Railway fastener image recognition method based on multi feature fusion

Ruxun Xu\textsuperscript{1,2,3,a}, Jinyue Han\textsuperscript{1,b}, Wenzhe Qi\textsuperscript{1,e}, Jianjun Meng\textsuperscript{1,2,3,d}, Hongqiang Zhang\textsuperscript{2,e}

\textsuperscript{1}Mechatronics T&R Institute, Lanzhou Jiaotong University, Lanzhou 730070, China; \textsuperscript{2}School of Mechatronic Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China; \textsuperscript{3}Engineering Technology Center for Informatization of Logistics and Transport Equipment, Lanzhou 730070, China; Email: \textsuperscript{a}xrx@mail.lzjtu.cn; \textsuperscript{b}hjygz123@126.com; \textsuperscript{c}qiwz@mail.lzjtu.cn; \textsuperscript{d}mengjj@mail.lzjtu.cn; \textsuperscript{e}1345014962@qq.com.

Abstract. In order to improve the accuracy of railway fastener image recognition and training efficiency in the process of machine learning. Two traditional feature extraction methods, MB-LBP features and PHOG features are fused to make up a new image features for training in this paper. At the same time, to solve the problem of low training speed caused by increased feature dimension, introducing Adaboost-SVM classifier, the classification efficiency of Adaboost classifier will be reduced after a certain set of conditions, then the remaining samples are sent to the SVM classifier for classification. The results show that the accuracy of the fusion image feature is 3.9% and 5.8% higher than that of the individual MB-LBP and PHOG, respectively, and the training speed is also significantly improved by using the Adaboost-SVM classifier.

1. Introduction
Fastenels play an important role in the safe driving of trains. According to the research, train derailment is mainly caused by the lack of rail fastener[1]. At present, with the growth of railroad traffic mileage rapidly, it is urgent to find an advanced alternative way. After entering the information age, the machine vision technology is advancing rapidly. Many research institutions both at home and abroad are committed to using machine vision to judge the state of railway fasteners, and have achieved certain results [2-3]. But in general, its recognition algorithm is still in the exploratory stage, and has some shortcomings, such as low recognition rate and low efficiency. For example, Documents[4]and[5],using neural network and the recognition algorithm of depth learning to detect the fastener, the detection object is aimed at a wide range of hexagonal fasteners used abroad, not suitable for the common type of fastener in our country. The nearest neighbor classifier based on Hamming distance is used for detection [6], but the accuracy is not increased.

This paper proposes a method of image recognition based on multi feature fusion in feature extraction stage. The regional projection method is used to locate the sleeper and railway fastener, and the MB-LBP (Multi-block Local Binary Pattern), the PHOG feature and the combination are extracted as the classification feature of the support vector machine (Support Vector Machine, SVM) classifier. Because the training speed of support vector machine is mainly related to the number of training samples and the selection of kernel functions. Therefore, in order to ensure the speed of the combination feature in training, the Adaboost-SVM hybrid classifier is introduced to improve the
training efficiency.

2. Location and Feature Extraction of Fasteners

Before fastener location and recognition, we need to preprocess the image first, so that we can get the image with obvious edges and a certain contrast. And the fastener is located under the complex background map.

2.1. Fastener Positioning Method

2.1.1. Gray scale projection integral method. The gray projection integration method is used to locate the image steps as follows[7].

   Step 1. Transforms the image into a gray image.

   Step 2. After the edge detection is done, the images are projected on the X axis Y axis (the location of the bigger luminance is the peak part).

   Step 3. We know the distance between fastener position and rail and sleeper intersection point through a priori data, and move to a certain distance to locate the center point of fastener.

   Figure 1. Projection location map of sleeper and rail

   In Figure 1, we can see that the peak coordinates of the vertical projection correspond to the trajectory portion of the contour. The horizontal projection of the left and right peaks corresponds to the edge of the sleeper. By extracting the intersection point between the vertical and horizontal coordinates, the coordinates of the image locating fastener are translated according to the relative position between the fastener and the track.

2.2. Method of Feature Extraction

The main feature of the fastener image is the symmetry of the main part of the fastener. At the same time, because the recognition of the fastener image is close to the face recognition process, two kinds of feature extraction methods widely used in the process of image processing are selected, and two improved algorithms, MB-LBP square method and PHOG, are selected in this paper.

LBP feature extraction method is widely applied in face recognition and image processing field. The main reason is that it has the following advantages:

(1) The process is simple. The method of using LBP only needs simple binary processing of the image and produces the corresponding vector and histogram. The processing steps are less, and the recognition rate is improved by a simple method.

(2) Excellent feature classification performance. LBP feature extraction method can highlight the texture features of the image in the process of processing, including the pixel points of the light and dark, the contour information of the image and its distribution in the graph. In addition, the LBP method has been improved since its birth, with the characteristics of rotation and gray scale invariance, which can effectively reduce the adverse effects caused by the image rotation and light illumination, thus improving the recognition rate of the algorithm.

2.2.1. Mb-lbp Method. In the process of extracting the contours of the fastener, the noise is easily
affected, and the multi block local two element model (MB-LBP) is an extension of the traditional LBP, which eliminates the deficiency of the traditional LBP mode, which is easily affected by the noise[8-9].

![Figure 2. Characteristic value of MB-LBP](image)

From Figure 2, we can see that the MB-LBP characteristic value of word block finally has a binary arrangement of the MB-LBP characteristic number from the upper left corner, the value is: 11000110=198. Expression is shown in formula (1).

\[
MB - LBP = \sum_{i=1}^{8} S(g_i - g_c)2^{i-1}
\]

(1)

In Formula 1, gC represents the average gray value of the central block. GI represents the average gray value of the surrounding 8 neighbourhoods. The formula of S (X) calculation method (2)

\[
S(X) = \begin{cases} 
1, & x \geq 0 \\
0, & \text{other}
\end{cases}
\]

(2)

The LBP feature extraction results are shown in Figure 3. From the graph, we can see the LBP feature extraction method, highlighting the details of the image.

2.2.2. Phog Method. The PHOG(pyramid histogram of oriented gradient) method is an optimization of the traditional HOG feature extraction method. Its principle is to extract the image features under a certain size of the image, and combine the extracted features in series to get the global feature [10]. As a feature extraction method with strong anti noise and certain anti rotation performance, PHOG has achieved good results in many image recognition algorithms for railway fasteners, [11-13].

Generally speaking, the image size of our pyramid histogram is fixed. But if we want to calculate the HOG feature of this picture, We can use different scale and hierarchical computing method to collect and stitching image features.
When the PHOG feature extraction is used, the classification of the number of images will affect the accuracy of recognition. When the image feature is divided into 3 layers, the better recognition effect can be obtained. Therefore, the selection parameters are: the number of layers \( L = 3 \), the region bins = 8, \( L = 3 \).

The zeroth level calculates the HOG characteristics of the whole graph, and the zero layer has \( 1 \times 8 \) dimensions.

The first level is divided into \( 2 \times 2 \) regions. The first layer has \( 2 \times 2 \times 8 \) dimensions.

The second level is divided into \( 4 \times 4 \) regions, and the second level regional features have \( 4 \times 4 \times 8 \) dimensions.

The third level is divided into \( 8 \times 8 \) regions, and the third level regional features have \( 8 \times 8 \times 8 \) dimensions.

The Phog feature \((1+4+16+64)\times 8\) at this time is 680 dimensions. Figure 5 is the histogram of fastener images after PHOG feature extraction. The eigenvalues are represented on different dimensions.

### 2.3. Feature Fusion

#### 2.3.1. Feature normalization and merger

When the features extracted from different ways are merged, if the features are not quantized to the unified interval, the data analysis results are often affected by the disunity of the data dimension.

![Figure 5. Pre normalized and post gradient descent](image)
For example, there are two variables to be optimized for X and Y. In Figure 6, \( \theta_1 \) and \( \theta_2 \) represent the coefficients of these two variables, respectively.

Z represents a function expressed by X and Y. A model that can be created (3).

\[
Z = (\theta_1X + \theta_2Y)
\] (3)

When the value range between X and Y is quite different (for example, the X range is 1-10 and the Y range is 400-500), the loss function may be the following form.

\[
J = (5\theta_1 + 400\theta_2 - Z_{correct})^2
\] (4)

At this time, the contour line of the loss function image is elliptic, and the process of finding the optimal solution is shown in Fig. 4 before normalization.

After normalization, the expression of loss function may be:

\[
J = (0.6\theta_1 + 0.65\theta_2 - Z_{correct})^2
\] (5)

From the above, we can see that after data normalization, it is beneficial to find optimal solution in a more direct way, thus improving the possibility of converging to optimal solution.

It is noted that the difference between the MB-LBP feature and the PHOG feature is huge. If the sample is trained directly, the loss function will be elliptical. In the process of finding the optimal solution of the training function, the gradient descent process of the loss function will produce a certain deviation, thus wasting a lot of time. Normalization of data characteristics and merging training can reduce training time effectively. In this paper, normalization of minimum and maximum values is used to normalize features.

The formula for quantifying characteristics is.

\[
Z = \frac{X_i - \min(X)}{\max(X) - \min(X)}
\] (6)

In the formula (6), Z represents normalized features, and Xi represents normalized features. Min (Xi) and max (Xi) represent the smallest and the largest values of all features, respectively.

The extracted features are re quantized, and the characteristic distribution is quantized in the range of [0,1]. Merge the normalized characteristics.

\[
CM = CP + CL
\] (7)

In formula 7, CP is the normalized PHOG eigenvalue and MB-LBP eigenvalue. CM is the eigenvalue after fusion. The fusion process is to splice two eigenvalues.

2.3.2. Adaboost-SVM classifier: In the process of classifying the merged features, the choice of classifier will have a greater impact on training speed and classification accuracy. The Adaboost classifier gives weights to the positive and negative samples. During the classification process, weak classifiers are used to make preliminary judgments on the samples, and the weights of different samples are changed according to the comparison between the judgment results and the actual results. Finally, the classification expression of strong classifier is determined. This method can achieve a
certain degree of classification accuracy by cascading the weak classifiers. However, as the number of layers increases, the remaining features of the samples to be trained have certain similarities.

The Adaboost-SVM hybrid classifier uses the approximate structure of the Adaboost classifier, as shown in Figure 7, after setting the overall parameters. If the input feature vector is determined to be a negative sample when passing through the first-level cascade structure, this feature vector will be removed from the training sequence and will not enter the next-level classification.

Therefore, the total number of samples classified by the classifier can be rapidly reduced. When the number of weak classifiers generated exceeds the preset value and the overall false positive rate is not less than the preset value, the remaining samples can be transferred to the SVM classifier for training. The following are the training steps for the Adaboost-SVM classifier.

Step 1. Select the detection rate $d$, And false alarm rate $f$, Set the overall false alarm rate $F$, The maximum number of weak classifiers $n_{\text{max}}$.

Step 2. Let $F_i = 1, i = 0$.

Step 3. Select the positive and negative samples for training the classifier.

Step 4. While ($F_i > F_i$)

\[
i = i + 1,
\]

(The number of weak classifiers), $F_i = F_{i+1}$.

While ($F_i > f \times F_i$) and ($n_i < n_{\text{max}}$);

\[
n_i = 0, n_{i+1} = n_{i+1},
\]

Use the adaboost algorithm to invoke positive and negative samples to train a classifier with features.

Step 5. Update the overall false positive rate $F_i$ of the strong classifier.

When the number of weak classifiers $n_t > n_{\text{max}}$ the overall false positive rate $F_i > F_t$, SVM classifier is used for classification.

3. Tests and Results

3.1. Positioning Test

In this paper, 1716 images in the sample database are located. The accurate positioning is 1683, the accuracy rate is 98.1%, and the total time consumption is 46.6s. This method mainly uses sleepers and rails for positioning. When the angle between the shooting device and the sleeper changes, it may cause mispositioning.

![Left fastener](image1)
![The original image](image2)
![Right fastener](image3)

Figure 7. Fastener location diagram

3.2. Feature Accuracy

A total of 1,000 positive samples and 716 negative samples were selected for testing and training. Use the resizer tool to uniformly modify the training sample to a size of 765×433. The positive sample is a good sample of railway fasteners. The test environment is inter corei5700 2GHZ, 1G memory, Windows7, matlab2010b PC.

Set the lowest classification accuracy $d=0.95$ and false alarm rate $f=0.35$ for each layer, and the maximum number of weak classifiers $n_{\text{max}}=255$. 

6
Table 1. Comparison of single feature and fusion feature

| feature    | SVM classifier | Adaboost-SVM classifier |
|------------|---------------|-------------------------|
|            | Test accuracy% | Training time (s)       | Test accuracy % | Training time (s) |
| PHOG       | 91.1          | 1883                    | 90.3           | 124               |
| MB-LBP     | 89.6          | 1517                    | 88.9           | 103               |
| Fusion     | 95.4          | 2422                    | 94.2           | 236               |

Table 2. Comparison of fusion feature

| method                          | Fusion features | PCA-HOG | Bag of words | Nearest neighbor |
|---------------------------------|-----------------|---------|--------------|------------------|
| Accuracy (%)                    | 94.2            | 81.9    | 83           | 89.1             |

Table 1 compares the single feature and the fusion feature. Experiments have tested the performance of the fusion feature and the single feature when using different classifiers. The results show that:

(1) Table 1 contrasts from top to bottom, that is, the test accuracy of the fusion feature is generally higher than that of the single feature when the same classifier is used. When the SVM classifier is used, the accuracy of the fusion feature reaches 95.4%. It shows that the accuracy of applying fusion features has been improved.

(2) Compared with the time of the left column and the right column of Table 1, the training time when using the Adaboost-SVM classifier is generally lower than the single use of the SVM classifier.

Table 2 compares the accuracy of fusion features with other feature extraction methods.

3.3. Concluding Remarks

(1) Aiming at the recognition of fastener images, this paper mainly improves the location of fasteners, the method of feature extraction and the selection of classifiers.

(2) The gray-scale projection integration method is used for positioning, which is easy to operate, with high positioning accuracy and fast operation speed.

(3) A feature extraction method based on multi-feature fusion is proposed. It combines the advantages of two practical feature extraction methods and more accurately describes the features of the image. The test shows that the accuracy of using fusion image features is 3.9% and 5.8% higher than the MB-LBP and PHOG features, respectively.

(4) After using Adaboost-SVM hybrid classifier, the actual classification of Adaboost followed by SVM classifier was used for final classification. This can increase the speed of training for high-dimensional multi-sample pictures.

References

[1] Islam S, Khan R A. Clustering and detection of good and bad rail line anchors from images[C]. 2015 18th International Conference on Computer and Information Technology (ICCIT),IEEE Conference Publications.2015, 222-226.

[2] Ruvo P D; Distante A; Stella E; Marino F. A GPU-based vision system for real time detection of fastening elements in railway inspection[C]. 2009 16th IEEE International Conference on Image Processing (ICIP). IEEE Conference Publications.2009, 2333 – 2336.

[3] Marino F, Distante A, Mazzeo P L, Stella E. A Real-Time Visual Inspection System for Railway Maintenance: Automatic Hexagonal-Headed Bolts Detection. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews).2007,37(3).418-428.

[4] Santu Y, Karakose M, Akin E. Random forest based diagnosis approach for rail fault inspection in railways[C]. 2016 National Conference on Electrical, Electronics and Biomedical Engineering (ELECO), IEEE Conference Publications.2016, 745 – 750.

[5] 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.

[6] Ma S, Research on Railway fastener recognition [D],Beijing JiaoTong University.2013.

[7] Zhang W Z, Chen Q, Du D, et. Gray projection integral method for line detection [J]. Journal of Tsinghua University (NATURAL SCIENCE EDITION)2005, 45(11), 1446-1449.
[8] Liao S C, Zhu X X, Lei Z, Zhang Lun, et. Learning Multi-scale Block Local Binary Patterns for Face Recognition[M]. http://www.cbsr.ia.ac.cn, 2007.

[9] Pan Q P, Yang W K, Sun C Y. License plate detection algorithm based on Haar and MB-LBP features. Journal of Southeast University (NATURAL SCIENCE EDITION). 2012, 42(S1), 74-77.

[10] Bosch A, Zisserman A, and MUNOZ X. Representing shape with a spatial pyramid kernel[C]. Proceedings of the ACM International Conference on Image and Video Retrieval, 2007.

[11] Yuan J, Wei B G, Wang L D. An Image Retrieval Method Synthesizing PHOGShape Description and Wavelet Pyramid Energy Distribution. Journal of Electronics. 2011, 39(9), 2114-2119.

[12] Sugiharto Aris; Harjoko Agus. Traffic sign detection based on HOG and PHOG using binary SVM and k-NN[C]. 2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE). 2016, 317-321.

[13] Wang H; Gao J; Tong L; Yu L. Facial expression recognition based on PHOG feature and sparse representation[C]. 2016 35th Chinese Control Conference (CCC). 2016. 3869 – 3874.