Optical Engineering

OpticalEngineering.SPIEDigitalLibrary.org

Railway equipment detection using exact height function shape descriptor based on fast adaptive Markov random field

Guodong Sun
Yang Zhang
Hanbing Tang
Huiming Zhang
Moyun Liu
Daxing Zhao

SPIE.

Guodong Sun, Yang Zhang, Hanbing Tang, Huiming Zhang, Moyun Liu, Daxing Zhao, "Railway equipment detection using exact height function shape descriptor based on fast adaptive Markov random field," *Opt. Eng.* **57**(5), 053114 (2018), doi: 10.1117/1.OE.57.5.053114.

Railway equipment detection using exact height function shape descriptor based on fast adaptive Markov random field

Guodong Sun,[a,*] Yang Zhang,[a,b] Hanbing Tang,[a] Huiming Zhang,[b] Moyun Liu,[a] and Daxing Zhao[a]
[a]Hubei University of Technology, School of Mechanical Engineering, Wuhan, China
[b]Nanjing University, Department of Computer Science, Nanjing, China

Abstract. This paper proposes a hierarchical feature-matching model for the typical faults detection, which is a big challenge in the trouble of a moving freight car detection system (TFDS) due to the constant color and complex background of images. The proposed model divides fault detection into two stages: image segmentation and parallel shape matching. In the process of segmentation, a fast adaptive Markov random field (FAMRF) algorithm is presented based on the image pyramid model and affinity propagation theory. In the process of shape matching, a shape descriptor named exact height function (EHF) is introduced on the basis of parallel dynamic programming. The experimental results indicate that the proposed hierarchical model combined with FAMRF and EHF can achieve automatic detection of an air brake system, bogie block key, and fastening bolt. The proposed model achieves high detection accuracy and great robustness, and it can be effectively applied to the fault detection in TFDS. [DOI: 10.1117/1.OE.57.5.053114]

Keywords: railway equipment detection; fast adaptive Markov random field; exact height function descriptor; parallel shape matching; affinity propagation.

Paper 180359 received Mar. 8, 2018; accepted for publication May 10, 2018; published online May 25, 2018.

1 Introduction

Fault inspection for railway equipment is a serious routine maintenance task to ensure the security of freight trains. For a long time, it has been performed by skilled workers, which is time-consuming and inefficient. With the rapid development of both image capturing devices and computing methods, we continue to focus on the image processing and machine learning algorithms that can be implemented more reliably and efficiently than manual detection.

Liu et al.[1] proposed a hierarchical inspection framework to detect the missing bogie block key (BBK) on freight trains under the complex environment. A cascaded detector trained by the AdaBoost is used to classify the fault region with a gradient-coded co-occurrence matrix (GCCM) and linear support vector machine (SVM). The proposed framework can inspect the missing BBK with high speed and accuracy. Cao et al.[2] presented the weighted margin sparse-embedded (WMSE) classifier for brake cylinder detection. The WMSE classifier takes advantages of VC-dimension minimization and weighted margin learning. The experiments show that the WMSE approach can obtain a much greater detection performance. To inspect typical trouble of a moving freight car detection system (TFDS) faults simultaneously, Sun et al.[3] proposed an automatic fault recognition system (AFRS) based on two convolutional neural network (CNN) models. The proposed system has outstanding performance on fault inspection of freight trains.

However, machine learning algorithms such as GCCM and WMSE need to build a sample database, which have direct impact on the inspection results. Moreover, with many hyperparameters needing to be optimized, the design of a good deep learning network architecture is complicated. The deep learning methods tend to overfit when training samples are insufficient. Complex networks lead to a sharp increase in training and recognition time. There are some good deep CNN models (VGGNet, GoogLeNet, ResNet, etc.) needing to be trained for several days or even weeks. The requirements of hardware (e.g., Titan X GPU) will increase at the same time.

Kim et al.[4] presented an automated inspection system for rolling stock brake shoes (BS). The distances between BS and wheels are measured by analysis module. The proposed system can measure all specifications of BS with high accuracy. Li et al.[5] described an automatic fault recognition method with two phases (localization and recognition) for BS key missing based on shape descriptors. This method is insensitive to noise and has good performance for complex background images. Xu[6] introduced an image recognition algorithm for BBK based on shape context (SC) and Gu[7] combined PCA with a SC shape descriptor to achieve high accuracy of fault inspection. In addition, some inspection algorithms are widely applied to inspect other railway equipment, such as dust collector,[8,9] BS wear,[9,10] fastening bolts,[9,11] locking plate,[12] angle cocks,[13] coupler yoke,[14] fastener,[15] and so on.

Due to the diversity and complexity of TFDS faults, an inspection algorithm for one type of fault does not suit another. Meanwhile, many inspection methods mentioned above analyze local features of fault regions and construct the classifiers to match target objects. Hence, these classifiers are insensitive to the size of targets. Global features obtained

*Address all correspondence to: Guodong Sun, E-mail: sgdeagle@163.com

0091-3286/2018/$25.00 © 2018 SPIE

by image segmentation should be considered to solve these problems.

Among the existing segmentation methods, the approaches with Markov random field (MRF) account for a large proportion. MRF-based methods can naturally represent the hierarchical structure, labeling accuracy, and spatial coherence within a Bayesian framework.[16,17] Dong et al.[18] proposed a label inference approach for segmenting natural images into perceptual regions using the MRF model. Chen et al.[19] put forward the multispectral image texture segmentation method using the multiresolution fuzzy MRF model for a variable scale in the wavelet domain. Therefore, a multiresolution MRF model based on a clustering method can be considered for TFDS detection.

Shape matching is also of great value in evaluating and comparing various algorithms. Contour-based methods extract contour features to describe objects such as SC,[5,20] inner distance shape context,[21] and height functions (HF).[22] Taking a sample point as the reference point, the HF is developed based on the distance between other sample points and the tangent line of the reference point. HF is insensitive to noise and occlusion, and also invariant to geometric transformations such as translation, rotation, and scaling.

In this paper, a multiresolution MRF model named fast adaptive MRF (FAMRF) is proposed to construct a hierarchical label space. Pixel position features of images are considered in the FAMRF. Multiscalar information in the wavelet domain is applied to improve the accuracy and efficiency. The histogram smoothing and affinity propagation based on a modified k-means algorithm are introduced to assign the number of layers and to achieve the adaptive image segmentation. A termination criterion of energy function is designed to modify the MRF model to improve the calculation speed. Subsequently, TFDS images are inspected based on the shape descriptor named exact height function (EHF). To achieve fast and accurate detection, the parallel matching framework is used to identify whether there is a fault in each label space.

The rest of this paper is organized as follows: Sec. 2 introduces the FAMRF hierarchical model. Section 3 focuses on the problem of parallel shape matching. The process of railway equipment detection is summarized in Sec. 4. The experimental results and analysis are showed in Sec. 5. This paper finally presents the conclusion and future work in Sec. 6.

2 Fast Adaptive Markov Random Field Model

As shown in Fig. 1, it is a challenge to perform such fault inspection because of the great change of illumination in an outdoor environment. A hierarchical model named FAMRF is designed to find an efficient solution for hierarchical segmentation. The FAMRF includes two issues: basic model construction and optimization. The basic model is constructed with wavelet-domain MRF (WMRF), histogram smoothing, and affinity propagation (AP)[23] theory. Figure 2 shows the diagram of the FAMRF hierarchy model.

2.1 *Wavelet-Domain Markov Random Field Model*

A majority of pixels in images are affected by the neighborhood pixels. In the MRF model, the prior information above is used to build the probability distribution of an interactive label to represent the spatial interaction of each pixel. Assume that digital image I has Np bands on the $N \times N$ grid S. The original image is decomposed into $l-1$ layers by wavelet transform, and the decomposed image is denoted as W. Each wavelet scale is expressed as n $(0 < n < l, n \in N_+)$. Each band of the W is decomposed by Haar

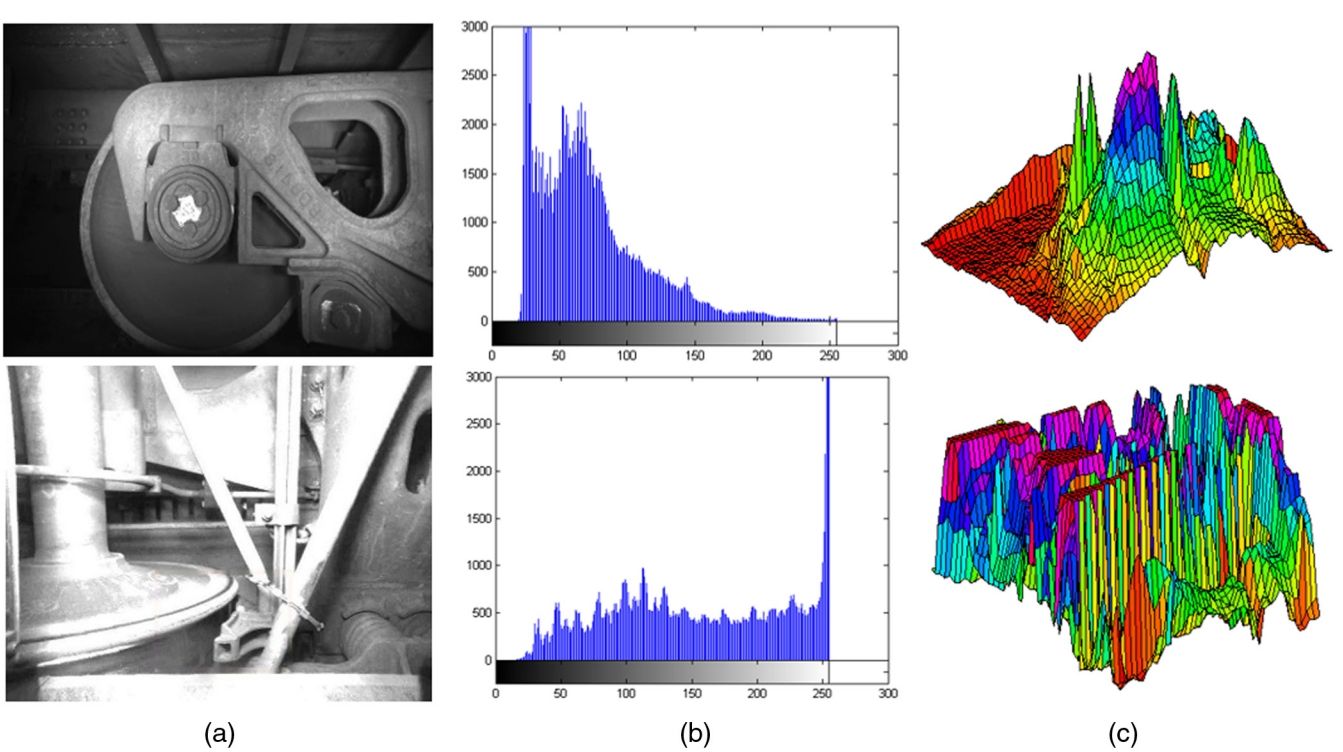

Fig. 1 Original images captured by TFDS and its background model (a) different fault images in TFDS, (b) gray histogram, and (c) background model.

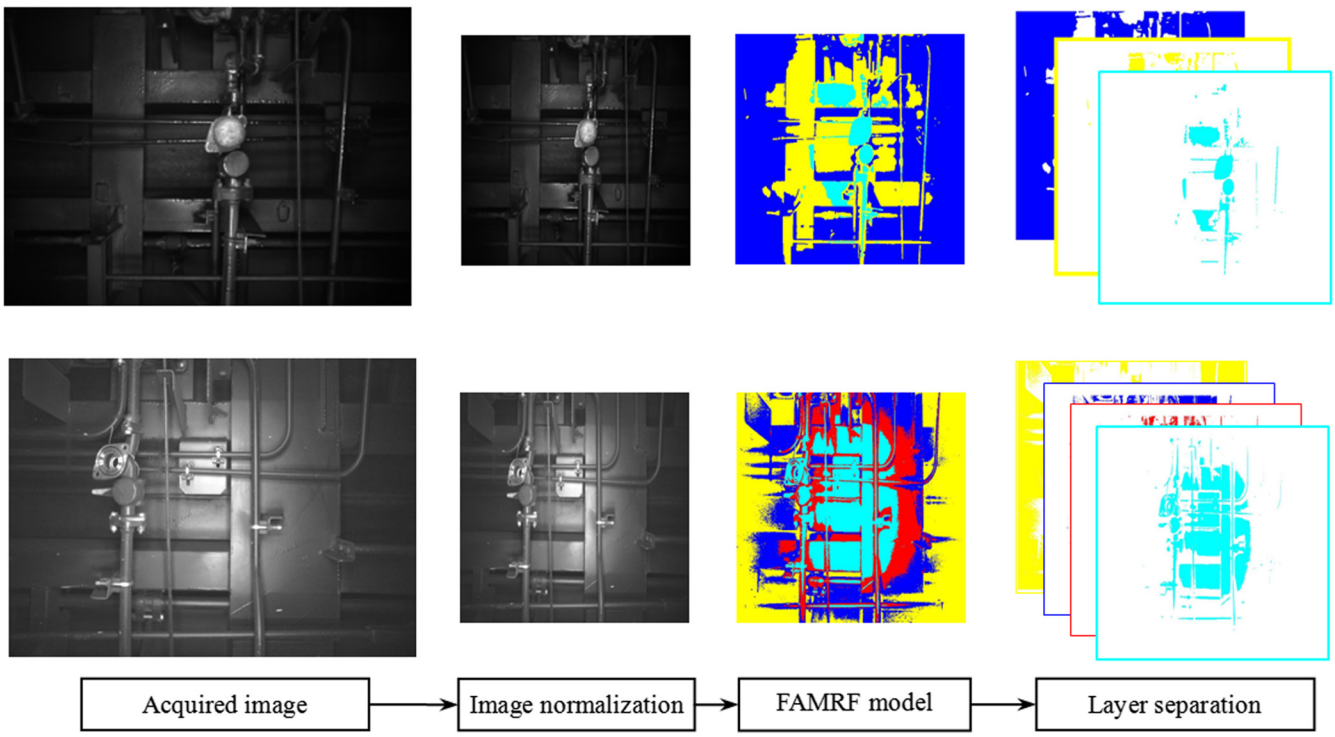

Fig. 2 Diagram of the FAMRF hierarchy model.

wavelet.[24] $W^{(b)}$ $(b \in \{1,2,\ldots,Np\})$ is the wavelet coefficient of b band image. Therefore, the original image corresponds to $n=0$ and the image sequence of l kinds of resolution is obtained. The feature vector of lowest resolution is given as

$$\mathbf{w}_{ij}^{n} = [\mathbf{w}_{ij}^{n,(1)}, \mathbf{w}_{ij}^{n,(2)}, \ldots, \mathbf{w}_{ij}^{n,(b)}, \ldots, \mathbf{w}_{ij}^{n,(Np)}]^T, \tag{1}$$

where $\mathbf{w}_{ij}^{n,(b)} = [w_{ij}^{LL,n,(b)}, w_{ij}^{LH,n,(b)}, w_{ij}^{HL,n,(b)}, w_{ij}^{HH,n,(b)}]^T$, and $w_{ij}^{LL,n,(b)}$, $w_{ij}^{LH,n,(b)}$, $w_{ij}^{HL,n,(b)}$, $w_{ij}^{HH,n,(b)}$ are the wavelet coefficients of pixel (i,j) in LL, LH, HL, and HH subband (as shown in Fig. 3) of b band image on scale n, respectively. The set of grid S in TFDS images can be defined as $S = \{S_1, S_2, \ldots, S_n, \ldots, S_{l-1}\}$, where S_n denotes the grid on scale n. And then, the label field of b band image is developed based on the grid set S_n at each resolution.

According to MRF theory, label field set $\mathbf{X}$ is composed of the pixel labels of the decomposed image, namely $\mathbf{X} = \{x_{S_1}, x_{S_2}, \ldots, x_{S_n}\}$. $x_{ij}^{(n)}$ is the label of pixel (i,j) in grid set S_n on scale n. The Potts' model[25,26] is selected to calculate the local constraint relationships between different labels. The local probability $P[x_{ij}^{(n)}|x_{N_{ij}}^{(n)}]$ of the label field set can be expressed as

$$P\left[x_{ij}^{(n)}|x_{N_{ij}}^{(n)}\right] = \frac{\exp\left[-\sum_{i'j'\in N_{ij}} V(x_{ij}, x_{i'j'})\right]}{\sum_{x_{ij}^{(n)}\in L} \exp\left[-\sum_{i'j'\in N_{ij}} V(x_{ij}, x_{i'j'})\right]},$$
$$V(x_{ij}, x_{i'j'}) = \begin{cases} \beta & (x_{ij} \neq x_{i'j'}) \\ 0 & (x_{ij} = x_{i'j'}) \end{cases}, \tag{2}$$

where L is the phase space[9] of the label field. N_{ij} is the neighborhood of pixel (i,j). V denotes multivariate potential function of the Potts' model. β is a given constant. $x_{N_{ij}}^{(n)}$ is the label of the neighborhood of pixel (i,j) on scale n.

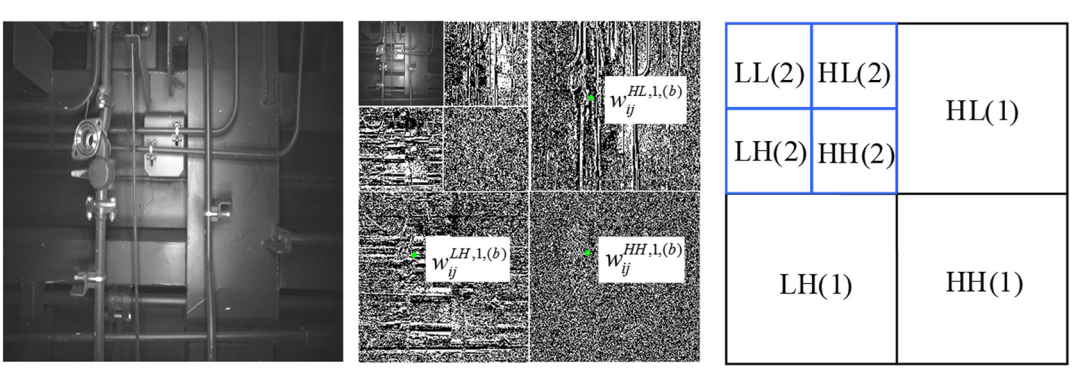

Fig. 3 Wavelet decomposition for TFDS images.

Each feature of the decomposed image is independent of any others under the condition of a given class. The Gauss model[26] is used to develop the feature field model, and a probability distribution of the feature field $\mathbf{W}$ is defined as follows:

$$f[w_{ij}^{(n)}|x_{ij}^{(n)}] = \sum_{m=1}^{M} \lambda_m^{(n)} f[w_{ij}^{(n)}|x_{ij}^{(n)} = m], \tag{3}$$

$$f[w_{ij}^{(n)}|x_{ij}^{(n)} = m] = \frac{1}{\sqrt{2\pi \sum_m^{(n)}}} \times \exp\left\{-\frac{1}{2}[w_{ij}^{(n)} - \mu_m^{(n)}]^T \left[\left|\sum_m^{(n)}\right|\right]^{-1} [w_{ij}^{(n)} - \mu_m^{(n)}]\right\}, \tag{4}$$

where m $(m = 1,2,\ldots,M)$ is the classified label. M is the total number of labels. $f[w_{ij}^{(n)}|x_{ij}^{(n)}]$ is a Gauss density probability function of pixels on scale n. $f[w_{ij}^{(n)}|x_{ij}^{(n)} = m]$ and $\lambda_m^{(n)}$ are a Gauss density probability function and the proportion of pixels of the class m on scale n, respectively. $\mu_m^{(n)}$ and $\sum_m^{(n)}$ are the mean vector and covariance matrix of the class m on scale n, respectively.

2.2 *Parameters Estimation*

Based on the theory above, the joint distribution of the label field $\mathbf{X}$ and the feature field $\mathbf{W}$ is shown as

$$P(W,X) = \prod_{n=0}^{l-1} \prod_{(i,j)\in S_n} f[w_{ij}^{(n)}|x_{ij}^{(n)}] P[x_{ij}^{(n)}|x_{N_{ij}}^{(n)}]. \tag{5}$$

Maximum a posteriori (MAP)[27,28] explicitly takes a certain prior distribution of signal in modeling, which results in a much more accurate estimation than that without the prior knowledge. Meanwhile, MAP-based approaches can effectively alleviate overfitting to yield satisfactory performance. So, it is used to estimate segmentation in a wavelet domain

$$\hat{x} = \arg\min_x \left\{\prod_{n=0}^{l-1} \prod_{(i,j)\in S_n} [E_{w_{ij}^{(n)}|x_{ij}^{(n)}} + E_{x_{ij}^{(n)}}]\right\}, \tag{6}$$

where $E_{w_{ij}^{(n)}|x_{ij}^{(n)}}$ and $E_{x_{ij}^{(n)}}$ denote the energy function of the feature field and label field, respectively. The Gauss model is used to estimate the feature field of each class in each scale, and the energy function of the feature field of class m is given as

$$E_{w_{ij}^{(n)}|x_{ij}^{(n)}=m} = -\frac{1}{2}[w_{ij}^{(n)} - \mu_m^{(n)}]^T \left[\left|\sum_m^{(n)}\right|\right]^{-1} [w_{ij}^{(n)} - \mu_m^{(n)}] + \ln\left[\sqrt{2\pi}\left|\sum_m^{(n)}\right|^{\frac{1}{2}}\right]. \tag{7}$$

The Potts' model is used to estimate the label field, and its energy function is given as $E_{x_{ij}^{(n)}} = \beta \times n_{ij}[x_{ij}^{(n)}]$, where $n_{ij}[x_{ij}^{(n)}]$ is the number of neighborhood locations whose phases are equal to that of $x_{ij}^{(n)}$ in the neighborhood of pixel (i,j). The label field energy is used to erase the isolated points in current segmentation to obtain better consistency.

The model parameters estimation cannot be performed by training data in unsupervised segmentation. This problem is solved by searching for a suboptimal solution using the expectation-maximization (EM)[27,29] method. The optimal segmentation results of each scale are obtained by the iterative conditional mode (ICM)[30] method using

$$\hat{x}_{ij}^{(n)} = \arg\min_{x_{ij}^{(n)}} \{E_{w_{ij}^{(n)}|x_{ij}^{(n)}} + E_{x_{ij}^{(n)}}\}. \tag{8}$$

To minimize the influence of fixed potential β, the variable weight $\alpha(t,n)$ is introduced to connect the feature field and label field model.[9]

$$\hat{x}_{ij}^{(n)} = \arg\min_{x_{ij}^{(n)}} \{\alpha(t,n) E_{w_{ij}^{(n)}|x_{ij}^{(n)}} + E_{x_{ij}^{(n)}}\}, \tag{9}$$

where $\alpha(t,n) = \frac{InC(n)}{t+1} + \frac{1}{M}$, and $InC(n)$ is the initial value of variable weights on the scale n, which is defined as $InC(n) = e \times 2n$. e is an irrational constant $(e = 2.71828\ldots)$.

2.3 *Initial Segmentation Algorithm*

The EM algorithm for image segmentation must rely on appropriate initial segmentation methods such as k-means[31] and fuzzy c-means. Unlike these methods that must specify the number of cluster categories in advance, affinity propagation[23] does not depend on a "good" initial cluster or group. Instead, AP obtains accurate solutions by approximating the NP-hard problems in a much more efficient and accurate way.[32] AP can use arbitrarily complex affinity functions as it does not need to search or integrate the parameter space.[32] Due to the flexibility of the AP method, an improved k-means based on affinity function is explored for effective image segmentation.

AP usually uses the negative squared Euclidean distance to measure the similarity $s(i,j)$. However, TFDS images contain so many pixels, which are generally >10,000 (100×100 pixels), that the computational complexity of affinity functions is extremely high. Therefore, the Euclidean distance is not appropriate for measuring similarity $s(i,j)$.

Traditionally, the histogram provides a visible clue for the probability density function of the objects that are distinguished from the background.[32] So, we use the Euclidean distance and the geodesic distance between points i and j along the probability density function of the histogram to measure similarity $s(i,j)$. Because of the noise in a TFDS image, its histogram varies greatly. A framework is established to smooth the histogram through diffusion-based kernel density estimation (KDE).[33] Figure 4 shows an example of smooth KDE.

2.4 *Termination Criterion for Iteration*

The number of iteration is usually used as the termination criterion in the ICM method. However, such prior knowledge cannot be obtained in unsupervised condition. Reduction of

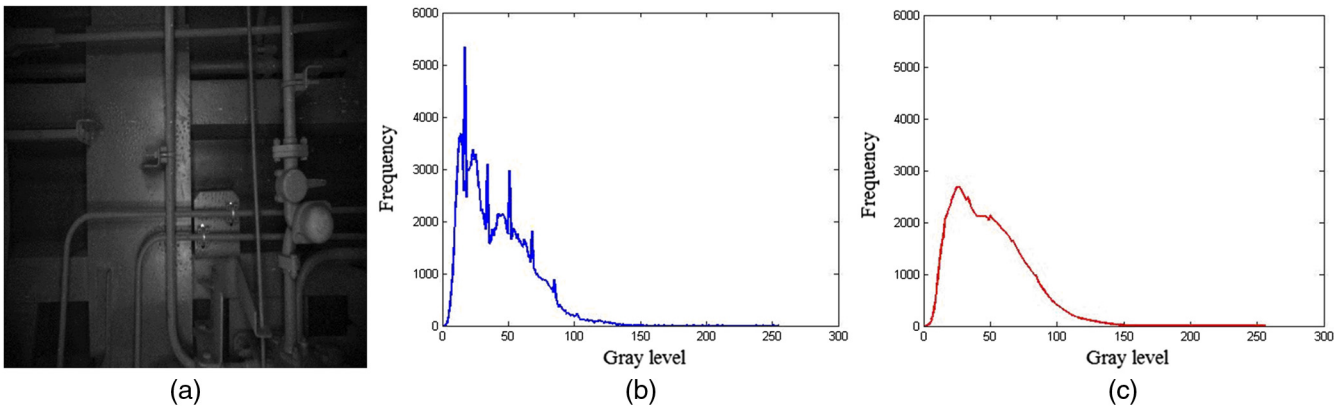

Fig. 4 An example smooth KDE conducted on the TFDS image histogram (a) normalized image, (b) original histogram, and (c) final histogram after smoothing by diffusion-based KDE.

pixel energy shows the convergence of the MRF-based segmentation algorithm. Therefore, according to the energy function of MRF, the standard deviation or its differential of the energy of different labels is calculated as a reference standard.

Standard deviation is considered as the iterative criterion, the energy of each pixel after t iterative estimation is $e_{\hat{x}_{ij}^{(n)(t)}}$. The total energy $E_{\hat{x}_{ij}^{(n)(t)}}$ on scale n and the mean energy $E\mu_m^{(n)(t)}$ in the class m after t iterative estimation can be, respectively, defined as

$$E_{\hat{x}_{ij}^{(n)(t)}} = \sum_{(i,j)\in S_n} e_{\hat{x}_{ij}^{(n)(t)}} = \sum_{m=1}^{M} \sum_{(i,j)\in S_n} e_{\hat{x}_{ij}^{(n)(t)}=m}, \tag{10}$$

$$E\mu_m^{(n)(t)} = \frac{\sum_{(i,j)\in S_n} e_{\hat{x}_{ij}^{(n)(t)}=m}}{N[\hat{x}_{ij}^{(n)(t)} = m]}, \tag{11}$$

where $N[\hat{x}_{ij}^{(n)(t)} = m]$ is the number of pixels in class m. And the standard deviation $E\sigma_m^{(n)(t)}$ is

$$E\sigma_m^{(n)(t)} = \sqrt{\frac{\sum_{(i,j)\in S_n}\{e_{\hat{x}_{ij}^{(n)(t)}=m} - E\mu[\hat{x}_{ij}^{(n)(t)} = m]\}}{N[\hat{x}_{ij}^{(n)(t)} = m]}}, \tag{12}$$

and the threshold of standard deviation $T_m^{(n)}$ at t iterative estimation can be defined as

$$T_m^n(t) = \left|1 - \frac{E\sigma_m^{(n)(t+1)}}{E\sigma_m^{(n)(t)}}\right|. \tag{13}$$

Standard deviation differential is considered as the iterative criterion, ICM algorithm stops iterating when the derivative of the threshold is less than a minimum value (i.e., $\varepsilon = 0.01$), that is $dT_m^n(t)/dt < \varepsilon$. The process for building the FAMRF model is summarized in Algorithm 1.

Taking TFDS image in Fig. 4(a) as an example, a standard deviation and its differential of different labels on different scales ($n = 2$) are calculated separately as the energy threshold. With an increase in the iteration times, $T_m^{(n)}$ will gradually decrease until the approach is stable and the termination criterion algorithm can be constructed.

Algorithm 1 Calculation steps for FAMRF model.

Input: Original image I, normalization size $N \times N$, scale n, potential β.

Calculation steps:

1. The original image is normalized as $N \times N$, and then decomposed into n scales by wavelet transform. The k-means algorithm improved by AP theory with the image histogram is used to obtain the initial segmentation results on the maximum scale n.
2. The MAP method with the EM algorithm is used to estimate the parameters of the Gauss model, and the new image segmentation is calculated by the ICM method.
3. Repeat step 2 to obtain the further segmentation results on the maximum scale until the threshold of standard deviation differential reaches the termination criterion. The segmentation results are calculated as the initial segmentation of the nearest scale $n - 1$.
4. Repeat step 3 until the final segmentation results on the minimum scale are obtained.

Output: Segmentation results I_s.

In Figs. 5(a)–5(c), the threshold of standard deviation tends to be stable when $t = 25$. If the standard deviation differential is used for iterative criterion, the decrease will be stable when $t = 10$, which indicates that the classification of pixels has been completed. Compared with the standard deviation, standard deviation differential can effectively reduce the computational complexity of the FAMRF model.

3 Parallel Shape Matching

After hierarchical segmentation (see details in Fig. 2), we use the EHF[34] shape descriptor to identify whether there is a fault in each layer. The parallel matching framework is introduced to improve the matching speed and efficiency.

In Fig. 6, $\mathbf{SX} = \{\mathbf{x}_i\}$ $(i = 1,2,\ldots,NS)$ denotes the equidistant sample points set of the outer contour of the target. The sample point $\mathbf{x}_i$ follows the contour anticlockwise. NS is the total number of sample points. For each sample point $\mathbf{x}_i$, its tangent line l_i is used as a reference axis l. The distance

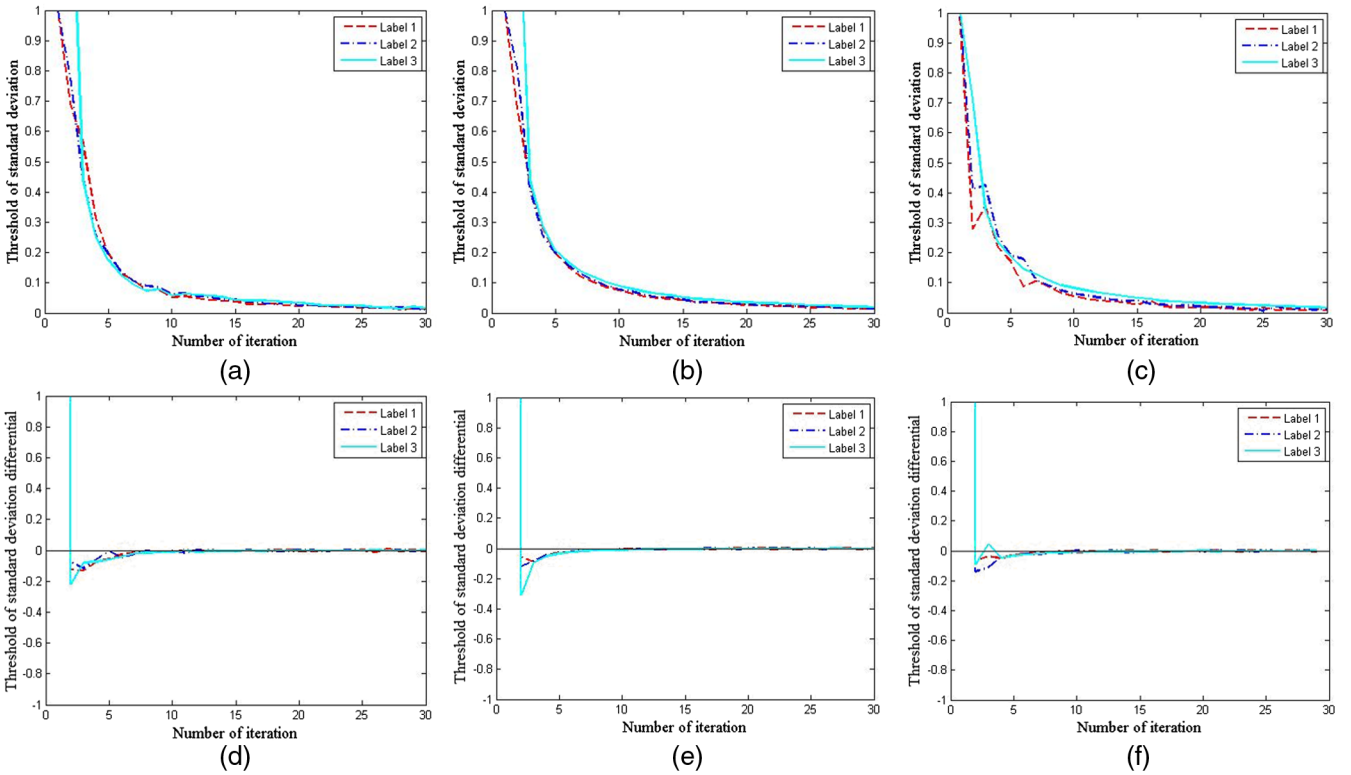

Fig. 5 The threshold of energy standard deviation and its differential of different labels on different scales ($n = 2$). (a) Standard deviation on scale 2, (b) standard deviation on scale 1, (c) standard deviation on scale 0, (d) standard deviation differential on scale 2, (e) standard deviation differential on scale 1, and (f) standard deviation differential on scale 0.

between the j'th $(j = 1,2,\ldots,NS)$ sample point $\mathbf{x}_j$ and the tangent line l_i is defined as the height value $H_{i,j}$. When the sample point is above the axis, its height value is positive. When the point is below the axis, its height value is negative. Therefore, the height value $H_{i,j}$ of point $\mathbf{x}_j$ to $\mathbf{x}_i$ is given as

$$H_{i,j} = \frac{det(\mathbf{x}_{i-1}, \mathbf{x}_j, \mathbf{x}_{i+1})}{|\mathbf{x}_{i-1}\mathbf{x}_{i+1}|}, \tag{14}$$

where *det* denotes the determinant. The default value of NS is 100.[22]

Because the reference axis is computed by three points $\mathbf{x}_{i-1}$, $\mathbf{x}_i$, and $\mathbf{x}_{i+1}$. Moreover, the accurate height values $h_{i,j}$ is $h_{i,j} = H_{i,j} - \Delta H$, where ΔH is the distance between axis l_i and line l. ΔH can also be computed by Eq. (14). The extract height function shape descriptor is defined as

$$\mathbf{h}_i = (h_{i,i+1}, \ldots, h_{i,NS}, h_{i,1}, \ldots, h_{i,i-1})^T. \tag{15}$$

The descriptor may be overly sensitive to local boundary deformations. A smoothing function[22] is used to solve this problem. Set a given positive integer lv $(1 < lv < NS)$ as smoothing level, the sequence of integers is divided into disjoint intervals $[1, lv]$, $[lv + 1, 2lv], \ldots,$ and then the mean value of the height values in each interval F_i^j as

$$F_i^j = \frac{1}{lv} \sum_{t=(j-1)lv+1}^{jlv} h_i^t, \tag{16}$$

where $j = 1,2,\ldots,D$ with $D = \lfloor NS/lv \rfloor$ (the integer part of NS/lv), t is an intermediate variable. Then, the matching cost $ct(p,q)$ between the point p on shape **SX** and the point q on shape **SY** is

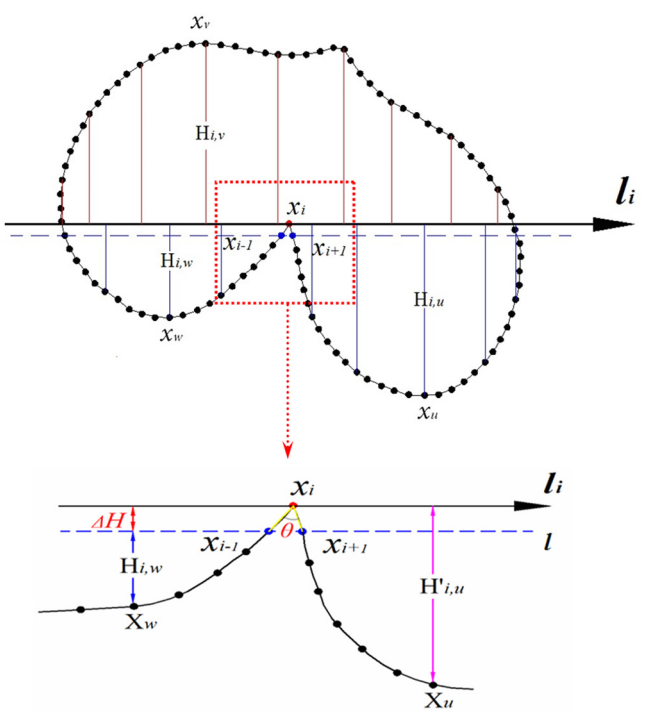

Fig. 6 The EHFs for the sample point.

$$ct(p,q)=\sum_{t=1}^{D}\lambda_t|F_p^t-F_q^t|, \tag{17}$$

where λ_t is the weight coefficient for every component of the height feature[22]

$$\lambda_t=\frac{1}{\min\{t,D-t+1\}}. \tag{18}$$

Dynamic programming (DP) has been widely used to search the optimal correspondence of contour points. Therefore, the optimal corresponding relationship g^* between shape **SX** and **SY** is calculated by the DP algorithm to minimize the overall matching cost[22]

$$\mathrm{DP}_{\min}(\mathbf{SX},\mathbf{SY})=\sum_{i=1}^{NS}ct[h_i,g^*(h_i)]. \tag{19}$$

It is essential to explore the minimum matching cost (MMC) between the template and the object contours for shape matching. The MMC between the normalized template (set as **Tmp**) and the object contours [set as $\mathbf{Con}(s)$, $s=1,2,\ldots,C$, where C is total number of the contours in detected images] is given as follows:

$$\mathrm{DP}_{\min}[\mathbf{Con}(s),\mathbf{Tmp}]=\sum_{i=1}^{NS}ct[h_i^s,g^*(h_i^s)], \tag{20}$$

where h_i^s is the exact height value of s'th object contours.

In a word, the target shape (set as **Tag**), namely the contour of detected objects, exhibits substantial similarity to the template, and the MMC is defined as

$$\mathrm{DP}_{\min}(\mathbf{Tag},\mathbf{Tmp})=\min\{\mathrm{DP}_{\min}[\mathbf{Con}(s),\mathbf{Tmp}]\}. \tag{21}$$

The process for parallel shape recognition is summarized in Algorithm 2.

Algorithm 2 Calculation steps for parallel shape matching.

Input: Segmented images I_s, normalized template I_t, smoothing level k

Calculation steps:

1. The contours of segmented images I_s and normalized template I_t are extracted, respectively, and the contours are sampled uniformly.
2. The EHF descriptors are used to represent sampled contour features of segmented images and the normalized template, and the feature reduction is performed based on the smoothing function.
3. The matching cost of each contour is calculated, and the parallel DP algorithm is used to find the optimal correspondence relationship between each contour in a parallel domain.
4. Find the minimum matching cost between the normalized template and object contours of target shape in segmented images.

Output: Result of parallel shape matching

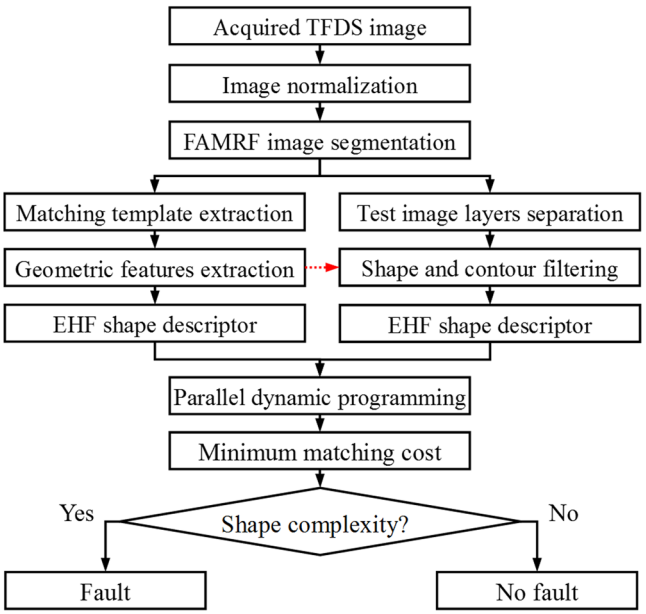

Fig. 7 Diagram of railway equipment detection based on FAMRF and EHF.

4 Railway Equipment Detection

Monochrome cameras are used to collect gray images in TFDS. So the color features are not helpful for fault inspection. In addition, image acquisition devices are installed outdoors. The images are easily affected by the varying illumination, weather, and other factors. As shown in Fig. 1(a), the acquired images are blurred or overexposed in different weather. The common faults in TFDS are mostly caused by the loss, damage, and position changes of small machine components. So, the differences between normal and fault images are not obvious, which increase the inspection difficulty and computational complexity.

Taking the associations between pixel spaces and the histogram information into consideration, the proposed FAMRF algorithm can effectively solve the problems caused by the constant color and complex background of TFDS images. The proposed EHF shape descriptor can accurately represent the contour feature of fault areas, which is helpful to improve the fault recognition accuracy. The process of railway equipment detection in detail based on FAMRF and the EHF shape descriptor is summarized in Fig. 7.

The test image is separated by the label after FAMRF. Note that the geometrical features of template contour are the divergence, which measure the compactness of a region, defined as $(\text{perimeter})^2/\text{area}$.[34] The shape complexity $C(\mathbf{X})$ presented in Fig. 7 is the one calculated by Eq. (22) described as (std means standard deviation)

$$C(\mathbf{X})=\frac{1}{D}\sum_{t=1}^{D}\mathrm{std}(F_1^t,F_2^t,\ldots,F_{NS}^t). \tag{22}$$

5 Experiments and Analysis

The experiment is developed under the PC condition of Intel Xeon E5-2630 V2 processor, 64G RAM, and Win7 OS. For the inspection of railway equipment, the default parameters are set as follows: normalization $N=400$, wavelet

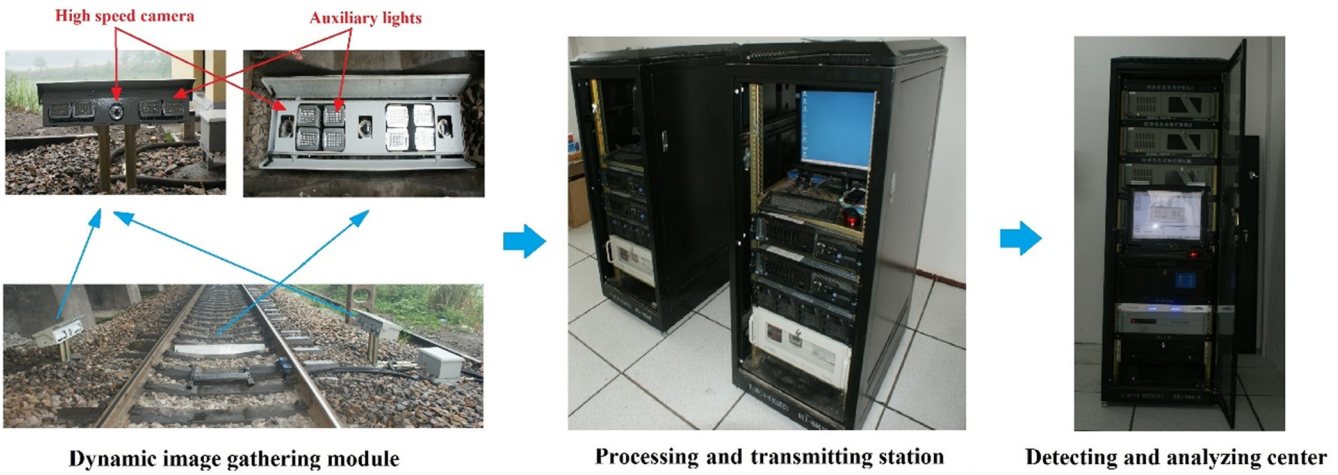

Fig. 8 Overview of the TFDS.

decomposition level $l = 3$, potential function $\beta = 5$, sampling number $NS = 100$, and smoothing level $lv = 10$. All images are resized to 512×700 pixels to improve the detection speed.

5.1 *System Overview*

As a real-time human–machine interaction fault detection system, TFDS integrates many technologies such as dynamic image acquisition, LED strobe control, big data real-time processing, network communication, and pattern recognition. As shown in Fig. 8, the hardware of TFDS includes three parts: dynamic image gathering module, processing and transmitting station, as well as detecting and analyzing center.

The dynamic image gathering module is mainly composed of a high-speed camera and the compensation light source installed on the track and the rail side. When freight trains pass through, the images are simultaneously captured by the dynamic image gathering module. Then these images are transmitted to processing and transmitting station and processed by the detecting and analyzing center to detect a possible fault of TFDS images. The filter on the compensation laser light source is used to eliminate the light interference. TFDS also contains an automatic vehicle number recognition system, which is used to receive the tag information installed at the bottom of the train. The preliminary classification of images can be achieved according to the tag information, which provides certain convenience for fault inspection.

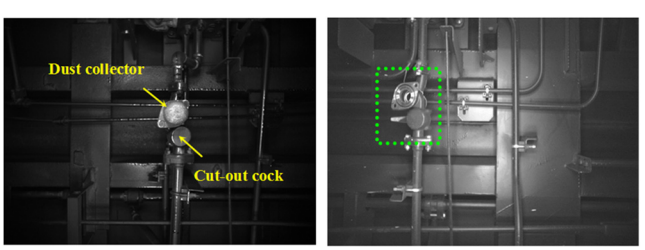

Fig. 9 Some typical samples used in air brake system inspection (a) normal air brake image and (b) dirt collector damaged and cut-out cock handle closed.

5.2 *Evaluation Metrics*

According to the provision of the Ministry of Railways, missing detection rate (MDR) and false detection rate (FDR) of a fault detection algorithm for TFDS fault images should be <5% and 10%, respectively. To evaluate the effectiveness of our algorithm, there are several indices defined as follows: total images, fault images, automatic detection,

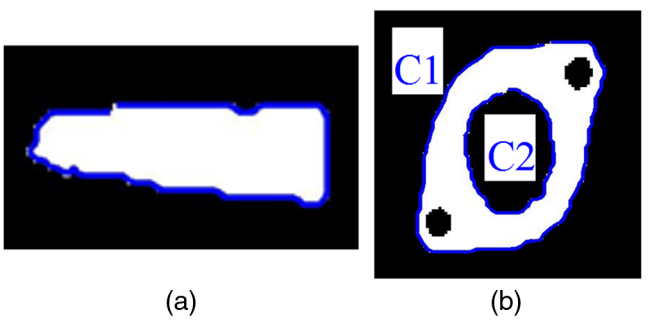

Fig. 10 Template extraction for air brake system (a) cutout cock handle template extraction and (b) dust collector template extraction.

Table 1 Detection results of air brake system.

	Cut-out cock handle closed		Dust collector damaged	
Algorithm	MDR ↓ (%)	FDR ↓ (%)	MDR ↓ (%)	FDR ↓ (%)
Cascade detector	7.88	15.29	1.88	8.82
HOG + Adaboost + SVM	2.59	9.41	0.47	2.59
Faster-RCNN (ZF)	1.18	4.00	0	14.94
Faster-RCNN (VGG16)	0.94	1.41	0	3.65
Faster-RCNN (VGG19)	1.29	0.47	0	3.06
R-FCN (ResNet-50)	0.83	2.59	0	19.41
SSD (VGG16)	0.12	23.06	0	26.71
FAMRF + EHF	1.29	5.41	1.06	2.82

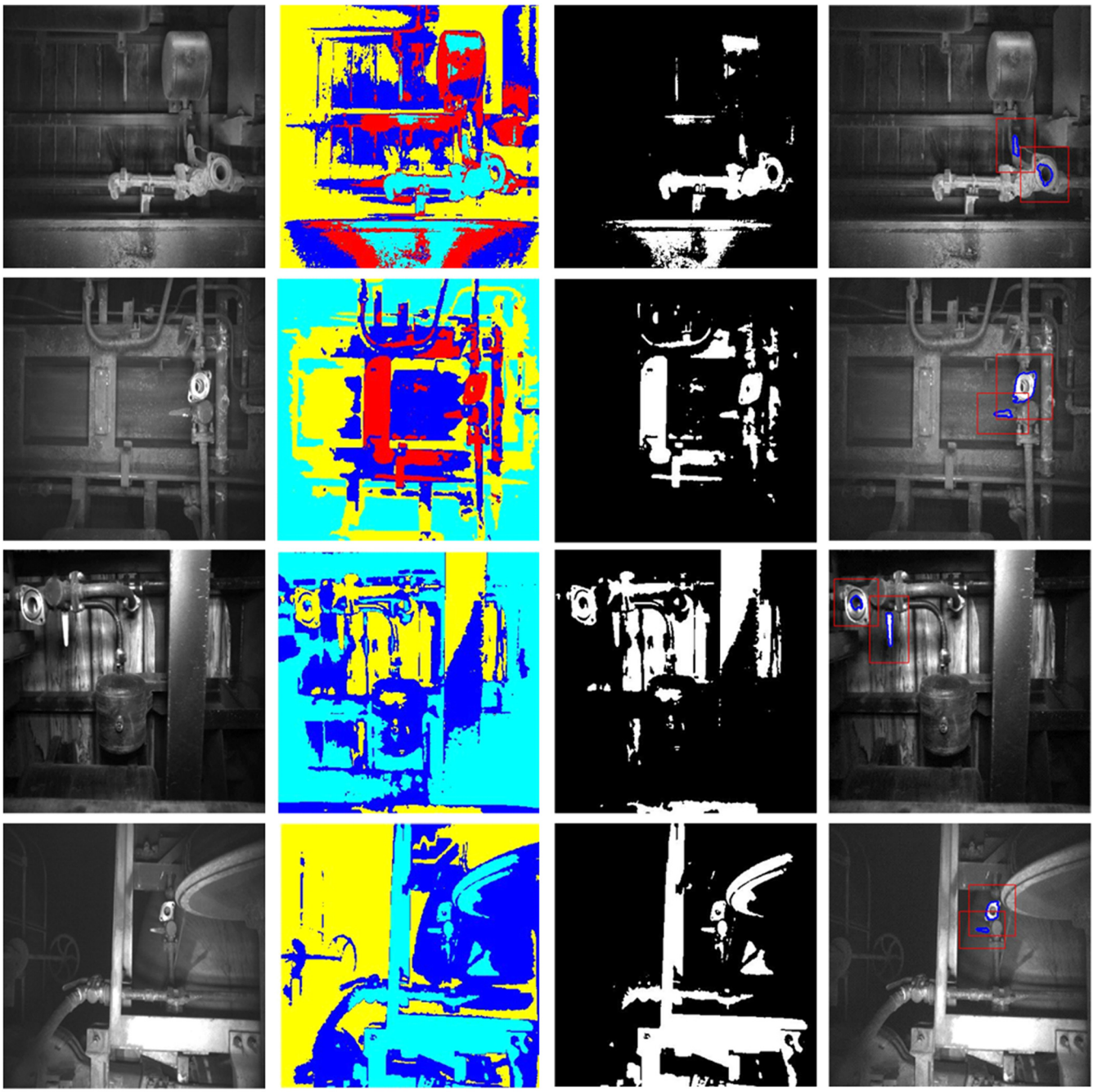

Fig. 11 Air brake system detection (a) normalized image, (b) FAMRF hierarchy model, (c) separated layer within two faults, and (d) detection result.

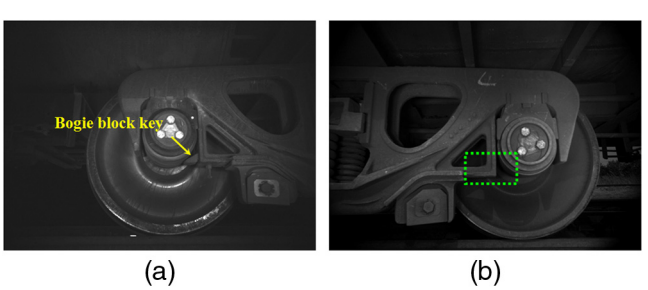

Fig. 12 Some typical samples used in BBK inspection (a) normal image and (b) BBK is missing.

correct detection, FDR, and MDR. Missing detection means the operation that a fault image is classified as a normal image. False detection means the operation that a normal image is detected as a fault image. And MDR and FDR are defined as

$$\mathrm{MDR} = (\text{Fault images} - \text{Correct detection})/\text{Total images}, \tag{23}$$

$$\mathrm{FDR} = (\text{Automatic detection} - \text{Correct detection})/\text{Total images}. \tag{24}$$

5.3 *Results and Analysis*

5.3.1 *Air brake system*

The air brake system is an important component in a freight train. It contains the cutout cock and dust collector. The cutout cock cuts off the air from the main air reservoir to the brake pipe, and it is also used to shut down the brake pipe. The dust collector is used to filter the impurities of compressed air. Normally, the handle of the cutout cock is invisible as shown in Fig. 9(a). When the train stops or the brake breaks down, the handle of the cutout cock is visible. As for the dust collector, the end cover of it is usually lost. These two defects often appear at the same time when the air brake system is malfunctioning, as shown in Fig. 9(b), which increases the difficulty of fault inspection.

For the faults that appear in the air brake system, the automatic detection process based on the proposed hierarchy model is described in Sec. 4. In particular, the fault area is manually selected as a matching template, as shown in Fig. 10. Meanwhile, the outer contour of each layer is extracted separately and screened by the geometric feature of the matching template. As shown in Fig. 10(b), the damaged dust collector contains two complete contours C1 and C2 after segmentation. C1 is the outside silhouette, and

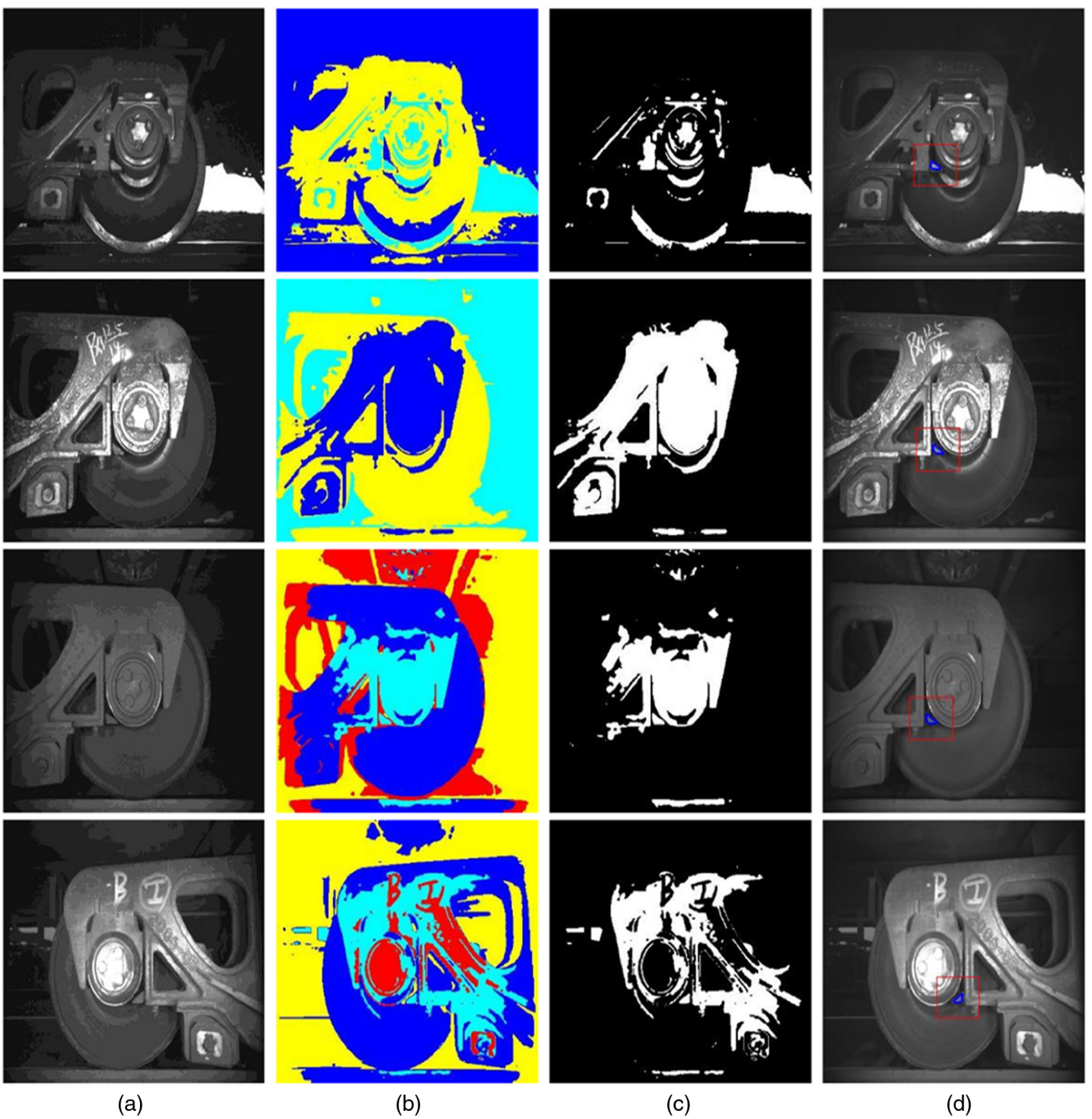

Fig. 13 BBK detection (a) normalized image, (b) FAMRF hierarchy model, (c) separated layer within BBK, and (d) detection result.

C2 is the inner silhouette. Both of them have a certain similarity in shape structure. And C2 is a better target when C1 cannot be clearly obtained by FAMRF.

To further illustrate the superiority of the proposed algorithm, we performed experiments with some widely used object detection methods such as a cascade detector[35] based on LBP (Local binary patterns), and histogram of oriented gradient (HOG) (to descript features) + Adaboost (to select features) with a linear SVM classifier,[36] single shot multibox detector (SSD)[37] based on the VGG16 model, region-based fully convolutional networks (R-FCN)[38] based on a ResNet-50 model, and Faster RCNN[39] based on different models (ZF, VGG16, and VGG19), respectively. The same TFDS images database are trained for fault recognition. The parameters of the above methods are all set as defaults. And all experiments of deep learning algorithms are conducted in Caffe[40] with a K40 GPU and the Ubuntu 16.04 OS. A total of 850 images including 179 images of the cutout cock handle closed and 52 images of the dust collector damaged are used for testing. The detection results are shown in Table 1. Some of the results are shown in Fig. 11.

It can be seen from Table 1 that our method has 1.29% MDR, and 5.41% FDR for the cutout cock handle database. And for the dust collector database, our method has 1.06% MDR, and 2.82% FDR. The proposed method is able to accurately inspect the faults of the air brake system in TFDS. In addition, the performance of the deep learning methods is comparable with our method. But the SSD and R-FCN methods are sensitive to the noise, and their FDR is too high. The Faster-RCNN method has a good performance. But this method needs to be trained for several days with a large number of training samples and needs high hardware requirements (see details in Sec. 5.3.4).

However, the FDR on the cutout cock handle is generally high. By analyzing Fig. 9, there are many pillar-shaped poles that are close to the cutout cock. After image segmentation by the FAMRF algorithm, the structures of the cutout cock handle are extremely similar to the poles, which make it difficult to distinguish through shape features. In addition, the method of HOG + Adaboost + SVM performs better than our algorithm. Because the outer contour of the dust collector changes little when the dust collector is damaged. The shape feature of the outer contour has not enough distinctive power in shape matching.

5.3.2 *Bogie block key*

The BBK is a very small part in a TFDS image, as shown in Fig. 12. Due to the clear contour of the TFDS image, we can easily obtain its shape features. The combination of the shape features and the geometry measurement can be used to classify different objects. Meanwhile, the shape features are insensitive to variable illumination, noise, and other interfering factors.

Because the triangle contour of the BBK is a relatively unique structure in complex background, shape features of the BBK can be described by the EHF shape descriptor. Based on shape features of the triangle contour and EHF shape descriptor, a normal target recognition mode is adopted to process the fault region. In this mode, the BBK region in a normal image is selected as a template. Some of the results are shown in Fig. 13. To further illustrate the superiority of the proposed algorithm, we also performed experiments using the above five methods. A total of 2897 images are used to inspect the BBK missing including 1257 fault images. The detection results are shown in Table 2.

The performances of our method are slightly better than others, but our method has a high FDR. As shown in Figs. 12 and 13, the main reason for the high FDR is that BBK is too small compared with other components. The similar shapes of other components have a strong influence on

Table 2 Detection results of BBK missing.

Algorithm	MDR ↓ (%)	FDR ↓ (%)
Cascade detector	2.11	1.31
HOG + Adaboost + SVM	0.90	2.14
Faster-RCNN (ZF)	1.14	0
Faster-RCNN (VGG16)	4.24	0.10
Faster-RCNN (VGG19)	6.31	0.03
R-FCN (ResNet-50)	3.59	0
SSD (VGG16)	1.93	0
FAMRF + EHF	0.76	1.52

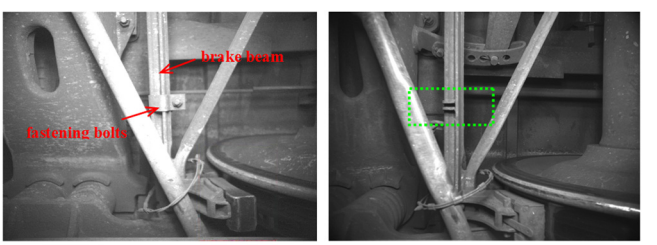

Fig. 14 Typical samples used in fastening bolt inspection (a) normal image and (b) fastening bolt missing.

Fig. 15 Contour extraction for fastening bolt inspection (a) template and (b) template screened by its contour.

shape matching for BBK. This is also the main reason for the high MDR of RCNN-based method especially Faster RCNN. For traditional Faster RCNN, the ROI pooling is performed on the final feature map layer (i.e., Conv5_3 in VGG16) to generate features of the region. However, the shallow convolution layer can extract some local features, such as texture and shape, whereas the deeper convolution layer learns more abstract features. With an increase in the convolution layers (e.g., ZF, VGG16, and VGG19 in Faster-RCNN), the feature of small objects is more difficult to be extracted, which leads to an increase in the MDR of the Faster RCNN method.

5.3.3 *Fastening bolt*

The fastening bolt on a brake beam is an essential component of a train brake device. When a train is braking, the brake beam will create a great horizontal force, which may break the fastening bolt or make it fall off, as shown in Fig. 14. The status of the fastening bolt after falling off is difficult to describe based on an exact feature model, which becomes the primary difficulty of fault inspection.[9]

For the fault inspection of the fastening bolt, the automatic detection process based on the proposed hierarchy model is described in Sec. 4. The normal fastening bolt

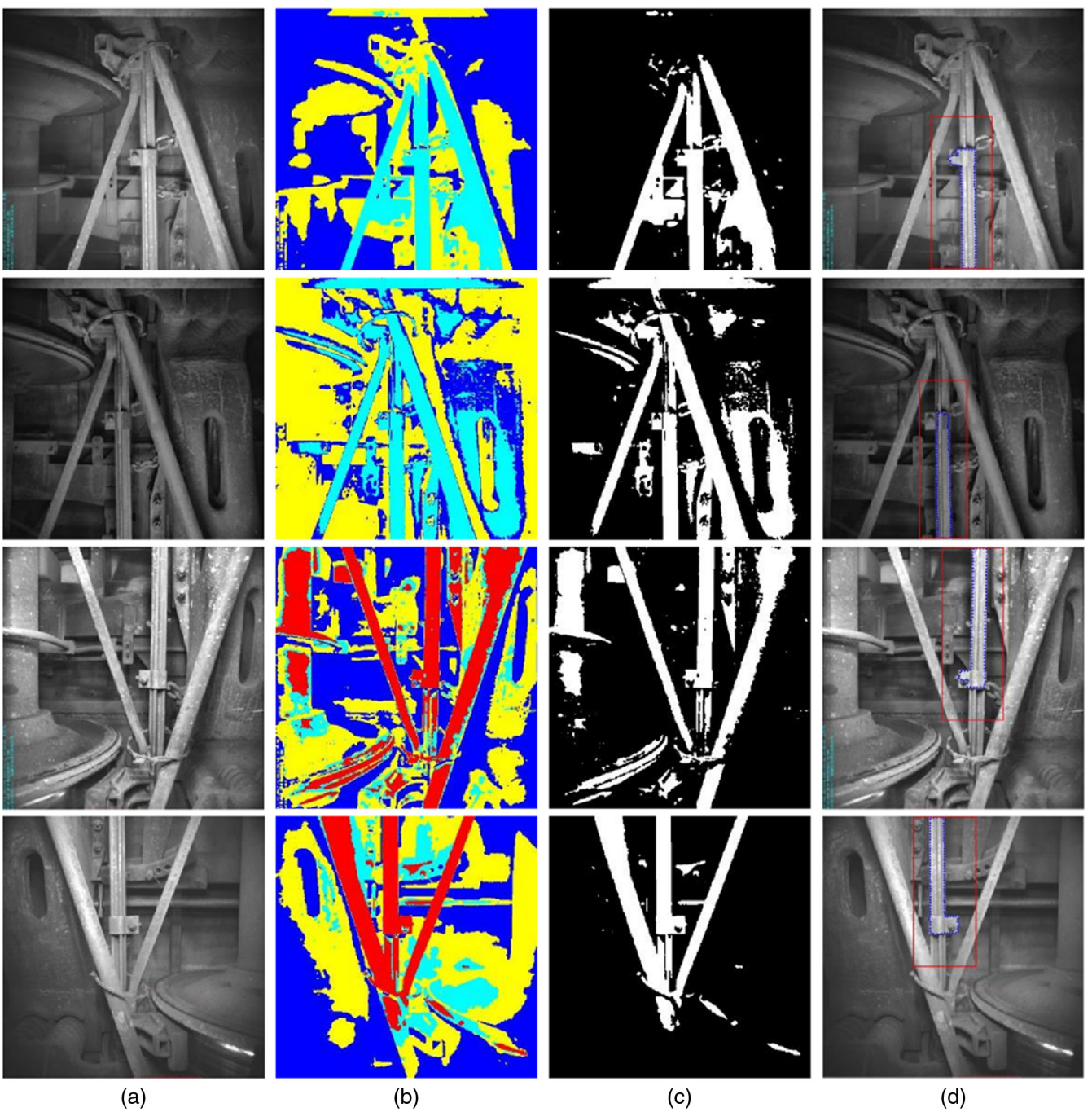

Fig. 16 Fastening bolt detection (a) normalized image, (b) FAMRF hierarchy model, (c) separated layer within fastening bolt, and (d) detection result.

Table 3 Detection results of fastening bolt missing.

Algorithm	MDR ↓ (%)	FDR ↓ (%)
Cascade detector	3.21	4.73
HOG + Adaboost + SVM	1.42	2.89
Faster-RCNN (ZF)	1.14	0
Faster-RCNN (VGG16)	0.05	0.05
Faster-RCNN (VGG19)	0.05	0.05
R-FCN (ResNet-50)	0.11	0.05
SSD (VGG16)	2.31	0.05
FAMRF + EHF	0.89	6.41

area is manually selected as the matching template as shown in Fig. 15. Some of the detection results of the fastening bolt are shown in Fig. 16. The detection results are detailed in Table 3. A total of 1902 images are used to inspect the missing fastening bolt including 367 fault images.

It can be seen from Table 3 that our method has 0.89% MDR, and 6.41% FDR. The proposed method can inspect the faults of the fastening bolt in TFDS. In addition, the performances of Faster-RCNN and R-FCN are comparable with our method. The SSD method is sensitive to the noise, and its MDR is too high. Unfortunately, the FDR of our method is generally high. By analyzing Figs. 14 and 15, there are no obvious changes between a normal and fault image. After segmentation, the structures of fault regions are extremely similar to normal images, which make it hard to distinguish by EHF.

5.3.4 *Comparison of computational complexity*

Table 4 lists a comparison of the training time and detection speed. To perform a freight train image with a size of 700×512 pixels, its detection speed is about 0.267 s including all steps with 12 workers in MATLAB R2016b. In Table 4, the training and computation speed of the cascade of LBP features and the HOG + Adaboost + SVM method are the fastest, but their accuracy is the lowest. In addition, the detection speed of the deep learning methods is satisfactory. But they always need too much time to train deep CNN, especially the SSD method without a pretraining model.

Table 4 Detection results of training and detection speed.

Algorithm	Training time/s	detection speed/s
Cascade detector	76	0.036
HOG + Adaboost + SVM	84	0.049
Faster-RCNN (ZF)	11,940	0.079
Faster-RCNN (VGG16)	29,160	0.239
Faster-RCNN (VGG19)	35,210	0.286
R-FCN (ResNet-50)	>36,000	0.179
SSD (VGG16)	>190,800	0.155
FAMRF + EHF	N/A	0.267

In summary, the proposed model based on FAMRF and the EHF shape descriptor can identify the typical faults above. Unlike machine learning methods, the manual annotation of images is not required. The algorithm can reduce the influence of single color, gray in-homogeneity, and complex background. Even if the relative position of the fault region is interfered by an artificial mark, the proposed algorithm can accurately inspect the faults in TFDS.

6 Conclusion

To ensure the rail traffic safety, a hierarchical feature-matching model is presented to automatically inspect the typical faults in TFDS. The proposed model consists of FAMRF image segmentation and EHF shape matching algorithms. TFDS images are first normalized and segmented by FAMRF with the Pyramid model and AP theory. Then, the images are separated by labels and filtered by the geometrical features of the template. Finally, the EHF shape description is used to realize parallel shape matching based on shape complexity. The detection results show that the proposed model can achieve automatic detection of the typical faults: air brake system malfunctions, BBK missing, and fastening bolt missing. The hierarchical feature-matching model is robust against changes of illumination in an outdoor environment and can achieve high detection accuracy. Further work will focus on the following three aspects to improve the speed, accuracy, and applicability of the proposed model.

1. The inspection speed should be further increased by reducing computational complexity of the hierarchical feature-matching model.
2. The detection accuracy could be improved by introducing machine learning algorithms into the model.
3. The model should be applied to other faults to improve the applicability.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant Nos. 51775177 and 51675166.

References

1. L. Liu, F. Zhou, and Y. He, "Automated visual inspection system for bogie block key under complex freight train environment," *IEEE Trans. Instrum. Meas.* **65**(1), 1–13 (2015).
2. Y. Cao et al., "Weighted margin sparse embedded classifier for brake cylinder detection," *Neurocomputing* **120**, 560–568 (2013).
3. J. Sun, Z. Xiao, and Y. Xie, "Automatic multi-fault recognition in TFDS based on convolutional neural network," *Neurocomputing* **222**, 127–136 (2017).
4. H. Kim and W. Y. Kim, "Automated inspection system for rolling stock brake shoes," *IEEE Trans. Instrum. Meas.* **60**(8), 2835–2847 (2011).
5. N. Li, Z. Wei, and Z. Cao, "Automatic fault recognition for brake-shoe-key losing of freight train," *Optik* **126**(23), 4735–4742 (2015).
6. W. Xu, "Research on typical train fault image recognition method based on shape descriptor," MS thesis, Department of Mechanical Engineering, Hubei University of Technology, Wuhan, China (2015).
7. C. Gu, "Image recognition based on object contour feature and its application in train bogie fault diagnosis," MS thesis, Department of Vehicle Engineering, Soochow University, Suzhou, China (2014).

8. F. Q. Zhou, R. Zou, and H. Gao, "Dust collector localization in trouble of moving freight car detection system," *J. Zhejiang Univ.–Sci. C* **14**(2), 98–106 (2013).
9. Y. Zhang, "Hierarchical feature matching of fault images in TFDS based on improved Markov random field and exact height function," MS thesis, Department of Mechanical Engineering, Hubei University of Technology, Wuhan, China (2017).
10. S. F. Lu et al., "Automated visual inspection of brake shoe wear," *Proc. SPIE* **9675**, 96752F (2015).
11. L. Liu, F. Zhou, and Y. He, "Automated status inspection of fastening bolts on freight trains using a machine vision approach," *Proc. Inst. Mech. Eng., Part F: J. Rail Rapid Transit* **230**(7), 1629–1641 (2015).
12. G. Nan and J. E. Liu, "A real-time visual inspection method of fastening bolts in freight car operation," *Proc. SPIE* **9675**, 96752G (2015).
13. F. Q. Zhou et al., "Automated visual inspection of angle cocks during train operation," *Proc. Inst. Mech. Eng., Part F: J. Rail Rapid Transit* **228**(7), 794–806 (2013).
14. C. Zheng and Z. Wei, "Automatic online vision-based inspection system of coupler yoke for freight trains," *J. Electron. Imaging* **25**(6), 061602 (2016).
15. H. Feng et al., "Automatic fastener classification and defect detection in vision-based railway inspection systems," *IEEE Trans. Instrum. Meas.* **63**(4), 877–888 (2014).
16. B. Peng, L. Zhang, and D. Zhang, "A survey of graph theoretical approaches to image segmentation," *Pattern Recognit.* **46**(3), 1020–1038 (2013).
17. W. Feng, J. Y. Jia, and Z. Q. Liu, "Self-validated labeling of Markov random fields for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(10), 1871–1887 (2010).
18. L. Dong, N. Feng, and Q. Zhang, "LSI: latent semantic inference for natural image segmentation," *Pattern Recognit.* **59**, 282–291 (2016).
19. M. Chen and J. Strobl, "Multispectral textured image segmentation using a multi-resolution fuzzy Markov random field model on variable scales in the wavelet domain," *Int. J. Remote Sens.* **34**(13), 4550–4569 (2013).
20. S. Belongie, J. Malik, and J. Puzica, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(4), 509–522 (2002).
21. H. Ling and D. W. Jacobs, "Shape classification using the inner distance," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(2), 286–299 (2007).
22. J. W. Wang et al., "Shape matching and classification using the height functions," *Pattern Recognit. Lett.* **33**(2), 134–143 (2012).
23. B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science* **315**(5814), 972–976 (2007).
24. Y. H. Zhang et al., "Global optimization of wavelet-domain hidden Markov tree for image segmentation," *Pattern Recognit.* **44**(12), 2811–2818 (2011).
25. C. Nieuwenhuis et al., "A survey and comparison of discrete and continuous multi-label approaches for the Potts model," *Int. J. Comput. Vision* **104**(3), 223–240 (2013).
26. H. Ayasso and A. Mohammad-Djafari, "Joint NDT image restoration and segmentation using Gauss-Markov-Potts prior models and variational Bayesian computation," *IEEE Trans. Image Process.* **19**(9), 2265–2277 (2010).
27. C. Santiago, J. C. Nascimento, and J.-S. Marques, "2D segmentation using a robust active shape model with the EM algorithm," *IEEE Trans. Image Process.* **24**(8), 2592–2601 (2015).
28. T. Yu and Y. H. Lai, "Generalized maximum a posteriori spectral amplitude estimation for speech enhancement," *Speech Commun.* **76**, 112–126 (2016).
29. G. Celeux, F. Forbes, and N. Peyrard, "EM procedures using mean field-like approximations for Markov model-based image segmentation," *Pattern Recognit.* **36**(1), 131–144 (2003).
30. N. Alajlan et al., "Fusion of supervised and unsupervised learning for improved classification of hyperspectral images," *Inf. Sci.* **217**(24), 39–55 (2012).
31. T. Kanungo et al., "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 881–892 (2002).
32. B. Foster et al., "Segmentation of PET Images for computer-aided functional quantification of tuberculosis in small animal models," *IEEE Trans. Biomed. Eng.* **61**(3), 711–724 (2014).
33. Z. Botev, J. Grotowski, and D. Kroese, "Kernel density estimation via diffusion," *Ann. Statist.* **38**(5), 2916–2957 (2010).
34. G. Sun et al., "Feature description of exact height function used in fast shape retrieval," *Opt. Precision Eng.* **25**(1), 224–235 (2017).
35. P. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. of the 2001 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR '01)*, pp. 511–518 (2001).
36. P. Dollár et al., "Fast feature pyramids for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(8), 1532–1545 (2014).
37. W. Liu et al., "SSD: single shot multibox detector," in *Proc. of the European Conf. on Comput. Vision*, pp. 21–37 (2016).
38. J. Dai et al., "R-FCN: object detection via region-based fully convolutional networks," arXiv:1605.06409 (2016).
39. S. Ren et al., "Faster R-CNN: towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, pp. 91–99 (2015).
40. Y. Jia et al., "Caffe: convolutional architecture for fast feature embedding," arXiv:1408.5093 (2014).

Guodong Sun is a full professor with the School of Mechanical Engineering, Hubei University of Technology. He received his BS and PhD degrees from Huazhong University of Science and Technology, Wuhan, China, in 2002 and 2008, respectively. He is the author of more than 20 journal papers. His current research interests include machine learning and imaging processing.

Yang Zhang received his BS degree and MS degree from Hubei University of Technology, Wuhan, China, in 2014 and 2017, respectively. Currently, he is pursuing the PhD with the Department of Computer Science, Nanjing University, Nanjing, China. His current research interests are machine learning and computer vision.

Hanbing Tang received his BS degree from Hubei University of Technology, Wuhan, China, in 2016. Currently, he is pursuing his MS degree with the School of Mechanical Engineering, Hubei University of Technology, Wuhan, China. His current research interests are machine learning and imaging processing.

Huiming Zhang received his BS degree from Northwestern Polytechnical University, Xi'an, China, in 2017. Currently, he is pursuing his MS degree with the Department of Computer Science, Nanjing University, Nanjing, China. His current research interests are machine learning and computer vision.

Moyun Liu is currently pursuing his BS degree with the School of Mechanical Engineering, Hubei University of Technology, Wuhan, China. His current research interests are machine vision and automatic control.

Daxing Zhao is a full professor with the School of Mechanical Engineering, Hubei University of Technology, Wuhan, China. He received his BS and MS degrees from Wuhan University of Technology, Wuhan, China, in 1983 and 2001, respectively, and his PhD from Northeastern University, ShenYang, China, in 2008. His current research interests include computer-aided design and imaging processing.