Frontiers of Information Technology & Electronic Engineering
www.zju.edu.cn/jzus; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Real-time monitoring of brake shoe keys in freight cars*

Rong ZOU[†1,2], Zhen-ying XU[1], Jin-yang LI[3], Fu-qiang ZHOU[2]

([1]*School of Mechanical Engineering, Jiangsu University, Zhenjiang 212013, China*)

([2]*MOE Key Laboratory of Precision Opto-Mechatronics Technology, Beihang University, Beijing 100191, China*)

([3]*MOE Key Laboratory of Modern Agricultural Equipment and Technology, Jiangsu University, Zhenjiang 212013, China*)

[†]E-mail: zr@ujs.edu.cn

Received Aug. 26, 2014; Revision accepted Nov. 15, 2014; Crosschecked Jan. 28, 2015

Abstract: Condition monitoring ensures the safety of freight railroad operations. With the development of machine vision technology, visual inspection has become a principal means of condition monitoring. The brake shoe key (BSK) is an important component in the brake system, and its absence will lead to serious accidents. This paper presents a novel method for automated visual inspection of the BSK condition in freight cars. BSK images are first acquired by hardware devices. The subsequent inspection process is divided into three stages: first, the region-of-interest (ROI) is segmented from the source image by an improved spatial pyramid matching scheme based on multi-scale census transform (MSCT). To localize the BSK in the ROI, census transform (CT) on gradient images is developed in the second stage. Then gradient encoding histogram (GEH) features and linear support vector machines (SVMs) are used to generate a BSK localization classifier. In the last stage, a condition classifier is trained by SVM, but the features are extracted from gray images. Finally, the ROI, BSK localization, and condition classifiers are cascaded to realize a completely automated inspection system. Experimental results show that the system achieves a correct inspection rate of 99.2% and a speed of 5 frames/s, which represents a good real-time performance and high recognition accuracy.

Key words: Condition monitoring, Feature expression, Brake shoe key, Machine vision
doi:10.1631/FITEE.1400305 **Document code:** A **CLC number:** TP277; U279.3

1 Introduction

The significance of employing advanced information technology for condition monitoring purposes in industries is highly appreciated. For a railway industry to successfully implement condition-based maintenance, effective non-destructive testing (NDT) methods which can predict or detect incipient faults in real time are required. NDT is a broad name describing a variety of methods and procedures concerned with the examination of all aspects of materials and structures such as uniformity, quality, and serviceability. Its fundamental principle is that the material being inspected will not be damaged in the inspection process. In NDT, image information is an important form of information (Rathod *et al.*, 2012). With the development of machine vision technology, visual inspection has gradually become one of the best methods for NDT. In this study, we focus on the visual inspection of the brake component condition.

Railway condition inspection routines are conducted periodically to ensure a proper condition of the structure or material (Márquez *et al.*, 2010; Oukhellou *et al.*, 2010; Zhou *et al.*, 2013; 2014). The traditional condition inspection process is performed by trained personnel who frequently searches for visual anomalies. This manual inspection is slow, laborious, and potentially hazardous, and the results are strictly dependent on the acuity, knowledge, and

* Project supported by the Special-Funded Programme on National Key Scientific Instruments and Equipment Development (No. 2012YQ140032), the National Natural Science Foundation of China (No. 51179076), the Jiangsu Province Postdoctoral Research Funding Plan (No. 1402012B), the Scientific Research Foundation of Jiangsu University for Senior Personnel (No. 14JDG134), and the Jiangsu Province Science and Technology Support Plan (Industrial) (No. BE2012149)

ORCID: Rong ZOU, http://orcid.org/0000-0002-2297-1348

endurance of qualified inspection personnel. Automated condition inspection technology in the transportation field can be divided into two categories: visual and non-visual methods. Of the non-visual methods, an acoustics-based fault diagnosis approach for inspecting roller bearings is undoubtedly one of the most successful methods (Anderson, 2007). However, the trackside acoustic detection system (TADS) is unable to detect other types of faults. With the development of machine vision technology, many visual inspection systems have been developed and applied to an increasing range of fault types. The vision-based condition inspection system developed by Marino *et al.* (2007) is used to detect automatically the absence of fastening bolts which secure rails to sleepers. The images of the rail are first obtained by a line scan camera. Then real-time wavelet transform of the images is used to extract the features. The converted data are put into a multi-layer perceptron neural network to achieve classification and identification (Mazzeo *et al.*, 2004). Zhang *et al.* (2011) used a structured light system and proposed a method based on dynamic structured light images to detect missing fastening components on high-speed railways. Artificial neural networks are also used to recognize the absence of fastening components (Zhang *et al.*, 2011). Yella *et al.* (2009) acquired sleeper images through a vision-based system and used a support vector machine (SVM) classifier to classify sleeper images. They achieved a classification accuracy of 90%. de Ruvo *et al.* (2009) presented a GPU-based vision system to recognize rail fastening elements. Their system can improve the performance of a quad-core CPU implementation by about 287%. Hart *et al.* (2008) proposed a multi-spectral visual trouble detection system, to detect the condition of a disc brake and its bearing performance. Kim and Kim (2011) used a binocular vision measurement system to measure accurately the thickness of the brake shoes of rolling stock.

In this paper, we propose a novel method for automated visual inspection of the brake shoe key (BSK) in freight cars. The BSK is a component of a freight car's braking system. Its absence will result in the break shoe falling off, which could cause the failure of the braking system and subject the freight car to the hazard of derailment. However, this issue has been seriously neglected for a long time. To the best of our knowledge, there are no publications on BSK condition inspection. The main contributions of this study are as follows:

1. An intelligent visual inspection system has been implemented to inspect BSK condition in freight cars, which substitutes for the traditional manual inspection. The system will effectively ensure the safety of freight cars and significantly decrease maintenance costs.

2. Variation in illumination is a challenging issue for practical visual inspection systems. The illumination insensitive features constructed by the gradient encoding histogram (GEH) are used to realize automated inspection of BSK condition in outdoor environments.

3. The proposed method is for BSK monitoring. If necessary, the system implemented can easily be adapted to other types of components.

4. A novel region-of-interest (ROI) segmentation algorithm is proposed to track the location of the target object, and provides the possibility for real-time and accurate condition inspection.

2 System overview

The entire automated visual inspection system is composed of a dynamic image gathering module, a data processing and transmitting station, and a detecting and analyzing center (Fig. 1). The data processing and transmitting station monitors the steel magnet group (C0–C10) at all times. When trains go through the steel magnet group, the signals gathered by various magnets are sent to the control room. The lighting sources and camera protection gates are then opened. High-speed cameras simultaneously capture dynamic images of the key parts of the moving freight cars. The acquired image data are firstly saved to the hard disks of local computers. Through the optic fiber network, images are transmitted to the remote monitoring servers, where they are analyzed by various terminals.

Fig. 2 illustrates the entire outdoor image acquisition device in the visual inspection system. BSK images of the moving freight cars are acquired by two high-speed CCD cameras with wide angle lenses (located on both sides of the box). As the angle and position of the cameras are fixed, the relative position

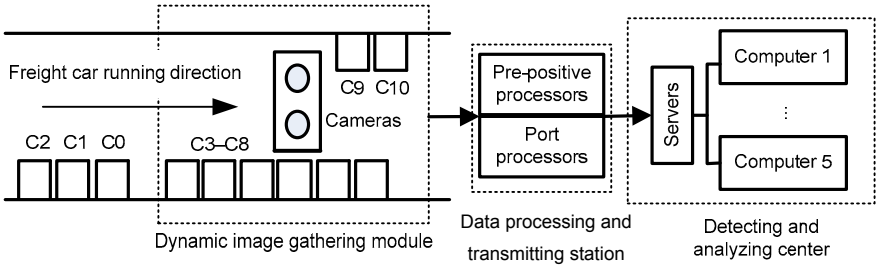

Fig. 1 Sketch of the system

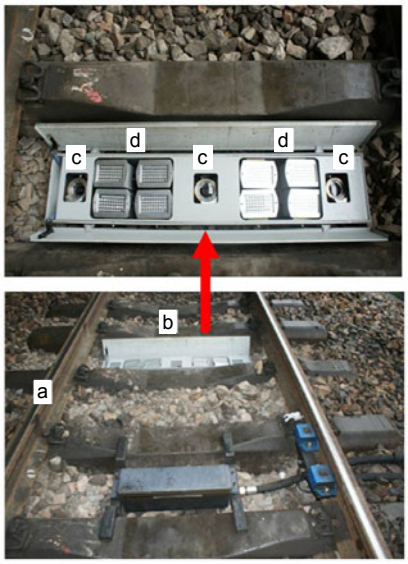

Fig. 2 The system used to acquire dynamic BSK images: a, railroad; b, bottom box; c, high-speed cameras; d, lighting sources

of the BSK is also fixed in the image. To adapt to outdoor environments, anti-sunlight cameras and strobe lights are used, and placed in the bottom box. The bottom box is installed in the gap between the railway sleepers, and fixed in the road bed surface. It uses a two-door opening and closing mode. Each door has its own state sensor for detecting the state of the door as open, closed, or abnormal. To keep the camera lens clean, a strong fan is mounted to the side of camera lens and keeps blowing. There is an alignment hole in the bottom of the bottom box which serves also as a drainage hole. In addition, active lighting and automated control of exposure time technologies are applied. However, due to the high-speed motion of freight cars and the extremely high dynamic range of illumination, BSK images are often underexposed or overexposed (Fig. 3) and their details are unclear.

After BSK images are acquired by the hardware devices, the inspection system should automatically segment the ROI and recognize the BSK condition in the ROI. Therefore, the performance of the system is strictly dependent on the performance of the ROI segmentation and BSK localization algorithms.

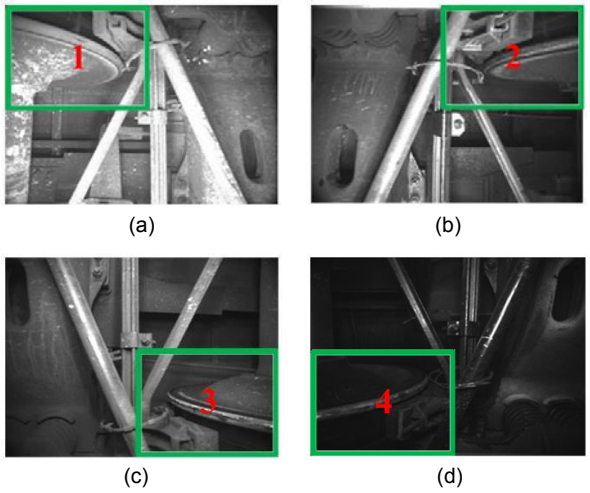

Fig. 3 Acquired BSK images

(a) First quadrant; (b) Second quadrant; (c) Third quadrant; (d) Fourth quadrant. Green rectangle: BSK region. References to color refer to the online version of this figure

3 Region-of-interest segmentation

In most intelligent inspection systems (Gualdi *et al.*, 2012; Milanés *et al.*, 2012), the first stage is to segment the ROI. Typically, a fixed ROI is determined according to specific content in images. The setting of the ROI decreases computation and the possibility of false localization. However, a fixed ROI does not apply to a change in target position. In Fig. 3, the BSK position is changing. If the images in Fig. 3 are divided into four regions of the same size,

namely quadrants 1, 2, 3, and 4, all four regions can become the ROI needed by the inspection system (Fig. 3). Taking into account that the quadrant region containing the BSK is significantly different from the other three quadrant regions in an image, a novel method of localizing and segmenting the ROI is proposed.

3.1 Feature expression for ROI localization

ROI localization needs to capture the global structural information and avoid excessive interference information from the texture. Furthermore, the ROI localization algorithm must be immune to variation in illumination. We propose multi-scale census transform (MSCT) solutions.

Census transform (CT) labels the pixels of an image by thresholding the 3×3 neighborhood of each pixel with the center value, and considers the result as a binary string or a decimal number. The conversion results reflect the structural properties of a sub-region. An original CT process is shown in Fig. 4.

MSCT is an extension of CT, which contains different sizes of neighborhoods and obtains more global structural information. In MSCT, the comparison operator between single pixels in CT is simply replaced with a comparison between average gray-values of sub-regions. Each sub-region is a square block containing neighboring pixels (or just one pixel in particular). The whole filter is composed of nine blocks (Fig. 5). We take the size s of the filter as a parameter, and $3s\times3s$ denotes the scale of the MSCT operator (a 3×3 MSCT is, in fact, the original CT). Note that the scalar values of averages over blocks can be computed very efficiently from the integral image. For this reason, MSCT can also be very fast: it incurs only a little more cost than the original 3×3 CT operator.

Fig. 6 gives examples of MSCT filtered ROI images by s=1, 3, 9. For a small scale, details of the local structure are well represented. It contains a large amount of useful discriminating information. Large scale MSCT filters use average values over the region. This is helpful to reduce noise and makes the representation more robust. Therefore, filters of various scales should be combined to achieve better performance.

In feature expression for ROI localization, denote $f_s(x, y)$ as an MSCT feature of scale $3s\times3s$ at location (x, y) computed from the original images. Then a histogram of the MSCT feature $f_s(\cdot, \cdot)$ over a certain image $I(x, y)$ can be defined as

$$H_s(\ell)=1_{\{f_s(x,y)=\ell\}},\quad \ell=0,1,\ldots,L-1, \qquad (1)$$

where 1_T is the indicator of set T, and ℓ is the label of the MSCT codes. Because all MSCT codes are 8-bit binary strings, there are a total of $L=2^8=256$ labels. Thus, the histogram has 256 bins.

The histogram of MSCT (HMSCT) reflects structural information in the image. Fig. 7 shows HMSCT features of four different quadrants in Fig. 3b, where s=1. Because each quadrant region contains different structures, the distribution of HMSCT also differs among quadrants.

HMSCT can encode global shape structure only in a small image patch. To capture the global structure of an image, we propose a spatial representation based on the spatial pyramid matching scheme (Lazebnik *et al.*, 2006). A spatial pyramid encodes a rough global structure of an image. The level 3 split in a spatial pyramid divides the image into $2^2\times2^2=16$ blocks (Fig. 8). To avoid artifacts created by non-overlapping division, the spatial pyramid structure has been improved by shifting the division (dash line blocks). This gives a total of 25 blocks in level 3. Similarly, level 2 has five blocks and level 1 has one.

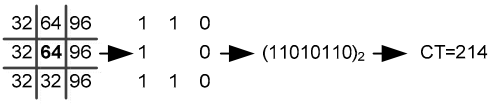

Fig. 4 Census transform

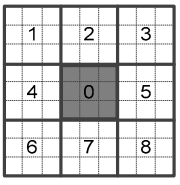

Fig. 5 Multi-scale census transform (MSCT) filter

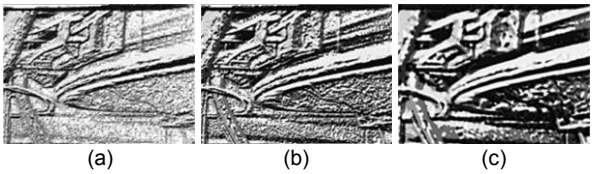

Fig. 6 MSCT filtered ROI images
(a) s=1; (b) s=3; (c) s=9

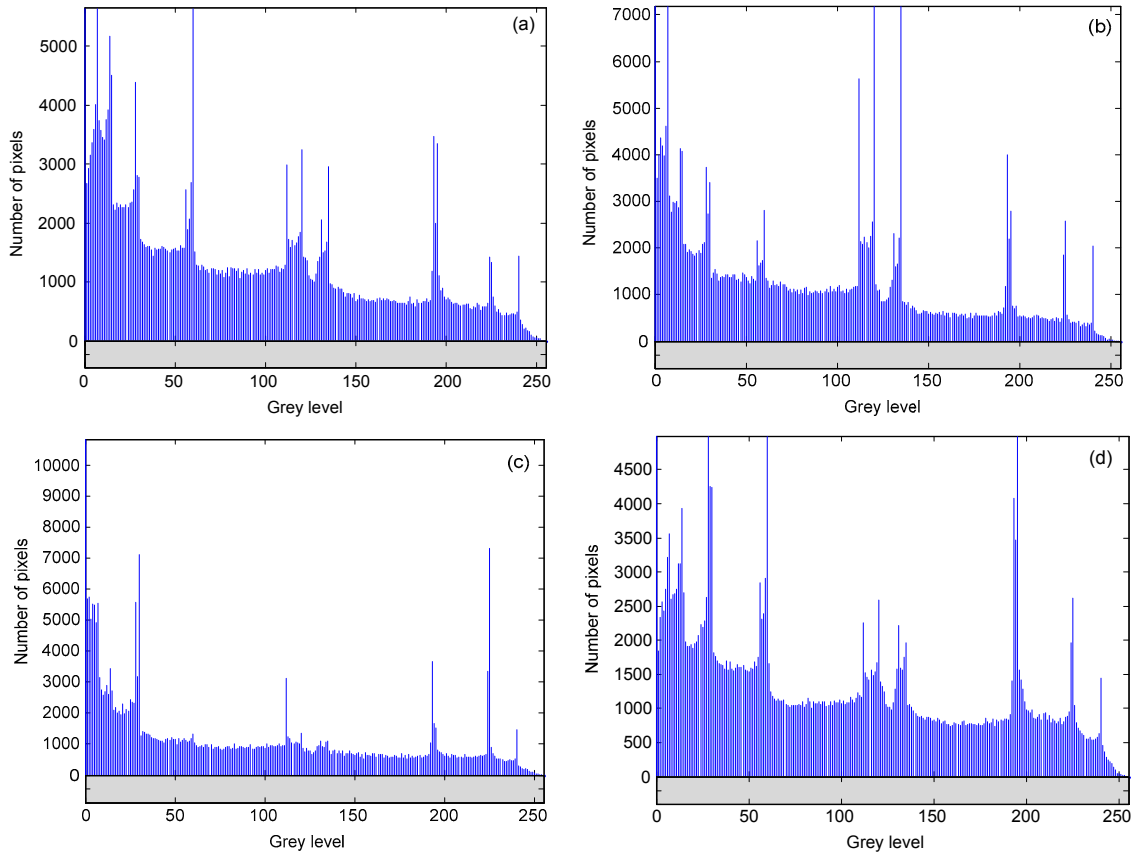

Fig. 7 Distribution of HMSCT in various quadrants of Fig. 3b

(a) First quadrant background; (b) Second quadrant BSK; (c) Third quadrant background; (d) Fourth quadrant background

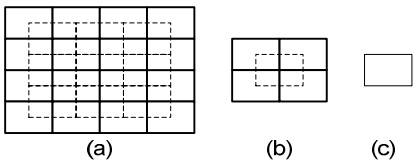

Fig. 8 Space pyramid structure

(a) Level 3; (b) Level 2; (c) Level 1. Dash line blocks represent the overlapping areas

The image is resized between different levels so that all blocks contain the same number of pixels. HMSCTs in all blocks are then concatenated to form an overall feature vector. To reduce computation, we use only a 2-level pyramid structure in practice.

3.2 Support vector machine classifier

The SVM classifiers were trained using the feature vectors generated by HMSCT. SVM is a learning algorithm based on small sample statistical theory and the maximum class interval (Guo *et al.*, 2012). It has advantages of theoretical completeness, short training time, and good generalization performance (Widodo and Yang, 2007). The SVM algorithm searches for the minimal VC-dimension of the system, given the training data. In the case of a two-class pattern recognition problem, the Vapnik-Chervonenkis (VC)-dimension reduction principle is implemented as a search for the optimal hyperplane (OHP) that separates the data. In brief, given training data $(\boldsymbol{x}_1, y_1)$, $(\boldsymbol{x}_2, y_2)$, …, $(\boldsymbol{x}_l, y_l)$, $\boldsymbol{x}_i \in \mathbb{R}^m$, $y_i \in \{-1, 1\}$ (i=1, 2, …, l), the OHP

$$\boldsymbol{\omega} \cdot \boldsymbol{x} + b_1 = 0 \tag{2}$$

can be found by the solution of the following optimization problem:

$$\min_{\boldsymbol{\omega}, b_1, \xi} \quad \frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}} \boldsymbol{\omega} + C\sum_{i=1}^{l} \xi_i \tag{3}$$

subject to

$$y_i(\boldsymbol{\omega}^{\mathrm{T}}\phi(\boldsymbol{x}_i)+b_1)\geq 1-\xi_i, \quad (4)$$

$$\xi_i\geq 0, \quad (5)$$

where $\boldsymbol{\omega}$, b_1, and ξ_i are the coefficients. Here, training vectors $\boldsymbol{x}_i$ are mapped into a higher dimensional space by function ϕ. SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space. $C>0$ is the penalty parameter of the error term. Furthermore,

$$K(\boldsymbol{x}_i,\boldsymbol{x}_j)\equiv\phi(\boldsymbol{x}_i)^{\mathrm{T}}\phi(\boldsymbol{x}_j) \quad (6)$$

is called the kernel function. In this study, linear SVM is used to ensure fast testing speed. Linear SVM adopts a linear kernel

$$K(\boldsymbol{x}_i,\boldsymbol{x}_j)\equiv\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j, \quad (7)$$

which is used to find a hyperplane parameter $\hat{\boldsymbol{\omega}}$ to minimize the sum of cost function and ℓ_2 regular item, namely

$$\hat{\boldsymbol{\omega}}=\arg\min_{\boldsymbol{\omega}}\left\{\frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}+C\sum_{i=1}^{N}\left(\max(1-y_i\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i,0)\right)^2\right\}, \quad (8)$$

where $\boldsymbol{X}\triangleq\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N\}$ is a set of feature vectors of training samples, N is the total number of trained samples, and $y\triangleq\{y_1,y_2,\cdots,y_N\}$ is the corresponding label of each training sample. The penalty parameter C, a parameter that can be adjusted by the user, can be used either to increase or decrease the penalty for classification errors.

SVMs were originally developed for binary classification. The ROI localization is a multi-class classification problem. Because the BSK appears only in four quadrants, it can be considered to be a 4-class classification problem (Fig. 9). We adopt the 1-against-all scheme. For a k-class problem, it constructs k binary SVMs, one for each class. Each class SVM is trained to separate its own data points from those of other classes (Kumar and Gopal, 2010). Due to the constraints on the location and number of ROIs in a BSK image, accurate ROI localization can be obtained without a large number of training samples. Finally, the quadrant region having the largest confidence level is labeled as the ROI and segmented. Fig. 10 gives an overview of the ROI segmentation architecture.

4 Brake shoe key localization in region-of-interest

Despite HMSCT features have been proved to be suitable for the ROI localization task, encoding features obtained from an original gray image cannot stand alone for object localization (Mu *et al.*, 2008). For BSK localization, a novel feature expression is built.

4.1 Feature expression for BSK localization

Recent psychology (Quinn *et al.*, 2001) and human vision (Elder and Velisavljević, 2009) studies

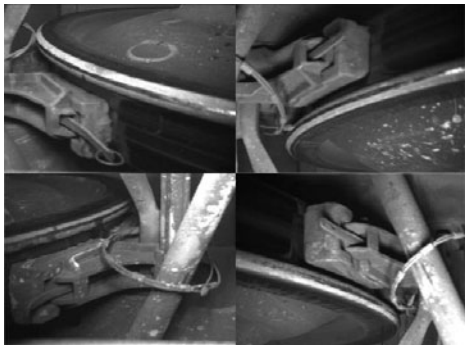

Fig. 9 BSK in four quadrants

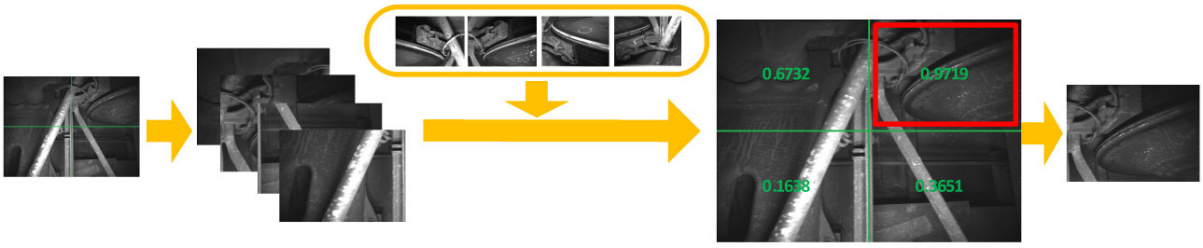

Fig. 10 ROI segmentation

have shown that humans apply shape information to distinguish object categories. For example, bikes or cups are recognized by their contour rather than by their color or texture. However, variation in illumination has a great influence on contour feature extraction. Histogram equalization (Lee *et al.*, 2012) and logarithmic transform (Cao *et al.*, 2012) are widely used to deal with the illumination problem. However, it is difficult for these image processing techniques to account for different lighting conditions. To solve the problem, we propose a solution of gradient domain. A gradient image may be regarded as an approximation of the edge image, which contains contour information of objects. In the gradient domain, the structure of the image is revealed. The huge success of histograms of oriented gradients (HOG) detectors (Dalal and Triggs, 2005) for pedestrian detection has confirmed this viewpoint. To enhance further the contour features of objects under illumination-variant conditions, the CT has been introduced into the gradient domain. Compared with gradient images, CT images have fewer image data (reduced to 8-bit data from the double-precision), and need not be normalized. The advantage of the gradient encoding method is that the contour can be stably depicted even under highly variable illumination conditions. Our novel illumination-insensitive feature descriptor is called the gradient encoding histogram (GEH). The GEH extraction algorithm is summarized in Algorithm 1 (Zhou *et al*., 2014).

The GEH extraction process of the BSK with different illuminations is shown in Fig. 11. Even in extreme lighting conditions, the contour features of the BSK can be reliably extracted by the GEH descriptor and the system has a good illumination-invariant performance. Furthermore, comparison of the generated gradient encoding image (Fig. 11c) with the original gray image (Fig. 11a), shows that the relative position of the BSK contour feature does not change through the expression of the GEH. Therefore, the GEH is not sensitive to deformation and texture changes in objects. Owing to the many excellent properties of the GEH, accurate and reliable BSK fault inspection becomes easier to achieve.

To localize the BSK, we use 128×128 pixels as the detection window size, which is a compact rectangle. The detection window is divided into 4×4 blocks (1024 pixels in each block) to describe accurately the structural features of the BSK in the detection window (Fig. 12). From these blocks, any adjacent 2×2 blocks are combined into one super-block, and can form a total of 3×3=9 super-blocks. The GEH of the BSK is extracted from each super-block. Thus, a 128×128 detection window contains 256×9 =2304 dimensional feature vectors (Fig. 12). Considering the requirement of a 3×3 pixel region for CT, the border of each detection window (one pixel width) is not included when the GEH descriptor is computed.

Algorithm 1 (Zhou *et al*., 2014) Gradient encoding histogram (GEH) extraction

Input: Image I

Output: GEH

Smoothen input image I by convolution with Gaussian kernel function

$$I_{\mathrm{G}}=I*G(x, y, \sigma), \tag{9}$$

where * is the convolution operator and

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{10}$$

is the Gaussian kernel function with standard deviation σ.

The smoothed image I_{G} is convolved with the derivative of the Gaussian kernel function in x and y directions to calculate the gradient

$$I_x=I_{\mathrm{G}}*G_x(x, y, \sigma), \tag{11}$$
$$I_y=I_{\mathrm{G}}*G_y(x, y, \sigma), \tag{12}$$

where $G_x(x, y, \sigma)$ and $G_y(x, y, \sigma)$ are the derivatives of the Gaussian kernel function in x and y directions, respectively.

For each pixel, compute the gradient magnitude I_{M}

$$I_{\mathrm{M}}=\sqrt{I_x^2+I_y^2}. \tag{13}$$

A threshold value is determined by halving the mean of the entire image gradient. To eliminate the influence of noise, gradient values less than the threshold become 0.

Compute census transform (CT) in the gradient image formed by the gradient amplitude

$$c(x,y)=\underset{(x',y')\in N}{\otimes}\varsigma(I_{\mathrm{M}}(x,y),I_{\mathrm{M}}(x',y')), \tag{14}$$

where $N(x, y)$ is a 3×3 local spatial neighborhood located at (x, y), $(x, y)\notin N(x, y)$, the value of the comparison function $\varsigma(I_{\mathrm{M}}(x, y), I_{\mathrm{M}}(x', y'))$ is 1 if $I_{\mathrm{M}}(x, y)<I_{\mathrm{M}}(x', y')$, and $\otimes$ is a concatenation operation.

Obtain the histogram of the CT image and form the expression of the GEH.

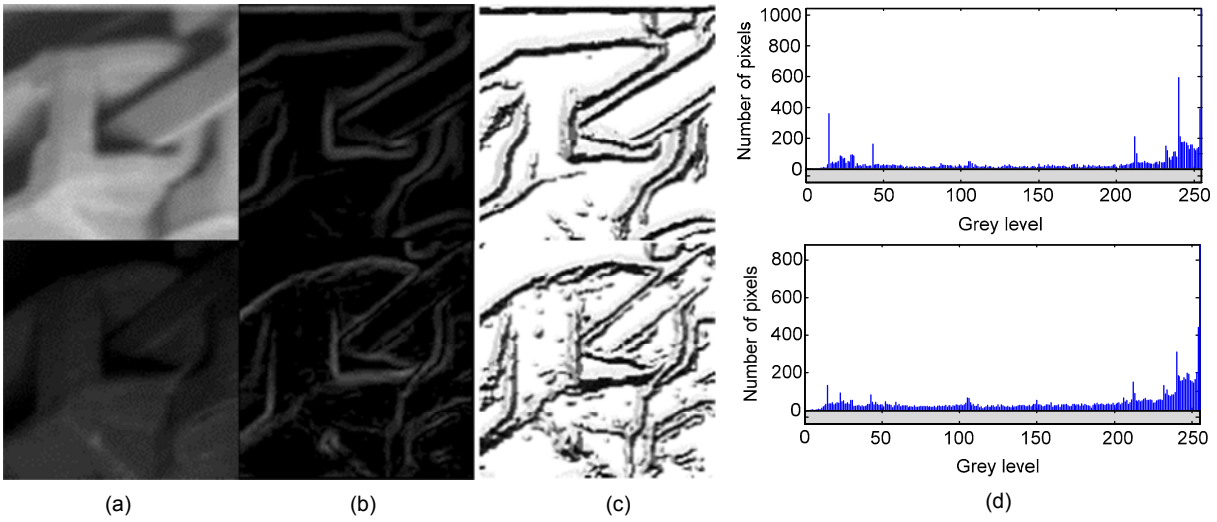

Fig. 11 Gradient encoding histogram extraction of BSK
(a) Original BSK images; (b) Gradient images; (c) Gradient encoding images; (d) Extracted GEH

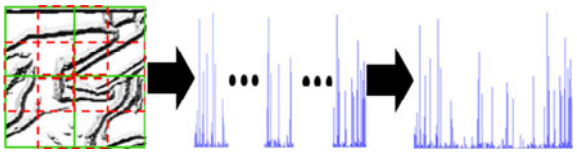

Fig. 12 Composition of GEH feature vectors for BSK

Because real-time inspection is essential in practical applications, integral histogram technology (Lampert *et al.*, 2009) is applied to speed up the calculation of the GEH. For each rectangular detection window $R(x, y, h, w)$ with upper-left corner coordinates (x, y), height h, and width w, the GEH can be formulated as

$$\boldsymbol{x}_{\mathrm{CT}}(R) \triangleq [h_{\mathrm{CT}}(R,0), h_{\mathrm{CT}}(R,1), \cdots, h_{\mathrm{CT}}(R,b-1)]^{\mathrm{T}}, \quad (15)$$

where $h_{\mathrm{CT}}(R, j)$ is the jth histogram in the detection window R, b is the total number of dimensions and b=256, and $h_{\mathrm{CT}}(R, j)$ is also formulated as

$$h_{\mathrm{CT}}(R,j)=II_{\mathrm{CT}}(x+w, y+h, j)+II_{\mathrm{CT}}(x, y, j) -(II_{\mathrm{CT}}(x+w, y, j)+II_{\mathrm{CT}}(x, y+h, j)), \quad (16)$$

where

$$II_{\mathrm{CT}}(x,y,j)=\sum_{y'=0}^{y'<y}\sum_{x'=0}^{x'<x} I_{\mathrm{CT}}(x',y',j), \quad (17)$$

and

$$I_{\mathrm{CT}}(x,y,j)=\begin{cases}1, & \mathrm{CT}(x,y)=j,\\ 0, & \text{otherwise.}\end{cases} \quad (18)$$

Unlike in conventional histogram computation, the integral histogram method does not repeat the histogram extraction for each possible region. The integral image makes the computation of the histogram use only three arithmetic operations and the time complexity becomes constant, namely $O(1)$, which greatly improves the operating speed of the algorithm and saves a lot of storage space.

4.2 Classifier construction

Accurate and reliable BSK localization is strictly dependent on effective training. In the training phase, the positive training sample set $\boldsymbol{P}$ is first established, in which the image size is 128×128 pixels (Fig. 11a). Simultaneously, the negative training sample set $\boldsymbol{N}$ is established, in which images do not contain the BSK. We randomly extract a 128×128 image size region from $\boldsymbol{N}$ and build a new negative sample set $\boldsymbol{N}_1$. Using $\boldsymbol{P}\cup\boldsymbol{N}_1$, we train a linear SVM classifier $\boldsymbol{H}_1$.

A bootstrap method is used in the training process to generate the final detector with a low false alarm rate. First, the linear SVM is solved to obtain the initial parameter $\hat{\boldsymbol{\omega}}_0^{\mathrm{T}}$ in $\boldsymbol{P}\cup\boldsymbol{N}_1$, which can generate classifier $\boldsymbol{H}_1$. 128×128 image regions are re-extracted randomly from the negative training set $\boldsymbol{N}$. They are then distinguished by classifier $\boldsymbol{H}_1$. Misclassified negative samples are collected to generate a new negative training set $\boldsymbol{N}_2$, and are added to set

$\boldsymbol{N}_1$. This expands set $\boldsymbol{N}_1$ to $\boldsymbol{N}_1 \cup \boldsymbol{N}_2$. After re-learning and re-training, we obtain a new mode parameter $\hat{\boldsymbol{\omega}}_1^{\mathrm{T}}$, and generate classifier $\boldsymbol{H}_2$. This process is repeated until the false alarm rate is lower than a fixed threshold or until the maximum number of repeats, k_1, is reached. The final model $\hat{\boldsymbol{\omega}}_{\mathrm{lin}}^{\mathrm{T}}$ is obtained and generates classifier $\boldsymbol{H}_{\mathrm{lin}}$. Finally, we apply a sliding window search method (Wojek *et al.*, 2008) to localize the BSK. The post-processing step is omitted to improve the detection speed because there is only one BSK in the ROI (Hoiem *et al.*, 2008). The region with maximum similarity is taken, which is considered to be the correct location of the BSK (Fig. 13).

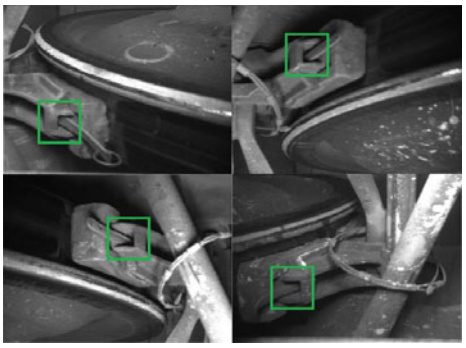

Fig. 13 Localization of BSK

5 Condition recognition

Unlike BSK localization, condition recognition is a specific fault discrimination process, which requires as much as possible to keep the discrimination information in the original image. Because encoding strategies based on gray images often achieve high recognition accuracy (Gu *et al.*, 2012), original BSK gray images are used in condition recognition rather than gradient images. We directly extract the smallest scale HMSCT feature from original gray images. The establishment of the feature vectors and fault classifier is similar to the above-mentioned BSK localization process. In the training phase, a positive training sample set $\boldsymbol{P}_{\mathrm{f}}$ is generated by 128×128 pixel images missing BSK (Fig. 14), and a negative training sample set $\boldsymbol{N}_{\mathrm{f}}$ is composed of BSK images of the same size (Fig. 15). Using $\boldsymbol{P}_{\mathrm{f}} \cup \boldsymbol{N}_{\mathrm{f}}$, we train a linear SVM classifier $\boldsymbol{H}_{f\mathrm{lin}}$. Finally, the localized BSK is verified by the fault classifier $\boldsymbol{H}_{f\mathrm{lin}}$. Therefore, in the inspection process, a classifier $\boldsymbol{H}_{\mathrm{lin}}$ for localizing the BSK region and a classifier $\boldsymbol{H}_{f\mathrm{lin}}$ for condition recognition are cascaded to discriminate accurately the BSK condition (Fig. 16). This is intended to avoid, as much as possible, the occurrence of errors and omissions, and ensure safe and efficient railroad operation.

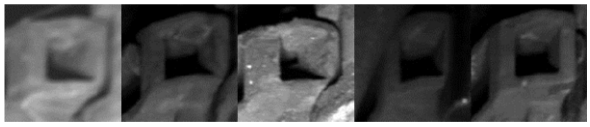

Fig. 14 Sample images in $\boldsymbol{P}_{\mathrm{f}}$

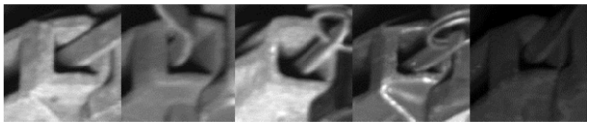

Fig. 15 Sample images in $\boldsymbol{N}_{\mathrm{f}}$

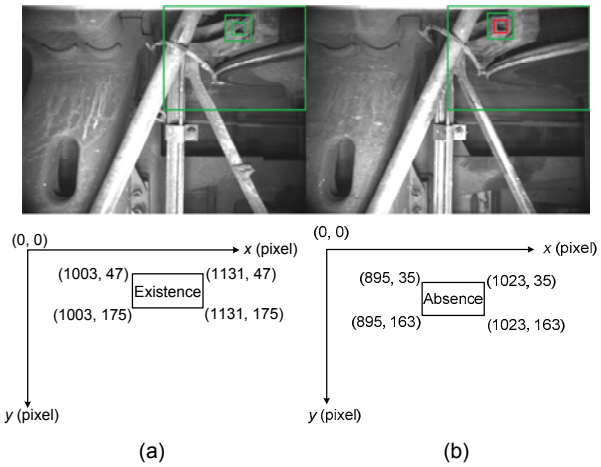

Fig. 16 Inspection of the BSK condition

(a) Existence; (b) Absence. Small red box marks the absence condition. References to color refer to the online version of this figure

6 Experimental results

We used a DALSA HM1400 high-speed digital CCD camera to acquire BSK images in our system (Fig. 2). The minimum exposure time of this camera is 4.7 μs. The highest frame rate is up to 64 frames/s in full resolution of 1400×1024 pixels. The camera is suitable for use with freight cars with a top speed of 120 km/h. A lens with a fixed focal length of 6 mm was also chosen. To reduce the impact of variation

on natural illumination, four compensation light source groups were installed in the bottom box.

6.1 Classifier training

To verify the performance of the visual inspection system for detecting a missing BSK fault, a total of 12 800 BSK images from 32 freight cars were collected in different places and under different illumination conditions, creating a database of BSK images (the BSK database), in which each class contained 3200 images. Some sample images from the BSK database are shown in Fig. 17.

For training the multi-class SVM classifier, 2400 sample images were extracted from the BSK database. There were 600 images in each class. Three hundred images from each class were used for training and the remaining 300 for testing. For the training sample set, we further extracted the quadrant region with 700×512 pixels containing the BSK, and established an ROI training set. As described earlier, the 1-against-all scheme was adopted to build the ROI classifier, and four binary SVMs were constructed. Taking into account the memory footprint and real-time requirements in practical applications, linear SVM was used. Note that the training images were scaled to 175×128 pixels to reduce computation. The generated ROI classifier is shown in Fig. 18.

Once the ROI region was segmented, the shape of the BSK was determined, and the corresponding BSK localization classifier was used. Considering the subsequent fault inspection process is identical in all ROIs, for ease of description, the effectiveness of

Fig. 17 Some sample images in the BSK database

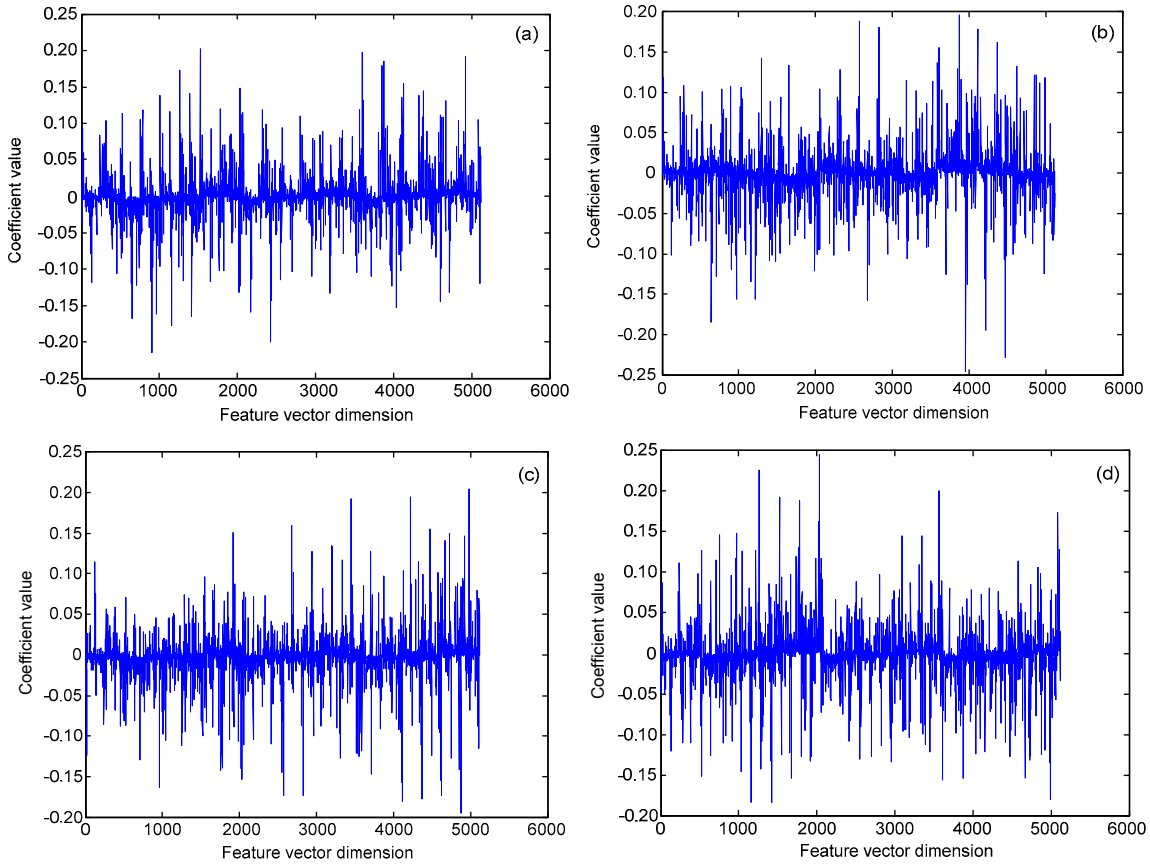

Fig. 18 Region-of-interest classifiers

(a) First quadrant; (b) Second quadrant; (c) Third quadrant; (d) Fourth quadrant

our method is proved by considering only the second quadrant region containing the BSK in the subsequent experimental section.

Similarly, 3200 sample images were obtained from the BSK database, and a training set and validation set were built, each containing 1600 sample images. To train classifier $\boldsymbol{H}_{\text{lin}}$, as described earlier, the size of the image patches was $n \times m$ pixels in the positive sample set containing 1600 image patches, where n and m depend on the size of the target object in the images. For the BSK, each sample was manually cut to 128×128 pixels, to contain a BSK region as compact as possible (Fig. 11a). In the negative sample set, 3700 background image patches without the BSK were randomly extracted from the training set, and the size of each image patch was also 128×128 pixels. The classifier $\boldsymbol{H}_{\text{lin}}$ was trained by the bootstrap method, which constantly updates the negative training set by introducing new negative samples. The bootstrap process was performed three times until the number of negative samples was 4000. The training process of classifier $\boldsymbol{H}_{f\text{lin}}$ was similar to that of classifier $\boldsymbol{H}_{\text{lin}}$. However, for classifier $\boldsymbol{H}_{f\text{lin}}$, the negative sample set contained 1300 existing BSK images with 128×128 pixels (Fig. 15), and the positive sample set contained 300 images missing the BSK with 128×128 pixels (Fig. 14). The generated classifiers $\boldsymbol{H}_{f\text{lin}}$ and $\boldsymbol{H}_{\text{lin}}$ are shown in Fig. 19. They consist of 2304-dimensional feature vectors.

6.2 Detection results

The performance of the classifier and inspection system was evaluated using the validation set. The validation set contained 1100 fault-free images and 500 fault images. For practical purposes, the speed and accuracy of the system are always important. So, our experiments focused on the following two aspects:

1. Detection speed

Real-time processing is a necessary property in most inspection systems. In our inspection system the inspection speed is very high. For 1400×1024 images, its speed is up to 5 frames/s (2.93 GHz, Intel Core i7-870, 4 GB RAM, and Win7 OS). The time consumption of the various main components of the algorithm is shown in Table 1.

As mentioned earlier, one of the reasons that inspection is fast is that we adopt a real-time ROI segmentation algorithm, which enables inspection to be implemented in a small ROI. In the ROI localization phase, we filtered 75% of candidate image regions and only 25% were needed for verification in the follow-on phase. The integral histogram and the absence of post-processing steps are also very important factors.

2. Detection accuracy

In Fig. 20, the recognition rates using the ROI validation set are given for six parameter configurations.

Table 1 Computing time distribution statistics

Main module	Time consumption (ms)	Percentage
ROI segmentation	90	48.1%
BSK localization		
Gradient image	15	8.0%
Integral image	16	8.6%
CT computing	31	16.6%
Slide window search	16	8.6%
Condition recognition	19	10.2%

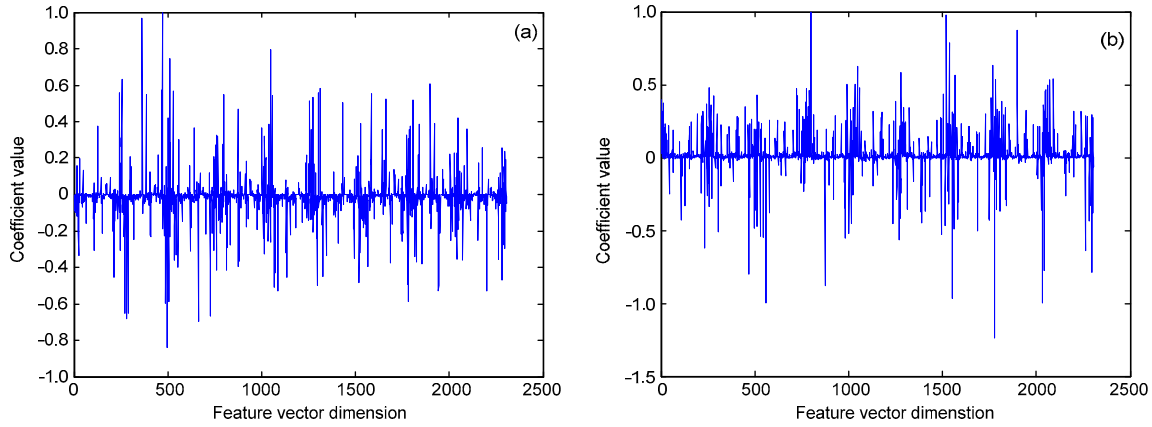

Fig. 19 Classifier $\boldsymbol{H}_{\text{lin}}$ of localizing BSK (a) and classifier $\boldsymbol{H}_{f\text{lin}}$ of condition recognition (b)

Applying a spatial pyramid matching scheme greatly improves the system performance (Level>0 vs. Level =0), while the multi-scale strategy enables the recognition rate to reach 100% after Level>0. Although the recognition rate of Level=1 is the same as that of Level=2, they have a time difference of at least 20 ms. In the actual system, we used the following parameters for the spatial pyramid: s=[1, 3], L=1.

The most straightforward way to evaluate the performance of multi-class classifiers is to use confusion matrix analysis. A confusion matrix from one run on the ROI validation set (s=[1, 3], L=1) is shown in Fig. 21. Our ROI classifier achieved an accuracy of 100%, and exhibited excellent performance.

The subsequent BSK localization is also essentially an object detection process. Generally, the detection accuracy of the metrics is represented by the receiver operating characteristic (ROC) curve (Milanés *et al.*, 2012). The curve is obtained by adjusting the threshold of the classifier. It expresses the correctly classified results for the positive samples and the misclassified results for all negative samples. An ideal decision variable would have an ROC curve that passed through (1, 1), which would correspond to a point in the top-left corner of the ROC axis. The ROC curve for the localization classifier $\boldsymbol{H}_{\text{lin}}$ in the validation set is shown in Fig. 22. The true detection rate had reached 100% when the false positive rate was more than 0.1. This indicates that classifier $\boldsymbol{H}_{\text{lin}}$ has very good classification performance. It also confirms the effectiveness of the training approach used by the classifier. Fig. 23 also shows detection error tradeoff (DET) curves of classifier $\boldsymbol{H}_{\text{lin}}$. DET curves can describe the missing rate versus false positives per window (FPPW). These terms are defined as follows:

$$\text{Missing rate}=\frac{\text{Num}_{\text{false}}}{\text{Num}_{\text{detPos}}+\text{Num}_{\text{false}}}, \tag{19}$$

$$\text{FPPW}=\frac{\text{Num}_{\text{false}}}{\text{Num}_{\text{totNeg}}}, \tag{20}$$

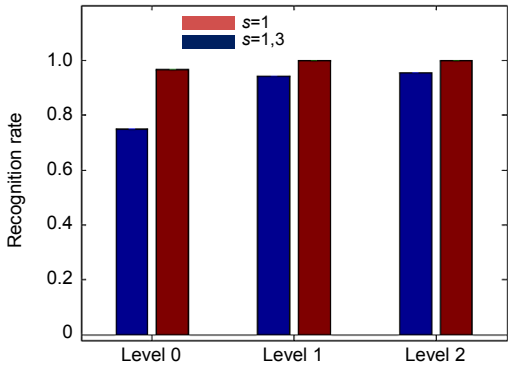

Fig. 20 ROI recognition rate

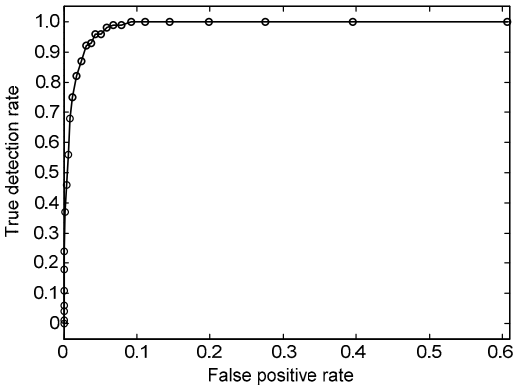

Fig. 22 ROC curve for classifier H_{lin}

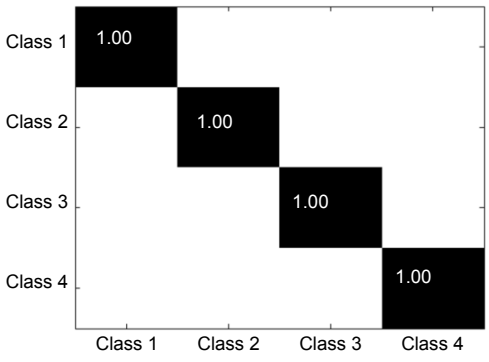

Fig. 21 Four categories confusion matrix

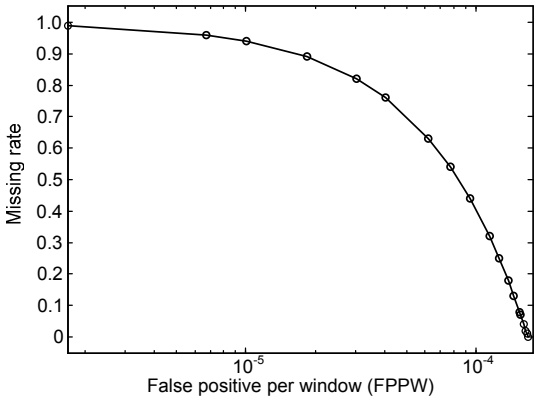

Fig. 23 FPPW results for classifier H_{lin}

where $\mathrm{Num}_{\mathrm{false}}$ is the number of false alarms, $\mathrm{Num}_{\mathrm{detPos}}$ is the number of detected positives, and $\mathrm{Num}_{\mathrm{totNeg}}$ is the total number of testing negative examples.

In Fig. 23, it is obvious that a low missing rate together with a low FPPW is favorable. When the false positive rate was about 1.68×10^{-4}, the missing rate was 0. This confirms the good performance of classifier $\boldsymbol{H}_{\mathrm{lin}}$.

Because the condition classifier $\boldsymbol{H}_{f\mathrm{lin}}$ verifies only a small BSK region in the images, we give evaluation results for classifier $\boldsymbol{H}_{f\mathrm{lin}}$ to intuitively reflect its performance. The average of evaluation results for a missing BSK was 1.64 and that for an intact BSK was −1.85 (Fig. 24). At the critical value 0 of the classification, they all have a large surplus. This indicates that classifier $\boldsymbol{H}_{f\mathrm{lin}}$ has a good classification performance for the BSK condition.

Table 2 shows the result statistics of BSK condition inspection in three different places for live testing. The hierarchical cascade detection framework, which includes the ROI classifier, the localization classifier $\boldsymbol{H}_{\mathrm{lin}}$, and the condition classifier $\boldsymbol{H}_{f\mathrm{lin}}$, effectively improved the inspection accuracy. For 5000 live BSK images in each place, our system had correct inspection rates above 99.2%, confirming the system's very high inspection reliability. Usually, visual inspection systems have an accuracy of 85%–95% (Marino *et al.*, 2007; Yella *et al.*, 2009; Zhou *et al.*, 2013). Under conditions of variable illumination, our system can achieve an accuracy rate of 99.2%, which is fully able to meet actual requirements.

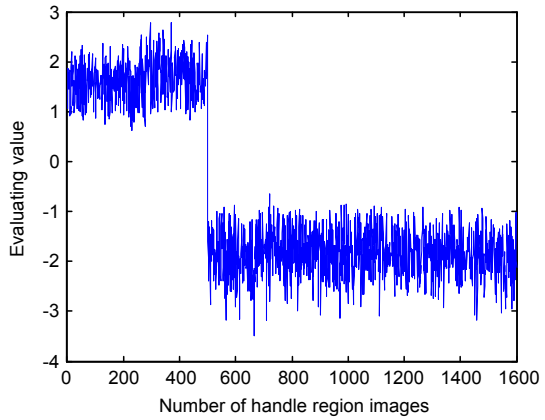

Fig. 24 Evaluation results for the classifier

Table 2 Fault inspection result statistics

Place	Number of correct inspections	Number of false inspections	Correct inspection rate
Wuhan	4962	38	99.2%
Shenyang	4970	30	99.4%
Beijing	4974	26	99.5%

Note: the total number of images for all the three places is 5000

7 Conclusions

This article presents a system for condition monitoring using a visual system. For missing BSK faults which could lead to serious accidents, we propose a novel real-time and accurate visual inspection method. Localizing the ROI quickly and accurately is the first task in practical inspection systems. To overcome illumination variation, we present an HMSCT feature descriptor and adopt an improved space pyramids scheme to realize reliable localization of the ROI. In the BSK localization stage, we show that the contour is the most important information source for object localization. To extract contour features effectively under illumination-variant conditions, a novel GEH feature descriptor based on the gradient field is proposed, which efficiently expresses the contour information of the BSK. Finally, in the condition recognition stage, a novel feature descriptor based on gray images is adopted. The whole system achieves the inspection of the BSK condition in a cascade framework. Experimental data proved that the proposed system has high speed and accuracy.

Another major contribution is the realization of a completely automated visual inspection. At present, in the railway transportation field many visual inspection systems are actually semi-automatic, and require some degree of manual operation. For example, in the system proposed by Yella *et al*. (2009) the detection region must be manually defined. Although satisfactory experimental results were obtained, our method may be further improved in some ways, e.g., by using multiple information sources.

References

Anderson, G.B., 2007. Acoustic detection of distressed freight car roller bearings. Proc. ASME/IEEE Joint Rail Conf. and Int. Combustion Engineer Division Spring Technical Conf., p.167-171. [doi:10.1115/JRC/ICE2007-40091]

Cao, X., Shen, W., Yu, L., *et al.*, 2012. Illumination invariant

extraction for face recognition using neighboring wavelet coefficients. *Patt. Recog.*, **45**(4):1299-1305. [doi:10.1016/j.patcog.2011.09.010]

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, p.886-893. [doi:10.1109/CVPR.2005.177]

de Ruvo, P., Distante, A., Stella, E., *et al*., 2009. A GPU-based vision system for real time detection of fastening elements in railway inspection. Proc. 16th IEEE Int. Conf. on Image Processing, p.2333-2336. [doi:10.1109/ICIP.2009.5414438]

Elder, J.H., Velisavljević, L., 2009. Cue dynamics underlying rapid detection of animals in natural scenes. *J. Vis.*, **9**(7):7.1-7.20. [doi:10.1167/9.7.7]

Gu, W., Xiang, C., Venkatesh, Y.V., *et al.*, 2012. Facial expression recognition using radial encoding of local Gabor features and classifier synthesis. *Patt. Recog*., **45**(1):80-91. [doi:10.1016/j.patcog.2011.05.006]

Gualdi, G., Prati, A., Cucchiara, R., 2012. Multistage particle windows for fast and accurate object detection. *IEEE Trans. Patt. Anal. Mach. Intell.*, **34**(8):1589-1604. [doi:10.1109/TPAMI.2011.247]

Guo, L., Ge, P.S., Zhang, M.H., *et al.*, 2012. Pedestrian detection for intelligent transportation systems combining AdaBoost algorithm and support vector machine. *Expert Syst. Appl.*, **39**(4):4274-4286. [doi:10.1016/j.eswa.2011.09.106]

Hart, J.M., Resendiz, E., Freid, B., *et al.*, 2008. Machine vision using multi-spectral imaging for undercarriage inspection of railroad equipment. Proc. 8th World Congress on Railway Research, p.1-8.

Hoiem, D., Efros, A.A., Hebert, M., 2008. Putting objects in perspective. *Int. J. Comput. Vis.*, **80**(1):3-15. [doi:10.1007/s11263-008-0137-5]

Kim, H., Kim, W., 2011. Automated inspection system for rolling stock brake shoes. *IEEE Trans. Instrum. Meas.*, **60**(8):2835-2847. [doi:10.1109/TIM.2011.2119110]

Kumar, M.A., Gopal, M., 2010. A comparison study on multiple binary-class SVM methods for unilabel text categorization. *Patt. Recog. Lett.*, **31**(11):1437-1444. [doi:10.1016/j.patrec.2010.02.015]

Lampert, C.H., Blaschko, M.B., Hofmann, T., 2009. Efficient subwindow search: a branch and bound framework for object localization. *IEEE Trans. Patt. Anal. Mach. Intell.*, **31**(12):2129-2142. [doi:10.1109/TPAMI.2009.144]

Lazebnik, S., Schmid, C., Ponce, J., 2006. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, p.2169-2178. [doi:10.1109/CVPR.2006.68]

Lee, P.H., Wu, S.W., Hung, Y.P., 2012. Illumination compensation using oriented local histogram equalization and its application to face recognition. *IEEE Trans. Image Process.*, **21**(9):4280-4289. [doi:10.1109/TIP.2012.2202670]

Marino, F., Distante, A., Mazzeo, P.L., *et al.*, 2007. A real-time visual inspection system for railway maintenance: automatic hexagonal-headed bolts detection. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, **37**(3):418-428. [doi:10.1109/TSMCC.2007.893278]

Márquez, F.P.G., Roberts, C., Tobias, A.M., 2010. Railway point mechanisms: condition monitoring and fault detection. *Proc. Inst. Mech. Eng. Part F: J. Rail Rapid Transit*, **224**(1):35-44. [doi:10.1243/09544097JRRT289]

Mazzeo, P.L., Nitti, M., Stella, E., *et al.*, 2004. Visual recognition of fastening bolts for railroad maintenance. *Patt. Recog. Lett.*, **25**(6):669-677. [doi:10.1016/j.patrec.2004.01.008]

Milanés, V., Llorca, D.F., Villagrá, J., *et al.*, 2012. Vision-based active safety system for automatic stopping. *Expert Syst. Appl.*, **39**(12):11234-11242. [doi:10.1016/j.eswa.2012.03.047]

Mu, Y., Yan, S., Liu, Y., *et al*., 2008. Discriminative local binary patterns for human detection in personal album. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, p.1-8. [doi:10.1109/CVPR.2008.4587800]

Oukhellou, L., Debiolles, A., Denœux, T., *et al.*, 2010. Fault diagnosis in railway track circuits using Dempster-Shafer classifier fusion. *Eng. Appl. Artif. Intell.*, **23**(1):117-128. [doi:10.1016/j.engappai.2009.06.005]

Quinn, P.C., Eimas, P.D., Tarr, M.J., 2001. Perceptual categorization of cat and dog silhouettes by 3- to 4-month-old infants. *J. Exp. Child Psychol.*, **79**(1):78-94. [doi:10.1006/jecp.2000.2609]

Rathod, V.R., Anand, R.S., Ashok, A., 2012. Comparative analysis of NDE techniques with image processing. *Nondestruct. Test. Eval.*, **27**(4):305-326. [doi:10.1080/10589759.2011.645820]

Widodo, A., Yang, B.S., 2007. Support vector machine in machine condition monitoring and fault diagnosis. *Mech. Syst. Signal Process.*, **21**(6):2560-2574. [doi:10.1016/j.ymssp.2006.12.007]

Wojek, C., Dorkó, G., Schulz, A., *et al.*, 2008. Sliding-windows for rapid object class localization: a parallel technique. Proc. 30th DAGM Symp. on Pattern Recognition, p.71-81. [doi:10.1007/978-3-540-69321-5_8]

Yella, S., Dougherty, M., Gupta, N.K., 2009. Condition monitoring of wooden railway sleepers. *Transpo. Res. Part C Emerg. Technol.*, **17**(1):38-55. [doi:10.1016/j.trc.2008.06.002]

Zhang, H., Yang, J., Tao, W., *et al.*, 2011. Vision method of inspecting missing fastening components in high-speed railway. *Appl. Opt.*, **50**(20):3658-3665. [doi:10.1364/AO.50.003658]

Zhou, F.Q., Zou, R., Gao, H., 2013. Dust collector localization in trouble of moving freight car detection system. *J. Zhejiang Univ.-Sci. C (Comput. & Electron.)*, **14**(2):98-106. [doi:10.1631/jzus.C1200223]

Zhou, F.Q., Zou, R., Qiu, Y., *et al*., 2014. Automated visual inspection of angle cocks during train operation. *Proc. Instit. Mech. Eng. Part F J. Rail Rapid Transit*, **228**(7):794-806. [doi:10.1177/0954409713495532]