

Research Article

Computer Vision-Based Structural Displacement Monitoring and Modal Identification with Subpixel Localization Refinement

Tao Liu ^{1,2}, Yu Lei ¹ and Yibing Mao ²

¹Department of Civil Engineering, Xiamen University, Xiamen 361005, China

²Technology Research Center for Reinforcement and Translation of Buildings, Xiamen University, Xiamen 361005, China

Correspondence should be addressed to Tao Liu; taoliu@xmu.edu.cn

Received 21 February 2022; Accepted 26 May 2022; Published 30 June 2022

Academic Editor: Thanh-Canh Huynh

Copyright © 2022 Tao Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In conventional structural health monitoring (SHM), the installation of sensors and data acquisition devices will affect the regular operation of structures to a certain extent and is also expensive. In order to overcome these shortcomings, the computer vision (CV-) based method has been introduced into SHM, and its practical applications are increasing. In this paper, CV-based SHM methods such as template matching and Hough circle transform are described. In order to improve the accuracy of pixel localization, the subpixel localization refinement method is introduced. The displacement monitoring experiment of an aluminum alloy cantilever with three targets is conducted by using the two CV-based SHM methods and the laser displacement sensors simultaneously. The displacement monitoring results of CV-based methods agree well with those measured by the laser transducer system in the time domain. After that, the first two modes of the cantilever are identified from the monitoring results. In addition, the experimental modes identified from the monitoring data and those calculated from the finite element model are also consistent. Therefore, the developed CV-based methods can obtain accurate displacement results in both time and frequency domains, which could be applied to complex structures with more monitoring targets.

1. Introduction

With the increasing applications of artificial intelligence (AI) in various industries, construction engineering and management are experiencing a rapid digital transformation [1], and civil engineering also benefits a lot from AI-based methods inevitably. Computer vision is an important branch of AI-based techniques and provides new ideas and novel solutions for SHM. In recent years, computer vision-based methods have been widely applied in several aspects, from laboratory tests to realistic structures, mainly including crack detection, displacement monitoring, system identification, vehicle load identification, and so on [2].

In order to evaluate the structural performance and safety condition, external loads and structural responses are often measured to obtain the dynamic and static characteristics of the structure. Displacement is of vital importance among all the measured indicators because other indicators such as the bearing capacity, the deflection, the load

distribution, and the modal parameters can be calculated indirectly from the displacement [3]. In conventional SHM systems, the cumbersome installation of expensive hardware is often required, such as different types of monitoring sensors and data acquisition devices [4], and megabytes of data need to be preprocessed and analyzed so as to evaluate the actual state of the structure [5].

In recent years, CV-based displacement monitoring methods have been emerging with the rapid development of computer vision and optical devices. Compared with conventional techniques, CV-based displacement monitoring methods are noncontact and have such advantages as real-time detection, long-term stability, remote monitoring capability, controllable image-taking speed, and data archiving applications [6]. However, CV-based methods face the challenge of long-distance monitoring, and the monitoring accuracy is often influenced by environmental factors, such as wind, vibration, temperature, airflow, and mist. Even so, CV-based displacement monitoring methods are

continuously improved and have been widely applied to identify the modal parameters of structures [7–11].

Regarding the CV-based displacement monitoring system, the videos of structural vibration are first collected by cameras and then discretized to successive frames. After that, the targets in each frame will be tracked by the target tracking algorithm to calculate the target displacement in the image coordinate system. Once the projection transforming from the world coordinate system to the image coordinate system and determining through camera calibration, the actual displacement of targets can be easily calculated from the successive image frames. The CV-based monitoring methods mainly include the following four aspects: camera calibration, feature extraction, target tracking, and displacement calculation. Target tracking is regarded as the key in the whole process and can be generally categorized into four types according to the features to be tracked: template matching, feature point matching, optical flow estimation, and shape-based tracking. In contrast, camera calibration and displacement calculation are relatively simple in practical applications.

Specifically, the template matching for target tracking has been applied to monitor the displacement response and identify the modal characteristics [7, 8, 12], the effectiveness of which is also validated by the experimental results. In terms of feature point matching, references [13, 14] used digital image processing techniques to extract the feature points and monitor the displacement. In the frequency domain, the natural frequency and mode shape of structures can be accurately identified. Many scholars have presented the practical applications of optical flow estimation for displacement monitoring and system identification. Khaloo and Lattanzi [15] employed the combination of parametric video stabilization, 3D denoising, and outlier robust camera motion estimation to mitigate the effects of camera motion and video encoding artifacts and assessed four canonical optical flow algorithms. Hoskere et al. [16] presented a new approach to facilitate the extraction of frequencies and mode shapes of full-scale civil infrastructure from video obtained by an unmanned aerial vehicle and directly addressed a number of difficulties associated with modal analysis of full-scale infrastructure using CV-based methods. Dong et al. [17] proposed a novel structural displacement measurement method using deep learning-based full-field optical flow methods and investigated the image collection, tracking, and nonuniform sampling in the experimental data so as to obtain more accurate displacement measurements. However, the monitoring accuracy of the above research only reaches the pixel level and the subpixel localization refinement is less studied. Regarding the shape-based tracking strategy for target tracking in displacement monitoring, Han et al. [18] used digital image processing techniques to extract the rings from the background and measure the displacement of a reinforced concrete frame structure model with infill walls in a large-scale shaking table test. Shan et al. [19] utilized a Canny-Zernike combination algorithm to attain the pixel edge of a circular target and conducted a free vibration test of a stayed-

cable model in the lab. Chen et al. [20] developed a CV-based dynamic displacement test method based on the circle detection algorithms and applied it to the large-scale shaking table tests. In addition, they suggested the applicable scopes of the Hough transform circle detection algorithm and optimal fitting circle detection method.

Although CV-based methods are increasingly applied in the field of SHM, there still lacks a thorough comparison of different target tracking strategies in both time and frequency domain. To this end, the accuracy comparison of displacement monitoring based on target tracking strategies of template matching and Hough circle transform is carried out in this paper. In addition, the quadratic surface fitting method is introduced to achieve the subpixel level accuracy. The displacement result monitored by the laser transducer system is regarded as the ground truth, and the modal parameters identified from the monitored displacement time history are compared with those calculated from the finite element analysis (FEA). The outline of this paper is as below: Section 2 introduces the general procedure of CV-based methods for displacement monitoring and the basic principles of template matching and Hough circle transform; Section 3 introduces the subpixel localization refinement based on quadratic surface fitting; Section 4 presents the validation experiment of displacement monitoring of a cantilever plate with three artificial targets and compares the displacement time history curves monitored by CV-based methods and laser transducer, as well as the identified and calculated modal parameters; Section 5 concludes the work of this paper and looks forward to the prospects of the future research.

2. General Procedure of CV-Based Displacement Monitoring

The general procedure of CV-based displacement monitoring includes the camera calibration, the selection of the region of interest (ROI), the feature extraction, the target tracking, and the displacement calculation [21]. The camera calibration determines the projection relationship between the world coordinate and image coordinate by modeling the camera as a pinhole. The ROI selection fixes the regions containing the unique feature or texture in the structural surface [22]. The feature extraction extracts the specific features from ROIs, such as the artificial targets attached to the structure or the texture and image feature of the structural surface. The target tracking intends to track the extracted features in the ROIs and determine the coordinates of the features in different frames. Finally, the image displacement can be transformed into the actual displacement based on the camera calibration.

In the abovementioned steps, target tracking is the key [23]. Target tracking algorithm should be chosen according to the features to be extracted and the convenience of measurement. In this paper, two visual tracking strategies, namely template matching and Hough circle transform, are described and selected for CV-based displacement monitoring.

2.1. Camera Calibration. In general, the camera should be calibrated before the target tracking to determine the relationship between the world coordinate and the image coordinate. The projection geometry of the object and its image counterpart is presented in Figure 1 [21]. According to the dimension of the calibration object, camera calibration methods include the self-calibration [24], 1D line based calibration [25], 2D plane-based calibration [22], 2.5D coding target-based calibration [26], and 3D reference object based calibration [27]. In the scope of SHM, simplified methods are often used to find the mapping relation between image coordinates and world coordinates. This paper uses the scaling factor (SF) [21] to calculate the actual displacement.

In the plane determined by the camera and the direction of ROI movement, the SF can be estimated by equation (1) when there is an angle between the optical axis and the normal direction of ROI movement [10]. In the expression, θ is the intersection angle, f is the focal length of the camera, D is the distance between the object and the camera, x_A^i and x_B^i are the image coordinates of point A and B. $I_A^i = x_A^i/d_{\text{pixel}}$ and $I_B^i = x_B^i/d_{\text{pixel}}$, in which d_{pixel} is the pixel size (mm/pixel).

$$SF = \frac{1}{I_A^i - I_B^i} \left(\frac{Dx_A^i}{f \cos^2 \theta - x_A^i \cos \theta \sin \theta} - \frac{Dx_B^i}{f \cos^2 \theta - x_B^i \cos \theta \sin \theta} \right). \quad (1)$$

If the angle θ is small enough ($\sin \theta \approx 0$), equation (1) can be further simplified into the following:

$$SF \approx \frac{D}{f} d_{\text{pixel}}. \quad (2)$$

2.2. Feature Extraction and Visual Tracking. In CV-based SHM, ROI is often used for feature extraction and visual tracking [28]. In terms of template matching, the template image that includes the artificial target is chosen in advance as the feature to be tracked. However, the Hough circle transform detects the circle shape in ROI. Regarding the visual tracking, the template matching algorithm determines the target locations in the successive frames, while the Hough circle transform detects the centroid of the circles. Figure 2 shows the schematic of the artificial target used in this paper, with a simple texture on the target surface. The distinct difference between the target texture and the red background is a benefit for the template matching, while the circle outline is suitable for the Hough circle transform. Target tracking obtains the image locations of the targets in different frames, and the actual displacements of the targets can be calculated by multiplying the image displacement by the scaling factor SF.

2.2.1. Template Matching. The basic idea for the template matching algorithm is sliding the template image on the successive frames and finding the locations of the most similar subimage, as shown in Figure 3. First, a subimage

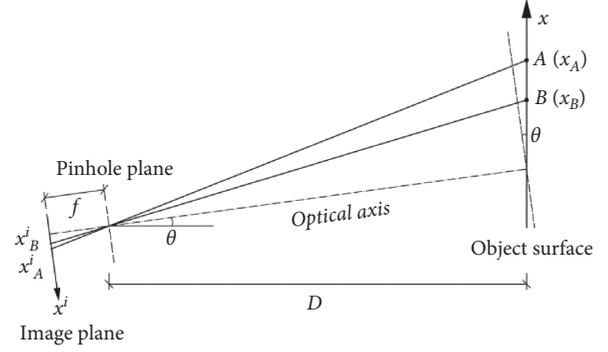


FIGURE 1: Determination of scaling factor.

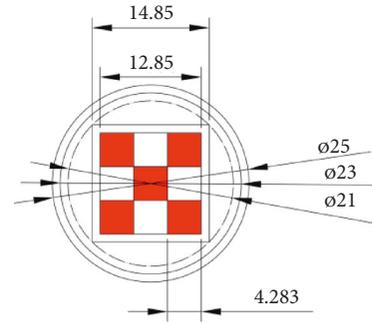


FIGURE 2: The schematic of the artificial target (mm).

($w \times h$ pixel) including the target is chosen as the template image from the first frame. Then, the similarity between the template image and the overlapped part is calculated when sliding the template image on the i -th frame ($W \times H$ pixel). After the similarity calculation, a $(W - w + 1) \times (H - h + 1)$ similarity matrix can be obtained. The extremum of the similarity matrix represents the best match according to different similarity measurement indicators and the target image location can be found by mapping the index of the extremum to the original frame.

In terms of similarity measurement, the following five indicators are often used [21]:

- (1) Sum of square difference:

$$\alpha_{\text{sq-diff}}(x, y) = \sum_{x', y'} (T(x', y') - I(x + x', y + y'))^2. \quad (3)$$

- (2) Normalized sum of square difference:

$$\alpha_{\text{nsq-diff}}(x, y) = \frac{\sum_{x', y'} (T(x', y') - I(x + x', y + y'))^2}{\sqrt{\sum_{x', y'} T(x', y')^2 \cdot \sum_{x', y'} I(x + x', y + y')^2}} \quad (4)$$

- (3) Normalized cross-correlation:

$$\alpha_{\text{ncorr}}(x, y) = \frac{\sum_{x', y'} (T(x', y') \cdot I(x + x', y + y'))}{\sqrt{\sum_{x', y'} T(x', y')^2 \cdot \sum_{x', y'} I(x + x', y + y')^2}} \quad (5)$$

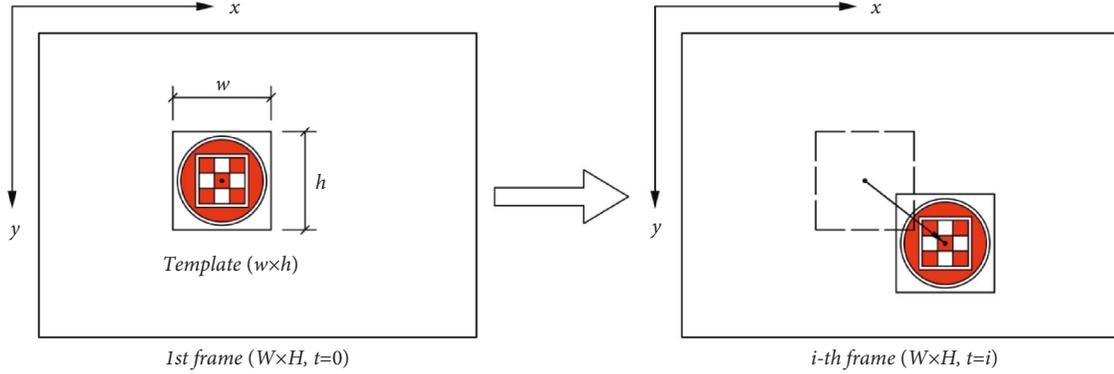


FIGURE 3: Basic principles of template matching.

(4) Correlation coefficient:

$$\alpha_{\text{ccoeff}}(x, y) = \sum_{x', y'} \left(T_I(x', y') \cdot I_I(x + x', y + y') \right). \quad (6)$$

(5) Normalized correlation coefficient:

$$\alpha_{\text{nccoeff}}(x, y) = \frac{\sum_{x', y'} \left(T_I(x', y') \cdot I_I(x + x', y + y') \right)}{\sqrt{\sum_{x', y'} T_I(x', y')^2 \cdot \sum_{x', y'} I_I(x + x', y + y')^2}}. \quad (7)$$

$T(x', y')$ is the gray value of the template image at (x', y') pixel; $I(x', y')$ is the gray value of the successive image at (x, y) pixel. In equations (6) and (7), $T_I(x', y')$ and $I_I(x + x', y + y')$ can be, respectively, expressed by the following:

$$T_I(x', y') = T(x', y') - \frac{1}{(w \cdot h)} \cdot \sum_{x'', y''} T(x'', y''),$$

$$I_I(x + x', y + y') = I(x + x', y + y') - \frac{1}{(w \cdot h)} \cdot \sum_{x'', y''} I(x + x'', y + y''). \quad (8)$$

In the above indicators, $\alpha_{\text{sq-diff}}$ and $\alpha_{\text{nsq-diff}}$ get the best match at the minimum, the rest of the indicators get the best match at the maximum [29]. $\alpha_{\text{sq-diff}}$ calculates the Euclidean distance between the template image and its overlapped counterpart as the similarity measurement, which is the most intuitive indicator and is easy to calculate. $\alpha_{\text{nsq-diff}}$ is the normalized version of $\alpha_{\text{sq-diff}}$, they all get the minimum when the subimage of the successive frame matches the template best. α_{nccorr} calculates the normalized correlation between the template and its overlapped part of the successive frame, the maximum indicates the best match. α_{ccoeff} removes the mean intensity of the template and the overlapped part in the successive image. α_{nccoeff} is the normalized version of α_{ccoeff} . All the above template matching methods have been implemented in the open software library OpenCV.

2.2.2. Hough Circle Transform. The Hough circle transform detects the circle outline of the artificial targets in the ROI of each frame by the following two steps: (1) finding the circle center and (2) determining the radius [30]. Because the center locates along the gradient direction of the edge, the edges of the target in the ROI are detected first. Then, the common intersection point of these gradient directions identifies the center (a, b) of a circle, as shown in Figure 4. Distance between the center and the edge point (x_i, y_i) can be calculated, and the most frequently occurred distance is taken as the radius according to the circle equation $(x - a)^2 + (y - b)^2 = r^2$. If any other centers are detected in an image, the radius can also be calculated in the same way. All the circles in the video frames can be tracked after the determination of their centers and radius.

In OpenCV, the Hough circle transform algorithm uses the Canny method to detect edges [31], and the grayscale gradient threshold must be set by the user in advance. The common intersection point of the gradient direction will be chosen as the actual center if the number of different gradient directions for the same center exceeds a predetermined threshold. To avoid detecting error centers in the neighborhood, the minimum distance between centers is also determined in advance. In the step of radius calculation, if the distance between the center and the edge point falls in the predetermined interval, the most frequently occurred distance is the radius. The additional constraints for center detection and radius calculation in OpenCV make the Hough circle transform preciser for target tracking.

3. Subpixel Localization Refinement

After the template image is extracted in the i -th image frame through the template matching algorithm by different similarity measurement indicators, the calculated horizontal and vertical distances toward the object in the first image frame will be integral numbers of a pixel. However, this assumption is not always true in real situation, which inevitably brings measurement error. In other words, the distance between the shifted template and the object image can be regarded as noninteger numbers of a pixel. In order to improve the accuracy of CV-based monitoring methods,

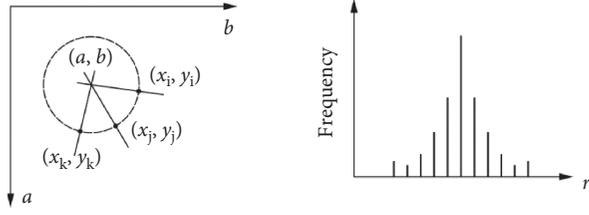


FIGURE 4: Basic principles of Hough circle transform.

quadratic surface fitting is adopted for subpixel localization refinement:

$$S(x, y) = ax^2 + bxy + cy^2 + dx + ey + f. \quad (9)$$

Considering the fitting accuracy and the complexity and quantity of computation, the point with the maximum similarity and the surrounding eight points are used to realize the quadratic surface fitting, namely:

$$\begin{bmatrix} S(-1, 1) & S(0, 1) & S(1, 1) \\ S(-1, 0) & S(0, 0) & S(1, 0) \\ S(-1, -1) & S(0, -1) & S(1, -1) \end{bmatrix} = \begin{bmatrix} \alpha(x_i - 1, y_j + 1) & \alpha(x_i, y_j + 1) & \alpha(x_i + 1, y_j + 1) \\ \alpha(x_i - 1, y_j) & \alpha(x_i, y_j) & \alpha(x_i + 1, y_j) \\ \alpha(x_i - 1, y_j - 1) & \alpha(x_i, y_j - 1) & \alpha(x_i + 1, y_j - 1) \end{bmatrix}. \quad (10)$$

where α is the similarity matrix and (x_i, y_j) are the coordinates of the maximum value in the similarity matrix.

Equation (11) can be rewritten into the following matrix form:

$$\begin{bmatrix} 1 & -1 & 1 & -1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 & -1 & 1 \\ 1 & -1 & 1 & 1 & -1 & 1 \end{bmatrix} \begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix} = \begin{bmatrix} \alpha(x_i - 1, y_j + 1) \\ \alpha(x_i, y_j + 1) \\ \alpha(x_i + 1, y_j + 1) \\ \alpha(x_i - 1, y_j) \\ \alpha(x_i, y_j) \\ \alpha(x_i + 1, y_j) \\ \alpha(x_i - 1, y_j - 1) \\ \alpha(x_i, y_j - 1) \\ \alpha(x_i + 1, y_j - 1) \end{bmatrix}. \quad (11)$$

The coefficients $a \sim f$ can be estimated through a pseudo-inverse computation, namely:

$$\begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix} = \begin{bmatrix} 1/6 & -1/3 & 1/6 & 1/6 & -1/3 & 1/6 & 1/6 & -1/3 & 1/6 \\ -1/4 & 0 & 1/4 & 0 & 0 & 0 & 1/4 & 0 & -1/4 \\ 1/6 & 1/6 & 1/6 & -1/3 & -1/3 & -1/3 & 1/6 & 1/6 & 1/6 \\ -1/6 & 0 & 1/6 & -1/6 & 0 & 1/6 & -1/6 & 0 & 1/6 \\ 1/6 & 1/6 & 1/6 & 0 & 0 & 0 & -1/6 & -1/6 & -1/6 \\ -1/9 & 2/9 & -1/9 & 2/9 & 5/9 & 2/9 & -1/9 & 2/9 & -1/9 \end{bmatrix} \begin{bmatrix} \alpha(x_i - 1, y_j + 1) \\ \alpha(x_i, y_j + 1) \\ \alpha(x_i + 1, y_j + 1) \\ \alpha(x_i - 1, y_j) \\ \alpha(x_i, y_j) \\ \alpha(x_i + 1, y_j) \\ \alpha(x_i - 1, y_j - 1) \\ \alpha(x_i, y_j - 1) \\ \alpha(x_i + 1, y_j - 1) \end{bmatrix}. \quad (12)$$

The extreme point of the quadratic surface equation can be then calculated when the derivatives with respect to x and y are both zero:

$$\begin{cases} \frac{\partial S(x, y)}{\partial x} = 2ax + by + d = 0, \\ \frac{\partial S(x, y)}{\partial y} = bx + 2cy + e = 0. \end{cases} \quad (13)$$

Finally, the refined location of the extreme value can be solved as follows:

$$\begin{cases} x_s = \frac{2c d - be}{b^2 - 4ac}, \\ y_s = \frac{2ae - b d}{b^2 - 4ac}. \end{cases} \quad (14)$$

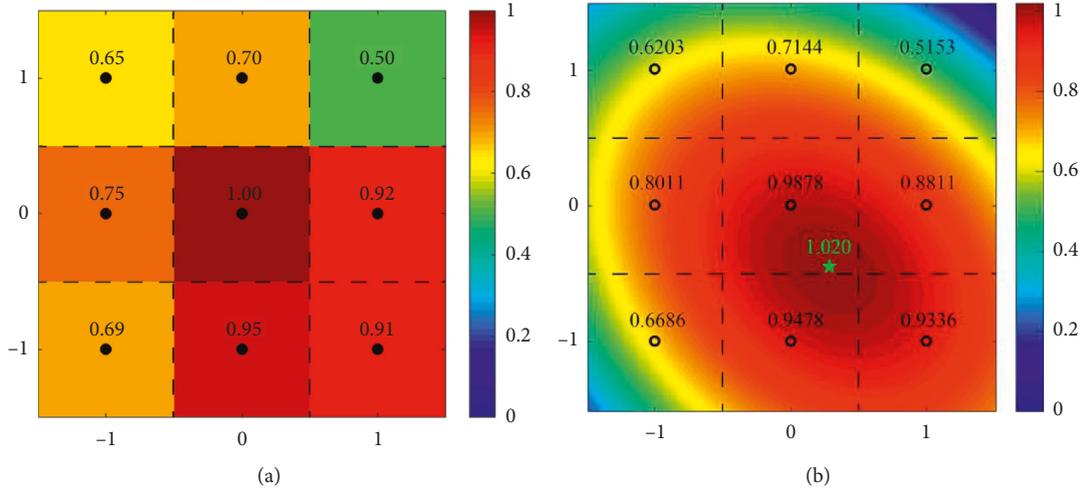


FIGURE 5: Example of the quadratic surface fitting refinement. (a) Nine normalized cross-correlation values for curve fitting. (b) The fitted quadratic surface.



FIGURE 6: Layout of experiment setup.

Since the values of the similarity matrix around the maximum are quite close to each other, a normalized operation should be performed first. In order to verify the subpixel localization refinement graphically, the original best-matching pixel point and the neighboring eight points are shown in Figure 5(a), in which the similarity indicator α_{ncorr} is used. The calculated quadratic surface is presented in Figure 5(b). The new extreme point (0.2798, -0.4549) marked with a solid green star can be considered as a better estimation of the best-matching location, and the above method can be easily applied to template matching and edge detection.

4. Experimental Setup and Monitoring Results

4.1. Experimental Setup. An experiment is designed to validate the CV-based displacement monitoring method, as shown in Figure 6. Three artificial targets are fixed on the top of an aluminum alloy cantilever plate. The left end of the cantilever plate is fixed. The length of the cantilever plate is

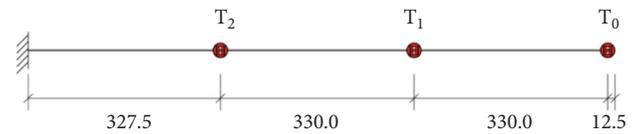


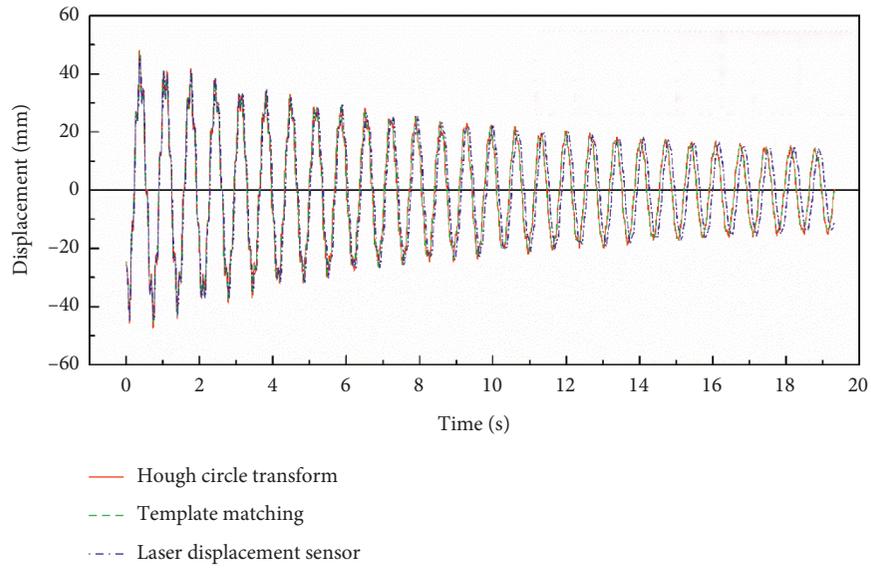
FIGURE 7: Locations of targets.

TABLE 1: Main experimental device.

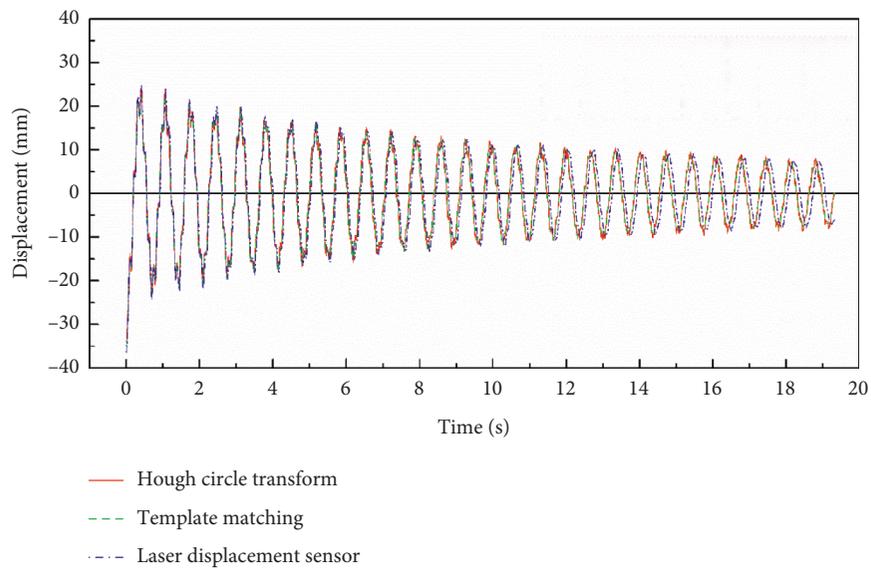
Main device	Model	Related parameters
Camera	iPhone 7 plus	FPS 60 Hz, Resolution 1920 × 1080
Laser displacement sensor	Panasonic HG-C1200	Sampling frequency 55.86 Hz
Data acquisition device	USB DAQ-280G	Resolution 24 bit

1 m, the size of the cross-section is (2 × 50) mm, the density is 2810 kg/m³, and the elastic module is 64.3 GPa. The artificial targets are three metal circle plates with a diameter of 25 mm, as shown in Figure 2. The targets are fixed on the top of the cantilever and their detailed locations are shown in Figure 7. A smartphone is used to collect the motion video in the horizontal direction, and the distance from the camera to the beam is 1.5 m. The laser measurement system is used as the reference, including three laser displacement sensors and a data acquisition device. The information of the main experimental instruments is listed in Table 1.

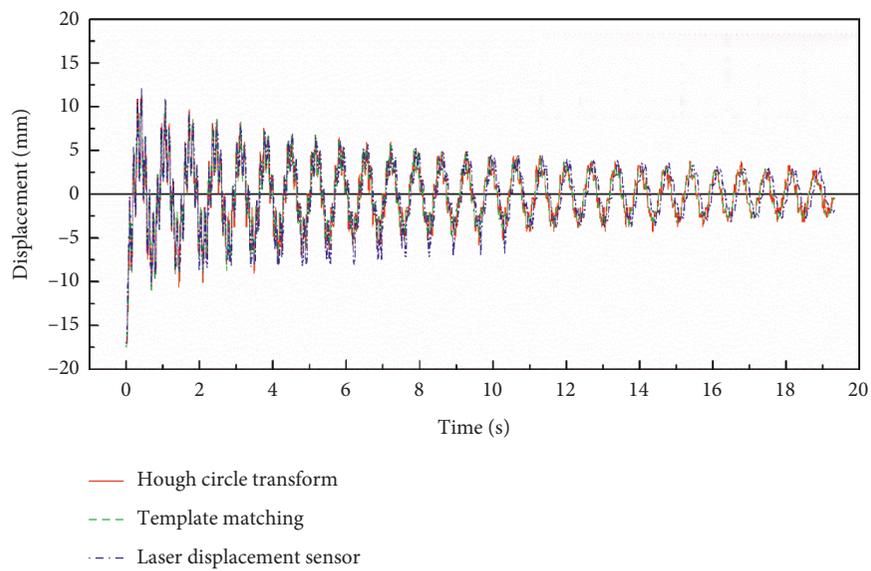
The free vibration of the cantilever plate is excited by the initial displacement so as to identify relatively high-order vibration modes. After releasing the initial displacements, the camera collects the vibration video and the measurement system starts to monitor the displacement of three targets simultaneously. The collected video is then analyzed by template matching algorithm and Hough circle transform algorithm, respectively, to obtain the target displacement and identify the modal parameters.



(a)

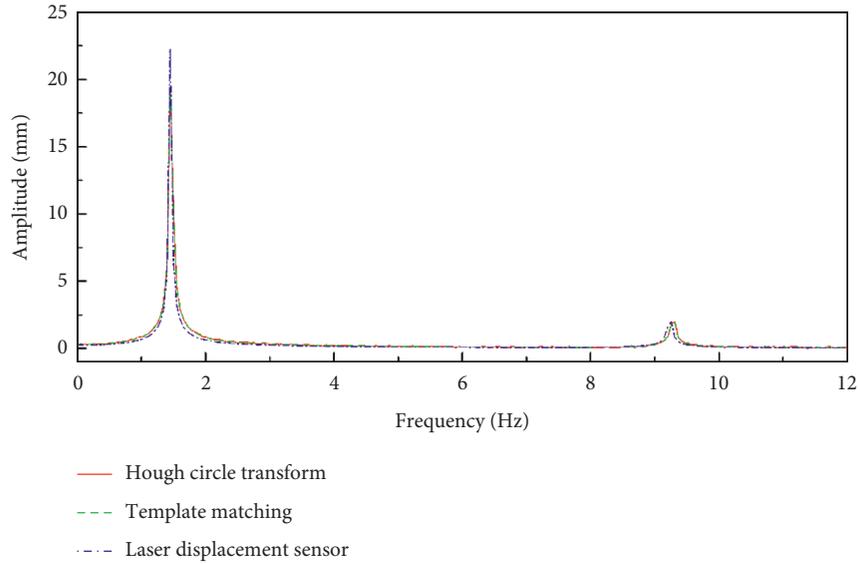


(b)

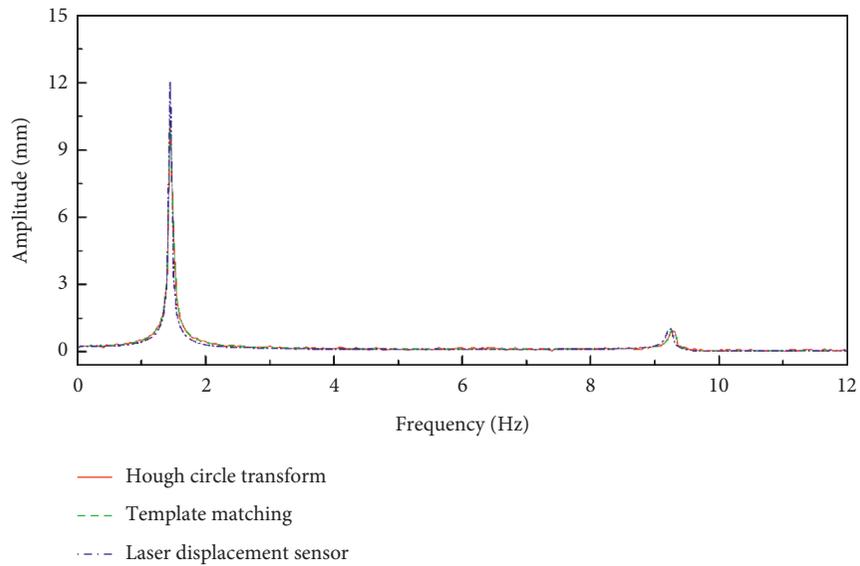


(c)

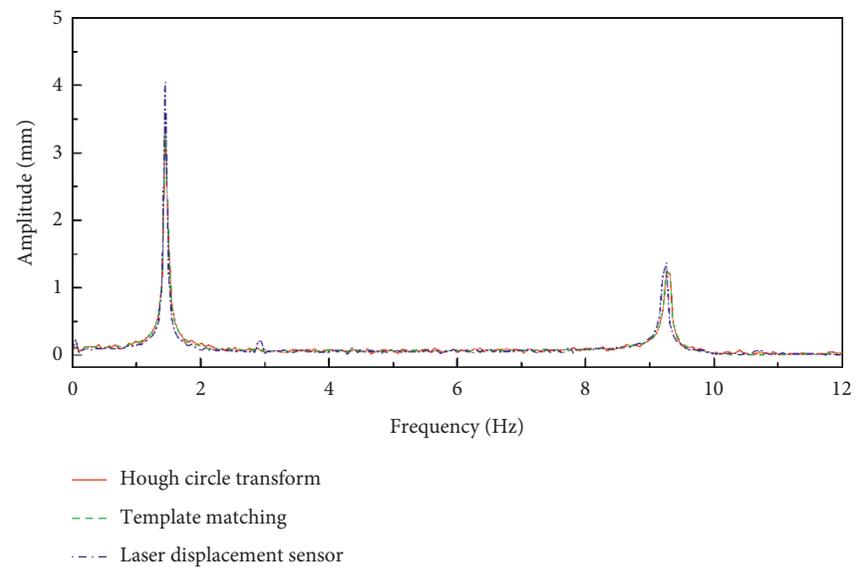
FIGURE 8: Displacement monitoring results of target: (a) T_0 , (b) T_1 and (c) T_2 .



(a)



(b)



(c)

FIGURE 9: Displacement amplitude spectrum of target: (a) T_0 , (b) T_1 and (c) T_2 .

TABLE 2: Comparison of identified and calculated natural frequencies.

Method	1 st natural frequency (Hz)	2 nd natural frequency (Hz)
Hough circle transform	1.449	9.278
Template matching	1.449	9.278
Laser displacement sensor	1.448	9.259
FEM results	1.468	9.220

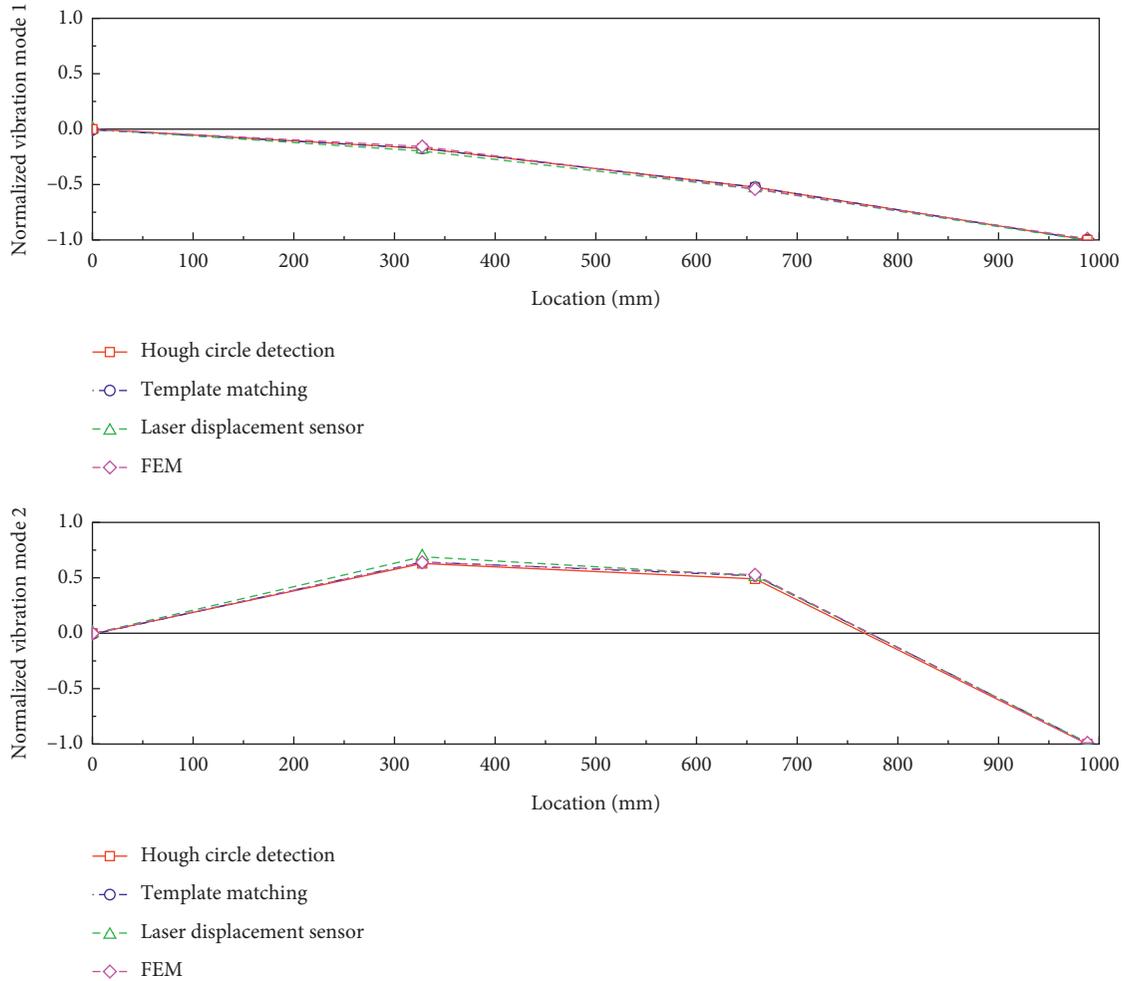


FIGURE 10: Comparison of the first two modes identified and calculated by different methods.

4.2. Displacement Monitoring Results. In the procedure of CV-based displacement monitoring methods, the precision of target tracking will directly influence the final monitoring result. When template matching is used to track the targets, different similarity measurement indicators may obtain different image locations of the targets for the same frame. In this experiment, the five different indicators expressed in equations (3)–(7) are used for the template matching. In this context, the mismatched frame means that the matched image range does not contain the complete target. When the template matching $\alpha_{sq-diff}$ is applied, three targets can be correctly detected in all the frames. For the template matching α_{ccoeff} , 20 error detections occur when tracking the target T_0 . Regarding the template matching methods with the normalized indicators, they all match the targets but

spend more time. Therefore, $\alpha_{sq-diff}$ is chosen as the similarity measurement indicator for template matching in this paper.

In terms of the Hough circle transform, some predetermined parameters will affect the results of circle detection. Thus, the parameters should be appropriately selected to guarantee that all the frames can be handled uniformly. After checking the tracking results, there are also some mismatched frames, which can be processed again by adjusting the parameter to make the tracking results more precise.

Since the actual size of the targets is known, the pixel size (mm/pixel) can be obtained after the targets are extracted in the frame, and then the displacement in the image coordinate (pixel) can be converted into that in the world

coordinate (mm) through scaling factor expressed by equation (2). The displacement responses of the three targets are monitored and plotted in Figure 8. Due to the difference between the sampling frequency of the camera and the laser sensor, there exists a slight time lag between the CV-based method and the laser measurement system with the increase in time. However, the displacement amplitude of these three targets measured by the CV-based methods and the laser sensors are consistent.

4.3. System Identification. The displacement time history of the target T_i is denoted by $x_i(t)$, which will be used to identify the structural mode. Based on the fast Fourier transform (FFT), $x_i(t)$ is transformed into a Fourier spectrum $X_i(f)$ where f is the frequency (Hz). With further mathematic operations, the amplitude spectrum $|X_i(f)|$ and the imaginary spectrum $\text{Im}[X_i(f)]$ can be obtained. Since the considered structure is simple, the natural frequencies can be obtained by the peak picking method and the vertical coordinates of the peaks are proportional to the corresponding modal shape. The normalized mode shape can be obtained after dividing by the modal displacement of T_0 . For the n -th order modal shape, the modal displacement of the i -th measurement point $S_i(f_n)$ can be determined according to the imaginary part spectrum:

$$S_i(f_n) = \begin{cases} |X_i(f_n)| & \text{Im}[X_i(f_n)] > 0 \\ 0 & \text{Im}[X_i(f_n)] = 0 \\ -|X_i(f_n)| & \text{Im}[X_i(f_n)] < 0. \end{cases} \quad (15)$$

Figure 9 shows the amplitude spectra of the monitoring results for three targets. In the frequency domain, the results of the CV-based methods agree well with those obtained by laser displacement sensors.

In addition, the theoretical modal frequencies are also calculated with the finite element method (FEM) and compared with those identified by measuring results, as listed in Table 2. The first two modal shapes identified and calculated by different methods are plotted in Figure 10.

5. Conclusions

In this study, the target tracking methods such as template matching and Hough circle transform used for CV-based SHM are introduced. Subpixel localization refinement based on quadratic surface fitting is introduced to improve the accuracy. The laboratory experiment is conducted to investigate the accuracy and robustness of the two tracking strategies. FFT is used to identify the structural modes through the peak picking method, and the identified experimental modes are compared with the theoretical modes calculated by FEM. The results of target tracking demonstrate that the similarity measurement indicator $\alpha_{\text{sq-diff}}$ is robust and time-saving for template matching. Although wrong detections may occur in a small number of frames, the Hough circle transform is still a good choice for tracking circle shape. Through comparison with the displacement response monitored by the laser measurement system,

template matching and Hough circle transform are capable of satisfying displacement measurement. In terms of system identification, the vibration modes identified and calculated by different methods are consistent. The experiment verifies that CV-based methods are robust and reliable in both time and frequency domain and could be applied for displacement monitoring and system identification.

Although good results can be obtained by the proposed method in an indoor validation experiment with a stable luminous environment, the camera of high quality and other algorithms eliminating negative effects of environments are required to monitor the displacement of targets in actual projects. In addition, the ROIs of image frames are chosen by manual in this paper, which will be complicated for complex structures with a large number of targets. Future research will focus on the self-adaptive selection of ROIs through the artificial intelligence method so as to extend the application scopes of the CV-based SHM method.

Data Availability

The data are available upon request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper was supported by the Open Funds of Xiamen Engineering Technology Center for Intelligent Maintenance of Infrastructures (grant number TCIMI201808).

References

- [1] Y. Pan and L. M. Zhang, "Roles of artificial intelligence in construction engineering and management: A critical review and future trends," *Automation in Construction*, vol. 122, p. 103517, 2021.
- [2] Y. Q. Bao and H. Li, "Artificial Intelligence for civil engineering," *China Civil Engineering Journal*, vol. 52, no. 5, pp. 1–11, 2019, (in Chinese).
- [3] X. W. Ye and C. Z. Dong, "Review of Computer Vision-based structural displacement monitoring," *China Journal of Highway and Transport*, vol. 32, no. 11, pp. 21–39, 2019, (in Chinese).
- [4] D. M. Feng and M. Q. Feng, "Computer vision for SHM of civil infrastructure: from dynamic response measurement to damage detection - A review," *Engineering Structures*, vol. 156, pp. 105–117, 2018.
- [5] H. W. Ma and Z. H. Nie, "Recent Advances and review of bridge safety monitoring," *Mechanics in Engineering*, vol. 37, no. 2, pp. 161–170, 2015, 181. (in Chinese).
- [6] Y. C. Tang, Y. F. Lin, X. Y. Huang, M. Yao, Z. Huang, and X. Zou, "Grand challenges of machine-vision technology in civil structural health monitoring," *Artificial Intelligence Evolution*, vol. 1, no. 1, pp. 8–16, 2020.
- [7] Y. Xu, J. Brownjohn, and D. Kong, "A noncontact vision-based system for multipoint displacement monitoring in a cable-stayed footbridge," *Structural Control and Health Monitoring*, vol. 25, no. 5, p. e2155, 2018.

- [8] D. M. Feng, M. Q. Feng, E. Ozer, and Y. Fukuda, "A vision-based sensor for noncontact structural displacement measurement," *Sensors*, vol. 15, no. 7, pp. 16557–16575, 2015.
- [9] J. G. Chen, T. M. Adams, H. Sun, E. S. Bell, and O. Buyukozturk, "Camera-based vibration measurement of the world war I memorial bridge in portsmouth, New Hampshire," *Journal of Structural Engineering*, vol. 144, no. 11, p. 04018207, 2018.
- [10] Y. D. Tian, J. Zhang, and S. S. Yu, "Rapid impact testing and system identification of footbridges using particle image velocimetry," *Computer-Aided Civil and Infrastructure Engineering*, vol. 34, no. 2, pp. 130–145, 2019.
- [11] S. W. Kim and N. S. Kim, "Dynamic characteristics of suspension bridge hanger cables using digital image processing," *NDT & e International*, vol. 59, pp. 25–33, 2013.
- [12] C. Z. Dong, X. W. Ye, and T. Jin, "Identification of structural dynamic characteristics based on machine vision technology," *Measurement*, vol. 126, pp. 405–416, 2018.
- [13] H. Yoon, H. Elanwar, H. Choi, M. Golparvar-Fard, and B. F. Spencer, "Target-free approach for vision-based structural system identification using consumer-grade cameras," *Structural Control and Health Monitoring*, vol. 23, no. 12, pp. 1405–1416, 2016.
- [14] T. Khuc and F. N. Catbas, "Completely contactless structural health monitoring of real-life structures using cameras and computer vision," *Structural Control and Health Monitoring*, vol. 24, no. 1, p. e1852, 2017.
- [15] A. Khaloo and D. Lattanzi, "Pixel-wise structural motion tracking from rectified repurposed videos," *Structural Control and Health Monitoring*, vol. 24, no. 11, p. e2009, 2017.
- [16] V. Hoskere, J. W. Park, H. Yoon, and B. F. Spencer, "Vision-based modal survey of civil infrastructure using unmanned aerial vehicles," *Journal of Structural Engineering*, vol. 145, no. 7, p. 04019062, 2019.
- [17] C. Z. Dong, O. Celik, F. N. Catbas, E. J. O'Brien, and S. Taylor, "Structural displacement monitoring using deep learning-based full field optical flow methods," *Structure and Infrastructure Engineering*, vol. 16, no. 1, pp. 51–71, 2020.
- [18] J. P. Han, Y. H. Zhang, and H. Y. Zhang, "Displacement measurement of shaking table test structure model based on computer vision," *Earthquake Engineering and Engineering Dynamics*, vol. 39, no. 04, pp. 22–29, 2019, (in Chinese).
- [19] B. Shan, S. Zheng, and J. Ou, "Free vibration monitoring experiment of a stayed-cable model based on stereovision," *Measurement*, vol. 76, pp. 228–239, 2015.
- [20] S. Chen, G. X. Chen, and X. J. Han, "Development for Vision-based displacement test method," *Journal of Vibration and Shock*, vol. 34, no. 18, pp. 73–78, 2015, (in Chinese).
- [21] C. Z. Dong and F. N. Catbas, "A review of Computer Vision-based structural health monitoring at local and Global levels," *Structural Health Monitoring*, vol. 20, no. 2, pp. 692–743, 2021.
- [22] Z. Y. Zhang, "A Flexible new technique for Camera Calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [23] Y. Xu and J. M. W. Brownjohn, "Review of machine-vision based methodologies for displacement measurement in civil structures," *Journal of Civil Structural Health Monitoring*, vol. 8, no. 1, pp. 91–110, 2018.
- [24] Z. Zhang, R. J. Zhao, E. H. Liu, K. Yan, and Y. Ma, "A single-image linear calibration method for camera," *Measurement*, vol. 130, pp. 298–305, 2018.
- [25] Z. Y. Zhang, "Camera calibration with one-dimensional objects," *Lecture Notes in Computer Science*, vol. 26, no. 7, pp. 161–174, 2002.
- [26] J. Zhang, H. Yu, H. Deng, Z. Chai, M. Ma, and X. Zhong, "A robust and rapid Camera Calibration method by one Captured image," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, pp. 4112–4121, 2019.
- [27] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987.
- [28] J. Lee, K.-C. Lee, S. Cho, and S.-H. Sim, "Computer Vision-based structural displacement measurement robust to Light-Induced image Degradation for in-Service Bridges," *Sensors*, vol. 17, no. 10, p. 2317, 2017.
- [29] OpenCV modules, "Template Matching," https://docs.opencv.org/3.4/d4/dc6/tutorial_py_template_matching.html.
- [30] H. K. Yuen, J. Princen, J. Illingworth, and J. Kittler, "Comparative study of Hough Transform methods for circle finding," *Image and Vision Computing*, vol. 8, no. 1, pp. 71–77, 1990.
- [31] OpenCV modules, "Hough Circle Transform," https://docs.opencv.org/master/d4/d70/tutorial_hough_circle.html.