Online Rail Surface Inspection Utilizing Spatial Consistency and Continuity

Jinrui Gan, Jianzhu Wang, Haomin Yu, Qingyong Li, *Member, IEEE*, and Zhiping Shi

***Abstract*—Rail surface inspection using visual inspection system is an important part of railway maintenance. However, accurate and efficient identification of possible defects remains challenging. This paper proposes a background-oriented defect inspector (BODI) to improve defect detection by considering specified characteristics of the track during inspection. Reformulating the inspection task in this manner offers a new way to model rail surface images. More specifically, BODI features a random sampling stage to obtain a compact background representation without any prior information. A sufficient number of random selections generates adequate and diverse background statistics, and defect-determination and a fusion of procedures then determine whether current pixel belongs to the background. Finally, a background update mechanism and parallelism ensure real-time applicability. The proposed BODI is evaluated on a working railway line. The experimental results demonstrate that it outperforms state-of-the-art methods.**

***Index Terms*—Computer vision, discrete surface defects, rail inspection, random sampling.**

I. Introduction

VISUAL inspection systems (VISs) [1]–[3] are widely used for railway examination because they can efficiently, precisely, and objectively evaluate large databases of acquired images. These systems benefit from the steady development of machine vision, and are playing increasingly significant roles in fault-detection tasks, such as track-profile measurement [4], bogie block key inspection [5], fastening elements inspection [6]–[8], and track surface inspection [9]. In these tasks, automated recognition is possible through the use of approaches focused on obtaining highly discriminative and local invariant features related to lighting conditions and local geometric constraints [10]. For instance, various texture analysis techniques can be combined with machine learning models [9], [11]–[14] for the classification, detection, and segmentation of rail images. These approaches, which rely on extracted global or local spatial patterns, perform well in their specified tasks. Giben *et al.* [15] and Gibert *et al.* [16] applied deep-learning-based frameworks to rail inspection, demonstrating their effectiveness in rail fastener assessment. This paper focuses on a promising and nondestructive VIS for inspecting rail surface discrete defect (RSDD).

RSDDs, a common and important form of failure, involve cracks in a rail surface, rolling contact fatigue due to high stress, spalling of the rail-head, and rail corrugation [17], [18]. They always indicate serious risks and are a major issue in rail safety [19]. Compared with the other rail-component inspection tasks mentioned above, RSDD inspection is more difficult because of the following two factors. The first is limited features and prior information. The most reliable feature to distinguish discrete defects from dynamic background is local gray-level information. This kind of fault always appears in a random or arbitrary manner, showing various appearances, shapes, and sizes. This limitation prevents inspection systems from using most object-recognition methodologies based on sophisticated texture and shape features [20]. The second complicating factor is limited number of training samples. Any railway line will include very few surface discrete defects available for training, which results in severe overfitting by machine learning models, especially those based on deep learning. Overall, a successful inspection system must address these obstacles.

In recent years, several RSDD inspection methods have been proposed. Li and Ren [20] have proposed the methods of LN+DLBP and MLC+PEME [21]. Both employ defect-enhancement and automatic-localization strategies. The method of LN+DLBP first enhances contrast between defects and background using the local normalization (LN) algorithm, and then identifies possible defects by the defect localization based on projection profile (DLBP) algorithm. Instead, the method of MLC+PEME uses local Michelson-like contrast (MLC) and proportion emphasized maximum entropy (PEME) algorithms. In addition, a new inverse Perona–Malik diffusion model presented for image enhancement [22] provides an adaptive threshold binarization to identify desired defects. These approaches, which focus on the contrast between defects and background, perform well on uniformly textured rail surface images. However, they fail to find defects with greatly varying appearances on dynamic backgrounds owing to heavy random noise caused by complex track conditions and camera quality [23]. Accordingly, our previous work [24] poses

Manuscript received December 25, 2017; revised March 5, 2018; accepted April 4, 2018. This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 2016JBZ006, in part by the Beijing Natural Science Foundation under Grant J160004, in part by the Shanghai Research Program under Grant 17511102900, in part by the National Natural Science Foundation of China under Grant 61632004, and in part by the Fundamental Research Funds under Grant 2017YJS061. This paper was recommended by Associate Editor Q. Wang. *(Corresponding author: Qingyong Li.)*

J. Gan, J. Wang, H. Yu, and Q. Li are with the Beijing Key Laboratory of Transportation Data Analysis and Mining, Beijing Jiaotong University, Beijing 100044, China (e-mail: liqy@bjtu.edu.cn).

Z. Shi is with the Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing 100048, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMC.2018.2827937

this problem in an alternative way to explore the characteristics of the background rather than the defects, and focuses on finding background modes for the RSDD inspection. The method shows promising performance, providing somewhat robust inspection. However, the obtained background modes are based on a mean shift algorithm that shifts this kernel iteratively to a higher-density region until convergence, which is computationally expensive.

Thus, effective RSSD inspection remains challenging. There are two points that should be noted.

1) Traditionally, VISs capture images of rails using a line-scan camera at fixed lengths (e.g., 1 m for an image) and then analyze these images individually for inspection. All the previously reported defect inspectors mentioned above focus on the appearance of defects in a single image, but ignore the spatial characteristics of the background. In other words, the rail is monitored and stored as individual images whose continuity is neglected. We therefore reformulate the problem of RSDD inspection considering spatio-temporal characteristics.
2) Our previous study analyzing the content of rail surface images [24] suggested that pixel intensity remains largely consistent in the longitudinal direction with only slight deviation, and outliers showing significant deviation always indicate true defects and other noise points. Defects occupying a small fraction of pixels therefore usually account for unusual observations in otherwise relatively homogeneous backgrounds. The defect pixels are wholly derived from background statistics. Therefore, we focus on distinguishing unusual pixels from typical backgrounds without reference to target signatures or subspaces. Unlike the background modes generated in our other work [24], this paper adopts a random strategy to obtain the background representation. The underlying idea is that it is more reliable to estimate the statistical distribution of a background pixel with a small number of close values than with a majority of all values.

Accordingly, we propose a novel method called the background-oriented defect inspector (BODI) for RSDD inspection. Specifically, the BODI builds a background model for each longitudinal line by randomly taking a collection of pixels along it. To capture the influence of random characteristics, our analysis employs a family of theories based on binomial distributions. Guided by the theory analysis results, several different random selection procedures and a fusion strategy are applied here to enhance the representation power of background. For robustness, adaptability, and real-time applicability, an update mechanism for the background models is utilized. RSDD inspection is therefore an online procedure. We also evaluate the proposed method on a working railway line. The experimental results show that our method outperforms state-of-the-art methods, thereby demonstrating its suitability for rail inspection.

Compared to the existing literatures, the main contribution of this paper is twofold. First, we reformulate the problem of RSDD inspection by exploring spatio-temporal characteristics of track and propose the BODI method. In this new perspective, we focus on the homogeneous background and treat a railway line as a whole, whereas previous methods handle visual features (e.g., contrast and wavelet features) of defects in a single image. Second, we propose a random strategy for generating and updating background model in BODI, which can largely reduce the computational cost compared with sophisticated algorithm (e.g., mean shift). Furthermore, we investigate a family of theories based on binomial distributions in order to capture the influence of random characteristics. These theories guide BODI to adopt several different random selection procedures and a fusion strategy.

The remainder of this paper is organized as follows. Section II briefly introduces the related works. Section III describes the proposed method in detail. The experimental setup and results are presented in Section IV, and the conclusions follow in Section V.

II. Related Works

The RSDD inspection problem is related to the topic of the saliency detection [25], if RSDDs are regarded as salient objects. Saliency detection has attracted a lot of interest in computer vision, because of its theoretical meaning for explaining human attention and applicable aims in complex vision problems such as scene understanding [26]. Generally speaking, most classic models described in literature aim to identify the possible salient pixels, blocks, superpixels, and regions from images first (i.e., compute a saliency map) and then integrate them to extract the entire salient object [27]–[29]. Another alternative is to harness the recently resurgence of the convolutional neural networks [30] to eliminate the need for hand-crafted features and the dependency on center bias knowledge [31], [32]. Please refer to [33] for a complete overview on the subject. Although the detection object in this paper is similar to salient object, it should be noted that RSDD inspection faces two challenging factors as aforementioned. Therefore, rather than these sophisticated algorithms, the specified characteristics of track will be completely explored in our method to meet the high inspection performance and real-time requirements in the rail inspection field.

Accordingly, the specified spatio-temporal characteristics of track are explored in this paper so that the inspection problem can refer to as video interestingness [34]–[37]. Further, the highly consistent background is expected in the rail surface [24], making RSDD inspection problem be related to the classic motion detection in video sequences [38]. The literature of the topic can be roughly divided into two categories according to their representation function for the background: 1) parametric [39], [40] and 2) nonparametric [41], [42] methods. For example, the most popular and widely used parametric approaches are based on Gaussian mixture models [43], [44], considering the dynamic and multimodel nature of scenes. Note that most techniques described in literature operate on each pixel independently and always assume an explicit probability density function (PDF). The other way is nonparametric method that has drawn much attention, because

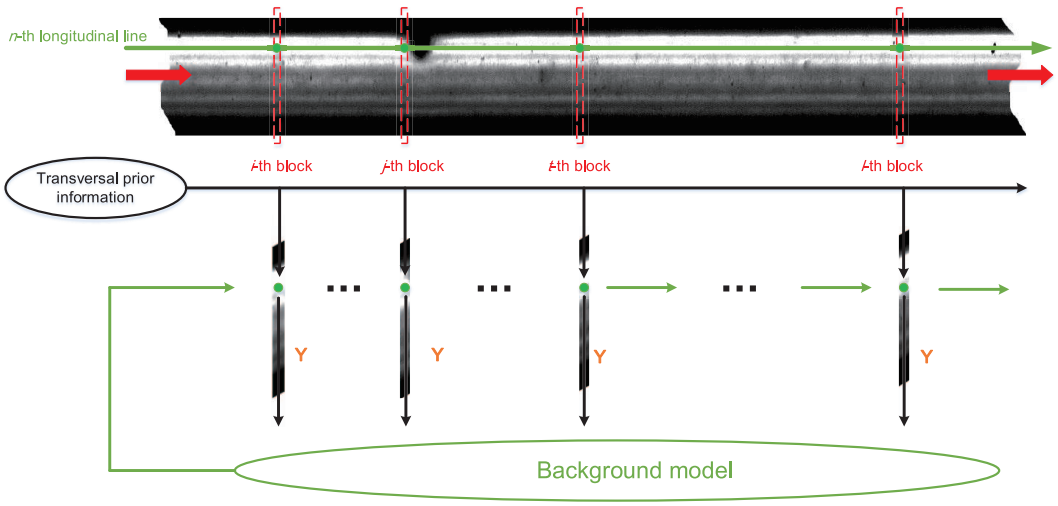


Fig. 1. Pipeline of RSDD inspection. Red dashed boxes represent blocks (here transversal lines), green solid circles represent pixels (or points), and each orange "Y" means the current pixel is classified as background. RSDD inspection is undertaken block by block. Our method uses the background model and transversal prior information to determine whether the current pixel in the current block belongs to the background. If it does, then its corresponding background will be updated; otherwise, the determining process turns to the next pixel along the longitudinal line. The same procedure is applied to other longitudinal lines. Note that the inspection procedure can be applied for rail line with any length.

it can avoid to find an appropriate shape for the PDF. The most popular way is to build a histogram or codebook for background values to determine whether or not a pixel belongs to a background [45], [46]. The recent nonparametric approach is visual background extractor (ViBe) proposed by Barnich and Van Droogenbroeck [47]. ViBe, which has significantly circumvented the difficulties of probability function estimation and dictionary learning, is an effective algorithm that can extract the target in the initial few frames.

III. Method

A. Reformulation and Overview

As our previous study [24] suggested, rail surface can be expected to have a highly consistent background, which would be broken by defects that would appear as foreground items. Any rail surface denoted by $\mathbf{B}$ can be further segmented into blocks of any width w, giving a spatial sequence of rail surface blocks, $\mathbf{B} = \{\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_t, \ldots\}$. Note that the length of $\mathbf{B}$ can theoretically be any value. Thus, $\mathbf{B}$ can represent any previously captured image [20]–[22], [24] or even the whole rail surface of a railway line being inspected.

Consider a classical formulation of anomaly detection, which equates anomalies to outliers. A statistical model $\Pr_Z(z)$ is postulated for the distribution of a measurement z under normal conditions. Abnormalities are defined under this model as measurements for which the probability is below a threshold. Here, we consider the RSDD inspection problem under an unsupervised manner, which is formulated as

$$\Pr\left(\mathbf{C}_t = 0 | \mathbf{B}_t, \{\mathbf{B}_i, \mathbf{C}_i\}_{i=1}^{t-1}\right) = \exp\{-D(\mathbf{B}_t, \mathbf{M}_t)\} \quad (1)$$

where the conditional probability $\Pr(x|y)$ measures the degree to which the current block $\mathbf{B}_t$ belongs to the background denoted by $\mathbf{C}_t = 0$, $\mathbf{M}_t$ is the representative background drawn from previous blocks and corresponding class labels $\{\mathbf{B}_i, \mathbf{C}_i\}_{i=1}^{t-1}$, and $D(x, y)$ can be any family of distance measurements. The formulation overcomes the limitations of these previous works focused on individual fixed rail surface images, and so provides broad opportunities for the rail inspection field. It is worth mentioning that subsequent algorithm designing is built into this formulation. For example, the following real-time implementation can also benefit from the updating of the representative background representation $\mathbf{M}_t$.

For a given pixel $p(x, y)$ in a block $\mathbf{B}_t$, where x and y denote the coordinates in the longitudinal and transversal directions, respectively, the pixels in the next positions along the longitudinal line can be expected to be similar unless they belong to a defect or noise. We thus build a background model for each longitudinal line to distinguish defects from the normal background model, and propose a new detection method (BODI). Fig. 1 illustrates the whole inspection procedure. In the work, we process pixels in a block independently and set w to 1. It means that each transversal line is a block. This is reasonable given that the entire longitudinal line along a rail's surface image is consistent. The whole inspection includes the following three main aspects.

1) *Preprocessing:* Given a block, the impact of unequal lighting and variations in reflections on the rail's surface should be first reduced to a certain extent. For each longitudinal line in the block, we first apply a *log* operator for all the pixels; then apply Z-scores standardization; finally normalize to [0, 255] by min–max normalization. Note that if the width w of the segmented block is small enough, we can hold a batch of segmented blocks at a time, and then apply the preprocessing strategy to these blocks.
2) *Background Representation and Defect Determination:* Our method essentially models the background with a set of randomly selected samples around the object pixel p. This pixel is then identified as background

or defect based on the background model and other important prior knowledge. More details, including real-time implementation, are described in the next section.

3) *Post-Processing:* Some small false-positive defects will be discarded at this stage. Note that the same batch strategy used in preprocessing is also adopted here, when w is small enough.

B. Background-Oriented Defect Inspector

Actually, a promising defect inspector (i.e., the definition of distance measurement shown in our reformulation) is assumed to be able to highlight the defects, so a key process is to choose a proper background representation. A simple way is to generate a global inspector always using all pixels detected as background pixels. As a result, the defect pixels contaminate the background representation. The aim of the defect inspector is to select adequate and representative background pixels from the rail surface image and define a proper measurement for the background while limiting the effect of defects. Our previous work [24] used longitudinal background modes in the feature space as the proper background representation. The strategy is computationally expensive because of mean shift, an iterative algorithm to achieve convergence. Here, the background representation is randomly selected to reduce the computational cost. Background statistics from the selected background pixel set are then adopted for the RSDD inspection.

To limit the impact of noise, we characterize each pixel $p(x, y)$ by a feature vector $\mathbf{F}(p) = [I, IMean, IMedian]^T$, where I, $IMean$, $IMedian$ denote the gray value, mean value, and median value of its spatial neighborhood, respectively. The background of each pixel p is modeled by a collection of N background pixels as follows:

$$\mathbf{M}(p) = \{\mathbf{F}_1, \ldots, \mathbf{F}_n, \ldots, \mathbf{F}_N\} \tag{2}$$

where the N background pixels in our method are randomly taken along the longitudinal, according to a uniform law. The background representation plays a significant role in the subsequent defect inspection, and the following part studies the relationship between background sample size N and the probability of striking defect pixels.

As mentioned above, true defect points usually occur in a small fraction along the longitudinal direction, and we experimentally find that defects points account for up to 4% of the total sample size. That is, the maximum probability of any randomly chosen pixel in the longitudinal direction being a defect pixel is $q = 4\%$. It is clear that the number of randomly chosen defect pixels L follows a binomial distribution with parameters N and q: i.e., $L \sim B(N, q)$. Thus, the probability of striking at least one defect pixel is represented as

$$\Pr(L \geq 1) = \sum_{L=1}^{N} \binom{N}{L} q^L (1-q)^{(N-L)} = Q. \tag{3}$$

It is easily verified that as N increases, the probability of striking defect pixels increases sharply. Therefore, fewer background samples are preferred to reduce the probability of striking defect pixels. However, the use of insufficient background samples will mean that the background is not properly represented, as the examined rail surface image will contain dynamic background and much noise. Overall, the background representation aims not to be free of defect pixels, but to show sufficient diversity. A fusion strategy for each random selection procedure is needed to balance these contradicting priorities. Subsequently, we further study the probability that all the selections strike defect pixels. Suppose that the number of selection procedures is K, then the number of procedures randomly striking defect pixels J also follows the binomial distribution with parameters K and Q computed in (3); i.e., $J \sim B(K, Q)$. Thus, the probability of all the procedures striking defect pixels is

$$\Pr(J = K) = \binom{K}{J} Q^J (1-Q)^{(K-J)} = P. \tag{4}$$

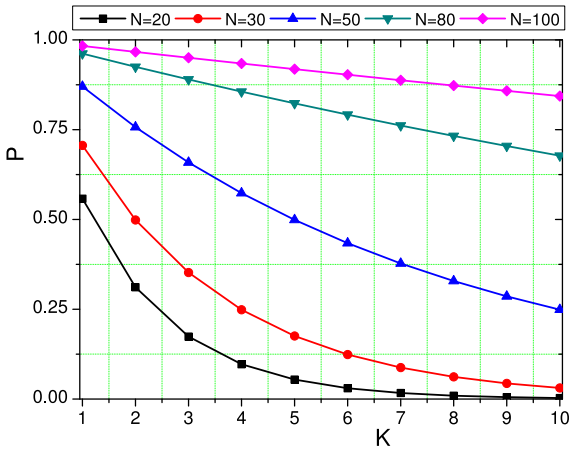


Fig. 2. Relationship between P and K under different background sample sizes N.

The relationship between P and K under different background sample sizes N is shown in Fig. 2. Different curvatures suggest that taking many random background samples requires more procedures that take random selections of background pixels. When K is sufficiently large, the probability of selecting defect pixels in these parallel procedures is clearly reduced. In this case, contamination by defect pixels in the selected background representation would be minimized so that defect pixels would be noticeable compared with the background pixels. Fusing all the different random selection procedures is beneficial for the diversity of background representation and minimizing the occurrence of defect pixels.

Comparison of the current pixel p with its corresponding background model $\mathbf{M}(p)$ determines its nature. Combining the characteristics of rail surface images with the above analysis improves defect determination for this kind of fault inspection task. Equation (2) can generate the background model $\mathbf{M}(p)$ for any pixel p. Here, we define a saliency function $S(p)$ for pixel p to measure its saliency value and accurately label it as background or defect. Intuitively, accurate inspection should greatly reduce the impact of irregularities; i.e., it should augment the contrast of true defects while suppressing the impacts of noise points. On the other hand, it should also reduce the response of points that are unlikely to be true defects. Thus, when considering each pixel independently, its

label should be based on the fitness between its appearance, background model, and spatial location. To this end, we define the following saliency function:

$$S(p) = \frac{1}{K}\sum_{k=1}^{K} \mathrm{LCI}^k(p, \mathbf{M}(p)) * \mathrm{TPI}^k(p) \tag{5}$$

where the final saliency function is defined as the average value under all the different random selection procedures; $\mathrm{LCI}(p, \mathbf{M}(p))$ is the longitudinal context information term, which measures the contrast between the pixel's appearance and its background model and encodes the longitudinal spatial consistency; and $\mathrm{TPI}(p)$ is the transversal prior information term, which measures the probability of true defects occurring at different locations in the transverse direction.

For the longitudinal context information term, two points should be noted. The first is that maximizing true defects while suppressing noise to a certain extent is to be encouraged. The second is that although a lower gray value prior of true defects is often suppressed by the factors mentioned above, large gray values are not the true defects. Accordingly, the longitudinal context information term is defined as

$$\mathrm{LCI}(p, \mathbf{M}(p)) = \frac{1}{N}\sum_{n=1}^{N}\left[(\mathbf{F}_n - \mathbf{F}(p))^T \mathbf{W} (\mathbf{F}_n - \mathbf{F}(p))\right]^{\frac{1}{2}} \mathrm{Pr}(p) \tag{6}$$

where $\mathbf{W}$ is the weight dialog matrix. The probability $\mathrm{Pr}(p)$ is expressed as

$$\mathrm{Pr}(p) = \frac{\sum_{n=1}^{N} \hat{\sharp}\{(\mathbf{F}_n - \mathbf{F}(p)) > \gamma\}}{d * N} \tag{7}$$

where $\hat{\sharp}$ is an element-wise operator and counts the elements that satisfy the condition, γ is a fixed parameter, and d is the dimension of feature vector $\mathbf{F}$. We set $\mathbf{W} = \mathrm{diag}([0.3, 0.35, 0.35])$ and $\gamma = 10$ experimentally. If point p occurs with a large vibration but has a large gray value, it is not the true defect point. Then $\mathrm{Pr}(p)$ is close to zero, which will alleviate its impact on accuracy. Instead of considering spatial near neighbors, which can cause large false negatives, the strategy considers its global background samples and accumulated weighted Euclidean distance. It can augment the discrimination between true defects and noise points, and the following experiments show that the strategy leads to good detection performances.

The transversal prior information term should encode the probability of defect occurrence in different locations. A general observation is that true defects most probably occur in the middle of a rail. Our previous work [24] surveying the centroids of true defects in the transverse direction also suggests that true defects are more likely to occur in the middle of a rail image, and rarely appear transversally away from the center. In other words, locations far from the middle are more likely to show false responses. To achieve the best tradeoff between robust inspection and alleviating the edge effect, the transversal prior information is also modeled by a simple parabola [24]. It should be noted that the transversal location of point in the transversal prior information term $\mathrm{TPI}(p)$ should be normalized to [0,1]. Interestingly, the transversal prior information term is related to spatial location rather than the background representation generated by different random selection procedures.

Given the saliency value, the final determining criterion is defined as

$$C(p) = \begin{cases} 1 & S(p) > T \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where T is a threshold parameter. If $C(p) = 1$, then the pixel p is labeled as a defect point, and vice versa. Different strategies for thresholding can be used, such as estimating a fixed threshold by cross validation, or calculating an adaptive threshold. In this paper a fixed threshold algorithm is adopted, similar to other fault inspection tasks such as can-end [48], LCD panel [49], and large-aperture optical elements [50]. Here, further segmentation by a fixed threshold from the saliency map yields precise shapes of defects, which is beneficial to quantitative evaluation and comparison with other well-established methods. Note that the parameter T would influence the performance of defect detection; its impact is discussed in the following experiments.

C. Real-Time Implementation

Remember that the RSDD inspection deals with the whole railway line, not simply individual rail images captured by a CCD camera. Our system can take advantage of the background's longitudinal consistency and the spatially continuous capturing of images to reduce redundant computation.

Suppose the current pixel p is assessed by the defined defect-determination process associated with its background model $\mathbf{M}(p)$. This background model is also applicable to the next pixel along the longitudinal line, and so does not need to be generated again. However, the greater the distance moved along the track, the less the initial background representation fits the current pixels. An update mechanism is therefore essential to account for changes in the background such as lighting or external interference. Neighboring background pixels share a similar temporal distribution especially those neighboring longitudinally. The update mechanism allows the spatial diffusion of information regarding the background evolution that relies on samples classified exclusively as background. The basic principle is that the background model of a pixel should contain samples from the recent past, but older samples should not necessarily be discarded [47]. The update mechanism combines the traditional conservative update policy and a random update scheme to avoid deadlock and everlasting ghosts [51] (i.e., defect pixels should never be included in the background models). Specifically, using a certain index $n \in 1, 2, \ldots, N$ (chosen uniformly at random), the corresponding background model sample $\mathbf{F}_n$ is replaced by the current pixel sample $\mathbf{F}(p)$. This allows the current pixel to be learned by the background model. Only a classified background pixel can be used in the random update of its model, and a random update scheme diffuses its neighboring values. As a result, the classified background pixel p will update its background model $\mathbf{M}(p)$ at random according to a uniform law.

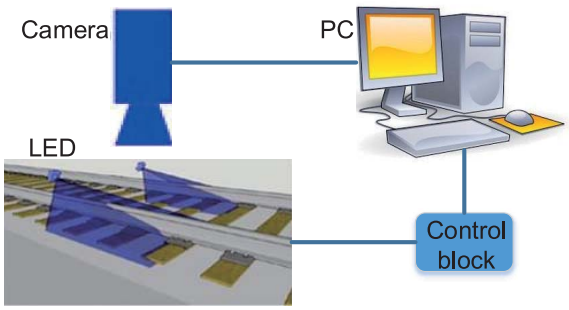


Fig. 3. Flow diagram of the automatic rail image acquisition system.

Note that different random selection procedures, and even different longitudinal lines, can be implemented using parallel computing to increase the inspection speed.

IV. Experimental Results and Analysis

This section details the experimental setup in our experiments, and reports the performance of the proposed method.

A. Experimental Setup

The performance of our method is tested using the automatic rail image acquisition system outlined in Fig. 3, which has an embedded control block automatically controlling LED light sources to adjust their brightness and ensure image quality. The employed CCD camera is an industrial Dalsa Spyder high-speed line-scan camera (maximum line rate of 65 000 lines/s) equipped with an automatic variable-rate filter attenuation system. A PC-Camlink frame grabber connecting the camera and an on-board computer is responsible for the high-speed coding and transmission of images captured by the camera. Several groups of LEDs are installed under a train carriage to eliminate most external stray light. In addition, the camera is triggered by a wheel encoder to ensure that each pixel of a captured image has the same physical size. The acquisition system is installed under a train carriage. It automatically generates one image per meter traveled by the testing train. The extracted images associated with distance information on the rail are automatically saved into several files. Note that each track image acquired by the system has 1-mile length and includes not only the rail surface area but also other components (e.g., fasteners, sleepers/ties, and ballast) on both sides of the rail. To reduce interferences of irrelevant information and to improve inspection performance, the exact rail surface area is extracted first in the subsequent image processing procedure. Here, we simply apply the algorithm based on a projection profile [20] for rail surface localization. Note that the image to be dealt with in the following is the extracted exact rail surface area.

The constructed acquisition system is used to capture rail images, which are used to evaluate the performance of the proposed method. Note that different random selection procedures and different longitudinal lines are implemented using parallel computing techniques in the following experiments.

B. Evaluation

The image-acquisition apparatus has a generally constant effect on the quality of the captured rail surface images, and any inadequacies would complicate and hinder the subsequent image processing module. Therefore, the inspection performance is determined mainly by the image analysis algorithms. Complete evaluation of the proposed method is undertaken by assessing its inspection performances from different perspectives (i.e., parameter sensitivity, reliability analysis, and robustness analysis) using different indexes.

To assess the parameter sensitivity, we used a validation set consisting of 40 images (each has 1 m) collected from common rails, containing at least one defect. Note that the true defects in the validation set are labeled by a human expert. Hence, two types of evaluation index from different perspectives (i.e., pixel-level and defect-level) are adopted here to quantitatively determine the parameters. Both are based on the common criteria of Precision, Recall, and F1-measure, which are widely used in pattern recognition and information retrieval communities [52].

Pixel-Level Index: Pre, *Rec*, and F1-measure (F) are defined in (9)

$$\begin{aligned} \text{Pre} &= \text{TPp}/(\text{TPp}+\text{FP}) \\ \text{Rec} &= \text{TPp}/(\text{TPp}+\text{FN}) \\ \text{F} &= 2*\text{Pre}*\text{Recall}/(\text{Pre}+\text{Rec}) \end{aligned} \tag{9}$$

where TPp denotes the number of correctly detected points, FP is the number of false defected points, and FN is the number of undetected defected points. There is an inverse relationship between *precision* and recall. That is, it is possible to increase one at the cost of reducing the other. Hence, F1-measure is introduced. The higher F1-measure is, the more detection result is consistent with human judgment.

Defect-Level Index: Pre′, *Rec′*, and F1-measure (F') are defined in (10)

$$\begin{aligned} \text{Pre}' &= \text{TPd}/P \\ \text{Rec}' &= \text{TPd}'/N \\ F' &= 2*\text{Pre}'*\text{Rec}'/(\text{Pre}'+\text{Rec}') \end{aligned} \tag{10}$$

where *TPd* denotes the number of correctly detected defects, *TPd′* is the number of correctly detected defects mapping to ground truth, P refers to the total number of detected defects, and N refers to the total number of labeled defects in ground truth. In this level, a detected defect is regarded as correct if it overlaps a labeled defect in the corresponding image more than 80%; otherwise, it is wrong. Note that the counted number in *Pre′* and *Rec′* are for the whole validation set.

Note that our method is applied for these individual images in the validation set. In general, each parameter can be chosen from a large range without greatly affecting the system's performance. Given that the greater the number of parallel selections and detection procedures K, the more adequate and diverse the background representation, we set $N = 20$ and $K = 5$ experimentally, considering the tradeoff between representation power and efficiency. As for the threshold parameter T, Fig. 4 shows the F1-measure curves according to T ranging

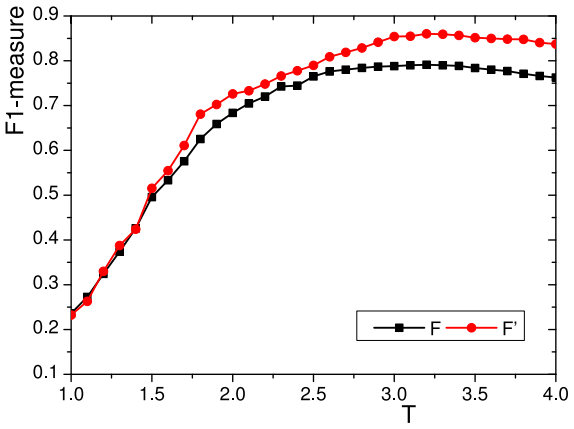


Fig. 4. Effects of the parameter T on F1-measures (F and F').

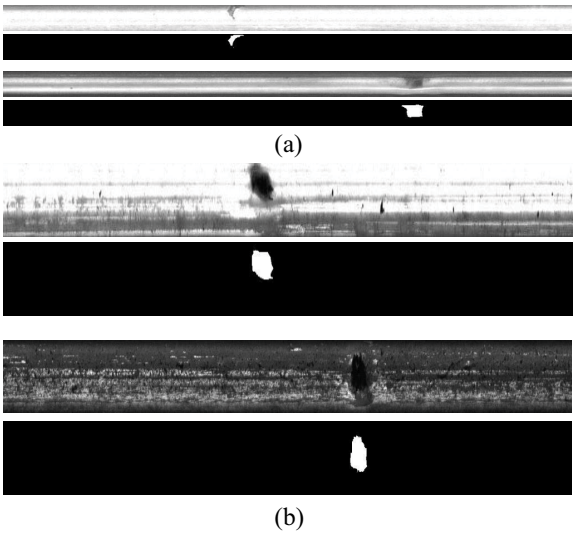


Fig. 5. Robustness validation. Inspection results for example rails with high brightness or variable reflectivity, taken from (a) common rails and (b) express rails.

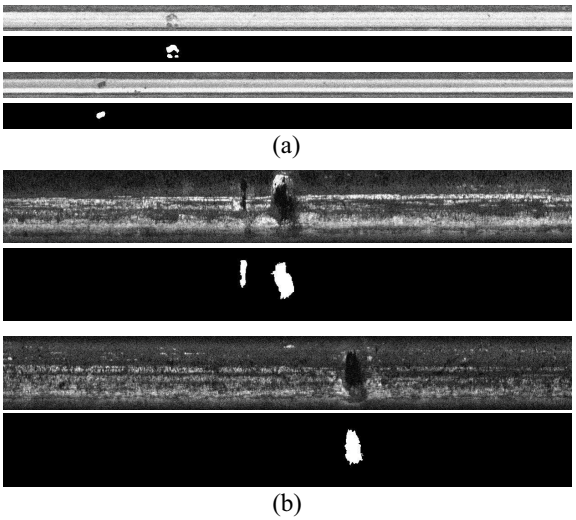


Fig. 6. Robustness validation. Inspection results for example rails corrupted by Gaussian white noise with $\sigma = 20$, taken from (a) common rails and (b) express rails.

from 1 to 4. It can be seen that very small number of T results in an inferior performance, because precision and recall have an inverse relationship. As expected, the F1-measures increase as T increases, and BODI achieves the best performance with $T = 3.2$, because this value results in a good balance between Precision and Recall. However, as T increases further, i.e., $T > 3.2$, the inspection performance begins to decrease. When T is large enough, the algorithm tends to increase Precision at the cost of largely reducing Recall. Although there exists some vibration as shown in Fig. 4, the performance of our method remains stable for a certain range of T. Whether a pixel is a true defect point depends on the learned "normal" background representation.

The largest standard deviation (std) of the above-mentioned F1-measures (F and F') does not exceed 1.0%. Other comparative methods based on machine vision show reasonable stability, because the property is intrinsically determined by the machine vision method [53]–[55]. Hence, the inspection results attained by our method is reliable.

Another important concern is the robustness of the proposed method under different circumstances. This is assessed using images obtained under extreme illumination conditions and images corrupted with Gaussian noise ($\sigma < 20$). The inspection results for these images shown in Figs. 5 and 6 demonstrate that our method still correctly extracts the defect regions in both cases. Two factors must be credited for the robustness inspection. The first originates from our background model behave, allowing the fusion strategy to enhance its representation power and the update mechanism to ensure its adaptability. The second factor is the defined defect determination. It integrates the longitudinal context information and transversal prior information to highlight defects while suppress other outliers.

C. Comparison and Analysis

1) Comparison With Established Methods: Here, we compare the inspection performance of the BODI method with other well-established methods, including LN+DLBP, MLC+PEME, the method of [22], and CFE. Note that their parameters are chosen in the same way as for the BODI method. Figs. 7 and 8 show comparative inspection results for example defects taken from common rails and express rails, respectively. These methods achieve good inspection results when the defects are uniform and the background is consistent. As compared with express rails, common rails have more consistent background, but more complex defects involving various appearances in its internal and sharing different characteristics with other defects. Consequently, MLC+PEME yields many false negatives for defects with a nonuniform appearance, because the threshold determined by maximizing entropy is not optimal. LN+DLBP is also not successful when handling various defects, and locates with bias. The method in [22] is sensitive to gradient, and thus cannot find accurate contours. The comparison results show that both BODI and CFE perform well, without false positives or false negatives.

TABLE I
INSPECTION STATISTICS FOR ACTUAL RAIL ASSESSMENT

Approach	Defect number	Correct detection	False detection	Missing detection	Detection rate
LN+DLBP		20	**712**	10	66.67%
MLC+PEME		29	79194	1	96.67%
Method in [22]	30	24	59991	6	80%
CFE		**30**	1244	**0**	**100%**
BODI		**30**	1809	**0**	**100%**

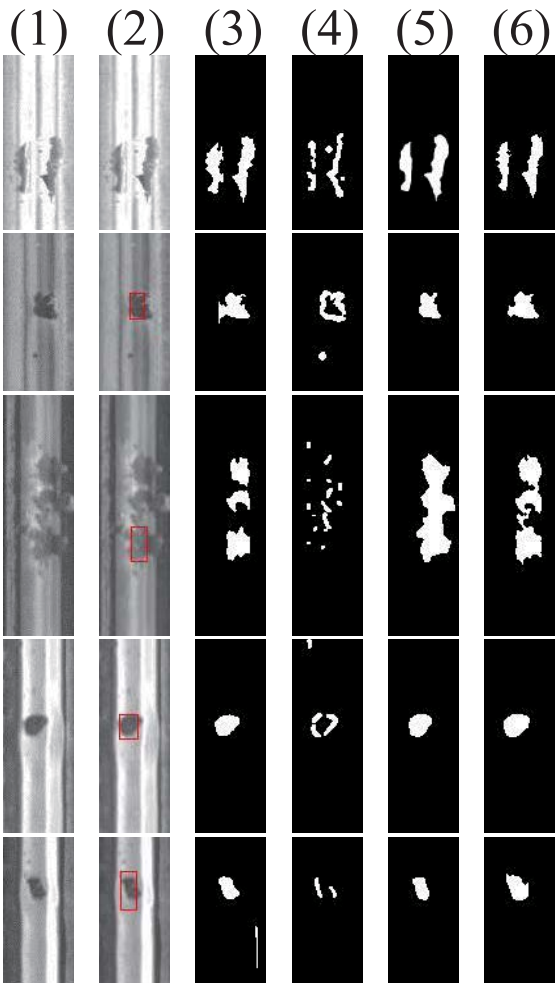


Fig. 7. Comparative inspection results for example defects taken from common rails. (1) Defects. Results for (2) LN+DLBP, (3) MLC+PEME, (4) the method of [22], (5) CFE, and (6) our method. (Note that defect detection in LN+DLBP is marked by a red box.)

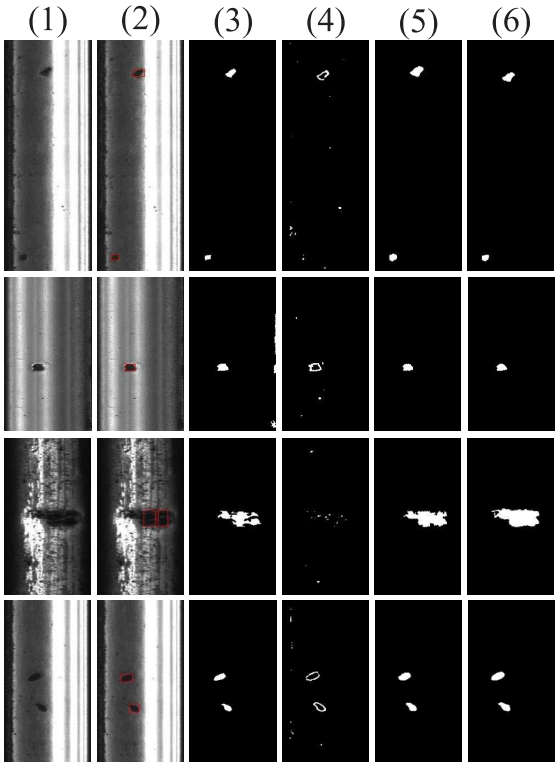


Fig. 8. Comparative inspection results for example defects taken from express rails. (1) Defects. Results for (2) LN+DLBP, (3) MLC+PEME, (4) the method of [22], (5) CFE, and (6) our method. (Note that defect detection in LN+DLBP is marked by a red box.)

TABLE II
PROCESSING SPEEDS (MILLISECONDS) OF VARIOUS METHODS: LN+DLBP, MLC+PEME, THE METHOD OF [22], AND CFE

	LN+DLBP	MLC+PEME	Method in [22]	CFE	BODI
Time	**15.76**	29.64	35.21	372.79	41.78

In order to quantitatively estimate the performance of the proposed method, we evaluate it on a working railway line from common rails. This line covers 45 *km* of 45366 continuous rail images. The evaluation indexes adopted here are common criteria, such as false detection, missing detection, and correct detection. Table I summarizes the comparison inspection results of these five methods. LN+DLBP has a low false detection, but misses some true defects, which is a potential safety risk. As expected, MLC+PEME and the method of [22] perform poorly in terms of false detection. The results show that the proposed method detects all defects successfully, achieving the best performance in terms of detection rate. This result demonstrates its potential applicability to rail inspection, and verifies that our strategy is effective. Furthermore, the selection of the threshold parameter T is robust in our method. However, BODI performs relatively poorly in terms of false detection, because it models pixels independently without considering the status of neighboring pixels.

2) Speed Comparison: We compare the computational times of each method during the field testing, and list the average execution times in Table II. All the methods were implemented on a computer with a 2.40 GHz Intel Xeon E5620 processor and 28 GB of random access memory.

TABLE III
INSPECTION STATISTICS FOR ACTUAL RAIL ASSESSMENT

Approach	Defect number	Correct detection	False detection	Missing detection	Detection rate
ViBe	30	23	6628	7	76.67%
BODI	30	**30**	**1809**	**0**	**100%**

It is worth mentioning that a update mechanism for background representation was also applied in CFE method to achieve a fair comparison. Specifically, the update strategy for the background modes in CFE method is similar to that in [56]. In contrast, the background model update mechanism in our method adopts random strategy. As expected, CFE has the highest computational cost. Although CFE performs well, its high computational complexity puts it at a disadvantage compared with other well-established methods. Note that the maximum speed of the test train carrying the automatic rail image acquisition system is 85 km/h. Accordingly, our algorithm can meet the real-time applicability in the rail inspection.

D. Discussion

If each transversal line is regarded as a video frame, the strategy employed by our method can be viewed as a classic model of computer vision; i.e., motion detection in a video sequence, where each transversal line corresponds to a video frame and RSDDs correspond to foregrounds. A recent nonparametric background modeling approach and most related model is the ViBe [47]. Here, we further discuss the differences between them.

ViBe models and updates the background with some randomly selected samples around each pixel, only using gray-level value; and determines the label of a pixel using a predefined number of samples within a certain distance. As we see, ViBe also has random characteristics. Barnich and Van Droogenbroeck pointed out that the strategy of employing random processes for the model's initialization and update mechanism proves to be powerful despite its unusual. In contrast, this paper theoretically analyzes random characteristics by considering a binomial distribution, and subsequently, guided by the analysis results, several random selection procedures and a fusion strategy are employed to enhance and diversify the background representation power, which is preferable for robust inspection.

We then consider the determination strategy. ViBe determines a point by its matching number, considering the locality of the current point. The strategy that considers near-neighbors can report many false detections. True defects are easily labeled as background, as other noise points appear very similar. In contrast, our strategy, which benefits from its global background samples and accumulated weighted Euclidean distance, can improve discrimination between true defects and noise points. Besides, it can also be easily extended. For example, it can incorporate the transversal prior information term to achieve robust inspection, as shown in (5).

Fig. 9 visualizes some example defects taken from common rails and express rails and corresponding inspection results of ViBe and our method. It can be seen that ViBe has large false negatives when the defects' internals have various appearances, because other noise points that similar to the true defect points confuse its determining procedure in ViBe. From the inspection results, BODI performs better with less false negatives and positives as compared with ViBe. Table III compares inspection results for BODI and ViBe on the working railway line. As expected, BODI is superior to ViBe in terms of correct detection, false detection, and missing detection, because its repeated parallel selection and detection procedure greatly reduce the probability of selecting defect pixels as the background in all procedures. These results indicate that BODI is less sensitive to contamination by defect points in the background statistics, because the defect pixels can be accentuated and false positives in the individual detection procedures can be suppressed by the fusion strategy. Furthermore, the defect determination procedure in BODI that combines longitudinal contextual information and transversal prior information is shown to be more suitable for the specified inspection task.

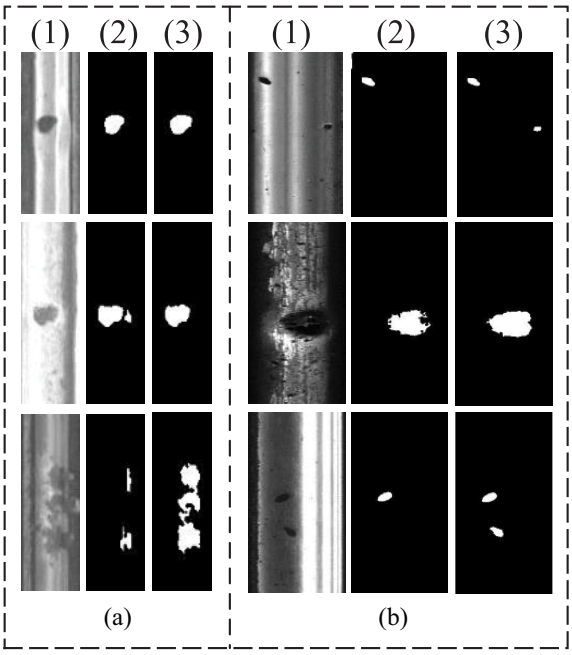


Fig. 9. Comparative inspection results of ViBe and our method for example defects taken from (a) common rails and (b) express rails. (1) defects. Results for (2) ViBe and (3) our method.

V. CONCLUSION

This paper presents a novel BODI method for RSDD inspection. In contrast to previous works on RSDD inspection that focus on fixed rail images captured by CCD camera, here we consider the specified spatial characteristics of the track and further reformulate this inspection problem, leading to an on-line detection. Instead of the background modes used in

previous work, we adopt a random strategy to generate background representations under the idea that it is more reliable to estimate the statistical distribution of a background pixel with a small number of close values than with a majority of all values. Specifically, our method randomly selects representative background pixels as the background representation, and it enhances the representation power by combining different random selections and a fusion strategy. At last, the defect determination is determined by a defined saliency function integrating longitudinal context information and transversal prior information. The experimental results show that our method has reliable detection performance and is capable of handling the effects of external interferences, such as resilience to light variations and noise, thus to obtain a robust inspection. Moreover, the BODI attains a 100% detection rate with low false positives in rail inspection field while proving its real-time applicability on the test train setup, indicating it is a promising RSDD inspection method.

In the future research, we will extend this paper in the following aspects. On one hand, our method builds a background model for each pixel, without considering the relationship between neighboring pixels. Therefore, our future research will develop a background model using super-pixels or considering the labels of neighboring pixels. Thus, encoding more spatial neighbor information should further improve the inspection accuracy. On the other hand, other different types of signal (e.g., ultrasound) will be carefully investigated and combined in the future work to find potential defects.

REFERENCES

[1] M. Trosino, J. J. Cunningham, and A. E. Shaw, III, "Automated track inspection vehicle and method," U.S. Patent 6 064 428, May 16, 2000.

[2] F. Marino, A. Distante, P. L. Mazzeo, and E. Stella, "A real-time visual inspection system for railway maintenance: Automatic hexagonal-headed bolts detection," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 3, pp. 418–428, May 2007.

[3] S. Minaeian, J. Liu, and Y.-J. Son, "Vision-based target detection and localization via a team of cooperative UAV and UGVs," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 7, pp. 1005–1016, Jul. 2016.

[4] C. Alippi, E. Casagrande, F. Scotti, and V. Piuri, "Composite real-time image processing for railways track profile measurement," *IEEE Trans. Instrum. Meas.*, vol. 49, no. 3, pp. 559–564, Jun. 2000.

[5] L. Liu, F. Zhou, and Y. He, "Automated visual inspection system for bogie block key under complex freight train environment," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 1, pp. 2–14, Jan. 2016.

[6] P. L. Mazzeo, M. Nitti, E. Stella, and A. Distante, "Visual recognition of fastening bolts for railroad maintenance," *Pattern Recognit. Lett.*, vol. 25, no. 6, pp. 669–677, 2004.

[7] Ç. Aytekin, Y. Rezaeitabar, S. Dogru, and İ. Ulusoy, "Railway fastener inspection by real-time machine vision," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 7, pp. 1101–1107, Jul. 2015.

[8] H. Feng *et al.*, "Automatic fastener classification and defect detection in vision-based railway inspection systems," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 4, pp. 877–888, Apr. 2014.

[9] C. Mandriota, M. Nitti, N. Ancona, E. Stella, and A. Distante, "Filter-based feature selection for rail defect detection," *Mach. Vis. Appl.*, vol. 15, no. 4, pp. 179–185, 2004.

[10] D. Mery *et al.*, "Modern computer vision techniques for X-ray testing in baggage inspection," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 4, pp. 682–692, Apr. 2017.

[11] X. Xie, "A review of recent advances in surface defect detection using texture analysis techniques," *Electron. Lett. Comput. Vis. Image Anal.*, vol. 7, no. 3, pp. 1–22, 2008.

[12] A. Caprioli, A. Cigada, and D. Raveglia, "Rail inspection in track maintenance: A benchmark between the wavelet approach and the more conventional Fourier analysis," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 631–652, 2007.

[13] M. Molodova, Z. Li, A. Nunez, and R. Dollevoet, "Monitoring the railway infrastructure: Detection of surface defects using wavelets," in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, 2013, pp. 1316–1321.

[14] A. K. Dubey and Z. A. Jaffery, "Maximally stable extremal region marking-based railway track surface defect sensing," *IEEE Sensors J.*, vol. 16, no. 24, pp. 9047–9052, Oct. 2016.

[15] X. Giben, V. M. Patel, and R. Chellappa, "Material classification and semantic segmentation of railway track images with deep convolutional neural networks," in *Proc. IEEE Int. Conf. Image Process.*, Québec City, QC, Canada, 2015, pp. 621–625.

[16] X. Gibert, V. M. Patel, and R. Chellappa, "Deep multitask learning for railway track inspection," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 1, pp. 153–164, Jan. 2017.

[17] D. F. Cannon, K.-O. Edel, S. L. Grassie, and K. Sawley, "Rail defects: An overview," *Fatigue Fracture Eng. Mater. Struct.*, vol. 26, no. 10, pp. 865–886, 2003.

[18] R. Clark, "Rail flaw detection: Overview and needs for future developments," *NDT E Int.*, vol. 37, no. 2, pp. 111–118, 2004.

[19] E. Resendiz, J. M. Hart, and N. Ahuja, "Automated visual inspection of railroad tracks," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 751–760, Jun. 2013.

[20] Q. Li and S. Ren, "A real-time visual inspection system for discrete surface defects of rail heads," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 8, pp. 2189–2199, Aug. 2012.

[21] Q. Li and S. Ren, "A visual detection system for rail surface defects," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 6, pp. 1531–1542, Nov. 2012.

[22] Z. He, Y. Wang, F. Yin, and J. Liu, "Surface defect detection for high-speed rails using an inverse P-M diffusion model," *Sensor Rev.*, vol. 36, no. 1, pp. 86–97, 2016.

[23] H. Trinh, N. Haas, Y. Li, C. Otto, and S. Pankanti, "Enhanced rail component detection and consolidation for rail track inspection," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Breckenridge, CO, USA, 2012, pp. 289–295.

[24] J. Gan, Q. Li, J. Wang, and H. Yu, "A hierarchical extractor-based visual rail surface inspection system," *IEEE Sensors J.*, vol. 17, no. 23, pp. 7935–7944, Dec. 2017.

[25] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[26] Q. Wang, Y. Yuan, P. Yan, and X. Li, "Saliency detection by multiple-instance learning," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 660–672, Apr. 2013.

[27] N. D. B. Bruce and J. K. Tsotsos, "Saliency based on information maximization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 155–162.

[28] T. Liu *et al.*, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 353–367, Feb. 2011.

[29] Q. Wang, Y. Yuan, and P. Yan, "Visual saliency by selective contrast," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 7, pp. 1150–1155, Jul. 2013.

[30] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[31] S. He, R. W. Lau, W. Liu, Z. Huang, and Q. Yang, "SuperCNN: A superpixelwise convolutional neural network for salient object detection," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 330–344, 2015.

[32] S. S. S. Kruthiventi, V. Gudisa, J. H. Dholakiya, and R. V. Babu, "Saliency unified: A deep architecture for simultaneous eye fixation prediction and salient object segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, 2016, pp. 5781–5790.

[33] A. Borji, M.-M. Cheng, Q. Hou, H. Jiang, and J. Li, "Salient object detection: A survey," *arXiv:1411.5878*, 2014. [Online]. Available: https://arxiv.org/abs/1411.5878

[34] S. Dhar, V. Ordonez, and T. L. Berg, "High level describable attributes for predicting aesthetics and interestingness," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Providence, RI, USA, 2011, pp. 1657–1664.

[35] Y.-G. Jiang *et al.*, "Understanding and predicting interestingness of videos," in *Proc. AAAI*, 2013, pp. 1113–1119.

[36] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proc. 14th ACM Int. Conf. Multimedia*, 2006, pp. 815–824.
[37] S.-H. Zhong, Y. Liu, F. Ren, J. Zhang, and T. Ren, "Video saliency detection via dynamic consistent spatio-temporal attention modelling," in *Proc. AAAI*, 2013, pp. 1063–1069.
[38] T. Bouwmans, "Recent advanced statistical background modeling for foreground detection—A systematic survey," *Recent Patents Comput. Sci.*, vol. 4, no. 3, pp. 147–176, 2011.
[39] S. Varadarajan, P. Miller, and H. Zhou, "Spatial mixture of Gaussians for dynamic background modelling," in *Proc. 10th IEEE Int. Conf. Adv. Video Signal Based Surveillance (AVSS)*, 2013, pp. 63–68.
[40] A. Staglianò, N. Noceti, A. Verri, and F. Odone, "Online space-variant background modeling with sparse coding," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2415–2428, Aug. 2015.
[41] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1778–1792, Nov. 2005.
[42] C. Benedek and T. Sziranyi, "Bayesian foreground and shadow detection in uncertain frame rate surveillance videos," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 608–621, Apr. 2008.
[43] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Fort Collins, CO, USA, 1999, p. 525.
[44] D.-S. Lee, "Effective Gaussian mixture learning for video background subtraction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 827–832, May 2005.
[45] Q. Zhu, Z. Song, Y. Xie, and L. Wang, "A novel recursive Bayesian learning-based method for the efficient and accurate segmentation of video with dynamic background," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3865–3876, Sep. 2012.
[46] M. Wu and X. Peng, "Spatio-temporal context for codebook-based dynamic background subtraction," *AEU Int. J. Electron. Commun.*, vol. 64, no. 8, pp. 739–747, 2010.
[47] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.
[48] T. Chen, Y. Wang, C. Xiao, and Q. M. J. Wu, "A machine vision apparatus and method for can-end inspection," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 9, pp. 2055–2066, Sep. 2016.
[49] Y. Gan and Q. Zhao, "An effective defect inspection method for LCD using active contour model," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 9, pp. 2438–2445, Sep. 2013.
[50] X. Tao, Z. Zhang, F. Zhang, and D. Xu, "A novel and effective surface flaw inspection instrument for large-aperture optical elements," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 9, pp. 2530–2540, Sep. 2015.
[51] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1337–1342, Oct. 2003.
[52] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.
[53] M. Graves and B. Batchelor, *Machine Vision for the Inspection of Natural Products*. London, U.K.: Springer-Verlag, 2003. [Online]. Available: https://www.springer.com/us/book/9781852335250
[54] C. Steger, M. Ulrich, and C. Wiedemann, *Machine Vision Algorithms and Applications*. Weinheim, Germany: Wiley, 2008.
[55] J. Beyerer, F. P. León, and C. Frese, "Machine vision: Automated visual inspection: Theory, practice and applications," *Consciousness Cogn.*, vol. 2, no. 2, pp. 89–108, 2015.
[56] Y. Liu, H. Yao, W. Gao, X. Chen, and D. Zhao, "Nonparametric background generation," *J. Visual Commun. Image Represent.*, vol. 18, no. 3, pp. 253–263, 2007.

Jinrui Gan is currently pursuing the Ph.D. degree in computer science and technology with the School of Computer Science and Technology, Beijing Jiaotong University, Beijing, China.

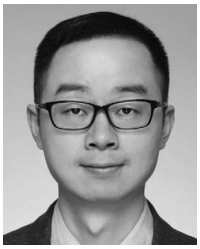
His current research interests include computer vision, image processing, pattern recognition, and machine learning.

Jianzhu Wang received the B.Sc. degree in mathematics and applied mathematics from Shandong Agricultural University, Tai'an, China, in 2015. He is currently pursuing the Ph.D. degree in computer science and technology with the School of Computer Science and Technology, Beijing Jiaotong University, Beijing, China.

His current research interests include computer vision and pattern recognition.

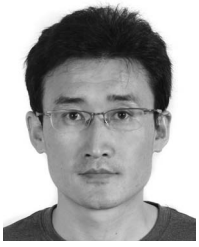
Haomin Yu received the B.Sc. degree in Internet of Things engineering from the Department of Computer Science and Technology, Anhui Normal University, Wuhu, China, in 2016. She is currently pursuing the Ph.D. degree in computer science and technology with the School of Computer Science and Information Technology, Beijing Jiaotong University, Beijing, China.

Her current research interests include computer vision and artificial intelligence.



Qingyong Li (M'09) received the B.Sc. degree in computer science and technology from Wuhan University, Wuhan, China, in 2001 and the Ph.D. degree in computer science and technology from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2006.

He is currently a Professor with Beijing Jiaotong University, Beijing. His current research interests include computer vision and artificial intelligence.



Zhiping Shi received the B.S. degree in engineering and the M.S. degree in application of computer science from the Inner Mongolia University of Technology, Hohhot, China, in 1995 and 2002, respectively, and the Ph.D. degree in computer software and theory from the Institute of Computing Technology, Chinese Academy of Science, Beijing, China, in 2005.

He is a Professor with the College of Information Engineering, Capital Normal University, Beijing. His current research interests include image understanding, machine learning, and formal method.