Lei Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a combined method of ant-based clustering and cerebellar model articulation controller for performance assessment for a fleet of machines. A novel ant-based clustering algorithm with kernel method is used to cluster machines in a fleet. The algorithm has two features. First, a projection based on kernel principal component analysis replaces random projection to improve the efficiency. Second, the clustering is performed on the feature space after kernel mapping to improve the clustering accuracy. The algorithm can cluster machines in a self-organizing way to achieve the horizontal assessment. The vertical assessment for the single machine based on CMAC is presented. Then, how to combine the vertical and horizontal assessment results is discussed. The outlier mining method to detect abnormal machines based on the clustering results is also proposed. Cluster-based global outlying factor is suggested to measure the outlying degree of abnormal machine. Finally, the case study on axial fans shows that the combined method can give a more comprehensive assessment for fans' performance monitoring.

1. Introduction

Intelligent Maintenance (IM) of machine is emerging as a replacement of traditional reactive maintenance style. Different from traditional "fail and fix (FAF)" maintenance practice, IM focuses on "predict and prevent (PAP)" methodology to achieve near-zero downtime performance. Usually, the machine and components go through a series of degradation states before failure. The predictive tools are needed to monitor degradation states rather than detect the faults [1]. Once the degradation has been detected, the remaining useful life of machine could be predicted. Therefore, the key challenges of implementing IM are machine degradation assessment and prediction.

Many efforts have been made to develop methods and tools for these challenges. Yan and Lee [2] proposed a logic regression method to assess performance degradation of an elevator door system. They [3] also presented the further study combining logic regression and fuzzy logic for nozzle life prediction in gas turbine. Logistic regression can easily represent the daily maintenance records as a dichotomous problem. However, this method is applicable only when

both normal and failure behavior are available. Yu et al. [4] presented a machine performance assessment approach based on Gaussian Mixture Model (GMM). Experimental results of real industrial run-to-failure bearing had shown that the model was efficient. Gaussian Mixture Model integrated with locality preserving projection was also suggested to assess the bearing performance degradation [5]. Mixture of Gaussian can be utilized to approximate an arbitrary distribution, while the number of the mixtures is not easy to be determined and the optimization parameters can be affected by different initialization methods. Lee [6] first proposed a pattern discrimination model based on the Cerebellar Model Articulation Controller (CMAC). Experiments on the stepping motor and the robot had proven the feasibility of the model. Lin and Wang [7] also used enhanced CMAC in performance analysis of rotating machinery. Zhang et al. [8] suggested a modified CMAC algorithm for performance degradation assessment of self-maintenance machine. Xu et al. [9] proposed a fuzzy based extension of CMAC to analyze two types of machine degradation severities. In one case, the network was trained by signals from different levels of machine degradation states. In the other case, the

TABLE 1: The features of vertical and horizontal assessment and prediction.

	Vertical assessment and prediction	Horizontal assessment and prediction
Machine type	Single machine	A fleet of machines
Compared data	Compared with historical data of itself	Compared with the data of the same (similar) machines
Analyzing domain	Time domain	Space domain (peer to peer)
Methods	CMAC, SOM, GMM, etc.	Cluster-based method

network was only trained by signals of normal state. The generalization ability of CMAC makes it very suitable for performance degradation assessment. It can be trained only using the data in the normal state. Wu et al. [10] proposed an online adaptive condition-based maintenance method with pattern discovery and fault learning capabilities. The method was mainly based on Self-Organizing Map (SOM). An experiment on the machine tool test bed validated the proposed approach. SOM provides a way of representing multidimensional feature space in a one- or two-dimensional space while preserving the topological properties of the input space. It is an unsupervised learning and no prior outputs are needed. Qiu et al. [11] proposed an adaptive wavelet filter method for hazard rate prediction of bearing. Further, they developed a robust performance degradation assessment method for rolling bearing [12]. This method was based on a combination of wavelet filter and SOM for fault identification and assessing performance degradation, respectively. Pan et al. [13] suggested a Fuzzy C-means (FCM) method for bearing performance degradation assessment. Lifting wavelet packet decomposition was used to create feature vectors.

Due to the complexity of the machine, the combination of multiple methods is usually used for identifying working condition. Lapira et al. [14] proposed a systematic framework that utilized multiregime modeling approach to trend and assess wind turbine degradation. Three methods composed by SOM, GMM, and neural network were combined and chosen to handle condition data with multiple operating regimes. Caesarendra et al. [15] proposed the combination between Relevance Vector Machine (RVM) and Logistic Regression (LR) for bearing performance assessment. Liao and Lee [16] presented a novel machine performance degradation scheme based on Fixed Cycle Features Test (FCFT). FCFT introduced a new testing method which obtained data during the transient periods of different working loads. Wavelet packet analysis, PCA, and mixture model were combined. A case study for chiller system was used as an example. A comparative study of maintenance data classification based on neural networks, logistic regression, and support vector machines was also provided [17].

Remaining Useful Life (RUL) prediction methods are important to extrapolate the machine's process behavior over time and predict its life in the future. Autoregressive Moving Average (ARMA) model is usually used for modeling and predicting future values in time series. Yu et al. [18] presented Elman Recurrent Neural Network (ERNN) for predicting the behavior of a boring process during its full life-cycle. This prediction was achieved by the fusion of the predictions of three principal components extracted from the spindle load. Liu et al. [19] proposed a match matrix method for

manufacturing process performance prediction and diagnosis. By constructing a match matrix, the similarity between two feature series can be represented by the best match index. Siegel et al. [20] suggested a robust regression curve fitting approach for Remaining Useful Life prediction for the helicopter oil-cooler bearings. Tran et al. [21] presented a three-stage method for both performance degradation assessment and RUL prediction. In the first stage, ARMA model was generated by using only normal operating data to identify the behavior of the complex system. In the second stage, the Cox' proportional hazard model was established to estimate the system survival function. In the last stage, SVM for regression [22], with multistep ahead prediction ability, was utilized to forecast the RUL. Huang et al. [23] suggested a new scheme for the prediction of a ball bearing's RUL based on self-organizing map (SOM) and back propagation neural network. Gebraeel et al. [24] also suggested a neural network approach for RUL predictions of bearing from vibration-based degradation signals. Two classes of neural network models, single-bearing model, and clustered-bearing model were developed.

A toolbox named as Watchdog Agent [25] has been developed, which integrated many algorithms mentioned above. The toolbox has been embedded in the Labview software for public usage.

Reviewing all the assessment and predicting methods above, we can see that most methods use time-domain comparison, which means machine's current condition is compared with its historical data to assess various degradation states. All these methods are used for single machine or system. This kind of methods can be concluded as vertical assessment and prediction. But in IM, there is also a need for peer to peer comparison for a fleet of machines. So-called "a fleet of machines" means that many machines, with the same or similar structures and working condition, compose a fleet. For example, many pumps in a factory or several elevators in a building can be taken as a fleet of machines. The running information of a fleet of machines can be exchanged or collected easily for analysis. Machines in the fleet can be compared to each other to assess their performance degradation. We call this comparison as horizontal assessment and prediction. Table 1 shows the features of vertical and horizontal assessment methods. The horizontal comparison of a fleet of machines presents the chance and innovation brought by Internet to the maintenance style. But now the relevant research and report on this field are few.

For this purpose, this paper proposed a horizontal method based on ant-based clustering. This kind of ant-based clustering algorithm imitates the ant's behavior of clustering their corpses [26]. The object with n attributes can

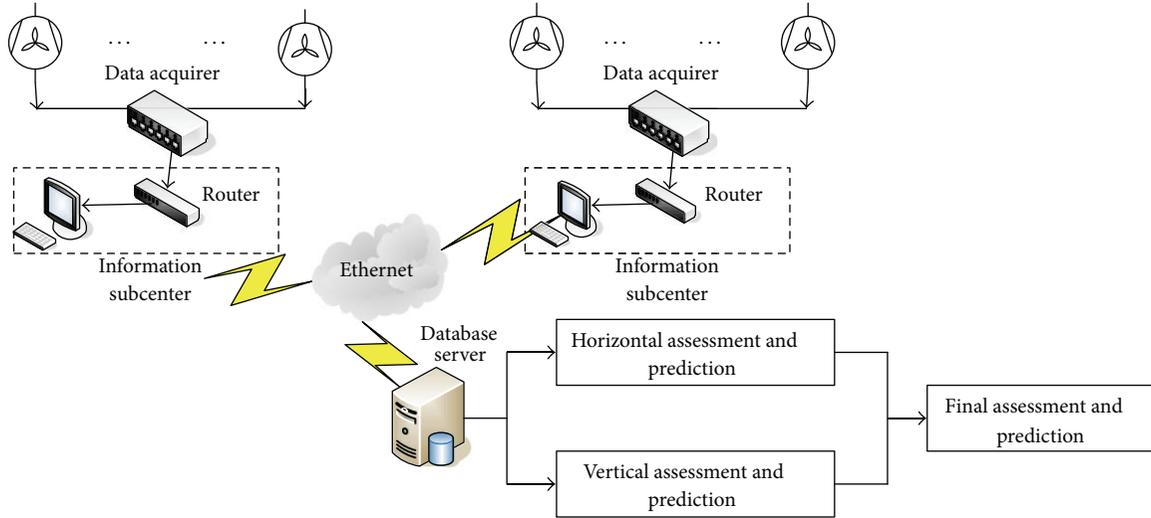


FIGURE 1: The system framework of proposed methods.

be looked as a point in n -dimensional space. The objects are projected into a low dimensional space (often a two-dimensional plane). Then the ants compute the similarity of the objects and decide to pick up or drop them. Compared with traditional clustering methods, such as partition-based methods, hierarchy-based methods, and density-based methods, ant-based clustering algorithm does not need any prior knowledge. This is an important merit which makes it suitable in our application. Because we usually do not know how many clusters there are in a fleet of machines, it should be created by a self-organizing way. Second, ant-based clustering process is visible on the projecting plane. This is very helpful in the monitoring of machine performance. The visible clustering results can be directly integrated into the interface of the monitoring software.

In our application, on one hand, the efficiency of the algorithm should be improved because some maintenance decisions or actions are dependent on the results of clustering analysis. On the other hand, the algorithm should have the ability to cluster the datasets with different structures. We have proposed a novel ant-based clustering algorithm using the kernel method to overcome these problems [27]. In this paper, we will show how this algorithm is applied in performance assessment for a fleet of machines.

The remainder of this paper is organized as follows. Section 2 described the system framework of proposed methods. Section 3 proposed the methods for performance assessment for a fleet of machines, including horizontal assessment, vertical assessment, and the fusion method. Section 4 gave the case study and analysis. Finally, Section 5 gave the conclusion and further research directions.

2. The System Framework of Proposed Methods

In this paper, we will propose a combined method of ant-based clustering and CMAC for performance assessment for

a fleet of machines. As shown in Figure 1, data acquirers can get the data of fans in monitoring nodes. The data can be collected and transferred by router to the database server by the Ethernet. The following analysis can be performed.

- (i) Machines in a fleet are clustered using ant-based clustering method.
- (ii) Performance assessment for the single machine is conducted based on CMAC.
- (iii) The horizontal and vertical assessments are combined to get the final conclusion.
- (iv) The outliers are detected based on outlying factor in a very small cluster.

3. The Assessment Method for a Fleet of Machines

3.1. Ant-Based Clustering Algorithm with the Kernel Method for Horizontal Assessment

3.1.1. The Basic Ant-Based Clustering Algorithm. The algorithm introduced by Lumer and Faieta [28] represents the basic ant-based clustering method. Some important concepts are firstly introduced through Figure 2.

The projection plane: the objects and ants are initially projected onto a two-dimensional plane. Each object or ant is projected randomly. The size of the plane can be determined based on the number of objects.

The local neighbourhood of object o_i : it is a neighbouring region of the object o_i and written as $\text{Neigh}(o_i)$. It is often a square with size $s \times s$ ($s = 2r + 1$), where r is the radius of $\text{Neigh}(o_i)$. The center of $\text{Neigh}(o_i)$ is the position of o_i .

The local similarity: it is the similarity of the object o_i with other objects in $\text{Neigh}(o_i)$. It is often measured by the distance

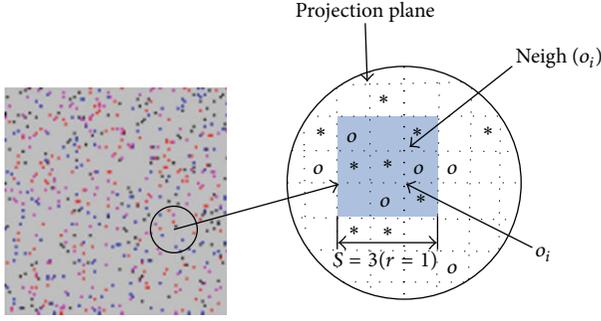


FIGURE 2: The local neighborhood of the object o_i .

between objects. In Figure 2, the object o_i is assumed to locate at the coordinate (x_i, y_i) . The local similarity of o_i is given by

$$f(o_i) = \begin{cases} \frac{1}{s^2} \sum_{o_j \in \text{Neigh}(o_i)} \left[1 - \frac{d(o_i, o_j)}{\alpha} \right], & \text{when } f > 0 \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where $d(o_i, o_j)$ is the distance between two objects. Typically Euclidean distance is used. α is a factor that defines the scale for dissimilarity.

The probability conversion function: it is a function that converts the local similarity of o_i into the probability of being picked up (or dropped) by ants. The probability that an ant will pick up or drop the object is

$$P_p(o_i) = \left(\frac{k_1}{k_1 + f(o_i)} \right)^2, \quad (2)$$

$$P_d(o_i) = \begin{cases} 2f(o_i) & \text{when } f(o_i) < k_2, \\ 1 & \text{when } f(o_i) \geq k_2, \end{cases} \quad (3)$$

where k_1, k_2 are two constants. k_1 and k_2 adjust the probabilities of picking up and dropping objects. $P_p(o_i)$ and $P_d(o_i)$ are compared with a random real number p ($p \in [0, 1]$), and the results determine whether the object o_i should be picked up or dropped.

The process of LF algorithm can be generalized as the following steps.

- (1) Projection: all objects and ants are randomly projected onto the grid.
- (2) Calculating the similarity: each ant calculates the object's similarity to others in the object's local neighbourhood.
- (3) The similarity is transformed to the probability to pick up or drop object.
- (4) Ants pick up or drop objects.
- (5) Ants move.
- (6) Repeat (2)–(5).

A number of modifications have been introduced to the basic LF algorithm to improve clustering quality and convergent speed. A latest review of ant-based clustering can be referred to [29].

3.1.2. Ant-Based Clustering Algorithm with the Kernel Method.

To improve the algorithm's efficiency and clustering quality, we incorporated the kernel method into ant-based clustering and created the novel ant-based clustering with the kernel method (ACK) [27]. The applications of kernels in ACK are shown in two ways. First, Kernel Principal Component Analysis (KPCA) is used to modify the random projection of all objects. Second, the Euclidean distance in the feature space is applied as a measure of the similarity between the objects.

Projection of the Objects Based on KPCA. In ant-based clustering algorithms, the objects are randomly projected onto the plane. One pattern corresponds randomly with a pair of coordinates. This random projection leads to few similarities between the objects in the local neighbourhood at the early stage. It takes a long time for an object to be similar to nearby objects from the inception of the algorithm. We have suggested a modified projection based on Principal Component Analysis (PCA) [30]. Then we applied KPCA to replace PCA. Compared with linear PCA, KPCA can extract features that are more useful for classification.

The Ant Movement Model Operating in the Feature Space.

In ACK algorithm, we applied the ant movement model suggested by Xu et al. [31], where each object is taken as an ant. Each ant has two states: movement and sleep. If the ant finds that a location is suitable for it to rest, it will stop moving and enter a sleeping state; otherwise, it will continue to move to another place. The fitness of the local neighbourhood is computed as (4), which is similar to (1) in Section 3.1.1. Consider

$$f(o_i) = \begin{cases} \frac{1}{s^2} \sum_{o_j \in L(o_i, r)} \left[1 - \frac{d(o_i, o_j)}{\alpha_i} \right], & \text{when } f > 0 \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

where $\alpha_i = (1/(N-1)) \sum_{j=1}^N d(o_i, o_j)$ is the average distance between o_i and other objects.

If we apply a kernel function to map the objects into the feature space, then the clustering can be performed according to the similarities of the objects in the feature space. The fitness becomes

$$f(o_i) = \begin{cases} \frac{1}{s^2} \sum_{o_j \in L(o_i, r)} \left[1 - \frac{d_F(o_i, o_j)}{\alpha_i} \right], & \text{when } f > 0 \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

where $\alpha_i = (1/(N-1)) \sum_{j=1}^N d_F(o_i, o_j)$ and $d_F(o_i, o_j)$ is the distance between o_i and o_j in the feature space. $d_F(o_i, o_j)$ can be obtained by the kernel function according to

$$d_F(o_i, o_j) = \sqrt{K(o_i, o_i) - 2K(o_i, o_j) + K(o_j, o_j)}. \quad (6)$$

A Gaussian kernel is applied in this study as follows:

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right). \quad (7)$$

```

0 /*initialization*/
1 All objects are placed on the grid based on KPCA
2 Initialize all parameters:  $r_1, t_{\max}, \alpha, \beta$ 
3 /*main loop*/
4 for  $t = 1$  to  $t_{\max}$  do
5   for  $i = 1$  to  $N$  do ( $N$  is the number of the objects)
6     compute  $f(o_i)$  and  $P_a(o_i)$ 
7     draw a random real number  $p \in (0, 1)$ 
8     if ( $p \leq P_a$ ) then
9       activate ant and move to next place
10    else
11      stay at current site and sleep
12    end if
13  end for
14 adjust  $r, \alpha, \beta$ 
15 end for
16 Output locations of all objects;

```

PSEUDOCODE 1: The pseudocode of the ACK algorithm.

For an ant, the probability of being activated by the local neighbourhood is

$$P_a(o_i) = \left(\frac{\beta}{\beta + f(o_i)} \right)^2, \quad (8)$$

where $\beta \in R^+$ is the threshold of the ant's active fitness. Equation (8) is also similar to (2) in Section 3.1.1. When $f \ll \beta$, $P_a(o_i)$ is close to 1. Thus, if the fitness of o_i is much smaller than the threshold, o_i has a high probability of being activated. The active o_i moves in the plane and searches for a more comfortable place to sleep. When $f \gg \beta$, $P_a(o_i)$ is close to 0. Therefore, o_i does not wake up and continues to sleep. The principle of the ant movement model is the same as that in the LF algorithm [28]. The big difference is that the passive movement of objects is transformed into active movement.

The pseudocode of the main body of the ACK algorithm is given in Pseudocode 1.

3.2. CMAC for Vertical Assessment and Comparison. CMAC neural network has been employed to evaluate machine performance degradation states [7, 9]. The main advantages of CMAC against other neural networks are its local generalization, extremely fast learning speed, and easy implementation in software and hardware.

As shown in Figure 3, CMAC can be considered as an associative memory network, which performs two mappings:

$$\mathbf{S} \longrightarrow \mathbf{A} \longrightarrow \mathbf{P}, \quad (9)$$

where \mathbf{S} is the m -dimensional input space. The input variable $\mathbf{S}_i = [s_1^i, s_2^i, \dots, s_m^i]$ ($i = 1, 2, \dots, N$). \mathbf{A} is an n -dimensional memory cell vector. Every input variable S_i in \mathbf{S} only activates g elements in \mathbf{A} . g is called as generalization parameter ($g = 3$ in Figure 3). \mathbf{P} is one-dimensional output space. The weights in the activated memory cells are added to create the output.

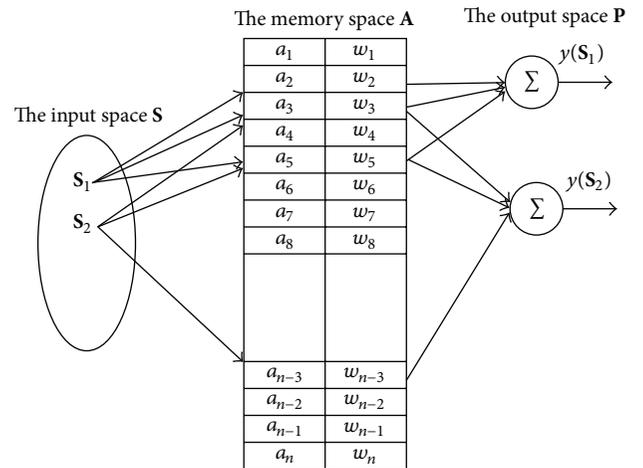


FIGURE 3: The structure of CMAC.

Before the mapping, each dimension s_k^i ($k = 1, 2, \dots, m$) in input variable \mathbf{S}_i should be quantized as

$$\hat{s}_k^i = \frac{s_k^i - s_k^{\min}}{r_k}, \quad (10)$$

where s_k^{\min} is the minimum of the k th dimension of \mathbf{S}_i ; r_k is the quantization parameter of the k th dimension of \mathbf{S}_i . The mapping address of s_k^i ($k = 1, 2, \dots, m$) will be determined according to its quantization value. Then, the first mapping combines all m -dimensional mapping address and projects the point \mathbf{S}_i in the input space into a binary associative vector \mathbf{A}_i . The elements in \mathbf{A}_i are defined as

$$a_{i,j} = \begin{cases} 1 & \text{if the } j\text{th element is activated} \\ & \text{by the } k\text{th sample,} \\ 0 & \text{otherwise,} \end{cases} \quad (11)$$

$$1 \leq j \leq n.$$

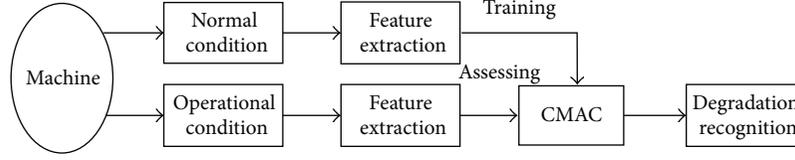


FIGURE 4: Using CMAC for machine degradation recognition.

The second mapping calculates the output of the network as scalar product of \mathbf{A}_i and the weight vector, \mathbf{W} , as shown in

$$Y_{i,r} = \mathbf{A}_i^T \cdot \mathbf{W} = \sum_{j=1}^n a_{i,j} w_j. \quad (12)$$

The weights are updated as

$$\begin{aligned} w_j(t+1) &= w_j(t) + \Delta w_j(t) \\ &= w_j(t) + \frac{\beta^* a_{i,j}}{g} \left(Y_{i,d} - \sum_{j=1}^n a_{i,j} w_j(t) \right), \end{aligned} \quad (13)$$

where i is the i th sample. $Y_{i,r}$ is the real output at the i th sample. w_j is the j th element in the weight vector \mathbf{W} . t is the t th cycle. β is the learning rate. $Y_{i,d}$ is the desired output at the i th sample. The initial value of w_j can be set to zero. The initial value of β should be in $(0, 2)$ and β can be reduced with the learning cycle increasing.

The mapping in CMAC model assures local generalization, which means that similar inputs create similar outputs while different inputs create nearly independent outputs. This generalization ability is a significant reason that CMAC can be used as a powerful tool for machine performance assessment. As shown in Figure 4, the features extracted from the normal condition of machine are used as the input of CMAC. The output in normal condition is set to 1. Then CMAC can be trained. After CMAC is trained well, the features in current operation can be inputted into CMAC. If the output is close to 1, this means that the running condition is near to normal. On the contrary, the running condition will be abnormal if the output deflects from 1 seriously. Therefore, the output of CMAC can be taken as a confidence value (CV), which can reflect the deflecting degree of machine's current condition from normal condition.

3.3. The Fusion of Horizontal and Vertical Assessment Results. When horizontal and vertical assessment results are gotten, they can be fused to create a more comprehensive assessment for machine performance. We will discuss two possible conditions. The first is that only one machine in the fleet has enough historical data. The second is that all machines' historical data in the fleet are collected sufficiently.

3.4. The Outlier Mining for Abnormal Machine Detection Based on Clustering Results. Cluster analysis can create some outliers as side products. As for the performance analysis of a fleet of machines, the outliers may represent the machines in fault states. Therefore, outlier detection is very important to

find abnormal machines. He et al. [32] presented a definition of cluster-based local outlier factor (CLOF) for identifying the physical significance of an outlier. The definition can be briefly described as follows.

Let the data set D be a set of points. After D is analyzed by a clustering algorithm, the clustering result is described as $C = \{C_1, C_2, \dots, C_k\}$, where

$$\begin{aligned} C_i \cap C_j &= \varphi, & C_1 \cup C_2 \cup \dots \cup C_k &= D, \\ & & 1 \leq i, j \leq k, i \neq j. \end{aligned} \quad (14)$$

k is the number of the clusters. They then defined Large Cluster (LC) and Small Cluster (SC) based on two parameters [32]. For the limit of the space, we will not describe in detail. Suppose that $C = \{C_1, C_2, \dots, C_k\}$ is the set of clusters with the sequence $|C_1| \geq |C_2| \geq \dots \geq |C_k|$, where that $|C_i|$ ($i = 1, 2, \dots, k$) is the number of the points in cluster C_i . As for each data point t , the CLOF of t is defined as

$$\text{CLOF} = \begin{cases} \frac{\min(\text{dis}(t, C_j))}{|C_i|}, & \text{where } t \in C_i, C_i \in SC, C_j \in LC, \\ \frac{\text{dis}(t, C_i)}{|C_i|}, & \text{where } t \in C_i, C_i \in LC, \end{cases} \quad (15)$$

where $\text{dis}(t, C_i)$ is the distance between t and C_i , which can be calculated by any distance formula in the clustering algorithm.

CLOF gives the way to measure the degree of the data point being an outlier. CLOF can indicate the degree of the abnormal machine departing from its nearest large cluster. But this nearest large cluster may represent machines with degradation performance. In fact, people are more interested in the degree of the outlier departing from machines with normal performance. A factor is needed to measure not only local but also global departing degree of outliers. So we suggest a definition of cluster-based global outlier factor (CGOF). Consider

$$\text{CGOF} = \begin{cases} \frac{\min(\text{dis}(t, C_j))}{|C_i|} * \overline{CV}(C_j), & \text{where } t \in C_i, C_i \in SC, C_j \in LC, \\ \frac{\text{dis}(t, C_i)}{|C_i|} * \overline{CV}(C_i), & \text{where } t \in C_i, C_i \in LC, \end{cases} \quad (16)$$

where $\overline{CV}(C_j)$ is the mean confidence value of all objects in the cluster C_j , which can be obtained by any vertical

TABLE 2: The comparisons of clustering time of three projection ways.

Algorithms	Random projection	PCA projection	KPCA projection
Clustering time(s)	192.36	132.43	94.66
Time saved (%)	\	31.15%	50.79%

assessing method. CGOF is gotten by CLOF multiplying a factor. The factor is the degree which describes that the outlier's nearest cluster degrades from the normal state. CGOF can measure not only local outlying degree but also global outlying degree of abnormal machine, which is very helpful to assess machine's performance.

4. Case Study and Analysis

We then used the ACK algorithm in a dataset which came from the monitoring data of the axial fans in one water cooling plant. There are 32 fans which are distributed in different places of the plant. The data of the fans are collected by the field bus to the server. These fans are of the same structure and work in the similar condition. They can compare each other to assess their performance. Three attributes to describe the fan's performance are gotten. They are the bearing temperature, the motor current, and the vibration of the spindle.

4.1. The Horizontal Assessment Based on ACK. We design two algorithms to compare with ACK. One applies random projection, and the other applies PCA projection. The clustering parts of two algorithms are the same as those in ACK. Figure 5 shows the comparison of three algorithms on the initial projections and the results after 600 cycles. Random projection needs a long time to create rough clusters. Some small rough clusters can be found after about 600 cycles. Compared with random projection, PCA and KPCA projection can create rough clusters at the initial stage. KPCA projection can create more separate and compressed clusters compared with PCA projection. These rough clusters can provide bases to create larger clusters around them, so the clustering time will be saved significantly.

Table 2 shows the comparison of clustering time. All the algorithms are run 10 times to get the average clustering time. The saved time is computed based on the time of random projection. Both PCA and KPCA projections can save time because of the creations of rough clusters at initial stage. KPCA can save more time compared with PCA. Compared with random projection, KPCA can save almost a half part of time.

The parameters of ACK in this case are listed as follows $t_{\max} = 2000$, $r = 2$, $\beta = 0.1$, $\sigma = 0.75$. Figure 6 shows the final clustering results. The fans are clustered to three different groups. The features of each group are listed in Table 3.

To compare and validate the clustering results, we use other three methods, K -means, KK -means (Kernel K -means), and SOM, to cluster the data. As for K -means and

KK -means, we set the number of cluster centers as three. SOM does not need this prior knowledge. Actually, SOM gets the number of the clusters as four. But there are only two fans in the fourth cluster. For certain points, these four methods get disputable results, which are listed in Table 4. We can see that ACK gets the similar results with KK -means. The reason is that these two algorithms cluster the data in the feature space after kernel mapping. Except for fans number 3 and number 10, the clustering results of ACK are consistent with the final possible conclusions.

Through above analysis, we can see that the performance assessment of the axial fans can be achieved by ant-based clustering algorithm in a self-organizing way. The number of the clusters is not needed as a prior knowledge here. In the present condition monitoring system, a lot of data created by a fleet of machines, like fans and motors, are not used sufficiently. The common style to process these data is showing them in the display screen, or giving the diagrams of certain important parameters. The comparison among machines in a fleet has not been performed.

Figure 7(a) shows the radar chart of the temperature feature of axial fans. Around the circumference is the number of fans, and the corresponding radial value is its temperature. The advantage of radar chart is that it can compare one single feature for machines clearly. Some special points, such as the minimum and maximum points, can be easily detected for each feature. But it cannot show the comparison of the general performance including all the features. Figure 7(b) shows the combination of ant-based clustering assessing results with the radar chart of the temperature. Different signs and colors are used to identify the different clusters. So we can easily recognize which group the fan belongs to and pay more attention to its difference with its similar ones. But this chart still cannot show the comprehensive performance of each machine. We will show the final assessing results of radar chart in Section 4.3 where the clustering results are fused with the vertical assessing results of the single machine.

4.2. The Vertical Assessment Based on CMAC. If the historical data of one fan for a long time can be gotten, CMAC can be used to assess the fan's performance degradation during its running. In this case, we can get the full monitoring data of fan number 7. Figure 8 shows the averaged features every day. The initial running states of 46 days are taken as the normal states. The Confidence Values (CVs) at these days are 1. Then the features and CVs are inputted to CMAC to train it. The training parameters in this case are $r = [1, 10, 1]$, $\beta = 1.0$, and $\text{Max_cycle} = 1000$. After CMAC is trained, the features at the operational condition are inputted to CMAC to get the CVs outputs. Figure 8 shows the CV outputs from CMAC. The CVs at different time can show the degree of performance degradation based on the normal state.

As for the single machine's performance assessment, the best way is that the machine's historical data can be collected. But the historical data may be not rich enough for every machine in a fleet because of all kinds of reasons. In our case, if only the data of fan number 7 is full and collected for a long time, can we assess the other fans' performance based

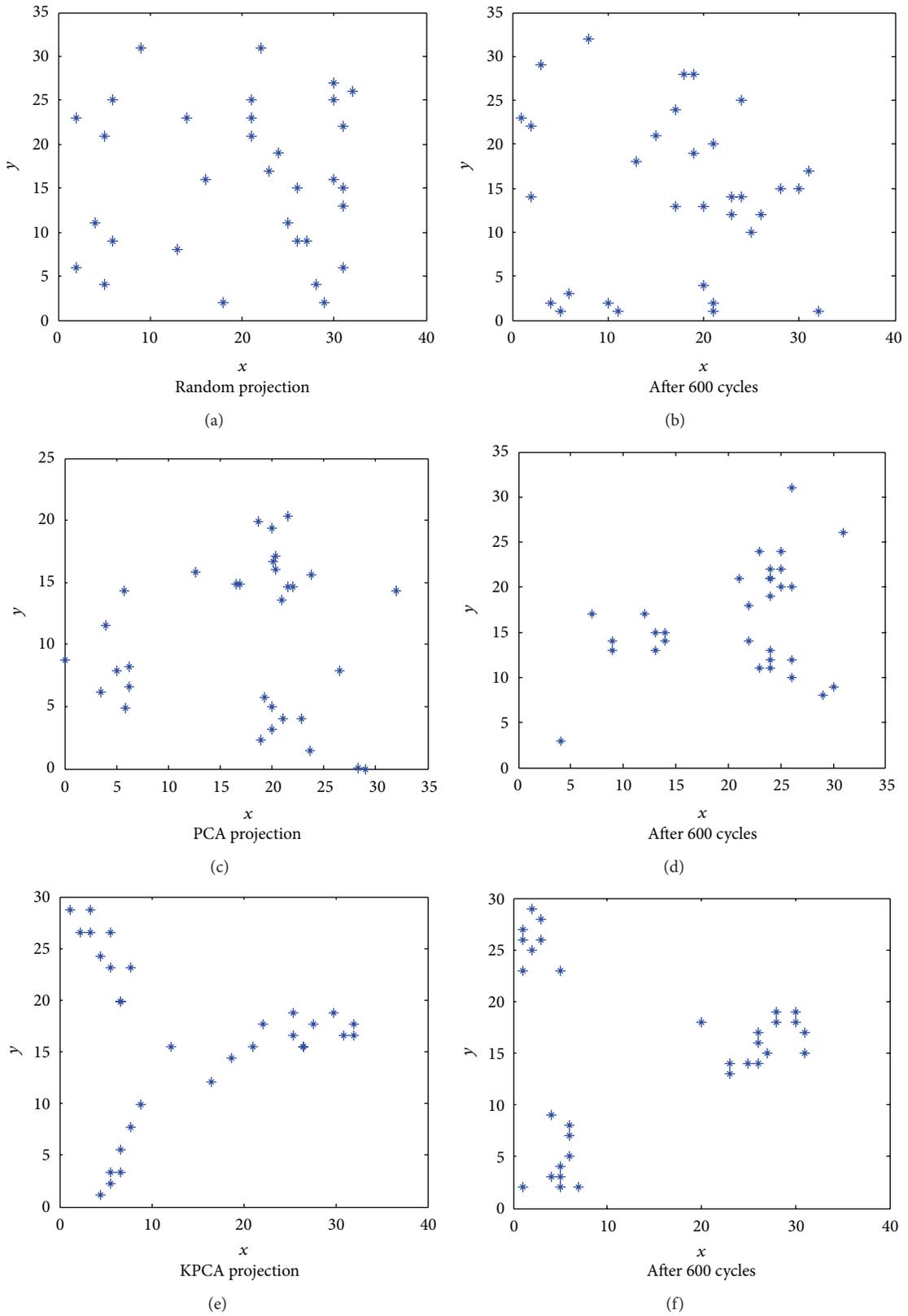


FIGURE 5: The comparison of three projection ways and clustering results after 600 cycles.

TABLE 3: The features of three clusters (L: low; M: medium; H: high).

Cluster	Number of fans	Members	Current (A)	Temperature (°C)	Vibration (mm)
C1	8	1, 4, 5, 11, 16, 17, 18, 19	200.10 (L)	44.80 (L)	6.24 (L)
C2	10	2, 3, 6, 7, 8, 20, 21, 27, 28, 29	222.59 (H)	50.76 (H)	7.45 (M)
C3	14	9, 10, 12, 13, 14, 15, 22, 23, 24, 25, 26, 30, 31, 32	208.95 (M)	48.25 (M)	9.09 (H)

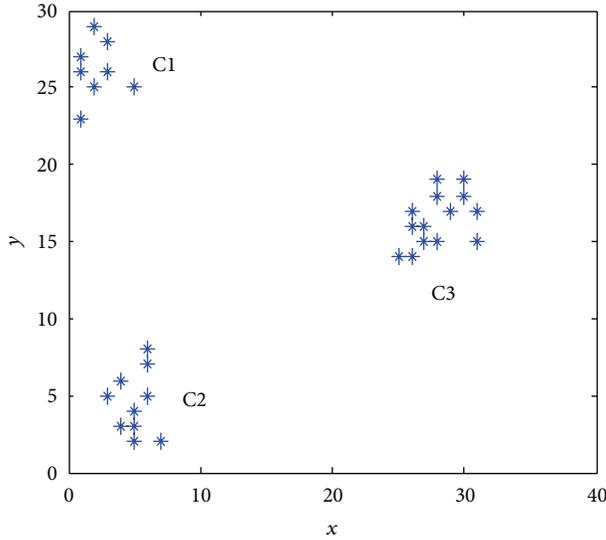


FIGURE 6: The final clustering results of ACK.

TABLE 4: The points with disputable clustering results in four methods.

Fan no.	ACK	K-means	KK-means	SOM	Final conclusion
3	2	2	3	3	2 or 3
10	3	1	3	1	3 or 1
25	3	1	3	2	3
27	3	3	3	4	3
31	3	1	3	3	3
32	3	3	3	4	3

on these data? In the next section, we will discuss how to fuse horizontal and vertical assessments to answer this question.

4.3. The Combination of Horizontal and Vertical Assessment Results. In Section 4.1, horizontal comparison of all fans is conducted using ant-based clustering algorithm. The used data are the average features at the 240th day (which is shown as the horizontal assessing time in Figure 8). In this section, we will discuss how to combine horizontal and vertical assessment results in two possible conditions. The first is only that fan number 7 has a full historical data. The second is that all fans' historical data are collected sufficiently.

4.3.1. The Fans Have no Enough Historical Data except Fan number 7. The conclusion of vertical assessment is that fan number 7 has a confidence value 0.804. The conclusion of horizontal assessment is that fan number 2 and number 7

are clustered in the same group. Then the simplest combined conclusion is the confidence value of fan number 2 is around 0.804. If we can have another fan's full historical data, for example, fan number 6, then the confidence value of fans number 2 can be gotten by averaging the confidence values of fan number 7, and fan number 6 (fan number 2, fan number 7 and fan number 6 are clustered in the same group). By this combination, although fan number 2 has no historical data, we can also assess its performance based on its same machine in a fleet.

Moreover, we can set up another CMAC model. Use the normal features of fan number 7 to train it (the output CV is 1); then the features of other fans at the horizontal assessing time are inputted to CMAC to get the corresponding CV outputs. Figure 9 shows a radar chart of all fans' CVs in the horizontal assessment time in Figure 8. Different clusters are fitted in different circles. The figure can give a clear display of the comprehensive performance comparison of fans. The confidence values smaller than 0.5 should be paid more attention because these machines are more likely to degrade to a failure state. Although this assessment is only based on one machine's normal data, it is still a good way to assess a fleet of machines especially in the case that the historical data for most machines are not sufficient.

4.3.2. All Fans' Historical Data Are Collected Sufficiently. If all fans' data can be collected sufficiently, the results of horizontal and vertical assessments can be used to test whether the assessing conclusion is consistent. For example, the conclusions of vertical assessment are that fan number 7 has a confidence value 0.804 and fan number 2 has a confidence value 0.798. The conclusion of horizontal assessment is that fans number 2 and number 7 are clustered in the same group. Then the results of horizontal and vertical assessments are consistent. If fan number 2 has a confidence value 0.398 (degraded seriously) while fans number 2 and number 7 are still clustered in the same group in horizontal assessments, the results of horizontal and vertical assessment are contradictory. Then further analysis should be conducted to make sure which conclusion is correct. For example, other clustering methods or vertical assessing methods should be used to analyze data again.

4.4. Discussion on Outlier Mining Based on Assessing Results. We use an example to show how to detect the outliers by ant-based clustering. During the operation of the fans, an artificial interference is added to three fans to make their vibration higher than normal operation. The clustering results of the fans are shown in Figure 10.

Three outliers are obviously clustered to a single cluster C4. C4 is a small cluster. Based on the analysis of Section

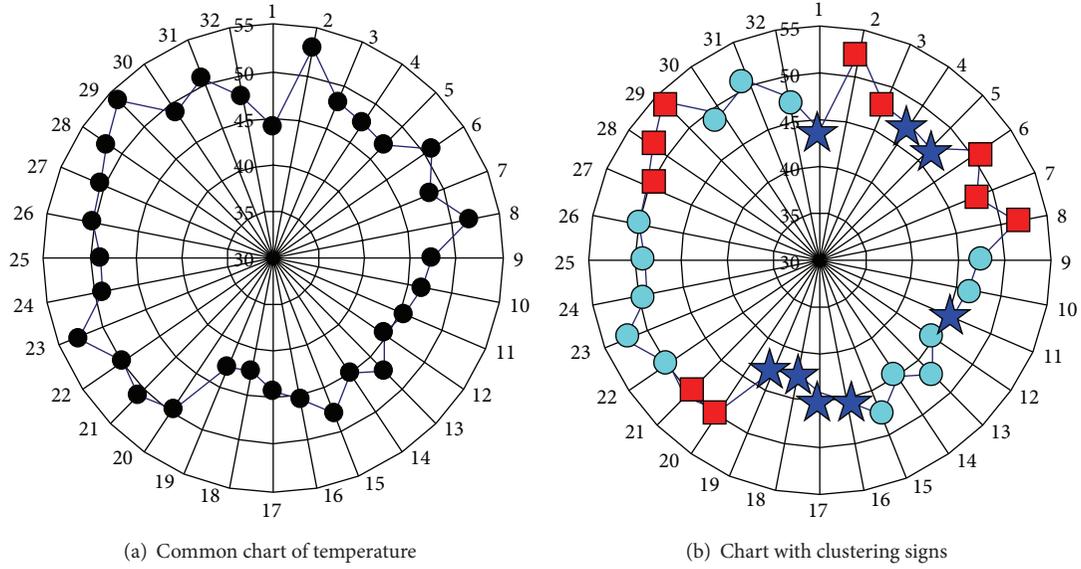


FIGURE 7: The comparison of radar charts of temperature.

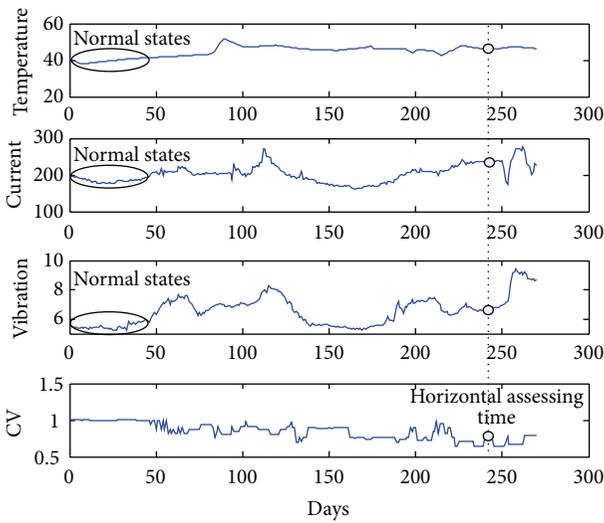


FIGURE 8: The vertical assessment of CMAC for fan number 7.

3.1.3, we get that the local neighborhood cluster of C4 is C3. Actually, based on the assessment results, the cluster C3 is a group with degradation state. The average CV of C3 is 0.677. Then the outlier factors can be computed using (15) and (16) in Section 3.4. The outlier factors are listed in Table 5. We can see that the factors can quantitatively represent the outlying degree of each outlier. The outliers should be paid more attention because they may mean that the abnormal behavior has happened.

5. Conclusions

A combined method of ant-based clustering and CMAC is suggested to assess the performance of machines in a fleet. Machines in a fleet can be compared to each other

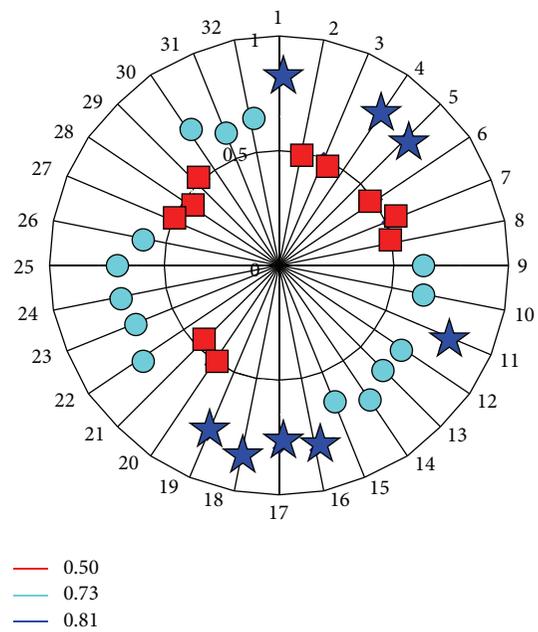


FIGURE 9: The confidence values comparison of fans.

because they have the same structures and work in the same condition. Ant-based clustering algorithm does not need any prior knowledge. It is proper to cluster machines in a self-organizing way. The application of kernel in ant-based clustering can improve the efficiency and clustering accuracy, which is achieved by modifying projection based on KPCA and clustering in feature space. The paper also described vertical assessment method based on CMAC for the single machine. The confidence value of the machine can indicate its degradation degree compared with historical normal data.

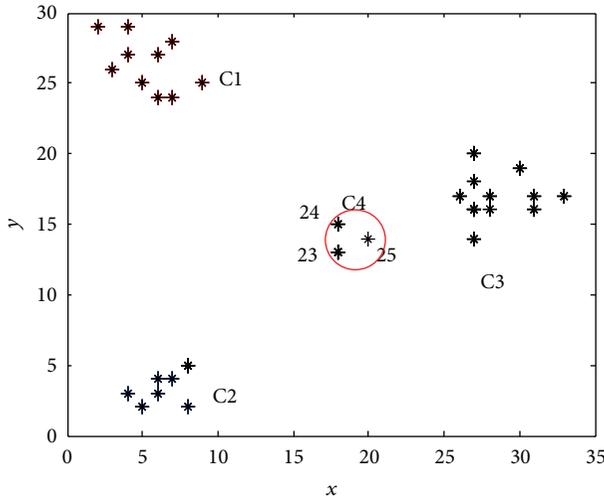


FIGURE 10: The clustering results with outliers.

TABLE 5: The outlying degree of three outliers.

Fan no.	CLOF	CGOF
23	5.1776	3.505
24	2.0163	1.365
25	3.5836	2.426

The suggested method was finally used for assessing the performance of axial fans. The results show that the method performs well. The method gives performance assessment for the machines in time domain and space domain. Even if the historical data of one machine is not available, the performance analysis can be conducted based on its similar ones in the fleet.

The following items are needed to be further studied:

- (i) the adaptive adjusting method of the parameters in ACK algorithm,
- (ii) the applications of the method in more datasets,
- (iii) the deep fusion of vertical and horizontal assessing results.

Acknowledgments

This paper is supported by the National Natural Science Foundation of China (no. 50705054) and the Key National Project of ITER China (no. 2011GB113005).

References

- [1] J. Lee, J. Ni, D. Djurdjanovic, H. Qiu, and H. Liao, "Intelligent prognostics tools and e-maintenance," *Computers in Industry*, vol. 57, no. 6, pp. 476–489, 2006.
- [2] J. Yan and J. Lee, "Degradation assessment and fault modes classification using logistic regression," *Journal of Manufacturing Science and Engineering*, vol. 127, no. 4, pp. 912–914, 2005.
- [3] J. H. Yan, N. Isobe, and J. Lee, "Fuzzy logic combined logistic regression methodology for gas turbine first-stage nozzle life prediction," *Applied Mechanics and Materials*, vol. 10-12, pp. 583–587, 2008.
- [4] G. Yu, J. Sun, and C. Li, "Machine performance assessment using Gaussian mixture model (GMM)," in *Proceedings of the 2nd International Symposium on Systems and Control in Aerospace and Astronautics (ISSCAA '08)*, pp. 1–6, December 2008.
- [5] J. Yu, "Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models," *Mechanical Systems and Signal Processing*, vol. 25, no. 7, pp. 2573–2588, 2011.
- [6] J. Lee, "Measurement of machine performance degradation using a neural network model," *Computers in Industry*, vol. 30, no. 3, pp. 193–209, 1996.
- [7] C.-C. Lin and H.-P. Wang, "Performance analysis of rotating machinery using enhanced cerebellar model articulation controller (E-CMAC) neural networks," *Computers and Industrial Engineering*, vol. 30, no. 2, pp. 227–242, 1996.
- [8] L. Zhang, J. Lee, Q. Cao, and L. Wang, "Advanced credit-assignment CMAC algorithm for robust self-learning and self-maintenance machine," *Tsinghua Science and Technology*, vol. 9, no. 5, pp. 519–526, 2004.
- [9] R.-Z. Xu, L. Xie, M.-C. Zhang, and C.-X. Li, "Machine degradation analysis using fuzzy CMAC neural network approach," *International Journal of Advanced Manufacturing Technology*, vol. 36, no. 7-8, pp. 765–772, 2008.
- [10] F. Wu, T. Wang, and J. Lee, "An online adaptive condition-based maintenance method for mechanical systems," *Mechanical Systems and Signal Processing*, vol. 24, no. 8, pp. 2985–2995, 2010.
- [11] H. Qiu, J. Lee, J. Lin, and G. Yu, "Robust performance degradation assessment methods for enhanced rolling element bearing prognostics," *Advanced Engineering Informatics*, vol. 17, no. 3-4, pp. 127–140, 2003.
- [12] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *Journal of Sound and Vibration*, vol. 289, no. 4-5, pp. 1066–1090, 2006.
- [13] Y. Pan, J. Chen, and X. Li, "Bearing performance degradation assessment based on lifting wavelet packet decomposition and fuzzy c-means," *Mechanical Systems and Signal Processing*, vol. 24, no. 2, pp. 559–566, 2010.
- [14] E. Lapira, D. Brisset, H. Davari Ardakani, D. Siegel, and J. Lee, "Wind turbine performance assessment using multi-regime modeling approach," *Renewable Energy*, vol. 45, pp. 86–95, 2012.
- [15] W. Caesarendra, A. Widodo, and B.-S. Yang, "Application of relevance vector machine and logistic regression for machine degradation assessment," *Mechanical Systems and Signal Processing*, vol. 24, no. 4, pp. 1161–1171, 2010.
- [16] L. Liao and J. Lee, "A novel method for machine performance degradation assessment based on fixed cycle features test," *Journal of Sound and Vibration*, vol. 326, no. 3-5, pp. 894–908, 2009.
- [17] J. Raza, L. Jayantha P., H. A. Atat, and L. Jay, "Methodology and theory a comparative study of maintenance data classification based on neural networks, logistic regression and support vector machines," *Journal of Quality in Maintenance Engineering*, vol. 16, no. 3, pp. 303–318, 2010.
- [18] G. Yu, H. Qiu, D. Djurdjanovic, and J. Lee, "Feature signature prediction of a boring process using neural network modeling with confidence bounds," *International Journal of Advanced Manufacturing Technology*, vol. 30, no. 7-8, pp. 614–621, 2006.

- [19] J. Liu, D. Djurdjanovic, J. Ni, N. Casotto, and J. Lee, "Similarity based method for manufacturing process performance prediction and diagnosis," *Computers in Industry*, vol. 58, no. 6, pp. 558–566, 2007.
- [20] D. Siegel, C. Ly, and J. Lee, "Methodology and framework for predicting helicopter rolling element bearing failure," *IEEE Transactions on Reliability*, vol. 61, no. 4, pp. 846–857, 2012.
- [21] V. T. Tran, H. T. Pham, B. S. Yang, and T. T. Nguyen, "Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and SVM," in *Engineering Asset Management and Infrastructure Sustainability*, J. Mathew et al., Ed., pp. 959–972, 2011.
- [22] V. N. Vapnik, S. Golowich, and A. Smola, "Support vector machine for function approximation regression estimation and signal processing," *Advances in Neural Information Processing System*, vol. 9, pp. 281–287, 1996.
- [23] R. Huang, L. Xi, X. Li, C. Richard Liu, H. Qiu, and J. Lee, "Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods," *Mechanical Systems and Signal Processing*, vol. 21, no. 1, pp. 193–207, 2007.
- [24] N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, "Residual life predictions from vibration-based degradation signals: a neural network approach," *IEEE Transactions on Industrial Electronics*, vol. 51, no. 3, pp. 694–700, 2004.
- [25] D. Djurdjanovic, J. Lee, and J. Ni, "Watchdog agent—an infotronics-based prognostics approach for product performance degradation assessment and prediction," *Advanced Engineering Informatics*, vol. 17, no. 3-4, pp. 109–125, 2003.
- [26] M. Dorigo, E. Bonabeau, and G. Theraulaz, "Ant algorithms and stigmergy," *Future Generation Computer Systems*, vol. 16, no. 8, pp. 851–871, 2000.
- [27] L. Zhang and Q. Cao, "A novel ant-based clustering algorithm using the kernel method," *Information Sciences*, vol. 181, no. 20, pp. 4658–4672, 2011.
- [28] E. Lumer and B. Faieta, "Diversity and adaptation in populations of clustering ants," in *Proceedings of the 3rd International Conference on Simulation of Adaptive Behavior: From Animals to Animats*, pp. 501–508, MIT Press/Bradford Books, Cambridge, Mass, USA, 1994.
- [29] J. Handl and B. Meyer, "Ant-based and swarm-based clustering," *Swarm Intelligence*, vol. 1, no. 2, pp. 95–113, 2007.
- [30] L. Zhang, Q. X. Cao, and J. Lee, "A modified clustering algorithm based on swarm intelligence," in *Proceedings of the 1st International Conference on Natural Computation (ICNC '05)*, vol. 3, pp. 535–542, Changsha, China, 2005.
- [31] X. Xu, L. Chen, and Y. Chen, "A4C: an adaptive artificial ants clustering algorithm," in *Proceedings of IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB '04)*, pp. 268–275, October 2004.
- [32] Z. He, X. Xu, and S. Deng, "Discovering cluster-based local outliers," *Pattern Recognition Letters*, vol. 24, no. 9-10, pp. 1641–1650, 2003.

Research Article

Improved Visual Hook Capturing and Tracking for Precision Hoisting of Tower Crane

Yanming Li, Shuangyuan Wang, and Bingchu Li

School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Correspondence should be addressed to Yanming Li; yml@sjtu.edu.cn

Received 12 April 2013; Accepted 27 May 2013

Academic Editor: Liang Gong

Copyright © 2013 Yanming Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To maintain safe operation of the tower crane, it is important to monitor the activities of the hook system. Visual monitoring and image recognition are the optimum methods for crane hook tracking and precision hoisting. High real-time performance and low computation requirements are required for tower crane hook capturing and tracking system which is implemented on the embedded Advanced RISC Machines (ARM) processor or Microcontrol Unit (MCU). Using the lift rope of a tower crane as the target object, a new high-performance hook tracking method suitable for ARM processor or MCU applications is presented. The features of the lifting process are analyzed, and an improved progressive probabilistic Hough transform (IPPHT) algorithm is proposed which can reduce capturing time by up to 80%. Combining color histogram with a binary search algorithm, an adaptive zooming method for precise hoisting is presented. Using this method the optimum zoom scale can be achieved within a few iterations.

1. Introduction

Tower cranes are important hoisting machinery widely used for construction. Advancing toward large-scale, high-speed, and automated operation leads to dramatic increases in losses caused by accidents that increase dramatically in the tower crane industry. Although many safe standards have been formulated, the recent spate of tower crane accidents has casted a spotlight on the need for tower crane safety [1]. Load hoisting is the most critical task performed by tower cranes, determining project quality, and it is during this task that accidents occur. Automatic monitoring and controlling of hoisting cannot only reduce the severity and numbers of tower crane accidents but also improve operation quality and efficiency.

Working safety in tower crane operation has recently been highlighted, and many types of safety devices have been designed. These devices can be divided into two classes: devices used to protect equipment from overloading [2, 3] and devices used to keep people away from hazard zone [4, 5]. Loads are often lifted over building constructions, thus resulting in blind zone for operators. Blind lifting is one of the main practices that adversely affect tower crane safety [6].

Hoisting process monitoring is a useful way to avoid blind lifting. However, little information can be provided to monitor the hoisting process using the aforementioned devices fore-mentioned. The hoisting process can be characterized by the motion of crane hooks; thus, capturing and tracking of crane hook motion are feasible methods for hoist-monitoring and evaluation.

Visual monitoring of working activities through the installation of surveillance cameras has become prevalent in the construction industry. A previous study detailed the development of a tower-crane-mounted live video system [7]. On-site testing proved that the system can improve site productivity and safety, but the imagery lacked the visual depth perception rendered by human eyes due to weak capacities for adaptive tracking and zooming in the loads. Visual tracking of a crane jib and trolley was used to monitor tower crane activities [8]. Rather than focusing on detecting a specific load carried by the tower crane, the hoisting activities of the crane were inferred using information gathered from both the crane jib and the trolley in conjunction with the layout plans of the site considered. In another study, a crossmark was labeled on the hook of a truck crane to facilitate recognition by camera, and the hook was tracked using a PID control

scheme [9]. The recognition performance was greatly affected by the rotation angle of the hook, and it was determined that the camera should be perpendicular to the crossmark. Continuously adaptive mean shift (Camshift) is an efficient tracking algorithm [10], and an improved Camshift tracking algorithm was proposed for a testing system that performs tower crane automatic tracking and zoom video monitoring [11]. The Camshift algorithm is too complex to implement using an Advanced RISC Machine (ARM) system, which is used in field activity monitoring, thus leading to low real-time performance for hook tracking. To measure the three-dimensional position of a moving crane hook, hook motion is tracked according to the variation in its area using optical flow and orientation code matching [12]. In addition to image processing techniques, virtual visualization and simulation are also often adopted for activity tracking and project management. Activities are reconstructed using a 3D or 4D model of the crane according to real-time monitoring data and building information modeling (BIM) [13–15]. Virtual visualization and simulation are advanced methods used for project management, but their output is too rough for activity tracking and precision hoisting due to a lack of detailed information regarding the activities performed by a crane and the crane's surroundings.

Currently, visual monitoring and image recognition are still the premier methods for crane hook tracking and precision hoisting. High real-time performance and low computation complexity are two key attributes of a practical tracking algorithm used for real-world applications. The controller of the tower crane hook capturing and tracking system is usually ARM-based or MCU-based. However, the operational capabilities of these processors are inferior to those of a personal computer; thus, more efficient algorithms are needed to guarantee high real-time performance. By selecting the lift rope of a tower crane as the captured object, a new highly efficient hook tracking method suited for ARM or MCU applications is presented in this paper. The hoisting features and reasons for selecting the rope as the captured object are analyzed, and an improved progressive probabilistic Hough transform (IPPHT) algorithm is proposed, which reduces the computational time by almost 80%. In this study, an adaptive zooming method for precision hoisting was developed by integrating color histograms with a binary search algorithm. The method was tested, and the results are herein presented.

2. System Architecture and Information Flow

Figure 1 is an illustration of the system architecture designed for the adaptive load capturing and tracking of a tower crane hook. The image information recorded by the camera is converted from analog to digital, and the IPPHT algorithm is then applied to capture the movement of the rope and to track the trajectory of the hook quickly and accurately. A Camshift-based adaptive zooming algorithm integrating color histograms with a binary search algorithm is implemented and executed to provide a detailed observation of hook activities and the surroundings, which is necessary for safe, precision

hoisting. The functions of capturing, tracking, and zooming are performed by a pan-tilt camera controlled by the PELCO-Protocol. The capturing and tracking video can be displayed for field monitoring and operation guidance; moreover, the imagery can be transmitted to a remote server for activity monitoring and project management after being compressed. The IPPHT-based capturing and tracking algorithm and adaptive zooming algorithm are the key components of the system, especially when the system is implemented using an ARM9 processor, which demands high computation efficiency. The algorithms are described in Section 3 and Section 4, respectively.

3. IPPHT-Based Algorithm for Rope Capturing and Hook Tracking

3.1. Analysis of Rope Features during Hoisting. One of the purposes of capturing and tracking hook movements is to maintain a suitable size for the hook and its load in the center of the visual field. There are two factors that make it difficult to select the hook and its load as the objects to be captured and tracked. One is the fact that many shape features generated by buildings are intermingled with images of the hook and its load; the other is the fact that the shape and size of the hook in visual field vary greatly with changes in the rotation and position of the hook when hoisting.

Figures 2(b) and 2(d) are images of the rope and hook obtained from the images as shown in Figures 2(a) and 2(c), respectively, using the Canny edge detection algorithm. Compared with the hook, the hoisting rope offers some advantages as the object to be captured and tracked. First, the hook is connected to the end of the rope; therefore, the hook position can be indicated using the coordinates of the end point of the rope. Second, the shape of the hoisting rope is linear; thus, the rope is easily recognized and captured compared to the complex shape of the hook. Finally, the rope end only slides vertically to the ground; therefore, it can be conveniently tracked. We observed the following features of the hoisting rope and hook.

- (1) The rope is nearly perpendicular to the ground with $\theta \leq \pm 20^\circ$, where θ is the swing angle of the rope.
- (2) As illustrated in Figure 2, the rope in the images processed using the Canny edge detection is represented by two parallel edges. However, it may become a single edge due to the noise generated on one side of the rope when the image is taken from a long distance.
- (3) As illustrated in Figure 2(b), a slight curve sometimes appears in the rope images due to wind effects or the swimming of suspension arm. As a result, the line detection algorithm may fail and many line segments with slopes may be generated.
- (4) In addition to the rope, many other line shapes generated by steel structures and buildings are also detected in the images.

3.2. Rope Capturing and Hook Tracking Based on Improved Progressive Probabilistic Hough Transform (IPPHT). The

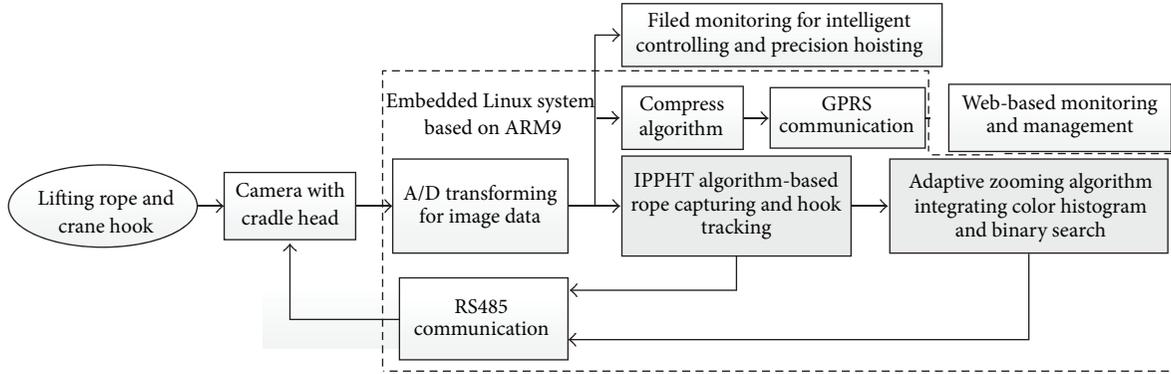


FIGURE 1: Overall architectures of adaptive load capturing and tracking system.

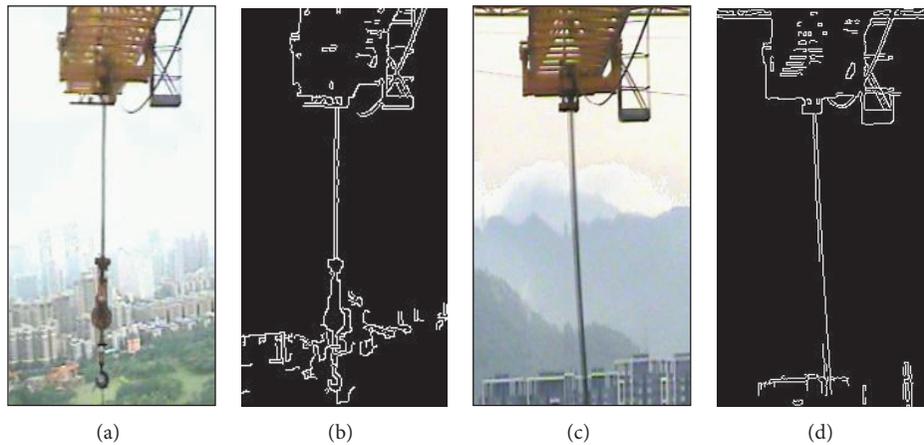


FIGURE 2: Images of rope and hook processed by Canny edge detection algorithm.

Hough transform (HT) is a popular and robust method for detecting lines in an image [16]. However, high computational and storage requirements are the drawbacks of the standard Hough transform (SHT) applied for real-time detection. Moreover, this simple transform fails to determine any information regarding the lengths or the start and end points of lines. Therefore, many variations of Hough's original transform have been proposed. The progressive probabilistic Hough transform (PPHT) can minimize the proportion of points that are used in voting when trying to maintain false-negative and false-positive detection rates at the level achieved by SHT [17]; thus, the requirements of computational and storage requirements can be reduced. On the other hand, the output of the PPHT can be denoted by the coordinates of the two end points of a line segment, which makes the transform suited for detecting the lift rope of a tower crane.

Although the computational and storage requirements are reduced, the PPHT is still too complex for the system shown in Figure 1, in which image recognition and tracking are implemented using an ARM-based system. Test reveals that the algorithm will spend 2 s to 3 s to detect image measuring a picture of 160×320 pixels using an S3C2440 processor at the basic frequency of 400 MHz. On the other hand,

the PPHT is not satisfactory for the detection of a rope with the features mentioned in Section 3.2. Thus, an improved progressive probabilistic Hough transform (IPPHT) is presented for hoist rope detection and hook capture.

3.2.1. Progressive Probabilistic Hough Transform with Angle Constraints. Let (ρ, θ) denote a point in the parameter space and (x, y) denote a point on a straight line in the image space; the parameter ρ represents the distance between the line and the origin, whereas θ is the angle of the vector from the origin to the closest point on the line. The basic mapping relation is expressed as follows [16]:

$$\rho = x \cos \theta + y \sin \theta. \quad (1)$$

Figure 3 illustrates the mapping relation of the HT for line detection. As shown in Figure 3, θ is also the angle of the line relative to the y -axis. Because the lift rope swings only from -20° to $+20^\circ$, a constraint of $\theta \in [-20^\circ, 20^\circ]$ is applied to (1) for the lift rope detection. Thus, (1) can be expressed as follows:

$$\rho = x \cos \theta + y \sin \theta, \quad -20^\circ \leq \theta \leq 20^\circ. \quad (2)$$

Let N_θ denote the number of bins for the angle, and let N_ρ denote the number of bins for the distance from the origin.

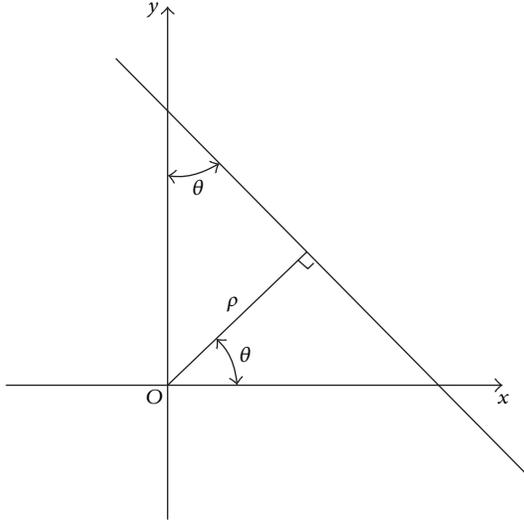


FIGURE 3: Mapping relation of HT and the angle of the line relative to the y axis.

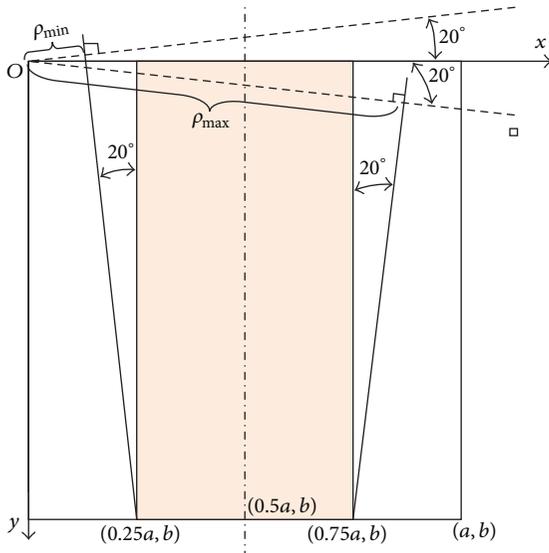


FIGURE 4: Ranges of ρ .

The parameter space is divided into $N_\rho \times N_\theta$ accumulators, which can be represented by the two-dimensional array $A[m][n]$. $a[i][j]$ is the element of the i th row and the j th column, which is used to store the vote number at the parameter pair $(\rho_0 + i\Delta\rho, \theta_0 + j\Delta\theta)$ and

$$\begin{aligned} \rho_0 &= [\rho_{\min}] = \max \{n \in Z \mid n \leq \rho_{\min}\}, \\ \Delta\rho &= 1, \quad \Delta\theta = 0.5^\circ, \\ \theta_0 &= -20^\circ. \end{aligned} \quad (3)$$

ρ_{\min} is the minimum of ρ . Figure 4 illustrates the range of ρ . Because the camera is mounted on the beam of the crane and the lift rope swings only from -20° to $+20^\circ$, the rope should appear in the middle zone of the visual field.

If the width of the image is a and image height is b , then a subzone from $0.25a$ to $0.75a$ along the x -direction can be selected as the rope detection zone. The zone can be derived from Figure 4 as follows:

$$\begin{aligned} \rho_{\min} &= 0.25a \cos 20^\circ - b \sin 20^\circ, \\ \rho_{\max} &= 0.75a \cos 20^\circ + b \sin 20^\circ. \end{aligned} \quad (4)$$

ρ_{\max} is the maximum of ρ . Thus,

$$\begin{aligned} N_\rho &= \left\{ n \in Z \mid n \leq \frac{(\rho_{\max} - \rho_{\min})}{\Delta\rho} \right\}, \\ N_\theta &= \frac{(20^\circ - (-20^\circ))}{\Delta\theta} = 80. \end{aligned} \quad (5)$$

The PPHT algorithm with angle constraint is improved as follows.

Step 1. Define a subzone as shown in Figure 4 for lift rope detection.

Step 2. Obtain an image with the Canny edge detection algorithm.

Step 3. Check the input image; if it is empty, then terminate.

Step 4. Vote into the accumulator with a single pixel randomly selected from the input image according to (2).

Step 5. Check if the highest peak in the accumulator that was modified by the new pixel is higher than predetermined threshold $\text{thr}(N)$. If it is not, then go to Step 4. The threshold $\text{thr}(N)$ is determined with a probability estimation which will be described in the next section.

Step 6. Search along a corridor specified by the peak in the accumulator, and find the longest segment that either is continuous or exhibits a gap not exceeding a given threshold. Eliminate noises and amend curves when searching.

Step 7. Remove the pixels in the segment from the input image.

Step 8. Cancel the accumulator votes of all the pixels from the line that have previously voted.

Step 9. If the line segment is longer than the minimum length, add it to the output list.

Step 10. Go to Step 3.

3.2.2. Termination Condition for Voting. In the IPPHT, the termination of voting is based on monitoring the polling process. The criterion is based on a measure of stability of the ranks of the highest peaks [17]. Let $C(\rho, \theta)$ denotes the voting number at parameter pair (ρ, θ) ; thus, $C(\rho, \theta)$ is an independent random variable with binomial distribution $B(N, p)$, where N is the number of edge pixels that have voted

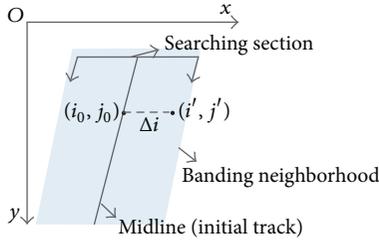


FIGURE 5: Search method.

so far and $\rho = 1/N_\rho$ is the probability of selecting a particular bin with a given θ . The termination condition for voting is

$$P(C(\rho, \theta) > \text{thr}(N)) < l. \quad (6)$$

In (5), $\text{thr}(N)$ is a threshold that needs to be determined, and $l = 5\%$ is a user parameter that indicates the number of false positives.

It has been proved that results identical to those of SHT can be obtained even if only a fraction of input points as low as 2% is used in the voting process of the PPHT [17]. Let $N = 50$ and $N_\rho \gg 10$ ($p = 1/N_\rho \ll 0.1$); thus, the binomial distribution $B(N, p)$ can be well approximated by a Poisson distribution $P(\lambda)$, where $\lambda = Np = N/N_\rho$. Therefore, the value of $\text{thr}(N)$ can be obtained using a single lookup in the Poisson distribution table.

3.2.3. Search Method of Line Segment for Noise Elimination and Curve Amendment. To eliminate edge noises and to amend the curve in the rope image mentioned in Section 3.1, a line segment search method is used. Figure 5 illustrates the search method. The line representing the lift rope in the processed image is nearly parallel to the y -axis, with an angle varying from -20° to 20° . Suppose $I(i_0, j_0)$ is a point on the line. A line segment perpendicular to the y -axis is selected as the search section. $I(i_0, j_0)$ is the midpoint of the search section. Define

$$T_{i_0, j_0} = \bigvee_{\substack{i_0-n \leq i \leq i_0+n \\ j=j_0}} P_{i, j}. \quad (7)$$

Then, T_{i_0, j_0} is selected as the condition to determine whether the search should end or continue.

If $T_{i_0, j_0} = 1$, then $i_0 = i_0 + di$, $j_0 = j_0 + dj$, and the search process continues.

If $T_{i_0, j_0} = 0$, then the search process ends, and $I(i_0, j_0)$ is the end point of the line.

In (7), $P_{i, j}$ is a Boolean variable. $P_{i, j} = 1$ indicates that a pixel exists at coordinates (i, j) , and $P_{i, j} = 0$ indicates that there is no pixel at coordinates (i, j) . n is the half width of the banding neighborhood.

The voting rule is modified relative to the rule in the PPHT in order to eliminate the edge noise and to amend the curve in the rope image mentioned in Section 3.1. As illustrated in Figure 5, point $I(i_0, j_0)$ is the midpoint of

the line, and point $I'(i', j')$ is a random point in the search section at point $I(i_0, j_0)$. Moreover,

$$\begin{aligned} i' &= i_0 + \Delta i, \\ j' &= j_0. \end{aligned} \quad (8)$$

Define $C(\rho, \theta)$ as the voting number at (ρ, θ) ; the initial value of $C(\rho, \theta)$ is $C(\rho, \theta) = 0$. Consider

$$\begin{aligned} A &= \{(\rho, \theta) \mid \rho = i_0 \cos \theta + j_0 \sin \theta \text{ and } -20^\circ \leq \theta \leq 20^\circ\}, \\ B &= \{(\rho, \theta) \mid \rho = i' \cos \theta + j' \sin \theta \text{ and } -20^\circ \leq \theta \leq 20^\circ\}. \end{aligned} \quad (9)$$

The new voting rule then is expressed as follows.

If $P_{i_0, j_0} = 1$, then the operation of $C(\rho, \theta) = C(\rho, \theta) + 1$ is executed for all elements in A .

If $P_{i', j'} = 1$, then the operations of $C(\rho, \theta) = C(\rho, \theta) - 1$ are executed for all elements in B , and at the same time the operation of $C(\rho, \theta) = C(\rho, \theta) + 1$ is executed for all elements in A .

3.3. Precision and Efficiency of IPPHT Compared to Those in PPHT. Figures 6(a) and 6(b) are two edge images processed with the Canny edge detection algorithm for the lift rope of a tower crane. Figures 6(c) and 6(d) are the detection results obtained from the images in Figures 6(a) and 6(b), respectively, using the PPHT. In Figure 6(c), the edge of the rope is a nonsingle-pixel edge and is regarded as two parallel lines. In Figure 6(d), only parts of the rope are detected due to the slight curve of the rope. Figures 6(e) and 6(f) are the detection results obtained from the images in Figures 6(a) and 6(b), respectively, using the IPPHT. The test conditions used for the PPHT and IPPHT are the same. The width of search section in the IPPHT is 5 pixels. The detection results presented in Figures 6(e) and 6(f) are more accurate than those presented in Figures 6(c) and 6(d). An added line segment generated by the steel structure is also detected, but it can be excluded by defining a valid image zone. The rope should appear at the center of the image along the x -direction due to the relative position of the camera and the rope; thus, only the line nearest to the center of the image is selected as the rope image. All other lines should be excluded.

The algorithms were implemented with an S3C2440 processor-based ARM system at 400 MHz using a Linux operation system. The computation times for images shown in Figures 6(c) and 6(d) were 2998 ms and 3001 ms, respectively. The computation times for the images shown in Figures 6(e) and 6(f) were 417 ms and 617 ms, respectively. Therefore, the IPPHT algorithm is suitable for real-time detection and monitoring.

3.4. Implementation of Hook Autotracking Based on IPPHT. The tracking process of the hook is implemented in the following steps:

- (1) detect the rope with the IPPHT;
- (2) to search the end point of the line segment downward along the line, take the end point as the starting

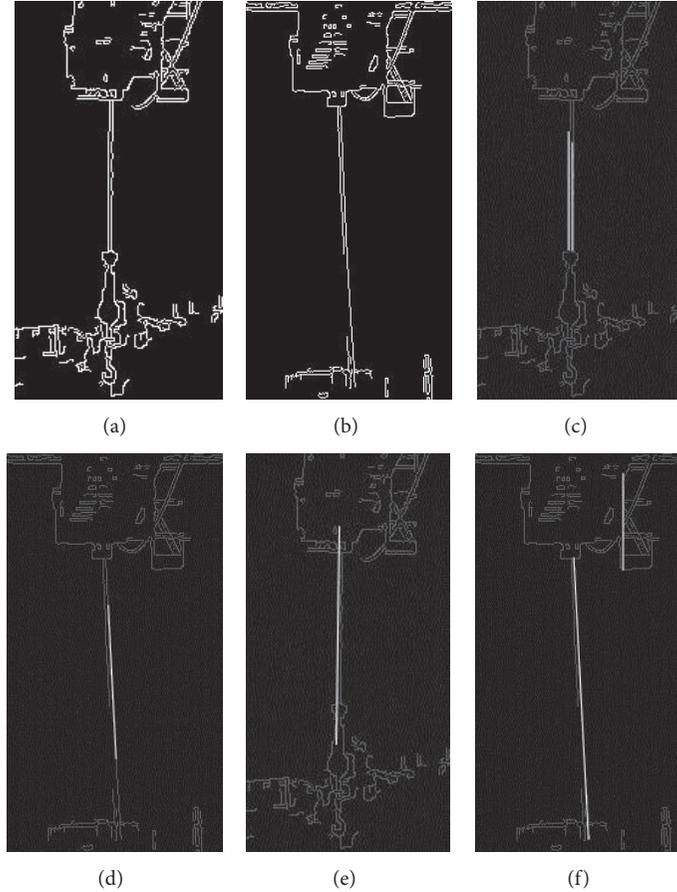


FIGURE 6: Detection results using both PPHT and IPPHT.

position of the hook, and then save the coordinates of the end point;

- (3) sample the hook position periodically in intervals of 1 s; then, adjust the pitch angle of the camera according to the change in the hook position using PID algorithm.

4. Adaptive Zooming for Precision Hoisting

The size of the hook that appears on the monitor should be adjusted to satisfy the requirement for precision hoisting. Object recognition must be performed before the hook size to be adjusted appears on the monitor. The shape of the hook system is intricate; moreover, its image shape on the monitor varies with the rotation and translation of the hook. Hook systems are often painted with striking colors, typically a combination of yellow and black, black and red, or yellow and red, to provide warning. Therefore, color is a suitable feature for hook system image retrieval. Color histograms have been used extensively for image retrieval and recognition due to their rotation- and translation-invariant properties. To improve the retrieval speed, a hook image zooming algorithm combining color histograms with a binary search was proposed.

Figure 7 illustrates the color histogram developed using a binary search. $P_0(x_0, y_0)$ is the end point of the lift rope, and A is the hook system image zone. B is a block used to determine the size of the hook system, $l = L/4$, and L is the width of the vision field. To eliminate the effects of an irregular rope shape, let $\Delta y = 10$ pixels and $h = 20$ pixels.

Suppose R_{ij} , G_{ij} , and B_{ij} are the R, G, and B component values of pixel point $P(i, j)$ in RGB color space respectively. Define the maximum and minimum component values of R_{ij} , G_{ij} , B_{ij} as M and m [18]:

$$\begin{aligned} M_{ij} &= \max(R_{ij}, G_{ij}, B_{ij}), \\ m_{ij} &= \min(R_{ij}, G_{ij}, B_{ij}), \\ C_{ij} &= M_{ij} - m_{ij}, \\ V_{ij} &= M_{ij}. \end{aligned} \quad (10)$$

Define

$$H_{ij}(\alpha, \beta) = \begin{cases} 1 & \text{if } H_{ij} \in [\alpha, \beta], \\ 0 & \text{if } H_{ij} \notin [\alpha, \beta]. \end{cases} \quad (11)$$

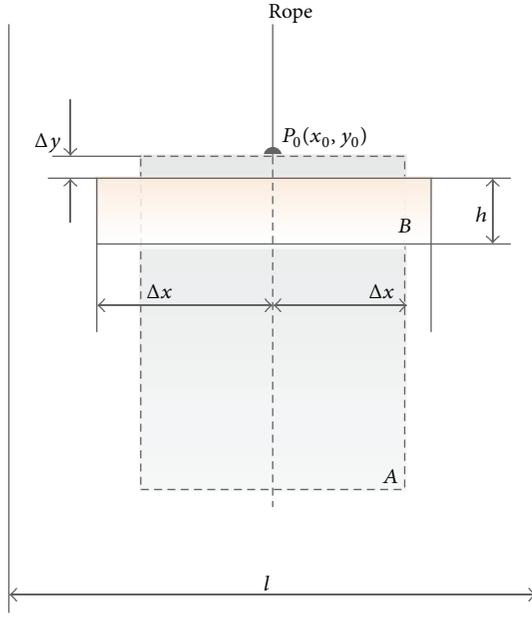


FIGURE 7: An illustration of color histogram based-binary search.



FIGURE 8: An instance of zooming result.

In (11) H_{ij} is the value of the component H in HSV color space at pixel point $P(i, j)$. H_{ij} is transformed from RGB color space [19]. Consider

$$H_{ij} = \begin{cases} \arccos \frac{(R_{ij} - G_{ij}) + (R_{ij} - B_{ij})}{2\sqrt{(R_{ij} - G_{ij})^2 + (R_{ij} - B_{ij})(G_{ij} - B_{ij})}}, & B_{ij} \leq G_{ij}, \\ 360 - \arccos \frac{(R_{ij} - G_{ij}) + (R_{ij} - B_{ij})}{2\sqrt{(R_{ij} - G_{ij})^2 + (R_{ij} - B_{ij})(G_{ij} - B_{ij})}}, & B_{ij} > G_{ij}, \end{cases} \quad (12)$$

Define

$$p = \sum_{j=y_0+\Delta y}^{y_0+\Delta y+h} \sum_{i=x_0-\Delta x}^{x_0+\Delta x} [a_0(i, j) + a_1(i, j) + a_2(i, j) + a_3(i, j)], \quad (13)$$

$$p_1 = \frac{p}{2\Delta x \cdot h}.$$

In (13),

$$a_0(i, j) = \begin{cases} 1 & \text{if } C_{ij} = 0 \text{ and } V_{ij} < 34 \\ 0 & \text{if } C_{ij} \neq 0 \text{ and } V_{ij} \geq 34, \end{cases}$$

$$a_1(i, j) = H_{ij}(55, 65), \quad (14)$$

$$a_2(i, j) = H_{ij}(115, 125),$$

$$a_3(i, j) = H_{ij}(100, 120).$$

The color of point $P(i, j)$ is determined as follows:

- if $a_0(i, j) = 1$, then the color of point $P(i, j)$ is black;
- if $a_1(i, j) = 1$, then the color of point $P(i, j)$ is tangerine;
- if $a_2(i, j) = 1$, then the color of point $P(i, j)$ is red;
- if $a_3(i, j) = 1$, then the color of point $P(i, j)$ is yellow.

The proportions of black, tangerine, red, and yellow colors in the zone are calculated using (13), and a binary search of the image zone of the hook system is then executed as follows:

- if $p_1 < 0.85$, let $\Delta x = \Delta x/2$, and then recalculate p and p_1 as indicated in (13);
- if $p_1 > 0.9$, let $\Delta x = 2\Delta x$, and then recalculate p and p_1 as indicated in (13);
- if $0.85 < p_1 < 0.9$, then end the search, and take $2\Delta x$ as the width of the hook system image.

Once the width of the hook system image in the current vision field is obtained, the camera is zoomed to make sure that $2\Delta x$ is one-tenth to one-eighth of the width of the vision field in the monitor. Figure 8 shows an instance of the width of the hook system being zoomed to one-eighth of the vision field, for which only four iterations are executed.

5. Conclusion

Automatic monitoring and controlling of load hoisting can not only reduce the severity and number of tower crane accidents but also improve operation quality and efficiency of tower crane. Using the lift rope of a tower crane as the object

to be captured, an IPPHT-based hook tracking algorithm was proposed and analyzed. The hook can be tracked quickly and accurately. Greatly reducing the computation complexity and storage requirements allows the algorithm to excel in ARM systems, which are widely adopted in the field control and monitoring of tower cranes. The algorithm can reduce the capture time to almost 80% that of the PPHT algorithm. A zooming algorithm in which color histograms are combined with a binary search algorithm was also proposed and analyzed to adjust the hook system to a suitable size on display monitors for precision and safe lifting. Only relevant colors are calculated, and the optimum zooming scale can be achieved in only a few iterations. The validity of the algorithms was tested.

Acknowledgments

This work was supported by the National High Technology Research and Development Program of China (Grant no. 2012AA041804), the National Natural Science Foundation of China (Grant no. 51275290), a Research Project of the State Key Laboratory of Mechanical System and Vibration MSV201206, and a Research Project of Shanghai Science and Technology Committee (11dz1121502).

References

- [1] V. W. Y. Tam and I. W. H. Fung, "Tower crane safety in the construction industry: a Hong Kong study," *Safety Science*, vol. 49, no. 2, pp. 208–215, 2011.
- [2] L. Y. Yuan, G. Q. Zhang, and H. B. Gao, "Implementation of a Multifunction Black-box for Tower Crane Based on ARM and μ COS-II," in *Proceedings of the 7th World Congress on Intelligent Control and Automation*, pp. 7681–7685, Chongqing, China, June 2008.
- [3] M. Abderrahim, E. Garcia, R. Diez, and C. Balaguer, "A mechatronics security system for the construction site," *Automation in Construction*, vol. 14, no. 4, pp. 460–466, 2005.
- [4] U.-K. Lee, J.-H. Kim, H. Cho, and K.-I. Kang, "Development of a mobile safety monitoring system for construction sites," *Automation in Construction*, vol. 18, no. 3, pp. 258–264, 2009.
- [5] R. Sacks, R. Navon, I. Brodetskaia, and A. Shapira, "Feasibility of automated monitoring of lifting equipment in support of project control," *Journal of Construction Engineering and Management*, vol. 131, no. 5, pp. 604–614, 2005.
- [6] A. Shapira and B. Lyachin, "Identification and analysis of factors affecting safety on construction sites with tower cranes," *Journal of Construction Engineering and Management*, vol. 135, no. 1, pp. 24–33, 2009.
- [7] A. Shapira, Y. Rosenfeld, and I. Mizrahi, "Vision system for tower cranes," *Journal of Construction Engineering and Management*, vol. 134, no. 5, pp. 320–332, 2008.
- [8] J. Yang, P. A. Vela, J. Teizer, and Z. K. Shi, "Vision-based crane tracking for understanding construction activity," in *Proceedings of the ASCE International Workshop on Computing in Civil Engineering*, pp. 258–265, June 2011.
- [9] S. Nara, D. Miyamoto, S. Takahashi, and S. Kaneko, "Visual feedback tracking of crane hook," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '06)*, pp. 2736–2742, chn, October 2006.
- [10] X. Liu, H. Chu, and P. Li, "Research of the improved camshift tracking algorithm," in *Proceedings of IEEE International Conference on Mechatronics and Automation (ICMA '07)*, pp. 968–972, Harbin, China, August 2007.
- [11] A.-Y. Wang and J. Yang, "Research on tower crane automatic tracking and zooming video monitoring system based on Improved Camshift algorithm," in *Proceedings of International Conference on Electric Information and Control Engineering (ICEICE '11)*, pp. 4434–4437, April 2011.
- [12] S. Takahashi and S. Kaneko, "Motion tracking of crane hook based on optical flow and orientation code matching," in *Proceedings of the 10th International Workshop on Advanced Motion Control (AMC '08)*, pp. 149–152, March 2008.
- [13] Y. Li and C. Liu, "Integrating field data and 3D simulation for tower crane activity monitoring," *Automation in Construction*, vol. 27, pp. 111–119, 2012.
- [14] M. Al-Hussein, M. Athar Niaz, H. Yu, and H. Kim, "Integrating 3D visualization and simulation for tower crane operations on construction sites," *Automation in Construction*, vol. 15, no. 5, pp. 554–562, 2006.
- [15] T. Huang, C. W. Kong, H. L. Guo, A. Baldwin, and H. Li, "A virtual prototyping system for simulating construction processes," *Automation in Construction*, vol. 16, no. 5, pp. 576–585, 2007.
- [16] J. Song and M. R. Lyu, "A Hough transform based line recognition method utilizing both parameter space and image space," *Pattern Recognition*, vol. 38, no. 4, pp. 539–552, 2005.
- [17] J. Matas, C. Galambos, and J. Kittler, "Robust detection of lines using the progressive probabilistic hough transform," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 119–137, 2000.
- [18] S. Jeong, C. S. Won, and R. M. Gray, "Image retrieval using color histograms generated by Gauss mixture vector quantization," *Computer Vision and Image Understanding*, vol. 94, no. 1–3, pp. 44–66, 2004.
- [19] R. Chakravarti and X. Meng, "A study of color histogram based image retrieval," in *Proceedings of the 6th International Conference on Information Technology: New Generations (ITNG '09)*, pp. 1323–1328, April 2009.

Research Article

EMD and Wavelet Transform Based Fault Diagnosis for Wind Turbine Gear Box

Qingyu Yang^{1,2} and Dou An²

¹ State Key Laboratory for Manufacturing System Engineering, Xi'an Jiaotong University, Xi'an 710049, China

² School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Correspondence should be addressed to Qingyu Yang; yangqingyu@mail.xjtu.edu.cn

Received 28 May 2013; Accepted 17 July 2013

Academic Editor: Cheng-liang Liu

Copyright © 2013 Q. Yang and D. An. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Wind turbines are mainly located in harsh environment, and the maintenance is therefore very difficult. The wind turbine faults are mostly from the gear box, and the fault signal is generally nonlinear and nonstationary. The traditional fault diagnosis methods such as Fast Fourier transform (FFT) and the inverted frequency spectrum identification method based on FFT are not satisfactory in processing this kind of signal. This paper proposes a hybrid fault diagnosis method which combines the empirical mode decomposition (EMD) and wavelet transform. The vibration signal is analyzed through wavelet transform, and the aliasing in high-frequency signals is then addressed by conducting EMD to the original signals. The experimental results based on a specific wind turbine gear box demonstrate that this method can diagnose the faults and locate their positions accurately.

1. Introduction

With the increasing integration of wind energy in power grid, wind turbine failure has become a significant issue. It has great impact on the whole power system and may cause great economic losses. Wind turbines are mostly located in harsh environment such as mountain areas or hilly regions, and the maintenance of the turbines is very difficult. As a result, the detection of early fault symptoms is of particular importance. Gear box is the core component in wind turbine and also the part with a high failure probability. Failure of gear box will affect the entire wind turbine. For example, the faulty part may impact other elements while rotating, and the failure information will be reflected in the gear box vibration signal [1]. The vibration signal is thus usually used for the diagnosis of gear box faults in wind turbines.

The popular techniques for analyzing the vibration signal include Fast Fourier transform (FFT), inverted spectrum, and signal modulation and demodulation. These methods have the advantages of high speed and relatively satisfactory accuracy and are often used in the actual fault diagnosis. However, in gear box failures, the vibration signals returning from the sensors tend to be nonlinear and nonstationary. The classical spectrum analysis methods have some obvious

disadvantages in processing these time-varying vibration signals, such as weak signal submerged by the side lobe of strong signal, low-frequency resolution, and relatively serious leakage of spectral side lobes [2]. The diagnostic results are therefore often undesirable.

Currently, wavelet transform and empirical mode decomposition (EMD) time-frequency analysis are commonly used for the processing of non-stationary signals. Wavelet transform can map any signal to a set of base functions obtained through dilation and translation of a mother wavelet, and it is able to achieve a reasonable decomposition of the signals in different frequency bands and at different time points. However, complex aliasing may exist in the high-frequency portion [3]. EMD, as a signal processing method for nonlinear and non-stationary signals, has a better performance than the conventional methods, but it has difficulty in isolating the signals within the second harmonic.

To address these problems, this paper proposes a hybrid fault diagnosis method combining the EMD and wavelet transform. First, the number of decomposition layers is determined according to the vibration characteristic frequencies of the gear box components, and the vibration signals will be decomposed through wavelet transform. Secondly, the fault information of different components is separated

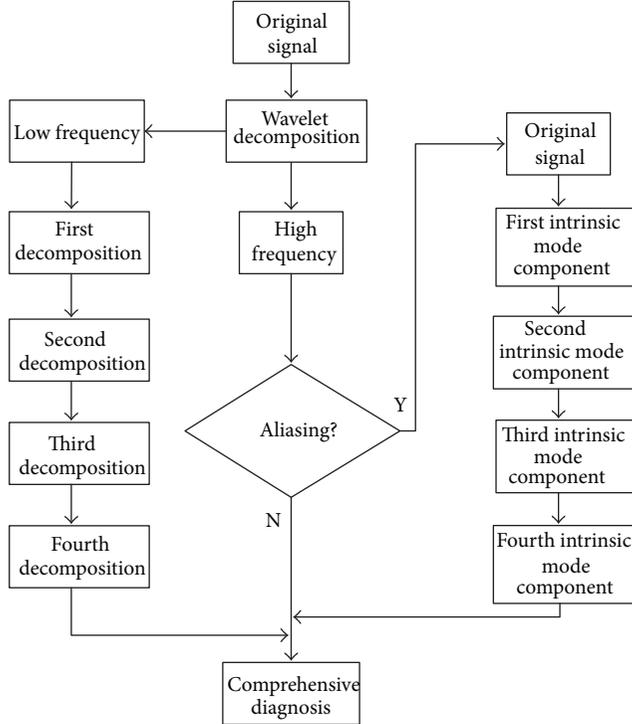


FIGURE 1: Flowchart of fault diagnosis process using the hybrid diagnosis method.

into different wavelet transform results, and the signals are then divided into high-frequency and low-frequency parts. For the low-frequency signals, the frequency bands are allocated purposely through wavelet transform. For the high-frequency signals, if aliasing appears, which indicates that the high-frequency energy in the vibration signals affects the use of wavelet transform, EMD is then used to decompose the original signals and analyze the first few decomposed intrinsic mode components. Finally, the fault conditions of the corresponding components can be determined according to their energy magnitudes.

2. Fault Diagnosis Based on EMD and Wavelet Transform

2.1. Empirical Mode Decomposition Method. Empirical Mode Decomposition (EMD), proposed by Huang et al. in 1998, is a new method for decomposing any signal into the composition of intrinsic mode components (also called IMF) based on an in-depth study of the instantaneous frequency of the signal [4]. The method considers that all the signals are made up of a series of intrinsic mode components, and each component contains part of the characteristics of the original signal. The biggest difference between EMD and wavelet transform is that wavelet transform needs to preset a specific wavelet basis function, and the decomposition depends on the basis function, whereas EMD obtains an adaptive basis function which varies in terms of different signals. At this point, EMD is an innovation for basis function [5–12].

After EMD, each intrinsic mode components satisfies two conditions: (1) in the entire data sequence, the number of poles (including the maximum and minimum points) and that of zeros must be equal or differ less one; and (2) at any time point, the mean of the upper envelope curve determined by the local maxima of the signal and that of the lower envelope curve determined by the local minima are both zero. For any original signal $x(t)$, the EMD process is as follows.

Step 1. Determine all the local extreme points of $x(t)$, and connect the maxima and minima with a cubic spline curve to form the upper and lower envelopes of the signal.

Step 2. Subtract the mean value $m(t)$ of the upper and lower envelopes from $x(t)$: $h_1(t) = x(t) - m(t)$, and examine whether $h_1(t)$ satisfies the condition of intrinsic mode component. If not, repeat the above operation until $h_1(t)$ is an intrinsic mode component, and it is denoted as $c_1(t) = h_1(t)$.

Step 3. Decompose $c_1(t)$ from $x(t)$, and obtain the remaining sequence $r_1(t)$: $r_1(t) = x(t) - c_1(t)$.

Step 4. Take $r_1(t)$ as a new “original” signal, and repeat the above operation to obtain successively the second, the third, until the n th intrinsic mode component, denoted as $c_1(t), c_2(t), \dots, c_n(t)$. The decomposition process ends until the preset stop criteria are met. Finally, the remainder of the original signal, $r_n(t)$, is left.

Through the above steps, the original signal is decomposed into a sum of n intrinsic mode components and a reminder:

$$x(t) = \sum_{i=1}^n c_i(t) + r_n(t). \quad (1)$$

EMD stipulates two theoretical stop criteria: (1) when the last intrinsic mode component or the remaining component is lower than the expected value; and (2) when the remaining component becomes a monotonic function, from which no intrinsic mode component can be decomposed.

In the actual screening process, it is difficult to guarantee the local mean value of the signal to be zero. Thus, the screening can be stopped by limiting the standard deviation between two successive processing results. The standard deviation is calculated as follows:

$$S_d = \sum_{t=0}^T \frac{|(h_{k-1}(t) - h_k(t))|^2}{h_k^2(t)}, \quad (2)$$

where T is the time span of the signal, $h_{k-1}(t)$ and $h_k(t)$ are two consecutive sequences in the process of screening intrinsic mode components, and the value of S_d is often controlled within the range of [0.2, 0.3].

2.2. Wavelet Transform Method. Wavelet transform [13] is essentially to decompose signal $x(t)$ into subsignals with different frequency bands through basis function $\varphi((t-b)/a)$, in which a is a scale factor that controls contraction and

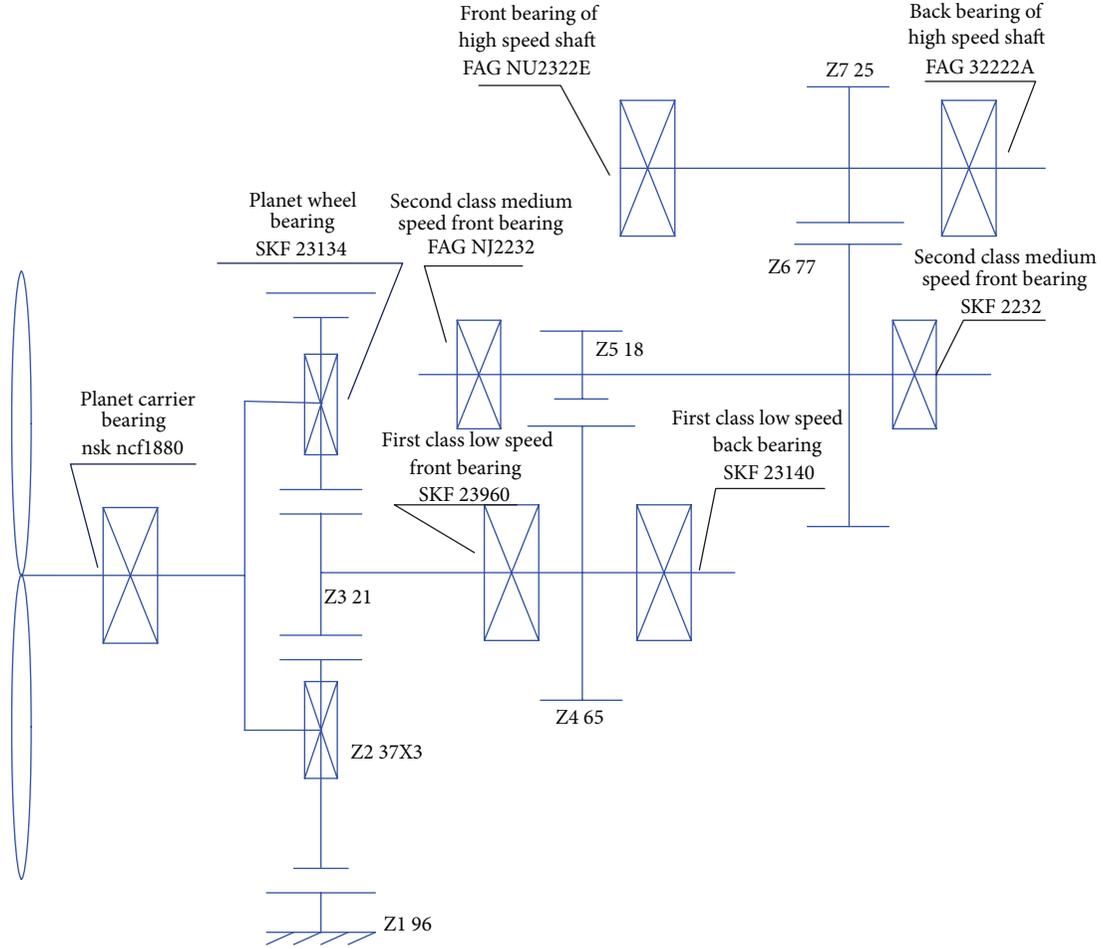


FIGURE 2: Structure of the gear box in the experiment.

stretch of the waveform, and b denotes the time-shift factor. The wavelet is actually a family of functions generated by the dilation and translation of basic wavelet or mother wavelet [11, 12]. The wavelet transform of signal $x(t)$ can be denoted as

$$WT_x(b, a) = a^{-1/2} \int_{-\infty}^{\infty} x(t) \varphi^* \left(\frac{t-b}{a} \right) dt. \quad (3)$$

Note that the wavelet transform of a signal is equivalent to observe the signal through the changes of wavelet scale factor and time-shift factor. The width of the wavelet function decreases, and the bandwidth increases when a decreases; otherwise, the width increases, and the bandwidth decreases.

The one-dimensional wavelet transform is usually conducted through Mallat pyramid decomposition as follows:

$$a_{j,k} = \sum_n h_{n-2k} a_{j-1,k}, \quad d_{j,k} = \sum_n g_{n-2k} a_{j-1,k} \quad (4)$$

$$j = 1, 2, \dots, J,$$

where a is the smoothed signal of the original signal, d is the detail signal, g is the impulse response of the band pass filter associated with the wavelet function, and h is the impulse response of the low-pass filter associated with the scaling function.

2.3. The Proposed Hybrid Diagnosis Method

2.3.1. Characteristics of Gear Box Vibration Signals. Wind turbine gear box raises the low speed of the wind wheel (typically 14–48 rpm) to a high generator speed (typically more than 1500 rpm), and the drive ratio is generally very high. For example, the drive ratio of gamesa 850 type wind turbine gear box is 62. To achieve this goal, multilevel drive is usually adopted, and planet wheel drive is used at the low-speed shaft end. The advantages of planet wheel drive are small size, light weight, and bigger drive ratio compared with that of parallel shaft drive. The vibration of gear box is mainly caused by the engagement of the gears. In the meshing of a pair of gears, the number of meshing teeth of a pair of gears changes from one to two and then back to one. The repetition of this forms an alternation of single-tooth and double-teeth meshing. Each alternation gives an impact to the gear, thus forming the gear mesh vibration. The mesh frequency and its harmonic components are usually denoted as follows:

$$x(t) = \sum_{m=1}^M X_m \cos(2\pi m f_z t + \phi_m), \quad (5)$$

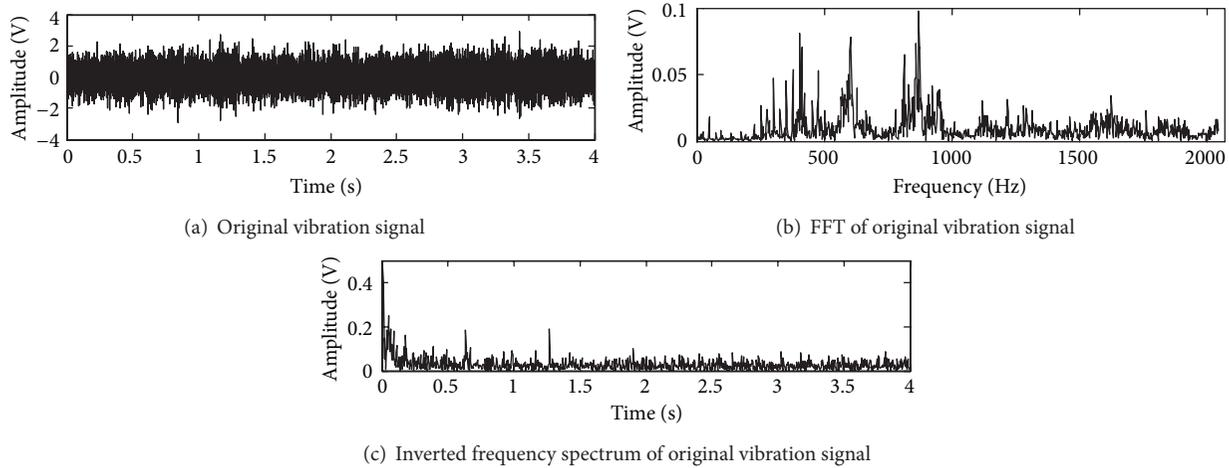


FIGURE 3: FFT and inverted frequency spectrum of original vibration signal.

where X_m is the m th order mesh frequency, ϕ_m is the initial phase of the harmonic components of X_m , and f_z is the gear mesh frequency.

Gears generate mesh vibration in both normal and abnormal circumstances, but the vibration levels are different. Hence, it is feasible to diagnose faults by using the gear mesh frequency and harmonic components of the vibration signal. It has been found that the spectrum of the vibration signal usually contains not only the gear mesh frequency and its harmonics but also peaks on both sides, which are caused by signal modulation [14, 15].

In the process of gear movement, frequency and amplitude modulation are generated at the same time. They have the same carrier frequency, their sideband frequencies are correspondingly equal, and their sidebands are symmetrical to carrier frequency. However, when they occur simultaneously, the superimposition may cause increase of some side-frequency amplitudes while decreases of some others due to the different phases of their edge frequencies. This is why the common spectrum of the signal does not always have symmetrical sidebands.

Apart from gear mesh vibration, bearing vibration is also a part of gear box vibration. The characteristic frequency range of rolling bearing vibration signal is generally higher than that of gear vibration signal, whereas the characteristic frequency of sliding bearing vibration signal is generally lower. Therefore, the fault information of different parts in the gear box distributes in different frequency bands.

2.3.2. Diagnosis Principle Using Hybrid Method. As mentioned in Section 2.2, wavelet transform of signal is decomposing the low-frequency approximation signal, and it does not decompose the high-frequency detail signal. This actually divides the frequency band of the signal at certain exponent intervals. The wavelet transform of high-frequency band signal achieves high time resolution and low-frequency resolution, while that of low-frequency band signal is on the contrary. The wavelet transform of the signal is equivalent

with using a low-pass filter and a plurality of band-pass filters without losing any original information. During gear box operation, the frequencies of each shaft and gear mesh are usually not the same, thus the spectrum of gear box vibration signal contains low-frequency and high-frequency components. Wavelet transform decomposes the original signal into a number of mutually independent bands. In another word, it divides the frequency information of different parts into the corresponding frequency bands, so that each of these bands carries the state information of a specific part. In addition, the orthogonality of the wavelet function ensures no redundancy and no omission of the state information, which is helpful in eliminating interference and accurately locating faults. In general, for the fault diagnosis using wavelet transform, the vibration frequency distribution of various gear box parts should be considered, and appropriate number of decomposition layers should be determined to make sure that the different layers of decomposition results contain the fault information of different gear box parts.

For the decomposition of high-frequency signals, frequency aliasing often exists because the sampling rate does not satisfy the sampling theorem. This may lead to fault frequency information and impact fault diagnosis. EMD is to decompose a signal into a series of intrinsic mode components and a remainder, and each decomposition is a redecomposition of the last remainder. The frequency range represented by the series of intrinsic mode components is not fixed. In the process of extracting intrinsic mode components, the part with relatively higher frequency is extracted every time, and the extreme points on each mean curve are getting fewer. The intrinsic mode component and the remainder on each layer do not necessarily split the entire signal bandwidth. The intrinsic mode components, $c_1(t)$, $c_2(t)$, \dots , $c_n(t)$, contain the components of the original signal at different bands from high frequency to low frequency, and each component contains certain inherent characteristics of the original signal. In general, the first few components decomposed by EMD contain the key information of the original signal, while the last few components represent the average trend of the signal.

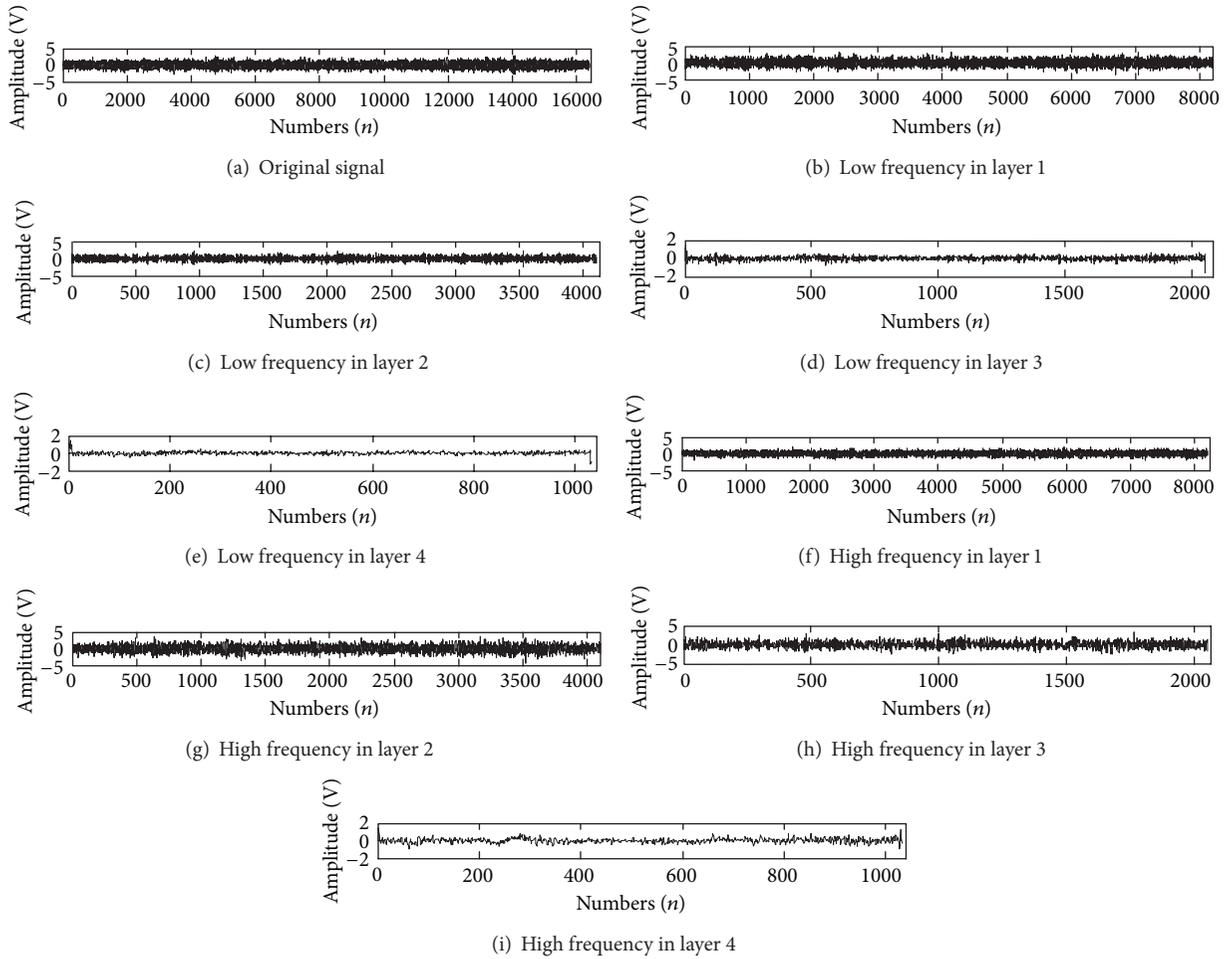


FIGURE 4: Time domain waveform and wavelet transform of vibration signal.

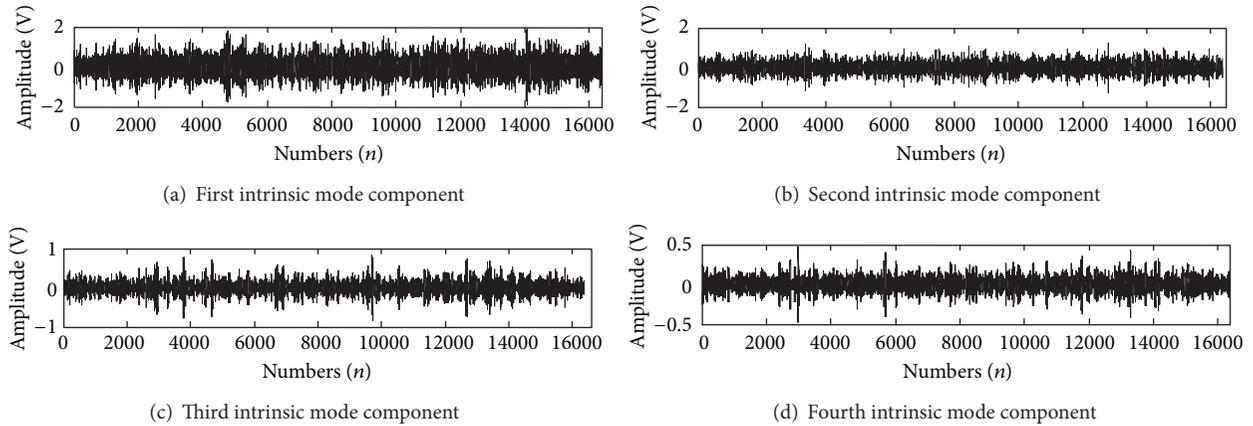


FIGURE 5: Four intrinsic mode components after EMD.

Hence, the first few intrinsic mode components are usually selected to be analyzed through EMD [16].

2.4. Workflow of the Hybrid Diagnosis Method. Figure 1 gives the workflow of the proposed hybrid diagnosis method. The first step is to determine the appropriate number of layers of

wavelet transform according to the vibration characteristic frequency of each element in the gear box. Conduct wavelet transforms to the original signal, dividing fault information into different layers. The decomposed signals are separated into low-frequency and high-frequency parts at each layer. The low-frequency parts can be analyzed directly for fault

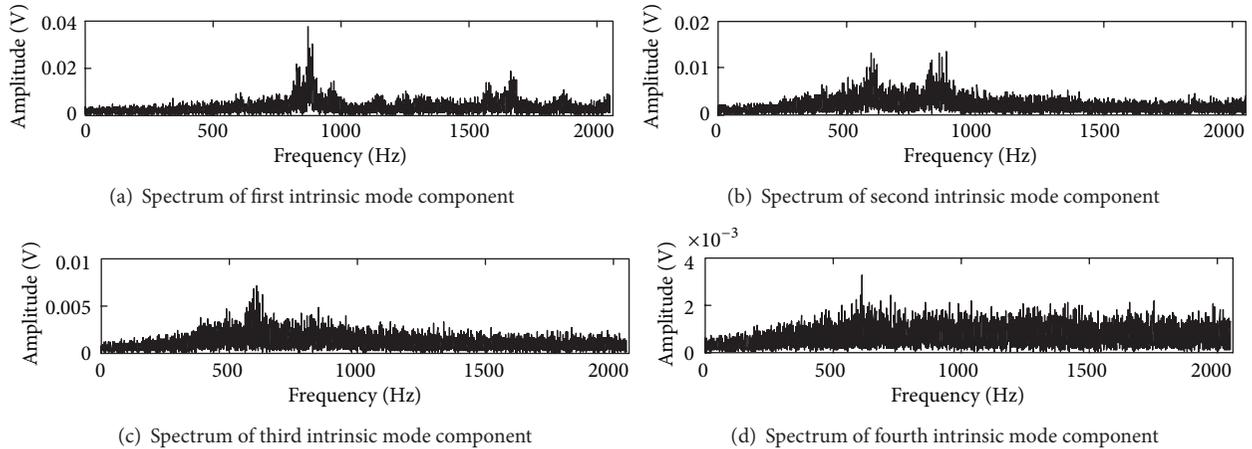


FIGURE 6: Spectrum of four intrinsic mode components.

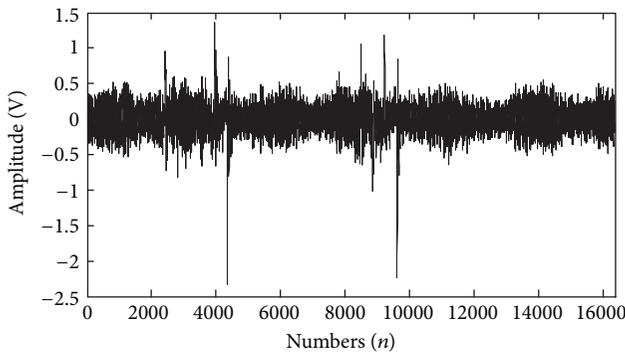


FIGURE 7: Regular vibration signal.

diagnosis. For the high-frequency parts, the signals can be analyzed directly if there is no aliasing. Whereas, the existence of aliasing suggests that the wavelet transform results are influenced by the energy of the high-frequency part, and EMD should be used to decompose the original signal into a series of intrinsic mode components. The first several intrinsic components that contain the key information of the signal are selected. Finally, the comprehensive fault diagnosis of the gear box is performed based on the wavelet transform results and first few intrinsic mode components.

3. Fault Diagnosis Instance and Performance Evaluation

An experiment to a specific wind turbine gear box with three-stage drive was conducted to evaluate the performance of the hybrid fault diagnosis method. Figure 2 illustrates the structure of the gear box in the experiment. The total drive ratio is 61.966. The first stage is a planetary gear drive. The shaft drive ratios from low to high are 5.576, 3.612, and 3.077, respectively. The rated input speed is 26.17 rpm. The vibration data comes from the sensor at the axial direction of gear box output. At the rated speed, the rotation frequency of the high-speed shaft is 27 Hz, and the gear mesh frequency is 675 Hz; the rotation frequency of the medium-speed shaft is 8.7 Hz, and the mesh

frequency of its front axle gear is 158 Hz; the mesh frequency of the front axle gear of the low-speed shaft is 51 Hz. In the process of signal sampling, the input speed is 28.91 rpm. The rotation frequency of high-speed shaft is 29.8 Hz, and the gear mesh frequency is 746 Hz; the rotation frequency of the medium-speed shaft is 9.7 Hz, and the mesh frequency of its front axle gear is 174 Hz; the mesh frequency of the front axle gear of the low-speed shaft is 56.4 Hz, and the rotation frequency of planet shaft is 0.49 Hz. The wavelet and EMD toolkits in MATLAB platform are used to conduct the simulation.

3.1. Fault Diagnosis with FFT and Inverted Frequency Spectrum. Before evaluating the performance of the proposed method, fault diagnosis based on the traditional spectrum analysis methods: FFT and inverted frequency spectrum have been conducted. Figures 3(a), 3(b), and 3(c) show the original vibration signal, FFT, and inverted frequency spectrum results of original signal, respectively.

FFT is an algorithm to compute the discrete Fourier transform and its inverse, which converts time domain to frequency domain. The frequency components of the original signal are reflected in the frequency domain. Notice that in Figure 3(b), mesh frequency of each shaft is mostly submerged by the noises and the side lobes; the sideband is unable to distinguish. Hence, the faults in the gear box cannot be diagnosed by only observing FFT of the original signal.

Inverted frequency spectrum, also called cepstrum, is the result of taking the inverse Fourier transform of the logarithm of a signal estimated spectrum. Inverted frequency spectrum converts the periodic signal and sidebands in FFT results to spectral lines, thus making it easier to detect the complex periodic component of the spectrum. In Figure 3(c), the frequencies of the first and second peak are 2 Hz and 0.82 Hz respectively, which are not the rotation frequency of any shafts of the gear box. Therefore, the location and the specific form of the faults cannot be diagnosed by observing the inverted frequency spectrum either.

3.2. Fault Diagnosis with Proposed Hybrid Diagnosis Method. Figure 4 presents the time-domain waveform and DB5

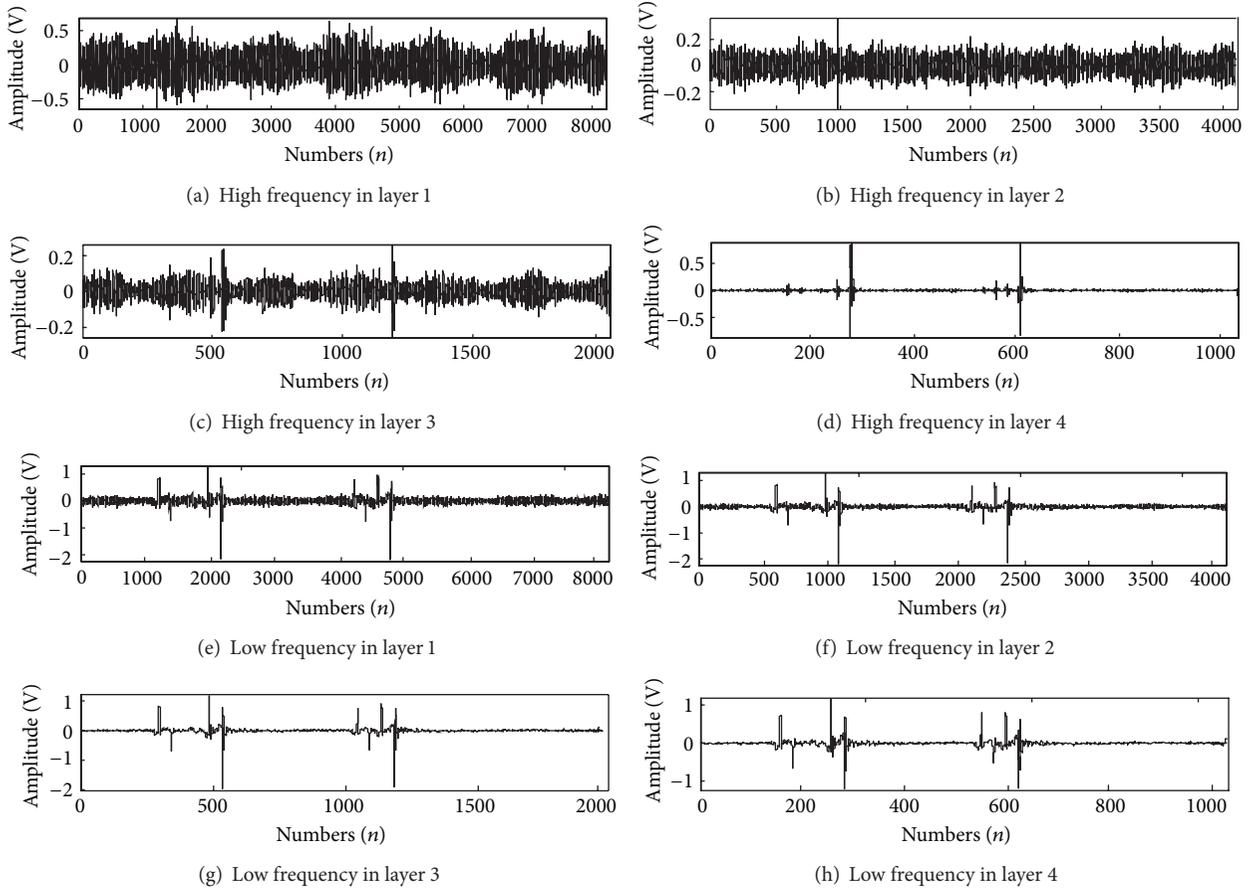


FIGURE 8: Wavelet transform of regular vibration signal.

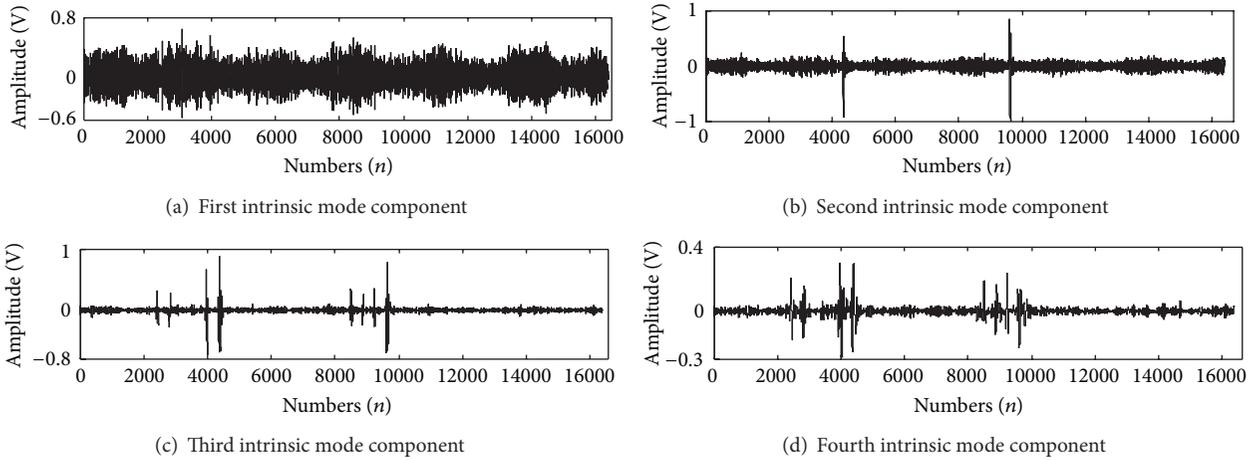


FIGURE 9: Four intrinsic mode components of EMD for regular vibration signal.

wavelet 4-layer decomposition results of the vibration signal. The sampling frequency is 4096 Hz. The number of sampling points is 16384, which is halved after wavelet transform. The signal is decomposed into low-frequency and high-frequency parts. The waveform is cluttered, and the vibration amplitude is large. Figures 4(d)-4(e) shows that the low-frequency part contains several relatively large impacts of

time-domain diagram of original signal. Figures 4(f)-4(i) show that the waveform of the high-frequency part is filled with random pulses. From the above characteristics, we can initially determine that the axial clearance of the planet carrier bearing is greater than normal levels. The excessive axial clearance of planet carrier bearing leads to the mesh collision of the gears and collision of box. Moreover, recall

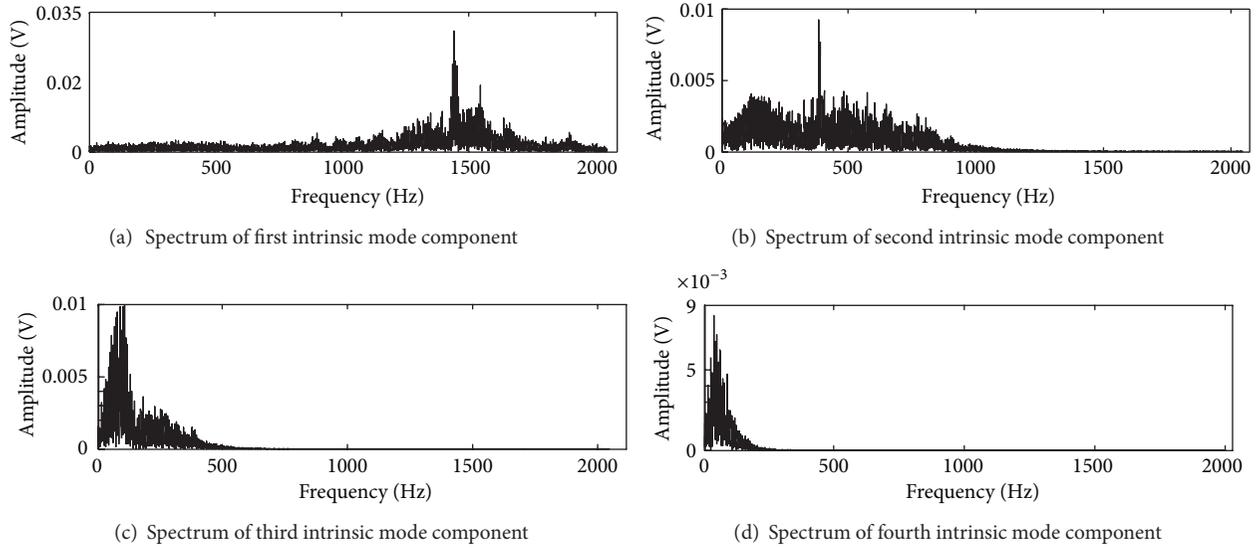


FIGURE 10: Spectrum of four intrinsic mode components.

that the rotation speed of planet shaft is the slowest in the gear box; thus it will generate low-frequency signals and the high harmonics of the signals when rotating. As a result of the excessive axial clearance of planet carrier bearing, the vibration signals will be largely covered by their high harmonics and look like a thick line, which are clearly reflected in Figures 4(a)–4(i). Consistent monotonicity of wavelet transform indicates that the high-frequency parts of all layers and the low-frequency part of the last layer compose the entire signal. The low-frequency parts of the other three layers exist as intermediate amount of decomposition results. Consistent monotonicity and asymptotic completeness of wavelet transform can be clearly seen from Figure 4. In addition, although the results of wavelet transform express frequency domain information, the data itself is from time domain. Each decomposed data in Figure 4 gives a different number of points, but the length of time is the same with the original signal. To further analyze the vibration signal, EMD is implemented to the original signal [17]. Figure 5 shows the first four intrinsic mode components, and Figure 6 gives the spectrum of each intrinsic mode component.

Notice that the component with the maximum energy in Figure 6(a) is at 870 Hz, which is the quintuple of the mesh frequency of front axle gear of medium-speed shaft 174 Hz. Meanwhile, equally spaced sidebands appear in both sides of 870 Hz, and the interval is 9.7 Hz, which is the rotation frequency of medium-speed shaft. Furthermore, the amplitude of high harmonics near 870 Hz has little difference with that of 870 Hz. Based on the above analysis, we can diagnose that the front axle of medium-speed shaft abrades seriously. Also, there exists energy peak near 600 Hz in Figures 6(b) and 6(c), which is a high-level component of the low-speed shaft gear mesh frequency. We can see clearly from Figures 6(c) and 6(d) that the signal energy distributes evenly especially in the band greater than 600 Hz, which is the high-frequency interference signals cause by vibration of gear box.

This indicates that the fault is caused by large axial clearance of gear box planet carrier bearing [18].

The gear box was opened for examination, and the faults were found to be located in planet carrier bearing and front axle of the medium-speed shaft. The axial clearance of planet carrier bearing was too large, and the gear was seriously abraded. These are consistent with the diagnostic results obtained from the proposed hybrid diagnosis method. Figure 7 illustrates the regular signal after replacement of the gear and adjustment of the bearing. Then, wavelet transform and EMD are also applied to the regular vibration signal. Figures 8, 9, and 10 show the results of wavelet transform, EMD, and spectrum of EMD, respectively. Notice that the amplitude of all the figures in Figure 8 is far lower than that of Figure 4, especially in Figures 8(c)–8(h). Figures 8(a)–8(c) present normal and cyclical fluctuations of the signals. Figure 9 illustrates the first four intrinsic mode components of EMD results, the characteristics of which are similar with that of wavelet transform. In Figure 10, the amplitude is low, the noise is relatively small, and no obvious characteristic component exists.

4. Conclusion

In this paper, a novel hybrid fault diagnosis method combining wavelet transform and EMD is proposed to address the non-linear and non-stationary fault signals of wind turbine gear box. The vibration signal is first analyzed using wavelet transform. The aliasing in the high-frequency signal is addressed by applying EMD to the original signal to deal with the interference of high-frequency energy in fault diagnosis with wavelet transform. The experiment results based on a wind turbine gear box with three-stage drive demonstrate that this method can diagnose the faults and locate them accurately.

Acknowledgment

The work was supported in part by the National Natural Science Foundation of China under Grant no. 61075001.

References

- [1] Z. He, Y. Zi, and X. Zhang, *Modern Signal Processing and Its Application in Engineering*, Xi'an Jiaotong University Press, 2007.
- [2] L. Wang, *Research on Intelligent Fault Diagnosis Technology of Complex Equipment*, Shenyang Normal University, 2011.
- [3] J.-S. Luo, D.-J. Yu, and F.-Q. Peng, "Morphological demodulation method based on emd and its application in mechanical fault diagnosis," *Journal of Vibration and Shock*, vol. 28, no. 11, pp. 84–86, 2009.
- [4] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society A*, vol. 454, no. 1971, pp. 903–995, 1998.
- [5] H. Li, H. Zheng, and L. Tang, "Study on gear fault diagnosis based on EMD and Power Spectrum," *Journal of Vibration and Shock*, vol. 25, no. 1, pp. 132–136, 2006.
- [6] L. Wang, C. Wang, and Z. Cai, "The diagnosis method for the rolling bearings tiny fault based on wavelet," *Chinese Journal of Applied Mechanics*, vol. 16, no. 2, pp. 95–99, 1999.
- [7] K. Qin, J. Chen, M. Jiang, and C. Chen, "A new method of fault feature extracting for rolling bearing—spectral-correlation density function," *Journal of Vibration and Shock*, vol. 20, no. 1, pp. 34–37, 2001.
- [8] Y. Tang and Q. Sun, "Study on vibration signal singularity analysis by wavelet for rolling element bearing defect detection," *Journal of Vibration*, vol. 15, no. 1, pp. 111–113, 2002.
- [9] G. Hu, Z. Ren, W. Huang, and G. Wang, "Location of slight fault in electric machine using trigonometric saline frequency modulation wavelet transforms," *Power System Technology*, vol. 27, no. 3, pp. 28–31, 2003.
- [10] Z.-K. Peng, Y.-Y. He, Q. Lu, W.-X. Lu, and F.-L. Chu, "Using wavelet method to analyze fault features of rub rotor in generator," *Proceedings of the Chinese Society of Electrical Engineering*, vol. 23, no. 5, pp. 75–79, 2003.
- [11] S.-X. Yang, J.-S. Hu, Z.-T. Wu, and G.-B. Yan, "The comparison of vibration signals time-frequency analysis between EMD-based HT and WT method in rotating machinery," *Proceedings of the Chinese Society of Electrical Engineering*, vol. 23, no. 6, pp. 102–107, 2003.
- [12] C. Fan, L. Zhang, Z. Wang, and S. Ji, "Research of fault diagnosis of rolling bearings based on EMD and power spectrum analysis method," *Journal of Mechanical Strength*, vol. 28, no. 4, pp. 628–631, 2006.
- [13] C. K. Chui, *An Introduction To Wavelets*, Academic Press, San Diego, Calif, USA, 1992.
- [14] Y. Guo, W. Yan, and Z. Bao, "Gear fault diagnosis of wind turbine based on discrete wavelet transform," in *Proceedings of the 8th World Congress on Intelligent Control and Automation (WCICA '10)*, pp. 5804–5808, July 2010.
- [15] X. Fan and M. J. Zuo, "Gearbox fault detection using empirical mode decomposition," in *Proceedings of the ASME International Mechanical Engineering Congress and Exposition (IMECE '04)*, pp. 37–45, November 2004.
- [16] Q. Miao, D. Wang, H.-Z. Huang, B. Zheng, and X. Fan, "Gearbox on-line condition monitoring using empirical mode decomposition," in *Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (DETC '09)*, pp. 665–670, September 2009.
- [17] W. J. Wang and P. D. McFadden, "Application of wavelets to gear box vibration signals for fault detection," *Journal of Sound and Vibration*, vol. 192, no. 5, pp. 927–939, 1996.
- [18] Y. Wang, Z. He, and Y. Zi, "A comparative study on the local mean decomposition and empirical mode decomposition and their applications to rotating machinery health diagnosis," *Journal of Vibration and Acoustics, Transactions of the ASME*, vol. 132, no. 2, Article ID 021010, 10 pages, 2010.

Research Article

Identification of Magnetizing Inrush and Internal Short-Circuit Fault Current in V/x-Type Traction Transformer

Y. H. Li,¹ X. J. Tian,¹ and X. Q. Li²

¹ School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China

² Department of Electrical Engineering, Beijing Institute of Petrochemical Technology, Beijing 102617, China

Correspondence should be addressed to X. J. Tian; tianxingjun66@163.com

Received 10 May 2013; Accepted 26 June 2013

Academic Editor: Liang Gong

Copyright © 2013 Y. H. Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper deals with the fault diagnosis methodology for the V/x-type traction transformer in railway passenger-dedicated lines. A kind of the protection methodology to combine empirical mode decomposition (EMD), energy weight, and information entropy is proposed. This method can sensitively reflect the dynamic information changes of traction transformer differential current, so it can not only effectively identify internal short-circuit fault current from magnetizing inrush but also better identify fault current accompanied with magnetizing inrush. In this method, the differential current is decomposed by EMD and the energy weight of each intrinsic mode function (IMF) is calculated, and the feature vector of the fault pattern recognition is obtained by constructing IMF energy entropy. The field-measured data illustrate that the new method not only has the advantages of high sensitivity, faster identification speed, and clear concept but also is available for the diagnosis of the complicated dynamic system.

1. Introduction

V/x-type traction transformer is one of the key equipments in the power distributing system of high speed train way. Fault protection is the important technology to guarantee the reliable operation of the V/x-type traction transformer in traction power supply system (TPSS). The application experiences show that the internal short-circuit fault will bring about serious consequences to system operation; thus the differential-fault-protection system of traction transformer must have the ability to correctly identify this fault and to carry out the real-time protection. However, unfortunately, for the traction transformer with large power, the great difficulty is that magnetizing inrush and fault current are of the approaching signal characteristics in signal shape. As the magnetizing inrush is the impulse current produced by the nonlinear magnetic flux-electric current characteristics while transformer switching without load, whose amplitude is close to the amplitude of internal short-circuit fault current, it is very difficult to distinguish them only according to the current amplitude value. This undoubtedly brings difficulties to the design of differential-fault-protection system, if the magnetizing inrush and fault current could not be quickly

and accurately distinguished by protection system, multiple sequential closing and charging for the traction transformer will be conducted, which will seriously shorten the service life of transformer. More thorny and serious problem is that internal short-circuit fault current which accompanies magnetizing inrush while switching without load; this condition should be regarded as fault behavior, but it is probably not identified. Therefore, to distinguish between magnetizing inrush and fault current quickly and correctly has become an urgent problem to be solved in differential-fault-protection system of traction transformer.

Currently, the second harmonic protection principle has been widely used in engineering applications as differential-fault-protection method of traction transformer. However, due to the resonance influence from compensation capacitor of high-voltage transmission line and the decline of magnetic saturation point of the traction transformer, it causes the difference weakening of the second harmonic characteristics between magnetizing inrush and fault current and finally results in the increment of the malfunction rate in differential-fault-protection. Therefore, in order to improve the differential-fault-protection performance of traction

transformer, we must search the new features and quickly and correctly distinguish magnetizing inrush from fault current.

According to the recognition principle of the fault feature, the differential-fault-protection methods of power transformer can be roughly divided into two types: model-based and signal-processing-based methods according to waveform characteristics. Model-based identification methods have flux linkages increment method [1], combined method of magnetizing inductance and loop equilibrium equation [2], equivalent circuit parameters method [3], magnetic flux symmetry property method [4], dynamic impedance characteristics [5], and so forth. Advantages of this type of methods is that it is able to distinguish magnetizing inrush from internal short-circuit fault current on the presupposition that the related electric parameters can accurately be acquired, but some main electrical parameters cannot be accurately measured in the practical applications. Diagnosis methods based on waveform characteristics signal processing are improvement or evolution of the second harmonic principle and the intermittent angle method, such as waveform distribution characteristics method [6], normalized grille curve method [7], wavelet transform method [8], neural network method [9], intelligent hybrid method [10], sine wave distortion characteristics method [11], voltage and current waveforms combined method [12], fundamental current amplitude method [13], mathematical morphology method [14], and so forth. Such methods have the advantage of clear principle, but they are subject to the effect of the nonperiodic component factors.

Traction transformer is different from power transformer in negative-sequence component and harmonics content of differential current. There is only little relevant literature about improved method of traction transformer differential-fault-protection, most of which are model-based method, such as unsaturated zone equivalent instantaneous inductance identification method [15], winding resistance and leakage inductance parameter method [16], and transformer structural modeling method [17], but these improved methods are also difficult to be applied in engineering because of the electrical measurement limits. In contrast, the waveform-based identification method of power transformer has more reference value because of its convenience in electrical measuring, and this method will be a better idea to improve differential-fault-protection of traction transformer by further upgrading algorithm.

Huang proposed a signal processing method based on empirical mode decomposition (EMD), whose basis functions are generated based on the signal itself characteristic scale, which have a clear physical meaning for dealing with nonlinear and nonstationary signals and have been applied in the field of mechanical vibration [18], radar location [19], frequency analysis of power system [20], and some other fields. Magnetizing inrush and fault current belong to nonlinear and nonstationary signals, and they have essentially different in the frequency components features, thus the protection method combining EMD, energy weight, and information entropy is suitable for the working state diagnosis of V/x-type traction transformer. This paper focuses on this new methodology; the differential current signals are first

decomposed into a number of IMF components under different time scales using EMD algorithm and the energy weight of each intrinsic mode function (IMF) and differential current signal is calculated in a power cycle. Then the IMF energy entropy is constructed based on energy weight which is sensitive to variation of differential current information, and this entropy value is defined as the feature vector of distinguishing between magnetizing inrush and internal short-circuit fault current.

2. Feature Extraction and Identification Algorithm

2.1. Decomposition Principle of EMD. The empirical mode decomposition (EMD) algorithm is a new signal analysis method, which can simultaneously and gradually decompose the signal according to different scales or frequency fluctuations or trends to produce a series of the narrow-band stationary data sequence with different characteristic scales; it is called intrinsic mode function (IMF). Thus, the essence of EMD algorithm is extracting the inherent fluctuation mode based on feature time scale of the data and then decomposing the data. For complex multicomponent signals, its decomposition should meet two conditions: (1) the number of its extreme points and zero-crossing points should be equal or differ by 1 at most; (2) at any extreme point the mean value of upper and lower envelopes curves remains zero. For an actual signal $x(t)$, the EMD algorithm is shown as follows.

Step 1. Determine all the extreme points of the signal $x(t)$.

Step 2. Calculate the upper and the lower envelopes of $x(t)$ using the cubic-spline function. In order to obtain easily the envelopes, the end effect needs to be curbed by adopting the mirror-extension extreme point method at both ends of the signal.

Step 3. Calculate the mean $m(t)$ of the upper and the lower envelopes, and define

$$h_1(t) = x(t) - m_1(t). \quad (1)$$

Step 4. Judge whether the $h_1(t)$ meets the IMF conditions. If the condition is not satisfied, let $x(t) = h_1(t)$ and return to Step 1; else set $h_1(t)$ as the first IMF component of the signal $x(t)$. After executing k loops, get $h_{1k}(t) = h_{1(k-1)}(t) - m_{1k}(t)$ and the $h_{1k}(t)$ meets the IMF condition. Define $c_1(t) = h_{1k}(t)$, $c_1(t)$ is called the first IMF component of signal $x(t)$. Generally, this calculation process of IMF component is called the screening method and the class Cauchy convergence criterion is adopted to judge each screening result. The equation of class Cauchy convergence criterion is

$$S_D = \sum_{j=0}^T \frac{[h_{k-1}(t) - h_k(t)]^2}{h_{k-1}^2(t)}. \quad (2)$$

In (2), the value of S_D is generally set from 0.2 to 0.3.

Step 5. Extract the $c_1(t)$ from the $x(t)$, and define

$$r_1(t) = x(t) - c_1(t). \quad (3)$$

After designating $r_1(t)$ as the new input signal and returning to Step 1, get the second IMF component. If the n th loop is carried, the n th IMF component is obtained. That is,

$$\begin{aligned} r_1(t) - c_2(t) &= r_2(t), \\ &\vdots \\ r_{n-1}(t) - c_n(t) &= r_n(t). \end{aligned} \quad (4)$$

When $r_n(t)$ is a monotonic function or $|r_n(t)|$ is close to the measurement error, the total loop ends. According to the EMD decomposition results, the signal $x(t)$ eventually can be expressed as

$$x(t) = \sum_{i=1}^n c_i(t) + r_n(t). \quad (5)$$

Therefore, the actual signal $x(t)$ is decomposed into n IMF components and a residual $r_n(t)$, and $r_n(t)$ is a trend or constant signal.

2.2. Principle of Energy Weight. For description of the differential current signal, the energy accompanied by time scale distribution is one of the most important parameters. When magnetizing inrush and fault current of traction transformer occur, the differential current signal will change greatly, whose typical feature is that the signal energy variation depends on the distribution of time and frequency, and each frequency component contains lots of features message. Therefore, the essential feature message of magnetizing inrush and fault current can be got by analyzing the differential current signal in the different operating conditions. It can be learned from decomposition principle of EMD that the IMF components achieved by dividing different time scale of signal are a group of narrow-band stationary series of local instantaneous frequency ordered from the highest to the lowest. Therefore, for the differential current signal, the IMF energy in different frequencies can reflect the running state of traction transformer.

To construct the energy feature vector at different frequencies which can better represent the differential current signals, the concept of energy weights is presented in this paper. Its essence is to describe the arithmetic square root of ratio of each IMF component energy and differential current signal energy in the data-window using statistical methods. Assuming that the single-phase differential current signal has k components (including the residual) treated by EMD, which indicates it has k multiscale components, the energy weight $\lambda_{c_i}(n)$ of a component c_i in the n th power cycle is

$$\lambda_{c_i}(n) = \frac{Ec_i(n)}{Ei(n)} = \sqrt{\frac{\sum_{j=1}^N |c_i(j)|^2}{\sum_{j=1}^N |i(j)|^2}}, \quad (6)$$

where $Ec_i(n)$ and $Ei(n)$ are, respectively, energy parameter of IMF component c_i and differential current signal in the n th power cycle; N is the data length in a power cycle.

Equation (6) shows that the energy weight $\lambda_{c_i}(n)$ reflects the correlation degree between IMF component c_i and differential current signal in a power cycle (representing all feature information under a single time scale), while the energy weight of whole IMF components represents all of the messages contained in the signal in many feature scales during a power cycle. So, the corresponding relationship between differential current signal and IMF component is established by energy weight, which provides basic conditions for further quantifying current characteristics.

2.3. Construction of Feature Vector. For the fault diagnosis, it is very important to select the parameters that can best reflect fault characteristics. Because the magnetizing inrush is the seriously distorted sine waveform while the fault current remains roughly sine waveform, the IMF component of two types of current will inevitably lead to differences in the number and energy weight. In order to better represent the essential characteristics of dynamic differential current in traction transformer, the concept of entropy in information theory is introduced to further describe the average information quantity of differential current. The entropy can provide useful information of the signal potential dynamic information, whose value is an accurate measure of average uncertainty and complexity of signal change [21]. The entropy $F(p)$ is defined as

$$F(p) = -\sum_{m=1}^{\text{Len}} (p_m \times \ln p_m), \quad (7)$$

where $\sum_{m=1}^{\text{Len}} p_m = 1$, Len is the length of p_m , p_m is the probability of signal, indicating the uncertainty of the signal, and its value is distributed between 0 and 1.

Obviously, the information entropy is a measure of the signal uncertainty degree, and the more uncertain and complex is the signal, the greater its entropy value is. After differential current signal of traction transformer is treated by EMD in a power cycle, IMF component at different time scale is utilized to ascertain the change information of differential current signal. Furthermore, if we define each time scale as an information source, the IMF component at single scale can be regarded as a message from the information source. Therefore, the IMF energy entropy is constructed by information entropy, EMD, and energy weight, which can sensitively reflect the weak and cramped change of differential current. Assuming that the current signal has k components (including the residual) treated by EMD in the n th power cycle, the total energy weight $\lambda_{\text{all}}(n)$ of differential current signal is

$$\lambda_{\text{all}}(n) = \sum_{i=1}^k \lambda_{c_i}(n). \quad (8)$$

The $\tilde{\lambda}_{c_i}(n)$ is got after $\lambda_{\text{all}}(n)$ treated by normalized processing, that is,

$$\tilde{\lambda}_{c_i}(n) = \frac{\lambda_{c_i}(n)}{\lambda_{\text{all}}(n)}. \quad (9)$$

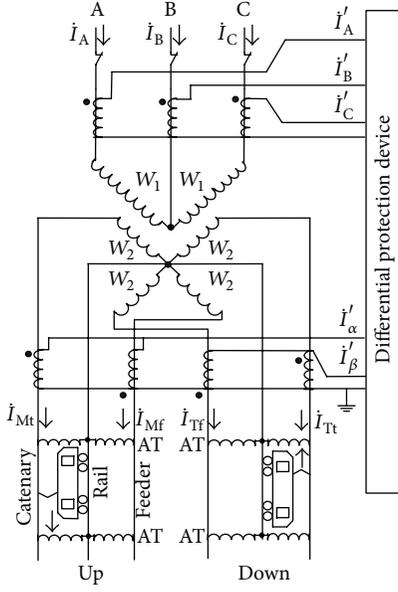


FIGURE 1: Differential-fault-protection wires of V/x-type traction transformers.

Therefore, the IMF energy entropy vector $J(n)$ of single-phase differential current signal in the n th power cycle can be constructed as follows:

$$J(n) = - \sum_{m=1}^k [\tilde{\lambda}_{c_i}(n) \times \ln \tilde{\lambda}_{c_i}(n)]. \quad (10)$$

It can be learned from (9) that the IMF energy vector $J(n)$ reflects the energy distribution of differential current signals at different scales, and the value of vector $J(n)$ is smaller for internal short-circuit fault current while the vector $J(n)$ is greater for magnetizing inrush.

3. Differential Current Signal and Magnetizing Inrush of Traction Transformer

3.1. Differential Current Signal. In the TPSS of railway passenger-dedicated lines, the most typical traction transformers are V/x-type wiring transformer group which is connected with two single-phase three-winding transformers by V-type. Such traction transformer has the advantages of high capacity utilization and two windings configured respectively, and its differential-fault-protection wires are shown in Figure 1. In addition, to increasing the power supply distance of feeder, the output of traction transformer is connected to catenary, feeder, and parallel autotransformer (AT), while the locomotive gets power from catenary and rail (center tap of the secondary side in AT), so the input voltage of locomotive pantograph is half of the total voltage.

According to the ampere-turns balance principle, the transmission ratio of current transformer, and designation of current direction, the extraction principle of differential

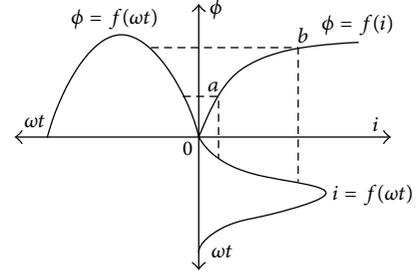


FIGURE 2: Waveform of magnetizing inrush.

current signals (treated in the differential-fault-protection system using software method) is

$$\begin{aligned} \Delta \dot{I}_A &= \dot{I}'_A - \frac{K_{i2}}{K_T K_{i1}} \dot{i}'_\alpha, \\ \Delta \dot{I}_B &= \dot{I}'_B - \frac{K_{i2}}{K_T K_{i1}} \dot{i}'_\beta, \end{aligned} \quad (11)$$

$$\Delta \dot{I}_C = \dot{I}'_C + \frac{K_{i2}}{K_T K_{i1}} (\dot{i}'_\alpha + \dot{i}'_\beta),$$

$$\dot{i}'_\alpha = \dot{I}_{Mt} - \dot{I}_{Mf}, \quad \dot{i}'_\beta = \dot{I}_{Tt} - \dot{I}_{Tf},$$

where $\Delta \dot{I}_A$, $\Delta \dot{I}_B$, and $\Delta \dot{I}_C$ are, respectively, the three-phase differential current signals; \dot{I}'_A , \dot{I}'_B , and \dot{I}'_C are, respectively, the three-phase ratio currents in high-voltage side; \dot{i}'_α and \dot{i}'_β are, respectively, the up and down ratio current in low-voltage side; \dot{I}_{Mf} , \dot{I}_{Mt} , \dot{I}_{Tf} , and \dot{I}_{Tt} are, respectively, the ratio current of up catenary, up feeder, down catenary, and down feeder; K_T is the ratio of turns W_1 in high-voltage side and turns W_2 in low-voltage side; K_{i1} and K_{i2} are, respectively, proportionality factor in high-voltage side and low-voltage side.

3.2. Nonlinear Generating Mechanism of Magnetizing Inrush. Magnetizing inrush of transformer is generated by the core excitation saturation factor. As shown in Figure 2, the magnetization curve of traction transformer magnetic materials shows the typical nonlinear feature. That is, in a certain voltage, excitation current waveform not only depends on the core flux density B_m but also depends on the core saturation degree. While $B_m < B_s$ (which is saturated magnetic flux density), the transformer magnetic circuit is in unsaturated condition and the functional relationship between excitation current i and magnetic flux ϕ is linear. While $B_m > B_s$, the magnetic circuit begins to saturate, and the functional relationship between i and ϕ becomes nonlinear because the increasing speed of i is faster than the increasing speed of ϕ . Due to the core excitation saturation, the waveform of excitation current will be distorted into peak wave, and the deeper core saturation degree will cause the more serious distortion of excitation current waveform.

Generally, there are mainly three operation conditions in traction transformer: transformer switching while without load; running with internal short-circuit fault; and switching without load but with internal short-circuit fault. This

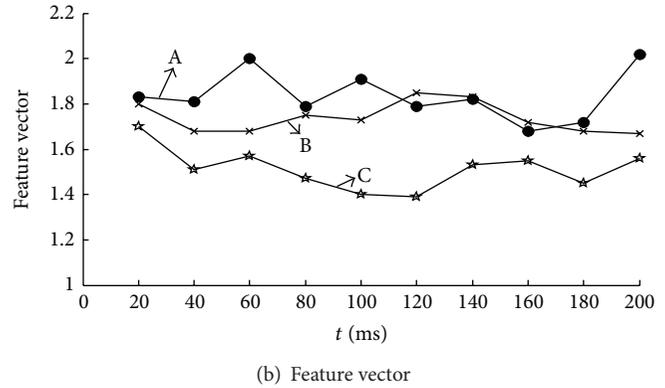
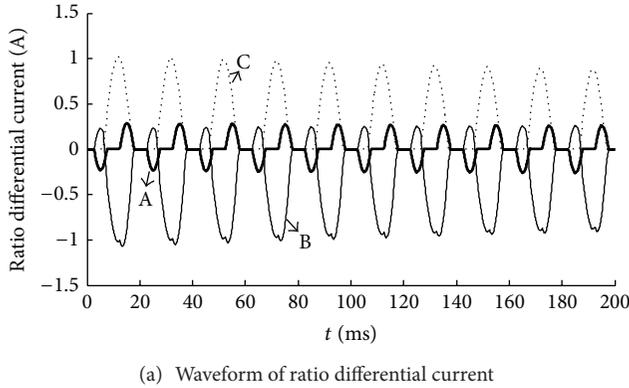


FIGURE 3: Magnetizing inrush of transformer switching on without load.

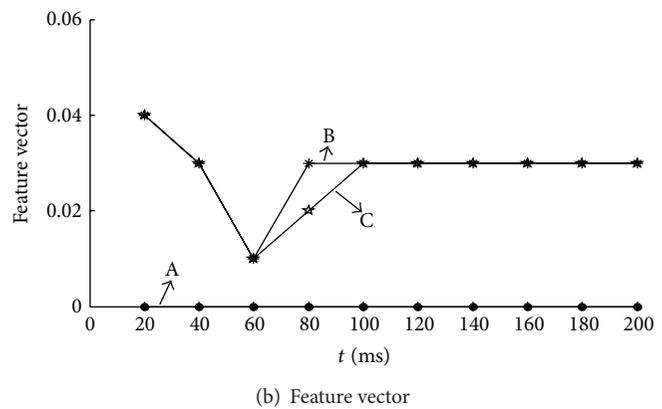
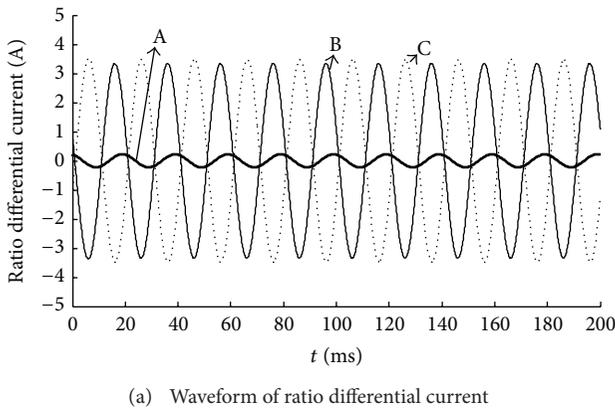


FIGURE 4: Internal short-circuit fault occurring in operation.

needs to be judged in differential-fault-protection system. The following is to test diagnosis effect and sensitivity of differential-fault-protection algorithm based on IMF energy entropy using field-measured differential current signal.

4. Diagnosis and Analysis of Actual Measured Current Signal

4.1. Magnetizing Inrush of Transformer Switching without Load. Figure 3(a) shows that waveform of three single-phase ratio differential currents while traction transformer switches without load (a power cycle is 20 ms and the transmission ratio of current transformer is 1000:1), and Figure 3(b) depicts change of its IMF energy entropy vector.

Figure 3(a) shows that the waveform of ratio differential current is distorted seriously while traction transformer switches without load, including unilateral and symmetrical magnetizing inrushes in waveform shape. As can be qualitatively analyzed in Figure 3(b), the most significant characteristic is that the each single-phase IMF energy entropy vector $J(n)$ is greater in a cycle, the minimum $J_C(n)$ is equal to 1.39, and the rest of vector value fluctuates from 1.40 to 2.02. And the other measured magnetizing inrush also meets this law.

4.2. Running with Internal Short-Circuit Fault. Figure 4(a) shows the waveform of three-phase ratio differential current

while the internal short-circuit fault of traction transformer occurs in running, and its IMF energy entropy vector is shown in Figure 4(b).

It can be seen in Figure 4(a) that a single-phase internal short-circuit fault of traction transformer happens in running, the three single-phase differential currents totally maintain the sine waveform shape, and the amplitude of fault phase is close to the amplitude of a nonfault phase (the reason is that current information is extracted by differential ways). As can be seen in Figure 4(b), each single-phase vector $J(n)$ is very small, and its value is almost close to 0.

4.3. Switching without Load but with Internal Short-Circuit Fault. Figure 5(a) shows waveform of three-phase ratio differential current while traction transformer switches without load but with internal short-circuit fault, whose IMF energy entropy vector is depicted in Figure 5(b). It can be seen in Figure 5(a) that the DC component of differential current increases because of the coefficient of magnetizing inrush and internal short-circuit current at the same time, and the ratio differential current magnitude is larger for fault phase and a nonfault phase. In Figure 5(b), the feature vector $J_A(n)$ of non-fault phase A is greater than 1.5 in 0.2 s, while $J_B(n)$ and $J_C(n)$ slowly reduce from a smaller value.

4.4. Result Analysis of Actual Measured Data. In recent five years, many magnetizing inrush cases and some internal

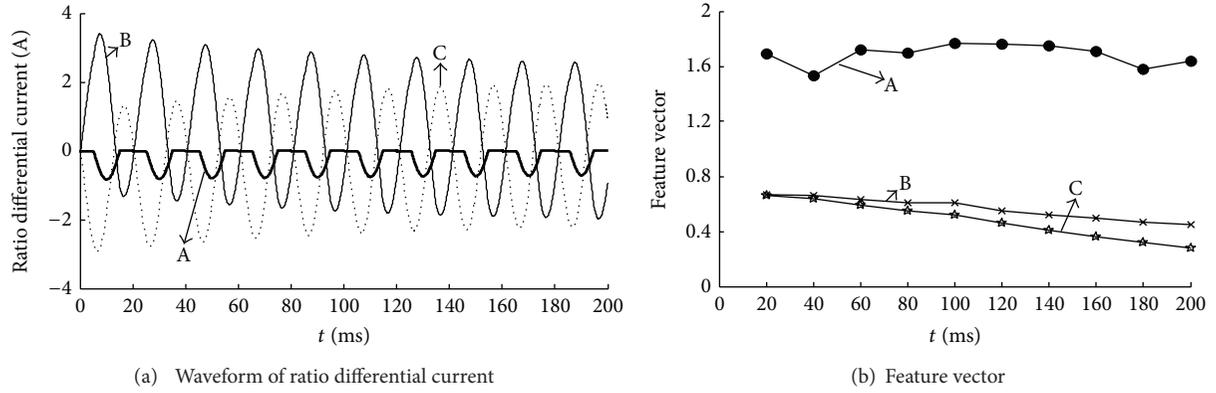


FIGURE 5: Internal short-circuit fault while switching without load.

TABLE 1: Feature distribution while transformer switching without load.

Cycle sequence	Phase A	Phase B	Phase C
1	[1.42, 1.98]	[1.32, 1.92]	[1.50, 1.94]
2	[1.33, 1.93]	[1.26, 1.95]	[1.32, 1.92]
3	[1.41, 2.00]	[1.28, 2.02]	[1.30, 2.04]
4	[1.35, 1.91]	[1.27, 1.89]	[1.28, 1.99]
5	[1.39, 1.91]	[1.29, 1.92]	[1.29, 2.10]

TABLE 2: Feature distribution while running with internal short-circuit fault.

Cycle sequence	Phase A	Phase B	Phase C
1	[0.00, 0.03]	[0.00, 0.31]	[0.01, 0.30]
2	[0.00, 0.02]	[0.00, 0.23]	[0.01, 0.22]
3	[0.00, 0.02]	[0.00, 0.19]	[0.01, 0.15]
4	[0.00, 0.01]	[0.00, 0.25]	[0.01, 0.23]
5	[0.00, 0.01]	[0.00, 0.28]	[0.01, 0.26]

short-circuit fault cases were recorded in China's railway passenger-dedicated lines. After these current waveform data are treated by new feature extraction algorithm, the distributions of IMF energy entropy vector are shown in Tables 1, 2, and 3, respectively.

Table 1 shows that the IMF energy vector $J(n)$ of three-phase ratio differential current has the outstanding features in each cycle of the first 0.1 s: (1) all the feature entropy values are larger, including the smallest $J_B(2) = 1.26$; (2) the feature vector shows good convergence.

It can be learned from Table 2 that the three-phases IMF energy entropy vector $J(n)$ are all smaller in the first 0.1 s, such as the maximum vector $J_B(1) = 0.31$.

As can be seen from Table 3, three-phase IMF energy entropy vector $J(n)$ has the following features in each cycle of the first 0.1 s: (1) one phase IMF energy entropy value is larger, two-phase feature entropy values are smaller, and its value shows good convergence; (2) the gap between the lower limit of maximum entropy value and the upper limit of the minimum entropy value is very great.

TABLE 3: Feature distribution while switching without load but with fault.

Cycle sequence	Phase A	Phase B	Phase C
1	[1.45, 2.01]	[0.10, 0.76]	[0.64, 0.72]
2	[1.28, 2.06]	[0.14, 0.74]	[0.62, 0.70]
3	[1.35, 2.05]	[0.13, 0.73]	[0.49, 0.69]
4	[1.28, 1.94]	[0.10, 0.74]	[0.53, 0.68]
5	[1.32, 1.96]	[0.07, 0.71]	[0.43, 0.68]

According to comprehensive analysis of the above field-measured experimental data, some important conclusions are obtained.

- (1) Frequency component of internal short-circuit fault current is less and the energy of fault inrush is mainly focused on the few dominant components, while the frequency component of magnetizing inrush is rich and the energy of magnetizing inrush is scattered.
- (2) In three operating conditions, the IMF energy entropy value is stable in 0.1 s, thus the vector value of the first cycle (20 ms) is adopted to set threshold value and diagnose fault in order to improve the speed of identification.
- (3) In the first cycle, the three-phase feature vector of magnetizing inrush (when transformer switches without load) meets the conditions $J(1) \geq 1.32$; three-phase feature vector of fault current (the internal short-circuit fault occurs in running) meets the conditions $J(1) \leq 0.31$; and three-phase feature vector while transformer switches without load but with internal short-circuit fault meets the conditions of one phase $J(1) \geq 1.45$ and the other two phases $J(1) \leq 0.76$. Therefore, the spatial distance between fault status and nonfault status in the first cycle is 0.56 ($1.32 - 0.76 = 0.56$).
- (4) Assuming that the threshold value $J_{set}(1)$ is set from 0.83 to 1.25 (choosing a reliable margin of 10%), fault current and magnetizing inrush can be correctly distinguished in the first cycle using the IMF energy

TABLE 4: Setting range of the tradition protection algorithm.

The first cycle	The second cycle	The third cycle	The fourth cycle	The fifth cycle
0.04	0.05	0.06	0.06	0.10

entropy algorithm. In addition, it is proved by some field-measured current data that this conclusion also applies to the larger capacity traction transformer.

4.5. Comparison with the Tradition Protection Algorithm. At present, the second harmonic protection algorithm adopted single-phase brake is used as the differential-fault-protection method of traction transformer. Using the second harmonic method in the above operation conditions, the setting scope of three-phase ratio differential current from the first cycle to the fifth cycle is got and listed in Table 4.

It can be seen from Table 4 that the setting range of the second harmonic protection algorithm in each cycle is very small, and the average setting range is 0.06 which is about one-tenth of the IMF energy entropy algorithm; thus it is the main reason of difficult setting for the traction transformer differential-fault-protection system. In addition, due to the fact that energy of the nonperiodic component will be diluted by the total energy, its influence on entropy values becomes smaller. Therefore, the new differential-fault-protection method based on IMF energy entropy in the accurate identification between magnetizing inrush and internal short-circuit current is not only better than the second harmonic method but also a good solution to the influence problems of non-periodic component.

5. Conclusion

The traditional differential-fault-protection performance of V/x-type traction transformer is seriously affected by the instability of the second harmonic component in magnetizing current, so a more reliable fault diagnosis methods need to be developed in engineering applications. According to the characteristics that there is a magnetizing inrush containing rich frequency components when traction transformer switches without load, while an internal short-circuit fault current is roughly a sine wave, the protection methodology combining empirical mode decomposition (EMD), energy weight, and information entropy is presented to identify magnetizing inrush and internal short-circuit fault current.

For the traction transformer, this method can sensitively reflect the dynamic information changes of differential current. If the IMF energy entropy values of three single phases are all large, it indicates that the traction transformer is switching without load (magnetizing inrush). If the IMF energy entropy values of three single phases are all small, it reflects that the traction transformer causes internal short-circuit fault in the operation. If the three single-phase feature entropy value shows such special phenomenon that a phase value is large while the two-phase value is small, it illustrates that the traction transformer is switching without load but with internal short-circuit fault.

It is crucial to achieve the new algorithm by selecting appropriate decomposition conditions. It has been found in experiment that, for the three-phase ratio differential current data of traction transformer, the EMD decomposition can be executed smoothly and quickly by using the mirror-extension extreme point method on border problem and the Cauchy convergence criterion on the termination conditions. Proved by some field-measured differential current data, the new algorithm based on IMF energy entropy has the advantages of high accuracy, faster identification speed, and clear concept; furthermore it is suitable to analyze complex and dynamic characteristics information of the traction transformer in railway passenger-dedicated lines.

Acknowledgment

This paper was supported by Scientific Research and Development Planning Projects "Study of Automation Systems Solutions and Technical Specifications in Traction Power Supply" in Ministry of Railways, China, under Grant 2007J022.

References

- [1] Y.-C. Kang, B.-E. Lee, T.-Y. Zheng, Y.-H. Kim, and P. A. Crossley, "Protection, faulted phase and winding identification for the three-winding transformer using the increments of flux linkages," *IET Generation, Transmission and Distribution*, vol. 4, no. 9, pp. 1060–1068, 2010.
- [2] X. Deng, C. Wang, and Z. Zhang, "UHV transformer protection based on equivalent circuit equation and excitation inductance," *Proceedings of the Chinese Society of Electrical Engineering*, vol. 32, no. 1, pp. 147–153, 2012.
- [3] F. de Leòn and J. A. Martinez, "Dual three-winding transformer equivalent circuit matching leakage measurements," *IEEE Transactions on Power Delivery*, vol. 24, no. 1, pp. 160–168, 2009.
- [4] Y. B. Zhao and Y. P. Lu, "New algorithm based on flux symmetry character for judging transformer inrush current," *Transactions of China Electrotechnical Society*, vol. 22, no. 12, pp. 66–71, 2007.
- [5] M.-L. Wu, Q.-M. Li, and Y.-B. Duan, "Identification of transformer magnetizing inrush current by use of dynamic impedance characteristics," *High Voltage Engineering*, vol. 33, no. 9, pp. 50–60, 2007.
- [6] X. Wang and Z. Wang, "Identification of transformer inrush currents based on waveform distribution characteristics," *Transactions of China Electrotechnical Society*, vol. 27, no. 1, pp. 148–154, 2012.
- [7] J. Ma, Z. Wang, Q. Yang, and Y. Liu, "Identifying transformer inrush current based on normalized grille curve," *IEEE Transactions on Power Delivery*, vol. 26, no. 2, pp. 588–595, 2011.
- [8] H.-Y. Li, X. Zhang, and Y.-B. Hou, "New method of inrush current discrimination based on wavelet transform," *High Voltage Engineering*, vol. 34, no. 3, pp. 500–503, 2008.
- [9] M. Tripathy, R. P. Maheshwari, and H. K. Verma, "Power transformer differential protection based on optimal probabilistic neural network," *IEEE Transactions on Power Delivery*, vol. 25, no. 1, pp. 102–112, 2010.
- [10] D. Barbosa, D. V. Coury, and M. Oleskovicz, "New approach for power transformer protection based on intelligent hybrid systems," *IET Generation, Transmission and Distribution*, vol. 6, no. 10, pp. 1009–1018, 2012.

- [11] J.-H. He, J.-Z. Li, B. Yao, Z.-J. Ou, and Y. Fan, "New approach of transformer inrush detected based on the sine degree principle of current waveforms," *Proceedings of the Chinese Society of Electrical Engineering*, vol. 27, no. 4, pp. 54–59, 2007.
- [12] X. Zeng, X. N. Lin, H. L. Weng, and X. F. Sheng, "Novel criterion of transformer protection based on time difference characteristic between phase voltage and differential current," *Automation of Electric Power Systems*, vol. 33, no. 3, pp. 79–83, 2009.
- [13] J. Ma, Z. Wang, and J. Wu, "A novel method to rapidly identify inrush current and internal fault current based on variation characteristic of fundamental current amplitude," *Transactions of China Electrotechnical Society*, vol. 24, no. 6, pp. 166–171, 2009.
- [14] J.-D. Huang and W.-Q. Luo, "New algorithm to identify inrush current based on improved mathematical morphology," *Proceedings of the Chinese Society of Electrical Engineering*, vol. 29, no. 7, pp. 98–105, 2009.
- [15] C.-L. Feng, B.-M. Ge, and D.-Q. Bi, "Identification of magnetizing inrush for three-phase V/V traction transformer," *Journal of the China Railway Society*, vol. 33, no. 6, pp. 35–40, 2011.
- [16] H.-C. Guan, S.-B. Gao, and Y.-S. Pan, "Research on SCOTT transformer protection based on parameter identification," *High Voltage Engineering*, vol. 34, no. 3, pp. 546–559, 2008.
- [17] L. Guo, S. Gao, and Q. Li, "Protection of impedance-matching balance traction transformer based on transformer model," *Journal of Southwest Jiaotong University*, vol. 41, no. 6, pp. 737–742, 2006.
- [18] H. Li, H. Wang, L. Bai, X. Luo, and P. Guo, "Dynamic characteristic information extraction of hydroturbine based on improved Hilbert-Huang transform method," *Proceedings of the Chinese Society of Electrical Engineering*, vol. 31, no. 2, pp. 78–84, 2011.
- [19] J. Niu, Y.-X. Liu, Y.-L. Qin, W.-D. Jiang, and X. Li, "A new method of radar micro-motion feature extraction of cone target based on empirical mode decomposition," *Acta Electronica Sinica*, vol. 39, no. 7, pp. 1712–1715, 2011.
- [20] C. Li, J. Liu, J. Yang, L. Yao, and M. Bazargan, "Frequency shift empirical mode decomposition for extracting low frequency oscillation parameters," *Automation of Electric Power Systems*, vol. 35, no. 20, pp. 1025–1032, 2011.
- [21] J. Wang, L. L. Cui, H. Q. Wang, and P. Chen, "Improved complexity based on time-frequency analysis in bearing quantitative diagnosis," *Advances in Mechanical Engineering*, vol. 2013, Article ID 258506, 11 pages, 2013.

Research Article

Thermal Performance Prediction of a Trapezoidal Cavity Absorber for a Linear Fresnel Reflector

Yanhua Lai, Tao Wu, Shuping Che, Zhen Dong, and Mingxin Lyu

School of Energy and Power Engineering, Shandong University, Jinan 250061, China

Correspondence should be addressed to Yanhua Lai; laiyh@sdu.edu.cn

Received 12 April 2013; Accepted 19 June 2013

Academic Editor: Cheng-liang Liu

Copyright © 2013 Yanhua Lai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A trapezoidal cavity absorber for a linear Fresnel reflector concentrator is analyzed and optimized via CFD simulation. The heat loss coefficient is introduced; the influences of ambient temperature, absorber temperature, cavity depth, inclination of side walls, insulation thickness, glass window, and emissivity of selective absorption coating have been studied. The results show that radiation dominates the cavity heat loss, and heat loss through the glass window is higher than through the insulation layer; among these factors, impact of emissivity of selective absorption coating and insulation layer is greater than that of the other factors. The simulation results via CFD prove that cavity with a 100 mm depth shows the best thermal performance among the parameters that have been taken into account.

1. Introduction

Concentrating solar collecting technology is necessary for the advanced solar thermal utilization which can obtain high temperature heat source. The solar concentrating collectors mainly consist of parabolic through collector (PTC), parabolic dish reflector (PDR), heliostat field collector (HFC), and linear Fresnel reflector (LFR) [1]. The most popular in commercial use is PTC which owes more than 10 years of successful operation to SEGS [2] located in southern California of America. But the costs (costs?) of manufacture, operation, and maintenance of PTCs are higher, which leads to a hot research on LFR. Compared to PTC, the main advantages of LFR [3, 4] are as follows:

- (1) fixed absorber tube with no need for flexible high pressure joints;
- (2) much cheaper planar or slight curvature reflecting mirrors and simple tracking systems;
- (3) efficient use of land since the primary reflecting mirrors can be placed one next to the other;
- (4) primary mirrors are mounted near ground so the cost of structural support is low, and the wind loads are small.

The LFR can be regarded as a decomposed PTC which consists of dozens of rows of primary reflecting mirrors and a stationary absorber (see Figure 1). Sun rays are reflected into the stationary receiver by the mirrors. Häberle et al. [5] had investigated the optical and thermal performance of Solarmundo LFR; the Solarmundo Fresnel collector has about 70% thermal performance of a parabolic trough (UVAC) per aperture area and is about 10% below the electricity costs of the whole system. The first person who brought LFR concept is Giorgio Francia [6] in the 1960s. In 1979, FMC company designed a 100 MWe LFR power plant for the Department of Energy, but the plan deadlocked over the funds. The research upsurge of LFR began in the 1990s. Israel company Paz built an LFR system with a Compound Parabolic Collector (CPC) as the secondary reflector [7]. The Belgium company Solarmundo built a similar LFR system [8] but huge; the primary reflecting mirror area is 2500 m², single mirror with a 0.5 m width. Germany carried out the “VDemo Fresnel” project [9]. The first demonstration LFR power plant Puerto Erradol had been built in Spain 2009.

One technical problem is shading and blocking between adjacent mirrors; a simple solution is to increase the height of receiver tower and to increase the gap between adjacent mirrors, which may increase the initial cost. Mills and Morrison [10] of University of Sydney brought up the concept

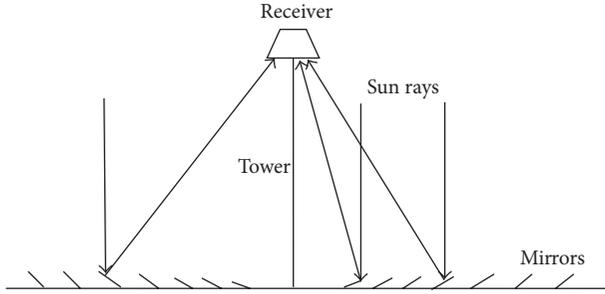


FIGURE 1: Schematic diagram of an LFR system.

of compact linear Fresnel reflector (CLFR), which has less site area. The optical efficiency is 74% at top. Solar Heat and Power Pty had built a 40 MWe CLFR solar heating system for preheating boiler feed water in Hunter Valley.

Receiving device is the key equipment of an LFR system [11]. Considering the requirement of the working fluid outlet temperature and concentration ratio, there are two main types of the receiving devices [12–14]. For the high temperature section (350°C~400°C), a metal glass vacuum tube and a CPC secondary reflector are used, which may be more expensive; for a lower temperature section (200°C~325°C), a trapezoid cavity receiver is used, which contains a tube bundle on upper surface of the cavity.

In this paper, we will analyze the thermal performance of a trapezoidal cavity for a small linear Fresnel receiver via CFD simulation and discuss the influence of cavity geometry, insulation thickness, emissivity of selective absorption coating, ambient temperature, and absorber temperature. The goal of this paper is to achieve a maximum thermal efficiency of the cavity, which can be regarded as the transfer coefficient between the outer surface of the cavity and the ambient air at steady state, and provide an optimization design of the cavity receiver for LFR.

2. Heat Transfer Model

2.1. Mathematical Model. The absorber is enclosed in order to minimize convection losses with ambient air. The bottom surface of the cavity is a glass window that allows the sun rays from the mirror array to enter the cavity to be absorbed by the absorber plate; there is a tube bundle upon the plate. Dey [15] studied four different arrangements of absorber plate and tube bundle, which is out of the scope of this work. The sides and top of the cavity are insulated on the outside to minimize conduction and convection losses to the environment. The absorber heats up due to the incident concentrated solar radiation and emits long wavelength radiation into the cavity. The emitted radiation is absorbed by the cavity sides and window; in the end, heat loss has been caused through insulation and window (see Figure 2).

In a Cartesian coordinate system, the continuity equation is as follows:

$$\frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} = 0. \quad (1)$$

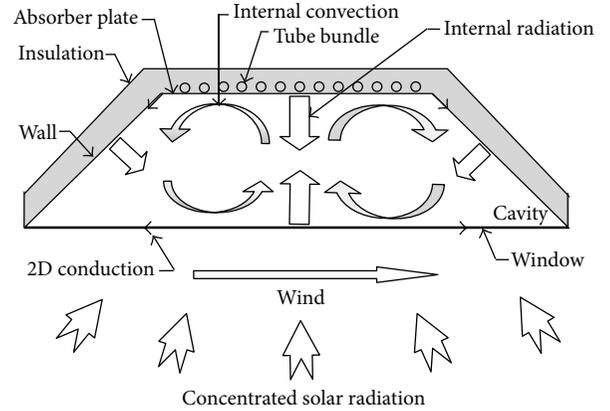


FIGURE 2: Cross-section of the absorber models of heat transfer.

The momentum conservation equation is

$$\frac{\partial(\rho u^2)}{\partial x} + \frac{\partial(\rho uv)}{\partial y} = -\frac{\partial P}{\partial x} + \frac{\partial}{\partial x} \left(\mu \frac{\partial u}{\partial x} \right) + \frac{\partial}{\partial y} \left(\mu \frac{\partial u}{\partial y} \right),$$

$$\frac{\partial(\rho uv)}{\partial x} + \frac{\partial(\rho v^2)}{\partial y} = -\frac{\partial P}{\partial y} + \frac{\partial}{\partial x} \left(\mu \frac{\partial v}{\partial x} \right) + \frac{\partial}{\partial y} \left(\mu \frac{\partial v}{\partial y} \right) - g\rho. \quad (2)$$

The energy equation is

$$\frac{\partial(\rho h)}{\partial x} + \frac{\partial(\rho uv)}{\partial y} = \frac{\partial}{\partial x} \left(k \frac{\partial T}{\partial x} \right) + \frac{\partial}{\partial y} \left(k \frac{\partial T}{\partial y} \right) + q_v, \quad (3)$$

where h is the enthalpy and q_v is the volumetric heat source.

The buoyancy-driven flow regime of natural convection should be decided by the Rayleigh number, which can be viewed as the ratio of buoyancy and viscosity forces times the ratio of momentum and thermal diffusivities,

$$Ra = GrPr = \frac{g\beta\Delta T l^3}{\nu^2} Pr = \frac{g\beta\Delta T l^3}{\alpha\nu}, \quad (4)$$

where Pr is the Prandtl number, g is gravity, l is characteristic length, α is dynamic viscosity, ν is kinematic viscosity, and β is thermal expansion coefficient.

Ra less than 10^8 indicates a buoyancy-induced laminar flow [16], with transition to turbulence occurring over the range of $10^8 < Ra < 10^{10}$.

2.2. Solution Method. To simplify the heat transfer model, several assumptions are used:

- (1) steady state;
- (2) 2D heat transfer model;
- (3) the absorber plate temperatures are fixed and resulting heat losses are calculated.

Notation for the cavity dimensions is shown in Figure 3. Fixed parameters in the modeling are shown in Table 1, and variable parameters are shown in Table 2.

TABLE 1: Constant parameters used in the simulation.

Constant parameter	Symbol	Value
Absorber plate width	W	160 mm
Side walls emissivity	ε_w	0.5
Glass window emissivity	ε_g	0.85
Glass window transmissivity	τ_g	0.95
Insulation thermal conductivity	λ_i	0.04 W/(m·K)

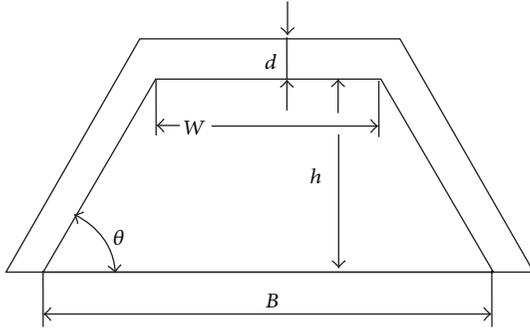


FIGURE 3: Schematic of cavity geometry.

The heat transfer model is solved by the commercial CFD software Fluent 6.3, taking into account all heat transfer mechanisms: radiation, convection, and conduction. A $0.5 \text{ m} \times 0.5 \text{ m}$ outer space is created for computing convection and radiation between the cavity and the environment. Heat transfer is fierce inside and around the cavity, so the grid is detailed there. Triangle mesh is used in the entire region, which is shown in Figure 4.

The absorber plate is modeled as isothermal surface. Surfaces of side walls, insulation, and glass window are modeled as combined convection/radiation boundary. The simulation model is laminar according to Ra number.

In this model, the glass window is semitransparent, so the discrete ordinates (DO) radiation model is chosen. All discretization is carried out using first order upwind. Air properties are modeled as piecewise linear. Minimum convergence criteria are set at 10^{-3} for continuity and velocity and 10^{-6} for DO and energy. First, 500-step iterations are carried out under default underrelaxation factors and then turn them down, proceeding 10000-step iterations, monitoring the temperature of outer surface of glass window, making sure there is no fluctuation.

3. Results and Discussion

3.1. Heat Loss Coefficient. When heat absorption and heat release reach a balance of the cavity receiver, as usual stating under steady state, the capacities that are absorbed and lost are equal:

$$q_{\text{loss}} = q_a = mC_P (T_o - T_i), \quad (5)$$

where m is the mass flow rate, T_i is the working fluid inlet temperature, and T_o is the working fluid outlet temperature.

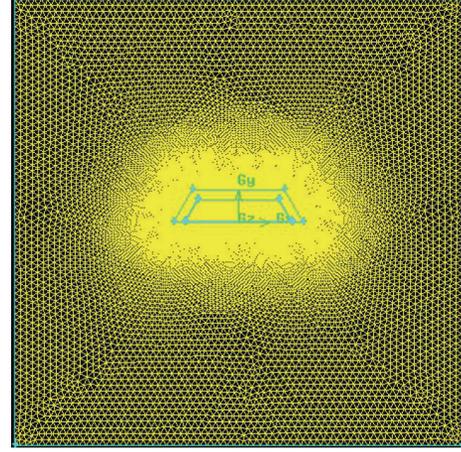
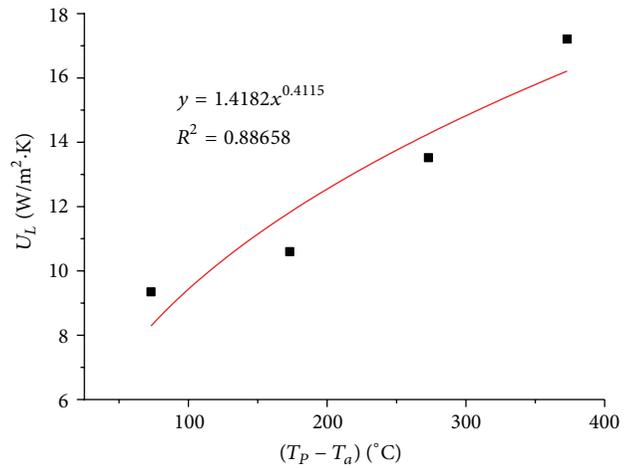


FIGURE 4: Mesh of the entire region.

FIGURE 5: Nonlinear curve fit of U_L ($h = 100 \text{ mm}$, $d = 20 \text{ mm}$).

Heat loss coefficient U_L is a key indicator of the thermal performance of the cavity receiver. As the U_L increases, the total heat lost increases. U_L is defined as

$$U_L = \frac{q_{\text{loss}}}{A_P (T_P - T_a)}, \quad (6)$$

where A_P is the area of absorber plate, namely, area of the side coated with absorbing coatings, T_P is the temperature of absorber plate, and T_a is the temperature of ambient air.

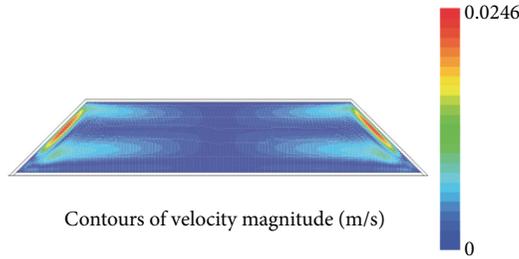
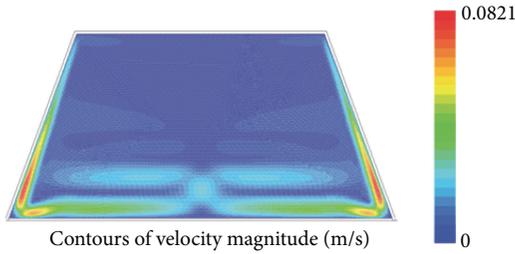
According to the literature [17], U_L can be predicted by the following empirical formula:

$$U_L = c \cdot (T_P - T_a)^d. \quad (7)$$

3.2. Heat Loss Simulated Result. The heat loss coefficient based on absorber plate aperture was calculated as a function of the difference between average tube temperature and ambient air temperature; see Figure 5. Figures 6 and 7 show the velocity magnitude inside the cavity, temperature of the absorber plate is 400°C , and width is 160 mm. These two

TABLE 2: Variable parameters used in the simulation.

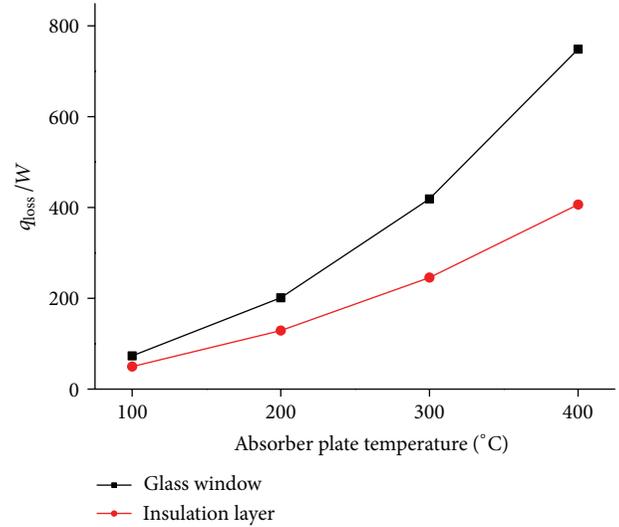
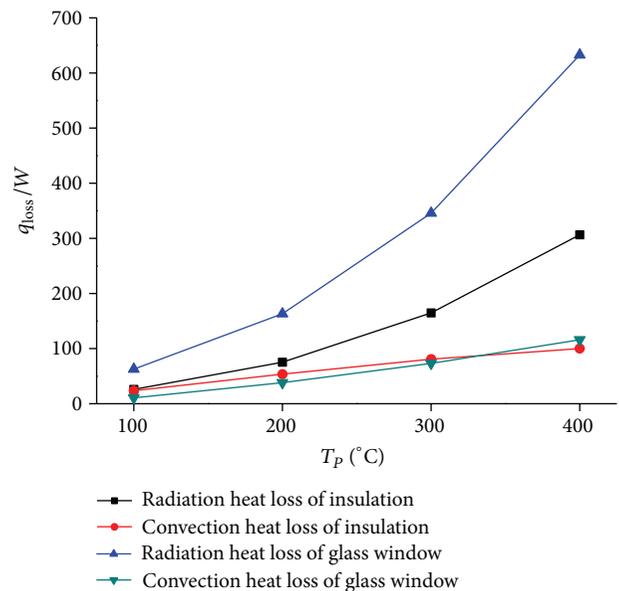
Varied parameters	Symbol	Value
Cavity depth	h	50 mm, 100 mm, and 150 mm
Side wall inclination	θ	30°, 45°, and 60°
Insulation thickness	d	0 mm, 10 mm, 20 mm, and 30 mm
Ambient temperature	T_a	270 K, 280 K, 290 K, 300 K, and 310 K
Absorber plate temperature	T_p	100°C, 200°C, 300°C, and 400°C
Emissivity of selective absorption coating	ε_p	0.05, 0.1, 0.2, 0.3, 0.4, and 0.5

FIGURE 6: Contours of velocity magnitude inside the cavity for $h = 50$ mm.FIGURE 7: Contours of velocity magnitude inside the cavity for $h = 150$ mm.

figures confirmed the convection heat transfer of inside walls through the air flow. Velocity near the side walls is the strongest among the inside area, the flow is enhanced when the cavity depth grows, and the biggest velocity occurs when cavity depth is 150 mm, which is 0.0821 m/s.

Heat losses through glass window are higher than through insulation layer, which can be observed in Figure 8. With the rise of the temperature of absorber plate, the difference is getting more and more obvious. When the temperature of the absorber is 100°C, heat losses through glass window are 1.48 times of that through insulation, when the temperature of the absorber reaches 400°C, that is, 1.84. The radiation losses dominate the thermal losses inside the cavity as presented in Figure 9, and the radiation losses through the glass window occupy the most of the total heat losses.

Figure 10 shows the heat loss coefficient U_L versus the absorber temperature. U_L increases with the increase of the absorber temperature. When the temperature of the absorber plate is 100°C, U_L is 9.35 W/(m²·K); as to 400°C of the absorber plate, U_L is 17.21 W/(m²·K), and 84.1% is decreased.

FIGURE 8: Schematic of heat loss versus T_p ($h = 100$ mm, $W = 160$ mm, and $d = 20$ mm).FIGURE 9: Schematic of heat loss versus T_p ($h = 100$ mm, $W = 160$ mm, and $d = 20$ mm).

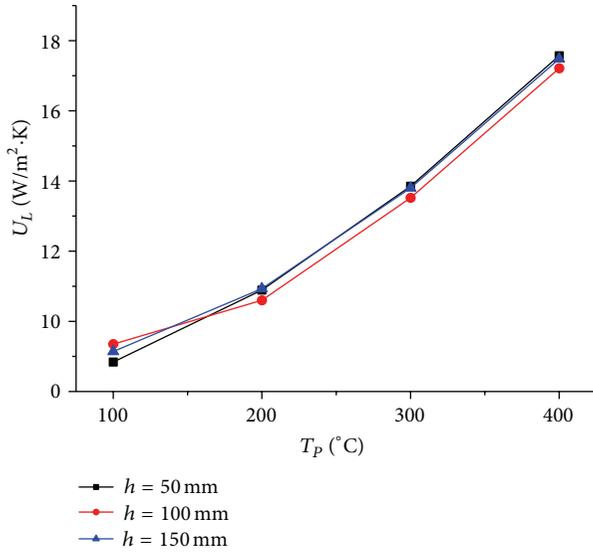


FIGURE 10: Schematic of U_L versus T_p ($W = 160$ mm, $d = 20$ mm).

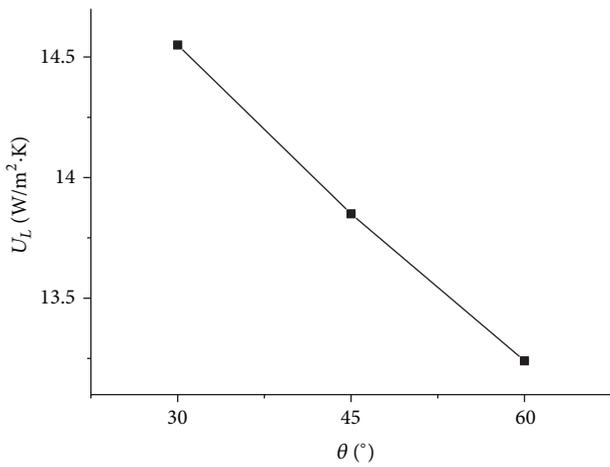


FIGURE 11: Schematic of U_L versus θ ($W = 160$ mm, $d = 20$ mm, and $h = 50$ mm).

We can also notice the influence of cavity depth h to the heat loss coefficient U_L from Figure 10. When the ambient temperature, insulation thickness, and the inclination of side walls are constant, U_L varied with cavity depth are small, but U_L with the 100 mm cavity depth are the smallest in general.

Hold other parameters' constant, such as width of absorber plate and cavity depth, and change the inclination of side walls; the influence is shown in Figure 11. U_L shows a downtrend with the increase of inclination of side walls; compared with U_L with the 30° inclination, 9% declination occurs with the 60° inclination of side walls; this is because areas of glass window and lateral wall surfaces are increased.

The insulation layer is an effective arrangement to minimize the heat losses through the cavity outer surface. We choose four insulation thicknesses for validating the influence of insulation: 0 mm, 10 mm, 20 mm, and 30 mm. The result is shown in Figure 12; as expected, the heat transfer

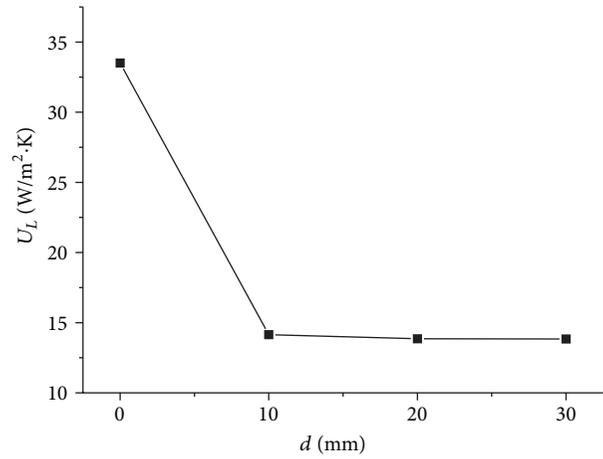


FIGURE 12: Schematic of U_L versus d ($W = 160$ mm, $h = 50$ mm, and $\theta = 45^\circ$).

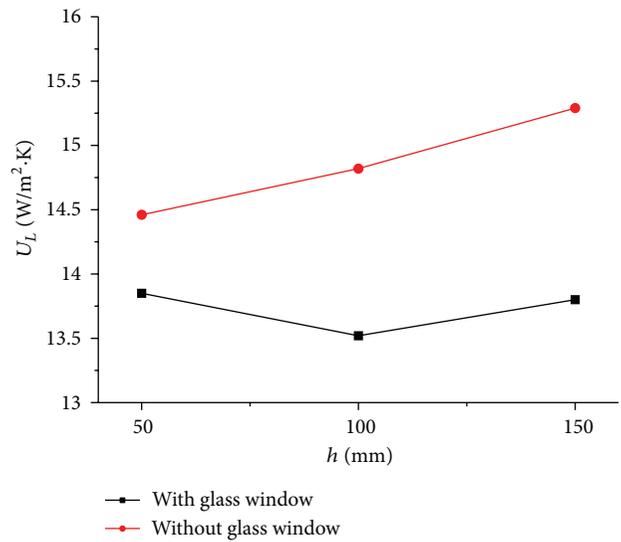


FIGURE 13: Schematic of U_L versus h ($W = 160$ mm, $d = 20$ mm).

coefficient decreases with the increase of insulation thickness. Heat losses coefficient with 10 mm insulation thickness is 58% lower than that of without insulation. But to continue increasing the thickness of insulation, the heat preservation effect improving is not obvious. Heat losses of the cavity with 30 mm insulation thickness only reduce by 2.2%, compared with 10 mm insulation thickness.

The main effect of glass window is to minimize the convection heat transfer between cavity inside and ambient air; the influence of glass window on total heat loss coefficient is shown in Figure 13. The effect of glass window is obvious; when cavity depth is 50 mm, 100 mm, and 150 mm, U_L of cavity with glass window has declined by 4.4%, 9.6%, and 10.8%, respectively. The concentrated solar beam causes an incidence angle on the outer surface of glass window that is greater than 0°, which means optical losses will be produced, so we should choose the glass window with high transmissivity and low reflectivity.

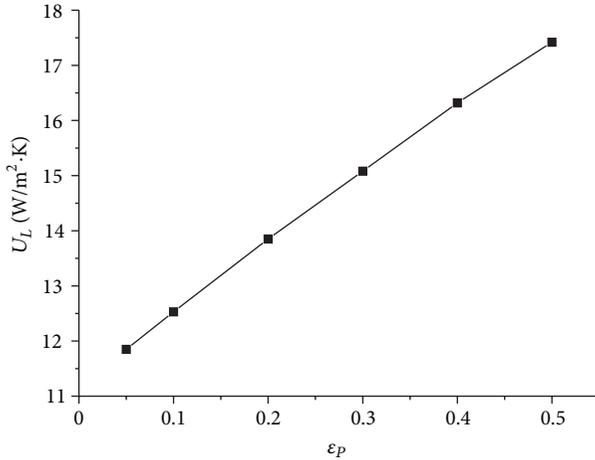


FIGURE 14: Schematic of U_L versus ϵ_p ($d = 20$ mm, $W = 160$ mm, $h = 50$ mm, and $\theta = 45^\circ$).

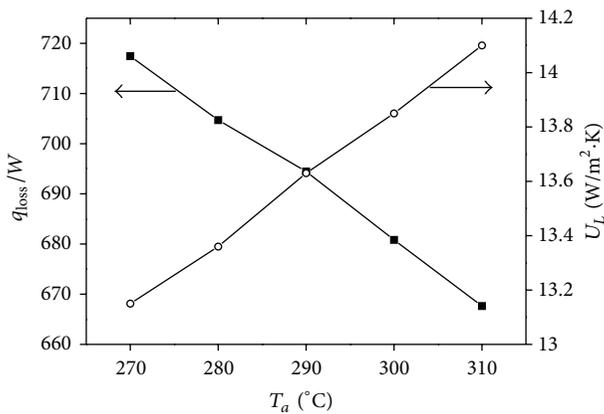


FIGURE 15: q_{loss} and U_L versus T_a ($d = 20$ mm, $W = 160$ mm, $h = 50$ mm, and $\theta = 45^\circ$).

The turning point of solar thermal utilizations is the popularization and use of selective absorption coating, which has a high absorptivity (>0.9) and low emissivity (<0.1). Since the first presentation by the experts of Israel and America in 1954, much research has been done so far [9]. The selective absorption coating can work under 550°C , which solves the bottleneck of concentrated solar energy utilization. The influence of selective absorption coating to U_L is shown in Figure 14; U_L increases nearly linearly with the increase of selective absorption coating emissivity of absorber. When the coating emissivity ϵ_p is 0.05, U_L is $11.85 \text{ W}/(\text{m}^2 \cdot \text{K})$, which is 47% lower than U_L with the 0.5 coating emissivity.

The declination of ambient temperature will lead to the increase of heat loss (see the q_{loss} curve in Figure 15); this makes sense, but U_L declines with the increase of ambient temperature. As mentioned in Section 2 of this paper, U_L can be regarded as the heat transfer coefficient between the outer surface of the cavity and the ambient air; according to the Newton cooling law, the smaller the temperature difference, the larger the heat transfer coefficient. The influence of

ambient temperature on U_L is small; when T_a increase from 270 K to 310 K, U_L just increases by 6.7%.

4. Conclusions

In this paper, a CFD performance prediction and optimum of the cavity receiver for linear Fresnel reflector concentrator have been made, considering the cavity configuration and other influence factors, and the effect on U_L is complex; main conclusions are as follows.

The most important factors which influence the heat loss coefficient are insulation and emissivity of selective absorption coating, which can reduce 58% and 47% of the heat loss coefficient, respectively; glass window plays the second position, which can reduce 10.8% of the heat loss coefficient at most when cavity depth is 150 mm. The influence of ambient temperature is small.

The heat loss coefficient shows a downtrend with the increase of inclination of side walls; the influence of cavity depth is negligible, which is consistent with the conclusion of Facão and Oliveira [18]. Among all the cavity models that are simulated in this work, the trapezoidal cavity with 100 mm cavity depth and 20 mm insulation thickness plays the best thermal performance.

Conflict of Interests

The authors declared that they have no conflict of interests related to this work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 51106091) and the Shandong University Innovation Foundations (Foundation no. 2012ZD018).

References

- [1] S. A. Kalogirou, "Solar thermal collectors and applications," *Progress in Energy and Combustion Science*, vol. 30, no. 3, pp. 231–295, 2004.
- [2] R. G. Cable, G. E. Cohen, D. W. Kearney, and H. W. Price, "SEGS plant performance 1989–1997," in *Proceedings of the International Solar Energy Conference*, pp. 445–452, June 1998.
- [3] H. Zhai, Y.-J. Dai, and J.-Y. Wu, "Investigation of concentrating solar photovoltaic/thermal system performance based on Fresnel lens," *Journal of Engineering Thermophysics*, vol. 28, no. 5, pp. 725–728, 2007.
- [4] N. Velázquez, O. García-Valladares, D. Saucedo, and R. Beltrán, "Numerical simulation of a Linear Fresnel Reflector Concentrator used as direct generator in a Solar-GAX cycle," *Energy Conversion and Management*, vol. 51, no. 3, pp. 434–445, 2010.
- [5] A. Häberle, C. Zahler, H. Lerchenmuller et al., "The Solar-mundo line focussing Fresnel collector. Optical and thermal performance and cost calculations," in *Proceedings of the International Symposium on Concentrated Solar Power and Chemical Energy Technologies*, Zürich, Switzerland, 2002.

- [6] D. T. Nelson, D. L. Evans, and R. K. Bansal, "Linear Fresnel lens concentrators," *Solar Energy*, vol. 17, no. 5, pp. 285–289, 1975.
- [7] D. Feuermann and J. Gordon, "Analysis of a two-stage linear Fresnel reflector solar concentrator," *Solar Energy Engineering*, vol. 113, p. 272, 1991.
- [8] D. Mills, "Advances in solar thermal electricity technology," *Solar Energy*, vol. 76, no. 1–3, pp. 19–31, 2004.
- [9] G. Morin, W. Platzer, M. Eck, R. Uhlig, A. Häberle, and M. Berger, "Road map towards the demonstration of a linear Fresnel collector using a single tube receiver," in *Proceedings of the 13th SolarPACES International Symposium*, Seville, Spain, 2006.
- [10] D. R. Mills and G. L. Morrison, "Compact linear fresnel reflector solar thermal powerplants," *Solar Energy*, vol. 68, no. 3, pp. 263–283, 2000.
- [11] F. J. Pino, R. Caro, F. Rosa, and J. Guerra, "Experimental validation of an optical and thermal model of a linear Fresnel collector system," *Applied Thermal Engineering*, vol. 12, pp. 1–9, 2012.
- [12] D. Chemisana and M. Ibáñez, "Linear Fresnel concentrators for building integrated applications," *Energy Conversion and Management*, vol. 51, no. 7, pp. 1476–1480, 2010.
- [13] C. Han, J. Ji, W. He, and G. Pei, "Theoretical study of parabolic trough solar concentrator," *Acta Energiæ Solaris Sinica*, vol. 30, no. 9, pp. 1182–1187, 2009.
- [14] P. L. Singh, R. M. Sarviya, and J. L. Bhagoria, "Thermal performance of linear Fresnel reflecting solar concentrator with trapezoidal cavity absorbers," *Applied Energy*, vol. 87, no. 2, pp. 541–550, 2010.
- [15] C. J. Dey, "Heat transfer aspects of an elevated linear absorber," *Solar Energy*, vol. 76, no. 1–3, pp. 243–249, 2004.
- [16] G. Li and H. Ma, "Numerical simulation of the closed cavity of high Rayleigh number laminar natural convection," *Journal of Huazhong University of Science and Technology*, vol. 21, pp. 14–17, 2004.
- [17] K. D. Masterson, "Selective surfaces for solar-thermal conversion," *Journal of Solid State Chemistry*, vol. 22, no. 1, pp. 41–49, 1977.
- [18] J. Facão and A. C. Oliveira, "Numerical simulation of a trapezoidal cavity receiver for a linear Fresnel solar collector concentrator," *Renewable Energy*, vol. 36, no. 1, pp. 90–96, 2011.

Research Article

Factors That Influence Failure Behaviour and Remaining Useful Life of Mining Equipment Components

Mark Ho and Melinda Hodkiewicz

The University of Western Australia, Australia

Correspondence should be addressed to Mark Ho; mho@mech.uwa.edu.au

Received 10 April 2013; Revised 13 June 2013; Accepted 16 June 2013

Academic Editor: Yixiang Huang

Copyright © 2013 M. Ho and M. Hodkiewicz. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mobile mining equipment often operates in harsh environments characterised by remote locations and highly variable rock and operating conditions. This research explores the hypothesis that the failure behaviour of mining equipment is influenced by the physical properties of the ore and waste. We describe a method of examining this relationship via data mining on maintenance records and apply it to the hydraulic cylinders of two classes of earthmoving mobile equipment. Failure data for the analysis are drawn from maintenance work orders from 14 sites mining for haematite iron, nickel sulphide, and coking and thermal coal. The results show that the distributions of the estimated life parameters for hydraulic cylinders on earthmoving equipment are distinctly different for haematite iron, coal, and nickel sulphide sites. Analysis of the relationship between selected physical properties identified the influence of rock impact hardness number, abrasion index, and absolute hardness of the ore as significant factors for these hydraulic cylinders. Their effects are significant when parameters are considered in combination, for example, rock impact hardness number and abrasion index, and vary according to the cylinder type and asset class. The engineering implications of these results are considered with respect to known failure modes of the cylinders.

1. Introduction

Reliability is of key importance in the mining industry, and there have been considerable efforts put in over the last 10 years to improve the reliability of assets. Inherent reliability is a function of design, but the achieved reliability is influenced by a variety of operating circumstances. These include organisational processes such as how the asset should be maintained, culture in how it is operated, site factors such as asset age profile and functional expectations, and environmental conditions primarily associated with ore type and location. The aim of this work is to test the hypothesis that the failure behaviour of components is influenced by the operating context, specifically the characteristics of the ore body, and to identify factors that might be driving that behaviour. This is of concern to the mining industry as many deposits of note are situated in locations with extreme environmental conditions such as dust and temperature.

2. Background

Reliability databases are used in the oil and gas, nuclear, chemical, and electronics industries for reliability assessments at the asset design stage, for benchmarking, and for reliability improvement programs. No such database exists in the mining industry; however, since 2010, CRC Mining has sponsored a research project to develop a framework for the sharing of failure data across organisations. In many cases in the mining industry, operators have only one or two high production assets such as excavators, shovels, crushers, or mills. With these small populations and limited failure data, it is a challenge to identify failure modes and predict failure behaviour. This impacts the engineer's ability to develop an effective maintenance program and accurately predict lifecycle costs and availability.

A shared database provides access to a larger number of failure events, and engineers can examine a wider range of

failure modes than those that they may have experienced at their site. Beyond the availability of compiled data and the availability of standard reliability statistics, the question arises as to how to use the data in site-specific reliability assessments. If data are sought for a single-asset class such as a make and model of haul truck in haematite iron ore, then is it appropriate to compile data from identical haul trucks moving gold, iron ore, nickel, coal, and other commodities? Or should there be a process to tailor the analysis depending on the particular commodity and other site-specific characteristics? If so, do we need to establish what effect(s) site-specific characteristics may have on the failure behaviour and determine how to enable the use of “compiled” failure data on specific sites? Can we then leverage the knowledge from a larger number of failures yet still be able to tailor the data using site-specific factors?

The proportional hazards model (PHM) is originally developed by Cox [1] to determine significant factors and their magnitude of impact on the measure of interest. In the field of reliability engineering, particularly for mining equipment, PHM modelling work has been performed to determine the effects of condition monitoring covariates on remaining useful life. Examples include studies where data have been used to identify significant condition monitoring covariates affecting the times to failure of transmissions [2] and wheel motors [3] on mine trucks, rail wagon bearings [4], circulating pumps in a petrochemical plant [5], and diesel engines [6]. The PHM has been also used to evaluate the effect of noncondition monitoring covariates on times to failure. Examples of these include evaluating the effect on times to failure of the following: cable material choice for power cables in underground loader hauler dumpers [7], a combination of internal (condition monitoring) and external (organisational and maintenance) covariates affecting hydraulic jack units in underground loader hauler dumpers [8], effect of design characteristics and the impact of strike action on aluminium reduction cells [9], and operating characteristics of hydrocarbon pipelines. Work using related methods (proportional covariate model) in conjunction with accelerated life tests has been used to estimate hazard rates in mechanical systems [10]. The use of the PHM will be adopted in this study.

Work with other modelling techniques has also been performed such as using linear or quadratic regression to determine the relationship between engine health and engine performance for gas turbines [11], the effect of design and operating conditions on times to failure for pumps in an oil refinery using MANOVA and multivariate statistics [12], and the effect of pit and operating characteristics on the incidence of tyre failure using the F -test [13].

From an engineering perspective, it is a plausible hypothesis that an asset operating in a highly silicified gold mine will demonstrate different failure behaviour than that of the same asset in a bituminous coal operation. However, there has been very little published academic work to look at this. In part, this can be explained by the challenges in obtaining data across a sufficiently large population of sites mining for different ore types. This study uses the data collected in the CRC Mining study to explore the influence of external

covariates by analysing individual subsystems across similar assets operating in multiple and diverse environments.

The challenge of dealing with data from different sites is compounded by the generally poor quality of reporting on failures generally found in computerised maintenance management systems and the considerable time required to clean the data [14, 15]. While it remains a possibility that the large original equipment manufacturers, especially those engaged in maintenance and repair contracts, have the data to perform this analysis, any work that has been done is not in the public domain. Given the absence of published work, a particular focus of this paper is setting out an appropriate methodology for preparation and analysis of the data.

3. Methodology

This section sets out the approach adopted to test the hypothesis that the failure behaviour of assets is influenced by the operating context, specifically the characteristics of the ore. The process involves the following key steps.

- (1) Select asset class, subsystem, and components.
- (2) Define “failure” and cleanse data.
- (3) Identify and characterise covariates of potential interest.
- (4) Examine characteristics of the dataset.
- (5) Explore effects of covariates on failure behaviour.
- (6) Select PHM models using model selection criteria.

3.1. Select Asset Class, Subsystem, and Components. With industry assistance, we compiled a confidential failure database. It includes data from 14 mine sites, 4 commodities, and 6 organisational entities. The largest set of data is for heavy mobile equipment, including front-end loaders (FELs), dozers, shovels, graders, scrapers, drills, and dump trucks. Some models of equipment are used across all of the different commodity groups and at many of the sites. Mobile equipment has a number of components with similar functions. An example is the hydraulic cylinders used on FELs and dozers in surface mining operations. The two classes of hydraulic cylinders of interest are the following: (1) lift cylinders used in lifting or lowering a boom to which the implement is attached and (2) tilt cylinders responsible for changing the angle of the implement.

3.2. Define Failure and Cleanse Data. For lift and tilt cylinders, events classified as failures include (1) major leaks preventing the equipment from further operation and (2) loss of function to causes other than leaks. In responding to these failure events, two assumptions are made, the first assumption is that repair restore the component to as good as new and that when a replacement cylinder is installed, a new (not refurbished) cylinder. In order to test whether the failures are independent and identically distributed, an examination of the cumulative failures versus time plots is performed. This is described in a subsequent section. Events involving these cylinders that are NOT classified as

TABLE 1: Weibull parameters and statistics of dataset.

Type	β (Fe)	β (coal)	β (Ni)	Number of failures Fe/coal/Ni	Number of suspensions Fe/coal/Ni
FEL lift cylinders	1.36	1.4	1.95	33/50/8	41/30/14
FEL tilt cylinders	1.07	1.25	1.32	30/52/6	41/32/12
Dozer lift cylinders	1.67	1.77	1.22	42/349/26	23/242/25
Dozer tilt cylinders	1.1	1.36	1.3	40/146/12	11/58/22

failures include (1) preventative replacement of cylinders (a suspension is recorded), (2) accidental damage (a suspension is recorded), (3) minor leaks under observation but not preventing further operation, and (4) repairs or adjustments to cylinder attachments such as mountings, pins, valves, hoses, and rods. In total, there were a total of 1342 records of interest, including 551 suspensions and 791 failure events.

3.3. Identify and Characterise the Covariates of Potential Interest. The external covariates examined in this paper include values that can be measured via analysis of data provided, or examining the conditions present at each mine site. The covariates can be broadly classified into two categories: characteristics of the mine and characteristics of the operation and maintenance of each piece of equipment. Covariates examined in this study originating from the characteristics of the mine include ore properties such as rock impact hardness number (RIHN), absolute hardness, density, abrasive index, and unconfined compressive strength, quartile indicators for concentrations of sodium, potassium and silicon compounds, and 3 ore-type indicators corresponding to each ore type. Covariates examined originating from the characteristics of the operation and maintenance include the equipment size, noncompliance to mechanical lubrication schedules, inspection intervals, and duty level of the equipment. It is important to note that this list of covariates is not a complete list of all possible covariates; however, the study was restricted to covariates whose values were directly observable via provided data or published reference material.

In instances where the exact value for that mine site was not known, a value in the middle of the range for that ore type was assumed for each group. Data were compiled (in order of preference) from the mine site's geotechnical database (where available), estimated parameters from equipment manufacturers, or commonly accepted ranges from published reference material [16].

3.4. Examine Characteristics of the Dataset

3.4.1. Examination of Weibull Parameters of Stratified Population. Table 1 shows the shape parameter (β) of the Weibull distribution for each population of FEL and dozer cylinders. It can be seen that the β values for lift cylinders for both types of equipment are similar for Fe and coal but distinctly different for Ni. For the tilt cylinders, the β values are similar for coal and nickel but slightly lower for Fe.

Other items of note in Table 1 are the low population of failures and high ratio of suspensions to failures (2:1) for cylinders on both FELs and dozers in nickel sulphide

operations. Cylinders for FELs in haematite operations also exhibit high suspension to failure ratios with 1.36 suspensions per failure. This results in wide confidence intervals when estimating parameters; however, this is somewhat mitigated by the fact that both the combined populations as well as most populations stratified by commodity all have statistically significant numbers of failures greater than 30. An exception to this is cylinders for FELs mining for nickel sulphide for which there are sparse failure data (8 lift cylinder failures and 6 tilt cylinder failures) which are outnumbered by suspensions (14 lift cylinder suspensions and 12 tilt cylinder suspensions).

3.4.2. Verification of Independent and Identically Distributed Data. Verification that datasets are independent and identically distributed (IID) is performed by plotting plots of times between consecutive failures and cumulative failure times against failure number [17]. Identical distributions can be verified by ensuring a single linear trend in the cumulative failure times.

Figures 1(a), 1(b), 1(c), and 1(d) show the cumulative failure times between consecutive failures. There are no significant changes in slope that would indicate an increasing or decreasing hazard rate. This absence of any significant change in slope of these curves does not guarantee that repairs are as good as new, but it is indicative of an absence of factors that indicate increasing or decreasing life of subsequent failures or lack of independence of the dataset.

3.5. Explore Effects of Covariates on Failure Behaviour. The relationships between ore characteristics and failure behaviour are explored statistically using proportional hazards modelling (PHM). For each cylinder type on each mobile equipment type, it is possible to construct a total of 15 one-covariate models (models with a single explanatory covariate) and 105 two-covariate models (models with two explanatory covariates). Due to the existence of only 3 ore types in this study, it is only possible to construct models with, at maximum, 2 covariates (corresponding to the 2 degrees of freedom available). This process was applied to both types of cylinders on each of the two types of mobile equipment leading to a total of 60 one-covariate models and 420 two-covariate models. The 480 models developed cover all possible combinations of the given covariates. All models were then tested for statistical accuracy and significance by the use of criteria in a model selection process.

The outputs from the proportional hazards model include the coefficients for each covariate, the standard error associated with each coefficient, and the baseline hazard rate.

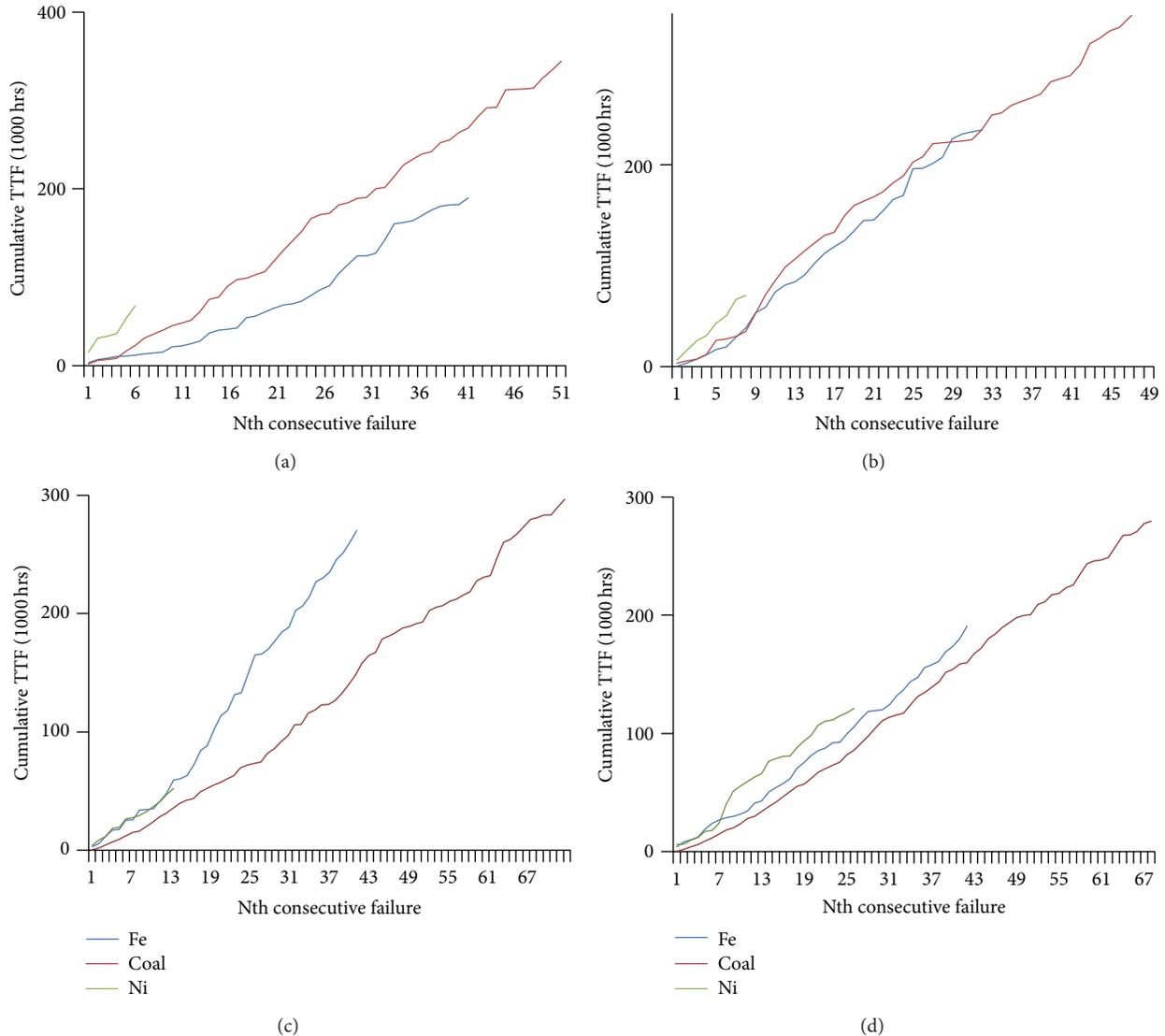


FIGURE 1: (a) FEL lift cylinder cumulative TTF. (b) FEL tilt cylinder cumulative TTF. (c) Dozer lift cylinder cumulative TTF. (d) Dozer tilt cylinder cumulative TTF.

The coefficient of each covariate is proportional to the impact that covariate has on the hazard rate. Thus, positive coefficients indicate that the covariate increases the hazard rate while negative coefficients indicate that the covariate decreases the hazard rate. The baseline hazard rate can be modified by the calculated coefficients, and observed values of each covariate to obtain the modified hazard function and survival curves. A parametric fit is performed to determine the Weibull parameters for the theoretical distribution that best fits each modified survival curve. These are the predicted parameters that each PHM would predict for each commodity. The actual parameters are obtained by performing a standard Weibull analysis directly to the dataset of each commodity.

3.6. Select PHM Models Using Model Selection Criteria. In order to assess the suitability of each PHM model, we consider the following factors. The accuracy for the model

is assessed by the average percentage deviation between the predicted and the actual MTTFs for all commodities. In this case, the lowest percentage deviation corresponds to the highest accuracy. Other measures of accuracy were considered, such as analysis of goodness of fit against the full Weibull curve for each commodity; however, the use of a point estimate (MTTF) comparison was selected due to its widespread use in industry. The significance of a covariate is assessed by comparing the magnitude of the coefficient with the magnitude of the standard error associated with that coefficient. Under this method, a higher ratio of coefficient to standard error indicates higher significance for that coefficient. The aim is to identify only models whose covariates most impact the hazard rate of the asset. Selection criteria for the models are as follows:

- (i) high accuracy (less than 10% deviation between actual and predicted MTTF),

- (ii) each coefficient having a ratio at least 1.28 times the standard error (80% confidence),
- (iii) hypothesis test for proportionality failing to reject the proportional hazards assumption (PHA). (A P value of 0.1 is used for this hypothesis test [18]. The null hypothesis states that the data are proportional, while the alternative hypothesis states that the data are not proportional.)

Due to the low complexity of the models (<2 covariates), it was not deemed necessary to consider the Akaike information criterion. Models that meet the model selection criteria are collated for similar subsystems across similar equipment types. Significant covariates or combinations of significant covariates are identified for further discussion and investigation.

4. Results

4.1. Relationships between Ore Characteristics and Failure Behaviour. A visual representation of the cylinders life in operations mining for different commodities was obtained by the Bayesian inference [19]. These representations show that the distributions of the Weibull scale parameter (η) of different commodities are different. The FEL lift cylinders for nickel sulphide operations in Figure 2(a) are significantly different from those in coal and haematite operations. The FEL tilt cylinders for all commodities in Figure 2(b) are different from each other, and dozer tilt cylinders for haematite operations in Figure 3(b) are different from those in coal operations. It is important to note that these distributions are not frequency distributions of the times to failure.

The PHM was applied to the cleansed data, and the selection process was used to identify appropriate models for lift and tilt cylinders on FELs and dozers. Table 2 shows a list of models that met the model selection criteria for accuracy and significance. A P value for the PHA hypothesis test is also included for each covariate. The P values were obtained using the `cox.zph` function in the survival package in *R* [20].

Table 2 is organised as follows. Columns 3 and 4 show the coefficient values for the covariates listed in Column 2. Columns 5 and 6 show the ratios of coefficient magnitude to standard error with higher ratios indicating stronger significance. The P value for PHA hypothesis testing is shown in Column 7 with values greater than 0.1, indicating a failure to reject the PHA. The deviation between the actual and predicted value for the MTTF is shown in Column 8. The MTTF calculation takes into account both β and η values [21] as follows:

$$MTTF = \int_0^{\infty} R(t) dt = \eta \left(\frac{1}{\beta} + 1 \right). \quad (1)$$

Table 2 shows that only 8 models from the 480 tested have significant relationships. These relationships apply to tilt and lift cylinders of FELs and lift cylinders of dozers. Most of the covariates are related to ore properties rather than ore types with absolute hardness and/or abrasion index appearing in four of the 8 models.

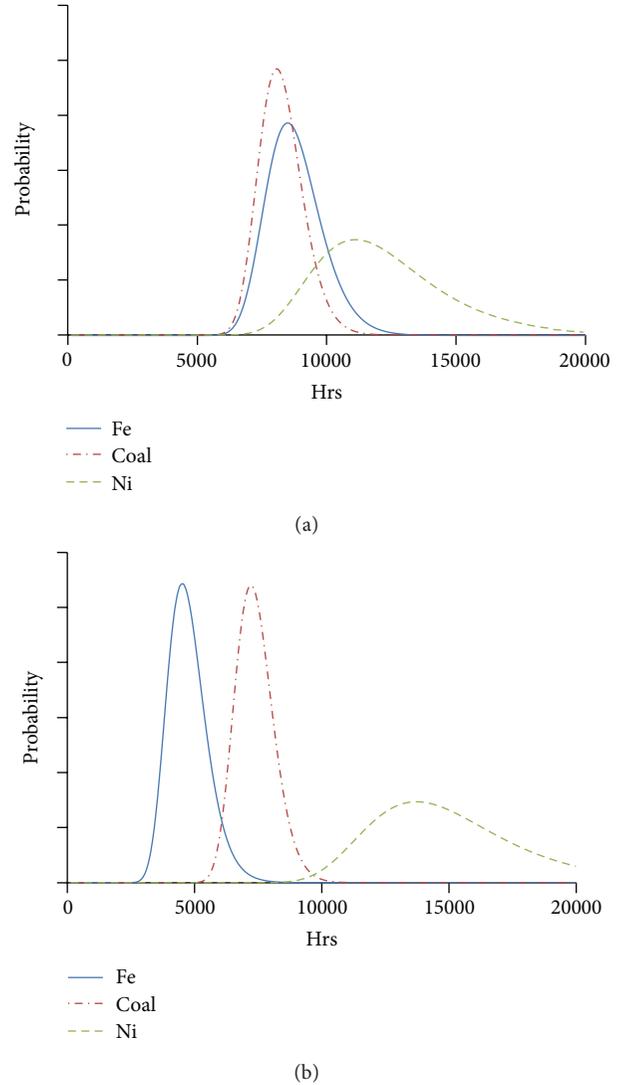


FIGURE 2: Distribution of η for FEL lift cylinders and FEL tilt cylinders.

The three models for dozer lift cylinders (Table 2: Rows 7 to 9) violate the proportionality assumption with their P values for the PHA hypothesis test (Table 2: Column 7) less than 0.1.

Of the models for FEL tilt cylinders, the models for abrasive index and absolute hardness (Table 2—Row 5) and ore indicators for coal and nickel sulphide (Table 2: Row 6) have P values for the PHA hypothesis tests close to the threshold of 0.1 (0.11, 0.11, and 0.124, resp.). The covariate models created for dozer tilt cylinders did not yield any relationships of significance as no covariate selection combination was able to meet the accuracy criteria.

Table 3 summarises the list of parameters, and their occurrences in models are identified as significant. The “number of occurrences” is the number of times that covariate was present in models that met the model selection criteria. The “effect on hazard rate” uses the sign of the calculated coefficient to determine whether the covariate positively or adversely impacts the hazard rate.

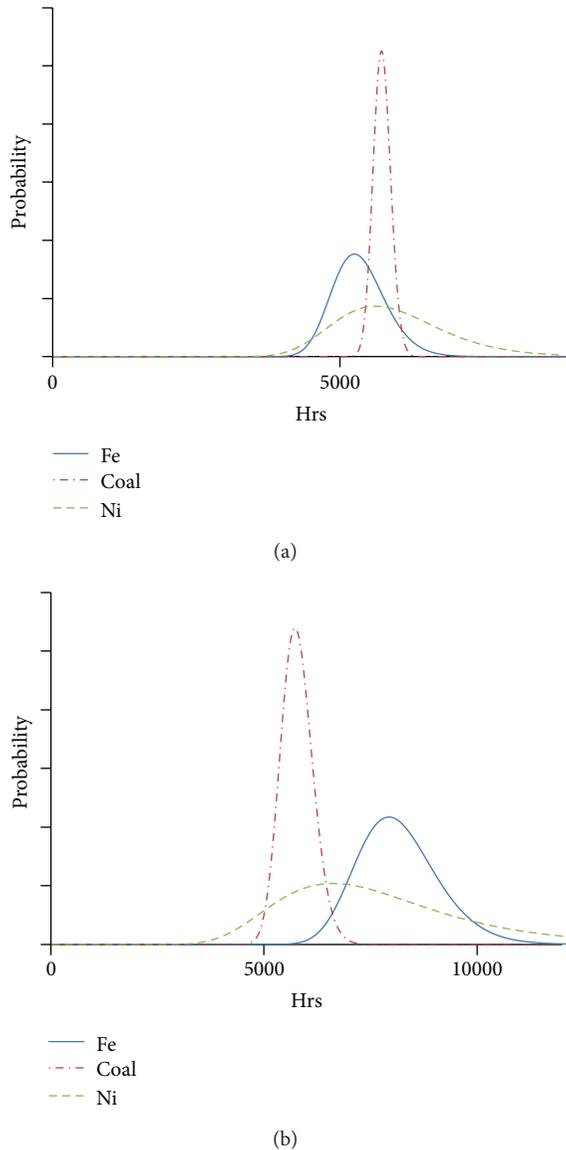


FIGURE 3: Distribution of η for dozer lift cylinders and dozer tilt cylinders.

In order to determine the magnitude of impact of each covariate relative to every other covariate, we normalise the measured covariate values in order to obtain a unit standard deviation. Table 4 shows the normalised coefficients for all models meeting the model selection criteria. It can be seen that the normalised covariates with the largest impact per standard deviation change in covariate values are the abrasive index and absolute hardness with ranges of 5.86 to 7.66 and -6.05 to -7.9 , respectively. This compares with covariates with less impact such as Si, with a normalised coefficient of 0.32.

5. Discussion

5.1. Analysis of Distributions and Weibull Parameters Based on Commodity Type. Examination of the distributions shown in Figure 2(a) to Figure 3(b) shows the following.

- (i) FEL Lift Cylinders. Cylinders in nickel sulphide operations have higher η and MTTF than those in haematite and coal operations. Coal and haematite operations have cylinders with comparable lifetime distributions.
- (ii) The high number of suspensions and low number of failures for FEL lift cylinders in nickel sulphide operations resulted in higher uncertainty in parameter estimation as evidenced by the wide dispersion of the distribution. Although haematite operations also have high numbers of suspensions, the 33 failures points were sufficient to give a distribution dispersion similar to that of the coal operations. This was also evident in FEL tilt cylinders.
- (iii) FEL Tilt Cylinders. Cylinders in nickel sulphide operations have higher η and MTTF than those in coal operations. Cylinders in haematite operations have the lowest η and MTTF out of all the three commodities.
- (iv) Dozer Lift Cylinders. Cylinders in haematite operations have lower η and MTTF than those in coal operations. The η and MTTF of nickel sulphide operations have a large variance and are not conclusively different from those in haematite or coal operations.
- (v) Dozer Tilt Cylinders. Cylinders in haematite operations have higher η and MTTF than those in coal operations. The η and MTTF of nickel sulphide operations have a large variance and are not conclusively different from those in haematite or coal operations.
- (vi) The variance of the distributions for nickel sulphide and haematite iron operations is larger than that for coal operations. This is due to the availability of a larger dataset for coal operations, leading to higher certainty (probability) of parameter estimates as well as the high number of suspensions for nickel sulphide operations.

Tables 3 and 4 show that absolute hardness and abrasive index were parameters that had the highest rate of occurrence in all of the significant models. They were also the parameters with the highest impact per standard deviation change of observed covariate values.

5.2. Exploring the Relationship to Specific Physical Properties.

The covariate models developed in this study identify, from a selected population of physical properties, those properties that impact the reliability of lift and tilt hydraulic cylinders on earthmoving mobile equipment. These properties are the abrasive index and absolute hardness of the mineral. To explore the relationship between what is observed in the data and what might be occurring in the physical environment, a review of common failure causes of hydraulic cylinders was conducted. The most common causes of cylinder failure (excluding rod breakage or alignment issues) are contamination, bearing and seal damage, chemical or heat degradation, and structural damage.

The coefficient calculated in this study for absolute hardness shows a decrease in hazard rate with absolute hardness.

TABLE 2: Models meeting the model selection criteria for tilt and lift cylinders on dozers and FELs.

Set	Covariates	Coefficient 1	Coefficient 2	Ratio Coef-SE1	Ratio Coef-SE2	PHA P value	MTTF deviation (%)
FEL-Lift	Abrasive index and absolute hardness	67.86	-0.22	1.77	1.8	0.597 0.587	3.81%
FEL-Lift	Nickel sulphide	-0.64	NA	1.60	NA	0.65	10%
FEL-Tilt	UCS and silicon %	-0.014	0.37	2.71	1.5	0.666 0.934	5.49%
FEL-Tilt	Abrasive index and absolute hardness	89.67	-0.29	2.16	2.21	0.11 0.11	5.3%
FEL-Tilt	Coal and nickel sulphide	0.42	-0.61	1.82	1.34	0.951 0.124	3.89%
Dozer-Lift	Absolute hardness and haematite	-0.046	2.82	2.84	3.02	<0.01 <0.01	6.3%
Dozer-Lift	Abrasive index and RIHN	8.25	-0.093	3.13	2.98	<0.01 <0.01	4.64%
Dozer-Lift	RIHN and haematite	-0.066	1.10	2.83	3.12	<0.01 <0.01	5.76%
Dozer-Tilt	No suitable models	NA	NA	NA	NA	NA	NA

TABLE 3: Covariates used in PHM that meet the model selection criteria.

Parameter	Number of occurrences	Effect on hazard rate
Absolute hardness	2	Higher scratch resistance decreases the hazard rates
UCS	1	Higher UCS decreases the hazard rate
Si	1	Higher silicon % increases the hazard rate
Abrasive index	2	Higher AI increases the hazard rate
Ore indicators	3	Coal mines have higher hazard rates Nickel mines have lower hazard rates Haematite mines have lower hazard rates for FELs but higher hazard rates for dozers
Most common model	Absolute hardness and abrasive index	

We postulate that the effect of a higher absolute hardness, which is the ability of a rock to resist scratching, decreases the ability of the material to create dust-sized particles. This results in a less dusty environment, which increases the reliability of a hydraulic cylinder due to less dust contamination of the hydraulic fluid or bearings. An increase in abrasiveness of the material appears to influence the reliability of the lift and tilt cylinders in some assets. In a high-dust environment, the ability of the particles to cause abrasive wear once in contact with the cylinder surfaces, bearings, or seals will accelerate deterioration.

The coefficient values and specific observed values for any commodity can be used to determine the magnitude of

impact on the hazard rate for that commodity. In the case of the abrasive index and absolute hardness model, the coefficient values are 67.86 and -0.22, respectively (Table 2: Row 2, Columns 3 and 4). A FEL in a nickel sulphide mining operation with an abrasive index of 0.04 and absolute hardness of 15 would experience a modifying factor to the baseline hazard rate of 0.55 (45% lower than the baseline hazard rate). The baseline hazard rate represents the hazard rate of a fictional ore with an abrasive index of 0 and absolute hardness of 0. Similarly, a FEL in a coal mining operation with a low abrasion index of 0.01 and low absolute hardness of 2 would experience a modifying factor to the baseline hazard rate of 1.27 (27% higher than the baseline hazard rate). This implies that the lift cylinder of a FEL in a nickel mining operation experiences a hazard rate that is 43% of the hazard rate experienced by an equivalent FEL in a coal mining operation.

It is possible to apply the modifying factor for any specific commodity to obtain the commodity-specific hazard rate, survival curves, and associated Weibull parameters (β and η). The practical use of these coefficients is to extract operation-specific parameters from pooled data. These operation-specific parameters such as the MTTF, β , or η can be used to benchmark existing operations or provide insight when developing a mine site with differing operating conditions.

5.3. Organisational and Maintenance Properties. The analysis examined covariates representing the maintenance and operating conditions of the organisation including equipment size, equipment duty level, inspection intervals, and compliance to mechanical lubrication schedules. As a proxy, the scheduled compliance to a 250-hour engine service task was used. This maintenance work does not cover the cylinders specifically, but it is an indication of organisational commitment to scheduled maintenance. There are no direct data on scheduled maintenance compliance for the lift and tilt cylinders. Of these covariates only, the equipment duty level

TABLE 4: Normalised coefficients for models meeting model selection criteria.

Set	Covariates	Normalised coefficient 1	Normalised coefficient 2
FEL-Lift	Abrasive index and absolute hardness	5.86	-6.05
FEL-Lift	Nickel sulphide	-0.64	NA
FEL-Tilt	UCS and silicon %	-0.59	0.32
FEL-Tilt	Abrasive index and absolute hardness	7.66	-7.9
FEL-Tilt	Coal and nickel indicators	0.42	-0.61

and compliance to mechanical lubrication schedules were found to be significant; however, they were not as significant as the ore properties. Equipment that was used for lower numbers of hours a day experienced higher hazard rates suggesting that deterioration due to time or exposure still occurs even when the equipment is not in use. Organisations with higher compliance to scheduled maintenance also experienced lower hazard rates. Further work in respect to developing and using organisational indicators as covariates is underway.

5.4. Sources of Uncertainty. There are a number of sources of uncertainty in this study due to the nature of failure data and also the simplifications inherent in compressing complex physical property distributions into a set of numbers. The original datasets contain data for all failure events, and there is considerable variability in the content of the fields and how data are represented. Extraction of failure and preventative replacement events was performed using an in-house data cleansing tool with the outputs cross-referenced against external data sources such as maintenance plans and external maintenance contracts. Incorporating data from these external data sources was done as follows.

Maintenance and repair contracts (MARC)s are specified time intervals wherein major maintenance and replacement activity is performed by an external contractor (most often the equipment vendor). Maintenance work performed under these contracts is recorded in the record system of the contractor rather than that of the mining company; it is therefore not included in the dataset. Data points immediately following a maintenance contract are removed as they also include times to failure/replacement of all maintenance work performed under the MARCs. Times between preventative replacements are compared against replacement intervals specified in the mining company's maintenance plan. Events marked as preventative replacements occurring significantly sooner than the planned replacement interval are treated as failures unless there is other evidence indicating that they were preventatively performed.

The dataset used includes 5 cases in which the number of failures is less than the number of suspensions. This can result in wider confidence intervals for both η and β values. Alternative methods of parameter estimation such as the Bayesian inference, as shown in Figure 2(a), or hypothesis testing can be used [22]. In addition, in two cases for nickel, there are low data populations resulting in poor fit. More data are being sought for nickel in order to improve confidence in estimated parameters.

5.5. Further Research. A work is currently underway to extend the range of components and asset classes under analysis. This will include mechanical, electrical/electronic, hydraulic, and structural subsystem components and asset classes from both fixed-plant and mobile equipments.

It was found that the assumption of proportionality was rejected for both types of cylinders for the dozer equipment class as shown in Table 2. Further work will be undertaken to more adequately model these assets by stratifying the asset class further or choosing an alternative modelling approach.

Approximations of dust levels at each mine site were inferred based on the ability of the material to resist fragmentation. This may not hold true due to other factors and dust-creation mechanisms such as the use of dust suppression, blasting techniques, presence of clay or other fine-grained materials, and nonimpact-related dust-creating mechanisms. Further work will be performed in assessing dust concentrations experienced at the rock interface on a mine site-specific basis. Additional environmental parameters such as average temperatures and chemical compositions of rocks will be added to determine whether other extreme environmental conditions have a significant influence on equipment failure.

It has been noted that the use of purely physical properties as covariates may not be sufficient to explain the failure behaviour of equipment. Additional covariates associated with organisational processes, site factors, and environmental conditions can be added as they become available. Organisational processes to be added include the culture in which the asset is operated (including functional role and level of loading) and site factors such as asset age profile and pit characteristics (e.g., single/double benching, blasting characteristics) that are controllable by the organisation. Maintenance factors to be added in future work include the effect of maintenance, inspection, or condition monitoring activities specific to the component. The framework supports the extension of the number of covariates to n -parameter models.

6. Conclusion

This study demonstrates that the distributions of the estimated life parameter for tilt and lift cylinders on dozers and FELs are distinctly different in location and shape for haematite iron, coal, and nickel sulphide operations. Further work is underway to extend the investigation to other components and assets.

The analysis uses proportional hazards modelling to extract context-specific information from compiled data.

The choice of model and the covariates used in the model also give an insight into the factors that influence the reliability of the asset and their relative magnitudes. Analysis of the relationship between selected physical properties identified the influence of the abrasive index and absolute hardness of the ore as significant factors influencing the lifetime of tilt and lift cylinders.

It was also found that the assumption of proportionality of hazard rates was found not to hold for a specific asset class subsequent to performing the appropriate hypothesis test. This resulted in the rejecting of all models for cylinders on the dozer equipment class.

This work is significant because of the questions it raises about factors such as ore properties that may influence asset component failure. With this understanding comes the opportunity to reduce failure rates and tailor equipment selection, operation, and maintenance activities to specific sites. This will be explored in future papers.

References

- [1] D. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society*, vol. 34, no. 2, pp. 187–220, 1972.
- [2] D. Lin, D. Banjevic, and A. K. S. Jardine, "Using principal components in a proportional hazards model with applications in condition-based maintenance," *Journal of the Operational Research Society*, vol. 57, no. 8, pp. 910–919, 2006.
- [3] A. K. S. Jardine, D. Banjevic, M. Wiseman, S. Buck, and T. Joseph, "Optimizing a mine haul truck wheel motors' condition monitoring program: use of proportional hazards modeling," *Journal of Quality in Maintenance Engineering*, vol. 7, no. 4, pp. 286–301, 2001.
- [4] A. Ghasemi and M. R. Hodkiewicz, "Estimating mean residual life for a case study of rail wagon bearings," *IEEE Transactions on Reliability*, vol. 61, no. 3, pp. 719–730, 2012.
- [5] P. J. Vlok, J. L. Coetzee, D. Banjevic, A. K. S. Jardine, and V. Makis, "Optimal component replacement decisions using vibration monitoring and the proportional-hazards model," *Journal of the Operational Research Society*, vol. 53, no. 2, pp. 193–202, 2002.
- [6] A. K. S. Jardine, P. Ralston, N. Reid, and J. Stafford, "Proportional hazards analysis of diesel engine failure data," *Quality and Reliability Engineering International*, vol. 5, no. 3, pp. 207–216, 1989.
- [7] D. Kumar, B. Klefsjö, and U. Kumar, "Reliability analysis of power transmission cables of electric mine loaders using the proportional hazards model," *Reliability Engineering and System Safety*, vol. 37, no. 3, pp. 217–222, 1992.
- [8] B. Ghodrati, U. Kumar, and F. Ahmadzadeh, "Remaining useful life estimation of mining equipment: a case study," in *Proceedings of the International Symposium on Mine Planning and Equipment (MPES '12)*, 2012.
- [9] J. Kalbfleisch, "Measuring the impact of an intervention on equipment lives," *Canadian Journal of Statistics*, vol. 10, no. 4, pp. 237–259, 1982.
- [10] Y. Sun, L. Ma, J. Mathew, W. Wang, and S. Zhang, "Mechanical systems hazard estimation using condition monitoring," *Mechanical Systems and Signal Processing*, vol. 20, no. 5, pp. 1189–1201, 2006.
- [11] Y. G. Li and P. Nilkitsaranont, "Gas turbine performance prognostic for condition-based maintenance," *Applied Energy*, vol. 86, no. 10, pp. 2152–2161, 2009.
- [12] M. Braglia, G. Carmignani, M. Frosolini, and F. Zammori, "Data classification and MTBF prediction with a multivariate analysis approach," *Reliability Engineering and System Safety*, vol. 97, no. 1, pp. 27–35, 2012.
- [13] P. F. Knights and A. L. Boerner, "Statistical correlation of off-highway tire failures with openpit haulage routes," *Mining Engineering*, vol. 53, no. 8, pp. 51–56, 2001.
- [14] B. Veron and M. R. Hodkiewicz, "Rapid availability model development using time delay accounting," in *Proceedings of the ICOMS Asset Management Conference*, Sydney, Australia, 2009.
- [15] K. Unsworth, E. Adriasola, A. Johnston-Billings, A. Dmitrieva, and M. Hodkiewicz, "Goal hierarchy: improving asset data quality by improving motivation," *Reliability Engineering and System Safety*, vol. 96, no. 11, pp. 1474–1481, 2011.
- [16] G. Australia, "Geochemical Atlas of Australia," 2013, <http://www.ga.gov.au/energy/projects/national-geochemical-survey/atlas.html>.
- [17] D. M. Loutit, R. Pascual, and A. K. S. Jardine, "A practical procedure for the selection of time-to-failure models based on the assessment of trends in maintenance data," *Reliability Engineering and System Safety*, vol. 94, no. 10, pp. 1618–1628, 2009.
- [18] P. M. Grambsch and T. M. Therneau, "Proportional hazards tests and diagnostics based on weighted residuals," *Biometrika*, vol. 81, no. 3, pp. 515–526, 1994.
- [19] M. Ho and K. Shen, "Analysis of failure data using proportional hazards and Bayesian inference," in *Proceedings of the 6th International Meeting on Mine Equipment Maintenance Meeting (MANTEMIN '11)*, Antofagasta, Chile, 2011.
- [20] "The Comprehensive R Archive Network," 2013, <http://cran.r-project.org/>.
- [21] M. Rausand and A. Hoyland, *System Reliability Theory: Models, Statistical Methods and Applications*, John Wiley & Sons, Hoboken, NJ, USA, 2nd edition, 2004.
- [22] J. Ansell and M. Phillips, *Practical Methods for Reliability Analysis*, Oxford Statistical Science Series, O.S. Publications, 1994.