

Internal Defects Detection Method of the Railway Track Based on Generalization Features Cluster

Fupei Wu (✉ wufupei@163.com)

Shantou University <https://orcid.org/0000-0002-7161-0047>

Xiaoyang Xie

Shantou University

Jiahua Guo

Shantou University

Qinghua Li

Shantou University

Original Article

Keywords: Railway track, Generalization features cluster, Defects classification, Ultrasonic image, Defect detection

Posted Date: April 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-399710/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Internal Defects Detection Method of the Railway Track Based on Generalization Features Cluster

Fupei Wu¹, Xiaoyang Xie¹, Jiahua Guo¹ and Qinghua Li^{1*}

Abstract

Many internal defects maybe arise in railway track working, which usually have different shapes and distribution rules. To solve the problem, an intelligent detection method is proposed for internal defects of railway track based on generalization features cluster in this paper. Firstly, defects are classified and counted according to their shapes and locations features. Then, generalized features of defects are extracted and formulated based on the maximum difference between different types of defects and the maximum tolerance among same types of defects. Finally, extracted generalized features are expressed by function constraints, and formulated as generalization feature clusters to classify and identify internal defects of the railway track. Furthermore, a reduced dimension method of the generalization features clusters is presented too in this paper. Based on the reduced dimension feature and strong constrained generalized features, the K-means clustering algorithm is developed for defects clustering, and good clustering results are achieved. To defects in the rail head region, its clustering accuracy is over 95%, and the Davies-Bouldin Index (DBI) index is small, which indicates the validation of the proposed generalization features with strong constraints. Experimental results show that accuracy of the proposed method based on generalization features clusters is up to 97.55%, and the average detection time is 0.12s/frame, which indicates it has good performance in adaptability, high accuracy and detection speed under the complex working environments.

Keywords: Railway track, Generalization features cluster, Defects classification, Ultrasonic image, Defect detection

1 Introduction

Railway rail transit system is an essential infrastructure for cargo and passenger [1]. The track quality will directly affect the operation safety of railway traffic [2]. As the growing demand for railway traffic density and running speed, the rail track wear is also increasing, which causes a higher probability of internal rail defect, and the complexity of defect is also various [3-4]. Therefore, its internal defects detection speed and accuracy are also should be improved.

Presently, most of the rail defect detection focuses on surface defect detection. For example, Yu et al. proposed a coarse-to-fine model (CTFM) to identify defects at different scales: sub-image level, region level, and pixel level [5]. Dubey [6] used a maximally stable extremal region technique to identify and visualize the geometrical features of the defect regions on the rail head surface in railway track images. Li presented an intelligent vision detection system (VDS) for discrete surface defects, which focuses on two critical issues of VDS: image enhancement and automatic thresholding [7]. Furthermore, Zhang et al. proposed an automatic railway visual detection system (RVDS) for surface defects and presented an algorithm for detecting and extracting region-of-interest which enables identification and segmentation of the defects from rail surface [8]. Compared with the internal defects of rail, the rail surface defect image is more effective to acquire by

high-speed camera, so the imaging standard is consistent. Internal rail defects cannot be visualized, while it can be detected using ultrasonic testing, eddy current testing and magnetic flux leakage testing [9]. Ultrasonic testing is widely used in rail internal defect detection because of its high penetration, excellent directivity and high sensitivity [10]. However, because of different ultrasonic acquisition equipments, it is difficult to have a unified imaging standard for rail ultrasonic B-scan image, but the feature expression of the defect is universal, which can be described by the generalized feature. B-scan image is imaged by echo of the ultrasonic probe, which can display a cross-section of rail. Few previous works can be found on internal defect detection of railway track based on ultrasonic image. Cygan [11] analyzed the advantages of B-scan image processing compared with A-scan signal analysis in rail internal defect detection. A-scan signal is a flaw detection method to evaluate the size and position of defects based on the amplitude and position of defect waves, but it is unable to determine the defect geometry directly. Huang [18] and Sun [29] used neural network model and deep learning to analyze rail B-scan images respectively and to achieve internal defects detection. Liang [12] put forward an improved imaging algorithm for rail defect identification, which is beneficial to the acquisition of a high-quality B-scan image and the design of high accuracy detection algorithm for rail internal defect image. However, there is still a challenge work to achieve high precision detection for different types of defects. Therefore, it is essential to develop an accuracy and high-speed detection algorithm for rail internal defects expressed by ultrasonic B-scan image.

The main types of rail internal defects include: (1)

*Correspondence: 17qhli@alumni.stu.edu.cn

¹ Key Laboratory of Intelligent Manufacturing Technology, Ministry of Education, Shantou University, Shantou 515063, China

Full list of author information is available at the end of the article

fatigue crack of the screw hole, (2) rail head defect, (3) crushing flaw of rail bottom, (4) transverse crack of rail bottom, and (5) material degradation of special parts [13]. Fatigue crack of screw hole is the crack located at different positions of screw hole. The internal defect of rail head mainly is the flaw, including black flaw and white flaw from the color showing in real environment. The crushing flaw of rail bottom includes transverse crack and longitudinal crack of rail bottom. To prevent such internal defects which affect the traffic working safety, nondestructive testing is usually carried out regularly to monitor track health.

Presently, X-ray can also be used to detect internal defects. For example, Cai et al. [14] proposed an X-ray image defect detection algorithm for casting based on Mask R-CNN, which provides a solution method for intelligent industrial defect detection. Guo et al. [15] proposed a welding defect detection method based on Fast R-CNN model with X-ray image and it achieved the expected experimental results. However, the mainstream method for detecting internal defects in railway track is still to analyze the ultrasonic images [16]. To detect such defects, the ultrasonic rail detection system developed by SPERRY company and TOKIMEC company can recognize and classify defects in real time. RTI company has developed an automatic defect identification system based on a neural network model, which has a learning function [17]. The Chinese Academy of Railway Sciences has developed a B-scan image rail defect classification system based on pattern recognition, whose recognition rate is about 95% [18]. However, the accuracy of the existing ultrasonic detection system is still not enough to meet the actual detection requirements.

To solve the problem of low detection accuracy, an internal defect detection method of railway track is proposed based on the generalization features cluster in this paper. Firstly, defects are classified and counted by analyzing their location and geometric features. Then, according to the maximum difference between different types of defects and the maximum tolerance of the same type of defects, the generalization features of defects are extracted. Finally, the generalization features clusters for various types of defects are established to classify the internal defects of railway track. On this basic, strong constrained generalization features clusters are formulated after dimension reduction, and K-means clustering algorithm is developed to cluster defects. Experimental results show that the proposed method can be used to detect internal defects with high accuracy and detection speed.

The rest of this paper is organized as follows. Firstly, the principle of ultrasonic image detection and defects' classification is analyzed in Section 2. Secondly, Section 3 presents generalization features clusters of different types of defects and K-means clustering algorithm based on strong constrained generalization features. Thirdly, the

experimental results and analysis of the proposed method are shown in Section 4. Finally, some conclusions are given in Section 5.

2 Principle of Ultrasound Image Detection and Defects Classification

2.1 The Principle of Internal Ultrasound Image Detection of Railway Track

Ultrasound has the advantages of fast spread speed, and broad applicability [19] in nondestructive testing. According to the propagation features of the ultrasonic wave in different medium, the probe emits a certain frequency of sound wave into the rail. If a rail defect arises, a defect wave will appear in front of the bottom wave, and the peak value of the bottom wave decreases or disappears. Furthermore, the size and position of the defect can be evaluated by the reflected signal. The ultrasonic echo signal is amplified, filtered and level converted by the ultrasonic receiver to draw and display the electrical signal in the form of a digital image, which is named as B-scan image. Therefore, the acquisition process of rail B-scan image is the collection process of rail defects data.

The acquisition mechanism of B-scan image in rail track is shown in Figure 1. The coupling liquid is sprayed on the contact surface between the probe and the rail to prevent the attenuation of ultrasonic signal energy, and water is usually selected as the coupling agent. The shape, position and depth of defects are detected by features of ultrasonic propagation, reflection and refraction in railway track [20]. Usually, ultrasonic pulse propagates inside the track to detect internal defects. If it encounters cracks and flaws with different acoustic impedance, it will produce a primary reflection and secondary reflection [21]. The defect shape in the railway track can be imaged and displayed [22] by analyzing the magnitude, quantity and waveform of the reflected wave.

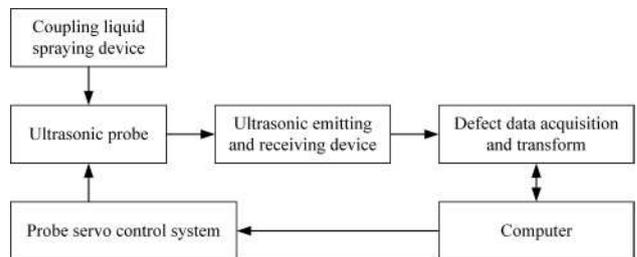


Figure 1 Rail B-scan image acquisition mechanism

The ultrasonic probes with different angles can detect defects in different locations of the railway track and distinguish them with different colors. As shown in Figure 2, the 0° probe generates ultrasonic longitudinal wave beams, which are used to detect horizontal cracks of screw holes shown as red in the B-scan image. The 70° probe generates ultrasonic shear wave beams and detects rail head flaws by primary or secondary waves. In addition, rail head flaws are shown as red, green and blue in B-scan

image. The 37° probe generates ultrasonic shear wave beams and detects other types of defects. Rail B-scan image is imaged by a reflection echo of 0°, 37°, 70° probe

[23]. As shown in Figure 3, the B-scan image of color ultrasound for normal railway track and the actual B-scan image containing defects are displayed, respectively.

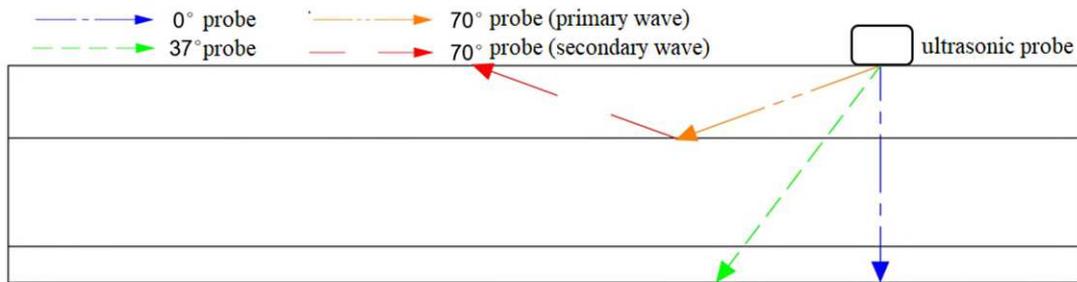
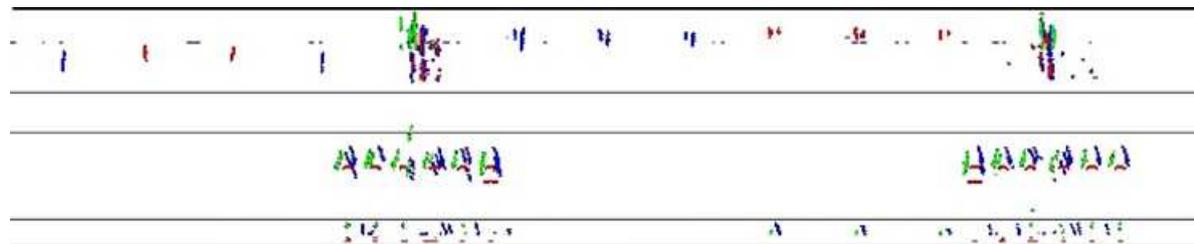


Figure 2 The reflection diagram of ultrasound in railway track



(a) The B-scan image for normal railway track



(b) The actual B-scan image with defects

Figure 3 Ultrasonic B-scan images in different situations

2.2 Classification of Railway Track Internal Defects

Color B-scan image can provide the projected-sectional features of defects with respect to the normal incident wave. In addition, it also shows the horizontal location and depth information of the internal defects in railway track. As shown in Table 1, 12 types of railway track internal defects can be detected by ultrasound. For example, the defect of type 1 is the inside flaw in rail head. Then, it can be detected by 70° probe and shown with red in ultrasonic B-scan image. To make all these types of defects clear, some defects are taken as an example in Figure 4. In Figure 4, the inside of rail head and the inside of screw hole are both near the center of railway. Therefore, defects of type 1, 2, 3 and 4 can be shown clearly in Figure 4, other types of defects also can be shown as similar way. However, it is noted that inverted cracks can be only found in one side of two screw holes which near the rail end face. Furthermore, defects are classified according to the ultrasonic imaging

mode, defect location, defect cause, defect features, and traditional railway track internal defect classification method [24].

The acquired ultrasonic image may have some differences with Figure 3(b). The acquisition process may cause random clutter, image coverage, image fracture, etc. As shown in Figure 5, all types of defects are labeled, which are shown in Table 1. There are various types of clutter in the black circle region, and the joint is in the yellow triangle region. The length of the rail head flaw is different, and it appears in pairs or separately. The lower crack of the screw hole is connected to its screw hole. There are differences in the thickness and length of rail bottom defect images contour. Furthermore, the defects are paired images contour or single images contour, which varies in shape and make it difficult to accurately describe defect features with accuracy models. Therefore, this paper proposes a detection method based on the generalization feature cluster to solve such a problem.

Table 1 Classification table of internal defects in railway track

Sequence number	Defect types	Detection probe	Image color
1	Inside flaw of rail head	70° probe	R
2	Middle flaw of rail head	70° probe	G
3	Outside flaw of rail head	70° probe	B
4	Upper crack on the inside of the screw hole	37° probe	G
5	Inverted upper crack on the inside of the screw hole	37° probe	G
6	Inverted lower crack on the inside of the screw hole	37° probe	G
7	Upper crack on the outside of the screw hole	37° probe	B
8	Lower crack on the outside of the screw hole	37° probe	B
9	Inverted lower crack on the outside of the screw hole	37° probe	B
10	Horizontal crack of screw hole	0° probe	R
11	Defect of rail bottom	37° probe	G、B
12	Longitudinal crack of rail bottom	37° probe	R

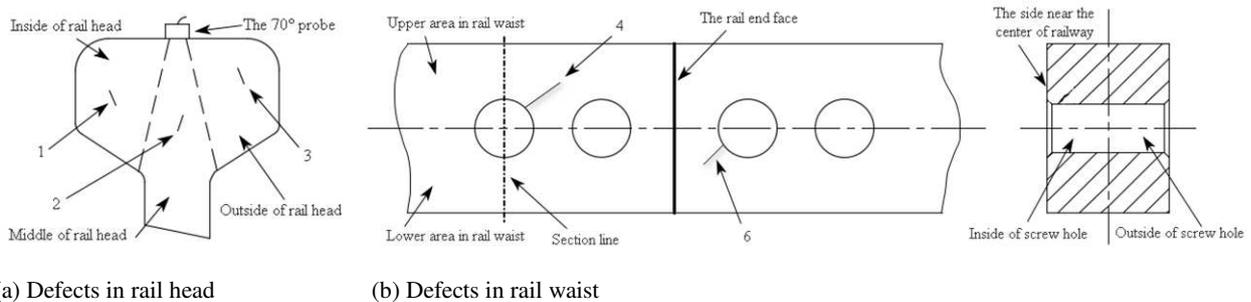


Figure 4 The position diagram of some defects

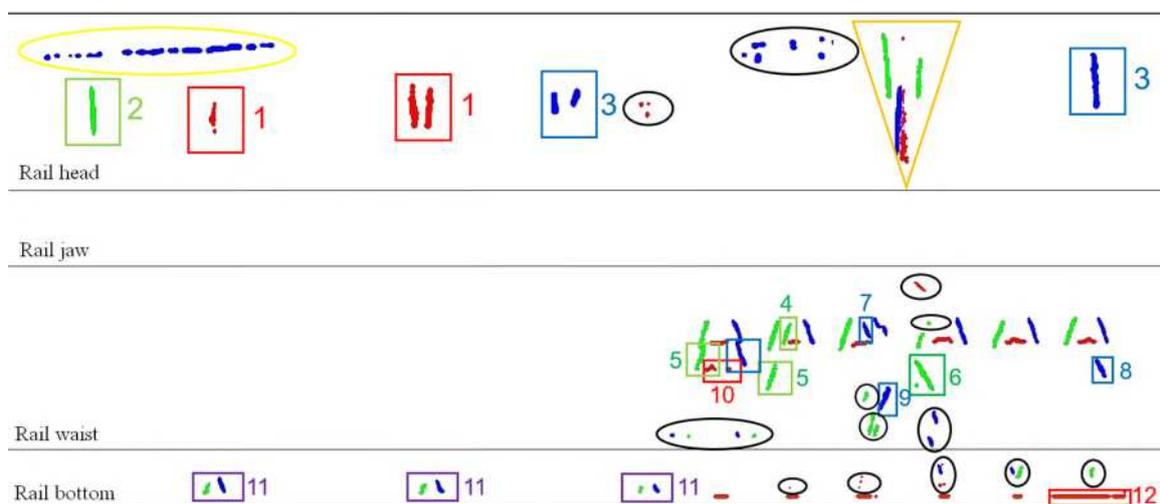


Figure 5 All kinds of defects and non-defects in B-scan image

3 Defects Feature Analysis and Detection

There are many types of defects and different patterns arising in railway track working. Building a precise defect feature model by using existing samples to detect such defects images often results in problems of over-constraint

or under-constraint [25]. This paper solves such problem as following. Firstly, the existing samples are analyzed to extract features for each type of defect. Secondly, extracted features are generalized to increase the fault tolerance and generalization of the features, which means that the generalized features not only show the general

characteristics of the same type of defects, but also eliminate the differences as much as possible, such as their positions, shapes and so on. Then, according to the correlation and non-correlation among defects, generalization features clusters are built. Lastly, the generalization features clusters are used to detect the B-scan image of the railway track. In general defect feature model, even in the same type of defects, the specific characteristics of one defect probably cannot detect another defect properly due to the differences in their locations and shapes. Specifically, there may be few feature constraints to identify defects, or overfull feature constraints that result in misjudgment for defects. However, in the proposed method, generalization feature clusters show the common features of the same type of defects and the non-correlation characteristics between different types of defects, which can effectively avoid the problem mentioned above, speed up the detection time and improve the detection accuracy.

Practically, because of the influence of different acquisition equipment, working environment, and other factors, the B-scan image usually has different features in shape, contour, position, and so on. It is difficult to describe defects accurately with traditional features or single feature due to its poor generalization and application ability. Therefore, multiple generalization features are combined to build a generalization feature cluster in this paper, which is used to achieve better defect detection and expand applicable range of the defect detection model.

3.1 Ultrasonic Image Preprocessing

As we know, image noise affects the detection accuracy and stability. It is necessary to eliminate the noise before defect features extracting. The image preprocessing flow chart is shown as Figure 6. Firstly, according to the color difference of defects, the image is separated into different color channels. Secondly, the morphological method is

used to remove image noise that emerged from image acquisition processes due to artificial or environmental factors. Specifically, binary B-scan images are dilated before being corroded, and the morphological operator is set according to the contour direction. As shown in Figure 7, They are two morphological operators for the processing of lower crack of the screw hole. Then, the standardization of the ultrasonic image is processed by the skeleton extraction algorithm [26] to eliminate the thickness problem of different image contour caused by the sensitivity of the ultrasonic probe.

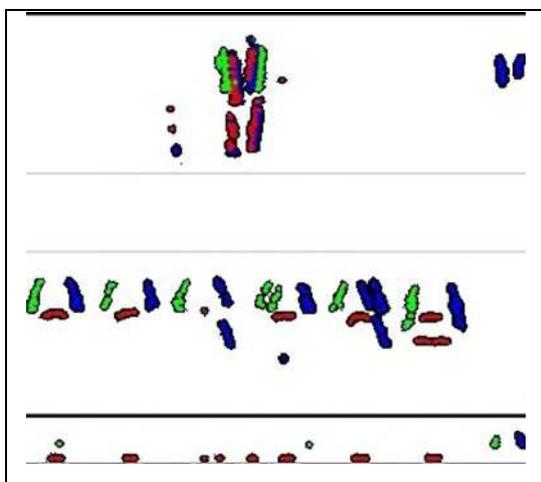


Figure 6 Flow chart of image preprocessing

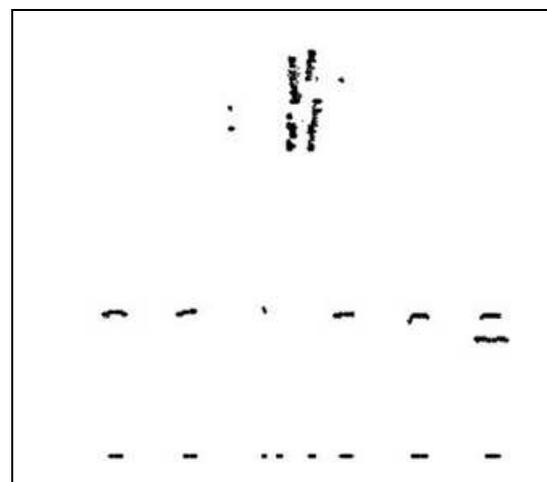
0	0	1	1	0	0
0	1	1	1	1	0
1	1	0	0	1	1
1	0	0	0	0	1

Figure 7 Morphological operators of the closed operation for inner and outer lower crack of screw hole

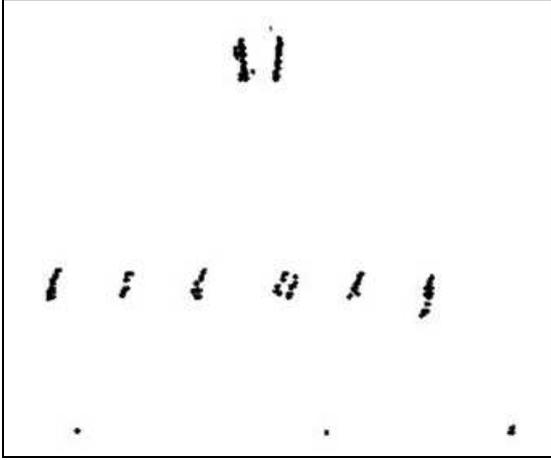
As shown in Figure 8(a), the image acquired by the train ultrasonic detector is shown in red, green and blue. Each color represents the defect detected by probes with different angles. Firstly, the image is separated into red, green and blue image by channel separation processing. Secondly, image denoising for each single channel image can eliminate the influence of small clutter on subsequent defect evaluation. Then, to improve the detection accuracy, the single channel image is refined to obtain the standardized monochrome ultrasonic image. Finally, the defect is located and detected by the proposed algorithm.



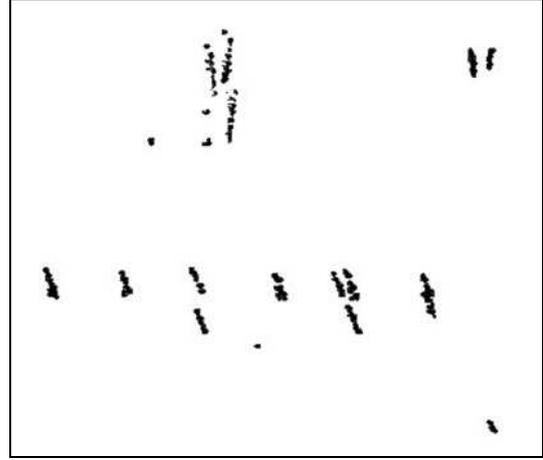
(a) Source image



(b) Red channel image



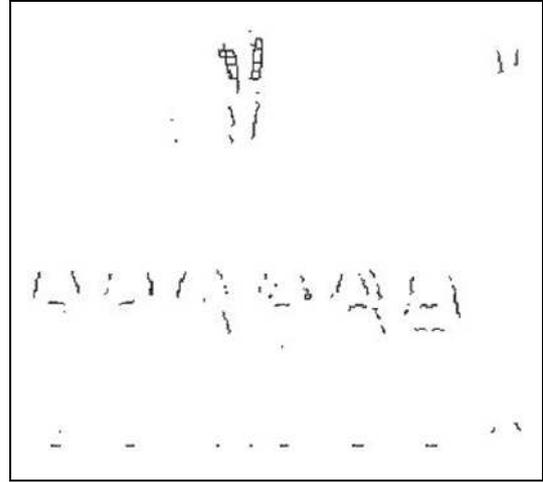
(c) Green channel image



(d) Blue channel image



(e) Morphological processing image



(f) Skeleton thinning image

Figure 8 Source image and preprocessing image

3.2 Model analysis of generalization features cluster

Generalization features cluster is defined as the overall evaluation of feature comparison, color recognition and Euclidean distance measurement. Its model can be expressed as follows.

$$\begin{cases} T_{min}^i < f_i(a_i) < T_{max}^i \\ C_i = R_i + G_i + B_i \\ D_{min}^i < \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} < D_{max}^i \end{cases} \quad (1)$$

In which, a_i is the feature of type i defect, such as its position region, area, length and slope. $f_i(a_i)$ means the feature function corresponding to a_i . T_{min}^i and T_{max}^i are the lower and upper thresholds of $f_i(a_i)$, respectively. In addition, their values are different for different defects. C_i is the color composition of type i defect. R_i , G_i and B_i represent the red, green and blue color components of type i defect, respectively. It should be noted that not all kinds of defects have these three channel colors. (x_i, y_i) and (x_0, y_0) are the centroid coordinates of type i defect and its corresponding reference, respectively. D_{min}^i and D_{max}^i are lower and upper limits of distance threshold,

respectively. Furthermore, the value of distance threshold is different to different defects.

Therefore, a defect can be classified as type i defect if it satisfies Eq. (1). For this way, every defect has the same evaluation method.

3.3 Generalization Features Clusters Analysis of Defects

Defects shown in Table 1 and Figure 5 are considered. Firstly, auxiliary lines in the image are analyzed and positioned to divide the image into three parts, including the rail head, rail waist and rail bottom. On this basic, different position region is used as original feature for rough classification. Secondly, to identify the defect more accurately, features of the defect, such as color, area, aspect ratio and location relationship, are analyzed and extracted. Then, to further increase the tolerance of extracted features to the same kind of defect, extracted features need to be transformed into generalization features. Finally, a suitable generalization features cluster composed of different

generalization features is used to detect the defect. The process of building generalization features clusters for the rail head, rail waist and rail bottom are analyzed and explained as follows.



Figure 9 Ultrasound image in rail head

Following, five generalization features, including position region, defect color, defect area, defect height ratio, and the distance between defect and joint, are selected and formulated to build a generalization features cluster. The gap between two rail joints is much larger than rail head flaw. The joints are detected by probes of 0° , 37° and 70° , and the image contour colors are red, green and blue. The ratio of their lengths to the depth of rail head is large. Rail head flaws can be divided into inner, middle, and outer flaws according to their different positions, which are distinguished by red, green and blue, respectively. In addition, flaw area in rail head is smaller than the joint area, and the ratio of its length to rail head height is small. After image preprocessing, a small number of larger clutters are also recognized as defects by generalization feature clusters. Usually, the larger clutter is caused by the separation of the ultrasonic probe from the track surface, which leads to a longer oblique line in the rail head region. Such generalization features can be expressed as follows.

Joint,

$$\begin{cases} 0 < y_c < h_1 \\ C_c = R + G + B \\ S_c \geq S_0 \\ \frac{l_c}{h_1} \geq l \\ \frac{H}{H_t} \geq H_{0d} \\ 0 < x_c < cols \end{cases} \quad (2)$$

Inner flaw of rail head,

$$\begin{cases} 0 < y_{r1} < h_1 \\ C_{r1} = R \\ S_{r1} < S_0 \\ \frac{l_{r1}}{h_1} < l \\ |x_{r1} - x_c| > m \end{cases} \quad (3)$$

Middle flaw of rail head,

$$\begin{cases} 0 < y_{g1} < h_1 \\ C_{g1} = G \\ S_{g1} < S_0 \\ \frac{l_{g1}}{h_1} < l \\ |x_{g1} - x_c| > m \end{cases} \quad (4)$$

3.3.1 Rail head

As shown in Figure 9, defects in rail head region include joints, inner flaw, outer flaw, middle flaw and clutter.

Outer flaw of rail head,

$$\begin{cases} 0 < y_{b1} < h_1 \\ C_{b1} = B \\ S_{b1} < S_0 \\ \frac{l_{b1}}{h_1} < l \\ |x_{b1} - x_c| > m \end{cases} \quad (5)$$

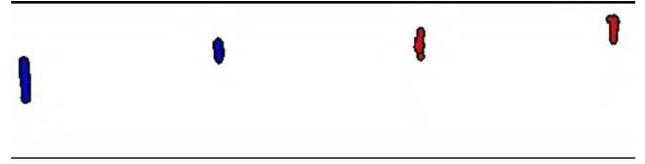
In Eq. (2) - (5), (x_c, y_c) , (x_{r1}, y_{r1}) , (x_{g1}, y_{g1}) and (x_{b1}, y_{b1}) are the centroid coordinates of joint, inner flaw, middle flaw and outer flaw in rail head, respectively. C_c , C_{r1} , C_{g1} and C_{b1} denote as the color composition of joint, inner flaw, middle flaw and outer flaw in rail head, respectively. S_c , S_{r1} , S_{g1} and S_{b1} are the area of joint, inner flaw, middle flaw and outer flaw in rail head, respectively. l_c , l_{r1} , l_{g1} and l_{b1} are the length of joint, inner flaw, middle flaw and outer flaw in rail head, respectively. In addition, S_0 is an area constant related to image resolution, H is the height of contour circumscribed rectangle for joint, H_t is the height of rail head region. H_{0d} is the ratio of the height of contour circumscribed rectangle for joint to the height of the rail head region. h_1 is the depth of the rail head. l is the constant of length proportional coefficient, which is set according to samples statistics and calculation. $cols$ is the width of detected image. m is the constant of horizontal distance between the defect and the joint. Generalization features clusters expressed by Eq. (2) - (5) can be used to detect joints and three kinds of rail head flaws.

Generally, generalized feature constraints of the rail head flaws include the longitudinal position of the flaw in rail head, the color composition of the rail head flaw, the area of the rail head flaw, the ratio of the length of the rail head flaw to the height of the rail head region, and the horizontal distance between the rail head flaw and the joint. The longitudinal position of the rail head flaw limits its longitudinal distribution in the rail head region. As shown in Figure 10, the position of the rail head flaw fluctuates up and down, but it is always in the rail head region. Therefore, the longitudinal position features have poor constraint but good generalization ability. The color composition of rail head flaw is an essential condition to distinguish its type and it has clear constraint directivity. The area of flaw reflects the size of the defect. While the area is not the same for each defect, it is generally

distributed in a specific range. After data statistics, its upper limit distribution can be defined by the area constant whose value setting is flexible. The ratio between the length of the rail head flaw and the height of the rail head region is an important criterion to distinguish the joint and the flaw, and it has a strong constraint and good features generalization ability. However, its threshold setting is critical. The horizontal distance constraint between the rail head flaw and the joint is to prevent the joint from being evaluated as the rail head flaw of the same color channel after channel separation. Generalized features constraints of joints are similar to those of rail head flaws.



(a) Rail head flaws at approximate height



(b) Rail head flaws at different height

Figure 10 Longitudinal distribution of rail head flaws

3.3.2 Rail waist

As shown in Figure 11, the ultrasound imaging of rail waist includes the screw hole, inner upper cracks of the screw hole, inner lower cracks of the screw hole, outer upper cracks of the screw hole, outer lower cracks of the screw hole, horizontal cracks of the screw hole, inner inverted cracks of the screw hole, outer inverted cracks of the screw hole and clutter.



Figure 11 Ultrasound image of the railway track waist

Following, seven generalization features, including position region, defect color, centroid coordinate, defect slope, defect length, ratio of defect area to defect length, and defect distance, are selected and formulated to build a generalization features cluster. Normal screw hole image contours are defined by semi-circular images composed of red, green and blue line segments, and one of them is marked with an ellipse in Figure 11. The image contour of upper crack in screw hole is located on the left or right side of the normal screw hole image contour, which is related to the defect position. The image contours of the lower crack and horizontal crack in screw hole both are located below the normal image contour of the screw hole. The image contour of inverted crack is located below the normal image contour of screw hole, and their directions are opposite. Usually, image contour of cluster is not a line segment, while it can be filtered by setting ratio constraint which is the ratio of image contour area to image contour length. For example, image contours on the inside of screw hole include four defects, including inner upper crack, inner lower crack and inner inverted crack of the screw hole, and the clutter. Generalization features clusters of image contours on the inside of screw hole can be expressed as follows.

The inner upper crack of the screw hole,

$$\left\{ \begin{array}{l} h_1 < y_{g2} < h_2 \\ C_{g2} = G \\ x_{g2} > x_g \\ k_g \cdot k_{g2} > 0 \\ l_1 < \frac{l_{g2}}{l_g} < l_2 \\ \frac{S_{g2}}{l_{g2}} < n \\ \sqrt{(x_g - x_{g2})^2 + (y_g - y_{g2})^2} < L_{min} \end{array} \right. \quad (6)$$

The inner lower crack of the screw hole,

$$\left\{ \begin{array}{l} h_1 < y_{g3} < h_2 \\ C_{g3} = G \\ y_{g3} > y_g \\ k_g \cdot k_{g3} > 0 \\ l_1 < \frac{l_{g3}}{l_g} \\ \frac{S_{g3}}{l_{g3}} < n \\ \sqrt{(x_g - x_{g3})^2 + (y_g - y_{g3})^2} < L_{min} \end{array} \right. \quad (7)$$

The inner inverted crack of the screw hole,

$$\left\{ \begin{array}{l} h_1 < y_{g4} < h_2 \\ C_{g4} = G \\ y_{g4} > y_g \\ k_g \cdot k_{g4} \leq 0 \\ l_1 < \frac{l_{g4}}{l_g} < l_2 \\ \frac{S_{g4}}{l_{g4}} < n \\ L_{min} < \sqrt{(x_g - x_{g4})^2 + (y_g - y_{g4})^2} < L_{max} \end{array} \right. \quad (8)$$

In Eq. (6) - (8), (x_g, y_g) is the centroid coordinates of screw hole. (x_{g2}, y_{g2}) , (x_{g3}, y_{g3}) and (x_{g4}, y_{g4}) are the centroid coordinates of inner upper crack, inner lower crack and inner inverted crack for screw hole, respectively. C_{g2} , C_{g3} and C_{g4} denote as the color composition of inner upper crack, inner lower crack and inner inverted crack for screw hole, respectively. k_g is the contour slope of screw hole. k_{g2} , k_{g3} and k_{g4} are the contour slope of inner upper crack, inner lower crack and inner inverted crack for screw hole, respectively. S_{g2} , S_{g3} and S_{g4} are the area of inner upper crack, inner lower crack and inner inverted crack for screw hole, respectively. l_g is the length of screw hole. l_{g2} , l_{g3} and l_{g4} are the length of inner upper crack, inner lower crack and inner inverted crack for screw hole, respectively. In addition, l_1 and l_2 are constants related to the ratios of defects length to the normal screw hole length. n is a constant related to the ratio of defect area to its length. h_2 is the height of rail waist region, which is expressed by the number of pixels. L_{min} and L_{max} are constants related to the distance between image contour centroids, which are the distance between the screw holes.

The combination of generalization features clusters expressed by Eq. (6) - (8) can detect three kinds of screw hole cracks, i.e. the inner upper crack, the inner lower crack and the inner inverted crack of screw hole. The outer upper

crack, the outer lower crack, the outer inverted crack and the horizontal crack of the screw hole can be detected by the same way.

The generalized features constraints of screw hole cracks can be summarized as following, the centroid height, the color composition, the position relationship between screw hole crack and its corresponding screw hole, the inclination relationship between screw hole crack and its corresponding screw hole, the length ratio between screw hole crack and its corresponding screw hole, the ratio of screw hole crack area and its length, the distance relationship between screw hole crack and its corresponding screw hole. Specifically, the centroid height of screw hole crack reflects its longitudinal position in screw hole. The color composition of screw hole crack reflects the angle of acquisition probe. The position relationship between the screw hole crack and its corresponding screw hole reflects that the screw hole crack appears above or below the screw hole. The inclination angle of the screw hole crack and its corresponding screw hole reflects whether they will be imaged by the same color channel. If the inclination direction is the same, they are imaged by the same color channel. Otherwise they are imaged by different color channels. Ratio of the screw hole crack length to the corresponding screw hole length reflects the size relationship of the screw hole crack. The ratio of crack area to its length approximately reflects the crack width, which is affected by the sensitivity of the probe. The distance between the screw hole crack and its corresponding screw hole is a significant constraint to evaluate the screw hole crack, and its threshold can be obtained by statistical methods.

3.3.3 Rail bottom

As shown in Figure 12, the defect of rail bottom contains longitudinal crack, transverse crack and clutter.



Figure 12 Ultrasound image of the rail bottom in railway track

Following, five generalization features of position region, defect color, defect slope, defect length and defect distance, all are selected and formulated to build a generalization features cluster. There are three types of image contours for transverse cracks, i.e. single blue oblique segment, single green oblique segment and blue-green oblique segment appearing in pairs. The longitudinal crack is a long line segment with red and horizon. Transverse cracks and longitudinal cracks are represented by generalization features cluster as follows. Transverse crack of rail bottom,

$$\left\{ \begin{array}{l} h_2 < y_i < h_3 \\ C_i = G|B|(G+B) \\ k_{i1} \cdot k_{i2} < 0 \\ l_i > 0 \\ \sqrt{(x_{i1} - x_{i2})^2 + (y_{i1} - y_{i2})^2} < l_{min} \end{array} \right. \quad (9)$$

Longitudinal crack of rail bottom,

$$\left\{ \begin{array}{l} h_3 < y_t < r_{ow} \\ C_t = R \\ k_t \approx 0 \\ l_t > T_{min} \end{array} \right. \quad (10)$$

In Eq. (9) and (10), (x_i, y_i) and (x_t, y_t) are the centroid coordinates of transverse and longitudinal crack in rail bottom, respectively. C_i and C_t denote as the color composition of transverse crack and longitudinal crack for rail bottom, respectively. l_i and l_t are the length of transverse crack and longitudinal crack for rail bottom, respectively. (x_{i1}, y_{i1}) and (x_{i2}, y_{i2}) are the centroid coordinates of two adjacent rail bottom defects respectively. k_{i1} and k_{i2} are the slope of two adjacent rail bottom defects respectively. In addition, if the distance between these two points is less than the setting distance I_{min} , the transverse crack of rail bottom is composed of blue and green defect contours in pairs, and the contours slope directions are opposite. k_t is the slope of longitudinal crack, and the image contour of longitudinal crack in rail bottom is approximately horizontal. r_{ow} is the rows of ultrasonic images, and T_{min} is the length threshold. The combination of generalization features clusters expressed by Eq. (9) - (10) can be used to detect transverse and longitudinal cracks in rail bottom.

3.4 Feature Clustering and Dimension Reduction

In the generalization features cluster, constraint condition of each feature is different. Some features have good constraint effect while some features have poor constraint effect. In this paper, by analyzing the constraint effect of each feature, the feature with strong constraint effect is selected. Then, on this basic, new constraint features are constructed to achieve dimension reduction and simplify defects identification. In the rail head region, S_c is the area of contour circumscribed rectangle, and H_{0d} is the ratio of the height for contour circumscribed rectangle to the height for rail head region, which both are strong constraint generalization features for evaluating the rail head flaw and joint. Usually, there are various kinds of screw hole cracks in the rail waist, which are evaluated by the position relationship between screw hole and defect. Therefore, it is difficult to constrain defects uniformly with the reduction dimension's generalized features. However, for a specific type of defect, it can be identified by using the reduced dimension generalization features cluster. For example, the contour centroid height and defect area are strong constraint generalization features to evaluate the horizontal crack of the screw hole. The centroid height and the contour circumscribed rectangle length are strong constraint generalization features to evaluate the lower crack of the screw hole. The angle formed by green and blue contours and the horizontal distance between two centroids are also strong constraint generalization features to evaluate the rail bottom defects.

In this paper, the K-means clustering method is also developed to classify defects in different regions with following ways.

(1) The number of defect types k is determined in different regions of the rail, and the strong constraint

generalization features of various kinds of defects are extracted respectively. Then, the data set is built by the strong constraint generalization feature parameter data. Moreover, normalization and standardization processing are required for the data with following expression,

$$\frac{x_{ij} - \mu_i}{\sigma_i} \quad \text{and} \quad \frac{x_{ij} - x_{imin}}{x_{imax} - x_{imin}} \quad (11)$$

In which, x_{ij} is the j th sample point of class i defect data, μ_i and σ_i are the mean and variance of class i defect data, respectively. x_{imax} and x_{imin} are the maximum and minimum of class i defect data, respectively.

(2) k clustering centers are selected randomly, which can refer to the initial setting threshold of related constraint features. Furthermore, the Euclidean distance from the rest of the data points to the clustering center is calculated, and then the nearest clustering point is selected as their types.

(3) All data points belonging to the same cluster are processed with centroid operation, then a new cluster center is calculated.

(4) Process of (2) to (3) is repeated until the cluster center no longer changes.

To evaluate the clustering, the DBI is used [27, 28]. DBI Index is the mean value of measuring the maximum similarity of each cluster, and it is an important index to measure the intra cluster distance and the cluster spacing. The smaller the value, the better the clustering effect. Assuming that there are m digital sequences which are clustered into n clusters, then m digital sequences are set as input matrix X , and n clusters are used as parameter to input algorithm. The DBI can be expressed as

$$DBI = \frac{1}{n} \sum_{i=1}^n \max \left(\frac{\bar{s}_i + \bar{s}_j}{\|c_i - c_j\|_2} \right) \quad (12)$$

In which, S_i is the average distance from the data of class i defect to the cluster center of its corresponding class. It can be expressed as

$$S_i = \left(\frac{1}{T_i} \sum_{j=1}^{T_i} |x_{ij} - C_i|^p \right)^{1/p} \quad (13)$$

In which, T_i is the number of data in class i , C_i is the cluster center of class i , p is used to calculate the Euclidean distance of two-dimensional data and $p=2$. The distance between cluster i defect and cluster j defect is defined as

$$C_{i,j} = \|C_i - C_j\|_p \quad (14)$$

DBI index can be defined with the same way.

$$R_{i,j} = \frac{S_i + S_j}{C_{i,j}} \quad (15)$$

Following, the $R_{i,j}$ of the maximum value for each cluster class i defect is calculated. Then,

$$D_i = \max R_{i,j}, i \neq j \quad (16)$$

D_i is the maximum similarity between cluster class i and other cluster classes. The DBI index is obtained by considering the mean of the maximum similarity of all classes,

$$DBI = \frac{1}{n} \sum_{i=1}^N D_i \quad (17)$$

Therefore, the algorithm proposed in this paper can be expressed in Figure 13. Firstly, various defects in ultrasound images are classified according to the classification information shown in Table I. Secondly, defect images are preprocessed to achieve image enhancement, including color channel separation, image

denoising, image standardization and image contour location. On this basic, the generalization features of various types of defects are analyzed. For example, in rail head region, after analyzing the influence of various defects features on the detection effect, five generalization features with better detection effect are selected to detect defects. Then, they are formulated and expressed by theoretical analysis. Finally, generalization features clusters of all kinds of defects are constructed by the adopted generalization features for defects identification and judgement. In addition, based on the proposed generalization features, K-means clustering algorithm is used to verify the validity of strong constraint generalization features based on dimension reduction and achieve the identification and classification of defects.

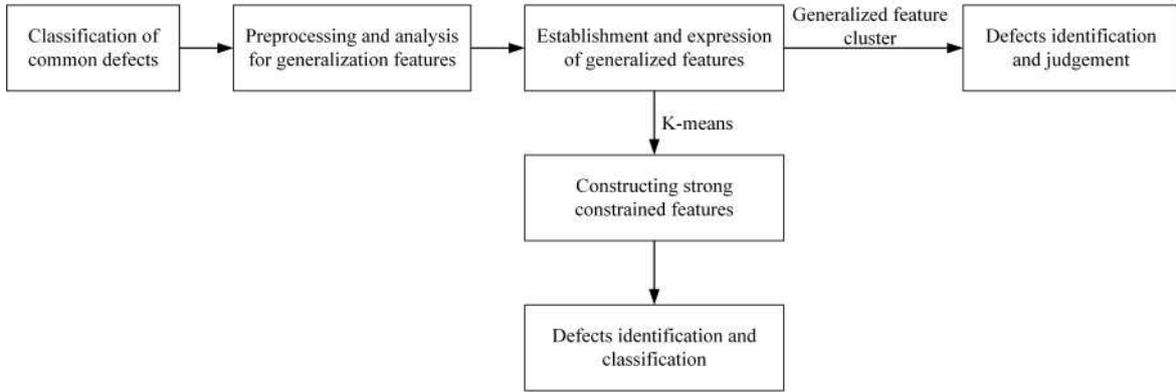


Figure 13 The flow chart of the proposed algorithm

4 Experimental Results and Analysis

4.1 Experimental Results and Analysis of The Proposed Method

To make the proposed algorithm clear, the detection method of rail head flaws is taken as an example to illustrate it again, and other types of defects detection is the same. Firstly, the image is preprocessed for image enhancement and removing noise. Detailed steps include color channel separation, image denoising, image standardization and image contour location. Secondly, generalized features of rail head flaws are formulated and expressed, including position region, defect color, defect area, defect height ratio, and the distance between defect and joint. Finally, these generalization features are used to construct an appropriate generalization features cluster for the identification and judgment of rail head flaws. On the other hand, after extracting all kinds of generalization features of flaws, the strong constraint generalization feature based on dimension reduction is analyzed and obtained, and K-means clustering algorithm is applied to identify and classify the rail head flaws.

Following, to evaluate the effectiveness of the proposed method, sample images and tested images are detected and

analyzed to verify its detection accuracy and speed. The cooperator, Guangdong Goworld Co., Ltd, provides 43 sample images with a resolution of 2000 * 400. Defects of sample images are labelled by three experienced track inspection workers before testing. Then, the proposed method is used to extract the generalization feature clusters from 43 sample images and detect defects. In this experiment, the misjudgment rate and the missed detection rate are also analyzed.

$$\text{Accuracy rate: } P_1 = \frac{e_e}{s} \quad (18)$$

$$\text{Misjudgment rate: } P_2 = \frac{e_r}{s} \quad (19)$$

$$\text{Missed detection rate: } P_3 = \frac{r_e}{s} \quad (20)$$

In Eq. (18)-(20), P_1 is the accuracy rate, P_2 is the misjudgment rate, and P_3 is the missed detection rate. s is the number of defects counted by manual inspection, e_e is the number of defects consistent with the results of the manual detection. e_r is the number of a type of defects

detected artificially but detected to be other type of defects by the proposed method. r_e is the number of defects detected artificially but detected to be normal by the proposed method. If there is no missed inspection and misjudgment, then $P_3 = P_2 = 0$, $P_1 = 1$. If there is no missed inspection and only false inspection, $P_3 = 0$, $P_2 = 1 - P_1$. Then 43 sample images and 953 testing images are detected, and the experimental results are shown in Table 2, Table 3 and Table 4, respectively.

Table 2 Test results of ultrasonic image for railway track sample images

Defect serial number	Defects number	Missed detection rate (P_3)	Misjudgment Rate (P_2)	Accuracy rate (P_1)
1	75	0	1.33%	98.67%
2	27	0	0	1
3	101	0	0.99%	99.01%
4	51	1.96%	00	98.04%
5	102	0.98%	0	99.02%
6	19	5.26%	0	94.74%
7	42	0	0	1
8	109	0.92%	0	99.08%
9	19	0	0	1
10	79	0	0	1
11	50	0	4.00%	96.00%
12	107	0	0	1
Total	781	0.38%	0.64%	98.98%

As shown in Table 2, 43 sample images are tested. Its detection accuracy of the proposed method is 98.98%, and the average detection speed is 0.17s per sample image. While the misjudgment rate is 0.64%, the missed detection rate is 0.38%.

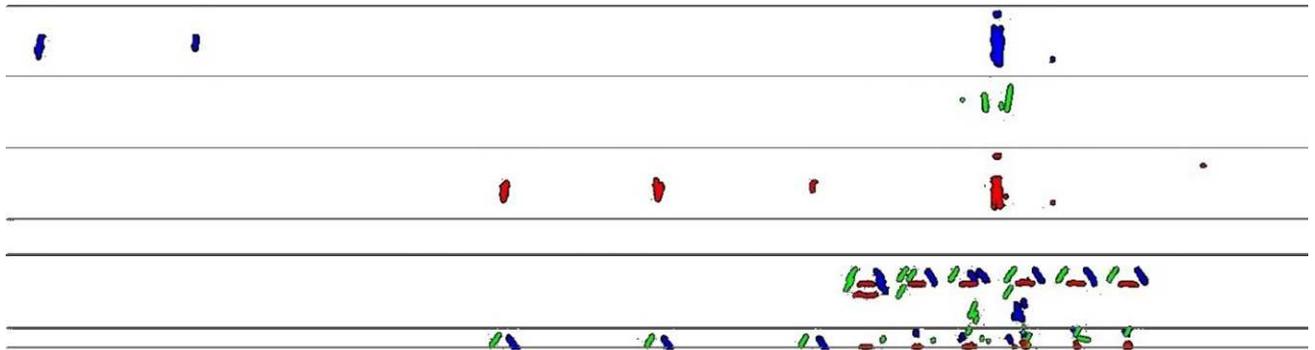
On the other hand, 953 railway track images are detected with the proposed method based on the generalization

features cluster extracted from the sample image, and its results are shown in Table 3. The detection accuracy is 97.55%, while the misjudgment rate is 1.14%, and the missed detection rate is 1.31%. in addition, the average detection time is 0.12s. The detection speed is faster than before, which validates the effectiveness of the proposed method.

The type I and type II track ultrasonic image are shown in Figure 14. Three channels of the rail head region are displayed respectively in type II track ultrasonic image that includes 7 boundary lines. In addition, channels of the rail head region are displayed too in type I track ultrasonic image that includes 5 boundary lines. 300 type II track testing images are detected by adjusting constraint parameters related to their position and region. Testing results are shown in Table 4.

Table 3 On-site ultrasound image test results for railway track

Defect serial number	Defects number	Missed detection rate (P_3)	Misjudgment Rate (P_2)	Accuracy rate (P_1)
1	347	1.44%	0.29%	98.27%
2	86	3.49%	0	96.51%
3	336	1.19%	0.30%	98.51%
4	132	0.76%	1.52%	97.73%
5	98	1.02%	2.04%	96.94%
6	13	0	15.38%	84.62%
7	197	0	1.02%	98.98%
8	166	1.81%	1.20%	96.99%
9	30	3.33%	10.00%	86.67%
10	154	0	1.30%	98.70%
11	215	2.79%	1.40%	95.81%
12	61	0	1.64%	98.36%
Total	1835	1.31%	1.14%	97.55%



(a) Type II track ultrasonic image



(b) Type I track ultrasonic image

Figure 14 Type I and type II railway track ultrasound image

Table 4 Type II railway track ultrasound image test results

Defect serial number	Defects number	Missed detection rate (P_3)	Misjudgment Rate (P_2)	Accuracy rate (P_1)
1	23	0	0	1
2	6	0	0	1
3	29	3.45%	0	96.55%
4	13	0	0	1
5	23	0	4.35%	96.65%
6	6	0	0	1
7	34	0	0	1
8	18	0	5.56%	94.44%
9	4	0	25.00%	75.00%
10	21	0	0	1
11	14	7.12%	0	92.88%
12	25	0	0	1
Total	216	0.93%	1.39%	97.69%

Through comparison among Table 2-Table 4, the detection accuracy rate is higher than 97% for different types of track defects image by using the generalization features cluster extracted from the same sample image. It means that the proposed method has good applicability and practicability for railway track defect detection. However, the rate of missing detection and misjudgment for testing image is slightly higher than that rate of the sample image. For example, the rate of misjudgment for the inner and outer inverted screw hole laceration was 15.38 %, 10.00% and 25.00%, respectively. The main reason is that the inverted crack of screw holes is located near the joints. The ultrasonic probe will be off the rail surface in a short time because of the gap at the joints. And the ultrasonic wave attenuates rapidly in the air, which results in the occurrence of a variety of clutter with different patterns and sizes in this region. Moreover, some clutter images contour near the joint are complex, and a few clutters are mistaken for inverted crack of screw hole because their contours conform to the features of generalization features clusters.

For these reasons, the rail waist region should be divided into the near joint region and far joint region, and the generalization features cluster should be extracted respectively, which will reduce the detection speed but improve the detection accuracy.

4.2 Comparison with Other Methods

To verify the detection performance of the proposed method, it is compared with two methods. One is the BP neural network method [18]. Another is the intelligent method of rail defect identification based on deep learning [29]. In Ref. [18], different neural network structures are used to identify different types of defects, and the initial weights are defined by empirical values. Then, momentum BP method and sigmoid function are used to train and output results, respectively. In Ref. [29], the method is proposed based on the framework of AlexNet convolutional neural network, and the softmax function is used to output results while the training is carried out in TensorFlow framework. The results are shown in Figure 15. The testing data and images of the proposed method all are from the EGT-60 double rail defect detector.

As shown in Figure 15, X-coordinates 1-12 are the number of the defect type corresponding to Table 1, and X-coordinates 13 is the average accuracy of each corresponding method. Ref. [18] only detects screw holes. And its detection accuracy for defect type 4-10 is 95%, which means that its average detection accuracy is also 95%. In addition, it needs to adjust the correlation thresholds of the hidden layer and output layer of neural network. Ref. [29] also classifies the screw holes as screw hole defects, and the detection accuracy of defect type 4-10 is 96%. That is, the average detection rate is 94.2%. While, by comparison, the average defect detection rate of the proposed method is up to 97.60%. Therefore, considering the main generalization features of the restraint defect, the proposed method performs better than these two methods obviously. The experimental results further validate the effectiveness of the proposed method for railway track internal defect detection.

4.3 Evaluating the Effectiveness of Strong Constraint Generalization Features after Reducing Dimension

To validate the effectiveness of the proposed clustering method that uses reduced dimension strong constraint generalization features, K-means clustering method is used to classify the rail head flaw and joint. Moreover, the DBI

index and correct and wrong classification index are also used to evaluate the clustering results. The results are shown in Figure 16. The number of samples are 120, 300 and 480 data points, respectively.

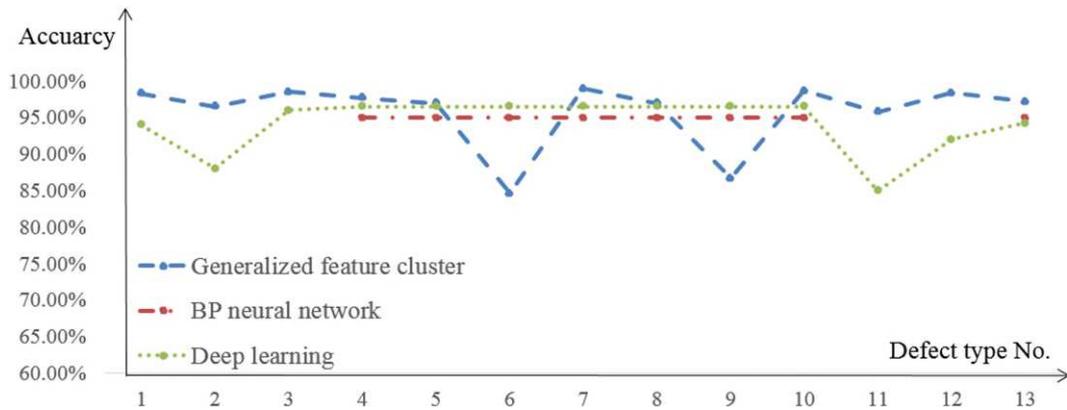
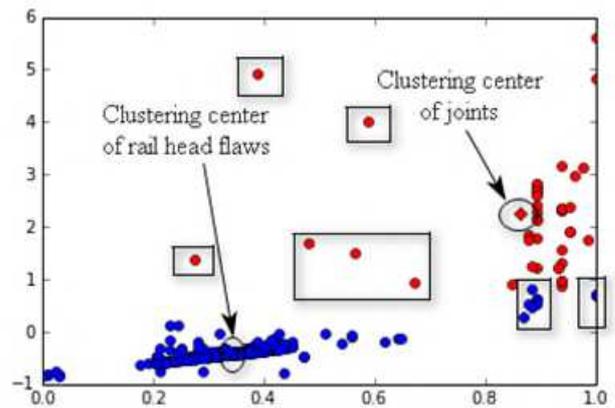


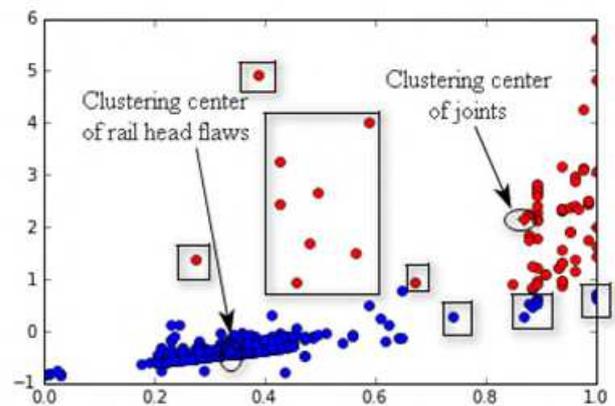
Figure 15 The comparison result of three methods

As shown in Figure 16, the blue and red dots are the clustering results of rail head flaw and joint, respectively. Moreover, the blue and red squares are the clustering centers of rail head flaw and joint, respectively. In addition, dots marked with rectangle are the data of error clustering. A few of them are considered as joints by K-means algorithm due to their locations. Such way usually cause misclassification. One of the reasons may be that they are noise with large image area and not removed during preprocessing. Another possible reason is that the contour of some flaws is changed after the morphological processing, which makes the area and height change too and results in misclassification.

Furthermore, from Figure 16, the distribution of the rail head flaw is relatively centralized and stable, and the distribution of the joint is relatively scattered. The reason is that joint is imaged by 70° end face wave with large amplitude and long displacement, which leads to the scattered situation of joint imaging. Therefore, joints present the feature of a broad span of data distribution.

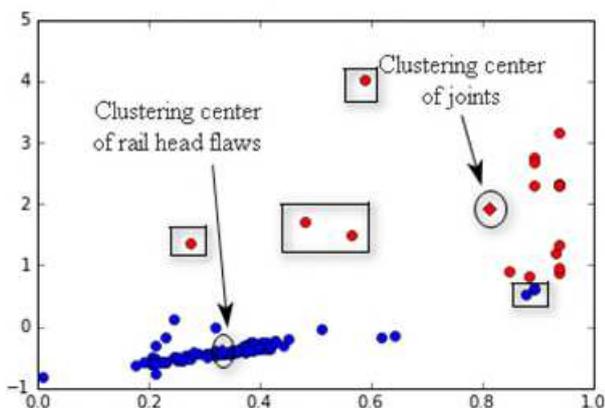


(b) Cluster results of 300 groups of data



(c) Cluster results of 480 groups of data

Figure 16 Clustering results of main defect in rail head region



(a) Cluster results of 120 groups of data

Table 5 shows detailed clustering results. From the cluster center, it shows that the cluster center of the rail head flaw is relatively stable, while the cluster center of

joint fluctuates. With the increase of the number of samples, the accuracy of clustering is gradually increasing. However, the number of error clustering is also increasing. The reason is that the K-means clustering algorithm only measures the similarity among data points, and it does not consider the correlation of features contained in the same set of data, such as the weight proportion of area parameter and height parameter. It can be solved by keeping the

stability of B-scan image acquisition to avoid the appearance of extreme data. In addition, it also can be inferred that the DBI index of several clustering results is smaller, which indicates that the clustering effect is improved, and the defect identification and classification can be achieved by strong constraint generalization features.

Table 5 Clustering results of rail head region defects

-		120 sets of data	300 sets of data	480 sets of data
Clustering center	Joints	[0.8132,1.9165]	[0.8646,2.2438]	[0.8695,2.1739]
	Rail head flaws	[0.3324, -0.3929]	[0.3388, -0.3922]	[0.3366, -0.3709]
Clustering center distance		2.3589	2.6879	2.6000
Number of positive samples		16	45	69
Number of negative samples		104	255	411
Correct clustering		114	287	461
Error clustering		6	13	19
Clustering accuracy		95.00%	95.67%	96.04%
DBI index		0.0816	0.0283	0.0521

5 Conclusions

To achieve rail internal flaw detection, an internal defect detection method based on generalization features cluster is proposed in this paper. The main conclusions are as follows:

(1) The proposed method can avoid the instability, over-constraint, or under-constraint of feature expression during detection processing. Experimental results show that the proposed method can also perform better than other methods with the accuracy of 97.55% and the average detection time of 0.12s/frame.

(2) K-means method is developed to cluster the strong constrained generalized features after dimension reduction. Besides, the DBI index and clustering accuracy are used to evaluate the results, and good clustering results are achieved with the accuracy of over 95%.

(3) Comprehensive experimental results show that the proposed method has higher detection accuracy and better application prospects than developed methods, which can be used to improve the operation safety of high-speed rail and rail transit systems.

The further work will be studying the internal defects detection method based on deep learning.

Acknowledgements

The authors would like to thank Guangdong Goworld Co. Ltd, China, for providing with experimental facilities and testing assistance.

Authors' Contributions

FW determined the overall framework and central idea of the article; XX was responsible for a part of experiments, and checked and approved the manuscript; JG was in charge of the literature search and wrote part of the manuscript; QL wrote part of the manuscript. All authors read and approved the final manuscript.

Authors' Information

Fupei Wu, received the scholar from Northeast Petroleum University in 2003, received the master from Shantou University in 2006, and received the Ph.D. from South China University of Technology in 2009. Now he works in Shantou University as an Assistant professor. His research interests include Automatic optical inspection and 3D measurement. E-mail: fpwu@stu.edu.cn.

Xiaoyang Xie, received a bachelor's degree from Shantou University in 2017. He is now studying for a master's degree in mechanical and electronic engineering at Shantou University. His research interests include machine vision and image processing. E-mail: 13xyxie1@stu.edu.cn.

Jiahua Guo, received a bachelor's degree from Hunan University of technology in 2016, a master's degree from Shantou University in 2019. His main research interests include machine vision, 3D reconstruction and measurement, and image processing. E-mail: 16jhguo@alumni.stu.edu.cn.

Qinghua Li, received a bachelor's degree from Chongqing University of post and telecommunications in 2017, a master's

degree from Shantou University in 2020. His research interests include machine vision and image processing. Corresponding author. E-mail: 17qhli@alumni.stu.edu.cn.

Funding

Supported in part by the National Natural Science Foundation of China (Grant No. 61573233), the Natural Science Foundation of Guangdong, China (Grant No. 2018A0303130188), the Guangdong Science and Technology Special Funds Project (Grant No. 190805145540361), and Special projects in key fields of colleges and universities in Guangdong Province (Grant No. 2020ZDZX2005).

Competing Interests

The authors declare no competing financial interests.

Author Details

¹ Key Laboratory of Intelligent Manufacturing Technology, Ministry of Education, Shantou University, Shantou 515063, China.

References

- [1] M Leboeuf. High-Speed Rail: Opportunities and threats. *Engineering*, 2016, 2(4): 36-50.
- [2] A Poudel, B Lindeman, R Wilson. Current practices of rail inspection using ultrasonic methods: A review[J]. *Materials Evaluation*, 2019, 77(7): 870-883.
- [3] X C Bian, H G Jiang, Y M Chen. Accumulative deformation in railway track induced by high-speed traffic loading of the trains[J]. *Earthquake Engineering and Engineering Vibration*, 2010, 9(3): 319-326.
- [4] H M Shi, Z J Yu, L Q Zhu, et al. Study on the monitoring method of longitudinal displacement for the High-speed continuous welded Rail[J]. *Chinese Journal of Scientific Instrument*, 2016, 37(4): 811-817. (in Chinese)
- [5] H M Yu, Q Y Li, Y Q Tan, et al. A Coarse-to-Fine model for rail surface defect detection[J]. *IEEE Transactions on Instrumentation and Measurement*, 2018, 68(3): 656-666.
- [6] A K Dubey, Z A Jaffery. Maximally stable extremal region marking-based railway track surface defect sensing[J]. *IEEE Sensors Journal*, 2016, 16(24): 9047-9052.
- [7] Q Y Li, S W Ren. A visual detection system for rail surface defects[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 2012, 42(6): 1531-1542.
- [8] H Zhang, X T Jin, Q M J Wu, et al. Automatic visual detection system of railway surface defects with curvature filter and improved gaussian mixture model[J]. *IEEE Transactions on Instrumentation and Measurement*, 2018, 67(7): 1593-1608.
- [9] G Y Tian, B Gao, Y L Gao, et al. Review of railway rail defect non-destructive testing and monitoring[J]. *Chinese Journal of Scientific Instrument*, 2016, 37(8): 1763-1780. (in Chinese)
- [10] S Mariani, T Nguyen, X Zhu, et al. Field test performance of noncontact ultrasonic rail inspection system[J]. *Journal of Transportation Engineering, Part A: Systems*, 2017, 143(5): 04017007.
- [11] H Cygan, L Girardi, Patrice Aknin, et al. B-scan ultrasonic image analysis for internal rail defect detection[J]. *WCRR 2003 Edinburgh Scotland*, 2003.
- [12] A Y Liang, S Sternini, L D S Francesco, et al. Rail flaw identification using ultrasonic imaging[R]. Transportation Research Board 97th Annual Meeting, 2018.
- [13] W J Wang, R Lewis, B Yang, et al. Wear and damage transitions of wheel and rail materials under various contact conditions[J]. *Wear*, 2016, 362-363: 146-152.
- [14] B Cai, K Shen, J L Fu, et al. Research on defect detection of X-ray DR images of casting based on Mask R-CNN[J]. *Chinese Journal of Scientific Instrument*, 2020, 41(3):61-69. (in Chinese)
- [15] W M Guo, K Liu, H F Qu. Welding defect detection of X-ray images based on Faster R-CNN model[J]. *Journal of Beijing University of Posts and Telecommunications*, 2019, 42(6): 20-28.
- [16] S Bahamon-Blanco, S Rapp, C Rupp, et al. Recognition of track defects through measured acceleration - part 1[J]. *IOP Conference Series Materials Science and Engineering*, 2019, 615: 012121.
- [17] Sperry B-Scan Dual Rail Inspection System, available: http://www.sperryrail.com/_customelements/uploadedResources/SperryDUALweb.pdf, 12.4.2016.
- [18] X Y Huang, Y S Shi, Y H Zhang, et al. BP neural network based on rail flaw classification of RFD car's B-scan data[J]. *China Railway*, 2018(3): 82-87. (in Chinese)
- [19] P W Loveday. Guided wave inspection and monitoring of railway track[J]. *Journal of Nondestructive Evaluation*, 2012, 31(4): 303-309.
- [20] D Mukherjee, Q M J Wu, G H Wang. A comparative experimental study of image feature detectors and descriptors[J]. *Machine Vision and Applications*, 2015, 26(4): 443-466.
- [21] S Kenderian, B B Djordjevic, R E Green Jr. Point and line source laser generation of ultrasound for inspection of internal and surface flaws in rail and structural materials[J]. *Research in Nondestructive Evaluation*, 2001, 13(4): 189-200.
- [22] Z D He, Y N Wang, J X Mao, et al. Research on inverse P-M diffusion-based rail surface defect detection[J]. *Acta Automatica Sinica*, 2014, 40(08): 1667-1679. (in Chinese)
- [23] A V Favorskaya, S V Kabisov, I B Petrov. Modeling of ultrasonic waves in fractured rails with an explicit approach[J]. *Doklady Mathematics*, 2018, 98(1): 401-404.
- [24] Q R Xu, Y S Shi, L H Xiong, et al. Research on detection technology of Top-Surface damage and magnetic leakage based on large flaw detection vehicle for rails[J]. *China Railway*, 2017(10): 39-44. (in Chinese)
- [25] R Stock, R Phipps. Rail grade dependent damage behaviour-Characteristics and damage formation hypothesis[J]. *Wear*, 2014, 314(1-2): 44-50.
- [26] C H Yang, C Liu. Skeleton extraction algorithm of optimized gradient path[J]. *Journal of Xiamen University(Natural Science)*, 2014, 53(2): 201-205. (in Chinese)
- [27] Z Y Wang, S Farhand, G Tsechpenakis. Fading affect bias: improving the trade-off between accuracy and efficiency in feature clustering[J]. *Machine Vision and Applications*, 2019, 30(2): 255-268.
- [28] S X Lu, G Y Lin, H L Liu, et al. A weekly load data mining approach based on hidden markov model[J]. *IEEE Access*, 2019, 7: 34609-34619.
- [29] C S Sun, J Liu, Y Qin, et al. Intelligent detection method for rail flaw based on deep learning[J]. *China Railway Science*, 2018, 39(5): 51-57. (in Chinese)

Figures

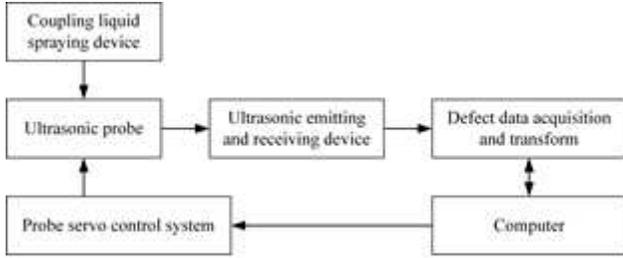


Figure 1

Rail B-scan image acquisition mechanism

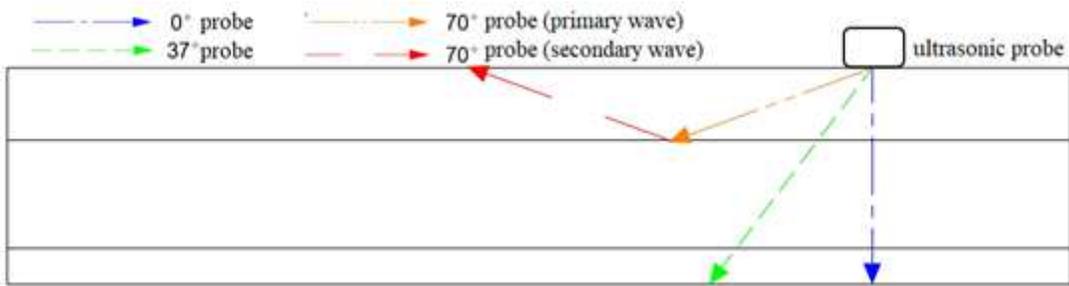


Figure 2

The reflection diagram of ultrasound in railway track

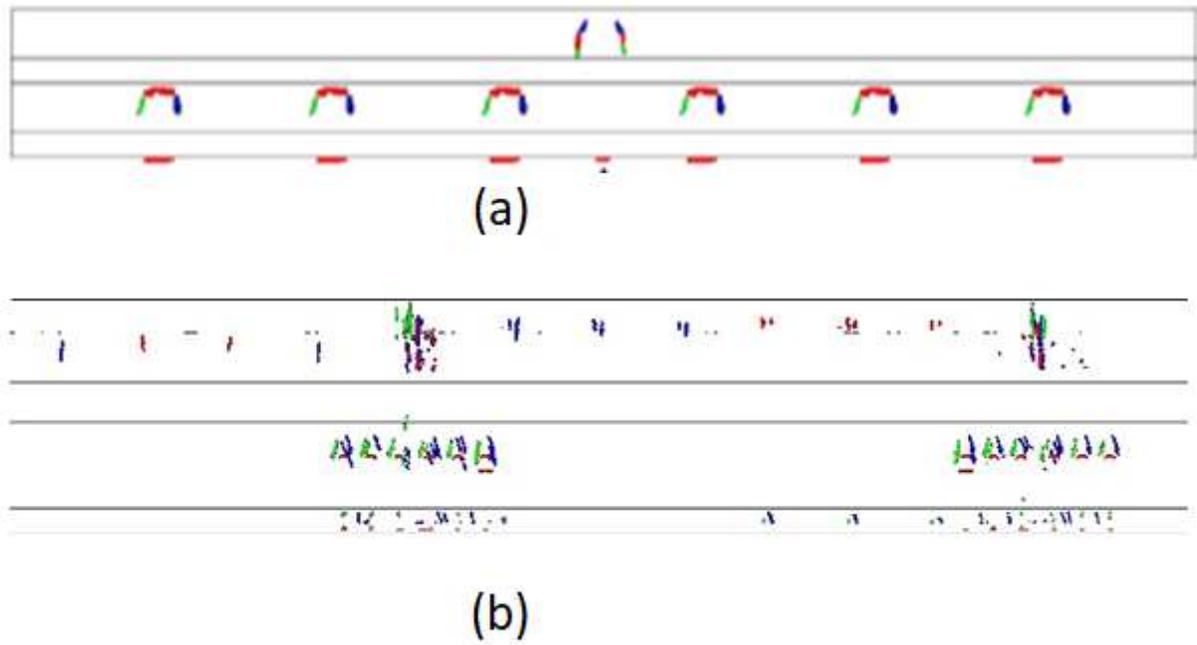


Figure 3

Ultrasonic B-scan images in different situations (a) The B-scan image for normal railway track (b) The actual B-scan image with defects

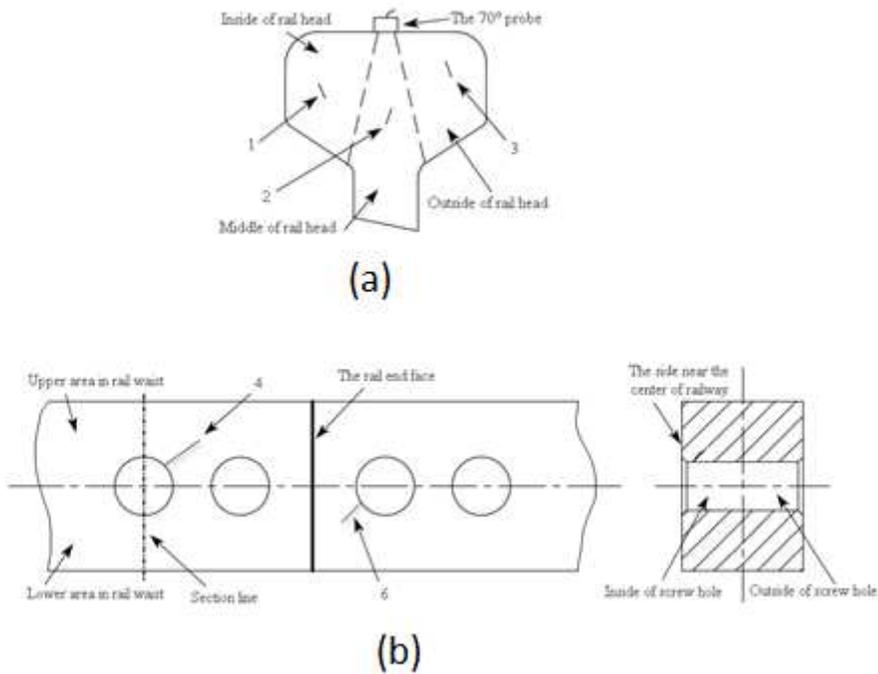


Figure 4

The position diagram of some defects (a) Defects in rail head (b) Defects in rail waist

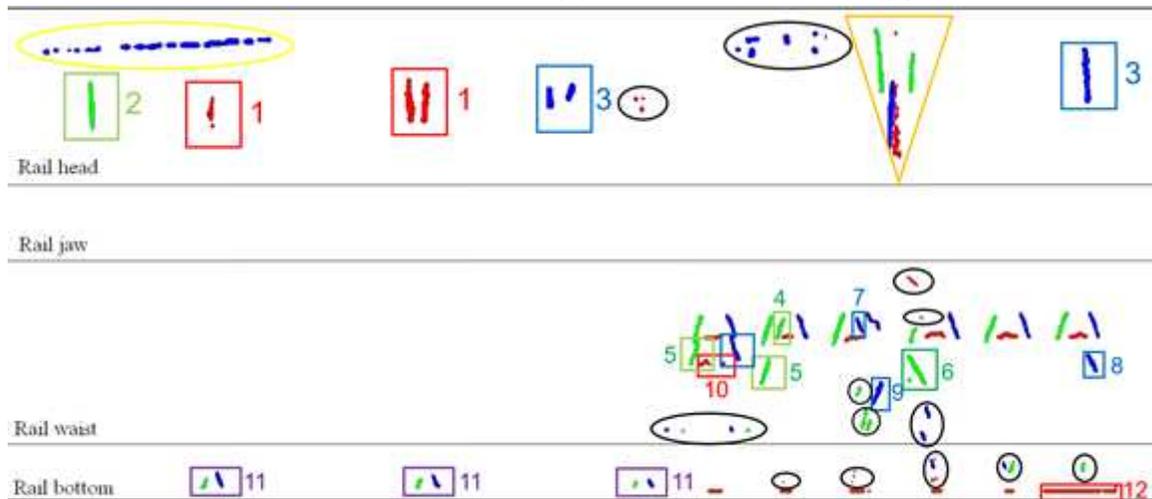


Figure 5

All kinds of defects and non-defects in B-scan image

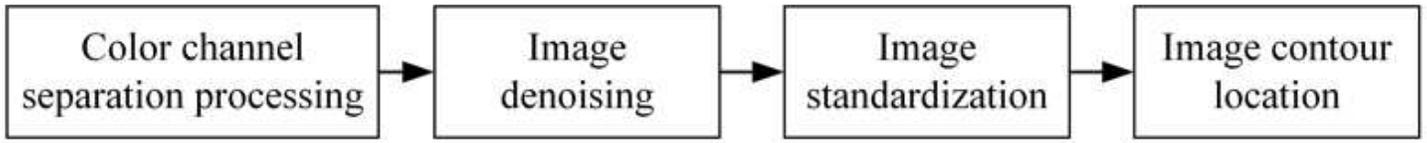


Figure 6

Flow chart of image preprocessing

0	0	1
0	1	1
1	1	0
1	0	0

1	0	0
1	1	0
0	1	1
0	0	1

Figure 7

Morphological operators of the closed operation for inner and outer lower crack of screw hole

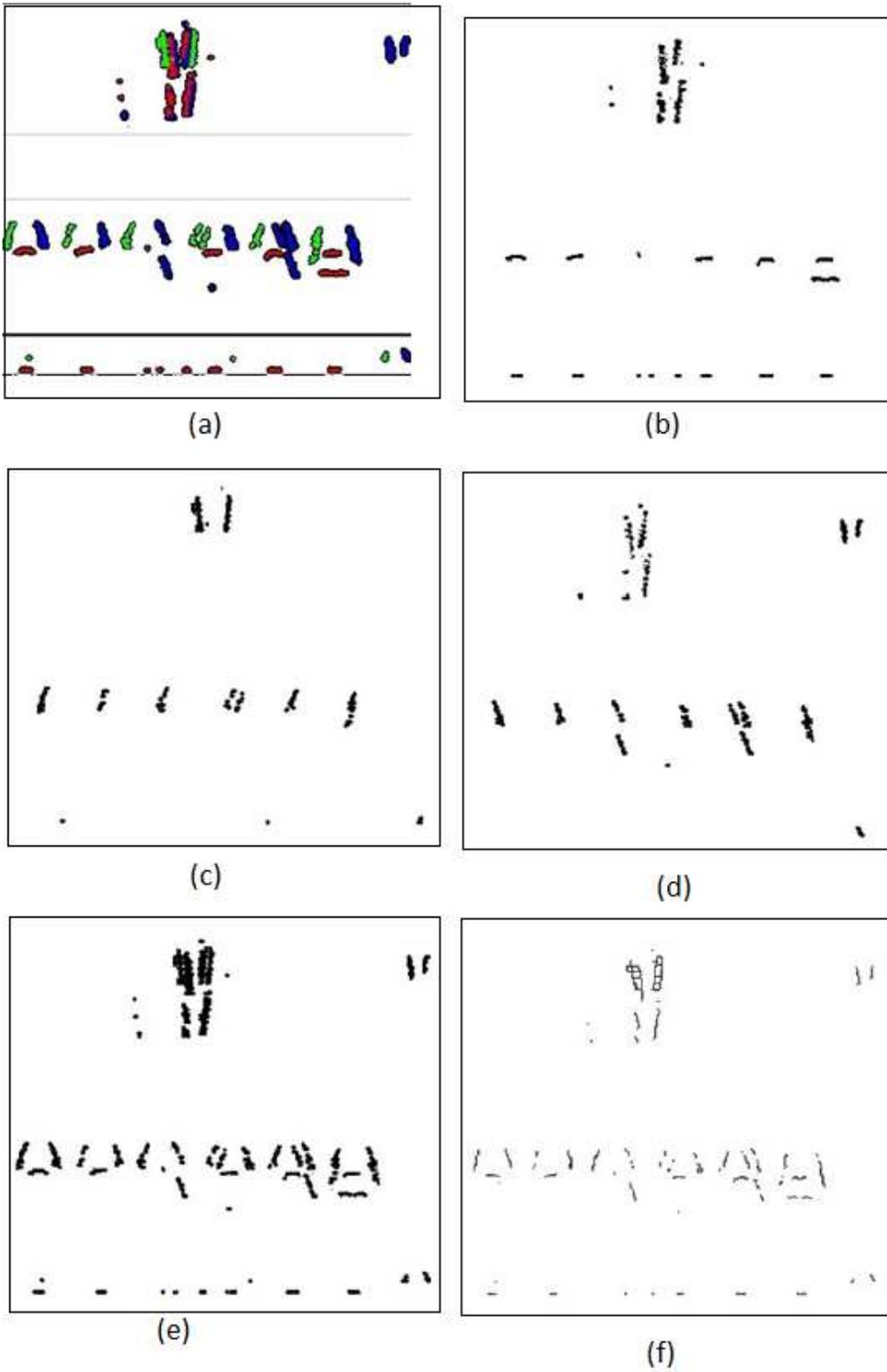


Figure 8

Source image and preprocessing image (a) Source image (b) Red channel image (c) Green channel image (d) Blue channel image (e) Morphological processing image (f) Skeleton thinning image



Figure 9

Ultrasound image in rail head



(a)



(b)

Figure 10

Longitudinal distribution of rail head flaws (a) Rail head flaws at approximate height (b) Rail head flaws at different height



Figure 11

Ultrasound image of the railway track waist



Figure 12

Ultrasound image of the rail bottom in railway track

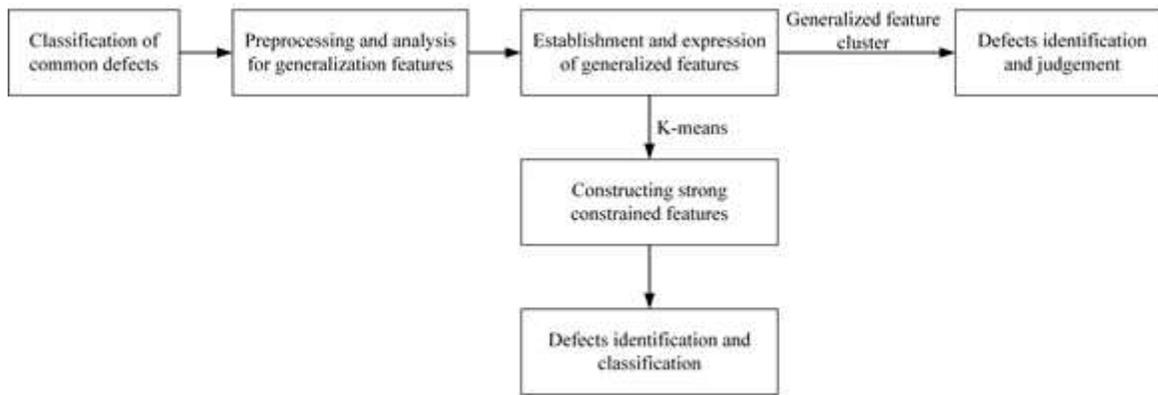


Figure 13

The flow chart of the proposed algorithm

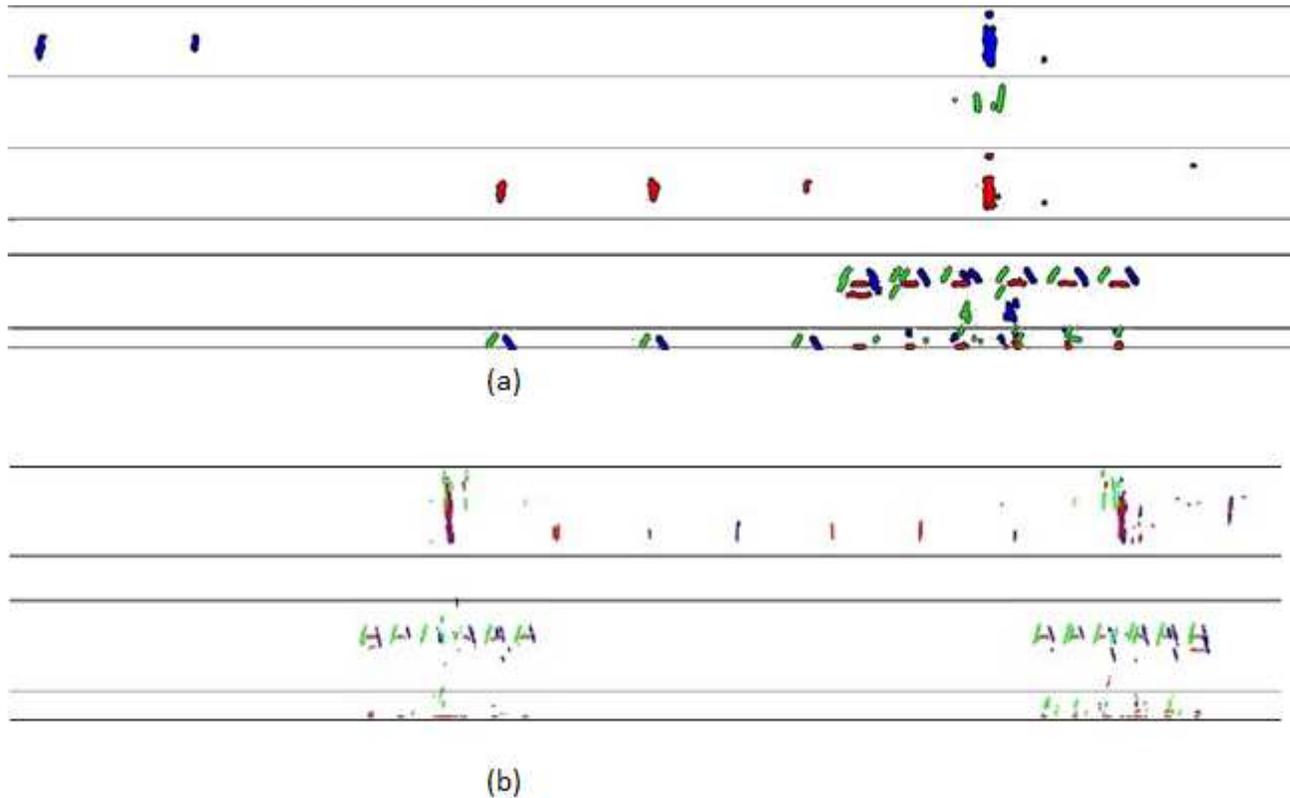


Figure 14

Type I and type II railway track ultrasound image (a) Type II track ultrasonic image (b) Type I track ultrasonic image

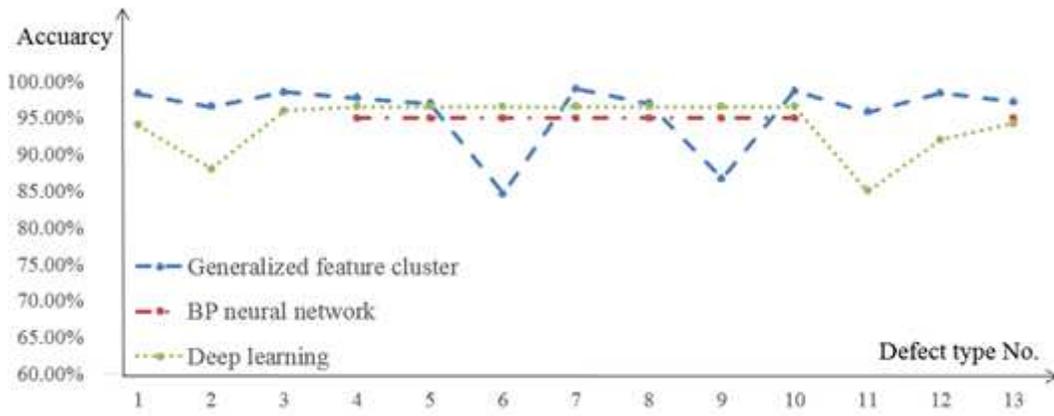


Figure 15

The comparison result of three methods

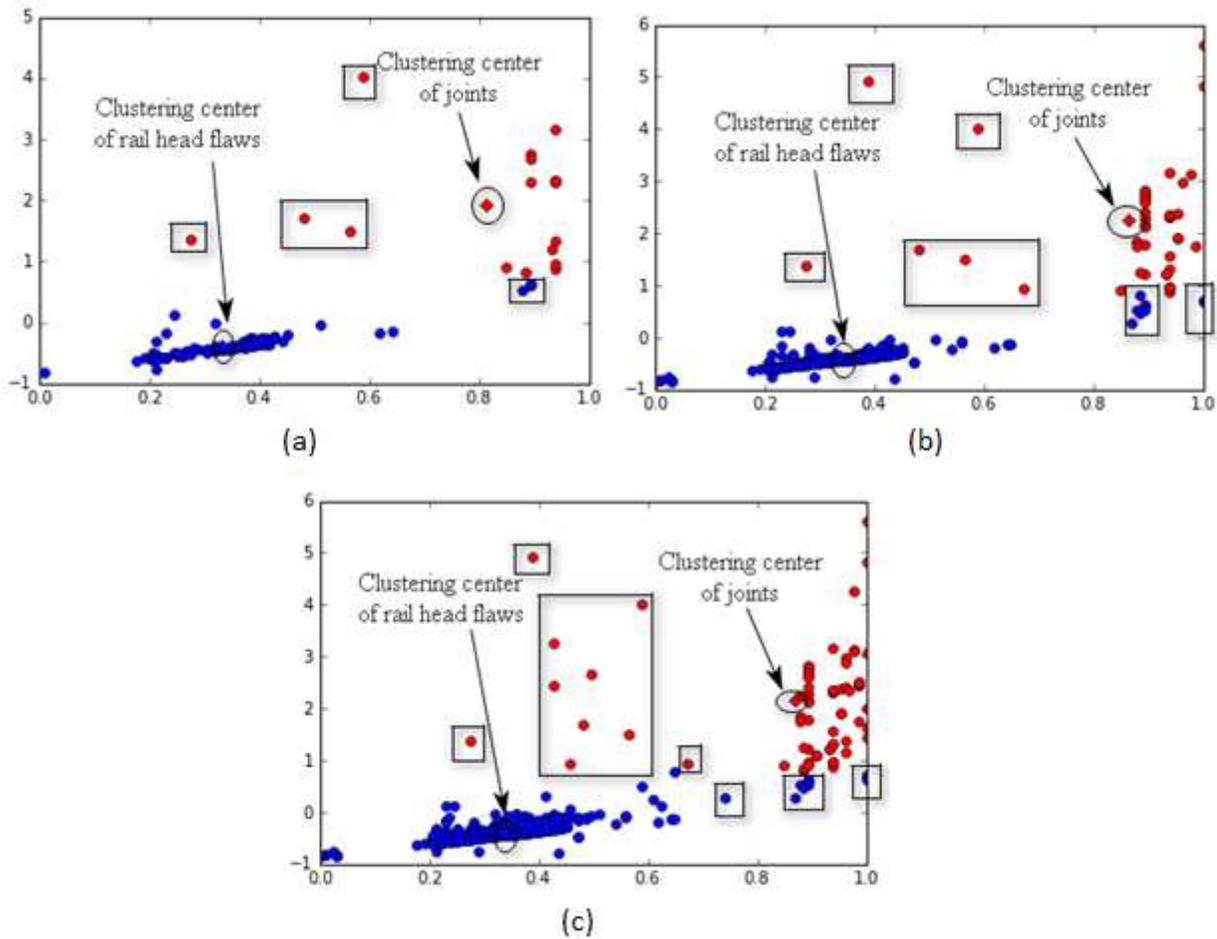


Figure 16

Clustering results of main defect in rail head region (a) Cluster results of 120 groups of data (b) Cluster results of 300 groups of data (c) Cluster results of 480 groups of data