IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 29, 2019, accepted February 13, 2019, date of publication February 19, 2019, date of current version March 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2900056

A Fast Bolt-Loosening Detection Method of Running Train's Key Components Based on Binocular Vision

JUNHUA SUN, (Member, IEEE), YANXIA XIE, AND XIAOQI CHENG
Key Laboratory of Precision Opto-Mechatronics Technology, Ministry of Education, Beihang University, Beijing 100191, China
Corresponding author: Junhua Sun (sjh@buaa.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61275162.

ABSTRACT Bolt-loosening can cause poor running quality of trains, even resulting in terrible accidents. Currently, existing bolt-loosening detection methods for running trains need 3D data of the whole train body, extremely decreasing the efficiency of fault detection. In this paper, we propose a fast bolt-loosening detection method for the running train's key components based on binocular vision. Since a train generally consists of many cars with the same structure, the position distribution of train's key components is regular and periodic. First, we propose a novel method to detect key component regions including bolts, taking full advantage of this periodic distribution rule. Second, the sub-pixel edges of the bolt cap and mounting surface in the localized regions are extracted and segmented, respectively, combining with the convolutional neural network (CNN). Finally, based on stereo matching and the binocular vision model, the 3D data of these edges are obtained to calculate the distance between the bolt cap and mounting surface. By comparing the calculated distance with the reference value, we can judge whether bolt-loosening has occurred. The experimental results indicate that multi-bolt looseness can be calculated simultaneously. The measurement repeatability and precision are superior to 0.03 and 0.08 mm, respectively, and the relative error is less than 1.42%.

INDEX TERMS Binocular stereo vision, bolt loosening, convolutional neural network, edge extraction.

I. INTRODUCTION

Faults of running trains can be generally divided into two categories: visible faults and hidden faults. Visible faults mainly include missing [1], [2], displacement [3] and deformation of larger parts. Hidden faults consist of various gradual faults, such as loose wire breakage, anti-loose mark ectopic and bolt-loosening [4]. Since bolt is an important component of the structural support, bolt-loosening without timely maintenance may bring about security risk. For example, brake power and transmission force would result in bolt's absence or breaking, causing serious traffic accidents. Therefore, it is necessary to develop a fast bolt-loosening detection method, for the safety of running trains.

At present, manual inspection based on Trouble of Running Freight Train Detection System (TFDS) [5], [6] is still the prevailing method for fault detection of running trains. TFDS captures key components' images of the running train firstly. And then, inspectors indoor determine whether faults exist by observing the images, consuming a lot of time and manpower. Aiming to avoid disadvantages of manual inspection, some methods based on features of 2D images [7]–[12] were introduced. Although these approaches could recognize certain visible faults, it cannot detect bolt-loosening due to lack of depth information.

In other fields, such as aerospace structure [13] and civil structure [14], several vision-based methods have been proposed to detect bolt-loosening. Kong and Lin [15] detected bolt loosening of steel joints, by eliminating misalignments between two images collected during different inspection periods. However, slight changes of lighting conditions and camera poses would have negative effect on the performance of bolt-loosening detection. For bridge system, Cha *et al.* first used the Hough transform and other image processing techniques to obtain simple damage-sensitive features. And then a liner support vector machine (LSVM) was trained to differentiate tight bolts from loose bolts [16]. Ramana *et al.* first used the Cascade detector to localize all bolts on images.

The associate editor coordinating the review of this manuscript and approving it for publication was Choon Ki Ahn.

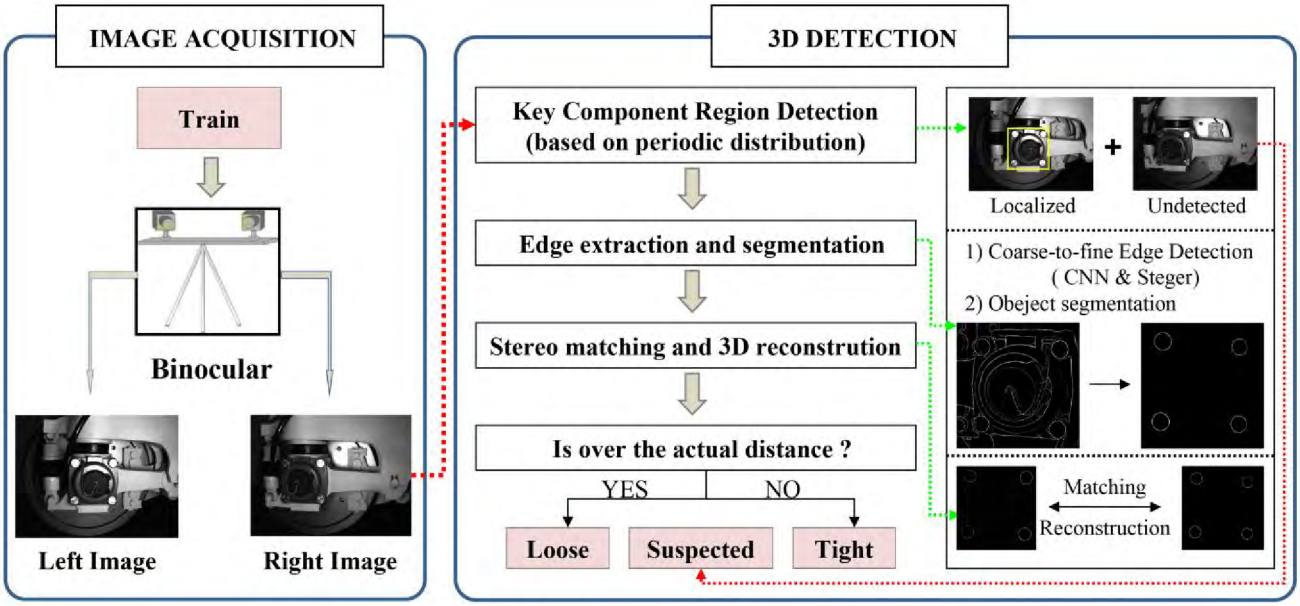

FIGURE 1. Principle of the bolt-loosening detection.

Then a LSVM is trained to separate loosened and tight structural bolts, by feeding into the bolt head dimensions and exposed shank length features [17]. Since bolt head dimensions and exposed shank length are calculated in 2D image plane rather than 3D space, the calculated values will be different once the angle and distance of the camera are changed. Namely, these methods can only detect distinctively-loosened bolts in condition of limited camera angles and distances [16], [17]. Compared with the static structure (such as bridge system and steel joints), trains have more complex structures and higher running speed. Therefore, adopting these methods above directly to detect bolt-loosening for running trains is not practical.

Recently, some inspection equipment manufacturers focused on the exploratory study about 3D detection of hidden faults such as bolt-loosening [18], [19]. They customized 3D vision sensors to obtain 3D data by scanning the whole train body, completing the fault recognition directly through 3D information. However, on one hand, these 3D data contain lots of useless information, extremely reducing the detection efficiency. On the other hand, since the train running isn't a rigorous one-dimensional motion, the collected 3D data would carry larger errors. In our previous work [20], we firstly localized target regions by a CNN-based target detection model. Then stereo matching and 3D reconstruction were performed to detect bolt-loosening faults. However, this method cannot judge whether a bolt region is omitted, laying a potential risk to the safety of running trains.

On basis of our previous work [20], a novel bolt-loosening detection method is proposed in this paper. When the train wheels pass the magnetic steels, the binocular stereo vision sensor is triggered to collect image sequence. Since a train generally consists of many cars with same structure, the position distribution of key component regions in the image sequence has a periodicity. With the periodic distribution rule, the current detection method has significant improvement over the previously proposed: 1) The omission rates of key components including bolts are reduced; 2) All undetected regions are outputted as suspected visible fault regions to guarantee the safety of running trains. Afterwards, the edges of bolt caps and mounting surface in the localized regions are extracted and segmented based on CNN. Then, stereo matching and 3D reconstruction is performed to obtain the 3D data of segmented edges. Finally, the calculated distance between the bolt cap and the mounting surface is compared with its reference distance to determine whether the bolts are loose or not, thus increasing the efficiency of bolt-loosening detection.

The rest of this paper is organized as follows: section II is an overview of our proposed method; the key methods involved are elaborated in section III; in section IV, the measurement system is described and the experimental results of bolt-loosening are given; finally, section V concludes this paper.

II. OVERVIEW OF THE BOLT-LOOSENING DETECTION METHOD

As shown in Fig. 1, our method is composed of two main modules: images acquisition and 3D detection of faults. The images acquisition module is designed to collect image sequence of running trains. When the train is passing, the binocular stereo vision sensor is triggered to acquire the left and right images of key components, which are sent to the 3D detection module at once.

The 3D detection module is used to measure bolts' looseness. Three key tasks are involved in this module:

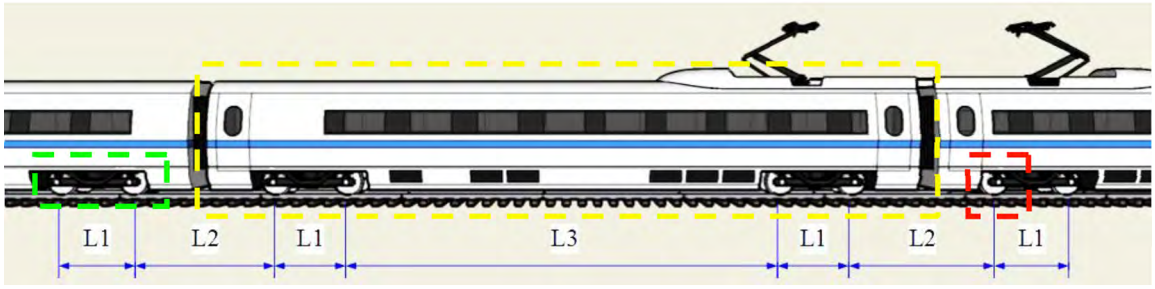

FIGURE 2. The train structure and its wheelbase distribution. The car, part and component are in yellow, green, red box, separately.

(i) key component regions detection, (ii) edge extraction and segmentation, and (iii) stereo matching and 3D reconstruction. Each task is a sub-task of the next. Correspondingly, three key methods are proposed as follows:

1) *KEY COMPONENT REGIONS DETECTION METHOD BASED ON THE PERIODIC DISTRIBUTION RULE*

This method is proposed to detect key component regions including bolts and to output all undetected regions as suspected visible fault regions.

2)*EDGE EXTRACTION AND SEGMENTATION*

This method uses a "coarse-to-fine" scheme to extract sub-pixel edges of bolt caps and their mounting surface in the localized regions, then segments them based on a novel shape feature descriptor.

3) *STEREO MATCHING AND 3D RECONSTRUCTION*

This method is introduced to calculate the distance between bolt caps and their mounting surface, based on epipolar constraint and binocular vision model. By comparing the calculated distance with reference value, we can judge whether the bolt is loose.

III. METHOD

A. KEY COMPONENT REGIONS DETECTION METHOD BASED ON THE PERIODIC DISTRIBUTION RULE

If the bolt-loosening region is failed to be detected, it may extremely affect the safety of running trains. Therefore, omission isn't allowed in the detection task. Combining with CNN for object detection [21]–[24], a novel detection method based on the periodic distribution rule is proposed. Taking full advantage of this rule, the method can output all undetected regions as suspected visible fault regions while ensuring high detection recall rates.

1) PERIODIC DISTRIBUTION OF WHEELBASE

In this method, we use the train's side frame to describe its structure's periodic distribution. As shown in Fig. 2, a train is generally formed of many cars with same structure. And a car usually consists many parts with various components. For any car of the train, the position distribution of four wheelbases satisfies

$$x(j) = \{L_1, L_3, L_1, L_2; j = 1, 2, 3, 4\} \tag{1}$$

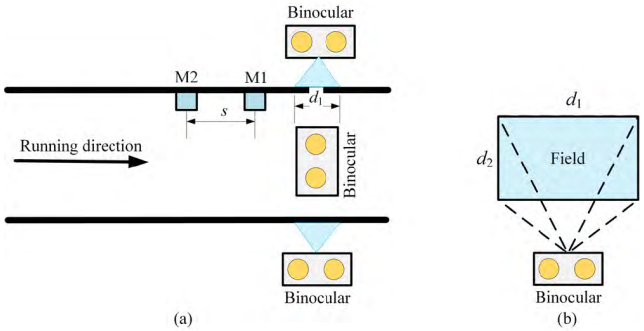

FIGURE 3. (a) Image acquisition method based on magnetic steel triggers. (b) The filed size d of the binocular stereo vision sensor. d_1 and d_2 are the length and width of the filed size. The filed size d satisfies $d = d_1 \times d_2$.

where, j indicates the index of the wheelbase. L_1 is the wheelbase of a bogie, L_2 is the nearest wheelbase between two adjacent cars, and L_3 is the adjacent wheelbase between two bogies.

If a train has C cars, the wheelbase distribution $\tilde{x}(j)$ can be periodically expressed as

$$\tilde{x}(j) = x((j))_4 \quad 0 < j \leq 4C,\ j \in N^+ \tag{2}$$

2) PERIODIC DISTRIBUTION OF IMAGE SEQUENCE

The image acquisition method mentioned in section II is based on magnetic steel triggers. As shown in Fig. 3 (a), two magnetic steels are mounted on the rail side at a distance of s. When the train wheels pass the triggers, M1 and M2 generate two pulse signals. One of them triggers the binocular stereo vision sensor for image acquisition afterwards. According to the pulse interval t_0 and distance s, the train's running speed is calculated. Depending on the shooting range d_1 in the running direction (see in Fig. 3 (b)), we can obtain t by $t = d_1 t_0 / s$.

Therefore, images of the train are sequentially collected in the running direction with a shooting interval of t. Fig. 4 shows several images $\left(n, \tilde{m}, \tilde{l}\right)$ of bogies in different cars. n and $\tilde{m}$ respectively indicate the car's numbering and part's numbering to which the image belongs. And component's numbering can be determined by $\tilde{l}$ as well.

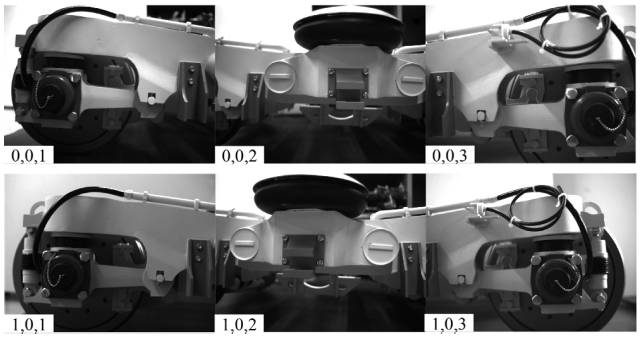

FIGURE 4. Images of the first bogie in different cars. Three images in the first row are different component images in the first car. Images in the second row are different component images in the second car.

a: COMPONENTS NUMBERING

Assume that different components in a part are numbered sequentially from zero. When images of components are collected in order of their position, their corresponding numberings are incremented by adding 1. Therefore, components' numberings in any wheelbase of a car can be given in Eq. (3). Its maximal numbering $l_{\max}$ depends on the fixed field size d of the binocular stereo vision sensor (see in Fig. 3 (b)).

$$\begin{cases} l_i(k) = \{k-1; k = 1, 2, \ldots, (l_{\max}(j)+1)\} \quad j \in [1,4] \\ l_{\max}(j) = \left\lceil \dfrac{\tilde{x}(j)}{d} w \right\rceil - 1 \end{cases} \tag{3}$$

where k is the component index. j indicates the index of the wheelbase, and w is the width of the train.

The components' numbering $l(k)$ in a car can be expressed as

$$l(k) = l_1(k) \cup l_2(k) \cup l_3(k) \cup l_4(k) \tag{4}$$

For a train formed of C cars with same structure, the position distribution of components is a sequence with period M. Its numbering $\tilde{l}(k)$ satisfies:

$$\begin{cases} \tilde{l}(k) = l((k))_M \; 0 < k \le M^* C, \quad k \in N^+ \\ M = \sum_{i=1}^{4} (l_{\max}(i)+1) \end{cases} \tag{5}$$

b: PARTS NUMBERING

When the numbering of components reaches $l_{\max}$, the images of current part have been completely obtained. For the following part, images can be acquired subsequently with the part's numbering adding by 1, and so on.

There are four wheelbases for any car of the train. The parts' numbering $m(p)$ can be expressed as

$$\begin{cases} m(1) = 0 \\ m(p+1) = (m(p)+1)^* \left(1 - \operatorname{sgn}\left(l_{\max}(p) - \tilde{l}(k)\right)\right) \\ + m(p+1)^* \operatorname{sgn}\left(l_{\max}(p) - \tilde{l}(k)\right) \\ p \in [1,3], \quad k \in [k_1, k_2] \\ k_1 = (l_{\max}(1)+1)^* \operatorname{sgn}(p-1) \\ + \dfrac{l_{\max}(2)+1}{(p-1)!}(p-1)(p-2) \\ k_2 = (l_{\max}(1)+1) + (l_{\max}(2)+1)^* \operatorname{sgn}(p-1) \\ + \dfrac{l_{\max}(2)+1}{(p-1)!}(p-1)(p-2) \end{cases} \tag{6}$$

where p indicates the index of the part, and sgn(.) is the sign function.

Similar to the wheelbase, the position distribution of parts shows periodicity as well. Its period is four. Therefore, its numbering $\tilde{m}(p)$ can be calculated by:

$$\tilde{m}(p) = m((p))_4 \quad 0 < p \le 4C, \; p \in N^* \tag{7}$$

c: CARS NUMBERING

Similarly, the numbering of the car adds by 1 once its parts' images are completely collected. Therefore, the cars' numbering $n(q)$ satisfies

$$\begin{cases} n(1) = 0 \\ n(q+1) = (n(q)+1)^* (1 - \operatorname{sgn}(3 - \tilde{m}(p))) \\ + n(q+1)^* \operatorname{sgn}(3 - \tilde{m}(p)) \; 0 < q < C, \quad p \in N^+ \end{cases} \tag{8}$$

Therefore, any image $\left(n(q), \tilde{m}(p), \tilde{l}(k)\right)$ can be uniquely determined by the numbering sequence of cars, parts and components. Besides, the image sequence $\left(n, \tilde{m}, \tilde{l}\right)$ has obvious periodicity.

3) KEY COMPONENT REGIONS DETECTION

In this paper, we design one CNN network architecture to detect all key component regions of trains. The network architecture is the same as the detection model introduced in our previous work [20]. Since the image sequence $\left(n, \tilde{m}, \tilde{l}\right)$ can be classified into 4 groups according to the various parts $\tilde{m}$, we respectively train the network with different kinds of images. Therefore, we can obtain 4 CNN-based detection models correspondingly.

Its detection principle is shown in Fig. 5. For any image $\left(n(q), \tilde{m}(p), \tilde{l}(k)\right)$ inputted at the predicted time, one detection model is selected depending on the value $\tilde{m}(p)$. Based on the regression theory [25], a series of 3×3 convolutional filters have been used to obtain predictions on six network layers. And then combine the selected detection model with those predictions. The following method of non-maximum suppression (NMS) eliminates highly duplicate detections and generates the final detection.

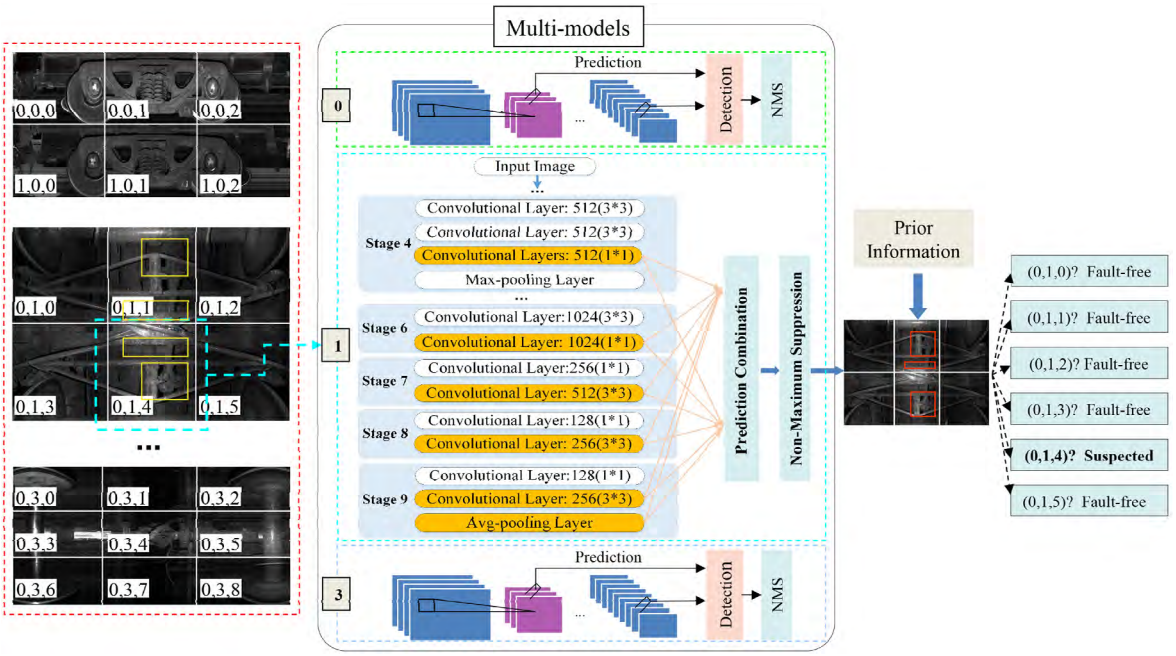

FIGURE 5. Principle of key component regions detection. Layers which are used to predict both location and confidences are shown in orange. And the regions to be detected and the localized regions are in yellow, red boxes, separately.

For each kind of image, the type and total number of key component regions are fixed. Based on the prior information, the total number of localized and undetected regions is automatically counted. The undetected regions are finally output as the suspected visible fault regions.

B. EDGE EXTRACTION AND SEGMENTATION

The method mainly consists of two pieces: edge extraction and edge segmentation. Edge extraction is to obtain sub-pixel edges of the localized key component regions. Based on a novel shape feature descriptor, edge segmentation is applied to separate edges of bolt caps and the mounting surface from others.

1) EDGE EXTRACTION

Fine edge extraction is the premise of accurate 3D measurement. In this paper, we adopt a ''coarse-to-fine'' scheme to extract sub-pixel edges. First, an CNN-based edge detection network is established and trained for pixel edge detection [26]–[29]. Then, edge center points are extracted based on Hessian Matrix Method proposed by Steger [30], [31]. Thus, we can obtain sub-pixel edges of the localized key component regions.

a: COARSE EXTRACTION

In this paper, another CNN network architecture is constructed to obtain pixel edges. It includes five output layers and a base network for classification with 5 stages (16 weight layers) [32], as shown in Fig. 6. The output layers are connected to 5 different layers of the base network respectively. And each output layer is associated with a classifier. Its object

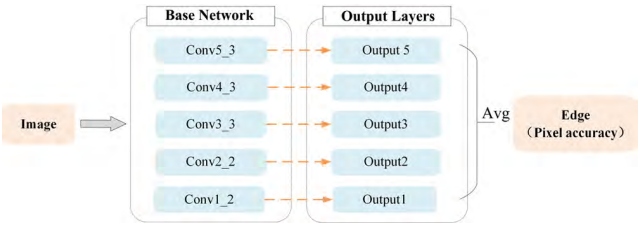

FIGURE 6. Architecture of the CNN-based edge detection network. The different convolutional layers in each stage are denoted as Conv(stage)_(layer). "Output" means the "output layer".

function $L_{output}(\mathbf{W}, \mathbf{w})$ is defined as

$$L_{output}(\mathbf{W}, \mathbf{w}) = \sum_{m=1}^{5} \alpha_m l_{output}^{(m)}\left(\mathbf{W}, \mathbf{w}^{(\mathbf{m})}\right) \tag{9}$$

where parameters of the CNN-based edge detection are denoted as $\mathbf{W}$. $\mathbf{w}$ indicates the weight of each classifier. α_m is scale factor. $l_{output}^{(m)}\left(\mathbf{W}, \mathbf{w}^{(\mathbf{m})}\right)$ is the class-balanced cross-entropy loss function proposed in [26], which aims to offset the heavily imbalance between edge pixels and non-edge pixels of any inputted image.

And then, parameters $\mathbf{W}$ and $\mathbf{w}$ are updated via minimizing the object function with back propagation algorithm. In this CNN-based network, different output layers can output different scale edge predictions. In virtue of continuous inheritance and learning, the accuracy of prediction in deeper layer is no longer lower than the early prediction.

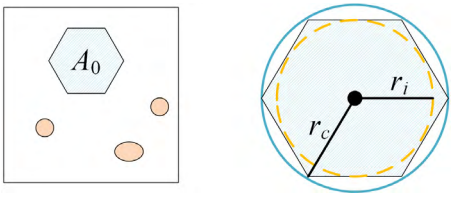

FIGURE 7. Schematic diagram of the shape feature descriptor.

Finally, multi-scale edge predictions $Y_{predict}$ are averaged to obtain pixel edges.

$$Y_{predict} = \text{Average}\left(Y^1_{output}, Y^2_{output}, \ldots, Y^5_{output}\right) \quad (10)$$

where Y^i_{output} is the edge prediction in the i-th output layer.

b: FINE EXTRACTION

For line edges, its first directional derivative (n_x, n_y) should vanish and second directional derivative should be a large absolute value [30]. Especially, the maximum and minimum values of second derivative correspond to the two eigenvalues of the Hessian matrix of the 2D image $f(x, y)$. Therefore, the direction (n_x, n_y) can be determined by the eigenvector corresponding to the eigenvalue of maximum absolute value of the Hessian matrix.

For any image point (x_0, y_0), there exists a variable t satisfying

$$\frac{\partial}{\partial x} f\left((tn_x + x_0), \quad (tn_y + y_0)\right) = 0 \quad (11)$$

If $(tn_x, tn_y) \in [-0.5, 0.5] \times [-0.5, 0.5]$ and the second derivative is larger than the given threshold, point $(tn_x + x_0), (tn_y + y_0)$ is declared as an edge center point. Hence, the sub-pixel edges could be extracted precisely.

2) EDGE SEGMENTATION

In this paper, we design a novel shape feature descriptor R to separate the sub-pixel edges of bolt caps and the mounting surface from others. The schematic diagram is shown in Fig. 7, which is given by

$$R = \text{sgn}\left(r_i / r_c - \sqrt{3}\alpha / 2\right) A_0 \quad (12)$$

where r_i and r_c are the radius of the edge region's inscribed circle and circumscribed circle respectively. A_0 indicates the region's area.

When the radius ratio $\boldsymbol{r_i/r_c}$ roughly equals to 0.866, the edge region is a regular pentagon hexagon. In this paper, we denote $\boldsymbol{R} \geq \mathbf{0}$ to indicate the region belongs to bolt cap, otherwise $\boldsymbol{R} < \mathbf{0}$. Because outer hexagon bolts are generally used for trains' structural support, the scale factor $\boldsymbol{\alpha}$ is set to 0.94.

C. STEREO MATCHING AND 3D RECONSTRUCTION

1) STEREO MATCHING

To get unique matching, points of the segmented edges in the left and right images are first extracted to form the candidate point set P_l and P_r. Then we perform the matching task based on the epiploar constraint [33]. Unlike our previous work [20], we use cross matching method to eliminate mismatch, which will only return consistent point pairs. For any point pairs, the following equation must be satisfied

$$\begin{cases} \mathbf{m}_R^T \mathbf{F} \mathbf{m}_L = 0 \quad with\ \mathbf{m}_L \in P_l\ and\ \mathbf{m}_R \in P_r \\ \mathbf{F} = \mathbf{K}_R^{-T} [\mathbf{t}]_\times \mathbf{R} \mathbf{K}_L^{-T} \\ P_l = \{(x_a, y_a) \,|a = 1, \ldots, N\} \\ P_r = \{(x_b, y_b) \,|b = 1, \ldots, M\} \end{cases} \quad (13)$$

where $\mathbf{m}_L$ is a point in the left image and its corresponding point in the right image is $\mathbf{m}_R$. The 3×3 matrix $\mathbf{F}$ is called the fundamental matrix. $\mathbf{R}$ and $\mathbf{t}$ are structure parameters (rotation and translation). $\mathbf{K}_L$ and $\mathbf{K}_R$ are camera intrinsic matrixes. And $[\mathbf{t}]_\times$ is the anti-symmetric matrix defined by $\mathbf{t}$, supposed to satisfy $\mathbf{t} \times \mathbf{r} = [\mathbf{t}]_\times \mathbf{r}$ for all 3D vector $\mathbf{r}$. Subscript a and b are index of points included in point set P_l and P_r, while the total number of point sets is denoted as N and M separately.

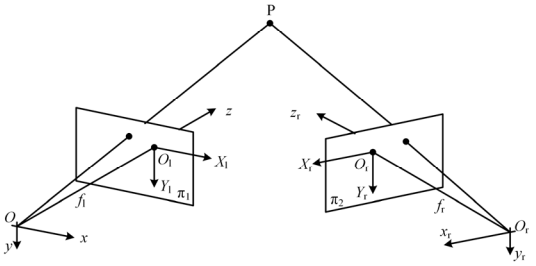

FIGURE 8. The binocular stereo vision model.

2) 3D RECONSTRUCTION

After establishing the stereo correspondence between P_l and P_r, the 3D coordinates of the spatial points can be solved according to the binocular stereo vision model [34]. The binocular stereo vision model is shown in Fig. 8. O-xyz is the left camera coordinate system, which locates at the origin of the word coordinate system. O_l-X_lY_l is the left image coordinate system. Similarly, the right camera coordinate system and its image coordinate system are O_r-$x_ry_rz_r$, and O_r-X_rY_r, respectively.

For the spatial point $\mathbf{P}$, its homogeneous coordinates $\tilde{\mathbf{M}} = (x, y, z, 1)^T$ can be solved by two camera projection models

$$\lambda_l \tilde{\mathbf{m}}_l = \mathbf{A}_l (\mathbf{I}|0) \tilde{\mathbf{M}}, \quad \mathbf{A}_l = \begin{bmatrix} f_l & 0 & 0 \\ 0 & f_l & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (14)$$

$$\lambda_r \tilde{\mathbf{m}}_r = \mathbf{A}_r (\mathbf{R}|\mathbf{T}) \tilde{\mathbf{M}}, \quad \mathbf{A}_r = \begin{bmatrix} f_r & 0 & 0 \\ 0 & f_r & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (15)$$

$$\mathbf{T} = \begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^T, \quad \mathbf{R} = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix} \quad (16)$$

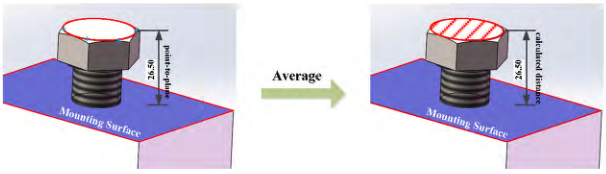

FIGURE 9. Distance between the bolt cap and its mounting surface.

where λ_l and λ_r are arbitrary scale factors. $\tilde{\mathbf{m}}_l = (X_l, Y_l, 1)^T$ and $\tilde{\mathbf{m}}_r = (X_r, Y_r, 1)^T$ are the homogeneous coordinates of image points correspondingly. The 3×3 matrix $\mathbf{R}$ is the rotation matrix from O-xyz to O_r-$x_r y_r z_r$, while the 3×1 vector $\mathbf{T}$ is the translation vector between the origins of two camera coordinate systems. Among them, t_x, t_y and t_z represent the translation amount in x_r, y_r and z_r directions, respectively. $\mathbf{A}_l$ and $\mathbf{A}_r$ are intrinsic matrixes, consisting of the effective focal length f_l and f_r.

Therefore, 3D coordinates of these points of the segmented edges can be solved as

$$\begin{cases} x = zX_l/f_l \\ y = zY_l/f_l \\ z = \dfrac{f_l\left(f_r t_x - X_r t_z\right)}{X_r\left(r_7 X_l + r_8 Y_l + f_l r_9\right) - f_r\left(r_1 X_l + r_2 Y_l + f_l r_3\right)} \\ \quad = \dfrac{f_l\left(f_r t_y - Y_r t_z\right)}{Y_r\left(r_7 X_l + r_8 Y_l + f_l r_9\right) - f_r\left(r_4 X_l + r_5 Y_l + f_l r_6\right)} \end{cases} \tag{17}$$

3) DISTANCE CALCULATION

After obtaining all the 3D coordinates of points on the mounting surface, its spatial plane equation can be approached with Least Square Fit Algorithm. As shown in Fig. 9, distance from the 3D point on the bolt cap to its mounting surface can be calculated. By averaging the point-to-plane distances, we can calculate the distance between the bolt cap and its mounting surface. By defining the distance between bolt cap and mounting surface for a tight bolt as reference value, we can finally judge whether a bolt is loose by comparing the calculated distance with reference value.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. SYSTEM STRUCTURE DESIGNING AND PARAMETERS SETTING

The system for bolt-loosening detection of trains is illustrated in Fig. 10. It mainly includes a bogie model, a binocular stereo vision sensor, and a computer with a single Titan X GPU.

Since bolt-loosening is the main reason for the failure of the bogie's axle box (AB), traction motor (TM) and gear box (GB), a bogie model of the high-speed train CRH380A is made to build the bolts dataset, whose size ratio to the real bogie is 1:5.

In this paper, the binocular stereo vision sensor is designed to collect the left and right images of the bogie model. It consists of 2 cameras (GC1380H) and 2 lenses (Cingeo1.4).

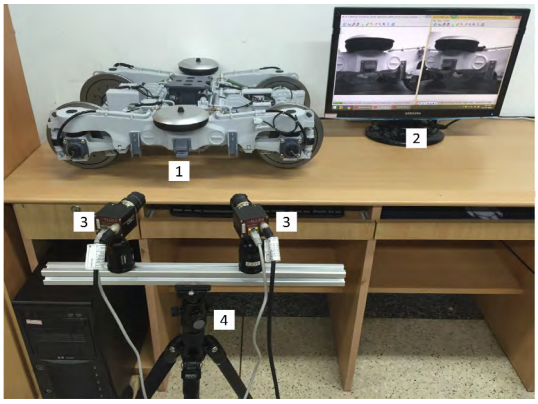

FIGURE 10. Measurement system. (1) Bogie. (2) Computer. (3) Camera. (4) Tripod.

The resolution of camera images is 1360pixel $\times$ 1024pixel, and the pixel size of its CCD is 6.45μm $\times$ 6.45μm. The focal length of lens is 12mm. Since the train passes through the fault detection system at the speed of 3km/h actually, the camera exposure time is set to 24μs to keep the image smear within 0.02mm. Moreover, since the object distance is below 500mm, its baseline length of the binocular stereo vision sensor is set to be 650mm. The angle between the optical axis and baseline is 45°. And its field size is set to 230mm $\times$ 170mm.

B. DATASET DESCRIPTION

In this study, we construct two types of CNN networks: one for the detection of key component regions and the other for the edge detection. Correspondingly, two datasets need to be built for CNN networks training.

1) DATASET FOR KEY COMPONENT REGION DETECTION

Because there isn't available public dataset for training the key component region detection network. In order to develop a dataset containing AB, TM and GB regions, the left and right images of the bogie are collected by the binocular stereo vision sensor under different lighting conditions. These images correspond to non-fault, bolt-loosening and bolt-missing respectively. We select images randomly to build training set, validation set and test set. The training and validation set own 1448 images (404 AB images, 522 TM images and 522 GB images). Each training image's format is the same as the PASCAL VOC dataset's [35]. There are 660 images in the test set. The proportions of the training and validation set, and test set can be seen in Table 1. Some images of the AB, TM and GB are shown in Fig. 11.

2) DATASET FOR EDGE DETECTION

For the CNN-based edge detection network, we train it on the public BSD500 dataset [36]. The dataset is composed of 200 train images, 100 validation and 200 test images. Each training image corresponds to a file containing edges labeled by annotators.

TABLE 1. The proportion of train and validation set, and test set.

Fault class	Region	Training and validation set	Test set
Non-fault or Bolt-loosening	AB	324	200
	TM	422	200
	GB	422	200
Bolt-missing	AB	80	20
	TM	100	20
	GB	100	20

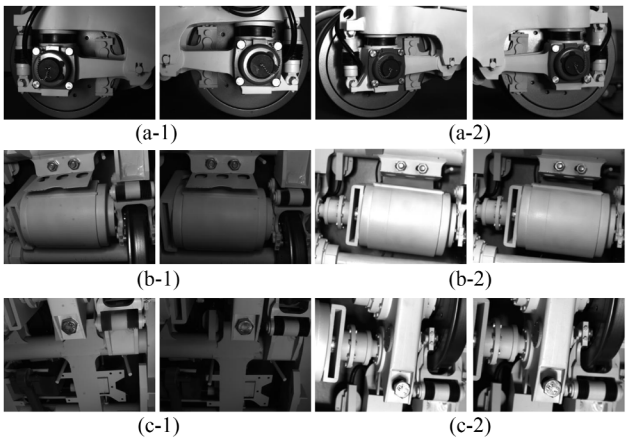

FIGURE 11. Some images of the bogie. (a-1) Non-fault image of AB. (a-2) Image of AB with bolt-loosening. (b-1) Non-fault image of TM. (b-2) Image of TM with bolt-loosening. (c-1) Non-fault image of GB. (c-2) Image of GB with bolt-loosening.

TABLE 2. Key component regions detection results of the images with non-fault or bolt-loosening.

Fault	Region	Total regions	Undetected	Suspected	False
Non-fault or Loosening	AB	200	0	0	0
	TM	200	0	0	0
	GB	200	0	0	0

C. KEY COMPONENT REGIONS DETECTION

1) DETECTION RESULTS

On the training set, we train the key component regions detection network introduced in section III.A. During the network training process, we firstly use the 10^{-3} learning rate for 10000 iterations, and then decay it to 10^{-4} and continue for another 10000 iterations training. Finally, we test the detection model on the test set (660 images) to evaluate its effectiveness. The localization results for some regions are shown in Fig. 12. The detection results of images with non-fault or bolt-loosening can be seen in Table 2. And Table 3 shows the detection results of images with bolt-missing.

As can be seen form Table 2, there is no omission or error for detection of key component regions. Our proposed method can simultaneously locate multiple key component regions with non-fault or bolt-loosening, and obtain 100% region detection recall rate.

TABLE 3. Key component regions detection results of the images with bolt-missing.

Fault	Region	Total regions	Undetected	Suspected	False
Missing	AB	20	0	20	0
	TM	20	0	20	0
	GB	20	0	20	0

TABLE 4. Key component regions detection results of the images with noise.

Noise level	Undetected			Recall		
	AB	TM	GB	AB	TM	GB
5	0	0	0	100%	100%	100%
10	0	0	0	100%	100%	100%
15	0	0	0	100%	100%	100%
20	0	0	0	100%	100%	100%
25	0	0	0	100%	100%	100%
30	0	1	2	100%	99.5%	99%
35	0	2	2	100%	99%	99%
40	0	9	6	100%	95.5%	97%
45	0	10	9	100%	95%	95.5%
50	0	13	12	100%	93.5%	94%

From Table 3, the key component regions with bolt-missing are all outputted as suspected visible fault regions. And there is no wrong output.

In summary, key component regions are either detected accurately for further diagnosis or output as suspected visible fault regions, which suggests that the omission rate is 0% in actual.

2) ROBUSTNESS TO NOISE

In this section, we add Gaussian noises of mean 0 and varying standard deviations (SD) to 600 test images. The key component regions detection results are shown in Table 4. The results show that when the SD of noise is within 50, the performance of the detection is still satisfactory. Although the detection recall of TM and GB decreases slightly when the SD of noise is greater than 25, it is still not lower than 93.5%.

In addition, noise has more negative effect on the detection of smaller regions (such as TM and GB) than on lager regions (such as AB). This is mainly because image features will be corrupted by noise. Under the same noise level, the loss of smaller regions' features is more serious. So, it is harder to detect the smaller region.

Since the noises in the Table 4 are sufficient to simulate the actual noises, our proposed detection method is robust enough to noise.

3) ROBUSTNESS TO LIGHTING CONDITION

Lighting condition is another important parameter to validate the performance of our approach. In this paper, we adopt the average intensity level of the image to quantify its lighting condition, and conduct a further experiment with varying lighting condition.

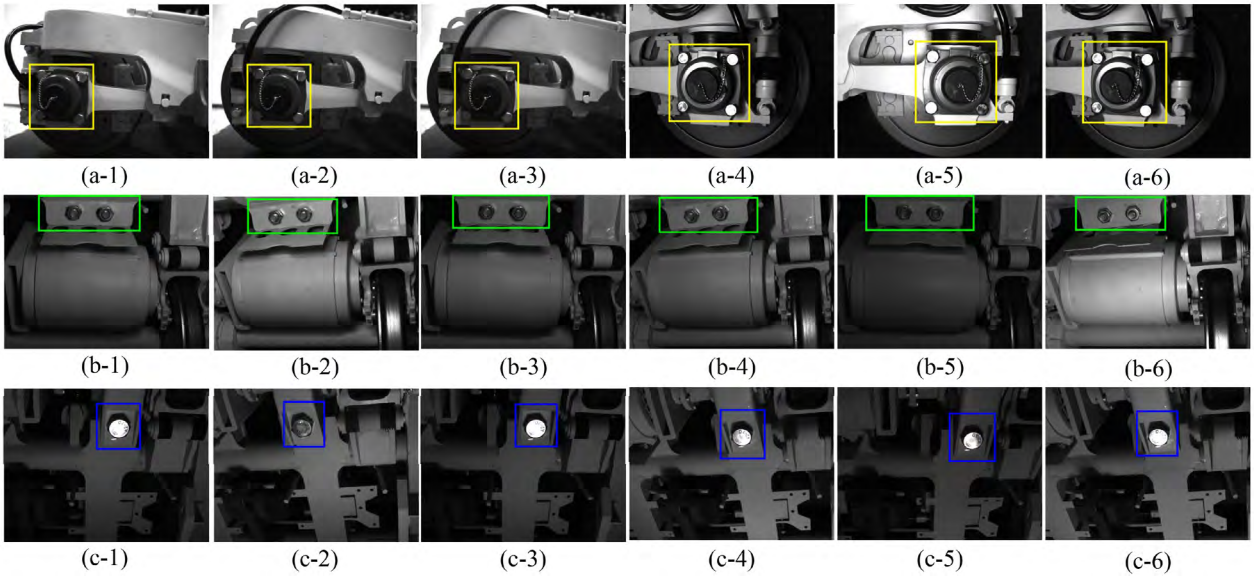

FIGURE 12. Display of key component region localization results. The detected AB, TM and GB are in yellow, green and blue boxes, separately. (a) Images with different lighting conditions include AB regions. (b) Images with different lighting conditions include TM regions. (c) Images with different lighting conditions include GB regions.

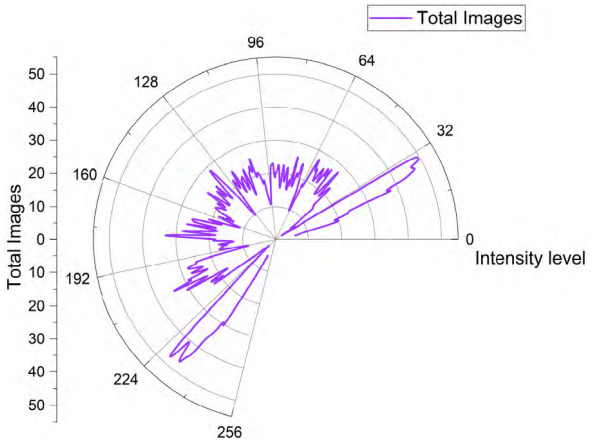

FIGURE 13. The average intensity distribution of 800 × 600 images with varying lighting conditions.

Firstly, we equally divide 256 intensity level into 8 levels. Different intensity levels represent varied lighting conditions. Then, adjust the average intensity of each test sample separately, so that it could be randomly distributed in the intensity interval $[i, i + 32)$. Hence, we can obtain 8×600 images with different lighting conditions. The average intensity distribution of these images is shown in Fig.13. Some images with different lighting conditions and their corresponding histograms are shown in Fig. 14. Finally, our proposed method is tested on 8×600 images to validate its robustness towards lighting condition.

The detection results are presented in the Table 5, As can be seen, our method can obtain 100% detection recall rate when the average intensity of each image is within the interval [32, 192). However, when the average intensity exceeds this interval, the detection recall rate decreases. Especially, the detection recall of GB drops to 57.5% when its average intensity is over 224. It is mainly due to image details loss caused by extremely low lighting condition (see in Fig.15(a)) or bright illumination (see in Fig. 15(b)).

TABLE 5. Key component regions detection results of images with varying lighting conditions.

Average gray	Undetected			Recall		
	AB	TM	GB	AB	TM	GB
[0, 32)	35	9	0	82.5%	95.5%	100%
[32, 64)	0	0	0	100%	100%	100%
[64, 96)	0	0	0	100%	100%	100%
[96, 128)	0	0	0	100%	100%	100%
[128, 160)	0	0	0	100%	100%	100%
[160, 192)	0	0	0	100%	100%	100%
[192, 224)	0	1	5	100%	99.5%	97.5%
[224, 256)	0	7	85	100%	96.5%	57.5%

All in all, without overexposure or underexposure, the performance of our method is satisfactory under varying lighting conditions.

D. EDGE EXTRACTION AND SEGMENTATION

After obtaining the location information of key component regions, edge extraction and segmentation are conducted. The CNN-based edge detection network is trained on the public BSD500 dataset. During network training process, the initializing learning rate is 10^{-6}. Momentum is 0.9 and the number of training iterations is 10000. After 5000 iterations training, we decay learning rate to 10^{-5}.

With the well-trained edge detection model, the edge detection result is shown in Fig. 16. For a clearer representation,

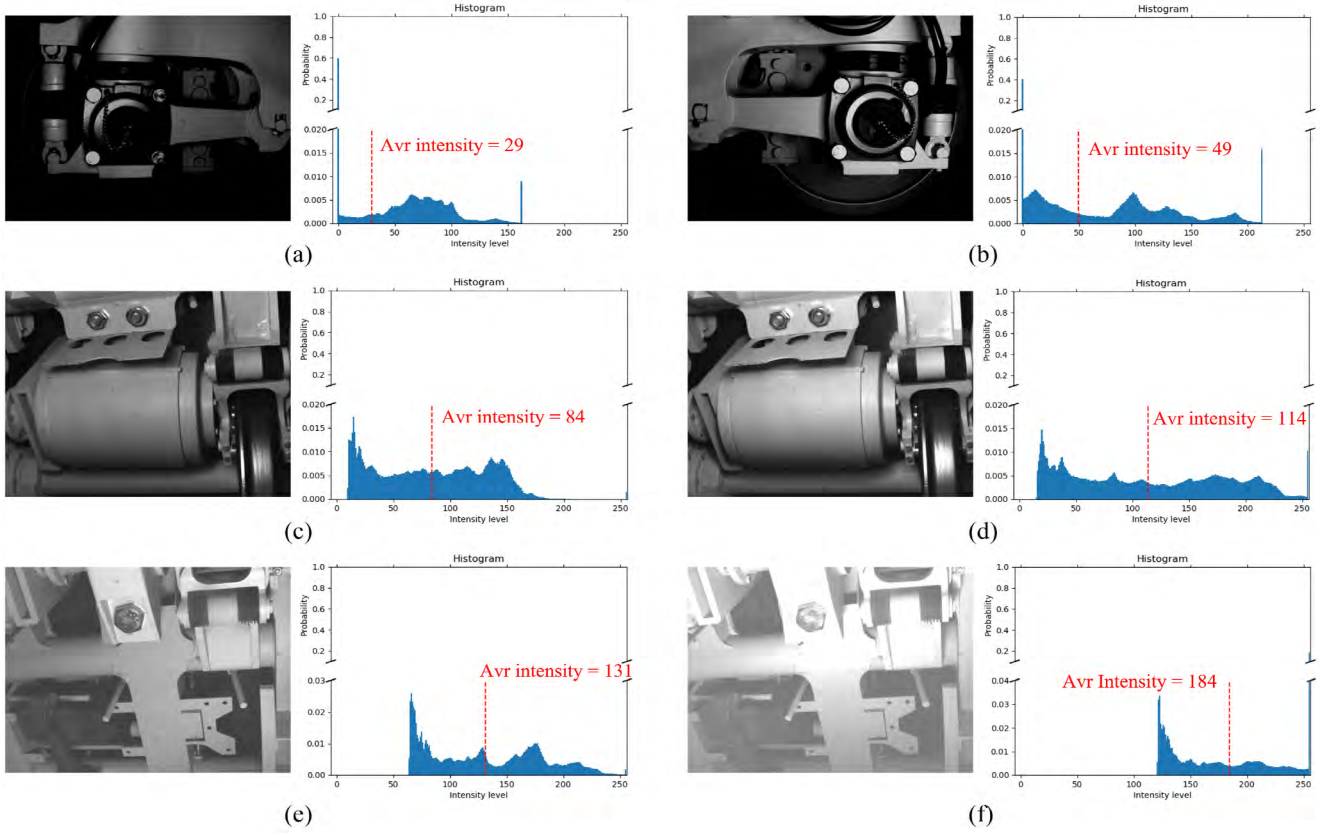

FIGURE 14. Some images with different lighting conditions and their corresponding histograms. The average intensity level of the image is marked with a red dotted line in its histogram. (a) The image of AB and its histogram with 29 average intensity. (b) The image of AB and its histogram with 49 average intensity. (c) The image of TM and its histogram with 84 average intensity. (d) The image of TM and its histogram with 114 average intensity. (e) The image of GB and its histogram with 131 average intensity. (f) The image of GB and its histogram with 184 average intensity.

FIGURE 15. Images with undetected key component regions. (a) The image under extremely low lighting condition. (b) The image under bright illumination.

the partial enlargement of the edge center points is shown in Fig. 17. Edge center points are marked with red dots.

From Fig. 16, good edge connectivity and few false edges demonstrate the superior performance of our method. In addition, our method can not only detect the completed edges of key components, but also segment the edges of bolt caps accurately.

E. BOLT-LOOSENING DETECTION

For a tight bolt in AB, TM and GB region, distance between the bolt cap and its mounting surface is 3.52mm, 6.40mm and 7.74mm respectively, measured by a Vernier caliper with precision of 0.02mm. We take these measured distances as reference values in this paper.

In order to detect single-bolt loosening and multi-bolt loosening, the distance between the bolt cap and its mounting surface is calculated based on the stereo matching and 3D reconstruction method introduced in section III.C. The stereo matching result is shown in Fig.18. By comparing the calculated distance with the reference value, we can judge whether the bolt is loose.

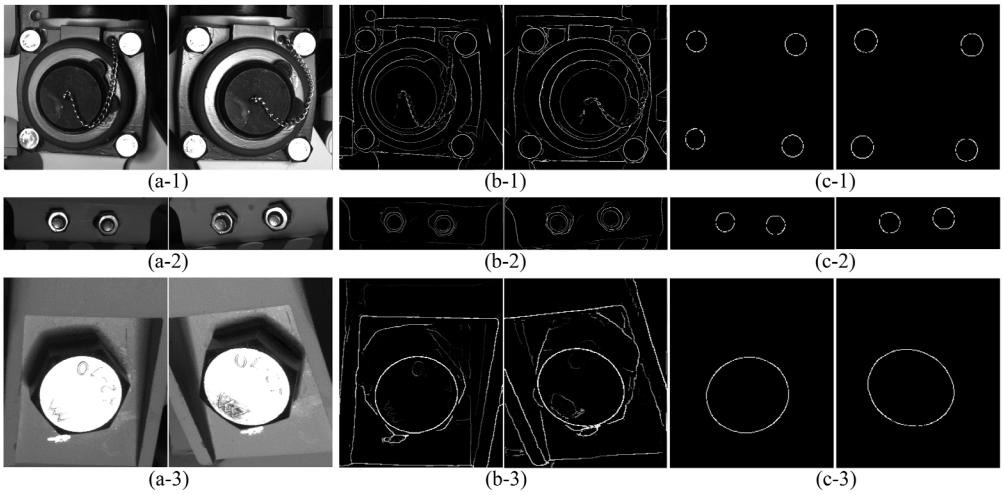

FIGURE 16. Results of edge extraction and segmentation. (a) Region of AB, TM and GB. (b) Edges with pixel accuracy in the AB, TM and GB region. (c) Bolt cap edges of the AB, TM and GB region.

TABLE 6. The experimental results of repeated measurements.

Region	Bolt	Calculated Distances (mm)										SD (mm)
		1	2	3	4	5	6	7	8	9	10	
AB	1	3.57	3.51	3.58	3.59	3.57	3.51	3.58	3.57	3.60	3.58	0.03
	2	3.50	3.57	3.59	3.58	3.57	3.57	3.59	3.58	3.57	3.58	0.03
	3	3.58	3.57	3.58	3.59	3.57	3.58	3.57	3.57	3.57	3.60	0.01
	4	3.58	3.59	3.57	3.51	3.57	3.59	3.58	3.57	3.57	3.59	0.02
TM	1	6.45	6.45	6.39	6.45	6.46	6.47	6.45	6.45	6.38	6.46	0.03
	2	6.46	6.49	6.45	6.47	6.45	6.47	6.45	6.49	6.46	6.45	0.02
GB	1	7.82	7.82	7.82	7.79	7.79	7.80	7.82	7.79	7.79	7.82	0.02

1) REPETITIVE EXPERIMENT

In order to evaluate the reliability of our bolt-loosening detection method, the distance for the same bolt in AB, TM and GB region are calculated for 10 times. The experimental results are presented in Table 6. As can be seen, the standard deviation (SD) of repeated calculation is less than 0.03mm. The calculation results are small in dispersion. It suggests that this method has good reliability.

2) SINGLE-BOLT LOOSENING

We select a bolt in AB, TM and GB regions respectively and loosen them within 0.00mm to 9.00mm at intervals of 3.00mm. The calculated results are given by Table 7. According to the Table 7, our method has a high computational accuracy when the looseness of a bolt is within 9mm. The absolute error and relative error of calculated distances are less than 0.07mm and 1.42% separately. It means that our proposed method can calculate the looseness of a single bolt accurately.

3) MULTI-BOLT LOOSENING

In the experiment, three randomly selected bolts in the AB regions are loosened by 3.00mm, 6.00mm and 9.00mm, while two bolts in TM region are released by 3.00mm, 6.00mm. And its distances are calculated by our proposed method. The experimental results are shown in Table 8. As can be seen, the method is still feasible and accurate when multiple bolts are loosened in the same area. The absolute error of experimental results is less than 0.08mm, and its relative error is kept within 1.23%. It means that our method can calculate the looseness of multiple bolts in the same area simultaneously as well.

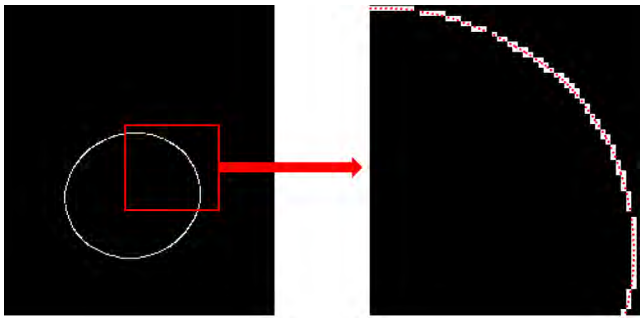

FIGURE 17. The partial enlargement of the GB's edge center points. The edge center points are marked with red dots.

4) TIME CONSUMPTION

In the task of fault detection, the time consumption is one of the most important indicators. According to the actual

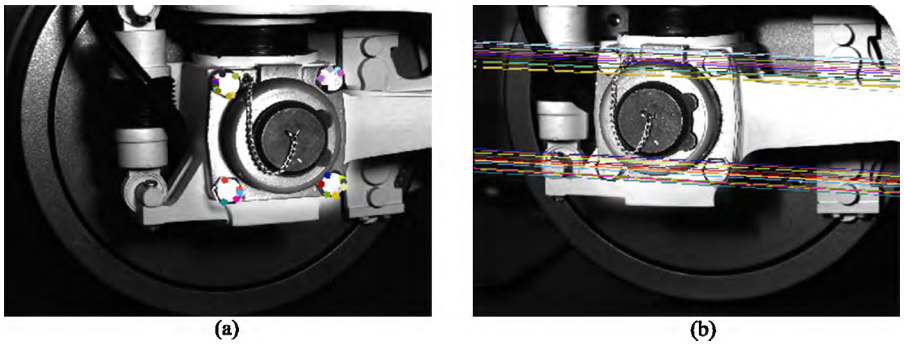

FIGURE 18. Stereo matching based on epiploar constraint. (a) The left image with AB region. Symbols '*' with varied colors are used to mark points on edges of bolt caps. (b) The right image with AB region. Lines with distinct colors are polar lines of the right image corresponding to the feature points with the same color on the left image.

TABLE 7. The calculated results for key component regions with single bolt loosening.

Region	Loose extent (mm)	CALCULATED DISTANCE (MM)	Absolute error (mm)	Relative error (%)
AB	0 (3.52)	3.57	0.05	**1.42%**
	3 (6.52)	6.57	0.05	0.77%
	6 (9.52)	9.58	0.06	0.63%
	9 (12.52)	12.57	0.05	0.40%
TM	0 (6.40)	6.47	**0.07**	1.09%
	3 (9.40)	9.45	0.05	0.53%
	6 (12.40)	12.46	0.06	0.48%
	9 (15.40)	15.45	0.05	0.32%
GB	0 (7.74)	7.79	0.05	0.65%
	3 (10.74)	10.80	0.06	0.56%
	6 (13.74)	13.80	0.06	0.44%
	9 (16.74)	16.81	**0.07**	0.42%

TABLE 8. The calculated results for key component regions with multi-bolt loosening.

Region	AB			TM	
Bolt	1	2	3	1	2
Reference Distance (mm)	6.52	9.52	12.52	9.40	12.40
Calculated Distance (mm)	6.60	9.59	12.58	9.45	12.46
Absolute error (mm)	**0.08**	0.07	0.06	0.05	0.06
Relative error (%)	**1.23**	0.74	0.48	0.53	0.48

requirements of the railway system in China, it is necessary to complete fault detection within 5 minutes for a motor train with 8 cars. Since the length of a car is about 25m, 250 images of the train's side frame are collected by the binocular stereo vision sensor with the field size 800mm × 600mm. In this paper, the average bolt-loosening detection time per image is 1.1s approximately. The total processing time is 4.6 minutes. Therefore, our proposed method can meet requirement of time consumption.

V. CONCLUSION

In this paper, a fast bolt-loosening detection method based on binocular vision is proposed to calculate bolts looseness for AB, TM and GB regions. The experimental results verify that our method can detect single-bolt loosening and multi-bolt loosening with high precision. Unlike the existing fault 3D detection techniques, our proposed method only reconstructs the 3D information of bolt caps and the mounting surface, avoiding scanning the entire train body. It improves the efficiency of 3D detection of bolt-loosening. Moreover, the key component regions detection method based on the periodic distribution can output all the undetected regions as suspected visible fault regions while ensuring high detection recall rates. The omission rate is 0% in actual, improving the reliability of fault detection.

Although our proposed method is ideal for automatic bolt looseness detection and has excellent performance, limitation still remains. In this development stage, the method can only calculate the looseness of the bolt with clean surface.

REFERENCES

[1] G. Sun, W. Feng, D. Zhao, and L. Yang, "An automatic fault recognition method for side frame key in TFDS," *Open Mech. Eng. J.*, vol. 9, no. 1, pp. 22–27, Feb. 2015.
[2] R. Zou, Z.-Y. Xu, J.-Y. Li, and F.-Q. Zhou, "Real-time monitoring of brake shoe keys in freight cars," *Frontiers Inf. Technol. Electron. Eng.*, vol. 16, no. 3, pp. 191–204, Mar. 2015.
[3] Z. Liu, D. Xiao, and Y. Chen, "Displacement fault detection of bearing weight saddle in TFDS based on hough transform and symmetry validation," *Proc. 9th Int. Conf. Fuzzy Syst. Knowl. Discovery (FSKD)*, Sichuan, China, May 2012, pp. 1404–1408.
[4] G. Nan and J. Yao, "A real-time visual inspection method of fastening bolts in freight car operation," *Proc. SPIE.*, vol. 9675, Oct. 2015, Art. no. 96752G.
[5] R. Liu, "Principle and application of TFDS," *Chin. Railways*, no. 5, pp. 26–27, Feb. 2005.
[6] J. Sun, Z. Xiao, and Y. Xie, "Automatic multi-fault recognition in TFDS based on convolutional neural network," *Neurocomputing*, vol. 222, pp. 127–136, Jan. 2017.
[7] X. Yang, L. Ye, and J. Yuan, "Research of computer vision fault recognition algorithm of center plate bolts of train," *Proc. 1st Int. Conf. Instrum., Meas., Comput., Commun. Control*, Beijing, China, Oct. 2011, pp. 978–981.
[8] Z. Hongjian, H. Ping, and Y. Xudong, "Fault detection of train center plate bolts loss using modified LBP and optimization algorithm," *Open Automat. Control Syst. J.*, vol. 7, no. 1, pp. 1916–1921, Oct. 2015.
[9] N. Li, Z. Wei, Z. Cao, and X. Wei, "Automatic fault recognition for losing of train bogie center plate bolt," in *Proc. IEEE 14th Int. Conf. Commun. Technol. (ICCT)*, Chengdu, China, Nov. 2012, pp. 1001–1005.

[10] 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. F, J. Rail Rapid Transit*, vol. 230, no. 7, pp. 1629–1641, Sep. 2016.
[11] Y. Dou, Y. Huang, Q. Li, and S. Luo, "A fast template matching-based algorithm for railway bolts detection," *Int. J. Mach. Learn. Cybern.*, vol. 5, no. 6, pp. 835–844, Dec. 2014.
[12] C. Li, Z. Wei, and J. Xing, "Online inspection system for the automatic detection of bolt defects on a freight train," *Proc. Inst. Mech. Eng. F, J. Rail Rapid Transit*, vol. 230, no. 4, pp. 1213–1226, Jun. 2016.
[13] Q. Li and X. Jing, "A second-order output spectrum approach for fault detection of bolt loosening in a satellite-like structure with a sensor chain," *Nonlinear Dyn.*, vol. 89, no. 1, pp. 587–606, Mar. 2017.
[14] Y.-J. Cha, W. Choi, G. Suh, S. Mahmoudkhani, and O. Büyüköztürk, "Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 9, pp. 731–747, 2018.
[15] X. Kong and J. Li, "Image registration-based bolt loosening detection of steel joints," *Sensors*, vol. 18, no. 4, p. 1000, Mar. 2018.
[16] Y.-C. Cha, K. You, and W. Choi, "Vision-based detection of loosened bolts using the Hough transform and support vector machines," *Automat. Construction*, vol. 71, pp. 181–188, Nov. 2016.
[17] L. Ramana, W. Choi and Y.-J. Cha, "Automated vision-based loosened bolt detection using the cascade detector," *Sensors Instrum.*, vol. 5, pp. 23–28, Apr. 2017.
[18] Harbin Kejia General Mechanical and Electrical Company. (2017). *Trouble of Moving EMU Image Detection System*. [Online]. Available: http://www.kejiajidian.cn/products_detail/productId=31.html
[19] New Vision. (2017). *Trouble of Moving EMU 3D Detection System (TEDS-3D)*. Accessed: Dec. 7, 2017. [Online]. Available: http://www.huaxingzhiyuan.com/Web/News/170.aspx
[20] Y. Xie and J. Sun, "On-line bolt-loosening detection method of key components of running trains using binocular vision," *Proc. SPIE*, vol. 10605, Nov. 2017, Art. no. 1060513.
[21] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1440–1448.
[22] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 580–587.
[23] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.
[24] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2016.
[25] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. ECCV*, Sep. 2016, pp. 21–37.
[26] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. J. Conf. Comput. Vis.*, Santiago, Chile, Dec. 2015, pp. 1–16.
[27] G. Bertasius, J. Shi, and L. Torresani, "DeepEdge: A multi-scale bifurcated deep network for top-down contour detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, vol. 52, no. 3, pp. 4380–4389.
[28] W. Shen, X. Wang, Y. Wang, X. Bai, and Z. Zhang, "DeepContour: A deep convolutional feature learned by positive-sharing loss for contour detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 3982–3991.
[29] J.-J. Hwang and T.-L. Liu, "Pixel-wise deep learning for contour detection," in *Proc. ICLR*, 2015.
[30] C. Steger, "An unbiased detector of curvilinear structures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 2, pp. 113–125, Feb. 1998.
[31] C. Steger, "Analytical and empirical performance evaluation of subpixel line and edge detection," *Empirical Eval. Methods Comput. Vis.*, Apr. 1998, pp. 188–210.
[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, 2015.
[33] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2003, ch. 9, sec. 2, pp. 241–246.
[34] G. Zhang, *Machine Vision*. Beijing, China: Science Press, 2005, ch. 5, sec. 1, pp. 99–101.
[35] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010.
[36] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.

JUNHUA SUN was born in Hubei, China, in 1975. He received the B.E. and M.E. degrees from the Beijing Institute of Machinery, China, in 1997 and 2003, respectively, and the Ph.D. degree in precision instrument and machinery from Beihang University, Beijing, China, in 2006, respectively, where he is currently a Professor and also a key member with the Key Laboratory of Precision Opto-Mechatronics Technology, Ministry of Education.

His research interests include machine vision, image processing and recognition, and machine learning.

YANXIA XIE was born in Hunan, China, in 1993. She received the B.E. degree from Chang'an University, Xi'an, China, in 2015, and the M.E. degree from Beihang University, Beijing, China, in 2018, where she is currently pursuing the Ph.D. degree with the Key Laboratory of Precision Opto-Mechatronics Technology, Ministry of Education.

Her research interests include image recognition and machine learning.

XIAOQI CHENG was born in Henan, China, in 1987. He received the B.E. degree from the Zhengzhou University of Aeronautics, Zhengzhou, China, in 2009, and the M.E. degree from Henan Polytechnic University, Jiaozuo, China, in 2012. He is currently pursuing the Ph.D. degree in measurement technology and instruments with the Key Laboratory of Precision Opto-Mechatronics Technology, Ministry of Education, Beihang University, Beijing, China.

His current research interests include machine vision, image recognition, and machine learning.