

# Detection of morphology defects in pipeline based on 3D active stereo omnidirectional vision sensor

Zhongyuan Yang<sup>1</sup>, Shaohui Lu<sup>2</sup>, Ting Wu<sup>2</sup>, Gongping Yuan<sup>2</sup>, Yiping Tang<sup>2</sup>✉

<sup>1</sup>Zhejiang Institute of Communications, No. 1515 Mogan Road, Hangzhou, People's Republic of China

<sup>2</sup>Zhejiang University of Technology, No. 288 Liuhe Road, Hangzhou, People's Republic of China

✉ E-mail: typ@zjut.edu.cn

**Abstract:** There are many kinds of defects in pipes, which are difficult to detect with a low degree of automation. In this work, a novel omnidirectional vision inspection system for detection of the morphology defects is presented. An active stereo omnidirectional vision sensor is designed to obtain the texture and depth information of the inner wall of the pipeline in real time. The camera motion is estimated and the space location information of the laser points are calculated accordingly. Then, the faster region proposal convolutional neural network (Faster R-CNN) is applied to train a detection network on their image database of pipe defects. Experimental results demonstrate that system can measure and reconstruct the 3D space of pipe with high quality and the retrained Faster R-CNN achieves fine detection results in terms of both speed and accuracy.

## 1 Introduction

Pipeline is widely used to transport crude oil, natural gas, refined oil product, CWM etc. However, an increasing number of these pipes become destroyed by ageing and internal damage comes to existence. If the damage grows large, serious accidents might occur leading to the heavy loss. For the fact that most pipes have a diameter of 700 mm or less, manual vision inspectors cannot enter these pipes directly. Therefore, automatic inspection by robots which overcome spatial constraint is feasible to measure the pipe and great effort has been done in this domain to improve the real time, accuracy, and robustness of the inspection algorithm.

Considerable work has been done in non-destructive detection and visual-based method for detection and modelling of deformations/dents in structures such as pipes. Non-destructive detection, for example ultrasound detection [1], magnetic flux leakage detection [2], eddy current detection and ray detection [3], is the branch of engineering concerned with non-contact method for measuring and evaluating defects in pipes. Ultrasound detection method determines the size and location of the defects through reflected ultrasound signal. By testing whether the magnetic field lines curved, magnetic flux leakage detection method find the internal defects of the pipe. Eddy current detection method determines the deformations based on changes of magnetic field intensity. Ray detection method is accomplished by real-time imaging of alpha rays, gamma rays and neutron rays. Depending on the transmitting and receiving of the signal repeatedly, all the methods above cannot provide any visual clues, thus leading to the low detecting efficiency.

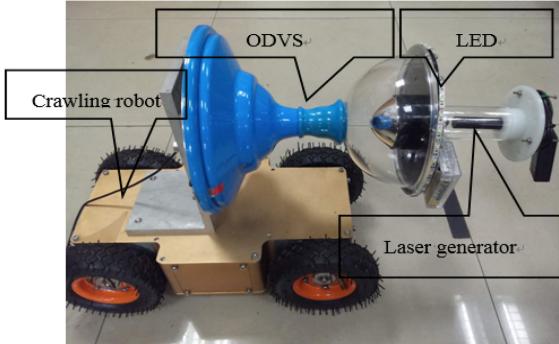
On the other hand, visual detection [4] is a new method forward used extensively to evaluate the condition or the quality of a component. Typically, automatic detection of robot equipped with cameras is desirable and easy to perform. However, many detection systems have to rotate the camera to record images because normal cameras with narrow view can see only a direction but a pipe is a cylindrical geometry. Promisingly, panoramic imaging technology and omnidirectional vision sensors with a wide-field of view afford available benefit to 3D measurement and reconstruction. Kannala *et al.* [5] introduced a pipe detection system using omnidirectional vision sensor that can efficiently take images of 360° in surrounding at a time. Kannala *et al.* [5] used the SFM (structure from motion) method to estimate the camera motion with 3D measurement. For the complicated background of internal pipes, it is not ideal to extract and track the feature points to acquire the

position relation between two images. A structured light visual method [6–10] which is a kind of active measurement has been proposed to solve the problem by already calibrated model. Zhang *et al.* [11] used circle structured light system to reconstruct the shape of a steel pipe. Laser is projected using a mirror and laser ring streaks recorded by CCD camera are displayed on the inner surface of the pipe. The proposed method extracts point cloud from single image without getting corresponding points. However, the position and orientation of the camera are asked to be constant while measurement. Accordingly, it is hard to integrate the measurement results if the camera moves.

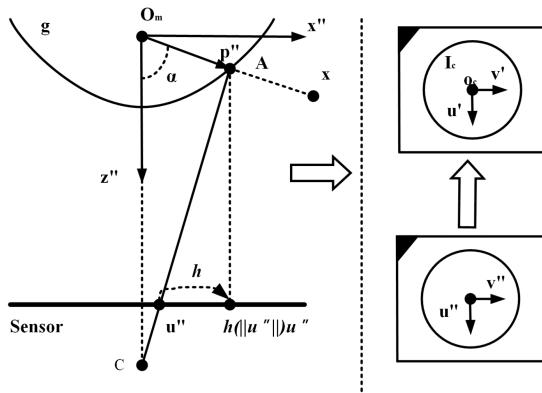
Two key problems of pipeline defect detection are as follows: (1) acquisition of texture information of inner pipe for defects detection such as crack, corrosion, roots. (2) 3D coordinate calculation of point cloud of the pipe for deformation detection. In this work, a vision detection system is proposed to realise both of these goals. The main techniques behind the system are as follows: (1) obtain laser slice panoramic images and texture panoramic images simultaneously; (2) 3D measurement and reconstruction of pipes in the basis of computer vision. (3) Four defects (crack, corrosion, roots, and dark branch) are detected and classified based on Faster R-CNN.

## 2 System overview

The active stereo omnidirectional vision sensor (ASODVS) system is composed of three sections such as crawling robot, microcontroller, omnidirectional vision sensor (ODVS) and omnidirectional laser generator, assembled at the top of the ODVS. With the exception of the crawling robot, other components are the original design of our laboratory. The omnidirectional laser generator contains one circular laser and a conical mirror, which reflects the laser light vertically into the inner wall of the pipe. Thus, the laser of the generator can cover a whole plane during one scanning. Fig. 1 shows the structure of the system. The detection robot travels along the pipe, thus during one strip, the whole 3D environment can be scanned. The laser points and the texture information of the inner pipe in panoramic images can be captured by the ODVS. The laser light reflects the real shape of the pipe so that it can be used to calculate the coordinates of the points of the inner wall. The texture panoramic obtains the texture and colour information and can be used for defects detection such as crack, corrosion, roots.



**Fig. 1** Pipeline visual inspection robot



**Fig. 2** Detection robot of pipe based on ASODVS

### 2.1 ODVS with single viewpoint

The custom-built ODVS is a key component of the system. An omnidirectional picture of the environment can be obtained by a hyperbolic mirror of the ODVS with a horizontal viewing angle of 360° and vertical viewing angle of 120°, respectively [12, 13]. The catadioptric properties of the hyperboloid mirror are used to achieve the ODVS single view characteristics. The angle of incidence is required to calculate the distance of each pixel. So camera calibration is inevitable.

By calibration, the relationship between the internal parameters: internal geometry and optical properties of the camera and the external parameters: coordinates of the camera in the real world can be established. The calibration algorithm can be obtained in [14].

The projection model proposed by Micusik *et al.* [15] is used to set up a map between the image plane and the real world, as shown in Fig. 2. It mainly consists of two different reference planes: sensor plane ( $u'', v''$ ) and image plane ( $u', v'$ ). The sensor plane, as shown in the left part of Fig. 2, is assumed to be perpendicular to the optical axis of the mirror. The intersection of the plane and the optical axis is the origin point. The image plane represented by pixel coordinates is related to the CCD video camera.  $O_m$  is the origin of the coordinate system, and the axis aligns with the optical axis of the mirror. In Fig. 2,  $X = [X, Y, Z]^T$  is a point in the 3D real-world coordinate system, and  $\mathbf{u}'' = (u'', v'')^T$  is a projection point of  $X$  in the sensor plane, and  $\mathbf{u}' = (u', v')^T$  is the corresponding pixel in the image plane.

By projective transformation matrix,  $X$  can be obtained according to  $A$  of the mirror.  $A$  is the focus of the optical centre  $C$  of the camera. In the image plane,  $\mathbf{u}' = (u', v')^T$  is the affine transformation of  $\mathbf{u}'' = (u'', v'')^T$ , which is an intersection of  $C$  and

the sensor plane. Classically, the relationship between the real-world 3D space and the pixels of the image plane can be established by the single-viewpoint catadioptric camera imaging model.

The transform function can be obtained by the following equation:

$$\lambda \mathbf{p}'' = \lambda \begin{pmatrix} \mathbf{X}''^T \\ \mathbf{Z}'' \end{pmatrix} = \lambda \begin{pmatrix} h(\|\mathbf{u}''\|) \mathbf{u}'' \\ g(\|\mathbf{u}''\|) \end{pmatrix} = \mathbf{P} \cdot \mathbf{X}, \quad \lambda > 0 \quad (1)$$

In this equation,  $\mathbf{P} = [\mathbf{R} | \mathbf{T}] \in \mathbb{R}^{3 \times 4}$  is the projective transformation matrix,  $\mathbf{R} \in \mathbb{R}^{3 \times 3}$  is the rotation matrix, and  $\mathbf{T} \in \mathbb{R}^{3 \times 1}$  is the space to mirror transformation matrix.

Mathematically, the transformation function from the sensor plane to the image plane is expressed by the following equation:

$$\mathbf{u}'' = \mathbf{A}\mathbf{u}' + \mathbf{t} \quad (2)$$

A modified perspective projection model is proposed by Scaramuzza *et al.* [16]. The transform relationships can be described by the following equation:

$$\mathbf{p}'' = \begin{bmatrix} \mathbf{u}'' \\ f(\|\mathbf{u}''\|) \end{bmatrix} \quad (3)$$

where  $f(\|\mathbf{u}''\|)$  is a function used to describe the rotational symmetry of hyperboloid mirrors.  $f$  can be derived by using Taylor expansion

$$f(\|\mathbf{u}''\|) = a_0 + a_1 \|\mathbf{u}''\| + a_2 \|\mathbf{u}''\|^2 + \dots + a_n \|\mathbf{u}''\|^N \quad (4)$$

where  $\|\mathbf{u}''\|$  is the distance from the point on the image plane to the centre of the panorama.

The best catadioptric camera imaging model is obtained by Scaramuzza and Micusik's models. However, in actual application, there are errors and deficiencies in the process of assembling ODVS. Equation (5) is simplified model using the model with errors presented by Scaramuzza

$$\lambda \mathbf{p}'' = \lambda \begin{pmatrix} \mathbf{u}'' \\ f(\|\mathbf{u}''\|) \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{A}\mathbf{u}' + \mathbf{t} \\ f(\|\mathbf{A}\mathbf{u}' + \mathbf{t}\|) \end{pmatrix} = \mathbf{P} \cdot \mathbf{X}, \quad \lambda > 0 \quad (5)$$

The ODVS is rotated around a circle and takes a series of panoramic images during the process. The mapping equations from 3D space locations to the pixels in the image plane are derived. As illustrated in Table 1, the calibration results of an ODVS are obtained. The angle of incidence of each pixel in the image plane can be obtained by the internal and external parameters of ODVS as shown in the following equation:

$$\tan \alpha = \frac{\|\mathbf{u}''\|}{f(\|\mathbf{u}''\|)} = \frac{\|\mathbf{u}''\|}{a_0 + a_1 \|\mathbf{u}''\| + a_2 \|\mathbf{u}''\|^2 + \dots + a_n \|\mathbf{u}''\|^N} \quad (6)$$

where  $\alpha$  is the angle of incidence,  $\|\mathbf{u}''\|$  is the distance from the points to the centre of the image plane, and  $a_0, a_1, a_2, a_N$  are internal and external parameters of ODVS. A mapping table is described between the image plane and the angles of incidence for less calculation.

**Table 1** ODVS calibration results

Calibration parameter	$\alpha_0$	$\alpha_2$	$\alpha_4$	$\mathbf{A}$	$\mathbf{t}$	Centre point
ODVS	-110.210	0.0023	-0.000	$\begin{bmatrix} 0.9999 & -10.1e-005 \\ -9.1e-005 & 1 \end{bmatrix}$	$\begin{bmatrix} 49.461 \\ -15.692 \end{bmatrix}$	$\begin{bmatrix} 587.096 \\ 348.682 \end{bmatrix}$

## 2.2 Omnidirectional laser generator and ASODVS + ODVS

This design goal is a rapid, direct and synchronisation acquisition all the panoramic space geometry, texture and colour information in pipeline. Therefore, it is necessary to integrate ASODVS with ODVS. Due to the limitation of pipeline space, it is necessary to design a compact and miniaturisation design. ASODVS + ODVS design is shown in Fig. 3a. The model of the ASODVS + ODVS is shown in Fig. 3b.

As shown in Fig. 3b, the infrared filter is installed in the middle between the camera unit 1 and the camera unit 2. The camera unit 2 is installed in the real focus of the hyperboloid mirror, and the camera unit 1 is installed in the virtual focus of the hyperboloid mirror. In the real focus, the laser slice panoramic images are imaged by ASODVS, while in the virtual focus, panoramic images with texture and colour are obtained by ODVS.

The advantage of this design is that two panoramic cameras share catadioptric mirror, so it has the same panoramic imaging parameters, easier to integrate panoramic colour texture image and laser panoramic image information, and their imaging does not interfere with each other. In addition, the colour texture panoramic images and the laser slice panoramic images can be processed at same time, and all the panoramic space geometry, texture and colour information in pipeline can be obtained fast, directly and synchronously.

As shown in Fig. 4, the central wavelength of the filter is 760 nm, thus, the visible light (red line in Fig. 3b) is transmitted through the filter, and the infrared light (blue line in Fig. 3b) is reflected on the filter. Thus, the visible light is reflected in the imaging unit 1 through the filter after the reflection of the hyperboloid mirror, and the infrared light emitted by the infrared laser transmitter is reflected in the image of the camera unit 2 after being reflected by the hyperbolic mirror. Due to sharing a hyperboloidal catadioptric mirror, and the midpoint filter configuration in hyperbolic mirror real focus and virtual focus, so as long as the parameters of the two cameras unit set (selected) agreement, the colour texture panoramic images can be correspond to the laser slice panoramic images, and synchronisation and consistency of the two kinds of panoramic images can be ensured.

In view of the above analysis, the imaging unit 1 and the camera unit 2 are selected for the same type of resolution.

Fig. 5 shows the laser slice panoramic image and texture panoramic image.

## 2.3 Panoramic images acquisition controlling system

In this work, industrial transportation pipeline is the detection object and the defect detection controlling system is designed as follows:

*Step 1:* the pipeline inspection robot is placed at the bottom of the pipe, the diameter  $d_0$  and length  $L_{L_0}$  of the pipeline is input. The initial value of the length recorder is  $L_0$ . The moving step of the robot is  $T_z$ . To ensure that the panoramic texture images are continuous (there are overlapping regions in two consecutive frames), the moving step is calculated as

$$T_z \leq \frac{d_0}{2} \left( \frac{1}{\tan \alpha_{\max}} + \frac{1}{\tan \alpha_{\min}} \right) \quad (7)$$

where  $\alpha_{\max}$  and  $\alpha_{\min}$  are the angle of elevation and depression of ODVS.

*Step 2:* moving distance of robot is calculated as

$$Z_m = L - L_0 \quad (8)$$

where  $L$  is current length of length recorder. When  $Z_m = nT_z$  ( $n = 0, 1, 2, \dots$ ), ODVS captures the panoramic texture image, saved as ' $Z_m$ '.

*Step 3:* the robot moves on uniformly,  $i = 0, \Delta Z(0) = Z_m$ .

*Step 4:* ODVS captures laser panoramic images at a fixed frequency, the current movement length is  $Z_m$  and the panoramic

images are saved as ' $Z_m + \text{laser}$ '.  
 $i = i + 1, \Delta Z(i) = Z_m, \Delta Z = \Delta Z(i) - \Delta Z(0)$ .

*Step 5:* if  $\Delta Z \leq T_z$ , jump to step 4.

*Step 6:* if  $Z_m \leq L_{L_0}$ , jump to step 2.

*Step 7:* Stop.

## 3 3D coordinates calculation based on ASODVS

ODVS takes panoramic image sequence with laser light of cross-section and then the 3D coordinates of the points can be calculated as following method. Fig. 6 shows the geometry measurement based on ASODVS, establish Gaussian coordinate with the origin of single-view point  $O_m$ , the spatial location of the point loud can be represented with  $(\alpha, \beta, \gamma)$ , where  $r$  is the distance between object point  $P$  and the origin  $O_m$ ;  $\alpha$  is the incident angle from point  $P$  to the origin  $O_m$ ;  $\beta$  is the azimuth angle of point  $P$  relative to the origin  $O_m$ . The above parameters satisfy

$$\begin{cases} r_p = \frac{H}{\cos \alpha_p} \\ \alpha_p = \arctan \left( \frac{\|u''\| \beta_p}{f(\|u''\| \beta_p)} \right) = \arctan \left( \frac{\|u''\| \beta_p}{a_0 + a_2 \|u''\| \beta_p^2} \right) \end{cases} \quad (9)$$

where  $\|u''\| \beta_p$  is the distance between point corresponding to object point  $P$  and the centre in imaging plane;  $H$  is the projection distance of point  $P$  and point  $O_m$  to  $Z$ -axis. Accuracy of the incident angles greatly depend on accuracy in calibration and extracting of laser projection, so extracting the incident point from panoramic image sequence is the key area of the proposed method.

The laser light extracting is the key to calculate 3D coordinate, frame difference method is used in this paper to get laser lines. Two consecutive panoramic images acquired as the robot marches along the pipe have obvious difference, if the absolute value of the difference in brightness is larger than the set threshold, then extract.

The laser light reflected from the internal pipe surface is possible discontinuity and with some width. Therefore, we use the Gaussian approximation method to extract the peak (the pixel that has highest intensity). The three brightest laser points in succession are selected and the sub-pixel offset  $d$  is calculated from these values as

$$d = \frac{\ln(f(i-1)) - \ln(f(i+1))}{2 \times [\ln(f(i-1)) - 2\ln(f(i)) + \ln(f(i+1))]} \quad (10)$$

where  $f(i)$  means intensity value at  $i$  in which  $i$  is an image coordinate of the observed peak intensity. As a result,  $(i+d)$  is obtained as the image coordinate of the laser light.

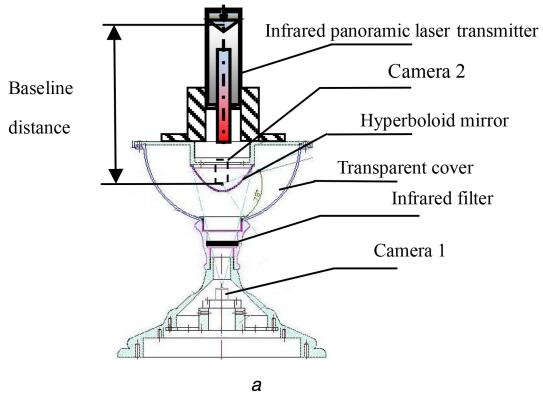
To connect the discontinuous edges, through (11) and (12), we compare the responsive intensity and gradient direction of the gradient operator to determine whether two points are on a line

$$|\nabla f(x, y) - \nabla f(x', y')| \leq T_\nabla \quad (11)$$

$$|\alpha(x, y) - \alpha(x', y')| \leq A_\alpha \quad (12)$$

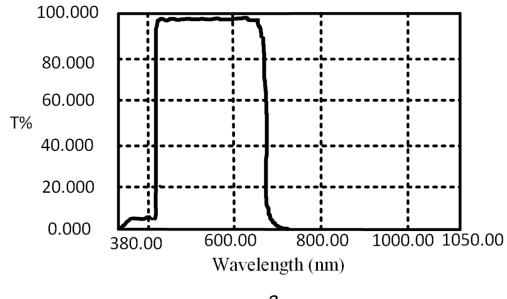
where  $\nabla f(x, y)$  means the gradient value of the boundary point,  $\nabla f(x', y')$  means the gradient value of the unconfirmed point,  $T_\nabla$  means the gradient threshold,  $\alpha(x, y)$  means the direction angle of the gradient vector of the boundary point,  $\alpha(x', y')$  means the direction angle of the unconfirmed point,  $A_\alpha$  means the direction angle threshold. When both of these equations are workable, the unconfirmed point belongs to the boundary and a complete laser light can be acquired. As a result, 3D coordinate of measurement point can be calculated by (2), the location is described in Cartesian coordinate as

$$\begin{cases} x = r \times \cos \alpha \times \sin \beta \\ y = r \times \cos \alpha \times \cos \beta \\ z = h \end{cases} \quad (13)$$



**Fig. 3** Design of the ASODVS + ODVS in this paper

(a) Schematic drawing of ASODVS + ODVS, (b) Model of the ASODVS + ODVS



**Fig. 4** Wavelength characteristics of infrared filters

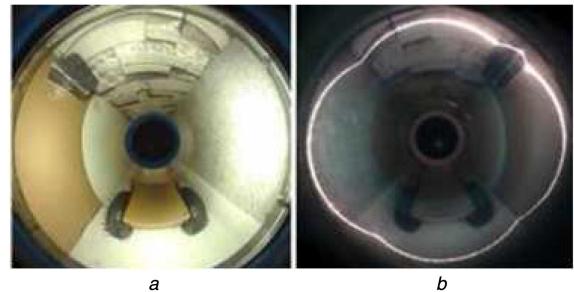
(a) Wavelength characteristics of infrared filter, (b) Transmission and reflection of infrared filter

## 4 Camera motion estimation

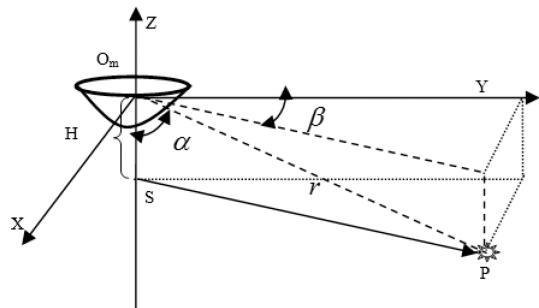
### 4.1 Corresponding point acquisition

In this work, speed-up robust features (SURF) algorithm is used to extract and match the feature points of panoramic texture images [17]. SURF algorithm is an improved algorithm than a scale invariant feature transform algorithm (scale-invariant feature transformation) with higher matching precision and better stability of [18].

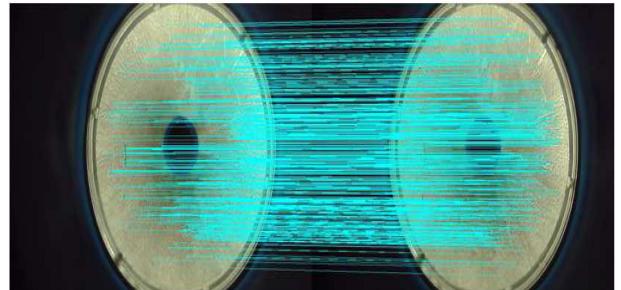
Although SURF has a higher matching accuracy, but the texture panoramic image taken by ODVS is with distortion, so there is still a false matching point. The random sample consensus [19] algorithm is used to remove the false matching points (Fig. 7).



**Fig. 5** Laser slice panoramic image and texture panoramic image  
(a) Omni-directional texture image, (b) Omni-directional infrared image



**Fig. 6** Point cloud geometry measurement based on ASODVS



**Fig. 7** Feature-mapping results of two adjacent panoramic texture images

### 4.2 Estimation of camera motion

The matrix that contains relative position and orientation between two observation points is calculated to estimate cameras motion.

Let the translation  $\mathbf{t}$  and the rotation  $\mathbf{R}$  relate the coordinate systems  $O_{m1}$  and  $O_{m2}$ , see Fig. 8. Let the projections of a 3D point  $P_i$  onto the mirrors be denoted as  $\mathbf{p}_i, \mathbf{p}'_i$ . The coplanarity of vectors  $\mathbf{p}_i, \mathbf{p}'_i$  and  $\mathbf{t} = [t_x, t_y, t_z]^T$  can be expressed in the coordinate system  $O_{m2}$  as

$$\mathbf{p}_i^T \mathbf{R}(\mathbf{t} \times \mathbf{p}_i) = 0 \quad (14)$$

where  $\times$  denotes the vector product. The coplanarity constraint (14) is rewritten in the matrix from as

$$\mathbf{p}_i^T E \mathbf{p}_i = 0 \quad (15)$$

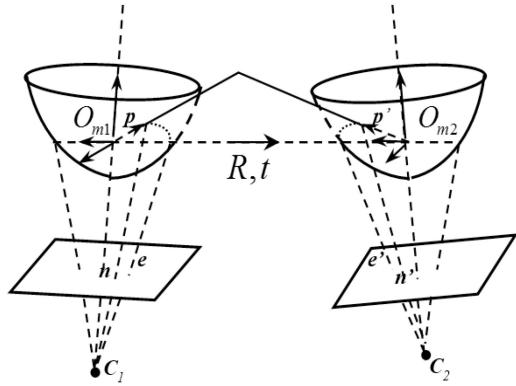
where

$$E = RS \quad (16)$$

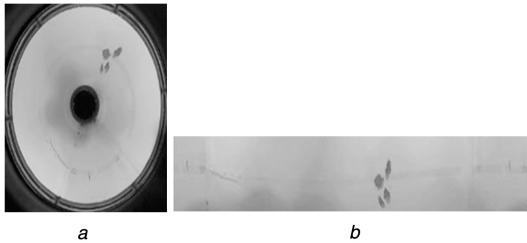
$$S = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \quad (17)$$

Equation (13) is transformed into (18)

$$\mathbf{u}_i^T \mathbf{e} = 0 \quad (18)$$



**Fig. 8** Opipolar geometry of ODVS



**Fig. 9** Cylindrical expansion of panoramic images

(a) The inner wall of pipeline panorama of ASODVS acquisition, (b) Unwrapping result for a panorama

where  $\mathbf{u}_i = [x_i x'_i, y_i x'_i, z_i x'_i, x_i y'_i, y_i y'_i, z_i y'_i, x_i z'_i, y_i z'_i, z_i z'_i]^T$ ,  $\mathbf{e} = [e_{11}, e_{12}, e_{13}, e_{21}, e_{22}, e_{23}, e_{31}, e_{32}, e_{33}]^T$ . Essential matrix  $\mathbf{e}$  is derived from simultaneously solving equations for more than eight pairs of corresponding ray vectors. The function in the following equation should be minimised:

$$\min \| \mathbf{U}\mathbf{e} \|^2 = 0 \quad (19)$$

where  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]^T$ . The matrix  $\mathbf{E}$  is derived from  $\mathbf{e}$  which is given as the eigenvector of the smallest eigenvalue of  $\mathbf{U}^T \mathbf{U}$ . The rotation matrix  $\mathbf{R}$  and translation vector  $\mathbf{t}$  can be derived from the matrix  $\mathbf{E}$ .

## 5 In-pipe defects detection based on Faster R-CNN

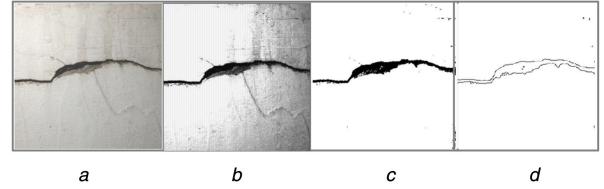
As mentioned above, the ODVS can obtain laser slice panoramic images and texture panoramic images in a single scan. The laser slice panoramic images are used for 3D coordinate calculation of the inner pipe point cloud, and the texture panoramic images are used for defects detection such as cracks, corrosion, roots, and dark branches.

In this section, we apply Faster R-CNN to train a detection network on our database of pipe defects, and identify four classes of defects and predict object bounds of them simultaneously.

### 5.1 Preprocessing of texture panoramic images

In order to eliminate the image distortion caused by panoramic imaging, the expansion algorithm is used for panoramic images.

Firstly, the coordinate system is set up by the centre of the panoramic image  $O^{**}(0,0)$ , the inner diameter is  $r$ , the external diameter is  $R$ , radius of the middle circle is  $r_1 = (r+R)/2$ , and azimuth is  $\beta = \tan^{-1}(y^*/x^*)$ . Then, the coordinate system of cylindrical expansion image is set up, the coordinate centre is  $O^{**}(0,0)$ ,  $x$ -axis  $X^{**}$ ,  $y$ -axis  $Y^{**}$ , the panoramic image is unfolded in a clockwise direction according to the azimuth angle. Establish a one-to-one correspondence between two points of panoramic image and expansion image.



**Fig. 10** Effect image after enhancing the brightness

(a) Original image, (b) Effect image after improving the contrast, (c) Effect image after binarisation, (d) Effect image after edge detection

The formulas are as follows, in the process of calculation, the gray value of non-integer coordinates is calculated by bilinear interpolation

$$x^* = y^*/(\tan(360x^{**}/\pi(R+r))) \quad (19)$$

$$y^* = (y^{**} + r)\cos\beta \quad (20)$$

where  $x^{**}$  and  $y^{**}$  are, respectively, the abscissa and ordinate of the unfolded image,  $x^*$  and  $y^*$  are, respectively, the abscissa and ordinate of the panoramic image,  $r_1$  and  $r_2$  are external diameter and inner diameter of expansion area.

The image preprocessing algorithms are used for image noise reduction, brightness and contrast adjustment as well as the suspected defect segmentation and extraction. The preprocessing algorithm references the existing literature [20]. Figs. 9 and 10 show the results of image preprocessing.

### 5.2 Faster R-CNN

In this section, we briefly introduce the key parts of the Faster R-CNN. We refer readers to the original paper [21] for more technical details. Faster R-CNN includes a region proposal network (RPN) and an object detection network. The architecture of Faster R-CNN is shown in Fig. 11. The top branch is Faster R-CNN [5] and the bottom branch is RPN.

In the RPN, the convolutional layers of a pre-trained network are followed by a  $3 \times 3$  convolutional layer. This corresponds to mapping a large spatial window or receptive field (e.g.  $228 \times 228$  for VGG16) in the input image to a low-dimensional feature vector at a centre stride (e.g. 16 for VGG16). Two  $1 \times 1$  convolutional layers are then added for classification and regression branches of all spatial windows.

In the training phase, the RPN takes an image of any size as input and outputs a set of rectangular object proposals, each with four coordinates of predicted object bounding box and 2 scores that estimate probability of object/not-object. By minimising the classification loss and regression loss for learning region proposals, the RPN are trained to generate high-quality proposals. In this work, cylindrical expansion images ( $1280 \times 240$ ) are as the input of the network.

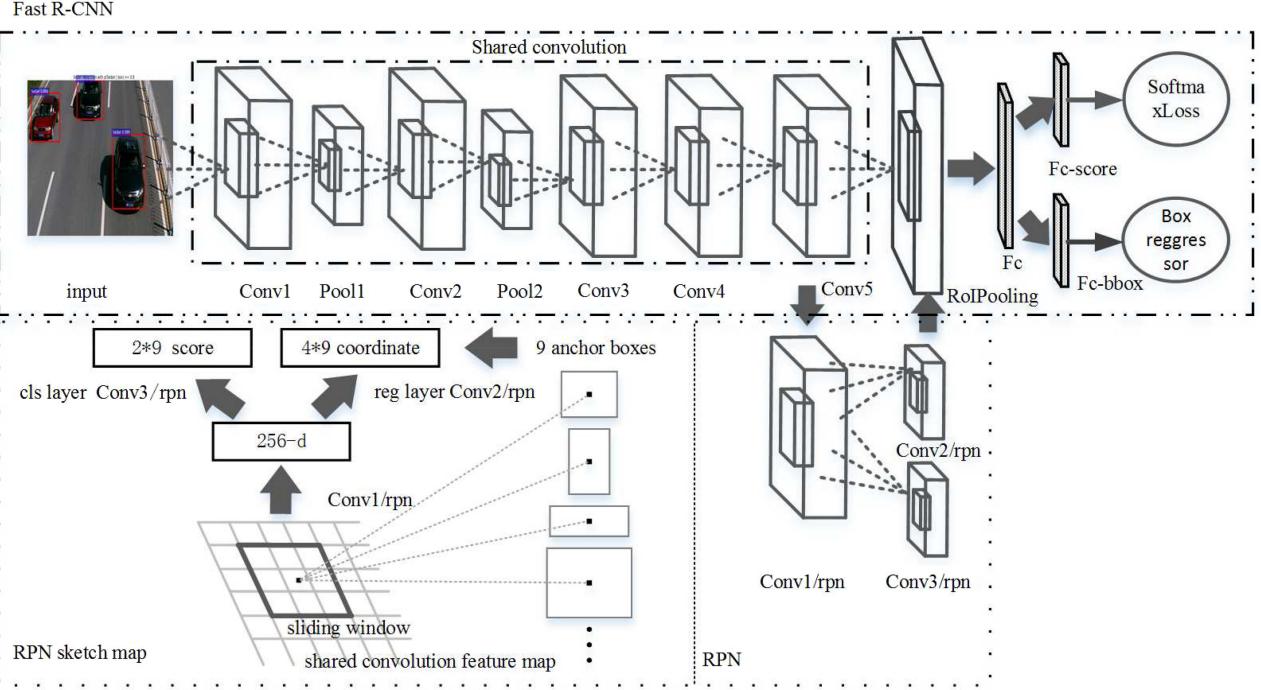
### 5.3 Training database

In order to highlight the advantages of panoramic imaging, the texture panoramic images are imputed to the Faster R-CNN after preprocessing. Fig. 12a shows the examples of the image database and Fig. 12b shows the examples of four defects. The dataset consists of about 3k trainval images and 1k test images over four object categories.

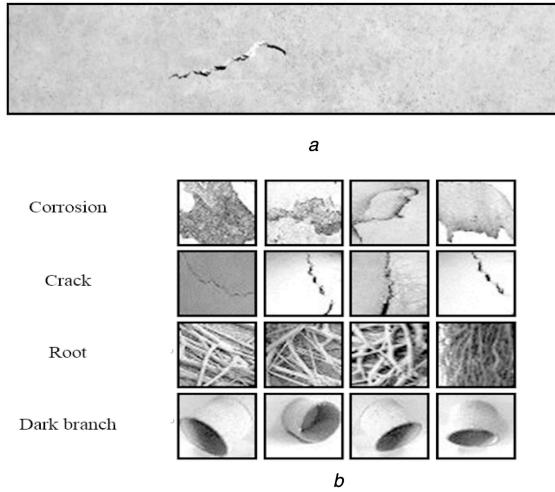
## 6 Experiments and analysis

### 6.1 Experiments setup

In this experiment, a CMOS camera with a resolution of  $1280 \times 720$  is connected to the microcontroller with the USB interface. The processing computer is equipped with Pentium IV CPU, 2G memory, and Windows 7 operating system. The detection robot is placed on the bottom of the pipe.

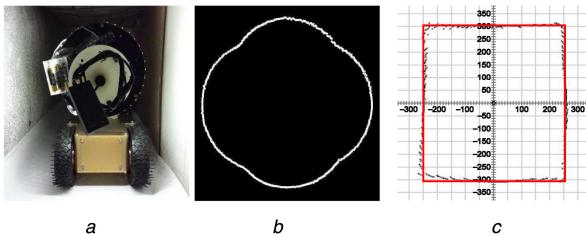


**Fig. 11** Architecture of Faster R-CNN



**Fig. 12** Cylindrical expansion of panoramic images

(a) Example of training image ( $1280 \times 720$ ), (b) Example of four kinds of detects



**Fig. 13** Vision measurement and results using inspection robot

(a) Inspection robot in the rectangular cavity, (b) Infrared centre points after extraction, (c) Accuracy of vision measurement data (mm)

After a series of parameters such as motion, step size, and speed are provided to the system, the pipe is scanned and reconstructed simultaneously.

## 6.2 Camera motion estimation

In order to verify the accuracy of camera motion estimation, three groups of experiments are carried out as follows:

**Table 2** Result of robot motion estimation

		Translation 10 mm no rotation	Translation 10 mm roll 3°	No translation roll 3°
translation	X, mm	0.60289	2.04095	1.74713
	Y, mm	0.97384	5.48832	8.41819
	Z, mm	9.93419	8.10635	5.10702
rotation	X-axis, °	-0.391441	-0.280016	-0.274791
	Y-axis, °	-0.197542	0.009074	0.197562
	Z-axis, °	0.059621	2.582041	2.654399

- (i) forward translation 1 cm, no rotation;
- (ii) roll 3°, forward translation 1 cm;
- (iii) roll 3°, no translation.

Table 2 shows the estimation results of the robot.

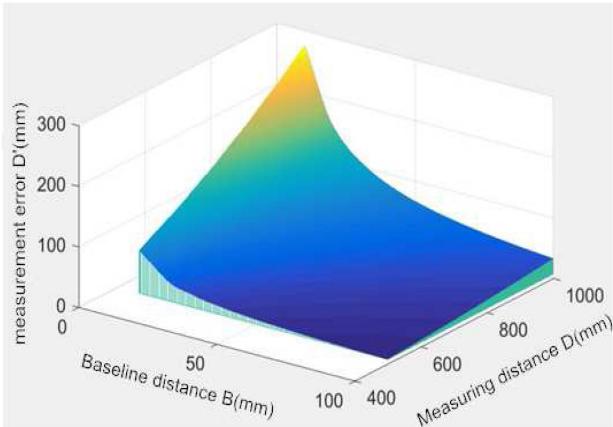
The proposed algorithm can accurately estimate angle of rotation. When ODVS rotates of 3°, the rotational component along the Z-axis is 2.6. The proposed algorithm can accurately estimate translation. When the ODVS translates 10 mm forward, the translation component along the Z-axis is approximately equal to 10 mm, and the translation component along the X- and Y-axis is about 0, respectively. When the ODVS translates forward 10 mm with 3° rotation, three coordinate axis translational component values are 2.0, 5.4, and 8.1 mm.

## 6.3 Measurement accuracy evaluation of ASODVS

In the visual measurement experiment, we place the inspection robot in the pipeline as shown in Fig. 13a. The pipeline is a rectangle with a length of about 3000 mm, and the size of the inner hole is 500 × 600 mm. The average speed is 300 mm/s and the sampling rate is 5 f/s.

Fig. 13b shows the point cloud data obtained from the laser slice panoramic image; Fig. 13c shows the reconstruction of a cross-section with the point cloud.

In order to intuitively show the difference between the visual measurement and the actual hole, the red line shown in Fig. 13c is the actual size of the pipe. Experimental results show that the maximum error is 11 mm and the relative error is 2.2%, which meet the national standard GB/T 27699-2011 [17].



**Fig. 14** System measurements errors

**Table 3** Comparison results between the measurement value and the ground truth (GT)

	Measurement value	GT
width, mm	486.6	500.0
height, mm	591.6	600.0
degree, °	4.8	5.0

**Table 4** Accuracy and speed comparison between three algorithms

	Selective search + Hog + SVM	Edge boxes + Hog + SVM	Faster R-CNN
accuracy, %	83.7	85.3	97
time, s	9.6	1.8	0.19

#### 6.4 Measurement analysis

For the design of the ASODVS, a catadioptric mirror machining error, taper machining error, laser projection error and calibration error will affect the measurement accuracy of different extent. On the visual measurement accuracy analysis, ASODVS calibration method and how to further improve the accuracy of visual measurement methods have been described in [18].

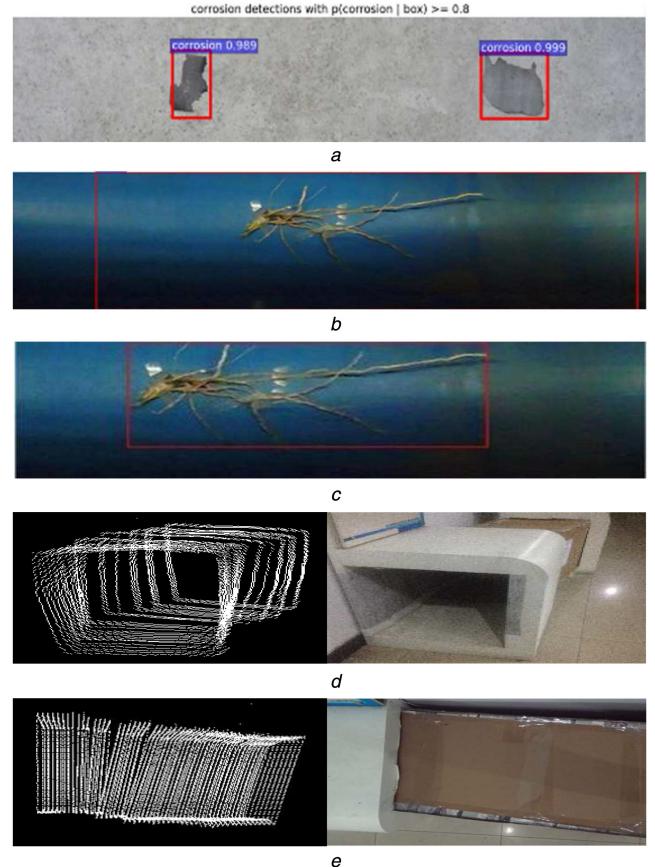
In this work, the method in [18] is introduced to simulate the measurement error of the ASODVS system. The simulation results are shown in Fig. 14. It is noteworthy that, in the actual measurement, in order to improve the accuracy of visual measurement, the appropriate baseline distance should be chosen according to the size of the pipeline, leading to minimum measurement error. In addition, increasing the resolution of ASODVS imaging chip will effectively improve the visual detection accuracy.

In this work, the pixel size of the imaging chip used in ASODVS is  $4 \text{ nm} \times 4 \text{ nm}$ , the quantisation error on the imaging plane is  $2.828 \text{ nm}$ , and the baseline distance of ASODVS is 76 mm. Table 3 shows the comparison between the measured value and the true value of the pipeline.

#### 6.5 Defects detection experiment of the pipe

Classically, it is required both of the quantitative analysis of pipeline structural defects and the quantitative analysis of functional defects. These methods have their own advantages. The quantitative detection is very helpful to detect the severity of pipeline injury, which estimates the remaining life of pipeline. The qualitative damage detection estimates the damage location and type.

In this work, the qualitative detection can be summarised as the ‘appearance’ detection of the pipeline, and the quantitative detection can be summarised as the ‘shape’ detection of the pipeline. The former is to detect and locate the defects through texture panoramic images based on Faster R-CNN.



**Fig. 15** Experimental results of pipeline defect detection and 3D reconstruction

(a) Defect detection results of the Faster R-CNN, (b) Inaccurate location of defects with the Faster R-CNN, (c) Result when select patches from the background of images in the training phrase and make them as negative samples, (d) Results of 3D reconstruction of main view, (e) Results of 3D reconstruction of top view

The test results using Faster R-CNN retrained on our database of four classes of defects are shown in Table 4. The database includes 3000 images for training and validation and 1000 images for test. Faster R-CNN method uses RPN network to replace the selective search and edge boxes algorithm for defect extraction, greatly improving the speed of defect detection. Table 4 shows the accuracy and speed comparison of this method with selective search and edge boxes. The method used in this work has the highest precision with 97%, and the recognition speed is much faster than the above two methods.

The recognition rate is quite good since our classification task is relatively easy for Faster R-CNN (Fig. 15a). However, there is a problem that parts of the test images that do not include target objects will also be classified wrongly as one of the four classes. To handle this problem, we randomly select patches from the background of images in the training phrase and make them as negative samples. Thus, when we use the retrained network for classification task, the network can work well. Fig. 15 shows the different results before and after retraining network. Fig. 15b is the cylindrical expansion image with root defect, and the real image coordinates of the external rectangle are (184, 10), (773, 321). However, Faster R-CNN detection result is (287, 6), (1001.452). Fig. 15c shows the fine result that the coordinates of the external rectangle are (186, 45), (772, 330).

The ‘shape’ detection is mainly to obtain the geometric data of the inner wall of the pipeline, and provide the data support for the quantitative detection and 3D reconstruction. The results of 3D reconstruction are shown in Figs. 15d and e.

## 7 Conclusion

In this paper, we improved the combination of ‘shape’ and ‘appearance’ detection method and qualitative and quantitative

analysis can solve the problem of the pipeline defect detection. In the entire pipeline disease detection and 3D modelling can be done in real time, without any human intervention, achieving the automatic defect detection and digital pipeline modelling.

The future work will focus on the quantification of defects. Faster R-CNN detects and locates the defects of the pipeline and 3D coordinate will be calculated of this region. The area, volume, width and height can be calculated precisely.

Considering the influence of pose and jitter on point cloud data stitching, this kind of problem belongs to the long underground pipe splicing fitting technique. The robot pose needs to be estimated and optimised using the adjacent panoramic texture information. According to the epipolar geometry of panoramic imaging model, the camera motion position can be estimated, then the camera motion pose is used to perform real-time splicing of the cloud data. The related research contents are introduced in details in subsequent paper.

## 8 Acknowledgments

This work was supported by the National Natural Science Foundation (grant no 61379078 project name: active three-dimensional panoramic vision sensing technology) and Zhejiang Technology Application Research Project (grant no 2017C31051 project name: study of some key technologies of underground infrastructure structural health monitoring), we wish to thank all colleagues that contributed to this work.

## 9 References

- [1] Masahiko, H., Hirotugu, O.: ‘Sh-wave emat technique for gas pipeline inspection’, *Nondestruct. Test. Eval. Int.*, 1999, **32**, (3), pp. 127–132
- [2] Tang, Y., Pan, M.C., Luo, F.L., et al.: ‘Detection of corrosion in pipeline using pulsed magnetic flux leakage testing’, *Comput. Meas. Control*, 2010, **18**, (1), pp. 38–43
- [3] Wang, Y., Wang, J.L.: ‘Optoelectronic inspection of in-pipe surfaces’, *J. Appl. Opt.*, 2008, **29**, (5), pp. 735–739
- [4] Yang, M., Su, T.C.: ‘Automated diagnosis of sewer pipe defects based on machine learning approaches’, *Expert Syst. Appl.*, 2008, **35**, (3), pp. 1327–1337
- [5] Kannala, J., Brandt, S.S., Heikkila, J.: ‘Measuring and modelling sewer pipes from video’, *Mach. Vis. Appl.*, 2008, **19**, (2), pp. 73–83
- [6] Osama, M., Tariq, S.E.: ‘Automated detection of surface defects in water and sewer pipes’, *Autom. Constr.*, 1999, **8**, (5), pp. 581–588
- [7] Yang, M.D., Su, T.C., Pan, N.F., et al.: ‘Feature extraction of sewer pipe defects using wavelet transform and co-occurrence matrix’, *Int. J. Wavelets Multiresolution Inf. Process.*, 2011, **9**, (2), pp. 211–225
- [8] Duran, O., Althoefer, K., Seneviratne, L.D., et al.: ‘Automated pipe defect detection and categorization using camera/laser-based profiler and artificial neural network’, *IEEE Trans. Autom. Sci. Eng.*, 2007, **4**, (1), pp. 118–126
- [9] Li, J.X., Wu, E.Q., Ke, Y.L.: ‘New inner surface profile sensor for mini-diameter pipes’, *J. Zhejiang Univ.*, 2006, **40**, (9), pp. 1619–1623
- [10] Matsui, K., Yamashita, A., Kaneko, T.: ‘3-D shape reconstruction of pipe with omni-directional laser and omni-directional camera’. Proc. Int. Conf. Asian Society for Precision Engineering and Nanotechnology, Kitakyushu, Japan, November 2009, pp. 1–5
- [11] Zhang, Y.H., Jin, C.Y., Wang, Y.: ‘Detection of pipe holes and three reconstruction base on circle-structured light’, *J. Beijing Univ. Chem. Technol.*, 2012, **39**, (5), pp. 113–117
- [12] Yamazawa, K., Yagi, Y., Yachida, M.: ‘Obstacle detection with omnidirectional image sensor hyperomni vision’. Proc. Int. Conf. Robotics and Automation, Aichi, Japan, May 1995, pp. 1062–1067
- [13] Tang, Y.P., Wang, Q., Zong, M.L., et al.: ‘Design of vertically aligned binocular omnistereo vision sensor’, *Chin. J. Sens. Actuators*, 2010, **2010**, (1), p. 624271
- [14] Yi, S., Choi, B., Ahuja, N.: ‘Real-time omni-directional distance measurement with active panoramic vision’, *Int. J. Control. Autom. Syst.*, 2008, **5**, (2), pp. 184–191
- [15] Micusik, B., Pajdla, T.: ‘Estimation of omnidirectional camera model from epipolar geometry’. Proc. Int. Conf. Computer Vision and Pattern Recognition, Madison, USA, June 2003, pp. 485–490
- [16] Scaramuzza, D., Martinelli, A., Siegwart, R.: ‘A toolbox for easily calibrating omnidirectional cameras’. Proc. Int. Conf. Intelligent Robots and Systems, San Diego, USA, October 2007, pp. 5695–5701
- [17] Zhu, Q.G., Wang, J., Zhang, P.Z., et al.: ‘Research on mobile robot localization based on Gaussian moment improved surf algorithm’, *Chin. J. Sci. Instrum.*, 2015, **36**, (11), pp. 2451–2457
- [18] Suo, C.B., Yang, D.Q., Liu, Y.P.: ‘Comparing sift, surf, brisk, orb and freak in some different perspectives’, *Beijing Surv. Mapp.*, 2014, **04**, (1), pp. 23–26
- [19] Zhou, J., Chen, L., Liu, Q., et al.: ‘Fast and accurate ransac based on optimal sequential probability test and local optimization’, *Chin. J. Sci. Instrum.*, 2012, **33**, (9), pp. 2037–2044
- [20] Yang, M.D., Su, T.C., Pan, N.F., et al.: ‘Systematic image quality assessment for sewer inspection’, *Expert Syst. Appl.*, 2011, **38**, (3), pp. 1766–1776
- [21] He, K., Zhang, X., Ren, S., et al.: ‘Deep residual learning for image recognition’. Proc. Int. Conf. Computer Vision and Pattern Recognition, Las Vegas, USA, June 2016, pp. 770–778