Rail Surface Defect Recognition Method Based on AdaBoost Multi-classifier Combination

Biao Yue^{1,2,3}, Yangping Wang^{1,2,3}, Yongzhi Min^{3,4}, Zhenhai Zhang^{3,4,5}, Wenrun Wang¹, Jiu Yong¹ 1. Experimental Teaching Cener on Computer Science, Lanzhou Jiaotong University, Lanzhou, China

yuebiao@mail.lzjtu.cn Tel:13919898968

2. National Virtual Simulation Experimental Teaching Center for Rail Transit Information and Control, Lanzhou Jiaotong

University, Lanzhou, China

1328396793@qq.com Tel:13519311970

3. Gansu Provincial Engineering Research Center for Artificial Intelligence and Graphic & Image Processing, Lanzhou, China 104053339@qq.com Tel:13519689184

4. School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou, China

5. Lanzhou Bocai Technology Co., Ltd, Lanzhou, China

Abstract- Rail surface defects have the characteristics of various types and complex morphological characteristics. It is difficult to obtain accurate classification results only by using a single classification method. Therefore, this paper presents a rail surface defect recognition method based on AdaBoost multi-classifier combination. Firstly, defect attributes are described by extracting geometric shape and gray level features of defect area, and Relief algorithm is used to select defect features and filter out features unrelated to classification. Then by using AdaBoost multiclassifier combination method and taking CART decision tree as a weak classification algorithm to design a combined classifier, rail surface defects classification is realized. The results show that this method can effectively identify three common types of defects: rail surface peeling block, tread crack and fish scale peeling crack. Key Words: Rail surface defects, Feature extraction, AdaBoost multi-classifier combination, CART decision tree

I. INTRODUCTION

Rail is an important component of railway line, and its damage condition directly affects the safety of railway transportation[1,2]. With the high-speed, heavy-duty and high-density operation of trains, rail surface defects caused by rolling contact fatigue are becoming more and more common. Timely detection of rail surface defects and maintenance measures can prevent them from further developing into internal defects, thus ensuring the safe operation of the railway. Machine vision technology, with its advantages of fast speed, high precision and non-contact, provides a new development direction for non-destructive detection of rail surface defects[3-5].

In recent years, many scientific research institutes have carried out in-depth research on machine vision inspection of rail surface defects. TANG et al. [6] extracted rail surface defect area by logic operation combination detection method, and classify defects by BP neural network. Andamuthu et al. [7] used the method of matching the template image with the measured image in order to reduce the influence of illumination on the rail area, and then extract the defect area by threshold segmentation. Jin Yan et al. [8] identified various defects on rail surface by combining neural network and morphological method. Tang Bo et al. [9] extracted 20dimensional feature vectors such as gray level, texture and geometric shape of defect image, and designed a defect image recognition classifier based on fuzzy set theory. Xu Ke et al. [10] decomposed the surface image of steel plate into subbands of different scales and directions by Tetrolet transform, extracted the Tetrolet high-pass sparse matrix features of subbands, and then realized classification and recognition of surface defects by using support vector machine. However, due to the variety of rail surface defects and the complexity of their morphological features, it is difficult to obtain accurate classification results only by using a single classification method, and the multi-classifier combination technology which combines multiple complementary classifiers can comprehensively inspect and learn defect samples from different angles, and can significantly improve the accuracy of defect classification.

At present, researchers use many classifier combination methods, such as Boosting, Bagging and Random Subspace[11,12]. Among them, Boosting algorithm is the most representative. The AdaBoost algorithm, which emerged in the subsequent improvement process, has the advantages of strong adaptability and high accuracy. It is the most popular Boosting algorithm at present and has been widely used in the field of image recognition. Therefore, this paper realizes rail surface defect classification based on AdaBoost algorithm. Firstly, Relief algorithm is used to select useful defect features and filter out features unrelated to classification based on defect features calculation. Then by using AdaBoost multi-classifier combination method and taking CART decision tree as a weak classification algorithm to design a combined classifier, rail surface defects classification is realized.

II. INTRODUCTION OF DEFECT TYPE AND HIGH FREQUENCY DEFECT AREA

Rail damage refers to rail fracture, cracks or other conditions that affect the performance of rail and endanger traffic safety. Through statistical analysis, it is found that rail damage in service can be mainly divided into contact fatigue damage, quenched rail damage, and damage caused by metallurgical quality and surface defects. Because of the characteristics of visual inspection technology, this paper only focuses on the detection of surface defects. Common rail surface defects are shown in Figure 1, which mainly include rail fracture, rail Corrugation, rail head crushing, stripping block, tread crack and fish scale peeling crack[13]. The location and shape of the defects are different because of the different reasons for the defects. By analyzing the area, perimeter, rectangularity and centroid of the defects, the types of rail surface defects can be determined.



Fig. 1 Common rail surface defects

According to the location of the stress on the rail caused by the train wheels passing through the rail, and the statistics of the location, degree and type of the damage in the past, the surface defect is mainly concentrated between the middle line of the rail tread and the working side[13]. As shown in Figure 2, it is a schematic diagram of the contact patch on the rail surface. The shadow part is the area where the defect is concentrated, P is the vertical force of the wheel, H is the lateral force of the wheel, N is the contact normal force and T is the contact tangential force.



Fig. 2 Schematic diagram of rail surface contact patch

III. DEFECT FEATURE EXTRACTION

Due to the shape of rail surface defects is complex and random, in order to identify the types of defects efficiently and accurately, it is necessary to extract a large number of features that can accurately describe the attributes of defects. Before defect feature extraction, this paper first preprocesses the collected rail image and segmentes the defect image. Then geometric and gray features are selected to describe defect attributes. The calculation methods of each feature are shown in Table 1.

Tab. 1 Defect features

	Geometric features				Gray Level features			
	Name	Expression	Significance	Name	Expression	Significance		
1	Area	$A_{0} = \sum_{i=1}^{M} \sum_{j=1}^{N} f(x_{i}, y_{j})$	Size of defect area	Gray mean	$\overline{f} = \sum_{b=0}^{L-1} b P(b)$	Overall level of gray level in defect area		
2	Perimeter	The sum of pixels on the boundary of defect region	Contour feature of defect area	Gray variance	$\sigma^2 = \sum_{b=0}^{L-1} (b - \overline{f})^2 P(b)$	Discrete degree of gray distribution in defect Region		
3	Rectangular degree	$R = \frac{A_0}{A_{\rm R}}$	Similarity with rectangular region	Energy	$E = \sum_{b=0}^{L-1} (P(b))^2$	The magnitude of information in defect region		
4	Length- width ratio	$r = \frac{L}{W}$	Slenderness and lengthening degree of defect area	Entropy	$H = -\sum_{b=0}^{L-1} P(b) \log_2(P(b))$	The richness of information content in defect areas		
5	Circularity	$D = \frac{C^2}{A_0}$	Boundary complexity degree of defect area	Skewness coefficient	$f_{\rm S} = \frac{1}{\sigma^3} \sum_{b=0}^{L-1} (b - \bar{f})^3 P(b)$	The asymmetry degree of gray histogram between defects		
6	Centroid	$\begin{cases} x_0 = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} x_i \\ y_0 = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} y_i \end{cases}$	Relative position in rail surface	Peak state coefficient	$f_{\rm K} = \frac{1}{\sigma^4} \sum_{b=0}^{L-1} (b - \bar{f})^4 P(b) - 3$	Intensity of gray scale around its mean		

Notes: (x_i, y_i) is the pixel of defect areas in the image, $f(x_i, y_j)$ is its gray value, M and N are the number of rows and columns in the defect region, respectively. A_0 and C are the area and perimeter of defect area respectively. A_R is the area of its smallest outer rectangle. L and W are the length and width of the smallest outer rectangle, respectively. P(b) represents the first-order probability distribution of pixel gray level, $P(b) = P\{f(x_i, y_j) = b\}$, $0 \le b \le L - 1$. b is the quantization level of gray level, the total is L level. P(b) = N(b)/M represents the first-order histogram of image gray level, N(b) is the total number of pixels with gray level b, M is the total number of pixels in the image.

In theory, corresponding to the complex defect morphology of rail surface, the more the number of features used to describe the defect attributes, the more accurate and comprehensive the defect identification. However, too many features will increase the complexity of training and decision algorithms. Therefore, Relief algorithm [14,15] is used to select defect features. The steps are as follows:

(1) Initialization: Let the defect sample set be S, features number be n, the weight vector of defect feature set be W, and the initial weight of each feature is 0, the number of random sampling is m, and the nearest neighbor number is k.

(2) A sample *D* is randomly extracted from defect sample set S, and the *k*-nearest neighbor method is used to find the farthest neighbor sample *H* and the nearest neighbor sample *M* from the same and different defect samples, respectively.

(3) For arbitrary feature j in defect feature set, the difference degree dif(D, H, j) and dif(D, M, j) between sample H, sample M and initial selected sample D are calculated, respectively.

$$dif(D, H, j) = \frac{D(j) - H(j)}{\max(j) - \min(j)}, j = 1, 2, \dots, N$$
(1)

Among them, D(j) and H(j) represent the eigenvalues of feature j in sample D and sample H, respectively. $\max(j)$ represents the maximum eigenvalue of feature j in defect sample set, $\min(j)$ is minimum eigenvalue, The calculation method of dif(D, M, j) is the same as that of dif(D, H, j).

(4) For arbitrary feature j in defect feature set, if the difference between sample H and D is less than that between sample M and D, it shows that this feature makes the difference between the same kind smaller and the difference between different kinds larger, therefore the weight of the feature should be increased. The formula for calculating the weight W(j) of j is as follows:

$$W(j) = W(j) - \operatorname{dif}(D, H, j) + \operatorname{dif}(D, M, j)$$
(2)

(5) Repeat step (2) ~ (4), calculate the weight of each feature in the feature set, and get the weight vector W.

In each cycle, each feature in the feature set is given the corresponding weight, and the value range of each element in the weight vector is [-1,1]. The bigger the weight, the stronger the ability of feature classification. According to the size of the weight, the feature with negative weight is sorted, that is, the feature independent of classification is filtered out, and only the feature that is effective to the sample is retained.

IV. DESIGN OF COMBINED CLASSIFIER BASED ON ADABOOST ALGORITHM

A. AdaBoost algorithm

AdaBoost algorithm [16] can construct a high-precision strong classifier, the key is the adoption of adaptive weight technology. Its basic idea is to give a training set and a weak learning algorithm. Firstly, each training sample in the training set is given an equal initial weight.

Then the weak learning algorithm is called to train the training set in *T*-round. After each round of training, the weight of the samples is updated according to the training results, and the samples that are misclassified in the training process are given a larger weight, which is equivalent to giving more explanations and expressions to the samples that are difficult to classify in the subsequent training process. After *T*-round training, a predictive function sequence can be obtained. Each prediction effect. The better weight is bigger, and vice versa. Finally, weighted voting is used to form the final classification decision function. The steps of the AdaBoost algorithm are as follows:

Suppose the training set is $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, $x_i \in X$, $y_i \in Y$, $i = 1, 2, \dots, N$, X represents sample space, Y represents the set of sample identifiers.

(1) Initialization: For samples $(x_i, y_i) \in S$, setting initial weights $D_1(x_i, y_i) = 1/N$, and set the number of iteration training $t = 1, 2, \dots, T$.

(2) According to the sample weight D_t of round t, weak classifier $h_t: X \to Y$ is obtained by calling weak learning algorithm.

(3) The prediction error ε_t of classification rule h_t on S is:

$$\varepsilon_t = \sum_{\forall y_i \neq h_t(x_i)} D_t(x_i, y_i)$$
(3)

If $\varepsilon_t \ge 1/2$ or $\varepsilon_t = 0$, then order T = t - 1, end the cycle. The evaluation factor $\beta_t = \varepsilon_t/(1 - \varepsilon_t)$ of classification rule h_t is further obtained from the calculation result of prediction error ε_t .

(4) Updating sample weights:

$$D_{t+1}(x_i, y_i) = \frac{D_t(x_i, y_i)}{Z_t} \times \begin{cases} \beta_t & \text{if } h_t(x_i) = y_i \\ 1 & \text{other} \end{cases}$$
(4)

Among them, Z_t is a normalized constant, make $\sum_{(x_i, y_i) \in S} D_{t+1}(x_i, y_i) = 1.$

(5) Repeat step (2) ~ (4) T-1 times to get T weak classifiers. Finally, the weak classifiers generated by each iteration are combined by weighted voting, and the strong classifier function H(x) is obtained.

$$H(x) = \arg \max_{y \in Y} \sum_{t:h_t(x)=y}^T \log \frac{1}{\beta_t}$$
(5)

B. CART Decision Tree Algorithms

CART decision tree[17] is a supervised learning algorithm proposed by Breman in 1984. Its basic process can be described as: through cyclic analysis of training data sets composed of test variables and target variables, a binary tree decision tree structure is formed[18]. This method chooses one attribute or combines multiple attributes among all classification attributes as splitting variables of tree nodes, and divides test variables into each branches. The degree of heterogeneity decline is used to judge whether the splitting nodes are generated or not. The lower the heterogeneity, the less the impurity content of the nodes, the higher the classification accuracy. CART decision tree algorithm uses Gini coefficient to characterize the heterogeneity of split nodes. If the training set T has n samples and the sample size is N, the Gini coefficient can be calculated as follows:

$$Gini(T) = 1 - \sum_{j=1}^{n} P_j^2$$
(6)

Among them, P_j is the proportion of *j* samples class to the total number of samples.

The classification attribute in CART decision tree algorithm is the rail surface defect feature extracted in this paper. If the eigenvalue *a* of feature *A* is selected to divided training set *T* into T_1 and T_2 , the split Gini coefficient is as follows:

$$Gini(T) = \frac{N_1}{N}Gini(T_1) + \frac{N_2}{N}Gini(T_2)$$
(7)

Among them, N_1 and N_2 are the number of samples in subset T_1 and T_2 of training samples respectively. By calculating, when Gini coefficient Gini(T) is minimized, the corresponding feature is the best segmentation feature and value. Then, according to the best segmentation feature and value, the current node is divided into two sub-nodes, and the samples in the training set are allocated to the two sub-nodes. The two parts obtained from the above steps are processed recursively according to the above splitting process, and a feature is re-selected to continue to be partitioned until all the features in the selected feature space are partitioned, then a sufficiently large decision tree can be generated.

C. Design of Combination Classifier

From the analysis of AdaBoost multi-classifier combination algorithm, we can see that after each iteration, the algorithm generates a new weak classifier, and constantly focuses on the difficult samples. The resulting series of weak classifiers do not repeat each other. Each weak classifier is good at dealing with the unique part of the sample set, which can ensure the complementarity of the combined classifier in the sample processing ability. Therefore, when using AdaBoost algorithm to combine multiple classifiers, the weak classification algorithm is required to be unstable. The CART decision tree algorithm is sensitive to the change of training sample set and can satisfy the requirement of weak classification algorithm instability of AdaBoost algorithm. Therefore, this paper takes CART decision tree as the weak classification algorithm of AdaBoost multi-classifier combination algorithm, realizes the combination of AdaBoost algorithm and CART decision tree, and records the combination algorithm as AdaBoost-CART.

The process of AdaBoost-CART classifier combination algorithm is shown in Figure 3. The solid line part of the graph is the training stage, and the dotted line part is the testing stage. The input of training phase is training sample set S and training times T of AdaBoost algorithm. Firstly, sample set S_1 and sample set weight D_1 are initialized. Then CART decision tree is established and pruned, weak classification rule h_1 and corresponding weight are obtained. According to the evaluation factor β_1 of classification rule h_1 , the weights of samples are updated to get S_2 and D_2 . The final combination classification rule h_{fin} is obtained by weighted voting after the above process is iterated T times.

In the testing stage, the test sample x is input into the trained combined classifier, and the classification results $h_1(x)$, $h_2(x)$, ..., $h_T(x)$ of weak classification rule h_1 , h_2 , ..., h_T are obtained respectively. After weighted voting, the highest number of votes is the category H(x) of test sample x.



combination algorithm

V. EXPERIMENTAL ANALYSIS OF DEFECT CLASSIFICATION

In order to test the validity of the multi-classifier combination algorithm, this paper uses 1200 rail images

collected from the railway site to build the defect sample bank. Among them, there are 173 images containing fish scale peeling crack defects, 241 images containing stripping block defects, 186 images containing tread crack defects, and the remaining 600 images of rail without defects. Gray and binary images of some defects are shown in Fig. 4.



Fig. 4 Partial defect pattern

The defect samples are randomly divided into two parts as training set and test set. The training set is used to train the classifier and the test set is used to test the error rate of the classifier. A total of 12 feature attributes of geometric shape and gray level of defects extracted from each rail image were tested.

A. Parameter analysis

Relief feature extraction algorithm has two important parameters, namely, the number of random sampling *m* and the number of adjacent samples k. Among them, parameter m is the number of samples randomly extracted from defect sample set in order to obtain stable results, and parameter k is used to characterize the distance between calculated samples. Empirical values m = 50 and k = 10 of parameters m and k are selected respectively. The parameter T in AdaBoost multiclassifier combination algorithm represents the number of training iterations. In theory, with the increase of training times T, the subsequent weak classifier creation process will continue to focus on samples with difficulty in classification, and the ultimate error recognition rate of the combined classifier will be lower. However, the creation of each weak classifier is time-sequential. As the number of training Tincreases, the complexity of classifier training increases.

In order to balance the error recognition rate and the training complexity of the multi-classifier combination algorithm, the iteration times T = 2, 10, 20, ..., 90, 100 are tested to calculate the error rate corresponding to different iterations and generate the error rate curve of multi-classifier combination algorithm as shown in Fig. 5. As can be seen from Fig. 5, with the increase of iteration times T, the error recognition rate of the classifier combination algorithm decreases gradually and to a greater extent when T is less than 30, while the change of the error recognition rate with iteration times T is greater than 30. Therefore, the number of iterations T is chosen as 30.



Fig. 5 The curve of error recognition rate with iteration times T

B. Recognition Results of Combination Classifier

After feature selection using Relief algorithm, the ranking order of feature weights in feature set from large to small is obtained, and the combined classifier is established by using 11 defect features with larger weights. Table 2 shows the recognition results of AdaBoost-CART combined classifier for fish scale peeling crack, stripping block and tread crack. The average recognition rate for different types of defects is 81.12%. Comparing the test results of BP^[6], SVM^[10], CART decision tree and AdaBoost-CART combined classifier, as shown in Table 3, we can see that the recognition rate of this method is higher than that of the other three methods. This is mainly because the multi-classifier combination algorithm will focus on the difficult samples in the training process. Compared with using CART decision tree alone, the recognition rate of AdaBoost-CART combined classifier is improved by 9.03%. Although the training time is increased by three times, the training process of classifier only needs to be executed once in practical application. The recognition process of test samples using the classifier completed by training is shorter, so the combined classifier has greater advantages.

Tab. 2 Statistics of classifier recognition results

Defect type	Total quantity	Error recogniti- on quantity	Recogniti -on quantity	Error recogniti- on rate (%)	Recogni- tion rate (%)
Scale peeling crack	248	16	202	6.45	81.45
Stripping block	275	20	219	7.27	79.64
Tread crack	203	14	167	6.89	82.27

Tab. 3 Test results comparison of different classification algorithms

	BP	SVM	CART	AdaBoost- CART
Recognition rate (%)	77.26	78.37	72.09	81.12
Error recognition rate (%)	7.15	6.92	7.43	6.87
Training time (s)	377	391	105	428

VI. CONCLUSION

In this paper, firstly, aiming at the complex and random shape of the defect region, the geometric and gray features of the defect region are extracted to describe the defect attributes. Then, aiming at the problem of information redundancy and feature duplication caused by the correlation between some features, Relief algorithm is used to select defect features and filter out features that are not related to classification. Finally, aiming at the problem of limited classification accuracy of single classifier, the method of multi-classifier combination is adopted, and the CART decision tree is used as the weak classification algorithm of AdaBoost multi-classifier combination algorithm to design the combined classifier, which realizes the classification and recognition of three common defects: scale peeling crack, stripping block and tread crack. The average recognition rate is 81.12%.

ACKNOWLEDGMENT

The authors would like to thank the National Natural Science Fund (No.61663022), the Program of Innovation team development plan of Ministry of education (IRT_16R36), Lanzhou Jiaotong University experimental teaching reform research project (2019007 and 2019026), Lanzhou Talent Innovation and Entrepreneurship Project 2015-RC-86, Lanzhou Jiaotong University innovation and entrepreneurship course construction project in 2019 (2019CXCYKC10) for financial support.

REFERENCES

- [1]. Molodova M,Li Z,Nunez A,et al.Automatic detection of squats in railway infrastructure [J].IEEE Transactions on Intelligent Transportation Systems,2014,15(5):1980-1990
- [2]. Gibert X,Patel V M,Chellappa R.Deep multitask learning for railway track inspection [J].IEEE Transactions on Intelligent

Transportation Systems, 2017, 18(1):153-164.

- [3]. Resendiz E,Hart J,Ahuja N.Automated visual inspection of railroad tracks [J].IEEE Transactions on Intelligent Transportation Systems, 2013, 14(2):751-760.
- [4]. Feng H,Jiang Z,Xie F,et al.Automatic fastener classification and defect detection in vision-based railway inspection systems[J].IEEE Transactions on Instrumentation and Measurement,2014,63(4): 877-888.
- [5]. Li Q,Ren S.A real-time visual inspection system for discrete surface defects of rail heads [J].IEEE Transactions on Instrumentation and Measurement,2012, 61(8): 2189-2199.
- [6]. TANG Xiangna, WANG Yaonan. Visual Inspection and Classification Algorithm of Rail Surface Defect[J]. Computer Engineering, 2013, 39(3): 25-30.
- [7]. A. Andamuthu, Vinothkumar K. A Robust Rail Surface Defect Analysis And Monitoring System[J]. Asian Journal of Information Science and Technology, 2013, 1(2): 15-21.
- [8]. JIN Yan, YANG Chang-hui, ZHANG Jian-xun. Defect Recognition of Steel Surface Based on Neural Networks and Morphology[J]. MACHINE TOOL & Comp. HYDRAULICS, 2010, 38(21): 26-28.
- [9]. TANG Bo, KONG Jianyi, WANG Xing-dong HOU Yu, CHEN Li. Recognition and classification for steel strip surface defect images based on rough set theory[J]. Journal of Image and Graphics, 2011, 16(7): 1213-1218.
- [10]. XU Ke,WANG Lei,WANG Jingyu. Surface Defect Recognition of Hot-rolled Steel Plates Based on Tetrolet Transform[J]. JOURNAL OF MECHANICAL ENGINEERING, 2016, 52(4):13-19.
- [11]. Zhang P B , Yang Z X . A Novel AdaBoost Framework With Robust Threshold and Structural Optimization[J]. IEEE Transactions on Cybernetics, 2018, 48(1):64-76.
- [12]. Fu Q, Jing B, He P, et al. Fault Feature Selection and Diagnosis of Rolling Bearings Based on EEMD and Optimized Elman_AdaBoost Algorithm[J]. IEEE Sensors Journal, 2018, 18(12):5024-5034.
- [13]. ZOU Dingqiang. Rail Failure Analysis and Damage Map[M]. BeiJing: China Railway Press,2011.
- [14]. Deng Z, Chung F L, Wang S. Robust Relief-Feature Weighting, Margin Maximization, and Fuzzy Optimization[J]. IEEE Transactions on Fuzzy Systems, 2010, 18(4):726-744.
- [15]. LI X L. The study and application of feature selection algorithms based on relief[D]. Dalian: Dalian University of Technology, 2013:51-58.
- [16]. Wu S, Nagahashi H. Parameterized AdaBoost: Introducing a Parameter to Speed Up the Training of Real AdaBoost[J]. IEEE Signal Processing Letters, 2014, 21(6):687-691.
- [17]. LV L, XIE Y W, HUANG X J, et al. Desertification Information Extraction Method Research based on the CART Decision Tree Classification[J]. Remote Sensing Technology and Application, 2017, 32(3):499-506.
- [18]. Jaworski M, Duda P, Rutkowski L. New Splitting Criteria for Decision Trees in Stationary Data Streams[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(6):2516-2529.