

# Unsupervised Saliency Detection of Rail Surface Defects Using Stereoscopic Images

Menghui Niu , Kechen Song , Liming Huang , Qi Wang , Yunhui Yan ,  
and Qinggang Meng , *Senior Member, IEEE*

**Abstract**—Visual information is increasingly recognized as a useful method to detect rail surface defects due to its high efficiency and stability. However, it cannot sufficiently detect a complete defect in the complex background information. The addition of surface profiles can effectively improve this by including a 3-D information of defects. However, in high-speed detection, the traditional 3-D profile acquisition is difficult and separate from the image acquisition, which cannot satisfy the above-mentioned requirements effectively. Therefore, an unsupervised stereoscopic saliency detection method based on a binocular line-scanning system is proposed in this article. This method can simultaneously obtain a highly precise image as well as profile information while also avoids the decoding distortion of the structured light reconstruction method. In our method, a global low-rank nonnegative reconstruction algorithm with a background constraint is proposed. Unlike the low-rank recovery model, the algorithm has a more comprehensive low rank and background clustering properties. Furthermore, outlier detection based on the geometric properties of the rail surface is also proposed in this method. Finally, the image saliency results and depth outlier detection results are associated with the collaborative fusion, and a dataset (RSSDS-113) containing the rail surface defects is established for the experimental verification. The experimental results demonstrate that our method can obtain a mean absolute error of 0.09 and area under the ROC curve of 0.94, better than 15 state-of-the-art algorithms.

**Index Terms**—Line-scanning system, low-rank nonnegative reconstruction, rail defects detection, saliency, stereoscopic.

Menghui Niu, Kechen Song, Liming Huang, Qi Wang, and Yunhui Yan are with the School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China, and also with the Key Laboratory of Vibration and Control of Aero-Propulsion Systems Ministry of Education of China, Northeastern University, Shenyang 110819, China (e-mail: menghuiniu@163.com; songkc@me.neu.edu.cn; lmhuang95@gmail.com; 1810109@stu.neu.edu.cn; yanyh@mail.neu.edu.cn).

Qinggang Meng is with the Department of Computer Science, Loughborough University, LE11 3TU Loughborough, U.K. (e-mail: q.meng@lboro.ac.uk).

## I. INTRODUCTION

**S**URFACE inspection plays a pivotal role in improving the rail manufacturing process. Different from the online rail surface inspection system [1], [2], the manufacturing process inspection of the rail products faces the following problems.

- 1) The working environment is harsh.
- 2) The appearance of the rolled rail surface is complicated.
- 3) The generation and distribution of defects are random.
- 4) The rail transmission has a high speed and vibration.
- 5) The surface detection system needs to work continuously.

Manual inspection is inefficient and has a low sensitivity [2], [32], which will slow down the entire manufacturing process. Compared with the manual inspection, nondestructive testing methods, such as ultrasonic testing, acoustic emission testing, and magnetic flux leakage [3], have a higher sensitivity and data interaction in the industrial inspection. However, most of them are time intensive [2] to the target motion and the detected defects lack the intuitive descriptions of defects' shapes. A machine vision has more advantages in surface defects testing. Especially with the development of the hardware and software, the application of machine vision technology in the railway measurement has attracted more attention.

Machine vision detection systems usually rely on analysis of images content [3], [4], and defects' characteristics [33], [34]. At present, most of them are based on grayscale images, which is due to the high detection speed and the low cost of the equipment. For some recent studies, such as mentioned in [3], the improved completed local binary patterns (ICLBP) and generalized completed local binary patterns (GCLBP) features of the grayscale images are constructed for the detection. In [4], the texture features of the defect-free grayscale images are analyzed for judging the steel strip defects. However, these methods of using grayscale images may cause a misjudgment in defect detection due to the lack of color information.

In some surface contour inspection fields, a 3-D profile reconstruction is one of the commonly used methods. Enzberg and Al-Hamadi [5] use a 3-D surface quality testing by establishing a surface approximation model. Lilienblum and Al-Hamadi [6] use the structured light and line-scanning stereo matching to obtain the 3-D information. However, the above-mentioned methods have to face the difficulty of the encode and decode structured light fringes, and the reconstruction accuracy depends on the resolution of stripes. Besides, the 3-D information and the image information cannot be uniformly used in the detection process that the two processes are independent.

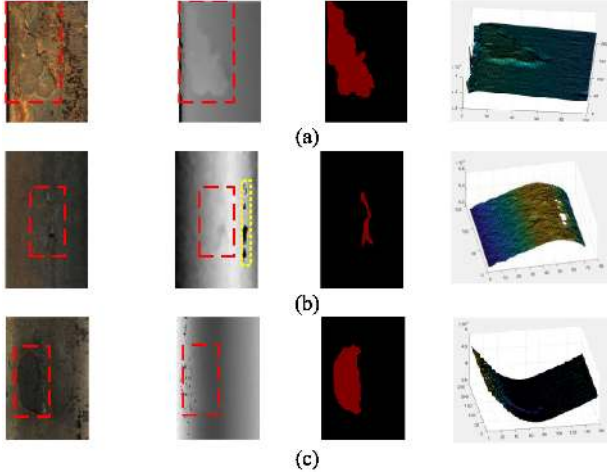


Fig. 1. Some defects' images of the rail surface. (a) Scar defect I. (b) Hole defect. (c) Scar defect II. From left to right: original image, depth map, ground truth, 3-D points cloud. The red box is the defect and the dark area in yellow box is the information missing area (which has false depth information).

The abovementioned image-based and 3-D profile-based detection methods can obtain impressive detection results but they still face the following problems.

- 1) Image information is susceptible to the illumination and shooting angles. The low-texture defects are difficult to be detected in the similar texture background area, as shown in Fig. 1(a). In particular, the grayscale images may detect more false defects due to the lack of color information.
- 2) In the profile detection, some local distortion or loss of depth information will lead to false detection, as shown in the yellow box area of Fig. 1(b). Besides, the defects with a small depth change are easily submerged in a large curvature background, as shown in Fig. 1(c).

In summary, the single information in the abovementioned description is not enough to detect defects. A multimodal fusion detection method with the image information and depth information is necessary.

In this article, a new stereoscopic saliency detection method based on a stereoscopic visual system is proposed. The visual system can quickly acquire the 2-D image and 3-D profile of a rail surface.

The main contributions of this article are as follows.

- 1) The application of a binocular line-scanning system in surface defect detection is a pioneering work that can serve as a reference for other industrial fields. Under the support of the system, an unsupervised saliency detection method based on the stereoscopic images is proposed. The method integrates a variety of information (2-D image and 3-D profile) and the rail data are used as a case for the experimental verification.
- 2) A global low-rank and nonnegative reconstruction (GLRNNR) saliency algorithm is constructed for the image defects detection. Compared with the traditional low-rank recovery (LRR) model, the algorithm incorporates

the background information constraints and nonnegative constraints, which can make our model has clustering properties under the constraint of global low rank.

- 3) According to the geometrical and depth information of the rail surface, a method for detecting the rail surface outlier by constructing the indirect plane is proposed. Then, the image saliency results and the depth saliency results are combined to get the final defect saliency map.
- 4) A total of 113 stereoscopic image pairs are collected and organized into the rail surface defects dataset (RSDDS-113), which include the left-camera images and the corresponding depth maps.

The rest of this article is organized as follows. Section II introduces some related work. Section III describes the hardware acquisition system. In Section IV, the 2-D image saliency detection method is explained and the 3-D profile outlier detection algorithm based on the indirect plane and geometric properties is elaborated, and then the abovementioned two detection results are fused. In Section V, 15 methods are used for experimental comparison and analysis. Finally, Section VI concludes this article.

## II. RELATED WORK

Saliency detection is inspired by the human visual system (HVS), to detect the region of interest with the visual uniqueness in the face of complex scenes [2], [7]–[9]. By selectively acquiring the relevant regional information, HVS can greatly avoid computational waste and reduce the difficulty of the image analysis.

In recent years, saliency models based on machine learning, the low-rank sparse principle [8], and graph theory [9] have achieved many good results. Especially, the LRR model is often used to distinguish the background and foreground, due to its not requiring a large number of labeled samples for training.

The LRR model (1) treats the background feature as a low-rank matrix  $L$ . The decomposed error is treated as foreground information  $S$  with a sparse noise.  $F$  denotes the original image feature matrix.  $\lambda$  is the balance factor, defined as follows:

$$\arg \min_{L, S} \text{rank}(L) + \lambda \|S\|_0 \text{ s.t. } F = L + S. \quad (1)$$

Since the original LRR model [8], [11] is a nondeterministic polynomial time problem, the kernel norm  $\|\cdot\|_*$  and 1-norm  $\|\cdot\|_1$  are used to solve (1). The abovementioned formula can be redescribed as follows:

$$\begin{aligned} & \arg \min_{L, S} \|L\|_* + \lambda \|S\|_1 \text{ s.t. } F = L + S \\ & \text{or} \\ & \arg \min_S \|Z\|_* + \lambda \|S\|_{2,1} \text{ s.t. } F = DZ + S \end{aligned} \quad (2)$$

where  $Z$  denotes the reconstruction coefficients of  $F$  with respect to the dictionary  $D$ . In some studies,  $D$  is played by an over-complete dictionary or directly replaced by  $F$ , namely  $F = FZ + S$ .

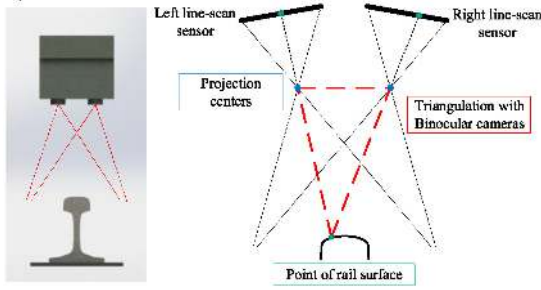


Fig. 2. Binocular stereo camera and triangulation.

The following problems usually exist in LRR models.

- 1) The construction of the over-complete dictionary  $D$  is tedious and difficult, especially when the number of samples is insufficient.
- 2) When the difference between the foreground and background is tiny, the foreground area also has a low-rank characteristic.
- 3) Finally, the negative values of  $Z$  lack a clear and reasonable guiding explanation, e.g., the cluster indication for the saliency analysis of the image.

In order to effectively solve the abovementioned problem, a GLRNNR saliency detection algorithm is used in image defects detection. The algorithm is based on the background *priori* constraint, and the nonnegative low-rank sparse constraint.

### III. HARDWARE ACQUISITION SYSTEM

In this article, a binocular line-scanning system is creatively implemented for surface detection. Compared with the existing visual inspection equipment, this system is based on a binocular stereo camera (BSC), which is produced by Chromasens. It can obtain 2-D and 3-D information at a high speed and a high resolution. Meanwhile, it can effectively preclude the nonuniform between the 2-D and 3-D information acquisition during the detection process and provides more effective information for the detection of surface defects.

The optical resolution of BSC can reach  $70 \mu\text{m}/\text{pixel}$ ; the maximum acquisition speed can be  $1.4 \text{ m/s}$ ; the maximum frame rate is  $21 \text{ kHz}$ ; and the line-scanning camera has 7142 pixels and an Red, Green and Blue color (RGB) three-channel sensor.

According to the principle of the triangulation and polar line correction, as shown in Fig. 2, the disparity map of the left and right cameras can be obtained by the stereo matching [35], [36]. Then, the corresponding depth map can be calculated based on the disparity map by combining the internal and external parameters of the cameras of BSC, as shown in Fig. 3.

In order to accelerate the calculation of the depth information, a simple and fast stereo matching algorithm called the semiglobal matching (SGM) [12] is used. The left and right consistency detection error of SGM is “10” and the matching disparity range is between  $-165$  and  $+165$ . Based on the abovementioned conditions, the BSC can provide the high-precision depth information with a resolution of  $14 \mu\text{m}$  in a range of  $52 \text{ mm}$ .

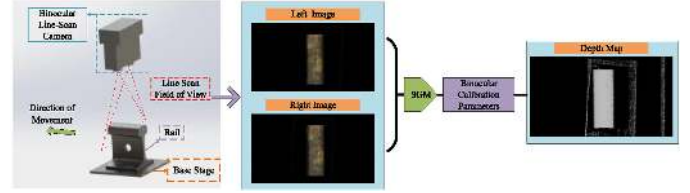


Fig. 3. Acquired depth information by the stereo matching.

## IV. PROPOSED METHOD

### A. Saliency Detection With GLRNNR

In general, the homogeneity region of the image has similar saliency information. In order to decrease the computational complexity, the simple linear iterative clustering [13] is used to segment the original image into  $n$  homogeneous regions. The  $m$ -dimensional feature information about each region is utilized to construct an overall image feature matrix  $F = \{f_1, f_2, \dots, f_n\} \in R^{m \times n}$ .

As previously stated, since the negative values in the coefficient matrix  $Z$  lack a reasonable explanation for the actual cluster, then according to the nonnegative low-rank and sparse graph (NNLRS) [14] model, the LRR model can be changed to the following representation with a nonnegative coefficient constraint:

$$\begin{aligned} & \arg \min_Z \|Z\|_* + \beta \|H\|_1 + \lambda \|S\|_{2,1} \\ & \text{s.t. } F = DZ + S, Z = H, H \geq 0. \end{aligned} \quad (3)$$

Otherwise, to improve the detection effect, some *priori* information [10], [23] are used, such as the boundary background *priori*, the center *priori*, and the color *priori*. However, the use of these *priori* information is separate from the solving process of LRR in some saliency models, which results in an inability to improve the results of LRR.

To solve the above-mentioned problems, a GLRNNR is utilized and transformed into the following equation optimization problem. In the process of the low-rank decomposition, not only the image boundary information is invoked as the dictionary  $B$  to carry on the background *priori* constraint but also the coefficient matrix  $Z$  is also nonnegative constrained. Meanwhile,  $L$  is used to ensure the low-rank property of the global background of the image

$$\begin{aligned} & \arg \min_{L, Z, S_1, S_2} \|L\|_* + \|Z\|_* + \alpha \|S_1\|_{2,1} + \beta \|S_2\|_{2,1} \\ & \quad + \lambda \|H_1\|_1 + \eta \|H_2\|_1 \\ & \text{s.t. } F = L + S_1, L = H_1, L = BZ + S_2, \\ & \quad Z = H_2, H_2 \geq 0 \end{aligned} \quad (4)$$

where  $L$  denotes the global low-rank term and  $S_1$  means the global sparse term.  $B$  stands for the background dictionary,  $L = BZ + S_2$  is the boundary background constraint reconstruction,  $Z$  indicates the reconstruction coefficient,  $S_2$  means the reconstruction error,  $H_2 \geq 0$  signifies the nonnegative

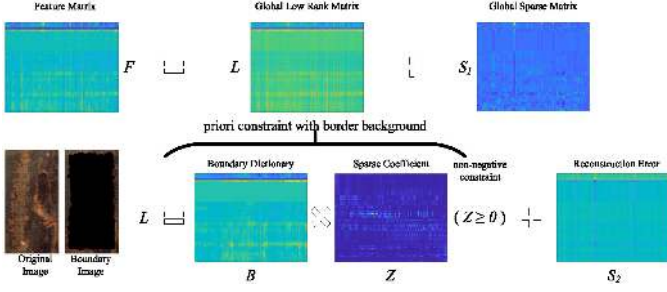


Fig. 4. GLRNNR decomposition with the boundary *priori* and nonnegative coefficient constraint.

TABLE I  
GLRNNR DECOMPOSITION WITH LADMAP

Algorithm	Algorithm for GLRNNR
Input:	feature matrix: $F$ , parameters: $\alpha > 0, \beta > 0, \gamma \geq 0, \eta > 0$ , maximum number of iterations: $\max\_iters$
Initialize:	$L_0 = S_{1,0} = H_{1,0} = S_{2,0} = Y_{1,0} = Y_{2,0} = Y_{3,0} = 0, Z = H_2 = Y_{4,0} = 0;$ $\mu_{1,0} = 0.3, \mu_{1,max} = 10^9, \mu_{2,0} = 0.2, \mu_{2,max} = 10^9, \kappa_0 = 1.2, \nu_1 = 10^{-5}, \nu_2 = 0.015, k = 0.$ $(\ F - L_k - S_{1,k}\ _F / \ F\ _F \geq \nu_1) \text{ or } (\ L_k - BZ_k - S_{2,k}\ _F / \ L_k\ _F \geq \nu_1)$
While:	$\text{or } (\mu_{1,k} \max(\ L_k - L_{k-1}\ _F, \ H_{1,k} - H_{1,k-1}\ _F, \ S_{1,k} - S_{1,k-1}\ _F) / \ F\ _F \geq \nu_2)$ $\text{or } (\mu_{2,k} \max(\ B\ _2 \cdot \ Z_k - Z_{k-1}\ _F, \ H_{2,k} - H_{2,k-1}\ _F, \ S_{2,k} - S_{2,k-1}\ _F) / \ L_k\ _F \geq \nu_2)$ $\text{or } (k < \max\_iters)$
Update	$(L, Z, S_1, S_2, H_1, H_2)$ as Step 1-6
Update	$(Y_1, Y_2, Y_3, Y_4)$ as Step 7
Update	$\mu_1, \mu_2$ as follow:
	$\mu_{1,k+1} = \min(\mu_{1,max}, \kappa_1 \cdot \mu_{1,k})$ $\kappa_1 = \begin{cases} \kappa_0 & \text{if } \mu_{1,k} \max(\ L_k - L_{k-1}\ _F, \ H_{1,k} - H_{1,k-1}\ _F, \ S_{1,k} - S_{1,k-1}\ _F) / \ F\ _F < \nu_2 \\ 1 & \text{others} \end{cases}$
	$\mu_{2,k+1} = \min(\mu_{2,max}, \kappa_2 \cdot \mu_{2,k})$ $\kappa_2 = \begin{cases} \kappa_0 & \text{if } \mu_{2,k} \max(\ B\ _2 \cdot \ Z_k - Z_{k-1}\ _F, \ H_{2,k} - H_{2,k-1}\ _F, \ S_{2,k} - S_{2,k-1}\ _F) / \ L_k\ _F < \nu_2 \\ 1 & \text{others} \end{cases}$
	$k = k + 1;$ end While
Output:	$(L, Z, S_1, S_2, H_1, H_2)$

constraint of  $Z$ , and  $\alpha, \beta, \lambda$ , and  $\eta$  are the balance factors, as shown in Fig. 4.

The abovementioned problem can be converted into the augmented Lagrangian function as follows:

$$\begin{aligned}
LA(L, Z, S_1, S_2) = & \|L\|_* + \|Z\|_* + \alpha \|S_1\|_{2,1} + \beta \|S_2\|_{2,1} + \lambda \|H_1\|_1 \\
& + \eta \|H_2\|_1 + \langle Y_1, F - L - S_1 \rangle + \langle Y_2, L - H_1 \rangle \\
& + \langle Y_3, L - BZ - S_2 \rangle + \langle Y_4, Z - H_2 \rangle \\
& + \frac{\mu_1}{2} \left( \|F - L - S_1\|_F^2 + \|L - H_1\|_F^2 \right) \\
& + \frac{\mu_2}{2} \left( \|L - BZ - S_2\|_F^2 + \|Z - H_2\|_F^2 \right)
\end{aligned} \quad (5)$$

where  $Y_1, Y_2, Y_3$ , and  $Y_4$  are the Lagrange multipliers, and  $\mu_1$  and  $\mu_2$  are the penalty for violating the linear constraints.  $L, Z, S_1$ , and  $S_2$  can be solved by alternating iterations with LADMAP [14], as given in Table I.

The detailed iterative update steps are as follows.

Step 1. *Updating L*: According to Zhuang *et al.* [14],  $\Psi_{(\tau)}(\cdot)$  expresses the singular-value shrinkage algorithm with soft threshold  $\tau$ , which is used to approximate the calculation of the kernel paradigm. The result of  $L$  obtained by the  $(k+1)$ th iteration is as follows:

$$L_{k+1} = \Psi_{\left(\frac{1}{2\mu_{1,k} + \mu_{2,k}}\right)} \left( \frac{\mu_{1,k}(T_{1,k} + T_{2,k}) + \mu_{2,k}T_{3,k}}{2\mu_{1,k} + \mu_{2,k}} \right) \quad (6)$$

$$\text{where } \begin{cases} T_{1,k} = F - S_{1,k} + Y_{1,k}/\mu_{1,k} \\ T_{2,k} = B \times Z_k + S_{2,k} - Y_{3,k}/\mu_{2,k} \\ T_{3,k} = H_{1,k} - Y_{2,k}/\mu_{1,k} \end{cases}$$

Step 2. *Updating H<sub>1</sub>*:

$$H_{1,k+1} = \Psi_{\left(\frac{\lambda}{\mu_{1,k}}\right)} (L_{k+1} - Y_{2,k}/\mu_{1,k}). \quad (7)$$

Step 3. *Updating S<sub>1</sub>*:

$$S_{1,k+1} = \Psi_{\left(\frac{\alpha}{\mu_{1,k}}\right)} (F - L_{k+1} + Y_{1,k}/\mu_{1,k}). \quad (8)$$

Step 4. *Updating Z*:

$$\begin{aligned}
Z_{k+1} = & \Psi_{\left(\frac{1}{\mu_{2,k} \|B\|_2^2}\right)} \\
& \times \left( Z_k + \frac{B^T (L_{k+1} - BZ - S_{2,k} + Y_{3,k}/\mu_{2,k})}{\|B\|_2^2} \right. \\
& \left. - \frac{(Z - H_{2,k} + Y_{4,k}/\mu_{2,k})}{\|B\|_2^2} \right). \quad (9)
\end{aligned}$$

Step 5. *Updating H<sub>2</sub>*: It is needed to perform the nonnegative constraints during the process of updating.

$$H_{2,k+1} = \max(\Psi_{\left(\frac{\eta}{\mu_{2,k}}\right)} (Z_{k+1} - Y_{4,k}/\mu_{2,k}), 0). \quad (10)$$

Step 6. *Updating S<sub>2</sub>*:

$$S_{2,k+1} = \Psi_{\left(\frac{\beta}{\mu_{2,k}}\right)} (L_{k+1} - BZ_{k+1} + Y_{3,k}/\mu_{2,k}). \quad (11)$$

Step 7. *Updating Y<sub>1</sub>, Y<sub>2</sub>, Y<sub>3</sub>, Y<sub>4</sub>*: The detailed iterative update steps for Lagrange multipliers ( $Y_1, Y_2, Y_3, Y_4$ ) are as follows:

$$\begin{cases} Y_{1,k+1} = Y_{1,k} + \mu_{1,k} (F - L_{k+1} - S_{1,k+1}) \\ Y_{2,k+1} = Y_{2,k} + \mu_{1,k} (L_{k+1} - H_{1,k+1}) \\ Y_{3,k+1} = Y_{3,k} + \mu_{2,k} (L_{k+1} - BZ_{k+1} - S_{2,k+1}) \\ Y_{4,k+1} = Y_{4,k} + \mu_{2,k} (Z_{k+1} - H_{2,k+1}) \end{cases} \quad (12)$$

Based on  $S_1$  and  $S_2$ , the saliency  $\text{Sal}$  of the  $i$ th superpixels region in the image is calculated as follows:

$$\text{Sal}_i = \frac{1}{Z_{\text{Sal}}} \cdot \|(S_1)_i + (S_2)_i\|_2^2 \quad (13)$$

where  $(S_1)_i$  and  $(S_2)_i$  are the  $i$ th column vectors of  $S_1$  and  $S_2$ , respectively.  $Z_{\text{Sal}}$  is the normalization coefficient.

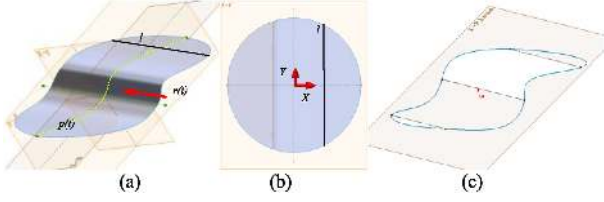


Fig. 5. (a) Spatial symmetric ruled surface region  $\Omega_S$ . (b) Projection on the  $X\_Y$  plane  $\Omega_{X\_Y}$ . (c) Indirect plane obtained by LSM.

Otherwise, a multiscale fusion method is used to obtain the final saliency map

$$\text{Sal\_map} = \sum_{j=1}^N w_j \cdot \text{Sal}_{(j)} \quad (14)$$

where the fusion is performed using  $N$  scales superpixel segmentation, and  $w_n$  and  $\text{Sal}_{(n)}$  are the corresponding weight and saliency results at the  $j$ th scale, respectively.

### B. Outlier Saliency Model With Depth

The ideal rail surface can be viewed as a ruled surface generated by a moving straight line

$$S(t, u) = p(t) + u \cdot r(t) \quad (15)$$

where  $S(t, u)$  is a point on the ruled surface. The directrix  $p(t)$  is the curved path of a moving line. The moving line is called a generator.  $r(t)$  means the unit vector of the generator passing through  $p(t)$ . If the ruled surface is a general cylinder,  $r(t)$  is fixed, e.g.,  $r(t) \equiv r_0$ .

It can be recognized that any rail cross section perpendicular to  $r(t)$  should have a similar profile. It is the key to detect the outlier of the profile.

Limited by the depth of the camera field, the indicative feature of  $r(t)$  may be missing in the image. It is particularly severe when the imaging plane is not perpendicular to the direction of the rail movement.

In order to discover  $r(t)$ , a method based on an indirect fitting plane is proposed in this article. The indirect plane ( $I$ -plane) is the approximate fitting plane of a spatially symmetric region on a ruled surface with the least square method (LSM), as shown in Fig. 5(c) and in Fig. 6(b).

In this method, it is assumed that there is a spatially symmetric area  $\Omega_S$  on a general cylindrical ruled surface and after a rigid transformation (translation and rotation),  $\Omega_S$  is converted into as shown in Fig. 5 and Fig. 6(a). The projection region of  $\Omega_S$  on the  $X\_Y$  plane is  $\Omega_{X\_Y} = \{(x, y) | x^2 + y^2 = r^2\}$ . The directrix of  $\Omega_S$  on the  $X\_Z$  plane is  $p(t)$ . The direction vector of the generator is  $r(t)$ , where  $r(t) \equiv \vec{V}_g$  and  $\vec{V}_g = (0, 1, \kappa_0)$ . Otherwise,  $l$  is one of the generators on  $\Omega_S$ . On surface  $\Omega_S$ , there are the following relationships:

$$\begin{cases} \int_l x \cdot y ds = 0 \\ \int_l y ds = 0 \\ \int_l y \cdot z_{pl} ds = 0 \end{cases} \quad \text{s.t. } l : z = \kappa_0 \cdot y + z_{pl}, z_{pl} \in p(t) \quad (16)$$

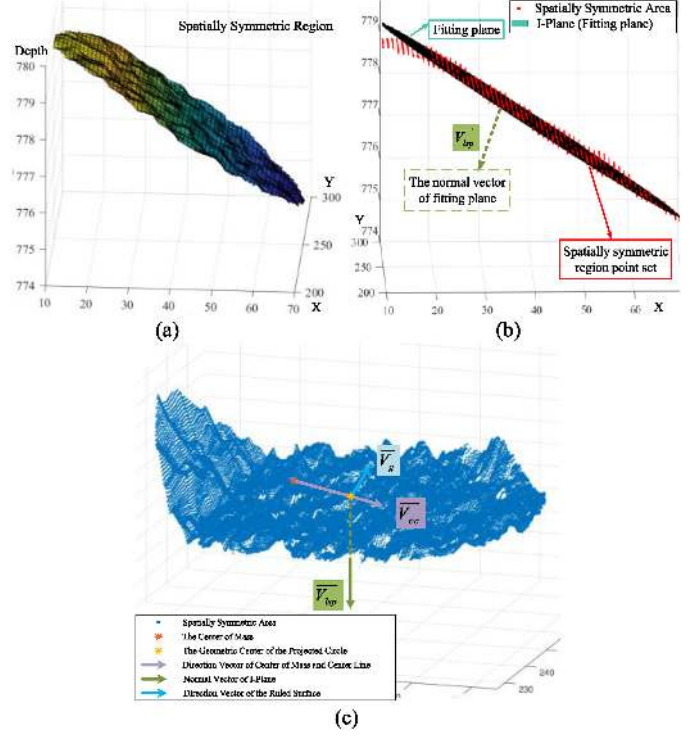


Fig. 6. (a) Spatially symmetric region. (b) Fit the  $l$ -plane with LSM. (c) Calculate the centroid of the mass and the direction  $\vec{V}_g$ .

where  $(x, y, z)$  are the points' coordinates on  $\Omega_S$ , and  $z_{pl}$  is the intersection of  $l$  and  $X\_Y$  plane.

The discrete sampling on surface  $\Omega_S$ , the sample point set  $P_S = \{(x_{P_S}, y_{P_S}, z_{P_S}) | (x_{P_S}, y_{P_S}, z_{P_S}) \in \Omega_S\}$  can be written as a discrete form, which is as follows:

$$\begin{cases} \sum x_{P_S} \cdot y_{P_S} = 0 \\ \sum y_{P_S} = 0 \\ \sum y_{P_S} \cdot z_{P_S} = \sum \kappa_0 \cdot y_{P_S}^2 \end{cases} \quad (17)$$

This article uses the LSM to find the  $I$ -plane  $a_0 \cdot x + a_1 \cdot y - z + a_2 = 0$

$$I_P = \arg \min_{a_0, a_1, a_2} \sum_{P_S} (a_0 \cdot x_{P_S} + a_1 \cdot y_{P_S} - z_{P_S} + a_2)^2. \quad (18)$$

Solving (18) by using  $\frac{\partial I_P}{\partial a_0} = 0$ ,  $\frac{\partial I_P}{\partial a_1} = 0$ ,  $\frac{\partial I_P}{\partial a_2} = 0$

$$\begin{bmatrix} \sum x_{P_S}^2 & \sum x_{P_S} \cdot y_{P_S} & \sum x_{P_S} \\ \sum x_{P_S} \cdot y_{P_S} & \sum y_{P_S}^2 & \sum y_{P_S} \\ \sum x_{P_S} & \sum y_{P_S} & n_{P_S} \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum x_{P_S} \cdot z_{P_S} \\ \sum y_{P_S} \cdot z_{P_S} \\ \sum z_{P_S} \end{bmatrix} \quad (19)$$

Known by (17) and (19),  $a_1 = \kappa_0$ . The normal vector of the  $I$ -plane is  $[a_0, \kappa_0, -1]$ , such that the  $I$ -plane is parallel to  $\vec{V}_g$ .

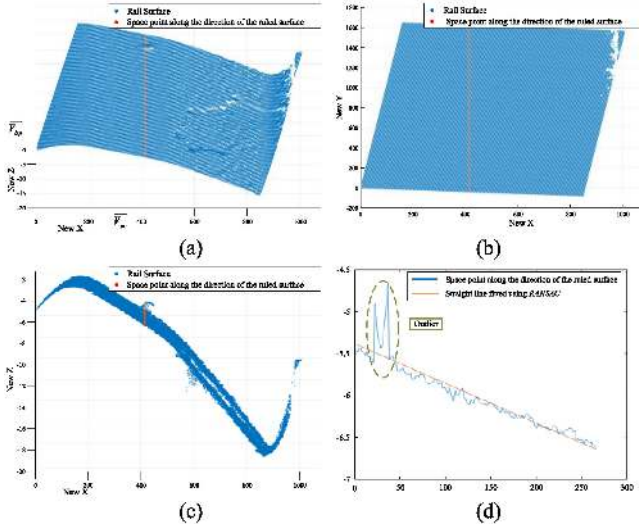


Fig. 7. (a) Three-dimensional point cloud profile in the coordinate system of  $(\vec{V}_{cc}, \vec{V}_g, \vec{V}_{lsp})$ . (b) Projection of point cloud on the new  $X-Y$  plane. (c) Projection of the profile along direction  $\vec{V}_g$ . (d) Outlier detection along direction  $\vec{V}_g$ .

The vector  $\vec{V}_{cc} = (1, 0, 0)$  is perpendicular to the direction vector  $r(t)$ .  $\vec{V}_g$  can be obtained by crossing the normal vector  $\vec{V}_{lsp} = [a_0, \kappa_0, -1]$  of the  $I$ -plane and the vector  $\vec{V}_{cc}$ , and  $\varepsilon$  is a normalized constant, as shown in Fig. 6(c).

$$\vec{V}_g = \varepsilon \cdot \vec{V}_{cc} \times \vec{V}_{lsp}. \quad (20)$$

For the calculation of  $\vec{V}_{cc}$  before rotation, it can be obtained by the center of distance mass  $P_m$  and geometric center  $P_g$ , where  $P_m = (\sum x_{P_S} \cdot d_{P_S}, \sum y_{P_S} \cdot d_{P_S}, \sum z_{P_S} \cdot d_{P_S}) / \sum d_{P_S}$  and  $d_{P_S}$  denotes the distance between  $\Omega_S$  and  $I$ -plane.  $P_g = (x_p, y_p, z_p)$  means the geometric center of  $\Omega_S$ .  $\vec{V}_{mg} = \vec{P}_g P_m$  is presented so that  $\vec{V}_{cc} = \vec{V}_{mg} - \vec{V}_{mg} \cdot \vec{V}_{lsp} \cdot \vec{V}_{lsp} / |\vec{V}_{lsp}|$ .

The original data are projected in the new coordinate system  $(\vec{V}_{cc}, \vec{V}_g, \vec{V}_{lsp})$ , as shown in Fig. 7(a). The point data perform the outlier detection of the new  $Z$ -value along  $\vec{V}_g$ , as shown in Fig. 7(d). It is worth noting that the actual surface is nonsmooth and the depth map contains the noise and errors, as shown in Fig. 7(c). Here, the RANSAC [15] algorithm is used to perform a line fit for  $Z$ -value along  $\vec{V}_g$ .  $\bar{z}_{lsp}$  represents the ideal  $Z$ -value online.

In order to weaken the influence of the noise and errors, the  $\vec{V}_g$  direction should be fine-tuned through setting the angle fine-tuning interval  $\varphi$  and minimizing the overall variance  $\xi = \arg \min_{\vec{\xi} \in \varphi} \text{var}(|z_{lsp} - \bar{z}_{lsp}|)_{\vec{V}_g + \vec{\xi}}$ .

Then, the saliency of the outlier of the rail surface is calculated as follows:

$$S_D = (|z_{lsp} - \bar{z}_{lsp}|)_{\vec{V}_g + \vec{\xi}} \quad (21)$$

where  $z_{lsp}$  represents the distance of the points on the ruled surface to the  $I$ -plane and  $\bar{z}_{lsp}$  indicates the ideal value along the direction  $\vec{V}_g + \vec{\xi}$ . It is worth noting that  $\Omega_S$  should be selected

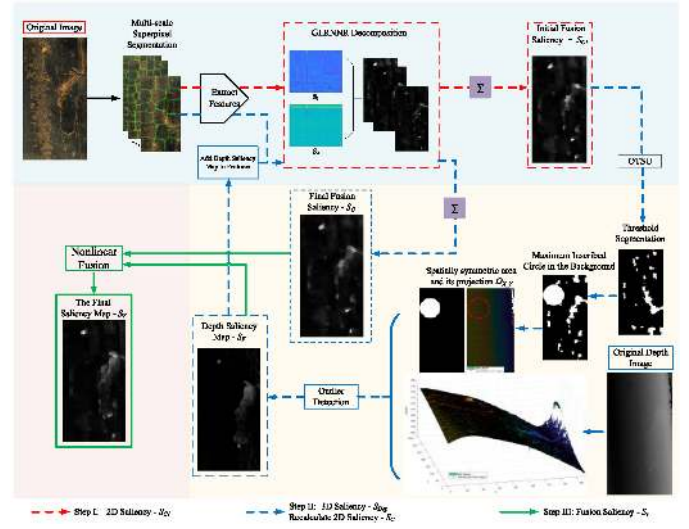


Fig. 8. Saliency detection of the rail surface defects combined with the 2-D and 3-D information.

as far as possible without defects or microdefects. As shown in Fig. 8, the 2-D saliency map is segmented into the foreground and background regions by Otsu [16]. Then, the largest inscribed circle of the background region is selected as  $\Omega_{X-Y}$ .

The main defect area is highlighted by the centroid constraint

$$S_{Dep}(x_{SD}, y_{SD}) = \frac{1}{Z_{Dep}} \times \exp \left[ - \left( \frac{(x_{SD} - x'_{SD})^2 + (y_{SD} - y'_{SD})^2}{\sigma_{xy}^2} \right) \right] \cdot S_D(x_{SD}, y_{SD}) \quad (22)$$

and

$$\begin{cases} x'_{SD} = \frac{x_{SD} \cdot S_D(x_{SD})}{\sum S_D} \\ y'_{SD} = \frac{y_{SD} \cdot S_D(y_{SD})}{\sum S_D} \end{cases} \quad (23)$$

$$\sigma_{xy} = \min(\text{Var}(x_{SD}), \text{Var}(y_{SD}))$$

where  $Z_{Dep}$  denotes the normalization constant,  $S_D(x_{SD}, y_{SD})$  means the pixel saliency value at  $(x_{SD}, y_{SD})$ , and  $\text{Var}(\cdot)$  represents the variance.

### C. Final Saliency Fusion

In order to effectively fuse the 2-D-based and 3-D-based saliency detection results, the collaborative fusion detection process, as shown in Fig. 8, is employed.

First, the initial 2-D saliency result  $S_{C1}$  is obtained by the algorithm of GLRNNR. Similar to [25], a 53-D feature vector is employed in GLRNNR.

Second,  $S_{C1}$  is subjected to the threshold segmentation by Otsu [16] to obtain  $\Omega_{X-Y}$  in the background. The 3-D detection

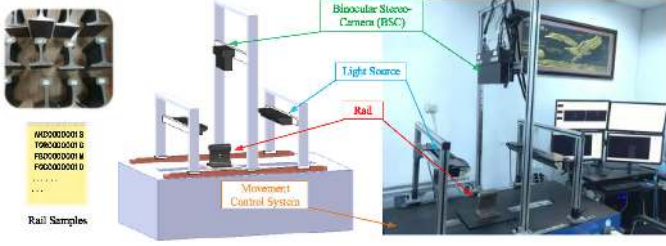


Fig. 9. Rail samples and the experimental equipment.

result  $S_{Dep}$  is obtained by the outlier detection, which is the pixel level.

According to the outlier saliency model, the detected defect may be incomplete and cover the background noise of a rough surface. Especially for the scar defect, the detection results are more obvious at the edge of the defect due to the interior of the defect is consistent with the background profile. With the superpixels segmentation and clustering properties of GLRNNR, it can expand the range of the defects and reduce the impact of the background noise.

Therefore,  $S_{Dep}$  will be seen as the 54th dimensional feature vector to recalculate the new 2-D result  $S_C$  using GLRNNR.

Finally,  $S_C$  and  $S_{Dep}$  results are nonlinearly combined as follows:

$$S_F = \frac{1}{3} \left( S_{Dep} + S_C + \sqrt{S_{Dep} \cdot S_C} \right). \quad (24)$$

## V. EXPERIMENT

### A. Dataset

1) *RSDDS-113*: The dataset samples are taken from an actual industrial production line of one section-steel factory. The 20 rail segments of them with the defects information are collected and employed to construct the rail surface defects dataset (RSDDS-113). Under the laboratory conditions, the data acquisition process is shown in Fig. 9.

The data of the samples cover all the positions of the rail, such as the waist surface, the tread surface, and the bottom surface. The types and locations of their surface defects are random. In the RSDDS-113 dataset, 113 pairs of them with the typical defects will be selected and employed. Every pair consists of the left-camera image and the corresponding depth image.

In the dataset, the main types of defects are rolling scar, corrosion, scratches, holes, pits, and so on, as shown in Fig. 10. The RSDDS-113 dataset and our codes are available at the Github homepage (<https://github.com/neu-rail-rsdds/rsdds>).

2) *Rail Surface Discrete Defect (RSDD) Dataset [2]*: The RSDD is a public railway image dataset, which is mainly composed of 2-D grayscale images captured from the express rails and heavy haul rails, including two subdatasets: Type-I and Type-II. The dataset is used to verify the applicability of GLRNNR for online railway images.

The Type-II dataset has a narrower and more consistent background than the Type-I dataset but more sophisticated defects are included, which are shown in Fig. 15.

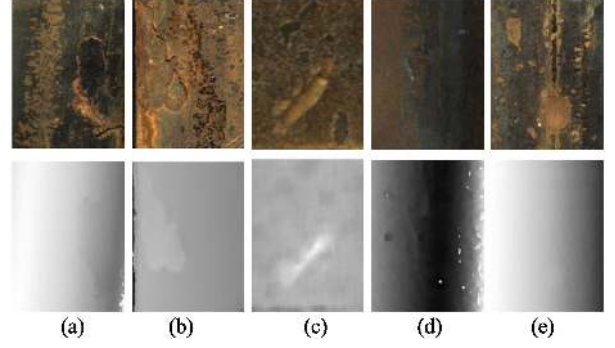


Fig. 10. Defect image (upper) and depth map (lower) of the rail surface. (a) Rolling scar. (b) Corrosion. (c) Scratches. (d) Holes. (e) Pits.

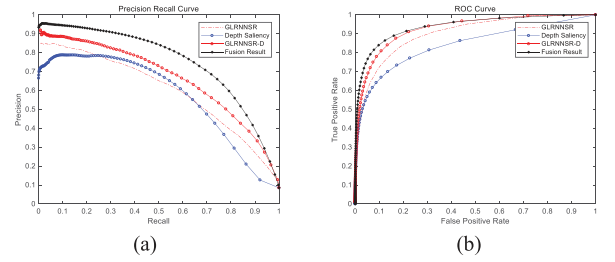


Fig. 11. (a) PR curve and (b) ROC curve for each step.

TABLE II  
MAE AND AUC OF EACH STEP OF OUR METHOD

Method	GLRNNR $S_{C1}$	Depth $S_{Dep}$	GLRNNR-D $S_C$	Final Fusion $S_F$
MAE	0.10	0.10	0.09	0.09
AUC	0.90	0.84	0.93	0.94

### B. Evaluation Metrics

For a comprehensive assessment, five evaluation metrics [17] are used for RSDD-113, including the precision–recall (PR) curve, the receiver operating characteristic (ROC) curve, area under the ROC curve (AUC), mean absolute error (MAE), and the  $F$ -measure. Two evaluation metrics (pixel-level index and defect-level index) [2] are employed for the RSDD dataset.

### C. Comparison With State-of-the-Art Methods

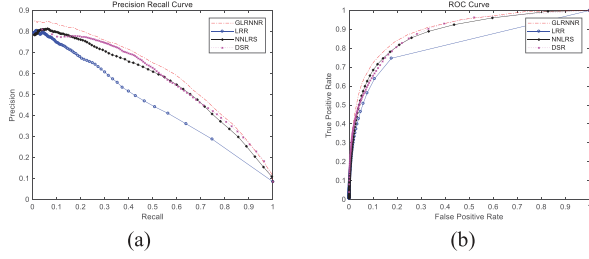
1) *RSDDS-113 Dataset*: In this article, the values of the parameters involved in this experiment are as follows:  $\alpha = 0.02$ ;  $\beta = 1.2$ ;  $\lambda = 0.02$ ;  $\eta = 1.2$  in (4). The scale of the superpixels in (14) is  $N = 5$ , namely 200–600 and  $w_n = 0.2$ , and  $\max\_iters = 200$  in Table I. Otherwise, the adjustment angle of the direction vector of the ruled surface is  $\varphi = (0, \pm 2 \sin(\pi/120), 0)$ .

The results of each step are evaluated by the above-mentioned indicators, as illustrated in Fig. 11 and Table II. The results demonstrate that each step is effective for generating the final saliency map.

As shown in Fig. 11, “depth saliency” denotes the result  $S_{Dep}$ , which is produced by the outlier detection. “GLRNNR”

**TABLE III**  
MAE AND AUC OF LRR, DSR, NNLSR, AND GLRNNR

Method	LRR	NNLSR	DSR	GLRNNR
MAE	0.09	0.11	0.12	0.10
AUC	0.85	0.88	0.88	0.90



**Fig. 12.** Comparison of LRR, DSR, NNLSR, and our GLRNNR. (a) PR curve. (b) ROC curve.

**TABLE IV**  
MAE AND AUC OF THE IMAGE SALIENCY METHODS AND OURS

Method	Itti	CA	CDCP	DRFI	DSR	GLC	HDCT	SMD	Zhou	MBS	Ours
MAE	0.3	0.2	0.2	0.09	0.12	0.12	0.12	0.11	0.11	0.15	0.09
AUC	0.7	0.9	0.8	0.86	0.88	0.89	0.90	0.85	0.81	0.90	0.94

**TABLE V**  
MAE AND AUC OF THE DEPTH SALIENCY METHODS AND OURS

Method	ACSD	CAIP-MB	DCMC	DES	LBE	Ours
MAE	0.17	0.08	0.14	0.45	0.29	0.09
AUC	0.73	0.70	0.78	0.70	0.78	0.94

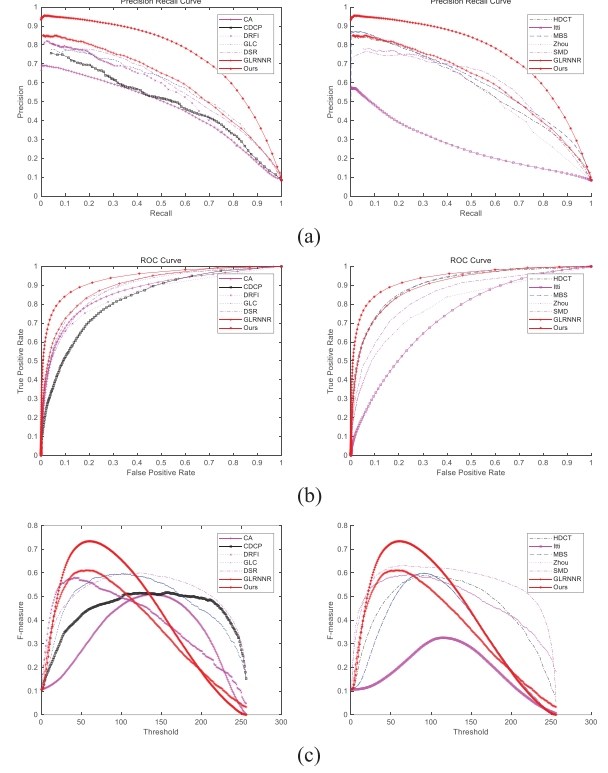
represents the saliency result  $S_{C1}$  generated using only a single color image. “GLRNNR-D” represents the saliency result  $S_C$  regenerated using  $S_{Dep}$  and the color image. “Fusion result” indicates the final merger  $S_F$  of  $S_C$  and  $S_{Dep}$ .

As illustrated in Table III and Fig. 12, compared with the algorithms, such as LRR [8], DSR [20], and NNLSR [14], the GLRNNR algorithm using only the color information is obviously superior to the abovementioned algorithm. Otherwise, it is also verified that GLRNNR with the nonnegative sparse constraint can improve the influence of the negative coefficients of DSR.

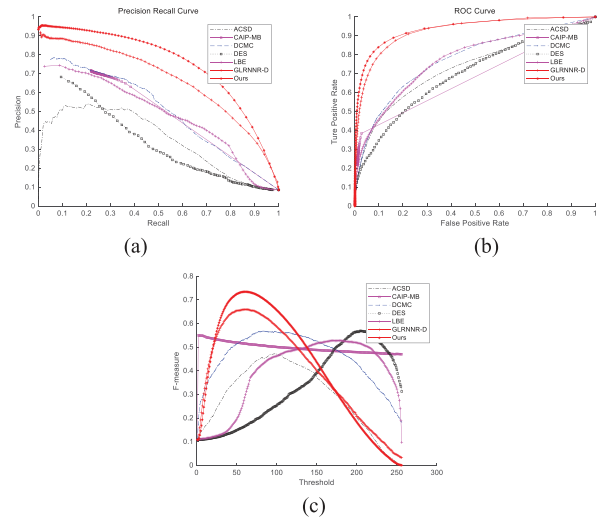
As shown in Figs. 13 and 14, 15 methods have been compared. It includes ten methods, which use only the color image, including CA [10], CDCP [11], DRFI [18], GLC [19], DSR [20], HDCT [21], Itti *et al.* [22], MBS [23], Zhou *et al.* [24], and SMD [25]. Besides, other five methods have been also compared, such as ACSD [26], CAIP-MB [27], DCMC [28], DES [29], and LBE [30], which use the image and depth information.

In order to verify the effectiveness of the proposed algorithm, “GLRNNR,” “GLRNNR-D,” and the final fusion result “Ours” are divided into groups for the experimental comparison, as shown in Figs. 13 and 14, and the MAE and AUC are also listed in Tables IV and V.

As shown in Fig. 13, the GLRNNR algorithm in this article can obtain a better result than the other methods without the depth information. The detection effect of “GLRNNR-D” is obviously better than ACSD, CAIP-MB, DCMC, DES, and LBE. The



**Fig. 13.** Comparison of CA, CDCP, DRFI, GLC, DSR, HDCT, Itti, MBS, Zhou, SMD, GLRNNR, and Ours. (a) PR curve. (b) ROC curve. (c) F-measure curve.



**Fig. 14.** Comparison of ACSD, CAIP-MB, DCMC, DES, LBE, GLRNNR-D, and Ours. (a) PR curve. (b) ROC curve. (c) F-measure curve.

MAE of “GLRNNR-D” is slightly worse than CAIP-MB in Fig. 14 and Table V but the AUC of “GLRNNR-D” is obviously higher than CAIP-MB. It is possible that CAIP-MB strengthens the requirement of accuracy and leads to a decrease in the recall rate.



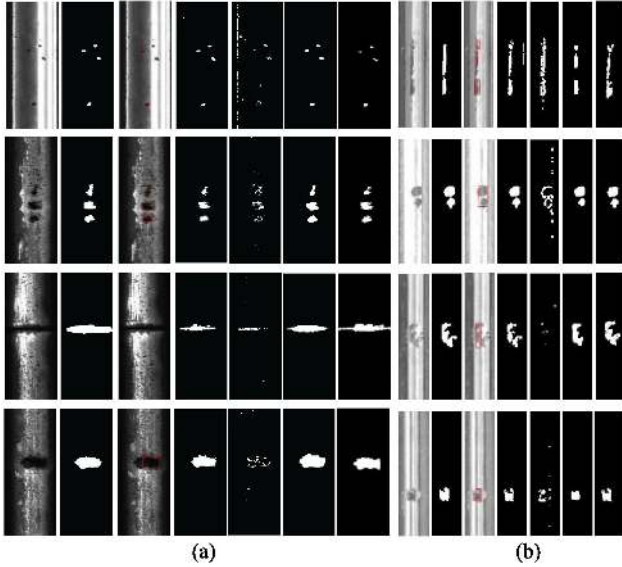


Fig. 15. Sample defect images and results of different methods for (a) Type-I and (b) Type-II of the RSDD dataset. From left to right, each column in (a) and (b) is original image, ground truth, LN + DLBP, MLC + PEME, PM, CTFM, and Ours.

The MAE is mainly for the regression problems but it is susceptible to the background noise. In Tables IV and V, the MAE of our proposed methods is equal to other excellent algorithms. To evaluate the effectiveness of defect detection, it is necessary to use some additional evaluation metrics, which are suitable for the segmentation and classification problems, such as the PR curve and AUC. It can be found that the PR curve and AUC of ours are excellent than other algorithms.

In the experimental comparisons, the final fusion results of our method are significantly better than the other 15 methods. Otherwise, when the resolution of the image is set to  $256 \times 512$  and the number of the superpixels is set to 200, “GLRNNR” takes 3.9 s, and the depth outlier detection takes 4.9 s, which does not use the graphics processing unit (GPU) acceleration.

In practical applications, offline and parallel processing methods are utilized. First, GLRNNR and 3-D outlier detection algorithms are, respectively, used for preprocessing detection. Then, if the defect requirements are not satisfied, the two pieces of information are combined using our fusion detection method in this article.

2) *RSDD Dataset*: For the RSDD dataset, the values of the parameters in (4) and (14) are the same as the RSDDS-113.

Considering the comparison with other known methods (LN+DLBP, MLC+PEME, PM, CTFM) in [2], Otsu [16] and active contours [31] are used to segment the saliency results into the foreground and background. As described in [2], the evaluation metrics of the RSDD dataset are pixel-level index (precision, recall, and  $F$ -measure) and defect-level index (precision', recall', and  $F$ -measure'). The results of the RSDD dataset are shown in Fig. 15 and the evaluation metrics of comparison are given in Tables VI and VII.

It can be ascertained that GLR is significantly better than the other algorithms in Type-I. Otherwise, the GLRNNR is equal to

TABLE VI  
EXPERIMENTAL RESULTS FOR THE TYPE-I RSDD DATASET

Method	pre (%)	rec (%)	F-measure	pre' (%)	rec' (%)	F-measure'
LN+DLBP	-	-	-	76.26	70.80	73.43
MLC+PEME	78.14	78.89	75.70	47.87	85.40	61.35
PM	-	-	-	41.19	72.94	52.65
CTFM	86.54	77.68	80.02	84.06	77.37	80.58
Ours	87.54	78.74	82.91	92.86	75.91	83.53

TABLE VII  
EXPERIMENTAL RESULTS FOR THE TYPE-II RSDD DATASET

Method	pre (%)	rec (%)	F-measure	pre' (%)	rec' (%)	F-measure'
LN+DLBP	-	-	-	88.89	71.82	79.45
MLC+PEME	73.88	83.05	76.05	58.68	91.71	71.56
PM	-	-	-	49.73	46.41	48.01
CTFM	84.12	73.25	76.45	85.83	83.98	84.89
Ours	80.31	77.86	79.06	94.81	80.66	87.16

CTFM at the accuracy and recall rate of the pixel-level index of Type-II but it is better than the other algorithms at the defect-level index.

## VI. CONCLUSION

In this article, a novel unsupervised stereoscopic saliency detection method for rail surface defects was proposed. It was based on a binocular line-scanning system, GLRNNR saliency algorithm, and depth outlier detection. First, a 2-D image and 3-D profile information of the rail surface were obtained by our developed binocular color line-scanning system. Second, utilizing LADMAP and GLRNNR, the algorithm could quickly obtain the saliency map of a 2-D image. Next, the outlier region of the 3-D profile was detected based on a 2-D saliency map and the surface characteristics. Meanwhile, the 3-D saliency map was also used as a feature to enhance the 2-D saliency map. Finally, the last 2-D saliency result and the 3-D result were nonlinearly fused. Our experimental results on the RSDDS-113 dataset outperformed the 15 state-of-the-art methods in the literature.

Furthermore, to verify the applicability to the online railway images, the gray image dataset RSDD was also used to compare with other known methods. The experimental results showed that the GLRNNR method proposed in this article is also suitable for general gray rail images and can obtain better detection results.

It is worth noting that the method in this article is only suitable for locating rail surface defects. More research is needed to determine the defect attributes. In addition, the limitations on the imaging hardware equipment hinder this method from being effectively applied to detect shallow internal defects. Therefore, in the future, we will look at the multimodal information fusion and defect classification.

## REFERENCES

- [1] J. Wang, Q. Li, J. Gan, H. Yu, and X. Yang, “Surface defect detection via entity sparsity pursuit with intrinsic priors,” *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 141–150, Jan. 2020.
- [2] H. Yu *et al.*, “A coarse-to-fine model for rail surface defect detection,” *IEEE Trans. Instrum. Meas.*, vol. 68, no. 3, pp. 656–666, Mar. 2019.

- [3] Q. Luo, Y. Sun, P. Li, O. Simpson, L. Tian, and Y. He, "Generalized completed local binary patterns for time-efficient steel surface defect classification," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 3, pp. 667–679, Mar. 2019.
- [4] H. Wang, J. Zhang, Y. Tian, H. Chen, H. Sun, and K. Liu, "A simple guidance template-based defect detection method for strip steel surfaces," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 2798–2809, May 2019.
- [5] S. von Enzberg and A. Al-Hamadi, "A multiresolution approach to model-based 3-D surface quality inspection," *IEEE Trans. Ind. Informat.*, vol. 12, no. 4, pp. 1498–1507, Aug. 2016.
- [6] E. Lilienblum and A. Al-Hamadi, "A structured light approach for 3-D surface reconstruction with a stereo line-scan system," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 5, pp. 1258–1266, May 2015.
- [7] Z. Li, L. S. Mihaylova, O. Isupova, and L. Rossi, "Autonomous flame detection in videos with a Dirichlet process Gaussian mixture color model," *IEEE Trans. Ind. Informat.*, vol. 14, no. 3, pp. 1146–1154, Mar. 2018.
- [8] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, 2012, pp. 853–860.
- [9] J.-G. Yu, J. Zhao, J. Tian, and Y. Tan, "Maximal entropy random walk for region-based visual saliency," *IEEE Trans. Cybern.*, vol. 44, no. 9, pp. 1661–1672, Sep. 2014.
- [10] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, Oct. 2012.
- [11] C. Zhu, G. Li, W. Wang, and R. Wang, "An innovative salient object detection using center-dark channel prior," in *Proc. IEEE Int. Conf. Comput. Vision Workshops*, Venice, Italy, 2017, pp. 1509–1515.
- [12] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.
- [13] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [14] L. Zhuang *et al.*, "Constructing a nonnegative low-rank and sparse graph with data-adaptive features," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3717–3728, Nov. 2015.
- [15] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.
- [16] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
- [17] R. Margolin, L. Zelnik-Manor, and A. Tal, "How to evaluate foreground maps," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Columbus, OH, USA, 2014, pp. 248–255.
- [18] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Portland, OR, USA, 2013, pp. 2083–2090.
- [19] N. Tong, H. Lu, Y. Zhang, and X. Ruan, "Salient object detection via global and local cues," *Pattern Recognit.*, vol. 48, pp. 3258–3267, 2015.
- [20] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *Proc. IEEE Int. Conf. Comput. Vision*, Sydney, NSW, Australia, 2013, pp. 2976–2983.
- [21] J. Kim, D. Han, Y.-W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform and local spatial support," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 9–23, Jan. 2016.
- [22] 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.
- [23] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Mech, "Minimum barrier salient object detection at 80 FPS," in *Proc. IEEE Int. Conf. Comput. Vision*, Santiago, Chile, 2015, pp. 1404–1412.
- [24] L. Zhou, Z. Yang, Z. Zhou, and D. Hu, "Salient region detection using diffusion process on a two-layer sparse graph," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5882–5894, Dec. 2017.
- [25] H. Peng, B. Li, H. Ling, W. Hu, W. Xiong, and S. J. Maybank, "Salient object detection via structured matrix decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 818–832, Apr. 2017.
- [26] R. Ju, L. Ge, W. Geng, T. Ren, and G. Wu, "Depth saliency based on anisotropic center-surround difference," in *Proc. IEEE Int. Conf. Image Process.*, Paris, France, 2014, pp. 1115–1119.
- [27] C. Zhu, G. Li, X. Guo, W. Wang, and R. Wang, "A multilayer backpropagation saliency detection algorithm based on depth mining," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, 2017, pp. 14–23.
- [28] R. Cong, J. Lei, C. Zhang, Q. Huang, X. Cao, and C. Hou, "Saliency detection for stereoscopic images based on depth confidence analysis and multiple cues fusion," *IEEE Signal Process. Lett.*, vol. 23, no. 6, pp. 819–823, Jun. 2016.
- [29] Y. Cheng, H. Fu, X. Wei, J. Xiao, and X. Cao, "Depth enhanced saliency detection method," in *Proc. Int. Conf. Internet Multimedia Comput. Service*, 2014, pp. 23–27.
- [30] D. Feng, N. Barnes, S. You, and C. McCarthy, "Local background enclosure for RGB-D salient object detection," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Las Vegas, NV, USA, 2016, pp. 2343–2350.
- [31] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
- [32] Q. Luo, X. Fang, L. Liu, C. Yang, and Y. Sun, "Automated visual defect detection for flat steel surface: A survey," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 3, pp. 626–644, Mar. 2020.
- [33] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1493–1504, Apr. 2020.
- [34] H. Dong, K. Song, Y. He, J. Xu, Y. Yan, and Q. Meng, "PGA-Net: Pyramid feature fusion and global context attention network for automated surface defect detection," *IEEE Trans. Ind. Informat.*, to be published, doi: 10.1109/TII.2019.2958826.
- [35] B. Jiang, J. Yang, Z. Lv, and H. Song, "Wearable vision assistance system based on binocular sensors for visually impaired users," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1375–1383, Apr. 2019.
- [36] J. Yang, K. Sim, X. Gao, W. Lu, Q. Meng, and B. Li, "A blind stereoscopic image quality evaluator with segmented stacked autoencoders considering the whole visual perception route," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1314–1328, Mar. 2019.



**Menghui Niu** received the B.E. degree in mechatronic engineering from the Henan University of Science and Technology, Luoyang, China, in 2013, and the M.S. degree in mechanical design and theory in 2016, from the School of Mechanical Engineering and Automation, Northeastern University, Shenyang, China, where he is currently working toward the Ph.D. degree in mechanical design and theory.

His currently research interests include machine vision and image processing.



**Kechen Song** received the B.S., M.S., and Ph.D. degrees in mechanical design and theory from the School of Mechanical Engineering and Automation, Northeastern University, Shenyang, China, in 2009, 2011, and 2014, respectively.

Between 2018 and 2019, he was an Academic Visitor in the Department of Computer Science, Loughborough University, Loughborough, U.K. He is currently an Associate Professor in the School of Mechanical Engineering

and Automation, Northeastern University. His research interest covers vision-based inspection system for steel surface defects, surface topography, image processing and pattern recognition.



**Liming Huang** received the B.S. degree in mechatronic engineering from the Shandong University of Science and Technology, Qingdao, China, in 2017.

He is currently working toward the M.S. degree in mechanical engineering with the School of Mechanical Engineering and Automation, Northeastern University, Shenyang, China.



**Qi Wang** received the B.S. and M.S. degrees in mechanical engineering from the University of Science and Technology Liaoning, Anshan, China, in 2015 and 2018, respectively. He is currently working toward the Ph.D. degree in mechanical design and theory with the School of Mechanical Engineering and Automation, Northeastern University, Shenyang, China.

His current research interests include image segmentation and thermal imaging defect detection.



**Yunhui Yan** received the B.S., M.S., and Ph.D. degrees in mechanical design and theory from School of Mechanical Engineering and Automation, Northeastern University, Shenyang, China, in 1981, 1985, and 1997, respectively.

He has been a teacher at Northeastern University since 1982, and became a Professor with the Department of Mechanical Engineering in 1997. During 1993–1994, he stayed in the Tohoku National Industrial Research Institute as a visiting scholar. His research interests include

intelligent inspection, image processing, and pattern recognition.



**Qinggang Meng** (Senior Member, IEEE) received the B.S. degree in electronic engineering and the M.S. degree in signal and image processing from the School of Electronic Information Engineering, Tianjin University, Tianjin, China, and the Ph.D. degree in computer science from Aberystwyth University, Aberystwyth, U.K., in 1987, 1990, and 2002, respectively.

He is currently a Professor with the Department of Computer Science, Loughborough University, Loughborough, U.K. His current research interests include biologically inspired learning algorithms and developmental robotics, service robotics, robot learning and adaptation, multi-UAV cooperation, human motion analysis and activity recognition, activity pattern detection, pattern recognition, artificial intelligence, and computer vision.

Prof. Meng is on the Editorial Boards of several journals, including the IEEE TRANSACTIONS ON CYBERNETICS.