Graph-domain features and their application in rotating machinery fault diagnosis

Yiyuan Gao1, 2, *, Wenliao Du2, Xiaoyun Gong2, Dejie Yu1

1 State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha, China
2 Mechanical and Electrical Engineering Institute, Zhengzhou University of Light Industry, Zhengzhou, China

*Corresponding author: yygao@hnu.edu.cn

Abstract. To more effectively extract the non-stationary and non-linear fault features of mechanical vibration signals, a novel fault diagnosis method for rotating machinery is proposed combining time-domain, frequency-domain with graph-domain features. Different from the conventional time-domain and frequency-domain features, the graph-domain features generated from horizontal visibility graphs can extract the fault information hidden in the graph topology. Aiming at the problem that too many features will lead to information redundancy, the Fisher score algorithm is applied to select several of sensitive features which are then fed into the support vector machine to diagnose the faults of rotating machinery. Experimental results indicate features extracted from the three domains can be used to obtain higher diagnosis accuracy than that extracted from any single domain or dual domains.

Keywords: Rotating machinery, fault diagnosis, horizontal visibility graph, graph-domain features.

1. Introduction
In recent years, graph signal processing [1] has provided a new perspective for mechanical fault diagnosis. Based on the graph, new vertex domain, graph spectral domain and vertex-graph spectral domain are developed and applied to the analysis of mechanical vibration signals, which correspond to the traditional time domain, frequency domain and time-frequency domain respectively. The corresponding and transformation relations of these six signal domains are shown in Figure 1. The premise of graph signal processing for mechanical fault diagnosis is to convert the vibration signal into an appropriate graph. To successfully introduce the graph Fourier transform, Ou et al. [2] first mapped a bearing vibration signal into a path graph based on the structural correspondence between time series and path graph. Researches show that horizontal visibility graph (HVG) can characterize time series more exactly and has better performance for retaining the dynamic characteristics of bearing vibration signals than path graph [3, 4]. Hence, vibration signals of rotating machinery are mapped into a HVG in this paper.

Researchers have also constructed many statistical features in the graph domains to express the structural and dynamic characteristics of graphs. At present, the existing graph-domain features...
include the vertex-domain features and graph spectral-domain features. The vertex-domain features, such as mean vertex degree, mean path length and clustering coefficient, have been widely applied in the dynamic analysis of temperature sequence, human heartbeat and other time series [5, 6]. There are two ways to obtain the graph spectral-domain features. The first is to construct the functions of the eigenvalues of the graph matrices, which are mainly used for describing the internal structural characteristics of the chemical molecules and molecular recognition [7, 8]; the other is to analogize the conventional frequency-domain features, which have been used for diagnosing the rolling bearing faults effectively [9].

![Graph spectral-domain features](image)

**Figure 1.** Corresponding and transformation relations of the six signal domains.

Feature extraction is significantly important for rotating machinery fault diagnosis. In addition to the traditional time-domain and frequency-domain features, the existing graph-domain features, including vertex-domain features and graph spectral-domain features, are also extracted after each vibration signal is mapped into a HVG. Too many features will lead to information redundancy, hence the Fisher score algorithm (FS) [10], one of the most widely used supervised feature selection algorithm, is used to measure the sensitivity of all features. Finally, several of more sensitive features are fed into the support vector machine (SVM) [11] to diagnose the rotating machinery faults. Experimental data demonstrate that the proposed method using features in time domain, frequency domain and graph domain has higher diagnostic accuracy than other methods using features in any single domain or dual domains. This is mainly because the feature of graph-domain has the ability to extract fault information hidden in graph topology effectively, and the combination of time-domain, frequency-domain and graph-domain features can provide more fault information.

The remainder of the paper is arranged as follows. The related theory is introduced in Sections 2 and 3. Then the proposed method is described in Section 4. In Section 5, the experimental results are analyzed. Finally, the conclusions of the paper are performed in Section 6.

2. Horizontal visibility graph

The horizontal visibility algorithm is proposed by Luque et al. [12], which can accurately map the dynamics characteristics of a random time series into the topological properties of a HVG. Generally, a graph consists of a finite amount of vertices and edges connecting the vertices. Let \( \{x_i\}_{i=1,2,\ldots,N} \) be a time series with \( N \) any real values, each vertex in the HVG corresponds to each value of the time series in turn. If the subsequent geometric criterion is satisfied in the time series, the vertices \( i \) and \( j \) in the HVG will be connected by an undirected edge.

\[
x_i, x_j > x_n, \quad \forall n \mid i < n < j.
\] (1)
A graph can be expressed by its adjacency matrix $A$. HVG is an undirected and unweighted graph, and hence HVG’s adjacency matrix is a symmetric 0-1 matrix. If there are an edge between any two vertices, the corresponding element in the adjacency matrix is set to 1, otherwise, 0.

Another commonly used matrix for graphs is the Laplacian matrix $L$. The Laplacian matrix $L$ is expressed as $L = D - A$, where $D$ is a diagonal matrix whose $i$th diagonal element equals the number of edges connected to the vertex $i$, that is, the sum of all the elements in the $i$th row or $i$th column of $A$.

In addition, the graph signal is interpreted as a mapping from a set of vertices in the HVG to a set of real numbers in the time series. The graph signal $f$ can be written in the form of a vector as Eq. (2).

$$ f = [f_1, f_2, \cdots, f_{N-1}, f_N]^T \in \mathbb{R}^N. \quad (2) $$

where the signal value $f_i$ in the graph signal (the $i$th data value of the time series) is indexed by the vertex $i$ in the HVG.

3. Graph-domain features

In this paper, the extracted graph-domain features include the vertex-domain features and graph spectral-domain features, which are described below.

3.1. Vertex-domain features

Degree, the number of edges connected to a vertex, is a simple and important concept among the attributes of individual vertices in graphs. The larger the degree of a vertex, the more important it is. If $k_i$ represents the degree of the vertex $i$, the mean vertex degree of the graph is defined as Eq. (3).

$$ K = \frac{1}{N} \sum_{i=1}^{N} k_i \quad (3) $$

The distance $d_{ij}$ between two vertices $i$ and $j$ in the graph is set to the amount of edges connecting the shortest path between these two vertices, and the mean path length shown in Eq. (4) represents the mean of the distances between any two vertices.

$$ P = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij} \quad (4) $$

The clustering coefficient reflects the tightness between the vertices of a graph. The local clustering coefficient $C_i$ is the clustering coefficient of each vertex in the graph, and the global clustering coefficient $C$ shown in Eq. (5) is the mean of the local clustering coefficients of all vertices.

$$ C_i = \frac{2E_i}{k_i(k_i - 1)} \quad C = \frac{1}{N} \sum_{i=1}^{N} C_i \quad (5) $$

where $E_i$ is the number of the closed triangle structures connected to the vertex $i$, and $k_i(k_i - 1)/2$ is the total amount of the triples connected to vertex $i$.

In the field of graph signal processing, there is also an important vertex-domain feature, namely the graph Laplacian quadratic form shown in Eq. (6), which provides a measure of the global smoothness of a graph signal.

$$ S = f^T L f \quad (6) $$

3.2. Graph spectral-domain features

At present, there are two ways to obtain the graph spectral-domain features. The first is to construct the functions of the eigenvalues of the graph matrices. $L$ and $L$ are the two most commonly used matrices. For the HVG mapped from a mechanical vibration signal, both $L$ and $L$ are real symmetric matrices. Therefore, $L$ and $L$ can be directly subjected to the standard orthogonal decomposition to obtain their eigenvalues and eigenvectors. $\lambda_i$ and $\mu_i$ represent the eigenvalues of $L$ and $L$ respectively, and the common graph spectral-domain features are defined in Table 1.

| Table 1. Common graph spectral-domain features. |
|-----------------------------------------------|
| Feature name | Feature expression | Feature name | Feature expression |
|---------------|---------------------|--------------|---------------------|

| Feature name | Feature expression | Feature name | Feature expression |
|---------------|---------------------|--------------|---------------------|

| Feature name | Feature expression | Feature name | Feature expression |
|---------------|---------------------|--------------|---------------------|

3
Graph energy \( A_1 = \sum_{i=1}^{N} |2_i| \)
Estrada index \( A_2 = \sum_{i=1}^{N} e^{\mu_i} \)
Laplacian energy \( L_1 = \sum_{i=1}^{N} |\mu_i - 2M/N| \)
Laplacian-energy-like invariant \( L_2 = \sum_{i=1}^{N} \sqrt{\mu_i} \)
The sum of powers of the Laplacian eigenvalues of graphs \( L_3 = \sum_{i=1}^{N} |\mu_i|^{13} \)

The other is to analogize the conventional frequency-domain features. Similar to the definition of the traditional Fourier transform, the graph Fourier transform of a graph signal \( f \) shown in Eq. (7) can be expressed as \( \hat{f} \) expanded with the eigenvectors of \( L \) [1].

\[
\hat{f}(\mu_i) = \langle f, \psi_i \rangle = \sum_{i=1}^{N} f_i \psi_{i}^* \tag{7}
\]
where \( \hat{f}(\mu_i) \) and \( \psi_i \) represent the graph spectrum coefficient and the eigenvector corresponding to the eigenvalue \( \mu_i \) of \( L \), respectively. By analogizing with the frequency-domain features in the reference [13], the corresponding 13 graph spectral-domain features are defined in Table 2.

**Table 2.** Thirteen graph spectral-domain features corresponding to frequency-domain features.

| Feature expression | Feature expression | Feature expression |
|-------------------|-------------------|-------------------|
| \( G_1 = \frac{1}{N} \sum_{i=1}^{N} \hat{f}(\mu_i) \) | \( G_2 = \frac{1}{N-1} \sum_{i=1}^{N} (\hat{f}(\mu_i) - G_1)^2 \) | \( G_3 = \sum_{i=1}^{N} (\hat{f}(\mu_i) - G_1)^3 \) |
| \( G_4 = \frac{\sum_{i=1}^{N} (\hat{f}(\mu_i) - G_1)^4}{NG_1^2} \) | \( G_5 = \sum_{i=1}^{N} \hat{f}(\mu_i) \) | \( G_6 = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{\mu_i - G_1}{NG_1^3} \right) \hat{f}(\mu_i) \) |
| \( G_6 = \frac{\sum_{i=1}^{N} (\mu_i - G_1)^3 \hat{f}(\mu_i)}{NG_1^3} \) | \( G_7 = \sum_{i=1}^{N} (\mu_i - G_1)^3 \hat{f}(\mu_i) \) | \( G_8 = \frac{1}{N} \sum_{i=1}^{N} (\mu_i - G_1) \hat{f}(\mu_i) \) |
| \( G_{11} = \sqrt{\frac{\sum_{i=1}^{N} \mu_i \hat{f}(\mu_i)}{\sum_{i=1}^{N} \hat{f}(\mu_i)}} \) | \( G_{12} = \sqrt{\frac{\sum_{i=1}^{N} \mu_i \hat{f}(\mu_i)}{\sum_{i=1}^{N} \mu_i ^2 \hat{f}(\mu_i)}} \) | \( G_{13} = \sqrt{\frac{\sum_{i=1}^{N} \mu_i ^2 \hat{f}(\mu_i)}{\sum_{i=1}^{N} \hat{f}(\mu_i) \sum_{i=1}^{N} \mu_i ^4 \hat{f}(\mu_i)}} \) |

### 4. The proposed fault diagnosis method for rotating machinery

Based on the time-domain, frequency-domain and graph-domain features, a novel fault diagnosis method is proposed for rotating machinery. The proposed method is described as follows:

1. **Signal collection.** The vibration signals of rotating machinery under different conditions are collected. The vibration signal of each condition is directly extracted. Next, each sample is mapped into a HVG, and then the corresponding graph signal indexed by the HVG can be obtained. Finally, the 24 graph-domain features including the 4 vertex-domain features \( (K, P, C, S) \) and the 20 graph spectral-domain features \( (A_{1-2}, L_{1-5}, G_{1-13}) \) are also extracted. For each sample, a total of 48 features are extracted from the time domain, frequency domain and graph domain. Notably, all features are normalized to eliminate the dimensional influence between feature data.

2. **Feature extraction.** First, 11 time-domain features and 13 frequency-domain features in the reference [13] are directly extracted. Next, each sample is mapped into a HVG, and then the corresponding graph signal indexed by the HVG can be obtained. Finally, the 24 graph-domain features including the 4 vertex-domain features \( (K, P, C, S) \) and the 20 graph spectral-domain features \( (A_{1-2}, L_{1-5}, G_{1-13}) \) are also extracted. For each sample, a total of 48 features are extracted from the time domain, frequency domain and graph domain. Notably, all features are normalized to eliminate the dimensional influence between feature data.

3. **Feature selection.** The FS algorithm is applied to calculate the sensitivity of all features, and several of more sensitive features are chosen as the fault features of rotating machinery. The training feature subsets and testing feature subsets are established using those selected features.

4. **Pattern recognition.** After feeding the training feature subsets into the SVM for training, the testing feature subsets are fed into the trained SVM model, and then the state of each testing sample is identified according to the output results.

Figure 2 illustrates the flow chart of the proposed method.
5. Experimental validations

5.1. Vibration data description
To verify the effectiveness of the proposed method, vibration signals of rotating machinery under eight working conditions are collected on the test bench of Hunan University. Bevel gears and roller bearings are selected as the types of gears and bearings respectively. In this experiment, vibration signals of different bearing and gear faults are picked up by the accelerometer placed on the bearing bases under the condition that the motor speed is 1200rpm, the sampling frequency is 10240Hz and the load is 4Nm. The fault sizes of bearing and gear are 0.2mm and 0.6mm, respectively. The amount of samples for each working condition is set to 50, and the length of each sample is set to 1024. In each working condition, 10 samples are randomly chosen as training samples, and the remaining 40 samples are testing samples. A total of 80 training samples and 320 testing samples are collected. The detailed description about eight working conditions is listed in Table 3.

| Bearing condition | Gear condition | Number of training/ testing samples | Condition label |
|-------------------|---------------|------------------------------------|-----------------|
| Normal            | Normal        | 10/40                              | 1               |
| Inner race fault  | Normal        | 10/40                              | 2               |
| Outer race fault  | Normal        | 10/40                              | 3               |
| Cage fault        | Normal        | 10/40                              | 4               |
| Normal            | Gear crack    | 10/40                              | 5               |
| Inner race fault  | Gear crack    | 10/40                              | 6               |
| Outer race fault  | Gear crack    | 10/40                              | 7               |
| Cage fault        | Gear crack    | 10/40                              | 8               |

5.2. Diagnosis results and analysis
The proposed method is applied to diagnose the different bearing and gear faults. The time-domain, frequency-domain and graph-domain features are all extracted, and several of more sensitive features selected by the FS algorithm are fed into the SVM to identify the eight working conditions. To further test the advantages of the proposed method, the features extracted from any single domain and dual domains are also used to identify the eight working conditions. There are a total of seven feature extraction methods, which are called time domain (T), frequency domain (F), graph domain (G), time and frequency domains (T+F), time and graph domains (T+G), frequency and graph domains (F+G), time and frequency and graph domains (T+F+G), respectively. To avoid the interference caused by unexpected situations, each feature extraction method was tested for 10 trials. Figure 3 shows the multi-class confusion matrix of the proposed method for the first trial. Figure 4 and Table 4 display the diagnosis results of different features with FS feature selection for 10 trials. 

![Figure 3](image-url)  
**Figure 3.** Multi-class confusion matrix of the proposed method for the first trial.  

![Figure 4](image-url)  
**Figure 4.** Diagnosis accuracy of different features with FS feature selection for each trial.  

**Table 4.** Diagnosis results of different features with FS feature selection for 10 trials.

| Different features     | Diagnosis results          |
|-----------------------|----------------------------|
|                       | Average accuracy (%)       | Standard deviation |
| Time domain (T)       | 74.50 (2384/3200)          | 1.80               |
| Frequency domain (F)  | 93.78 (3001/3200)          | 0.60               |
| Graph domain (G)      | 86.34 (2763/3200)          | 1.08               |
Time and Frequency domains (T+F)  94.41 (3021/3200)  0.60
Time and graph domains (T+G)  89.34 (2859/3200)  1.51
Frequency and graph domains (F+G)  96.97 (3103/3200)  0.49
Time and Frequency and graph domains (T+F+G)  97.72 (3127/3200)  0.42

It can be seen from Figure 3 that only a small amount of testing samples in conditions 1, 5 and 8 are misdiagnosed, and all testing samples in other conditions are diagnosed correctly. It can be observed from Figure 4 that features extracted from time domain, frequency domain and graph domain can be used to obtain higher diagnosis accuracy than that extracted from any single domain or dual domains in each trial. As can be seen from Table 4, the proposed method (i.e. features extracted from time domain, frequency domain and graph domain) obtains the highest average accuracy and the smallest standard deviation. These results indicate that the proposed method is effective and superior in diagnosing different faults of key rotating components.

In the absence of feature selection, the diagnosis results of each feature extraction method for 10 trials are listed in Table 5. Comparing Table 5 with Table 4, it can be concluded that for all feature extraction methods, the FS features selection is helpful for improving the average diagnosis accuracy, and reducing the standard deviation. The reason is that too many features may result in information redundancy, hence feature selection is necessary and important.

Table 5. Diagnosis results of different features without feature selection for 10 trials.

| Different features               | Diagnosis results |
|---------------------------------|-------------------|
|                                 | Average accuracy (%) | Standard deviation |
| Time domain (T)                 | 70.38 (2252/3200)   | 2.58               |
| Frequency domain (F)            | 89.38 (2860/3200)   | 1.59               |
| Graph domain (G)                | 80.03 (2561/3200)   | 2.54               |
| Time and Frequency domains (T+F)| 91.31 (2922/3200)   | 1.54               |
| Time and graph domains (T+G)    | 82.75 (2648/3200)   | 2.29               |
| Frequency and graph domains (F+G)| 90.78 (2905/3200)  | 3.28               |
| Time and Frequency and graph domains (T+F+G)| 91.56 (2930/3200)| 2.74               |

To further research the impact of different training samples on all feature extraction methods, 2, 4, 6 and 8 samples are also chosen as training samples for each working condition, and remaining samples are testing samples. The FS feature selection and SVM classification are used for fault diagnosis in all cases. The average accuracy of each feature extraction method for 10 trials under different training and testing samples is drawn in Figure 5. As the amount of training samples increases, the average accuracy of each feature extraction method is gradually improved. Our proposed method using features extracted from time domain, frequency domain and graph domain, still has the highest average accuracy even with fewer training samples.

Figure 5. Average accuracy of different features under different training and testing samples.
6. Conclusions
Combining time-domain, frequency-domain and graph-domain features, a new fault diagnosis method for rotating machinery is proposed in this paper. The following conclusions are drawn by analyzing the experimental results:

1. The graph-domain features generated from horizontal visibility graphs can extract the fault information hidden in the graph topology effectively.

2. Features extracted from time domain, frequency domain and graph domain can be used to obtain higher diagnosis accuracy than features extracted from any single domain or dual domains.

3. Too many features can result in information redundancy. The FS feature selection method is helpful for improving the classification performance of SVM.

Acknowledgments
This work was financially supported by the National Natural Science Foundation of China (No. 51875182) and Foundation for University Key Teacher of Henan Educational Department (No. 2018GGJS091).

References
[1] D.I. Shumun, S.K. Narang, P. Frossard, A. Ortega, P. Vanderghynst, The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains, IEEE Signal Proc. Mag. 30 (2013) 83-98.
[2] L. Ou, D.J. Yu, H.J. Yang, A new rolling bearing fault diagnosis method based on GFT impulse component extraction, Mech. Syst. Signal Process. 81 (2016) 162–182.
[3] Y.Y. Gao, D.J. Yu, Total variation on horizontal visibility graph and its application to rolling bearing fault diagnosis, Mech. Mach. Theory 147 (2020) 103768.
[4] Y.Y. Gao, D.J. Yu, H.J. Wang, Fault diagnosis of rolling bearings using weighted horizontal visibility graph and graph Fourier transform, Measurement 149 (2020) 107036.
[5] L. Zhou, Z.Q. Gong, R. Zhi, G.L. Feng, An approach to research the topology of chinese temperature sequence based on complex network, Acta Physica Sinica 57 (2008) 7380-7389.
[6] Z.G. Shao, Network analysis of human heartbeat dynamics, Appl. Phys. Lett. 96 (2010) 073703.
[7] E. Estrada, Characterization of the folding degree of proteins, Bioinformatics 18 (2002) 697-704.
[8] D. Stevanovic, A. Ilic, On the Laplacian coefficients of unicyclic graphs, Linear Algebra Appl. 430 (2009) 2290-2300.
[9] L. Ou, D.J. Yu, Path graph Fourier transformation and its applications to rolling bearing fault diagnosis, J. Mech. Eng. 51 (2015) 76–83 (in Chinese).
[10] X. He, D. Cai, P. Niyogi, Laplacian score for feature selection, Advances in Neural Information Processing Systems, Vancouver, 2005, pp. 507-514.
[11] P. Gangsar, R. Tiwari, Comparative investigation of vibration and current monitoring for prediction of mechanical and electrical faults in induction motor based on multiclass-support vector machine algorithms, Mech. Syst. Signal Process. 94 (2017) 464-481.
[12] B. Luque, L. Lacasa, F. Ballesteros, J. Luque, Horizontal visibility graphs: Exact results for random time series, Phys. Rev. E 80 (2009) 046103.
[13] F. Li, J.X. Wang, B.P. Tang, D.Q. Tian, Life grade recognition method based on supervised uncorrelated orthogonal locality preserving projection and K-nearest neighbor classifier, Neurocomputing 138 (2014) 271-282.