Journal of Civil Structural Health Monitoring (2023) 13:1–14
https://doi.org/10.1007/s13349-022-00616-x

ORIGINAL PAPER

Vibration-based and computer vision-aided nondestructive health condition evaluation of rail track structures

Shaohua Wang[1] · Hao Zheng[1] · Lihua Tang[1] · Zhaoyu Li[1] · Renda Zhao[2] · Yuqian Lu[1] · Kean C. Aw[1]

Received: 20 April 2022 / Revised: 29 July 2022 / Accepted: 30 July 2022 / Published online: 13 August 2022
© The Author(s) 2022

Abstract

In railway engineering, monitoring the health condition of rail track structures is crucial to prevent abnormal vibration issues of the wheel–rail system. To address the problem of low efficiency of traditional nondestructive testing methods, this work investigates the feasibility of the computer vision-aided health condition monitoring approach for track structures based on vibration signals. The proposed method eliminates the tedious and complicated data pre-processing including signal mapping and noise reduction, which can achieve robust signal description using numerous redundant features. First, the method converts the raw wheel–rail vibration signals directly into two-dimensional grayscale images, followed by image feature extraction using the FAST-Unoriented-SIFT algorithm. Subsequently, Visual Bag-of-Words (VBoW) model is established based on the image features, where the optimal parameter selection analysis is implemented based on fourfold cross-validation by considering both recognition accuracy and stability. Finally, the Euclidean distance between word frequency vectors of testing set and the codebook vectors of training set is compared to recognize the health condition of track structures. For the three health conditions of track structures analyzed in this paper, the overall recognition rate could reach 96.7%. The results demonstrate that the proposed method performs higher recognition accuracy and lower bias with strong time-varying and random vibration signals, which has promising application prospect in early-stage structural defect detection.

Keywords Rail track structure · Health condition monitoring · Vibration signal processing · Image feature extraction · Visual Bag-of-Words model

1 Introduction

With the rapid development of urban rail transportation, ballastless track has been widely used in many metro lines because of its advantages of low periodic maintenance and greater vehicle stability [1]. For metro shield tunnels, the track structures consist of rail, fasteners, track slab, subgrade and tunnel lining. The track slab and subgrade are usually constructed as monolithic cast-in-place concrete structures on the tunnel lining. However, under continuous cyclic trainloads, cracks often occur at the interface between the subgrade and the tunnel lining near the expansion joints of track slab [2]. Over time, this type of cracks will further propagate along the cross-sectional interface, ultimately affecting damage to other vehicle components or track structures [3]. Before such damage issues can be detected visually, the track structures have not been able to achieve normal service performance, which will endanger the operation of trains. Therefore, health condition monitoring and damage detection of rail track structures, especially the early-stage damage identification, is critically important [4–6].

As the vibration properties (e.g., the natural frequency, amplitude, and damping ratio) of structural system vary with the mechanical parameters and boundary conditions, many researchers attempted to utilize the variation of vibration properties to detect the structural health condition [7–10]. And, the commonly used methods for vibration signal analysis included empirical mode decomposition (EMD) [11, 12], short-time Fourier transform (STFT) [13, 14], wavelet transform (WT) [15, 16], etc. In general, the traditional vibration

Shaohua Wang and Hao Zheng have contributed equally to this work and share first authorship.

✉ Lihua Tang
l.tang@auckland.ac.nz

[1] Department of Mechanical and Mechatronics Engineering, The University of Auckland, 1010 Auckland, New Zealand

[2] School of Civil Engineering, Southwest Jiaotong University, Chengdu 610031, China

signal-based health condition evaluation in the frequency domain required pre-processing of the raw signals. Hence, complex mapping and transformations were performed to filter out noise and redundant information to highlight the information related to the damage. Finally, eigenvalues or feature vectors that best characterize the damage were extracted.

However, the noise is a challenging issue to deal with. In practical engineering, wheel–rail impact vibration contains multiple excitation sources, with time-varying and random characteristics, which means that there are numerous environmental noises in the vibration signals. These strong environmental noises tend to mask the subtle variations caused by damage in the vibration signals [7, 17]. For such complex signals, traditional signal processing methods have shown shortcomings with low efficiency and poor generality. To address these issues, researchers developed some signal visualization methods by converting the vibration signals into two-dimensional grayscale images for analysis. The image features could then be extracted directly for damage detection and recognition without noise reduction [18–20]. These analysis methods demonstrated better robustness in strong noises environments [21, 22]. Specifically, Do et al. [23] used the scale invariant feature transform (SIFT) algorithm to extract the local feature vector of 2D images based on vibration signals for the detection and diagnosis of asynchronous motor faults. Zheng et al. [24] proposed a novel FAST-Unoriented-SIFT algorithm for extracting planetary gear fault feature values, which was more efficient and had more extracted features compared to SIFT algorithm. Meanwhile, bag-based representations have been widely used to compute the similarity between digital objects by characterizing the frequency of occurrence of object features [25]. Among them, Visual Bag-of-Words (VBoW) models have been effectively used for image feature clustering and classification tasks [26]. Qi et al. [27] used VBoW model to extract surface features of soil to efficiently characterize textural information. Yang et al. [28] extracted the vibration signal features of rotating machinery based on the VBoW model for detection and achieved good results. Zheng et al. [29] found that VBoW model had good recognition efficiency and accuracy when training small datasets of planetary gear fault features.

Inspired by the aforementioned research, in this work, a vibration-based and computer vision-aided health condition evaluation method for rail track structures is proposed. Different from the other traditional methods, this method could evade the noise reduction process, directly extract numerous redundant features, and use the redundancy of features to achieve a more robust signal description. The rest of the paper is organized as follows. Section 2 focuses on the methodology used in this work. It consists of three parts: grayscale image processing of rail vibration signals, image feature extraction and the establishment of VBoW model with optimal parameters. Section 3 describes the vibration signal acquisition during field test. Section 4 analyzes and discusses the experiment results. Specifically, the optimal parameters are first obtained from the training dataset by fourfold cross-validation. Then, the reliability and real-time of the proposed method are verified using the testing dataset. Finally, the advantages of the method used in this work is highlighted by performance comparison. Conclusive remarks are given in Sect. 5. By employing the vibration signal visualization, encoding and classification methods, the feature information of different track structural health conditions could be effectively identified.

2 Methodology

2.1 Vibration signal visualization

This section introduces the method to convert the raw vibration signals to the pixels of the grayscale image [30]. Firstly, the amplitude of one-dimensional time-domain vibration signal **s** is normalized to the range of [1, 255], and the normalized signal s_{norm} is obtained according to Eq. (1).

$$\mathrm{s}_{\mathrm{norm}} = \frac{127}{\max(|\mathrm{s}|)} \cdot \mathrm{s} + 128 \tag{1}$$

where $\mathrm{s} = [s_1, s_2, \cdots, s_n]$ is the raw vibration signal, n represents the length of the signal, and s_n is the magnitude of the n-th value of the signal.

Then, the normalized signal $\mathbf{s}_{norm}$ is sequentially mapped to one pixel of the two-dimensional grayscale image **G** as the grayscale value. The grayscale value of each pixel of the grayscale image **G** is listed according to Eq. (2).

$$\mathrm{G}(x, y) = \mathrm{s}_{\mathrm{norm}}(N \cdot (x-1) + y), x \in \{1, 2, \ldots, M\}, y \in \{1, 2, \ldots, N\} \tag{2}$$

where $\mathbf{G}(x, y)$ denotes the grayscale value of the pixel in the x-th row and y-th column of the grayscale image **G**. $\mathrm{s}_{\mathrm{norm}}(i)$ denotes the i-th value of $\mathbf{s}_{\mathrm{norm}}$. The obtained grayscale image **G** has $M \times N$ pixel points (i.e., M and N represent the numbers of the rows and columns of the grayscale image, respectively). The conversion method is shown in Fig. 1.

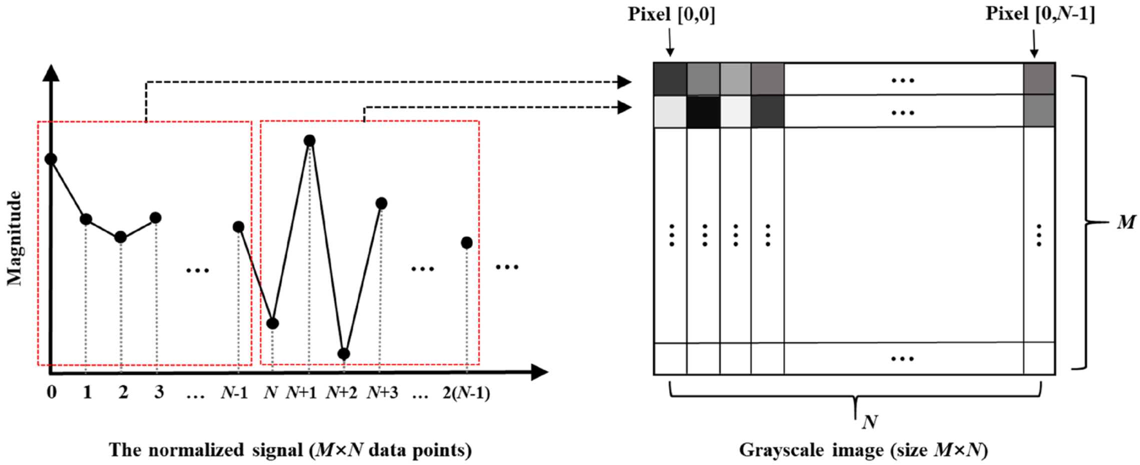

Pixel [0,0]
Pixel [0,N-1]
Magnitude
0 1 2 3 ... N-1 N N+1 N+2 N+3 ... 2(N-1)
...
M
N
The normalized signal (M×N data points)
Grayscale image (size M×N)

Fig. 1 Conversion of vibration signal to grayscale image

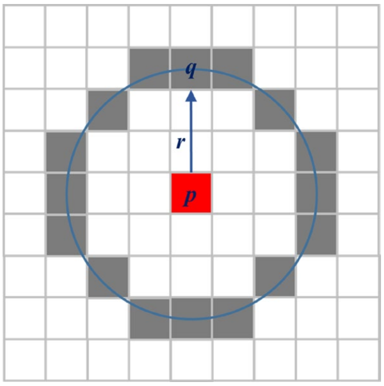


Fig. 2 A diagram of FAST feature point detection

2.2 Image feature extraction

The image feature extraction method utilized in this work combines the features from Accelerated Segment Test (FAST) algorithm [31] and the excellent feature description capability of the Unoriented Scale Invariant Feature Transform (SIFT) algorithm [32]. The FAST-Unoriented-SIFT algorithm has shown good performance and robustness in noisy environment [24].

2.2.1 FAST feature point detection

The pixel points with large difference in grayscale value from the surrounding pixels are indicated as feature points [31]. The status $S_{p\to q}$ of center pixel p with respect to pixel q is calculated according to Eq. (3). p is a candidate feature point. q is any pixel on a circle around p with r being the radius, as shown in Fig. 2.

$$S_{p\to q} = \begin{cases} d_q, G_q \le G_p - t_p \\ s_q, G_p - t_p \le G_q \le G_p + t_p \\ b_q, G_p + t_p \le G_q \end{cases} \tag{3}$$

where G is the grayscale value of pixel point, n_c is the corner detector on this circle and t_p is the threshold value. The status d_q or b_q represents that the point pixel on this circle is darker or brighter than the center pixel respectively, and s_q represents that the point pixel is similar to the center pixel. In other words, if the number of statuses d_q and b_q in $S_{p\to q}$ is greater than the corner detector n_c, the center pixel p of this circle is considered as a feature point.

In this work, $r=4$ and $n_c=12$ were taken according to Ref. [31]. The threshold value t_p determines the number of extracted feature points, and the related analysis is shown in Sect. 4.1.1.

2.2.2 Unoriented-SIFT feature description

After detecting the feature points, 16 × 16 window around one feature point p is firstly taken, and then the gradient modulus and direction of each pixel point within this window are calculated according to Eqs. (4) and (5).

$$\mathrm{m}(x, y) = \sqrt{(\mathrm{G}(x+1, y) - \mathrm{G}(x-1, y))^2 + (\mathrm{G}(x, y+1) - \mathrm{G}(x, y-1))^2} \tag{4}$$

$$\theta(x, y) = \arctan \frac{\mathrm{G}(x, y+1) - \mathrm{G}(x, y-1)}{\mathrm{G}(x+1, y) - \mathrm{G}(x-1, y)} \tag{5}$$

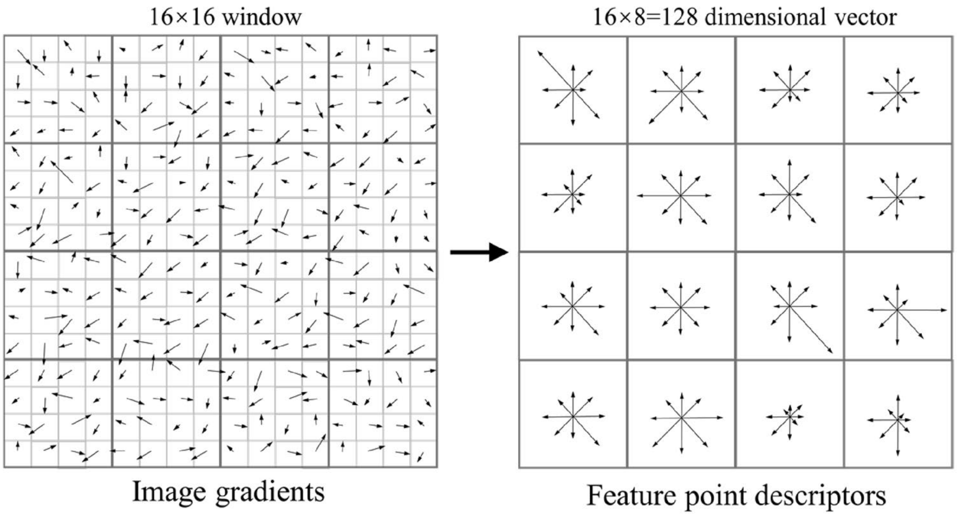


Fig. 3 The process of Unoriented-SIFT feature description

where m(x, y) and $\theta(x, y)$ denote the modulus and direction of the gradient, respectively.

After that, the window is divided into 16 cubes of 4×4, and the modulus of the gradient in each cube in 8 directions (45 degrees in each direction) is counted. An 8-dimensional description vector is then obtained for each cube, and a 128 (16×8) dimensional description vector d is finally obtained for the whole window, which is the Unoriented-SIFT description vector of feature point p. The process of Unoriented-SIFT feature description is shown in Fig. 3, where the direction of the arrow represents the gradient direction of the pixel, and the length of the arrow represents the modulus of the gradient.

2.3 VBoW modeling

In the VBoW model, each image feature (i.e., the 128-dimensional description vector d mentioned in Sect. 2.2.2) is quantized as one word, and each grayscale image is viewed as a bag full of words. Then, similar image features are clustered into one class by k-means algorithm, and the vector of the clustering center is quantized as one keyword. Therefore, the VBoW modeling process includes keywords acquisition and grayscale image representation.

2.3.1 Keywords acquisition

By assuming that N image features $\mathrm{D}_N=\{\mathrm{d}_1, \mathrm{d}_2, \ldots, \mathrm{d}_N\}$ are extracted from one grayscale image, the k-means algorithm eventually partitions D_N into κ clusters $\{\mathrm{C}_i | i = 1, 2, \ldots, \kappa\}$ by minimizing the squared error according to Eq. (6).

$$e=\sum_{i=1}^{\kappa}\sum_{\mathrm{d}_j\in\mathrm{C}_i}\left\|\mathrm{d}_j-\mu_i\right\|^2 \tag{6}$$

where $\mathrm{C}_{i'} \cap \mathrm{C}_{i''\neq i'} = \emptyset$ and $\mathrm{D}_N = \cup_{i=1}^{\kappa}\mathrm{C}_i$. d_j denotes one image feature in C_i. μ_i denotes the cluster center in C_i. The smaller the value of e is, the better the clustering result is. The calculation process is shown in Algorithm 1.

Algorithm 1 General framework of keywords acquisition

```
Input: the dataset $\mathbf{D}_N=\{\mathbf{d}_1,\mathbf{d}_2,\ldots,\mathbf{d}_N\}$, the number of cluster centers $\kappa$, the permissible error $e_\mathrm{p}$
Output: the clustering result (i.e., keywords acquisition)
Initialize cluster centers $[\boldsymbol{\mu}_1;\boldsymbol{\mu}_2;\cdots;\boldsymbol{\mu}_\kappa]$
do
for $i$=1: $\kappa$
    for $j$ =1: $N$
        if $\left\|\mathbf{d}_j-\boldsymbol{\mu}_i\right\|^2=\min_i\left\{\left\|\mathbf{d}_j-\boldsymbol{\mu}_i\right\|^2\right\}$
            $w_{ji}$=1 (i.e. $\mathbf{d}_j\in\mathbf{C}_i$)
        else
            $w_{ji}$=0 (i.e. $\mathbf{d}_j\notin\mathbf{C}_i$)
        end if
    end for
        Update the cluster centers $\boldsymbol{\mu}_i=\frac{\sum_{j=1}^{N}w_{ji}\mathbf{d}_j}{\sum_{j=1}^{N}w_{ji}}$
        Calculate the squared error $e=\sum_{i=1}^{\kappa}\sum_{\mathbf{d}_j\in\mathbf{C}_i}\left\|\mathbf{d}_j-\boldsymbol{\mu}_i\right\|^2$
    until $e<e_\mathrm{p}$
end for
Obtain keywords $\boldsymbol{\mu}=[\boldsymbol{\mu}_1;\boldsymbol{\mu}_2;\cdots;\boldsymbol{\mu}_\kappa]$ of VBoW model
```

2.3.2 Grayscale image representation

After the keyword acquisition, the frequency of keywords in each grayscale image is counted to obtain the word frequency vector. In the training set, the codebook vector is obtained by averaging the word frequency vectors represented by all grayscale images related to the same health condition, i.e., each codebook vector describes one health condition of the track structure. By calculating the Euclidean distance between the word frequency vector extracted from each grayscale image in the testing set and the codebook vectors of three health conditions in the training set, each word frequency vector is categorized to one health condition according to the minimum Euclidean distance to the codebook vector. The process is shown in Fig. 4.

2.3.3 Algorithm of selecting optimal number of keywords

The number of keywords κ is a critical hyper-parameter of VBoW model. To optimize the recognition accuracy and stability of the proposed health condition evaluation method, fourfold cross-validation is used to select the optimum of κ during training step. Also, the value function $\varepsilon(\kappa)$, positively correlated with recognition rate and negatively correlated with standard deviation, is constructed to evaluate the merit of the selection of κ (see Eq. (7)).

$$\varepsilon(\kappa)=1-\frac{P_\kappa}{e^{S_\kappa}} \tag{7}$$

where P_κ and S_κ denote the overall recognition rate and standard deviation of the fourfold cross-validation results. The smaller the value of $\varepsilon(\kappa)$ is, the better the recognition

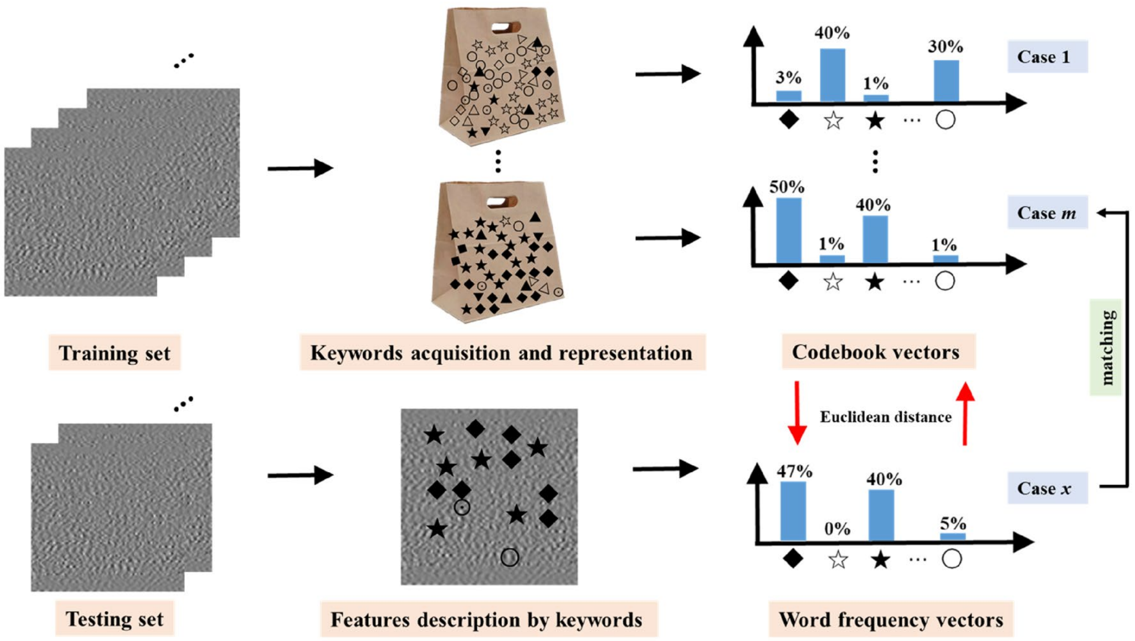


Fig. 4 The process of grayscale images representation (Notes: ◆,☆,★ … denote keywords)

accuracy and stability of the model corresponding to this hyper-parameter are.

To improve the computational efficiency, firstly the recommended data ranges of κ are obtained by equal-interval down-sampling method. Next, the value function for each κ in these ranges are calculated, and the κ with the minimum value function is the optimum κ_{opt}. The overview of the algorithm to seek the optimal number of keywords is shown in Algorithm 2.

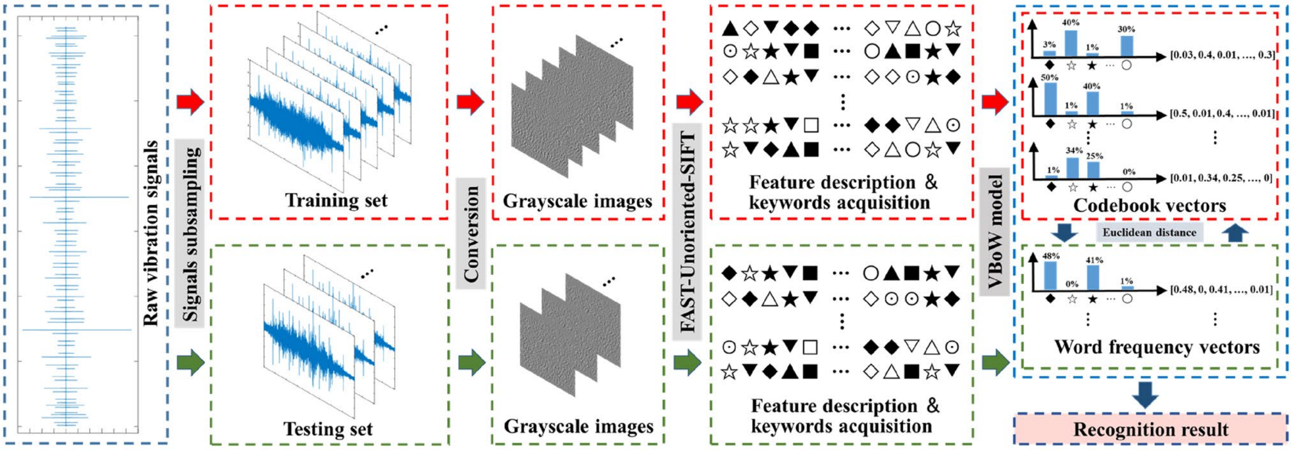


Fig. 5 The proposed rail track structural health condition evaluation method

Algorithm 2 General framework of the optimal parameter selection algorithm

```
Input: The training set, the maximum number of keywords $\kappa_{\max}$, the minimum of value function $\varepsilon_{\min}$,
the minimal difference of two adjacent value function $\varepsilon_{\mathrm{adj}}$, the interval of down-sampling $M$
Output: The optimal number of keywords $\kappa_{\mathrm{opt}}$
Surrogate model building: Using the training set, build the proposed health condition evaluation model in Fig.4
M-time interval down-sampling $\boldsymbol{\kappa}(tM)$, $t$=floor($\kappa_{\max}/M$)
for $i = 1: t-1$
    Find the value function $\varepsilon(\boldsymbol{\kappa}(iM))$: using the surrogate model, calculate $\varepsilon(\boldsymbol{\kappa}(iM))$ by Eq.(7)
    if $\varepsilon(\boldsymbol{\kappa}(iM)) < \varepsilon_{\min}$ and $\varepsilon(\boldsymbol{\kappa}((i+1)M)) < \varepsilon_{\min}$ and $|\varepsilon(\boldsymbol{\kappa}((i+1)M))-\varepsilon(\boldsymbol{\kappa}(iM))| < \varepsilon_{\mathrm{adj}}$ then
        Store the data range from $\boldsymbol{\kappa}(iM)$ to $\boldsymbol{\kappa}((i+1)M)$
    end if
    for $j= iM : (i+1)M$
        if $\varepsilon(\boldsymbol{\kappa}(j)) < \varepsilon_{\min}$
            Store $\boldsymbol{\kappa}(j)$
        else
            $j=j+1$
        end if
        Update the optimum of $\kappa_{\mathrm{opt}}$
    end for
end for
```

2.4 Rail track structural condition identification

In this work, the proposed method of rail track structural health condition consists of the following steps: Firstly, the raw vibration signals are subsampled and converted into grayscale images using the method in Sect. 2.1, which are divided into two parts, i.e., the training set and the testing set. Then, the FAST-Unoriented-SIFT algorithm in Sect. 2.2 is used to extract features from the two datasets. After that, the VBoW model is built, respectively for the training set and the testing set using the method in Sect. 2.3. Specifically, based on vibration signals conversion and feature extraction, the optimum of keywords κ_{opt} is obtained by the algorithm described in Sect. 2.3.3. By calculating the Euclidean distance between the word frequency vector extracted from each grayscale image in the testing set and the codebook vectors of three health conditions in the training set, each word frequency vector is categorized to one health condition according to the minimum Euclidean distance to the codebook vector. The proposed rail track structural condition evaluation process is summarized in Fig. 5.

3 Field test and vibration signal acquisition

The tested metro line has a total length of 80 km and the maximum operating speed is 80 km/h. To be specific, the tested rail track system is located in a single-hole, single-line circular shield tunnel. Two rails of approximately 60 kg/m

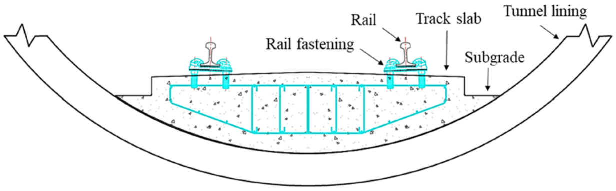


Fig. 6 Cross-sectional layout of rail track structure

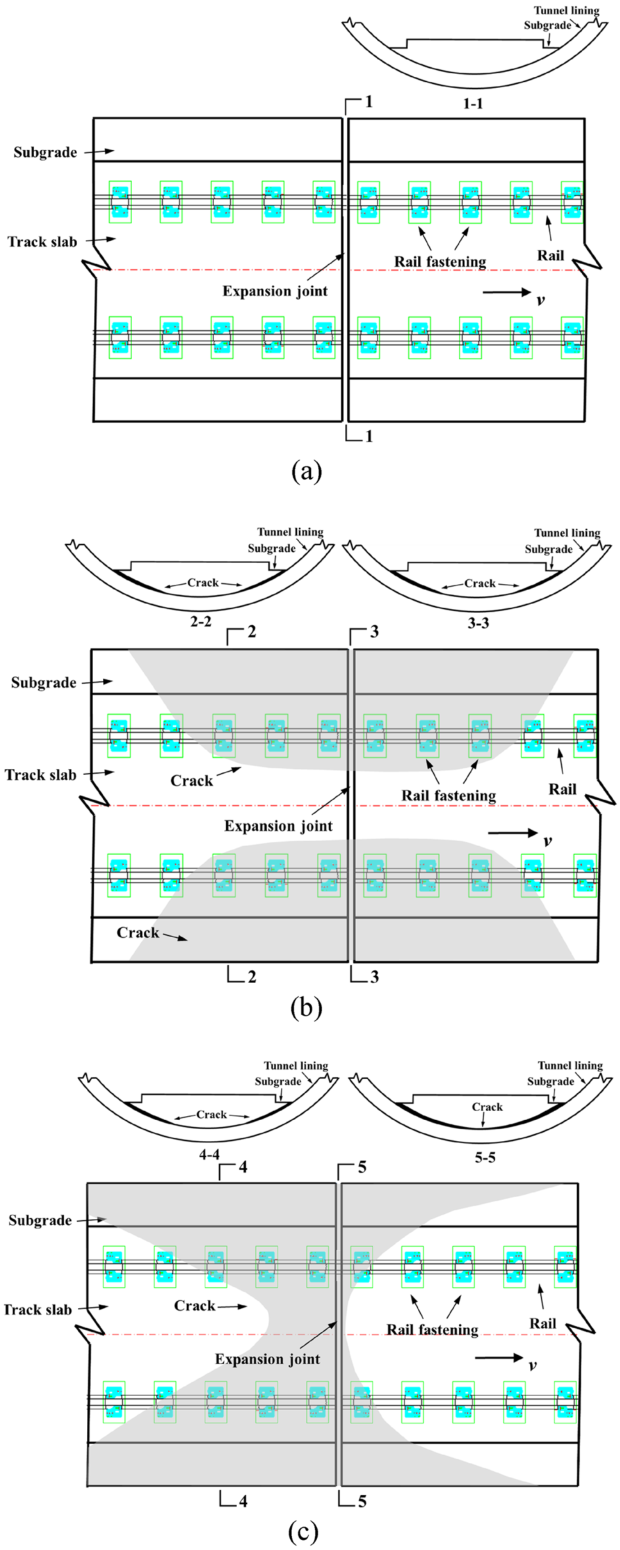


Fig. 7 Schematics of rail tracks in different structural health conditions (top view): **a** normal; **b** minor damage; **c** severe damage

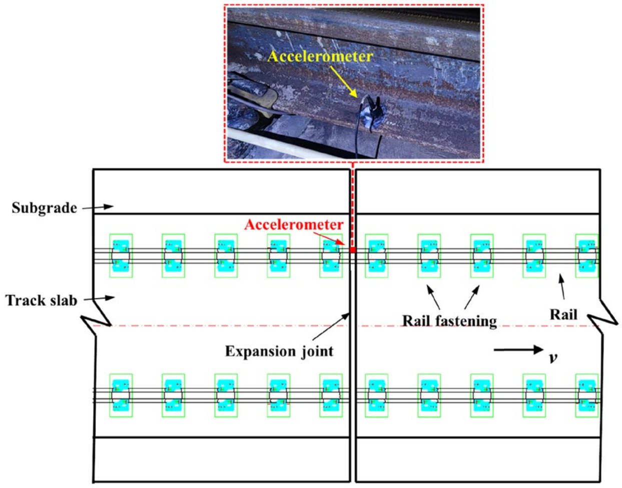


Fig. 8 Accelerometer for rail track vibration measurement

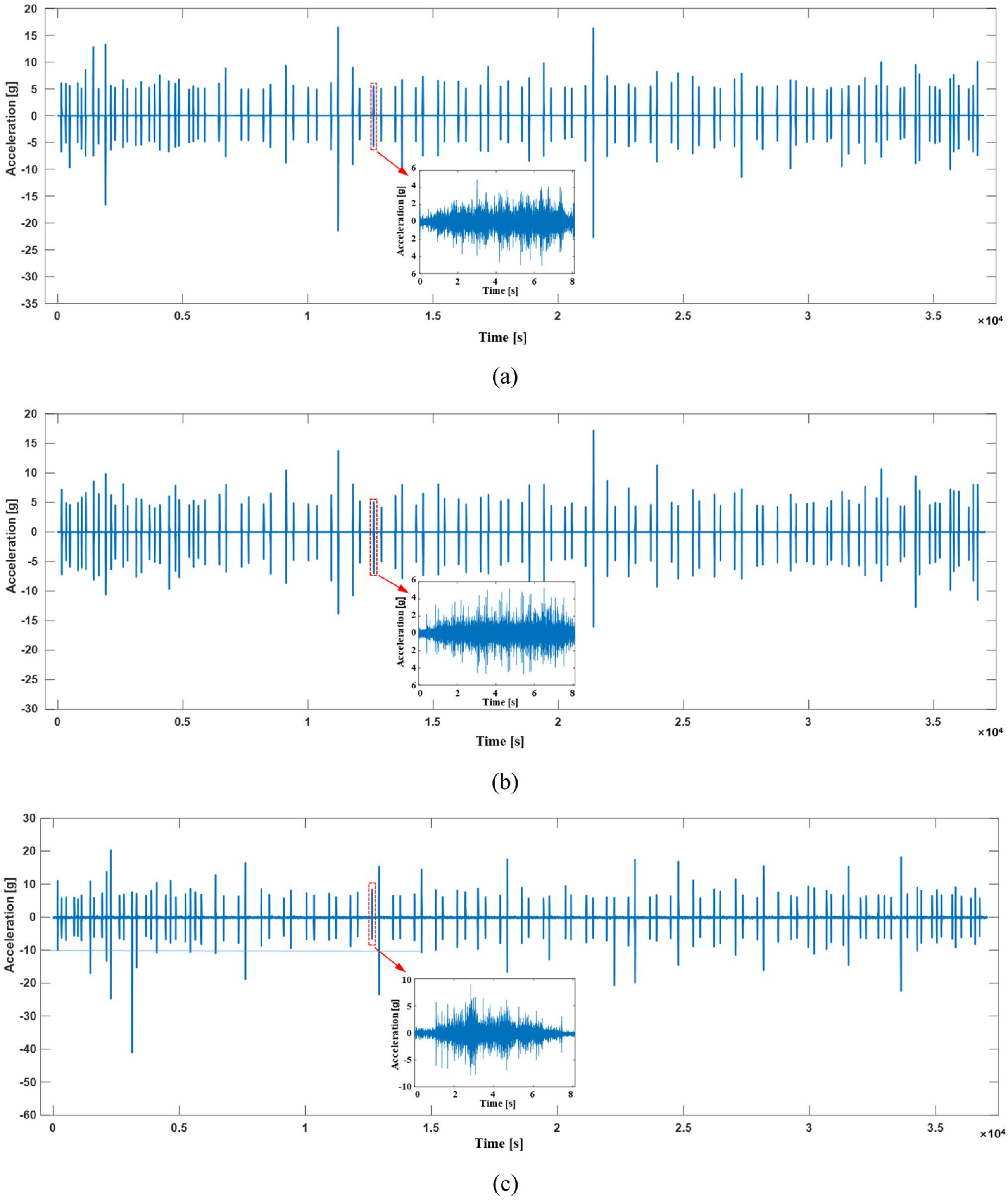
are supported by rail fasteners system spaced at 0.6 m intervals on both sides along the line. Clips and anchor bolts of rail fastening system are used to hold the rails firmly on the track slab. Both the track slab and the subgrade are laid on the tunnel lining with cast-in-place reinforced concrete, with 20 mm expansion joints at every 12.5 m track slab. The cross-sectional layout of the railway track structures is shown in Fig. 6.

In this work, the damage type of rail track structure is the separation crack between concrete subgrade and tunnel lining, which mainly occurs near the expansion joint at the end of the track slab. Therefore, the entire field test includes the measurement of geometrical parameters of the separation crack and the vibration signal acquisition.

3.1 Separation crack measurement

First, the geometric parameters of the separation crack in the same straight track were measured, including the penetration depth along the direction perpendicular to the tunnel lining and crack opening displacement, which were measured by feeler gauge and steel ruler, respectively. With the expansion joint being the center (denoted as location of “0”), the geometrical parameters of the separation crack at a few locations on both sides of the center along the direction of the rail track were measured and recorded. The locations and geometric parameters of cracks are shown in Table 1.

With the continuous periodic loading from passing trains, the separation crack between concrete subgrade and tunnel lining will further propagate along the cross-sectional interface. The crack propagation is affected by the component of the train loading normal to the propagation direction. When the direction of the train loading is close to 90º to the crack propagation direction (i.e., close to penetration), the train loading influences the crack propagation significantly [33, 34]. Once the crack is fully penetrated, the concrete subgrade above the cracked area lacks the connection with the tunnel lining, leading to the significant decrease in the bearing capacity of the subgrade and the rapid crack propagation along the longitudinal direction of the track. Therefore, the penetration status is crucial in terms of crack propagation. The classification criteria for the rail track structural



(a)

(b)

(c)

Fig. 9 Vibration signals: **a** normal; **b** minor damage; **c** severe damage

Table 1 Locations and geometric parameters of cracks

	Left hand			Right hand		
	Location (m)	Penetration depth (m)	Crack opening displacement (mm)	Location (m)	Penetration depth (m)	Crack opening displacement (mm)
Minor damage	0	2.9	1	0	2.9	1
	+ 1.2	2.9	1	+ 1.2	2.9	0.7
	+ 2.4	2.2	0.75	+ 2.4	1.0	0.7
	− 1.2	2.9	0.8	− 1.2	2.9	1
	− 2.4	1.0	0.8	− 2.4	/	/
Severe damage	0	Fully penetrated	1.7	0	Fully penetrated	3.3
	+ 1.2	1.45	1.3	+ 1.2	1.45	1.5
	+ 2.4	0.9	0.6	+ 2.4	0.2	0.8
	− 1.2	2.5	1.5	− 1.2	Fully penetrated	3
	− 2.4	1.7	1.2	− 2.4	2.0	1.5

Table 2 Specification of accelerometer

	Sensitivity	Sensing range	Sampling frequency	Weight
Accelerometer	1 mV/m·s^{-2}	± 1000 m·s^{-2}	0.5 ~ 10,000 Hz (± 10%)	5.5 g

health conditions is therefore based on whether penetration is reached, i.e., the minor damage corresponding to the pre-penetration status and the severe damage to the fully penetrated status, as indicated in Fig. 7b and c, respectively.

3.2 Vibration signal acquisition

After determining the damage types of the rail track structures, vibrations of the rail track with passing trains were measured. Three accelerometers were installed on the rail flange at the location of the expansion joints under three structural health conditions respectively, as shown in Fig. 8.

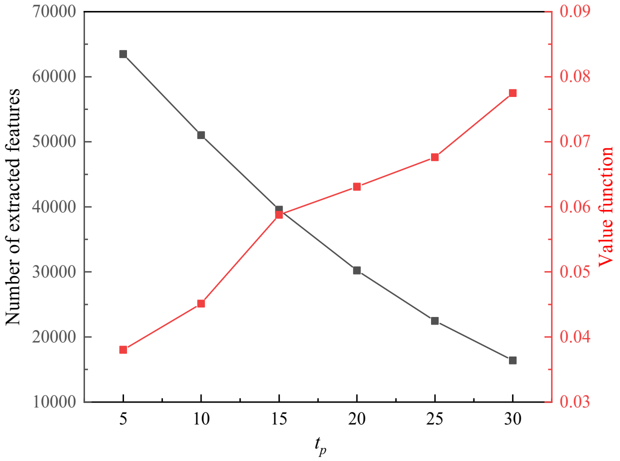


Fig. 10 Number of extracted features and value function with varying t_p

The vibration signals are acquired by the dynamic signal test and analysis system (model: DH5902N, DongHua Testing Technology Co., Ltd), with 16 data acquisition channels. The accelerometers used in the test is piezoelectric

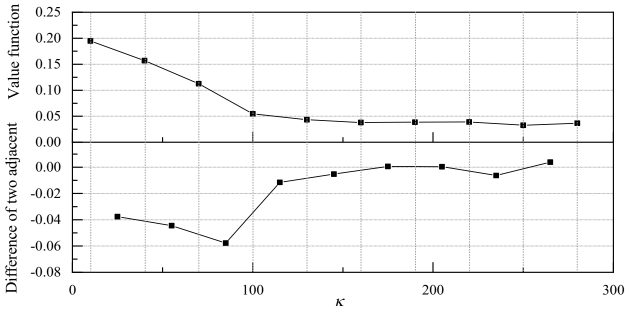


Fig. 11 Value function and difference of two adjacent value functions with $M = 30$

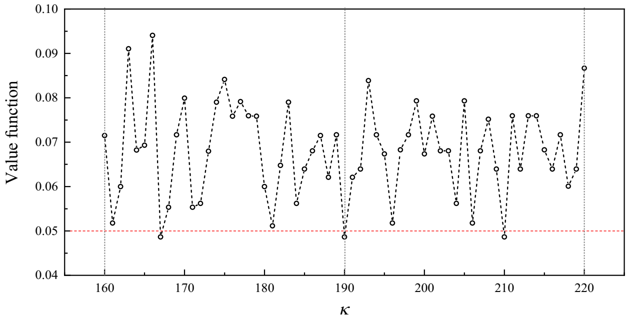


Fig. 12 Comparison of the value function in the recommended data ranges with $M = 30$

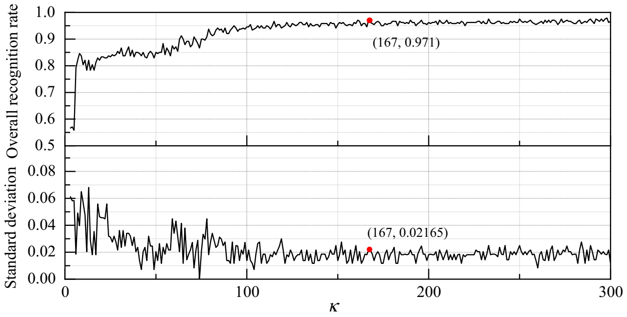


Fig. 13 The overall recognition rate and standard deviation of training set with varying κ

type (Model: DH1A108E, DongHua Testing Technology Co., Ltd) and the specification is listed in Table 2.

In this field test, the vehicle operating speed was 35 ~ 50 km/h. When one train passed through a section of rail track, the vibrational displacement of that section was very significant within about 7.5 s. While the sampling frequency of vibration signals in this field test was 1000 Hz, considering that the rail would vibrate freely for a short time after the train passes, the corresponding time for vibration signal truncation in this work was 8.1 s, i.e., 8100 points for each segmented vibration signal. The time-domain acceleration signals of rail vibrations under three health conditions are shown in Fig. 9.

Overall, compared with the normal case, the vibration amplitude of minor damage has a slight increase when the same train passing, but the change is not obvious. However, the signal amplitude has a relatively obvious increase in the severe damage case than the other two cases. It shows that there are differences in vibration signals under different rail track structural health conditions, but the vibration signal is not sensitive to this early-stage structural damage.

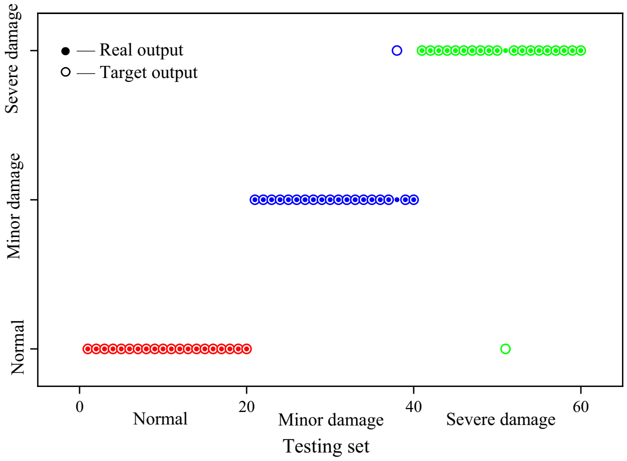


Fig. 14 The recognition result of testing set

4 Results and discussion

In this work, 100 datasets corresponding to each health condition (totally 100 × 3 datasets for three health conditions) are utilized for identification. Specifically, 80 datasets from each health condition (80 × 3 datasets totally) are utilized for training and fourfold cross-validation to ensure the robustness of the algorithm in Sect. 4.1. Then, the remaining 20 datasets from each health condition (20 × 3 datasets totally) are utilized for testing to verify the accuracy and effectiveness in Sects. 4.2 and 4.3. Finally, in Sect. 4.4, the performance of the proposed method is compared with other methods using the same 100 × 3 datasets.

4.1 Optimal parameter selection

The selection of two parameters is important to build the VBoW model, i.e., the threshold value for feature extraction and the optimal number of keywords.

4.1.1 Selection of threshold value

The selection of the threshold value t_p (see Eq. (3)) will not only affect the number of extracted features, but also can affect the clustering in the VBoW model. The calculation results of number of extracted features and value function with varying t_p are shown in Fig. 10, where κ is 167, and this parameter is selected optimally as shown in Sect. 4.1.2.

It is found that as t_p increases, fewer features are extracted, and the value function increases accordingly. Also, obviously, when $t_p = 5$, the value function is lower and more features could be extracted as compared to the other selected values. Therefore, $t_p = 5$ is selected for further analysis.

4.1.2 Selection of the optimal number of keywords

In this work, the data range of κ is initially set to [10, 300], the minimum of value function $\varepsilon_{\min}$ is set to 0.05, the minimal difference of two adjacent value functions ε_{adj} is set to 0.001 and the interval M is set to 30 by considering the computational cost. To ensure that the overall trend of the

Table 3 Real-time analysis of testing set

	Signal conversion (s)	Feature extraction (s)	Identification (s)	Total (s)
Normal	0.043	0.039	0.063	0.145
Minor damage	0.042	0.028	0.067	0.137
Severe damage	0.042	0.040	0.055	0.137

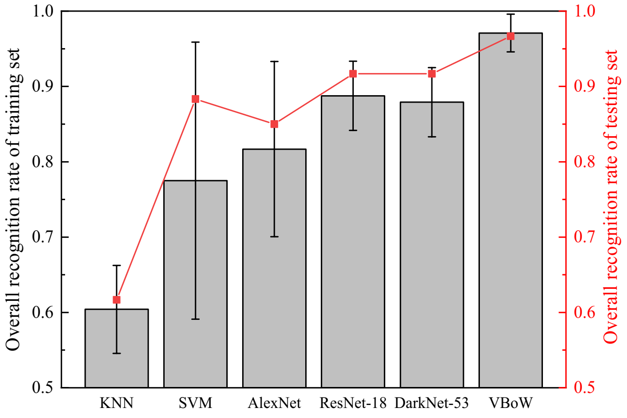


Fig. 15 Comparison of recognition results with different algorithms and models

value function varying with the number of keywords is less affected by local fluctuations, it is necessary to employ average filtering on the results of each value function. Specifically, the value function corresponding to each keyword is the averaged value function of five adjacent keywords. The calculated result is shown in Fig. 11. Overall, the $\varepsilon(\kappa)$ decreases as the value of κ increases. When κ is greater than 150, the change of the value function tends to level off. Moreover, to obtain the recommended data ranges of the value function, the difference between two adjacent value functions is further analyzed. According to the algorithm described in the Sect. 2.3.3, the recommended data ranges of κ are [160, 190] and [190, 220].

By comparing all the value function in above recommended data ranges (Fig. 12), the minimum value function is lower than $\varepsilon_{\min}$ when the κ is taken as 167, 190 and 210. By considering the computational cost, $\kappa_{opt}=167$ is selected and regarded as the optimum.

4.1.3 Training results validation

To further illustrate the recognition accuracy and stability of the proposed health condition evaluation method under κ_{opt} taken in Sect. 4.1.2, the overall recognition rate (P_κ in Eq. (7)) and standard deviation (S_κ in Eq. (7)) when κ is set to [3, 300] are shown in Fig. 13. Although value-sweeping calculation is not used in our method, here it is used to demonstrate the algorithm performance and to show the validity of the proposed method for selecting κ_{opt}.

In general, with the increase of κ, the P_κ of the algorithm increases and S_κ decreases. As can be seen in Fig. 13, when κ is greater than 150, the algorithm can maintain a high recognition rate and low standard deviation, indicating that the algorithm is suitable for the recognition of rail track structural health condition and had high stability. The overall recognition rate of fourfold cross-validation on the training set reaches 97.1% when κ is 167 as selected in the previous work. Although the overall recognition rate could reach 97.9% when κ is taken as 298, it does not improve much, and the computational cost is significantly increased. Therefore, the κ_{opt} selected by the proposed method can ensure the recognition accuracy and stability of the proposed health condition evaluation method.

4.2 Testing results

To verify the applicability of the recognition performance of the method used in this work, the remaining 20×3 signals are tested and analyzed here. The test result is shown in Fig. 14. Specifically, the horizontal axis represents the 60 signals in the testing set while the vertical axis represents the actual three rail track structural health conditions. Moreover, the solid points represent the real output of each sample, and the circles represent the target output of the test samples. The test results show that 58 of the 60 testing samples are correctly identified, with an overall recognition rate of 96.67%.

4.3 Real-time analysis

The operating efficiency of this recognition method directly affects its practical engineering value. To illustrate the effectiveness of the proposed method, the real-time performance of test set is analyzed. The computational platform consists of an i7-10750H CPU and a NVIDIA GeForce RTX 2060 graphics processing unit. The whole recognition process based on VBoW model is implemented with Matlab R2020 under Windows 10. In addition, the recognition process has three steps, i.e., signal conversion, image feature extraction and health condition identification. The average time for completing each step is shown in Table 3.

As shown in Table 3, for all the three health conditions of the rail track structures in this work, the average time for the whole recognition process using VBoW model is relatively small, within 0.15 s. In Sect. 3.2, each segmented vibration signal sample is 8.1 s. Therefore, VBoW model is very efficient and has the ability to meet the real-time requirement.

4.4 Performance comparison by different methods

To further illustrate the superiority of the proposed health condition evaluation method, it is compared with the representative traditional learning-based classification algorithms and the deep learning models. Specifically, the traditional learning-based classification algorithms are selected as: Support Vector Machine (SVM) [35, 36] and K-Nearest Neighbor algorithm (KNN) [37, 38], and the deep learning models are selected as: AlexNet [39], ResNet-18 [40], and DarkNet-53 [41]. To be more detailed, Grid Searching (GS) technique is adopted to optimize the parameters (e.g., penalty coefficient and the kernel function parameter) of the SVM model, and the number of nearest neighbors in this work is set to be 1 for KNN, which is optimal.

All the above six comparison methods are run in the same MATLAB environment, accepting other default parameter values. In addition, all methods are subjected to fourfold cross-validation on 80×3 training signals and performance validation on 20×3 testing signals. The calculation results are shown in Fig. 15. The overall recognition rate and standard deviation are taken to reflect the reliability and stability, respectively.

As shown in Fig. 15, the overall recognition rates of KNN and SVM algorithms for wheel–rail vibration signals are only about 60% and 77%, respectively. Although the overall recognition rates of AlexNet, ResNet-18 and DarkNet-53 algorithms exceed 80%, this is insufficient to meet the requirements of real-time monitoring in real-life applications. It can be found that our method has a higher recognition rate of 97.1% as compared to the other five models and algorithms. Furthermore, our method obtains a lower standard deviation, demonstrating more stable recognition behavior. In addition, the overall recognition result by 96.7% from the testing set also revalidates the better performance of the proposed method.

5 Conclusions and future work

In this work, a novel computer vision-aided method to evaluate the nondestructive health condition of rail track structures based on vibration signals has been proposed. Specially, the method used in this work does not require tedious noise reduction processing and redundant feature elimination. By directly converting the raw vibration signals into grayscale images, we have adopted multi-dimensional feature vectors of images instead of one-dimensional feature array of traditional signal processing methods. To quickly extract numerous features from the vibration signals, the FAST-Unoriented-SIFT algorithm has been utilized. Meanwhile, the VBoW model with optimal keyword has been proposed for well describing and identifying grayscale images features. Finally, the overall recognition rate of the proposed model in the testing set is 96.7% (i.e., 58 out of 60). In addition, by comparison to the traditional learning-based classification algorithms and the representative deep learning models, it is found that the proposed method is more suitable for the effective identification of strong time-varying and random vibration signals and has promising prospects for practical structural health monitoring applications.

Although the proposed method can achieve satisfactory results in this work for the damage identification under different rail track structural health conditions, due to the learning capacity, all these popular algorithms based on Visual Bag-of-Words model are restricted to relatively short signal segments. However, high-dimensional signals and features can better capture the valuable information of dynamic system under a more complex environment. On the other hand, in rail track structural health monitoring, to obtain more spatial information, numerous sensors need to be fixed along the track line for real-time monitoring, which not only results in data overload but also increases operation and maintenance costs. In contrast, placing mobile sensor networks on vehicle components (e.g., axlebox) for health monitoring of rail track structures have greater potential for low-cost monitoring applications. However, the rail track vibration signal data collected by the mobile sensor networks is spatio-temporal, and is subjected to vehicle interference noise. Recent researches [42, 43] proposed some methods where sparse vibration data based on mobile sensor networks were used to successfully achieve feature identification of bridge structures. Therefore, future work can be focused on investigating a more robust extraction method of the high-dimensional feature information of wheel–rail dynamic system around the method of VBoW model and sparse representation using vehicle mobile sensor networks.

Acknowledgements This work was supported by the Ph.D. scholarship from the China Scholarship Council (Grant Nos. 202007000010 and 202006420001), Catalyst Seeding General Grant administered by the Royal Society of New Zealand (Contract 20-UOA-035-CSG), International Innovation Cooperation Programme of Department of Science and Technology of Sichuan Province, China (No. 2020YFH0066), and Grant No. 109 [2021] of Division of American and Oceanian Affairs, China Scholarship Council.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Matias SR, Ferreira PA (2020) Railway slab track systems: review and research potentials. Struct Infrastruct Eng 16:1635–1653
2. Yan QX, Sun MH, Qing SY, Deng ZX, Dong WJ (2020) Numerical investigation on the damage and cracking characteristics of the shield tunnel caused by derailed high-speed train. Eng Fail Anal 108:104205
3. Shen SL, Wu HN, Cui YJ, Yin ZY (2014) Long-term settlement behaviour of metro tunnels in the soft deposits of Shanghai. Tunn Undergr Space Technol 40:309–323

4. Symonds N, Corni I, Wood RJK, Wasenczuk A, Vincent D (2015) Observing early stage rail axle bearing damage. Eng Fail Anal 56:216–232
5. Sainz-Aja J, Pombo J, Tholken D, Carrascal I, Polanco J, Ferreno D, Casado J, Diego S, Perez A, Abdala JE, Esen A, Cebasek TM, Laghrouche O, Woodward P (2020) Dynamic calibration of slab track models for railway applications using full-scale testing. Comput Struct 228:106180
6. Yang F, Gao MY, Wang P, Zuo JY, Dai J, Cong JL (2021) Efficient piezoelectric harvester for random broadband vibration of rail. Energy 218:119559
7. Deraemaeker A, Reynders E, De Roeck G, Kullaa J (2008) Vibration-based structural health monitoring using output-only measurements under changing environment. Mech Syst Signal Process 22:34–56
8. Chen M, Zhai WM, Zhu SY, Xu L, Sun Y (2021) Vibration-based damage detection of rail fastener using fully convolutional networks. Vehicle Syst Dyn. https://doi.org/10.1080/00423114.2021.1896010
9. Zhang C, Zhao WJ, Wang WG, Zhang J (2021) Vision-based tire deformation and vehicle-bridge contact force measurement. Measurement 183:109792
10. Das S, Saha P, Patro S (2016) Vibration-based damage detection techniques used for health monitoring of structures: a review. J Civ Struct Heal Monit 6:477–507
11. Li YB, Xu MQ, Liang XH, Huang WH (2017) Application of bandwidth EMD and adaptive multiscale morphology analysis for incipient fault diagnosis of rolling bearings. IEEE Trans Ind Electron 64:6506–6517
12. Niu YB, Ye Y, Zhao WJ, Shu JP (2021) Dynamic monitoring and data analysis of a long-span arch bridge based on high-rate GNSS-RTK measurement combining CF-CEEMD method. J Civ Struct Heal Monit 11:35–48
13. Lei YG, Lin J, He ZJ, Zi YY (2011) Application of an improved kurtogram method for fault diagnosis of rolling element bearings. Mech Syst Signal Process 25:1738–1749
14. Ahmadi HR, Mahdavi N, Bayat M (2021) A novel damage identification method based on short time Fourier transform and a new efficient index. Structures 33:3605–3614
15. Hu ST, Shi WZ, Lu C, Chen Y, Chen G, Shen GT (2021) Rapid detection of cracks in the rail foot by ultrasonic B-scan imaging using a shear horizontal guided wave electromagnetic acoustic transducer. Ndt&E Int 120:102437
16. Qu H, Li M, Chen G, Sneed LH, Anderson NL (2016) Multi-bandwidth wavelet transform of impact echo signals for delamination detection of thin concrete slabs. J Civ Struct Heal Monit 6:649–664
17. Sohn H (2007) Effects of environmental and operational variability on structural health monitoring. Philos Trans R Soc A 365:539–560
18. Lu C, Wang Y, Ragulskis M, Cheng YJ (2016) Fault diagnosis for rotating machinery: a method based on image processing. PLoS ONE 11:e0164111
19. Ruiz M, Mujica LE, Alferez S, Acho L, Tutiven C, Vidal Y, Rodellar J, Pozo F (2018) Wind turbine fault detection and classification by means of image texture analysis. Mech Syst Signal Process 107:149–167
20. Dai XW, Gao ZW (2013) From model, signal to knowledge: a data-driven perspective of fault detection and diagnosis. IEEE Trans Ind Inform 9:2226–2238
21. Gao ZW, Cecati C, Ding SX (2015) A Survey of fault diagnosis and fault-tolerant techniques-part i: fault diagnosis with model-based and signal-based approaches. IEEE Trans Ind Electron 62:3757–3767
22. Shahriar MR, Ahsan T, Chong U (2013) Fault diagnosis of induction motors utilizing local binary pattern-based texture analysis. Eurasip J Image Video Process 2013:1–11
23. Do V, Chong UP (2011) Signal model-based fault detection and diagnosis for induction motors using features of vibration signal in two-dimension domain. Stroj Vestn-J Mech Eng 57:655–666
24. Zheng H, Cheng G, Li Y, Liu C (2019) A new fault diagnosis method for planetary gear based on image feature extraction and bag-of-words model. Measurement 145:1–13
25. Subudhi BN, Ghosh S, Shiu SCK, Ghosh A (2016) Statistical feature bag based background subtraction for local change detection. Inform Sci 366:31–47
26. Silva FB, Werneck RD, Goldenstein S, Tabbone S, Torres RD (2018) Graph-based bag-of-words for classification. Pattern Recogn 74:266–285
27. Qi L, Adamchuk V, Huang HH, Leclerc M, Jiang Y, Biswas A (2019) Proximal sensing of soil particle sizes using a microscope-based sensor and bag of visual words model. Geoderma 351:144–152
28. Yang SH, Lu GL, Wang AQ, Liu J, Yan P (2019) Change detection in rotational speed of industrial machinery using Bag-of-Words based feature extraction from vibration signals. Measurement 146:467–478
29. Zheng H, Cheng G, Li Y, Liu C (2020) A fault diagnosis method for planetary gear under multi-operating conditions based on adaptive extended bag-of-words model. Measurement 156:107593
30. Wen L, Li XY, Gao L, Zhang YY (2018) A new convolutional neural network-based data-driven fault diagnosis method. IEEE Trans Ind Electron 65:5990–5998
31. Rosten E, Porter R, Drummond T (2010) Faster and better: a machine learning approach to corner detection. IEEE Trans Pattern Anal 32:105–119
32. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vision 60:91–110
33. Huang Z, Fu HL, Chen W, Zhang JB, Huang HW (2018) Damage detection and quantitative analysis of shield tunnel structure. Automat Constr 94:303–316
34. Xu LH, Ma M (2020) Study of the characteristics of train-induced dynamic SIFs of tunnel lining cracks based on the modal superposition approach. Eng Fract Mech 233:107069
35. Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2:1–27
36. Gui GQ, Pan H, Lin ZB, Li YH, Yuan ZJ (2017) Data-driven support vector machine with optimization techniques for structural health monitoring and damage detection. KSCE J Civ Eng 21:523–534
37. Murphy OJ (1990) Nearest neighbor pattern-classification perceptrons. Proc IEEE 78:1595–1598
38. Zhang SC, Cheng DB, Deng ZY, Zong M, Deng XL (2018) A novel kNN algorithm with data-driven k parameter computation. Pattern Recogn Lett 109:44–54
39. Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. Commun Acm 60:84–90
40. He KM, Zhang XY, Ren SQ, Sun J (2016) Deep residual learning for image recognition. In: Proc Cvpr IEEE, pp 770–778
41. Redmon J, Farhadi A (2018) Yolov3: An incremental improvement. arXiv preprint
42. Sadeghi Eshkevari S, Pakzad SN, Takáč M, Matarazzo TJ (2020) Modal identification of bridges using mobile sensors with sparse vibration data. J Eng Mech. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001733
43. Sadeghi Eshkevari S, Matarazzo TJ, Pakzad SN (2020) Bridge modal identification using acceleration measurements within moving vehicles. Mech Syst Signal Process 141:106733

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.