Article

Displacement Identification by Computer Vision for Condition Monitoring of Rail Vehicle Bearings

Lei Lei [1], Dongli Song [1,*], Zhendong Liu [2] , Xiao Xu [1] and Zejun Zheng [1]

[1] State Key Laboratory of Traction Power, Southwest Jiaotong University, Chengdu 610036, China; leilei309@my.swjtu.edu.cn (L.L.); XuXiao@my.swjtu.edu.cn (X.X.); zhengjuner@my.swjtu.edu.cn (Z.Z.)

[2] Department of Engineering Mechanics, KTH Royal Institute of Technology, 10044 Stockholm, Sweden; zhendong@kth.se

* Correspondence: sdlcds@swjtu.edu.cn; Tel.: +86-028-8646-6026

Abstract: Bearings of rail vehicles bear various dynamic forces. Any fault of the bearing seriously threatens running safety. For fault diagnosis, vibration and temperature measured from the bogie and acoustic signals measured from trackside are often used. However, installing additional sensing devices on the bogie increases manufacturing cost while trackside monitoring is susceptible to ambient noise. For other application, structural displacement based on computer vision is widely applied for deflection measurement and damage identification of bridges. This article proposes to monitor the health condition of the rail vehicle bearings by detecting the displacement of bolts on the end cap of the bearing box. This study is performed based on an experimental platform of bearing systems. The displacement is monitored by computer vision, which can image real-time displacement of the bolts. The health condition of bearings is reflected by the amplitude of the detected displacement by phase correlation method which is separately studied by simulation. To improve the calculation rate, the computer vision only locally focuses on three bolts rather than the whole image. The displacement amplitudes of the bearing system in the vertical direction are derived by comparing the correlations of the image's gray-level co-occurrence matrix (GLCM). For verification, the measured displacement is checked against the measurement from laser displacement sensors, which shows that the displacement accuracy is 0.05 mm while improving calculation rate by 68%. This study also found that the displacement of the bearing system increases with the increase in rotational speed while decreasing with static load.

Keywords: displacement detection; bearing system; experimental platform; computer vision; phase correlation; GLCM; condition monitoring

1. Introduction

Axle box bearings of rail vehicle, as a key component of railway running gear, are used to adapt the rotational movement of wheelsets into a longitudinal motion of the car body along the track. Any fault state of the axle box bearing seriously challenges the running safety of rail vehicles [1]. Different from other kinds of bearings, axle box bearings work in a harsh condition and are heavily subjected to wheel–rail contact force, dynamic vibration generated by the car body and frame, meshing excitation during the gear engagement in the power transmission of gear box, and the excited dynamic load generated by the bearing itself [2]. Therefore, it is necessary to monitor the health condition of the bearings to ensure the running safety of rail vehicles.

At present, the condition monitoring and health diagnosis methods of bearings in railway applications are mainly categorized into two groups: on-board monitoring based on vibration or temperature, and trackside monitoring based on acoustics. The on-board monitoring requires additional detecting equipment installed on the bogie, which greatly increases the manufacturing cost. The acoustic signals measured by trackside monitoring are seriously affected by ambient noise [3,4]. In order to reduce monitoring cost and further

improve monitoring accuracy and reliability, it is necessary to develop a new method for bearing monitoring.

In contrast, there are many monitoring technologies for structural health monitoring which can potentially apply to bearing condition monitoring. Structural health monitoring is widely used for the assessment of structural performance and safety state by monitoring, analyzing, and identifying various loads and structural responses of the target structures [5–7]. Displacement is an important index for structural state evaluation and performance evaluation [6] because the displacement can be further converted into a corresponding physical index for structural safety assessment. Static and dynamic characteristics of the structure, such as bearing capacity [8], deflection [9], deformation [10], load distribution [11], load input [12], influence line [13], influence surface [14], modal parameters, etc. [15,16], can thus be reflected by the structural displacement. Among them, structural displacement monitoring based on computer vision has attracted more and more attention because it has many advantages, e.g., non-contact, high accuracy, time and cost saving, multi-point monitoring, etc. [17]. Computer vision monitoring methods of structural displacement have been applied to many tasks of bridge health monitoring, Yoon et al. [18] used a UAV(Unmanned Aerial Vehicles) to carry a 4K camera to monitor the displacement of a steel truss bridge, and obtained the absolute displacement of the structure without the influence of UAV movement. Ye et al. [19] used a computer-controlled programmable industrial camera to monitor the behavior of an arch bridge under vehicle load, obtaining the influence line of its structural displacement, and realizing the real-time online displacement monitoring of multiple bridges. Tian et al. [20] combined the acceleration sensor and visual displacement measurement method to carry out an impact test of a structure and construct the frequency response function of the structure, to realize the estimation of the structure's mode, mode shape, damping, and modal scale factor, and to realize the impact displacement monitoring of the pedestrian bridge. Besides the bridge health monitoring, many other engineering applications also use this method to monitor and identify structural displacements. For example, Chang et al. used structural displacement monitoring, feature extraction, and the support vector machine of computer vision to form vibration monitoring systems for the on-site diagnosis and performance evaluation of industrial motors and carried out preventive maintenance experiments [21]. Liu studied a track displacement monitoring system in which a fixed camera at the trackside was used for imaging and then the actual displacement of the track was calculated through a digital image processing algorithm, which realized an accurate non-contact measurement of track displacement [22]. Based on the above studies using computer vision to monitor structural displacement in different engineering applications, it can be seen that it is effective, convenient, and accurate to monitor structural states by detecting displacement signals.

Based on the existing bearing monitoring are some shortcomings of traditional methods (including installing additional sensing devices on the bogie which increases manufacturing cost; trackside monitoring is susceptible to ambient noise, etc.). In this article, a displacement monitoring method based on computer vision to monitor the vertical displacement of the axle box bearing of the rail vehicles under simulated real working conditions in proposed, in order to realize the non-contact, high accuracy of axle box bearings of condition monitoring, the realization of ultimate axle box bearing fault diagnosis, and preventive maintenance. Firstly, a portable camera is used to image the platform and detect the displacement amplitude of the bolts, which is used to calculate the state of the bearing through the phase correlation method. Next, the displacement amplitudes of the bearing system in the vertical direction are derived by comparing the correlations of the image's gray-level co-occurrence matrix (GLCM). Finally, for verification, the measured displacement is checked against the measurement from a laser displacement sensor. In the following sections, the proposed approach is used to monitor the displacement of several sets of the platform under different working conditions, and the experimental results are analyzed. Finally, the associated open research challenges are discussed.

2. Experimental Setup of Bearing Experimental Platform

The platform was built to simulate the real operation of the bearing, and at the same time, it can apply radial load (simulating the weight of the train) and stimulate the track irregularity to the tested bearing. The structure of the platform is shown in Figure 1. It mainly consists of three parts (shown in the dotted box in Figure 2): (1) the power input; (2) the excitation and expansion platform; (3) the whole bearing system. The platform adopts a horizontal structure, and the tested bearings are placed on both sides of the spindle. The spindle is connected to the motor through a universal joint coupling (simulating the speed of the train axle); the side view of the whole bearing system is shown in Figure 3. The bearing system is fixed on the excitation platform, and the end cover and the shell of bearing box are connected by eight bolts (bolts on which this article focuses) evenly distributed along the circumference. The radial load (simulated vehicle axle load) is applied to the bearing under test by two sets of vertically mounted springs. The radial load is loaded onto the two bearing under test. By applying vertical excitation to the test system (simulating the vertical irregularity of the actual line), the excitation platform is bolted to bearing bracket. The inertial force of the equivalent parts is reasonably simulated by the mass of the spindle and the bearing bracket. In this way, the bearing under the combined action of the radial load and excitation platform can simulate the working condition of the train operation so as to monitor the service state of the axle box bearing of a high-speed train.

The electric machinery speed of the platform is 0~1460 r/min, the excitation frequency of the excitation platform is 0~50 Hz, and the vertical loading range of the tested bearing is 0~2000 kg. The parameters of the axle box bearing under test are shown in Table 1. To the platform can be added a temperature sensor, acoustic sensor, vibration sensor, etc., which have collected data including images, sound, temperature, and vibration.
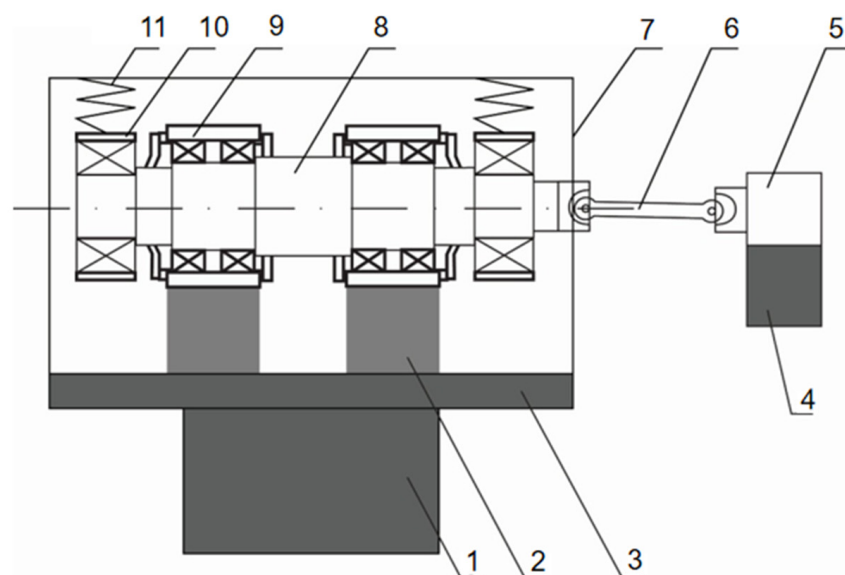
1. excitation platform; 2. bearing bracket; 3. expansion platform; 4. the motor support;
5. electric machinery; 6. coupling; 7. door frame; 8. the spindle; 9. support bearing;
10. test bearing; 11. spring loaded.

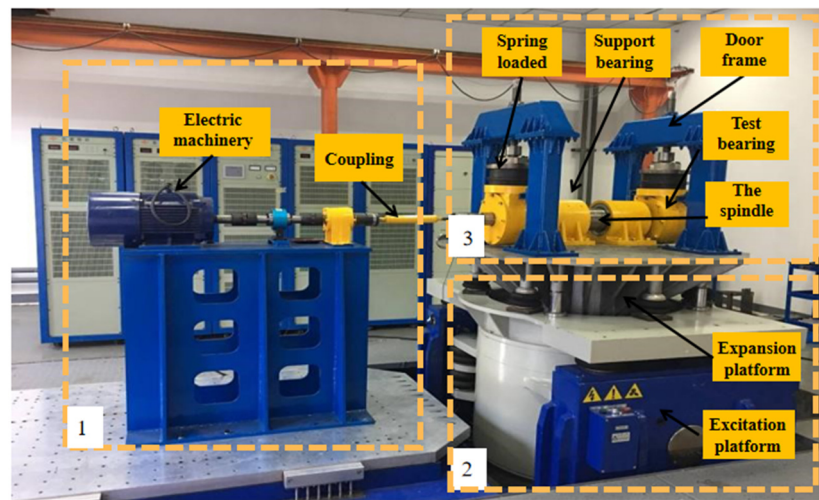**Figure 1.** Structure of the bearing system experimental platform.
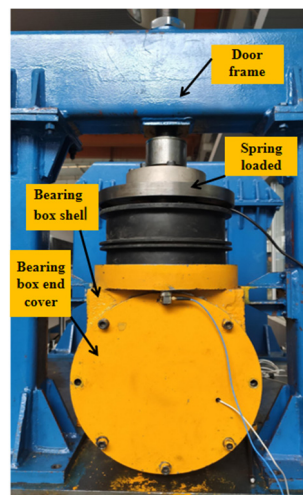
Figure 2. The design of the experimental platform.

Figure 3. Side view of the bearing system.

Table 1. Axle box bearing dimension parameters.

Bearing Dimension Parameter	Parameter Value
bearing outside diameter/mm	240
bearing inner diameter/mm	130
bearing width/mm	180.5
roller diameter/mm	27
number of rollers	17×2

3. Methodology

The general framework of the displacement monitoring system based on computer vision is shown in Figure 4. This section mainly describes the identification principles and methods of camera calibration, object tracking and displacement identification. The flow chart of displacement identification in this article is shown in Figure 5. Video input was collected by the portable camera (the video capture scheme is shown in Figure 6), and then captured the video into images according to the frame; all the images constitute the displacement image set. The displacement image of the first frame is marked as the sample image, and the coordinates of the three bolt points in the sample image are automatically located by the positioning algorithm. Then, images with the same coordinate position as the sample image are intercepted from the displacement image set and the sample image

are calculated by the phase correlation method. Finally, the displacement amplitudes of the bearing system in vertical direction are derived by comparing the correlations of the image's gray-level co-occurrence matrix (GLCM).

The portable camera was positioned on the central axis on the side of the platform; sample video images of the displacement of the platform for a short period of time are shown in Figure 7. The collected image contents include the bearing end cover and bolts, door frame, and expansion platform. The whole bearing system is fixed on the vibration platform, and the bearing end cover and bolts, door frame and expansion platform are fixed rigidly by bolts; therefore, the displacement amplitude of each place in the image collected in Figure 7 is the same. Therefore, in order to improve the calculation rate, and achieve the purpose of real-time monitoring and identification, the local images of the three bolts under the end cover of bearing box (as shown in the red rectangle in Figure 7) were used instead of the whole image to calculate the vertical displacement amplitudes of the three bolts over time. The displacement amplitudes in vertical direction of the three bolts over time were calculated individually. The displacement amplitudes of the bearing system in vertical direction were derived by comparing the correlations of the image's gray-level co-occurrence matrix (GLCM).
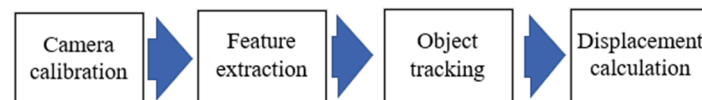

Figure 4. General framework of computer vision displacement monitoring.
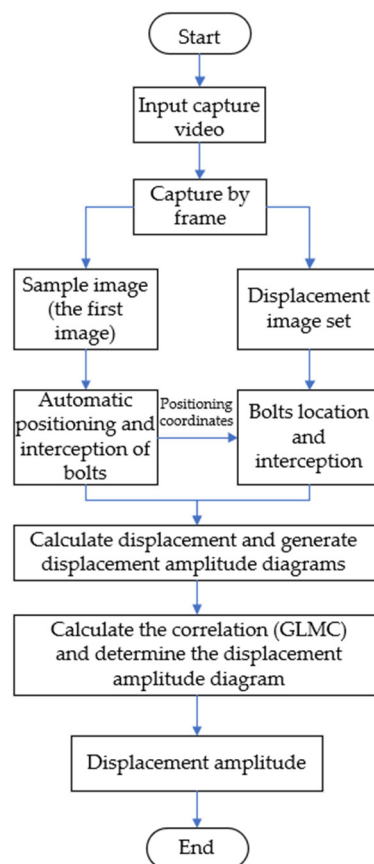

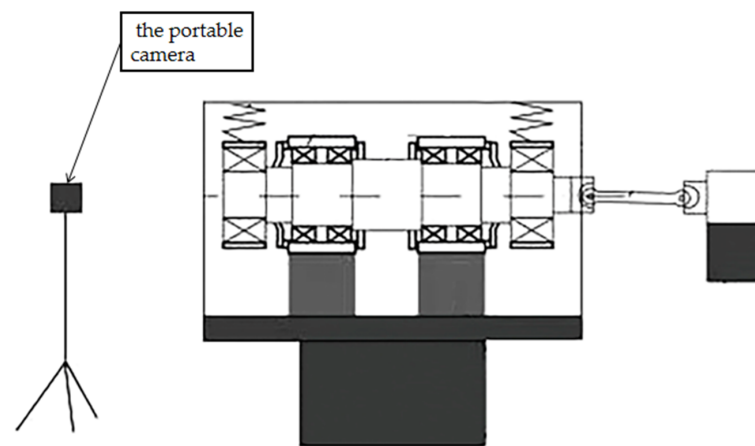**Figure 5.** Flow chart of displacement identification.

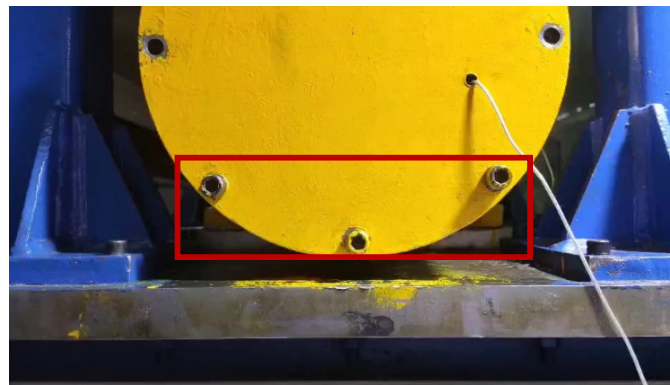**Figure 6.** Displacement video capture scheme.

Figure 7. Image of bearing system experimental platform.

3.1. Camera Calibration

For the displacement monitoring of the platform, the measurement task was limited to one-dimensional displacement; thus, this article adopted a simplified camera calibration method: the scale factor method [23].

As shown in Figure 8, when the camera optical axis is perpendicular to the structural plane, and the optical axis is in line with the normal of the structural plane, the calculation formula of scale factor *e* is:

$$e = \frac{D}{d} \tag{1}$$

Or

$$e = \frac{Z}{f} d_{pixel} \tag{2}$$

where D is the size of the selected object in the structure plane, d is the corresponding number of pixels in the image plane, f is the focal length of the lens, Z is the distance from the camera to the structure plane, and d_{pixel} is the pixel size.
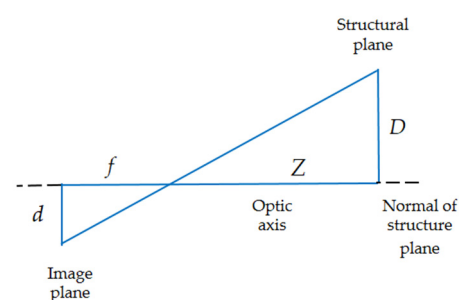

Figure 8. Calculation diagram of scale factor.

3.2. Object Tracking and Feature Extraction

The displacement calculation method used in this article is the phase correlation method; the method needs two Fourier transforms and one inverse Fourier transform. Therefore, the amount of calculation load increases with the input images and decreases with the operation speed. Therefore, it cannot meet the needs of real-time displacement monitoring. The simple compressed image size will lead to a decrease in the accuracy of the calculation of displacement, because the displacement amplitude of the captured image (as shown in Figure 9) is the same at every point. Observe the captured image, which contains the lower half of the end cover of the bearing box of the bearing system and the bolts above, part of the expansion platform and the door frame. Among them, the bolt features on the bearing end cover are obvious, the geometric features are stable, and the contour points are distributed uniformly along a central point, which reduces the calculation error of the power spectrum (Equation (12)) caused by the small and concentrated image pixel value gradient in the later phase correlation method to calculate the displacement. Therefore, the local image of the three bolts is selected for calculation of the phase correlation method. Figure 9 shows the displacement image of the platform under different ray directions due to different time periods (the difference can be seen clearly by the portion of the rectangular box in each drawing). Due to the influence of the change of image gray value caused by different ray direction on the positioning of bolts of end cover, this article compares two positioning methods: template matching positioning method and contour feature positioning method.
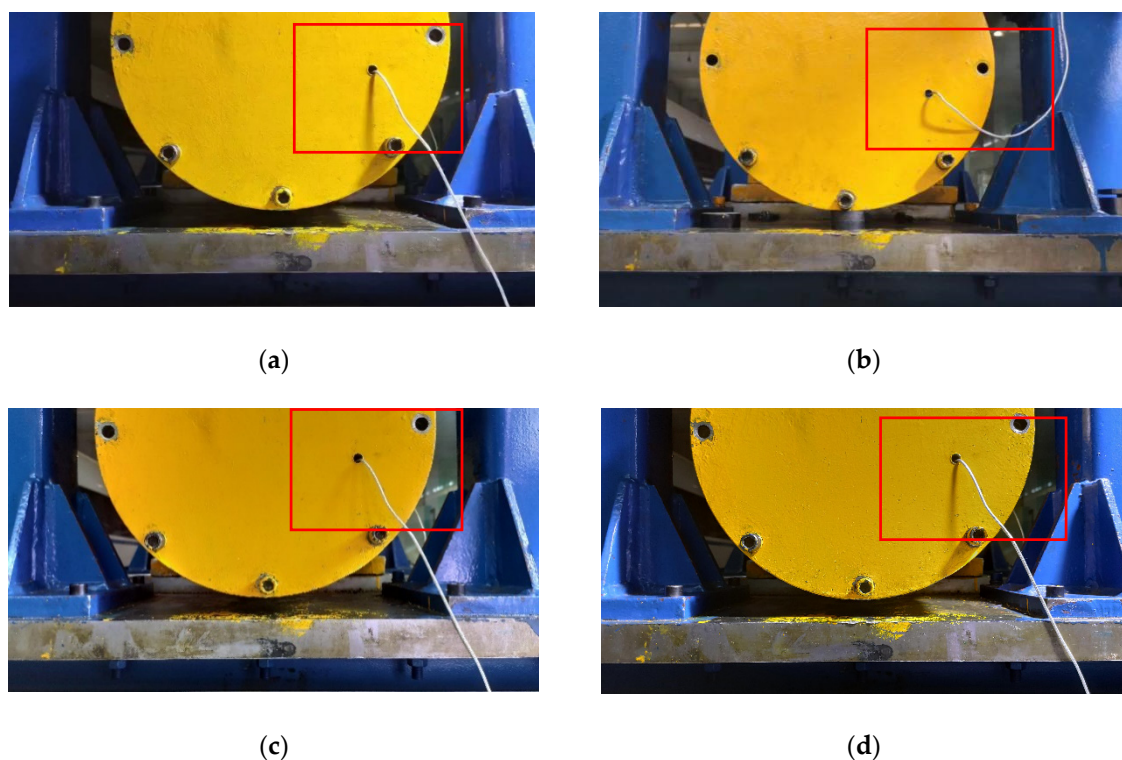

(**a**)

(**b**)

(**c**)

(**d**)

Figure 9. Images of the platform in different ray directions: (**a**) image under strong supplementary light source condition; (**b**) image under strong natural light conditions; (**c**) Images under low natural light conditions; (**d**) image under weak supplementary light source condition.

3.2.1. Template Matching Positioning Method

The template matching and positioning method is shown in Figure 10 (taking one of the three bolts as an example). The main process is to intercept the grayscale image of the target bolt (the size of the captured image is $\sqrt{n} \times \sqrt{n}$) subset and the template in a part of the image in advance, and then traverse the matrix with the size of $\sqrt{n} \times \sqrt{n}$

in the grayscale pixel matrix of the image to be fixed. The two matrices are reduced to one dimension to perform correlation operation to complete the global search. The measurement index of the search is the normalized correlation coefficient r.

$$r = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \cdot \sqrt{n \sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}} \tag{3}$$

where x_i represents the element corresponding to the reduction in one dimension of the template matrix, and y_i represents the element corresponding to the reduction in one dimension of the truncated matrix with the size of $\sqrt{n} \times \sqrt{n}$ in the grayscale pixel matrix of the image to be determined.
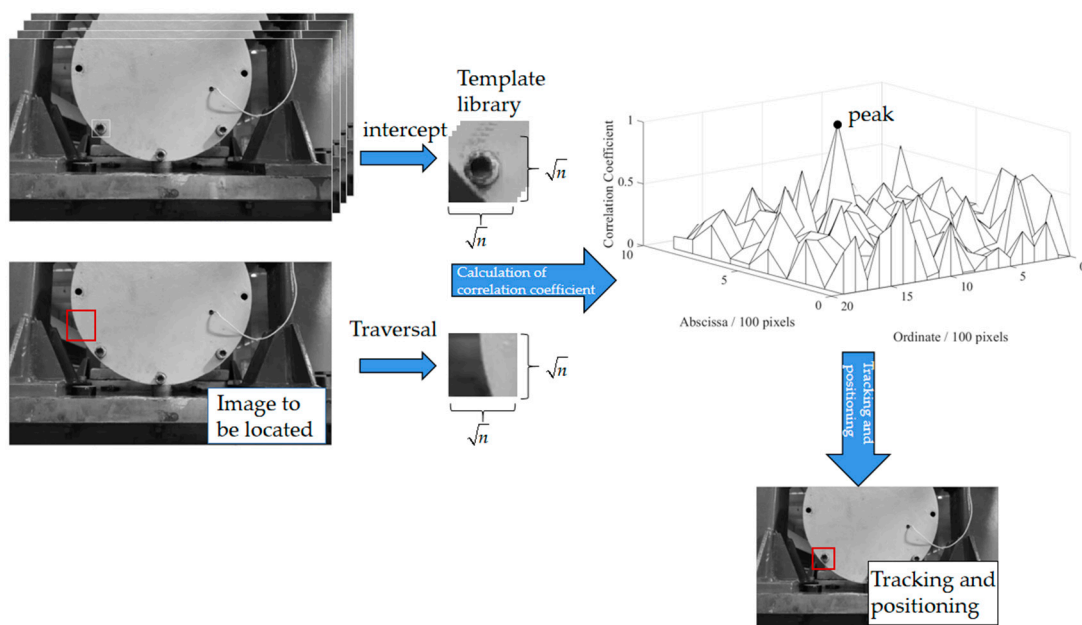

Figure 10. Template matching process.

The correlation coefficient r ranges from 0 to 1, and the peak position in the correlation coefficient map in Figure 10 is the matching position. In Figure 10, the abscissa corresponds to the abscissa of the image, the ordinate corresponds to the ordinate of the image, and the vertical coordinate is the correlation coefficient. The horizontal and vertical coordinates are in pixels, and the vertical coordinates are unified dimensions. Figure 11 shows the effect picture of the image taken under different illumination conditions in Figure 9. When the size of the image to be positioned is input as 960×544, the size of the template is set as 51×51 and the number of templates is 1; the calculation time is shown in Table 2 (the computer CPU used for calculation was Inter(R) Core(TM) I7-4790 CPU @ 3.60 GHz).

Table 2. Calculation times of template matching.

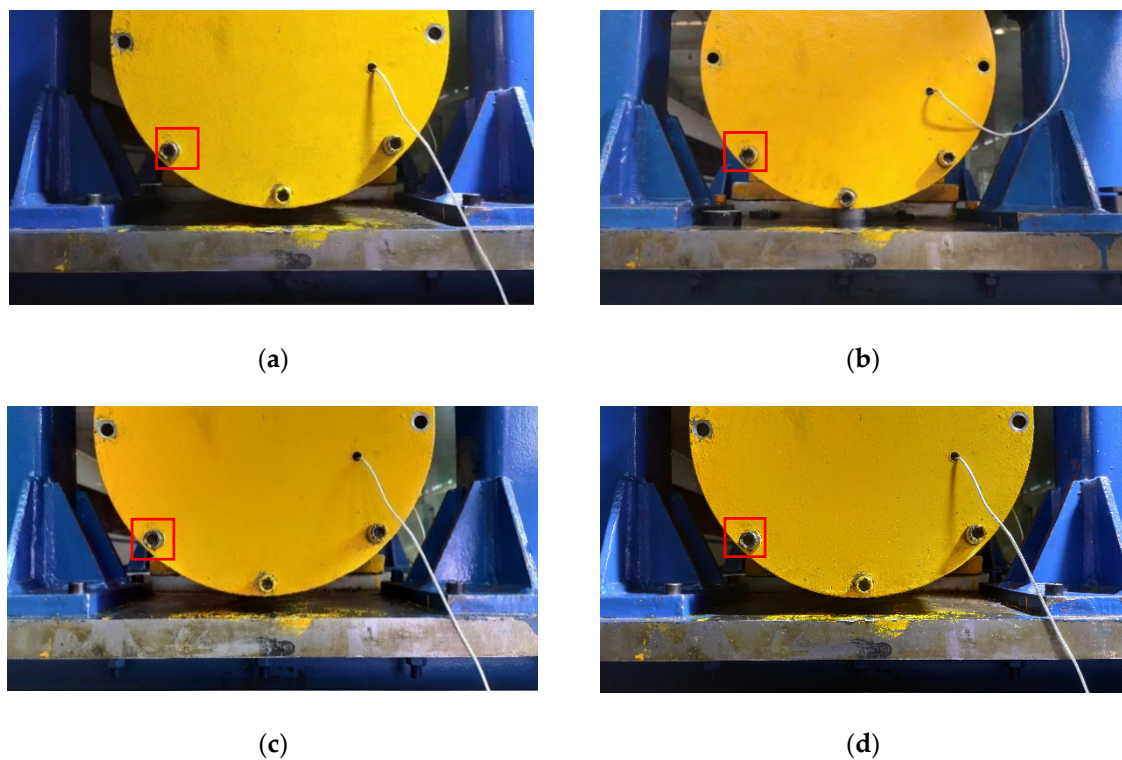| Image Number	Processing Time (s)
1	305.0
2	305.0
3	305.0
4	303.0
mean value	304.5

(a)

(b)

(c)

(d)

Figure 11. Template matching effect image under different illumination conditions (as Figure 9): (**a**) image under strong supplementary light source condition; (**b**) image under strong natural light conditions; (**c**) images under low natural light conditions; (**d**) image under weak supplementary light source condition.

According to the positioning effect image in Figure 11 and the calculation time in Table 1, it can be seen that the template matching method has a good positioning effect and high robustness to the illumination conditions of the image. However, this method needs to traverse the entire image removal matrix and calculate multiple correlation coefficients, so the calculation time is relatively long, and the average positioning time is 304.5 seconds per piece, which is difficult to meet the real-time displacement identification requirements of the platform.

3.2.2. Contour Feature Positioning Method

The contour feature location method skips the gray pixel matrix of the image and directly uses the outstanding contour features in the image, establishing the geometric model through auxiliary graphics to complete the positioning. In order to more conveniently observe and select the contour features used in positioning, the image is first detected by canny edge, as shown in Figure 12. We can observe that the obvious and easy-to-detect contour features in the image are the two horizontal lines, L_1 and L_2 running through the whole image, and the contour features of these two lines are selected as the reference lines for the establishment of the geometric model. Observed that bolts P_1 and P_3 were symmetrically distributed about bolt P_2, and bolt P_2 was just on the center line L_3 of the circular contour. L_1, L_2, L_3, P_1, P_2, P_3 and the circular contour were extracted from the edge image Figure 12 to establish the geometric model, as shown in Figure 13.
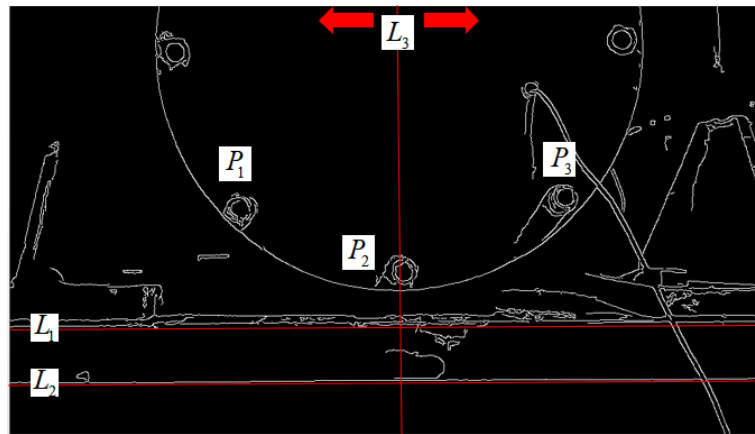
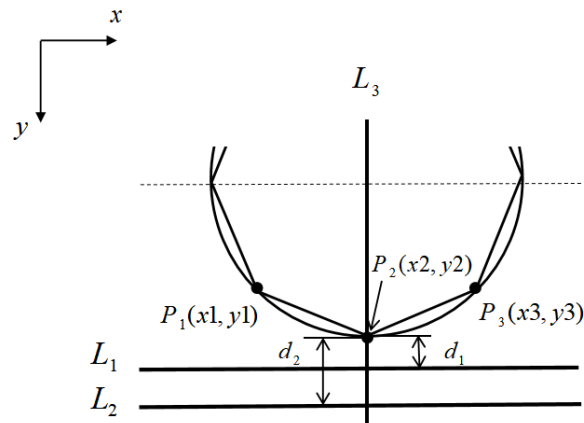**Figure 12.** Effect of canny edge detection.

Figure 13. Simplified geometric model of end cover and bolt.

After the geometric model shown in Figure 13 is established, the positioning algorithm flow according to the established geometric model is shown in Figure 14. The specific steps are as follows.
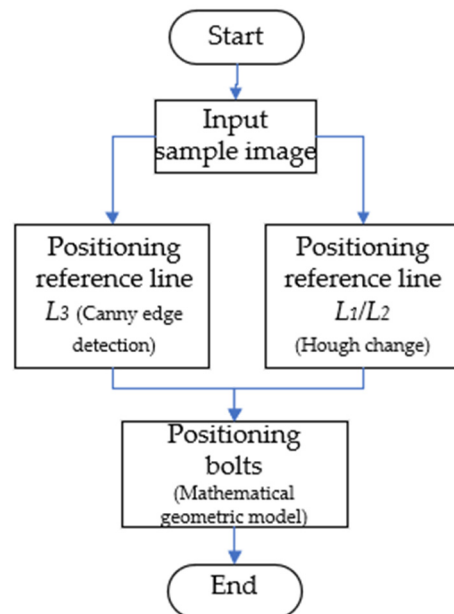

Figure 14. Bolt positioning flow chart.

- Positioning reference line L_3: canny edge detection for the image obtains the pixel matrix A of the image, as shown in Figure 12. The value of 255 in the matrix is the edge point, then the x-coordinate value of the reference line L_3 is determined by Equation (4):

$$X = \frac{X2 - X1}{2} + X1 \qquad (4)$$

where $X1$ and $X2$ are the number of columns where the first edge point is located when traversing from the middle of the matrix to both sides in the first row of matrix A (as shown by the arrow in Figure 12).

- Positioning reference line L_1/L_2: Hough transform is adopted to detect the straight line, and the edge line of the excitation platform is positioned as the reference line, but there are two upper and lower edge lines L_1 and L_2, as shown in Figure 13. It can be observed that, when positioned as the lower edge line L_1, the value of the pixel matrix of the grayscale image below the edge line is obviously smaller than that located at the upper edge line L_2. According to Equation (5), whether the positioned edge line is the upper edge line L_1 or the lower edge line L_2 can be determined:

$$\begin{cases} L_1 & f(\frac{M_2+M_1}{2} + a, X) < b \\ L_2 & f(\frac{M_2+M_1}{2} + a, X) \geq b \end{cases} \qquad (5)$$

where M_1 and M_2 are the vertical coordinates of the two endpoints of the detected line, and $f(x, y)$ is the gray value of the image at (x, y). a is the error value of the positioning edge line, and is the empirical parameter, b is the gray value below the edge line L_2, and is an empirical parameter. In this article, $a = 6$, $b = 40$.

When the positioning reference line is determined to be L_1 or L_2, the vertical distance between the reference line and the bolt P_2 is d_1 or d_2:

$$d_1 = H * k_1 \qquad (6)$$

$$d_2 = H * k_2 \qquad (7)$$

where H is the height of the image; H = 544 in this article. k_1 and k_2 are empirical parameters, $k_1 = 0.1$; $k_2 = 0.25$.

- Positioning bolts: after positioning the reference line, and according to the geometric relationship shown in Figure 13, the locations of the center points of the bolt are at the positions $P_1(x1, y1)$, $P_2(x2, y2)$, $P_3(x3, y3)$. When the positioning edge line is L_1:

$$\begin{cases} x1 = X * (1 - k_3) \\ x2 = X \\ x3 = X * (1 + k_3) \\ y2 = \frac{M_2+M_1}{2} - d_1 \\ y1 = y3 = \frac{M_2+M_1}{2} - d_1 - k_4 * H \end{cases} \qquad (8)$$

When the positioning edge line is L_2:

$$\begin{cases} x1 = X * (1 - k_3) \\ x2 = X \\ x3 = X * (1 + k_3) \\ y2 = \frac{M_2+M_1}{2} - d_2 \\ y1 = y3 = \frac{M_2+M_1}{2} - d_2 - k_4 * H \end{cases} \qquad (9)$$

where k_3 and k_4 are empirical parameters.

After positioning to the center point of the bolts, the images of the bolts can be captured according to a certain size of the rectangular box. The positioning effect pictures are shown

in Figure 15, and the calculation time is shown in Table 3 (the computer CPU used for calculation was Inter(R) Core(TM) I7-4790 CPU @ 3.60 GHz).
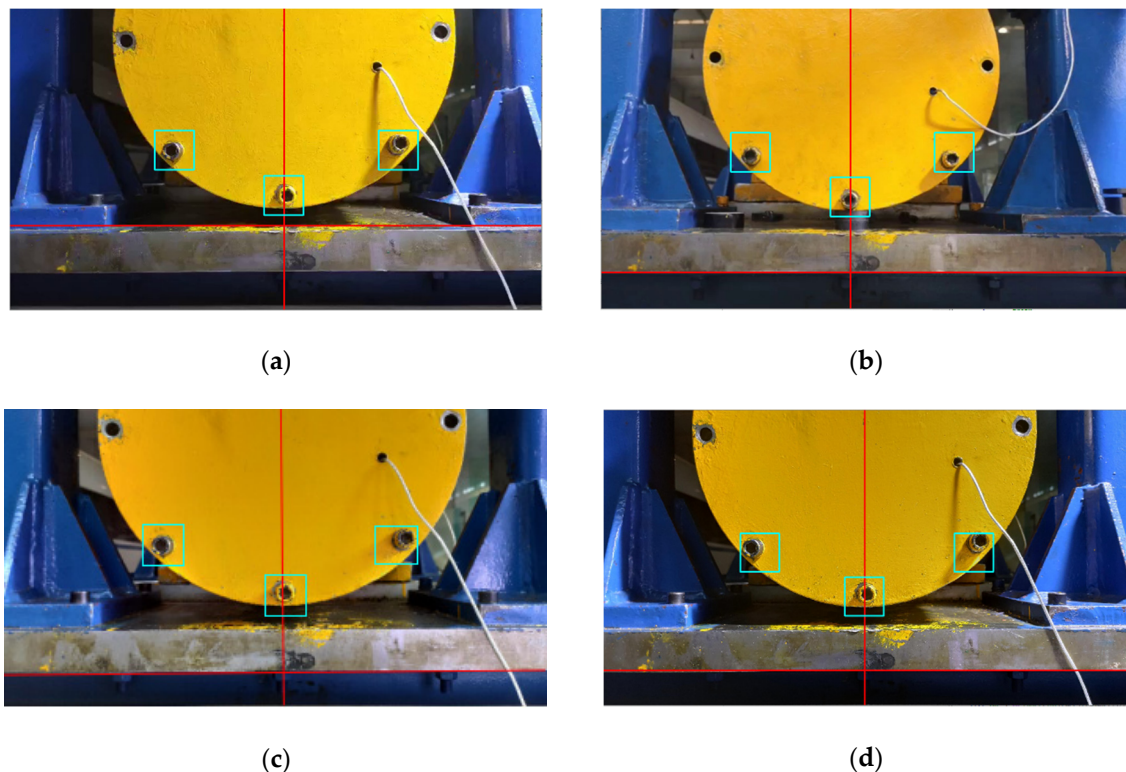

(**a**)

(**b**)

(**c**)

(**d**)

Figure 15. Contour feature positioning effect image under different illumination conditions (as Figure 9): (**a**) image under strong supplementary light source condition; (**b**) image under strong natural light conditions; (**c**) images under low natural light conditions; (**d**) image under weak supplementary light source condition.

Table 3. Calculation times of contour feature positioning.

Image Number	Processing Time (s)
1	0.02
2	0.02
3	0.02
4	0.02
mean value	0.02

The red line in Figure 15a is the case of the reference line L_1 and L_3 of the positioning, and the case of the reference line L_2 and L_3 of the positioning is shown in Figure 15b–d. In the green rectangles are the locations of P_1, P_2 and P_3 bolts. The method of using contour lines to establish reference lines to assist positioning has a good effect. Moreover, it can accurately identify and position objects under different lighting conditions and different surface textures, reflecting the robustness of the algorithm to surface texture changes and light intensity of objects. The average calculation speed is 0.02s to complete the positioning, which can meet the requirements of real-time monitoring and identification of the displacement of the platform.

Two different positioning methods of template matching and contour feature are compared. As can be seen from the positioning renderings in Figures 11 and 15, the positioning effect of the two methods is good, and the robustness to ray intensity is very good, which can realize the automatic positioning requirements of bolts under different lighting conditions. According to the positioning times required by the two positioning methods in Tables 2 and 3, the average positioning time of the template matching method

is 304.5 s. The average of the contour feature method is 0.02 s, which greatly reduces the positioning time. Although the two positioning methods can achieve the same positioning effect, considering the requirements of real-time monitoring, this article adopted the second positioning method by contour feature for bolt positioning interception, and then carried out subsequent displacement calculation.

3.3. Displacement Calculation

In this article, the phase correlation method is used to calculate the displacement of the collected video images. The phase correlation algorithm [24] is a frequency-domain correlation algorithm based on Fourier power spectrum. This method only takes the phase information in the cross-power spectrum, which reduces the dependence on the image content. In addition, the obtained correlation peak is sharp and prominent, the detection range of displacement is large, and the matching accuracy is high. The image gray scale is less dependent and has a certain anti-interference ability. Assuming that $f_2(x, y)$ and $f_1(x, y)$ are two image signals, and $f_2(x, y)$ is obtained by $f_1(x, y)$ translation (dx, dy), which satisfies the following formula:

$$f_2(x, y) = f_1(x - dx, y - dy) \tag{10}$$

It is reflected in the frequency domain in the form of:

$$F_2(\mu, v) = F_1(\mu, v) * e^{-i*2\pi*(\mu*dx+v*dy)} \tag{11}$$

The cross-power spectrum of $f_2(x, y)$ and $f_1(x, y)$ can be obtained from the above formula:

$$H(\mu, v) = \frac{F_1 * F_2^*}{|A_1| * |A_2^*|} = e^{-i*2\pi*(\mu*dx+v*dy)} \tag{12}$$

where F^* is the conjugate of F.

The inverse Fourier transform of the cross-power spectrum can obtain a Dirac function (pulse function) and find the offset by finding the coordinates of the peak. However, this method can only obtain the displacement of the pixel level. Then, the peak position can be found according to the above, and a weighted mean of response size can be processed in $a \times a$ form centered on this position. The following formula can be applied to obtain the precision position at the sub-pixel level:

$$x = \frac{\sum_{a \times a} i f(i, j)}{\sum_{a \times a} f(i, j)} \tag{13}$$

$$y = \frac{\sum_{a \times a} j f(i, j)}{\sum_{a \times a} f(i, j)} \tag{14}$$

The final (x, y) is the subpixel displacement between the two images.

For the selection of the calculated displacement amplitudes, this article proposes a new method to convert the displacement amplitudes into images and calculate the correlation of the GLCM of the displacement amplitudes. The displacement amplitudes with the greatest correlation, namely the most periodic and obvious, are selected to represent the displacement of the platform.

In image processing, when not only the distribution of gray level but also the relative position of pixels in the image should be considered, the GLCM of the image is usually generated [25]. Let Q be an operator to define the relative position of two pixels, and consider an image f with L possible gray levels. Let G be a matrix, whose element g_{ij} is the number of pixels with grays of z_i and z_j appearing in the position indicated by Q in f, where $1 \leq i, j \leq L$. The matrix formed in this way is called GLCM. Figure 16 shows an example of how the GLCM is constructed using $L = 8$ and the position operator Q defined by "one pixel to its right". The array on the left is the small image in consideration, and the array on the right is the matrix G.
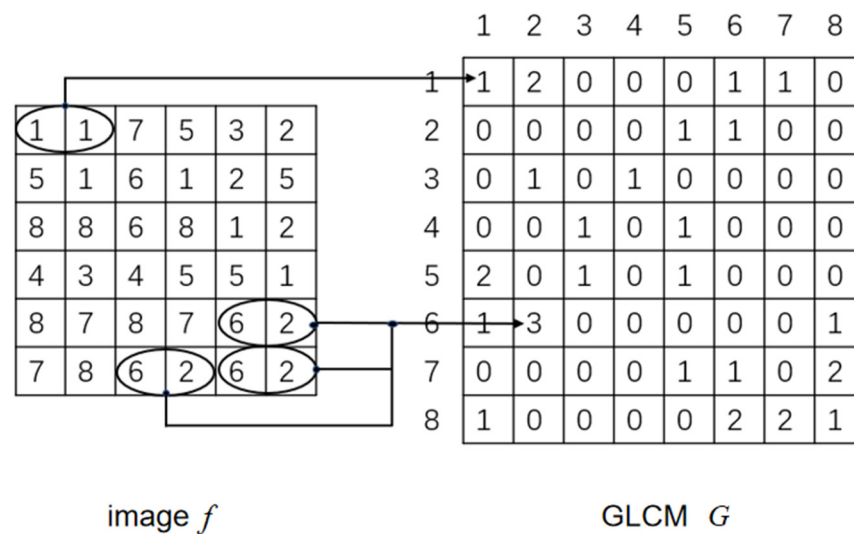
Figure 16. How to generate the gray-level co-occurrence matrix (GLCM).

Element (1,1) of the GLCM G is 1, because in f, the pixel with value of 1 to the right of a pixel with value of 1 appears only once. Similarly, the element of G (6,2) is 3, because in f, the right value of a pixel with value of 6 appears three times with value of 2, and the possible gray level of the image determines the size of the matrix G. The total number of pixel pairs n that satisfy Q is equal to the sum of the elements of G. As a result:

$$p_{ij} = g_{ij}/n \tag{15}$$

Correlation is a descriptor of the characteristics of the GLCM and is a measure of how closely a pixel is related to its neighbors on the entire image. The range is $[1, -1]$, which corresponds to perfect positive correlation and perfect negative correlation. The correlation is calculated as follows:

$$\sum_{i=1}^{K}\sum_{j=1}^{K}\frac{(i - m_r)(j - m_c)p_{ij}}{\sigma_r\sigma_c} \tag{16}$$

In Equation (16), the quantity used is defined as follows:

$$m_r = \sum_{i=1}^{K} i \sum_{j=1}^{K} p_{ij} \tag{17}$$

$$m_c = \sum_{j=1}^{K} j \sum_{i=1}^{K} p_{ij} \tag{18}$$

$$\sigma_r^2 = \sum_{i=1}^{K} (i - m_r)^2 \sum_{j=1}^{K} p_{ij} \tag{19}$$

$$\sigma_c^2 = \sum_{j=1}^{K} (j - m_c)^2 \sum_{i=1}^{K} p_{ij} \tag{20}$$

where $K \times K$ is the size of the GLCM.

4. Experimental Results and Analysis

4.1. Measured Results

Double row cylindrical roller bearings were used in the experiment; the displacement image of platform in stable state for 30 s was collected under the working conditions of rotation speed $n = 0$, static load $F = 0$ kg and excitation frequency $f = 6$ Hz. The displacement was calculated by using the phase correlation method of local images proposed in this

article. The frame rate of the portable camera used in this test was 30 frames s^{-1}; and the portable camera was placed on the central axis of the platform, and the camera calibration was carried out in accordance with Section 3.1. The computer CPU used for calculation was Inter(R) Core(TM) I7-4790 CPU @ 3.60 GHz. The sampling frequency in the experiment was more than twice that of the excitation frequency of the platform; therefore, the influence of temporal aliasing effect and rolling shutter effect on the image was not considered. According to the method proposed in Section 3.3., the vertical displacement amplitude of the input three bolts and the whole image were calculated, respectively. The four displacement amplitudes are shown in Figure 17, and the calculation times are shown in Table 4. The displacement amplitudes calculated from the three bolts were generated by the GLCM of position operator Q defined as "one pixel to the right of the bolt". By calculating the correlation of GLCM, the amplitudes with the greatest correlation, namely, the strongest periodicity, were selected to represent the vertical displacement amplitudes of the platform. The correlation of the three bolts according to their respective GLCM is shown in Table 5.
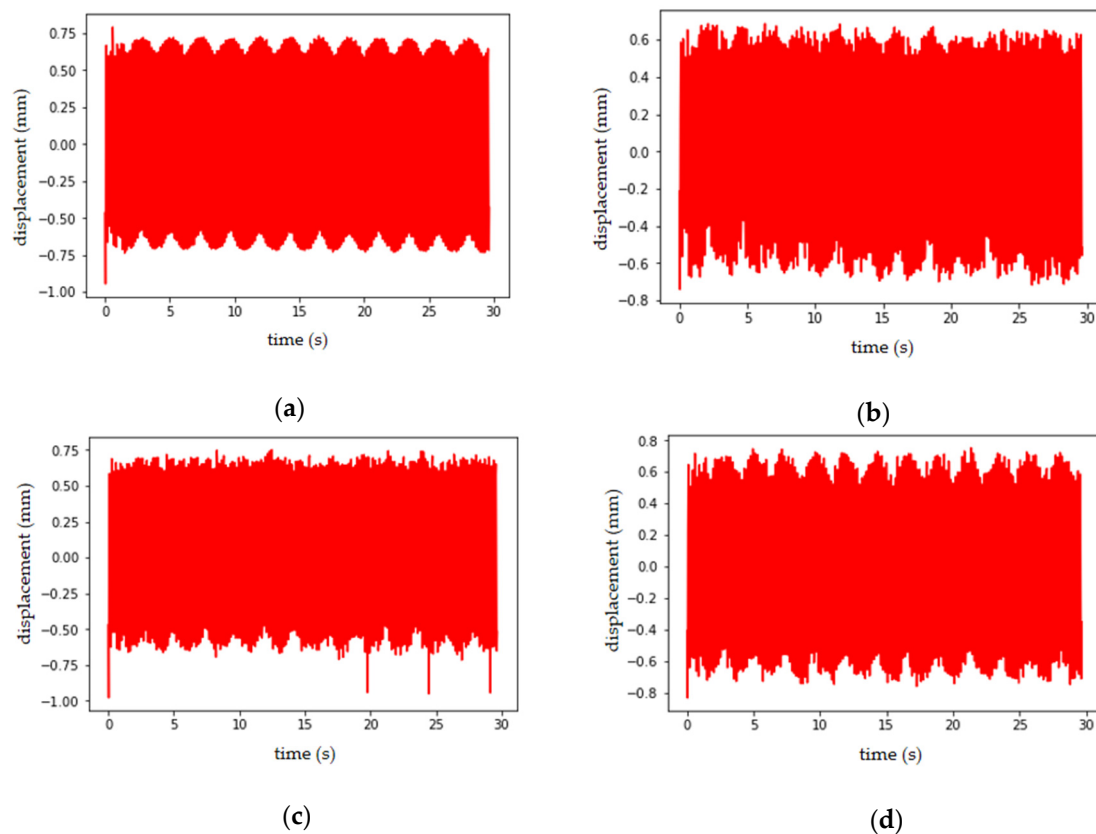

(a)

(b)

(c)

(d)

Figure 17. Vertical displacement amplitude diagram: (**a**) the vertical displacement amplitudes obtained when the input is a whole image; (**b**) the vertical displacement amplitudes obtained when the input is bolt P_1; (**c**) the vertical displacement amplitudes obtained when the input is bolt P_2; (**d**) the vertical displacement amplitudes obtained when the input is bolt P_3.

Table 4. Calculation times of input different images.

| Input Image	Operating Time (s)
Whole image	24.199
Bolt P_1	6.895
Bolt P_2	7.252
Bolt P_3	7.000
Bolts P_1, P_2, P_3	7.221

Table 5. Correlation of the GLCM of different vertical displacement amplitudes.

Displacement Amplitude Diagram	Correlation (GLCM)
Bolt P_1	0.9098
Bolt P_2	0.9986
Bolt P_3	0.9991

By observing Table 4, it can be seen that the vertical displacement amplitudes calculated from the local images of input bolt P_3 have the strongest periodicity; thus, the displacement amplitudes at bolt P_3 are shown as the vertical displacement amplitudes of the platform.

4.2. Verification of Results

In order to verify the effectiveness and accuracy of the method in this article, the real displacement amplitude of the platform was collected by using a laser displacement sensor under the same working condition as in Section 4.1. Moreover, the comparison and verification were made from the peak mean value of the displacement amplitude (as shown in Table 6) and the spectrum diagram (as shown in Figure 18).
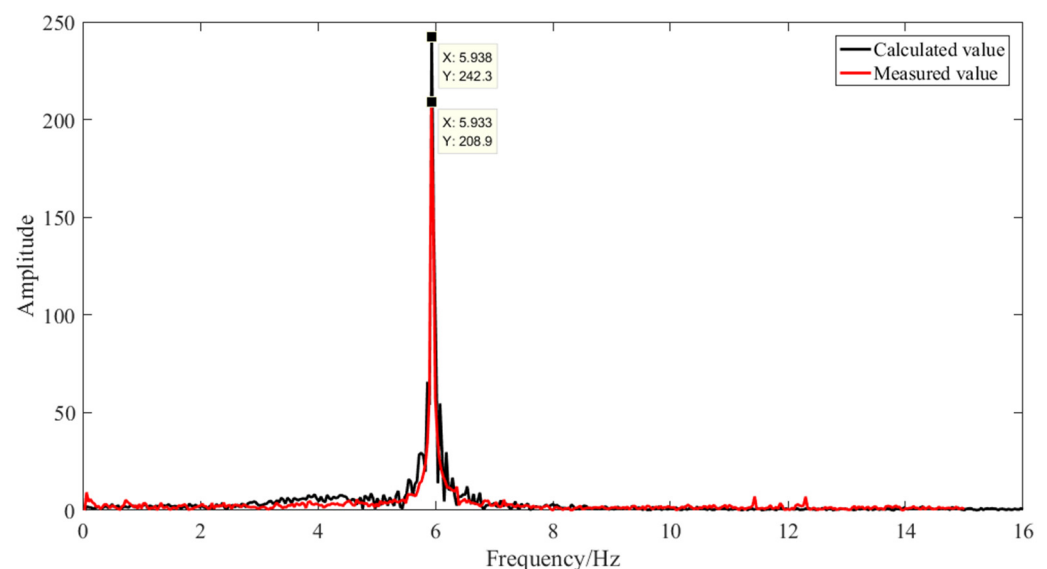
Figure 18. Spectrum comparison diagram.

Table 6. Peak mean values of displacement amplitudes of different displacement images.

Displacement Amplitude Diagram	Peak Mean (mm)
Bolt P_1	0.6041
Bolt P_2	0.6367
Bolt P_3	0.6465
Actual measurement	0.6973

By comparing the spectrum diagram shown in Figure 18, it can be seen that the frequency of the vertical displacement amplitude of the platform calculated from the local image of the bolt P_3 selected by the method described in Section 4.1. differs little from the actual measured value. From Table 5, the average peak of the three bolts' displacement amplitudes and laser displacement sensor displacement amplitude were contrasted; bolt P_3 is closer to the value measured by the sensor, and the error is 0.05 mm, which verifies that the selection of displacement amplitude based on the correlation of the GLCM is effective.

In this article, local (bolt) images were input to replace the whole image for displacement calculation, and GLCM correlation of different displacement amplitudes was

calculated by converting the displacement amplitudes into images, and the method of final calculation result of displacement amplitudes with the maximum correlation was selected. By observing Tables 4 and 6, it can be seen that the calculation rate can be increased by 68% if the local image is input instead of the whole image, and the calculated displacement error can be guaranteed to be 0.05 mm.

4.3. Analysis of Displacement Amplitude under Different Working Conditions

The method verified in this article is used to calculate the displacement amplitude of the platform under different working conditions. The average values of the peak displacement amplitudes are shown in Table 7.

Table 7. Mean peak values of displacement amplitudes under different working conditions.

Working Condition				Peak Mean (mm)
Number of Groups	**Load (kg)**	**Rotation Speed (n/min)**	**Excitation (Hz)**	
1	600	500	6	0.738
2	600	800	6	0.757
3	600	1100	6	0.855
4	1200	500	6	0.707
5	1200	800	6	0.718
6	1200	1100	6	0.737

In Table 7, by comparing different working conditions, it is found that when static load and excitation frequency are unchanged, the vertical displacement amplitude of the platform tends to increase with the rotating speed. The vertical displacement amplitude of the platform tends to decrease with the static load when the rotational speed and excitation frequency remain unchanged.

5. Discussion

In this article, the vertical displacement of the platform was monitored by computer vision. The local image of the bolts on the bearing end cover were used instead of the whole image for calculation. Although the calculation speed can be increased by 68% with this method and an accuracy of 0.05 mm can be guaranteed, there are still several challenges that remain open for future investigation, and some critical challenges are discussed in detail below: (1) Limited frame rate and temporal aliasing effect: temporal aliasing is caused by the sampling rate (i.e., number of frames per second) of a scene being too low compared to the transformation speed of objects inside of the scene; this causes objects to appear to jump or appear at a location instead of smoothly moving, which can cause errors in the calculation of displacement. When the excitation frequency of the platform gradually increases and gradually exceeds the sampling rate of the camera, we cannot blindly adopt the camera with high frame rate. Therefore, the time aliasing effect needs to be studied and solved in the next stage of research. (2) Rolling shutter effect: most cameras use complementary metal oxide semiconductor (CMOS) sensors. The CMOS sensor uses a sequential readout, scanning each line exposed at different times to obtain the image, resulting in geometric distortion, especially when the relative velocity between the camera and the object is high. As the excitation frequency of the platform increases, in order to reduce the error, additional research is needed to eliminate the shutter curtain effect in the case that the speed of the structure is large relative to the camera.

6. Conclusions and Future Work

This article proposed to use displacement signal to monitor the state of rail vehicle bearings. With the help of an experimental platform of bearing system, a method of displacement monitoring by computer vision detection was explored to identify the displacement. The vertical displacement of all the components in the whole bearing system

was the same, to improve the calculation rate and meet the purpose of real-time displacement monitoring; the bolts on the axle end cap were used for displacement identification. Two positioning methods were compared: the template matching method and contour feature method. Considering localization accuracy and localization efficiency, the localization method based on contour feature was chosen. The displacement amplitudes of the bearing system in the vertical direction were derived by comparing the correlations of the image's gray-level co-occurrence matrix (GLCM). The measured displacement of the laser displacement sensor was compared with the calculated displacement in the frequency domain and the mean value of the peak displacement to verify the accuracy of the proposed method.

The following conclusions and findings are drawn:

1. Contour feature to locate the bolt is much faster than using the template matching method. The locating rate is 0.024 s/sheet on average, and the bolt is more robust to illumination conditions;
2. By replacing the whole image with a local image, the phase correlation method can improve the calculation rate by 68% with an accuracy of 0.05 mm;
3. According to the correlation of GLCM of the displacement amplitude image, the displacement amplitude graph is the closest to the real value;
4. The method of replacing the whole image with the local (bolt) images to calculate the displacement proposed in this article was used to calculate the displacement of six groups of bearing platforms under different test conditions;
5. It was found that the vertical displacement amplitude of the bearing system increases with the increase in rotating speed and decreases with the increase in static load;
6. It is feasible and effective to introduce displacement signals to monitor the state of bearings;
7. At the same time, practical considerations and limitations of the proposed method were discussed, including the issues of the temporal aliasing effect, and rolling shutter effect. In addition, the following aspects are determined for future research: displacement amplitudes of the bearing system in the vertical direction should be derived by comparing more parameters of the image's gray-level, rather than just by comparing co-occurrence matrix (GLCM); studying the influence of the acquisition distance on the accuracy of displacement calculation by setting the distance between different portable equipment placement points and the platform; diagnosing bearing faults by the displacement signals.

Author Contributions: Conceptualization, L.L. and D.S.; methodology, L.L. and D.S.; software, L.L.; validation, L.L., Z.Z. and X.X.; formal analysis, L.L. and X.X.; investigation, L.L.; resources, L.L.; data curation, L.L.; writing—original draft preparation, L.L.; writing—review and editing, L.L. and Z.L.; visualization, X.X.; supervision, L.L.; project administration, D.S.; funding acquisition, D.S. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, Y.J. *Research on Intelligent Fault Diagnosis Technique of Axle Box Bearing of High Speed Train*; Southwest Jiaotong University: Chengdu, China, 2017.
2. Niu, Z.H. *Vibration Characteristics and Experimental Study of Axle Box System of High Speed Train*; Jilin University: Jilin, China, 2019.

3. Liu, L.; Song, D.N.; Geng, Z.L.; Zheng, Z.J. A Real-Time Fault Early Warning Method for a High-Speed EMU Axle Box Bearing. *Sensors* **2020**, *3*, 823. [CrossRef] [PubMed]

4. Montalvo, J.; Tarawneh, C.; Lima, J.; Cuanang, J.; Santos, N.D.L. Estimating the Outer Ring Defect Size and Remaining Service Life of Freight Railcar Bearings Using Vibration Signatures. In Proceedings of the 2019 Joint Rail Conference, Salt Lake City, UT, USA, 9–12 April 2019.

5. Gul, M.; Dumlupinar, T.; Hattori, H.; Catbas, N. Structural Monitoring of Movable Bridge Mechanical Components for Maintenance Decision-making. *Struct. Monit. Maint.* **2014**, *1*, 249–271. [CrossRef]

6. Gul, M.; Catbas, F.N.; Hattori, H. Image-based Monitoring of Open Gears of Movable Bridges for Condition Assessment and Maintenance Decision Making. *J. Comput. Civil Eng.* **2015**, *29*, 04014034. [CrossRef]

7. Garcia-palencia, A.; Santini-bell, E.; Gul, M.; Catbas, N.A. FRF-based Algorithm for Damage Detection Using Experimentally Collected Data. *Struct. Monit. Maint.* **2015**, *2*, 399–418. [CrossRef]

8. Ojin, T.; Carey, C.H.; Obrien, E.J.; Doherty, C.; Taylor, S.E. Contactless Bridge Weigh-in-motion. *J. Bridge Eng.* **2016**, *21*, 04016032.

9. Moreu, F.; Li, J.; Jo, H.; Kim, R.E.; Scloa, S.; Spencer, B.F.; Lafave, J.M. Reference-free Displacements for Condition Assessment of Timber Railroad Bridges. *J. Bridge Eng.* **2016**, *21*, 04015052. [CrossRef]

10. Xu, Y.; Brownjohn, J.; Kong, D.A. Non-contact Vision-based System for Multipoint Displacement Monitoring in a Cable-stayed Footbridge. *Struct. Control Health Monit.* **2018**, *25*, 1–23. [CrossRef]

11. Hester, D.; Brownjohn, J.; Bocian, M.; Xu, Y. Low Cost Bridge Load Test: Calculating Bridge Displacement from Acceleration for Load Assessment Calculations. *Eng. Struct.* **2017**, *143*, 358–374. [CrossRef]

12. Celik, O.; Dong, C.Z.; Catbas, F.N. A Computer Vision Approach for the Load Time History Estimation of Lively Individuals and Crowds. *Comput. Struct.* **2018**, *200*, 32–52. [CrossRef]

13. Catbas, N.; Zaurin, R.; Gul, M.; Gokce, H.B. Sensor Networks, Computer Imaging, and Unit Influence Lines for Structural Health Monitoring: Case Study for Bridge Load Rating. *J. Bridge Eng.* **2012**, *17*, 662–670. [CrossRef]

14. Kguc, T.; Catbas, F.N. Structural Identification Using Computer Vision-based Bridge Health Monitoring. *J. Struct. Eng.* **2018**, *2*, 04017202.

15. Dong, C.Z.; Celik, O.; Catbas, F.N. Marker Free Monitoring of the Grandstand Structures and Modal Identification Using Computer Vision Methods. *Struct. Health Monit.* **2019**, *18*, 1491–1509. [CrossRef]

16. Yang, Y.; Dorn, C.; Mancini, T.; Taiken, Z.; Kenyon, G.; Farrar, C.; Mascarenas, D. Blind Identification of Full-field Vibration Modes from Video Measurements with Phase-based Video Motion Magnification. *Mech. Syst. Signal Process.* **2017**, *85*, 567–590. [CrossRef]

17. Ye, X.W.; Dong, C.Z.; Liu, T. A Review of Machine Vision-based Structural Health Monitoring: Methodologies and Applications. *J. Sens.* **2016**, 7103039. [CrossRef]

18. Yoon, H.; Shin, J.; Spencer, B.F. Structural Displacement Measurement Using an Unmanned Aerial System. *Comput. -Aided Civil Infrastruct. Eng.* **2018**, *33*, 183–192. [CrossRef]

19. Ye, X.W.; Dong, C.Z.; Liu, T. Image-based Structural Dynamic Displacement Measurement Using Different Multi-object Tracking Algorithms. *Smart Struct. Syst.* **2016**, *17*, 935–956. [CrossRef]

20. Tian, Y.; Zhang, J.; Yu, S. Rapid Impact Testing and System Identification of Footbridges Using Particle Image Velocimetry. *Comput.-Aided Civil Infrastruct. Eng.* **2019**, *34*, 130–145. [CrossRef]

21. Chany, C.Y.; Chang, E.C.; Huang, C.W. In Situ Diagnosis of Industrial Motors by Using Vision-based Smart Sensing Technology. *Sensors* **2019**, *19*, 5340. [CrossRef]

22. Liu, Y.Q. Track Displacement Monitoring System Based on Image Processing. *Comput. Appl. Softw.* **2019**, *36*, 247–250.

23. Ye, X.W.; Dong, C.Z. Review of Computer Vision-based Structural Displacement Monitoring. *China J. Highw. Transp* **2019**, *32*, 22–39.

24. Balci, M.; Foroosh, H. Subpixel Estimation of Shifts Directly in the Fourier Domain. *IEEE Trans. Image Process.* **2006**. [CrossRef] [PubMed]

25. Rafael, C.G.; Richard, E.W. *Digital Image Processing*, 3rd ed.; Pulishing House of Eletronics Industry: Beijing, China, 2017; pp. 536–539.