



## **University of Huddersfield Repository**

Tang, Xiaoli

Investigation of Data Compression Methods for Intelligent Machine Condition Monitoring

### **Original Citation**

Tang, Xiaoli (2020) Investigation of Data Compression Methods for Intelligent Machine Condition Monitoring. Doctoral thesis, University of Huddersfield.

This version is available at <http://eprints.hud.ac.uk/id/eprint/35227/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: [E.mailbox@hud.ac.uk](mailto:E.mailbox@hud.ac.uk).

<http://eprints.hud.ac.uk/>

# **INVESTIGATION OF DATA COMPRESSION METHODS FOR INTELLIGENT MACHINE CONDITION MONITORING**

**Xiaoli Tang**

A thesis submitted to the University of Huddersfield in partial fulfilment  
of the requirements for the degree of Doctor of Philosophy



July 2020

---

## **COPYRIGHT**

---

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the “Copyright”) and s/he has given The University of Huddersfield the right to use such copyright for any administrative, promotional, educational and/or teaching purposes.
- ii. Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the University Library. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.
- iii. The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the “Intellectual Property Rights”) and any reproductions of copyright works, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions

---

## ABSTRACT

---

Condition monitoring (CM) delivers significant benefits to the industry by minimising breakdown losses and enhancing the safety and high-performance operation of machinery. However, the use of data acquisition systems with multiple sensors and high sampling rates leads to massive data and causes considerably high cost for purchasing and deploying hardware for data transmission, storage and processing. Hence, data compression is crucial and important to reduce the data size and speed up the calculation for the development of intelligent machine CM systems. Although data compression has received high attention in many fields, few researchers have focused on their research in the field of machine CM. Therefore, this PhD research concentrates on investigating novel and high-performance data compression algorithms according to the characteristics of one-dimensional (1D) and two-dimensional (2D) signals to solve the bottleneck of the massive data transmission, and hence improve the performance of remote and real-time machine CM systems.

The research is carried out according to a compound experimental and analytic route based on a wireless sensor network. To demonstrate the effectiveness of data compression based techniques for CM, the prototype of an intelligent wireless sensing system is developed using cost-effective micro-electromechanical systems (MEMS) accelerometers and the Bluetooth low energy (BLE) communication module. Moreover, various waveform parameters with low cost computing in time and frequency domains are investigated and identified that RMS is the most effective parameter to give good indication for the leakage in a piping system, showing that data compression via statistics is effective and thus indicates that the performance of data compression for CM highly depends on applications.

Subsequently, high-performance but high-complexity methods are proposed based on dimension reduction, sparse representation, feature extraction and advanced compressive sensing (CS) for fault diagnosis of rotating machinery with 1D or 2D signals, which have the potentials to be implemented on MEMS modules in a wireless sensor network (WSN) in future. Firstly, a compression scheme based on dimension reduction is proposed to extract the periodic characteristics of the 1D vibration signal. Recurrence plot (RP) of vibration phase space trajectory and its quantification indicators, as well as principal component analysis (PCA), are combined to realize feature extraction, compression and fault classification for a tapered roller bearing system.

Secondly, a two-step compression method is performed on 1D vibration signals based on frequency shift, adaptive sparse representation and CS is explored to overcome the problem of the large quantity of data storage for ball bearing fault diagnosis. Simultaneously, this compression method has the capability to reconstruct envelope signals with noise elimination.



Then, for 2D thermal images captured from a two-stage reciprocating compressor, the dense scale-invariant feature transform (SIFT) features indicating edge information are extracted and represented as a sparse matrix by sparse coding. The compressed features are used for the classification of six different types of faults with the support vector machine (SVM).

Finally, the advanced CS technique is exploited on pre-processing the 2D thermal images of gearboxes to realise intelligent fault classification with high accuracy of more than 99.81% by a typical deep learning algorithm, namely convolutional neural network (CNN). The CNN calculation speed is dramatically accelerated with compressed images. All these proposed approaches are evaluated by simulations and experiments, which verifies that they can reliably detect the fault types or classify different fault types with very high accuracy. Besides, the proposed data compression based intelligent CM approaches provide theoretical bases for maintenance-free CM systems because data compression can save the transmission bandwidth and power consumption for remote and real-time machine CM systems.

**Key Words:** Intelligent machine condition monitoring (CM), Data compression, Wireless sensor network (WSN), Compressive sensing (CS), Deep learning

---

## TABLE OF CONTENTS

---

<b>COPYRIGHT.....</b>	<b>I</b>
<b>ABSTRACT.....</b>	<b>II</b>
<b>TABLE OF CONTENTS .....</b>	<b>IV</b>
<b>LIST OF FIGURES .....</b>	<b>VII</b>
<b>LIST OF TABLES .....</b>	<b>XII</b>
<b>DECLARATION.....</b>	<b>XIII</b>
<b>DEDICATIONS AND ACKNOWLEDGEMENTS.....</b>	<b>XIV</b>
<b>LIST OF ABBREVIATIONS .....</b>	<b>XV</b>
<b>PUBLICATION LIST .....</b>	<b>XVIII</b>
<b>CHAPTER 1 INTRODUCTION .....</b>	<b>1</b>
1.1 Industry 4.0 and Machinery Maintenance .....	2
1.2 Machine Condition Monitoring .....	5
1.2.1 Visual Inspection.....	5
1.2.2 Oil Analysis.....	5
1.2.3 Performance Trend.....	6
1.2.4 Non-destructive Testing .....	6
1.2.5 Electrical Signals.....	8
1.2.6 Vibration and Acoustic Signals.....	8
1.3 Importance of Data Compression in Condition Monitoring .....	9
1.4 Research Aim and Objectives.....	10
1.5 Organization .....	11
<b>CHAPTER 2 LITERATURE REVIEW .....</b>	<b>13</b>
2.1 Data Compression .....	14
2.1.1 Traditional Data Compression Methods.....	15
2.1.2 Compressive Sensing .....	20
2.2 Summary .....	29
<b>CHAPTER 3 LEAKAGE DETECTION IN PIPING SYSTEMS WITH AN INTELLIGENT WIRELESS VIBRATION SENSING SYSTEM .....</b>	<b>30</b>
3.1 Vibration Mechanisms of Piping Systems.....	31
3.1.1 Microscopic Interpretation .....	31
3.1.2 Natural Frequencies .....	31

3.1.3	Turbulent Flow .....	33
3.1.4	Frictional Effects .....	33
3.2	Design of the CM System with MEMS .....	34
3.2.1	MEMS Sensor Nodes .....	34
3.2.2	Wired MEMS Real-Time CM System .....	38
3.2.3	Sensor Calibration .....	42
3.2.4	Leakage Detection in Domestic Piping Systems .....	46
3.3	Improvement of the CM System with Data Compression .....	53
3.3.1	Wireless Sensor Node Design .....	53
3.3.2	Mobile Monitoring Application Design .....	55
3.3.3	Evaluation of the Designed Remote CM System .....	57
3.4	Summary .....	63
<b>CHAPTER 4 DIMENSION REDUCTION BASED COMPRESSION IN FAULT DIAGNOSIS OF TAPERED ROLLER BEARINGS.....</b>		<b>64</b>
4.1	Introduction .....	65
4.1.1	Rolling Element Bearings .....	65
4.1.2	Tapered Roller Bearings .....	67
4.2	Compression Method.....	68
4.2.1	Recurrence Plot .....	69
4.2.2	Recurrence Quantification Analysis.....	73
4.2.3	Principal Component Analysis.....	75
4.2.4	Proposed Compression Method .....	77
4.3	Experiments.....	78
4.3.1	Experimental Setup .....	79
4.3.2	Fault Frequencies Estimation and Effect of Clearance Change .....	82
4.3.3	Results and Discussion.....	84
4.4	Summary .....	101
<b>CHAPTER 5 SPARSE REPRESENTATION BASED COMPRESSIVE SENSING IN FAULT DIAGNOSIS OF BALL BEARINGS.....</b>		<b>102</b>
5.1	Introduction .....	103
5.2	Compression Method.....	104
5.2.1	Methodology .....	104
5.2.2	Reconstruction Performance Evaluation Indicators .....	107
5.3	Simulation Study .....	109
5.3.1	Model Simulation.....	109
5.3.2	Results and Discussion.....	111
5.4	Experiments.....	117

5.4.1	Experimental Setup .....	118
5.4.2	Results and Discussion.....	120
5.5	Summary .....	130
<b>CHAPTER 6 FEATURE EXTRACTION BASED COMPRESSION OF 2D SIGNALS IN FAULT DIAGNOSIS OF RECIPROCATING COMPRESSORS.....</b>		<b>132</b>
6.1	Introduction .....	133
6.2	Compression Method.....	136
6.2.1	Dense SIFT Descriptor.....	136
6.2.2	Sparse Coding .....	137
6.2.3	Support Vector Machine .....	138
6.3	Experiments.....	139
6.3.1	Experimental Setup .....	139
6.3.2	Results and Discussion.....	142
6.4	Summary .....	149
<b>CHAPTER 7 COMPRESSIVE SENSING BASED COMPRESSION OF 2D SIGNALS FOR FAULT DIAGNOSIS OF GEARBOXES.....</b>		<b>151</b>
7.1	Introduction .....	152
7.2	Compression Method.....	154
7.2.1	Convolutional Neural Network .....	154
7.2.2	Proposed CS-based CNN .....	157
7.3	Experiments.....	158
7.3.1	Test Facilities .....	158
7.3.2	Fault Simulation .....	160
7.3.3	Test Procedure.....	161
7.3.4	Results and Discussion.....	163
7.4	Summary .....	175
<b>CHAPTER 8 CONCLUSIONS AND FUTURE WORKS.....</b>		<b>177</b>
8.1	Review of Objectives and Achievements .....	178
8.2	Conclusions .....	180
8.3	Contributions to Knowledge.....	181
8.4	Future Work .....	182
<b>REFERENCE.....</b>		<b>183</b>

## LIST OF FIGURES

Figure 1.1 The development of industrial revolutions.....	2
Figure 1.2 Maintenance strategies .....	3
Figure 1.3 Structure of the thesis .....	12
Figure 2.1 Simulated data acquisition process.....	14
Figure 2.2 Matrix representation of signal compression .....	21
Figure 2.3 Schematic diagram of the single-pixel compressive camera.....	25
Figure 2.4 DMD micromirror structure .....	26
Figure 3.1 Representation of fluid molecule motion in a pipe from a micro perspective .....	31
Figure 3.2 Pipe vibration modes: (a) longitudinal modes, and (b) circumferential modes .....	32
Figure 3.3 Simulation of turbulence at a valve .....	33
Figure 3.4 Simulated frictional effects of the fluid against the pipe wall.....	34
Figure 3.5 MEMS applications .....	34
Figure 3.6 Typical structure of a sensor node .....	35
Figure 3.7 Typical MEMS accelerometer structures: (a) piezoresistive sensor, and (b) capacitive sensor .....	35
Figure 3.8 ARM Cortex-M series products and advantages .....	36
Figure 3.9 Battery products: (a) primary batteries, and (b) secondary batteries.....	38
Figure 3.10 Evaluation boards: (a) Feather M0 Adalogger, and (b) ADXL345 .....	39
Figure 3.11 Diagram of the ADXL345 evaluation board .....	39
Figure 3.12 Physical connection of a sensor node .....	40
Figure 3.13 Arduino interface for programming to drive the sensor node .....	41
Figure 3.14 Improved wired sensor node structure .....	41
Figure 3.15 Designed data acquisition GUI.....	42
Figure 3.16 Experimental devices for sensor calibration.....	43
Figure 3.17 Spectra of two accelerometers for sensor calibration.....	44
Figure 3.18 Amplitudes and RMS values at different frequencies .....	45
Figure 3.19 Spectra of ADXL345 with frequency calibrated .....	45
Figure 3.20 Experimental platform structure diagram.....	46
Figure 3.21 Waveforms of four different leakage conditions at Position 1 .....	47
Figure 3.22 Vibration RMS values and thresholds setting at Position 1 .....	48

Figure 3.23 Vibration spectra of four different leakage conditions at Position 1 .....	49
Figure 3.24 Particular frequency analysis at Position 1 .....	49
Figure 3.25 Waveform of four different leakage conditions at Position 2 .....	50
Figure 3.26 Vibration RMS values and threshold setting at Position 2 .....	51
Figure 3.27 Vibration spectra of four different leakage conditions at Position 2 .....	52
Figure 3.28 Particular frequency analysis at Position 2 .....	52
Figure 3.29 Node structure and data acquisition process .....	53
Figure 3.30 Physical connection of the designed wireless sensor node .....	54
Figure 3.31 Copper base: (a) front view, and (b) side view .....	55
Figure 3.32 Leakage detector node: (a) internal structure, and (b) appearance .....	55
Figure 3.33 Android Studio software interface .....	56
Figure 3.34 Mobile application interfaces: (a) installation interface, and (b) icon interface .....	56
Figure 3.35 Mobile application interfaces of the Leak Detector: (a) connection interface, (b) line chart display interface, (c) calibration interface, and (d) process bar display interface .....	57
Figure 3.36 Acceleration RMS in three different directions .....	58
Figure 3.37 Relation between vibration RMS values and flow rates .....	59
Figure 3.38 Leakage real-time monitoring displayed with the line graph: (a) closed, (b) small leakage, (c) medium leakage, and (d) large leakage .....	59
Figure 3.39 Leakage real-time monitoring displayed with the bar process chart: (a) closed, (b) small leakage, (c) medium leakage, and (d) large leakage .....	60
Figure 3.40 Pipe temperature change with the flow rate .....	60
Figure 3.41 Infrared thermometer connected with Feather M0 Bluefruit LE .....	61
Figure 3.42 On rotor sensing application for rotating machines: (a) APP icon, (b) hardware design, (c) installation, and (d) monitoring interface .....	62
Figure 3.43 Design of package for on rotor sensing application: (a) box body, (b) box cover, (c) internal structure, and (d) assemble .....	62
Figure 4.1 Typical types of bearings applied in industry .....	65
Figure 4.2 Schematic diagram of a rolling element bearing .....	66
Figure 4.3 Tapered roller bearing structure and components .....	67
Figure 4.4 Damaged tapered roller bearings with different faults .....	68
Figure 4.5 Two steps in the sequence of partitions .....	71
Figure 4.6 Flow chart of the proposed dimension reduction based compression method .....	77
Figure 4.7 Tapered roller bearing test rig .....	79
Figure 4.8 Diagram of the tapered roller bearing structure .....	81
Figure 4.9 Tapered roller bearings with seeded defects .....	81
Figure 4.10 Typical life curve affected by the installation clearance .....	82

Figure 4.11 Geometric dimensions of TIMKEN 31308 .....	83
Figure 4.12 Envelope signals of the healthy and faulty bearings .....	85
Figure 4.13 Envelope spectra of the healthy and faulty bearings .....	86
Figure 4.14 Mutual information and FNN results of the three cases .....	87
Figure 4.15 State space plots for the healthy and faulty bearings .....	88
Figure 4.16 Colour and white-black recurrence plots for the healthy and faulty bearings.....	90
Figure 4.17 Scatter plot of the healthy and faulty bearings .....	90
Figure 4.18 Envelope signals of the outer race fault bearing at different clearances .....	91
Figure 4.19 Envelope spectra of the outer race fault bearing at different clearances .....	92
Figure 4.20 Mutual information and FNN results for the outer race fault bearing at difference clearances .....	93
Figure 4.21 State space plots for the outer race fault bearing at different clearance .....	93
Figure 4.22 Colour and white-black recurrence plots for the outer race fault bearing at different clearance .....	95
Figure 4.23 Scatter plot of different clearances of the outer race fault bearing.....	96
Figure 4.24 Envelope signals of the roller fault bearing at different clearances .....	97
Figure 4.25 Envelope spectra of the roller fault bearing at different clearances .....	98
Figure 4.26 Mutual information and FNN results for the roller fault bearing at difference clearances .....	98
Figure 4.27 State space plots for the roller fault bearing at different clearance .....	99
Figure 4.28 Colour and white-black RPs for the roller fault bearing at different clearance..	100
Figure 4.29 Scatter plot of different clearances of the roller fault bearing .....	101
Figure 5.1 Structure of a ball bearing .....	103
Figure 5.2 Flow chart of the proposed two-step compression based ball bearing fault detection method.....	105
Figure 5.3 Hilbert transform principle from the perspective of signal processing .....	106
Figure 5.4 Simulated impulse signal and system resonance .....	110
Figure 5.5 Simulated bearing signals with and without noise .....	111
Figure 5.6 Spectra of the signal before and after envelope.....	112
Figure 5.7 Relation between $PRD$ and parameter $c$ .....	113
Figure 5.8 Comparison of original envelope and reconstruction at $SNR = -15$ dB.....	113
Figure 5.9 Comparison of original envelope and reconstruction at: (a) $SNR = 0$ dB, (b) $SNR = -10$ dB.....	115
Figure 5.10 Fault features selected for $F-SNR$ calculation: (a) $SNR = 0$ dB, (b) $SNR = -5$ dB, (c) $SNR = -10$ dB, and (d) $SNR = -15$ dB.....	117
Figure 5.11 Schematic diagram of the ball bearing test rig .....	118
Figure 5.12 The bearings with defects: (a) outer race faults, and (b) inner race faults .....	119

Figure 5.13 Original signals of bearings with outer race faults: (a) time waveforms, and (b) spectra .....	121
Figure 5.14 Spectra of bearings with outer race faults: (a) before envelope, and (b) envelope spectra .....	122
Figure 5.15 Original and reconstructed envelope spectra of bearings: (a) healthy, (b) large outer race fault, (c) medium outer race fault, and (d) small outer race fault .....	123
Figure 5.16 Original and reconstructed envelope signals of bearings: (a) healthy, (b) large outer race fault, (c) medium outer race fault, and (d) small outer race fault .....	123
Figure 5.17 Fault feature selection for envelope spectra of bearings with outer race faults: (a) large, (b) medium, and (c) small .....	125
Figure 5.18 Original signals of bearings with inner race faults: (a) time waveforms, and (b) spectra .....	126
Figure 5.19 Spectra of bearings with inner race faults: (a) before envelope, and (b) envelope spectra .....	127
Figure 5.20 Original and reconstructed envelope spectra of bearings: (a) healthy, (b) large inner race fault, (c) medium inner race fault, and (b) small inner race fault .....	128
Figure 5.21 Original and reconstructed envelope signals of bearings: (a) healthy, (b) with a large fault, (c) with a medium fault, and (d) with a small fault .....	128
Figure 5.22 Fault feature selection for envelope spectra of bearings with inner race faults: (a) large, (b) medium, and (c) small .....	130
Figure 6.1 Schematic diagram of a two-stage reciprocating compressor .....	133
Figure 6.2 Internal compression structure of a two-stage reciprocating compressor .....	134
Figure 6.3 Relationship between pressure and volume in a cylinder during working.....	135
Figure 6.4 128-dimensional dense SIFT descriptor .....	137
Figure 6.5 Flow chart of the proposed fault classification method.....	139
Figure 6.6 Structure of the two-stage reciprocating compressor test rig .....	140
Figure 6.7 Simulated defects of the reciprocating compressor: (a) AFB, (b) ASW, (c) DVL and (d) ICL .....	141
Figure 6.8 A sample photo taken from the top view .....	143
Figure 6.9 Selected images for six different cases: (a) BL, (b) AFB, (c) ASW100, (d) ASW150, (e) DVL and (f) ICL.....	144
Figure 6.10 Examples of the resized grayscale image under the different conditions: (a) BL, (b) AFB, (c) ASW100, (d) ASW150, (e) DVL and (f) ICL.....	145
Figure 6.11 2D rotationally symmetric Gaussian lowpass filter and its gradients .....	146
Figure 6.12 Filtered vertical and horizontal vertical and horizontal Gaussian filters.....	146
Figure 6.13 Extracted dense SIFT features from original images under the different conditions: (a) BL, (b) AFB, (c) ASW100, (d) ASW150, (e) DVL and (f) ICL.....	147
Figure 6.14 Extracted dense SIFT features from resized images under the different conditions: (a) BL, (b) AFB, (c) ASW100, (d) ASW150, (e) DVL and (f) ICL.....	148
Figure 6.15 Confusion matrix of the fault classification accuracy for testing results .....	149



Figure 7.1 Structure of a CNN example .....	154
Figure 7.2 An example of the convolutional layer.....	155
Figure 7.3 An example of the polling layer .....	156
Figure 7.4 Flow chart of the proposed CS-based CNN method .....	157
Figure 7.5 Two-stage helical gearbox test rig: (a) schematic diagram and (b) actual test rig	159
Figure 7.6 Schematic diagram of lubricant levels: (a) inside view, (b) axial view, and (c) side view.....	160
Figure 7.7 Tooth breakage fault simulation: (a) healthy pair, (b) 50% tooth breakage, and (c) 100% tooth breakage.....	161
Figure 7.8 Timeline recording of the experimental process .....	162
Figure 7.9 Oil temperatures of the GB1 for different cases under various loads .....	163
Figure 7.10 Surface topography of the GB1 house.....	164
Figure 7.11 Images for BL under different loads: (a) 0%, (b) 30%, (c) 70%, and (d) 100%	164
Figure 7.12 Images for different oil level under 100% loads: (a) BL (b) OS2000, and (c) OS1500 .....	165
Figure 7.13 Images for different cases under 100% loads: (a) BL, (b) VIS100, (c) TB050, and (d) TB100.....	166
Figure 7.14 Images for VIS100 under different loads: (a) 0%, (b) 30%, (c) 70%, and (d) 100% .....	166
Figure 7.15 Images for TB050 under different loads: (a) 0%, (b) 30%, (c) 70%, and (d) 100% .....	167
Figure 7.16 Images for TB100 under different loads: (a) 0%, (b) 30%, (c) 70%, and (d) 100% .....	167
Figure 7.17 Equalized grayscale images for different cases under 100% loads: (a) BL, (b) OS2000, (c) OS1500, (d) VIS100, (e) TB050, and (f) TB100.....	169
Figure 7.18 Equalized grayscale images for OS2000 under different loads: (a) 0%, (b) 30%, (c) 70%, and (d) 100% .....	169
Figure 7.19 CS based images for different cases under 100% loads: (a) BL, (b) OS2000, (c) OS1500, (d) VIS100, (e) TB050, and (f) TB100 .....	170
Figure 7.20 CS based images for OS2000 under different loads: (a) 0%, (b) 30%, (c) 70%, and (d) 100%.....	170
Figure 7.21 Features of the first convolutional layer .....	171
Figure 7.22 Features of the second convolutional layer .....	172
Figure 7.23 Features of the third convolutional layer.....	172
Figure 7.24 Features of the fully connected layer.....	173
Figure 7.25 CNN training process .....	174
Figure 7.26 Confusion matrix of the fault classification accuracy for testing results .....	175

## LIST OF TABLES

Table 2.1 Comparison between random and deterministic sensing matrixes .....	22
Table 3.1 Typical accelerometers and parameters.....	36
Table 3.2 Comparison of traditional wireless transmission technologies.....	37
Table 3.3 Wire connection between ADXL345 and Feather M0 Adalogger .....	40
Table 3.4 Parameters of ADXL345 and CA-YD-182A .....	43
Table 3.5 Spectra details of CA-YD-182A and ADXL345.....	44
Table 3.6 Flow representations with various tap rotation angles .....	46
Table 4.1 Specification of the TIMKEN 31308 .....	79
Table 4.2 Clarke AC motor .....	79
Table 4.3 Specification of data acquisition equipment .....	80
Table 4.4 Estimated fault frequencies of 31308.....	83
Table 4.5 Extra frequencies caused by various reasons for different defects .....	84
Table 4.6 Values of $x$ and $y$ calculated with the exploratory experiment.....	84
Table 5.1 Key parameters for simulation.....	109
Table 5.2 Main indicator values compared with the signal with noise .....	115
Table 5.3 Specification of sensors .....	118
Table 5.4 Ball bearing specification.....	119
Table 5.5 Main indicator values for outer race faults .....	124
Table 5.6 Main indicator values for inner race faults .....	129
Table 6.1 Specification of the two-stage reciprocating compressor .....	140
Table 6.2 Parameters of the FLIR ONE Pro .....	140
Table 6.3 Experimental parameters and dataset grouping for training and testing.....	142
Table 7.1 Detailed parameters of the applied CNN layers.....	158
Table 7.2 Specifications of the test facilities .....	159
Table 7.3 Specifications of MILLGEAR 320 EP and MILLGEAR 100 EP .....	161
Table 7.4 Dataset grouping for CNN training and testing .....	162

---

## DECLARATION

---

This dissertation is submitted for the degree of Doctor of Philosophy at the University of Huddersfield. I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of Huddersfield.

This work is original except where acknowledgement and references are made to the previous work. Neither this nor any substantially similar dissertation has been or is being submitted for a degree, diploma or other qualification at any other university.

---

## DEDICATIONS AND ACKNOWLEDGEMENTS

---

This research is carried out at the Centre for Efficiency and Performance Engineering (CEPE), School of Computing and Engineering at University of Huddersfield, UK for a PhD degree. It is financially funded by the Fee-waiver Scholarship scheme from the University of Huddersfield and the State Scholarship Fund by the China Scholarship Council (CSC). I sincerely appreciate the support and funding from CEPE, the University of Huddersfield and CSC during my PhD study.

I would like to express my gratitude to my main supervisor Professor Andrew D. Ball, Pro Vice-Chancellor for Research and Enterprise at the University of Huddersfield, for his encouragement and trust during my research. With his support, I got more opportunities to broaden my mind in this study.

I really appreciated the invaluable advice and guidance from my co-supervisor Professor Fengshou Gu, the Head of Measurement and Data Analytics Research Group (MDARG) and the Deputy Director of the Centre for Efficiency and Performance Engineering (CEPE) in my research. His rich experience in the field of machine condition monitoring significantly inspires me in my academic research.

Then, I would like to thank all members in CEPE group and all my friends for their generous help in my research and my life, especially Dr Guojin Feng, Dr Khalid Rabeyee, Mr Xiuquan Sun and Ms Haiyang Li for their advice and discussion in the experiments.

Finally, I special appreciate my family. Many thanks for the support, understanding and encouragement from my parents and my elder sister, as well as my parent-in-law. And many thanks my husband, Mr Yuandong Xu, for his patient discussion and valuable advice in my research, and also sincere trust and care in my life.

---

## LIST OF ABBREVIATIONS

---

1D	One-dimensional
2D	Two-dimensional
ADC	Analog-to-digital converter
AE	Acoustic emission
AFB	Air filter blockage
AI	Artificial intelligent
AR	Autoregressive model
ANN	Artificial neural network
ASW	Asymmetrical stator windings
BCS	Bayesian compressive sensing
BLE	Bluetooth low energy
BP	Basis pursuit
CB	Conventional bispectrum
CCA	Curvilinear component analysis
CDA	Canonical discriminant analysis
CM	Condition monitoring
CMOS	Complementary metal oxide semiconductor
CNN	Convolutional neural network
CR	Compression ratio
CS	Compressive sensing
DCT	Discrete cosine transform
DDMA	Discriminant diffusion maps analysis
DFT	Discrete Fourier transform
DMD	Digital micromirror device
DNN	Deep neural network
DoG	Difference of Gaussians
DPCM	Differential pulse code modulation
DVL	Discharge valve leakage
DWT	Discrete wavelet transform

ECG	Electrocardiogram
ECT	Eddy-current testing
EEMD	Ensemble empirical mode decomposition
EMD	Empirical mode decomposition
F-SNR	Fault feature signal-to-noise ratio
FFT	Fast Fourier transform
FNN	False nearest neighbour
GUI	Graphical user interface
HOS	Higher order spectra
HT	Hilbert transform
I <sup>2</sup> C	Inter-integrated circuit
ICA	Independent component analysis
ICL	Intercooler leakage
IMF	Intrinsic mode function
IoT	Internet of Things
IR	Infrared radiation
KNN	K-nearest neighbour
KPCA	Kernel principal component analysis
KS-LLE	Kernel sparse LLE
LDA	Linear discriminate analysis
LLE	Locally linear embedding
LPP	Locality preserving projection
LVQ	Learning vector quantisation
MEMS	Micro-electromechanical systems
MMV-CS	Multiple measurement vector CS
MP	Matching pursuit
MRI	Magnetic resonance imaging
MSB	Modulation signal bispectrum
MT	Magnetic particle testing
NDT	Non-destructive testing
NPC	Non-deterministic polynomial complete
OMP	Orthogonal matching pursuit
PC	Principal component
PCA	Principal component analysis
PRD	Percent root mean square difference

PT	Penetrant testing
RAM	Random-access memory
ReLU	Rectified linear unit
RF	Radio frequency
RFT	Remote filed testing
RIP	Routing information protocol
RKMFA	Regularized kernel marginal Fisher analysis
RLC	Run length coding
RMS	Root mean square
RMSE	Root mean square error
RNG	Random-number generator
ROM	Read-only memory
RP	Recurrence plot
RQA	Recurrence quantification analysis
RT	Radiographic testing
SD	Secure digital
SIFT	Scale-invariant feature transform
SPI	Serial peripheral interface
STFT	Short time Fourier transform
SVD	Singular value decomposition
SVM	Support vector machine
USB	Universal serial bus
UT	Ultrasonic testing
WPT	Wavelet packet transform
WSN	Wireless sensor network
WT	Wavelet transform

---

## PUBLICATION LIST

---

- [1] **Xiaoli Tang**, Yuandong Xu, Fengshou Gu, Andrew D. Ball, and Guangbin Wang. "Fault detection of rolling element bearings using the frequency shift and envelope based compressive sensing." In *2017 23rd International Conference on Automation and Computing (ICAC)*, pp. 1-6. IEEE, 2017.
- [2] Yuandong Xu, **Xiaoli Tang**, Fengshou Gu, Andrew D. Ball, and James Xi Gu. "Early detection of rolling bearing faults using an auto-correlated envelope ensemble average." In *2017 23rd International Conference on Automation and Computing (ICAC)*, pp. 1-6. IEEE, 2017.
- [3] Zainab Mones, Guojin Feng, **Xiaoli Tang**, Usama Haba, Fengshou Gu, and A. D. Ball. "A comparative study of gravitational acceleration cancellation from on-rotor MEMS accelerometers for condition monitoring." In *24th International Congress on Sound and Vibration, ICSV*. 2017.
- [4] Yuandong Xu, Pieter van Vuuren, **Xiaoli Tang**, Fengshou Gu, and Andrew Ball. "A Robust Method to Detect Faults of Rolling Bearings Using Ensemble Average Autocorrelation Based Stochastic Subspace Identification." In *30th International Congress & Exhibition on Condition Monitoring and Diagnostic Engineering Management*. 2017.
- [5] Zainab Mones, Qiang Zeng, Lei Hu, **Xiaoli Tang**, Fengshou Gu, and Andrew D. Ball. "Planetary gearbox fault diagnosis using an on-rotor MEMS accelerometer." In *2017 23rd International Conference on Automation and Computing (ICAC)*, pp. 1-6. IEEE, 2017.
- [6] **Xiaoli Tang**, Zainab Mones, Xianghong Wang, Fengshou Gu, and Andrew D. Ball. "A Review on Energy Harvesting Supplying Wireless Sensor Nodes for Machine Condition Monitoring." In *2018 24th International Conference on Automation and Computing (ICAC)*, pp. 1-6. IEEE, 2018.
- [7] **Xiaoli Tang**, Xianghong Wang, Robert Cattley, Fengshou Gu, and Andrew D. Ball. "Energy harvesting technologies for achieving self-powered wireless sensor networks in machine condition monitoring: A review." *Sensors* 18, no. 12 (2018): 4113.
- [8] Khalid Rabeyee, **Xiaoli Tang**, Yuandong Xu, Dong Zhen, Fengshou Gu, and Andrew D. Ball. "Diagnosing the Change in the Internal Clearances of Rolling Element Bearings based on Vibration Signatures." In *2018 24th International Conference on Automation and Computing (ICAC)*, pp. 1-6. IEEE, 2018.
- [9] Khalid Rabeyee, **Xiaoli Tang**, Fengshou Gu, and A. D. Ball. "The Effect of Wear Evolution on Vibration-based Fault Detection in Tapered Roller Bearings." *International Journal of Condition Monitoring* 9, no. 1 (2019): 18-23.
- [10] Baoshan Huang, Guojin Feng, **Xiaoli Tang**, James Xi Gu, Guanghua Xu, Robert Cattley, Fengshou Gu, and Andrew D. Ball. "A Performance Evaluation of Two Bispectrum Analysis Methods Applied to Electrical Current Signals for Monitoring Induction Motor-Driven Systems." *Energies* 12, no. 8 (2019): 1438.
- [11] Bei Zhang, Qichang Zhang, Wei Wang, Jianxin Han, **Xiaoli Tang**, Fengshou Gu, and Andrew D. Ball. "Dynamic Modeling and Structural Optimization of a Bistable Electromagnetic Vibration Energy Harvester." *Energies* 12, no. 12 (2019): 2410
- [12] Rongfeng Deng, **Xiaoli Tang**, Baoshan Huang, Xiuquan Sun, Yongbo Li, Fengshou Gu, and Andrew D. Ball. "Diagnosing Starved Lubrication of Helical Gearboxes using Infrared



- Thermal Imaging." In *2019 4th International Conference on Maintenance Engineering (IncoME-IV 2019)*, pp. 1-11, 2019.
- [13] **Xiaoli Tang**, Ke Li, Pieter A. van Vuuren, Junfeng Guo, Funso Otuyemi, Fengshou Gu, and Andrew D. Ball. "Condition Monitoring of Lubricant Shortage for Gearboxes based on Compressed Thermal Images." In *2019 32nd International Congress and Exhibition on Condition Monitoring and Diagnostic Engineering Management (COMADEM 2019)*, pp. 1-11, 2019.
- [14] Rongfeng Deng, **Xiaoli Tang**, Lin Song, Abdullahi Abdulmumeen, Fengshou Gu, and Andrew D. Ball. "An Evaluating Study of Using Thermal Imaging and Convolutional Neural Network for Fault Diagnosis of Reciprocating Compressors." In *2019 32nd International Congress and Exhibition on Condition Monitoring and Diagnostic Engineering Management (COMADEM 2019)*, pp. 1-11, 2019.

---

## CHAPTER 1 INTRODUCTION

---

*With the increase of the mechanical complexity and scale, the equipment management and maintenance are facing more and more challenges. Condition monitoring (CM) always plays a vital role in preventing the unexpected failure of machines and saving substantial maintenance costs.*

*This chapter describes the applications and development of machine CM in Industry 4.0. It also presents the importance of data compression for one-dimensional (1D) and two-dimensional (2D) signals in efficient CM. The lack of attention to data compression in CM activates the motivation of this research. To fulfil the research aim, the objectives are detailed in this chapter. Finally, the organisation of this thesis is outlined with a structure diagram.*

## 1.1 Industry 4.0 and Machinery Maintenance

The fourth industrial revolution, well known as Industry 4.0, is an intelligent era using information technology to achieve faster innovation in manufacturing and promote industrial development [1]. The development history of the industrial revolution is illustrated in Figure 1.1. Industry 4.0 characterizes a fusion of emerging technologies in the fields of physics, digital processing and biology, such as artificial intelligent (AI), robotics, Internet of Things (IoT), autonomous vehicles, 3D printing, nanotechnology, biotechnology, material science and energy storage [2]. It has the potential to raise global income levels and improve the quality of life because technology creates new products and services with high efficiency and productivity.

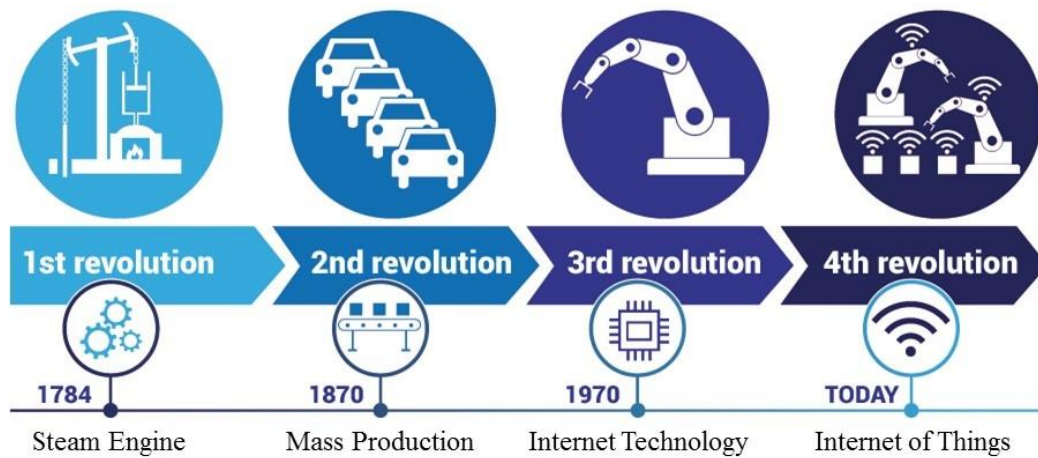


Figure 1.1 The development of industrial revolutions

Machinery plays an essential role in the industry. With the advent of Industry 4.0, the traditional industry is transforming into the direction of automation, digitization, intelligence and precision, which requires more scientific and efficient maintenance management of mechanical devices in various fields. During the operation of machines, it is necessary to strengthen their maintenance management to guarantee overall performance and optimize their operation quality. Moreover, mechanical failure can lead to large-scale downtime or catastrophic failure, which also may threaten human lives. Scientific and efficient maintenance can not only prolong the lifecycle of machines but also repair or replace faulty components in the early stage of failure to reduce the probability of disaster, save expenses and improve work efficiency [3].

Machinery maintenance is a series of organised activities implemented on a machine or a facility to keep it operating in an optimal state with high production output and minimum cost. It can be divided into unplanned maintenance and planned maintenance according to whether it is pre-planned, which is as shown in Figure 1.2. For machines that have an emergency or identified failure during machine operation, unplanned maintenance is prevailingly performed. Planned maintenance consists of corrective maintenance, improvement maintenance, preventive maintenance and predictive maintenance. Corrective maintenance aims to repair, replace or restore the faulty part

to avoid the repetitive occurrence from the source to maximize the effectiveness of plants and eliminate crashes. Improvement maintenance has the purpose of reducing or even eliminating maintenance requirements. It is subdivided into engineering services (facility construction), design-out maintenance (redesigning facilities to tolerate frequent occurrence of failure) and shutdown improvement maintenance (performed in total stoppage situation of the facility plant). Preventive maintenance is a series of activities that are performed on machinery before the occurrence of a failure, including replacement, adjustment and major overhaul. Predictive maintenance aims at detecting abnormal physical conditions of equipment and performing appropriate maintenance to maximise its service lifespan. Condition-based maintenance is the commonly used maintenance method because it can remotely detect and locate the faults in an early stage for large-scale devices in real time, especially with the aids of advanced technology.

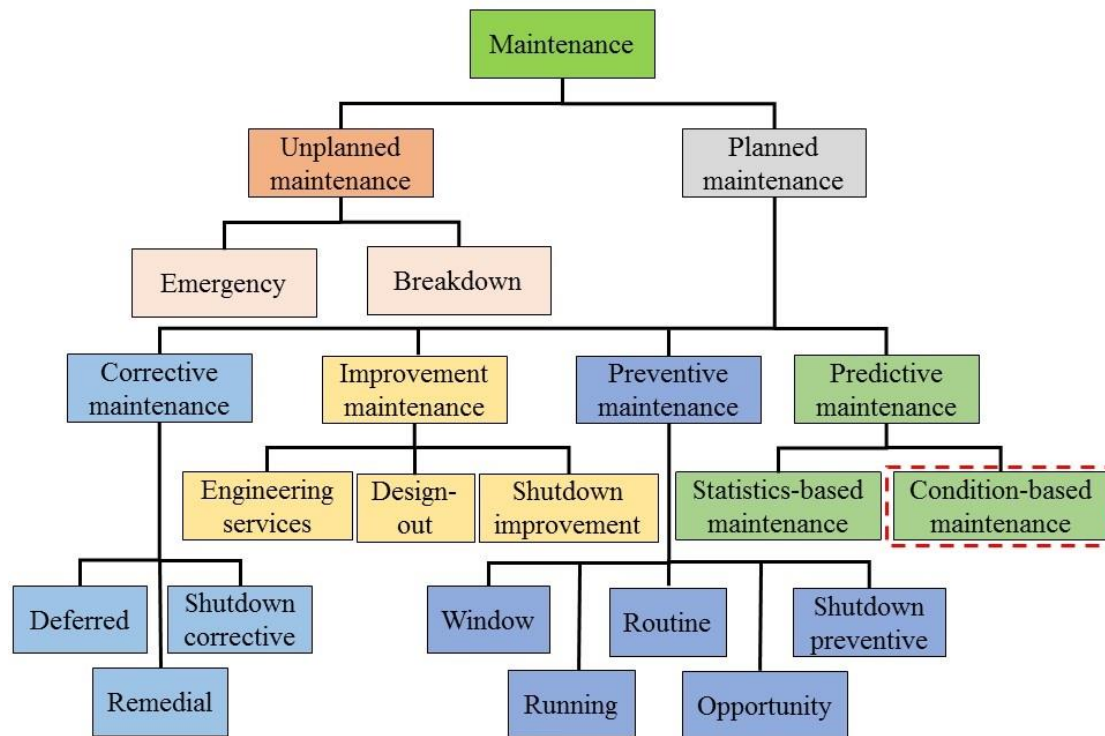


Figure 1.2 Maintenance strategies

In the condition-based maintenance process, the innovation of technology combination reduces the investment of human resources and realised more efficient and effective management. For example, AI has made remarkable progress in recent years driven by the growth of computing and the availability of massive data. In September 2019, the robotics design company Boston Dynamics simultaneously released two promotional videos of the biped robot Atlas which consistently performed beautiful gymnastics and parkour movements, as well as a quadruped robot working in the harsh environments. Customers can perform practical tests by adding specialized sensors and software onto Spot to monitor construction sites and provide remote inspections for natural gas piping systems, oil and power facilities.

Moreover, intelligent condition-based maintenance is becoming increasingly attractive with the support of wireless networks and IoT to improve the development progress of Industry 4.0 [4], [5]. The wireless sensor network (WSN) is a distributed intelligent network system composed of one or multiple sensor nodes which can monitor physical or environmental conditions (such as pressure, temperature, motion or pollutants, etc.) in different locations. Currently, with the rapid development of the micro-electromechanical systems (MEMS), system on chip, wireless communication and low-power embedded technology, WSNs have been widely applied in intelligent transportation, environmental monitoring, machine condition monitoring and so on [6], [7]. Considering the development and broad applications of WSNs, it is necessary to introduce their advantages and disadvantages as follows.

(1) Advantages:

- ✓ Economic. The cost of MEMS is much lower than traditional sensors and data acquisition devices.
- ✓ Flexible. Every sensor node in a WSN can move in any way and keep in touch with other nodes dynamically because of its self-organization.
- ✓ Dynamic. Sensor node failures caused by environmental factors or energy depletion and adding a new sensor node will affect the structure of the WSN. However, WSNs can adapt to this challenge through dynamic system reconfiguration.
- ✓ Integratable. The sensor nodes have the characteristics of low power consumption, small size and economic price, and can realise the integration of components and functions.

(2) Disadvantages:

- ✓ Lower transmission rate and limited bandwidth compared to wired networks. Hence, it may cause data loss if large amounts of data are transferred in real time from multiple sensors simultaneously.
- ✓ More complex to configure but less safe than wired networks.
- ✓ Easily affected by distance and surroundings. For example, Wi-Fi routers can work at the traditional 2.4GHz, reach up to about 46m indoors (with walls and furniture) and 92m outdoors.

However, intelligent applications, like data fusion for various sensors, bring not only opportunities for CM of machinery but also challenges of the transmission, storage and processing of massive data. In recent years, the amount of data is exploding at an exponential rate rather than a linear rate. The data exponential growth will reduce the efficiency of CM and even paralyze intelligent online monitoring systems. Therefore, data compression becomes a crucial task to ensure the normal operation of CM systems and save a lot of expenses.

## 1.2 Machine Condition Monitoring

Machine condition monitoring is an important action to monitor the health status of a machine by continuously observing changes of its state variables, such as temperature, pressure, vibration, acoustic and so on, to identify whether the machine is healthy or not. As soon as the state variables change, the machine should be diagnosed and repaired immediately. Otherwise, it will cause catastrophic failure to result in huge economic losses and even endanger human life.

Rotating machinery plays a crucially essential role in Industry 4.0. Its condition monitoring has attracted more and more attention from many researchers in recent years. They explore efficient data processing methods to effectively reduce machine downtime and maintenance costs. According to sensing technology, the common machine CM techniques include visual inspection, oil analysis, performance trend, non-destructive testing, electrical signal based methods, vibration and acoustic signal based methods.

### 1.2.1 Visual Inspection

Visual inspection is a basic method to monitor the condition of a machine through the senses of engineers, such as vision, hearing, smell and feel [8]. Visual inspection does not rely on complex sensors and acquisition equipment and experienced engineers are able to quickly and accurately identify faults that occur in machines. However, it is not appropriate to perform a manual visual inspection in some hazardous work environments, such as in the wind turbine field. Sometimes, artificial vision is difficult to meet the requirements, such as during mass production. Additionally, it needs quantifiable parameters to support the inspection and be recorded by engineers for comparison with future inspections. In general, severe defects can be inspected by engineers so that the early stage of faults can be ignored during the inspection. Finally, although it saves costs in the purchase and installation of sensors and acquisition equipment, long-term employment and training of a large number of engineers is costly and inefficient. With the development of science and technology, machines can replace the human eye to make measurements and detection in visual inspection, which is invaluable for defect detection. In fact, machine vision inspection is a non-intrusive CM method. It is characterized by increasing productivity and production automation. Besides, it is easy to implement information integration in machine vision inspection. However, machine vision inspection is usually limited to surface condition check and vulnerable to resolution. As a result, many other advanced technologies have been developed to monitor the condition of machines.

### 1.2.2 Oil Analysis

The oil analysis based CM method obtains information on the lubrication and the wear status of a machine by analysing the performance change of the medium such as lubricating oil of the monitored equipment to evaluate its working condition and predict its fault type [9].

Oil analysis is further divided into oil physical and chemical property analysis and oil debris analysis. The performance of the lubricant directly affects the wear state of mechanical friction. Monitoring the performance of the lubricant is to detect the operation state of the system and prevent the failure caused by poor lubrication. Oil debris analysis primarily detects the source of failure caused by friction based on the material composition of the abrasive particles. The common analytical techniques include spectroscopic analysis, ferrographic analysis, magnetic plugs and so on [10], which are suitable for analysing different sizes of debris. However, instruments for oil analysis are expensive and it is difficult to monitor the status of lubricants online, especially in harsh environments. Furthermore, the sampling and testing processes are complicated, so that oil samples are susceptible to contamination during these processes and affects the analysis results.

### 1.2.3 Performance Trend

Performance trend monitoring is another basic technique for CM. The commonly used measurements of performance include pressure, speed, displacement and temperature, etc., which can be applied to monitor conditions of the rotating machines.

The advantages of performance trend for CM include low sensor cost, low sampling frequency and easy understanding by engineers. Therefore, it is suitable for long-term online CM. However, some failures occur slowly, like fatigue wear. Early stage of faults has very little effect on these performance measurements so that they change slightly in the early stage of faults. It is difficult to determine whether the small changes of the captured performance are caused by ambient or early stage of faults. In addition, instruments require to be calibrated to guarantee the accuracy of the measurements. Many manufacturers equip some sensors to measure machine performance trends during production, such as Rolls-Royce Holdings PLC routinely monitoring aircraft engines under long-term service.

Among these measurements, temperature is one of the most commonly used ones in CM of rotating machinery, as any instability or failure can cause temperature changes. If the influence of ambient on the change of the measured temperature can be eliminated, the temperature can characterise the changes of the mechanical operating states. Infrared thermal imaging technology displays the heat distribution of the object surface in a non-contact way, which can overcome the problem of instability caused by single-point measurement. Actually, it is also a non-destructive testing technique.

### 1.2.4 Non-destructive Testing

Non-destructive testing (NDT) refers to changes in the reaction of heat, sound, light, electricity and magnetism caused by abnormalities or defects in the internal structure of the material without impairing or affecting the performance of the tested objects [11], [12]. It uses physical or chemical methods to check the structure, nature, quantity, shape, size, location and other changes of the internal and surface of the tested object by means

of modern technology and equipment. NDT is an indispensable and effective tool for industrial development. The traditional NDT techniques include radiographic testing (RT), ultrasonic testing (UT) [13], [14], magnetic particle testing (MT) and liquid penetrant testing (PT) [15], [16]. In addition, there are eddy-current testing (ECT) [17], acoustic emission detection (AE) [18], [19], infrared radiation and thermal testing (IR) [20] and remote field testing (RFT) methods to detect defects of the objects. These NDT techniques are widely used in aerospace, military, automotive, electric power, railway, metallurgy, machinery and other industries [21]–[26].

Compared to the destructive testing, NDT has the following characteristics: the first one is non-destruction because it does not damage the structure of the objects; the second one is comprehensiveness because it can not only work on the raw material test and production inspection but also test the equipment in service. It focuses on the inspection range of weld surface defect inspection, status check, cavity inspection, assembly inspection and residue inspection. But these NDT techniques have their own advantages and disadvantages. For example, RT can display the internal defects of the weld as a 2D image and can be recorded. However, its shortcomings are that it has requirements for the thickness of the steel plate, high cost, great damage to human and environmental pollution. UT is a low-cost and no-polluting method, which is suitable for all kinds of environments, but it is not convenient to read data. MT is cheap and easy to operate, but it is only suitable for defect detection of magnetic or magnetic materials on the surface or near surface of magnetically permeable materials and requires an external power supply. PT is easy to operate and suitable for a wide range of environments but can only detect defects on the surface of the workpiece. ECT is expensive and only suitable for detecting small workpieces, surfaces and near-surface defects, although it is portable and can detect many types of materials.

As described previously, all objects above absolute zero ( $-273^{\circ}\text{C}$ ) emit infrared radiation with the wavelength range of  $0.78\mu\text{m}$  to  $1000\mu\text{m}$  which can be captured by infrared thermography technology. As a non-destructive method, the thermal camera uses an infrared detector and an optical imaging objective lens to receive the infrared radiation energy distribution map of the target to obtain an infrared thermal image [27]. This thermal image corresponds to the heat distribution field of the object. In other words, thermal cameras work by converting invisible infrared energy from an object into a visible thermal image. The different colours of the thermal image represent the different temperatures of the object. This advanced technique captures the thermal field distribution of the surface of the object, rather than a single point temperature, and therefore has fairly high stability. Because of the characteristics of the thermal cameras and their non-contact measurement, they have been widely used in various fields, such as disease diagnosis, circuit board and power transformation box detection, geothermal exploration, atmospheric and ocean monitoring, etc. In recent years, it has been applied to mechanical fault monitoring in the industrial field [27], [28]. Many studies have shown the effectiveness and practicability of this technology in CM. In future, thermal



cameras can be equipped with devices like robots and drones for large-scale equipment CM.

#### 1.2.5 Electrical Signals

The electrical signals, such as current, voltage and power, are the non-intrusive signatures which are widely used to diagnose the faults of induction motor-based drive systems [29]. The mechanical failures can cause changes in load torque and speed of rotating machines. And the speed change will induce additional signals related to the fault frequency in the electrical signals [30], which provides an effective method for mechanical fault diagnosis. The traditional method applied for the electrical signal processing is the spectrum analysis based on discrete Fourier transform (DFT). However, sometimes the fault features are too weak to extract, or the influence of noise may lead to high diagnostic error. To overcome these issues, the discrete wavelet transform (DWT), empirical mode decomposition (EMD), bispectrum (CB) and modulation signal bispectrum (MSB) methods are widely adopted to extract effective diagnostic features for CM [29].

There are several advantages of using electrical signals for machine CM. The first one is the easy installation. The sensors can be installed with non-contact and non-intrusive relative to machinery through remote implementation, which means it can be achieved remote online CM with electrical signatures. The second benefit is the low cost. The instruments that collect electrical signals are much cheaper than most of other sensors like accelerometers and AE sensors. Therefore, electrical signals are receiving more and more attention in the field of machine CM in recent years.

#### 1.2.6 Vibration and Acoustic Signals

In the last few decades, vibration signals have been extensively applied for fault diagnosis of rotating machines because the captured responses can give accurate indications of internal system conditions.

The traditional methods used for feature extraction and signal analysis include time domain statistics, fast Fourier transform (FFT) and short time Fourier transform (STFT), envelope spectrum analysis and wavelet transform. In addition, some statistic calculation algorithms have been particularly proposed to deal with these non-linear and non-stationary signals, such as singular value decomposition (SVD) [31], autoregressive model (AR) [32], principal component analysis (PCA), independent component analysis (ICA), kernel principal component analysis (KPCA) [33], modulation signal bispectrum [34] and so on [35].

One of the disadvantages of vibration signals for CM is the sensor installation, especially on large machinery in harsh environments. The acquisition of acoustic signals can be at a distance from the objects with the non-contact collection method. The acoustic signals can detect the very early stage of faults with sophisticated signal processing techniques [36] because they are extremely susceptible to environmental

noise contamination. Acoustic emission signals represent the acoustic radiation (elastic waves) generated when the internal structure of a solid material changes irreversibly. Generally, AE signals are in high-frequency regions (above 100 kHz), which can provide reliable features for early fault diagnosis of rotating machinery. AE devices are extremely expensive and the captured signals lead to a significantly large amount of data because of the high sampling rate (over megahertz). Vibration and acoustic signals also have a large amount of data due to the high sampling rate (tens of thousands hertz) required for the reliable and accurate fault features. The next subsection will highlight the importance of data compression in the field of machine CM, which has only received very little attention.

### **1.3 Importance of Data Compression in Condition Monitoring**

Machine CM plays a vital role in the industry, which aims to reduce the maintenance cost and improve the efficiency and productivity of machines. Various types of signals captured with the high sampling frequency are processed to extract diagnostic features with sophisticated signal processing techniques. Long-term online CM inevitably results in an explosive growth of the data sets, consuming substantial resources of storage and processing. Therefore, one of the big challenges in the field of machine CM is data compression for the cost-effective purpose of CM.

Data compression refers to reducing the amount of data to save the storage space and improve the efficiency of transmission, storage and processing. It is a technique to reorganise data according to advanced algorithms to reduce data redundancy and storage space, which includes lossy compression and lossless compression. Lossy compression is that the reconstructed compression data is not completely identical to the original one. Some information is discarded to exchange for a higher compression ratio. On the contrary, the reconstructed compression data obtained by lossless compression is identical to the original one. But its compression ratio is much lower than that of the lossy compression.

The necessity of data compression for machine CM is summarized as follows.

- (1) Data volume exponential growth. With the development of technology, the accuracy and measurement range of sensing and acquisition instruments have continuously increased, leading to an exponential growth in data volume. AI has started to be applied to the field of machine CM, which requires various types of data from multiple channels simultaneously with a high sampling rate.
- (2) Limitation of wireless transmission. The MEMS based sensing and acquisition modules are much cheaper than traditional acquisition devices. Additionally, wireless sensing systems have the characteristics of better flexibility, scalability and maintainability in machine CM. WSNs have been developed rapidly and applied widely in recent years, but data transmission is limited by the transmission bandwidth, speed and energy consumption.

- (3) Limitation of storage. The collected data can not only be used for real-time online CM but also can be stored for comparison of characteristic changes. Therefore, long-term storage will take up a lot of space.
- (4) Data redundancy. It is generally known that the more features extracted, the higher the accuracy of fault diagnosis. However, more features lead to complex calculations and greater exploration space. Hence, more features may be just effective data duplication, resulting in data redundancy and wasting space.

Data compression can significantly reduce the cost of storage, data transmission time and communication bandwidth, which has received widespread attention in many fields, such as digital communication, medicine, WSNs, satellite imagery and so on. However, most of researchers have not recognised the importance of data compression in the field of machine CM in recent years. But it is significantly important to compress the data to effectively and efficiently promote the implementation and optimization of online and remotely intelligent condition monitoring with the wireless transmission in industrial applications. Therefore, the motivation of this research is to investigate the useful algorithms to compress 1D or 2D data and eliminate data redundancy to develop the intelligent CM for future usage after or even before data acquisition.

## **1.4 Research Aim and Objectives**

Based on the previous description, this research aims to investigate the effective and efficient data compression algorithms, mainly applied to 1D and 2D signals, for intelligent fault diagnosis of machines, which have potentials to be implemented in embedded hardware to fundamentally reduce the amount of data. To accomplish this aim, the following objectives are set in the implementation of this research:

- (1) Review the popular advanced literature of data compression algorithms especially applied in the field of machine CM. Divide these compression techniques into the traditional post-acquisition compression and advanced pre-acquisition compression methods for further verification of application in CM.
- (2) Build a wireless sensor node with MEMS sensors to collect 1D characteristic signals and analyse the acquired signals to extract features based on compressed statistics for the remote condition monitoring in real time.
- (3) Verify the availability and effectiveness of the developed intelligent WSN node with compression function by experimental studies.
- (4) Investigate the dimension reduction based compression methods to save storage space for 1D diagnostic signals of CM and simultaneously distinguish the fault types with the compressed features.
- (5) Propose sparse representation based compression algorithms like compressive sensing for 1D diagnostic signals with a large compression ratio which has the potential to be implemented in the hardware for the pre-acquisition compression.
- (6) Develop the compression method according to extracting features from 2D thermal images for fault diagnosis.

- (7) Exploit compressive sensing with an intelligent classification method on 2D thermal images captured by monitoring the condition of machines to differentiate various types of faults. The combined method has the potential to reduce the cost of data collection, transmission and storage.

## 1.5 Organization

This research involves several different compression techniques applied to different types of signals for mechanical components in the industrial field. The main structure of this thesis is organised as shown in Figure 1.3. It will be divided into eight chapters. The main contents of the remaining chapters are listed as follows.

**Chapter 2** overviews the recent development of compression technologies and applications in the field of machine CM, especially the application of WSNs and some innovative intelligent methods. Furthermore, the advanced compressive sensing algorithms are introduced and discussed, which provides potential pre-acquisition compression methods for machine CM.

**Chapter 3** expounds the importance of CM of piping systems and analyses the principle of vibration caused by leakage from the macroscopic and microscopic views. Then, the designed wired and wireless sensing systems with MEMS are introduced to achieve leakage detection with less wireless transmitted data in terms of the statistics based compression method.

**Chapter 4** proposes a compression method based on dimension reduction to effectively extract the periodic characteristics of the 1D vibration signal of the tapered roller bearing system. Recurrence plot and its quantification analysis, as well as principal component analysis, are combined to implement feature compression and classification.

**Chapter 5** investigates a two-step adaptive compression method based on sparse representation and compressive sensing to reduce the amount of data used for ball bearing fault diagnosis. In addition, the proposed compression method has the ability to eliminate noise and reconstruct signals.

**Chapter 6** explores a compression method based on sparse coding to reduce the dense scale-invariant feature transform (SIFT) features related to edge information extracted from the 2D thermal images of a reciprocating compressor to diagnose its various types of faults.

**Chapter 7** exploits the compressive sensing method on the 2D thermal images of gearboxes to realize intelligent fault diagnosis with the aid of the convolutional neural network (CNN).

**Chapter 8** reviews the achievements for the objectives mentioned in Chapter 1. Then, it concludes this research on data compression method investigation for intelligent machine CM and make some suggestions for future work in the related research fields.

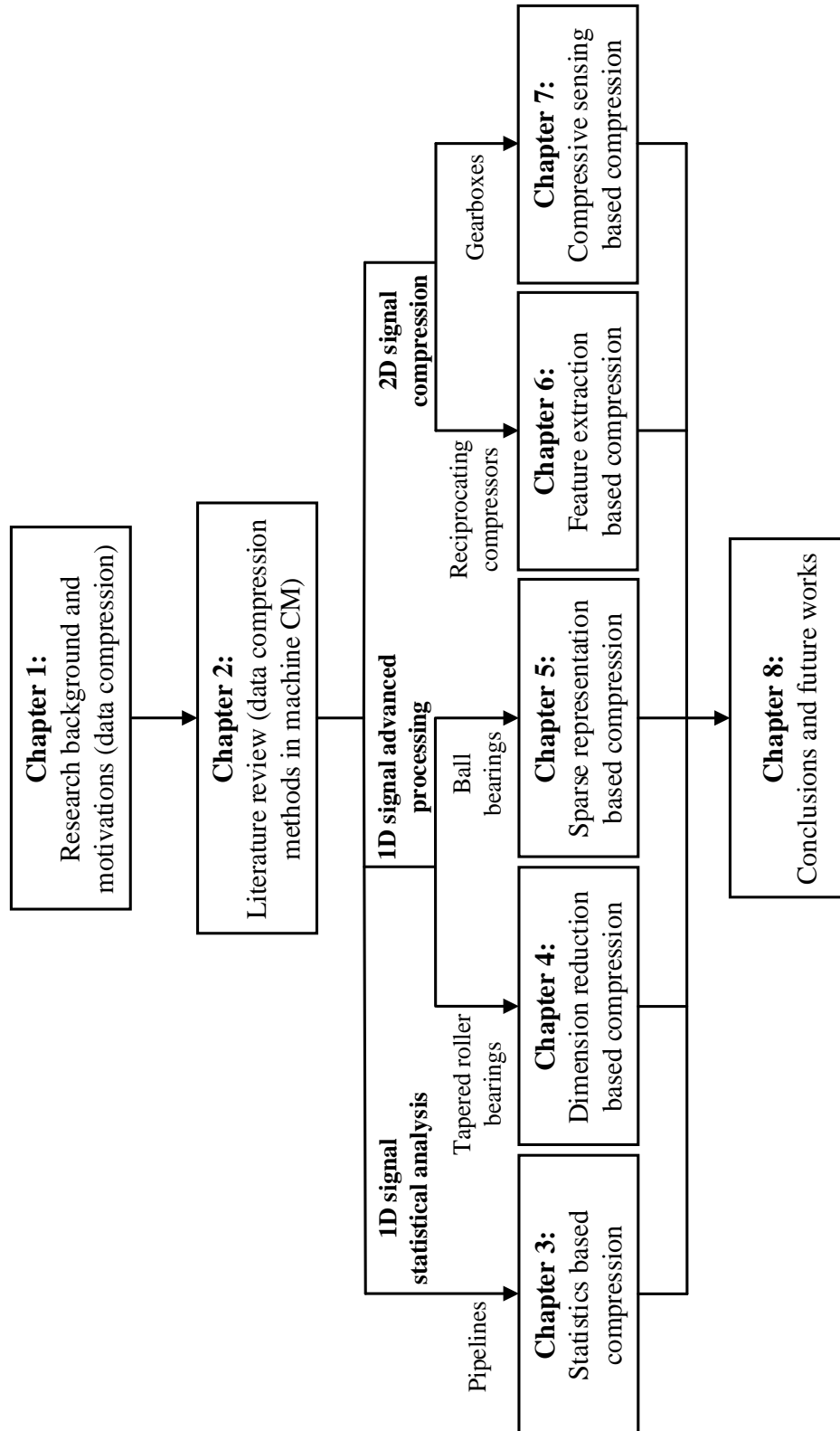


Figure 1.3 Structure of the thesis

---

## CHAPTER 2 LITERATURE REVIEW

---

*According to the importance and necessity of data compression in machine CM, this chapter overviews and summarizes the state-of-the-art research on data compression techniques in this field. It gives an overview of data compression methods in machine condition monitoring. Especially, an advanced random sampling based compression method, compressive sensing, is introduced and reviewed with the wide applications in various fields, particularly in machine condition monitoring.*

## 2.1 Data Compression

Data compression is not essential, but it is a critical procedure in signal processing to overcome the overwhelming data expansion issues in Industry 4.0. It uses a few bits to express the signal representation to achieve lossy or lossless compression because most of the actual data in physical world contain large quantities of redundancy information.

Data acquisition is a process of converting analogue signals into digital signals and transmitting them. According to the well-known Nyquist-Shannon sampling theorem, an analogue signal waveform may be uniquely reconstructed without any loss from the samples acquired if the sampling rate is equal to or greater than twice the highest frequency component of this analogue signal. Figure 2.1 (a) states an analogue signal waveform with the frequency of 1 Hz. Figure 2.1 (b) exhibits the sampling pulses with the sampling rate of 10 Hz. The sampled signal is displayed in Figure 2.1 (c), which is composed of discrete points. These points can be transmitted according to certain transmission rules after quantisation and encoding.

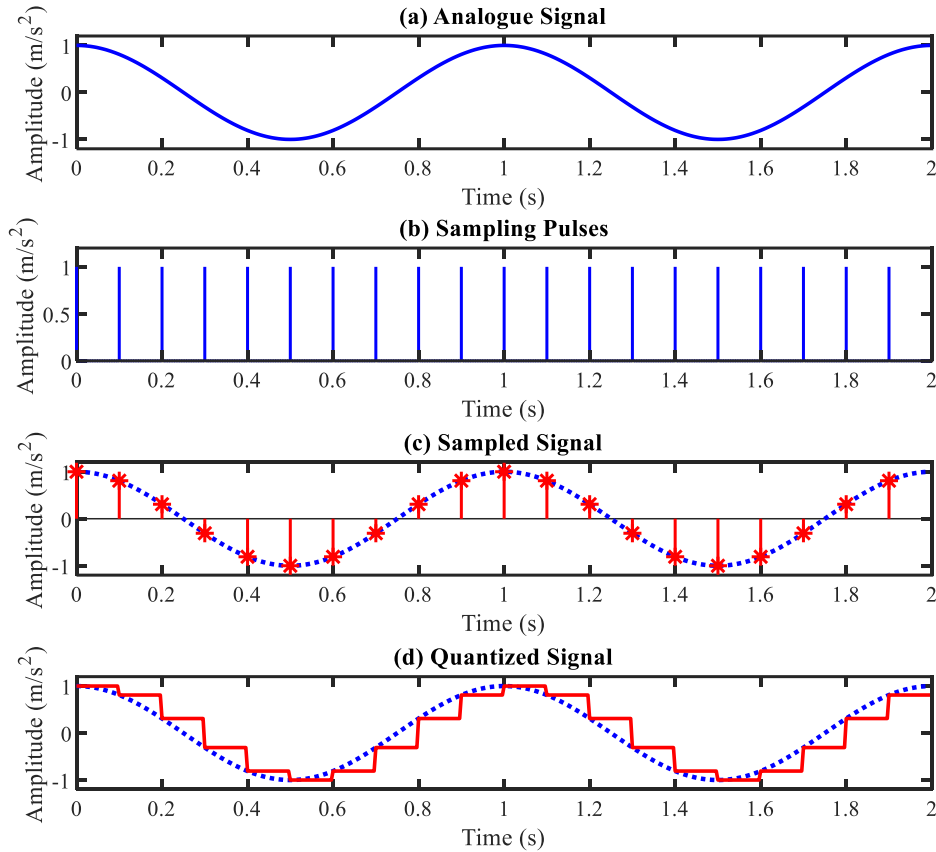


Figure 2.1 Simulated data acquisition process

Generally, a discrete signal obtained at a higher sampling rate can represent the waveform more accurately, which also leads to massive data and complicated computation. In the field of signal processing, data compression can be implemented during the sampling process before data transmission or at the receiving end after the signal is sampled and transmitted. At the signal acquisition end, there are two main

types of data compression. The first one is the traditional method that data is uniformly sampled based on the Nyquist-Shannon sampling theorem and then quantized. The quantized signals can be compressed before data compression, i.e. in the process of encoding. The other one is compressive sensing. It makes a breakthrough compared to the traditional sampling theorem through the random sample instead of the uniform sample, which has been rapidly developed and widely applied in the last decade. At the receiving end, the received transmission data is successfully decoded and then stored for further signal processing. As an alternative, the decoded signal can be compressed and stored to save hard disk storage space. Some researchers have studied some effective data compression methods at the acquisition and receiving ends in the field of machine CM, which will be detailed and summarized as follows.

### 2.1.1 Traditional Data Compression Methods

Traditional data compression methods can be divided into pre-transmission compression and post-transmission compression. They require to strictly follow the Nyquist-Shannon sampling theorem when the signals are sampled.

#### 2.1.1.1 Pre-transmission Data Compression Methods

With the development and application of wireless networks, condition monitoring of machinery working under harsh environments needs to be implemented with the wireless transmission on a large scale. Compared with the wired transmission, wireless transmission has the advantages of low cost, stable and robust performance, flexible networking, good extensibility, and low maintenance cost. In the meantime, it is limited by the transmission bandwidth, power consumption and computing ability of nodes in the process of the wireless node application in machine CM in the past decades. Therefore, some researchers are committed to studying the uniformly sampled signal to achieve compression in the encoding process at the data acquisition end before wired or wireless transmission, which is often referred to as source code. Source code primarily adopts the statistical characteristics of signal sources to resolve their correlations and then eliminate redundant information from signal sources to improve the effectiveness of communication through data compression. Differential pulse code modulation (DPCM), Huffman coding, run length coding (RLC), transform coding and other techniques are commonly used as the analogue signal coding techniques.

DPCM encodes the difference (namely prediction error) between the sampled value and the predicted one instead of encoding each sampled value independently [37]. The range of the prediction error is smaller than that of the sampled value because the correlation between the sampled value and the predicted value is strong. Therefore, fewer bits will be used to encode the prediction error to reduce the bit rate. Chan *et al.* [37] studied a method to achieve twice compression of the bearing vibration signals before wireless transmission with downsampling of intrinsic mode function (IMF) decomposed by EMD and DPCM. Three experiments about the typical, misalignment and unbalance shafts were carried out and the results revealed that only the proposed



compression method achieved a higher performance and compression ratio than the wavelet transform (WT) for machine CM.

Huffman coding is a compression technique used to reduce the number of bits needed to send or store. It is based on the frequency of occurrence of a data item (pixel in images). The principle is to use a lower number of bits to encode the data that occurs more frequently for lossless data compression [38]. Datta *et al.* [38] proposed a hybrid data compression technique to achieve remote CM of motors through DPCM along with Huffman coding. The sampled data was coded by DPCM firstly and then by Huffman coding to implement 20%-50% reduction in the amount of data transmitted. It has been demonstrated by decompression that the proposed method was lossy. In the following two years, these researchers developed corresponding software to receive, decompress and process the compressed and transmitted signals, which successfully detected the broken bar fault of induction motors [39]. Unfortunately, it only applied the simulated signals while practical experimental data was not tested. But the development of this scheme has promoted the progress of data compression in remote machine CM.

If Huffman coding is intuitively based on probability to optimize the length of the data to achieve data compression, then arithmetic coding is truly lossless compression method based on information entropy. Huang *et al.* [40] investigated a lossless compression method with arithmetic coding applied on the quantitated parameters and errors of 2D discrete cosine transform (DCT) signals. It showed that the data volume only accounted for 32.47% of the original mechanical vibration signal was saved to achieve decompression without loss.

Moreover, an efficient under-sampled method based on the fusion of kurtogram, band-pass filter, band-pass sampling and the Hilbert transform (HT) was proposed to reduce the amount of wireless data transmission by over 80% [41]. The under-sampled data had the ability to realize the fault detection of motor bearings. More importantly, the authors designed a WSN prototype with a wireless module working at 433 MHz and evaluated its practicability with both simulation and experimental signals.

#### 2.1.1.2 Post-transmission Data Compression Methods

The realization of lossless compression of complex data before transmission is the necessary requirement for the wireless machine CM system in future. At present, main pre-transmission data compression methods studied by researchers concentrate on the techniques applied to the traditional communication system data compression because of low computational complexity and easy implementation. However, the signal characteristics representing various types of machine faults are different from traditional communication systems. Fault information is easily seriously affected by noise or even lost during transmission. Many complicated compression algorithms have been proposed according to the signal characteristics used to monitor the machine condition. Unfortunately, most of them cannot be implemented on a large scale in

hardware at the acquisition end due to the limitation of storage capacity, computing ability, power consumption and the cost. Accordingly, they are commonly used at the receiving end (i.e. post-transmission) to effectively reduce the required storage space, especially for long-term online machine CM although they cannot be performed from the source of data acquisition. But with the development and advancement of the MEMS technology and nanotechnology, mature compression technologies have a great opportunity to realize on wireless sensor nodes. These post-transmission data compression algorithms are mainly based on transform coding, feature extraction, basic statistics, sparse representation, data or feature dimension reduction, and artificial intelligence.

Wavelet transform is a transformation-based technique similar to DCT. The signal is described in the wavelet domain and the area where energy concentrated can be definitely identified. In 2004, Peng *et al.* [42] gave a detailed overview of the wide applications of WT in machine CM, including data compression. The principle of WT compression is essentially the sparsity of the signal. Most wavelet coefficients are very small and can be ignored with a few bits representing the raw signal, which will not generate a significant error in the stage of reconstruction. Simultaneously, it stated that WT based methods are more suitable for compression of non-stationary mechanical vibration signals. Wavelet packet transform (WPT) is also a technique that decomposes signals into wavelet packet bases and coefficients in the wavelet domain. The small coefficients below the given threshold will be set to zero to sparse the coefficient matrix. Hence, how to determine the threshold based on characteristics of vibration signals has become a hot topic in recent years. For example, Hao *et al.* [43] carried out the wavelet packet decomposition of the bearing fault signal to obtain 16 sub-bands and calculated the energy of each sub-band. Then the energy threshold was set according to the compression ratio to sparse the coefficient matrix. Finally, the sparse wavelet coefficients were encoded, which can be used for further signal reconstruction with the aid of the wavelet basis. These methods are more significant in hardware implementation before encoding at the acquisition end in theory. In fact, authors implemented them by processing the collected vibration signals. In addition, the adaptive wavelet basis is difficult to select and there are some inherent deficiencies such as border distortion, energy leakage [44]. Compared with wavelet compression, Guo *et al.* [44] developed a compression method based on the selected IMFs decomposed by the ensemble empirical mode decomposition (EEMD). The experiments performed on both a small electric motor in a laboratory and a real traction motor of a train demonstrated that only a few extremes with large amplitudes reserved can represent the original vibration signals because they had sufficient faulty impulses to indicate the bearing fault type of a motor, which means most of the data has been discarded to compress the tremendous amount of samples.

The basic multivariable statistics of signals used for describing machine condition include mean, standard deviation, root mean square, kurtosis, peak, maximum,

skewness, energy, entropy, and some frequency domain features, like shaft frequency, fault frequencies and harmonics [45]. These extracted features obtain useful information for fault diagnosis and also reduces the amount of data used to detect faults of machines.

However, dimensions of features extracted are still high and require to be further reduced and fused with other compression algorithms [46], such as the commonly used PCA, canonical discriminant analysis (CDA) and linear discriminate analysis (LDA). Moreover, the correlation between features can be eliminated to reduce the information redundancy and improve the calculation efficiency and fault diagnosis accuracy. For example, Sawitri *et al.* [47] applied PCA as the compression method to reduce the dimension of features extracted from the time domain and frequency domain (like mean, root mean square (RMS), peaks, entropy, variance, etc.) denoised current signals with both WT and PCA methods. The compressed feature parameters were utilized for the classification of 10 different conditions of the induction motor with support vector machine (SVM) to realize the average identification index of 83.51% with 14 principal components of 25 characteristic parameters. Chopade *et al.* [48], Dias *et al.* [49] and Ruiz-Carcel *et al.* [50] effectively and efficiently utilized PCA to reduce the dimension of features in the time, frequency and wavelet domains which were extracted from the current signals, position measurements and vibration signals to diagnose the mechanical fault types, respectively. PCA was also used to compress the first 32 vibration envelope harmonic features of a two-stage reciprocating compressor by Smith *et al.* [51]. The compressed signals were as the input of the Naïve Bayes classifier to achieve a higher classification success rate of 83.3% instead of 75% with the raw envelope harmonic features for five different conditions of a two-stage reciprocating compressor. Additionally, both PCA and CDA methods were applied to reduce the dimensionality of the data from two planetary gearboxes [52]. The data were collected from two gearboxes applied in the bucket wheel drive unit. Furthermore, Arellano-Espitia *et al.* [53] reduced the dimension of 15 statistical vibration features to a 2D feature set with PCA and LDA methods. But these linear techniques represent high-dimensional features in low-dimensional space by a linear transformation. Compared with linear techniques, a non-linear technique autoencoder was applied in [53] with the achievement of higher classification accuracy for four different electromechanical system fault conditions.

Although the linear compression approaches have high calculation efficiency, sometimes the feature representation obtained by nonlinear compression methods is more effective to distinguish various faults so that they attract widespread attention in recent years. A manifold is a low-dimensional local structure that appears in a high-dimensional sample space. Manifold learning as an advanced compression technique can recover low-dimension manifold structures from high-dimensional samples. Commonly used non-linear dimension reduction methods based on manifold learning techniques for data compression include autoencoder, KPCA, locality preserving

projection (LPP), locally linear embedding (LLE) and isometric feature mapping (Isomap), etc. [54]. The following examples are the usage of non-linear manifold learning methods combining with other algorithms to implement mechanical fault diagnosis with dimension-reduced data. The autoencoder, as a non-linear smart compression technique, can reduce the input high-dimensional features with correlation and redundancy to a low-dimensional space through non-linear transform coding [55]. Calle *et al.* [56] also investigated the difference of fault features obtained after the dimension reduction by four compression algorithms: PCA, LDA, Relief and autoencoders. The researchers concluded that the best compression method for CM was confirmed because of the diversity of fault characteristics. But LDA and Relief had more prominent advantages on compressing their experimental data of bearings and gears. Additionally, the KPCA algorithm can map signals from low-dimensional to high-dimensional through non-linear kernel transformation. Then it is reduced to a low dimension to produce a different effect based on the added non-linear transformation. An adaptive KPCA method was applied to convert both statistical and energy features extracted from the wavelet sub-bands of audio signals decomposed with WPT into linear uncorrelated components to recognise the tool wear conditions in [57]. Curvilinear component analysis (CCA) focuses on small distances in the output space in the progress of manifold learning. Prieto *et al.* [58] studied CCA method on multiple time-based statistical features to extract new CCA features and visualized with 2D exhibition, as well as fault classification with a hierarchy of multilayer for bearings. To effectively extract optimal features and reduce dimensions from high-dimensional samples, Jiang *et al.* [59] proposed a regularized kernel marginal Fisher analysis (RKMFA) method to realize intelligent bearing fault diagnosis with the combination of the K-nearest neighbour (KNN) classifier. LPP as a typical manifold learning technique is to construct the distance relationship between the samples in the high-dimensional space and keep it in the mapping results so that the local neighbourhood structures of the samples are preserved while reducing the sample dimension and redundancy. Therefore, some researchers applied LPP to reduce the dimension of vibration features generated according to WPT for further classification with the SVM method to detect the tool wear status [60]. Wang *et al.* [61] also studied LPP for feature dimension reduction compared with PCA and KPCA to select and fuse the top ranked features with the vibration signals of a spiral bevel gear. In addition, they also compared the effectiveness of KPCA, LLE and Isomap on the selection and fusion of non-linear features, and demonstrated that KPCA performed the best effects on tool wear detection [62]. Isomap aims to find a low-dimensional space to ensure that the geodetic distance between the samples is approximately the same in the raw high-dimensional space and the dimension reduced space. It needs to calculate the global optimal solution for all samples resulting in the heavy calculation and low efficiency. The improved and commonly used algorithm is LLE because it discards the global optimal dimension reduction of all samples and only focuses on the local optimality. Some researchers improved the LLE method as the kernel sparse LLE (KS-LLE) to adaptively determine

neighbours of nonlinear system samples and their weights with kernel sparse representation [54]. The experimental results of gearbox fault diagnosis validated that the proposed KS-LLE method was more effective than the traditional manifold learning technique on both feature extraction and dimension reduction. In addition, Huang *et al.* [63], [64] verified that the discriminant diffusion maps analysis (DDMA) method was more suitable to reduce the redundancy of features and preserve the fault characteristics to monitor the condition of tool wear.

Actually, most of the previously introduced references or researches use the compressed features as input for some machine learning methods like SVM, KNN and MLP to achieve the intelligent classification. In order to implement high classification accuracy, feature extraction and dimension reduction are two necessary preprocessing steps to obtain high performance of machine CM. In conclusion, these data dimension reduction techniques can solve the following problems: Firstly, it can alleviate the dimension disaster induced by the increase of data dimensions. As the data dimension increases, the memory and processing ability required to process the dataset rises exponentially. Besides, high-dimensional dataset normally has a higher sparsity which can be compressed; Secondly, dimension reduction can preserve fault related information while compressing data, so that information loss is minimized in this process. Finally, high-dimensional data has a complex structure and difficult to understand, but dimension reduction makes useful information visible in only two or three dimensions. However, the post-transmission data compression methods based on information statistics, feature extraction, sparse representation, dimension reduction and so on are limited by the wireless transmission. These compression methods are mostly applied to the raw vibration, electric signals, etc. with high sampling rates. Therefore, limited by the computing ability and power consumption of wireless sensor nodes, they can only be performed at the post-transmission end. Fortunately, they help to effectively and efficiently classify the different fault types and save the storage space. In future, an advanced compression method called compressive sensing has the potential to implement data compression on the hardware and break the traditional uniform sampling theorem at the pre-transmission end to achieve the on-board CM, wireless transmission of compressed data and reconstruction of raw signals.

### 2.1.2 Compressive Sensing

Compressive sensing (CS), proposed by Donoho *et al.* [65], is different from the traditional Nyquist-Shannon sampling theorem because it achieves randomly sampling instead of the uniform sampling and uses far fewer samples to reconstruct the original signals through optimization if the raw signals are sparse or compressible. Compared with the traditional data compression methods, CS has a high probability to achieve pre-acquisition compression on the hardware in the field of machine CM. Suppose a discrete-time signal  $x(n), n = 1, 2, \dots, N$  with the length of  $N$ , it can be represented with a basis matrix of  $\Psi = [\psi_1, \psi_2, \dots, \psi_N]$ , where  $\psi_i, i = 1, 2, \dots, N$  are the basis vectors with the size of  $N \times 1$ . Then, signal  $x$  can be expresses as

$$x = \sum_{i=1}^N s_i \psi_i \text{ or } x = \Psi s \quad (2-1)$$

where  $s_i, i = 1, 2, \dots, N$  are the weighting coefficients and  $s$  is the coefficient vector. This process is the sparse representation of the signal. If only  $k$  ( $k \ll N$ ) of the coefficients are nonzero and another  $N - k$  are zero, it can be said that signal  $x$  is  $k$ -sparse. If  $k$  of the coefficients are large and another  $N - k$  of the coefficients are very small ones, signal  $x$  is compressible [65]. In reality, most of digital signals are sparse or compressible in a specific transform domain.

### 2.1.2.1 Encoding process

In the encoding process, a sparse  $N$ -points signal  $x$  is compressed by a random matrix  $\Phi$  to get the signal  $y$  with the length of  $M$  ( $M < N$ ) and signal  $y$  will be sent at the transmitting end. Specifically,  $\Phi$  is called measurement matrix with the size of  $M \times N$ . So that compressed signal  $y$  is defined as

$$y = \Phi x \quad (2-2)$$

Then, signal  $y$  will be transferred with only a few samples, which is efficient and stable especially for the wireless transmission. CS has been extensively applied in image or signal processing and approximation calculation [66], [67]. However, most of the signals are not sparse or compressible in the time domain. Usually, it needs to involve basis vectors  $\Psi$  to enhance the sparsity characteristics. Then, signal  $y$  can be expressed as

$$y = \Phi x = \Phi \Psi s = \Theta s \quad (2-3)$$

where,  $\Psi$  is the transformation matrix with the size of  $N \times N$ ;  $\Theta$  is the sensing matrix with the size of  $M \times N$  and  $s$  is the sparse representation of signal  $x$  with  $k$  zeros or negligibly small coefficients. In other words, signal  $x$  is called  $k$ -sparse in the  $\Psi$  domain, which can be compressed and has the potential to be reconstructed without or with little loss. There are several common transform domains exploited for a sparse representation of a signal, including time-domain representation, discrete Fourier transform, discrete cosine transform, discrete wavelet transform and so on. Figure 2.2 shows the theory of signal representation and compression with pixel matrixes [68].

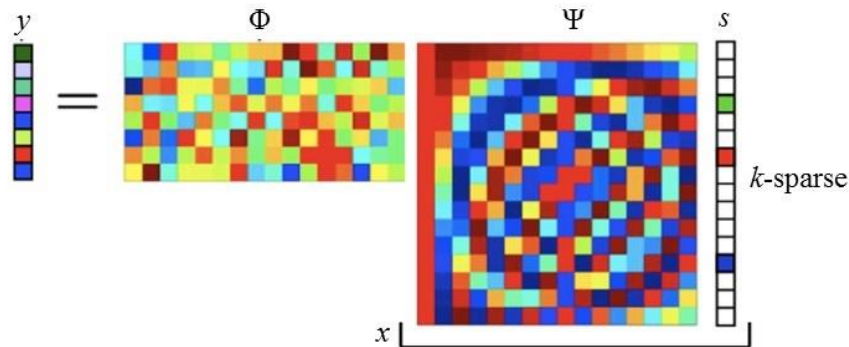


Figure 2.2 Matrix representation of signal compression

Furthermore, a stable measurement matrix should be designed reasonably to implement the signal recovery. In fact, Candes and Tao [69] have proposed that sensing matrix  $\Theta$  must meet the routing information protocol (RIP) defined as follows, then the signal can be successfully reconstructed.

**RIP definition:** for any  $k$ -sparse signal and the constant  $\delta_k \in (0,1)$ , sensing matrix  $\Theta$  should satisfy with Equation (2-4).

$$1 - \delta_k \leq \frac{\|\Theta x\|_2^2}{\|x\|_2^2} \leq 1 + \delta_k \quad (2-4)$$

The equivalent condition of RIP is that measurement matrix  $\Phi$  and sparse base matrix  $\Psi$  are irrelevant. In addition to the sparsity, the incoherence of these two matrices is the second essential requirement for CS to achieve signal reconstruction with high probability. Hence, measurement matrix  $\Phi$  should be incoherent with orthogonal base matrix  $\Psi$  to make the sensing matrix  $\Theta$  fulfil with RIP [69], [70] if  $M \geq ck\log(N/k)$ , where  $c$  a small constant [65]. There are some commonly used matrixes, such as Gaussian random matrix, Bernoulli random matrix, partially orthogonal matrix, Toeplitz and some deterministic sensing matrixes [71]. Table 2.1 compares the characteristics of random and deterministic sensing matrixes [72].

Table 2.1 Comparison between random and deterministic sensing matrixes

Random Sensing Matrix	Deterministic Sensing Matrix
<ul style="list-style-type: none"> <li>• Outside the mainstream of signal processing: worst case signal processing</li> <li>• Less efficient recovery time</li> <li>• No explicit constructions</li> <li>• Larger storage</li> <li>• Looser recovery bounds</li> </ul>	<ul style="list-style-type: none"> <li>• Aligned with the mainstream of signal processing: average case signal processing</li> <li>• More efficient recovery time</li> <li>• Explicit constructions</li> <li>• Efficient storage</li> <li>• Tighter recovery bounds</li> </ul>

In general, the Gaussian random matrix is usually used as the measurement matrix to randomly sample the signal.

#### 2.1.2.2 Decoding process

Decoding is a process of reconstruction in CS algorithms. In other words, the signal reconstruction process is the process of solving signal  $s$  when the compressed signal  $y$  and sensing matrix  $\Theta$  are known at the receiving end. After obtaining  $s$ , the original signal  $x$  can be recovered from signal  $s$  according to Equation (2-3). However, the number of equations ( $M$ ) is much less than that of variables ( $N$ ) in Equation (2-3). Therefore, it is difficult to reconstruct the original signal because the equations have no deterministic solution. Fortunately, since signal  $s$  is  $k$ -sparse, it is hopeful to get an optimal solution from  $M$  measurements with high probability if sensing matrix  $\Theta$  satisfies the RIP. In [73], [74], it has been demonstrated that a sparse signal can be

reconstructed with tolerable distortion from very few measurements via the minimization strategies.

Actually, the sparse signal  $s$  means that the non-zero elements in  $s$  are the least. To reconstruct this sparse signal is to solve the  $l_0$ -norm of vector  $s$ . Based on the algorithm proposed by Candes *et al.* [69],  $k$ -sparse coefficients  $s$  can be reconstructed through  $l_0$ -norm.

$$\hat{s} = \arg \min \|s\|_0 \quad s.t. \quad \Theta s = y \quad (2-5)$$

Nevertheless,  $l_0$ -norm minimization is a non-deterministic polynomial complete (NPC) problem in general. To recover  $k$ -sparse or compressible signals of length  $N$ ,  $l_1$ -norm minimization is used effectively to solve this optimization problem because it aims to solve the convex optimization problem which is shown in Equation (2-6).

$$\hat{s} = \arg \min \|s\|_1 \quad s.t. \quad \Theta s = y \quad (2-6)$$

The  $l_1$ -norm minimization is known as a fundamental convex optimization algorithm (convert  $l_0$ -norm to  $l_1$ -norm to solve by the linear programming), called basis pursuit (BP), to recover sparse solutions. In addition to the BP algorithm, the gradient projection method and the minimum regression method are another two main convex optimization algorithms. The greedy algorithm achieves the approximation of the signal vector by selecting the appreciate atoms for further iteration. Matching pursuit (MP), orthogonal matching pursuit (OMP) and their extended algorithms [75] are also widely exploited to solve this optimization problem. Due to the effectiveness of MP and OMP in solving convex optimization problems, their implementation procedures are briefly introduced.

Suppose  $\Theta$  in Equation (2-3) is an overcomplete dictionary matrix (the number of atoms is much larger than the length of signal  $y$ , i.e.  $M \ll N$ ) whose each column vector represents an atom. The given signal  $y$  can be represented as a sparse linear combination  $s$  of these atoms. In other words, it satisfies the following relationship:

$$\|y - \Theta s\|_2 \leq \epsilon \quad (2-7)$$

where  $\epsilon$  is a very small constant. The estimated optimal signal  $s$  can be obtained by MP and OMP algorithms directly.

#### (1) Matching pursuit

The basic idea of MP is to select an atom from the dictionary matrix  $\Theta$  that matches signal  $y$  the best. Then, the residual of signal  $y$  and the selected atom is calculated. After that, the second atom is selected to match the calculated residual. With the iterations of the aforementioned procedure, signal  $y$  can finally be represented by the linear combination of all selected atoms plus the last residual value. If the final residual is in a negligible range, signal  $y$  is approximately represented as a linear combination of these selected atoms. The algorithm steps are as follows:

**Input:** sensing matrix  $\Theta$ , measurement  $y$  and sparsity  $k$



**Output:** estimation  $\hat{s}$  and residual  $r_n$

**Initialization:** initial residual  $r_0 = y$  and iteration index  $t = 0$

**Step 1:** Calculate the inner product of  $r_0$  and each atom (column) of  $\Theta$  and select the atom with the maximum absolute inner product value as the best match atom in this iterative operation, which is expressed in Equation (2-8).

$$|\langle r_0, \theta_{c_0} \rangle| = \sup_{i=1,2,\dots,N} |\langle r_0, \theta_i \rangle| \quad (2-8)$$

where  $c_0$  is the column index of the dictionary matrix;  $\langle \cdot \rangle$  is the inner product function and  $\sup(\cdot)$  is the function to get the maximum value.

**Step 2:** Decompose  $y$  into the two parts consisting of the vertical projection and the residual value of the best match atom  $\theta_{c_0}$  as shown in the following equation:

$$y = s_1 \theta_{c_0} + r_1 \quad (2-9)$$

where  $s_1$  is the weight of  $y$  in the vertical direction of  $\theta_{c_0}$  and  $r_1$  is the updated residual.

**Step 3:** Set  $t = t + 1$  and perform the same decomposition of **Step 1** and **Step 2** for the updated residual  $r_1$ .

**Step 4:** When the iteration meets the stopping criterion,  $y$  can be expressed as:

$$y = \sum_{t=0}^n \hat{s}_t \theta_{c_t} + r_n \quad (2-10)$$

where  $\hat{s}_t$  is the estimated weight and all weights generate the estimated  $\hat{s}$ .

However, the residual can only be confirmed to be orthogonal to the currently selected atom, which will result in repeatedly selected atoms. In addition, more iterations are required and the selected atoms are not optimal for reconstruction. As a result, the OMP algorithm was investigated to overcome this issue.

## (2) Orthogonal matching pursuit

In the OMP process, the residual is always orthogonal to the already selected atom. This means that an atom will not be selected twice, and the result will converge in a limited number of iterations.

**Input:** sensing matrix  $\Theta$ , measurement  $y$  and sparsity  $k$

**Output:** estimation  $\hat{s}$  and residual  $r$

**Initialization:** initial residual  $r_0 = y$ , index set  $\Lambda_0 = \emptyset$ , selected atom set  $\varphi_0 = \emptyset$  and iteration index  $t = 1$

**Step 1:** Calculate the inner product of  $r_0$  and each atom of  $\Theta$  and select the atom with the maximum absolute inner product value as the best match atom in this iterative operation. Simultaneously, find the atom column index  $c_t$  in  $\Theta$ , which is expressed as:

$$c_t = \arg \max_{i=1,2,\dots,N} |\langle r_{t-1}, \theta_{c_t} \rangle| \quad (2-11)$$

**Step 2:** Update the index set  $\Lambda_t = \Lambda_{t-1} \cup \{c_t\}$  and record the selected atom set  $\varphi_t = [\varphi_{t-1}, \theta_{c_t}]$ .

**Step 3:** Get  $\hat{s}_t$  by the methods of the least squares as expressed:

$$\hat{s}_t = \arg \min \|y - \varphi_t \hat{s}\|_2 \quad (2-12)$$

**Step 4:** Update the residual  $r_t = y - \varphi_t \hat{s}_t$  and set  $t = t + 1$ .

**Step 5:** Determine whether  $t > k$  ( $k$  is the sparsity of  $s$ ) is satisfied. If it is, stop the iteration and get the estimation  $\hat{s}$ ; otherwise, repeat **Step 1** to **Step 5**.

According to the description of MP and OMP methods to reconstruct the representation coefficients or weights, signal  $x$  also can be recovered with the orthogonal transform basis  $\Psi$  as follows.

$$\hat{x} = \Psi \hat{s} \quad (2-13)$$

As the kernel of CS, many researchers investigated related reconstruction algorithm in recent years [76]. Nevertheless, these reconstruction algorithms have many drawbacks such as high computational complexity and reconstruction distortion, so that reconstruction algorithm also needs to be improved.

### 2.1.2.3 CS applications

#### (1) Single-pixel compressive digital camera

A single-pixel camera was developed according to the compressive sensing theory by the scientists at Rice University to solve the “inefficiencies” of modern digital camera. It is a very famous CS application. Its schematic diagram is shown in Figure 2.3 [77], which only acquires  $M$  random pixel values instead of  $N$  ( $M < N$ ) to make high-quality photo storage more convenient with the smaller size.

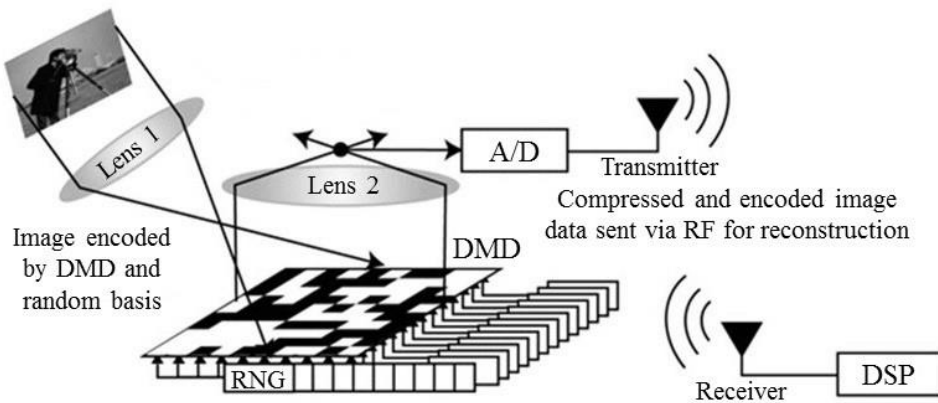


Figure 2.3 Schematic diagram of the single-pixel compressive camera

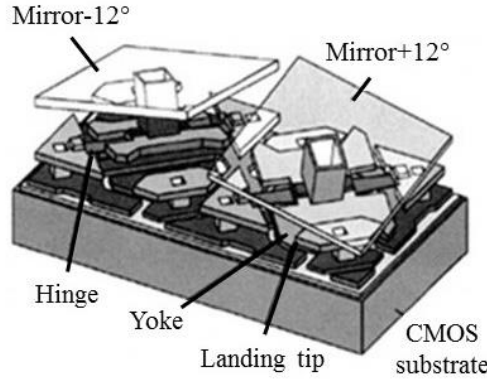


Figure 2.4 DMD micromirror structure

The first lens can be used to project the image at a suitable size on the digital micromirror device (DMD). DMD, instead of the traditional charge-coupled device or complementary metal oxide semiconductor (CMOS) sensors, consists of an array of  $N$  tiny mirrors with two directions:  $+12^\circ$  (reflecting the image towards the photo detector) and  $-12^\circ$  (reflecting the image away from the photo detector) for each one. The direction of each mirror is independently controlled by a random-number generator (RNG). After that, the second lens is employed to focus the image reflected by the DMD on a single-pixel sensor. Repeat this process to get only  $M$  pixel points transferred via radio frequency (RF) technique. At last, reconstruct the original image with tolerable distortion at the receiver end [78], [79]. Figure 2.4 displays the structure of a DMD micromirror, which will provide a reference for the hardware development of CS-based wireless sensing systems for machine condition monitoring.

## (2) Medical applications

Biological signals such as electrocardiogram (ECG) signals of the human body have the sparse features in time, frequency or wavelet domain so that the signals are compressible with CS to save the storage space. On the other hand, due to the damage to the human body caused by radiation, it is necessary to reduce the number of measurements while guaranteeing sufficient image quality. Traditionally, the medical imaging compression needs expensive instruments to acquire the entire signal and then throw away most of the useless information through compression. CS as an emerging compression method breaks the traditional Nyquist-Shannon sampling theorem, so it has been widely applied in medical imaging applications. In addition, the decrease of transferred data is conducive to the development of wireless transmission for medical images. For example, Wang *et al.* [80] proposed a compressed sensing encoder with high energy efficiency for wireless ECG systems that only uses one Analog-to-digital converter (ADC) instead of 64 channels. Craven *et al.* [81] also concentrated on researching of CS for ECG monitoring and proposed a CS architecture to successfully increase the compression ratio based on the combination of a novel redundancy removal scheme with quantization and Huffman entropy coding. Furthermore, CS is also investigated and applied for magnetic resonance imaging (MRI) signals. Chen *et al.* [82]

mainly dedicated efforts to reconstructing the images from undersampled  $k$ -sparse data relying on the wavelet tree structure which has remarkable improvements compared to the traditional compressive sensing on MRI signals.

### (3) WSN applications

With the development of smart facilities, a wireless sensor network consists of spatially distributed autonomous devices using sensors to monitor physical or environmental conditions, like temperature, humidity, pressure, location, sound, etc. [83]. However, WSNs have the disadvantages of power supply and transmission bandwidth limitation. The energy consumption is mainly affected by the number and also the period of data transferred, which means that the transmission of large amounts of data over a long period will limit the random distribution of sensors because of a considerable amount of energy consumption. Accordingly, dramatically compressing the size of data before transmission is one of the most effective and efficient ways to set up the sensor nodes without external power supply and short-term maintenance.

Due to the large demands for transmission bandwidth and data transmission power, CS, as an effective paradigm, has been exploited and further researched extensively to achieve a low sample rate but high sensing quality in WSNs. For example, Nguyen *et al.* [84] employed a high-efficient method to reduce the energy consumption for WSNs through the integration of random walk and CS to significantly prolong the network lifetime. In [85], Chen *et al.* utilized network coding and spatio-temporal compression, combined with CS to develop a clustered compression scheme which ensures the reconstruction of original data with useful signal characteristics. Wang *et al.* [86] proposed a CS-based approach to monitoring the wireless access points of vehicular networks, which successfully achieves the required estimation accuracy with low communication cost.

### (4) Machine CM applications

Condition monitoring and fault diagnosis are significantly essential in the mechanical field. Signals containing fault information, like vibration signals and acoustic emission signals, require a high sampling rate according to the Nyquist–Shannon sampling theorem to ensure the effectiveness of the acquired signals. It results in a significant increase in the volume of data transmission, storage and processing. While, CS can compress the data to save the cost of transmission and storage for the complex machinery. For example, Wong *et al.* [87] emulated a CS mechanism through resampling the acquired the time waveform of vibration signals of roller element bearings with a random Bernoulli matrix. They have concluded through SVM results that CS helped significantly reduce the wireless transmission bandwidth requirement with slight performance degradation. Yuan *et al.* [88] proposed a CS based bi-step compression method to diagnose faults of aircraft onboard and simultaneously transmit the compressed features to the ground for further investigation and maintenance. Its effectiveness has been evaluated by the hydraulic plunger pump with two common fault

types and rolling bearings working under fluctuant conditions [89]. In addition, the authors have compared the accuracy of classification with CS compressed data and the raw data with other current approaches, and the compressed data can be reconstructed with little loss for other application in [89]. Tang and Wang *et al.* investigated several compression methods based on CS framework for fault detection of bearings, including constructing an over-complete dictionary on the basis of faulty vibration characteristics [90], detecting harmonics of faulty bearings with CS based compressed vibration signals [91], applying CS on peaks calculated by the peak-hold-down-sample algorithm [92] and combining statistical features in time domain with CS realizing two-stage compression [93]. Ahmed and Nandi *et al.* also made significant contributions to the data compression with CS in CM of bearings in recent years. They compared the classification performance of with CS based compressed data directly using the logistic regression classifier and with features extracted by PCA or LDA from the CS based compressed data in [94]. Furthermore, the extracted PCs and discriminant components were combined by CCA to express their correlation for the next step of classification with SVM, which showed very high classification accuracy with only three-stage processing [95], [96]. Additionally, they proposed the multiple measurement vector CS (MMV-CS) approach to extract features from the CS based compressed vibration signals of bearings. Then the extracted features were selected by combining with different algorithms, like Fisher score [97], [98], Laplacian score [98], [99], neighbourhood component analysis etc. [100], sparse autoencoder [101] and Relief-F, Pearson correlation coefficients and Chi-Square [98] for further feature dimension reduction. Finally, only a few features were preserved as the input of the various intelligent classifiers such as SVM, multi-class SVM, artificial neural networks (ANN) and deep neural networks (DNN). All developed approaches combined by these algorithms can effectively distinguish different fault types of bearings with the significantly reduced data volume. Ma *et al.* [102] proposed a Bayesian CS framework to compress vibration signals via the K-SVD dictionary learning method, then reconstruct original signals with the Laplace prior model. We also proposed frequency shift and envelope analysis based CS scheme to monitor the condition of rolling element bearings with the data compressed by several thousands of times compared with the raw signals [103]. Some researchers focused on the reconstruction algorithms to recover the received compressed data for other future analysis. For instance, a novel algorithm was proposed for the reconstruction of compressed vibration signals with CS and achieved remote machine CM [104]. Moreover, its advancements in wireless communication were introduced for detecting bearing and gear faults in rotating machines. In [70], the authors improved the OMP algorithm and exploited it as the stabled reconstruction method which could effectively improve its convergence speed in CS. Additionally, the investigated CS method has the potential to compress sparse sound vibration data of the aircraft for structural CM.

The CS applications on machine CM mainly include signal compression, feature extraction, fault classification and raw signal reconstruction with the acquired data at the receiving end. Its application at the acquisition end is limited by signal sparsity and commutating performance, so the CS based hardware implementation is a crucial challenge for machine CM at present [72].

## **2.2 Summary**

This chapter makes the extensive literature review of data compression methods including traditional uniform sampling and advanced non-uniform sampling methods performed at both data acquisition and receiving ends for machine CM. The traditional pre-transmission data compression methods can reduce the amount of data before transmission to save transmission bandwidth and power consumption of the sensor nodes, as well as the storage space. However, most of them cannot compress data according to the characteristics of signals, and useful information is easily lost during compression process or the compressed data cannot be recovered at the receiving end. The post-transmission data compression methods cannot solve the limitation of wireless transmission and power consumption, but they will reduce the amount of data to save the storage space. CS as an advanced compression method has the potential to implement sampling with a lower rate than the Nyquist-Shannon sampling theorem at the acquisition end. The introduction and overview of CS show its superiority in a wireless real-time machine CM system compared with other traditional compression methods. The review provides the basis for developing and improving data compression algorithms for CM of various machines in this thesis.

---

## **CHAPTER 3    LEAKAGE DETECTION IN PIPING SYSTEMS WITH AN INTELLIGENT WIRELESS VIBRATION SENSING SYSTEM**

---

*WSNs attract more and more attentions in Industry 4.0 in these years because of their portability, wide distribution, programmability, remote controllability and stability. Piping systems are broadly applied in industry and home. The leakage induced by corrosion, weather, or man-made damage can lead to serious consequences such as explosion disasters, significant damage of industrial equipment and waste of resources. As a non-intrusive detection method, vibration shows great potential for detecting leakage in piping systems. Therefore, an intelligent wireless vibration sensing system is developed to monitor the condition of piping systems efficiently and efficiently based on data compression. Statistical features are extracted on the sensor node to reduce the amount of data for transmission, hence reducing power consumption and increasing the battery life of the wireless sensor node.*

### 3.1 Vibration Mechanisms of Piping Systems

Fluid supply systems have extremely complicated vibration sources due to their complex structures with valves, corners, clamps at different positions, as well as the different size and materials of pipes, etc. When the fluid passes through different parts of a piping system, the vibration frequency and amplitude caused by various vibration sources are different. Therefore, the comprehension of the vibration mechanisms of piping systems is significantly important and essential to monitor the status of the piping systems. So that the vibration mechanisms of the fluid supply piping systems will be presented and discussed from different views as follows.

#### 3.1.1 Microscopic Interpretation

The fluid is composed of a large number of molecules. From a microscopic point of view, the generation of vibration closely relates to the force of molecules acting on the pipe wall.

Generally, molecules randomly collide against the pipe wall with most of them moving towards the same direction with the fluid flow. In this process, a small portion of the kinetic energy of the fluid molecules is converted into heat energy and dissipated in the environment by the heat transfer effects of the pipes. Simultaneously, most kinetic energy will be converted into potential energy in the form of pressure oscillations [105] which can provide dynamic forces from the inner wall of the pipe. After that, the pipe wall can expand at a slight deformation under the action of pressure, then vibrations are induced on the external surface of the pipe. According to the kinetic analysis, when the fluid flow is fast, the molecules hit randomly against the pipe wall at a high speed resulting in a large vibration amplitude because of the high pressure inside the pipe wall.

Figure 3.1 describes the pipe wall vibration principle from a micro perspective. The black arrow represents the fluid flow direction and blue arrows point to the random motion direction of each fluid molecule representing with blue dots.

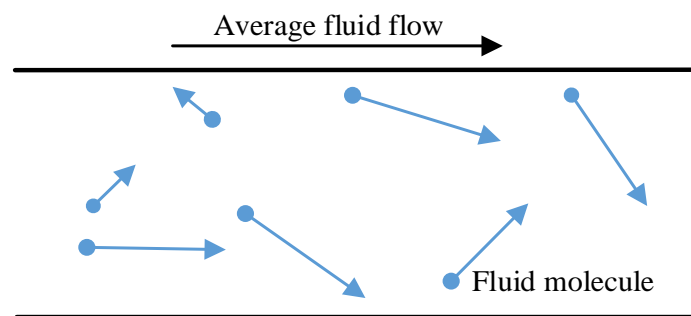


Figure 3.1 Representation of fluid molecule motion in a pipe from a micro perspective

#### 3.1.2 Natural Frequencies

Natural frequencies, also called resonance frequencies, are specific frequencies, at which the system could vibrate with a larger amplitude compared to other frequencies



under the small driving force due to the energy accumulation effects. In general, a system has numerous modes that can be characterised by the modal parameters of natural frequencies, modal shapes and damping ratios. However, the longitudinal modes and the circumferential modes are mainly considered in vibration of the pipe as illustrated in Figure 3.2, where  $m$  is the number of longitudinal half-waves in the longitudinal mode shape,  $n$  is the number of circumferential waves in the circumferential mode shape. As the fluid flow generates a wide range of excitation, the longitudinal mode and the circumferential mode can be excited at a low frequency and a high frequency respectively, which can cause a more pronounced vibration amplitude.

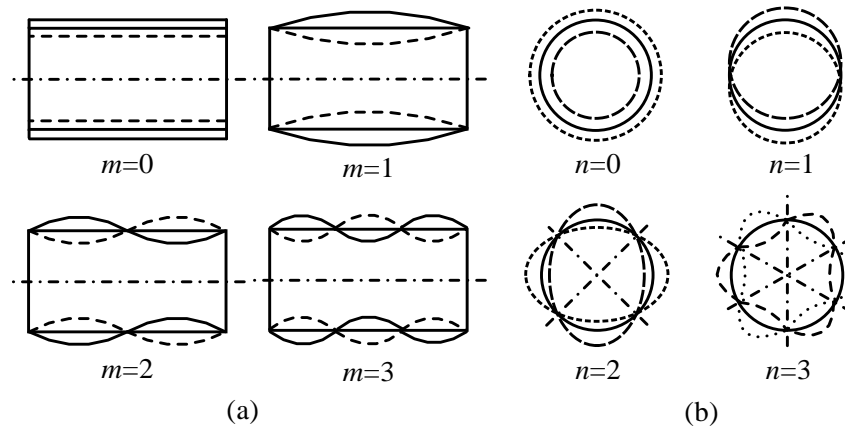


Figure 3.2 Pipe vibration modes: (a) longitudinal modes, and (b) circumferential modes

To calculate the natural frequencies of an infinitely long pipe with the cylindrical shell, Blevins proposed the following Equations (3-1) and (3-2) [106]–[108].

$$f_i = \frac{\lambda_i}{2\pi R} \left[ \frac{E}{\gamma(1-\nu^2)} \right]^{1/2} \quad (3-1)$$

$$\lambda_i = \frac{1}{12^{1/2}} \frac{h}{R} \frac{i(i^2-1)}{(1+i^2)^{1/2}} \quad i = 2, 3, 4 \dots \quad (3-2)$$

where,

$f_i$  is the natural frequency of the pipe, Hz;

$\lambda_i$  is the frequency factor, dimensionless;

$R$  is the mean radius of the pipe wall, inches;

$\gamma$  is the mass density of the pipe material, lb-sec<sup>2</sup>/in<sup>4</sup>;

$\nu$  is the Poisson's ratio, dimensionless;

$h$  is the pipe wall thickness, inches;

$i$  is the mode number, integer.

When the fluid flows through a pipe to cause excitations around natural frequencies of the piping system, the vibration amplitude can be enlarged to give more features for the piping system condition monitoring.

### 3.1.3 Turbulent Flow

The laminar flow inside water pipes can be destroyed by the small whirlpool when the flow rate significantly increases. Irregular fluid motion or mixing produces a velocity component that is perpendicular to the axial direction, which generates the turbulent flow or turbulence inside the pipe. Typically, fluid flow near a solid surface or fluid flow with a high liquid viscosity is laminar. However, most fluid flows are turbulent.

Generally, turbulence occurs at high Reynolds numbers and is dominated by inertial forces [109]. The Reynolds number is an essential parameter used to estimate the flow patterns as shown in Equation (3-3).

$$Re = \frac{\rho v L}{\mu} \quad (3-3)$$

where,

$\rho$  is the density of the fluid, kg/m<sup>3</sup>;

$v$  is a fluid velocity, m/s;

$L$  is a characteristic linear dimension, m;

$\mu$  is the dynamic viscosity of the fluid, Pa·s or N·s/m<sup>2</sup> or kg/(m·s).

When the Reynolds number is greater than 2000, the effect of the inertia force can be higher than the viscous force, making the fluid flow relatively unstable, thereby increasing the small change in velocity as well.

Figure 3.3 clearly shows that turbulence is formed due to disordered and irregular flow fields caused by the irregular shape of the valve. The fundamental characteristic of turbulence is the randomness of fluid movements. Both transverse pulsations and reverse movements relative to the fluid motion of a turbulent flow change rapidly with time. Therefore, pipes are easier to vibrate and deform because of heat transfer, kinetic energy conversion, etc. induced by the random motion of fluid flows.

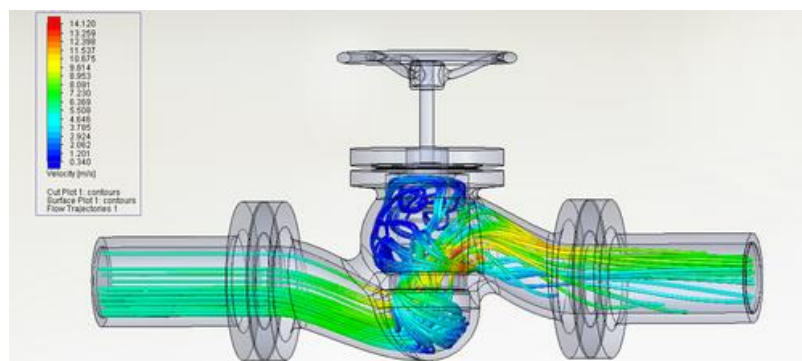


Figure 3.3 Simulation of turbulence at a valve

### 3.1.4 Frictional Effects

Another cause of pipe wall vibration is the internal friction [110] generated by the relative motion of the fluid inside of a pipe to the pipe wall. Internal friction, as an

inherent physical property of fluids, exists in the form of friction, which is produced by the resistance as deformations occur in the internal fluid. The friction between the pipe wall and fluid flow generates a force of friction inducing vibration. Frictional effects of the fluid against the pipe wall is simulated in Figure 3.4, which illustrates that frictional effects of liquid motion are the result of momentum transfer and cohesion function between the liquid molecules.

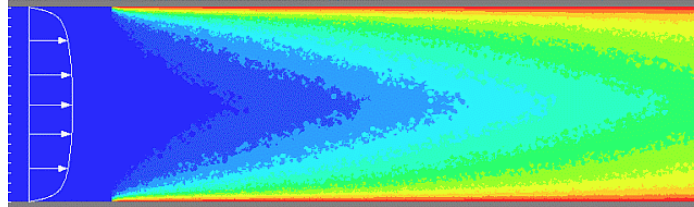


Figure 3.4 Simulated frictional effects of the fluid against the pipe wall

## 3.2 Design of the CM System with MEMS

### 3.2.1 MEMS Sensor Nodes

MEMS is a technology of micro high-tech devices by merging sensing, controller and communication functions based on the micro and nanotechnology. A MEMS sensor system is highly integrated with micromechanical and microelectronic functions because they are typically made of silicon-based materials and the semiconductor integrated circuits. It extensively plays an essential role in industrial, aerospace, military engineering, medical, automotive and other environments as shown in Figure 3.5 with characteristics of tiny size, low cost, high reliability, easy integration, mass production and so on compared with the traditional acquisition systems [111]. Therefore, MEMS has been successfully utilized in a variety of fields in recent years, also including machine CM.

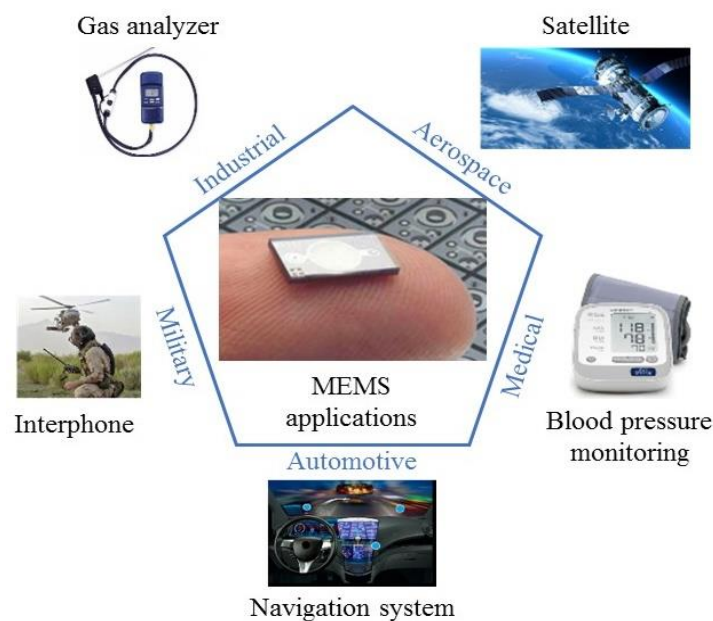


Figure 3.5 MEMS applications

In general, a sensor node consists of four main components, a sensing unit, a processing unit, a communication unit and a power unit, to achieve the function of data acquisition. A typical structure of a sensor node is shown in Figure 3.6.

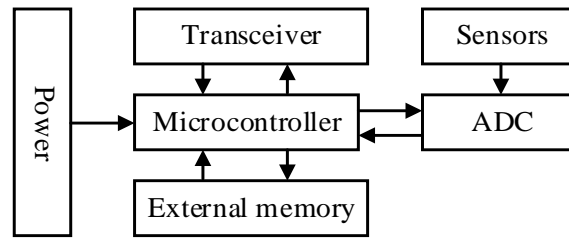


Figure 3.6 Typical structure of a sensor node

### 3.2.1.1 Sensing Unit

The function of a sensing unit is detecting different physical quantities in the surrounding environment with a variety of sensors, such as temperature, humidity, pressure and vibration. For example, MEMS accelerometers are classified into piezoresistive sensors and capacitive sensors according to the working principle. They are developed and manufactured to capture vibration signals [112]. Figure 3.7 gives the structures of these two different types of accelerometers [3], [4].

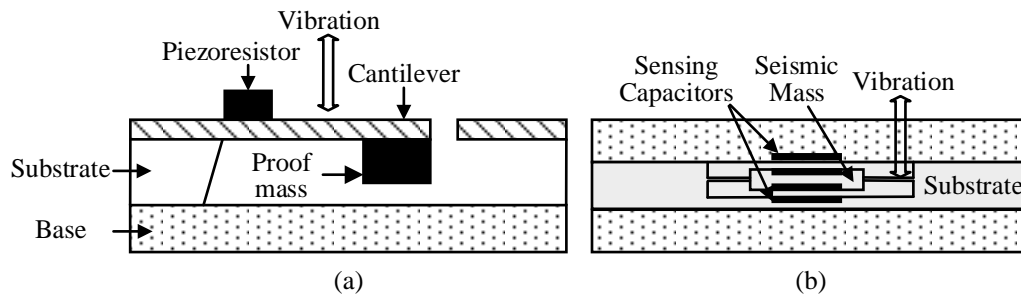


Figure 3.7 Typical MEMS accelerometer structures: (a) piezoresistive sensor, and (b) capacitive sensor

For the piezoresistive sensors shown in Figure 3.7 (a), as the sensor vibrates together with the objects, the inertia of the mass will cause a bending in cantilever, which leads to the stress making the piezoresistor resistance change. Accordingly, the output will be related to the acceleration of objects. The principle of the capacitive sensor shown in Figure 3.7 (b) is similar to the piezoresistive one. When the sensor vibrates with the shake of the objects, the movement of the seismic mass caused by the inertia will change the electric capacity between the mass and those two fixed sensing capacitors, which can indicate the acceleration of objects [112].

For the designer, some parameters should be carefully considered before selection of the appropriate sensing components, like the sensitivity, resolution, range and bandwidth. Table 3.1 shows the principal parameters of commonly used accelerometers manufactured by some famous companies. The least significant bit every gravity (LSB/g) gives the sensitivity of digital results of accelerometer analogue output after ADC.

Table 3.1 Typical accelerometers and parameters

Company	MEMS	Range ( $g=9.8 \text{ m/s}^2$ )	Sensitivity (LSB/g)	Resolution (bits)	Max. Data Rate (Hz)
Analogue Devices Inc.	ADXL345	$\pm 2/4/8/16$	256	10-13	3200
	ADXL313	$\pm 0.5/1/2/4$	1024	10-13	3200
Bosh Sensortec	BMA280	$\pm 2/4/8/16$	4096	14	2000
InvenSense	MPU-6500	$\pm 2/4/8/16$	16384	16	4000
Kionix Inc.	KX022	$\pm 2/4/8$	16384	16	1600
	KX122	$\pm 2/4/8$	16384	16	25600
STMicroelectronics	LIS2DS12	$\pm 2/4/8/16$	16384	16	6400
	LIS3DSH	$\pm 2/4/6/8/16$	16384	16	1600

### 3.2.1.2 Processing Unit

A processing unit consists of the memory and the microcontroller with the primary functions of storage and computation, respectively. Due to its flexibility, a microcontroller can connect with other units through programming.

ARM Cortex-M is a group of ARM processor cores, which can combine with a random-access memory (RAM), read-only memory (ROM), register and other peripherals to integrate into a microcontroller. After ARM11, this product was named by “Cortex”, which can be divided into three series: A (for those high-end operating systems and applications based on virtual memory), R (for real-time system) and M (for microcontrollers). ARM Cortex-M products and their advantages are described in Figure 3.8.

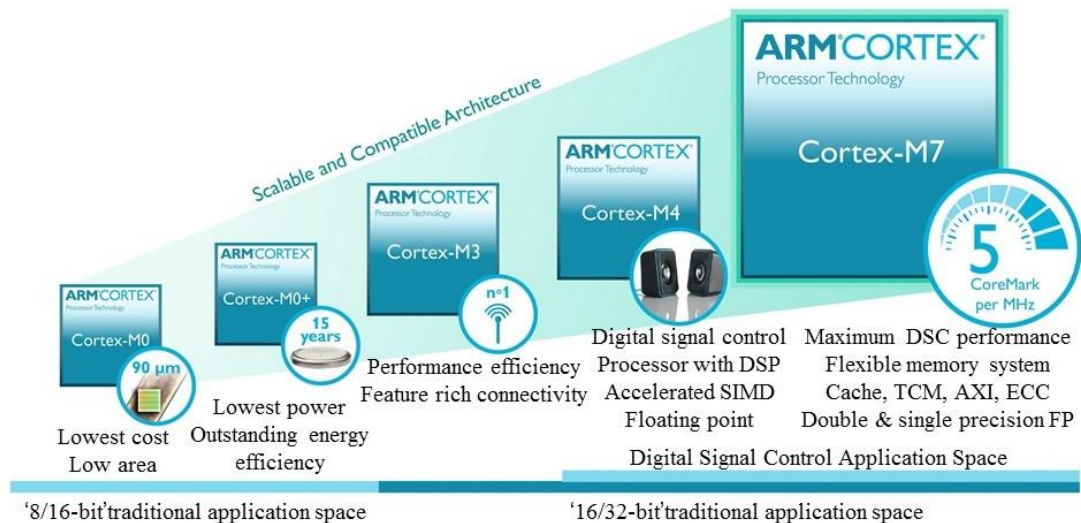


Figure 3.8 ARM Cortex-M series products and advantages

Cortex-M0, with small chip size and low power consumption, is the smallest processor among ARM series. Compared with the Cortex-M0, M0+ improves performance with better compatibility and further reduces power consumption in the condition of

retaining all data instructions. However, Cortex-M3 not only completely exploits hardware-based interrupt handling but also has high computational performance. Cortex-M4 strengthens the computational capabilities by adding functions of digital signal control, floating point, parallel computing and so on. Finally, Cortex-M7 is the most high-performance member in the ARM Cortex-M family. It can significantly enhance the performance of the operation and digital signal control for high-end embedded applications, such as automotive electronics, smart home and other industrial applications.

### 3.2.1.3 Communication Unit

A communication unit has the function of information communication and transformation with a base station or neighbour sensors [113]. The wireless universal transmission media include RF, optical communication (like the laser) and infrared radiation, etc. In general, idle mode or sleep mode will be activated to reduce power consumption when the sensor node has no task. Table 3.2 compares the functions of traditional wireless transmission technologies [114], which can provide a reference for the choice of transmission way.

Table 3.2 Comparison of traditional wireless transmission technologies

Technique	Transmission Rate	Transmission Distance	Power Consumption			Features
			Sleep ( $\mu$ W)	Transmit (mW)	Receive (mW)	
Low Power Wi-Fi	54Mbps	1km	300	350	270	High speed, high reliability and high-power consumption
BLE 5.0	2Mbps	Up to 300m in theory	-	-	-	High speed, long distance, wide bandwidth, ultra-low power consumption and high compatibility
BLE 4.2	1Mbps	Up to 100m, Normally operate within 10m	8	60	53	Low power consumption, low cost, high security and low latency
ZigBee	250kbps	10 to 100m	4	72	84	Low power consumption, low cost, low complexity and self-organization
EnOcean	125kbps	Up to 30m	0.60	99.0	72.0	Energy harvesting based and ultra-low power consumption
Z-wave	40kbps	Indoor: 30m/40m Outdoor: 100m	3	70	65	RF-based, low cost, low power consumption, low radiation, anti-interference and high reliability
ANT	60kbps	30m at 0dBm	3	110	75	Ultra-low power consumption, high flexibility and proprietary

A Bluetooth low energy (BLE) wireless transmission module was selected because of its advantageous characteristics, including low cost, high transmission rate, convenient networking, low power consumption, high communication security and better transmission performance.

#### 3.2.1.4 Power Unit

The power consumption should be seriously taken into consideration in the progress of development of sensor nodes. The function of a power unit is supplying power for a sensor node with a battery or a universal serial bus (USB) interface. Batteries can be mainly categorised into the primary batteries and secondary ones based on the characteristics of reuse. The primary battery cannot be recharged and will be discarded until the energy depletion, such as the dry cell and zinc-air battery. It is really a waste of resources and easy to pollute environment. On the contrary, the secondary battery can be recharged with a long service life, like the storage cell and lithium cell, which are widely used in mobile phones, automobiles and other industrial products. Some typical battery products are displayed in Figure 3.9. There is no doubt that a rechargeable battery will be chosen to charge the sensor node, which also may be charged by the energy harvested from the monitored system or ambient environment.



Figure 3.9 Battery products: (a) primary batteries, and (b) secondary batteries

### 3.2.2 Wired MEMS Real-Time CM System

#### 3.2.2.1 Wired Sensor Node Design and Drive

According to the units introduced in the previous subsection, the evaluation boards of Feather M0 Adalogger and ADXL345 shown in Figure 3.10 from Adafruit industry are selected as the microcontroller and sensing unit, respectively.



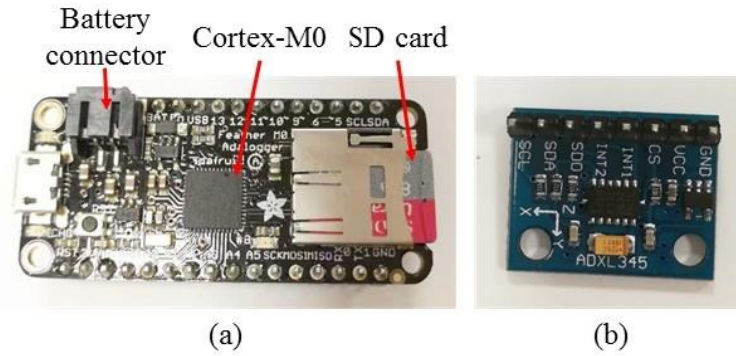


Figure 3.10 Evaluation boards: (a) Feather M0 Adalogger, and (b) ADXL345

The core of Feather M0 Adalogger has an ATSAM21G18 ARM Cortex-M0 processor with 256K of FLASH and 32K of RAM [115]. A secure digital (SD) card inserted into the card slot on the board can store the data collected instead of the serial port communication. As shown in Figure 3.11, ADXL345 is a three-axis digital accelerometer with the board size of 25mm×19mm×3.14mm. Its resolution is 13 bits with the corresponding measurement range of  $\pm 16g$ . The digital filter, ADC, power management and other components are integrated into this evaluation board. It is accessible through either a serial peripheral interface (SPI) or inter-integrated circuit (I<sup>2</sup>C) digital interface to communicate with the processing unit.

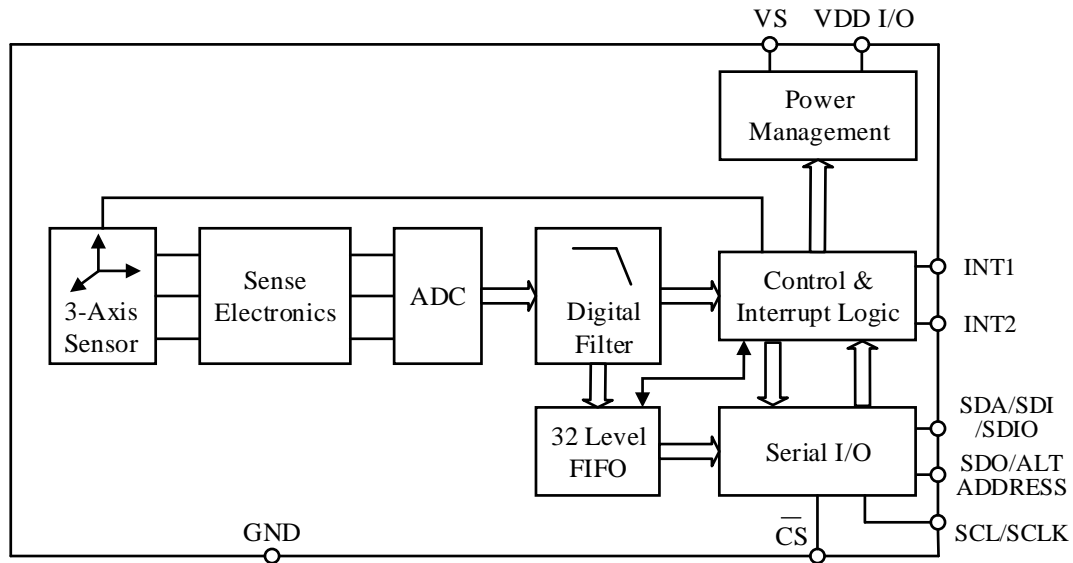


Figure 3.11 Diagram of the ADXL345 evaluation board

The data transmission rate of I<sup>2</sup>C is 100 kbps, 400 kbps and 3.4 Mbps for standard mode, fast mode and high-speed mode, respectively. While it often goes over 10 Mbps at the SPI transfer mode. Table 3.3 lists the wire connection way between ADXL345 evaluation board and Feather M0 Adalogger with the SPI mode. Figure 3.12 describes the physical connection of the sensor node.



Table 3.3 Wire connection between ADXL345 and Feather M0 Adalogger

Function	ADXL345	Feather M0 Adalogger
Power Supply	+3.3V	3V
Ground	GND	GND
Clock	SCL	SCK
Chip Select	CS	Pin 10
Data Transmission	SDA (Serial data in)	MOSI (Master out and slave in)
Data Transmission	SDO (Serial data out)	MISO (Master in and slave out)
Interrupt	INT1	Pin 12



Figure 3.12 Physical connection of a sensor node

After the successful wire connection, the sensor node can be driven by an open-source software named Arduino with the interface shown in Figure 3.13.

Through programming with this software, the microcontroller can give instructions to the sensor, receive and process the data coming from the sensor. The serial transmission rate, acceleration range, sampling frequency and the cut-off frequency of the high-pass filter can be set manually according to the requirements. The maximum acceleration range and sampling frequency are  $\pm 16g$  and 3200 Hz, respectively. It means the sensor node is designed successfully if the data is received without data loss by the computer through the micro serial port.

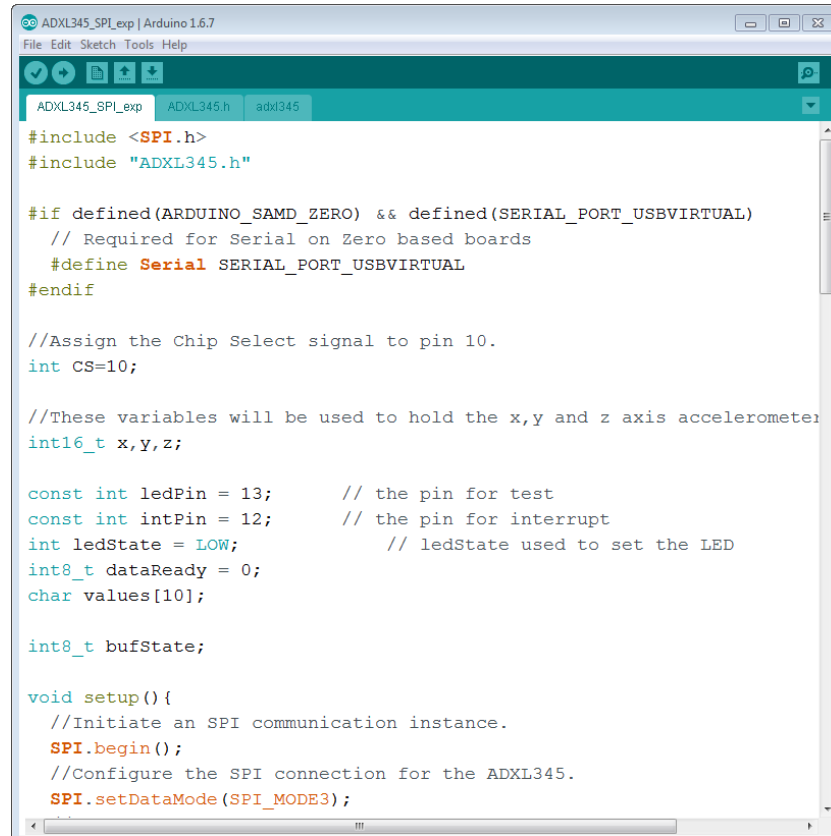


Figure 3.13 Arduino interface for programming to drive the sensor node

Unfortunately, data loss exists in the progress of testing. The reason is the wrong data resolution protocol of the micro serial port in Feather M0 Adalogger. It has been solved by adding a communication module to collect data from the mini serial port instead of the micro serial port. The structure of the improved wired sensor node is illustrated in Figure 3.14.

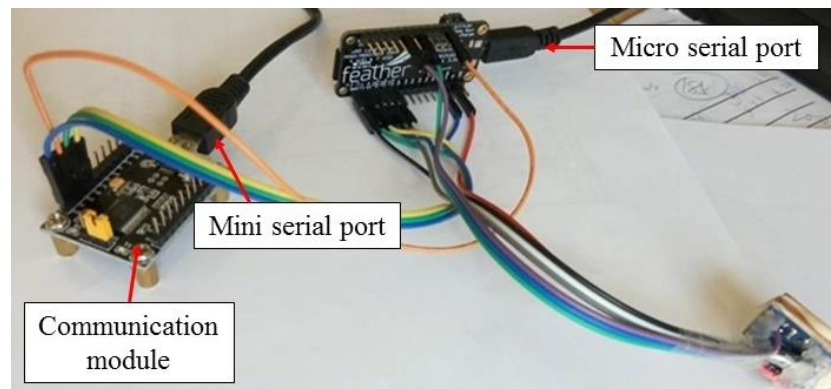


Figure 3.14 Improved wired sensor node structure

### 3.2.2.2 Data Acquisition GUI Design

The designed sensor node requires a data acquisition interface to monitor the condition of objects in real time and save data for further analysis. The MATLAB graphical user interface (GUI) is a user-friendly tool to design such a CM system. As shown in Figure 3.15, the designed data acquisition GUI has the functions of port scan, parameter setting,

raw data saving, status display, time waveforms and spectra display, as well as RMS value calculation. Especially, the instruction of a cut-off frequency setting will be sent to Feather M0 Adalogger directly to operate a high-pass filter. The filtered signal can be clearly illustrated in both time domain and frequency domain.

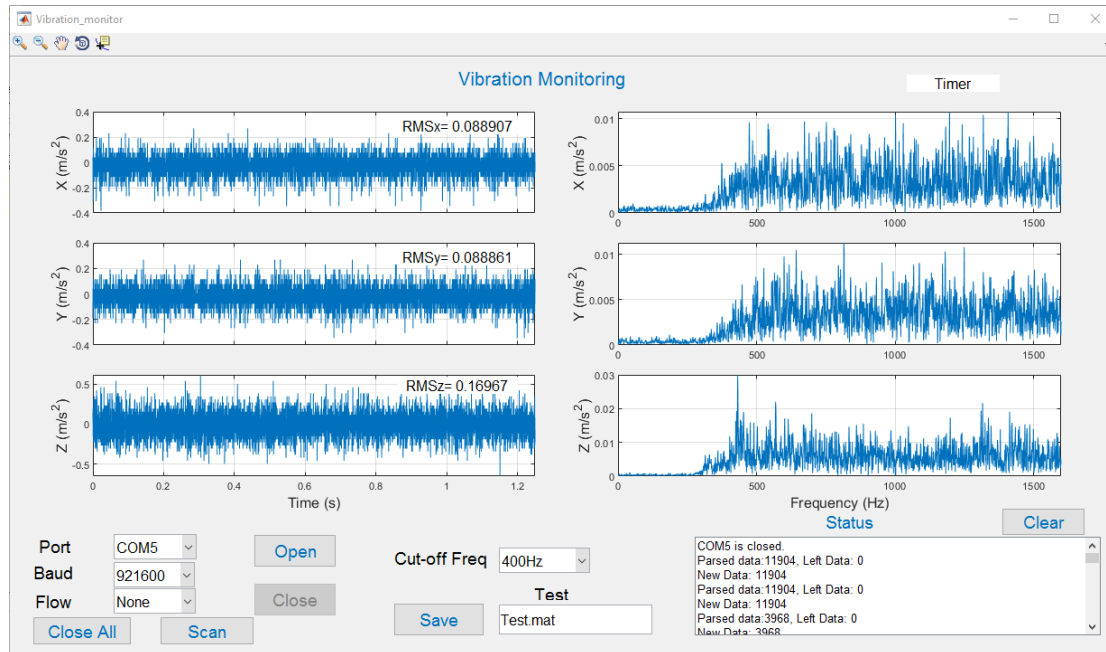


Figure 3.15 Designed data acquisition GUI

### 3.2.3 Sensor Calibration

#### 3.2.3.1 Experimental Devices and Setup

For the designed sensor node, it is necessary to calibrate the accuracy of the MEMS accelerometer. A commonly used piezoelectric accelerometer, CA-YD-182A with high accuracy and sensitivity, is selected to be the reference for ADXL345 because of its high sensitivity, stability and robustness. As illustrated in Figure 3.16, both ADXL345 and CA-YD-182A are fixed in a manufactured metal base and installed on the top of the shaker. They can measure the vibration of the shaker activated by a digital signal generator.

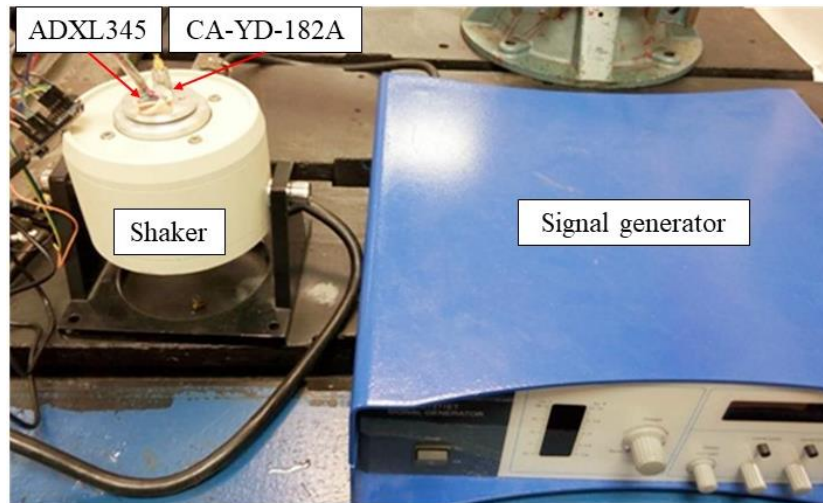
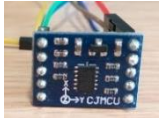



Figure 3.16 Experimental devices for sensor calibration

Table 3.4 compares the parameters of ADXL345 and CA-YD-182A. It is clear that the performance of CA-YD-182A is better than that of ADXL345. However, ADXL345 is superior to CA-YD-182A in terms of economic, programmability and integration, which benefits for production of portable and low-cost products.

Table 3.4 Parameters of ADXL345 and CA-YD-182A

Accelerometer	ADXL345 	CA-YD-182A 
Sensitivity	256 LSB/g (or 3.9 mg/LSB)	2.05 mV/(m/s <sup>2</sup> )
Max. Range	±16g	2500 m/s <sup>2</sup>
Bandwidth	0.1 to 1600 Hz	1 to 10,000 Hz
Price	£0.99	£190.56

In this calibration test, the sampling frequency and measurement range of the designed sensor node is set to 1600 Hz and  $\pm 2g$ , respectively. Because the shaker shakes in the vertical direction, so that only signal in Z-axis will be saved and analysed. For CA-YD-182A, it is a one direction accelerometer with the frequency of the data sampling is set to 96 kHz. The signal generator supplies a series of excitation signals, including 25.3 Hz, 50.0 Hz, 99.7 Hz, 149.9 Hz, 199.9 Hz, 300.3 Hz, 399.9 Hz, 499.9 Hz, 600.0 Hz, 650.0 Hz and 700.1 Hz. The sampling duration is 20 seconds.

### 3.2.3.2 Results and Discussion

As the sampling frequency set to collect signals from CA-YD-182A is too high, a low-pass filter with the cut-off frequency of 800 Hz is applied. Figure 3.17 displays the spectra of vibration signals collected by ADXL345 and CA-YD-182A lined with blue and red colours, respectively. The difference in vibration frequencies obtained from

these two accelerometers is getting larger and larger. Moreover, the amplitude of vibration collected by the MEMS accelerometer is a little smaller.

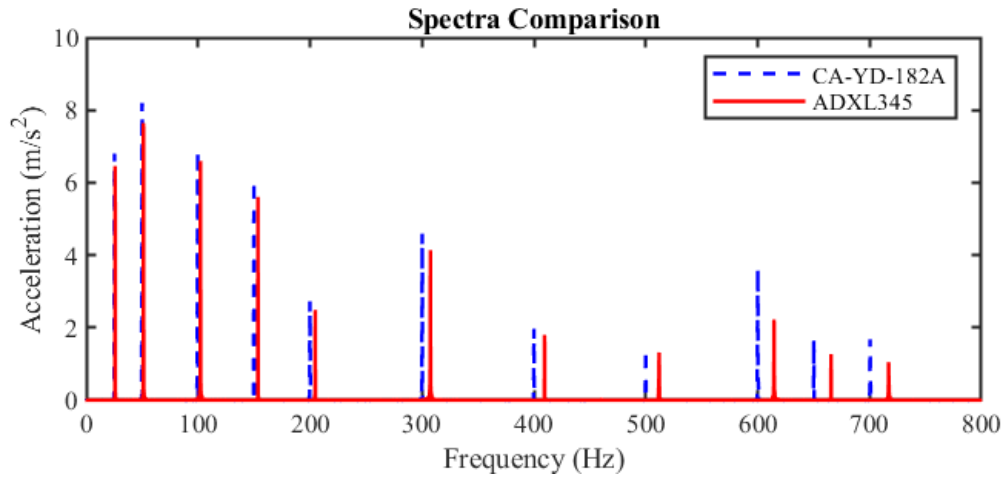


Figure 3.17 Spectra of two accelerometers for sensor calibration

Both frequency and amplitude values of vibration collected by two accelerometers are recorded in Table 3.5. The calculated frequency error for CA-YD-182A is 0.00% to 0.05%. On the contrary, it changes between 2.37% and 2.46% for ADXL345. It demonstrates that ADXL345 has lower precision than CA-YD-182A due to the accuracy and error of production, but ADXL345 can be used for vibration monitoring because of the frequency error can be fixed to approximately 2.40% according to the linear variation tendency.

Table 3.5 Spectra details of CA-YD-182A and ADXL345

Generator Frequency (Hz)	CA-YD-182A			ADXL345		
	Frequency		Amplitude Value (m/s <sup>2</sup> )	Frequency		Amplitude Value (m/s <sup>2</sup> )
	Value (Hz)	Error (%)		Value (Hz)	Error (%)	
25.3	25.3	0.00	6.802	25.9	2.37	6.458
50.0	50.0	0.00	8.193	51.2	2.40	7.642
99.7	99.7	0.00	6.889	102.1	2.41	6.593
149.9	149.9	0.00	6.071	153.5	2.40	5.602
199.9	199.9	0.00	2.726	204.7	2.40	2.491
300.3	300.3	0.00	4.593	307.7	2.46	4.135
399.9	400.1	0.05	1.969	409.7	2.45	1.799
499.9	500.0	0.02	1.399	512.0	2.42	1.317
600.0	600.1	0.02	3.621	614.6	2.43	2.224
650.0	650.1	0.02	1.657	665.6	2.40	1.272
700.1	700.3	0.03	1.680	717.0	2.41	1.047

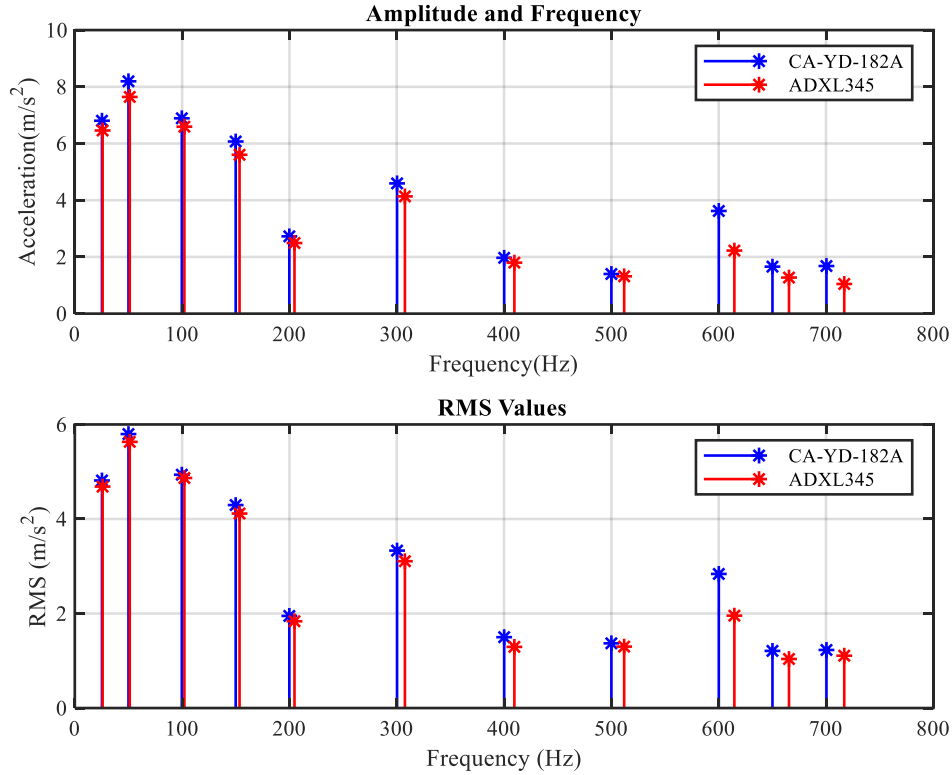


Figure 3.18 Amplitudes and RMS values at different frequencies

Figure 3.18 shows the good consistency of frequency, amplitude and RMS values of CA-YD-182A and ADXL345, excluding the amplitude and RMS values at 600 Hz. Both the peak amplitudes and RMS values of ADXL345 are smaller than those of CA-YD-182A. But the RMS values are much closer for these two sensors. It is clear that performance of ADXL345 is a little worse at higher frequencies.

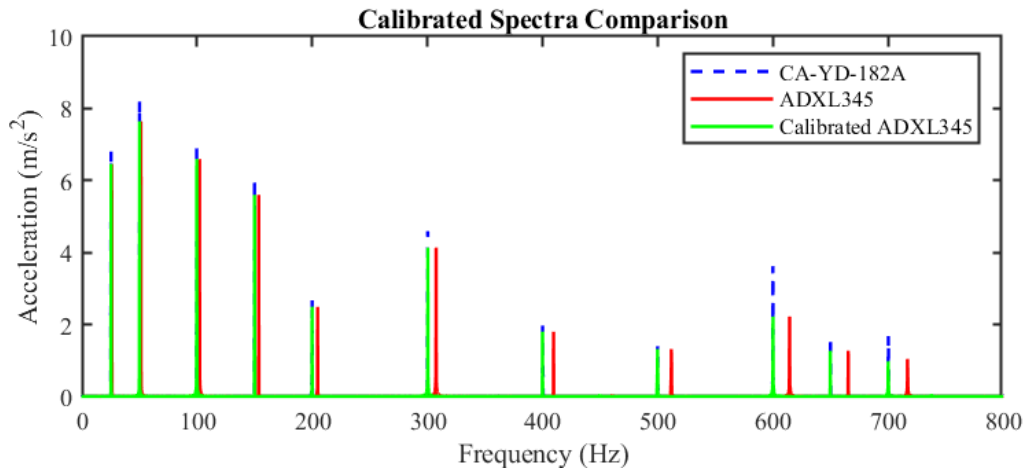


Figure 3.19 Spectra of ADXL345 with frequency calibrated

Based on the results discussed for Table 3.5, ADXL345 can be calibrated with the frequency shown in Figure 3.19. Although performance of ADXL345 is not as good as CA-YD-182A, it is still widely applied due to its advantages of high economy, easy

programming and integration. As a result, the MEMS accelerometer ADXL345 can be used for vibration monitoring instead of piezoelectric sensors to save cost.

### 3.2.4 Leakage Detection in Domestic Piping Systems

#### 3.2.4.1 Test Facilities and Experiments

The experimental platform is a domestic piping system with copper pipes (copper pipe diameter is 15mm) installed on the wall in the laboratory. The structure of pipes at the corner and near taps is very complex so that it is easier to generate turbulences leading to large vibration amplitudes in these positions. Besides, the long straight pipe is prone to the pipe resonance. As a result, two detection positions, Position 1 and Position 2, were selected between the valve and two taps (Taps 1 and 2) as shown in Figure 3.20. Position 1 is very close to Tap 1 at the corner with ADXL345 and CA-YD-182A installed on the top and bottom of the pipe, respectively. Position 2 is located at the middle of the long straight pipe with a remote distance to the valve and Tap 2, where two accelerometers were installed neighbourly. The different size of leaks was simulated by independently opening Tap 1 or Tap 2 with different angles to detect the signals at Position 1 and Position 2, respectively, which is shown in Table 3.6.

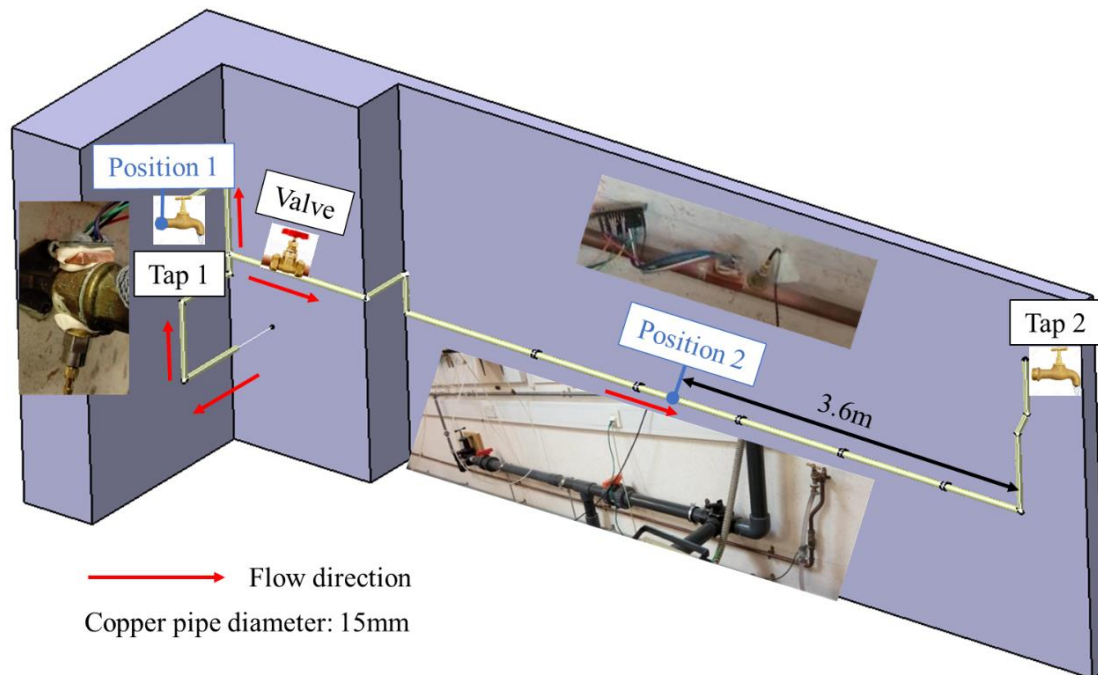


Figure 3.20 Experimental platform structure diagram

Table 3.6 Flow representations with various tap rotation angles

Flow Rate	Rotation Angle of Tap 1 (Position 1)	Rotation Angle of Tap 2 (Position 2)
No Flow	0°	0°



<b>Small Flow</b>	30°, 60°, 90°, 120°	15°, 30°, 45°, 60°, 75°, 90°, 105°
<b>Medium Flow</b>	150°, 180°, 210°, 240°, 270°, 300°, 330°, 360°	120°, 135°, 150°, 165°, 180°, 195°, 210°, 225°, 240°
<b>Large Flow</b>	450°, 540°, 630°, 720°	270°, 300°, 330°, 360°

The vibration signals were synchronously collected by both CA-YD-182A with the DAQ system and ADXL345 with the designed sensor node at the sampling frequency of 96 kHz and 3200 Hz, respectively. Each test was carried out three times with 20 seconds of data recorded.

### 3.2.4.2 Results and Discussion

#### (1) Experimental results at Position 1 (MEMS)

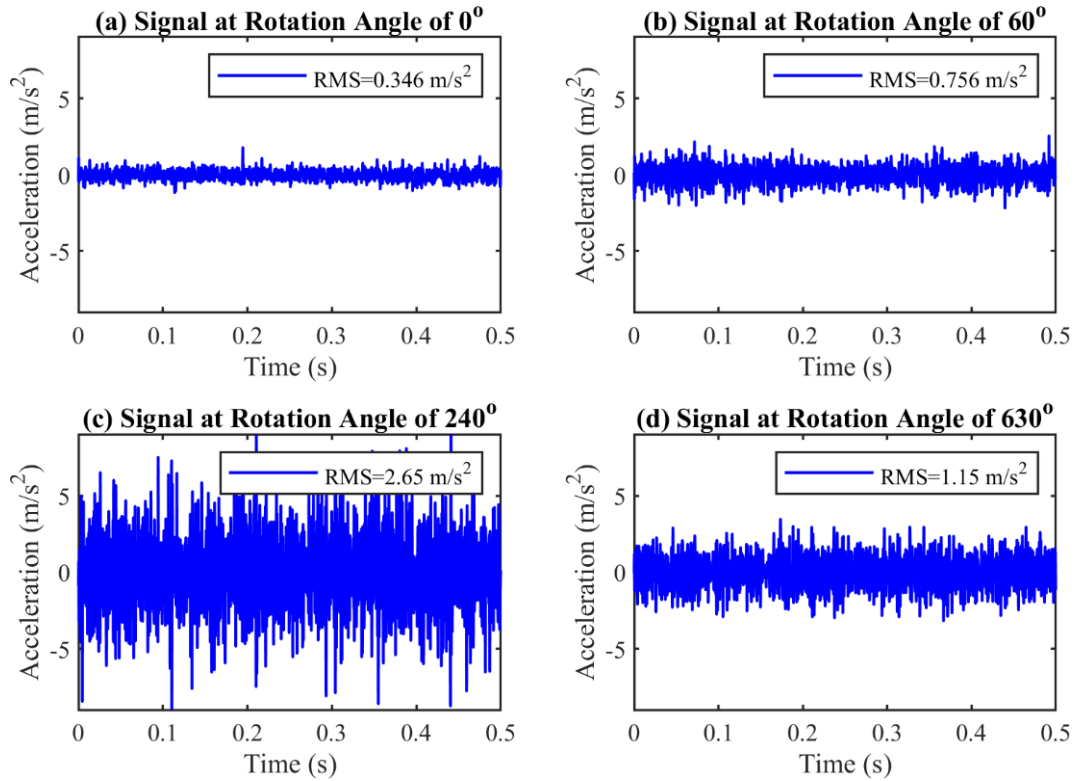


Figure 3.21 Waveforms of four different leakage conditions at Position 1

Figure 3.21 illustrated vibration waveforms at Position 1 of the pipes under four different conditions, representing zero leakage, small leakage, medium leakage and large leakage, respectively. It can be seen that the RMS value of zero leakage is the smallest. RMS value at the condition with a rotation angle of 60° is about twice as high as that with a rotation angle of 0°. The vibrational RMS value increases to a significantly high value, 2.65 m/s<sup>2</sup>, at the condition with a rotation angle of 240°. However, this value declines to 1.15 m/s<sup>2</sup> at the condition with a rotation angle of 630°.



It means that RMS values increase initially then slightly decrease with the increase of leakage speeds, which displays that vibrational RMS values have nonlinear correlations with the rotation angle of taps and this is described as shown in Figure 3.22.

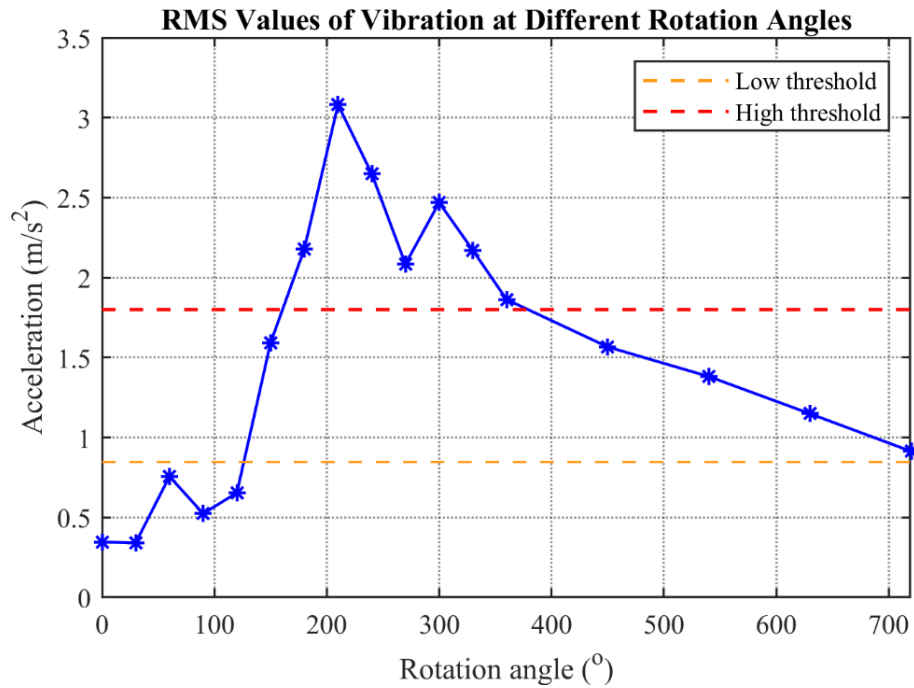


Figure 3.22 Vibration RMS values and thresholds setting at Position 1

The vibrational RMS value increases slightly when the rotation angle of the tap is less than  $120^\circ$ , then significantly rises until the rotation angle of the tap is up to  $210^\circ$ . It indicates that when the spindle of the tap goes up together with the handle, the resonance of the piping system excited by the force of water molecules acting on the pipe wall becomes more and more apparent. However, it will gently decrease when the rotation angle is too large because turbulence is relatively smaller when the gap inside of the tap nearly constant. As shown in Figure 3.22, two threshold lines, the yellow dot line and red dot line with the values of  $0.8 \text{ m/s}^2$  and  $1.8 \text{ m/s}^2$  respectively, are set to distinguish the leakage of the piping system qualitatively. But it is difficult to recognise the small and large leakage, especially.

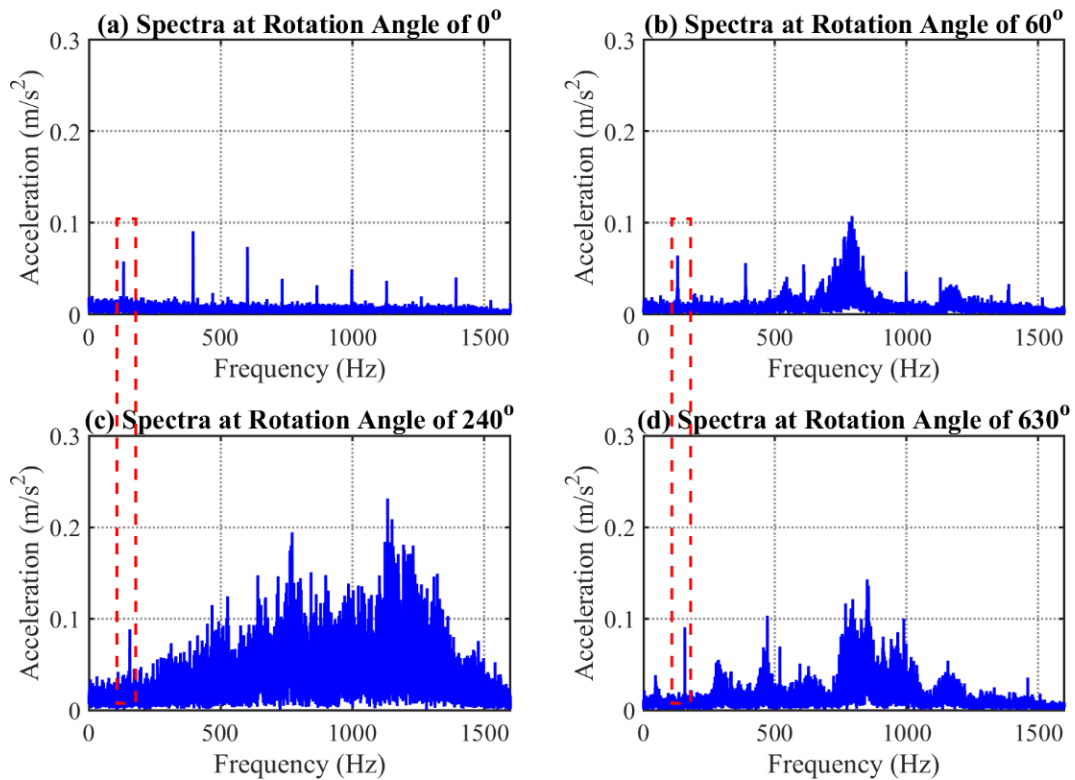


Figure 3.23 Vibration spectra of four different leakage conditions at Position 1

Therefore, the spectra of vibration signals are illustrated in Figure 3.23. Both periodic responses induced by pipe resonances and wideband contents caused by turbulence and friction effects are displayed in spectra. The red dot rectangles indicate a particular frequency band from 130 Hz to 180 Hz for these four cases.

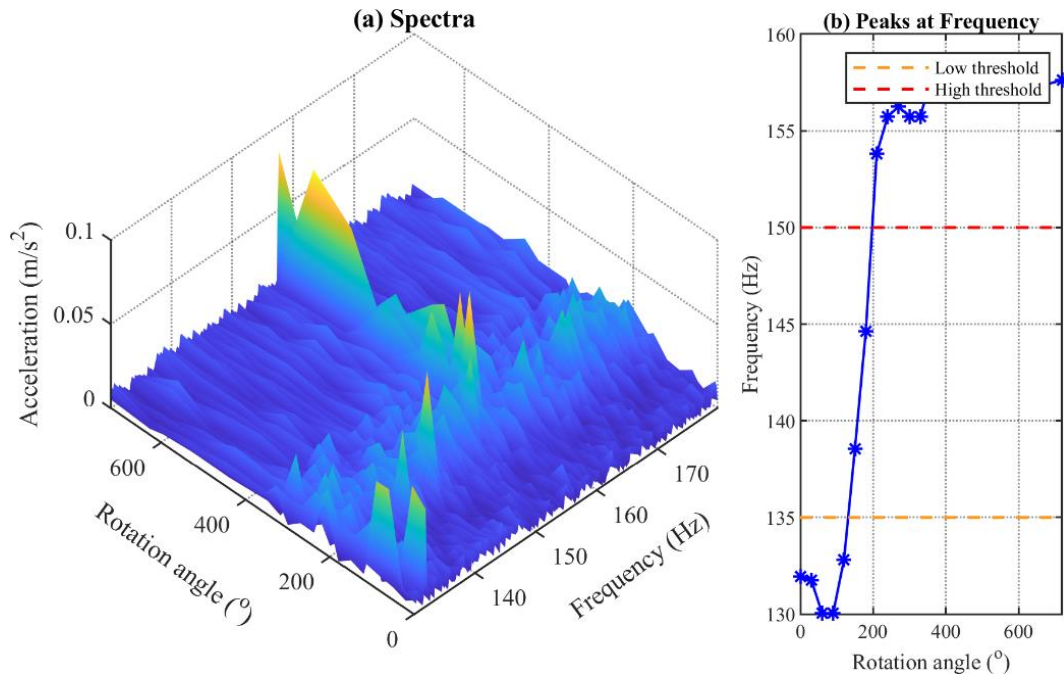


Figure 3.24 Particular frequency analysis at Position 1

Figure 3.24 further analyses the particular frequency band displayed with red dotted frames. Figure 3.24 (a) shows that the highest amplitude shifts with the change of

rotation angles in the frequency band of 130 Hz to 180 Hz. Figure 3.24 (b) displays the frequency with peaks in this particular frequency band for each rotation angle. The peak value can clearly distinguish leakage when the rotation angle is lower than  $135^\circ$  or higher than  $200^\circ$ . However, considering complexity of the structure of the piping system at Position 1, this shifting frequency may be caused by turbulence and will be unstable and unfixed. Furthermore, sensor installation is complicated at this position. As a result, Position 1 is not suitable for leakage detection with the designed CM system.

## (2) Experimental results at Position 2 (MEMS)

Position 2 locates in the middle of the long straight pipes. Tap 2 is a little tight, so the maximum angle rotated can only reach up to  $360^\circ$ . Figure 3.25 randomly illustrates four vibration signals of different leakage conditions with various rotation angle at Position 2. The RMS value significantly varies in figures (a), (b) and figures (c), (d) which represent zero, small, medium and large leakage, respectively. A plotted curve for the various leaks shows the relationship between vibration RMS values and rotation angles in Figure 3.26.

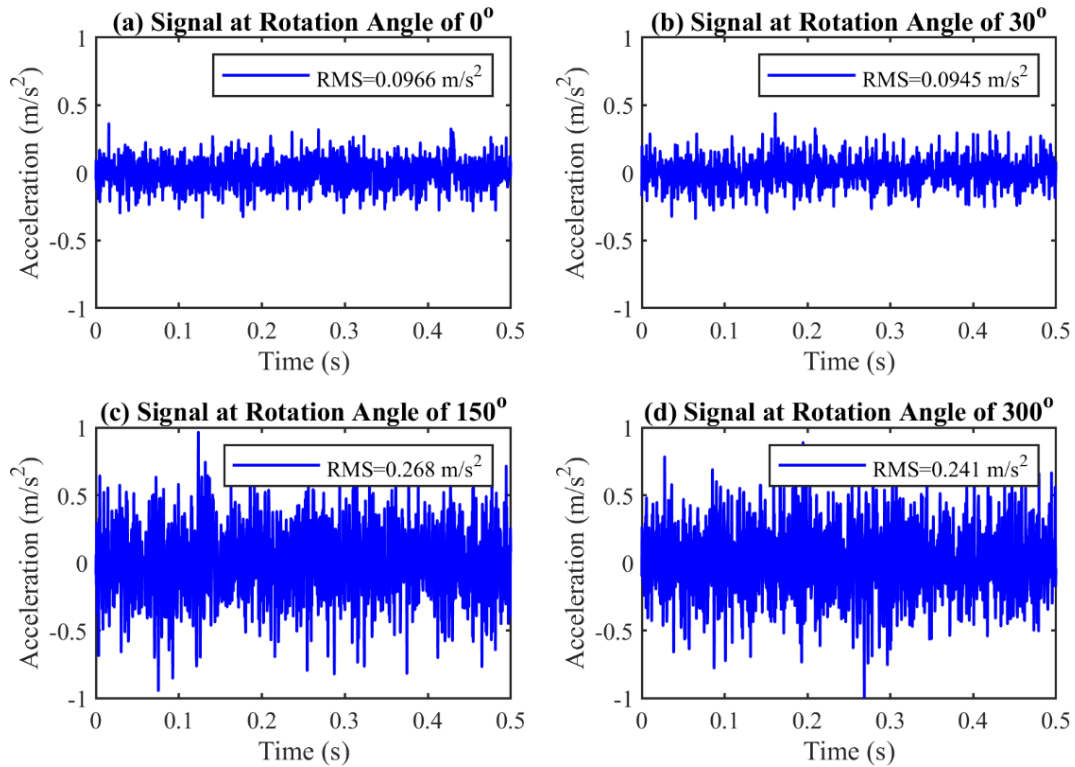


Figure 3.25 Waveform of four different leakage conditions at Position 2

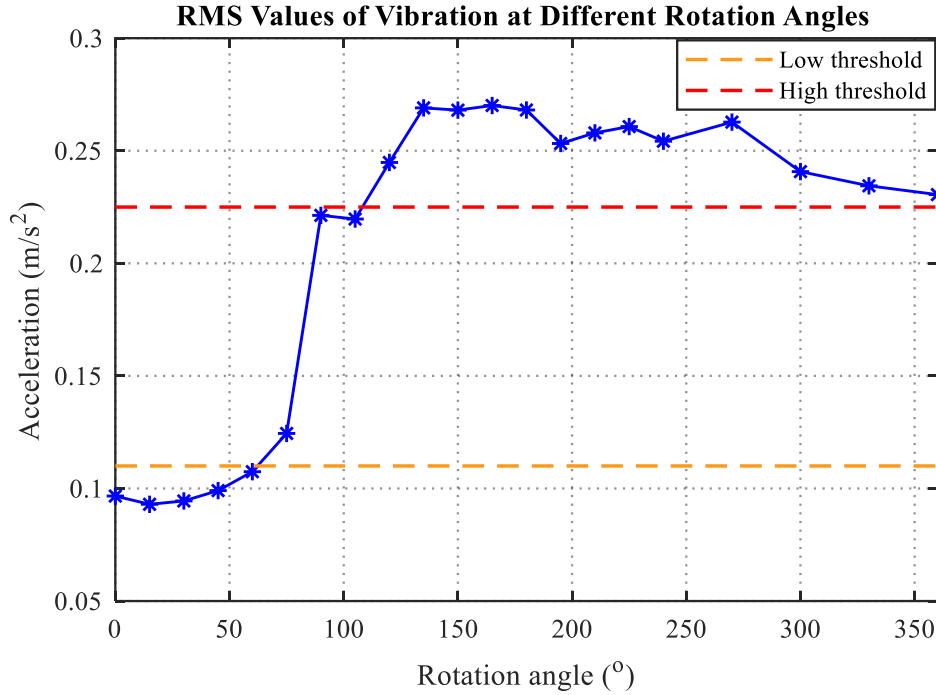


Figure 3.26 Vibration RMS values and threshold setting at Position 2

It demonstrates that vibration RMS value increases approximately with the rising of the rotation angle in Figure 3.26. Two threshold lines with the values of  $0.11 \text{ m/s}^2$  and  $0.225 \text{ m/s}^2$  are set to show different levels of leakage quantity. The vibration amplitude significantly increases as the rise of the rotation angle, especially at the rotation angle greater than  $60^\circ$ . This is because the resonance of the pipe is excited by a force generated by the water flow with the response frequency band highlighted by the red dotted frames between 210 Hz and 280 Hz in Figure 3.27 (c) and (d). However, the resonance amplitude is very small at the small leakage because the energy is too small to visibly activate resonance of the pipe, which is shown in Figure 3.27 (b).

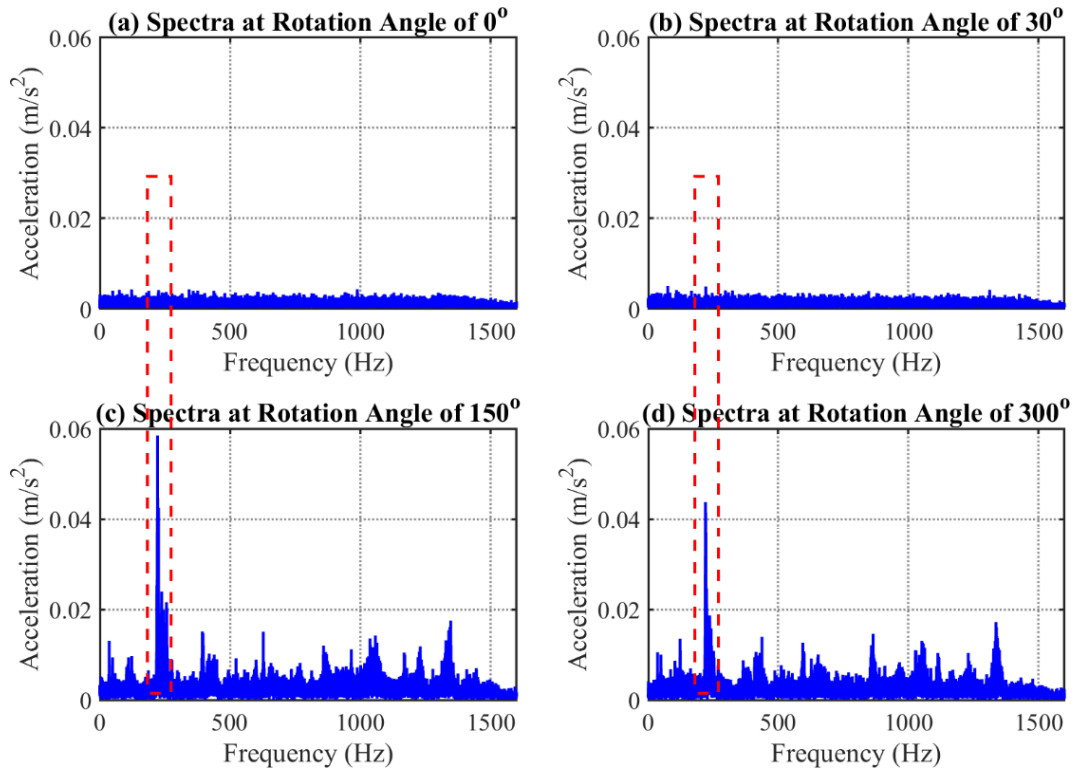


Figure 3.27 Vibration spectra of four different leakage conditions at Position 2

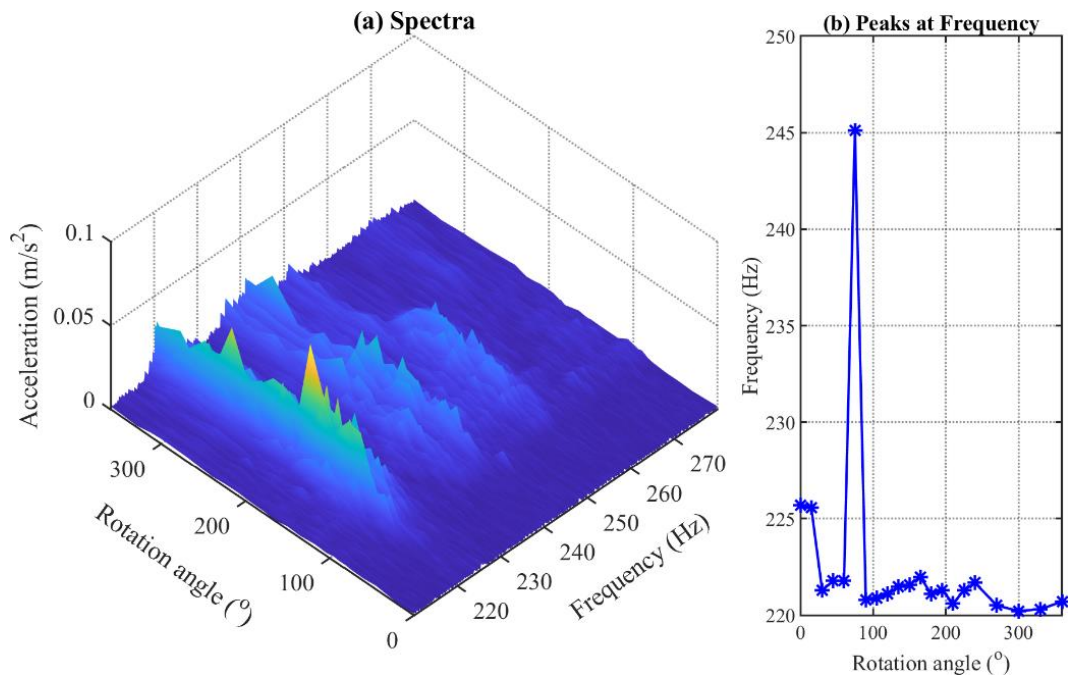


Figure 3.28 Particular frequency analysis at Position 2

Figure 3.28 indicates that peaks of frequency amplitudes appear between 220 Hz and 225 Hz, which can be considered to be the natural frequency of the pipe system. However, there is no reliable relationship between leakage quantity and rotation angle. As a result, the vibration RMS values at Position 2 can be used to monitor condition of the piping system.

### 3.3 Improvement of the CM System with Data Compression

Improvements in the condition monitoring system include designing a wireless sensor node, developing a mobile monitoring application, as well as the evaluation of the designed condition monitoring system. First, the specific statistics, RMS value, was demonstrated that it is useful for the leakage diagnostic of the piping system in the previous subsection. A compression method based on vibrational statistics supplies the potential to achieve remote control and observation of the equipment condition with mobile devices based on wireless transmission in practice. Moreover, the reduction in the amount of data transmitted considerably reduces power consumption, which supplies possibilities for achieving the maintenance-free condition monitoring system according to energy harvesting.

#### 3.3.1 Wireless Sensor Node Design

A WSN node includes four indispensable components: sensor, processor, battery, and communication module. Figure 3.29 highlights the structure of a node and functions of each component. The sensor module can collect raw signals and transmit the analogue-to-digital converted signals to the processor. Pre-processing and feature extraction are carried out in the processor unit to reduce the quantity of data transmitted and stored as much as possible because the power consumed by data processing is much less than that of data transmission [114]. The BLE module is responsible for the data packaging and transmission to the smartphone at an effective distance via Bluetooth. The lithium polymer battery supplies power for the other three units. At the receiving end, a mobile application is developed to monitor the real-time signal of the object. Furthermore, the node can achieve to directly upload data to the cloud database by the smartphone or other portable devices with the assistant of Wi-Fi if it is necessary.

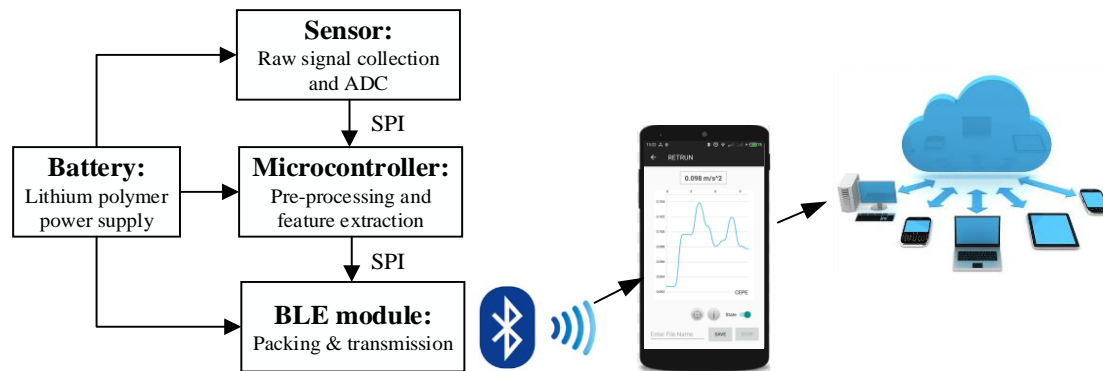


Figure 3.29 Node structure and data acquisition process

The designed wireless sensor node connection is shown in Figure 3.30. The accelerometer, ADXL345, is connected with Adafruit Feather M0 Bluefruit LE module by the SPI connection method. Adafruit Feather M0 Bluefruit LE is an all-in-one module which combines portable microcontroller cores (Cortex M0) with BLE transmission based on USB and battery charging. The advantages of this module



include its thin thickness, small size, light weight, good compatibility with Arduino, low power consumption and low cost. The rechargeable battery can supply power for the Feather M0 Bluefruit LE module. The accelerometer can be powered with continuous and stable energy by the Feather M0 Bluefruit LE module directly.

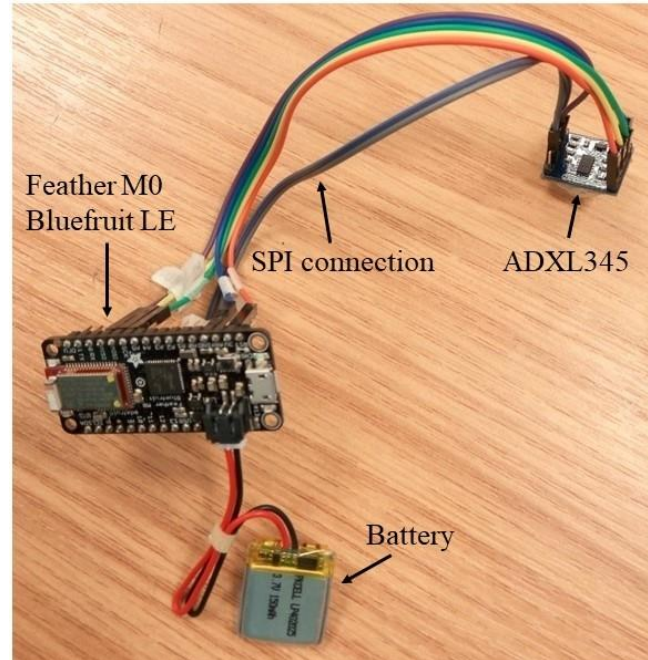


Figure 3.30 Physical connection of the designed wireless sensor node

As described in the previous subsection, the first challenge of the system improvement is the node driven by the programme in Arduino. The sampling frequency and acceleration measure range of ADXL345 are set to 3200 Hz and  $\pm 4g$ , respectively. To avoid the effect of the environmental noise caused by other behaviours, such as door closing and car passing, a high-pass filter with the threshold of 100 Hz is added to filter the signal with the frequency below 100 Hz. Furthermore, the vibration RMS value was demonstrated in the previous analysis that it is an effective and efficient indicator for CM of the piping system.

Another challenge is installation of the accelerometer, which aims to ensure to capture an obvious vibration signal. It is well known that a sensor can effectively measure the vibration signals when contact way and transmission media are valid to minimize the loss in the process of transmission. Because the experiment was carried out with a standard domestic water piping system with copper, so the accelerometer base designed with copper can tightly attach to the surface of the copper pipe with the aid of silicon which increases the contract area and is helpful for signal transmission. Moreover, the thickness of the base is only 2 mm to effectively reduce energy attenuation during signal transmission. The copper base manufactured by the technicians at the University of Huddersfield is shown in Figure 3.31.

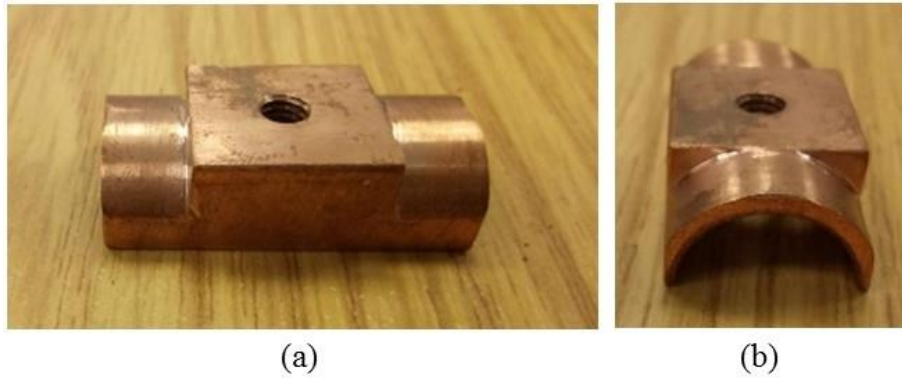


Figure 3.31 Copper base: (a) front view, and (b) side view

The accelerometer, ADXL345, was attached to a copper sheet by ceramic adhesive with functions of fixation and insulation. Then, the sensor together with a copper sheet was screwed onto the copper base through the screw rods. The integrated node was fixed on the surface of the copper pipe with two clips and covered with a white junction box as shown in Figure 3.32. The switch button can control the on and off status of the node.

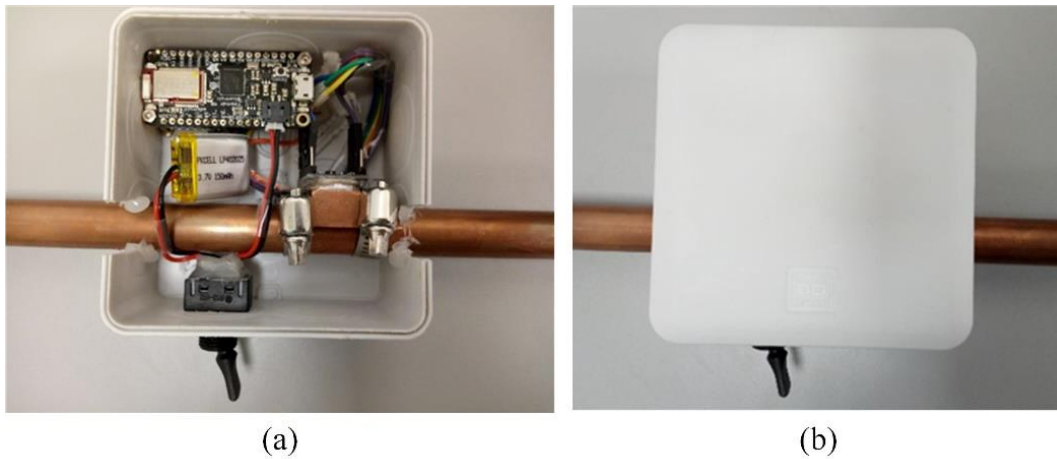


Figure 3.32 Leakage detector node: (a) internal structure, and (b) appearance

### 3.3.2 Mobile Monitoring Application Design

In order to monitor condition of the object with the designed wireless sensor node in real time, a monitoring application is designed for the Android operation system with the Android Studio software as shown in Figure 3.33.



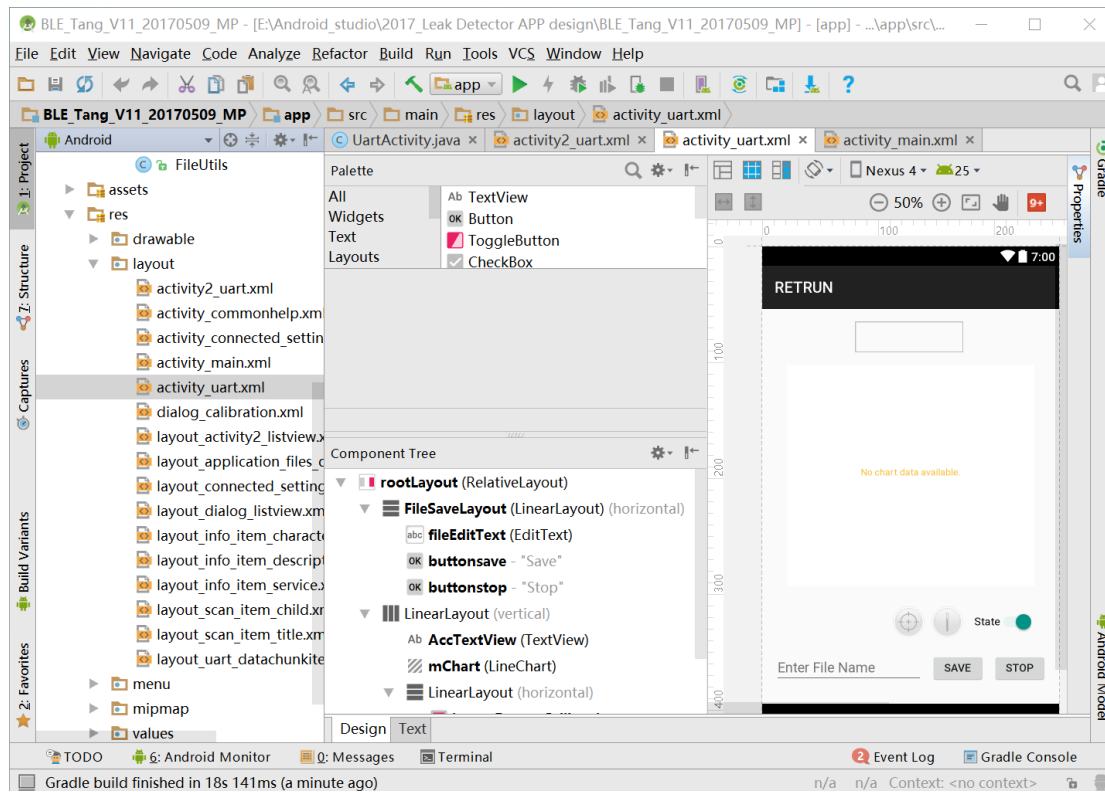


Figure 3.33 Android Studio software interface

An Android application named “Leak Detector” was developed and its installation process and icon are shown in Figure 3.34 (a) and (b), respectively.

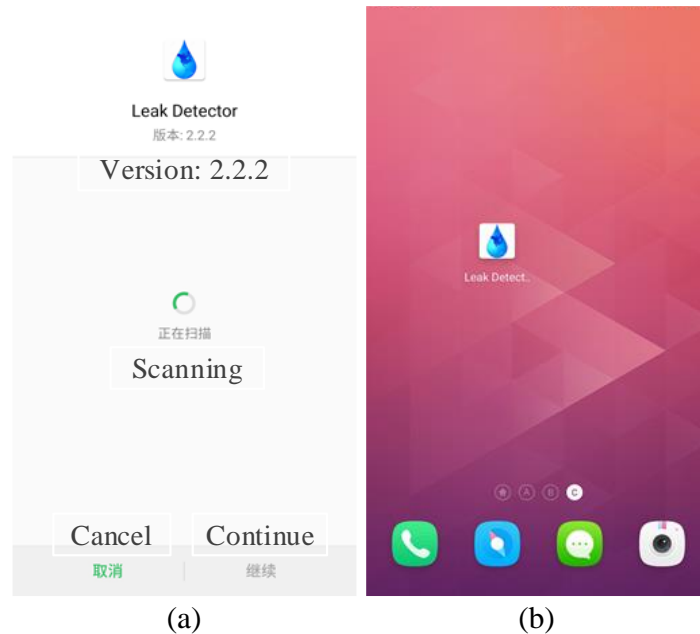





Figure 3.34 Mobile application interfaces: (a) installation interface, and (b) icon interface

As illustrated in Figure 3.35 (a), Adafruit Feather M0 Bluefruit LE can be discovered by the Bluetooth function of the mobile once the node is powered. Then, click the “CONNECT” button and the green LED on the Feather M0 Bluefruit LE board will

blink, indicating that wireless connection is being requested. After connection is successful, the points can be drawn with hollow circles connected by the blue lines to display the real-time values in the line chart as shown in Figure 3.35 (b). Additionally, the display frame above the line chart will clearly show the real-time RMS value of vibration signals. At the bottom of this interface, there is a function of saving datasets for further analysis. The switch button “State” aims to control whether the line chart is hidden or not. Button  is the key to the interface conversion from interface (b) to interface (c). Figure 3.35 (c) is a calibration panel which helps set thresholds of large, medium, and small flow rates. Button  is the path to switch to the process bar display interface which distinguishes the level of current water flow according to the thresholds. Similar to Figure 3.35 (b), Figure 3.35 (c) also has a function of vibration value display on the top of the process bar chart. Button  is applied to switch back to the line chart. Finally, there is a “RETURN” button on the top of the screen, which has the function of exit to the first connection interface.

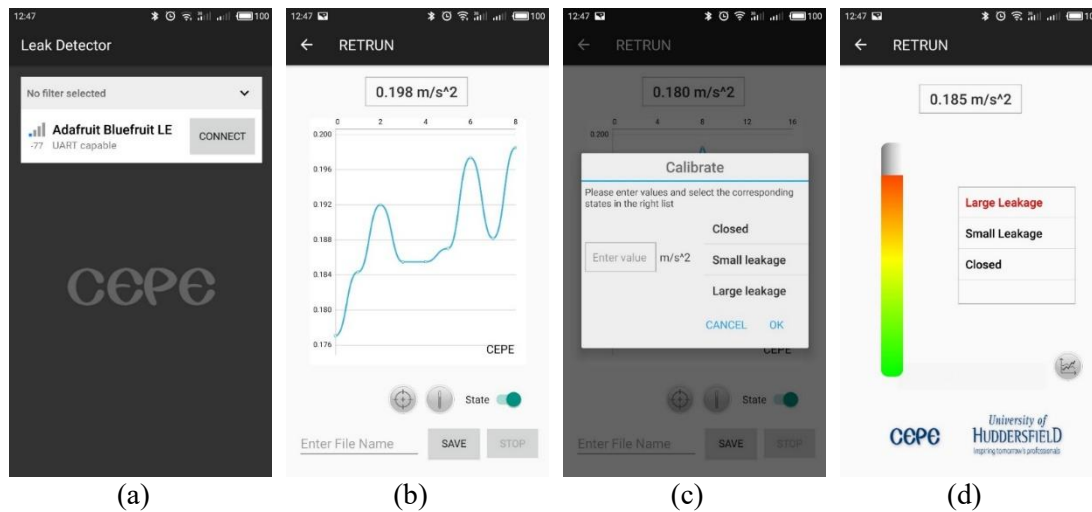


Figure 3.35 Mobile application interfaces of the Leak Detector: (a) connection interface, (b) line chart display interface, (c) calibration interface, and (d) process bar display interface

Other sensors with either analogue or digital output also can be set as the sensing unit to connect to this microcontroller unit. Besides, a high-pass filter and RMS calculating are programmed in the processing unit to suppress the effect of low-frequency signals and reduce the amount of data transmission. Data can be received via mobile Bluetooth and observed on the mobile application in real time.

### 3.3.3 Evaluation of the Designed Remote CM System

The designed CM system can be validated by carrying out some experiments about remote leakage detection with the domestic piping systems in real time. The installation location of the wireless sensor node was the middle of the long straight pipe described in Figure 3.20 because it is easily prone to the pipe resonance.

The acceleration caused by the pipe wall vibration in three directions (X, Y, Z) were acquired with the sample rate of 3200 Hz separately. Then the RMS values of these three directions were calculated at various flow rates, which is shown in Figure 3.36. It can be seen that the RMS value in the Y direction is the best indicator to accurately distinguish the flow rates, following by it in the X direction, and the Z direction is the last. Two thresholds of 1 L/min and 2 L/min were set manually to distinguish small, medium and large leakage, respectively.

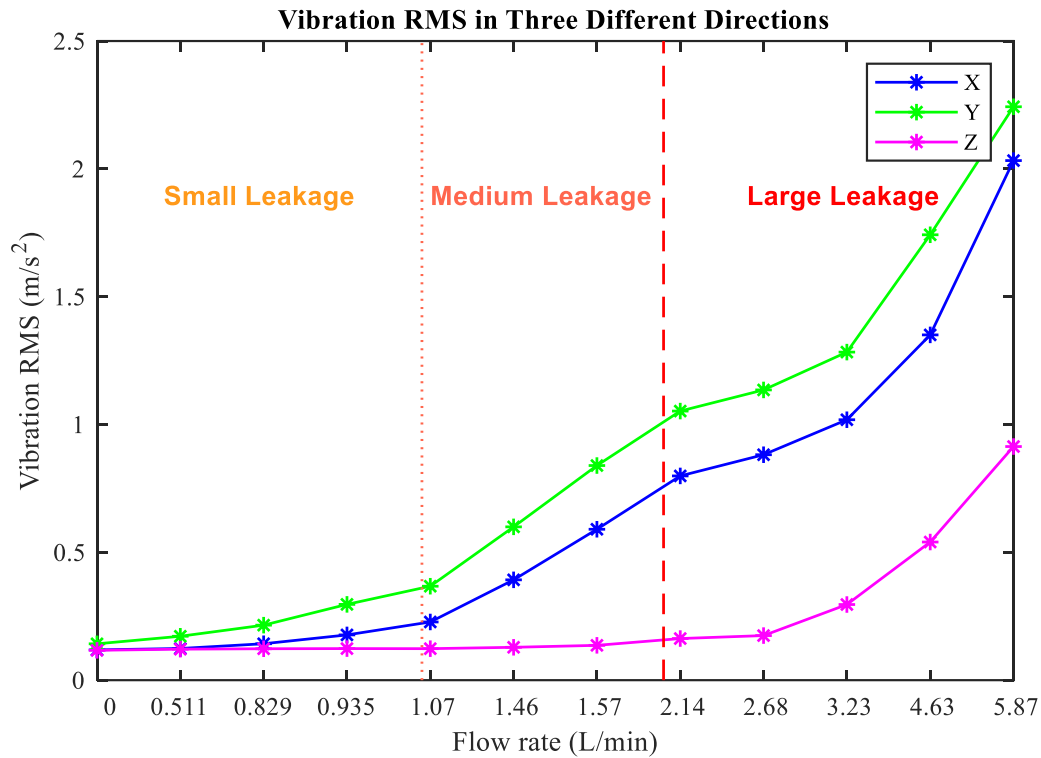


Figure 3.36 Acceleration RMS in three different directions

However, it may misjudge the condition of the piping system if only RMS values in a single direction are employed when the installation direction of the accelerometer is improper. So that a total vibration RMS value for these three directions is calculated as observed in Figure 3.37, which is much more stable and robust to detect leakage of the piping system.

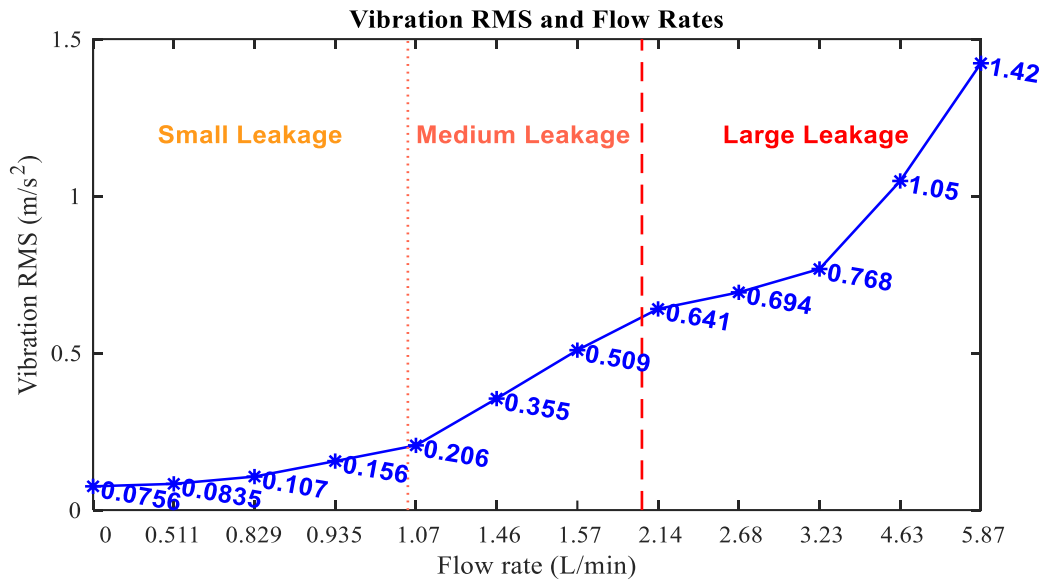


Figure 3.37 Relation between vibration RMS values and flow rates

A monitor video was taken to express how to use this designed CM system to effectively and efficiently detect the water leakage in real time when the node locates at about 6 to 7 meters far from the leak position and the monitoring location.

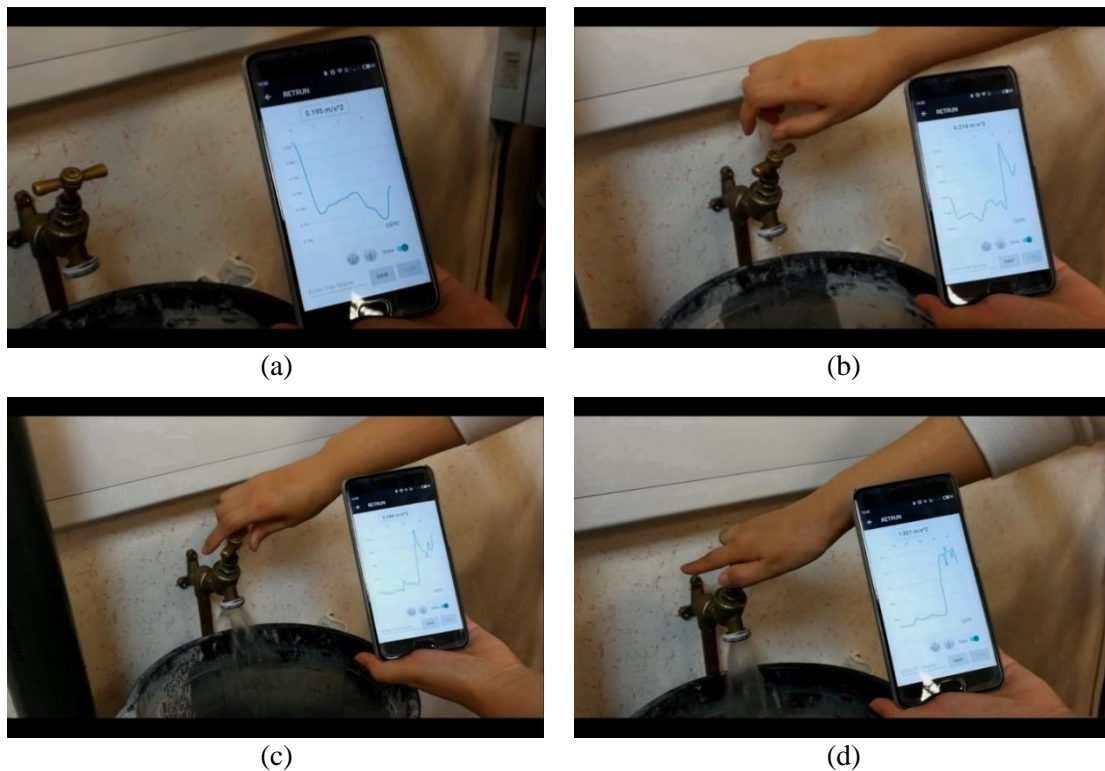


Figure 3.38 Leakage real-time monitoring displayed with the line graph: (a) closed, (b) small leakage, (c) medium leakage, and (d) large leakage

The line graphs and bar process charts are shown in Figure 3.38 and Figure 3.39 with closed, small leakage, medium leakage and large leakage states represented from (a) to (d), respectively. The acceleration amplitudes are about 0.195 m/s<sup>2</sup>, 0.214 m/s<sup>2</sup>, 0.346 m/s<sup>2</sup>, and 1.521 m/s<sup>2</sup> for these four different conditions in the line charts, which can

clearly distinguish whether there is a leak or not, and even can effectively distinguish the amount of the leak.

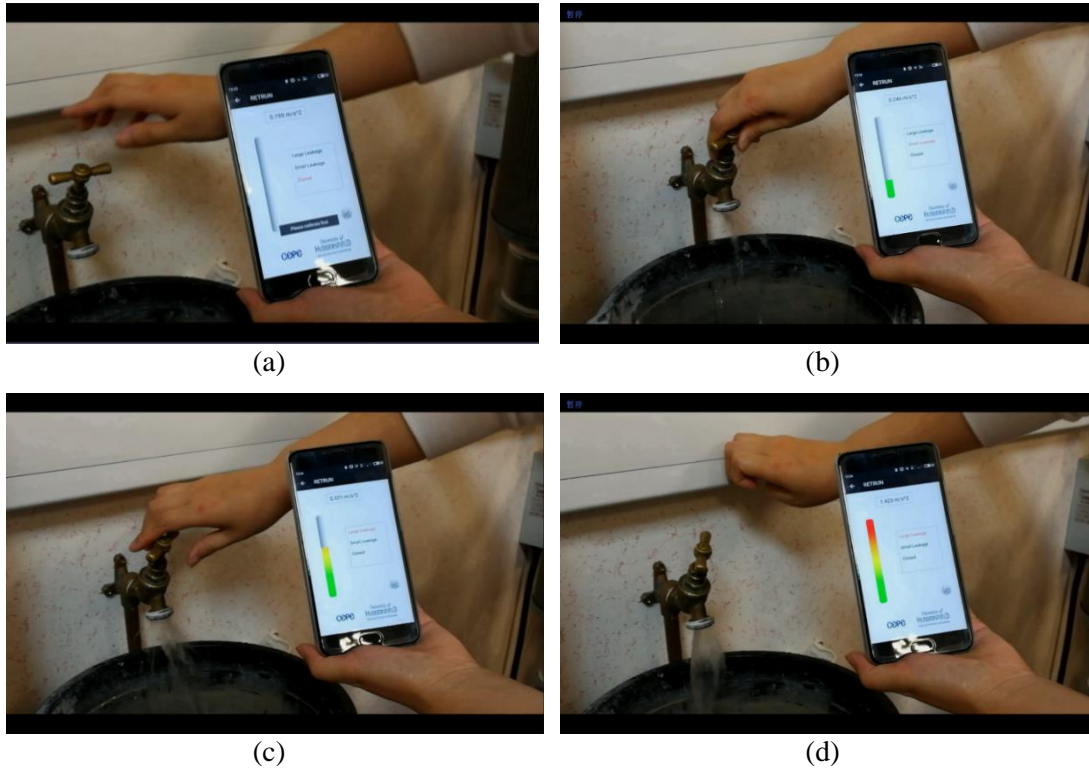


Figure 3.39 Leakage real-time monitoring displayed with the bar process chart: (a) closed, (b) small leakage, (c) medium leakage, and (d) large leakage

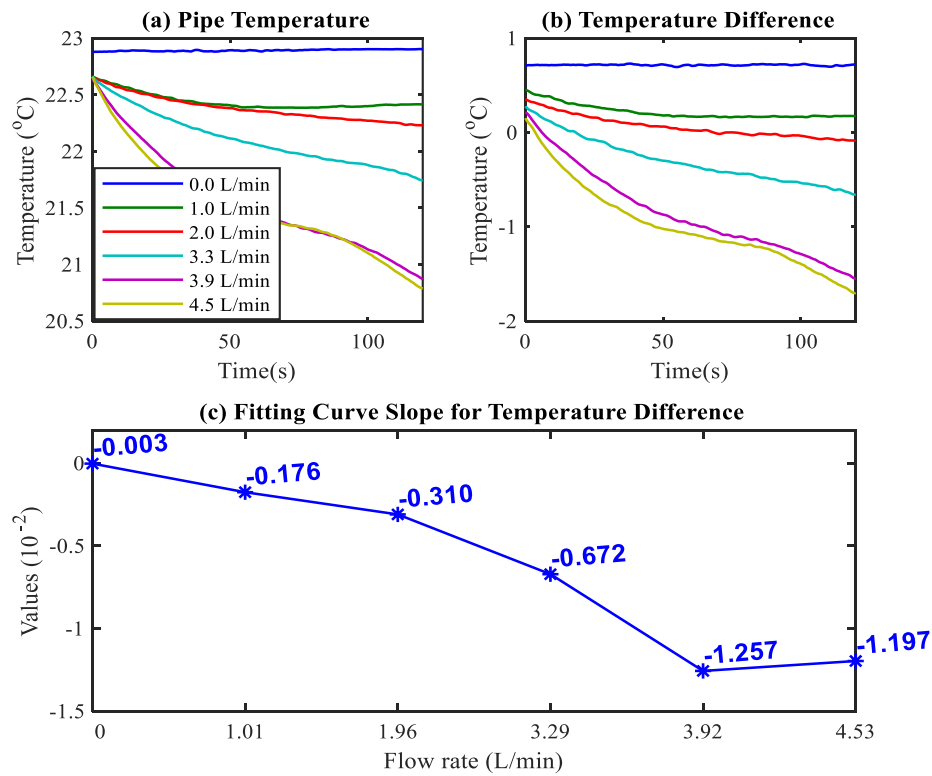


Figure 3.40 Pipe temperature change with the flow rate

From the temperature, water flow coming from outside will change the temperature of the domestic pipe wall mainly because of thermal conduction and convection when a leak occurs since the temperature outside is significantly different from the room temperature. Figure 3.40 illustrates the pipe temperature and temperature difference from ambient temperature captured by the MEMS infrared thermometer called MLX90614 GY-906 connected with Feather M0 Bluefruit LE as shown in Figure 3.41. The infrared thermometer works based on thermal radiation of the pipes. It is apparent that speed of temperature change is closely related to the flow rate. The faster the water flows, that is to say, the bigger the leakage, the faster the pipe temperature changes. This phenomenon also is helpful for the effective detection of leaks.

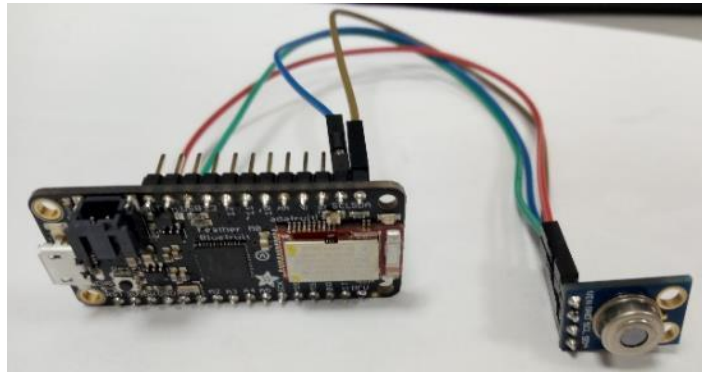


Figure 3.41 Infrared thermometer connected with Feather M0 Bluefruit LE

In the future work, some new functions need to be investigated in this wireless sensor node to achieve more accurate, convenient and broad application on CM in various fields. For example,

- ✓ Combine other features to assist vibrations to achieve efficient CM, such as detecting the ambient and object temperatures by infrared thermometers with the contactless method of installation.
- ✓ Optimize the node design to reduce power consumption and achieve to upload the object condition to the cloud to support remote CM.
- ✓ Develop the node to be a maintenance-free CM system according to energy harvesting, which will be applied extensively in CM, especially for mechanical fault detection.
- ✓ It can be a potential application of on rotor sensing in CM of rotating machinery to reduce or even avoid the effect of the transfer path on the fault signal in future. Figure 3.42 illustrates our preliminary demo of on rotor sensing application, including the mobile application, hardware and its installation, as well as the monitoring interface. The designed cover box is detailed in Figure 3.43.



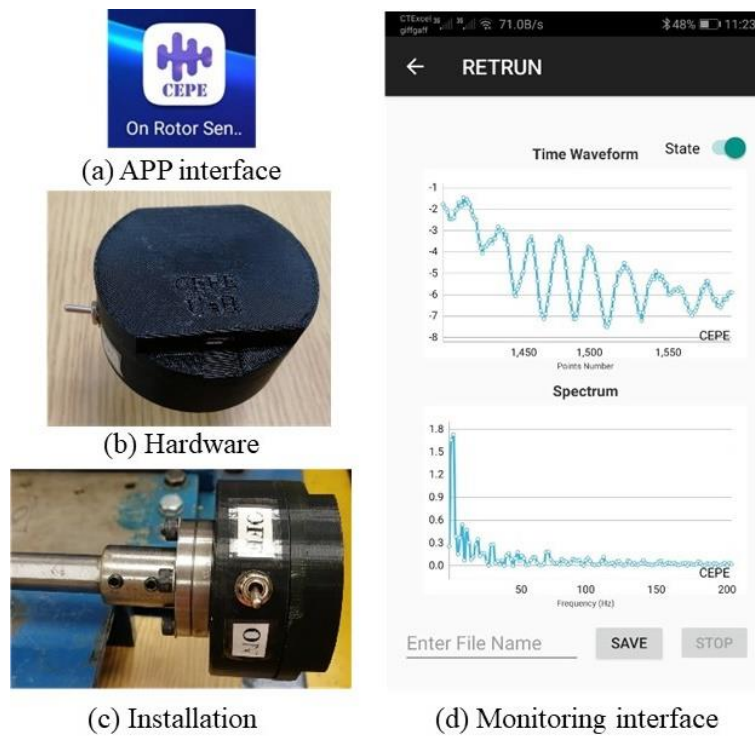


Figure 3.42 On rotor sensing application for rotating machines: (a) APP icon, (b) hardware design, (c) installation, and (d) monitoring interface

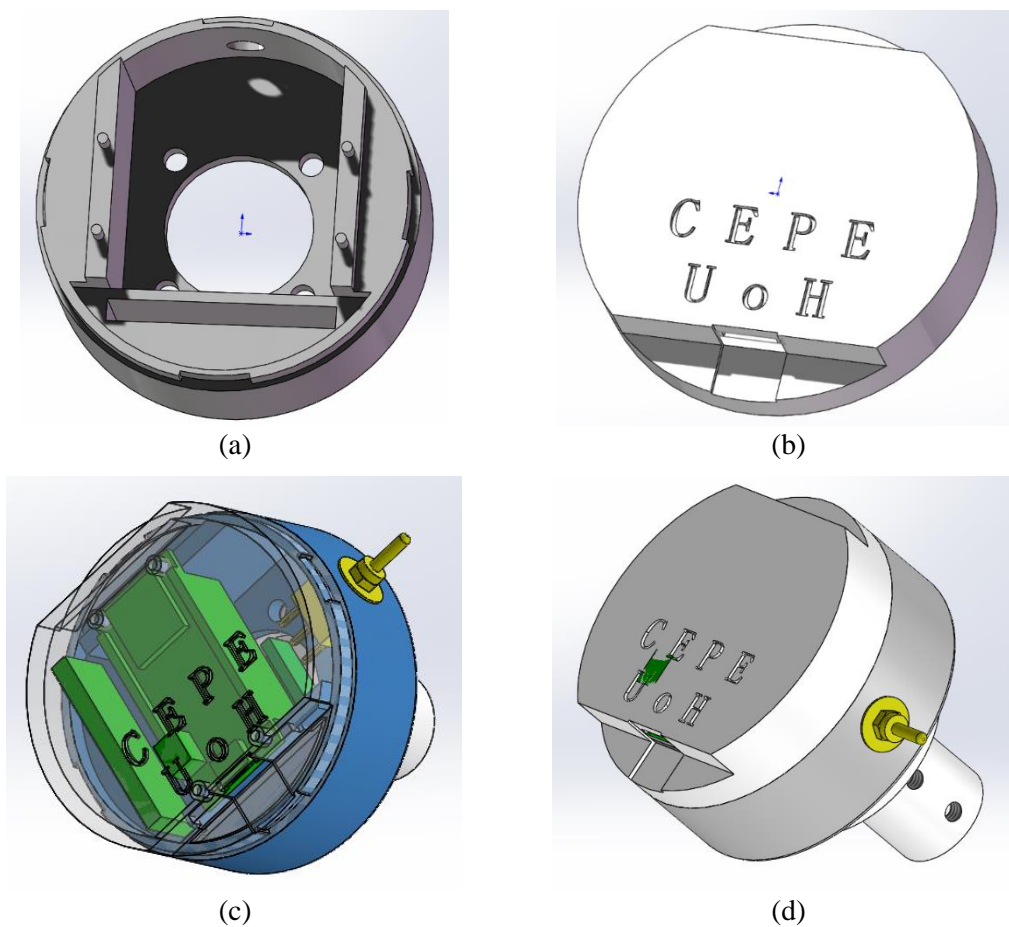


Figure 3.43 Design of package for on rotor sensing application: (a) box body, (b) box cover, (c) internal structure, and (d) assemble

### 3.4 Summary

In this chapter, an intelligent vibration sensing system, consisting of both the sensor node and the signal monitored mobile APP, was successfully designed based on the basic data compression method of the specific statistics (RMS values of vibration at three directions). The designed CM system can effectively and efficiently detect and assess leakage of the domestic piping system in real time according to the vibration mechanisms of the piping system. The intelligent wireless sensor node locates in the middle of the long straight pipe, where is easily prone to the pipe resonance induced by the water flow. The developed leak detector successfully detects leakage in piping systems using the compressed data. The detector consumes limited power because the amount of data transmitted is greatly reduced with an effective compression method.

Besides, the developed leak detector not only successfully detects leakage but also saves power because the amount of data transmitted is greatly reduced with an effective compression method. Finally, even the designed CM system works very well, it can be further improved in power supply and remote monitoring, as well as applied as on rotor sensing for fault detection of rotating machinery in future.



---

## **CHAPTER 4    DIMENSION REDUCTION BASED COMPRESSION IN FAULT DIAGNOSIS OF TAPERED ROLLER BEARINGS**

---

*Rolling element bearings are the essential parts of rotating machines, of which the faults can cause severe failures or even breakdown of a machine. The use of multiple sensors and the high-speed data acquisition device lead to a manifest disadvantage of high costs in purchasing hardware for data collection, transmission, storage and processing. As a result, this chapter and the next chapter investigate the compression algorithms for fault diagnosis of rolling element bearings based on 1D vibration signals. In this chapter, the recurrence quantification analysis upon the recurrence plot was developed to extract features and reduce data dimensions. The dimension reduction based compression method is efficient and effective to diagnose faults of tapered roller bearings.*

## 4.1 Introduction

### 4.1.1 Rolling Element Bearings

Rolling element bearings, whose faults can cause serious failures or even breakdown of a machine, are the essential components of rotating machines to carry a load by placing rolling elements between inner and outer races. Rolling element bearings, including ball bearings, cylindrical roller bearings, spherical roller bearings, tapered roller bearings, needle roller bearings and CARB toroidal roller bearings, etc. are common types of bearings applied in the industry with the structures as shown in Figure 4.1.



Figure 4.1 Typical types of bearings applied in industry

Early fault detection is a technology to monitor condition of rotating machinery during operation to prevent machine failures. System overload, shaft imbalance, insufficient lubrication, fatigue wear, and overheating are common factors to induce failures of the inner race, the outer race, roller elements and the cage of a bearing.

The acquired bearing vibration signals usually exhibit typical modulation characteristics for fault diagnosis of rolling element bearings. Especially, the high-frequency resonance technique treats the measured vibration responses as a result of the modulation between the periodic impulses and structural resonances. The bearing fault signal can be obtained by demodulation from a high-frequency band of the collected system vibration signal. The theoretical fault frequency of rolling element bearings can be calculated according to the rotation speed, number of balls, diameters of balls and the pitch circle, and the contact angle [116] for the judgement of fault types.

The outer race fault frequency, fundamental cage frequency, roller fault frequency and inner race fault frequency are calculated as shown from Equations (4-1) to (4-4) in

theory, respectively. The roller diameter and pitch circle diameter in these equations are displayed in Figure 4.2.

Outer race fault frequency  $f_o$ :

$$f_o = \frac{N_r}{2} f_s \left( 1 - \frac{D_r}{D_p} \cos \varphi \right) \quad (4-1)$$

Fundamental cage frequency  $f_{cage}$ :

$$f_{cage} = \frac{1}{2} f_s \left( 1 - \frac{D_r}{D_p} \cos \varphi \right) \quad (4-2)$$

Roller fault frequency  $f_r$ :

$$f_r = \frac{D_p}{2D_r} f_s \left( 1 - \frac{D_r^2}{D_p^2} \cos^2 \varphi \right) \quad (4-3)$$

Inner race fault frequency  $f_i$ :

$$f_i = \frac{N_r}{2} f_s \left( 1 + \frac{D_r}{D_p} \cos \varphi \right) \quad (4-4)$$

where,

$N_r$  is the number of rollers, dimensionless;

$f_s$  is the shaft rotating frequency, Hz;

$D_r$  is the roller diameter, mm;

$D_p$  is the pitch circle diameter, mm;

$\varphi$  is the contact angle, degree.

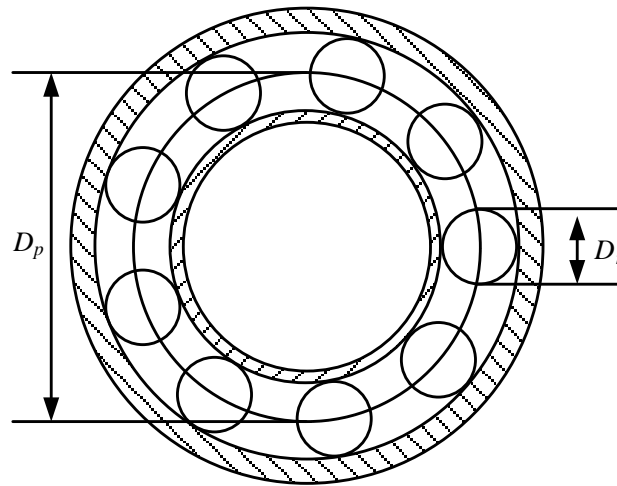


Figure 4.2 Schematic diagram of a rolling element bearing

Using these equations, the theoretical fault frequency of most rolling element bearings can be calculated as a reference to diagnose bearing fault types. The tapered roller bearings with different types of faults are investigated in this chapter because of their wide heavy-duty applications.

#### 4.1.2 Tapered Roller Bearings

Tapered roller bearings were designed and applied since the early 20th century. They have a tapered inner race and an outer race with tapered rollers arranged between two ring raceways as displayed in Figure 4.3. It mainly bears combined (simultaneously acting radial and axial) loads operating at a high speed. The bearing capacity of the bearing depends on the contact angle of the outer race. The larger the angle, the greater the load carrying capacity. It can be divided into single row, double row and four-row tapered roller bearings according to the number of rows of rolling elements in the bearing. Tapered roller bearings are a type of rolling element bearings that is widely used in pairs for better radial load handling in various heavy-duty applications, such as vehicle wheels, differentials, pinion shafts, machine tool spindles, construction machinery, agricultural machinery, railway vehicle gear reduction devices, rolling mills, etc.

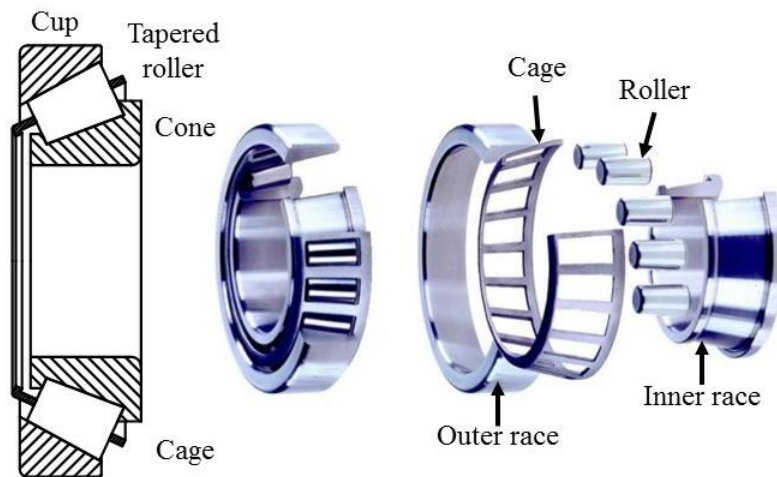


Figure 4.3 Tapered roller bearing structure and components

However, tapered roller bearings are prone to have a shortened lifespan or even be damaged directly if they undergo enormous amounts of stress under conditions of inadequate or starved lubricant, incorrect installation, or misalignment during operation. Figure 4.4 lists some common faults of tapered roller bearings, like bruising, etching, line spalling, cage deformation, roller binding and skewing, inadequate lubrication, peeling, excessive preload or overload and misalignment [117]. They are mainly caused by the debris contamination coming from the contaminated lubricant, a leaky seal or other fatigued parts, water or moisture contamination, the improper tool used for installation, starved lubricant, tremendously high stress and human error.



Figure 4.4 Damaged tapered roller bearings with different faults

Clearance between the roller and outer raceway needs to be adjusted when the single row tapered roller bearing is installed. Bearings that are improperly handled during installation or that have large clearance error are highly susceptible to damage when subjected to heavy loads. In order to ensure the bearing can work stably, it is a crucial task to monitor their working state.

However, the fault signal is often overwhelmed by noise in the low-frequency band, and it is necessary to demodulate the fault signal from a high-frequency band of the vibration signal which is captured at a sampling frequency of up to tens of thousands hertz. Both use of multiple sensors and high-speed data acquisition show considerable shortages of high system cost involved in purchasing hardware for data transmission, storage and processing. Therefore, a dimension reduction based feature compression method is developed for fault classification of tapered roller bearings and validated by some experiments.

## 4.2 Compression Method

In a mechanical system, possible states can be represented in the phase space with a corresponding phase space point. Each point of the phase space represents the position and momentum of each particle in the system [118]. Tapered roller bearings are usually used for the rotating machinery to withstand loads. There will be a noticeable periodical fault signal when a fault occurs. Hence, the signal can be converted into the high-

dimensional phase space to analyse. Although valuable information in the high-dimensional space is rich, it is not conducive to observation and analysis. A recurrence plot can be applied to observe the periodic information of a phase space trajectory in two dimensions. Recurrence quantitative measurements can evaluate the periodic information of this complicated system. However, the dimension of the extracted features is also too high to analyse further. Besides, the collected signal will be divided into segments after preprocessing due to the length of data collection and the limitation of computing ability. As a result, principal component analysis is used for dimension reduction by decreasing the number of principal components to classify the types of faults.

#### 4.2.1 Recurrence Plot

Recurrence, proposed by Henri Poincaré in 1890 [119], is a fundamental characteristic of a dynamical system, which is useful for exploring all relevant features of behaviours of the system [120]. Recurrence plot (RP) is an advanced graphical tool to visualize the binary texture of time series of a dynamical system with lines and single dots, which was investigated by Eckmann *et al.* in 1987 [121]. It is easily constructed to implement the pattern analysis of a complex system with the aid of some recurrence quantification [122]. In theory, RP depicts pairs of times at which the phase space trajectory of the dynamical system visits roughly the same place in the phase space [120]. In other words, RP allows investigating the  $m$ -dimensional phase space trajectory through its 2D representation by dimension reduction.

Let  $u(t), t = 1, 2, \dots, N$ , be a time series with the length of  $N$  and its generated phase space trajectory can be expressed as [123]

$$\begin{cases} x(1) = \{u(1), u(1 + \tau), \dots, u(1 + (m - 1)\tau)\} \\ \dots \\ x(i) = \{u(i), u(i + \tau), \dots, u(i + (m - 1)\tau)\} \\ \dots \\ x(N - (m - 1)\tau) = \{u(N - (m - 1)\tau), u(N - (m - 2)\tau), \dots, u(N)\} \end{cases} \quad (4-5)$$

where  $x(i), i = 1, 2, \dots, N - (m - 1)\tau$ , is the points of the trajectory generated in the phase space;  $\tau$  and  $m$  are the time delay and embedding dimension of the reconstructed phase space, respectively.

Therefore, the 2D recurrence matrix  $R(i, j)$  is expressed in Equation (4-6).

$$R(i, j) = \begin{cases} 1 & \text{if } \|x(i) - x(j)\| \leq \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad (4-6)$$

where  $x(i)$  and  $x(j)$  are the points in the phase space trajectory at which the system is situated at times  $i$  and  $j$ , respectively [124];  $\varepsilon$  is a recurrence threshold predefined according to system characteristics. Equation (4-6) is also simply represented as

$$R(i, j) = \Theta(\varepsilon - \|x(i) - x(j)\|) \quad (4-7)$$

where  $\Theta(\cdot)$  is the Heaviside step function. If the Euclidean distance of these two system states at times  $i$  and  $j$  in the phase space is very close to  $\varepsilon$ ,  $R(i, j) = 1$  and the coordinates  $(i, j)$  will be a black dot in the grayscale recurrence graph; otherwise, it will be displayed as a white dot instead.

However, there are three principal challenges in this process: time delay parameter ( $\tau$ ) estimation, optimal embedding dimension ( $m$ ) determination and predefined threshold ( $\varepsilon$ ) selection. In [125], Chen and Yang mentioned that the time delay parameter and embedding dimension can be estimated with mutual information method and false nearest neighbour algorithm, respectively. Marwan *et al.* [120] reviewed various methods of threshold selection which provides an opportunity for the optimal threshold selection in this research.

#### 4.2.1.1 Time Delay Parameter Estimation

There are three main methods for determining the reconstruction delay time: autocorrelation function, average displacement and mutual information methods. Autocorrelation function method is not suitable for nonlinear analysis and average displacement method is mainly based on experience with randomness. Although the mutual information method requires complex calculations, it is suitable for both linear and nonlinear systems from the perspective of information theory. In 1985, both Shaw [126] and Fraser *et al.* [127] proposed that mutual information is a good criterion for the selection of time delay in construction phase space from time-series data.

Suppose that  $Q$  and  $S$  are two systems, then the mutual information is defined as

$$I(Q, S) = H(Q) + H(S) - H(Q, S) \quad (4-8)$$

where  $H(Q)$  and  $H(S)$  are the entropy of the systems  $Q$  and  $S$ , respectively;  $H(Q, S)$  is the joint entropy of the system  $Q$  and  $S$ .

Let  $S$  be the time series signal  $\{u(t)\}, t = 1, 2, \dots, N$ ,  $Q$  be a time series signal  $\{u(t + \tau)\}$  of  $S$  with time delay  $\tau$ . Then,  $Q$  and  $S$  can generate a 2D reconstruction image with the mutual information expressed as

$$I(Q, S) = \sum_i \sum_j P_{sq}(s_i, q_j) \log_2 \left[ \frac{P_{sq}(s_i, q_j)}{P_s(s_i)P_q(q_j)} \right] \quad (4-9)$$

where  $P_{sq}(s_i, q_j)$  is the joint distribution probability in the reconstructed image when  $S = s_i$  and  $Q = q_j$ ;  $P_s(s_i)$  and  $P_q(q_j)$  are edge distribution probabilities.

Fraser *et al.* [127] proposed a recursive approach to obtain mutual information  $I(Q, S)$ . It is based on the probability of the edge distribution to divide the partitions and judge for each partition. Firstly, the probability of each side is equally divided according to the edge distribution. And the reconstructed image is divided into four partitions displayed with red lines in Figure 4.5. For each partition, it needs to determine whether there is a substructure or it is already sparse. If there is no substructure or it is sparse, it does not require further division. Otherwise, the partition should be further divided into

two equal parts according to the edge probability until the substructure can no longer be divided. Finally, mutual information can be calculated based on the divided partitions. They also demonstrated that the first minimum in the mutual information of these two systems provides the best available systematic criterion for selecting time delays for phase space trajectory [127].

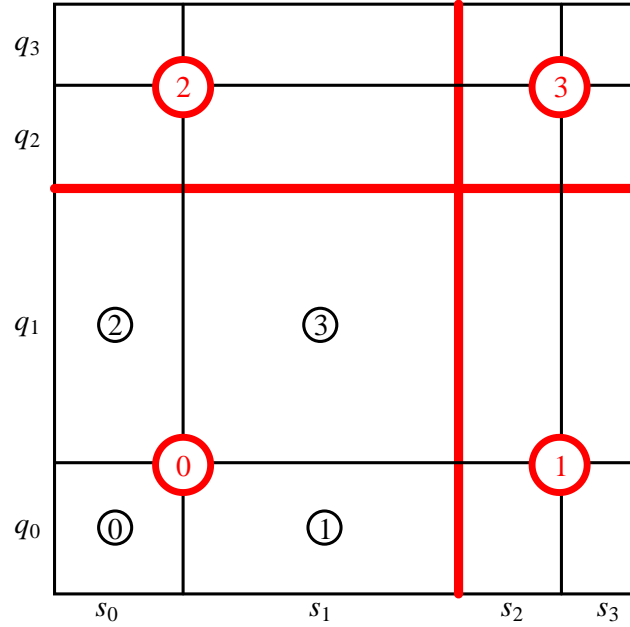


Figure 4.5 Two steps in the sequence of partitions

This method has been successfully used for various systems to determine the time delay parameter. Therefore, it is selected to estimate the time delay for the nonlinear vibration signals of the tapered roller bearing system.

#### 4.2.1.2 Embedding Dimension Determination

Embedding dimension is a crucial parameter in the process of constructing the phase space trajectory. Many neighbours will be false when an embedding dimension is too low. Otherwise, the neighbours are all true if an embedding dimension is appropriate or higher. However, a high embedding dimension leads to the generated trajectory redundant and excessive computation. Therefore, an acceptable minimum embedding dimension should be determined by looking at the behaviour of near neighbours with the false nearest neighbour (FNN) algorithm, which was proposed by Kennel *et al.* [128].

FNN is a method to expand the time series into a multidimensional phase space, which simplifies modelling and prediction. It iteratively increases the dimensions of the phase space and measures changes of the neighbour relationship at each step. Two points that are not adjacent in the high-dimensional phase space may become adjacent when they are projected into the same coordinate, which are called false neighbours. As the embedding dimension increases, false neighbours are gradually eliminated.



In the  $m$ -dimensional phase space, each phase point vector  $x(i) = \{u(i), u(i + \tau), \dots, u(i + (m - 1)\tau)\}$  has the  $r$ th nearest neighbour point  $x^{NN}(i)$  with the square of the Euclidean distance defined as

$$D_m(i, r) = \|x(i) - x^{NN}(i)\| \quad (4-10)$$

where  $\|\cdot\|$  is the Euclidean norm. When the dimension of the phase space increases from  $m$  to  $m + 1$ , the square of the distance between two points changes to

$$D_{m+1}^2(i, r) = D_m^2(i, r) + [x(i + m\tau) - x^{NN}(i + m\tau)]^2 \quad (4-11)$$

If  $D_{m+1}^2(i, r)$  is much larger than  $D_m^2(i, r)$ , it may be caused by high-dimensional chaos attracting two non-adjacent points to become adjacent points when projected onto low-dimensional orbits. Therefore, these points are false neighbours. The criterion for the false neighbour elimination is stated as

$$\sqrt{\frac{D_{m+1}^2(i, r) - D_m^2(i, r)}{D_m^2(i, r)}} = \frac{|x(i + m\tau) - x^{NN}(i + m\tau)|}{D_m(i, r)} > \epsilon \quad (4-12)$$

where  $\epsilon$  is the threshold. Kennel *et al.* found that for  $\epsilon \geq 10$  the false neighbours are clearly identified. However, this criterion calculates an incorrect estimate when the signal contains noise or the sample length is limited. Then, the nearest false neighbour can be determined according to the criteria described in Equation (4-13).

$$\frac{D_{m+1}(i)}{D_A} \geq 2 \quad (4-13)$$

where  $D_A$  is the average distance between each point and all other points. It can be defined as

$$D_A^2 = \frac{1}{N} \sum_{i=1}^N [x(i) - \bar{x}]^2 \quad (4-14)$$

where  $\bar{x} = \frac{1}{N} \sum_{i=1}^N x(i)$ .

Finally, the embedding dimension  $m$  can be determined according to the nearest false neighbour with FNN.

#### 4.2.1.3 Predefined Recurrence Threshold Selection

From Equations (4-6) and (4-7), it is manifest that the recurrence threshold  $\epsilon$  should be chosen with special attention because if  $\epsilon$  is too small, the potential recursive features of the system will not appear properly. On the other side, if  $\epsilon$  is too large, it will result in many artefacts and tangential motion (misjudging simple continuous points on the trajectory) [120]. Moreover, noise affects the existing structure of the system signal. Therefore, it is foremost to select an appropriate threshold for various signals.

Some methods for recurrence threshold selection have been proposed and improved by researchers in these years. For example, it should not exceed 10% of mean or the maximum phase space diameter [120], [129]–[131]; it is selected in terms of the

recurrence point density [132]; also, it considers the standard deviation of the time series [133] and so on.

For rotating machinery with tapered roller bearings, fault features may be submerged by noise and be not apparent. According to this feature, it was found that the corresponding standard variance of the phase space trajectory for different cases is effective as the recurrence threshold in this study. The results will be presented in the next subsection to validate its availability on vibration signals of rotating machinery.

#### 4.2.2 Recurrence Quantification Analysis

Recurrence quantification analysis (RQA) is an optimal quantification method to analyse nonlinear recurrence plots of dynamical systems. The quantification measurements are defined based on recurrence density, diagonal and horizontal (vertical) structures to estimate chaos transitions presented by its phase space trajectory. All commonly used measurements will be introduced below and applied to the proposed compression method.

##### (1) Recurrence rate ( $RR$ )

The basic and simplest measure of RQA is the recurrence rate indicating the density of recurrence points and expressed as shown in Equation (4-15).

$$RR = \frac{1}{N^2} \sum_{i,j=1}^N R(i, j) \quad (4-15)$$

where  $R(i, j)$  is the recurrence point defined in Equations (4-6) and (4-7);  $N$  is the number of points on the phase space trajectory, also the length of the squared recurrence matrix. This indicator summarizes the proportion of the recurred states in the system.

##### (2) Determinism ( $DET$ )

$DET$  represents the percentage of recurrence points forming diagonal lines in RP.

$$DET = \frac{\sum_{l=l_{min}}^N lP(l)}{\sum_{l=1}^N lP(l)} \quad (4-16)$$

where  $P(l)$  is the probability of the length of diagonal lines with length  $l$ , which can be calculated with the histogram of length  $l$ ; and  $l_{min}$  is the minimum length of diagonal lines,  $l_{min} = 2$  in this study. Usually, a periodic signal has long diagonal lines; chaotic signals have very short ones; and stochastic signals have no diagonal lines [124]. As a result,  $DET$  will be large for a periodic system and very small for a stochastic system.

##### (3) Laminarity ( $LAM$ )

Similar to  $DET$ ,  $LAM$  is defined as the percentage of recurrence points generating vertical lines in RP as shown in Equation (4-17).

$$LAM = \frac{\sum_{v=v_{min}}^N vP(v)}{\sum_{v=1}^N vP(v)} \quad (4-17)$$

where  $P(v)$  is the probability of the length of vertical lines with length  $v$ , which can be determined by the histogram of the length  $v$ ; and  $v_{min}$  is the minimum length of vertical lines,  $v_{min} = 2$  in this study. It is a typical measure to analyse the intermittency and laminar states [134].

(4) Ratio (*RATIO*)

*RATIO* denotes the ratio between *DET* and *RR* and expressed as

$$RATIO = N^2 \frac{\sum_{l=l_{min}}^N lP(l)}{(\sum_{l=1}^N lP(l))^2} \quad (4-18)$$

(5) Averaged diagonal line length (*L*)

Moreover, the averaged diagonal line length is calculated in Equation (4-19) to measure the predictability time of the system.

$$L = \frac{\sum_{l=l_{min}}^N lP(l)}{\sum_{l=l_{min}}^N P(l)} \quad (4-19)$$

(6) Trapping time (*TT*)

Same to *L*, the averaged vertical line length reveals the trapping time of the system described as

$$TT = \frac{\sum_{v=v_{min}}^N vP(v)}{\sum_{v=v_{min}}^N P(v)} \quad (4-20)$$

(7) Longest diagonal line (*L<sub>max</sub>*)

*L<sub>max</sub>* is longest diagonal line on RP, which will be helpful to detect the periodicity of the phase space trajectory.

$$L_{max} = \max(\{l_i; i = 1, 2, \dots, N_l\}) \quad (4-21)$$

(8) Longest vertical line (*V<sub>max</sub>*)

Relative to the longest diagonal line, the longest vertical line *V<sub>max</sub>* is calculated according to the following equation.

$$V_{max} = \max(\{v_i; i = 1, 2, \dots, N_v\}) \quad (4-22)$$

(9) Divergence (*DIV*)

*DIV* is the inverse of *L<sub>max</sub>*, which would be an estimator for the Lyapunov exponent to the system.

$$DIV = \frac{1}{L_{max}} \quad (4-23)$$

(10) Diagonal entropy (*ENTR*)

*ENTR* is expressed according to the Shannon entropy to describe the complexity of RP as regards the diagonal lines. For the dynamical system with large noise, the complexity of the system is high, and the diagonal distribution is reduced leading to a larger value of *ENTR*.

$$ENTR = -\sum_{l=l_{min}}^N p(l)\ln(p(l)) \quad (4-24)$$

#### (11) Trend (*TREND*)

*TREND* represents the relationship between the density of recurrence points in a line parallel to the mean diagonal and its distance to the mean diagonal. It is expressed as

$$TREND = \frac{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2)(RR_i - \langle RR_i \rangle)}{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2)^2} \quad (4-25)$$

where  $RR_k = \frac{1}{N-k} \sum_{j=i=k}^{N-k} R(i, j)$ ;  $\langle \cdot \rangle$  is the function of calculating the average value;  $\tilde{N}$  is the maximal number of diagonal lines parallel to the mean diagonal. *TREND* is able to provide the stationary information of the dynamical system.

#### (12) RP entropy (*Entropy*)

In order to identify the distribution of points on the RP image, the entropy of RP is defined as

$$Entropy = -\sum_{i,j=1}^N p(R(i, j))\ln(p(R(i, j))) \quad (4-26)$$

where  $p(R(i, j))$  is the probability distribution of  $R(i, j)$ .

All these twelve measures or indicators will be used to extract the periodicity, stability, pattern characteristics from the RP image to analyse the phase space trajectory of the dynamical system. In general, RP reduces the system signal from multidimensional to two dimensions. Then, RQA extracts 12 features as features from the 2D signal of RP to represent the phase space trajectory. Although the number of extracted features is small, the number of segments divided is large due to the long collected data. Additionally, some of these features have a strong correlation, which is redundant for fault diagnosis. Eliminating the feature redundancy can improve the efficiency and accuracy of fault detection. Fortunately, the principal component analysis has the ability to reduce the dimension of features.

### 4.2.3 Principal Component Analysis

Principal component analysis is a statistical technique to implement exploration and visualization of high-dimensional datasets [135]. Its main objectives are data preprocessing (like denoising and feature extraction) and data compression (like dimension reduction). PCA can synthesize potentially high-dimensional variables into linearly orthogonal and independent low-dimensional ones called principal components (PCs) [136]. The new low-dimensional datasets, especially the first principal component, retain most of the information of the original data for the application.

The principle of PCA used for dimension reduction is introduced in detail. In the process of projecting  $N$ -dimensional features to  $k$ -dimensional ( $N > k$ ), PCA requires to find a set of orthogonal coordinate axes sequentially from the original space. The first new coordinate axis is selected as the direction with the largest variance in the

original data. Then, the second new one is chosen as the direction in which the variance is the largest in the plane orthogonal to the first new coordinate axis. The third new one is the axis with the largest variance in the direction orthogonal to the plane of the first two axes. With the same criteria,  $N$ -dimensional coordinate axes can be obtained. Finally, it was surprisingly found that most of the  $N$  coordinate axes were included in the first  $k$  coordinate axes, and the subsequent axes contained almost zero variance. Therefore, only the first  $k$  coordinate axes will be reserved because they keep most of the features of the original data. This is the procedure to achieve the dimension reduction of data features.

From a statistical point of view, the mean of the sample  $X$  is expressed as

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4-27)$$

The covariance between sample  $X$  and sample  $Y$  is described as

$$\text{Cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (4-28)$$

The covariance matrix of  $N$  samples is shown as

$$\text{Cov}(X_1, X_2, \dots, X_N) = \begin{bmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_N) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) & \cdots & \text{Cov}(X_2, X_N) \\ \cdots & \cdots & \cdots & \cdots \\ \text{Cov}(X_N, X_1) & \text{Cov}(X_N, X_2) & \cdots & \text{Cov}(X_N, X_N) \end{bmatrix} \quad (4-29)$$

In fact, the eigenvectors of the covariance matrix of the data is the  $N$ -dimensional coordinate axes. The matrix composed of eigenvectors corresponding to the  $k$  largest eigenvalues (i.e. the  $k$  largest variances) is the  $k$ -dimensional coordinate axes [136].

Let  $A$  be the covariance matrix, vector  $\vec{v}$  is the eigenvector of matrix  $A$ , then

$$A\vec{v} = \lambda\vec{v} \quad (4-30)$$

where  $\lambda$  is the eigenvalue corresponding to the eigenvector  $\vec{v}$ . A set of eigenvectors of a matrix is a set of orthogonal vectors. Moreover, the matrix  $A$  can be decomposed into

$$A = Q\Sigma Q^{-1} \quad (4-31)$$

where  $Q$  is a matrix composed of eigenvectors of the matrix  $A$ ;  $\Sigma$  is a diagonal matrix and the elements on the diagonal are the eigenvalues;  $Q^{-1}$  is the inverse matrix of  $Q$ .

PCA algorithm can be realized according to covariance matrix decomposed with the singular value decomposition method as shown in Equation (4-32).

$$A = U\Sigma V^T \quad (4-32)$$

where  $U$  and  $V$  are the left and right singular orthogonal vector matrices, respectively;  $V^T$  is the transposed matrix of  $V$ ; the elements on the diagonal of the matrix  $\Sigma$  are called singular values (recorded as  $\lambda_i, i = 1, 2, \dots, n$ ) arranged in descending order, and the other values are zero. To reduce the dimension of the features,

the vectors corresponding to the  $k$  largest singular values are selected as the PCs based on the following criterion:

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i} \geq 99\% \quad (4-33)$$

#### 4.2.4 Proposed Compression Method

As the system response of the rotating machinery is nonlinear with strong periodic characteristics, analysing the phase space trajectory of the vibration signal is an effective and efficient way to monitor the condition of rotating machines. But the high-dimensional trajectory is difficult to visualize and analyse. RP and RQA can reduce the dimension and extract features of the vibration signals, respectively. PCA is also applied to reduce the dimension of extracted features. A similar approach has been applied in other fields, such as the medical field [137]–[141], electric and electrical field [142], [142], dynamics [143], chemical field [144] and so on. The procedure of the proposed dimension reduction based compression method for fault diagnosis of tapered roller bearings in the mechanical field is described as displayed in Figure 4.6.

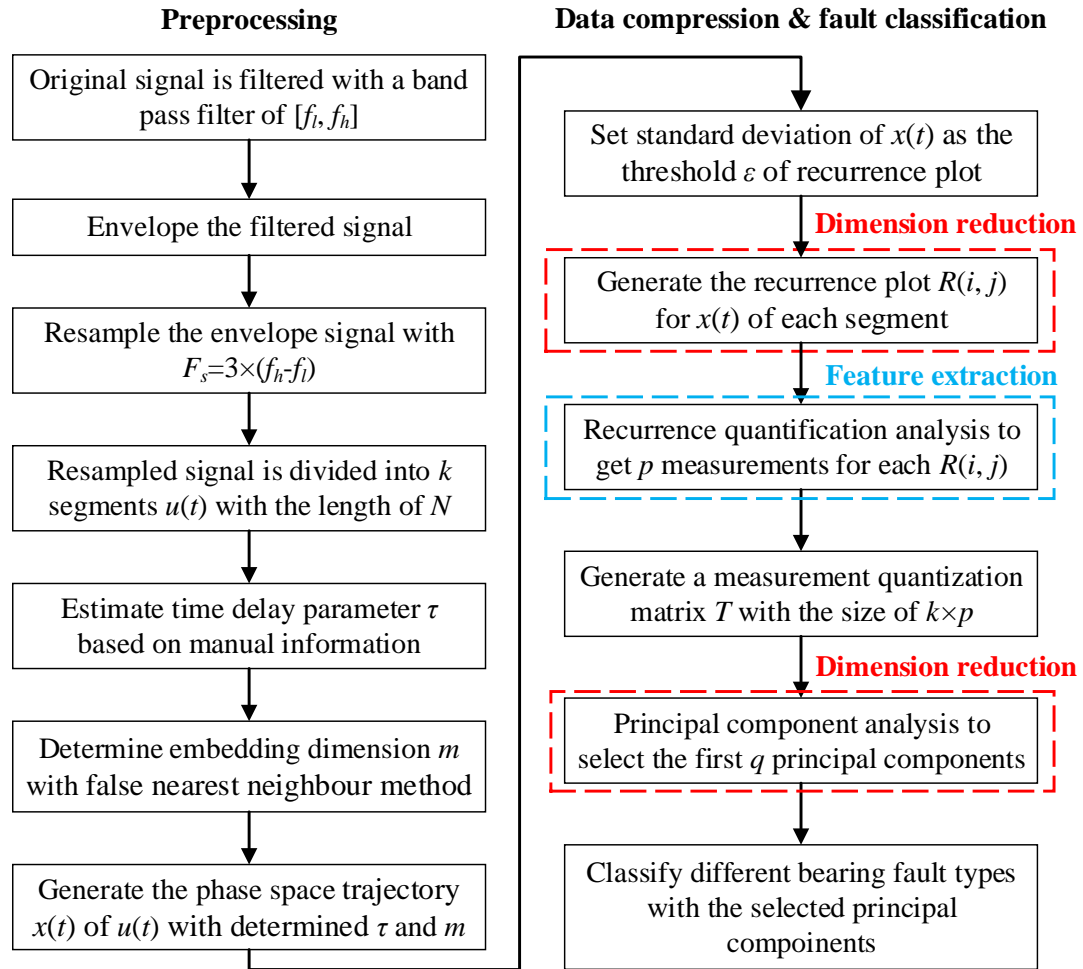


Figure 4.6 Flow chart of the proposed dimension reduction based compression method

The proposed compression method includes the preprocessing, data compression and fault classification with the vibration signals. Normally, the bearing fault frequency is modulated into a high-frequency band because of the modulation characteristics of the rotating mechanical system. As a result, a high sampling frequency (at least twice the highest modulation frequency) is required to acquire valuable information to monitor the condition of rotating machines according to the Nyquist sampling theorem.

In order to avoid the problem of high-dimensional calculation of large amounts of data, it is necessary to pre-process the collected vibration data in advance. The frequency band of  $[f_l, f_h]$  is around the carrier frequency of the bearing system, which is rich of the modulated fault information. Then, a band-pass filter is used to obtain this narrow but informative band signal. After that, Hilbert transform is applied on this filtered signal to get the envelope signal. The fault frequency is converted from the high frequency band to a low frequency band by demodulation in the envelope signal. As a result, the envelope signal can be resampled with the sampling frequency of  $F_s$  (where  $F_s = 3 \times (f_h - f_l)$ ) which not only meets the requirement of the Nyquist sampling theorem but also reduces the amount of data. The resampled signal is divided into  $k$  segments with the length of  $N$ . To obtain more abundant information related to the failure cycle, the high-dimensional phase space

However, it is difficult to directly observe the useful information from the high-dimensional signals even they have abundant information. So the dimensionality of the signal is reduced from  $m$  to 2 for each segment by generating a 2D recurrence graph according to the adaptive recurrence threshold  $\epsilon$ . The recurrence threshold is estimated by the corresponding standard variance of the phase space trajectory for different cases. This is the first step to compress the high-dimensional phase space trajectory of roller bearing signals. But it is difficult to directly observe accurate periodic information from the 2D recurrence graph. Hence,  $p$  features representing periodicity, stationarity, etc. are extracted from the 2D image of every segment by RQA to quantify fault characteristics. Therefore, the amount of data used for fault diagnosis is compressed to only  $k \times p$ , where  $k$  is the number of segments. Finally, the PCA method is applied to select the first  $q$  principal components to reduce the size of the extracted features from  $k \times p$  to  $k \times q$  (where  $q < p$ ). These  $k \times q$  components can be successfully used for the fault classification of tapered roller bearings.

### 4.3 Experiments

To validate the proposed dimension reduction based compression method, some experiments of tapered roller bearings were carried out. The test facilities and seeded faults will be introduced at first. Then, the experimental results will be discussed to verify the proposed compression method.

### 4.3.1 Experimental Setup

#### 4.3.1.1 Test facilities

A simple test rig was constructed to investigate the dynamics of tapered roller bearings because of their wide application in industry. Figure 4.7 shows the schematic diagram of the test rig, consisting of an induction motor, a flexible coupling, a ball bearing, a shaft and a tapered roller bearing. The tapered bearing with the model of TIMKEN 31308 is the test object with the specification listed in Table 4.1 [145]. The Optima controller accurately controls the speeds of the AC motor. The parameters of the motor are described in Table 4.2. A K-type thermocouple was mounted on the surface below the bearing house to monitor the operating temperature of the tapered roller bearing, which can avoid too high working temperature. In addition, a Hengstler incremental encoder was installed at the fan end of the motor and connected by an elastic pipe. The encoder signal allows the estimation of the accurate speeds in various working conditions. Then, a charge type piezoelectric accelerometer (CA-YD-104T) was directly installed on the surface of the bearing house which is circled with a red ellipse in Figure 4.7. A charge amplifier with the model of Type 2635 from Brüel & Kjær company was used to convert the charge output from accelerometers to voltage, allowing the collection by the data acquisition hardware. The 16-bit NI 6221 from the National Instruments with the maximum sampling rate of 250 kS/s was applied to collect the output of the charge amplifier. The principal parameters of the sensors and data acquisition devices are expressed in Table 4.3.

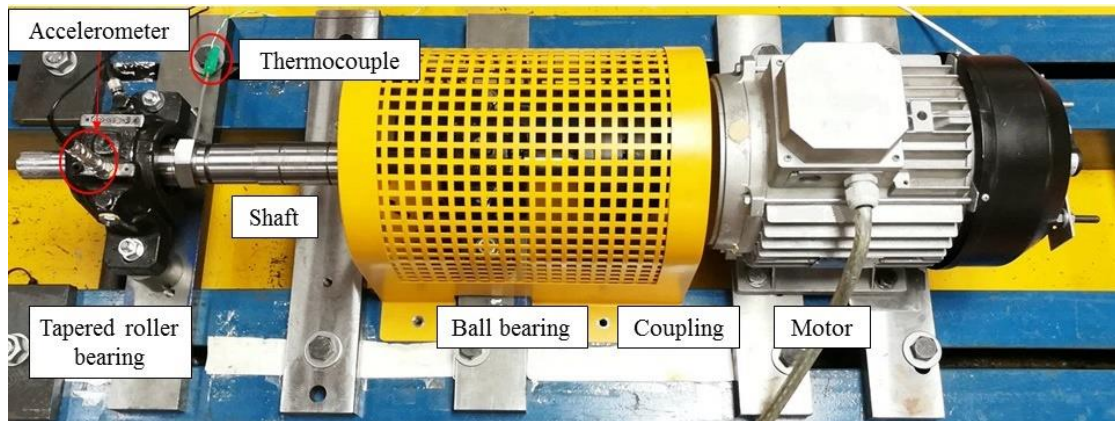


Figure 4.7 Tapered roller bearing test rig

Table 4.1 Specification of the TIMKEN 31308

Parameters	Value
Bore	40.0 mm
Cup Outer Diameter	90.00 mm
Cone Width	23.00 mm
Cup Width	17.00 mm
Bearing Width	25.25 mm
Number of Rollers	15

Table 4.2 Clarke AC motor

Parameters	Values
Rotating Speed	1420 RPM
Power	4 kW 3 phases
Frequency	50 Hz
Rated Voltage	230-400 V
Rated Current	15.9-9.2 A
Connection	Delta



Table 4.3 Specification of data acquisition equipment

Device	Parameters	Values
K-type Thermocouple	Maximum Temperature	+1100 °C
	Probe Diameter	3 mm
	Probe Length	150 mm
	Probe Material	Stainless steel
CA-YD-104T Accelerometer	Sensitivity	35 pC/g
	Measurement Range	800 g
	Frequency Response	1~7000 Hz
	Operating Temperature	-20~+120 °C
Hengstler Incremental Encoder	Number of Pulses	100
	Absolute Maximum Speed	Max. 6000 rpm
	Supply Voltage	5 VDC $\pm$ 10%
	Operating Temperature	-10~60 °C
	Vibration Proof	100 m/s <sup>2</sup> (10~2000 Hz)
	Shock Resistance	1000 m/s <sup>2</sup> (6ms)
Type 2635 Charge Amplifier Brüel & Kjær	Max. Input	10 <sup>5</sup> pC
	Sensitivity	Up to 10 V/pC
	Acceleration Output	0.1 mV to 1 V/ms <sup>-2</sup>
	Frequency Range	2 Hz TO 100 kHz
NI 6221 National Instruments	ADC Resolution	16 bits
	Sample Rate	250 kS/s
	Input Range	$\pm$ 0.2 V, $\pm$ 1 V, $\pm$ 5 V, $\pm$ 10 V
	Number of Channels	8 differential or 16 single ended

This test rig was designed to investigate the effect of clearance changes in the tapered roller bearing on fault diagnosis according to the independent characteristics of the inner and outer races of this bearing. Figure 4.8 explains the details of the tapered roller bearing test rig. A cylinder structure was designed to push the inner ring axially and a screw nut was used to support the cylinder upon the screws on the shaft. Meanwhile, there are two grub screws to prevent the cylinder from moving due to the rotation of the shaft. The clearance changes were precisely measured with a slip metric gauge box set with the type of Matrix Pitter 8075 C [146], [147]. The larger the measured clearance shown in red arrows, the smaller the bearing clearance.

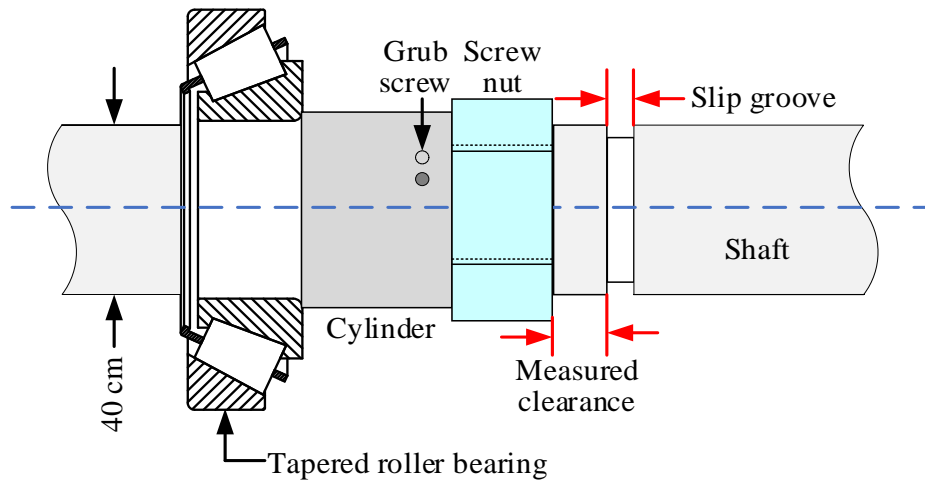


Figure 4.8 Diagram of the tapered roller bearing structure

#### 4.3.1.2 Defect seeding

To simulate failure of the tapered roller bearing, the defects with the length of 4mm were created artificially on the outer race and the roller of two TIMKEN 31308 bearings, respectively, which are displayed in Figure 4.9. The bearing geometry is described in Table 4.1. One healthy bearing will be used as the baseline to compare with the faulty ones.

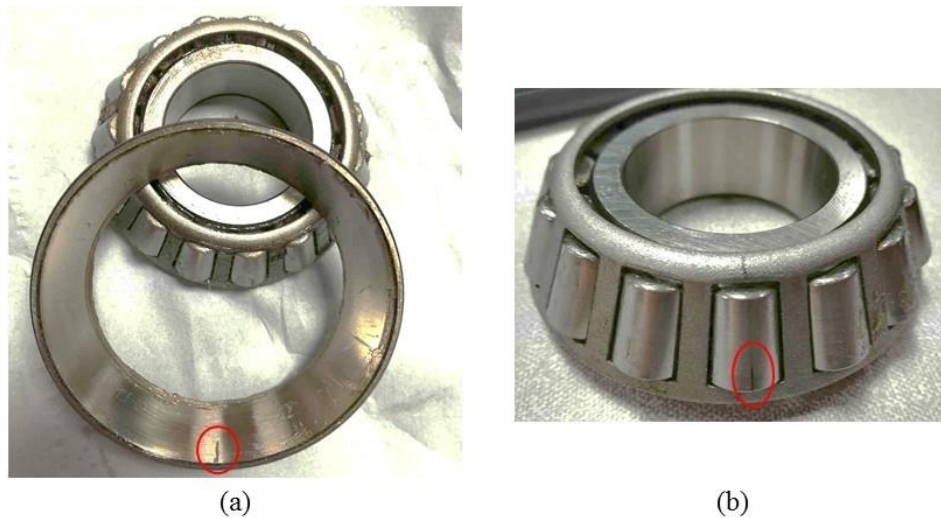


Figure 4.9 Tapered roller bearings with seeded defects

However, the contact angle of the tapered roller bearing is usually recommended between  $10^\circ$  and  $30^\circ$  when it operates. There is no fixed angle to calculate the fault frequency. Hence, the fault frequencies of the outer race and roller, as well as the effect of the bearing clearance are considered to be confirmed with the experiments.

#### 4.3.1.3 Test procedure

Any unreasonable installation method can lead to improper clearance between bearing components. The improper clearance will shorten the lifespan of the tapered roller bearings. Figure 4.10 illustrates the life curve of the TIMKEN 31308 tapered roller

bearing affected by the installation clearance. When the axial clearance between the components is very close to zero, the rated life of the bearing is very high. It will decrease when the axial clearance increases in the direction of both preload and endplay. However, zero axial clearance is very difficult to achieve with the manual installation.

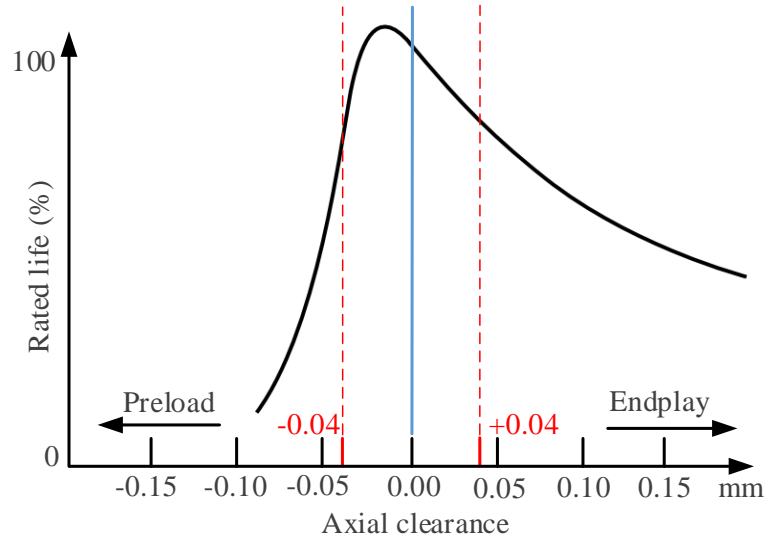


Figure 4.10 Typical life curve affected by the installation clearance

In this experiment, two objectives will be achieved: diagnosis of different faulty bearings and influence of bearing component clearance on fault diagnosis. To achieve the first objective, the healthy bearing and faulty bearings were successively installed at zero axial clearance in the test rig. For the achievement of the second objective, bearings with the outer race fault and roller fault were installed at +0.04 mm clearance in the endplay direction; then, screwing the nut to the preload direction to make the clearance to be +0.02 mm, 0 mm and -0.02 mm in order. The AC motor operated at the full constant speed of about 1500 rpm without any additional load. The operation speed, temperature and vibration signals were collected by the sampling frequency of 50 kHz and recorded for 30 seconds with three repetitions. All recordings of these tests started when the temperature observed by the thermocouple was up to 30°C because the temperature has a large influence on the bearing component clearance.

#### 4.3.2 Fault Frequencies Estimation and Effect of Clearance Change

As mentioned previously, the contact angle of the tapered roller bearing is difficult to determine for the fault frequency calculation. The fault frequencies of 31308 tapered roller bearings can be approximately estimated with the geometric dimensions shown in Figure 4.11 which are obtained using AutoCAD to measure the company's recommended bearing structure. The estimated bearing fault frequencies at the rotating speed of 1500 rpm are listed in Table 4.4.

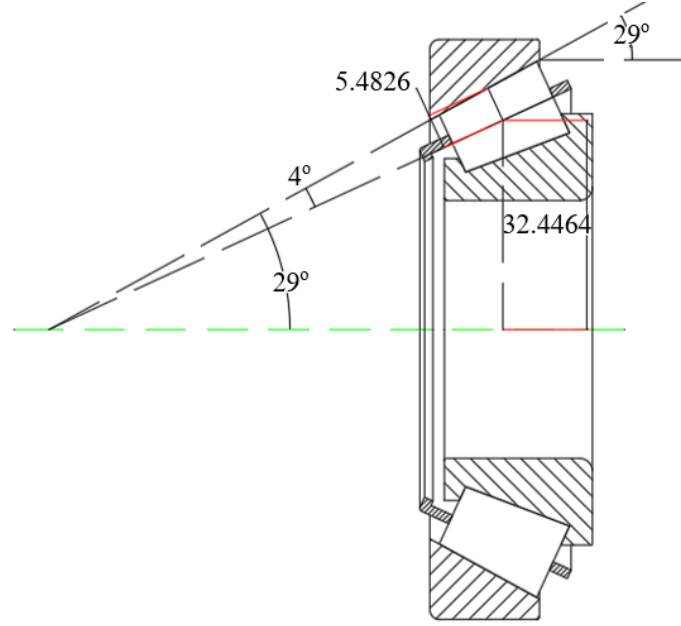


Figure 4.11 Geometric dimensions of TIMKEN 31308

Table 4.4 Estimated fault frequencies of 31308

Parameter	Outer Race Fault $f_o$	Cage Fault $f_{cage}$	Roller Fault $f_r$	Inner Race Fault $f_i$
Frequency (Hz)	159.79	10.65	72.36	215.21

However, contact between the rollers and raceways is not a stable line contact in the procedure of mechanical operation. Besides, the contact angle will also change randomly. Hence, fault frequencies cannot match the experimental results, especially the roller fault frequency.

To determine the fault frequency of the bearing more accurate, an exploratory experiment was carried out on two bearings with the independent large outer race fault and roller fault. The experiment is to examine the parameters used for the frequency calculation. Because the contact angle and the pitch radius are difficult to measure, the outer race fault and roller fault are used to calculate the contact angle and the ratio of roller and pitch diameters. Let  $x = \frac{D_r}{D_p}$ ,  $y = \cos\alpha$ , then

$$x = \frac{f_{sr}}{2f_r} \left[ 1 - \left( 1 - \frac{2f_o}{N_r f_{so}} \right)^2 \right] \quad (4-34)$$

$$y = \left( 1 - \frac{2f_o}{N_r f_{so}} \right) / x \quad (4-35)$$

where  $f_{so}$  and  $f_{sr}$  are the shaft frequency when the fault type is the outer race fault and roller element fault, respectively;  $f_r$  is the roller fault frequency and  $f_o$  is the outer race fault frequency;  $N_r$  is the number of rollers. As explored by Su *et al.* in [148], some extra frequencies will occur caused by various reasons, which are described in Table

4.5. The roller fault frequency usually appears in the form of twice the frequency, which should be about 134 Hz in this test.

Table 4.5 Extra frequencies caused by various reasons for different defects

Causes	Extra Frequencies for Different Defects		
	Outer Race Fault	Inner Race Fault	Roller Fault
Stationary Loading	–	$f_s$	$f_{cage}$
Shaft Unbalance	$f_s$	–	$f_s - f_{cage}$
Roller Errors	$f_{cage}$	$f_s - f_{cage}$	–
Transmission Path	–	$f_s$	$f_{cage} \& 2f_r$

The parameters  $x$  and  $y$  are calculated based on Equations (4-34) and (4-35) with two different experimental cases. It is apparent that there is a certain error between the theoretical and practical values by comparing Table 4.4 and Table 4.6.

Table 4.6 Values of  $x$  and  $y$  calculated with the exploratory experiment

Parameter	Outer Race Fault (Hz)		Roller Fault (Hz)		$x$	$y$	Contact Angle
	$f_o$	$f_{so}$	$f_r$	$f_{sr}$			
Values	156.92	24.98	134.12	24.98	0.1813	0.8956	26.4°

### 4.3.3 Results and Discussion

The experiments are mainly divided into two tests: the first test monitors the healthy and faulty bearings operating at zero clearance between different components; the other test investigates the effect of different clearances on fault diagnosis of the tapered roller bearing. They will be analysed with the proposed compression method based on the traditional signal processing method. The results will be displayed and discussed in detail as follows.

#### 4.3.3.1 Healthy and faulty bearing diagnosis

The first test aims to distinguish the healthy and faulty bearings at the clearance of components of 0mm. The captured datasets sampled with the sampling frequency of 50kHz and recorded for 30 seconds are filtered by a band-pass filter with the frequency band of [3800, 4500] Hz. Then, the fault signals can be demodulated from the filtered vibration signals through the envelope analysis. Figure 4.12 displays parts of the envelope signals for the healthy and faulty tapered roller bearings. It is obvious that the envelope amplitude of the healthy bearing is smaller than that of the faulty ones. Additionally, there is no apparent periodicity for the healthy bearing. However, the envelope signals of the outer race fault bearing and the roller fault bearing have the significantly obvious periodicity of about 0.04 and 0.1 seconds, respectively.

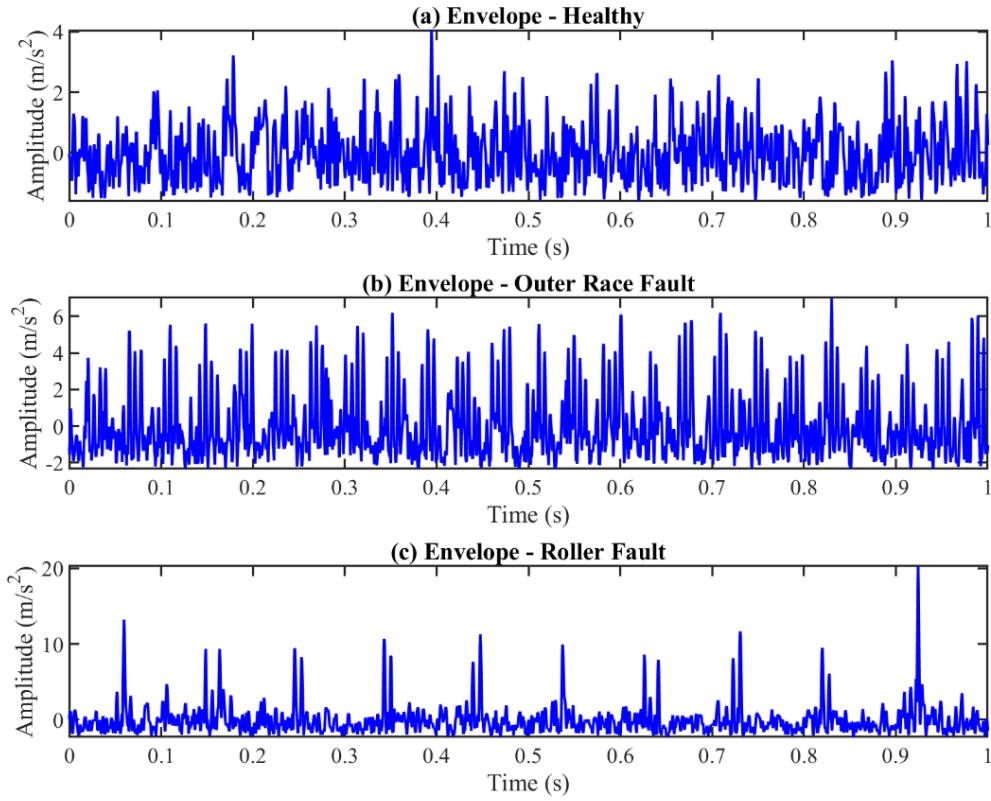


Figure 4.12 Envelope signals of the healthy and faulty bearings

In order to further diagnose the types of bearing faults, the spectra of the envelope signals are calculated with the fast Fourier transform as shown in Figure 4.13. For the healthy bearing, only the shaft frequency,  $f_s=24.9$  Hz, clearly exists in Figure 4.13 (a), which states that the imbalance occurs induced by the manufacture or installation error. In Figure 4.13 (b), in addition to the shaft frequency, the outer race fault ( $f_o=156.9$  Hz) and its second harmonic, as well as their left and right sidebands related to the shaft frequency are apparently illustrated. However, the spectrum of the roller fault bearing is very complex in Figure 4.13 (c). As mentioned in Table 4.5, the signal of the roller fault bearing contains the cage fault frequency ( $f_{cage}=10.5$  Hz) and the roller fault frequency ( $f_r=134.1$  Hz) in the spectrum. Besides, the harmonics and sidebands are also extremely intensive which complicates the signal periodicity.

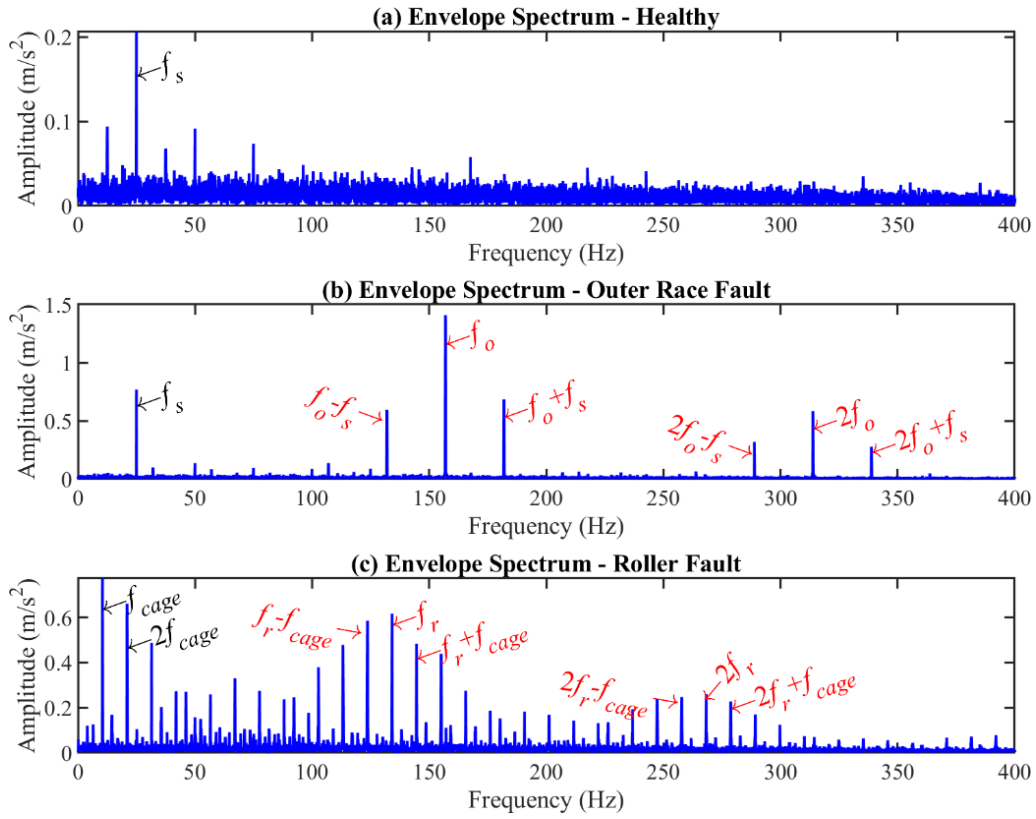


Figure 4.13 Envelope spectra of the healthy and faulty bearings

The trajectory in a phase space presents all possible states of a dynamical system. In order to extract more information to diagnose the fault types of the tapered roller bearing with fewer samples and features, a dimension reduction based compression method is proposed by combining the RP, RQA and PCA methods.

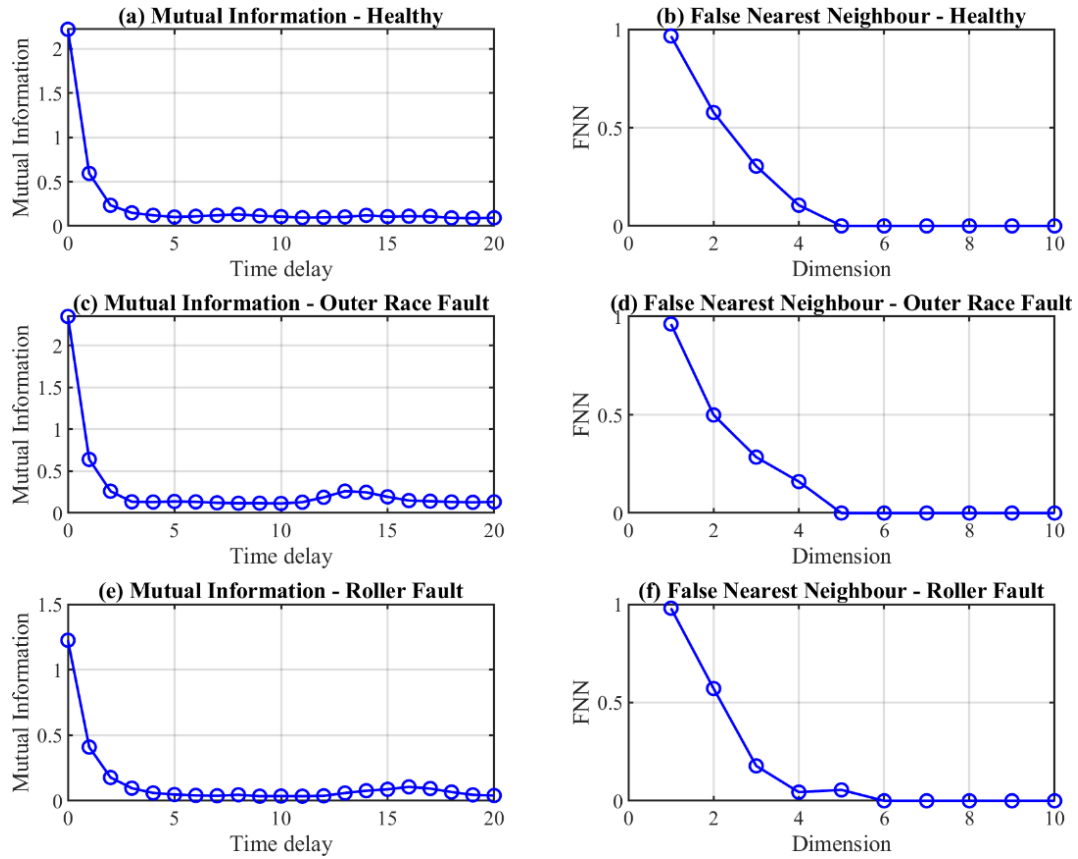


Figure 4.14 Mutual information and FNN results of the three cases

The vibration signal should be pre-processed as Figure 4.6 shown. After the resample, the amount of data will be reduced to 2100 per second with the time period of 30 seconds. Then, the resampled signal is divided into 63 segments with the length of 1000 points for every segment. For each segment, a phase space will be created according to the determined time delay and embedding dimension. Figure 4.14 (a), (c) and (e) illustrate the mutual information between a time series signal and its time-delayed signals for the healthy and faulty bearings, respectively. The mutual information can be regarded as the amount of information about another random system contained in a random system. The time delay  $\tau$  can be selected in terms of mutual information, which is selected as the set of [3 3 3] for these three cases. With the aid of the delayed time selected, the embedding dimension  $m$  of the phase space can be confirmed by the FNN as shown in Figure 4.14 (b), (d) and (f). The dimension that none or just a few false nearest neighbours existing is the optimal dimension to choose. As a result, the dimension is chosen as the set of [5 5 6] for these three cases. With the determined time delay and dimension, the phase space trajectory can be generated. The first three-dimensional trajectories of the healthy and faulty bearings are as shown in Figure 4.15.



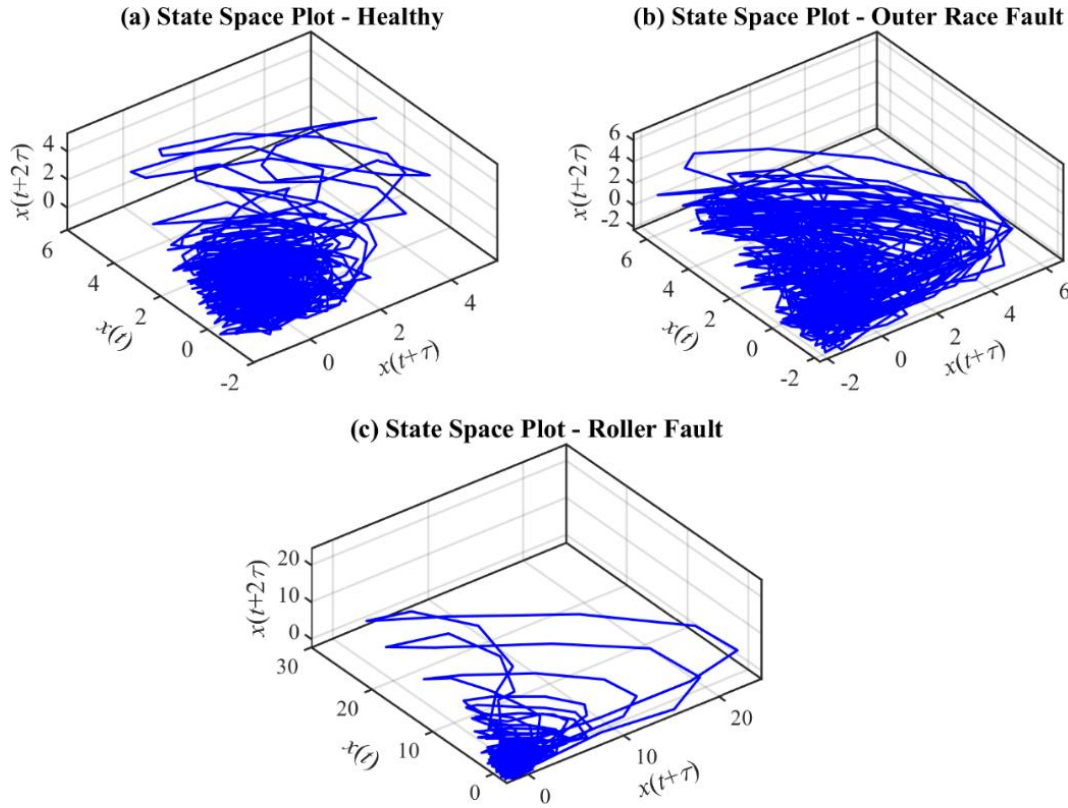
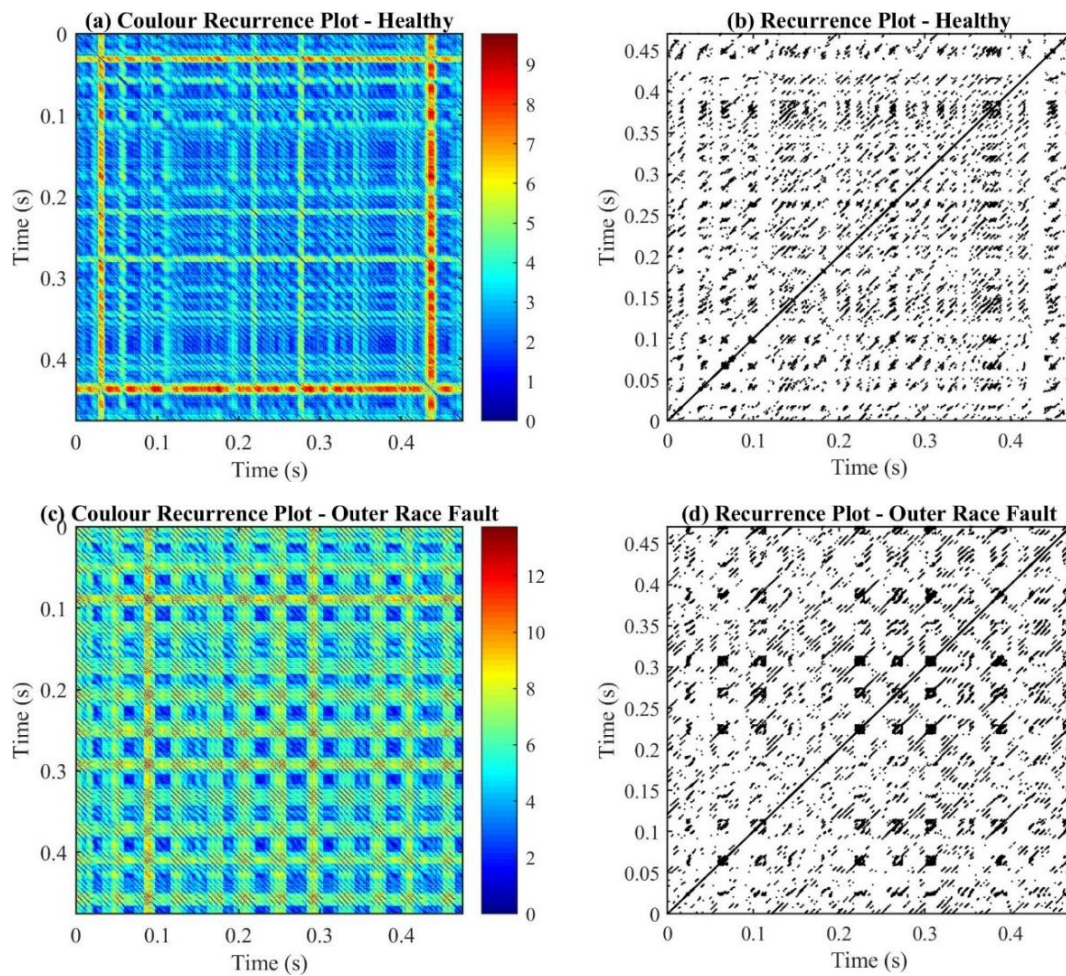


Figure 4.15 State space plots for the healthy and faulty bearings

To reduce the dimension of the trajectories, the 2D colour recurrence figures are plotted as illustrated in Figure 4.16 (a), (c) and (e). Two axes of the recurrence plots represent time. The colour recurrence plot is the real visual representation of the recurrence matrices. The colour values demonstrate that the Euclidean distance of the healthy bearing is shorter than the faulty ones. Moreover, the appearance of the periodic patterns is very different for these three cases, which indicates substantial differences in the periodicity of the vibration signals. However, it is difficult to get more information from these colour recursive figures. Therefore, the white-black recursive graphs are plotted with the recurrence threshold setting as the standard deviation of the trajectories [1.0906, 1.8187, 2.5510] in Figure 4.16 (b), (d) and (f) for these three different cases. It can be seen that the structural patterns of the recurrence plots of the faulty bearings are much more complicated than that of the healthy one, and the roller fault bearing is the most complicated.

The recursive characteristics of the healthy bearing are distributed in Figure 4.16 (b). It can be seen that the main diagonal is highlighted in the black colour. But there are no other long lines parallel to the main diagonal. Only some very short diagonal lines are visible possibly caused by the background noise or the slight defect. Most areas of the recurrent plot are the single black points. It expresses that the periodicity of this trajectory is very weak, and this result is consistent with the signal characteristics of this healthy bearing as shown in Figure 4.13 (a). Compared with the healthy bearing, the texture in the recursive graph of the outer race fault bearing illustrated in Figure 4.16 (d) is much more transparent to show the periodical information as displayed in

Figure 4.13 (b). Firstly, the diagonal lines except the main one is a little longer and more regular than those of the healthy bearing, which represents the transient behaviour like periodic impulse response. In addition, the pattern distribution is relatively uniform, indicating that the periodicity is relatively stable. Figure 4.16 (f) draws the white-black recursive graph of the roller fault bearing. It can be seen that the large diagonal structures occur as the laminar segments, which states the periodicity related to the cage frequency. Moreover, the distance between diagonal lines has small fluctuations explaining multiple periodicity emerges in this vibration signal. Finally, the recursive pictures specify that the pattern implies sufficient information to determine whether the bearing is healthy or faulty. However, RQA can be relied on to get more accurate states of the bearing instead on observing the recurrence plots. Simultaneously, it achieves the feature extraction process to reduce the number of parameters to be saved.



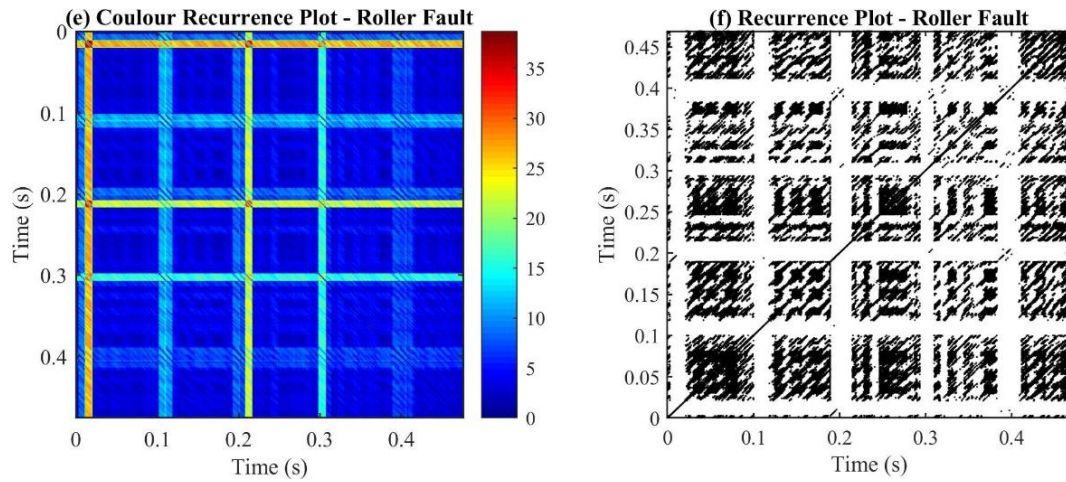


Figure 4.16 Colour and white-black recurrence plots for the healthy and faulty bearings

Twelve recursive quantization parameters introduced in the previous subsection are extracted from the white-black recurrence plot of each segment to generate a feature matrix with the size of  $63 \times 12$ . Nevertheless, according to the definition of the RQA indicators, it is obvious that there is redundancy between these parameters. As a result, the PCA algorithm is used to decompose the features and obtain irrelevant PCs and reduce the dimension of the signal characteristics. Then, the first three PCs account for more than 99% are selected as values for the three axes to be plotted to classify different faults as shown in Figure 4.17. It is obviously verified that these three bearing states can be clearly distinguished. The points of the healthy bearing (blue star points) and the outer race fault bearing (red star points) gather very well because the periodical information of the dynamic system response output is relatively stable. Only these black points of the roller fault bearing are a little discrete because the fault frequencies involved in the roller fault bearing is relatively more. But they can be well separated from the blue and red star points to effectively implement the fault diagnosis of tapered roller bearings.

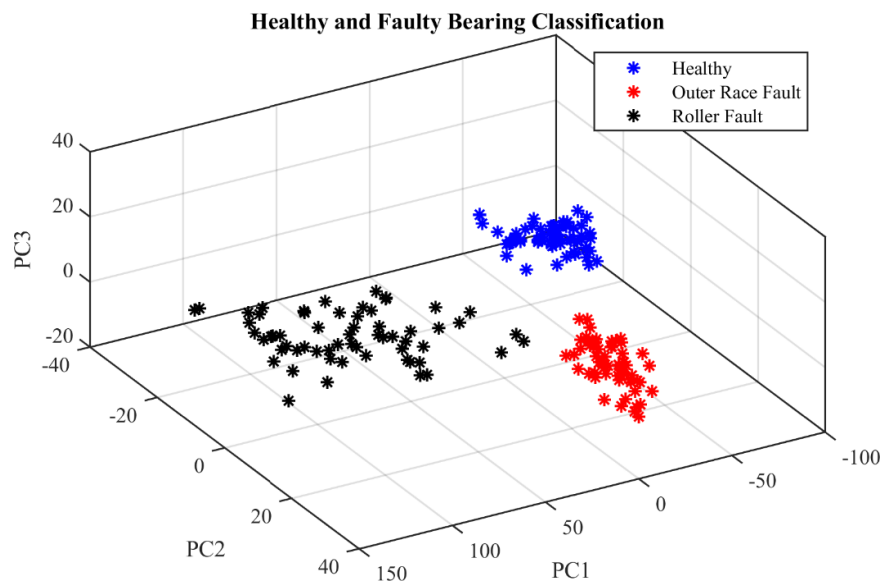


Figure 4.17 Scatter plot of the healthy and faulty bearings



#### 4.3.3.2 Effect of clearance on bearing defect diagnosis

To investigate the effect of clearance on bearing defect diagnosis, some tests to change the clearance between components for the outer race fault bearing and the roller fault bearing were carried out. The results will be discussed as follows.

##### (1) Outer race fault bearing

The data sampling rate and the recorded period are the same as the previous experiment settings. The envelope signals of the filtered signals for different outer race fault bearing clearances are shown in Figure 4.18. As the clearance decreases, the periodicity becomes more and more obvious.

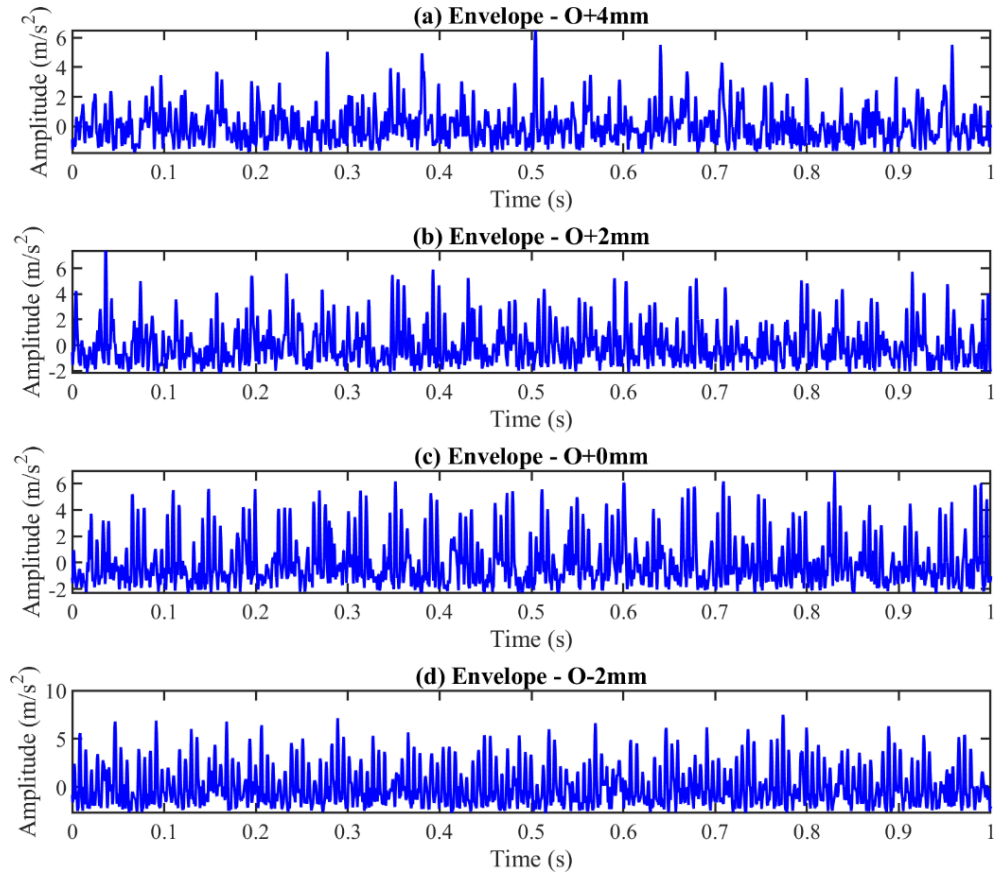


Figure 4.18 Envelope signals of the outer race fault bearing at different clearances

The envelope spectra of these four different clearances are illustrated in Figure 4.19. The shaft frequency  $f_s=24.9$  Hz clearly emerges together with its harmonics, which demonstrates the misaligned shaft. Moreover, the outer race fault frequency and its second harmonic are very clear and recorded as  $f_o$  and  $2f_o$  in Figure 4.19. Besides, the left and right sidebands of the fault frequency and the second harmonic are also very apparent. More importantly, the outer race fault frequency decreases and its amplitude increases as the clearance reduces. This phenomenon indicates that the peak value of the fault frequency cannot be used to diagnose the severity of the defects when the bearing clearance changes because the clearance between the bearing components also affects the fault frequency value and its amplitude.

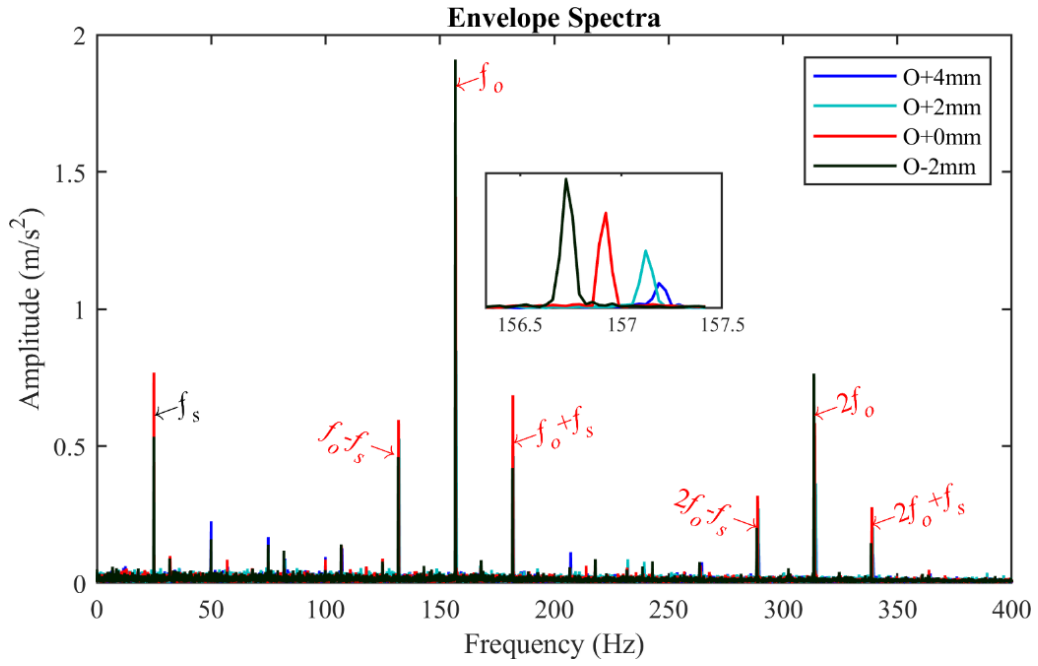


Figure 4.19 Envelope spectra of the outer race fault bearing at different clearances

To reduce the number of samples for the fault diagnosis, the proposed dimension reduction based compression methods is applied for the envelope signals. The relationship between the mutual information and time delay are described in the left columns in Figure 4.20. The determined time delays of [2, 3, 2, 3] and dimensions of [5, 5, 6, 5] are used to create the phase space trajectories for these four cases with different clearances. The generated first three-dimensional phase space trajectories are displayed in Figure 4.21.

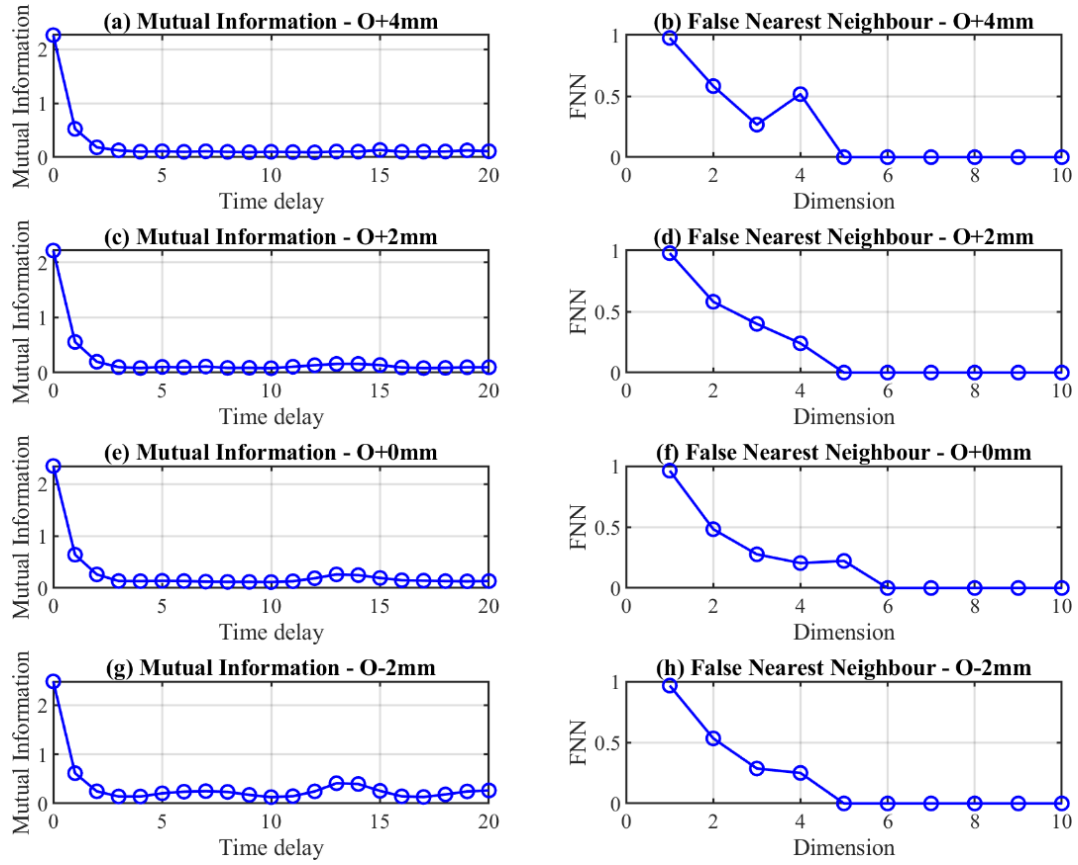


Figure 4.20 Mutual information and FNN results for the outer race fault bearing at difference clearances

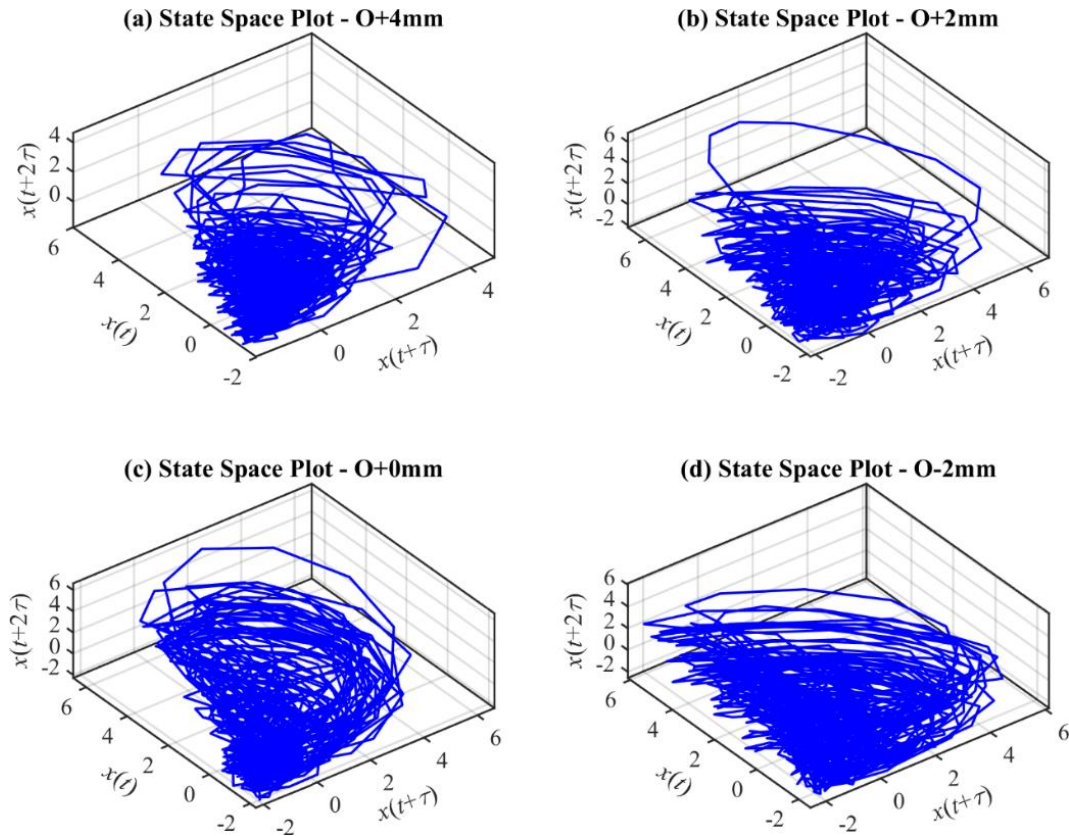
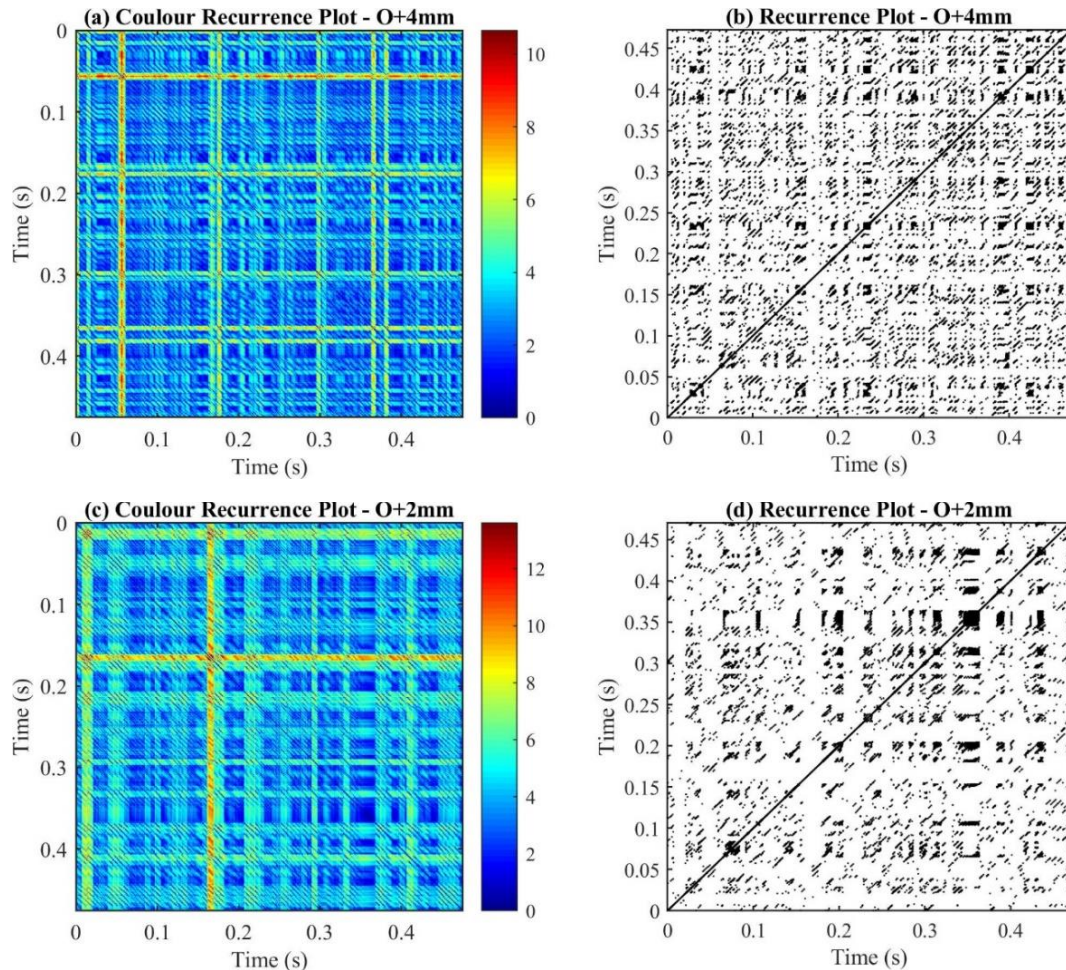


Figure 4.21 State space plots for the outer race fault bearing at different clearance

To extract features and reduce the dimension of the trajectories, the 2D colour and white-black recursive graphs are plotted in the left and right column of Figure 4.22, respectively. The white-black recurrence plots are determined with the thresholds of [1.0946, 1.4916, 1.8177, 1.9341] for these four cases. All four white-black recursive figures contain the diagonal lines representing the periodicity of signals. There are many short diagonals and single black dots in the case of clearance of +4 mm as shown in Figure 4.22 (b), which means the signal is a little chaotic and with the background noise. Figure 4.22 (d) and (f) are very similar with less single black dots and have some longer diagonal lines compared to Figure 4.22 (b). Some very small orthogonal segments appear in these two cases indicating that the states at these locations change slowly or does not change. The recurrence plot for the case with the clearance of -2 mm is displayed in Figure 4.22 (h). The longer diagonal lines are clearly visible, which explains the periodicity of the signal is significant. But the recursive figures are difficult to interpret the fault diagnosis of different clearances of the outer race fault bearing. Hence, RQA indicators are calculated to quantify the features in the RP figures.





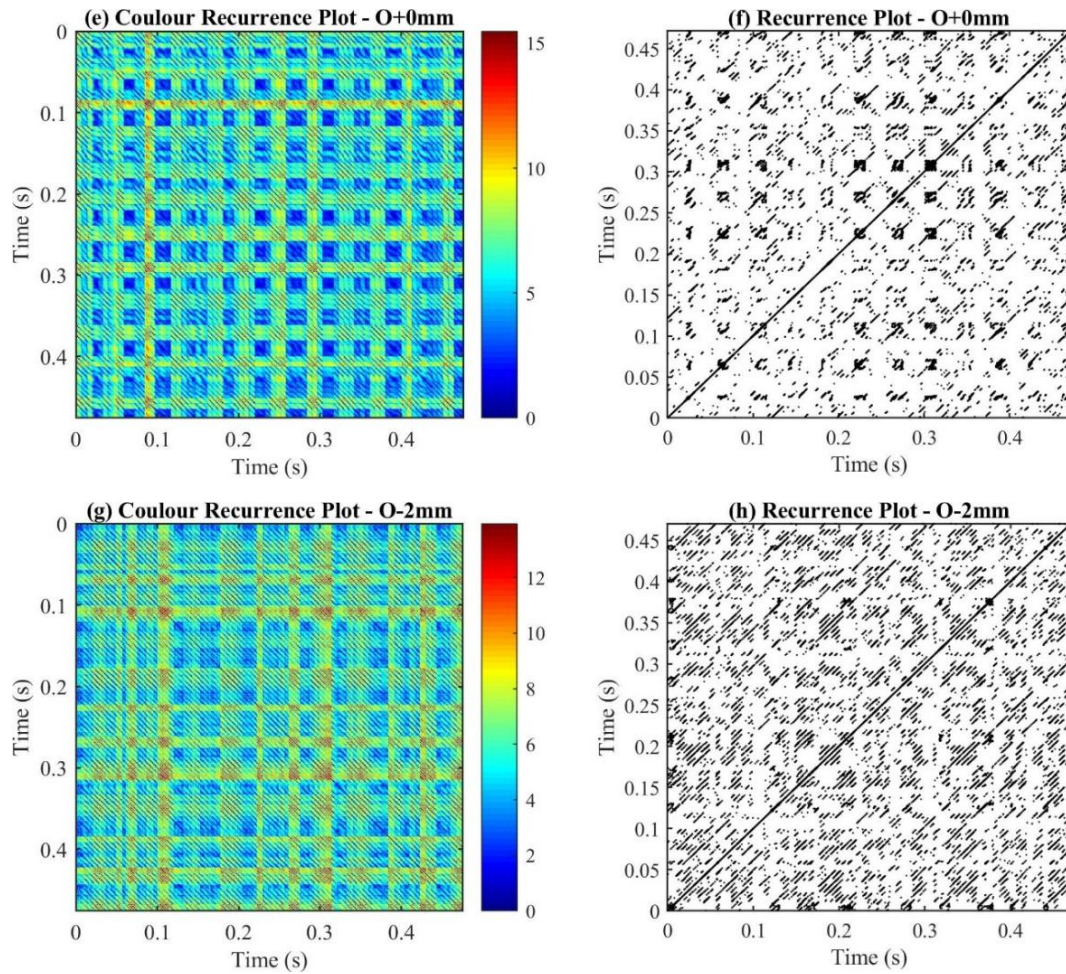


Figure 4.22 Colour and white-black recurrence plots for the outer race fault bearing at different clearance

As presented in the previous test, twelve recursive quantization parameters are extracted to form a  $63 \times 12$  feature matrix. Then, the PCA method is applied on the extracted feature matrix to reduce the redundancy of the features. The size of the features is reduced to  $63 \times 3$  because the first three PCs account for more than 99% of all information. The selected PCs for these four cases are plotted in Figure 4.23. It can be seen that these four cases can be clearly distinguished, which is used to determine whether the bearing clearance is normal or not to avoid misjudging the severity of bearing defects.



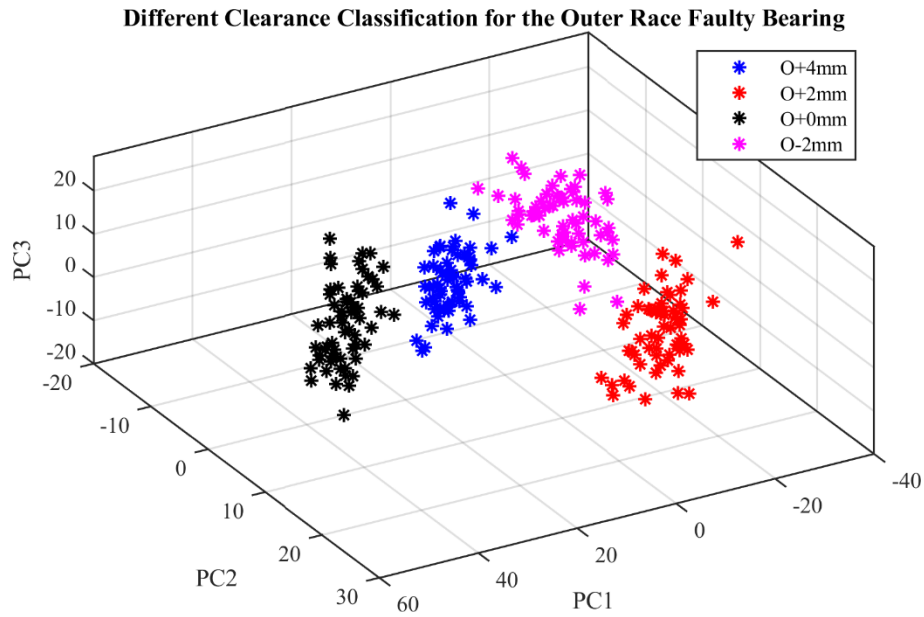


Figure 4.23 Scatter plot of different clearances of the outer race fault bearing

## (2) Roller fault bearing

Similarly, some experiments about four different clearances were performed with a roller faulty tapered roller bearing. Figure 4.24 illustrates the envelope signals of these four cases. The periodicity in Figure 4.24 (a) is inconspicuous which may be induced by the poor contact between the faulty roller and the inner and outer races. There are significant impact signals in Figure 4.24 (b) to (d) generated by the excitation of the faulty roller. But the level of the SNR of the signal in Figure 4.24 (b) is lower than another two.

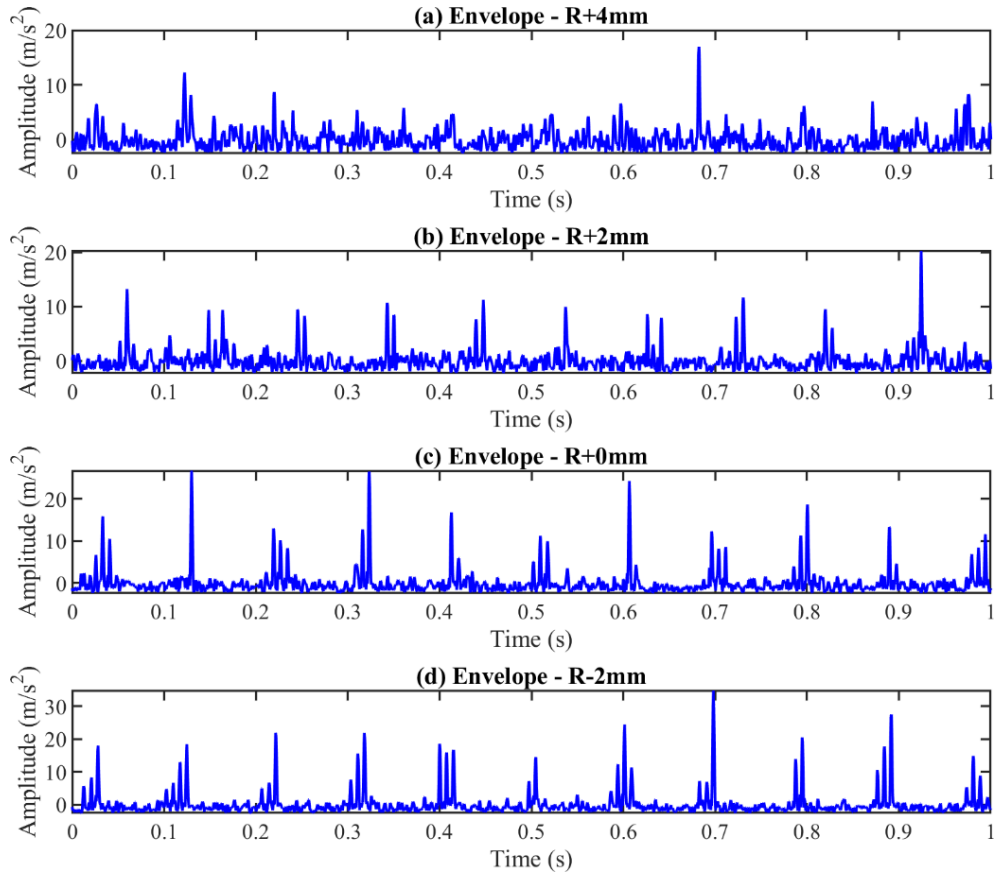


Figure 4.24 Envelope signals of the roller fault bearing at different clearances

Figure 4.25 states the envelope spectra of the roller fault bearing at four different clearances. The rotating frequency  $f_s$  is not very conspicuous because its magnitude is smaller than that of the cage fault frequency  $f_{cage}$  and its harmonics. Moreover, the roller fault frequency  $f_r$  and its second harmonic, as well as many sidebands of them related to the cage fault frequency are distinctly illustrated. The frequency band of [133.8, 134.6] Hz is enlarged to detail the roller fault frequency. The fault frequency peak of the clearance of +4 mm did not appear, which may be induced by the non-contact of the contact surface because of large clearance. Both the fault frequency magnitude and fault frequency value increase as the clearance decreases.

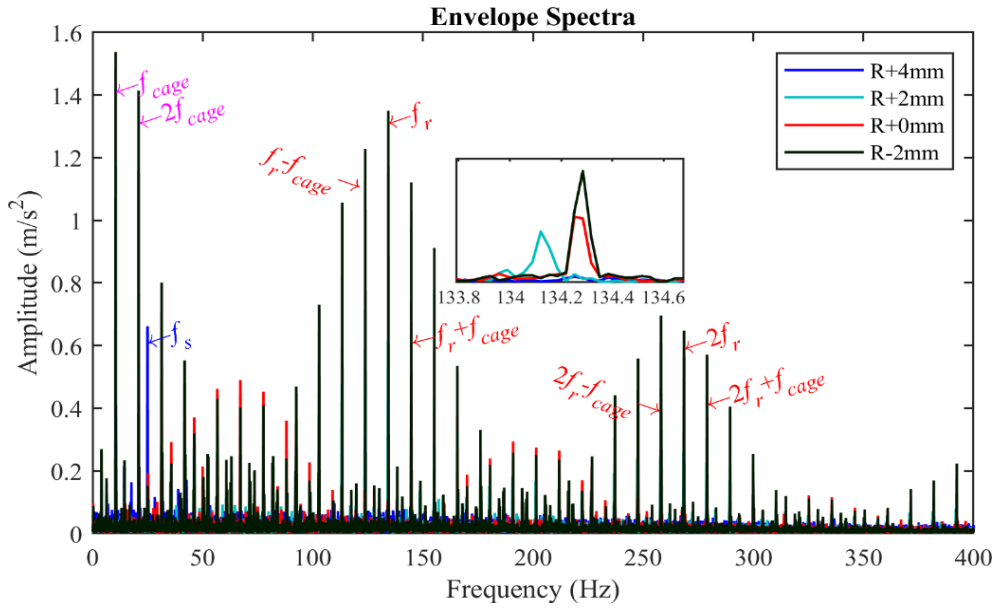


Figure 4.25 Envelope spectra of the roller fault bearing at different clearances

As with the previous procedure, time delays and dimensions are explored according to the mutual information and FNN to create the trajectories of these four cases. The time delays and dimensions are determined as [4, 5, 4, 5] and [6, 6, 6, 6] based on the results displayed in Figure 4.26. The first three-dimensional phase space trajectories of these four cases are plotted in Figure 4.27.

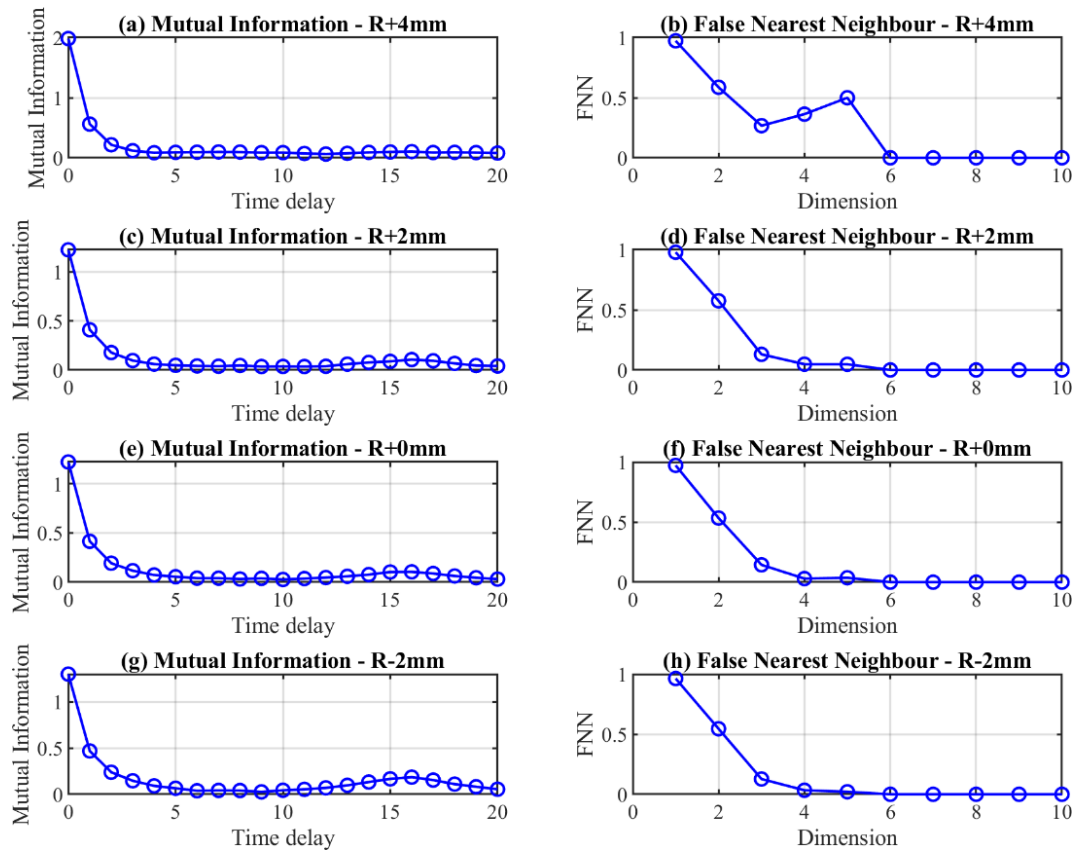


Figure 4.26 Mutual information and FNN results for the roller fault bearing at difference clearances

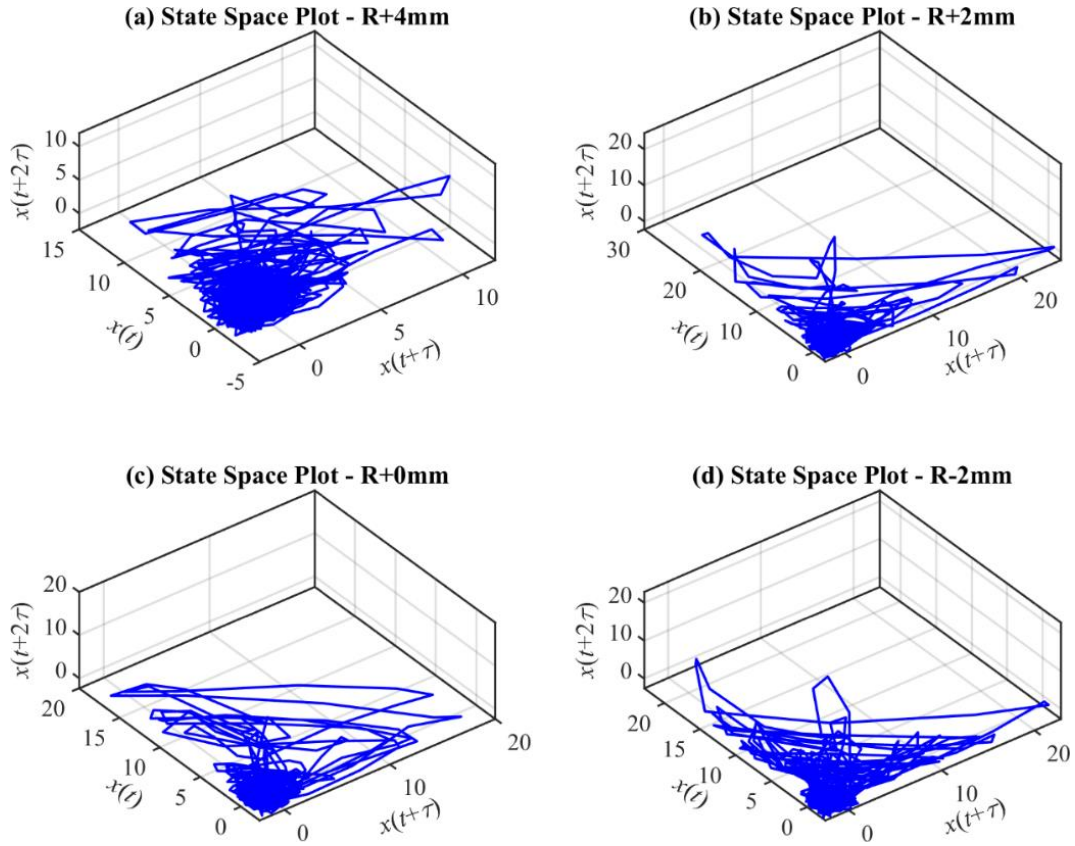


Figure 4.27 State space plots for the roller fault bearing at different clearance

Figure 4.28 presents the colour and the white-black recurrence plots for these four different clearances of the roller fault bearing. The recurrence thresholds are set as [1.9380, 2.5577, 2.3800, 2.9056] for these four cases, respectively. They differ from the recursive maps of the outer race fault bearing because the diagonal lines exist as the laminar segments. Actually, these blocked square structures can reflect the cage fault frequency because they appear in a period of about 0.1 second. There are large white bands between the diagonal structures to form rectangular structures. This phenomenon indicates that the states of these locations rarely occur. It is very easy to distinguish the states of the clearance +4 mm from that of -2 mm because the density of the points is dramatically different. The general trend is that the fault display is clearer as the clearance is reduced. However, it is very difficult to distinguish all these four cases with RPs because their structures are hard to interpret.



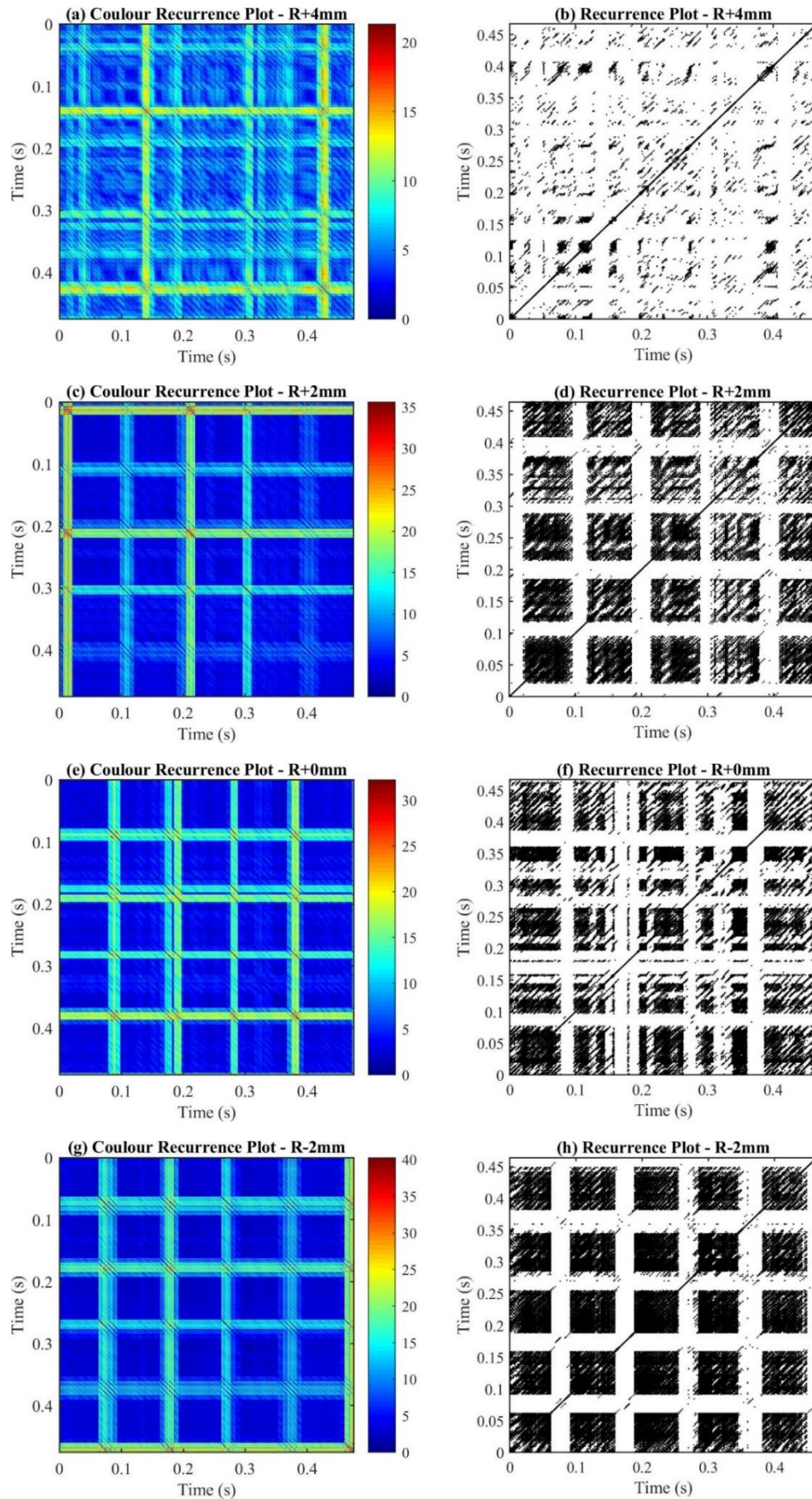


Figure 4.28 Colour and white-black RPs for the roller fault bearing at different clearance

A  $63 \times 12$  feature matrix is also generated with twelve extracted recursive quantization indicators for these four cases. The PCA algorithm is applied to eliminate the redundancy between these features and reduce the size of the matrix to  $63 \times 3$ , simultaneously. The selected PCs are applied for the classification with the result as shown in Figure 4.29. It is obvious that the cases with the clearances of +4 mm and -2 mm are classified very well, but the other two cases have some overlaps with each other. But their cluster centres have a certain distance. It is not very easy to accurately identify a small clearance between the roller fault bearing components.

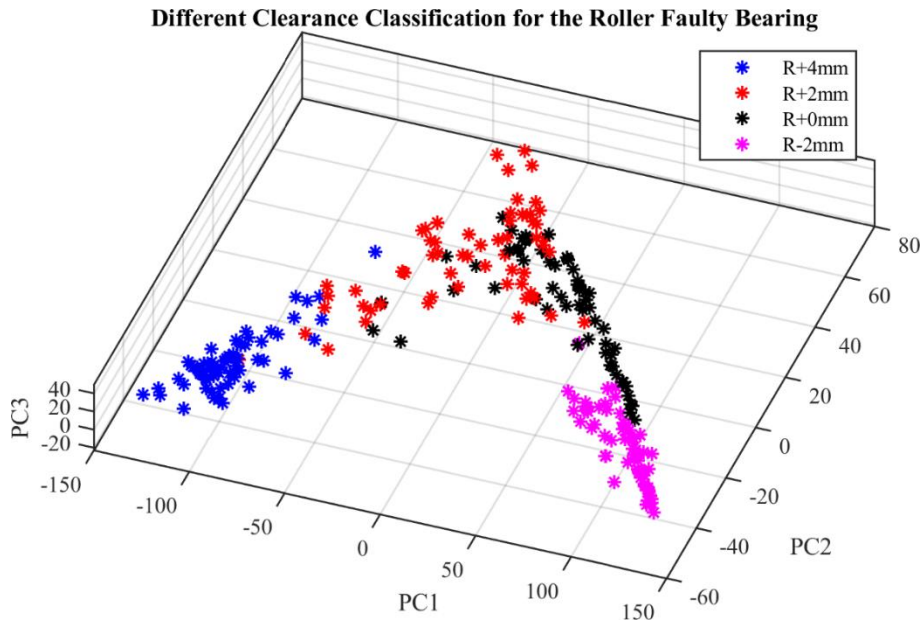


Figure 4.29 Scatter plot of different clearances of the roller fault bearing

#### 4.4 Summary

This chapter proposed a dimension reduction compression method for the fault diagnosis of tapered roller bearings. The time delay and dimension are the critical parameters in phase space trajectory, which are determined by the mutual information and FNN respectively. Then, twelve RQA indicators are extracted in the 2D RPs which are obtained by dimension reduction of phase space trajectories. The final step is to utilise the PCA to reduce the redundancy of the extracted features and classify the fault conditions. The results show that the bearings conditions can be successfully distinguished with only 189 data points whilst the raw vibration signal has a length of 1,500,000 samples.

In addition, the internal clearance of the tapered roller bearing affects the results of fault diagnosis, but the compressed features can help determine whether the bearing clearance is normal or not to avoid misjudging the severity of bearing defects. Only the classification result of the roller fault bearing is not very accurate because of complex periodic information, which mean the proposed compression method requires to be further improved for its wide applicability in bearing fault diagnosis.

---

## **CHAPTER 5    SPARSE REPRESENTATION BASED COMPRESSIVE SENSING IN FAULT DIAGNOSIS OF BALL BEARINGS**

---

*This chapter investigates a sparse representation based compressive sensing method for fault diagnosis of ball bearings with high efficiency during the long-term condition monitoring. The two-step adaptive compression method based on frequency shift and compressive sensing was developed to significantly reduce the data sets of vibration signals. The performance of the proposed compression method was examined by both simulated vibration signals and experimental data. Moreover, indicators of reconstruction performance were studied to explain the effectiveness of this sparse representation based compressive sensing method.*

## 5.1 Introduction

Ball bearings are a typical type of rolling element bearings and widely applied in machines with rotating components, such as water pumps, automotive industries, home appliances and aerospace. A ball bearing with the structure as shown in Figure 5.1, consisting of the outer race, cage, balls and inner race, has functions of rotational friction reduction and supporting radial and axial loads [149], [150]. However, the faults on the outer race, the cage, rollers or the inner race will result in failure of ball bearings, which will cost an enormous amount of maintenance expenditure because of inevitable damages of a machine during operation. Therefore, it is essential to monitor the condition of ball bearings in a rotating machine, especially when it works under a heavy load in a harsh environment in a long time.

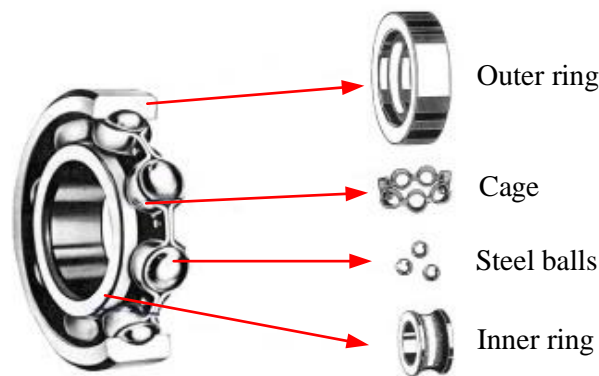


Figure 5.1 Structure of a ball bearing

Generally, according to the location where the fault occurs, the ball bearing component faults can be categorized as the outer race faults, cage faults, ball faults and inner race faults. The main causes of these faults include imperfections and irregularities in the manufacturing process, collisions during transportation, installation errors and wear and tear over long periods of use [150]. Many researchers have measured various types of signals to monitor the condition of ball bearings, like temperature [151], electrical characteristics [152], acoustic [153], [154], acoustic emission [155], [156] and vibration [157]–[160] signals. Vibration signals are commonly used ones because they are effective and reliable for bearing fault diagnosis and the transducers are relatively convenient to install. Once the localised fault occurs, a periodic vibration will be generated during the operation of the bearing. The vibration is closely related to the rotation speed and the ball number of the ball bearing, the size and depth of the defects as well. The relations between these typical fault frequencies and parameters of bearings have been introduced in the previous chapter.

However, the vibration response of the system is a modulation signal with the fault features in a low frequency range modulated by the carrier signal into a high-frequency band [161]. Therefore, according to the Nyquist-Shannon sampling theorem, the signal must be collected with a higher sampling frequency, then the fault signal can be effectively demodulated with high accuracy for fault identification of bearings. As a



result, it is necessary to reduce the hardware cost in the processes of data transfer, storage and processing through preventing large amounts of data generated during online condition monitoring with multiple sensors and high sampling frequency.

Compressive sensing is one of the most famous achievements in the field of signal processing in this century, and it has been effectively applied in the fields of MRI and image processing. It has the ability to reduce the amount of acquired data from the source according to the principle of non-uniform sampling, which is different from the Nyquist-Shannon sampling theorem with the uniform sampling. Besides, it has a noise reduction ability under the premise of meeting the conditions. Furthermore, when the original signal is sparse or compressible and the compression matrix satisfies the RIP protocol, the original signal can be reconstructed with or without loss [65]. In this chapter, a two-step adaptive compression algorithm is proposed to realise the fault detection of ball bearings with much fewer data according to the sparse representation based CS.

## **5.2 Compression Method**

The vibration signal acquired from the bearing system usually exhibits as typical modulation characteristics with a periodic pulse excitation signal modulated by the system resonance into a high-frequency band. Usually, the sampling frequency of the system response signal is up to tens of thousands of Hz or even more. In order to solve this problem, a two-step adaptive compression method based on frequency shift and CS is proposed to reduce the amount of data used for fault diagnosis of ball bearings because it has the potential to achieve data reduction from the data acquisition system and also has the ability of noise reduction. Additionally, some performance measurements will be used to evaluate the reconstructed signals which are reconstructed with the greedy compression recovery algorithm.

### **5.2.1 Methodology**

The proposed fault diagnosis method includes the following three steps: the frequency shift and envelope based compression step, the CS based compression step and the fault diagnosis step, the details of which is shown in Figure 5.2.

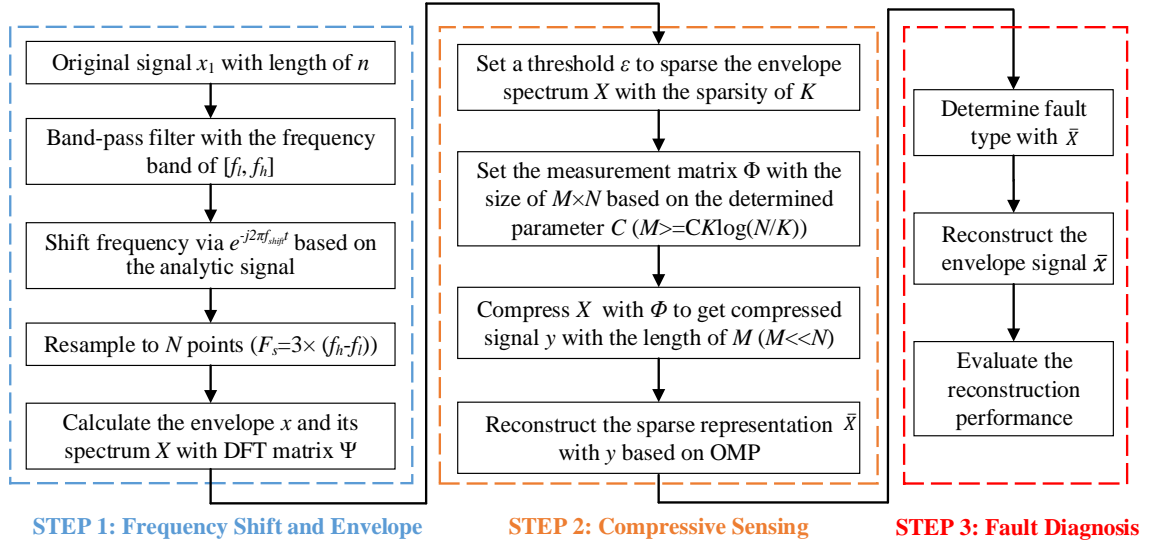


Figure 5.2 Flow chart of the proposed two-step compression based ball bearing fault detection method

The original signal  $x_1$  with the length of  $n$  is a modulated signal with the fault information is modulated to the high-frequency band. This optimal high frequency band is chosen as  $[f_l, f_h]$  to filter the original signal. The band-pass filtered vibration signal can be shifted from a high frequency band  $[f_l, f_h]$  to the lower one  $[f_l - f_{shift}, f_h - f_{shift}]$  with  $f_{shift}$  moving. Equations (5-1) and (5-2) show the process that signal  $f(t)$  shifts  $f_{shift}$  to the low frequency band in the frequency domain if it is multiplied by  $e^{-j2\pi f_{shift}t}$  in the time domain.

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(f) e^{j2\pi f t} df \leftrightarrow S(f) = \int_{-\infty}^{\infty} s(t) e^{-j2\pi f t} dt \quad (5-1)$$

$$\begin{aligned} g(t) = s(t) e^{-j2\pi f_{shift}t} &\leftrightarrow G(f) = \int_{-\infty}^{\infty} g(t) e^{-j2\pi f t} dt \\ &= \int_{-\infty}^{\infty} s(t) e^{-j2\pi f_{shift}t} e^{-j2\pi f t} dt \\ &= \int_{-\infty}^{\infty} s(t) e^{-j2\pi (f + f_{shift})t} dt \\ &= S(f + f_{shift}) \end{aligned} \quad (5-2)$$

However, the negative frequency component will affect the frequency shift result. Therefore, it is necessary to convert the signal into an analytic signal and then perform frequency shifting to generate an accurate frequency shift signal consisting of fault information.

In mathematics and signal processing, an analytic signal is a complex function with no negative frequency components [162]. Suppose that  $s(t), t \in \mathbf{R}$  is a real-valued signal, then the analytic signal  $z(t)$  can be defined as Equation (5-3):

$$z(t) = s(t) + j\hat{s}(t) \quad (5-3)$$

where  $j$  is the imaginary unit with  $j^2 = -1$  and  $\hat{s}(t)$  is the Hilbert transform of  $s(t)$ .

From the perspective of signal processing, Hilbert transform can be regarded as a signal through a linear time-invariant system with an impulse response of  $h(t) = \frac{1}{\pi t}$  as shown in Figure 5.3 (a), which can also be recorded as  $H[s(t)]$  and calculated as the following equation.

$$H[s(t)] = \hat{s}(t) = s(t) * \frac{1}{\pi t} \quad (5-4)$$

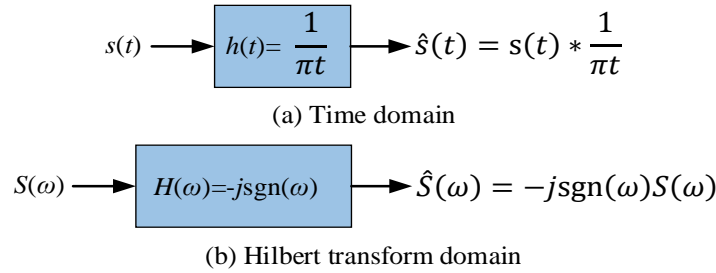


Figure 5.3 Hilbert transform principle from the perspective of signal processing

The frequency response of the impulse response  $h(t) = \frac{1}{\pi t}$  is usually used as a Hilbert filter as shown in Equation (5-5) and Figure 5.3 (b).

$$h(t) = \frac{1}{\pi t} \leftrightarrow H(\omega) = H\left(\frac{1}{\pi t}\right) = -j\text{sgn}(\omega) \quad (5-5)$$

where  $\text{sgn}(\omega)$  is a symbolic function and defined as:

$$\text{sgn}(\omega) = \begin{cases} 1, & \omega \geq 0 \\ -1, & \omega < 0 \end{cases} \quad (5-6)$$

As a result, the Fourier transform of  $\hat{s}(t)$  is shown in Equation (5-7) if the Fourier transform of  $s(t)$  is  $S(\omega)$ .

$$\hat{S}(\omega) = -j\text{sgn}(\omega)S(\omega) \quad (5-7)$$

Then, Fourier transform of  $z(t)$  can be expressed as follows:

$$\begin{aligned} Z(\omega) &= S(\omega) + j\hat{S}(\omega) \\ &= S(\omega) + j(-j\text{sgn}(\omega)S(\omega)) \\ &= S(\omega) + \text{sgn}(\omega)S(\omega) \\ &= 2S(\omega)u(\omega) \end{aligned} \quad (5-8)$$

where  $u(\omega)$  is a unit step function with  $u(\omega) = 1, \omega > 0$ . Therefore, the Fourier transform of an analytic signal has no negative frequency components. As a result, the frequency shift by the exponential function according to Equation (5-2) can be achieved on the analytic signals. After shifting frequency with an exponential function, the real part of the complex signal is the real-valued signal after the frequency shift.

Then, the vibration signal is resampled to get  $N$  points based on the bandwidth of  $(f_h - f_l)$ . The compressed signal is demodulated to get its envelope signal  $x$  and the envelope spectrum  $X$  with the DFT matrix  $\Psi$ . The fault frequency and its harmonics

can be recognised from signal  $X$  with the high noise level. This is the first compression step which compresses the original signal from the length of  $n$  to  $N$  (that  $N = 3 \times (f_h - f_l)$ ).

The second compression step is mainly achieved on the compressed envelope spectrum with the CS method. According to the requirements aforementioned in Chapter 2, the signal to be compressed with CS must be sparse or compressible. However, most of the practical vibration responses of mechanical systems are the non-sparse and non-compressible signals. Hence, they need to be converted to other sparse or compressible domains with sparse representation according to the Equation (2-3). In the first compression step, the obtained envelope spectrum is a compressible signal. In order to get a better compression ratio and performance, the envelope spectrum signal can be further sparse to obtain accurate sparsity of the signal via the investigated adaptive signal sparse method. Firstly, all amplitude peaks with values of  $A_p$  will be picked out and aligned in descending order. Then, a threshold value  $\varepsilon$  is determined according to  $\varepsilon = \kappa |\max(A_p) - \min(A_p)|$  [163], the threshold coefficient  $\kappa$  is set to 1%. After that, if there are five consecutive peaks satisfy Equation (5-9), then the first of them,  $A_{p_i}$  will be set as the peak threshold. Any peak with the value higher than the peak threshold will be reserved and other amplitudes of envelope spectrum will be set to zero to sparse the envelope spectrum signal with the sparsity of  $K$ .

$$|A_{p_i} - A_{p_{i+1}}| < \varepsilon, i = 1, 2, \dots, m - 1 \quad (5-9)$$

where  $m$  is the number of amplitude peaks. Then, the size of the measurement matrix can be determined to  $M \times N$  according to Equation (5-10).

$$M \geq cK \log(N/K) \quad (5-10)$$

where  $c$  is a constant [65]. The sparse envelope spectrum signal is compressed with a Gaussian random matrix  $\Phi$  into the observed signal  $y$  with the length of  $M$ . According to the RIP protocol, the sparse representation  $\hat{X}$  can be reconstructed via the optimization algorithm OMP once the observed signal  $y$  is available.

The last step is the fault diagnosis and reconstruction performance estimation. The reconstructed signal  $\hat{X}$  contains faulty signal components which are important to detect and diagnose the fault type of ball bearings. Moreover, the envelope signal  $\hat{x}$  can be reconstructed and denoised compared with the original envelope signal. The performance of reconstruction can be measured by some measurements introduced in the next subsection.

### 5.2.2 Reconstruction Performance Evaluation Indicators

The compression process and the reconstructed signal should be evaluated by some traditional statistical performance indicators, such as compression ratio, mean square error and percent root mean square difference [164]. Moreover, a more reasonable and

reliable method, fault feature signal-to-noise ratio (F-SNR) will be customized to evaluate the quality and efficiency of the reconstructed envelope spectrum for bearing fault diagnosis.

#### 5.2.2.1 Compression Ratio

Compression ratio ( $CR$ ) is the ratio of the length of the original signal to the length of the compressed signal, which can be calculated as shown in Equation (5-11).

$$CR = \frac{N}{M} \quad (5-11)$$

where  $N$  is the length of the original signal and  $M$  is the length of the compressed signal. For the same set of data, the larger the value of  $CR$ , the less the amount of data needs to be stored after compression.

#### 5.2.2.2 Root Mean Square Error

Root mean square error ( $RMSE$ ) indicating the difference between the reconstructed and the original signals is defined by

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\bar{x}_i - x_i)^2}{N}} \quad (5-12)$$

where,  $\bar{x}_i$  is the  $i$ th element of the reconstructed signal and  $x_i$  is the  $i$ th element of the original signal of length  $N$ . The higher the  $RMSE$  means the worse performance of signal reconstruction with CS.

#### 5.2.2.3 Percent Root Mean Square Difference

Percent root mean square difference ( $PRD$ ) is mainly used to measure the distortion of the reconstructed signal compared with the original signal. It is given as shown in Equation (5-13).

$$PRD = \sqrt{\frac{\sum_{i=1}^N (\bar{x}_i - x_i)^2}{\sum_{i=1}^N (x_i)^2}} \times 100 \quad (5-13)$$

#### 5.2.2.4 Fault Feature Signal-to-noise Ratio

$F$ -SNR is a newly developed indicator to evaluate the performance of signals based on the fault features. The fault frequency and its harmonics are picked out as the representative of the fault signals, and other remaining signals are treated as noise.  $F$ -SNR is expressed as Equation (5-14).

$$F\text{-SNR} = 20\log_{10} \frac{A_{\text{signal}}}{A_{\text{noise}}} = 20\log_{10} \frac{A_{\text{fault}}}{A_{\text{other}}} \quad (5-14)$$

where  $A_{\text{fault}}$  is the mean of those amplitudes of fault frequency and its harmonics;  $A_{\text{other}}$  is the mean of amplitudes of the remaining signals. It can be seen from the definition of this indicator that it is more reasonable and reliable to evaluate the performance of the signal used for fault detection.

## 5.3 Simulation Study

### 5.3.1 Model Simulation

To validate the proposed two-step compression algorithm for fault diagnosis of ball bearings, the vibration signal of a rolling element bearing system with local defect is simulated as expressed in Equation (5-15) [165], [166].

$$x_1(t) = h(t) * u(t) + n(t) \quad (5-15)$$

where,

$h(t)$  is the response of the system indicating the system resonant behaviours simulated as shown in Equation (5-16);

$u(t)$  is the periodic impulses produced by the elements passing through the defects;

$n(t)$  is the Gaussian white noise signal generated by the complicated environments, data acquisition systems and complex vibration signal transmission path.

$$h(t_i) = e^{-\tau 2\pi f_c t_i} \cos(2\pi f_c t_i) \quad (5-16)$$

where,

$\tau$  is the oscillation damping factor;

$f_c$  is one frequency of the system response which is also regarded as the carrier frequency;

$t_i$  is the time period of system response to an impulse excitation, which means it is the same as the period of the fault impulse,  $\frac{1}{f_o}$ , where  $f_o$  is the fault frequency.

Some key parameters of the model simulation are listed in Table 5.1. The sampling frequency,  $F_s$  and the natural frequency of the system  $f_c$  are set to be 25,600 Hz and 2500 Hz, respectively. The outer race fault is simulated with the fault frequency of 89.80 Hz. The simulated data length is 5 seconds to ensure the number of impulses is enough to operate. Part of the simulated impulse signal and the system resonance waveform are both illustrated in Figure 5.4 to express these two signals more clearly.

Table 5.1 Key parameters for simulation

Parameters	Value
Sampling Frequency, $F_s$	25,600 Hz
Natural Frequency, $f_c$	2500 Hz
Fault Frequency, $f_o$	89.80 Hz
Data Length, $t$	5 seconds

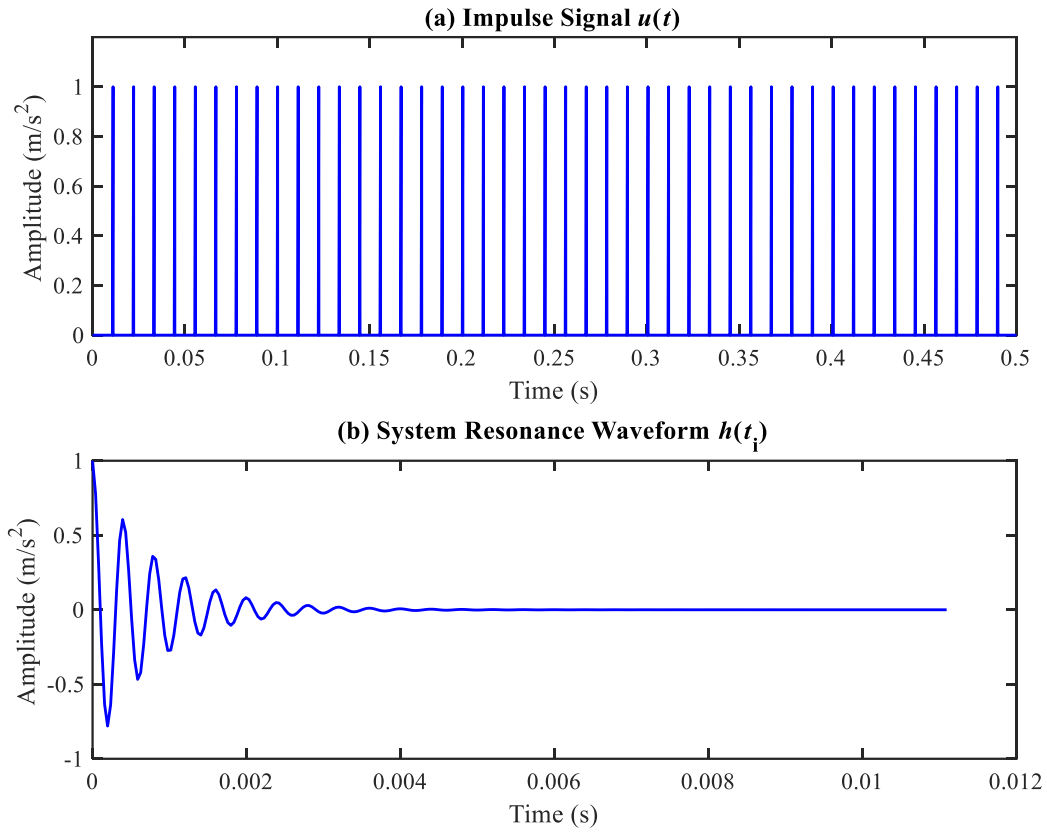


Figure 5.4 Simulated impulse signal and system resonance

Therefore, the signal waveform of  $h(t) * u(t)$  as the output of the ideal system without any noise is simulated through the convolution, which is plotted with the red colour in Figure 5.5 (a). It is a very clear periodic impulse signal. However, it is impossible that a realistic mechanical system signal has no noise because it is always affected by the ambient environment and the passing path. As a result, the Gaussian white noise  $n(t)$  is added to this clean signal waveform to generate a signal  $x_1(t)$  with the SNR of -15 dB, which is illustrated with the blue line and labelled as “With noise”. It is significantly different between the signals with noise and without noise. In particular, periodic features cannot be visually observed.

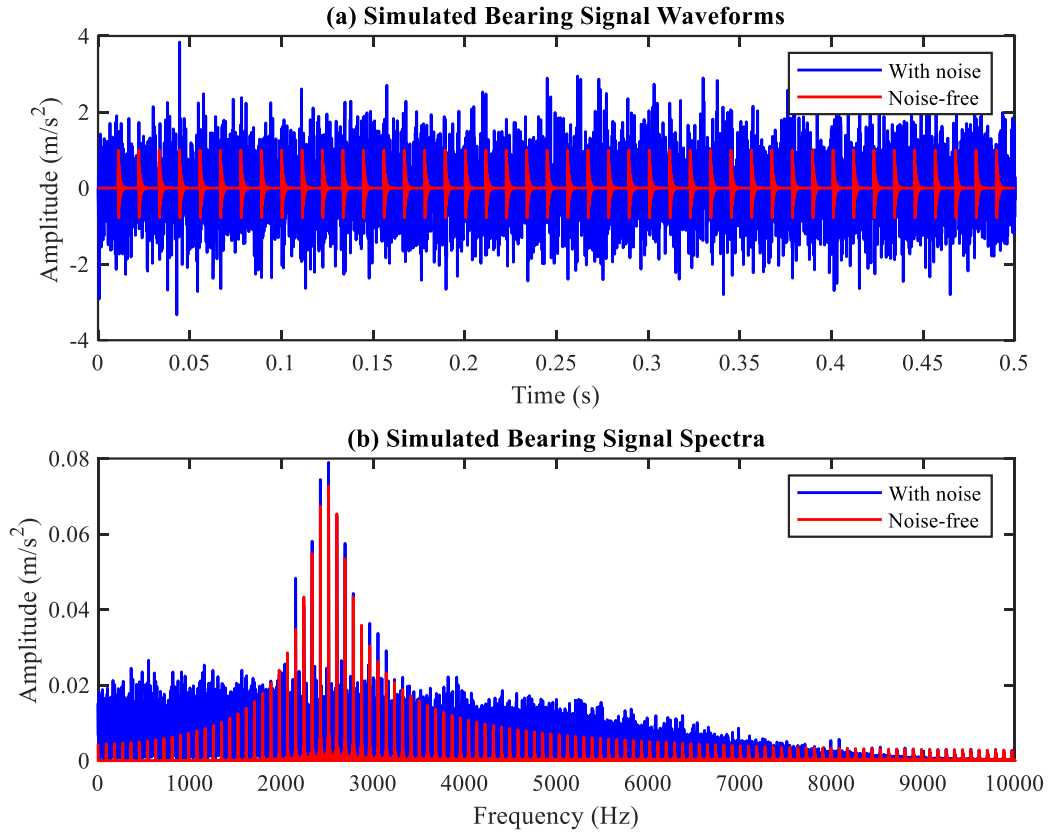


Figure 5.5 Simulated bearing signals with and without noise

Moreover, the spectra of these two signals are also expressed in Figure 5.5 (b). It is clear that the energy and the fault information is mainly concentrated in the frequency band of [2000 Hz, 3000 Hz] because the carrier frequency is set to 2500 Hz in this simulated signal. This simulated signal with the outer race fault will be applied to investigate the proposed two-step compression method for the fault diagnosis.

### 5.3.2 Results and Discussion

#### 5.3.2.1 Step One: Frequency Shift and Envelope

In order to reduce the sampling frequency, the signal in the high-frequency band of [2000, 3000] Hz is converted into the analytic signal according to Equations (5-3) and (5-4). Then, the analytic signal can be shifted to a lower frequency band of [0, 1000] Hz with the exponential function  $e^{-j2\pi f_{shift}t}$  (where  $f_{shift} = 2000$  Hz) according to the Equation (5-2). Now, the fault information has been shifted to the frequency band of [0, 1000] Hz, so that it is not necessary to sample the signal with a very high sample rate. Then, the real part of the shifted signal can be selected to resample with the sample frequency of only 3000 Hz based on  $F_{s0} = 3 \times (f_h - f_l)$  meeting the requirements of Nyquist-Shannon sampling theorem, where  $f_h = 3000$  Hz and  $f_l = 2000$  Hz in this case. The envelope of this resampled signal, regarded as  $x(t)$  is the effective information related to the fault.



This is the first compression step of the proposed method which compressed the signal  $x_1(t)$  from the length of  $t \times F_s$  ( $5 \times 25600$ ) to  $t \times F_{s0}$  ( $5 \times 3000$ ) to get the envelope signal  $x(t)$  with the compression ratio up to 8.5. The spectra of the signal after the frequency shift and downsampling and its envelope spectrum written as  $X(f)$  are illustrated in Figure 5.6. The outer race fault frequency  $f_{outer}$  and its harmonics are quite clear in its envelope spectra.

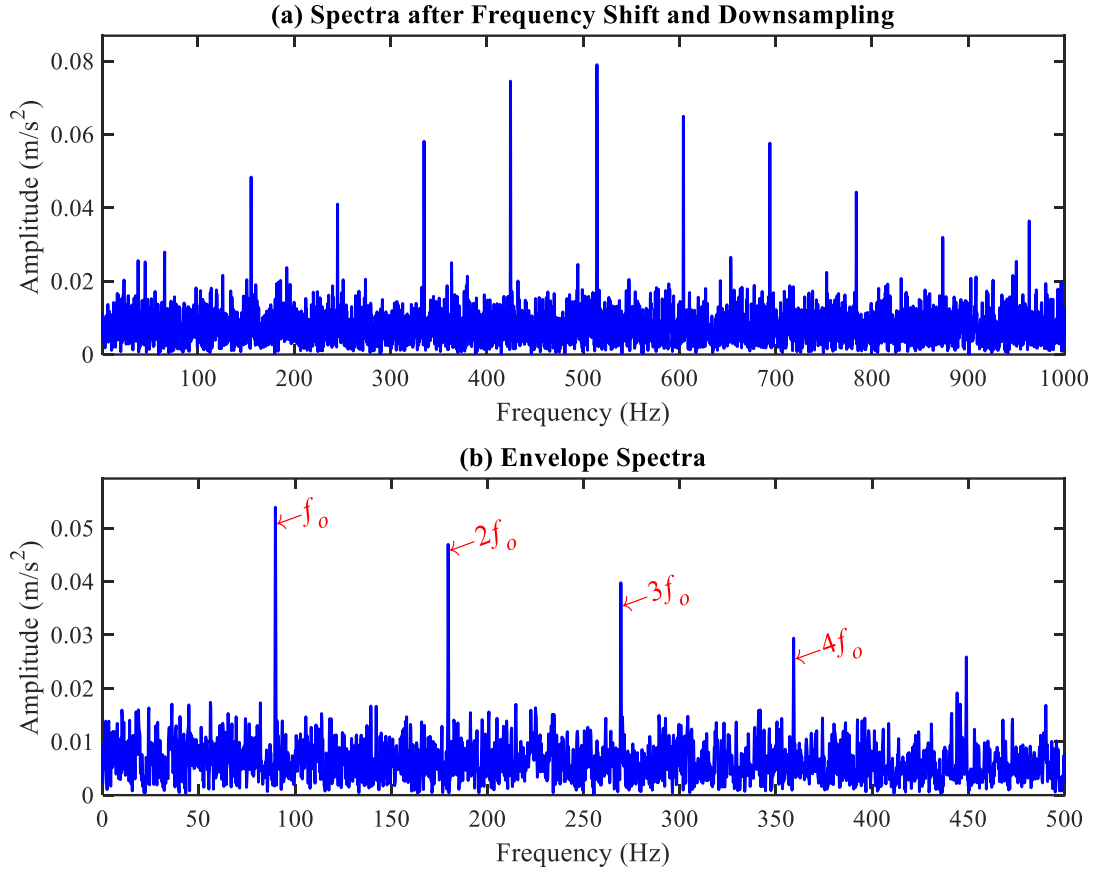


Figure 5.6 Spectra of the signal before and after envelope

### 5.3.2.2 Step Two and Three: Compressive Sensing and Fault Diagnosis

Based on the result of the first compression step, the number of data can be further reduced by CS, which has the potential to achieve in data acquisition system in future. Moreover, CS has the abilities to reduce the noise and reconstruct the signal before compression with the tolerable distortion.

In order to further compress the envelope signal  $x(t)$  with CS, it is necessary to sparse it according to the requirements of the CS theory. Fortunately,  $x(t)$  is compressible in the frequency domain, so that the spectrum  $X(f)$  is obtained to prepare for compression. However, the signal  $X(f)$  is not sparse, which makes it difficult to determine some parameters in the CS algorithm implementation process. According to the method described in Equation (5-9), the sparsity of the envelope spectrum can be obtained by the proposed adaptive signal sparse method. The sparsity is set to  $K = 16$  for this envelope spectrum  $X(f)$  after sparse. Furthermore, the parameter  $c$  in Equation (5-10)

should be confirmed for further determination of the measurement matrix. In order to confirm the value of parameter  $c$  used in Equation (5-10),  $c$  is set from 0.4 to 16 with an interval of 0.1. The relation between  $PRD$  and  $c$  is investigated in Figure 5.7. The  $PRD$  value dramatically decreases at first and then keep constant when  $c > 2$ . It demonstrates that setting the value of  $c$  to 2 is the best choice to trade off the compression performance and computing time.

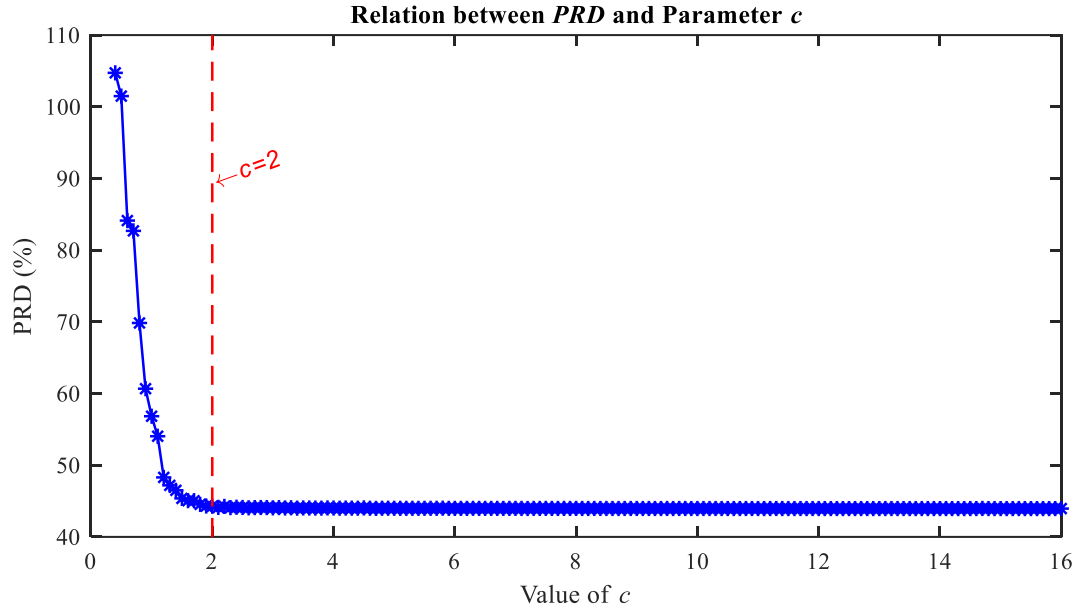


Figure 5.7 Relation between  $PRD$  and parameter  $c$

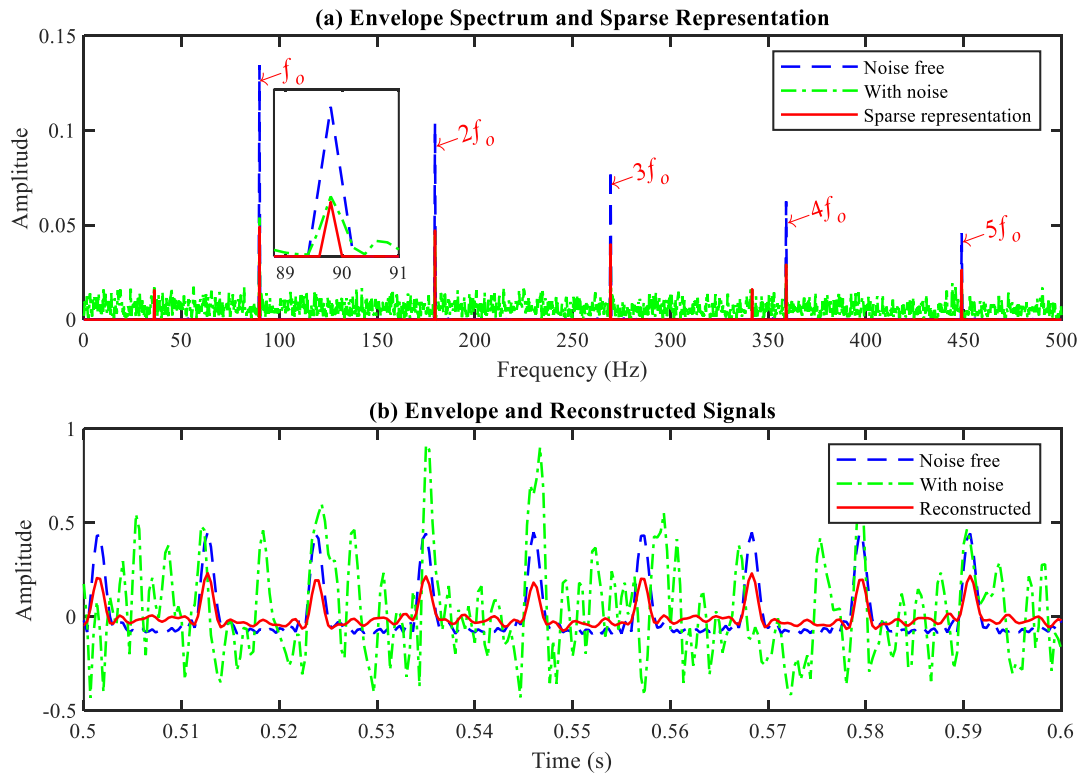


Figure 5.8 Comparison of original envelope and reconstruction at  $SNR = -15$  dB

According to Equation (5-10),  $M$  can be set to the minimum integer greater than  $cK\log(N/K)$ , which should be  $M = 96$  here. Hence, the measurement matrix  $\Phi$  can be set to the size of  $M \times N = 96 \times 15000$  to get the compressed envelope spectrum as the observation  $y$  with the size of  $M \times 1$ . The sparse representation of the envelope spectrum  $X(f)$  can be calculated via the optimization algorithm OMP. To eliminate the effects of randomness and validate the reliability and robustness of the proposed algorithm, the simulation was carried out 100 times with the average result shown in Figure 5.8.

The envelope spectrum of the signal without noise expresses the outer race fault frequency and its harmonics clearly with the blue line. While the traditional envelope spectrum of the signal with noise (SNR = -15 dB) has the broadband noise even the fault frequency and harmonics are evident to detect the fault. Compared to the traditional envelope spectrum plotted with the green colour, the recovered sparse representation of the envelope signal  $x(t)$  is more effective and efficient to detect the faults of ball bearings according to the fault frequency and its harmonics with much less noise as shown in Figure 5.8. It means that only  $M$  values are transmitted and stored for the fault diagnosis of ball bearings, which saves a lot of resources and cost in industrial applications. Moreover, this adaptive compression step with CS has the function of noise elimination result in a lower level noise of the sparse representation signal for the better performance of fault detection.

The last function of this method is to reconstruct the envelope signal according to Equation (2-6) with the DFT matrix. The reconstructed envelope signal is illustrated with the red line in Figure 5.8. Compared with the original envelope signal with noise in the green colour, the reconstructed envelope signal is an obvious periodic signal. Its periodicity almost overlaps with the envelope signal without noise plotted with the blue colour. This phenomenon further illustrates that CS can eliminate noise. At the same time, the reduced amplitude of the reconstructed envelope signal also indicates that CS has an inevitable loss in the process of signal reconstruction. However, the loss is tolerable because it does not affect the performance of fault diagnosis and reduce the impact of the noise. This is one of the important reasons for applying CS to ball bearing fault diagnosis because fewer points can help reconstruct the envelope signal with little loss. Figure 5.9 illustrates another two results to verify the effectiveness and reliability of the proposed method when the original signals are at the level of SNR = 0 dB and -10 dB, respectively.

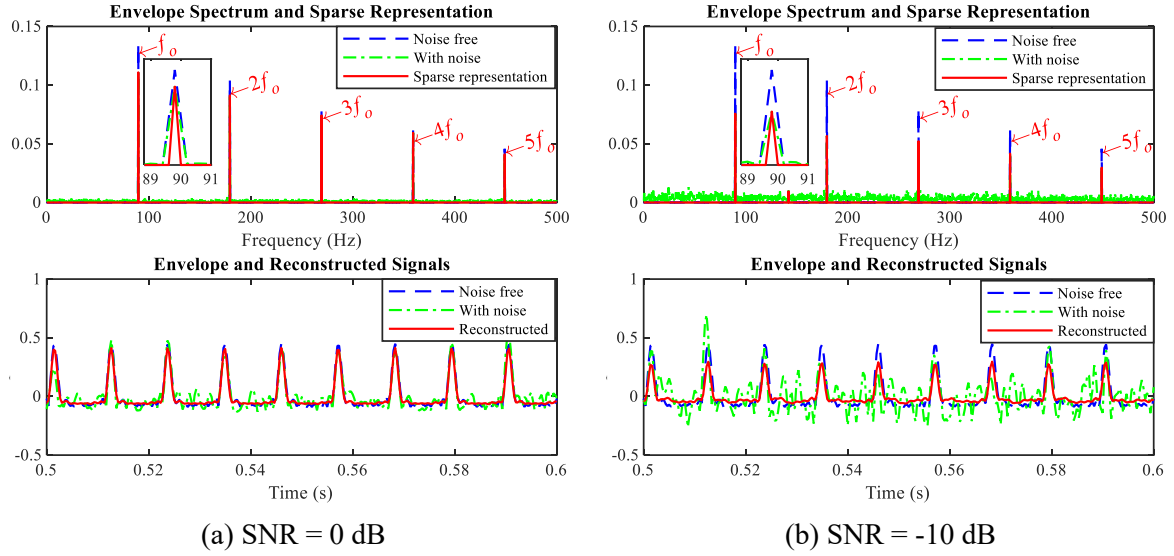


Figure 5.9 Comparison of original envelope and reconstruction at: (a) SNR = 0 dB, (b) SNR = -10 dB

### 5.3.2.3 Performance Evaluation

To evaluate the performance of the proposed two-step compression method in ball bearing fault diagnosis, the main indicators are used for different signals with various noise level. In the first compression step, the amount of data is compressed from 128,000 to 15,000 with a compression ratio of about 8.53 according to the frequency shift, downsampling and envelope application. In the second compression step, CS plays a major role in reducing the number of sample points. It further compresses the amount of data from 15,000 to only 96 with a compression ratio of up to 156.25. But the second compression step is adaptive because the length of the observation is related to the signal sparsity. In total, the amount of data is compressed from 128,000 to 96 by combining these two compression steps, which achieves significantly considerable data reduction with the compression ratio of more than 1333 times.

Table 5.2 Main indicator values compared with the signal with noise

Indicators		Original Signal SNR			
		0 dB	-5 dB	-10 dB	-15 dB
$K$		16	16	16	16
$M$		96	96	96	96
$CR$	Step 1	8.53	8.53	8.53	8.53
	Step 2	156.25	156.25	156.25	156.25
	2 Steps	1333.3	1333.3	1333.3	1333.3
$RMSE$ (m/s <sup>2</sup> )	Compared with noisy signal	0.05	0.08	0.14	0.24
$PRD$ (%)		35.31	58.66	85.03	96.44
$RMSE$ (m/s <sup>2</sup> )		0.03	0.04	0.06	0.09

Indicators		Original Signal SNR			
		0 dB	-5 dB	-10 dB	-15 dB
<i>PRD</i> (%)	Compared with noise-free signal	17.36	27.22	43.00	59.08
<i>F-SNR</i> (dB)	Original envelope spectrum	30.47	24.91	17.86	11.39
	Reconstructed envelope spectrum	+Inf	58.65	52.36	43.04

Table 5.2 listed the performance indicators of signals at different noise levels in the second compression step. The sparsity  $K$  depends on the number of fault frequency and harmonic components in the envelope spectra because the fault frequency and harmonic components will be picked out in the process of finding the peaks meeting the requirements of Equation (5-9). The number of observations or measurements  $M$  is determined according to Equation (5-10). It is evident that both  $K$  and  $M$  are the same for signals with different levels of noise because they have the same fault signals and the fault characteristics are obvious after sparse processing. Referring to Equation (5-11), the compression ratio  $CR$  can be calculated for each compression step and the two-step combined as listed in Table 5.5.

Additionally, the reconstructed envelope signal can be evaluated from the time domain and the frequency domain. In the time domain, the reconstructed envelope signal is compared with the envelope of the original noisy signal and noise-free signal, respectively. Both  $RMSE$  and  $PRD$  values increase as the original signal noise level decreases because the impact of noise on the fault signal is larger and larger. However, the reconstructed envelope signal is much closer to the envelope of the noise-free signal, which demonstrates that the CS method has the ability to eliminate the noise. In the frequency domain, the new developed indicator  $F-SNR$  is used on the original envelope spectra and the reconstructed envelope spectra. As shown in Figure 5.10, all points selected for the  $F-SNR$  calculation are displayed manifestly. The original and reconstructed envelope spectra for signals with the SNR of 0dB, -5dB, -10dB and -15dB are presented in Figure 5.10 (a), (b), (c) and (d), respectively. The blue line and the red dot line separately plot the original signals before selecting fault features and the remaining signals, which means the difference between the signals drawn by the blue and red lines represents the selected fault features.

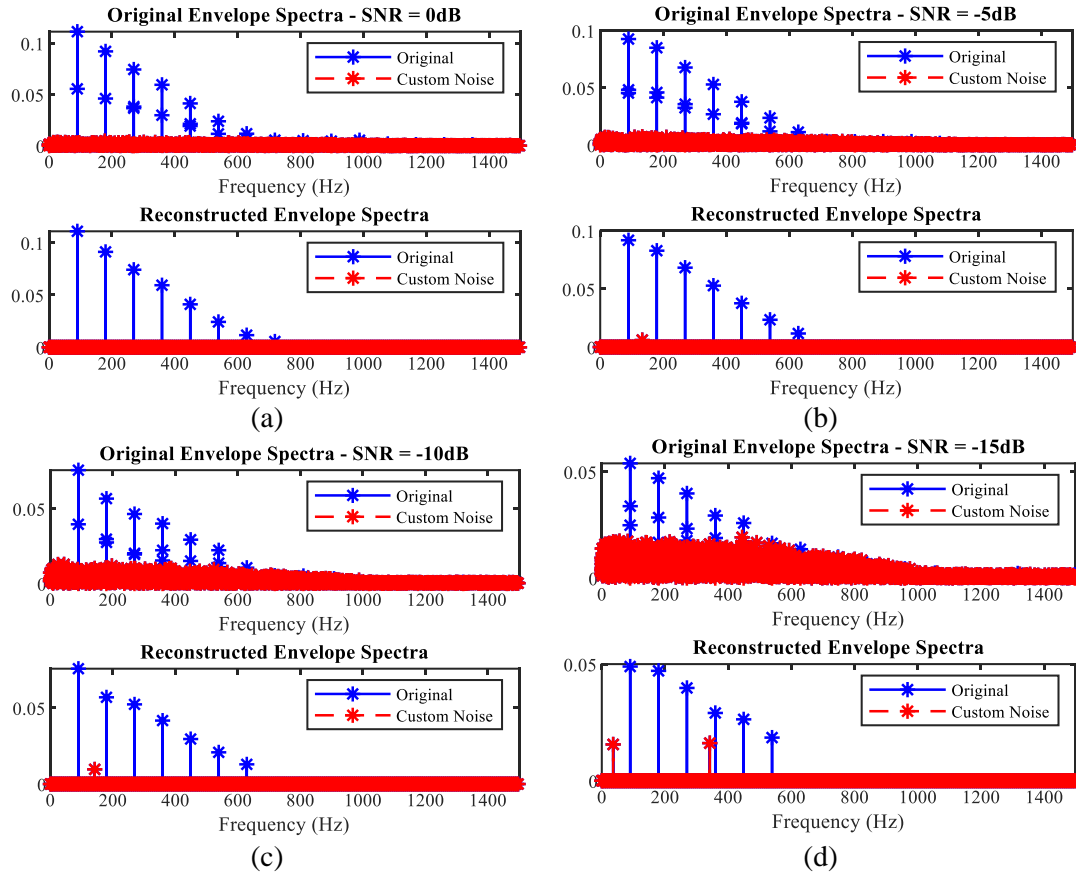


Figure 5.10 Fault features selected for  $F$ -SNR calculation: (a) SNR = 0 dB, (b) SNR = -5 dB, (c) SNR = -10 dB, and (d) SNR = -15 dB

According to Equation (5-14), the mean of amplitudes of the fault features and the remaining signals are  $A_{fault}$  and  $A_{other}$ , respectively. From the results of  $F$ -SNR, the fault components of the reconstructed envelope spectra are more abundant than the noise information. But the noise in the original envelope spectrum is significantly larger than that of the reconstructed envelope spectrum. Therefore, the proposed two-step compression method can be demonstrated to be effective and efficient with the simulated ball bearing signal. It not only reduces the amount of data with the compression ratio of more than 1500 times but also reconstructs the envelope signal with noise elimination.

## 5.4 Experiments

For the validation of the proposed two-step adaptive compression method, the simulation study was performed with the satisfactory results in both compression ratio and fault diagnosis. Then, it needs to be further verified by the experimental data because the mechanical system and the ambient environment may change during the actual application process and also the system response is much more complicated than the simulation. Therefore, an outer race failure test was carried out on a ball bearing test rig with the data acquired to verify the effectiveness of the proposed compression method.

### 5.4.1 Experimental Setup

#### 5.4.1.1 Test Facilities

As described in Figure 5.11, the ball bearing test rig mainly consists of an Optima controller, an induction motor, supporting bearings, flexible couplings, the shaft and a DC generator. The Optima controller supplies 3-phases electricity to the 4kW induction motor and in the meantime the controller is also connected with the DC generator to control the load. The specification of the motor is the same as the specification listed in Table 4.2. Besides, considering the position of the faulty ball bearing, a piezoelectric accelerometer with the model number of CA-YD-185 was installed in the vertical direction on the top of the motor drive end bearing housing. In addition, a Hengstler incremental encoder was installed on the fan end of the motor, connected by an elastic pipe. Their specifications are listed in Table 5.3. Moreover, a 16-channel data acquisition system from Sinocera Piezotronics, Inc. was used to collect data with the sampling frequency of 96 kHz at 24-bit resolution.

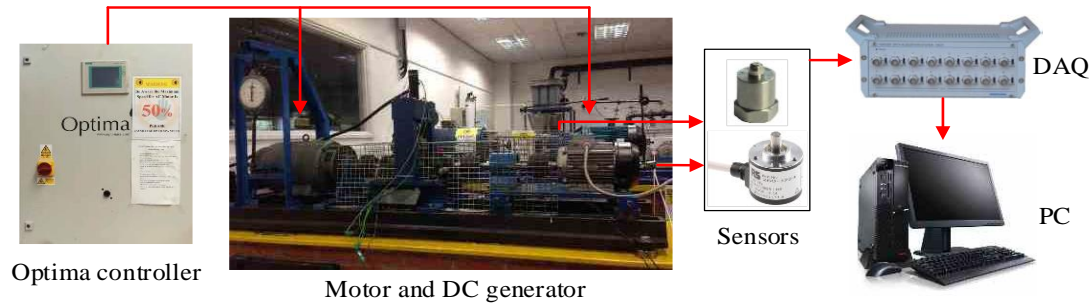


Figure 5.11 Schematic diagram of the ball bearing test rig

Table 5.3 Specification of sensors

Sensors	Parameters	Values
Accelerometer (CA-YD-185)	Voltage Sensitivity	5.106 mV/ms <sup>-2</sup>
	Power Supply	12~24 VDC
	Frequency Range	0.5~5000 Hz
	Accelerometer Limit	1000 m/s <sup>2</sup>
	Temperature Range	-40~120 °C
	Excitation Current	2~10 mA
Hengstler Incremental Encoder	Number of Pulses	100
	Absolute Maximum Speed	Max. 6000 rpm
	Supply Voltage	5 VDC±10%
	Operating Temperature	-10~60 °C
	Vibration Proof	100 m/s <sup>2</sup> (10~2000 Hz)
	Shock Resistance	1000 m/s <sup>2</sup> (6 ms)

#### 5.4.1.2 Defect Seeding

To simulate the failure occurred on the ball bearing, different sizes of the defects were created artificially on the ball bearings. The tested ball bearing is the type of 6206ZZ from NSK with the geometry enumerated in Table 5.4. Seven identical ball bearings were prepared, one of which was a healthy one and the other six were seeded small, medium and large defects on the outer and inner races, respectively. These six ball bearings with defects are illustrated in Figure 5.12.

Table 5.4 Ball bearing specification

Parameters	Values
Pitch Diameter, $D_p$	46.4 mm
Roller Diameter, $D_r$	9.53 mm
Shaft Rotating Frequency, $f_s$	24.90 Hz
Roller Number, $N_r$	9
Contact Angle, $\varphi$	0
Outer Race Fault Frequency, $f_o$	88.86 Hz
Inner Race Fault Frequency, $f_i$	134.80 Hz



(a)



(b)

Figure 5.12 The bearings with defects: (a) outer race faults, and (b) inner race faults

The theoretical outer race fault frequency  $f_o$  and inner race fault frequency  $f_i$  can be calculated according to the Equation (4-4). When the test rig operates at the speed of



24.90 Hz,  $f_o$  and  $f_i$  are 88.86 Hz and 134.80 Hz, respectively, which are the particular frequencies to use for the identification of faults and shown in Table 5.4.

#### 5.4.1.3 Test Procedure

In this test, the system was operated at the full constant speed of the motor and 100% load. The shaft rotating frequency is 24.9 Hz. Both the vibration signal and encoder signal were collected with the sampling frequency of 96 kHz and recorded for 5 seconds with three sets. All tests are divided into three groups: the first one is the healthy case regarded as the baseline; the second one is the outer race faults with the large, medium and small defects tested, respectively; the third one is the inner race faults also with the large, medium and small defects tested, separately. All datasets were saved as the .mat file format.

### 5.4.2 Results and Discussion

To validate the effectiveness the proposed two-step compression method on actual data, two groups of tests were conducted in this research: one for outer race faults with different sizes and the other for inner race faults. The cases of healthy, with large, medium and small defects are abbreviated as ‘Healthy’, ‘Large’, ‘Medium’ and ‘Small’, respectively. The results will be separately discussed from outer race faults and inner race faults.

#### 5.4.2.1 Ball Bearings with Outer Race Faults

In this test, the datasets were recorded for 5 seconds with sampling frequency of 96 kHz for these four different sizes of outer race faults. The time waveforms of these four cases are illustrated with different colours in Figure 5.13 (a). The amplitudes dramatically reduces as the size of the faults decreases, which also can be distinguished with the RMS values. Besides, their spectra are also clearly described in Figure 5.13 (b). The frequency band of [2000, 3000] Hz with large vibration amplitudes is better to use for the processing of fault diagnosis because the fault signal is modulated into this band.

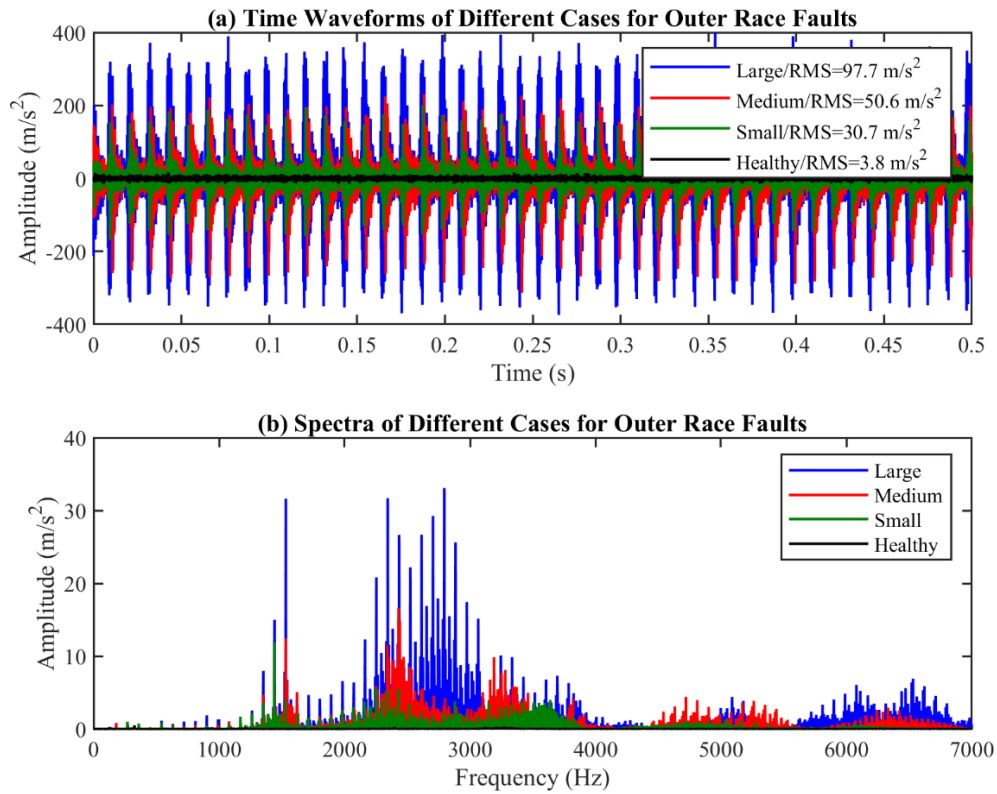


Figure 5.13 Original signals of bearings with outer race faults: (a) time waveforms, and (b) spectra

As introduced from Equations (5-2) to (5-8), the frequency in this special band can be shifted to a lower band of [0 1000] Hz to reduce the sampling frequency by analytic function. After the successful shifting with an exponential function, the sampling frequency can be reduced to 3000 Hz because the valid frequency bandwidth is only 1000 Hz. The spectra of the shifted and down sampled signal are illustrated in Figure 5.14 (a). As a result, the number of sampling points decreases from 480,000 to 15,000 with the compression ratio of 32 in the first compression step. Then, envelope, as the demodulated method, is applied to this shifted signal and its spectrum are expressed in Figure 5.14 (b). The fault frequency around 90.2 Hz and its four harmonics are significantly explicit that these bearings have the outer race faults with different sizes of the defects. The larger the size of the defect, the bigger the magnitude of the fault frequency, which is expressly described in the zoom window in Figure 5.14 (b). The values of the fault frequency for three fault cases are not exactly the same due to the manufacturing error and speed control, so that fault frequency has a minor shift when the bearing has a large fault.

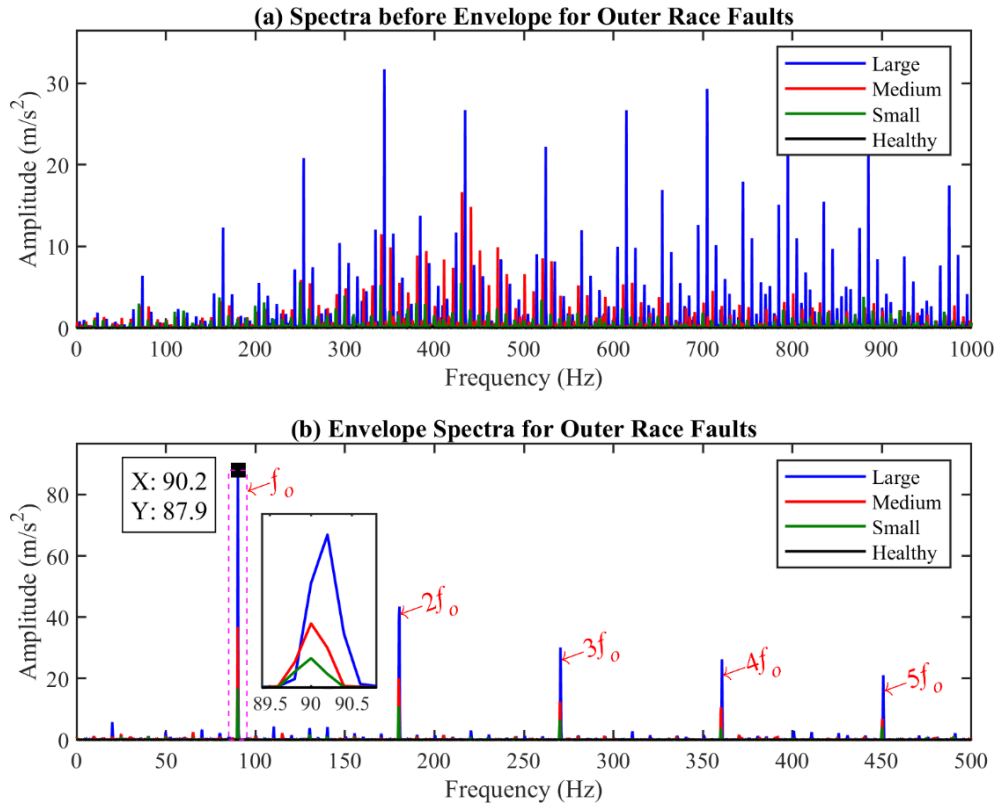


Figure 5.14 Spectra of bearings with outer race faults: (a) before envelope, and (b) envelope spectra

The second compression step is to use a sparse representation based CS method to compress further the data of the signals. The envelope spectra are converted into the sparse signals by the proposed adaptive sparse method defined in Equation (5-9). The sparsity of the four cases of health, large, medium and small faults are 36, 14, 18 and 22, respectively, as shown in Table 5.5. When the defect size is small, the fault characteristics are easily submerged by noise, which is not conducive to the fault feature extraction. Therefore, the sparsity increases as the size of the fault decrease. On the other hand, the sparsity determines the number of observations, which means the bearing with a small defect or early fault requires more data to detect. After the second compression step, the number of observations and compression ratios for these four cases are listed in Table 5.5.

Moreover, the original envelope spectra and the sparse representations of envelope signals recovered by the OMP method are compared with the results depicted in Figure 5.15. It is obvious that the vibration amplitude of healthy is much smaller than others with faults. Also, there is random noise without any outer race fault frequency. However, the outer race fault features, including the fault frequency and its harmonics, can effectively and efficiently state that the bearing has the outer race fault. Furthermore, according to the magnitudes of the fault features, the defect size can be apparently distinguished. For the reconstructed sparse representations of the envelope spectra, they have minor losses compared with the original ones. But they have an outstanding ability to eliminate noise.

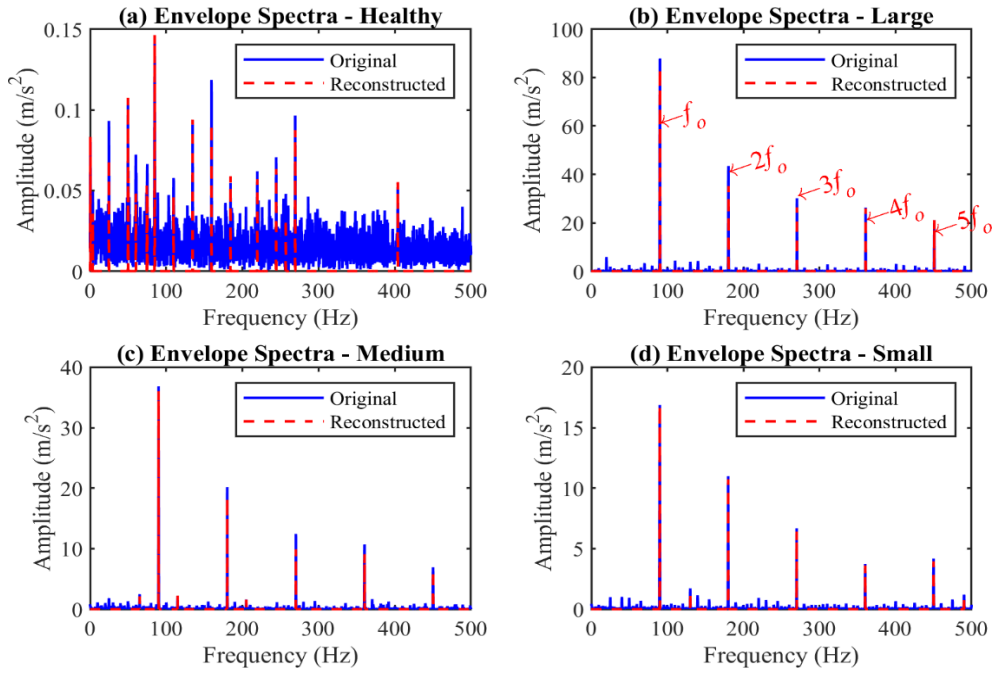


Figure 5.15 Original and reconstructed envelope spectra of bearings: (a) healthy, (b) large outer race fault, (c) medium outer race fault, and (d) small outer race fault

Besides, the compressed signals with CS have a characteristic that they can be reconstructed with a tiny loss or even without loss if the signal is sparse enough. Figure 5.16 describes the reconstruction effect of the envelope signals compared with the original ones. The envelope signal of the healthy bearing is not sparse so that its reconstructed envelope timewave has the smaller magnitudes than the original one. But the faulty cases have excellent reconstruction results, especially for the bearing with a large outer race fault, which can be evaluated with the performance indicators.

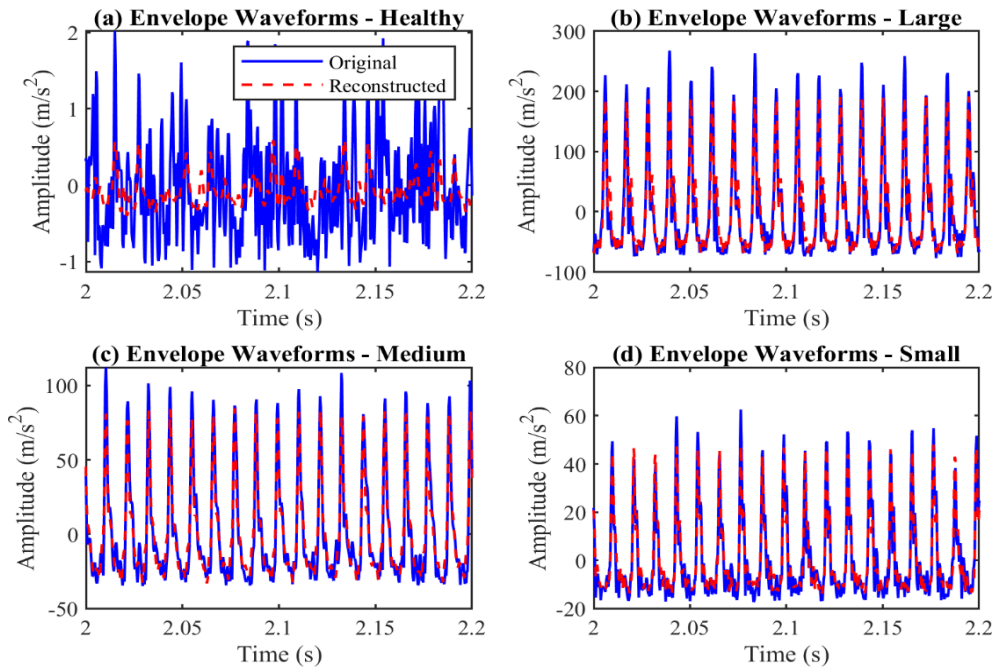


Figure 5.16 Original and reconstructed envelope signals of bearings: (a) healthy, (b) large outer race fault, (c) medium outer race fault, and (d) small outer race fault

Table 5.5 described some primary indicators of evaluating the performance of the proposed two-step compression method for bearings with outer race faults. As with the processing of the simulated signal, the envelope spectra are sparse after the first compression step. The sparsity  $K$  becomes larger as the fault characteristics decrease. The number of observation points has the same trend as the sparsity.  $CR$  in the first compression step is primarily related to the selected band bandwidth. It is closely related to the sparsity of the compressed signal. The values of  $CR$  are these two compression steps and two steps combined for these four different cases are comprehensively compared in Table 5.5.  $CR$  can reach more than 5,000 for the signal of the bearing with a large fault because of the apparent fault characteristics. Although  $RMSE$  is used to evaluate the difference between the original and reconstructed signals, the large  $RMSE$  value does not indicate that the reconstruction effect is poor since the signal amplitude of the bearing with a larger fault is much larger than the smaller ones. In contract,  $PRD$  is a better indicator because it describes the distortion of the reconstructed signal relative to the original one. However, since the proposed algorithm has a good noise reduction effect, the distortion of the reconstructed signal has not a strong relationship with the defect size.

The proposed performance evaluation indicator,  $F$ -SNR, aims to measure the strength of the effective fault features compared to the undesired signals. As shown in Figure 5.17, the blue points are the selected fault features and the red points are the remaining undesired signals. In terms of the  $F$ -SNR values, the faulty bearing signals have apparent fault features. Additionally, the  $F$ -SNR values decreases as the size of the defect reduces. Moreover, the reconstructed envelope spectra have much more pronounced fault features compared with the original ones, which explains that the proposed compression method has noticeable capability of noise reduction.

Table 5.5 Main indicator values for outer race faults

Indicators		Defect Size			
		Healthy	Large	Medium	Small
$K$		36	14	18	22
$M$		189	85	106	125
$CR$	Step 1	32	32	32	32
	Step 2	79.37	176.47	141.51	120.00
	2 Steps	2539.7	5647.1	4528.3	3840.0
$RMSE$ (m/s <sup>2</sup> )		0.62	41.59	16.30	6.20
$PRD$ (%)		94.01	50.32	46.87	36.64
	Original envelope spectrum	-	35.61	33.92	31.96

Indicators		Defect Size			
		Healthy	Large	Medium	Small
$F$ -SNR (dB)	Reconstructed envelope spectrum	-	+Inf	52.89	48.81

To sum up, it is demonstrated that the proposed two-step compression method based on the sparse representation has significant ability to compress a large amount of data with  $CR$  of up to thousands of times. Additionally, the proposed method can eliminate noise to preserve the fault characteristics within the bearing signals from different defect cases. Furthermore, the newly developed performance indicator,  $F$ -SNR, is more conducive to measuring the effectiveness of the signal for fault diagnosis.

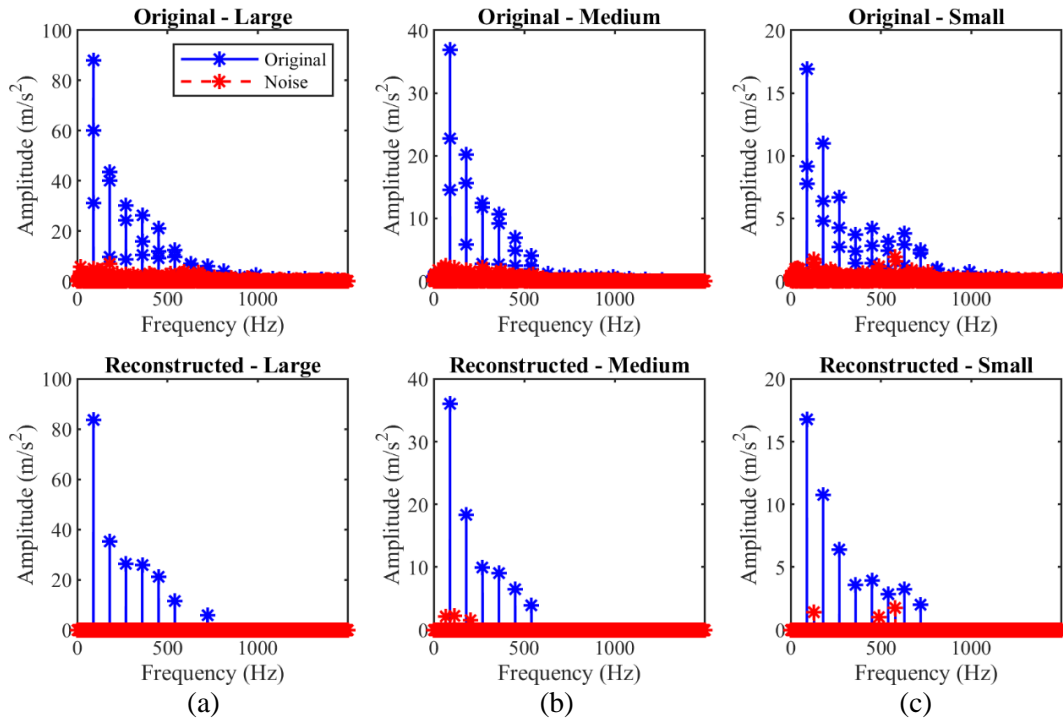


Figure 5.17 Fault feature selection for envelope spectra of bearings with outer race faults: (a) large, (b) medium, and (c) small

#### 5.4.2.2 Ball Bearings with Inner Race Fault

In order to prove the wide application of the proposed method, another series of tests on the bearings with inner race faults were performed in this research. Figure 5.18 (a) illustrates the time waveforms and spectra of four different cases. The amplitude of the vibration signal is positively related to the size of the bearing fault. The spectra of these cases are displayed in Figure 5.18 (b) and the energy mainly focused in the frequency band of [2000, 4000] Hz which is rich of the fault information.

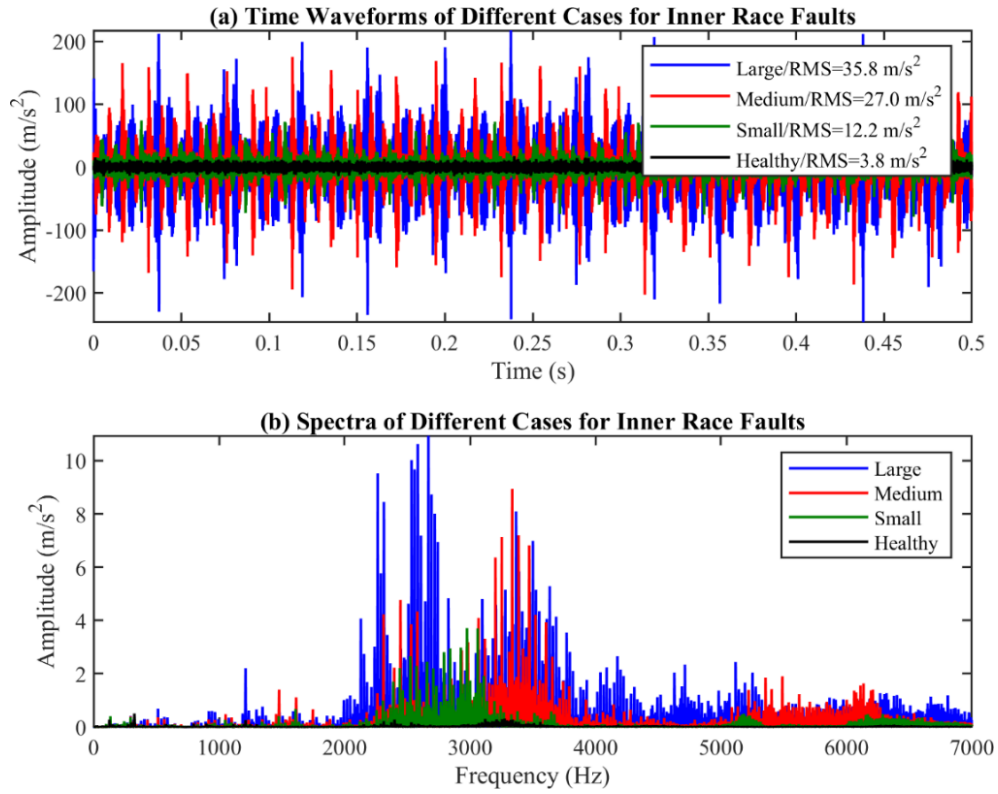


Figure 5.18 Original signals of bearings with inner race faults: (a) time waveforms, and (b) spectra

To reduce the amount used for the classification, only the frequency of [2000 Hz, 3000 Hz] was shifted to a lower band of [0 Hz 1000 Hz] with the sampling rate reduced to 3000 Hz by the analytic function. The shifted spectra and envelope spectra are illustrated in Figure 5.19 (a) and (b), respectively. The inner race fault frequency is apparent in the enlarged box. Their frequency values are not exactly the same because of manufacturing errors and rotating speed fluctuation. In addition, the rotating frequency and its harmonics, as well as the sidebands of the fault frequency are evident in the envelope spectra.

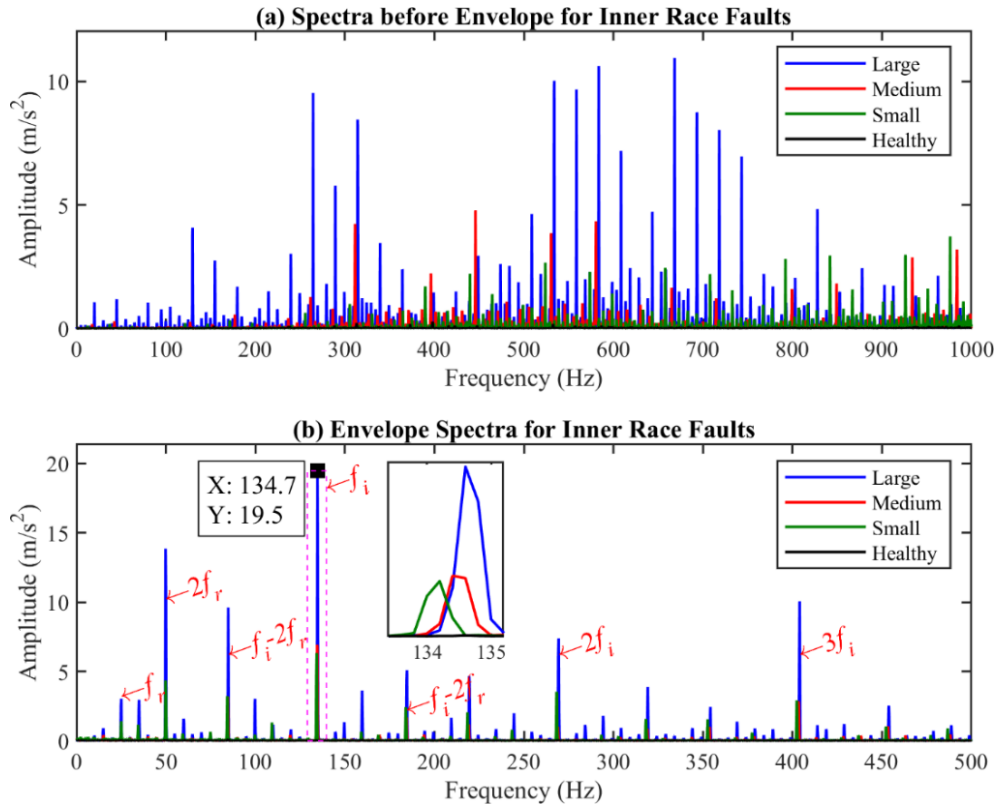


Figure 5.19 Spectra of bearings with inner race faults: (a) before envelope, and (b) envelope spectra

As described in the previous subsection, the second compression step is based on the sparse representation of envelope spectra. The sparsity of these four cases is determined as 35, 26, 28 and 42 as listed in Table 5.6 respectively according to the proposed adaptive sparse method expressed in Equation (5.9). The sparsity of signals will determine the number of observations in the second compression step, which will be shown with the compression ratios together in Table 5.6. The original envelope spectra and their recovered sparse representations by the OMP approach are displayed in Figure 5.20. It is manifest that both the original envelope spectra and the reconstructed sparse representation effectively explain the fault type of the bearing is the inner race fault, but the amplitude of the fault frequency cannot indicate the severity of the fault because of the selection of the frequency band. However, the reconstructed signals have only minor losses but less noise due to the noise elimination function of CS.

Furthermore, the envelope waveforms are recovered by the measurement matrix. The original and reconstructed envelope waveforms of these four cases are illustrated in Figure 5.21 with the blue and red lines, respectively. For the healthy case, the signal is not sparse because it is rich in noise, and a part of the noise is eliminated with CS, which leads to a significant difference between the original envelope waveform and the reconstructed one. But the faulty cases have much more excellent reconstruction results, especially for the bearing with a large outer race fault, which can be evaluated with the performance indicators as listed in Table 5.6.



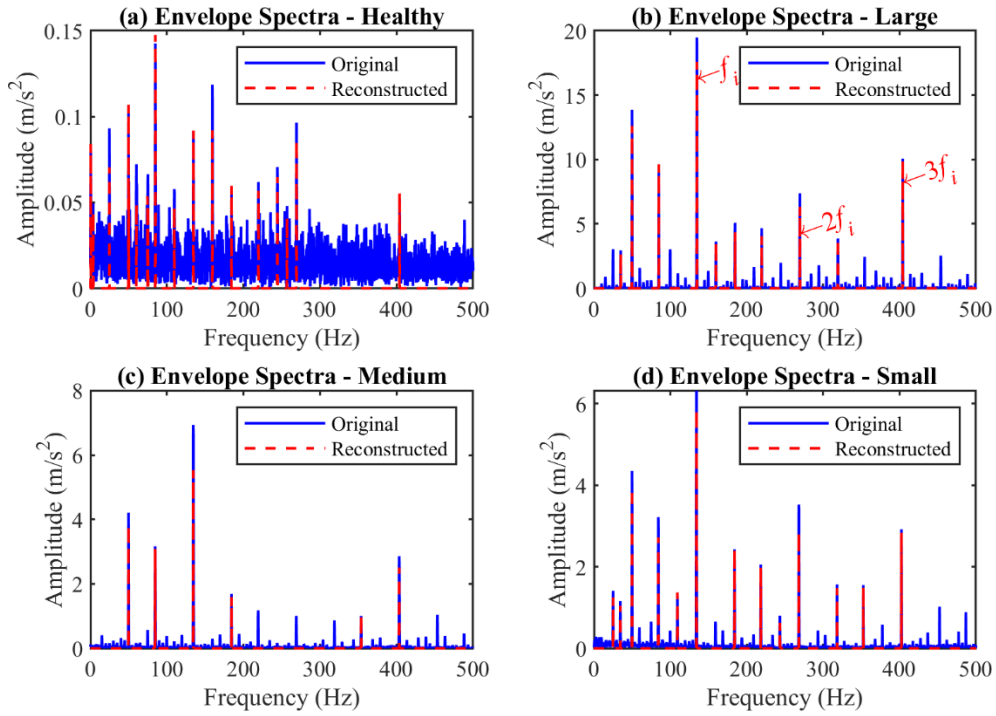


Figure 5.20 Original and reconstructed envelope spectra of bearings: (a) healthy, (b) large inner race fault, (c) medium inner race fault, and (b) small inner race fault

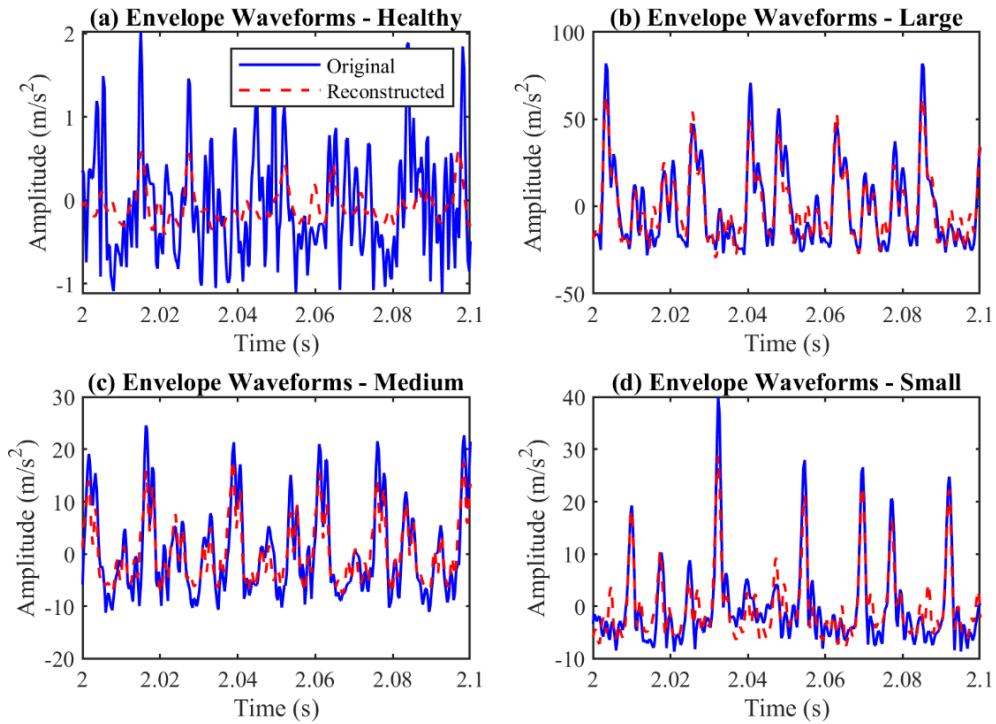


Figure 5.21 Original and reconstructed envelope signals of bearings: (a) healthy, (b) with a large fault, (c) with a medium fault, and (d) with a small fault

From the results shown in Table 5.6, it is noticeable that the  $CR$  of the inner race fault bearing signal is smaller than that of the outer race fault one which is caused by the low sparsity of the complex inner race fault signal. It is pronounced that both  $RMSE$  and  $PRD$  are not sensitive enough to evaluate the severity of the fault, but they can

distinguish the healthy and faulty bearings. In contrast,  $F$ -SNR can help distinguish the severity of the fault with a minor difference of the reconstructed envelope spectrum because of the low noise level.

Table 5.6 Main indicator values for inner race faults

Indicators		Defect Size			
		Healthy	Large	Medium	Small
$K$		35	26	28	42
$M$		185	144	153	215
$CR$	Step 1	32	32	32	32
	Step 2	81.08	104.17	98.04	69.77
	2 Steps	2594.6	3333.3	3137.3	2232.6
$RMSE$ (m/s <sup>2</sup> )		0.62	13.05	5.47	4.94
$PRD$ (%)		94.01	55.38	67.95	58.24
$F$ -SNR (dB)	Original envelope spectrum	-	24.31	25.06	23.63
	Reconstructed envelope spectrum	-	35.41	34.14	33.56

To calculate  $F$ -SNR, the blue is the selected fault features, and the red dots represent the remaining signals that are not relating to the fault characteristics defined in Equation (5-14). The value of  $F$ -SNR for reconstructed envelope spectrum is larger than the original one, which introduces that the fault characteristics of the reconstructed signal are more obvious, and the compression method has the denoising function, especially when the fault is severe.

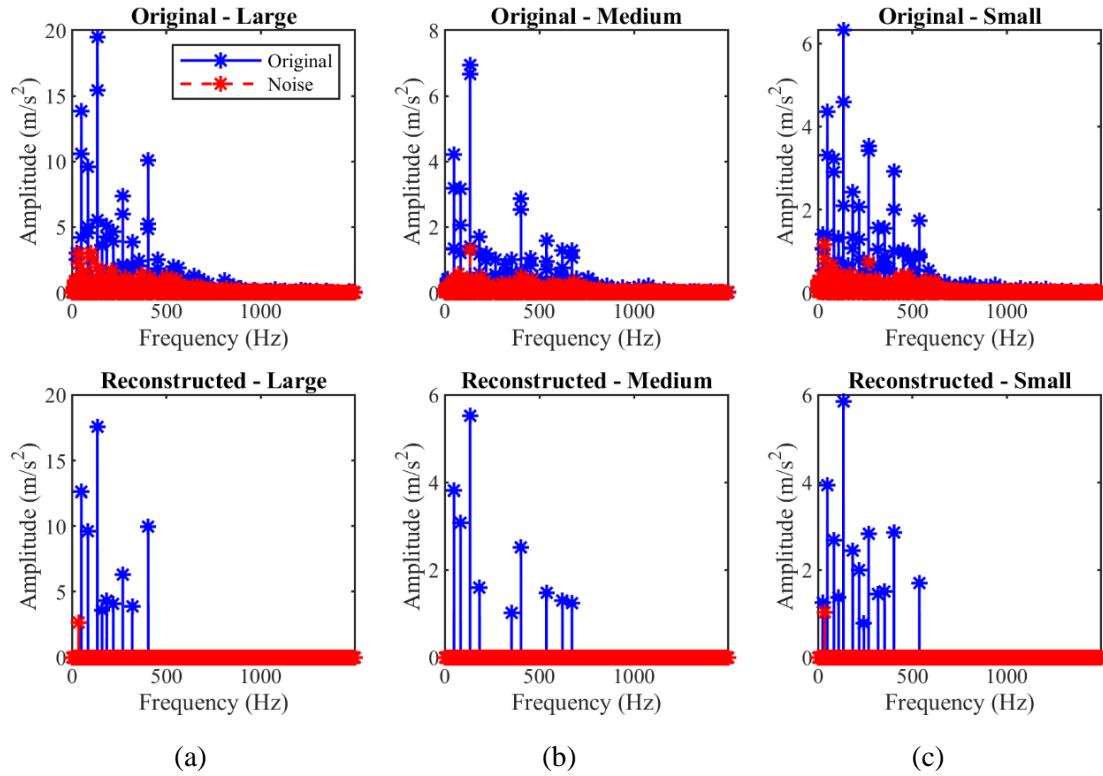


Figure 5.22 Fault feature selection for envelope spectra of bearings with inner race faults: (a) large, (b) medium, and (c) small

Although the compression ratio and reconstruction performance of the inner race fault bearing is not as good as the outer race fault one, the proposed approach compresses the original signal up to thousands of times and the fault can be determined to be bearing inner race fault by the sparse representation of the compressed signal, which provides a potential application to dramatically reduce the data volume in the field of wireless rotating machine CM.

## 5.5 Summary

A two-step adaptive compression method based on frequency shift and CS is proposed to reduce the size of datasets for fault diagnosis of ball bearings, which can be achieved at the acquisition end to significantly save the transmission bandwidth, storage space, and power consumption.

Besides, the original envelope signal can be reconstructed with much less noise to approximatively display the timewave of the faults. Moreover, an adaptive sparse signal method is investigated according to the fault mechanism of ball bearings. Furthermore, a new developed reconstruction performance evaluation indicator is defined following the effectiveness of fault characteristics on fault detection.

Both simulation and experimental results demonstrate that the defects on the ball bearing outer race and inner race can be accurately and efficiently detected with a very small amount of data (less than 200 data points) transferred and stored with the compression ratio of up to thousands of times. In conclusion, this chapter provides a

theoretical basis for wireless sensor nodes to further realize the online condition monitoring of rotating machines in real time in future.

---

## **CHAPTER 6    FEATURE EXTRACTION BASED COMPRESSION OF 2D SIGNALS IN FAULT DIAGNOSIS OF RECIPROCATING COMPRESSORS**

---

*2D thermal images contain a wealth of information relating to machine conditions. However, the transmission and storage of substantial images are less efficient and more cost-intensive than 1D signals. This chapter and the next chapter will focus on 2D signal compression in intelligent fault diagnosis.*

*This chapter employs the dense SIFT to extract the features from the 2D infrared thermal images. To reduce the dimensionality of the massive feature sets, the sparse coding including dictionary learning and sparse representation were applied. Then, the sparse features are smartly classified by SVM to achieve effective diagnosis of various compressor faults.*

## 6.1 Introduction

Temperature is an effective indicator to monitor the condition of a machine because any change of the system will cause fluctuations in temperature. Usually, a thermocouple can be used to measure the temperature at a location of the system. However, the single point temperature is not reliable and stable because it is easily affected by the environment. It will result in non-uniform temperature distribution on the surface of the housing of machines if they work in an abnormal condition. IRT is a non-contact and non-invasive NDT method that can conveniently and accurately record changes in the temperature field distribution of surfaces of machines when the parts of the running machine are defective [167].

Reciprocating compressors change the cylinder volume cycle and achieve air pressurization and delivery through the reciprocating motion of the piston in the cylinder [168]. Figure 6.1 illustrates the schematic diagram of a two-stage reciprocating compressor. They are widely used in petroleum pipelines, chemical plants, refineries, new energy, food and other industries because of their wide application pressure range, high compression efficiency and strong adaptability.

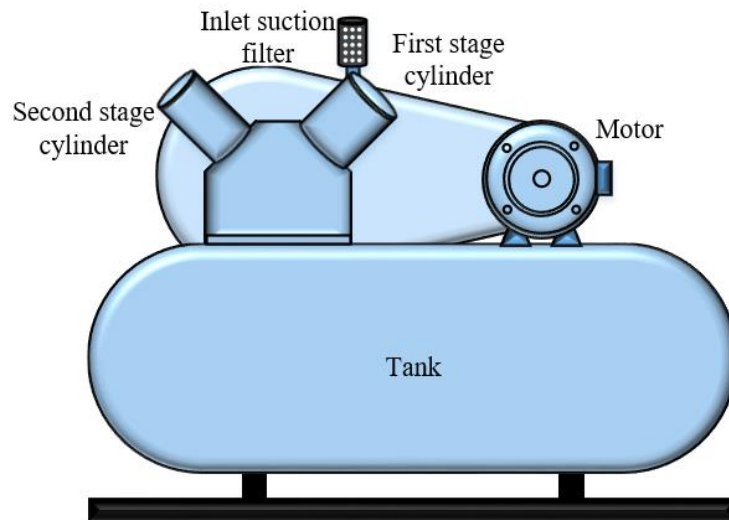


Figure 6.1 Schematic diagram of a two-stage reciprocating compressor

The compression structure of the reciprocating compressor is very important, consisting of cylinders, piston components, intake and discharge valves and other components as shown in Figure 6.2. In addition, it describes how the two-stage reciprocating compressor works. With the help of parts such as cylinders, pistons and valves, the compression process of a reciprocating compressor can be divided into four processes of compression, discharge, expansion and suction.

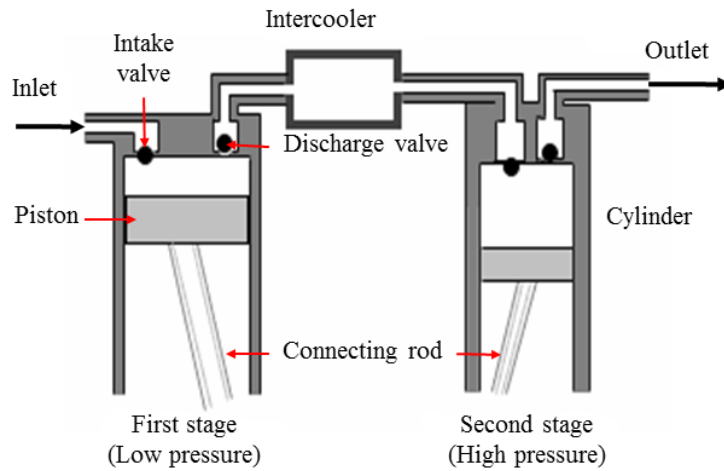


Figure 6.2 Internal compression structure of a two-stage reciprocating compressor

The relationship between pressure and volume in a cylinder during operation is a cyclic process as shown in Figure 6.3. From point A to point B, as the piston moves up, the volume of the cylinder gradually shrinks which means the air begins to be compressed. The air in the cylinder cannot be returned to the inlet pipe because the intake valve is closed. The air pressure of the outlet pipe is higher than that inside the cylinder so that the air in the cylinder cannot flow out from the discharge valve. As the volume of the cylinder shrinks, the air pressure continuously increases. When the pressure rises to the value of point B, the pressure of the compressed air rises to be higher than that of the air in the outlet pipe. The discharge valve opens, and the compressed air is discharged into the outlet pipe until the piston moves to the top dead centre, i.e. point C. Then, the piston starts moving to the down. As the piston moves downward, the volume of the cylinder increases, the inside pressure drops, and the gas remaining in the cylinder expands continuously. After the pressure in the cylinder is reduced to less than the pressure in the inlet pipe, i.e. point D, the intake valve opens and the gas enters the cylinder until the piston moves to the bottom dead centre which is represented with point A. Then, the piston begins to move up again and repeat the previously mentioned process. The piston continuously reciprocates in the cylinder to allow the cylinder to suction and discharge the air.

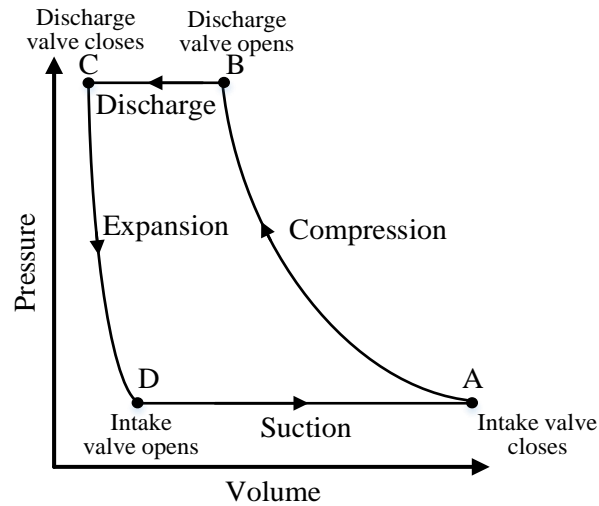


Figure 6.3 Relationship between pressure and volume in a cylinder during working

Although a reciprocating compressor has a wide pressure range, high efficiency and maintainability, some faults, like intercooler leakage, valve leakage, air filter blockage, asymmetrical stator winding of the motor, etc., will crucially affect its efficiency and even cause serious explosion accidents due to its complicated structure [168]. Hence, condition monitoring and fault diagnosis of reciprocating compressors are essential in industrial applications. Many efforts have been made by researchers with temperature, pressure, instantaneous angular speed, vibration and acoustic signals in these decades. For example, Tran *et al.* [168] investigated a fault diagnosis method of reciprocating compressor valves with Teager-Kaiser energy operator based on three signals of pressure, vibration and acoustic signals. Elhaj *et al.* [169] simulated multiple operation conditions of a two-stage reciprocating compressor with the numerical simulation and found that both the instantaneous angular speed fluctuation and pressure waveform are sensitive to its fault detection. Yang *et al.* [170] achieved reliable and flexible condition classification of a small reciprocating compressor with the methods of SVM and artificial neural network, etc. based on the noise and vibration signals. The results demonstrated that SVM and learning vector quantisation (LVQ) have highly accurate classification of reciprocating compressor faults. Haba *et al.* [171], [172] and Ogbulafor *et al.* [173] developed MSB based and wavelet transform based methods to effectively and efficiently detect different faults of the two-stage reciprocating compressor, respectively. Feng *et al.* [174], [175] detect the leakage faults of the two-stage reciprocating compressor with the on-rotor MEMS accelerometers based on the vibration signals. However, the sensors installed to collect these signals should be contacted with the surface of the reciprocating compressors or even invade the inside of the compressor, which will damage the structure of the compressor. Moreover, the noise of the reciprocating compressor is extremely large which is easy to affect the extraction of the fault signatures.

According to the ideal gas flow, the relationship between the temperature, pressure and volume of gas in a confined space is expressed as  $PV = nRT$ , where  $P$ ,  $V$  and  $T$  are the



pressure, volume and absolute temperature;  $n$  is the number of moles of gas and  $R$  is the ideal gas constant [176]. Therefore, the change of the air pressure and volume in the cylinder will inevitably lead to a change in temperature, resulting in the change of the temperature distribution of the surface of the reciprocating compressor because of the thermal conduction. The infrared thermography technique has sufficient ability to capture the change of the temperature distribution field of the reciprocating compressor surface so that it has been used to monitor the condition of compressors [177], [178]. But thermal images need to be stored for a long time to compare with the previous state, which consumes a lot of storage space. In this chapter, an infrared thermal camera FLIR ONE Pro is used to capture the thermal videos during the operation of the reciprocating compressor. The dense SIFT features are extracted from the recorded thermal videos and compressed as the sparse representations of different conditions for fault classification.

## 6.2 Compression Method

### 6.2.1 Dense SIFT Descriptor

The SIFT, known as one of the most robust methods for feature extraction in image processing, was proposed and improved by David G. Lowe [179] in 1999. It describes the image local features and remains invariant to image rotation, scaling, translation and brightness variations in scale space, which demonstrates that these changes will not substantially affect the extracted features. As a result, it leads to wide applications on object recognition and classification. The process of obtaining SIFT descriptors is briefly introduced as follows [180], [181].

The scale space  $L(x, y, \sigma)$  of an input image  $I(x, y)$  is created by convoluting with a Gaussian kernel function  $G(x, y, \sigma)$  as shown in Equation (6-1).

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (6-1)$$

where,  $\sigma$  is the scale space factor reflecting the degree of image blur and the Gaussian filter is as defined in Equation (6-2).

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (6-2)$$

Then, points of interest regarded as keypoints are identified as the local extrema of the Difference of Gaussians (DoG), denoted by  $D(x, y, \sigma)$  and stated in Equation (6-3).

$$D(x, y, \sigma) = L(x, y, k_i\sigma) - L(x, y, k_j\sigma) \quad (6-3)$$

where  $k_i\sigma$  and  $k_j\sigma$  represent the variable scales. In order to select the main points, the obtained keypoints require to be further reduced by curve fitting the DoG function to remove the points with low contrast and unstable edges. The magnitude  $m(x, y)$  and orientation  $\theta(x, y)$  of each main point  $L(x, y)$  can be calculated by means of gradients of its neighbouring pixels according to Equations (6-4) and (6-5), respectively.

$$m(x, y) = \sqrt{[L(x+1, y) - L(x-1, y)]^2 + [L(x, y+1) - L(x, y-1)]^2} \quad (6-4)$$

$$\theta(x, y) = \arctan \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \quad (6-5)$$

The main orientation of the feature point is determined by the peak of the orientation histograms taking the orientation  $\theta(x, y)$  as the horizontal coordinate and gradient magnitude as the vertical coordinate. As a result, each SIFT feature point is represented by location, scale and orientation, also shown with  $(x, y, \sigma, \theta)$ . There are  $4 \times 4 = 16$  histograms in each patch, and each one has with 8 bins. Therefore, a 128-dimensional vector is formed as the SIFT descriptor.

However, the SIFT descriptor is apparently effective and reliable for areas with clearly detectable and distinguishing texture features. For the platter regions, dense SIFT provides robust and sufficient feature measurements through selecting intensive keypoints with a sliding window to calculate the dense SIFT descriptors instead of extracting from scale-space extrema, which was proposed by Liu *et al.* [182]. In this study, an image is divided into patches with the size of  $16 \times 16$  pixels. As shown in Figure 6.4, each patch contains 16 blocks with the size of  $4 \times 4$  pixels and each block has gradients of 8 orientations so that a 128-dimensional dense SIFT feature descriptor can be obtained to represent this patch. Furthermore, an image  $X_i$  with  $m$  patches can be represented with a matrix  $Y_i = [y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(m)}] \in R^{128 \times m}$  which consists of  $m$  dense SIFT descriptor vectors  $y_i$ .

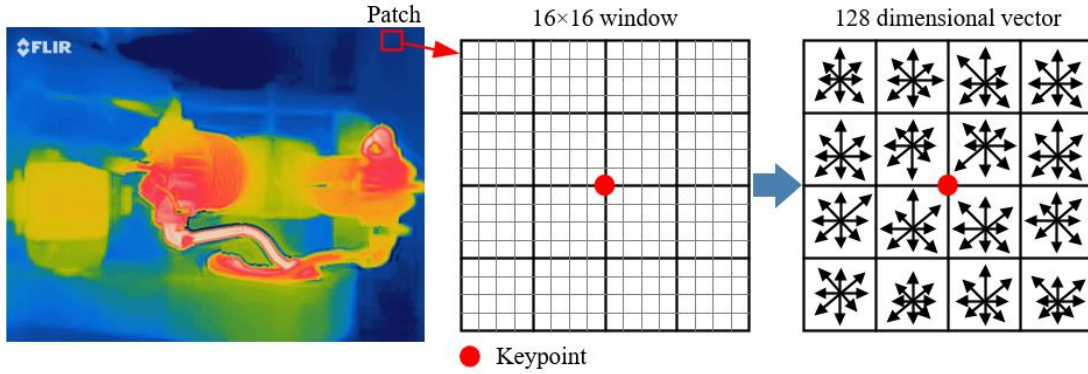


Figure 6.4 128-dimensional dense SIFT descriptor

### 6.2.2 Sparse Coding

In order to reduce the dimension of the features, an unsupervised learning method, well known as sparse coding or sparse dictionary learning, was applied in large-scale visual recognition [183]. It includes two processes: dictionary learning and sparse representation. Dictionary learning is a dimensionality reduction representation of massive data sets or information compression. Moreover, it attempts to learn the basic features from the samples to effectively explain structures and patterns hidden in the data sets, especially the over-complete dictionaries. The sparse representation uses a small number of learned dictionary atoms and non-zero coefficients to represent more

information, which not only saves storage space but also speeds up the calculation. As a result, the sparse coding enables the automatic selection of features and broadly applied in blind source separation, audio signal processing, image feature extraction, pattern recognition and so on.

Assumed that a video  $\mathbf{X}$  has  $N$  frames and it can be expressed as  $\mathbf{X} = \{X_1, X_2, \dots, X_N\}$  and represented with the extracted dense SIFT features as  $\mathbf{Y} = \{Y_1, Y_2, \dots, Y_N\}$ .  $K$  dense SIFT descriptors are randomly selected from  $\mathbf{Y}$  for the training process, i.e., dictionary learning to obtain a set of bases  $[\phi_1, \phi_2, \dots]$  to represent  $\mathbf{Y}$  by minimizing the Equation (6-6).

$$\min_{\Phi, \alpha^{(k)}} \sum_{k=1}^K \left\{ \|\mathbf{y}^{(k)} - \Phi \alpha^{(k)}\|^2 + \lambda |\alpha^{(k)}| \right\} \quad (6-6)$$

where  $\Phi$  is the array of  $\phi_k$  and  $\alpha^{(k)}$  is the weight of  $\phi_k$ .  $\Phi$  and  $\alpha^{(k)}$  are calculated by iterative calculation to convergence. In this study, the number of the elements for the trained dictionary is set to 2048, so that the dimension of  $\Phi$  is  $128 \times 2048$ .

Given a new image, a sparse vector  $\alpha$  is obtained by solving the Equation (6-6) with the fixed trained dictionary  $\Phi$ . The process of obtaining a sparse representation  $A_i = [a_i^{(1)}, a_i^{(2)}, \dots, a_i^{(K)}] \in R^{2048 \times K}$  of the input image is called coding.

### 6.2.3 Support Vector Machine

In 1995, Corinna and Vladimir proposed a nonlinear SVM model [184], [185], which is a traditional and effective supervised learning model used for solving small samples, nonlinear and high dimensional data classification and recognition analysis. The training dataset is designed as  $(X_i, c_i), i = 1, 2, \dots, N$ , where  $X_i$  is an input image represented with the sparse feature representation  $A_i$  and  $c_i \in C = \{1, 2, \dots, M\}$  is the classification labels. Equation (6-7) shows the model training principle of SVM.

$$\min_{w, b, \xi} \left( \frac{1}{2} \|w_m\|^2 + C \sum_{i=1}^N \xi_i \right) \text{ s.t. } f(c_i)(w_m a_i + b_m) \geq 1 - \xi_i, \xi_i \geq 0 \quad (6-7)$$

where  $f(c_i) = \begin{cases} 1, & c_i = m \\ -1, & \text{others} \end{cases}$  and  $\xi$  is a threshold.

Some new samples as the inputs are tested with the trained model to estimate their class predictions. The assessment criteria of the predicted classification, called *accuracy*, can be given by

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6-8)$$

where  $TP$ ,  $TN$ ,  $FP$  and  $FN$  represent the number of true positives, true negatives, false positives and false negatives in the prediction results, respectively. The fault classification method is implemented with SVM based on the dense SIFT features and corresponding sparse representations, whose flow chart is clearly shown in Figure 6.5.

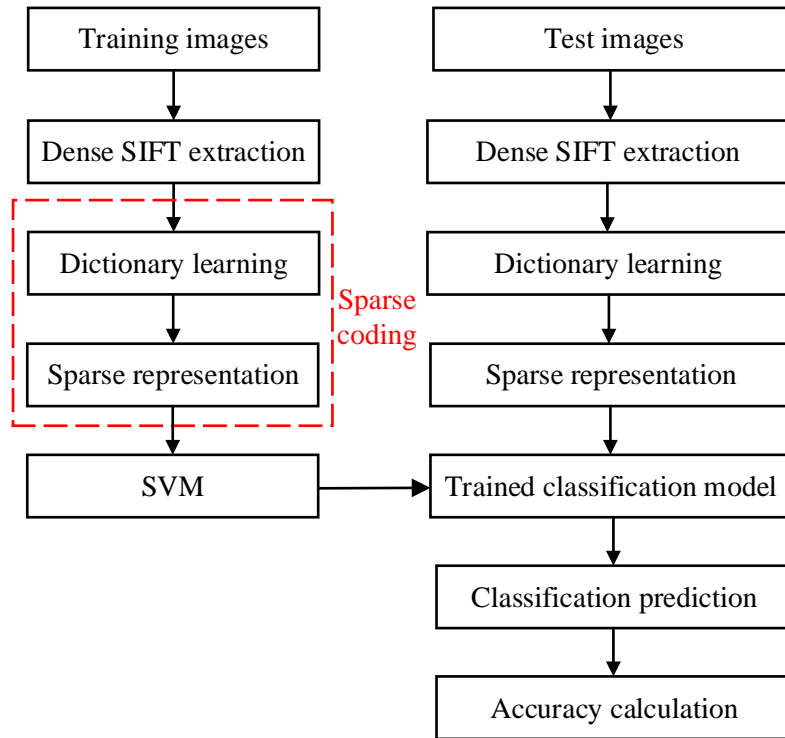


Figure 6.5 Flow chart of the proposed fault classification method

## 6.3 Experiments

### 6.3.1 Experimental Setup

#### 6.3.1.1 Test Facilities

To validate the compression method proposed in the previous subsection, some experiments were carried out on a two-stage and single action V-shaped Broom Wade TS9 reciprocating compressor with the structure shown in Figure 6.6. The compressor is driven by a three-phase motor via the transmission belt. The air enters the first stage cylinder through the air filter, and the compressed air goes into the second stage cylinder to be further compressed and then enters into the tank. The specification of the reciprocating compressor is listed in Table 6.1.

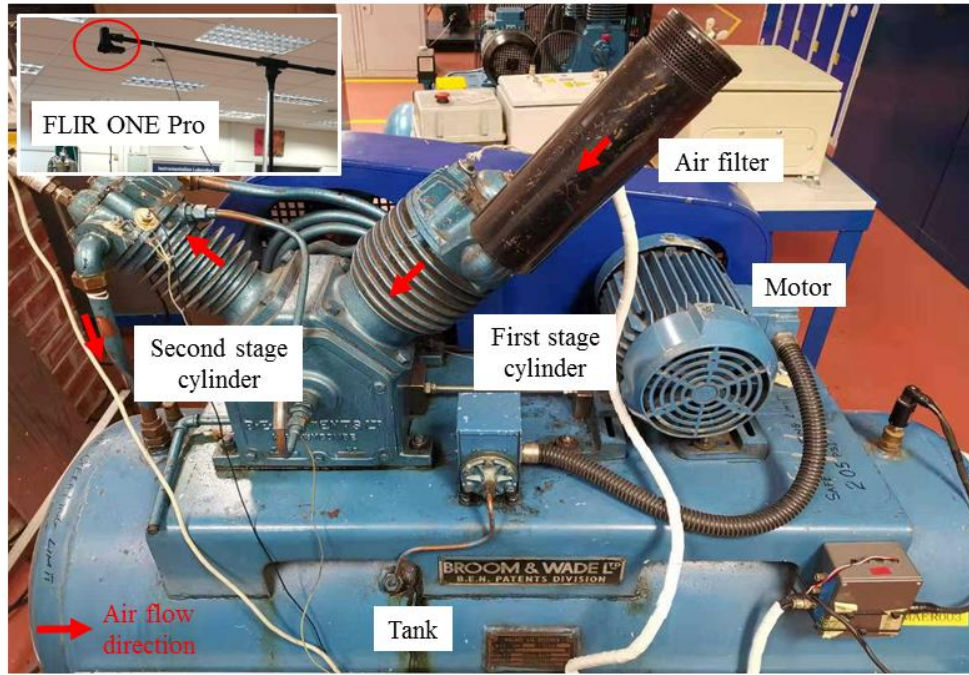


Figure 6.6 Structure of the two-stage reciprocating compressor test rig

Table 6.1 Specification of the two-stage reciprocating compressor

Parameters		Values
Electric Motor	Speed	1420 rpm
	Power	2.5/3 kW
Piston Diameter	First Stage	93.6 mm
	Second Stage	55.6 mm
Max. Working Pressure		1.38 MPa (200 psi)
Piston Stroke		76 mm
Crankshaft Speed		420-440 rpm

A thermal imaging infrared camera, called FLIR ONE Pro, was used in this experiment because it has improved temperature measurement accuracy and better visual details. Moreover, it has an adjustable connector to fit with a variety of handheld devices. Besides, its size is very small which is very portable. The specification of the FLIR ONE Pro is detailed in Table 6.2. In this experiment, it was fixed on the tripod at a distance of 93 cm from the V-shaped corner between the first and second stage cylinders to capture the temperature field distribution of this two-stage reciprocating compressor surface in the top view.

Table 6.2 Parameters of the FLIR ONE Pro

Parameters	Values
Thermal Pixel Size	12 $\mu\text{m}$
Thermal Resolution	19,200 pixels (160×120 pixels)

Visual Resolution	1440×1080 pixels (W×H)
Thermal Sensitivity	70 mK
Accuracy	$\pm 3^{\circ}\text{C}$ or $\pm 5\%$
Spectral Range	8-14 $\mu\text{m}$
Object Temperature Range	$-20^{\circ}\text{C}$ to $400^{\circ}\text{C}$
HFOV / VFOV	$50^{\circ} \pm 1^{\circ}$ / $43^{\circ} \pm 1^{\circ}$
Frame Rate	8.7 Hz
Focus	Fixed 15 cm – infinity

### 6.3.1.2 Defect Seeding

To simulated failures commonly occur in the two-stage reciprocating compressor, some defects were seeded or simulated artificially in different parts of the compressor. The first fault is air filter blockage (AFB), which was performed through sealing a quarter of the air filter inlet holes as shown in Figure 6.7 (a). The occurrence of this failure will seriously affect the amount of inlet air. The next two faults are forming asymmetrical stator windings (ASW) through adding a phase winding resistance with different resistance values. The ASW faults will make the motor work in an unstable state. Figure 6.7 (b) displays the external resistor bank consisting of seven resistors, five resistance values of  $0.1\ \Omega$  and another two of  $0.5\ \Omega$ . These two sets of fault simulation are achieved by connecting the resistance values of  $1\ \Omega$  ( $0.1 \times 5 + 0.5 = 1\ \Omega$ ) and  $1.5\ \Omega$  ( $0.1 \times 5 + 0.5 \times 2 = 1.5\ \Omega$ ) to a phase of motor power supply, respectively. The third one is the discharge valve leakage (DVL) in the second stage simulated with a 2 mm diameter hole drilled on the discharge valve as illustrated in Figure 6.7 (c). The second stage discharge pressure and volumetric efficiency will be low if the discharge valve is faulty and remains open during operation. The last one is the intercooler leakage (ICL), which was simulated by loosening the screw nut with two circles, but the leakage is very small because of the seal function of the PTFE tape.

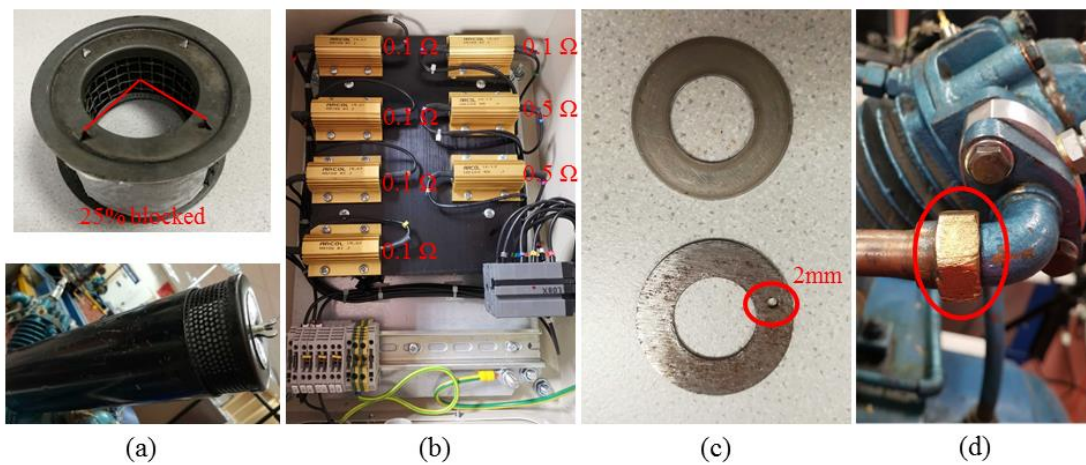


Figure 6.7 Simulated defects of the reciprocating compressor: (a) AFB, (b) ASW, (c) DVL and (d) ICL



### 6.3.1.3 Test Procedure

The reciprocating compressor was operated to constantly compress air from the pressure in the tank of 0 psi to 120 psi (i.e. 0.83 MPa). The first test was implemented in the normal working condition of the reciprocating compressor, which is a healthy case and can be regarded as the baseline relative to other faulty cases (denoted by BL). The other five cases including AFB, ASW with 1  $\Omega$  (denoted by ASW100) and ASW with 1.5  $\Omega$  (denoted by ASW150), DVL and ICL were also performed on consecutive days with different ambient temperatures as shown in Table 6.3. The time taken for each test is listed in Table 6.3, which is slightly different from the actual time required for the pressure to rise from 0 to 120 psi. Only the DVL case took a longer time to finish which may be due to the big leak hole on the discharge valve. During the operation of the compressor, the thermal camera recorded the temperature field distribution of the focusing area.

Table 6.3 Experimental parameters and dataset grouping for training and testing

Labels	Running Period (H:M:S)	Ambient Temperature ( $^{\circ}\text{C}$ )	Frames	Training Group	Testing Group
BL <sup>1</sup>	00:10:14	23.8	3293	Randomly select 1600 frames	Select the remaining 1600 frames
AFB <sup>2</sup>	00:10:12	24.1	3411		
ASW100 <sup>3</sup>	00:10:15	23.3	3829		
ASW150 <sup>4</sup>	00:10:11	23.7	3822		
DVL <sup>5</sup>	00:10:31	23.2	3487		
ICL <sup>6</sup>	00:10:12	23.7	3425		

Note: 1. BL: baseline (healthy case); 2. AFB: air filter blockage; 3. ASW100: asymmetrical stator winding with 1.0  $\Omega$ ; 4. ASW150: asymmetrical stator winding with 1.5  $\Omega$ ; 5. DVL: discharge valve leakage; 6. ICL: intercooler leakage.

However, a one-meter extension cable was added to connect the infrared thermal camera with the mobile phone for the convenient recording operation. The frame rates of the recorded videos are lower than the specified 8.7 Hz which may be limited by the performance of the extension cable. Additionally, the length of the recordings for different cases varies. Therefore, different frames of videos were recorded for these six cases. In order to facilitate processing, 3200 images are randomly selected for each case with half of the randomly selection used to train and another half to test. The specific dataset grouping is list in Table 6.3.

### 6.3.2 Results and Discussion

To clearly observe the temperature field distribution of the compressor surface from the thermal images, a photo was taken from the top view with the natural light to exhibits areas of interest as shown in Figure 6.8, which is the area of the thermal imaging video taken. The position of the camera remained unchanged during the process of all experiments to minimise the impact of changes in the focusing area on the classification

results. Deng *et al.* [186] in our research group have demonstrated that the changes of position have very little effect on the classification results with the CNN method. Moreover, the dense SIFT descriptor has the ability to maintain invariant to rotation, scaling, translation and brightness variations of images, which states that these changes will not affect the extracted features.

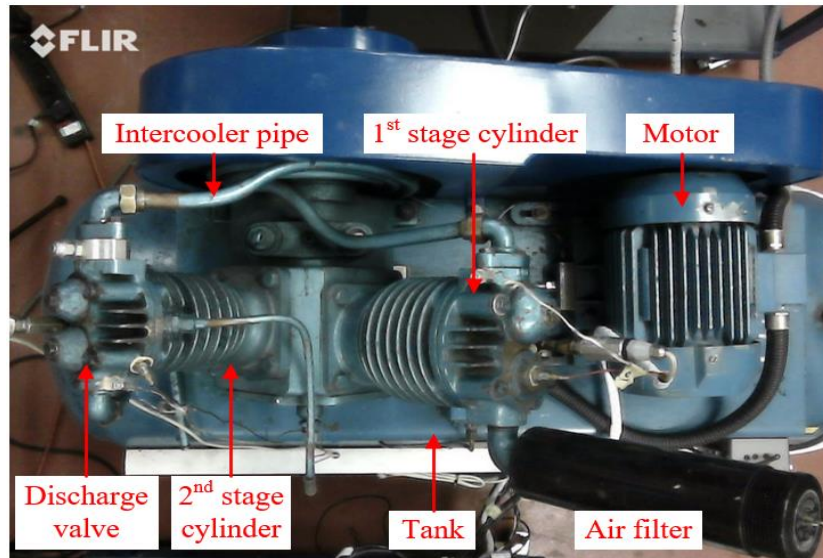


Figure 6.8 A sample photo taken from the top view

#### 6.3.2.1 Visualization analysis

The last captured image was selected to state the temperature field distribution for each case as illustrated in Figure 6.9. Figures (a) to (f) express six different cases of BL, AFB, ASW100, ASW150, DVL and ICL, respectively. The dust and debris are inevitable in the factory. When the compressor operates for a long time, it is unescapable to accumulate a large amount of dust, debris or oil to clog the inlet of the air filter and reduce air inflow speed. A clogged air filter will induce pressure drop in the first cylinder which may result in the temperature decrease and compressor output reduced by more than 2%. Asymmetrical stator winding is a common stator winding failure in motors. When the fault occurs in the early stage, the motor can continue to run with the imbalanced current and low output power leading to more heat generated on the motor. However, long-term operation under the condition with the ASW fault will cause motor breakdown and even irreversible damage. The third failure commonly occurred in the reciprocating compressor is the discharge valve leakage, which may be induced by the corrosive contaminants, improper lubrication, severe fatigue, spring failure, etc. It will result in low discharge pressure and volumetric efficiency. Moreover, the discharge temperature will be higher because some air in the tank will slip back into the cylinder from the leaked discharge valve. An intercooler exists between two stages for a two-stage reciprocating compressor. Its leakage caused by corrosion or poor sealing will reduce the air volume, which results in pressure reduction and temperature decrease in the second stage cylinder. All these different types of faults will cause changes in the pressure and temperature in the compressor, and the temperature field



distribution of the compressor wall will also gradually change with the thermal conduction. The last one of the 3200 images randomly selected from the captured thermal videos was chosen for each case to express the thermal field distribution as shown in Figure 6.9. Figure (e) states that the reciprocating compressor with the DVL fault generates much more heat than other faults. However, the other five cases are difficult to distinguish. For further analysis, the image resolution is compressed from  $1440 \times 1080$  pixels to  $160 \times 120$  pixels (same to the thermal resolution of the camera) with the cubic interpolation. Simultaneously, the values of several points are marked in the resized grayscale images. Unfortunately, different fault types cannot be definitely identified from neither the RGB images nor corresponding grayscale images. Therefore, the intelligent method will be applied to extract features to distinguish different types of failures.

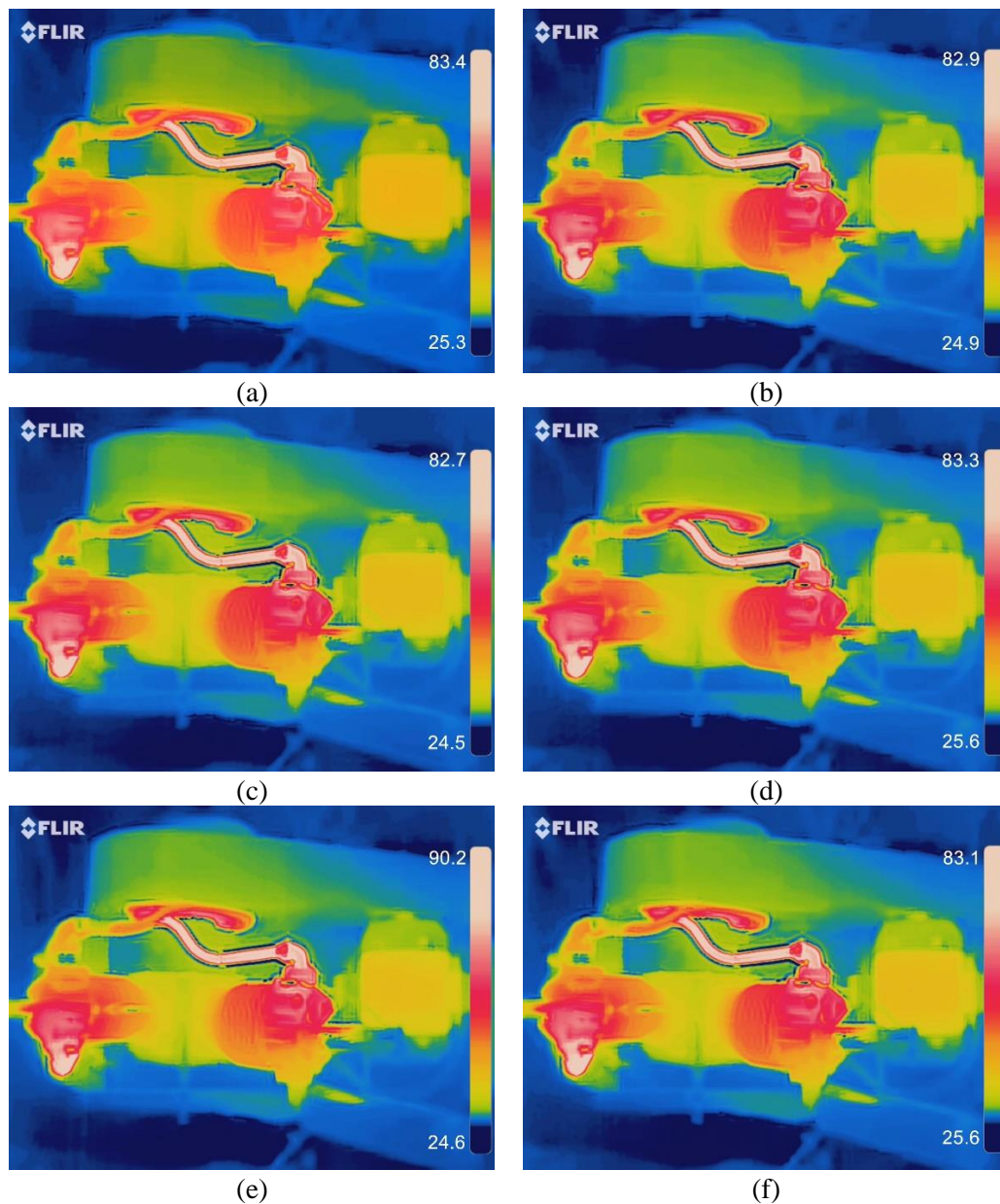


Figure 6.9 Selected images for six different cases: (a) BL, (b) AFB, (c) ASW100, (d) ASW150, (e) DVL and (f) ICL

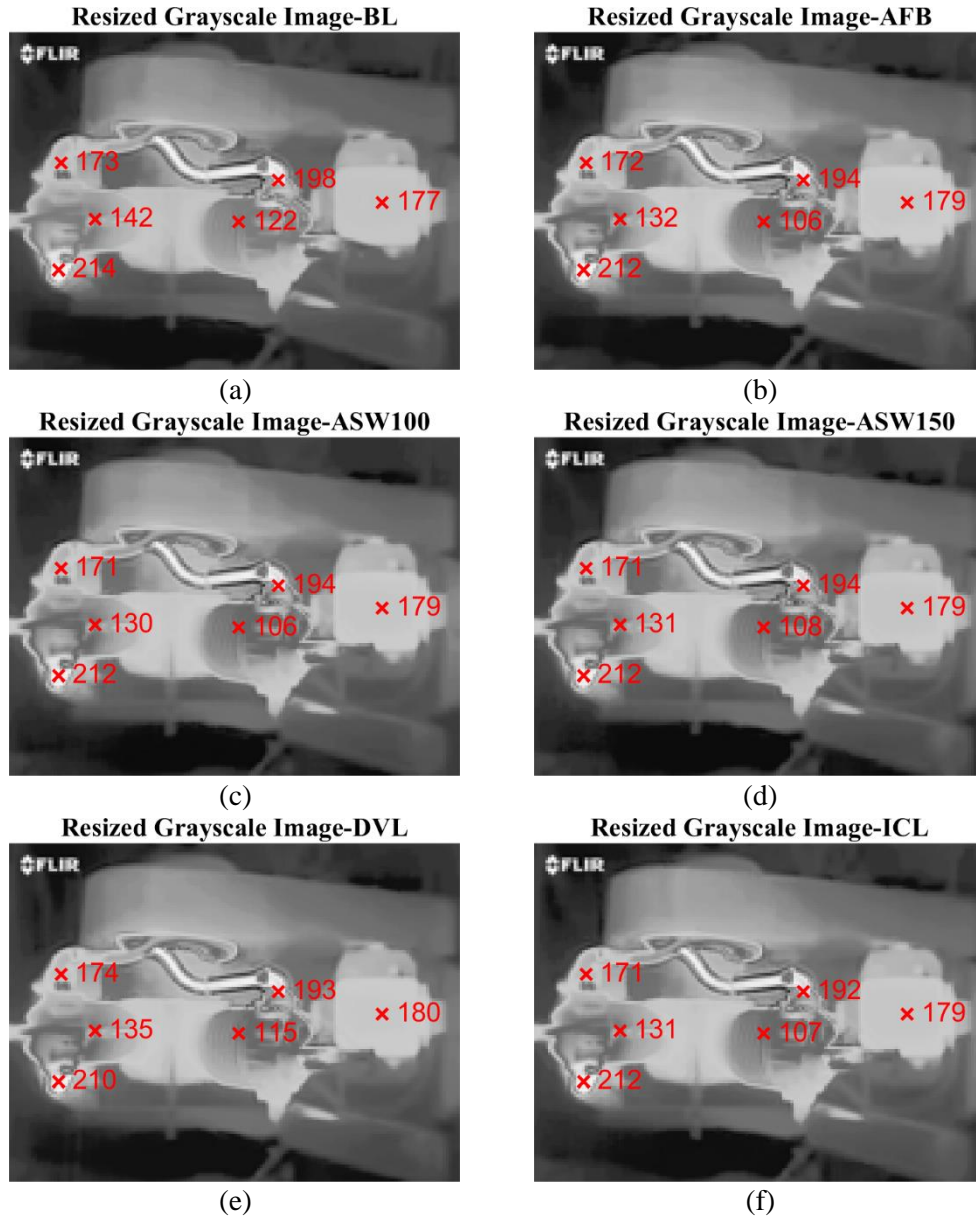


Figure 6.10 Examples of the resized grayscale image under the different conditions: (a) BL, (b) AFB, (c) ASW100, (d) ASW150, (e) DVL and (f) ICL

### 6.3.2.2 Intelligent analysis

After the previous preprocessing with the grayscale transform and resize, the data amount of each image has been reduced from  $1440 \times 1080 \times 3$  to  $160 \times 120$ . These are the common and basic compression methods in image processing, but the computational complexity is acceptable for conventional computers instead of using GUIs. Unfortunately, it is impossible to observe features from both the original and compressed images to distinguish the healthy case and the other five faulty cases. Therefore, the 128-dimensional dense SIFT features in the previous subsection are extracted from the compressed grayscale thermal images and sparsely represented based on sparse coding to implement the feature extraction and compression, as well as the fault type classification.

In the process of extracting dense SIFT features, a rotationally symmetric Gaussian lowpass filter with the size of  $5 \times 5$ , together with its horizontal and vertical gradients are shown in Figure 6.11. They are applied to obtain the numerical gradient of the resized grayscale images, respectively. The filtered results of an image for the BL case are illustrated in Figure 6.12. The horizontal and vertical gradients represent the directional gradients in the  $x$  and  $y$  directions separately. Based on these two directional gradients, the gradient magnitude and direction are calculated and displayed in the second row of Figure 6.12.

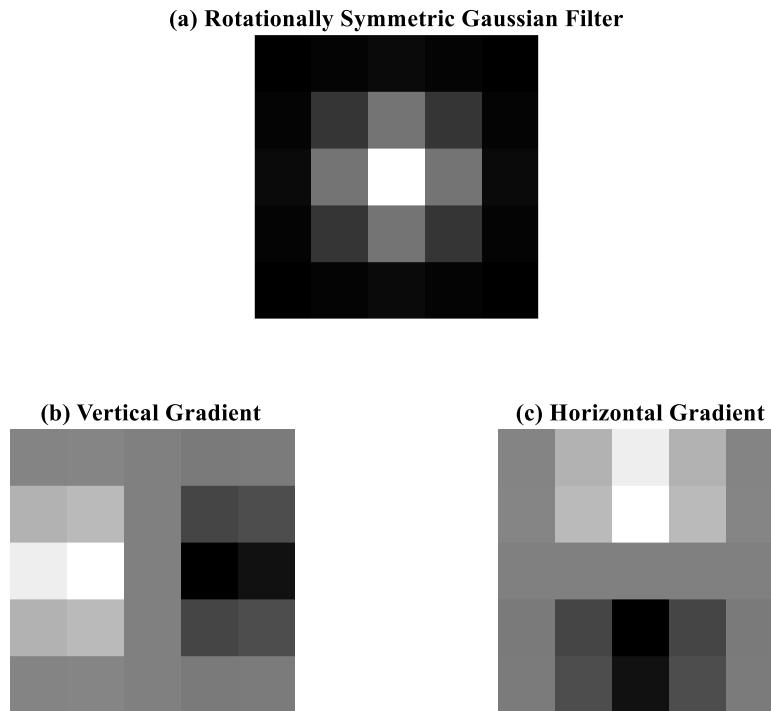


Figure 6.11 2D rotationally symmetric Gaussian lowpass filter and its gradients

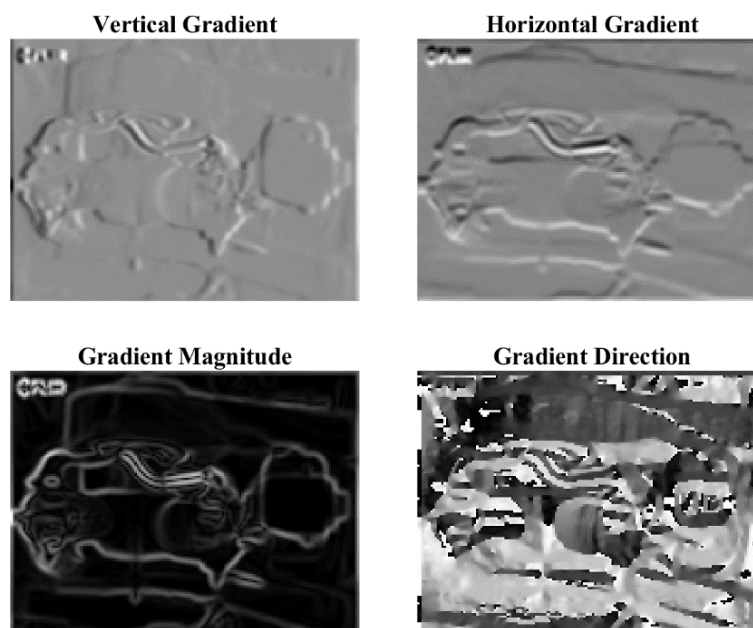


Figure 6.12 Filtered vertical and horizontal vertical and horizontal Gaussian filters



All these figures preserve unambiguous edge information. The 128-dimensional dense SIFT features are generated by calculating the preserved information in 8 different orientations for 16 blocks as described in Figure 6.4. In order to clearly observe the extracted SIFT features, these features of the original scale grayscale images are rearranged and reconstructed according to the corresponding positions of the patches and original images for these six different cases as shown in Figure 6.13. The feature size of each image is  $6030 \times 128$ . To visualize the extracted features with 2D images, the representative vector with the largest amplitude is selected from 8 different orientations as the grey value for each  $4 \times 4$  block. Therefore, the displayed images look like sharpened effects. If the mean values are utilized instead of the maximum values, the effect will be smooth but the outline will not be obvious.

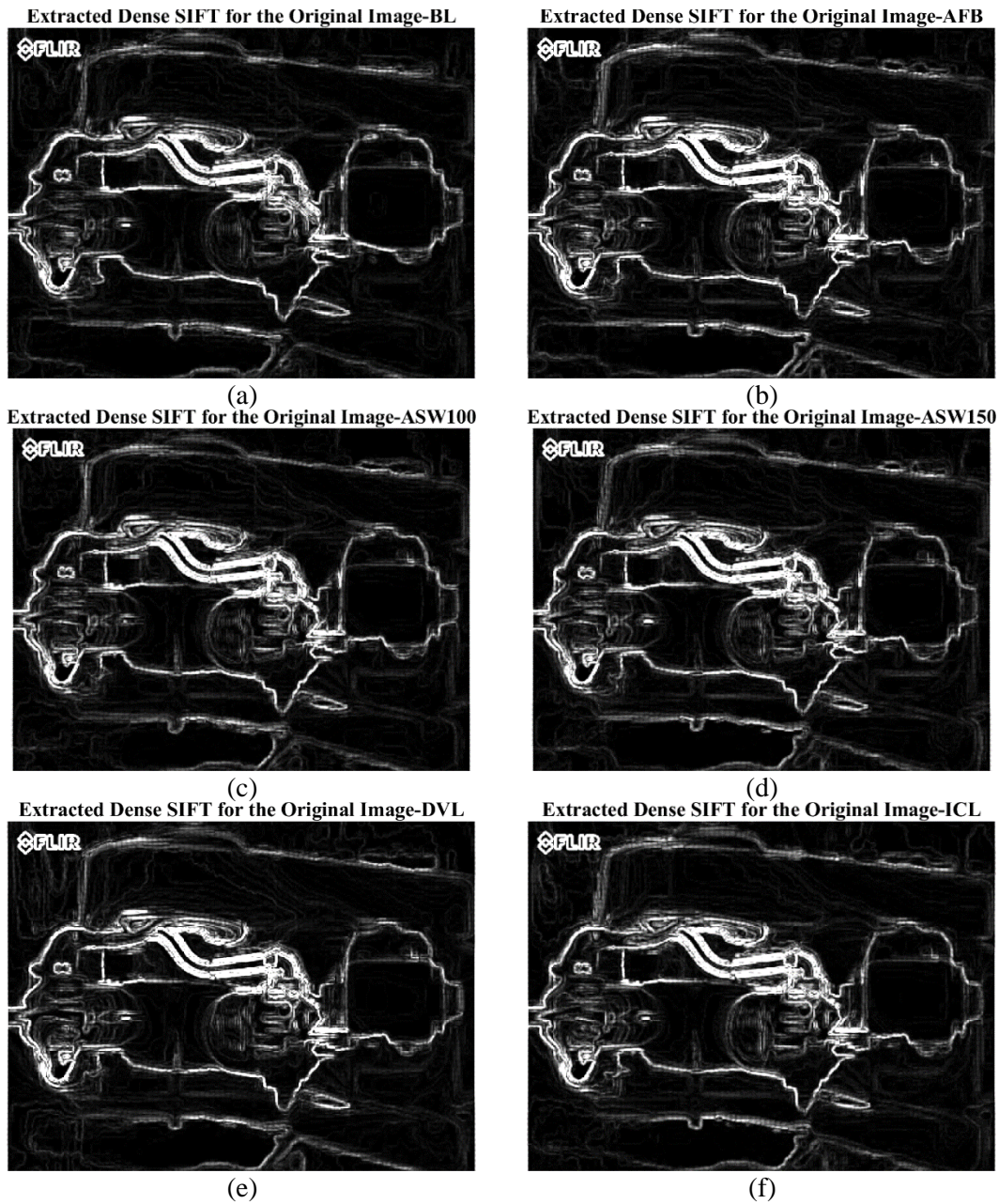


Figure 6.13 Extracted dense SIFT features from original images under the different conditions: (a) BL, (b) AFB, (c) ASW100, (d) ASW150, (e) DVL and (f) ICL

Actually, the resized images are applied to extract the dense SIFT features for further analysis. The resized image can be divided into 70 patches of  $16 \times 16$  without overlap, the size of the extracted features of each image is  $70 \times 128$ . As described previously, the maximum vector of 8 orientations is chose as the representation of each block to approximately observe the extracted features in Figure 6.14. The compressor contour is faintly visible, but the extracted edge information is very blurred. Even the graphics of adjacent patches cannot be perfectly connected. Fortunately, SVM as a generalized linear classifier can classify data in a binary way based on supervised learning.

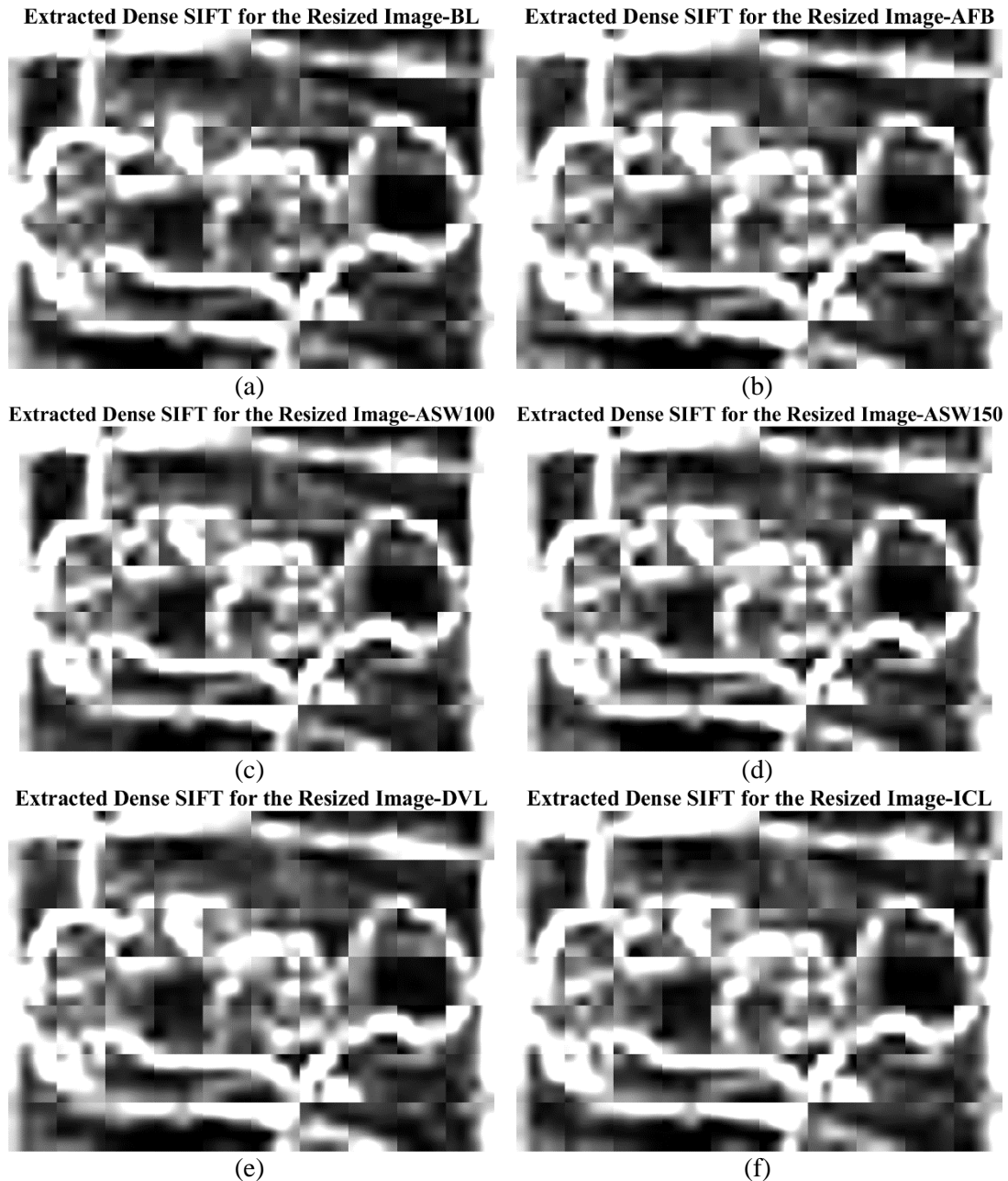


Figure 6.14 Extracted dense SIFT features from resized images under the different conditions: (a) BL, (b) AFB, (c) ASW100, (d) ASW150, (e) DVL and (f) ICL

It is manifest that the extracted features are a massive dataset which can be compressed to reduce the dimension and accelerate calculation with the dictionary learning and

sparse coding. The size of the learned dictionary is  $128 \times 2048$ . To improve the accuracy of training, images at different scales forming a multi-resolution image pyramid is provided to the machine, so that the machine can perceive the objects at different scales. In this study, the three-layer pyramid image feature sets are input to obtain the sparse representations, half of which will be applied to classify various types of faults with unsupervised SVM method and another half for evaluating the performance of the extracted and sparse features.

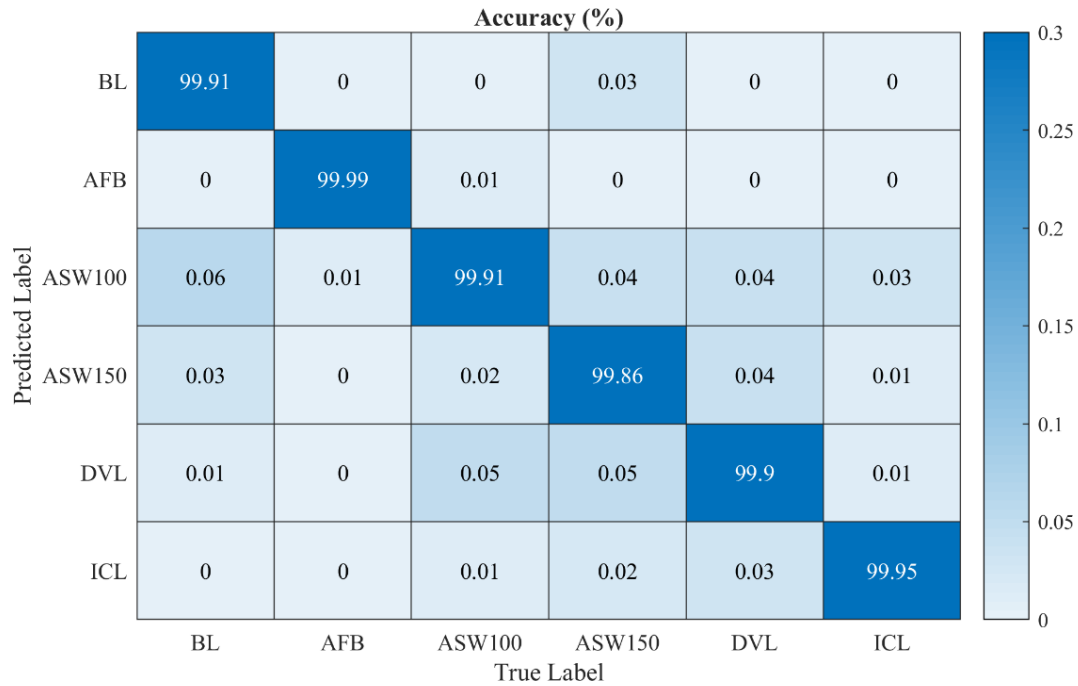


Figure 6.15 Confusion matrix of the fault classification accuracy for testing results

The fault classification accuracy of thermal images for six different cases using SVM is shown with a confusion matrix in Figure 6.15. The  $x$  and  $y$  axes represent the real and predicted fault types, respectively. For example, the probability of AFB being correctly judged is 99.99%, and the probability of being misjudged as ASW100 is 0.01%. The classification accuracy of each case is displayed with diagonal values. It is significantly apparent that the dense SIFT features extracted from thermal images have sufficient information for further classification of six different cases based on sparse coding for the reciprocating compressor. The probability of the highest classification error is 0.14%.

## 6.4 Summary

This chapter presents an effective method to diagnose the various types of faults in reciprocating compressor by thermal imaging analysis. The differences in temperature distributions presented by thermal images explain the changes in heat generation and heat conduction caused by the air compression and component friction. Thermal imaging features extracted by dense SIFT can be represented as a significantly sparse matrix with the sparse coding. The compressed image features with fewer dimensions can lead to efficient fault diagnoses by the SVM algorithm. The proposed compression

method can achieve an effective and efficient classification with an accuracy more than 99.86% for six different cases of the two-stage reciprocating compressor.

---

## CHAPTER 7    COMPRESSIVE SENSING BASED COMPRESSION OF 2D SIGNALS FOR FAULT DIAGNOSIS OF GEARBOXES

---

*As an essential part used for power transmission in many industrial applications, the gearboxes are vulnerable to various types of faults due to the long-time and heavy-duty operations. The faults occurred can cause the alteration of heat characteristics including heat sources, heat transfer paths and temperature field distribution. Thus, temperature information measured on gearbox housing can be a promising technique to monitor the gearboxes.*

*A high dimensional temperature signal captured by thermal imaging remotely, revealing heat distribution of an object, has the capability to extract more sufficient and accurate signatures for diagnosing the gearbox failures than the single point temperature measurement. However, the 2D images occupy a large transmission bandwidth and storage space, especially for wireless transmission. As a result, they are desired to be compressed whilst reserving meaningful features for fault diagnosis. This chapter investigates an intelligent CS-based CNN method on 2D images for the fault and severity classification of gearboxes.*



## 7.1 Introduction

Helical gearboxes are widely used for the power, speed and motion transmissions in industrial machinery [187], [188], such as wind turbines, ships, vehicles, trains and so on. Considering the relative motion of mating gears and bearings inside gearboxes, sufficient and accurate lubrication is critically necessary for gearboxes to maintain the smooth operation and guarantee their long service life. However, in the harsh working environments, oil loss due to leakage and evaporation is inevitable, especially the applications in the heavy-duty machinery. The loss of lubricant, also known as starved lubrication in the gearboxes leads to severe friction or even wear among the mating components which will cause an abnormal rise in the temperature of the mating surface [189]. Furthermore, the gearbox may suffer various types of fatal failures such as tooth surface wear or tooth breakage due to insufficient lubrication, incorrect assembly or overloads etc. [190], [191]. Broken teeth will produce the increased impulse, vibration and large noise and seriously influence other teeth, which shortens the service life of the gears, especially operating with high speed and heavy load [192]. Additionally, viscosity is an important measurement to evaluate lubricant viscosity. For high-speed operation and overloaded machinery, the lubricant oil with low viscosity is normally used for the gear lubrication because it can create and maintain a lubrication film with a proper thickness between two moving metal surfaces [193]. Its degradation and improper use will affect the production and lubrication of the oil film resulting in gear wear in a long-term operation of the machine. It is noticeable that various faults of the gearbox will have an important impact on its life and lead to significant economic losses. Therefore, it is significantly essential to monitor the condition of a gearbox and make fault diagnosis of gears with insufficient lubrication, oil degradation or tooth breakage.

To diagnose various faults in a gearbox, some researchers have investigated different methods based on the vibration, temperature, current and voltage, acoustic emission and sound signals in recent years [30], [194]–[196]. In 2004, Banks *et al.* [194] preliminarily gave an indicator to detect lubricant shortage in a closed splash lubricated system based on the vibration characteristics of the front and rear differentials extracted from the frequency band ranging from 15 kHz to 24 kHz. Lee *et al.* [195] compared the typical vibration with current signals on the detection of the gearbox lubrication. With the optimized classification model, the detection accuracy of gearbox lubrication levels has reached 96% by adopting the current signals. Some researchers [30], [196], [197] at our research centre investigated the effects of gearbox lubricant shortage with the oil temperature, vibration response, motor current and voltage. Features extracted from these signals are effective to indicate the oil shortage in a two-stage helical gearbox. Marques *et al.* [193], [198]–[200] investigated the power loss model for each component and concluded that lubricants with the same viscosity grade may have different power loss due to the base lubricant type and additives. Some researchers [201]–[204] and Vuuren *et al.* [202] achieved online estimation of lubricant viscosity of gearboxes with current and voltage signals. Hamel *et al.* [205] examined the

influence of oil film thickness on helical gear spall fault detection with acoustic emission signals and demonstrated that the lubricating conditions will seriously affect or even prevent the defect detection with acoustic emission signals. Sun *et al.* [206] investigated the influence of viscosity and overlap ratios on vibration characteristics of helical gears and found that low lubricant viscosity will cause a higher amplitude of vibration. For the tooth breakage fault detection, many researchers have built dynamic models to study the friction mechanism of the gear teeth and the effects of lubrication and broken teeth of the system and validated the with experiments [207]–[209].

However, the cost of transducers is expensive. Moreover, the installation of the transducers is complicated and inconvenient, or even impossible under harsh working environment, especially measure the oil temperature of gearboxes with thermocouples. Additionally, the measurement of oil or housing temperature is a single-point temperature measurement which has low stability and reliability for fault diagnosis of gearboxes operating under complicated conditions, such as variable loads. In contrast, a 2D thermal video revealing the temperature field distribution of the gearbox housing can provide more valuable information to achieve an accurate fault diagnosis and classification of machinery. Moreover, the temperature measurement based on thermal imaging has the advantages of non-contact, non-intrusive, high reliability and stability, which can benefit a robust condition monitoring. Touret *et al.* [210] made a review to introduce most of the researches on gearbox defect detection with the contact and non-contact (e.g. thermography) temperature sensors and stated that temperature has obvious potential to instead of vibration or acoustic on machine condition monitoring. Janssens *et al.* [211] have successfully classified eight different fault conditions of bearings based on features of standard deviation, the Gini coefficient and the Moment of Light extracted from thermal images to achieve an accuracy of 88.25%.

With the development of artificial intelligence, smart condition monitoring of machines has attracted more and more attention in the industry [114]. For example, Janssens *et al.* [212] successfully detected the bearing faults and predicted oil-level based on infrared thermal images with CNN. Li *et al.* [213] developed an IRT-CNN method to remotely monitor the condition of gearboxes online with infrared thermal images. However, it can be seen that the thermography technique is not widely used in the condition monitoring of gearboxes at present, which is probably because the 2D imaging signals have higher requirements on transmission bandwidth and storage space. In addition, the computational complexity of the CNN is particularly high, and the size of the captured 2D images is too large to slow down the calculation speed, which will affect the potential possibilities of intelligent online condition monitoring. Therefore, it is necessary to compress 2D images before regarding it as the input layer of the CNN. Compressive sensing as an algorithm that can fundamentally and effectively reduce the sampling frequency has been introduced in detail in the previous overview. Lu *et al.* [214] proposed a novel approach to achieve physics based compressive sensing with

fewer sensors and limited collected data, which was applied in manufacturing process monitoring with less costs and high efficiency.

In this research, CS will be applied for the 2D thermal image compression and the compressed images will be the input layer of the CNN to distinguish different faults and their severity of a two-stage helical gearboxes.

## 7.2 Compression Method

### 7.2.1 Convolutional Neural Network

Convolutional neural networks have been developed in recent years and have attracted widespread attention as an efficient identification method to recognise 2D graphics with characteristics of displacement, scaling and other forms of distortion invariance. CNN has become one of the research hotspots in many scientific fields because it avoids the complicated pre-processing of images and directly uses the original ones as inputs. It is different from other neural networks, CNN continuously extracts features from local to global through various filters to implement object recognition.

A typical CNN consists of an input layer, a hidden layer and an output layer as illustrated in Figure 7.1. Generally, 2D (grayscale images) or 3D (colour images) images can be processed as an input layer, and the out layer will generate classification labels according to the logic functions or softmax functions. The hidden layer is composed of one or multiple convolutional layers, pooling layers and fully connected layers, which will be introduced as follows.

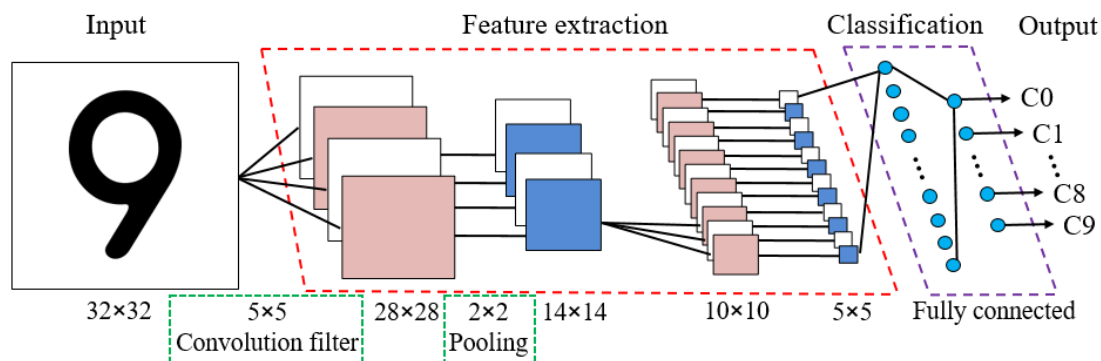


Figure 7.1 Structure of a CNN example

#### (1) Convolutional layer

The purpose of the convolution is to extract features from the input image. It can learn the characteristics of the image from a small piece of the input image and preserve the spatial relationship between pixels. The filter, a small matrix, is one of the most important parts in CNN. The convolution feature matrix can be produced by sliding the filter over the image and calculating the dot product. Different filters can detect different features, such as edges, curves and so on, from the same image. It means that the more filters are used, the more features are extracted. It is well known that

convolution features are determined by three parameters: depth, stride and zero-padding. The depth corresponds to the number of convolution filters; the stride is the number of pixels per sliding filter matrix; the zero-padding refers to padding at the edge of the input image matrix with zero values so that the edges of the input image matrix can be filtered. Equation (7-1) expresses the convolution result of the  $i$ -th filter kernel and the  $j$ -th local region in layer  $l$ .

$$y_i^{l+1}(j) = W_i^l * x^l(j) + b_i^l \quad (7-1)$$

where the notation  $*$  is the function of the dot product;  $W_i^l$  and  $b_i^l$  represent the weights and bias of the  $i$ -th filter kernel in layer  $l$ ;  $x^l(j)$  denotes the  $j$ -th local region in layer  $l$ ;  $y_i^{l+1}(j)$  is the convolution result of  $i$ -th filter kernel and the  $j$ -th local region in layer  $l$  and will be the input of layer  $l + 1$  after rectification.

A nonlinear activation function called rectified linear unit (ReLU) is used to maintain the nonlinear properties of the extracted convolution characteristics after convolution in the convolutional layer as described in Equation (7-2).

$$z_i^{l+1}(j) = \max(0, y_i^{l+1}(j)) \quad (7-2)$$

where  $y_i^{l+1}(j)$  is the convolution results;  $z_i^{l+1}(j)$  is the rectified convolution results and is also the final output of this convolutional layer. Other nonlinear functions, such as hyperbolic tangent or Sigmoid, can also implement nonlinear transformation of features, but ReLU is better in most cases because of its low computational complexity, fast training speed and more robustness to various disturbances. Figure 7.2 illustrates the calculation process of the convolutional layer. In this example, the size of the zero padding is set to 1, the filter size is  $3 \times 3$  and the stride is set to 2. It can be seen that  $W_0$  and  $W_1$  are two filters which means the depth is 2 here. The bias of these two corresponding filters are 0 and 1, respectively. The convolution results are rectified with ReLU function to get the output matrix. Therefore, the output matrix with the size of  $3 \times 3 \times 2$  will be the input of next layer for this example.

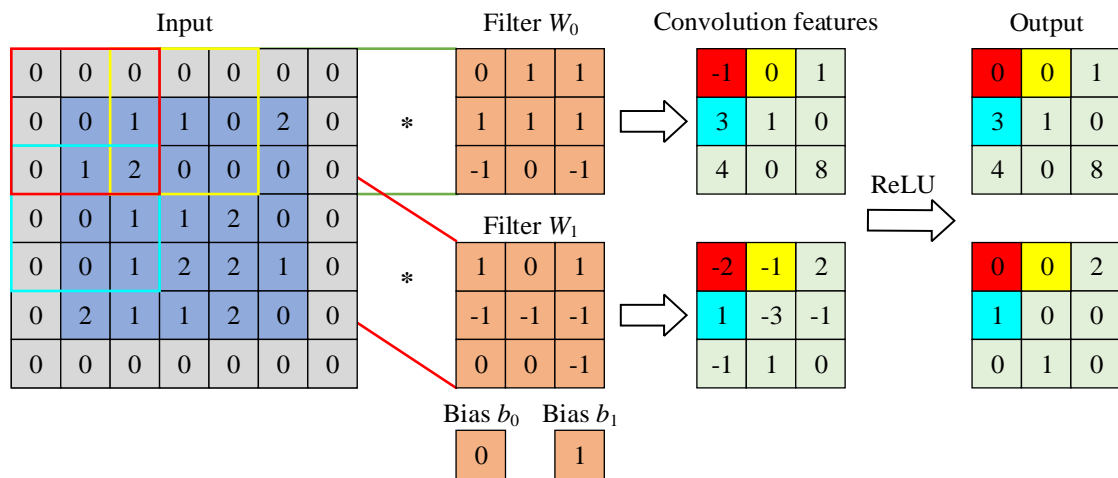


Figure 7.2 An example of the convolutional layer

## (2) Pooling layer

To effectively reduce the computational complexity, another efficient tool in CNN is pooling. Pooling is a downsampling or subsampling process, which only preserves important information and reduces pixel information to reduce the dimension of the feature map. There are several common methods to perform pooling process, such as maximization, averaging, summing and so on. In general, max pooling is the most common and effective method. It retains the maximum value for each sub-region which means keeping the best match result for this area because the larger the value, the better the filter and input image match. Hence, the addition of the pooling layer can greatly reduce the cost of the computation and ensure the validity of the extracted features.

$$q_i^{l+1}(j) = \max_{(j-1)w+1 \leq t \leq jw} (p_i^l(t)) \quad (7-3)$$

where  $p_i^l(t)$  is the value of the  $t$ -th neuron,  $(j-1)w+1 \leq t \leq jw$ , in the  $i$ -th frame of layer  $l$ ;  $w$  represents the width of the pooling region and  $q_i^{l+1}(j)$  denotes the value of the neuron in layer  $l+1$ . An example of max pooling with the filter size of  $2 \times 2$  and the stride of 2 is depicted in Figure 7.3.

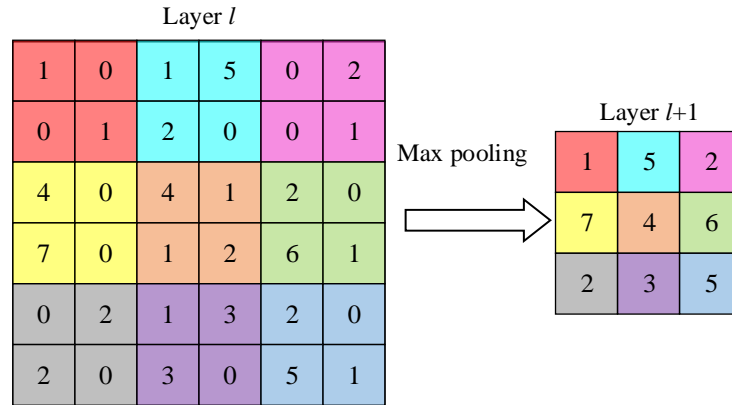


Figure 7.3 An example of the pooling layer

## (3) Fully connected layer

The convolutional layers and the pooling layers extract local features of the object, and the fully connected layer is to reassemble the obtained local features into a global feature through the weight matrix. The weight matrix is multiplied with the extracted local features to map the  $N$  features into  $K$  real values. The function softmax normalizes these  $K$  real values to  $K$  probabilities with the sum of 1. The output values can be calculated as

$$u_j = s_j \cdot v + b_j = s_{j1} \cdot v_1 + s_{j2} \cdot v_2 + \cdots s_{jN} \cdot v_N + b_j \quad (7-4)$$

where  $s_j$  is the weight of the  $j$ -th category feature;  $v$  and  $u_j$  are the input and output values, respectively;  $b_j$  is the bias. The output probabilities  $\hat{h}_j$  can be expressed as

$$\hat{h}_j = \text{softmax}(u_j) = \frac{e^{u_j}}{\sum_{k=1}^K e^{u_k}} \quad (7-5)$$

The output probability is the result of the intelligent classification of the CNN, e.g. the final result of the output layer.

The CNN, as an intelligent classification network, has good fault tolerance, parallel processing capabilities and self-learning abilities. It is especially suitable for image processing, such as 2D and 3D images with motion, scaling and other forms of distortion invariance, to automatically extract features from the images. It extracts features through training data instead of manual extraction. Local filters can be shared in CNN to reduce the number of network parameters and the network complexity. This feature makes CNN much closer to the real biological neural network and a very popular intelligent classification tool.

### 7.2.2 Proposed CS-based CNN

Although the CNN avoids complex pre-processing of images, the original images are too large to slow down the calculation speed because the captured images are redundant and compressible, such as background. Therefore, the captured thermal images can be compressed with CS before entering the input layer. In future, a CS-based infrared thermal camera has the potential to be developed and applied according to the principle of single-pixel compressive digital camera designed by the researchers at the Rice University.

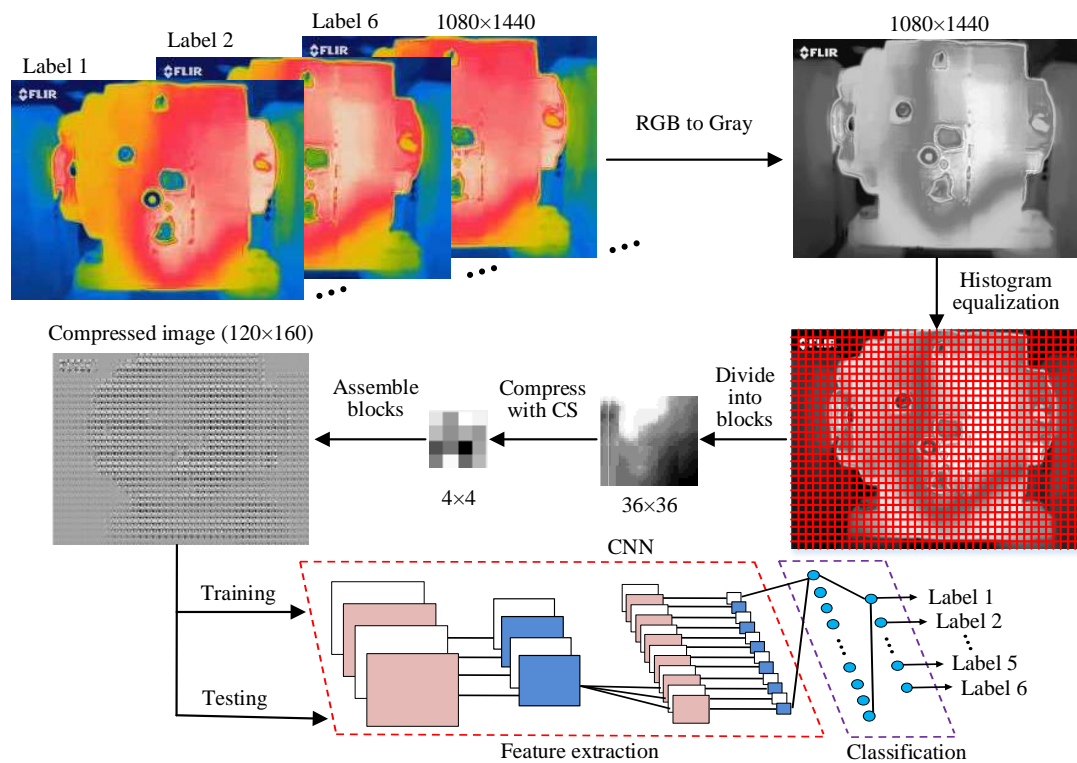


Figure 7.4 Flow chart of the proposed CS-based CNN method

Figure 7.4 gives a detailed and comprehensive depiction of the flow chart of the proposed CS-based CNN algorithm for the fault classification of gearboxes with 2D thermal images. The process can be detailed as the following steps:

**Step 1:** convert the captured thermal images from RGB to grayscale images to reduce the dimension of the images from 3D to 2D and label them according to different cases;

**Step 2:** calculate the histogram equalization of each grayscale image to enhance contrast and strengthen the edge information of the image;

**Step 3:** divide the equalized images into blocks with the size of  $36 \times 36$  and further compress the blocks to  $4 \times 4$  by CS, so that the compression ratio is 81 calculated based on Equation (5-11);

**Step 4:** assemble all compressed blocks for each image to reconstruct a compressed image based on original image size and the division rules;

**Step 5:** train a CNN with the parameters listed in Table 7.1 and use part of the reconstructed compressed grayscale 2D images to form the input layer;

**Step 6:** validate the trained CNN with the rest of the compressed images and classify different fault types to generate the classification accuracy.

Table 7.1 Detailed parameters of the applied CNN layers

Layer	Name	Filter Size	Filter Number	Stride
1	Input layer ( $160 \times 120$ pixels)			
2	Convolutional layer+Batch normalization+ReLU	$3 \times 3$	8	[1, 1]
3	Pooling layer	$2 \times 2$		[2, 2]
4	Convolutional layer+Batch normalization +ReLU	$3 \times 3$	16	[1, 1]
5	Pooling layer	$2 \times 2$		[2, 2]
6	Convolutional layer+Batch normalization +ReLU	$3 \times 3$	32	[1, 1]
7	Fully connected layer			
8	Output layer (Softmax layer+Classification layer)			

## 7.3 Experiments

### 7.3.1 Test Facilities

In order to validate the availability of the proposed intelligent CS-based CNN method to detect and diagnose different faults in a gearbox, a series of experiments were carried out on a two-stage helical gearbox test rig. Figure 7.5 illustrates the schematic diagram of the test rig system, which is mainly comprised of a 15 kW induction motor with the delta connection, a DC load generator, two helical gearboxes installed back-to-back and a sensorless variable speed drive control system for determining the speed and load of the test rig. A thermal imaging infrared camera, FLIR ONE Pro with the key



specifications shown in Table 6.2, was placed in front of the gearbox 1 (abbreviated as GB1) at a distance of 25cm from the gearbox housing surface. As a comparison, the oil temperature of the gearbox was measured by a K-type thermocouple installed as shown in Figure 7.5. The data was collected by the data acquisition system YE6232B and a computer. Table 7.2 lists the specifications of the most test facilities applied in this test.

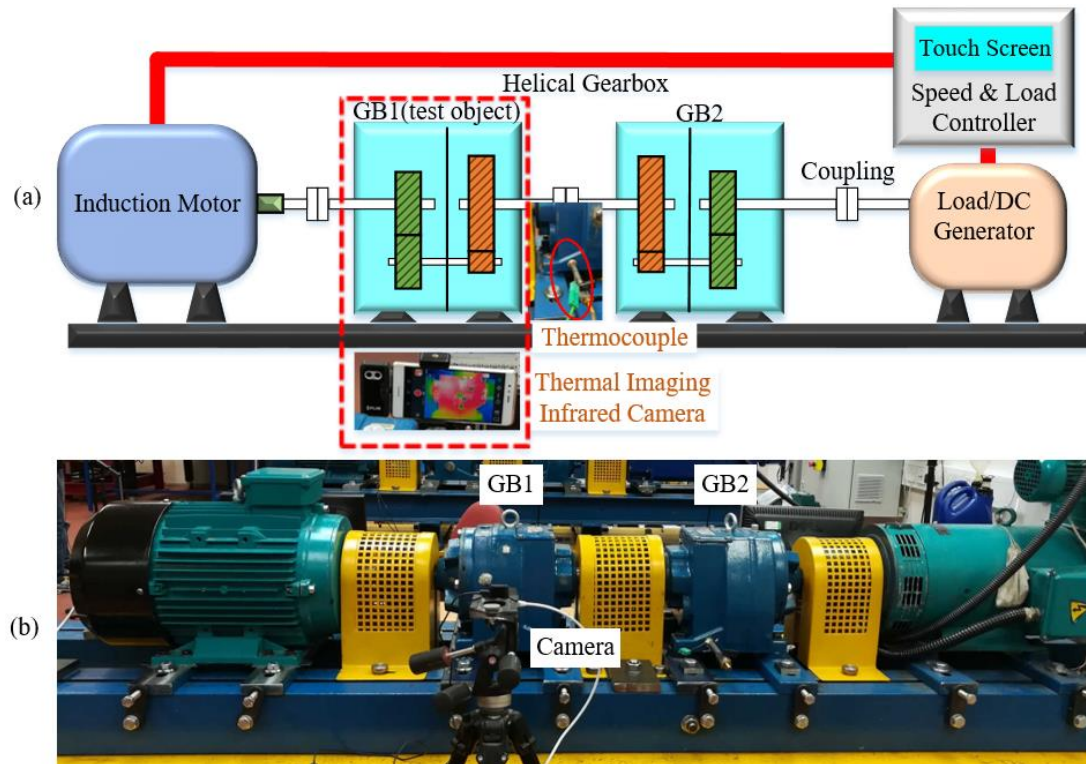


Figure 7.5 Two-stage helical gearbox test rig: (a) schematic diagram and (b) actual test rig

Table 7.2 Specifications of the test facilities

Facilities	Parameters	Values
AC Motor (Brook Crompton)	Motor Type	T-DA160LA
	Frequency	50 Hz
	Voltage	280-415 $\Delta$ /660-720 Y (V)
	Current	29.63-27.13/17.06-15.64 (A)
	Power	15 kW
	Rated Speed	1460 rpm
Gearbox (Radicon Company)	Model No.	M07223.6BRC-1
	Ratio	3.678/1
	Oil Type	EP 320
	Oil Volume	2600 mL
DAQ YE6232B	Channels	16
	A/D Bits	24
	Input Mode	V/IEPE



	IEPE Power Supply	4 mA/+24 VDC
	Signal Input Range	$\leq \pm 10$ VP
	Signal Frequency Range	DC-30 kHz
	Sample Rate	Max. 96 kHz/CH, parallel
K-type Thermocouple	Maximum Temperature	+1100 °C
	Probe Diameter/Length	3 mm/150 mm
	Probe Material	Stainless steel

### 7.3.2 Fault Simulation

#### (1) Oil shortage

In the harsh working environments, oil loss due to leakage and evaporation is inevitable, especially the applications in the heavy-duty machinery. Therefore, the gearbox fault of oil shortage was simulated with the oil level referenced in [30]. The standard oil quantity of this gearbox is 2600 mL recommended by the manufacturer and this case is regarded as the healthy condition or baseline, denoted as BL-2600mL in Figure 7.6. Then, 600 mL oil was taken out from GB1, which is the first faulty case of the oil shortage and denoted as OS-2000mL. Moreover, the second faulty case of the oil shortage was simulated by further removing 500 mL oil to make the quantity to 1500mL, denoted as OS-1500mL. The oil level and gear immersion states for each case are separately illustrated from the inside, axial and side views in Figure 7.6, which shows that the input gear and output gear are less lubricated as the oil levels are getting lower as the oil is removed. The lubricant used in this study is the MILLGEAR 320 EP with the typical characteristics listed in Table 7.3.

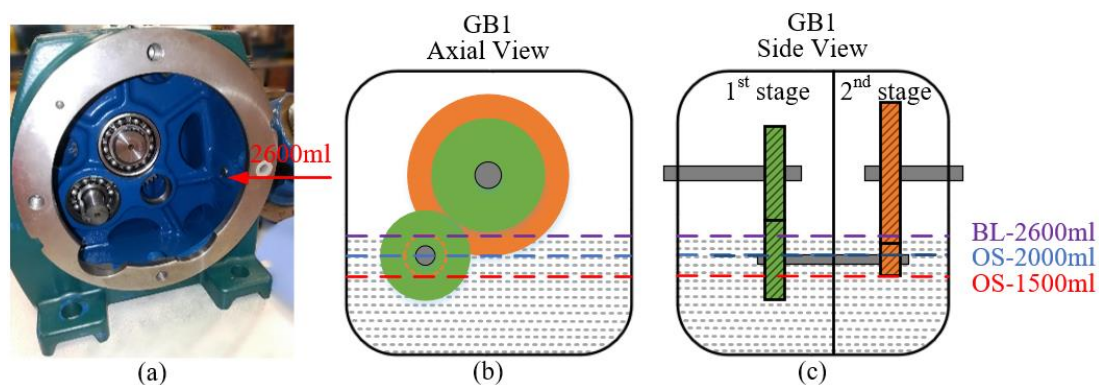


Figure 7.6 Schematic diagram of lubricant levels: (a) inside view, (b) axial view, and (c) side view

#### (2) Oil degradation

Generally, there are several reasons leading to low oil viscosity. The first one is adding a lower viscosity fluid than the recommended requirement, i.e. improper use of the lubricant. The second one is accidentally mixed with non-lubricants like solvents and diesel fuel into the lubricant. The third one is the loss or shearing down of the viscosity-

index improver which is a well-known additive. When the lubricant operates at an extreme high environment, the high temperature causes a decrease in viscosity. Therefore, the viscosity of the lubricant will decrease as the oil age increases, which is a process of oil degradation. As a result, the lubricant with low viscosity, MILLGEAR 100 EP, was used to simulate the oil degradation in this test. Both the specifications of MILLGEAR 320 EP and MILLGEAR 100 EP are introduced and compared in Table 7.3.

Table 7.3 Specifications of MILLGEAR 320 EP and MILLGEAR 100 EP

<b>MILLGEAR EP</b>	<b>320</b>	<b>100</b>
Specific Gravity @ 15°C	0.901	0.885
Kinematic Viscosity @ 100°C/cSt	23.5	10.9
Kinematic Viscosity @ 40°C/cSt	320	100
Viscosity Index	92	93
Pour Point (°C)	-9	-9
Flash Point (°C)	200	200

### (3) Tooth breakage

Gear tooth breakage is caused by overload, fatigue or chemical attack. Gears work in pairs with the drive gear connected to the power source and the driven gear meshed with the drive gear. In this study, 50% and 100% of the tooth face were artificially removed from two driven gears in the width direction, respectively, which is depicted in Figure 7.7.

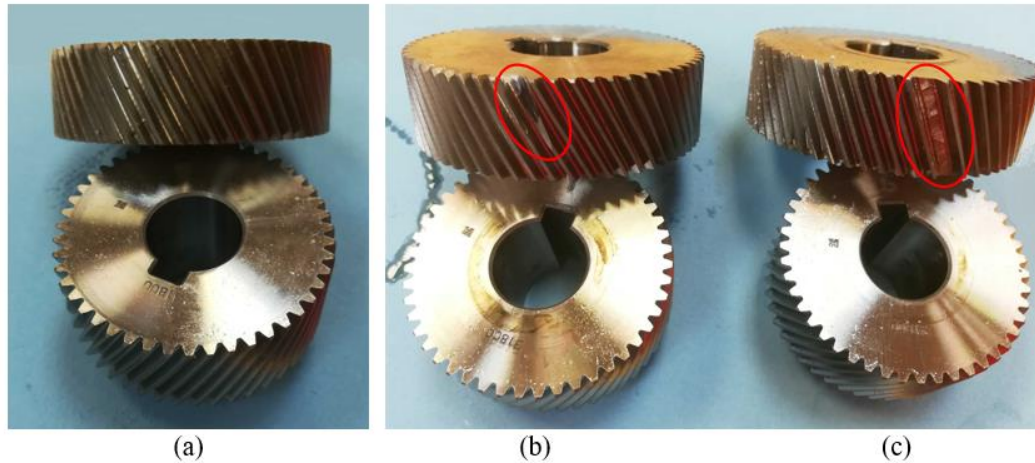


Figure 7.7 Tooth breakage fault simulation: (a) healthy pair, (b) 50% tooth breakage, and (c) 100% tooth breakage

### 7.3.3 Test Procedure

According to the description of the fault simulation, it is obvious that six different cases should be carried out with the two-stage helical gearbox test rig: the healthy case with 2600 mL MILLGEAR 320 EP lubricant and the healthy gears (denoted as BL for the

result display), two oil shortage cases (denoted as OS2000 and OS1500), a low-viscosity case (written as VIS100) and two tooth breakage cases (marked as TB050 and TB100). All these tests were performed at a constant speed (the full speed of the driving motor, 1475 rpm) and four different loads (0%, 30%, 70% and 100% load of the DC motor). The experimental operation process for each case is explained as follows.

**Step 1:** warm up the test rig with 75% of the full speed and 80% load, and observe the oil temperature of GB1 measured by the K-type thermocouple until the temperature reaches 32 °C (the room temperature was about 21 °C);

**Step 2:** set 100% of the full speed and 0% load, run 20 minutes and begin to collect the oil temperature and record the temperature filed distribution of the GB1 house with the thermal video for 4 minutes. One minute after the recording, the operation will be changed to 100% of the full speed and 30% load;

**Step 3:** repeat **Step 2** with the entire process taking 25 minutes in total and record the data during this period. Then, repeat operation and recording of the other two sets of loads.

The time it takes to run a case is 100 minutes plus the warm-up time period for each case as shown in Figure 7.8. All cases were performed based on this timeline recording.

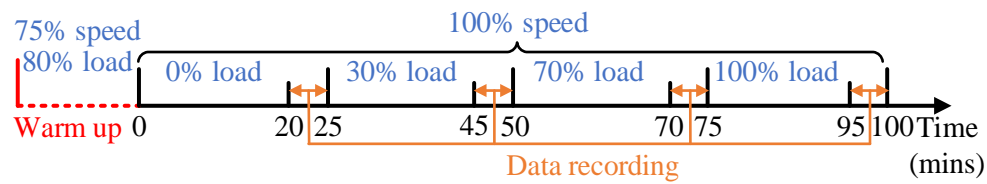


Figure 7.8 Timeline recording of the experimental process

Table 7.4 Dataset grouping for CNN training and testing

Labels	Training Group		Testing Group	
	Speed	Load	Speed	Load
BL	100%	0% (800 frames), 100% (800 frames)	100%	30% (800 frames), 70% (800 frames)
OS2000				
OS1500				
VIS100				
TB050				
TB100				

Since the camera storage speed is less than 8.7 frames/s and the frame rate is not constant for every case, the first 800 frames are selected as the training or testing dataset for each set (24 sets, 6 cases×4 loads). Table 7.4 lists the 24 sets of data divided into the training and the testing groups for the CNN.

### 7.3.4 Results and Discussion

To illustrate the superiority of thermal imaging based CM, the results will be delivered and discussed in terms of oil temperature of GB1 and the temperature field distribution of the GB1 housing surface.

#### 7.3.4.1 Oil Temperature based CM

The oil temperature of GB1 was measured and observed in real time to monitor the condition of the machine running. The variation in oil temperature may preliminarily indicate the gearbox failure. Figure 7.9 depicts the average oil temperature in GB1 for different cases, including BL, OS2000, OS1500, VIS100, TB050 and TB100 under various loads. It is apparent that the oil temperature has an increasing trend with the rise of the load. Meanwhile, the temperature has a significant decrease when the quantity of the lubricant gradually reduces, and the viscosity of the lubricant is lower than the recommended one. The temperature drop of the broken tooth fault is not obvious, especially for the case of TB100, the temperature cannot be distinguished from the case of BL. The case of TB050 has a little clear difference in the upward trend compared with the case of BL.

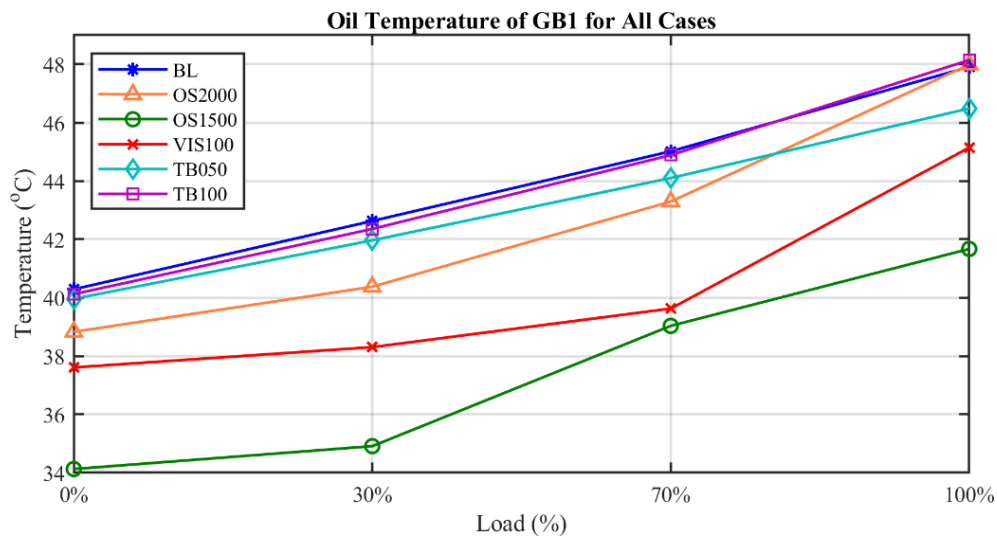


Figure 7.9 Oil temperatures of the GB1 for different cases under various loads

However, the oil temperature, as 1D information, can only be used to approximately indicate the variation either in loads or different severity of the failure. When a gearbox is operated under variable loads with failure, it is difficult to determine the cause of the variation in oil temperature by this 1D temperature array, such as oil temperature or housing temperature. As a mature technology, thermal imaging has been considered in the field of machine condition monitoring by capturing temperature field distribution of objects. The next subsection will display and discuss the results and effects of oil shortage, oil degradation and tooth breakage in the gearbox detected with the temperature field distribution of the gearbox housing surface.



### 7.3.4.2 Temperature Field Distribution based CM

#### (1) RGB images

In order to visually observe the temperature field distribution, the last image is selected from each video recorded under variable loads and various oil levels as shown in Figure 7.11 and Figure 7.12, respectively. Figure 7.10 depicts the surface topography of the GB1 house as a comparison of captured thermal images. It is noted that the recorded images are flipped for the actual position.



Figure 7.10 Surface topography of the GB1 house

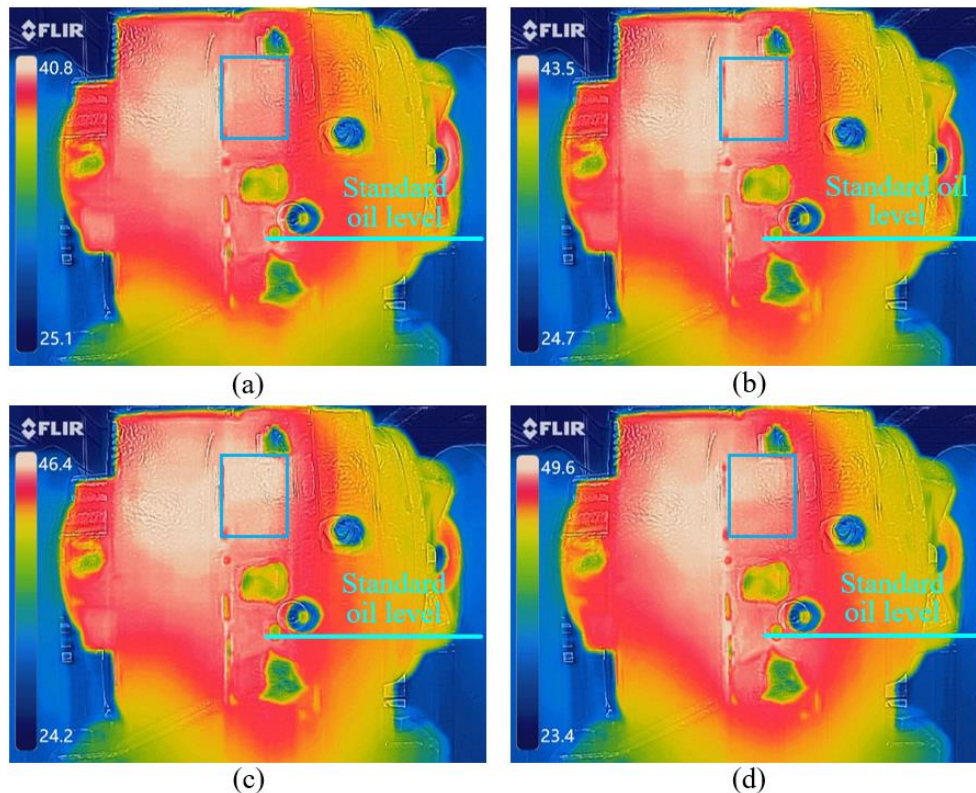


Figure 7.11 Images for BL under different loads: (a) 0%, (b) 30%, (c) 70%, and (d) 100%

For the case with the standard oil level in GB1, i.e. BL, four images divided from the recorded videos for four different loads of 0%, 30%, 70% and 100% are shown in Figure 7.11 (a) to (d), respectively. The oil level recommended by the manufacturer is

marked with the line in each figure. The temperature field distributions of these four different loads looks nearly same, only subtle differences exhibits as observed in the blue boxes. The high-temperature areas mainly concentrate in the upper middle part of the gearbox, where the second-stage gears locate. The temperature rises as load increases, which is consistent with the previous results made by the oil temperature. Additionally, the hottest place does not change a lot, but the high-temperature region spread more and more along with the increasing loads. However, these differences in the thermal images provide useful information to differentiate different loads.

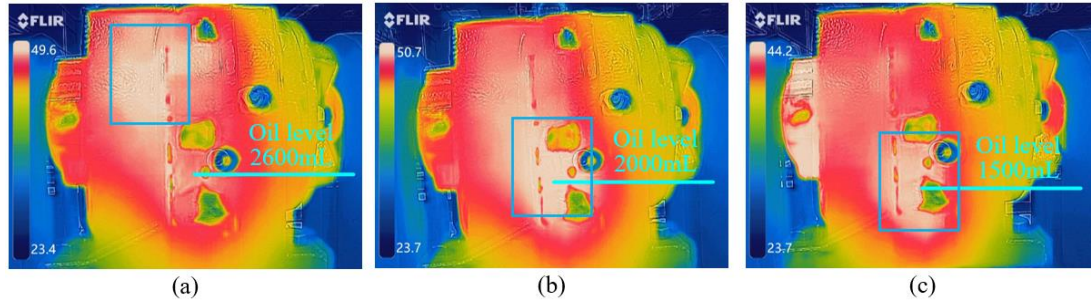


Figure 7.12 Images for different oil level under 100% loads: (a) BL (b) OS2000, and (c) OS1500

Figure 7.12 depicts the thermal images of the tested gearbox operating under three different cases with different oil levels (the corresponding oil level is marked with lines) with 100% load. It can be clearly noticed that the high-temperature area shifts from the upper to the lower of the gearbox housing. Furthermore, the heat produced by the friction of components also conducted along the shaft when the lubricant particularly starves, so the temperature of the shaft and bearing connecting with the GB2 is also very high. In addition, the highest temperature under a starved lubrication condition shown in the thermal imaging is lower than that under the normal oil level. Figure 7.13 displays other three cases compared with BL, but it is difficult to visualize the differences except the highest temperature of the tooth breakage cases.



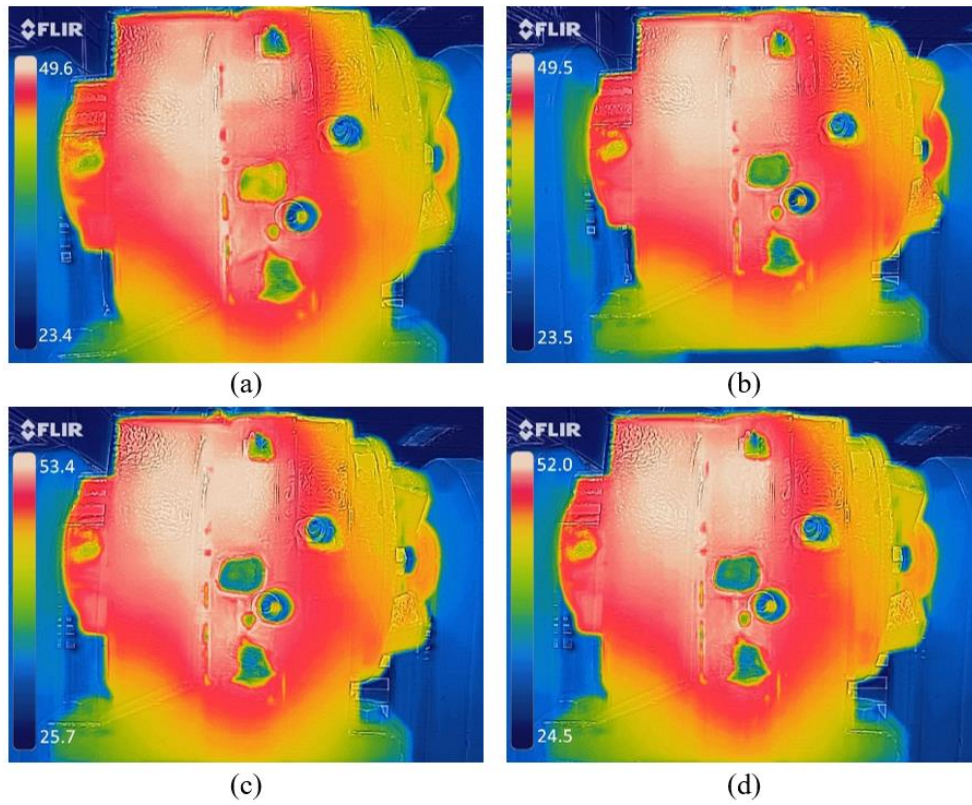


Figure 7.13 Images for different cases under 100% loads: (a) BL, (b) VIS100, (c) TB050, and (d) TB100

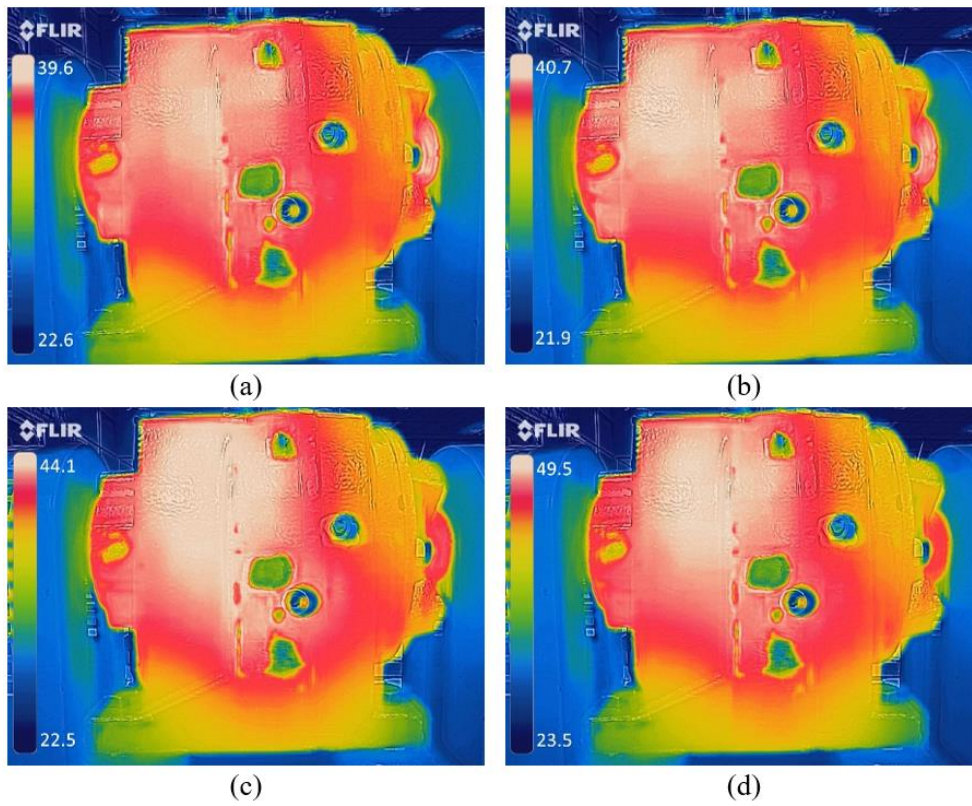


Figure 7.14 Images for VIS100 under different loads: (a) 0%, (b) 30%, (c) 70%, and (d) 100%

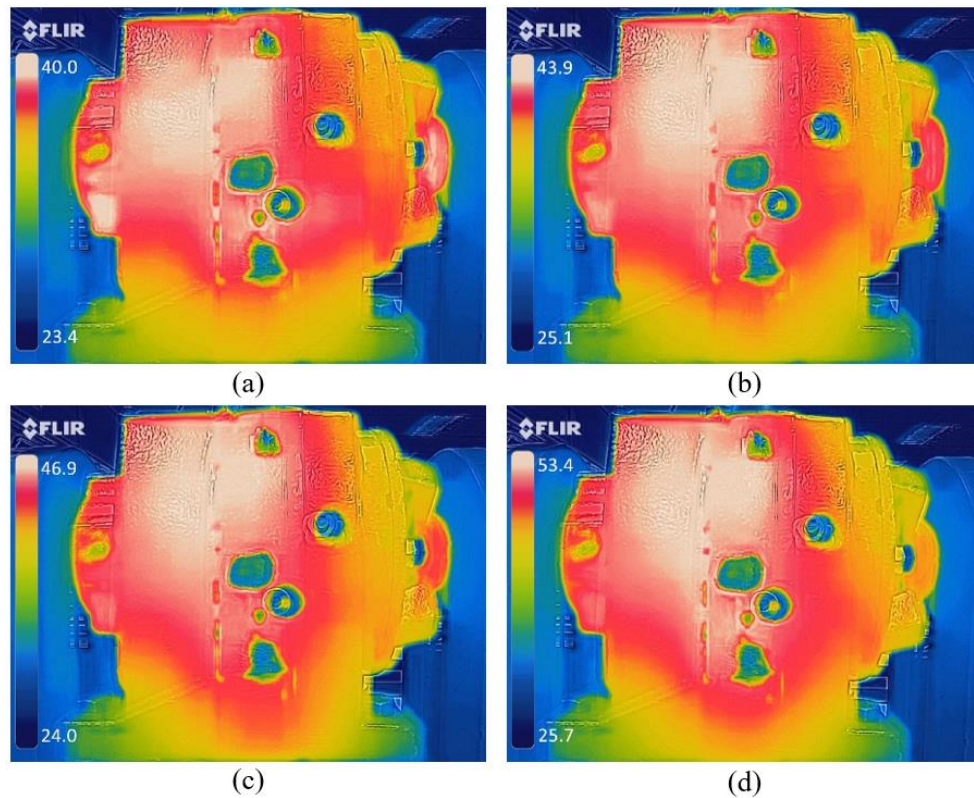


Figure 7.15 Images for TB050 under different loads: (a) 0%, (b) 30%, (c) 70%, and (d) 100%

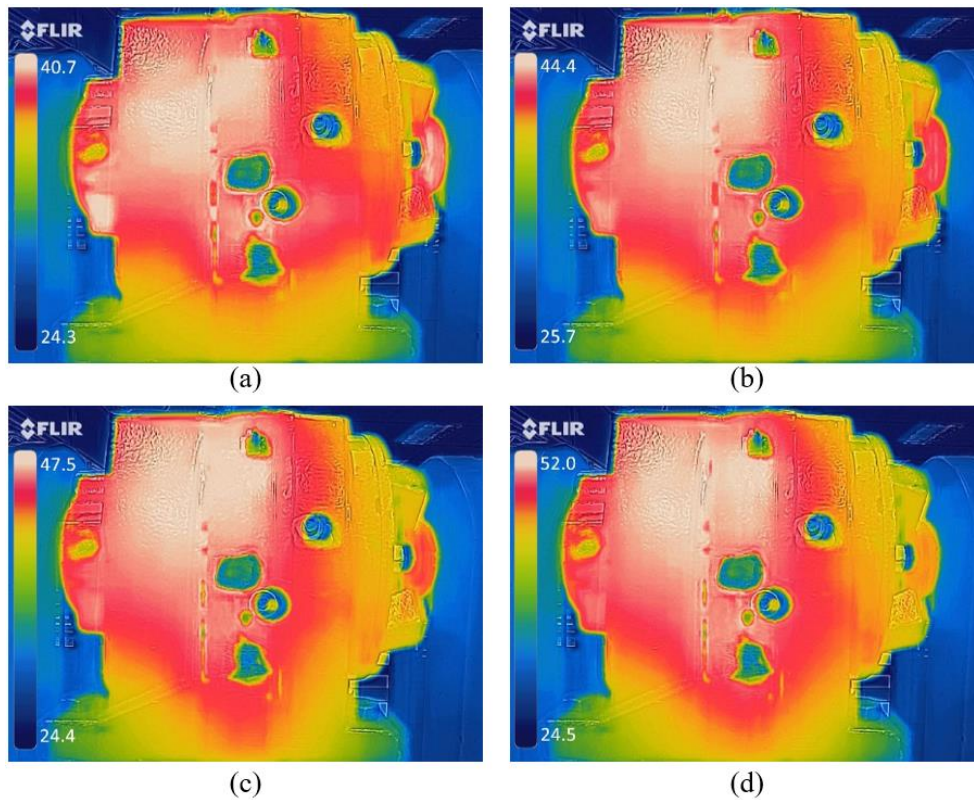


Figure 7.16 Images for TB100 under different loads: (a) 0%, (b) 30%, (c) 70%, and (d) 100%

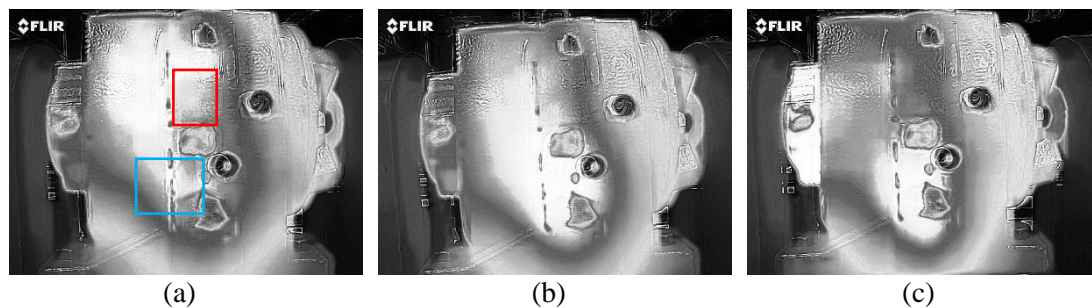
Four images selected from VIS100, TB050 and TB100 cases under different loads are illustrated in Figure 7.14 to Figure 7.16, respectively. All these three cases have similar



trend. The highest temperature in this shooting area rises as the load increases. In the meantime, the high-temperature area extends and changes slightly, but it is not easy to detect and compare with human observation. Moreover, the captured colour images have three channels of R, G and B with the size of 1080×1440 (height×width), which means they need a big storage space and will result in complex computations. Therefore, it is necessary to significantly reduce the size of the images before inputting into the CNN.

## (2) Equalized grayscale images

The first compression step is to reduce the number of channels. All selected images for different cases are converted into grayscale images from RGB images to reduce the channel number from 3 to 2. The images can be represented with 2D matrices instead of three-dimensional matrices. In order to enhance contrast in images especially the edge information, a well-known image processing technique, histogram equalization, was applied in this study. The equalized grayscale images for six different cases are selected and shown in Figure 7.17. It is apparent that the high-temperature areas of the cases of OS2000 and OS1500 in Figure 7.17 (b) and (c) have a significantly obvious downward shift as the oil level gradually decreases. Because the oil absorbs a lot of heat generated by friction so that the oil temperature rises. In addition, the temperature at the location of the bearing that connects the shaft to GB2 also rises due to the heavy load the insufficient lubrication under 100% load. The case using low-viscosity lubricants to simulate oil degradation, the high-temperature area is concentrated in the position of the secondary gears in the gearbox. When the viscosity of the lubricant is low, the thickness of the formed oil film is too thin to perform good lubrication, resulting in an increase of the heat generated by the friction of the tooth surfaces. Comparing the contents in the blue box of Figure 7.17 (a) and (d), this low viscosity case, VIS100, can also be distinguished from the case of BL. For the tooth breakage cases, their temperature field distributions have no much differences from the case of BL. Only the temperature field distributions at the position of the first stage gears, i.e. faulty gears, are slightly expanded as shown in the red rectangles of Figure 7.17 (a), (e) and (f).



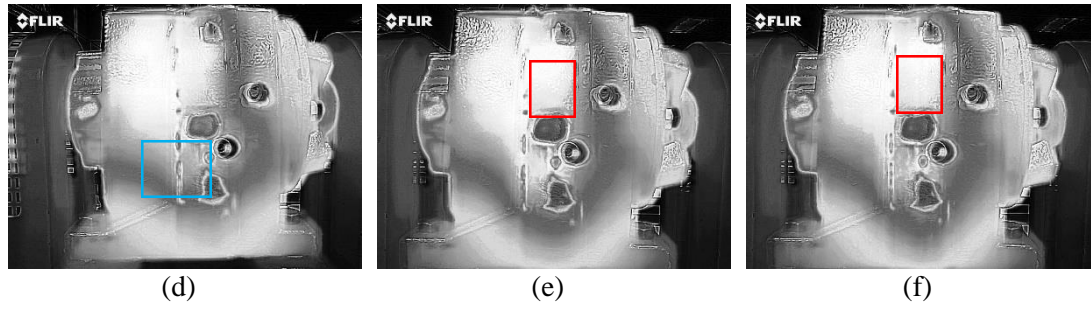


Figure 7.17 Equalized grayscale images for different cases under 100% loads: (a) BL, (b) OS2000, (c) OS1500, (d) VIS100, (e) TB050, and (f) TB100

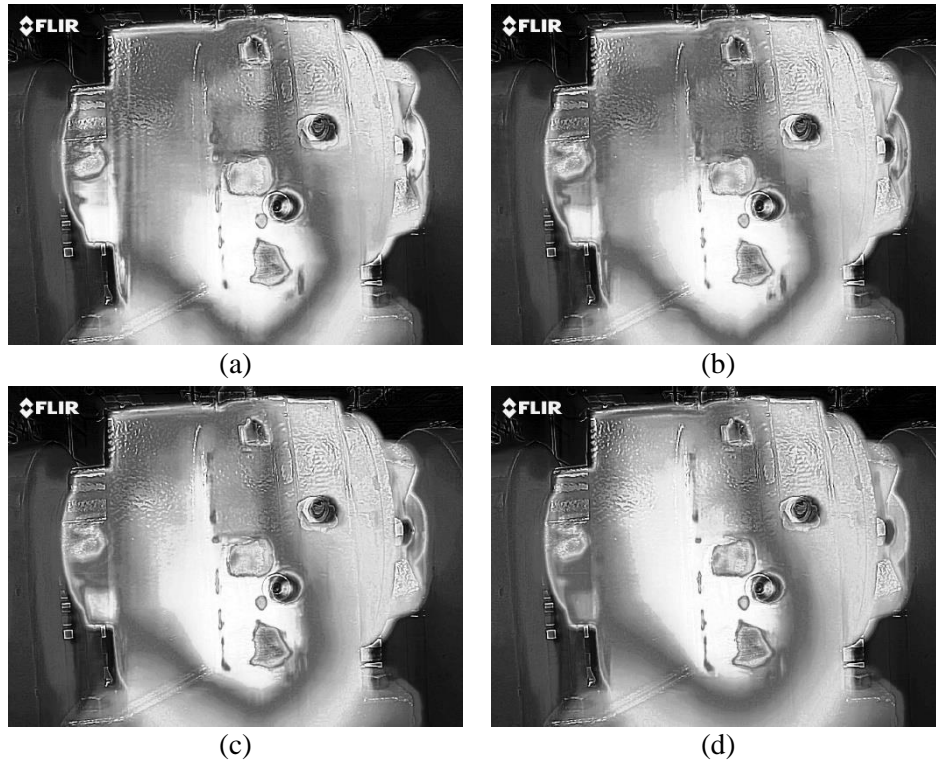


Figure 7.18 Equalized grayscale images for OS2000 under different loads: (a) 0%, (b) 30%, (c) 70%, and (d) 100%

Figure 7.18 illustrates the equalized grayscale images for the case of OS2000 under four different loads. When the load is light, the temperature field distribution is very similar to the case of OS1500, but the high-temperature area is a little big and the temperature at the bearing position is lower. As the load increases, the high-temperature area gradually expands and the boundary is much clearer. As a result, two situations with one light load and one heavy load as a group for one case are used for training the CNN in this study. But before training, the size of the images should be further reduced with the CS method.

### (3) CS based images

CS will work well because there is some redundant and unhelpful information in the images, such as background. The second compression step is to apply CS to achieve random sampling so that the original image has the potential to be reconstructed with

high accuracy. As the flow chart described in Figure 7.4, the equalized grayscale images with the size of  $1080 \times 1440$  pixels are divided into 1200 small blocks with the size of  $36 \times 36$  pixels. Each block is compressed to  $4 \times 4$  pixels with CS. Then, the compressed blocks are reassembled based on the division rule. The size of the reassembled image will be compressed to  $120 \times 160$  pixels with the compression ratio of 81.

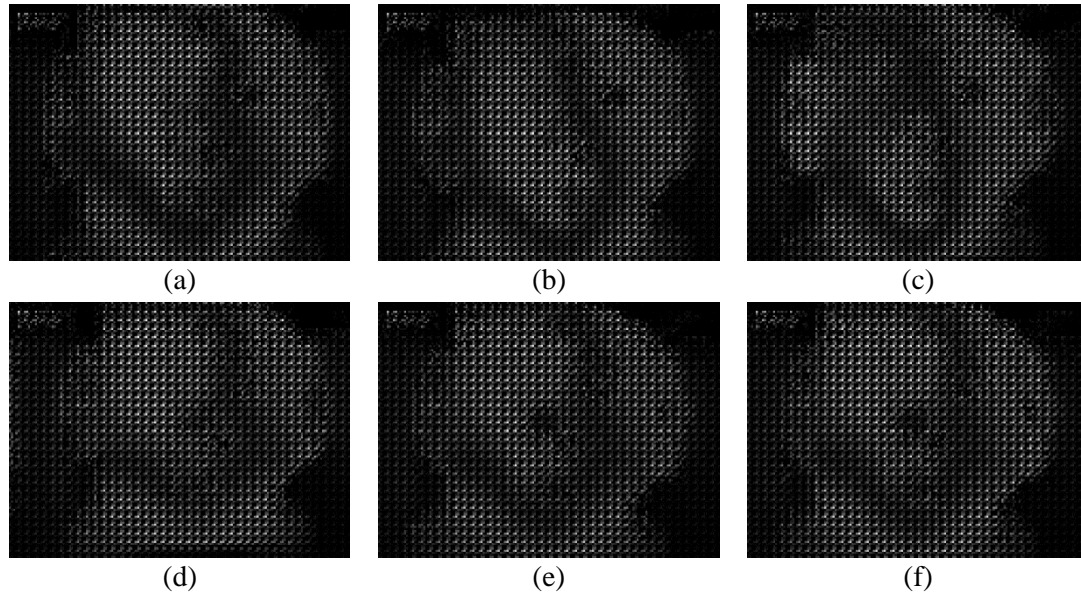


Figure 7.19 CS based images for different cases under 100% loads: (a) BL, (b) OS2000, (c) OS1500, (d) VIS100, (e) TB050, and (f) TB100

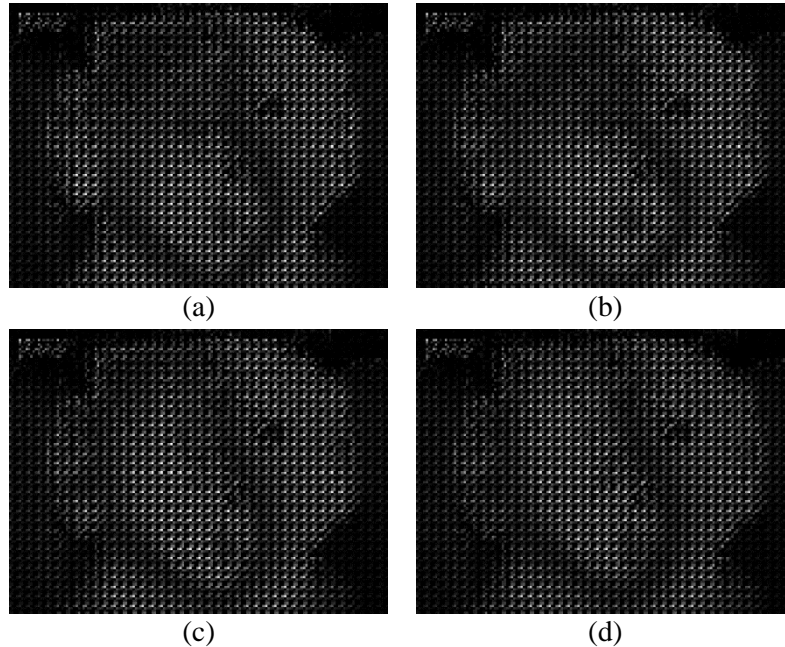


Figure 7.20 CS based images for OS2000 under different loads: (a) 0%, (b) 30%, (c) 70%, and (d) 100%

Figure 7.19 and Figure 7.20 display the CS based images for different cases under 100% load and OS2000 under different loads, respectively. Since the edge information of the original images are very clear, it is possible to roughly observe the change of the edge region for different cases and loads after random sampling. However, the background

information is all nearly black now. This phenomenon indicates that the effective information for fault diagnosis is preserved during the compression process, while the amount of data used for fault detection is compression by a large compression ratio (compress  $3 \times 81$  times with two compression steps in total). These CS based images will be the input of the CNN. In the CNN, the features will be extracted and trained to obtain a net for the fault classification.

#### (4) Intelligent CNN classification

It has been discussed that the thermal images provide valuable information extracted from the 2D temperature field distribution to distinguish different loads for various cases from Figure 7.11 to Figure 7.20. Some cases can be clearly detected with visualization based on the changes in temperature field distribution. But sometimes it is affected by the variable load. Therefore, it is necessary to use an intelligent classification method to detect various types of faults instead of observation. In this study, all images are divided into two groups as listed in Table 7.4: the training group and the testing group. The training process acts in terms of the layers with parameters as shown in Table 7.1. The obtained characteristics with the convolution filters in three convolutional layers are displayed in Figure 7.21 to Figure 7.23, respectively. These images in the first convolutional layer mostly contain edge and brightness features as shown in Figure 7.21. Some more complex features are constructed in the latter two convolutional layers as illustrated in Figure 7.22 and Figure 7.23. Each image represents the result of convolution of a convolution filter with the input image. These convolution filters extract features of various cases from different angles and brightness.

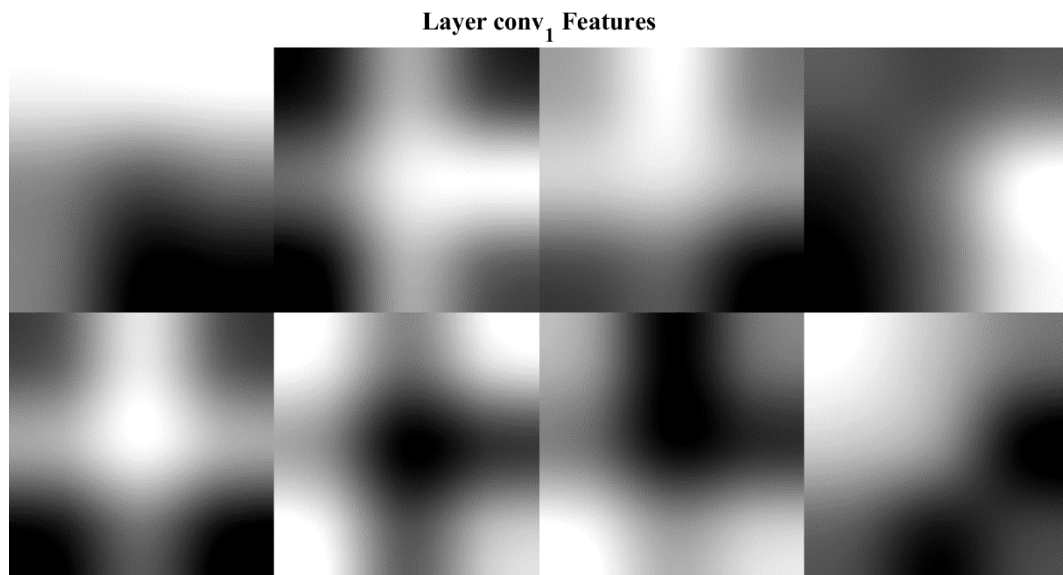


Figure 7.21 Features of the first convolutional layer

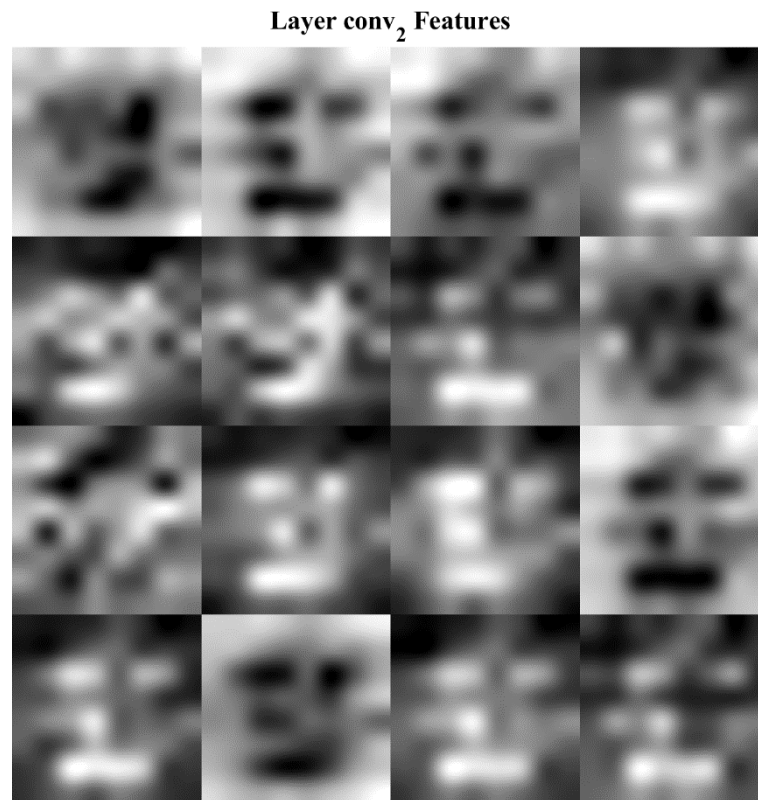


Figure 7.22 Features of the second convolutional layer

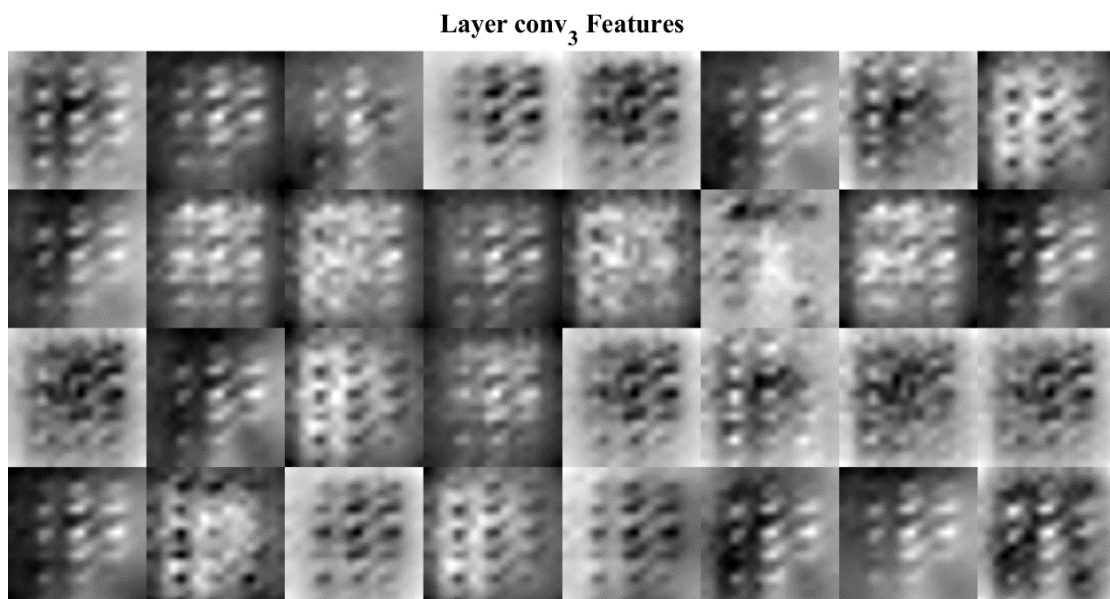


Figure 7.23 Features of the third convolutional layer

Only one fully connected layer is used in this network. The fully connected layer is set towards the end of the trained CNN and before the softmax function application. It combines the features learned in the previous layers in a high-level to get more accurate features to represent various cases different from other cases. This layer is the only and final fully connected layer and the images generated in this layer correspond to six different cases: BL, OS2000, OS1500, VIS100, TB050 and TB100 as shown in Figure



7.24, respectively. Although the features constructed in the fully connected layer are not clearly observed, the difference in outline and edge information is obvious.

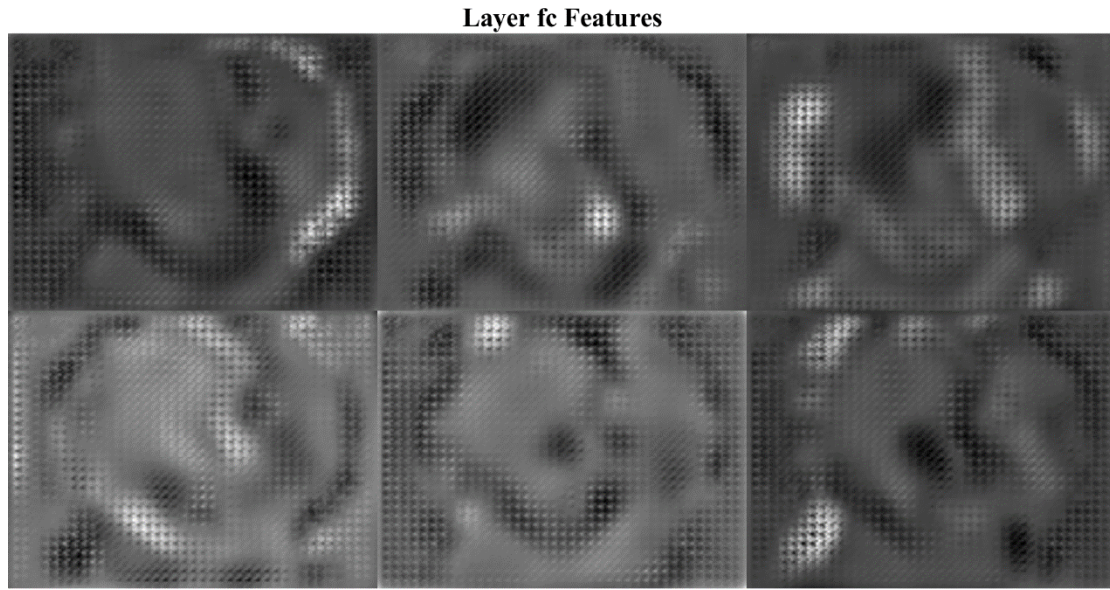


Figure 7.24 Features of the fully connected layer

The network training process is monitored as shown in Figure 7.25. The top figure describes the relationship between classification accuracy and the number of iterations. The light blue solid line and the black dotted line represent the training and validation results, respectively. The blue solid line draws the smoothed training process. The bottom graph illustrates the relationship between loss and the number of iterations in the training process. The cross entropy loss is the loss function used in the final classification layer which is expressed in Equation (7-6).

$$loss = -\sum_{i=1}^P \sum_{j=1}^Q t_{ij} \ln(y_{ij}) \quad (7-6)$$

where  $P$  and  $Q$  are the number of samples and classes, respectively;  $t_{ij}$  is the indicator that the  $i$ th sample belongs the  $j$ th class;  $y_{ij}$  is the result of the softmax function in this study, i.e. the probability that the  $i$ th input of the network is associated with the  $j$ th class.

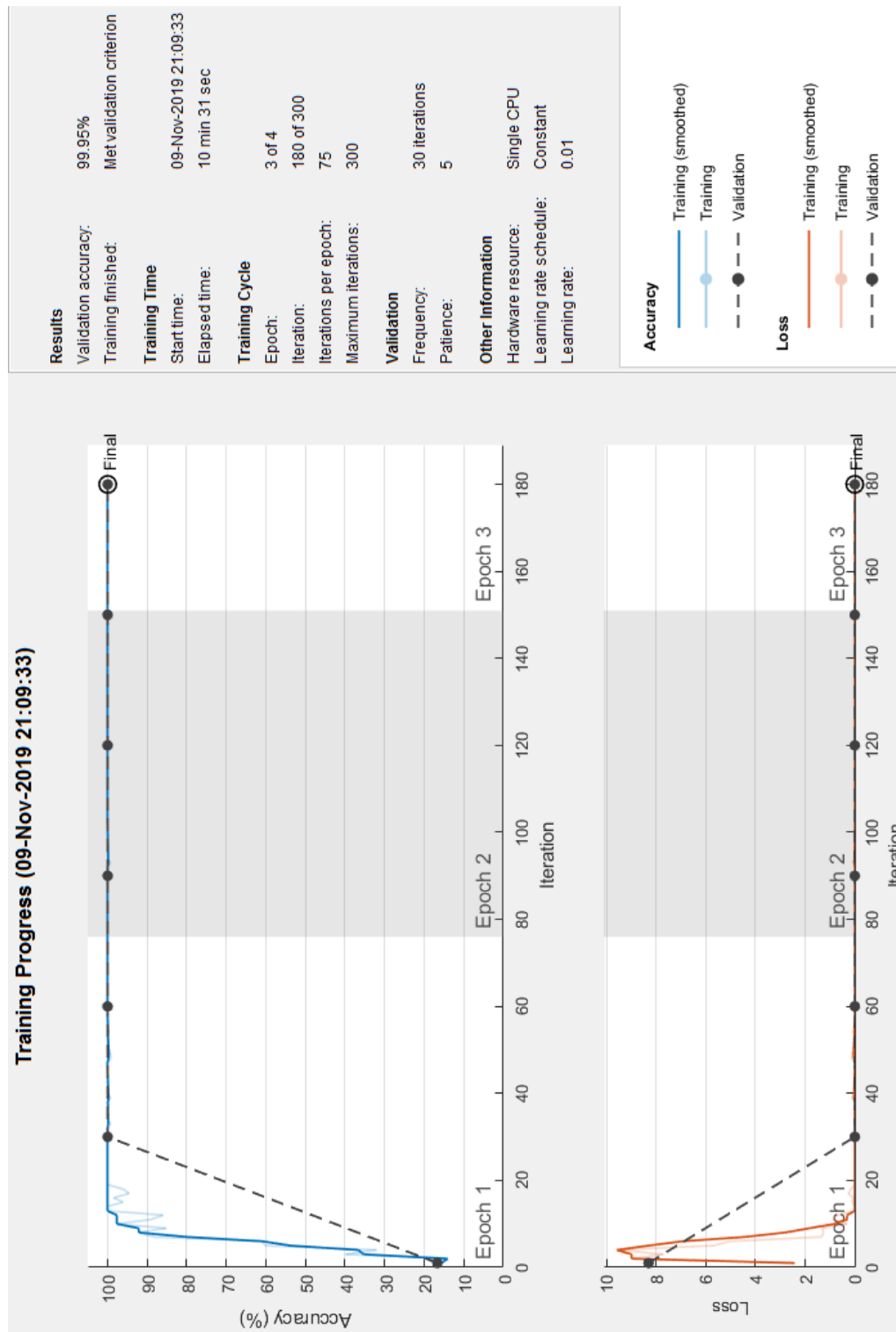


Figure 7.25 CNN training process

It is manifest that the trend of accuracy changes in inversely proportional to the trend of loss. In the training process, the accuracy tends to be stable after 20 iterations indicating that this process has converged. The validation accuracy is up to 99.95%.

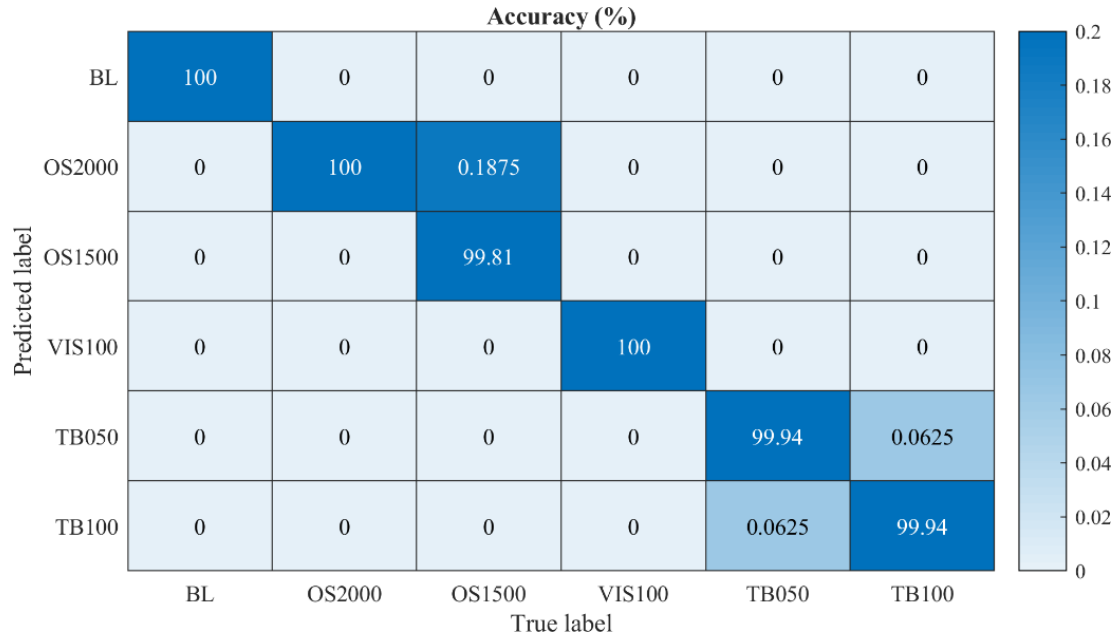


Figure 7.26 Confusion matrix of the fault classification accuracy for testing results

Two loads for each case are applied to train the network, and another two are used to test the trained network to prevent overfitting. The classification accuracy result of the test is displayed with a confusion matrix in Figure 7.26. The x-coordinate and y-coordinate respectively represent the actual labels and predicted labels of the images, which are the input images to test the effectiveness of the trained network. It can be seen that the cases of BL, OS2000 and VIS100 can be accurately recognised in this study. The case of OS1500 is correctly identified with the probability of 99.81% and as OS2000 with the identification error of 0.1875%. The probability that TB050 and TB100 are recognised each other is only 0.0625% because the same type of fault with different severity results in similar thermal field distribution.

In this study, the image resolution is compressed from  $1440 \times 1080$  pixels to  $160 \times 120$  pixels with the CS method. However, the gearbox fault information is preserved in the process of compression. It is evident that the CS based thermal images can be used for classification in a variety of cases under variable loads with high accuracy. In addition, verification with variable loads not only prevents overfitting, but also effectively demonstrates the robustness of this CS-based CNN classification method to ambient changes.

## 7.4 Summary

This chapter presents an effective CS based compression method to diagnose the gearbox faults (lubricant starvation, oil degradation and tooth breakage) using thermal imaging analysis. The difference in temperature field distribution presented by thermal images explains the changes in heat generation and heat conduction from the mating components to the gearbox housing. Moreover, it is more reliable and robust using a 2D thermal imaging signal to detect the various gearbox faults than adopting a 1D temperature signal. The input thermal images are compressed from  $1440 \times 1080 \times 3$



(RGB) to  $160 \times 120$  (grayscale) with a compression ratio of 243. The compressed images based CNN lead to effective and efficient fault diagnosis. The intelligent method upon CS achieves a highly accurate classification with the probability of more than 99.81%.

Finally, CS has the potential to be implemented in the thermal cameras to randomly sample the thermal field distribution of the objects. The thermal camera can be installed to a drone or robot and the captured compressed thermal images can be uploaded to the cloud platform for analysis and backup during the inspection.

---

## CHAPTER 8      CONCLUSIONS AND FUTURE WORKS

---

*This chapter summarizes the achievements of the objectives and draws comprehensive conclusions of this research on data compression approaches applied for 1D and 2D signals in the intelligent machine condition monitoring. Moreover, the contributions to knowledge are also presented. Finally, some meaningful suggestions for future research efforts on data compression usage in CM systems are proposed based on the research outcomes in this thesis.*

## 8.1 Review of Objectives and Achievements

This research aims to investigate the effective and efficient data compression algorithms mainly applied to 1D and 2D signals for intelligent fault diagnosis of machinery, which have potentials to be implemented in hardware to reduce the data volume for intelligent CM. The main objectives listed to achieve this purpose in Section 1.4 are all fulfilled and summarized as follows.

**Objective 1:** Review the popular advanced literature of data compression algorithms especially applied in the field of machine CM. Divide these compression techniques into the traditional post-acquisition compression and advanced pre-acquisition compression methods for further verification of application in CM.

**Achievement 1:** The data compression algorithms including traditional uniform sampling and advanced non-uniform sampling methods at both data acquisition and receiving ends for machine CM are reviewed in Section 2.1.1 and Section 2.1.2, respectively. Their advantages and disadvantages are discussed. In particular, as an advanced compression method, CS shows the potential to implement sampling with a lower rate than the Nyquist-Shannon sampling theorem at the acquisition end. The overview of CS applications delivers its superiority in a wireless real-time machine CM system compared with other traditional compression methods.

**Objective 2:** Build a wireless sensor node with MEMS sensors to collect 1D characteristic signals and analyse the acquired signals to extract features based on compressed statistics for the remote condition monitoring in real time.

**Achievement 2:** In Chapter 3, both the wired and wireless vibration sensing systems are successfully designed with MEMS modules based on the statistics of 1D vibration signals. The main purpose of the wired sensor node is to assist the design of the wireless sensing system. Both of them achieve a sampling rate of 3200 Hz. A data acquisition GUI and an Android APP are designed with MATLAB and Android Studio, respectively, to monitor the wired and wireless transmitted data.

**Objective 3:** Verify the availability and effectiveness of the developed intelligent WSN node with compression function by experimental studies.

**Achievement 3:** To verify the availability and effectiveness of the designed sensing system, the sensor calibration is performed in Section 3.2.3 first with a shaker test. The calibration results show that the ADXL345 has an error of about 2.4% in the frequency domain due to the accuracy of the on-chip clock system. The wired sensing system is also tested with raw vibration signals of the domestic piping system and used to monitor the condition and save data with the designed MATLAB GUI. Furthermore, in Section 3.3, it is verified that the designed wireless sensing system is effective and efficient to monitor the leakage of the domestic piping system and its size with RMS values of vibration signals remotely in real time in the lab. Besides, it saves power consumption of the sensor node because the amount of data transmitted is greatly reduced with an

effective statistical compression method. As a result, the developed intelligent WSN node with compression function is demonstrated.

**Objective 4:** Investigate the dimension reduction based compression methods to save storage space for 1D diagnostic signals of CM and simultaneously distinguish the fault types with the compressed features.

**Achievement 4:** A dimension reduction based compression method for the fault diagnosis of tapered roller bearings is proposed in Chapter 4. The RPs are generated to display the periodic texture information by the defined threshold  $\varepsilon$  from the phase space trajectory formed by the time delay and the embedding dimension. Then, twelve RQA indicators are extracted as features from the 2D RPs which are the dimension reduction results of phase space trajectories of vibration signals. The PCA method is applied to reduce the dimension and the redundancy of the extracted indicating features for the classification of different fault types. The effectiveness of this algorithm has been evaluated with the vibration signals of the tapered roller bearings with the outer race fault and the inner race fault respectively.

**Objective 5:** Propose sparse representation based compression algorithms like compressive sensing for 1D diagnostic signals with a large compression ratio which has the potential to be implemented in the hardware for the pre-acquisition compression.

**Achievement 5:** In Chapter 5, a two-step adaptive compression method based on frequency shift and CS is proposed to reduce the data volume used for fault diagnosis and noise elimination of ball bearings, which has the potential to be achieved at the acquisition end to directly save the power consumption, transmission bandwidth and storage space. Both the numerical and experimental results demonstrate that the defects on the ball bearing outer race and inner race can be accurately and efficiently detected with a very small amount of 1D vibration data transmitted and stored with the compression ratio of up to thousands of times. This method provides a theoretical basis for wireless sensor nodes to further realize the online CM of rotating machines in real time in future.

**Objective 6:** Develop the compression method according to extracting features from 2D thermal images for fault diagnosis.

**Achievement 6:** Chapter 6 presents an effective compression method to diagnose the various types of faults in the reciprocating compressor by compressed thermal images. The features extracted from captured thermal images by dense SIFT can be represented as a sparse matrix with sparse coding. The compressed features with less dimension will lead to efficient fault diagnosis by the classification with SVM which achieves an accuracy of more than 99.86% for six different cases of the two-stage reciprocating compressor tested in the lab.

**Objective 7:** Exploit compressive sensing with an intelligent classification method on 2D thermal images captured by monitoring the condition of machines to differentiate

various types of faults. The combined method has the potential to reduce the cost of data collection, transmission and storage.

**Achievement 7:** Chapter 7 states an effective CS based compression method to diagnose the gearbox lubricant starvation, oil degradation and tooth breakage faults with 2D thermal images which represent temperature field distribution of the gearbox surface. The converted grayscale images are compressed from the size of  $1440 \times 1080$  to  $160 \times 120$  with CS based on the divided blocks. The CS-based compressed and assembled images are as the input of the CNN to achieve a highly accurate classification with the probability of more than 99.81%. Besides, it can be concluded by comparison that using a 2D thermal imaging signal for fault diagnosis is much more reliable and robust than a 1D temperature signal.

## 8.2 Conclusions

According to the investigations and discussion made in the aforementioned chapters, this research about data compression in machine CM can be summarized with the following key conclusions:

**Conclusion 1:** The extensive literature review expresses the importance and necessity of investigating data compression methods based on statistics, dimension reduction, sparse representation and feature extraction in the intelligent machine CM filed.

**Conclusion 2:** A cost-effective wireless CM system is successfully designed based on the statistical characteristics of the system vibration responses. Its effectiveness was examined and evaluated by a domestic piping system. Simultaneously, it is demonstrated that the middle of the long straight pipe is a good position to install the accelerometer for CM because it is easily prone to the pipe resonance by the small and random excitations from the water flow.

**Conclusion 3:** The healthy and faulty bearings can be successfully classified with the compressed features through the proposed dimension reduction compression method. Although the clearance between the components of the tapered roller bearing affects the results of the fault diagnosis, the compressed features can help determine whether the bearing clearance is normal or not to avoid misjudging the severity of bearing defects.

**Conclusion 4:** A two-step adaptive compression method based on frequency shift and CS is investigated to reduce the dataset size for diagnosis of ball bearing faults. Its effectiveness is verified with both simulation and experimental results that only very few data can detect the defects on ball bearings via fault frequencies

**Conclusion 5:** For 2D thermal images captured from different cases of a two-stage reciprocating compressor, the dense SIFT features are extracted and represented as a sparse matrix with the sparse coding. The compressed image leads to effective fault classification with SVM, which achieves an accuracy of more than 99.86% for six different cases of the reciprocating compressor.

**Conclusion 6:** The 2D thermal images are compressed with CS to reduce the feature size of the input layer of the CNN to speed up fault type classification of gearboxes. The trained CNN achieved a highly accurate classification with the probability of more than 99.81% using the compressed thermal images.

### 8.3 Contributions to Knowledge

The works on data compression for machine CM performed in this research have been introduced and discussed in previous chapters. The contributions of these works to knowledge are summarized as follows.

**Contribution 1:** This research reviews and summarizes both the traditional and the advanced data compression methods used for pre-transmission and post-transmission of data in the field of machine CM. This has not been studied in the previous research and even data compression has not attracted widespread attention in machine CM. It paves the way to significantly efficient CM techniques under large quantities of measurements.

**Contribution 2:** A wireless CM system with cost-effective MEMS accelerometers was designed to detect leakage in water piping systems, and statistics based data compression approach was proposed for reducing data transmission load. The investigation in prototyping the leak detector denotes that data compression via waveform statistics can be effective and efficient in CM.

**Contribution 3:** Two novel methods for compressing 1D vibration signal are developed to distinguish various bearing faults efficiently.

- A dimension reduction compression method based on PCA is proposed to reduce the dimension of the RQA indicators. The compressed features retain the periodic fault information and lead to the successful detection and diagnosis of bearing faults.
- A two-step adaptive compression method based on frequency shift and CS is developed to compress and then reconstruct envelope signals with high SNR. The developed compression strategy achieves efficient diagnosis of rolling element bearing faults.

**Contribution 4:** Another two novel approaches for compressing 2D thermal images are developed to intelligently monitor the machine conditions.

- The dense SIFT features are extracted from the 2D thermal images of a two-stage reciprocating compressor and represented as a sparse matrix with the sparse coding. The compressed features with fewer dimensions lead to effective fault classification with SVM.
- The advanced CS method is applied to the pre-processed 2D thermal images captured from the gearbox. The compressed images are used as the input of the CNN to classify six different fault types, which speeds up the calculation of the CNN while does not affect the classification accuracy.

## 8.4 Future Work

**Recommendation 1:** Optimize the proposed compression algorithm by exploring effective signal denoising, adaptive optimal frequency band selection and signal sparse representation methods to improve the efficiency of data compression for better performance of machine CM.

**Recommendation 2:** Improve the designed wireless sensing system through integrating other types of sensors, like temperature, and trying other wireless transmission ways, like Wi-Fi. Then consider uploading the states and raw data to the cloud platform for sharing to realize remote and real-time CM of machinery.

**Recommendation 3:** Investigate energy harvesting techniques to provide power to the wireless sensor nodes and hence achieve maintenance-free machine condition monitoring systems [114]. Simultaneously, optimize the working condition of the wireless sensor nodes to reduce the node energy consumption through reducing the amount of data transmitted, organising communication between multiple nodes and considering intermittent data collection.

**Recommendation 4:** Implement the proposed data compression algorithms on the wireless sensor nodes, especially the CS based method for 1D signals. The CS on hardware at the acquisition end can break through the conventional Nyquist sampling theory, which can significantly reduce the data collected but allow reconstructing the original signals with high accuracy.

**Recommendation 5:** Attempt to apply the CS theory to thermal imaging device based on the single-pixel compressive digital camera and CS based MRI equipment. If the CS based infrared thermal camera can be developed, it will have wide applications, especially in the field of machine CM, which can be installed on a drone or robot and the captured compressed thermal images can be uploaded to the cloud platform for analysis and backup during the inspection in future.



## REFERENCE

- [1] P. K. Muhuri, A. K. Shukla, and A. Abraham, 'Industry 4.0: A Bibliometric Analysis and Detailed Overview', *Eng. Appl. Artif. Intell.*, vol. 78, pp. 218–235, Feb. 2019.
- [2] G.-Y. Lee, M. Kim, Y.-J. Quan, M.-S. Kim, T. J. Y. Kim, H.-S. Yoon, S. Min, D.-H. Kim, J.-W. Mun, J. W. Oh, I. G. Choi, C.-S. Kim, W.-S. Chu, J. Yang, B. Bhandari, C.-M. Lee, J.-B. Ihn, and S.-H. Ahn, 'Machine Health Management in Smart Factory: A Review', *J. Mech. Sci. Technol.*, vol. 32, no. 3, pp. 987–1009, Mar. 2018, doi: 10.1007/s12206-018-0201-1.
- [3] X. Tian, 'Enhanced Information Extraction from Noisy Vibration Data for Machinery Fault Detection and Diagnosis', Ph.D. Thesis, University of Huddersfield, Huddersfield, UK, 2017.
- [4] Vanraj, D. Goyal, A. Saini, S. S. Dhami, and B. S. Pabla, 'Intelligent Predictive Maintenance of Dynamic Systems Using Condition Monitoring and Signal Processing Techniques - A Review', in *2016 International Conference on Advances in Computing, Communication, Automation (ICACCA) (Spring)*, 2016, pp. 1–6, doi: 10.1109/ICACCA.2016.7578870.
- [5] Z. Li, Y. Wang, and K.-S. Wang, 'Intelligent Predictive Maintenance for Fault Diagnosis and Prognosis in Machine Centers: Industry 4.0 Scenario', *Adv. Manuf.*, vol. 5, no. 4, pp. 377–387, Dec. 2017, doi: 10.1007/s40436-017-0203-8.
- [6] D. Bhattacharyya, T. Kim, and S. Pal, 'A Comparative Study of Wireless Sensor Networks and Their Routing Protocols', *Sensors*, vol. 10, no. 12, pp. 10506–10523, Nov. 2010, doi: 10.3390/s101210506.
- [7] P. Tiwari, V. P. Saxena, R. G. Mishra, and D. Bhavsar, 'Wireless Sensor Networks: Introduction, Advantages, Applications and Research Challenges', vol. 14, Apr. 2015.
- [8] K. G. Bøving, Ed., 'Chapter 36 - Visual inspection', in *NDE Handbook*, Butterworth-Heinemann, 1989, pp. 379–396.
- [9] T. H. Loutas, D. Roulias, E. Pauly, and V. Kostopoulos, 'The Combined Use of Vibration, Acoustic Emission and Oil Debris On-Line Monitoring towards a More Effective Condition Monitoring of Rotating Machinery', *Mech. Syst. Signal Process.*, vol. 25, no. 4, pp. 1339–1352, May 2011, doi: 10.1016/j.ymssp.2010.11.007.
- [10] M. Kumar, P. Shankar Mukherjee, and N. Mohan Misra, 'Advancement and Current Status of Wear Debris Analysis for Machine Condition Monitoring: A Review', *Ind. Lubr. Tribol.*, vol. 65, no. 1, pp. 3–11, 2013.
- [11] M. Maizuar, L. Zhang, S. Miramini, P. Mendis, and C. Duffield, 'Structural Health Monitoring of Bridges Using Advanced Non-Destructive Testing Technique', in *ACMSM25*, Singapore, 2020, pp. 963–972, doi: 10.1007/978-981-13-7603-0\_91.
- [12] S. Gholizadeh, 'A Review of Non-Destructive Testing Methods of Composite Materials', *Procedia Struct. Integr.*, vol. 1, pp. 50–57, Jan. 2016, doi: 10.1016/j.prostr.2016.02.008.
- [13] A. Tatarinov, A. Rumjancevs, and V. Mironovs, 'Assessment of Cracks in Pre-Stressed Concrete Railway Sleepers by Ultrasonic Testing', *Procedia Comput. Sci.*, vol. 149, pp. 324–330, Jan. 2019, doi: 10.1016/j.procs.2019.01.143.
- [14] J. Blitz and G. Simpson, *Ultrasonic Methods of Non-Destructive Testing*, vol. 2. Springer Science & Business Media, 1995.
- [15] K. G. Bøving, Ed., 'Chapter 4 - Dye Penetrant Examination: Penetrant Flaw Detection', in *NDE Handbook*, Butterworth-Heinemann, 1989, pp. 32–41.
- [16] W. C. Morrey, 'Penetrant Testing in the Quality Control of Nuclear Power Plant Construction', *NDT Int.*, vol. 10, no. 1, pp. 9–12, Feb. 1977, doi: 10.1016/0308-9126(77)90106-7.
- [17] C. Yang, W. Xu, Y. Chen, B. Guo, and D. Shan, 'Restoration of Fatigue Damage in Steel Tube by Eddy Current Treatment', *Int. J. Fatigue*, vol. 124, pp. 422–434, Jul. 2019, doi: 10.1016/j.ijfatigue.2019.03.012.
- [18] A. S. N. Huda, S. Taib, K. H. Ghazali, and M. S. Jadin, 'A New Thermographic NDT for Condition Monitoring of Electrical Components Using ANN with Confidence Level Analysis',

- ISA Trans.*, vol. 53, no. 3, pp. 717–724, May 2014, doi: 10.1016/j.isatra.2014.02.003.
- [19] T. Jayakumar and C. K. Mukhopadhyay, ‘NDT Techniques: Acoustic Emission’, in *Reference Module in Materials Science and Materials Engineering*, Elsevier, 2016.
  - [20] S. Doshvarpassand, C. Wu, and X. Wang, ‘An Overview of Corrosion Defect Characterization Using Active Infrared Thermography’, *Infrared Phys. Technol.*, vol. 96, pp. 366–389, Jan. 2019, doi: 10.1016/j.infrared.2018.12.006.
  - [21] C. Meola, G. M. Carlomagno, A. Squillace, and A. Vitiello, ‘Non-Destructive Evaluation of Aerospace Materials with Lock-in Thermography’, *Eng. Fail. Anal.*, vol. 13, no. 3, pp. 380–388, 2006.
  - [22] F. Ciampa, P. Mahmoodi, F. Pinto, and M. Meo, ‘Recent Advances in Active Infrared Thermography for Non-Destructive Testing of Aerospace Components’, *Sensors*, vol. 18, no. 2, p. 609, 2018.
  - [23] A. Katunin, K. Dragan, and M. Dziendzikowski, ‘Damage Identification in Aircraft Composite Structures: A Case Study Using Various Non-Destructive Testing Techniques’, *Compos. Struct.*, vol. 127, pp. 1–9, 2015.
  - [24] M. Neslušan, J. Čížek, K. Zgútová, P. Kejzlar, J. Šrámek, J. Čapek, P. Hruška, and O. Melikhova, ‘Microstructural Transformation of a Rail Surface Induced by Severe Thermoplastic Deformation and Its Non-Destructive Monitoring via Barkhausen Noise’, *Wear*, vol. 402, pp. 38–48, 2018.
  - [25] P. Nowak, M. Nowicki, A. Juś, and R. Szewczyk, ‘Utilization of Eddy Current Tomography in Automotive Industry’, *Acta Phys. Pol. A*, vol. 131, no. 4, pp. 1168–1170, 2017.
  - [26] M. R. Jolly, A. Prabhakar, B. Sturzu, K. Hollstein, R. Singh, S. Thomas, P. Foote, and A. Shaw, ‘Review of Non-Destructive Testing (NDT) Techniques and Their Applicability to Thick Walled Composites’, *Procedia CIRP*, vol. 38, pp. 129–136, 2015.
  - [27] S. Bagavathiappan, B. B. Lahiri, T. Saravanan, J. Philip, and T. Jayakumar, ‘Infrared Thermography for Condition Monitoring – A Review’, *Infrared Phys. Technol.*, vol. 60, pp. 35–55, Sep. 2013, doi: 10.1016/j.infrared.2013.03.006.
  - [28] V. T. Tran, B.-S. Yang, F. Gu, and A. Ball, ‘Thermal Image Enhancement Using Bi-dimensional Empirical Mode Decomposition in Combination with Relevance Vector Machine for Rotating Machinery Fault Diagnosis’, *Mech. Syst. Signal Process.*, vol. 38, no. 2, pp. 601–614, Jul. 2013, doi: 10.1016/j.ymssp.2013.02.001.
  - [29] B. Huang, G. Feng, X. Tang, J. X. Gu, G. Xu, R. Cattley, F. Gu, and A. D. Ball, ‘A Performance Evaluation of Two Bispectrum Analysis Methods Applied to Electrical Current Signals for Monitoring Induction Motor-Driven Systems’, *Energies*, vol. 12, no. 8, p. 1438, Jan. 2019, doi: 10.3390/en12081438.
  - [30] N. Hamad, X. Sun, R. Zhang, S. Abusaad, F. Gu, and A. D. Ball, ‘Diagnosing Lubricant Shortages in Gearboxes Using Instantaneous Phases from Electrical Signals’, in *Proceedings of the 24th International Conference on Automation & Computing*, Newcastle University, Newcastle upon Tyne, UK, 2018, p. 6.
  - [31] Y. Jiang, B. Tang, Y. Qin, and W. Liu, ‘Feature Extraction Method of Wind Turbine based on Adaptive Morlet Wavelet and SVD’, *Renew. Energy*, vol. 36, no. 8, pp. 2146–2153, 2011.
  - [32] X. Wang and V. Makis, ‘Autoregressive Model-based Gear Shaft Fault Diagnosis Using the Kolmogorov–Smirnov Test’, *J. Sound Vib.*, vol. 327, no. 3–5, pp. 413–423, Nov. 2009, doi: 10.1016/j.jsv.2009.07.004.
  - [33] Q. He, F. Kong, and R. Yan, ‘Subspace-based Gearbox Condition Monitoring by Kernel Principal Component Analysis’, *Mech. Syst. Signal Process.*, vol. 21, no. 4, pp. 1755–1772, May 2007, doi: 10.1016/j.ymssp.2006.07.014.
  - [34] X. Tian, F. Gu, I. Rehab, G. Abdalla, and A. Ball, ‘A Robust Fault Detection Method of Rolling Bearings Using Modulation Signal Bispectrum Analysis’, presented at the 28th International Congress of Condition Monitoring and Diagnostic Engineering Management (COMADEM 2015), Buenos Aires, Argentina, 2015.
  - [35] R. Teti, K. Jemielniak, G. O’Donnell, and D. Dornfeld, ‘Advanced monitoring of machining

- operations', *CIRP Ann.*, vol. 59, no. 2, pp. 717–739, Jan. 2010, doi: 10.1016/j.cirp.2010.05.010.
- [36] N. Baydar and A. Ball, 'A Comparative Study of Acoustic and Vibration Signals in Detection of Gear Failures Using Wigner–Ville Distribution', *Mech. Syst. Signal Process.*, vol. 15, no. 6, pp. 1091–1107, Nov. 2001, doi: 10.1006/mssp.2000.1338.
- [37] J. C. Chan and P. W. Tse, 'A Novel, Fast, Reliable Data Transmission Algorithm for Wireless Machine Health Monitoring', *IEEE Trans. Reliab.*, vol. 58, no. 2, pp. 295–304, Jun. 2009, doi: 10.1109/TR.2009.2020479.
- [38] J. D. Das, S. Chowdhuri, and J. Bera, 'A Novel Hybrid Huffman Coding Compression Technique for Secured Condition Monitoring of Induction Machine', in *Proceedings of The 2014 International Conference on Control, Instrumentation, Energy and Communication (CIEC)*, 2014, pp. 467–471, doi: 10.1109/CIEC.2014.6959132.
- [39] J. Datta, J. Bera, and S. Chowdhuri, 'Development of Remote Monitoring Analysis and Reporting System for Industrial Machines', in *2016 2nd International Conference on Control, Instrumentation, Energy Communication (CIEC)*, 2016, pp. 466–470, doi: 10.1109/CIEC.2016.7513776.
- [40] Q. Huang, M. Ros, Y. Zhang, and Z. Su, 'Mechanical Vibration Data Lossless Compression with 2D Block Adaptive Quantization', in *2018 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*, 2018, pp. 297–301, doi: 10.1109/SDPC.2018.8664997.
- [41] S. Lu, P. Zhou, X. Wang, Y. Liu, F. Liu, and J. Zhao, 'Condition Monitoring and Fault Diagnosis of Motor Bearings Using Undersampled Vibration Signals from a Wireless Sensor Network', *J. Sound Vib.*, vol. 414, pp. 81–96, Feb. 2018, doi: 10.1016/j.jsv.2017.11.007.
- [42] Z. K. Peng and F. L. Chu, 'Application of the Wavelet Transform in Machine Condition Monitoring and Fault Diagnostics: A Review with Bibliography', *Mech. Syst. Signal Process.*, vol. 18, no. 2, pp. 199–221, Mar. 2004, doi: 10.1016/S0888-3270(03)00075-X.
- [43] W. Hao and G. Jinji, 'The Research of Optimal Selection Method for Wavelet Packet Basis in Compressing the Vibration Signal of a Rolling Bearing in Fans and Pumps', *J. Phys. Conf. Ser.*, vol. 364, p. 012033, May 2012, doi: 10.1088/1742-6596/364/1/012033.
- [44] W. Guo and P. W. Tse, 'A Novel Signal Compression Method based on Optimal Ensemble Empirical Mode Decomposition for Bearing Vibration Signals', *J. Sound Vib.*, vol. 332, no. 2, pp. 423–441, Jan. 2013, doi: 10.1016/j.jsv.2012.08.017.
- [45] Y. Zhang, P. Hutchinson, N. A. J. Lieven, and J. Nunez-Yanez, 'Adaptive Event-Triggered Anomaly Detection in Compressed Vibration Data', *Mech. Syst. Signal Process.*, vol. 122, pp. 480–501, May 2019, doi: 10.1016/j.ymsp.2018.12.039.
- [46] J. J. Saucedo-Dorantes, M. Delgado-Prieto, R. A. Osornio-Rios, and R. de Jesus Romero-Troncoso, 'Multifault Diagnosis Method Applied to an Electric Machine based on High-Dimensional Feature Reduction', *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 3086–3097, May 2017, doi: 10.1109/TIA.2016.2637307.
- [47] D. R. Sawitri, I. K. E. Purnama, and M. Ashari, 'Detection of Electrical Faults in Induction Motor Fed by Inverter using Support Vector Machine and Receiver Operating Characteristic', *J. Theor. Appl. Inf. Technol.*, vol. 40, no. 1, pp. 15–22, Jun. 2012.
- [48] S. A. Chopade, J. A. Gaikwad, and J. V. Kulkarni, 'Bearing Fault Detection Using PCA and Wavelet based Envelope Analysis', in *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, 2016, pp. 248–253, doi: 10.1109/ICATCCT.2016.7912002.
- [49] C. G. Dias and F. H. Pereira, 'Broken Rotor Bars Detection in Induction Motors Running at Very Low Slip Using a Hall Effect Sensor', *IEEE Sens. J.*, vol. 18, no. 11, pp. 4602–4613, Jun. 2018, doi: 10.1109/JSEN.2018.2827204.
- [50] C. Ruiz-Carcel and A. Starr, 'Data-based Detection and Diagnosis of Faults in Linear Actuators', *IEEE Trans. Instrum. Meas.*, vol. 67, no. 9, pp. 2035–2047, Sep. 2018, doi: 10.1109/TIM.2018.2814067.
- [51] A. Smith, F. Gu, and A. D. Ball, 'An Approach to Reducing Input Parameter Volume for Fault Classifiers', *Int. J. Autom. Comput.*, vol. 16, no. 2, pp. 199–212, Apr. 2019, doi: 10.1007/s11633-

- 018-1162-7.
- [52] R. Zimroz and A. Bartkowiak, 'Two Simple Multivariate Procedures for Monitoring Planetary Gearboxes in Non-Stationary Operating Conditions', *Mech. Syst. Signal Process.*, vol. 38, no. 1, pp. 237–247, Jul. 2013, doi: 10.1016/j.ymssp.2012.03.022.
  - [53] F. Arellano-Espitia, J. J. Saucedo, M. D. Prieto, and R. A. Osornio-Rios, 'Autoencoder based Feature Reduction Analysis Applied to Electromechanical Systems Condition Monitoring', in *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2019, pp. 891–897, doi: 10.1109/ETFA.2019.8869371.
  - [54] C. Sun, P. Wang, R. Yan, R. X. Gao, and X. Chen, 'Machine Health Monitoring based on Locally Linear Embedding with Kernel Sparse Representation for Neighborhood Optimization', *Mech. Syst. Signal Process.*, vol. 114, pp. 25–34, Jan. 2019, doi: 10.1016/j.ymssp.2018.04.044.
  - [55] G. Aydemir, 'Deep Learning based Spectrum Compression Algorithm for Rotating Machinery Condition Monitoring', in *Volume 1: Development and Characterization of Multifunctional Materials; Modeling, Simulation, and Control of Adaptive Systems; Integrated System Design and Implementation*, San Antonio, Texas, USA, 2018, p. V001T03A027, doi: 10.1115/SMASIS2018-8137.
  - [56] K. López de Calle, S. Ferreiro, A. Arnaiz, and B. Sierra, 'Dynamic Condition Monitoring Method based on Dimensionality Reduction Techniques for Data-Limited Industrial Environments', *Comput. Ind.*, vol. 112, p. 103114, Nov. 2019, doi: 10.1016/j.compind.2019.07.004.
  - [57] Z. Li, R. Liu, and D. Wu, 'Data-Driven Smart Manufacturing: Tool Wear Monitoring with Audio Signals and Machine Learning', *J. Manuf. Process.*, vol. 48, pp. 66–76, Dec. 2019, doi: 10.1016/j.jmapro.2019.10.020.
  - [58] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, 'Bearing Fault Detection by a Novel Condition-Monitoring Scheme based on Statistical-Time Features and Neural Networks', *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3398–3407, Aug. 2013, doi: 10.1109/TIE.2012.2219838.
  - [59] L. Jiang, T. Shi, and J. Xuan, 'An Intelligent Fault Diagnosis Method of Rolling Bearings based on Regularized Kernel Marginal Fisher Analysis', *J. Phys. Conf. Ser.*, vol. 364, p. 012042, May 2012, doi: 10.1088/1742-6596/364/1/012042.
  - [60] G. F. Wang, Y. W. Yang, Y. C. Zhang, and Q. L. Xie, 'Vibration Sensor based Tool Condition Monitoring Using v Support Vector Machine and Locality Preserving Projection', *Sens. Actuators Phys.*, vol. 209, pp. 24–32, Mar. 2014, doi: 10.1016/j.sna.2014.01.004.
  - [61] J. Wang, Y. Zheng, L. Duan, J. Xie, and L. Zhang, 'Virtual Sensing for Gearbox Condition Monitoring based on Extreme Learning Machine', *J. Vibroengineering*, 2017.
  - [62] J. Wang, J. Xie, R. Zhao, L. Zhang, and L. Duan, 'Multisensory Fusion based Virtual Tool Wear Sensing for Ubiquitous Manufacturing', *Robot. Comput.-Integr. Manuf.*, vol. 45, pp. 47–58, Jun. 2016, doi: 10.1016/j.rcim.2016.05.010.
  - [63] Y. Huang, X. F. Zha, J. Lee, and C. Liu, 'Discriminant Diffusion Maps Analysis: A Robust Manifold Learner for Dimensionality Reduction and Its Applications in Machine Condition Monitoring and Fault Diagnosis', *Mech. Syst. Signal Process.*, vol. 34, no. 1, pp. 277–297, Jan. 2013, doi: 10.1016/j.ymssp.2012.04.021.
  - [64] Y. Huang, X. Liu, C. Liu, and Y. Li, 'Intrinsic Feature Extraction Using Discriminant Diffusion Mapping Analysis for Automated Tool Wear Evaluation', *Front. Inf. Technol. Electron. Eng.*, vol. 19, no. 11, pp. 1352–1361, Nov. 2018, doi: 10.1631/FITEE.1601512.
  - [65] R. G. Baraniuk, 'Compressive Sensing [Lecture Notes]', *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007, doi: 10.1109/MSP.2007.4286571.
  - [66] D. L. Donoho, 'De-Noising by Soft-Thresholding', *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, 1995.
  - [67] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed. Academic press, 2009.
  - [68] J. Stinnett and J. Gillenwater, 'NIR Single Pixel Camera', *Connex. Httpcnx Orgcontentcoll1052511*, 2008.

- [69] E. J. Candes and T. Tao, 'Decoding by Linear Programming', *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005, doi: 10.1109/TIT.2005.858979.
- [70] S. Ji, L. Huang, J. Wang, J. Shen, and J.-U. Kim, 'An Improved Reconstruction Methods of Compressive Sensing Data Recovery in Wireless Sensor Networks', *Int. J. Secur. Its Appl.*, vol. 8, no. 1, pp. 1–8, Jan. 2014, doi: 10.14257/ijisia.2014.8.1.01.
- [71] T. L. Nguyen and Y. Shin, 'Deterministic Sensing Matrices in Compressive Sensing: A Survey', *Sci. World J.*, vol. 2013, 2013.
- [72] M. M. Abo-Zahhad and A. I. Hussein, 'Compressive Sensing Algorithms for Signal Processing Applications: A Survey', *Commun. Netw. Syst. Sci.*, pp. 197–216, 2015.
- [73] D. L. Donoho, 'Compressed Sensing', *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [74] E. J. Candes and T. Tao, 'Near-Optimal Signal Recovery from Random Projections: Universal Encoding Strategies?', *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [75] D. Needell and R. Vershynin, 'Uniform Uncertainty Principle and Signal Recovery via Regularized Orthogonal Matching Pursuit', *Found. Comput. Math.*, vol. 9, no. 3, pp. 317–334, 2009.
- [76] J. Haupt and R. Nowak, 'Signal Reconstruction from Noisy Random Projections', *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4036–4048, 2006.
- [77] D. Takhar, J. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. Kelly, and R. G. Baraniuk, 'A New Compressive Imaging Camera Architecture Using Optical-Domain Compression', *ISTSPIE Comput. Imaging IV*, vol. 6065, no. 606509, p. 1, 2006.
- [78] A. Thompson, 'Compressive Single-Pixel Imaging', in *2nd IMA Conference on Mathematics in Defence*, Defence Academy of the United Kingdom, Swindon, 2011, p. 6.
- [79] M. F. Duarte, M. A. Davenport, D. Takbar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, 'Single-Pixel Imaging via Compressive Sampling', *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, 2008.
- [80] Y. Wang, S. Doleschel, R. Wunderlich, and S. Heinen, 'High Energy Efficient Analog Compressed Sensing Encoder for Wireless ECG System', *Microelectron. J.*, vol. 56, no. Supplement C, pp. 10–16, Oct. 2016, doi: 10.1016/j.mejo.2016.08.002.
- [81] D. Craven, B. McGinley, L. Kilmartin, M. Glavin, and E. Jones, 'Energy-Efficient Compressed Sensing for Ambulatory ECG Monitoring', *Comput. Biol. Med.*, vol. 71, no. Supplement C, pp. 1–13, Apr. 2016, doi: 10.1016/j.combiomed.2016.01.013.
- [82] C. Chen and J. Huang, 'Exploiting the Wavelet Structure in Compressed Sensing MRI', *Magn. Reson. Imaging*, vol. 32, no. 10, pp. 1377–1389, Dec. 2014, doi: 10.1016/j.mri.2014.07.016.
- [83] Wikipedia contributors, 'Wireless Sensor Network', *Wikipedia*, 18-Nov-2019. [Online]. Available: [https://en.wikipedia.org/w/index.php?title=Wireless\\_sensor\\_network&oldid=926811057](https://en.wikipedia.org/w/index.php?title=Wireless_sensor_network&oldid=926811057).
- [84] M. T. Nguyen and K. A. Teague, 'Compressive Sensing based Random Walk Routing in Wireless Sensor Networks', *Ad Hoc Netw.*, vol. 54, no. Supplement C, pp. 99–110, Jan. 2017, doi: 10.1016/j.adhoc.2016.10.009.
- [85] S. Chen, C. Zhao, M. Wu, Z. Sun, H. Zhang, and V. C. M. Leung, 'Compressive Network Coding for Wireless Sensor Networks: Spatio-Temporal Coding and Optimization Design', *Comput. Netw.*, vol. 108, no. Supplement C, pp. 345–356, Oct. 2016, doi: 10.1016/j.comnet.2016.09.007.
- [86] H. Wang, Y. Zhu, and Q. Zhang, 'Compressive Sensing based Monitoring with Vehicular Networks', in *2013 Proceedings IEEE INFOCOM*, 2013, pp. 2823–2831, doi: 10.1109/INFCOM.2013.6567092.
- [87] M. Wong, M. Zhang, and A. Nandi, 'Effects of Compressed Sensing on Classification of Bearing Faults with Entropic Features', in *Signal Processing Conference (EUSIPCO), 2015 23rd European*, 2015, pp. 2256–2260.
- [88] H. Yuan, C. Lu, Z. Wang, and F. Li, 'Making Health Management More Concise and Efficient:

- Aircraft Condition Monitoring based on Compressed Sensing', p. 8, 2016.
- [89] H. Yuan and C. Lu, 'Rolling Bearing Fault Diagnosis under Fluctuant Conditions based on Compressed Sensing', *Struct. Control Health Monit.*, vol. 24, no. 5, p. e1918, May 2017, doi: 10.1002/stc.1918.
  - [90] G. Tang, Q. Yang, H.-Q. Wang, G. Luo, and J. Ma, 'Sparse Classification of Rotating Machinery Faults based on Compressive Sensing Strategy', *Mechatronics*, vol. 31, pp. 60–67, Oct. 2015, doi: 10.1016/j.mechatronics.2015.04.006.
  - [91] G. Tang, W. Hou, H. Wang, G. Luo, and J. Ma, 'Compressive Sensing of Roller Bearing Faults via Harmonic Detection from Under-Sampled Vibration Signals', *Sensors*, vol. 15, no. 10, pp. 25648–25662, Oct. 2015, doi: 10.3390/s151025648.
  - [92] H. Wang, Y. Ke, G. Luo, and G. Tang, 'Compressed Sensing of Roller Bearing Fault based on Multiple down-Sampling Strategy', *Meas. Sci. Technol.*, vol. 27, no. 2, p. 025009, Feb. 2016, doi: 10.1088/0957-0233/27/2/025009.
  - [93] H. Wang, Y. Ke, G. Luo, L. Li, and G. Tang, 'A Two-Stage Compression Method for the Fault Detection of Roller Bearings', *Shock Vib.*, vol. 2016, pp. 1–11, 2016, doi: 10.1155/2016/2971749.
  - [94] H. O. A. Ahmed, M. L. D. Wong, and A. K. Nandi, 'Compressive Sensing Strategy for Classification of Bearing Faults', in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 2182–2186, doi: 10.1109/ICASSP.2017.7952543.
  - [95] H. O. A. Ahmed and A. K. Nandi, 'Intelligent Condition Monitoring for Rotating Machinery Using Compressively-Sampled Data and Sub-space Learning Techniques', in *Proceedings of the 10th International Conference on Rotor Dynamics – IFToMM*, Rio de Janeiro, Brazil, 2018, vol. 61, pp. 238–251, doi: 10.1007/978-3-319-99268-6\_17.
  - [96] H. O. A. Ahmed and A. K. Nandi, 'Three-Stage Hybrid Fault Diagnosis for Rolling Bearings with Compressively Sampled Data and Subspace Learning Techniques', *IEEE Trans. Ind. Electron.*, vol. 66, no. 7, pp. 5516–5524, Jul. 2019, doi: 10.1109/TIE.2018.2868259.
  - [97] H. O. A. Ahmed and A. K. Nandi, 'Multiple Measurement Vector Compressive Sampling and Fisher Score Feature Selection for Fault Classification of Roller Bearings', in *2017 22nd International Conference on Digital Signal Processing (DSP)*, 2017, pp. 1–5, doi: 10.1109/ICDSP.2017.8096076.
  - [98] H. Ahmed and A. K. Nandi, 'Compressive Sampling and Feature Ranking Framework for Bearing Fault Classification with Vibration Signals', *IEEE Access*, vol. 6, pp. 44731–44746, 2018, doi: 10.1109/ACCESS.2018.2865116.
  - [99] H. O. A. Ahmed, M. L. D. Wong, and A. K. Nandi, 'Classification of Bearing Faults Combining Compressive Sampling, Laplacian Score, and Support Vector Machine', in *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, 2017, pp. 8053–8058, doi: 10.1109/IECON.2017.8217413.
  - [100] H. Ahmed and A. Nandi, 'Three-Stage Method for Rotating Machine Health Condition Monitoring Using Vibration Signals', in *2018 Prognostics and System Health Management Conference (PHM-Chongqing)*, 2018, pp. 285–291, doi: 10.1109/PHM-Chongqing.2018.00055.
  - [101] H. O. A. Ahmed, M. L. D. Wong, and A. K. Nandi, 'Intelligent Condition Monitoring Method for Bearing Faults from Highly Compressed Measurements Using Sparse Over-Complete Features', *Mech. Syst. Signal Process.*, vol. 99, pp. 459–477, Jan. 2018, doi: 10.1016/j.ymssp.2017.06.027.
  - [102] Y. Ma, X. Jia, Q. Hu, D. Xu, C. Guo, Q. Wang, and S. Wang, 'Laplace Prior-based Bayesian Compressive Sensing Using K-SVD for Vibration Signal Transmission and Fault Detection', *Electronics*, vol. 8, no. 5, p. 517, May 2019, doi: 10.3390/electronics8050517.
  - [103] X. Tang, Y. Xu, F. Gu, and G. Wang, 'Fault Detection of Rolling Element Bearings Using the Frequency Shift and Envelope based Compressive Sensing', in *Proceedings of the 23rd International Conference on Automation & Computing, (University of Huddersfield, 7-8 September 2017)*, 2017, doi: 10.23919/IConAC.2017.8082063.
  - [104] Y. Wang, J. Xiang, Q. Mo, and S. He, 'Compressed Sparse Time-Frequency Feature Representation via Compressive Sensing and Its Applications in Fault Diagnosis', *Measurement*, vol. 68, pp. 70–81, May 2015, doi: 10.1016/j.measurement.2015.02.046.

- [105] M. Ismail, R. A. Dziyauddin, and N. A. A. Salleh, 'Performance Evaluation of Wireless Accelerometer Sensor for Water Pipeline Leakage', in *2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS)*, 2015, pp. 120–125, doi: 10.1109/IRIS.2015.7451598.
- [106] S. M. Price and D. R. Smith, 'Sources and Remedies of High-Frequency Piping Vibration and Noise', in *Proc. of the 28th Turbomachinery Symposium*, 1999, pp. 189–212.
- [107] J. Wachel, S. J. Morton, and K. E. Atkins, 'Piping Vibration Analysis', presented at the Proceedings of the 19th turbomachinery symposium, 1990, pp. 119–134.
- [108] R. D. Blevins and R. Plunkett, 'Formulas for Natural Frequency and Mode Shape', *J. Appl. Mech.*, vol. 47, p. 461, 1980.
- [109] Wikipedia contributors, 'Turbulence', *Wikipedia*, 19-Dec-2019. [Online]. Available: <https://en.wikipedia.org/w/index.php?title=Turbulence&oldid=931600892>. [Accessed: 24-Dec-2019].
- [110] V. Y. Rudyak and A. Belkin, 'Fluid Viscosity under Confined Conditions', in *Doklady Physics*, 2014, vol. 59, pp. 604–606.
- [111] N. Maluf and K. Williams, *Introduction to Microelectromechanical Systems Engineering*, 2nd ed. Boston: Artech House, 2004.
- [112] M. Bao, *Analysis and Design Principles of MEMS Devices*. Elsevier, 2005.
- [113] A. El Kateeb, A. Ramesh, and L. Azzawi, 'Wireless Sensor Nodes Processor Architecture and Design', in *Advanced Information Networking and Applications-Workshops, 2008. AINAW 2008. 22nd International Conference on*, 2008, pp. 892–897.
- [114] X. Tang, X. Wang, R. Cattley, F. Gu, and A. D. Ball, 'Energy Harvesting Technologies for Achieving Self-Powered Wireless Sensor Networks in Machine Condition Monitoring: A Review', *Sensors*, vol. 18, no. 12, p. 4113, Dec. 2018, doi: 10.3390/s18124113.
- [115] L. Fried, 'Adafruit Feather M0 Basic Proto: Power Management', *Adafruit*, 23-Dec-2019. [Online]. Available: <https://learn.adafruit.com/adafruit-feather-m0-basic-proto/power-management>. [Accessed: 09-Sep-2016].
- [116] T. Igarashi and H. Hamada, 'Studies on the Vibration and Sound of Defective Rolling Bearings: First Report: Vibration of Ball Bearings with One Defect', *Bull. JSME*, vol. 25, no. 204, pp. 994–1001, 1982.
- [117] 'Bearing Damage Analysis for Tapered Roller Bearings', *Know Your Parts*, 14-Feb-2018. [Online]. Available: <https://www.knowyourparts.com/technical-resources/hubs-bearings/damage-analysis-tapered-roller-bearings/>. [Accessed: 19-Sep-2019].
- [118] J. L. Lebowitz, 'Boltzmann's Entropy and Time's Arrow', *Phys. Today*, vol. 46, pp. 32–32, 1993.
- [119] H. Poincaré, 'Sur Le Problème Des Trois Corps Et Les Équations De La Dynamique', *Acta Math.*, vol. 13, no. 1, pp. A3–A270, 1890.
- [120] N. Marwan, M. Carmen Romano, M. Thiel, and J. Kurths, 'Recurrence Plots for the Analysis of Complex Systems', *Phys. Rep.*, vol. 438, no. 5, pp. 237–329, Jan. 2007, doi: 10.1016/j.physrep.2006.11.001.
- [121] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, 'Recurrence Plots of Dynamical Systems', *EPL Europhys. Lett.*, vol. 4, no. 9, p. 973, Nov. 1987, doi: 10.1209/0295-5075/4/9/004.
- [122] T. D. Pham, 'Fuzzy Recurrence Plots', *EPL Europhys. Lett.*, vol. 116, no. 5, p. 50008, Dec. 2016, doi: 10.1209/0295-5075/116/50008.
- [123] J. Yu and D. Liang, 'Impact Localization System of Composite Structure based on Recurrence Quantification Analysis by Using FBG Sensors', *Opt. Fiber Technol.*, vol. 49, pp. 7–15, May 2019, doi: 10.1016/j.yofte.2019.01.024.
- [124] R. Zarghami, F. Sharifi, and N. Mostoufi, 'Investigating the Hydrodynamics of High Temperature Fluidized Bed by Recurrence Plot', *Exp. Therm. Fluid Sci.*, vol. 83, pp. 88–99, May 2017, doi: 10.1016/j.expthermflusci.2016.12.017.
- [125] Y. Chen and H. Yang, 'Multiscale Recurrence Analysis of Long-Term Nonlinear and Nonstationary Time Series', *Chaos Solitons Fractals*, vol. 45, no. 7, pp. 978–987, Jul. 2012, doi:

- 10.1016/j.chaos.2012.03.013.
- [126] R. Shaw, *The Dripping Faucet as a Model Chaotic System*. Aerial Pr, 1984.
  - [127] A. M. Fraser and H. L. Swinney, 'Independent Coordinates for Strange Attractors from Mutual Information', p. 7.
  - [128] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, 'Determining Embedding Dimension for Phase-Space Reconstruction Using a Geometrical Construction', p. 9.
  - [129] J. P. Zbilut and C. L. Webber Jr, 'Embeddings and Delays as Derived from Quantification of Recurrence Plots', *Phys. Lett. A*, vol. 171, no. 3–4, pp. 199–203, 1992.
  - [130] M. Koebbe, G. Mayer-Kress, and J. Zbilut, 'Use of Recurrence Plots in the Analysis of Time-Series Data', *Proc SFI Stud. Sci. Complex.*, 1994.
  - [131] G. M. Mindlin and R. Gilmore, 'Topological Analysis and Synthesis of Chaotic Time Series', *Phys. Nonlinear Phenom.*, vol. 58, no. 1–4, pp. 229–242, 1992.
  - [132] J. P. Zbilut, J.-M. Zaldivar-Comenges, and F. Strozzi, 'Recurrence Quantification based Liapunov Exponents for Monitoring Divergence in Experimental Data', *Phys. Lett. A*, vol. 297, no. 3–4, pp. 173–181, 2002.
  - [133] M. Thiel, M. C. Romano, J. Kurths, R. Meucci, E. Allaria, and F. T. Arecchi, 'Influence of Observational Noise on the Recurrence Quantification Analysis', *Phys. Nonlinear Phenom.*, vol. 171, no. 3, pp. 138–152, 2002.
  - [134] S. Schinkel, O. Dimigen, and N. Marwan, 'Selection of Recurrence Threshold for Signal Detection', *Eur. Phys. J. Spec. Top.*, vol. 164, no. 1, pp. 45–53, Oct. 2008, doi: 10.1140/epjst/e2008-00833-5.
  - [135] I. T. Jolliffe and J. Cadima, 'Principal Component Analysis: A Review and Recent Developments', *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.*, vol. 374, no. 2065, p. 20150202, Apr. 2016, doi: 10.1098/rsta.2015.0202.
  - [136] P. Sanguansat, *Principal Component Analysis*. BoD – Books on Demand, 2012.
  - [137] I. Gruszczyńska, R. Mosdorf, P. Sobaniec, M. Żochowska-Sobaniec, and M. Borowska, 'Epilepsy Identification based on EEG Signal Using RQA Method', *Adv. Med. Sci.*, vol. 64, no. 1, pp. 58–64, 2019.
  - [138] M. Borowska, E. Brzozowska, P. Kuć, E. Oczeretko, R. Mosdorf, and P. Laudański, 'Identification of Preterm Birth based on RQA Analysis of Electrohysterograms', *Comput. Methods Programs Biomed.*, vol. 153, pp. 227–236, Jan. 2018, doi: 10.1016/j.cmpb.2017.10.018.
  - [139] S. Zeemering, P. Bonizzi, B. Maesen, R. Peeters, and U. Schotten, 'Recurrence Quantification Analysis Applied to Spatiotemporal Pattern Analysis in High-Density Mapping of Human Atrial Fibrillation', in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 7704–7707, doi: 10.1109/EMBC.2015.7320177.
  - [140] V. Barroso-García, G. C. Gutiérrez-Tobal, L. Kheirandish-Gozal, D. Álvarez, F. Vaquerizo-Villar, P. Núñez, F. del Campo, D. Gozal, and R. Hornero, 'Usefulness of Recurrence Plots from Airflow Recordings to Aid in Paediatric Sleep Apnoea Diagnosis', *Comput. Methods Programs Biomed.*, vol. 183, p. 105083, Jan. 2020, doi: 10.1016/j.cmpb.2019.105083.
  - [141] Q.-Z. Liang, X.-M. Guo, W.-Y. Zhang, W.-D. Dai, and X.-H. Zhu, 'Identification of Heart Sounds with Arrhythmia based on Recurrence Quantification Analysis and Kolmogorov Entropy', *J. Med. Biol. Eng.*, vol. 35, no. 2, pp. 209–217, 2015.
  - [142] M. Arefi and B. Chowdhury, 'Coherency Detection of Generators Using Recurrence Quantification Analysis', *Electr. Power Syst. Res.*, vol. 169, pp. 162–173, Apr. 2019, doi: 10.1016/j.epsr.2018.12.017.
  - [143] R. Mosdorf and G. Górski, 'Identification of Two-Phase Flow Patterns in Minichannel based on RQA and PCA Analysis', *Int. J. Heat Mass Transf.*, vol. 96, pp. 64–74, May 2016, doi: 10.1016/j.ijheatmasstransfer.2016.01.015.
  - [144] Y. Hou, C. Aldrich, K. Lepkova, and B. Kinsella, 'Detection of under Deposit Corrosion in a CO<sub>2</sub> Environment by Using Electrochemical Noise and Recurrence Quantification Analysis',



- Electrochimica Acta*, vol. 274, pp. 160–169, Jun. 2018, doi: 10.1016/j.electacta.2018.04.037.
- [145] ‘Timken Part Number X31308’, *The Timken Company*. [Online]. Available: <https://cad.timken.com/item/all-product-types/tapered-roller-bearings-ts-tapered-single-metric-2/x31308-y31308-1>. [Accessed: 18-Oct-2019].
- [146] K. Rabeyee, X. Tang, Y. Xu, D. Zhen, F. Gu, and A. D. Ball, ‘Diagnosing the Change in the Internal Clearances of Rolling Element Bearings based on Vibration Signatures’, in *2018 24th International Conference on Automation and Computing (ICAC)*, 2018, pp. 1–6, doi: 10.23919/ICoNAC.2018.8749121.
- [147] K. Rabeyee, X. Tang, F. Gu, and A. Ball, ‘The Effect of Wear Evolution on Vibration-based Fault Detection in Tapered Roller Bearings’, *Int. J. Cond. Monit.*, vol. 9, pp. 18–23, 2019.
- [148] Y.-T. Su and S.-J. Lin, ‘On Initial Fault Detection of a Tapered Roller Bearing: Frequency Domain Analysis’, *J. Sound Vib.*, vol. 155, no. 1, pp. 75–84, 1992.
- [149] Wikipedia contributors, ‘Ball Bearing’, *Wikipedia*, 14-Sep-2019. [Online]. Available: [https://en.wikipedia.org/w/index.php?title=Ball\\_bearing&oldid=915667521](https://en.wikipedia.org/w/index.php?title=Ball_bearing&oldid=915667521). [Accessed: 16-Sep-2019].
- [150] P. K. Kankar, S. C. Sharma, and S. P. Harsha, ‘Fault Diagnosis of Ball Bearings Using Machine Learning Methods’, *Expert Syst. Appl.*, vol. 38, no. 3, pp. 1876–1886, Mar. 2011, doi: 10.1016/j.eswa.2010.07.119.
- [151] T. Jafar and K. M.M., ‘Experimental Testing and Thermal Analysis of Ball Bearings’, doi: 10.1016/j.triboint.2012.10.009.
- [152] F. Immovilli, A. Bellini, R. Rubini, and C. Tassoni, ‘Diagnosis of Bearing Faults in Induction Machines by Vibration or Current Signals: A Critical Comparison’, *IEEE Trans. Ind. Appl.*, vol. 46, no. 4, pp. 1350–1359, Jul. 2010, doi: 10.1109/TIA.2010.2049623.
- [153] N. Tandon and A. Choudhury, ‘A Review of Vibration and Acoustic Measurement Methods for the Detection of Defects in Rolling Element Bearings’, *Tribol. Int.*, vol. 32, no. 8, pp. 469–480, Aug. 1999, doi: 10.1016/S0301-679X(99)00077-8.
- [154] S. Kumar, D. Goyal, and S. S. Dhami, ‘Statistical and Frequency Analysis of Acoustic Signals for Condition Monitoring of Ball Bearing’, *Mater. Today Proc.*, vol. 5, no. 2, Part 1, pp. 5186–5194, Jan. 2018, doi: 10.1016/j.matpr.2017.12.100.
- [155] A. Choudhury and N. Tandon, ‘Application of Acoustic Emission Technique for the Detection of Defects in Rolling Element Bearings’, *Tribol. Int.*, vol. 33, no. 1, pp. 39–45, Jan. 2000, doi: 10.1016/S0301-679X(00)00012-8.
- [156] Z. Wang, X. Wu, X. Liu, Y. Cao, and J. Xie, ‘Research on Feature Extraction Algorithm of Rolling Bearing Fatigue Evolution Stage based on Acoustic Emission’, *Mech. Syst. Signal Process.*, vol. 113, pp. 271–284, Dec. 2018, doi: 10.1016/j.ymssp.2017.08.001.
- [157] G. Feng, D. Zhen, X. Tian, F. Gu, and A. D. Ball, ‘A Novel Method to Improve the Resolution of Envelope Spectrum for Bearing Fault Diagnosis based on a Wireless Sensor Node’, in *Vibration Engineering and Technology of Machinery*, J. K. Sinha, Ed. Springer International Publishing, 2015, pp. 765–775.
- [158] Y. Xu, P. van Vuuren, X. Tang, F. Gu, and A. Ball, ‘A Robust Method to Detect Faults of Rolling Bearings Using Ensemble Average Autocorrelation based Stochastic Subspace Identification’, in *COMADEM*, 2017.
- [159] X. Tian, J. X. Gu, I. Rehab, G. M. Abdalla, F. Gu, and A. Ball, ‘A Robust Detector for Rolling Element Bearing Condition Monitoring based on the Modulation Signal Bispectrum and Its Performance Evaluation Against the Kurtogram’, *Mech. Syst. Signal Process.*, vol. 100, pp. 167–187, 2018.
- [160] F. Hou, J. Chen, and G. Dong, ‘Weak Fault Feature Extraction of Rolling Bearings based on Globally Optimized Sparse Coding and Approximate SVD’, *Mech. Syst. Signal Process.*, vol. 111, pp. 234–250, Oct. 2018, doi: 10.1016/j.ymssp.2018.04.003.
- [161] P. D. McFadden and J. D. Smith, ‘Vibration Monitoring of Rolling Element Bearings by the High-Frequency Resonance Technique - A Review’, *Tribol. Int.*, vol. 17, no. 1, pp. 3–10, Feb.

- 1984, doi: 10.1016/0301-679X(84)90076-8.
- [162] J. O. Smith, *Mathematics of the Discrete Fourier Transform (DFT): With Audio Applications*. Julius Smith, 2007.
  - [163] X. Zhang, 'Application Research on Compressed Sensing in Health Monitoring of Rotating Machinery', National University of Defense Technology, 2015.
  - [164] R. Khanam and S. N. Ahmad, 'Selection of Wavelets for Evaluating SNR, PRD and CR of ECG Signal', *Int J Eng Sci Innov Technol*, vol. 2, pp. 112–119, 2013.
  - [165] Y. Xu, X. Tang, F. Gu, A. Ball, and J. X. Gu, 'Early Detection of Rolling Bearing Faults Using an Auto-Correlated Envelope Ensemble Average', in *Proceedings of the 23rd International Conference on Automation & Computing, (University of Huddersfield, 7-8 September 2017)*, 2017.
  - [166] Y.-F. Wang and P. J. Kootsookos, 'Modeling of Low Shaft Speed Bearing Faults for Condition Monitoring', *Mech. Syst. Signal Process.*, vol. 12, no. 3, pp. 415–426, 1998.
  - [167] A. R. Mohanty, *Machinery Condition Monitoring: Principles and Practices*. Boca Raton, Florida: CRC Press, 2014.
  - [168] V. T. Tran, F. AlThobiani, and A. Ball, 'An Approach to Fault Diagnosis of Reciprocating Compressor Valves Using Teager–Kaiser Energy Operator and Deep Belief Networks', *Expert Syst. Appl.*, vol. 41, no. 9, pp. 4113–4122, Jul. 2014, doi: 10.1016/j.eswa.2013.12.026.
  - [169] M. Elhaj, F. Gu, A. D. Ball, A. Albarbar, M. Al-Qattan, and A. Naid, 'Numerical Simulation and Experimental Study of a Two-Stage Reciprocating Compressor for Condition Monitoring', *Mech. Syst. Signal Process.*, vol. 22, no. 2, pp. 374–389, Feb. 2008, doi: 10.1016/j.ymssp.2007.08.003.
  - [170] B.-S. Yang, W.-W. Hwang, D.-J. Kim, and A. Chit Tan, 'Condition Classification of Small Reciprocating Compressor for Refrigerators Using Artificial Neural Networks and Support Vector Machines', *Mech. Syst. Signal Process.*, vol. 19, no. 2, pp. 371–390, Mar. 2005, doi: 10.1016/j.ymssp.2004.06.002.
  - [171] U. Haba, G. Feng, A. Shaeoub, X. Peng, F. Gu, and A. Ball, 'Motor Current Signature Analysis for the Compound Fault Diagnosis of Reciprocating Compressors', *Int. J. COMADEM*, vol. 20, no. 3, 2017.
  - [172] U. Haba, K. Brethee, O. Hassin, F. Gu, and A. D. Ball, 'An Investigation into Vibration Response for Condition Monitoring of Reciprocating Compressor based on Modulation Signal Spectrum Analysis', *Int. J. COMADEM*, vol. 21, no. 3, pp. 31–37, 2018.
  - [173] U. E. Ogbulafor, G. Feng, Z. Mones, F. Gu, and A. Ball, 'Application of Wavelet Packet Transform and Envelope Analysis to Non-stationary Vibration Signals For Fault Diagnosis of a Reciprocating Compressor', 2017.
  - [174] G. Feng, N. Hu, Z. Mones, F. Gu, and A. D. Ball, 'An Investigation of the Orthogonal Outputs from an on-Rotor MEMS Accelerometer for Reciprocating Compressor Condition Monitoring', *Mech. Syst. Signal Process.*, vol. 76–77, pp. 228–241, Aug. 2016, doi: 10.1016/j.ymssp.2015.12.033.
  - [175] M. Zainab, F. Guojin, T. Xiaoli, H. Usama, G. Fengshou, and B. Andrew, 'A Comparative Study of Gravitational Acceleration Cancellation from on-Rotor MEMS Accelerometers for Condition Monitoring', in *Proceedings of the 24th International Congress on Sound and Vibration, ICSV 2017*, 2017.
  - [176] Wikipedia contributors, 'Ideal Gas Law', *Wikipedia*, 01-Nov-2019. [Online]. Available: [https://en.wikipedia.org/w/index.php?title=Ideal\\_gas\\_law&oldid=924105155](https://en.wikipedia.org/w/index.php?title=Ideal_gas_law&oldid=924105155). [Accessed: 02-Nov-2019].
  - [177] A. A. Ohliger, 'Overview of Infrared in the Petroleum Industry', in *Thermosense XXV*, 2003, vol. 5073, pp. 208–212, doi: 10.1117/12.499019.
  - [178] S. S. Singh and A. K. Singh, 'Condition Monitoring of Reciprocating Air Compressor Using IR Thermography', *IJPAP Vol4901 January 2011*, Jan. 2011.
  - [179] D. G. Lowe, 'Object Recognition from Local Scale-Invariant Features', in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999, vol. 2, pp. 1150–1157 vol.2,

- doi: 10.1109/ICCV.1999.790410.
- [180] R. Azhar, D. Tuwohingide, D. Kamudi, Sarimuddin, and N. Suciati, 'Batik Image Classification Using SIFT Feature Extraction, Bag of Features and Support Vector Machine', *Procedia Comput. Sci.*, vol. 72, pp. 24–30, Jan. 2015, doi: 10.1016/j.procs.2015.12.101.
  - [181] H. A. Alberry, A. A. Hegazy, and G. I. Salama, 'A Fast SIFT based Method for Copy Move Forgery Detection', *Future Comput. Inform. J.*, vol. 3, no. 2, pp. 159–165, Dec. 2018, doi: 10.1016/j.fcij.2018.03.001.
  - [182] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman, 'SIFT Flow: Dense Correspondence across Different Scenes', in *Computer Vision – ECCV 2008*, 2008, pp. 28–42.
  - [183] Z. Ji, J. Theiler, R. Chartrand, and S. P. Brumby, 'SIFT-based Sparse Coding for Large-scale Visual Recognition [J]', *SPIE Def. Secur. Sens.*, 2013.
  - [184] C. Cortes and V. Vapnik, 'Support-Vector Networks', *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.
  - [185] A. Widodo and B.-S. Yang, 'Support Vector Machine in Machine Condition Monitoring and Fault Diagnosis', *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2560–2574, Aug. 2007, doi: 10.1016/j.ymssp.2006.12.007.
  - [186] D. Rongfeng, T. Xiaoli, S. Lin, A. Abdullahi, G. Fengshou, and B. Andrew, 'An Evaluating Study of Using Thermal Imaging and Convolutional Neural Network for Fault Diagnosis of Reciprocating Compressors', in *Proceedings of COMADEM 2019*, 2019.
  - [187] P. Qian, X. Ma, D. Zhang, and J. Wang, 'Data-Driven Condition Monitoring Approaches to Improving Power Output of Wind Turbines', *IEEE Trans. Ind. Electron.*, pp. 1–1, 2018, doi: 10.1109/TIE.2018.2873519.
  - [188] Y. Li, K. Feng, X. Liang, and M. J. Zuo, 'A Fault Diagnosis Method for Planetary Gearboxes under Non-Stationary Working Conditions Using Improved Vold-Kalman Filter and Multi-Scale Sample Entropy', *J. Sound Vib.*, vol. 439, pp. 271–286, Jan. 2019, doi: 10.1016/j.jsv.2018.09.054.
  - [189] H. S. J. Rashid, C. S. Place, D. Mba, R. L. C. Keong, A. Healey, W. Kleine-Beek, and M. Romano, 'Reliability Model for Helicopter Main Gearbox Lubrication System Using Influence Diagrams', *Reliab. Eng. Syst. Saf.*, vol. 139, pp. 50–57, Jul. 2015, doi: 10.1016/j.res.2015.01.021.
  - [190] K. F. Brethee, D. Zhen, F. Gu, and A. D. Ball, 'Helical Gear Wear Monitoring: Modelling and Experimental Validation', *Mech. Mach. Theory*, vol. 117, pp. 210–229, Nov. 2017, doi: 10.1016/j.mechmachtheory.2017.07.012.
  - [191] R. Zhang, F. Gu, H. Mansaf, T. Wang, and A. D. Ball, 'Gear Wear Monitoring by Modulation Signal Bispectrum based on Motor Current Signal Analysis', *Mech. Syst. Signal Process.*, vol. 94, pp. 202–213, Sep. 2017, doi: 10.1016/j.ymssp.2017.02.037.
  - [192] J. Lin and M. J. Zuo, 'Gearbox Fault Diagnosis Using Adaptive Wavelet Filter', *Mech. Syst. Signal Process.*, vol. 17, no. 6, pp. 1259–1269, Nov. 2003, doi: 10.1006/mssp.2002.1507.
  - [193] P. M. T. Marques, C. M. C. G. Fernandes, R. C. Martins, and J. H. O. Seabra, 'Efficiency of a Gearbox Lubricated with Wind Turbine Gear Oils', *Tribol. Int.*, vol. 71, pp. 7–16, Mar. 2014, doi: 10.1016/j.triboint.2013.10.017.
  - [194] J. C. Banks, K. M. Reichard, and M. S. Brought, 'Lubrication Level Diagnostics Using Vibration Analysis', in *2004 IEEE Aerospace Conference Proceedings (IEEE Cat. No.04TH8720)*, 2004, vol. 6, pp. 3528–3534 Vol.6, doi: 10.1109/AERO.2004.1368169.
  - [195] C.-Y. Lee, C.-C. Kuo, R. Liu, I.-H. Tseng, and L.-C. Chang, 'Detection of Gearbox Lubrication Using PSO-based WKNN', *Meas. Sci. Rev.*, vol. 13, no. 3, pp. 108–114, Jun. 2013, doi: 10.2478/msr-2013-0019.
  - [196] K. F. Brethee, F. Gu, and A. D. Ball, 'Condition Monitoring of Lubricant Starvation based on Gearbox Vibration Signatures', *Int. J. COMADEM*, vol. 20, no. 3, Sep. 2017.
  - [197] B. Elforjani, Y. Xu, K. Brethee, Z. Wu, F. Gu, and A. Ball, 'Monitoring Gearbox Using a Wireless Temperature Node Powered by Thermal Energy Harvesting Module', in *2017 23rd International Conference on Automation and Computing (ICAC)*, 2017, pp. 1–6, doi: 10.23919/IConAC.2017.8081994.

- [198] C. M. C. G. Fernandes, P. M. T. Marques, R. C. Martins, and J. H. O. Seabra, 'Gearbox Power Loss. Part I: Losses in Rolling Bearings', *Tribol. Int.*, vol. 88, pp. 298–308, Aug. 2015, doi: 10.1016/j.triboint.2014.11.017.
- [199] C. M. C. G. Fernandes, P. M. T. Marques, R. C. Martins, and J. H. O. Seabra, 'Gearbox Power Loss. Part II: Friction Losses in Gears', *Tribol. Int.*, vol. 88, pp. 309–316, Aug. 2015, doi: 10.1016/j.triboint.2014.12.004.
- [200] C. M. C. G. Fernandes, P. M. T. Marques, R. C. Martins, and J. H. O. Seabra, 'Gearbox Power Loss. Part III: Application to a Parallel Axis and a Planetary Gearbox', *Tribol. Int.*, vol. 88, pp. 317–326, Aug. 2015, doi: 10.1016/j.triboint.2015.03.029.
- [201] S. Abusaad, K. Brethee, M. Assaeh, R. Zhang, F. Gu, and A. D. Ball, 'The Detection of Lubricating Oil Viscosity Changes in Gearbox Transmission Systems Driven by Sensorless Variable Speed Drives Using Electrical Supply Parameters', *J. Phys. Conf. Ser.*, vol. 628, p. 012078, Jul. 2015, doi: 10.1088/1742-6596/628/1/012078.
- [202] P. A. Van Vuuren, Y. Xu, F. Gu, and A. D. Ball, 'Monitoring Gearbox Oil Viscosity by Means of Motor Current Signal Analysis', presented at the WCCM 2017 - 1st World Congress on Condition Monitoring 2017, 2017.
- [203] R. Zhang, X. Gu, F. Gu, T. Wang, and A. Ball, 'Gear Wear Process Monitoring Using a Sideband Estimator based on Modulation Signal Bispectrum', *Appl. Sci.*, vol. 7, no. 3, p. 274, 2017.
- [204] R. Zhang, F. Gu, H. Mansaf, T. Wang, and A. D. Ball, 'Gear wear monitoring by modulation signal bispectrum based on motor current signal analysis', *Mech. Syst. Signal Process.*, vol. 94, pp. 202–213, Sep. 2017, doi: 10.1016/j.ymssp.2017.02.037.
- [205] M. Hamel, A. Addali, and D. Mba, 'Investigation of the Influence of Oil Film Thickness on Helical Gear Defect Detection Using Acoustic Emission', *Appl. Acoust.*, vol. 79, pp. 42–46, May 2014, doi: 10.1016/j.apacoust.2013.12.005.
- [206] X. Sun, T. Wang, R. Zhang, F. Gu, and A. Ball, 'Influence of Frictional Excitations on the Diagnostic Features of Helical Gears under Different Operating Conditions', in *15th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, CM 2018/MFPT 2018*, 2018, pp. 522–531.
- [207] H. Ma, J. Zeng, R. Feng, X. Pang, Q. Wang, and B. Wen, 'Review on Dynamics of Cracked Gear Systems', *Eng. Fail. Anal.*, vol. 55, pp. 224–245, Sep. 2015, doi: 10.1016/j.engfailanal.2015.06.004.
- [208] H. Jiang and F. Liu, 'Dynamic Features of Three-Dimensional Helical Gears under Sliding Friction with Tooth Breakage', *Eng. Fail. Anal.*, vol. 70, pp. 305–322, Dec. 2016, doi: 10.1016/j.engfailanal.2016.09.006.
- [209] N. Baydar, Q. Chen, A. Ball, and U. Kruger, 'Detection of Incipient Tooth Defect in Helical Gears Using Multivariate Statistics', *Mech. Syst. Signal Process.*, vol. 15, no. 2, pp. 303–321, Mar. 2001, doi: 10.1006/mssp.2000.1315.
- [210] T. Touret, C. Changenet, F. Ville, M. Lalmi, and S. Becquerelle, 'On the Use of Temperature for Online Condition Monitoring of Geared Systems – A Review', *Mech. Syst. Signal Process.*, vol. 101, pp. 197–210, Feb. 2018, doi: 10.1016/j.ymssp.2017.07.044.
- [211] O. Janssens, R. Schulz, V. Slavkovikj, K. Stockman, M. Loccufier, R. Van de Walle, and S. Van Hoecke, 'Thermal Image based Fault Diagnosis for Rotating Machinery', *Infrared Phys. Technol.*, vol. 73, pp. 78–87, Nov. 2015, doi: 10.1016/j.infrared.2015.09.004.
- [212] O. Janssens, R. V. de Walle, M. Loccufier, and S. V. Hoecke, 'Deep Learning for Infrared Thermal Image based Machine Health Monitoring', *IEEEASME Trans. Mechatron.*, vol. 23, no. 1, pp. 151–159, Feb. 2018, doi: 10.1109/TMECH.2017.2722479.
- [213] Y. Li, J. X. Gu, D. Zhen, M. Xu, and A. Ball, 'An Evaluation of Gearbox Condition Monitoring Using Infrared Thermal Images Applied with Convolutional Neural Networks', *Sensors*, vol. 19, no. 9, p. 2205, Jan. 2019, doi: 10.3390/s19092205.
- [214] Y. Lu and Y. Wang, 'Monitoring Temperature in Additive Manufacturing with Physics-based Compressive Sensing', *J. Manuf. Syst.*, vol. 48, pp. 60–70, Jul. 2018, doi: 10.1016/j.jmsy.2018.05.010.