Unsupervised Saliency Detection of Rail Surface Defects using Stereoscopic Images

Menghui Niu, Kechen Song, Liming Huang, Qi Wang, Yunhui Yan, 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 that a complete defect in complex background information. The addition of surface profiles can effectively improve this by including three-dimensional (3D) information of defects. However, in high-speed detection, traditional 3D profile acquisition is difficult and separate from image acquisition, which cannot satisfy the above requirements effectively. Therefore, an unsupervised stereoscopic saliency detection method based on a binocular line-scanning system is proposed in this paper. This method can simultaneously obtain a highly precise image as well as profile information, while also avoid decoding distortion of the structured light reconstruction method. In this method, a global low-rank non-negative reconstruction algorithm with a background constraint is proposed. Unlike the low-rank recovery (LRR) model, the algorithm has more comprehensive low-rank and background clustering properties. Furthermore, outlier detection based on geometric properties of the rail surface is also proposed in this method. Finally, image saliency results and depth outlier detection results are associated with collaborative fusion. Finally, a data set (RSDDS-113) containing rail surface defects is established for experimental verification. The experimental results demonstrate that our method can obtain an MAE of 0.09 and AUC of 0.94, better than 15 state of the art algorithms.

Index Terms—Rail defects detection, saliency, low rank non-negative reconstruction, stereoscopic, line-scanning system

I. INTRODUCTION

Surface inspection play a pivotal role in improving the rail manufacturing process. Different from the online rail surface inspection system [1]-[2], the manufacturing process inspection of rail products faces the following problems: the working environment is harsh, the appearance of the rolled rail surface is complicated, the generation and distribution of defects are random, the rail transmission has high-speed and vibration, the surface detection system needs to work continuously.

Manual inspection is inefficient and low-sensitivity [2] [33], which will slows down the entire manufacturing process. Compared with manual inspection, non-destructive testing (NDT) methods such as ultrasonic testing (UT), acoustic emission (AE) testing, and magnetic flux leakage (MFL) [3] have higher sensitivity and data interaction in industrial inspection. However, most of them are time-intensive [2] to the target motion, and the detected defects lack intuitive descriptions of defects shapes. Machine vision have more advantages in surface defects testing. Especially, with the development of hardware and software conditions, the application of machine vision technology in railway measurement has attracted more attention.

The machine vision detection system mostly rely on the analysis of images content [3] [4] and defects characteristics [34] [35]. At present, most of them are based on grayscale images, which is due to the high detection speed and the low cost of equipment. For some classic applied research, such as mentioned in [3], ICLBP and GCLBP features of grayscale images are constructed for detection. And 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 misjudgment of pseudo defects due to the lack of more color information.
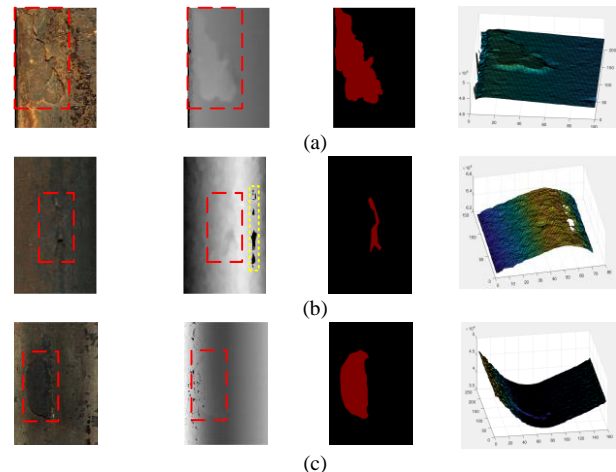

Fig. 1. Some defects images of rail surface: (a) Scar defect I, (b) Hole defect, (c) Scar defect II. From left to right: Original image, Depth map, Ground truth, 3D points cloud. The red box is the defect, and the yellow box is the information missing area.

In some surface contour inspection fields, three-dimensional (3D) profile reconstruction is one of the commonly used methods. Sebastian [5] performs 3D surface quality testing by establishing a surface approximation model. Erik [6] uses structured light and line-scanning stereo matching to obtain 3D information. However, the above methods have to face the difficulty of encode and decode structured light fringes, and the

reconstruction accuracy depends on the resolution of stripes. Besides, the 3D information and the image information can not be uniformly used in the detection process that the two processes are independent.

The above image-based and 3D profile-based detection methods can obtain impressive detection results, but they still face the following problems to be solved.

1) Image information is susceptible to illumination and shooting angles. The low-texture defects are difficult to be detected in the approximate background area, as shown in Fig. 1(a). In particular, grayscale images may detect more pseudo defects due to the lack of color information.

2) In 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, defects with small depth change are easily submerged in a large curvature background, as shown in Fig. 1(c).

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

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

The main contributions of this paper are as follows:

1) The application of a binocular line-scanning system in surface defects 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 stereoscopic images is proposed. The method integrates a variety of information (2D image and 3D profile), and the rail data is used as a case for experimental verification.

2) A global low-rank and non-negative reconstruction (GLRNNR) saliency algorithm is constructed for image defects detection. Compared to the traditional low-rank recovery (LRR) model, the algorithm incorporates background information constraints and non-negative constraints which can make the result more global low-rank, background priori and have clustering properties.

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) 113 stereoscopic image pairs are collected and organized into the rail surface defects data set (RSDDS-113), which include the left-camera images and the corresponding depth maps.

The structure of the article is as follows. Some related works are introduced in Section II. The hardware acquisition system is described in Section III. In Section IV, the 2D image saliency detection method is explained, and the 3D profile outlier detection algorithm based on the indirect plane and geometric properties is elaborated, then the above two detection results are fused. In Section V, 15 methods are used for experimental comparison and analysis. The main work of this paper is summarized in Section VI.

II. RELATED WORK

Saliency detection is the simulation of the human visual system (HVS), detecting the region of interest with visual uniqueness in the face of complex scenes [2], [7]-[9]. By selectively acquiring relevant regional information, HVS can greatly avoid computational waste and reduce the difficulty of 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 low-rank recovery (LRR) model is often used to distinguish the background and foreground, due to it does not require a large number of labeled samples for training.

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

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

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

$$\arg\min_{L,S}\ \|L\|_* +\lambda\|S\|_1\ \ \text{s.t. } F = L+S$$
$$or \tag{2}$$
$$\arg\min_{S}\ \|Z\|_* +\lambda\|S\|_{2,1}\ \ \text{s.t. } F = DZ+S$$

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 $F = FZ+S$.

The following problems usually exist in LRR models.

Firstly, the construction of the over-complete dictionary D is tedious and difficult, especially when the number of samples is insufficient.

Secondly, when the difference between the foreground and background is tiny, the foreground area also has a low-rank characteristic.

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 above problem, a global low-rank and non-negative reconstruction (GLRNNR) saliency detection algorithm is used in image defects detection, the algorithm based on the background priori constraint and the non-negative low-rank sparse constraint.

III. HARDWARE ACQUISITION SYSTEM

In this paper, a binocular line-scanning system is creatively implemented at surface detection. Compared with the existing visual inspection equipment, this system is based on a binocular stereo-camera (BSC) which produced by Chromasens. It can obtain 2D and 3D information at high speed and high resolution.

Meanwhile, it can effectively preclude the non-uniform between 2D and 3D information acquisition during the detection process and provides more effective information for the judgment of surface defects.
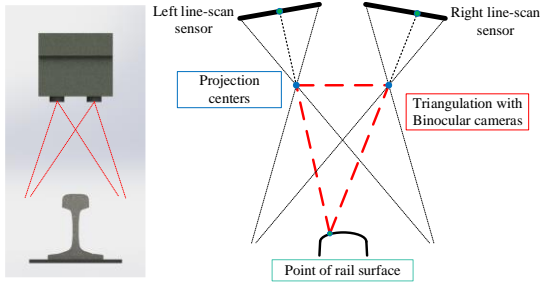

Fig. 2. Binocular stereo camera and triangulation

The optical resolution of BSC can reach "70 μm/pixel", the maximum acquisition speed can be reached "1.4 m/s", the maximum frame rate is "21 kHz", the line-scanning camera has "7142 pixels" and RGB three-channel sensor.

According to the principle of triangulation and polar line correction as shown in Fig. 2. The disparity map of the left and right cameras can be obtained by stereo matching [36] [37]. 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.
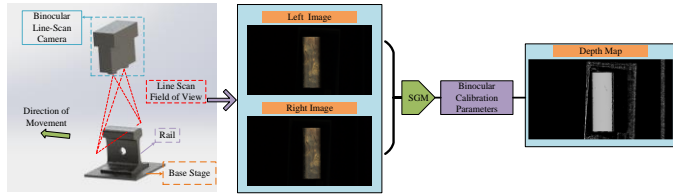

Fig. 3. Acquire depth information by stereo matching

In order to accelerate the calculation of depth information, a simple and fast stereo matching algorithm called SGM (Semi-Global Matching) [12] is used. The left and right consistency detection error of SGM is "10", and the matching disparity range is "(-165, +165)". Based on the above conditions, the BSC can provide high-precision depth information with a resolution of "14 μm" in a range of "52 mm".

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 SLIC (Simple Linear Iterative Clustering) [13] is used to segment the original image into n homogeneous regions. m-dimensional feature information about each region is utilized to construct an overall image feature matrix $F = \{f_1, f_2, \cdots, f_n\} \in R^{m \times n}$.

As previously stated, since the negative values in the coefficient matrix Z lack a reasonable explanation for the actual cluster, then according to the NNLRS (Nonnegative Low-Rank and Sparse Graph) [14] model, the LRR model can be changed to the following representation with a non-negative coefficient constraint:

$$\arg\min_{Z} \|Z\|_* + \beta\|H\|_1 + \lambda\|S\|_{2,1} \tag{3}$$
$$\text{s.t. } F = DZ + S, \ Z = H, \ H \geq 0.$$

Otherwise, to improve the detection effect, some priori information [10], [23] is used, such as the boundary background priori, the center priori, the color priori, etc. 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 problems, a global low-rank and non-negative reconstruction (GLRNNR) is utilized and transformed into the following equation optimization problem. In the process of 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 non-negative constrained. Meanwhile, L is used to ensure the low-rank property of the global background of the image.

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

where, L denotes the global low-rank term, S_1 means the global sparse term. B states that the background dictionary, $L = BZ + S_2$ is the boundary background constraint reconstruction, Z indicates the reconstruction coefficient, and S_2 means reconstruction error, $H_2 \geq 0$ signifies the non-negative constraint of Z, and $\alpha, \beta, \lambda, \eta$ are the balance factors, as shown in Fig. 4.
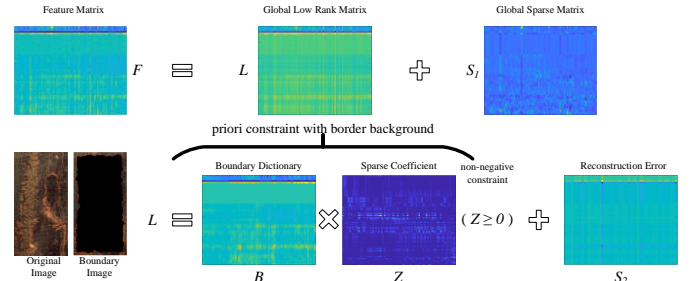

Fig. 4. GLRNNR decomposition with boundary priori and non-negative coefficient constraint

The above problem can be converted into the augmented Lagrangian function, as follow:

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

where Y_1, Y_2, Y_3 and Y_4 are the Lagrange multipliers, μ_1 and μ_2 are the penalty for violating the linear constraints. (L, Z, S_1, S_2) can be solved by alternating iterations with LADMAP [14], as shown in Table I.

The detailed iterative update steps are as follows:

Step 1. Updating L: According to [14], $\Psi_{(\tau)}(\cdot)$ expresses the singular value shrinkage algorithm with soft threshold τ, 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:

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

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_1:

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

Step 3. Updating S_1:

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

Step 4. Updating Z:

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

Step 5. Updating H_2: It is need to perform non-negative constraints during the process of updating.

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

Step 6. Updating S_2:

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

Step 7. Updating Y_1, Y_2, Y_3, Y_4: The detailed iterative update steps for Lagrange multipliers $\left(Y_1, Y_2, Y_3, Y_4\right)$ are as follow:

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

Based on S_1 and S_2, the saliency Sal of the i-th superpixels region in the image is calculated as follows:

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

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

Otherwise, a multi-scale fusion method is used to obtain the final saliency map:

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

where, the fusion is performed using N scales super pixel segmentation, w_n and $Sal_{(n)}$ are the corresponding weight and saliency results at the j-th scale, respectively.

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:	$L_0 = S_{1\_0} = H_{1\_0} = S_{2\_0} = Y_{1\_0} = Y_{2\_0} = Y_{3\_0} = 0, Z = H_2 = Y_{4\_0} = 0$; $\mu_{1\_0} = 0.3, \mu_{1\_max} = 10^9, \mu_{2\_0} = 0.2, \mu_{2\_max} = 10^9, \kappa_0 = 1.2, \nu_1 = 10^{-5}, \nu_2 = 0.015, k = 0.$
While:	$\left(\|F - L_k - S_{1\_k}\|_F/\|F\|_F \geq \nu_1\right) or \left(\|L_k - BZ_k - S_{2\_k}\|_F/\|L_k\|_F \geq \nu_1\right)$ $or\left(\mu_{1\_k}\max\left(\|L_k - L_{k-1}\|_F, \|H_{1\_k} - H_{1\_k-1}\|_F, \|S_{1\_k} - S_{1\_k-1}\|_F\right)/\|F\|_F \geq \nu_2\right)$ $or\left(\mu_{2\_k}\max\left(\|B\|_2\cdot\|Z_k - Z_{k-1}\|_F, \|H_{2\_k} - H_{2\_k-1}\|_F, \|S_{2\_k} - S_{2\_k-1}\|_F\right)/\|L_k\|_F \geq \nu_2\right)$ $or(k < \text{max}\_iters)$

Update $\left(L, Z, S_1, S_2, H_1, H_2\right)$ as **Step 1-6**
Update $\left(Y_1, Y_2, Y_3, Y_4\right)$ as **Step 7**
Update μ_1, μ_2 as follow:

$\mu_{1\_k+1} = \min(\mu_{1\_max}, \kappa_1\cdot\mu_{1\_k})$
$\kappa_1 = \begin{cases} \kappa_0 & if\ \mu_{1\_k}\max\left(\|L_k - L_{k-1}\|_F, \|H_{1\_k} - H_{1\_k-1}\|_F, \|S_{1\_k} - S_{1\_k-1}\|_F\right)/\|F\|_F < \nu_2 \\ 1 & others \end{cases}$
$\mu_{2\_k+1} = \min(\mu_{2\_max}, \kappa\cdot\mu_{2\_k})$
$\kappa_2 = \begin{cases} \kappa_0 & if\ \mu_{2\_k}\max\left(\|B\|_2\cdot\|Z_k - Z_{k-1}\|_F, \|H_{2\_k} - H_{2\_k-1}\|_F, \|S_{2\_k} - S_{2\_k-1}\|_F\right)/\|L_k\|_F < \nu_2 \\ 1 & others \end{cases}$
$k = k+1$;
end While
Output: $\left(L, Z, S_1, S_2, H_1, H_2\right)$

B. The 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 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 paper. 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).

In this method, it is assumed that there is a spatially symmetric area Ω_S on a general cylindrical ruled surface, and after a rigid transformation (translation and rotation), Ω_S is converted into as shown in Fig. 5. The projection region of Ω_S

on the X_Y plane is $\Omega_{X\_Y} = \{(x,y) \mid x^2 + y^2 = r^2\}$. The directrix of Ω_S on 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 Ω_S. On surface Ω_S, there are the following relationships:

$$\begin{cases} \int_l x \cdot y \ ds = 0 \\ \int_l y \ ds = 0 \\ \int_l y \cdot z_{pl} \ ds = 0 \end{cases} \quad (16)$$

$$s.t. \ l : z = \kappa_0 \cdot y + z_{pl}, \ z_{pl} \in p(t)$$

where, (x, y, z) are the points coordinates on Ω_S, and z_{pl} is the intersection of l and X_Y plane.

Discrete sampling on surface Ω_S, the sample point set $P_S = \{(x_{P_S}, y_{P_S}, z_{P_S}) \mid (x_{P_S}, y_{P_S}, z_{P_S}) \in \Omega_S\}$ can be written as discrete form 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 $\frac{\partial I_P}{\partial a_0} = 0, \frac{\partial I_P}{\partial a_1} = 0, \frac{\partial I_P}{\partial a_2} = 0$, and there:

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

Known by (17) and (19), $a_1 = \kappa_0$. The normal vector of I-Plane is $[a_0, \kappa_0, -1]$ such that I-Plane is parallel to $\vec{V_g}$.
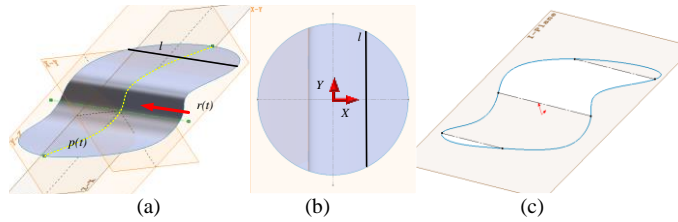
Fig. 5. (a) Spatial symmetric ruled surface region Ω_S (b) Projection on X_Y plane Ω_{X_Y} (c) Indirect plane obtained by *LSM*

The vector $\vec{V_{cc}} = (1, 0, 0)$ is perpendicular to the direction vector $r(t)$. $\vec{V_g}$ can be obtained by a crossing the normal vector $\vec{V_{lsp}} = [a_0, \kappa_0, -1]$ of I-Plane and the vector $\vec{V_{cc}}$, ε is a normalized constant.

$$\vec{V_g} = \varepsilon \cdot \vec{V_{cc}} \times \vec{V_{lsp}} \quad (20)$$

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, $P_m = \left(\sum x_{P_S} \cdot d_{P_S}, \sum y_{P_S} \cdot d_{P_S}, \sum z_{P_S} \cdot d_{P_S} \right) / \sum d_{P_S}$, d_{P_S} denotes the distance between Ω_S and I-plane. $P_g = (x_p, y_p, z_p)$ means the geometric center of Ω_S. $\vec{V_{mg}} = \overrightarrow{P_g P_m}$ 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 is projected in 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 non-smooth and the depth map contains noise and errors, as shown in Fig. 7(c). Here, the *RANSAC* [15] algorithm is used to perform line fit for *Z-value* along $\vec{V_g}$. $\overline{z_{lsp}}$ represents the ideal *Z-value* on line.

In order to weaken the influence of noise and errors, the $\vec{V_g}$ direction should be fine-tuned through setting the angle fine-tuning interval φ and minimizing the overall variance $\vec{\xi} = \arg\min_{\vec{\xi} \in \varphi} \mathrm{var}\left(\left| z_{lsp} - \overline{z_{lsp}} \right| \right)_{\vec{V_g} + \vec{\xi}}$.
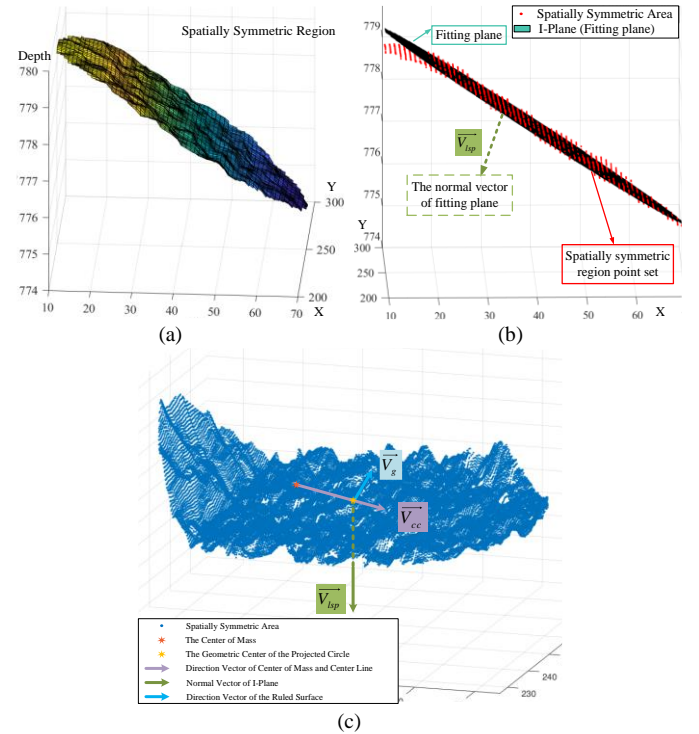
Fig. 6. (a) spatially symmetric region, (b) fit the I-Plane with *LSM*, (c) calculate the centroid of mass and the direction $\vec{V_g}$

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

$$S_D = \left(\left| z_{lsp} - \overline{z_{lsp}} \right| \right)_{\vec{V_g} + \vec{\xi}} \quad (21)$$

where, z_{lsp} represents the distance of points on the ruled surface to the I-Plane, and $\overline{z_{lsp}}$ indicates the ideal value along

the direction $\vec{V_g} + \vec{\xi}$. It is worth noting that Ω_S should be selected as far as possible without defects or micro defects. As shown in Fig. 8, the 2D saliency map is segmented into foreground and background regions by *OTSU* [16]. Then, the largest inscribed circle of the background region is selected as Ω_{X_Y}.

The main defect area is highlighted by centroid re-constraint:

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

and

$$\begin{cases} x'_{S_D} = \sum \dfrac{x_{S_D} \cdot S_{D\left(x_{S_D}\right)}}{\sum S_D} \\ y'_{S_D} = \sum \dfrac{y_{S_D} \cdot S_{D\left(y_{S_D}\right)}}{\sum S_D} \end{cases} \tag{23}$$

$$\sigma_{xy} = \min\left(Var(x_{S_D}), Var(y_{S_D}) \right)$$

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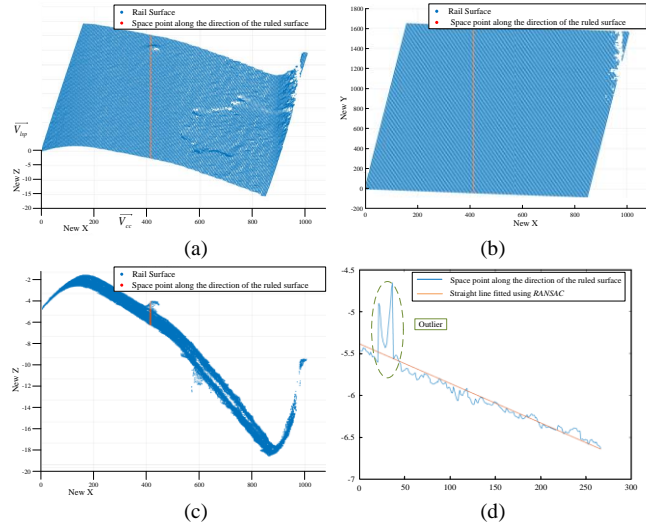

Fig. 7. (a) 3D point cloud profile in coordinate system of $\left(\vec{V_{cc}}, \vec{V_g}, \vec{V_{lsp}}\right)$, (b) The projection of Point Cloud on the New X_Y plane, (c) The projection of profile along direction $\vec{V_g}$. (d) Outlier detection along direction $\vec{V_g}$

C. The Final Saliency Fusion

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

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

Secondly, S_{C1} is subjected to threshold segmentation by *OTSU* to obtain Ω_{X_Y} in the background. The 3D detection result S_{Dep} is obtained by outlier detection, which are pixel level.

According to the outlier saliency model, the detected defect may be incomplete and cover the background noise of 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 defects and reduce the impact of background noise.
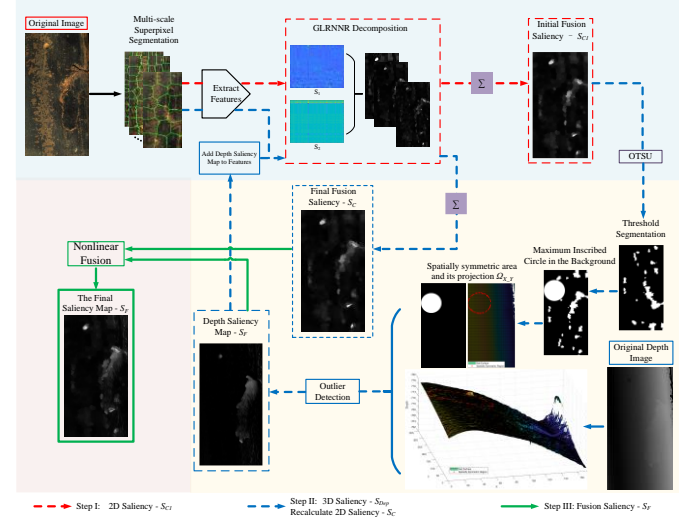

Fig. 8. The saliency detection of rail surface defects combined with 2D and 3D information

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

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

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

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 defects information are collected and employed to construct the rail surface defects data set (RSDDS-113). Under laboratory conditions, the data acquisition process is shown in Fig. 9.
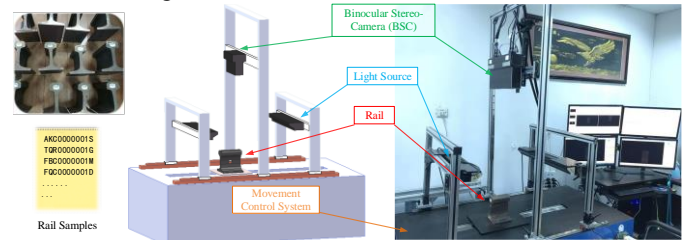

Fig. 9. The rail samples and the experimental equipment

The data of samples cover all the positions of the rails 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 typical defects

will be selected and employed. Every pair is consisting of the left camera image and the corresponding depth image.

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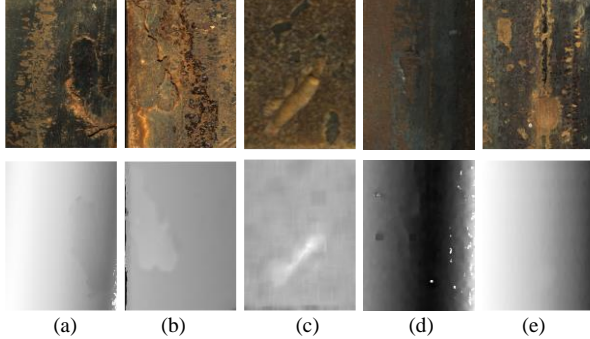

Fig. 10. The defect image (upper) and depth map (lower) of the rail surface: (a) rolling scar (b) corrosion (c) scratches (d) holes (e) pits

2) Rail Surface Discrete Defect (RSDD) Dataset [32]

The RSDD is a public railway image data set, which is mainly composed of two-dimensional grayscale images captured from express rails and heavy haul rails, including two sub-data sets: 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 consistent background than Type-I dataset, but more sophisticated defects are included, which as shown in Fig.15.

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) [32] are employed for RSDD dataset.

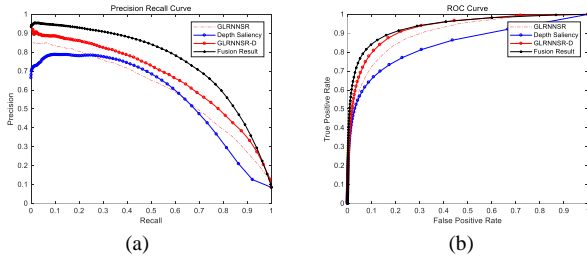*C. Comparison with State-of-the-Arts*

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

1.) *RSDDS-113 dataset:* In this paper, 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 super-pixels 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 indicators as shown in Fig. 11 and Table II. The results demonstrate that each step is effective for generating the final saliency map.

TABLE II
THE 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

As shown in Fig. 11, "Depth Saliency" denotes the result S_{Dep} , which produced by outlier detection. "GLRNNR" 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} .

TABLE III
THE MAE AND AUC OF LRR, DSR, NNLRSR AND GLRNNR

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

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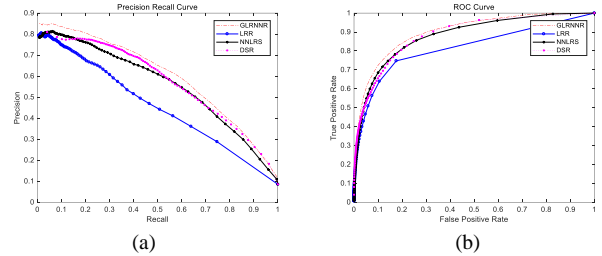

Fig. 12. The comparison of LRR, DSR, NNLRS and our GLRNNR, (a) The PR curve, (b) The ROC curve

As shown in Fig. 13 and Fig. 14, 15 methods are compared. It includes 10 methods which use only color image, such as CA [10], CDCP [11], DRFI [18], GLC [19], DSR [20], HDCT [21], Itti [22], MBS [23], Zhou [24] and SMD [25]. Besides, other 5 methods also be compared, such as ACSD [26], CAIP-MB [27], DCMC [28], DES [29] and LBE [30], which use image and depth information.

TABLE IV
THE 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

In order to verify the effectiveness of the proposed algorithm, "GLRNNR", "GLRNNR-D" and the final fusion result "Ours" are divided into groups for experimental comparison, as shown in Fig. 13 and Fig. 14, and the MAE and AUC will also be shown in Table IV and Table V.

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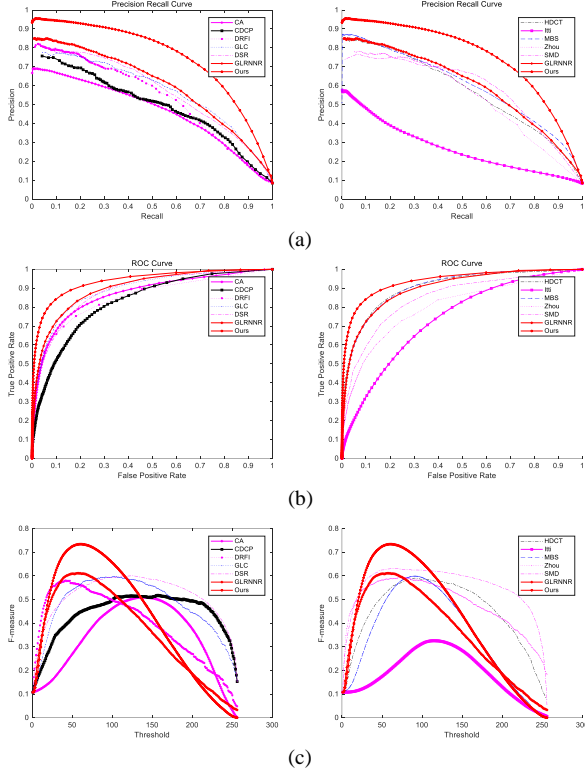

Fig. 13. The comparison of CA, CDCP, DRFI, GLC, DSR, HDCT, Itti, MBS, Zhou, SMD, GLRNNR and OURS. (a) The PR curve, (b) The ROC curve, (c) The F-measure curve

The MAE is mainly for the regression problems, but it is susceptible to background noise. In Tables IV and Tables V, the MAE of ours 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 segmentation and classification problems, such as PR curve and AUC. It can be found that the PR curve and AUC of ours are excellent than other algorithms.
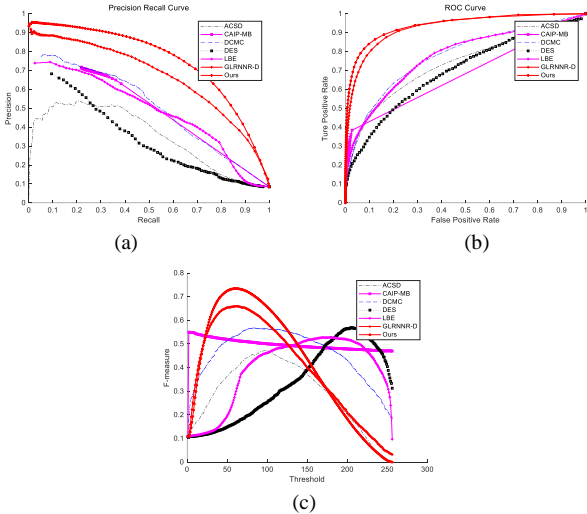

Fig. 14. The comparison of ACSD, CAIP-MB, DCMC, DES, LBE, GLRNNR-D and OURS. (a) The PR curve, (b) The ROC curve, (c) The F-measure curve

In the experimental comparisons, the final fusion "Ours" is significantly better than the other 15 methods. Otherwise, when the resolution of the image is set to 256×512 and the number of super pixels is set to 200, "GLRNNR" takes 3.9 seconds, and the depth outlier detection takes 4.9 seconds, which does not use GPU acceleration.

TABLE V
THE 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

In practical application, offline and parallel processing methods are utilized. Firstly, GLRNNR and 3D outlier detection algorithms are respectively used for pre-processing 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.) *Rail Surface Discrete Defect (RSDD) Dataset:* For the RSDD dataset, the values of the parameters in (4) and (14) is the same as the *RSDDS*-113.
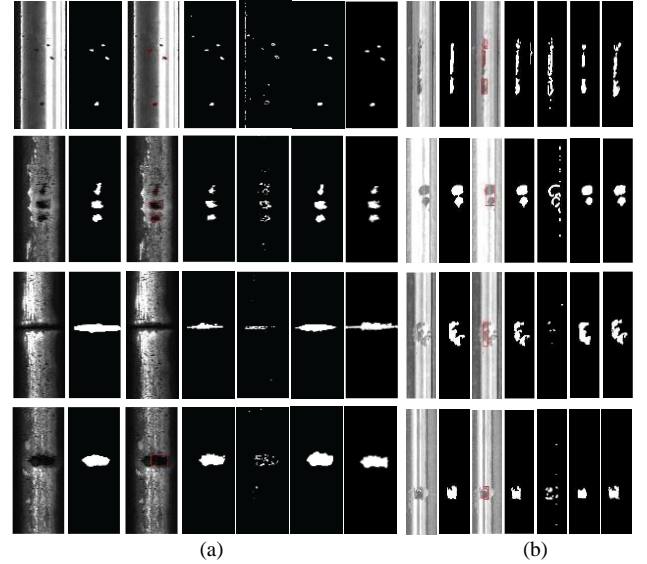

Fig. 15. The sample defect images and results of different methods for the Type-I (a) and Type-II (b) of RSDD dataset. From left to right, each column in (a) and (b) is: Original image, Ground truth, LN + DLBP, MLC + PEME, PM, CTFM and Ours.

Considering the comparison with other known methods (LN+DLBP, MLC+PEME, PM, CTFM) in [32], the OSTU and active contours [31] are used to segment the saliency results into the foreground and background. As described in [32], the evaluation metrics of RSDD dataset are Pixel-Level index (precision, recall, F-measure) and Defect-Level index (precision', recall', F-measure'). The results of the RSDD dataset are shown in Fig.15 and the evaluation metrics of comparison are shown in Table VI and Table VII.

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

It can be ascertained that GLR is significantly better than other algorithms in TYPE-I. Otherwise, the GLRNNR is equal to CTFM at the accuracy and recall rate of the pixel level index

of TYPE-II, but it is better than other algorithms at the defect level index.

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

VI. CONCLUSION

This paper proposes a novel unsupervised stereoscopic saliency detection method for rail surface defects. It is based on a binocular line-scanning system, global low-rank and non-negative reconstruction (GLRNNR) saliency algorithm, and depth outlier detection. Firstly, a 2D image and 3D profile information of the rail surface are obtained by the binocular color line-scanning system. Secondly, utilizing LADMAP and GLRNNR, the algorithm can quickly obtain the saliency map of a 2D image. Next, the outlier region of the 3D profile is detected based on a 2D saliency map and the surface characteristics. Meanwhile, the 3D saliency map is also used as a feature to recalculate the 2D saliency map. Finally, the last 2D saliency result and the 3D result are nonlinearly fused. Our experimental results on the RSDDS-113 dataset outperform 15 state of the art methods in the literature.

Furthermore, to verify the applicability to online railway images, the gray image data set RSDD is also used to compare with other known methods. The experimental results show that the GLNNRR method proposed in this paper is also suitable for general gray rail images and can obtain better detection results.

It is worth noting that the method in this paper is only suitable for locating rail surface defects, more research is needed to determine the defect attributes. In addition, limitations on imaging hardware equipment, hinder this method being effectively applied to detect shallow internal defects. Therefore, in the future, we will look at 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," in IEEE Transactions on Industrial Informatics. doi: 10.1109/TII.2019.2917522.

[2] H. Yu et al., "A Coarse-to-Fine Model for Rail Surface Defect Detection," in IEEE Transactions on Instrumentation and Measurement, vol. 68, no. 3, pp. 656-666, March 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," in IEEE Transactions on Instrumentation and Measurement, vol. 68, no. 3, pp. 667-679, March 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," in IEEE Transactions on Industrial Informatics, 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," in IEEE Transactions on Industrial Informatics, 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," in IEEE Transactions on Instrumentation and Measurement, 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," in IEEE Transactions on Industrial Informatics, vol. 14, no. 3, pp. 1146-1154, March 2018.

[8] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, 2012, pp. 853-860.

[9] J. Yu, J. Zhao, J. Tian and Y. Tan, "Maximal Entropy Random Walk for Region-Based Visual Saliency," in IEEE Transactions on Cybernetics, vol. 44, no. 9, pp. 1661-1672, Sept. 2014.

[10] S. Goferman, L. Zelnik-Manor and A. Tal, "Context-Aware Saliency Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 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," 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, 2017.

[12] H. Hirschmuller, "Stereo Processing by Semiglobal Matching and Mutual Information," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 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," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 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," in IEEE Transactions on Image Processing, 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, "ACM, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[16] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," in IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62-66, Jan. 1979.

[17] R. Margolin, L. Zelnik-Manor and A. Tal, "How to Evaluate Foreground Maps," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014.

[18] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng and S. Li, "Salient Object Detection: A Discriminative Regional Feature Integration Approach," 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, 2013.

[19] N. Tong H. Lu Y. Zhang X. Ruan, "Salient object detection via global and local cues" Pattern Recognition 2014.

[20] X. Li, H. Lu, L. Zhang, X. Ruan and M. Yang, "Saliency Detection via Dense and Sparse Reconstruction," 2013 IEEE International Conference on Computer Vision, Sydney, NSW, 2013.

[21] J. Kim, D. Han, Y. Tai and J. Kim, "Salient Region Detection via High-Dimensional Color Transform and Local Spatial Support," in IEEE Transactions on Image Processing, 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," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 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," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015.

[24] L. Zhou, Z. Yang, Z. Zhou and D. Hu, "Salient Region Detection Using Diffusion Process on a Two-Layer Sparse Graph," in IEEE Transactions on Image Processing, 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," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pp. 818-832, 1 April 2017.

[26] R. Ju, L. Ge, W. Geng, T. Ren and G. Wu, "Depth saliency based on anisotropic center-surround difference," 2014 IEEE International Conference on Image Processing (ICIP), Paris, 2014.

[27] C. Zhu G. Li X. Guo W. Wang R. Wang A Multilayer Backpropagation Saliency Detection Algorithm Based on Depth Mining Springer International Publishing Cham pp. 14-23 2017.

[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," in IEEE Signal Processing Letters, vol. 23, no. 6, pp. 819-823, June 2016.

[29] Y. Cheng H. Fu X. Wei J. Xiao X. Cao "Depth enhanced saliency detection method" Proceedings of International Conference on Internet Multimedia Computing and Service pp. 23-28 2014.

[30] D. Feng, N. Barnes, S. You and C. McCarthy, "Local Background Enclosure for RGB-D Salient Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016.

[31] T. F. Chan, L. A. Vese, "Active contours without edges," IEEE Transactions on Image Processing, Volume 10, Issue 2, pp. 266-277, 2001.

[32] H. Yu, Q. Li, Y. Tan, J. Gan, J. Wang, Y.-A. Geng, L. Jia, "A coarse-to-fine model for rail surface defect detection", IEEE Trans. Instrum. Meas., vol. 68, no. 3, pp. 656-666, Mar. 2016.

[33] Q. Luo, X. Fang, L. Liu, C. Yang and Y. Sun, "Automated Visual Defect Detection for Flat Steel Surface: A Survey," in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 3, pp. 626-644, March 2020.

[34] Y. He, K. Song, Q. Meng and Y. Yan, "An End-to-End Steel Surface Defect Detection Approach via Fusing Multiple Hierarchical Features," in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 4, pp. 1493-1504, April 2020, doi: 10.1109/TIM.2019.2915404.

[35] 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," in IEEE Transactions on Industrial Informatics, doi: 10.1109/TII.2019.2958826.

[36] B. Jiang, J. Yang, Z. Lv and H. Song, "Wearable Vision Assistance System Based on Binocular Sensors for Visually Impaired Users," in IEEE Internet of Things Journal, vol. 6, no. 2, pp. 1375-1383, April 2019, doi: 10.1109/JIOT.2018.2842229.

[37] 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," in IEEE Transactions on Image Processing, vol. 28, no. 3, pp. 1314-1328, March 2019, doi: 10.1109/TIP.2018.2878283.

Liming Huang received the B.S. degree from Shandong University of Science and Technology, Qingdao, China, in 2017. He is currently pursuing the M.S. degree in Northeastern University, Shenyang, China. His current research interests include saliency detection, deep learning.

Qi Wang received his B.S., M.S. degrees in School of Mechanical Engineering and Automation, University of Science and Technology LiaoNing, China, in 2015 and 2018, respectively. He is currently pursuing the PhD degree with School of Mechanical Engineering and Automation, Northeastern University, Shenyang, China. His research interests include image segmentation and thermal imaging defect detection.

Menghui Niu received the B.E. degree from the School of Henan University of Science and Technology, Luoyang, China, in 2013, and the M.S. degree from the School of Mechanical Engineering and Automation, Northeastern University, Shenyang, China, in 2016, where he is currently pursuing the Ph.D. degree. His research interests include machine vision and image processing.

Yunhui Yan received the B.S., M.S. and Ph.D. degrees in School of Mechanical Engineering and Automation, Northeastern University, Shenyang, China, in 1981, 1985 and 1997, respectively. He has been a teacher in Northeastern University of China since 1982, and became as professor in 1997. During 1993-1994, he stayed in the Tohoku National Industrial Research Institute as a visiting scholar. His research interest covers intelligent inspection, image processing and pattern recognition.

Kechen Song received the B.S., M.S. and Ph.D. degrees in 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, UK. 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.

Qinggang Meng (Senior Member, IEEE) received his B.S. and M.S. degrees in electronic engineering from Tianjin University, Tianjin, China, and the Ph.D. degree in intelligent robotics from the Department of Computer Science at Aberystwyth University, UK. He is currently a Professor with the Department of Computer Science, Loughborough University, 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 IEEE Transactions on Cybernetics.