

PAPER • OPEN ACCESS

## Cost-effective, vision-based multi-target tracking approach for structural health monitoring

To cite this article: Tong Wu *et al* 2021 *Meas. Sci. Technol.* **32** 125116

View the [article online](#) for updates and enhancements.

You may also like

- [Binocular vision-based 3D method for detecting high dynamic and wide-range contouring errors of CNC machine tools](#)  
Xiao Li, Wei Liu, Yi Pan *et al.*
- [Vision based monitoring and characterisation of combustion flames](#)  
G Lu, G Gilibert and Y Yan
- [The bounds on tracking performance utilising a laser-based linear and angular sensing and measurement methodology for micro/nano manipulation](#)  
Leon Clark, Bijan Shirinzadeh, Yanling Tian *et al.*

# Cost-effective, vision-based multi-target tracking approach for structural health monitoring

Tong Wu<sup>1</sup> , Liang Tang<sup>1,\*</sup>, Shuai Shao<sup>1</sup> , Xiang-Yu Zhang<sup>1</sup>, Yi-Jun Liu<sup>1</sup> and Zhi-Xiang Zhou<sup>2</sup>

<sup>1</sup> College of Civil Engineering, Chongqing Jiaotong University, Chongqing, People's Republic of China

<sup>2</sup> College of Civil and Transportation Engineering, Shenzhen University, Shenzhen, People's Republic of China

E-mail: [tangliang@cqjtu.edu.cn](mailto:tangliang@cqjtu.edu.cn)

Received 12 July 2021, revised 26 August 2021

Accepted for publication 9 September 2021

Published 27 September 2021



CrossMark

## Abstract

The displacement response of structures is an important parameter in structural health monitoring (SHM). Displacement responses can be applied in both structural performance monitoring and structural dynamic characteristics monitoring. To overcome the shortcomings of traditional contact sensors, a vision-based multi-point structural displacement measurement system equipped with an inexpensive surveillance camera and a consumer camera was developed herein. In addition, to reduce the computing time of target tracking, an improved region-matching algorithm based on the prior knowledge of structural deformation was proposed. Numerical results revealed that the improved region-matching algorithm could save computing time without reducing location accuracy. Moreover, static and dynamic loading tests were conducted on a scale model of a suspension bridge to validate the effectiveness of the proposed vision-based measurement system. Displacement responses and modal parameters obtained from the vision-based measurement system were compared with those of traditional contact sensors, and a satisfactory consistency was obtained. Hence, the proposed vision-based measurement system could be a cost-effective alternative to conventional displacement sensors and accelerometers for SHM.

Keywords: structural health monitoring, vision-based method, displacement measurement, modal parameter identification, adaptive hierarchical localization algorithm

(Some figures may appear in colour only in the online journal)

\* Author to whom any correspondence should be addressed.



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

## 1. Introduction

Bridge structures are subjected to various external loads, such as vehicles, wind, and earthquakes. Structural health monitoring (SHM) is used to monitor and evaluate the performance of bridges, find out potential safety troubles in time, and take corresponding maintenance and reinforcement [1, 2]. The displacement response of structures under loads is an important parameter for SHM, and it reflects the overall stiffness of a structure.

Traditional contact displacement sensors, such as dial indicators and linear variable differential transformers (LVDTs), are generally installed on the bottom face of bridges. Although contact displacement sensors can yield accurate measurement results, brackets are often installed under bridges to obtain relatively fixed points of these structures, and wiring is required for data transmission and instrument power supply; hence, the actual operation is very complicated. The total station and level can only measure static displacements of the structure without attaining real-time displacement responses under different loads [3–5].

To overcome the shortcomings of traditional contact sensors, different non-contact measurement methods, such as GPS, laser Doppler vibrometry (LDV), radar interferometry, and vision-based method, have been developed. The GPS technology can measure both static and dynamic responses with a maximum measurement distance of up to 30 km. However, it is mainly used in long-span bridges and high-rise structures because of its low measurement accuracy and low sampling frequency [6–8]. LDV has a high measurement accuracy; however, it can only measure structural displacements in short distances. Moreover, the cost of LDV is high; thus, it is very difficult to realize multi-point synchronous displacement measurements [9–11]. Radar interferometry can achieve remote measurements with a good resolution; however, it requires reflective surfaces mounted on structures [12]. With the continuous development of computer vision technologies and image acquisition equipment, structural displacement monitoring methods based on computer vision have attracted great attention in civil engineering because they can be applied for large-range structural monitoring with high accuracy [12]. In these methods, videos of monitoring structures are captured by video cameras and motion trajectories of measuring points are calculated by target tracking. Subsequently, displacements of monitoring structures are determined by a scaling factor of image coordinates to physical coordinates. Cameras are generally fixed far away from structures, eliminating the requirement of fixed support points used in contact displacement monitoring methods. In addition, as the field of view (FOV) of cameras can cover multiple measurement areas on structures, it is easy to realize multi-point measurements at a low cost.

The advantages of vision-based methods have been verified in numerous applications [13–20]. For example, Busca *et al* [13] measured the vertical displacement of a bridge by tracking a target plate fixed on the bridge using a vision sensor.

A square optical target was used for tracking, and dynamic displacements of multiple points on the bridge were simultaneously measured. Ribeiro *et al* [15] measured the dynamic displacement of a railway bridge based on the random sampling consistency algorithm. Based on the robust direction code matching algorithm, a vision sensor system for real-time displacement measurement was developed by tracking a natural target on the bridge. Feng *et al* [16] proposed an advanced template matching algorithm for real-time displacement extraction from video images. Bhowmick *et al* [17] measured full-field displacement responses of continuous vibration edges of structural members by calculating horizontal and vertical components of the Lagrangian displacement of each pixel in each frame edge with sub-pixel accuracy.

Optical targets, such as LED lamps and laser spots, are often tracked to obtain high precision measurement results even in a dark environment. Dong *et al* [21] used LED lamps and black spots as tracking targets for structural dynamic displacement measurements of a simply-supported rectangular steel beam. Oh *et al* [22] measured the vibration of a three-story shear frame by tracking multi-LED targets. Ye *et al* [23] compared three types of image processing algorithms for structural dynamic displacement measurements and measured the displacement influence lines of an arch bridge by tracking optical targets. Wahbeh *et al* [24] developed a target equipped with low-power LEDs and measured the displacement of a long-span bridge by the proposed vision-based system. Park *et al* [25] used a motion capture system equipped with colorful LED lamps to measure three-dimensional displacements of a three-story frame. Maksymenko *et al* [26] measured the displacement of a concrete beam by tracking a laser spot mounted on the structure. Zhang *et al* [27] proposed a new method based on laser projection and image processing technologies for monitoring medium-small span bridges. Miguel *et al* [28] developed a novel sensing device equipped with a laser beam, a video camera, and LED lights to monitor displacements and rotations of bridges and structures. Tian *et al* [29] conducted field, remote, and multi-point deflection measurements of the Wuhan Yangtze River Bridge by tracking six LED targets mounted on the structure.

In the present work, to improve efficiency and reduce the cost, a vision-based multi-point structural displacement measurement system equipped with an inexpensive surveillance camera and a consumer camera was designed and an improved template matching algorithm was proposed. Both static and dynamic displacements were measured, and modal parameters were identified in a non-contact, cost-effective, and time-saving way. Static and dynamic loading tests were conducted on a 1/30 scale model of the Taohuayu Yellow River Highway suspension bridge to validate the effectiveness of the proposed vision-based measurement system. Displacement results obtained from the vision-based measurement system and LVDTs as well as modal parameter results obtained from the vision-based measurement system and accelerometers were compared.

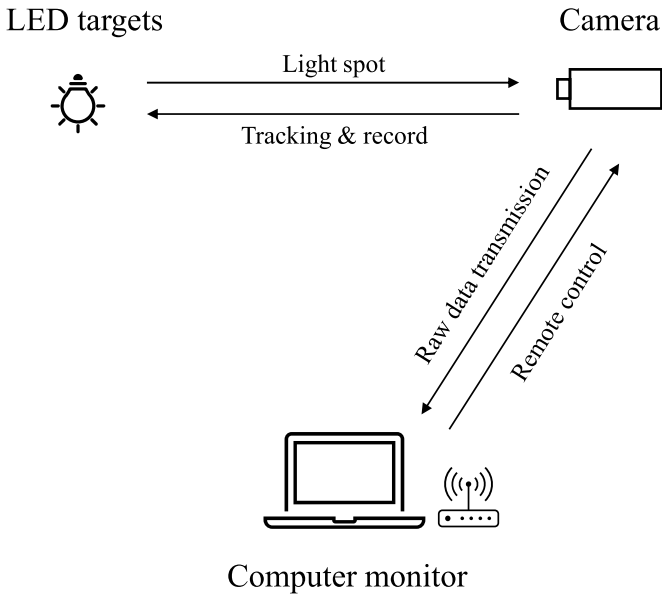


Figure 1. Construction of vision-based measurement system.

## 2. Design and principle of vision-based measurement system

### 2.1. Construction of vision-based measurement system

A vision-based measurement system is generally composed of an optical target consisting of multiple LED lamps, a camera, and a computer (figure 1). Multiple LED lamps are fixed laterally along the longitudinal direction of a bridge, and the camera is set on one side of the bridge. By adjusting the position of the camera, all measuring points are placed in its FOV. The position of the optical target changes with the deformation of the bridge.

The bridge vibration video is captured by the camera and stored in the computer. The video is then converted into grayscale images frame by frame. LED lamps are detected as multi-light spots in grayscale images. The displacement of a measuring point is calculated by tracking the change in the center coordinates of a light spot.

All light spots in an image are first identified and marked, and the template matching algorithm is performed for each spot.

In figure 2, the single spot number is  $T_i^j$ , where  $j$  is the number of measuring points in a single image and  $i$  is the frame number of the image in the video. The center coordinates of target  $T_1^j$  is  $(x_1^j, y_1^j)$ , whereas the center coordinates of the same target  $T_i^j$  in the  $i$ th image is  $(x_i^j, y_i^j)$ .

Hence, the pixel displacement of the target can be calculated as

$$\begin{cases} \Delta x = x_i^j - x_1^j \\ \Delta y = y_i^j - y_1^j \end{cases} \quad (1)$$

For most bridges, especially small and medium-sized bridges, on the one hand, it can assume that there are no transverse deformation and longitudinal deformation since the vertical deformation is much larger than the deformation of other two direction. On the other hand, the vertical displacement under design load is in the level of millimeter. Therefore, only the vertical position of LED lamps changes at a small scale. Considering  $\Delta x = 0$ , the pixel displacement of target  $T_i^j$  can be simplified as

$$\Delta y = y_i^j - y_1^j \quad (2)$$

In order to obtain structural displacements from pixel displacements, a scaling factor is used to establish the relationship between pixel coordinates and physical coordinates [16]. In figure 3, when the optical axis of the camera is perpendicular to the object surface, the scaling factor can be calculated as

$$SF = \frac{H}{h} \quad (3)$$

or

$$SF = \frac{Z}{f} d_{\text{pixel}} \quad (4)$$

where  $H$  is the physical length on the object plane,  $h$  is the corresponding pixel length on the image plane,  $d_{\text{pixel}}$  is the pixel size ( $\text{mm pixel}^{-1}$ ),  $Z$  is the distance between the camera and the object, and  $f$  is the focal length.

The actual displacement can be calculated as

$$\Delta Y = SF \times \Delta y \quad (5)$$

where SF is the scaling factor between image and physical dimensions.

### 2.2. Improved template matching algorithm based on leaping computation

The traditional template matching algorithm searches a template in an image based on the global correlation calculation method, and its criterion for judgment is the normalized correlation coefficient  $N(i, j)$ .

$$N(i, j) = \frac{\sum_{m=1}^M \sum_{n=1}^n (M_{i,j}(m, n) - \bar{M}_{i,j})(T(m, n) - \bar{T})}{\sqrt{\sum_{m=1}^M \sum_{n=1}^n (M_{i,j}(m, n) - \bar{M}_{i,j})^2} \sqrt{\sum_{m=1}^M \sum_{n=1}^n (T(m, n) - \bar{T})^2}} \quad (6)$$

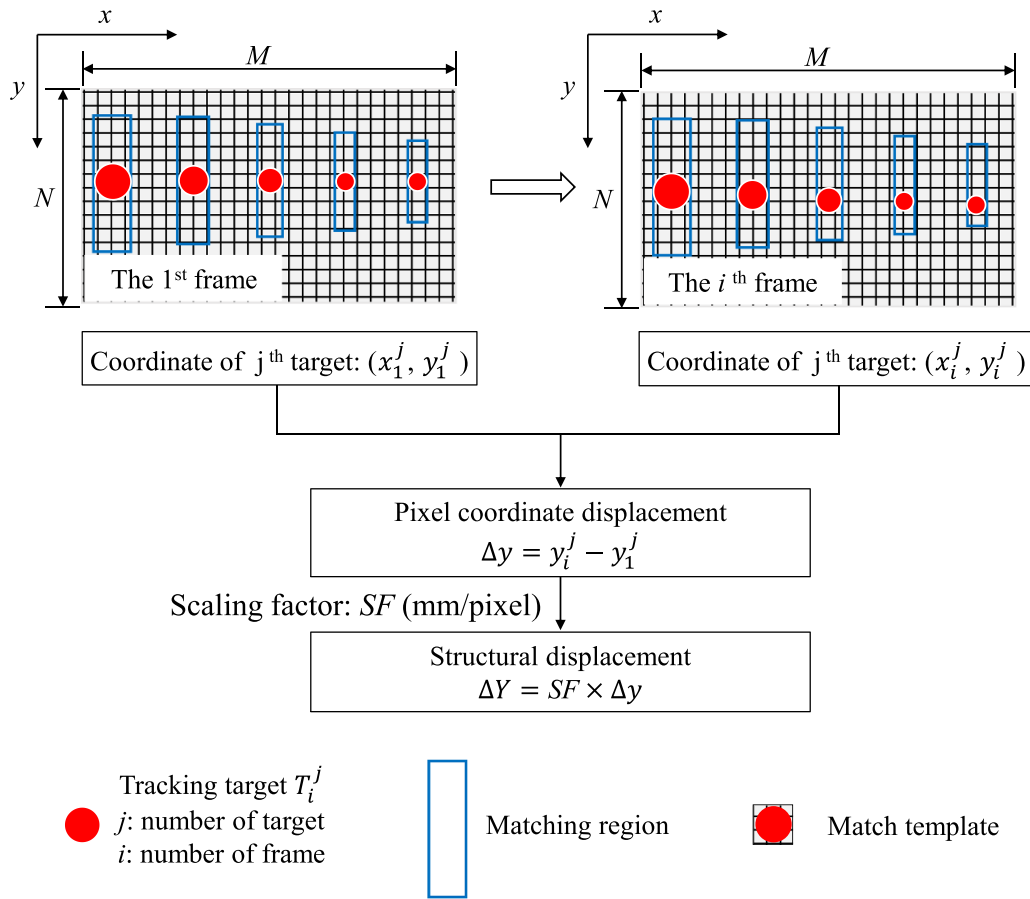


Figure 2. Multi-point displacement measurement process.

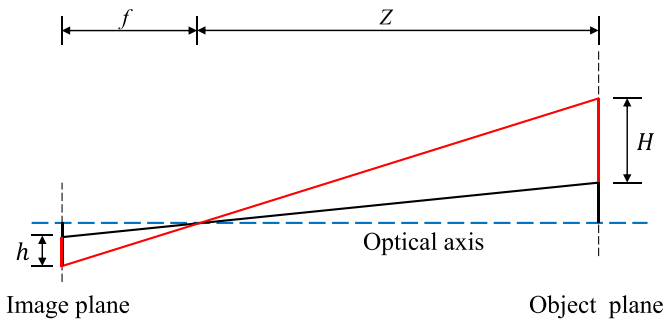


Figure 3. Scaling factor calculation method.

where  $M_{i,j}(m,n)$  is the template selected from the first image,  $\bar{M}_{i,j}$  is the average gray value of all pixels in  $M_{i,j}(m,n)$ ,  $T(m,n)$  is the image at time  $t$ ,  $\bar{T}$  is the average gray value of pixels in  $T(m,n)$  covered by  $M_{i,j}(m,n)$ ,  $(i,j)$  is the coordinates of template translation,  $M$  and  $N$  are the width and height of the template, respectively. Hence, when  $N(i,j) = 1$ , the best matching point is attained.

The traditional template matching algorithm searches the target object pixel by pixel in the matching region and calculates the correlation coefficient  $N(i,j)$  of each pixel. Therefore,

as shown in figure 4, an adaptive hierarchical localization algorithm based on leaping computation is proposed in this paper to reduce the cost and time of matching calculation. The step size and matching area of each layer are reduced gradually. The flowchart of the proposed algorithm is described below.

Step 1: Take the upper left corner pixel of the original image (matching region  $R_1$ ) as the origin and  $D_1$  as the calculation step size. Calculate the correlation coefficient of each discrete pixel in this layer. Select a new matching region  $R_2$  with point  $(x_1,y_1)$ , where the largest correlation coefficient is the center and  $S_1$  is the radiation distance.

Step 2: Take the upper left corner pixel of the matching region  $R_2$  as the origin and  $D_2$  as the calculation step size. Calculate the correlation coefficient of each discrete pixel in this layer. Select a new matching region  $R_3$  with point  $(x_2,y_2)$ , where the largest correlation coefficient is the center and  $S_2$  is the radiation distance.

Step 3: Take the upper left corner pixel of the matching region  $R_3$  as the origin and  $D_3$  as the calculation step size. Calculate the correlation coefficient of each discrete pixel in this layer. Select a new matching region  $R_4$  with point  $(x_3,y_3)$ , where the largest correlation coefficient is the center and  $S_3$  is the radiation distance.

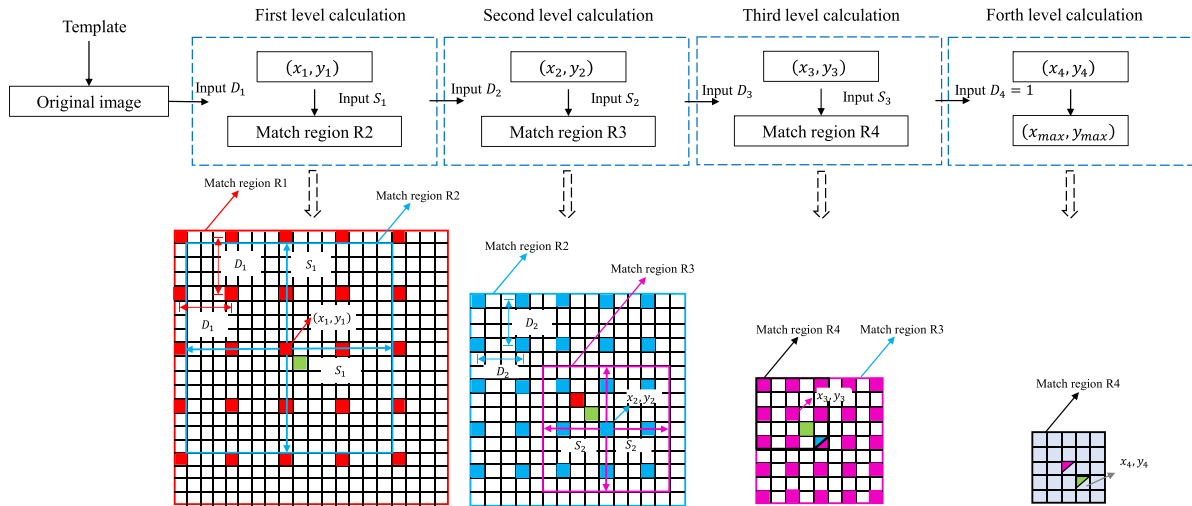


Figure 4. Flowchart of the adaptive hierarchical localization algorithm.

Step 4: Take the upper left corner pixel of the matching region  $R_4$  as the origin and  $D_4 = 1$  as the calculation step size. Calculate the correlation coefficient of each pixel in this layer. Select the pixel coordinates of the maximum correlation coefficient  $(x_4, y_4)$  as the best matching point.

It should be noted that the selection of step size mainly depends on the matching region size and the template size. The step size must be less than the template size so that no pixels are missed during correlation calculations. Moreover, in order to improve calculation efficiency, the step size should not be less than 1/60 of the matching region size. Hence, based on these two basic rules, the step size can be selected as 1/30–1/60 of the matching region size.

In order to verify the accuracy of the proposed algorithm, template matching was conducted for two images with different resolutions: (a) original image size =  $300 \times 300$  pixels and template size =  $10 \times 10$  pixels and (b) original image size =  $1536 \times 1536$  pixel and template size =  $50 \times 50$  pixels. The adaptive hierarchical localization algorithm was carried out in one layer, two layers, three layers, and four layers, and the corresponding results are presented in table 1. It is noticeable that the adaptive hierarchical localization algorithm greatly reduced the computing time.

According to the vibration range of a structure, a calculation program is written to find and locate the template as a vertical strip in a small matching region.

For example, in figure 5, target objects of  $25 \times 25$  pixels and matching regions of  $225 \times 30$  pixels,  $425 \times 30$  pixels,  $625 \times 30$  pixels,  $825 \times 30$  pixels are trimmed from an image of  $1080 \times 1920$  pixels, and the corresponding object location calculation results are listed in table 2.

When the matching area size was doubled, tripled, and quadrupled, the computing time increased by about 13%, 20%, and 50%, respectively.

### 2.3. Spot center location method based on Gaussian surface fitting

The template matching method generates pixel-level displacements, which are not accurate for practical applications; hence, the Gaussian surface fitting method is used to achieve sub-pixel-level displacements.

The bridge deformation video is converted into grayscale images frame by frame. Let the matrix  $g(x, y)$  denote the gray values of different pixels in a grayscale image; hence, an image with  $M \times N$  pixels can be expressed as

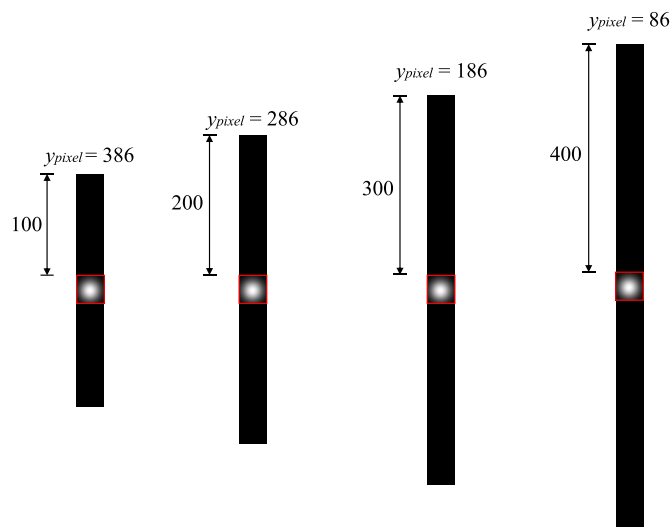
$$G = \begin{bmatrix} g(0,0) & g(0,1) & \cdots & g(0,N-1) \\ g(1,0) & g(1,1) & \cdots & g(1,N-1) \\ \vdots & \vdots & \ddots & \vdots \\ g(M-1,0) & g(M-1,1) & \cdots & g(M-1,N-1) \end{bmatrix}. \quad (7)$$

LED lamps fixed laterally along the longitudinal direction of the bridge are detected as light spots in a grayscale image. In figure 6, the gray value of a light spot decreases gradually from the center to the edges; thus, the ideal distribution of the gray value is a two-dimensional Gaussian distribution (figure 7). In order to accurately measure the change in the center position of the light spot, the center position should be accurately measured.

When the FOV is fixed, the most effective way to improve the location accuracy is to enhance the resolution of an image acquisition equipment (increase the number of pixels). However, in order to track the center position of a light spot with consumer cameras, smartphones, and surveillance cameras available in the market, the accuracy of target location in an image needs to be improved by an image processing algorithm. In the field of computer vision and pattern recognition, numerous sub-pixel location algorithms have been developed to locate the center position of a light spot. Spot

**Table 1.** Efficiency analysis of the adaptive hierarchical localization algorithm.

Image size (pixels)	Computing layer	Step size of each layer (pixels)	Radiation distance (pixels)	Coordinates of the best matching point	Computing time (s)	Increment (%)
(300 × 300)	1	1	—	(180, 180)	0.567	100.00
	2	6, 1	50	(180, 180)	0.143	25.22
	3	8, 6, 1	80, 40	(180, 180)	0.119	20.99
	4	8, 6, 4, 1	100, 60, 30	(180, 180)	0.098	17.28
(1536 × 1536)	1	1	—	(700, 700)	65.567	100.00
	2	20, 1	150	(700, 700)	3.267	4.98
	3	25, 15, 1	200, 120	(700, 700)	2.239	3.41
	4	28, 12, 6, 1	250, 100, 30	(700, 700)	0.563	0.86



**Figure 5.** Location calculation of different matching regions.

**Table 2.** Influence of matching region on computing time.

Case No.	Template size (pixels)	Matching region size (pixels)	Center coordinates of the target (pixels)	Computing time (s)	Increment (%)
1	25 × 25	225 × 30	(468, 510)	0.047	100.00
2	25 × 25	425 × 30	(468, 510)	0.053	112.76
3	25 × 25	625 × 30	(468, 510)	0.056	119.14
4	25 × 25	825 × 30	(468, 510)	0.068	144.68

center extraction methods can be divided into intensity-based and threshold-based algorithms [30]. Intensity-based algorithms include the gray centroid method, the Gaussian surface fitting method, and the paraboloid fitting method [31, 32], and threshold-based algorithms include the ellipse fitting method and the Hough transformation method. It is reported that intensity-based algorithms have better accuracy than threshold-based algorithms [30]. Among intensity-based algorithms, the anti-noise ability of the gray centroid algorithm is weak; thereby, its accuracy is limited. Although the anti-noise ability of the weight centroid algorithm has

been improved, its stability is still insufficient. The Gaussian surface fitting algorithm has high precision and good stability; however, its operation is very complex. The paraboloid fitting algorithm is a simplified form of the Gaussian surface fitting algorithm and has relatively poor accuracy and stability. Therefore, the Gaussian surface fitting algorithm is generally used to calculate the center position of a light spot.

The least square method is used for two-dimensional Gaussian fitting, and the maximum gray value point obtained by surface fitting is considered as the gray center point on the

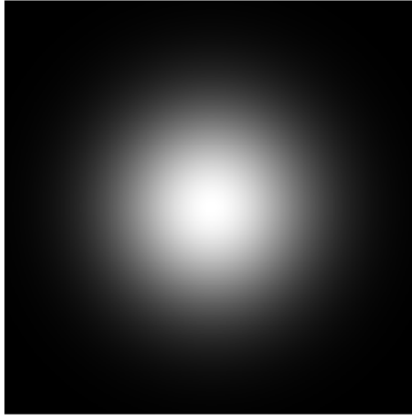


Figure 6. Grayscale image of a light spot.

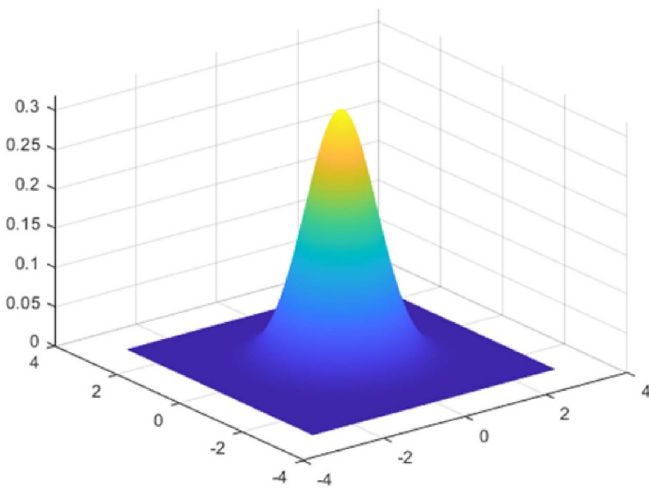


Figure 7. Surface of spot gray value distribution.

surface. The equation of two-dimensional Gaussian distribution can be expressed as

$$I(x,y) = A \exp\left(-\frac{(x-x_0)^2}{2\sigma_x^2} - \frac{(y-y_0)^2}{2\sigma_y^2}\right), \quad (8)$$

where  $I(x,y)$  is the illuminance of a light spot on the  $(x,y)$  plane,  $A$  is the light intensity amplitude,  $\sigma_x$  and  $\sigma_y$  are the standard deviations in  $x$ - and  $y$ -directions, respectively. Now, taking logarithms on both sides of equation (8),

$$\ln I(x,y) = \ln A - \frac{(x-x_0)^2}{2\sigma_x^2} - \frac{(y-y_0)^2}{2\sigma_y^2}. \quad (9)$$

Further, introducing  $F = \ln I(x,y)$ ,  $a = -\frac{1}{2\sigma_x^2}$ ,  $b = -\frac{1}{2\sigma_y^2}$ ,  $c = \frac{x_0}{\sigma_x^2}$ ,  $d = \frac{y_0}{\sigma_y^2}$ ,  $e = \ln A - \frac{x_0^2}{2\sigma_x^2} - \frac{y_0^2}{2\sigma_y^2}$  into equation (9),

$$F = ax^2 + by^2 + cx + dy + e, \quad (10)$$

where  $[a,b,c,d,e]$  are to be estimated. In addition,  $(x_i, y_i, F(x_i, y_i)) \in B$  represents all points in the template for Gaussian fitting. According to equation (10), the objective

function could be determined (equation (11)). Moreover, the parameters  $[a,b,c,d,e]$  are estimated by the least square method.

$$\varepsilon^2 = \min \sum (ax^2 + by^2 + cx + dy + e - F)^2. \quad (11)$$

Let  $\frac{\partial \varepsilon^2}{\partial a} = 0$ ,  $\frac{\partial \varepsilon^2}{\partial b} = 0$ ,  $\frac{\partial \varepsilon^2}{\partial c} = 0$ ,  $\frac{\partial \varepsilon^2}{\partial d} = 0$ ,  $\frac{\partial \varepsilon^2}{\partial e} = 0$ ; thus,

$$\begin{cases} 2 \sum (ax^2 + by^2 + cx + dy + e - F)x^2 = 0 \\ 2 \sum (ax^2 + by^2 + cx + dy + e - F)y^2 = 0 \\ 2 \sum (ax^2 + by^2 + cx + dy + e - F)x = 0 \\ 2 \sum (ax^2 + by^2 + cx + dy + e - F)y = 0 \\ 2 \sum (ax^2 + by^2 + cx + dy + e - F) = 0. \end{cases} \quad (12)$$

Equation (12) can be further transformed into

$$\begin{cases} \sum (ax^4 + by^2x^2 + cx^3 + dx^2y + ex^2) = \sum x^2 F \\ \sum (ax^2y^2 + by^4 + cxy^2 + dy^3 + ey^2) = \sum y^2 F \\ \sum (ax^3 + bxy^2 + cx^2 + dxy + ex) = \sum x F \\ \sum (ax^2y + by^3 + cxy + dy^2 + ey) = \sum y F \\ \sum (ax^2 + by^2 + cx + dy + e) = \sum F. \end{cases} \quad (13)$$

Equation (13) is a positive definite matrix, which can be solved by the Householder transformation method. The center position and light amplitude of a light spot can be expressed as

$$\begin{cases} x_0 = -c/2a \\ y_0 = -d/2b \\ A = \exp(e - c^2/4a - d^2/4b). \end{cases} \quad (14)$$

### 3. Laboratory test of a self-anchored suspension bridge scale model

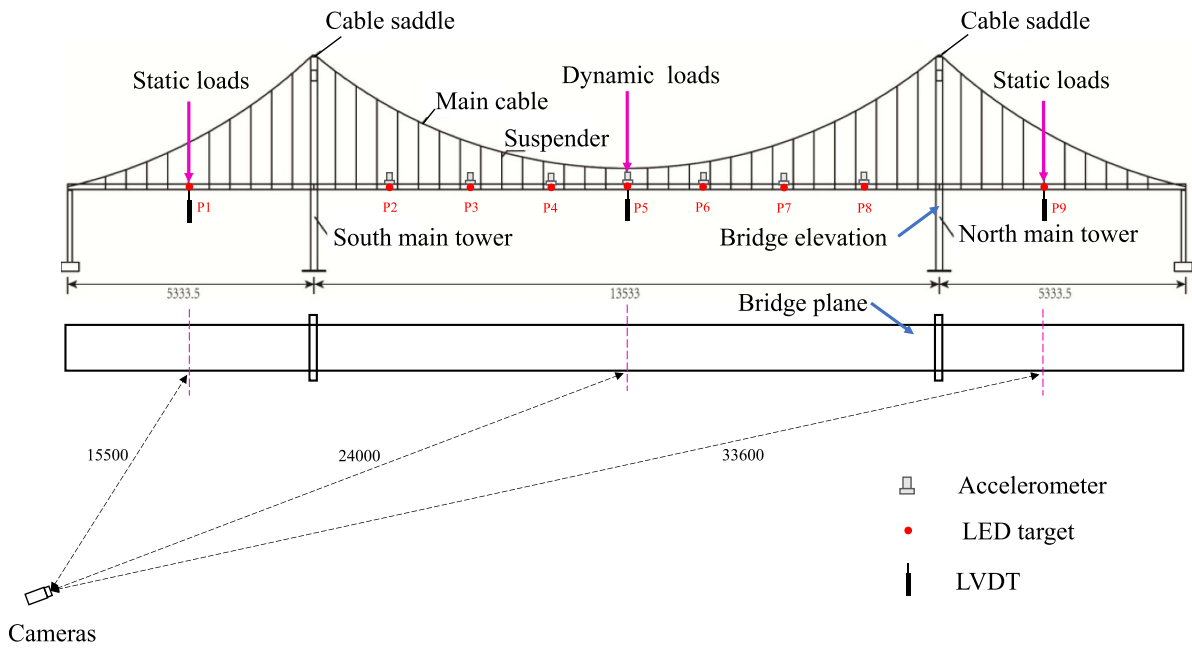
#### 3.1. Experimental model

Static and dynamic loading tests were carried out on a 1/30 scale model of the Taohuayu Yellow River Highway suspension bridge to measure its displacement responses. The span of the model bridge was 5.3335 m + 13.533 m + 5.3335 m = 24.200 m, and its width was 1.2 m. A counterweight of 52 C30 concrete deck slabs was laid on the main girder. Moreover, 10 and 29 pairs of hangers existed in two side spans and the middle span, respectively. The distance between adjacent hangers was 0.45 m, the distance between hangers on two side spans was 0.933 m, and the distance between side hangers and the bridge end was 0.817 m. The heights of the north and south main towers were 3.87 m and 3.96 m, respectively. A stiffening beam with a 2 mm thick diaphragm was placed every 450 mm, and these diaphragms were welded steel box girders. Stiffening beams and diaphragms were made of Q345D steel, the main cable was made of 16 steel wires with a diameter of 3 mm and a cross-sectional area of 117.78 mm<sup>2</sup>, hangers were made



**Table 3.** Structural parameters of the model bridge.

Basic parameter		Unit	Value
Span	Left side span	m	5.3335
	Main span	m	13.533
	Right side span	m	5.3335
Tower	North tower height	m	3.87
	South tower height	m	3.96
Stiffening girder	Moment of inertia	mm <sup>4</sup>	3.63562 × 10 <sup>8</sup>
	Elastic modulus	GPa	206
	Mass density	kg m <sup>-3</sup>	7850
Deck slab	Length × width × height	m	1.16 × 0.45 × 0.2
	Elastic modulus	GPa	30
	Mass density	kg m <sup>-3</sup>	2500
Main cable	Cross-sectional area	mm <sup>2</sup>	117.78
	Elastic modulus	GPa	195
	Mass density	kg m <sup>-3</sup>	7850
Number of girders	Left side span	—	10
	Main span	—	29
	Right side span	—	10



**Figure 8.** Experimental model (unit: mm).

of steel wires with a diameter of 4 mm. Moreover, Q345D steel was used for the two main towers, and 4 mm and 3 mm thick steel plates were used for tie beams and tower columns, respectively. The bottom of each main tower was connected with a 5 mm thick bottom steel plate anchored to the ground by a 5 mm thick stiffening plate, which was welded with the main tower and the bottom steel plate. The bottom steel plate was anchored to the ground by four pieces  $\varphi 50$  bolts. The structural parameters of the model bridge are listed in table 3.




Seven LED lamps were fixed on the main span at equal distances along the span direction of the bridge. One accelerometer was attached at each LED lamp position, and one LVDT was fixed at the middle of the main span. One LED lamp and one LVDT were fixed at the middle of two side spans.

Measuring points from south to north were numbered as P1–P9 (figure 8).

The sampling frequency of contact sensors (LVDTs and accelerometers) was 100 Hz. Displacement and acceleration responses were collected by a dynamic signal testing and analysis system (Model #DH5902N, Donghua Testing Technology Co. Ltd, China). Displacement responses of the model bridge were obtained by tracking the changes in the center positions of LED lamps using a DAHUA network surveillance camera and a Sony 4K camera. The specific model and technical parameters of each data acquisition equipment are listed in table 4.

Static loading tests were mainly carried out on the two side spans, and a static concentrated load (two people) was applied

**Table 4.** Technical specifications of data acquisition equipment.

Device type	Model	Technical parameters	Unit price (\$)
LVDT	 DH - LVDTC20	Range: 0–40 mm Resolution: $\leq 1 \mu\text{m}$ Sampling rate: 100 Hz	308.80
Accelerometer	 DH - 1A202E	Range: $\pm 5 \text{ g}$ Axial sensitivity: $100 \text{ mV m s}^{-2}$ Frequency range: 0.2–1500 Hz Sampling rate: 100 Hz	462.60
LED lamp		Power: 0.3 W Diameter: 12 mm Working voltage: 12 V	0.15
Network surveillance camera	 DAHUA 4843F1-ZYL-PV	Maximum resolution: $2688 \times 1520$ pixels Focal length: 2.7–12 mm Frame rate: 25 fps Image sensor: CMOS Image sensor size: $7.18 \text{ mm} \times 5.32 \text{ mm}$	154.00
Consumer camera	 SONY FDR AX-700	Maximum resolution: $3840 \times 2160$ pixels Focal length: 9.3–111.6 mm Frame rate: 100 fps Image sensor: CMOS Image sensor size: $13.2 \text{ mm} \times 8.8 \text{ mm}$ .	1420.32
Dynamic signal testing and analysis system	 DH 5902N	Sampling rate: 100 Hz	23 160.00

in two levels at the middle of each side span (numbered as S1 and S2). During dynamic loading tests, the bridge was excited in two ways: continuous jumping and running of one person. The continuous jumping scenario was numbered as J1, and low-speed and high-speed running scenarios across the bridge were numbered as R1 and R2, respectively. The specific descriptions of each scenario are listed in table 5. The DAHUA camera and the Sony camera were fixed on the stable ground in front of the north side of the model bridge, and the straight-line distance between the cameras and the middle of the bridge was 24.0 m. Moreover, the cameras were fixed on tripods, and

their optical axes were kept perpendicular to the bridge facade. In order to prevent disturbances caused by manual operation, the cameras were controlled by a Bluetooth remote control device. Figure 9 is the test site photo.

### 3.2. Test results and analysis

**3.2.1. Multi-spot center location.** When the recorded video was converted into grayscale images frame by frame, it was found that the gray value of LED lamps in a grayscale image was significantly larger than those of other areas (figure 10).

**Table 5.** Experimental scenarios.

Case No.	Type	Image acquisition equipment	Measured quantity
S1	Static test	DAHUA camera	Displacement
S2		DAHUA camera	Displacement
J1	Dynamic test	DAHUA and Sony cameras	Displacement and acceleration
R1		DAHUA and Sony cameras	Displacement and acceleration
R2		DAHUA and Sony cameras	Displacement and acceleration



**Figure 9.** Test site photo.

In this experiment,  $5 \times 5$  square filter window templates were used for the median filtering of images. Median filtering is a common nonlinear filtering method. In this technique, the gray value of every pixel point in an image is replaced with an intermediate value after sorting the gray values of all pixels in the neighbouring window of a point. Median filtering can effectively remove random noise and pulse interference in an image and keep edge information.

The center pixel coordinates of a light spot at different times under scenario J1 are displayed in figure 11. The coordinates and gray values of each pixel in region of interest were calculated by  $I(x,y) = A \exp\left(-\frac{(x-x_0)^2}{2\sigma_x^2} - \frac{(y-y_0)^2}{2\sigma_y^2}\right)$ . The coordinates of the center point on the surface were determined by the least square method. The displacements of each measuring point at different times are presented in figures 12 and 13. The vibration displacement time histories of the targets were obtained by concatenating the processed results of each frame with time. If the coordinates of the  $j$ th light spot in the  $i$ th frame are  $(x_i^j, y_i^j)$ , its pixel displacement time history would be  $h_x(y_i^j, t)$ .

**3.2.2. Displacement measurement under static loading scenarios.** Static loading tests were carried out to measure the displacements of P1 and P9. The distances between these two points and the camera were 15.5 m and 33.6 m, respectively. According to camera specifications, the scaling factors of these two points were  $5.5 \text{ mm pixel}^{-1}$  and  $11.9 \text{ mm pixel}^{-1}$ , respectively. The concentrated load was applied to the middle of the side span in two stages. After the first loading stage, data were collected for about 10 s. Similarly, after the second loading stage, data acquisition was performed for 35 s. Two stages were also adopted during unloading. LVDTs and the DAHUA camera were used to record the displacement curves of P1 and P9 in the loading process (figures 14 and 15), and displacement curves detected by the camera were well consistent with those collected by LVDTs.

The values in each loading stage were calculated by averaging all measured values. The displacement measurement errors between LVDTs and the camera in different loading stages are listed in table 6. It is noticeable that the maximum errors of P1 and P9 were 5.54% and 8.45%, respectively; thus, fulfilling the requirements of engineering measurements.



(a)



(b)

Figure 10. Grey image: (a) original image and (b) after median filtering.

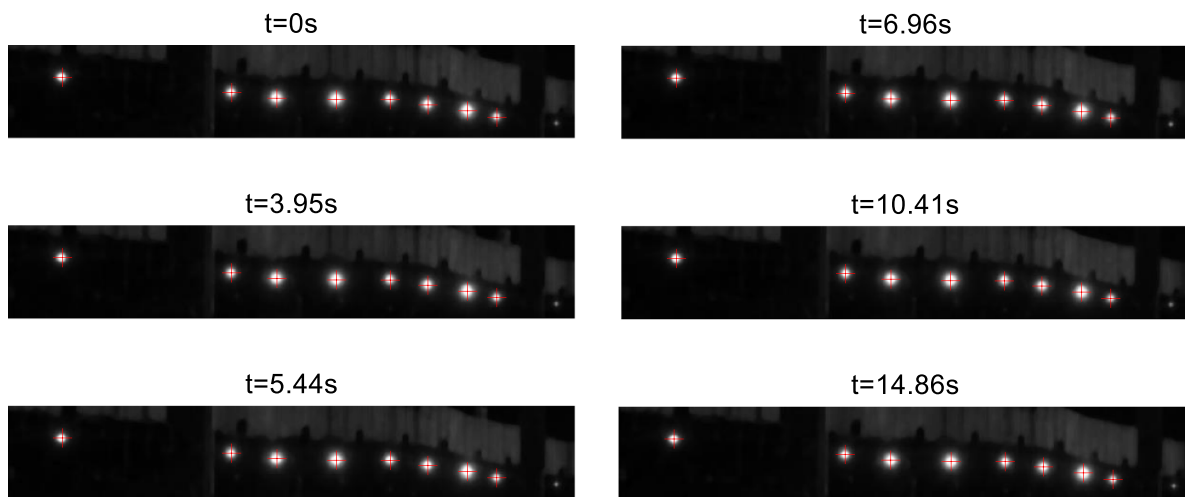


Figure 11. Spot center marking at different times by the Sony camera.

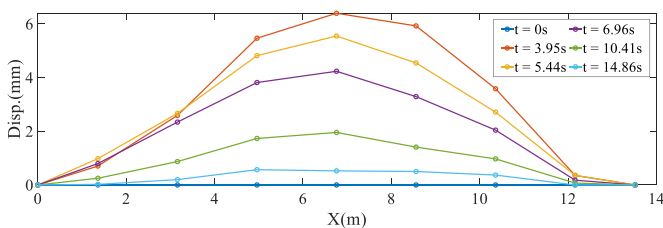


Figure 12. Displacement results of P2–P8 at different times by the Sony camera.

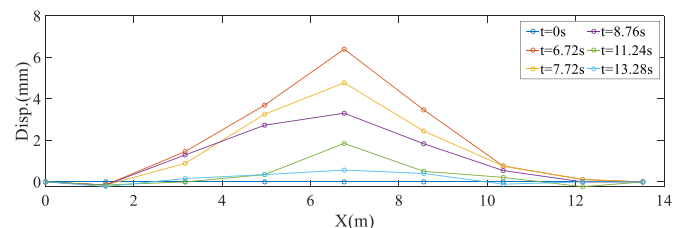


Figure 13. Displacement results of P2–P8 at different times by the DAHUA camera.

3.2.3. *Displacement measurement under dynamic loading scenarios.* Dynamic loading tests were carried out on the main span of the model bridge. The bridge was excited by different means, and the dynamic responses of P2–P8 were measured. The displacement time histories of each measuring point were calculated by multiplying their pixel displacement time histories with the scaling factors. The displacement time histories of P5 under scenarios J1, R1, and R2 are displayed

in figures 16, 17, and 18, respectively. The maximum displacements of scenarios J1 (after each excitation), R1, and R2 are listed in table 7. The maximum errors between the displacement values collected by LVDTs and the Sony camera in scenarios J1, R1, and R2 were 2.90%, 1.06%, and 0.87%, respectively.

It is noticeable from figure 16 that a high degree of consistency was achieved between the displacement results of the

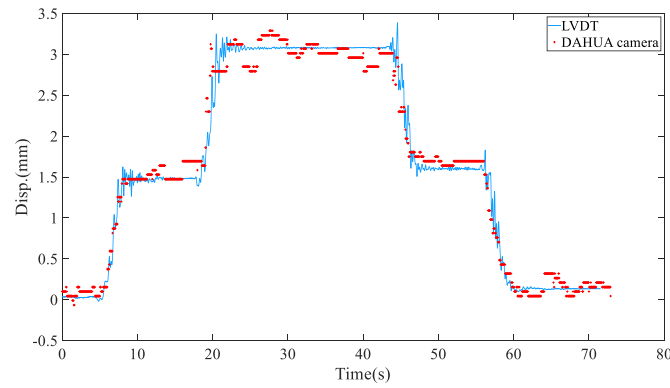


Figure 14. Displacement results of P1 under scenario S1.

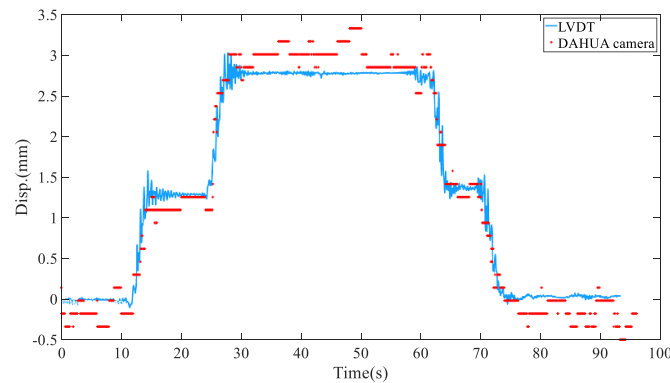


Figure 15. Displacement results of P9 under scenario S2.

Table 6. Displacement results and relative errors under different static loading scenarios.

Case No.	Loading level	LVDT displacement (mm)	DAHUA camera	
			Displacement (mm)	Error (%)
S1	Loading level I	1.485	1.565	5.38
	Loading level II	3.081	3.015	2.14
	Unloading level I	1.606	1.695	5.54
S2	Loading level I	1.284	1.351	5.22
	Loading level II	2.769	3.003	8.45
	Unloading level I	1.355	1.314	3.03

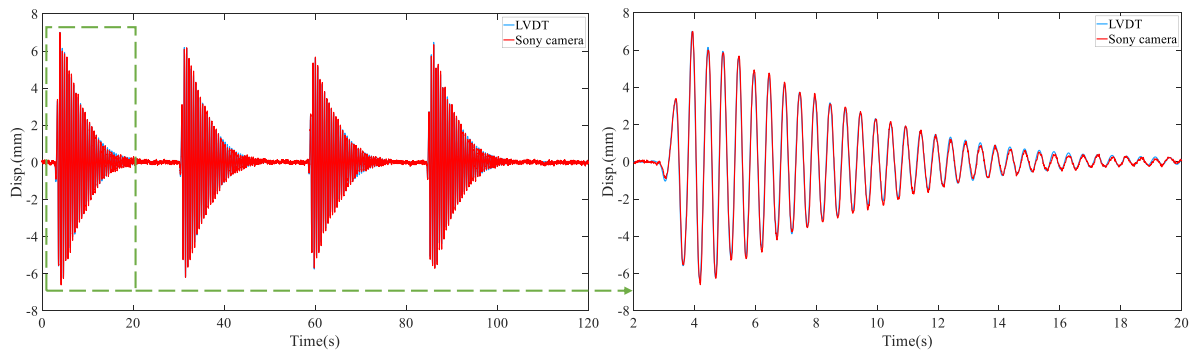
Sony camera and LVDTs. However, the displacement values collected by the DAHUA camera were inconsistent with those of the Sony camera and LVDTs because the sampling frequency of the DAHUA camera was only 25 FPS, which was 1/4 of those of the Sony camera and LVDTs; hence, the displacement measurement reliability of the DAHUA camera was poor. The data obtained from the DAHUA camera was further used for modal analysis.

3.2.4. Modal analysis of scenario J1. The natural frequency of the model bridge was obtained by performing the fast Fourier transformation of the displacement time histories of P2–P8 under scenario J1. The amplitude-frequency curves of P2–P8 obtained based on the acceleration results measured by accelerometers and the displacement results measured by the Sony camera and the DAHUA camera are displayed in

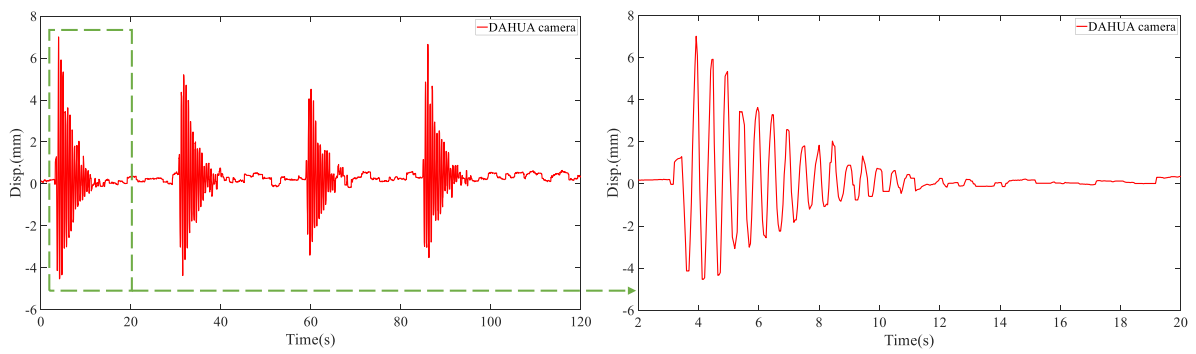
figures 19–24. The peak values of the amplitude-frequency curves were easily detected, and a high consistency was attained between the frequency of each measuring point. Figures 25 and 26 compare the amplitude-frequency curves of P5 obtained from the displacement results of LVDTs and the Sony camera.

The relative errors in the frequencies measured by accelerometers, the Sony camera, and the DAHUA camera are listed in table 8. It is noticeable that the values of the Sony camera and accelerometers were very close and the maximum relative error was only 0.93%. The maximum relative error between the results of the DAHUA camera and accelerometers was 1.82%.

The modal shapes of the model bridge are displayed in figures 27–29. In comparison to accelerometers, the Sony camera and the DAHUA camera yielded better results. Table 9

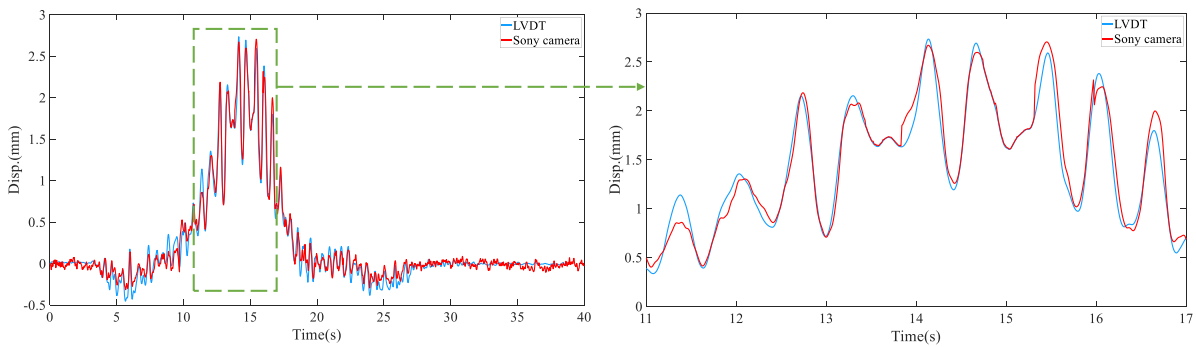


(a)

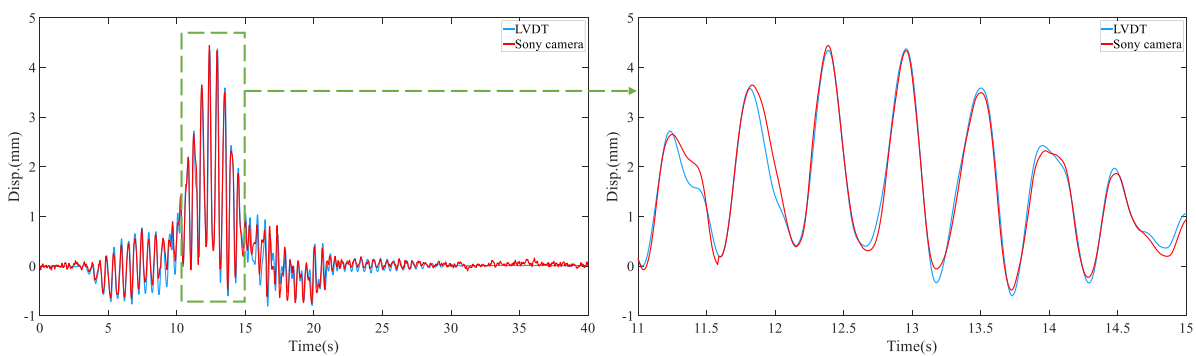


(b)

**Figure 16.** Displacement time histories of P5 in scenario J1 collected by (a) LVDTs and the Sony camera and (b) the DAHUA camera.



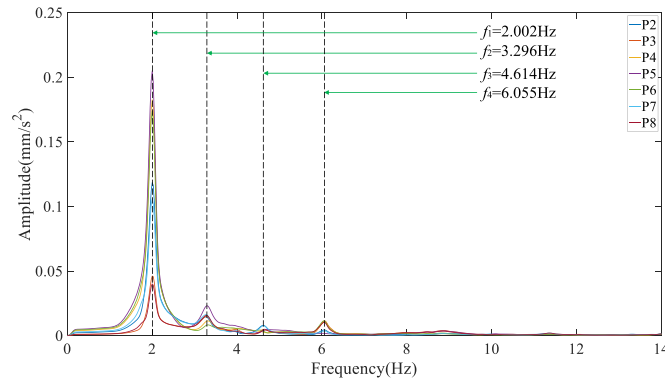
**Figure 17.** Displacement time histories of P5 in scenario R1.



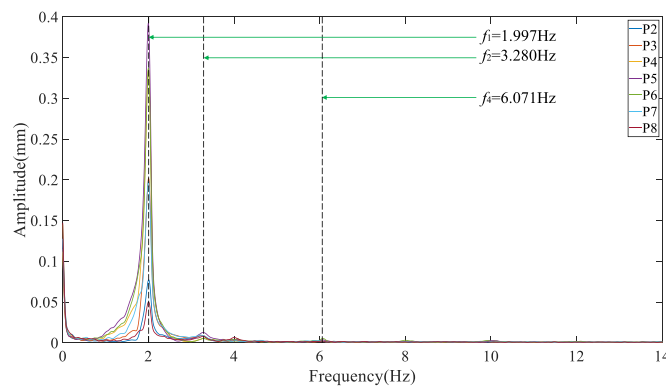
**Figure 18.** Displacement time histories of P5 in scenario R2.

**Table 7.** Displacement measurement results and relative errors under different dynamic loading scenarios.

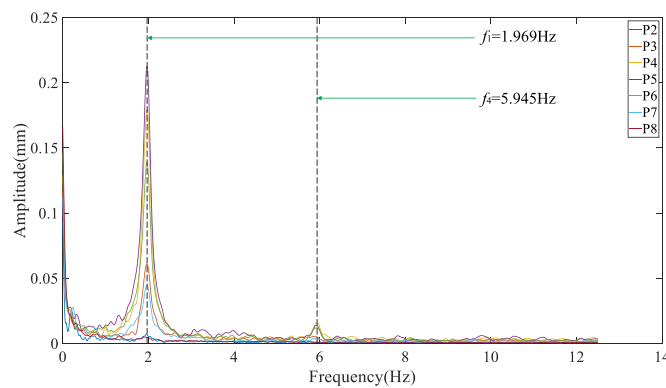
Case No.	Excitation sequence	LVDT displacement (mm)	Sony FDR AX-700	
			Displacement (mm)	Error (%)
J1	1st	7.012	7.041	0.41
	2nd	6.192	6.119	1.18
	3rd	5.317	5.471	2.90
	4th	5.788	5.743	0.78
R1	—	2.733	2.704	1.06
R2	—	4.348	4.310	0.87



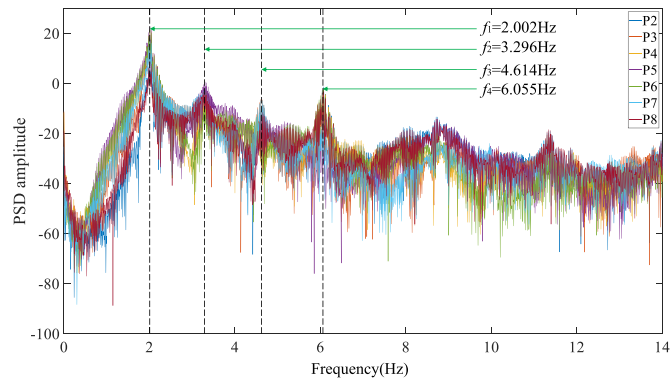
**Figure 19.** Amplitude spectra of P2–P8 measured by accelerometers.



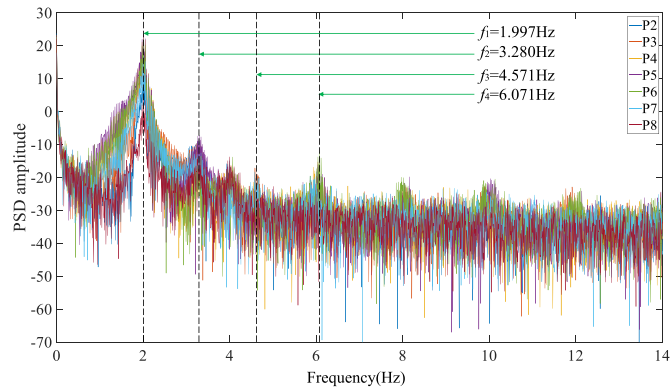
**Figure 20.** Amplitude spectra of P2–P8 measured by the Sony camera.



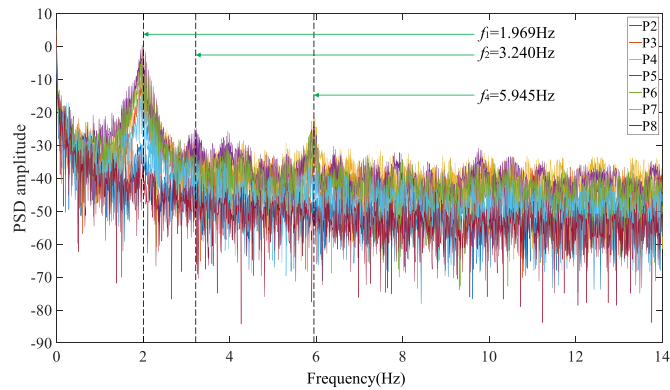
**Figure 21.** Amplitude spectra of P2–P8 measured by the DAHUA camera.



**Figure 22.** Power spectra of P2–P8 measured by accelerometers.



**Figure 23.** Power spectra of P2–P8 measured by the Sony camera.



**Figure 24.** Power spectra of P2–P8 measured by the DAHUA camera.



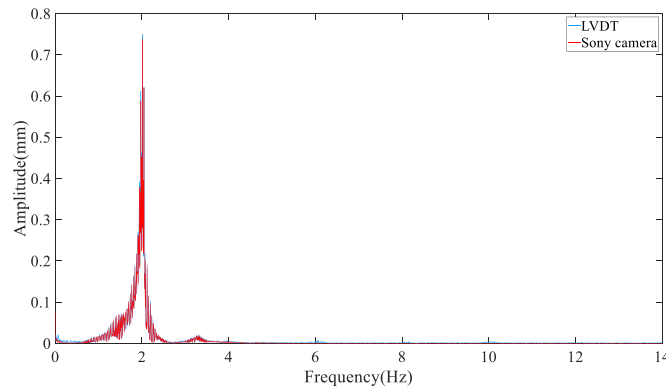


Figure 25. Amplitude spectra of P5 obtained from the displacement results of LVDTs and the Sony camera.

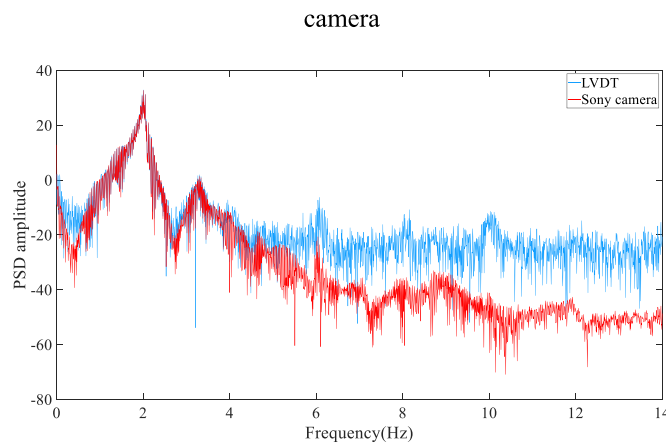


Figure 26. Power spectra of P5 obtained from the displacement results of LVDTs and the Sony camera.

Table 8. Natural frequencies and relative errors.

Mode order	Accelerometer	Sony camera		DAHUA camera	
	Frequency (Hz)	Frequency (Hz)	Error (%)	Frequency (Hz)	Error (%)
1	2.002	1.997	0.25	1.969	1.65
2	3.296	3.280	0.49	3.240	1.70
3	4.614	4.571	0.93	—	—
4	6.055	6.071	0.26	5.945	1.82

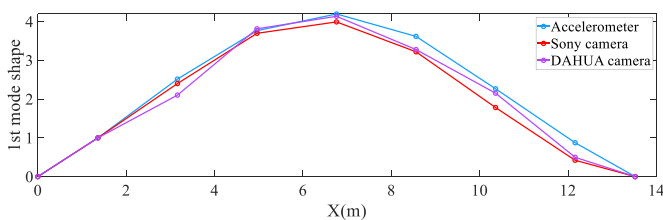


Figure 27. First-order mode.

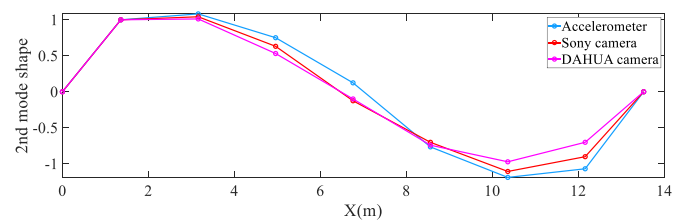


Figure 28. Second-order mode.

reveals that the cost of the proposed system is much cheaper than that of the conventional system. Moreover, according to table 4, as the cost of LED lamps is much cheaper than that

of contact sensors, the spatial resolution of vibration modes can be improved by increasing the density of LED lamps the bridge facade with little cost increased.

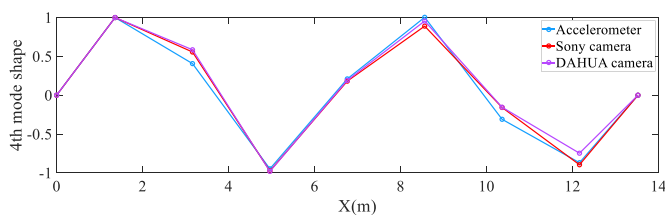


Figure 29. Fourth-order mode.

Table 9. Cost comparison of the proposed system and the conventional system.

Item	System composition	Total price (\$)
Conventional system	Accelerometer $\times 7$ , LVDT $\times 1$ , Dynamic signal testing and analysis system $\times 1$	26 707.00
Proposed system	LED lamp $\times 7$ , DAHUA camera $\times 1$ , Sony camera $\times 1$	1575.37

#### 4. Conclusions

In order to improve efficiency and reduce costs, a vision-based measurement system equipped with an inexpensive surveillance camera and a consumer camera was developed. In addition, an improved target tracking algorithm based on the prior knowledge of structural deformation was proposed for displacement measurement and modal parameter identification. Static and dynamic loading tests were conducted on a 1/30 scale model of the Taohuayu Yellow River Highway suspension bridge to validate the effectiveness of the proposed vision-based measurement system. Displacement results obtained from the vision-based measurement system and LVDTs as well as modal parameter results obtained from the vision-based measurement system and accelerometers were compared. The main observations of this investigation are summarized below.

- The improved template matching algorithm based on leapfrog computation greatly saved computing time. According to the characteristics of bridge deformation, the matching region was pre-selected to reduce computing time. When the size of the matching region was quadrupled, the computing time increased by about 50%.
- In static loading tests, a good consistency between displacement results obtained from the inexpensive surveillance camera and LVDTs was noticed, and the maximum error was calculated as 8.45% at a distance of 33.6 m.
- In dynamic loading tests, dynamic displacements obtained from the consumer camera and LVDTs maintained great consistency with a maximum error of 2.90%.
- Low-order frequencies and mode shapes were obtained by the modal analysis of vibration displacement time histories. The maximum errors of low-order frequencies for the consumer camera and the surveillance camera were 0.93%

and 1.82%, respectively. Hence, natural frequencies and mode shapes identified by the cameras were in good agreement with those detected by the accelerometers.

Therefore, the proposed vision-based measurement system could offer a non-contact, cost-effective, and time-saving alternative to conventional displacement sensors and accelerometers for SHM. Although dynamic displacement results obtained from the inexpensive surveillance camera were not ideal, modal parameter identification results were found to be satisfactory.

#### Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

#### Funding

The authors acknowledge the financial support for this research from the National Natural Science Foundation of China (Grant Nos. 51708068 and 51778094) and the Graduate Education Innovation Fund project (Grant No. 2021B0004). The statements made herein are solely the responsibility of the authors.

#### Conflict of interest

The authors declare no known competing financial interests or personal relationships that could influence the work reported in this paper.

#### ORCID iDs

Tong Wu  <https://orcid.org/0000-0002-9726-0794>

Shuai Shao  <https://orcid.org/0000-0002-9243-6666>

#### References

- Hao W, Li A, Guo T and Tao T 2014 Establishment and application of the wind and structural health monitoring system for the Runyang Yangtze river bridge *Shock Vib.* **2014** 421038
- Feng D and Feng M 2015 Model updating of railway bridge using *in situ* dynamic displacement measurement under trainloads *J. Bridge Eng.* **20** 04015019.1–04015019.12
- Psimoulis P A and Stiros S C 2013 Measuring deflections of a short-span railway bridge using a robotic total station *J. Bridge Eng.* **18** 182–5
- Fukuda Y, Feng M Q and Shinozuka M 2010 Cost-effective vision-based system for monitoring dynamic response of civil engineering structures *Struct. Control Health Monit.* **17** 918–36
- Kohut P, Holak K, Uhl T, Ortyl Ł, Owerko T, Kuras P and Kociertz R 2013 Monitoring of a civil structure's state based on noncontact measurements *Struct. Health Monit.* **12** 411–29
- Gentile C and Bernardini G 2009 An interferometric radar for non-contact measurement of deflections on civil

- engineering structures: laboratory and full-scale tests *Struct. Infrastruct. Eng.* **6** 521–34
- [7] Brown C J, Karuma R, Ashkenazi V, Roberts G W and Evans R A 1999 Monitoring of structures using the global positioning system *Struct. Build.* **134** 97–105
- [8] Meng X, Dodson A H and Roberts G W 2007 Detecting bridge dynamics with GPS and triaxial accelerometers *Eng. Struct.* **29** 3178–84
- [9] Xia H, Roeck G D, Zhang N and Maeck J 2003 Experimental analysis of a high-speed railway bridge under Thalys trains *J. Sound Vib.* **268** 103–13
- [10] Pieraccini M, Parrini F, Fratini M, Atzeni C, Spinelli P and Micheloni M 2007 Static and dynamic testing of bridges through microwave interferometry *NDT&E Int.* **40** 208–14
- [11] Nassif H H, Gindy M and Davis J 2005 Comparison of laser doppler vibrometer with contact sensors for monitoring bridge deflection and vibration *NDT&E Int.* **38** 213–8
- [12] Ye X W, Dong C Z and Liu T 2016 A review of machine vision-based structural health monitoring: methodologies and application *J. Sens.* **5** 7103039
- [13] Busca G, Cigada A, Mazzoleni P and Zappa E 2014 Vibration monitoring of multiple bridge points by means of a unique vision-based measuring system *Exp. Mech.* **54** 255–71
- [14] Busca G, Cigada A, Mazzoleni P, Zappa E and Franzi M 2012 Cameras as displacement sensors to get the dynamic motion of a bridge: performance evaluation against traditional approaches *Proc., 6th Int. Conf. Bridge Maintenance, Safety and Management, Int. Association for Bridge Maintenance and Safety (IABMAS)* pp 2835–41
- [15] Ribeiro D, Calçada R, Ferreira J and Martins T 2014 Non-contact measurement of the dynamic displacement of railway bridges using an advanced video-based system *Eng. Struct.* **75** 164–80
- [16] Feng D, Feng M Q, Ozer E and Fukuda Y 2015 A vision-based sensor for noncontact structural displacement measurement *Sensors* **15** 16557–75
- [17] Bhowmick S, Nagarajaiah S and Lai Z 2020 Measurement of full-field displacement time history of a vibrating continuous edge from video *Mech. Syst. Signal Process.* **144** 106847
- [18] Mas D, Ferrer B, Acevedo P and Espinosa J 2016 Methods and algorithms for video-based multi-point frequency measuring and mapping *Measurement* **85** 164–74
- [19] Son K-S, Jeon H-S, Park J-H and Park J W 2015 Vibration displacement measurement technology for cylindrical structures using camera images *Nucl. Eng. Technol.* **47** 488–99
- [20] Trebuña F and Hagara M 2014 Experimental modal analysis performed by high-speed digital image correlation system *Measurement* **50** 78–85
- [21] Dong C Z, Ye X W and Jin T 2017 Identification of structural dynamic characteristics based on machine vision technology *Measurement* **126** 405–16
- [22] Oh B K, Hwang J W, Kim Y, Cho T and Park H S 2015 Vision-based system identification technique for building structures using a motion capture system *J. Sound Vib.* **356** 72–85
- [23] Ye X W, Dong C Z and Liu T 2016 Image-based structural dynamic displacement measurement using different multi-object tracking algorithms *Smart Mater. Struct.* **17** 935–56
- [24] Wabbeh A M, Caffrey J P and Masri S F 2003 A vision-based approach for the direct measurement of displacements in vibrating systems *Smart Mater. Struct.* **12** 785–94
- [25] Park S W, Park H S, Kim J H and Adeli H 2015 3D displacement measurement model for health monitoring of structures using a motion capture system *Measurement* **59** 352–62
- [26] Maksymenko O P, Sakharuk O M, Ivanytskyi Y L and Kun P S 2020 Multilaser spot tracking technology for bridge structure displacement measuring *Struct. Control Health* **28** 2675
- [27] Zhang L X, Liu P, Yan X and Zhao X F 2020 Middle displacement monitoring of medium–small span bridges based on laser technology *Struct. Control Health* **27** 2509
- [28] Vicente M, Gonzalez D, Minguez J and Schumacher T 2018 A novel laser and video-based displacement transducer to monitor bridge deflections *Sensors* **18** 970
- [29] Tian L and Pan B 2016 Remote bridge deflection measurement using an advanced video deflectometer and actively illuminated LED targets *Sensors* **16** 1344
- [30] Shortis M R, Clarke T A and Short T 1994 Comparison of some techniques for the subpixel location of discrete target images *Proc. SPIE* **2350** 239–50
- [31] Dong H and Wang L 2012 Non-iterative spot center location algorithm based on Gaussian for fish-eye imaging laser warning system *Optik* **123** 2148–53
- [32] Tang Y Q, Gu G H, Qian W X, Chen Q C and Zhang J 2017 Laser spot center location algorithm of four-quadrant detector based on Gaussian distribution *Infrared Laser Eng.* **46** 206003