



# Robust and versatile vision-based dynamic displacement monitoring of natural feature targets in large-scale structures

Shengfei Zhang<sup>1</sup> · Qiang Han<sup>1</sup> · Kejie Jiang<sup>1</sup> · Xinzheng Lu<sup>2</sup> · Guoquan Wang<sup>3</sup>

Received: 3 July 2023 / Accepted: 22 May 2024  
© Springer-Verlag GmbH Germany, part of Springer Nature 2024

## Abstract

Dynamic displacement response is an essential indicator for assessing structural state and performance. Vision-based structural displacement monitoring is considered as a promising approach. However, the current vision-based methods usually only focus on certain application scenarios. This study introduces a Sparse Bayesian Learning-based (SBL) algorithm to enhance robustness, accuracy, and computational efficiency in target tracking. Furthermore, a robust and versatile Vision-based Dynamic Displacement Monitoring System (VDDMS) was developed, capable of monitoring displacements of varying application scenarios. The robustness of the proposed algorithm under changing illumination conditions is validated through a specially designed indoor experiment. The feasibility of field application of VDDMS is confirmed through an outdoor shear wall shaking table test. Furthermore, a large-scale bridge shaking table test is conducted to evaluate the reliability and versatility of VDDMS in monitoring natural feature targets on large structures subjected to different seismic excitations. The root mean square error, when compared to laser displacement sensors, ranges from 0.2% to 2.9% of the peak-to-peak displacement. Additionally, VDDMS accurately identifies multi-order frequencies in bridge structures. The study investigates the influence of initial template selection on accuracy, highlighting the significance of distinctive texture features. Moreover, two error evaluation schemes are proposed to quickly assess the reliability of vision-based displacement sensing technologies in various application scenarios.

**Keywords** Dynamic displacement monitoring · Vision-based sensing technology · Illumination changes · Large-scale structures · Shaking table tests · Natural feature targets

## 1 Introduction

Structural Health Monitoring (SHM) aims to provide valuable information for assessing structural integrity and making maintenance decisions by measuring the structural response [1]. Among various structural indices used in SHM, displacement response plays a crucial role [2]. Monitoring the dynamic displacements of structures under different load types offers valuable insights into their condition and

behavior. These dynamic displacements allow for calculating important structural properties such as bearing capacity, deflection, deformation, load distribution, and modal parameters. Furthermore, they can be converted into physical indicators for assessing structural safety.[3] However, traditional displacement measurement methods, which involve placing a limited number of sensors on the surface of a structure, such as Laser Displacement Sensors (LDS) and wireless accelerometers, have limitations. They are cumbersome to install, expensive to maintain, and provide measurements only at discrete points, limiting spatial resolution [4].

In recent years, vision-based displacement sensing technology has emerged as a promising alternative [5]. This technology initially tracks the movement trajectory of the measured target in video and subsequently determines the dynamic displacement of the structure by analyzing the positional relationship between the camera and the structure. [6] Vision-based displacement sensing technology offers advantages such as long-distance capability, non-contact

✉ Qiang Han  
qhan@bjut.edu.cn

<sup>1</sup> National Key Laboratory of Bridge Safety and Resilience, Beijing University of Technology, Beijing 100124, China

<sup>2</sup> Key Laboratory of Civil Engineering Safety and Durability of China Education Ministry, Department of Civil Engineering, Tsinghua University, Beijing 100084, China

<sup>3</sup> Department of Earth and Atmospheric Sciences, University of Houston, Houston, TX 77204, USA

operation, and wide range coverage [7–9]. It has garnered significant interest from researchers and engineers in the field of SHM and has been further developed to achieve modal identification [10, 11], model updating [12], damage detection [13, 14], load identification [15, 16], and cable force estimation [17], and other applications.

The core of vision-based displacement sensing technology lies in target tracking algorithms. However, commonly used algorithms have certain limitations. Optical flow methods [18–22] and phase-based motion magnification methods [23, 24] are effective for minor displacement changes but may fail with large target displacements and exhibit limited robustness. Optical flow methods are sensitive to illumination changes and prone to accumulating errors, while phase-based motion magnification methods are prone to noise from non-measurement target movements. Feature point matching methods [25–27] can track larger displacements but require adjusting multiple parameters and thresholds. Among these, correlation-based template matching methods [28–31] stand out due to their robustness, versatility, and minimal user intervention requirement. However, this method typically necessitates mounting targets on the structure and can only extract displacements with integer pixel accuracy [32]. Several scholars have proposed enhanced techniques to address these limitations. Feng & Feng [33] employ upsampling techniques through Fourier transform, which results in a substantial computational burden without significant accuracy improvement. Pan et al. [34] significantly enhance measurement accuracy using the Inverse-Compositional-Gauss–Newton (IC-GN) nonlinear optimization algorithm, yet the iterative and interpolation procedures involved result in high computational costs. Zhang et al. [35] introduce the Modified Taylor Approximation-based sub-pixel refinement (MTA) algorithm as an additional step after correlation-based matching. This algorithm demonstrates excellent computational efficiency and accuracy, but it remains sensitive to changes in illumination or noise. The existing correlation-based template matching approaches still lack the desired robustness and efficiency.

In addition, despite the attention received by vision-based displacement sensing technologies in SHM, their practical applications are still in their infancy. Most research has primarily focused on specific scenarios, such as measuring the dynamic displacement of small-scale structures or the quasi-static displacement localized areas of larger structures [36]. However, for comprehensive SHM, it is crucial to extend the application of monitoring dynamic displacements to full-scale structures, particularly for modal analysis of slender bridge structures. Before vision-based displacement sensors can fully replace traditional sensors in the SHM field, it is essential to conduct further research that explores their application on larger full-scale structural targets and in more challenging environmental conditions.

The study aims to address these gaps through the following objectives:

- (1) Introducing a Sparse Bayesian Learning-based (SBL) algorithm to enhance robustness, accuracy, and computational efficiency in target tracking. Furthermore, developing a robust and versatile Vision-based Dynamic Displacement Monitoring System (VDDMS) capable of monitoring displacements of varying application scenarios.
- (2) Verifying the effectiveness of the proposed algorithm and VDDMS through a specially designed indoor test and an outdoor shear wall shaking table test.
- (3) Conducting a shaking table test on a large-scale (1:40) steel arch bridge model. Employing a single entry-level consumer camera, VDDMS monitors the displacement and frequencies of the bridge under different seismic excitations and lighting conditions, targeting the natural texture features of the structure's surface.
- (4) Examining the effect of initial template selection on the accuracy of VDDMS and proposing two fast and convenient error assessment schemes suitable for field applications.

The organization of this paper is as follows: Sect. 2 explains the theoretical framework of the VDDMS, including the introduction of the SBL algorithm. In Sect. 3, two experiments are conducted to verify the validity of the proposed algorithm and system. Section 4 encompasses the shaking table test conducted on the bridge. Finally, Sect. 5 summarizes the conclusions drawn from this study and outlines potential directions for future work.

## 2 Theoretical framework

The robust and versatile Vision-based Dynamic Displacement Monitoring System (VDDMS) framework is composed of three phases, as depicted in Fig. 1. In phase 1, a camera calibration process is performed to acquire camera distortion parameters and a projection matrix that includes both camera intrinsic and extrinsic parameters. Phase 2 involves tracking selected targets using the proposed SBL algorithm. In phase 3, the physical displacement of the targets in world coordinates is calculated. The VDDMS enables efficient and robust monitoring of dynamic displacements at multiple points. This approach allows for accurate and reliable displacement measurements in various scenarios, providing a practical and accessible solution for monitoring dynamic movements.

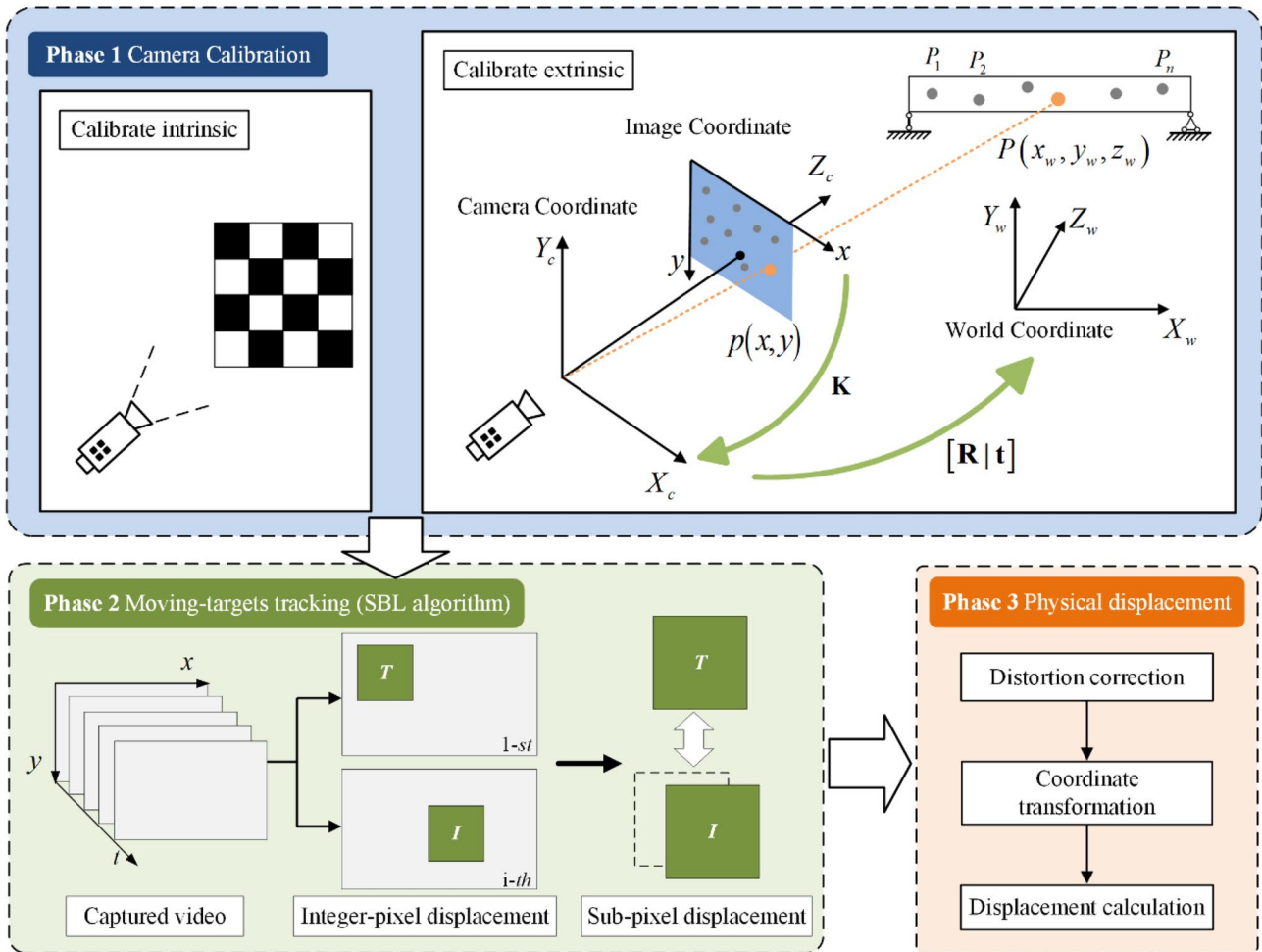


Fig. 1 Framework of the Vision-based Dynamic Displacement Monitoring System (VDDMS)

2.1 Camera calibration

(1) Full camera calibration process

The camera calibration process establishes the projection relationship between three-dimensional (3D) world coordinates and two-dimensional (2D) image coordinates. The transformation from image coordinates  $\mathbf{x} = (x, y)$  to world coordinates  $\mathbf{X} = (X, Y, Z)$  is represented by Eq. (1)

$$s\mathbf{x} = \mathbf{K}[\mathbf{R}|\mathbf{t}]\mathbf{X} \Leftrightarrow s \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x \gamma c_x \\ 0 f_y c_y \\ 0 0 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

where  $s$  is the scale factor,  $\mathbf{K}$  is the camera intrinsic matrix related to the lens, and  $[\mathbf{R}|\mathbf{t}]$  is the camera extrinsic matrix related to the relative position between the camera and the measurement object.

The intrinsic matrix is unchanged so long as the lens focal length does not change. However, the extrinsic matrix must be recalibrated whenever the camera position is changed. Therefore, this study proposes to divide camera calibration into two steps. The intrinsic camera matrix and distortion parameters are estimated using a checkerboard method [37]. The extrinsic camera matrix is obtained through the Perspective-n-point method [38] using 2D-to-3D point correspondences.

(2) Simplified camera calibration process

In scenarios where the full camera calibration process poses challenges, two simplified calibration procedures can be used. The first is the planar homography matrix method [38], simplifies Eq. (1) using the planar homography matrix  $\mathbf{H}$ , as shown in Eq. (2). The

Perspective-n-point method can be applied to solve this equation.

$$s\mathbf{x} = \mathbf{H}\mathbf{X} \Leftrightarrow \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (2)$$

The second is the scale factor method, which is the most frequently used method in other studies. [39] The scale factor method estimates calibration parameters using the physical size  $D$  of the selected object and its corresponding pixel number  $d$  in the image plane. The scale factor method can be represented by the following equation.

$$s = \frac{D}{d} \quad (3)$$

The scale factor method, while simple, has several limitations [40]. The choice of camera calibration process depends on the specific characteristics of the camera, lens, and motion involved. Here are some suggestions for selecting the camera calibration process: (1) For scenarios where the measured structure exhibits one-dimensional motion and the camera is positioned perpendicular to the movement plane, the scale factor method is a suitable choice for camera calibration. However, it is important to note that when working with targets at different positions, recalibration might be necessary. (2) When measuring a structure that moves on a 2D plane and the distortion of the lens is small, the homography matrix method is suitable. The camera tilt angle does not affect the measurement results. (3) In scenarios requiring 3D displacement measurement, a large field of view, or significant lens distortion, a full camera calibration process is necessary [41]. This comprehensive calibration accounts for the lens distortion and ensures accurate and reliable measurements for each target in the complex environments.

By considering these suggestions and selecting the appropriate camera calibration process, the VDDMS can adapt to different scenarios and provide reliable and precise displacement measurements for dynamic movements. In the experimental part of this study, three camera calibration processes are used separately.

## 2.2 The SBL algorithm

### 2.2.1 Initial integer-pixel displacement

The framework of the Sparse Bayesian Learning-based (SBL) targets tracking algorithm, shown in Fig. 2, comprises two main steps: initial integer-pixel displacement estimation and further sub-pixel refinement. In the initial integer-pixel displacement estimation, correlation-based template matching is

used. The selected template is slid on another image, and the best match is found by calculating the correlation between the template and the overlapping region during the sliding process. The zero-mean normalized cross-correlation coefficient (ZNCC) is utilized to quantify the correlation strength between the variables in this paper. Compared with other correlation coefficients, it provides more accurate and reliable results and is insensitive to offsets and scale changes in the intensity of the target area [42]. The ZNCC is calculated as follows.

$$ZNCC(u, v) = \frac{\sum_{x=x_1}^{x_2} \sum_{y=y_1}^{y_2} (T(x, y) - \bar{T})(I^i(x+u, y+v) - \bar{I}^i)}{\sqrt{\sum_{x=x_1}^{x_2} \sum_{y=y_1}^{y_2} (T(x, y) - \bar{T})^2} \sqrt{\sum_{x=x_1}^{x_2} \sum_{y=y_1}^{y_2} (I^i(x+u, y+v) - \bar{I}^i)^2}} \quad (4)$$

where  $T(x, y)$  and  $I^i(x, y)$  represent the grayscale intensity of the first frame and  $i$ -th frame, respectively. The  $u$  and  $v$  values denote integer-pixel displacement change.  $(x_1, y_1)$  and  $(x_2, y_2)$  are the coordinates of the top-left and bottom-right corners of the template in the first frame.  $\bar{T}$  and  $\bar{I}^i$  denote the mean intensity values of the template and the overlapping region, respectively.  $\bar{T} = \frac{1}{\mathcal{A}} \sum_{x=x_1}^{x_2} \sum_{y=y_1}^{y_2} T(x, y)$  and  $\bar{I}^i = \frac{1}{\mathcal{A}} \sum_{x=x_1}^{x_2} \sum_{y=y_1}^{y_2} I^i(x+u, y+v)$ , where  $\mathcal{A}$  represents the area of the template.

While the ZNCC is at its maximum, indicating the best match in overlapping region. While sliding the template over the entire image can be time-consuming, an adaptive matching region algorithm is proposed to improve efficiency in this paper. The algorithm performs the template matching process on a local region instead of the entire image. The center of the sliding matching region for the template in each frame corresponds to the center of the best match in the previous frame. This region is larger than the template, with its size denoted as  $\lambda \cdot \mathcal{A}$ . If the maximum ZNCC of the current frame is less than 0.8, increase the value of  $\lambda$  and repeat the template matching process until the local region expands to the image boundary. This adaptive matching region technique improves the efficiency of the matching process while reducing the probability of false matches.

### 2.2.2 Refined sub-pixel displacement

The correlation-based template matching can only estimate the pixel-level displacement changes. Further refinement is required to obtain more accurate and reliable sub-pixel displacement changes. The relationship between the intensity of a physical point in the template of the first frame and the intensity of the corresponding point in the best matching region of the  $i$ -th frame can be expressed as:

$$T(x, y) = I^i(x+u + \Delta u, y+v + \Delta v) \quad (5)$$

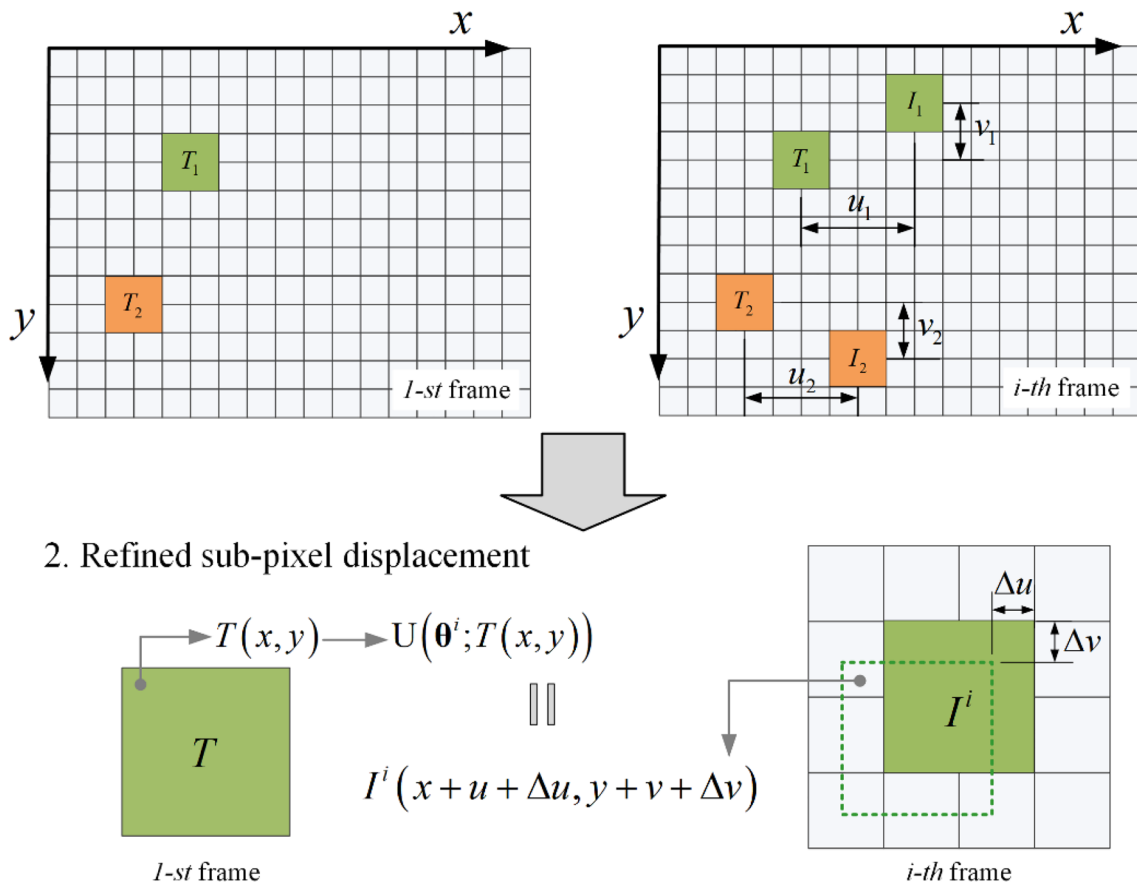


Fig. 2 Framework of the Sparse Bayesian learning-based targets tracking (SBL) algorithm

where  $u$  and  $v$  are the initial integer-pixel displacement components,  $\Delta u$  and  $\Delta v$  are the sub-pixel displacement components, respectively.

To account for grayscale changes caused by illumination variations or overexposure/underexposure, a nonlinear brightness variation model is introduced as follows:

$$U(\theta^i; T(x, y)) = \theta_1^i \cdot T(x, y)^2 + \theta_2^i \cdot T(x, y) + \theta_3^i \quad (6)$$

where  $\theta^i$  is a parameter vector that describes the grayscale transformation between the initial template and the best matching region of the  $i$ -th frame. More complex nonlinear brightness variation models can be formulated by modifying this equation.

The first-order Taylor expansion of Eq. (5) at  $(x + u, y + v)$  is as follows

$$\theta_1^i \cdot T(x, y)^2 + \theta_2^i \cdot T(x, y) + \theta_3^i - \Delta u \cdot I_{x+u}^i - \Delta v \cdot I_{y+v}^i = I^i(x + u, y + v) \quad (7)$$

where  $I_{x+u}^i$  and  $I_{y+v}^i$  are the spatial gradients of the  $i$ -th frame at  $(x + u, y + v)$ , calculated using the gray gradient algorithm based on the Barron operator [43].

A sub-pixel displacement regression reference model is established using the sparse Bayesian learning scheme [44, 45]. It assumes that the grayscale change of each pixel point within the template is consistent. Let  $n$  represent the index of the pixel points, ranging from  $n = 1$  to  $N$ , where  $N$  is the total number of pixels in the template. The grayscale data on both sides of Eq. (7) are represented as input samples  $\mathbf{x}_n$  and target values  $t_n$ , respectively. The training dataset  $\mathcal{D} = \{\mathbf{x}_n, t_n\}_{n=1}^N$  is then used in the sparse Bayesian learning scheme, and noise  $\epsilon_n$  is introduced:

$$t_n = y(\mathbf{x}_n, \mathbf{w}) + \epsilon_n = w_0 + \sum_{j=1}^{M-1} w_j \cdot \phi_j(\mathbf{x}_n) + \epsilon_n \quad (8)$$

where  $y(\mathbf{x}_n, \mathbf{w})$  represents the objective function,  $\mathbf{w} = [w_0, w_1, \dots, w_{M-1}]$  represents an unknown parameter vector, where parameter  $w_j$  either sub-pixel displacement or illumination change parameters, and  $\phi_j(\mathbf{x}_n)$  is the basic function.

Assuming  $\epsilon_n$  to be a Gaussian error vector in the sparse Bayesian framework, it can be expressed as  $\epsilon_n \sim \mathcal{N}(0, \sigma^2)$ . The likelihood of the complete data set can be written as

$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{t} - \Phi(\mathbf{x})\mathbf{w}\|_2^2\right\} \quad (9)$$

where  $\Phi(\mathbf{x})$  is the design matrix, and  $\Phi(\mathbf{x}) = [\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \dots, \Phi(\mathbf{x}_N)]^T$ ,

$$\Phi(\mathbf{x}_n) = [\phi_0(\mathbf{x}_n), \phi_1(\mathbf{x}_n), \dots, \phi_{M-1}(\mathbf{x}_n)]^T.$$

In order to avoid overfitting and promote sparsity, the prior distribution of  $\mathbf{w}$  of Eq. (9) is assumed to be a zero-mean Gaussian distribution, and a separate hyperparameter  $\alpha_j$  is introduced for each  $w_j$ . The prior distribution of  $\mathbf{w}$  is given by

$$p(\mathbf{w}|\boldsymbol{\alpha}) = \prod_{j=0}^{M-1} \mathcal{N}(w_j|0, \alpha_j^{-1}) \quad (10)$$

where  $\alpha_j$  represents the precision of the corresponding parameter  $w_j$ .

Through the Bayesian inference, the distribution of the posterior parameter  $\mathbf{w}$  can be derived as follows:[44]

$$p(\mathbf{w}|\mathbf{t}, \mathbf{x}, \boldsymbol{\alpha}, \beta) = 2\pi^{-\frac{N+1}{2}} \left|\sum\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{w} - \mathbf{m})^T \sum (\mathbf{w} - \mathbf{m})\right\} \quad (11)$$

where the posterior parameter distribution is denoted by  $\mathbf{w} \sim \mathcal{N}(\mathbf{m}, \sum)$ , and  $\beta = \sigma^{-2}$ . The posterior mean  $\mathbf{m}$  and covariance  $\sum$  are as follows:

$$\mathbf{m} = \beta \sum \Phi^T \mathbf{t} \quad (12)$$

$$\sum = (\beta \Phi^T \Phi + \mathbf{A})^{-1} \quad (13)$$

with  $\mathbf{A} = \text{diag}(\boldsymbol{\alpha})$ .

The values of  $\boldsymbol{\alpha}$  and  $\beta$  are determined using the maximum likelihood method of the second kind.[46] In this method, the marginal likelihood function is maximized by integrating the weight vector. It can be expressed as:

$$p(\mathbf{t}|\mathbf{x}, \boldsymbol{\alpha}, \beta) = \int p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) p(\mathbf{w}|\boldsymbol{\alpha}) d\mathbf{w} \quad (14)$$

Directly maximizing the marginal likelihood function is computationally complex. Therefore, the logarithm of the marginal likelihood function is maximized, which can be expressed as:

$$\ln p(\mathbf{t}|\mathbf{x}, \boldsymbol{\alpha}, \beta) = -\frac{1}{2} \{N \ln(2\pi) + \ln |\mathbf{C}| + \mathbf{t}^T \mathbf{C}^{-1} \mathbf{t}\} \quad (15)$$

with  $\mathbf{C} = \beta^{-1} \mathbf{I} + \Phi \mathbf{A}^{-1} \Phi^T$ .

By setting the derivative of the marginal likelihood function to zero, the following reestimation equations are obtained:

$$\alpha_j^{new} = \gamma_j / m_j^2 \quad (16)$$

$$(\beta^{new})^{-1} = \frac{\|\mathbf{t} - \Phi \mathbf{m}\|^2}{N - \sum_j \gamma_j} \quad (17)$$

with  $\gamma_j = 1 - \alpha_j \sum_{ij}$ .

In the final step, the optimized parameters  $\boldsymbol{\alpha}$  and  $\beta$  are substituted into Eq. (12) and Eq. (13). The resulting posterior mean  $\mathbf{m}$  is considered to be the precise value for the parameter vector  $\mathbf{w}$ , which includes sub-pixel displacement and illumination change parameters.

For a more detailed explanation of the proposed SBL algorithm procedure, please consult Fig. 3. The process follows the steps outlined below.

- (1) Choose a template from the initial frame of the video;
- (2) Read the current frame of the video, then determine the search region for the current frame based on the position of the template from the previous frame;
- (3) Perform template matching based on correlation using Eq. (4);
- (4) If  $ZNCC_{\max} > 0.8$ , determine Integer-pixel displacements  $u$  and  $v$ , otherwise, expand the search region and go back to step 3;
- (5) Set initial values for  $\boldsymbol{\alpha}$  and  $\beta$ ;
- (6) Calculate the mean  $\mathbf{m}$  and covariance  $\sum$  of the posterior probabilities using Eq. (12) and Eq. (13);
- (7) Update hyperparameters  $\boldsymbol{\alpha}$  and  $\beta$  using Eq. (16) and Eq. (17);
- (8) Cycle steps 6 and 7 until reaching the maximum number of cycles or convergence;
- (9) Extract sub-pixel displacement from the mean vector  $\mathbf{m}$ .

The implementation of the proposed method, as presented in this study, is carried out using the Python 3.8 programming language in conjunction with the open-source computer vision library OpenCV 4.5.

### 2.3 Physical displacement estimation

When a full camera calibration process is employed, it can provide the essential camera distortion parameters required for correcting lens distortion in displacement measurements. However, directly correcting the raw video would impose a significant computational burden. Therefore, in this study, the proposed approach involves running the target tracking algorithm first, followed by the correction of the coordinate points representing the center position of the target in each frame. The correction process is as follows:

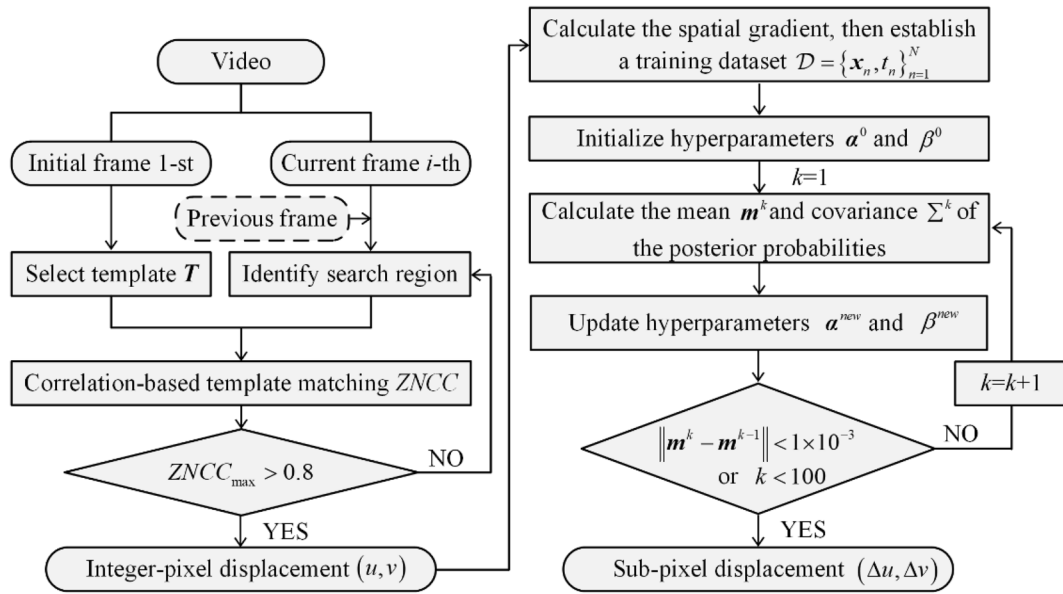


Fig. 3 Flowchart of the proposed SBL algorithm

$$\begin{aligned}
 x_c &= x(1 + k_1r^2 + k_2r^4 + k_3r^6) + 2p_1xy + p_2(r^2 + 2x^2) \\
 y_c &= y(1 + k_1r^2 + k_2r^4 + k_3r^6) + 2p_2xy + p_1(r^2 + 2x^2)
 \end{aligned}
 \tag{18}$$

where  $(x, y)$  are the image coordinates,  $(x_c, y_c)$  are the corrected coordinates,  $r$  is the Euclidean distance of the distorted point to the distortion image center,  $k_i$  and  $p_i$  are the distortion parameters.

Next, the image coordinates are transformed into world coordinates based on the camera calibration results. It is important to note that recovering the out-of-plane (Z-axis) coordinates of the measured structure using a single camera is theoretically impossible. [47] Therefore, in this study, the assumption is made that the Z value is a constant. Based on Eq. (1), the modified transformed equation when Z is set to 0 can be expressed as follows:

$$\begin{bmatrix} X \\ Y \\ s \end{bmatrix} = \begin{bmatrix} -r_{11} & -r_{12} & x' \\ -r_{21} & -r_{22} & y' \\ -r_{31} & -r_{31} & 1 \end{bmatrix}^{-1} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}
 \tag{19}$$

with  $[x' \ y' \ 1]^T = \mathbf{K}^{-1}[x \ y \ 1]^T$ .

To estimate the physical displacement, the initial template center coordinates are subtracted from the world coordinates of the center of the best match in each frame. This calculation yields the physical displacement change of the measured target, which can be expressed as follows:

$$\begin{bmatrix} dX^i \\ dY^i \end{bmatrix} = \begin{bmatrix} X^i \\ Y^i \end{bmatrix} - \begin{bmatrix} X^1 \\ Y^1 \end{bmatrix}
 \tag{20}$$

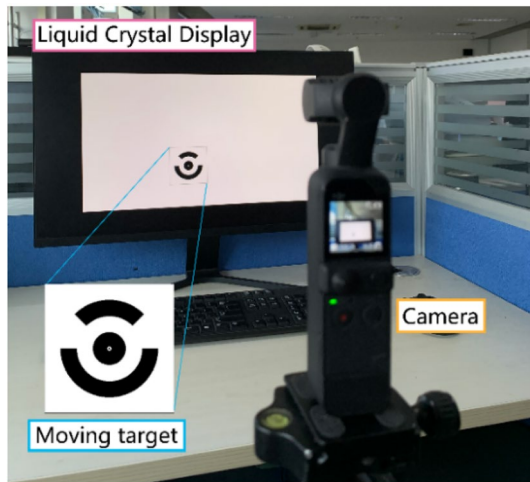
### 3 Verification

#### 3.1 Indoor experiment

To verify the robustness and computational efficiency advantages of the SBL algorithm, we conducted a novel experiment was conducted for dynamic displacement monitoring. The experiment involved the synthesis of a video depicting target movement, which was then played on a liquid crystal display (LCD). By leveraging the precise physical distance between adjacent pixels, the displacement of the moving target on the LCD could be accurately controlled. The VDDMS system was employed to measure the physical displacement of the moving target.

The experiment employed an RMMNT27NQ LCD model, featuring a resolution of  $2560 \times 1440$  pixels with a pixel pitch of 0.233 mm. The moving target consisted of a logo pattern containing a Schneider code [48], measured  $69.9 \times 69.9$  mm. The target exhibited horizontal motion at a speed of 4.66 mm/s over a duration of 10 s. As shown in Fig. 4, the experimental setup included compact entry-level action camera, DJI Pocket 2, featuring a resolution of  $1920 \times 1080$  pixels and a frame rate of 60 fps. The camera was positioned perpendicularly to the target, and the calibrated scale factor was established at 0.48 mm/pixel.

The experiment comprised four subtests, except for Subtests (a), each introducing specific variations. Subtests (b) and (c) involved gradual changes in the gray intensity of the moving target to simulate varying illumination conditions. In subtest (d), Gaussian noise with an increasing standard deviation up to 0.2 was applied to the target during



**Fig. 4** The indoor experimental setup

its movement. The variations in the moving target for each subtest are depicted in Fig. 5. All subtests use a consistent initial template of size  $145 \times 145$  pixels. The MTA algorithm [35], IC-GN algorithm [34], Kanade-Lucas-Tomasi (KLT) algorithm [49], and the SBL algorithm were employed for the four subtests. The MTA and IC-GN algorithms were initialized using the same ZNCC-based integer-pixel level displacement search process as the SBL algorithm. All experiments in this study were performed on a computer equipped with an AMD Ryzen 7 5800X @3.80 GHz CPU.

To evaluate the measurement error globally, the Root Mean Square Error (RMSE) was calculated using the following equation:

$$RMSE = \frac{1}{n} \sqrt{\sum_{i=1}^n (d_{gt} - d_v)^2} \quad (21)$$

where  $n$  is the number of displacement data points,  $d_{gt}$  are the ground truth of displacement values, and  $d_v$  is the value measured by the vision-based displacement sensors.

Figure 6 and Table 1 illustrate the error and computational speed of each target tracking algorithm. In subtest

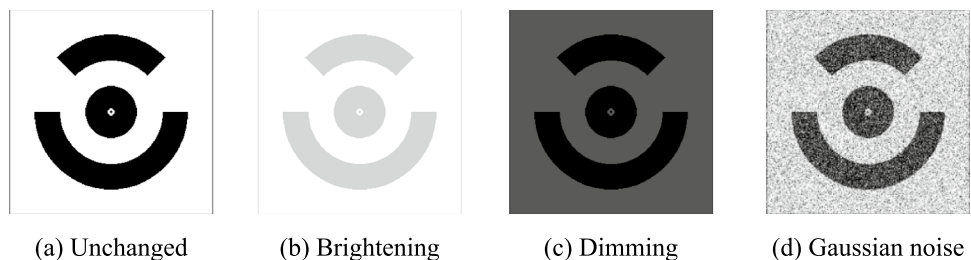
(a), the MTA, IC-GN, and SBL algorithms have consistent accuracies with an RMSE of 0.05 mm. The RMSE, which is 1/10 pixel when applying the inverse transformation of the scale factor, demonstrates the effectiveness of the algorithms employed in this experiment. The KLT algorithm exhibited an increasing error over time, with an RMSE of 0.34 mm. The KLT algorithm is prone to errors when tracking fast targets, and these errors gradually accumulate over time. This accumulation of errors is a result of calculating displacements between consecutive frames in the KLT algorithm[27].

In subtests (b) and (c), as the gray intensity variation of the moving object increases, the MTA algorithm is more affected, leading to a larger error range. The IC-GN algorithm show good robustness for small brightness changes, but when significant dimming occurs, the iterative optimization process of the IC-GN algorithm deviates from the correct direction, causing a sudden increase in error range. In contrast, the SBL algorithm consistently maintains the error within a controllable range regardless of the intensity variation. In subtest (d), the accuracy of the SBL algorithm decreases slightly but remains at a reasonable level of accuracy.

Among the four algorithms, the KLT algorithm stands out for its high computational efficiency, processing an average of 42.3 frames per second (fps). On the other hand, due to the interpolation and iterative calculations involved, the IC-GN algorithm has a significantly lower average computation speed of only 0.24 fps. The MTA and SBL algorithms exhibit comparable computation speeds of 16.4 and 15.8 fps, respectively. It is worth noting that the computation speed is influenced by the size of the template used. Opting for a smaller template size can enhance the speed of the SBL algorithm.

In summary, the SBL algorithm demonstrates greater robustness in terms of measurement accuracy across different subtests, while maintaining satisfactory computational efficiency. Its ability to consistently keep the error within a controllable range makes it a promising choice for vision-based displacement sensing technologies.

**Fig. 5** The variations in the moving target for each subtest





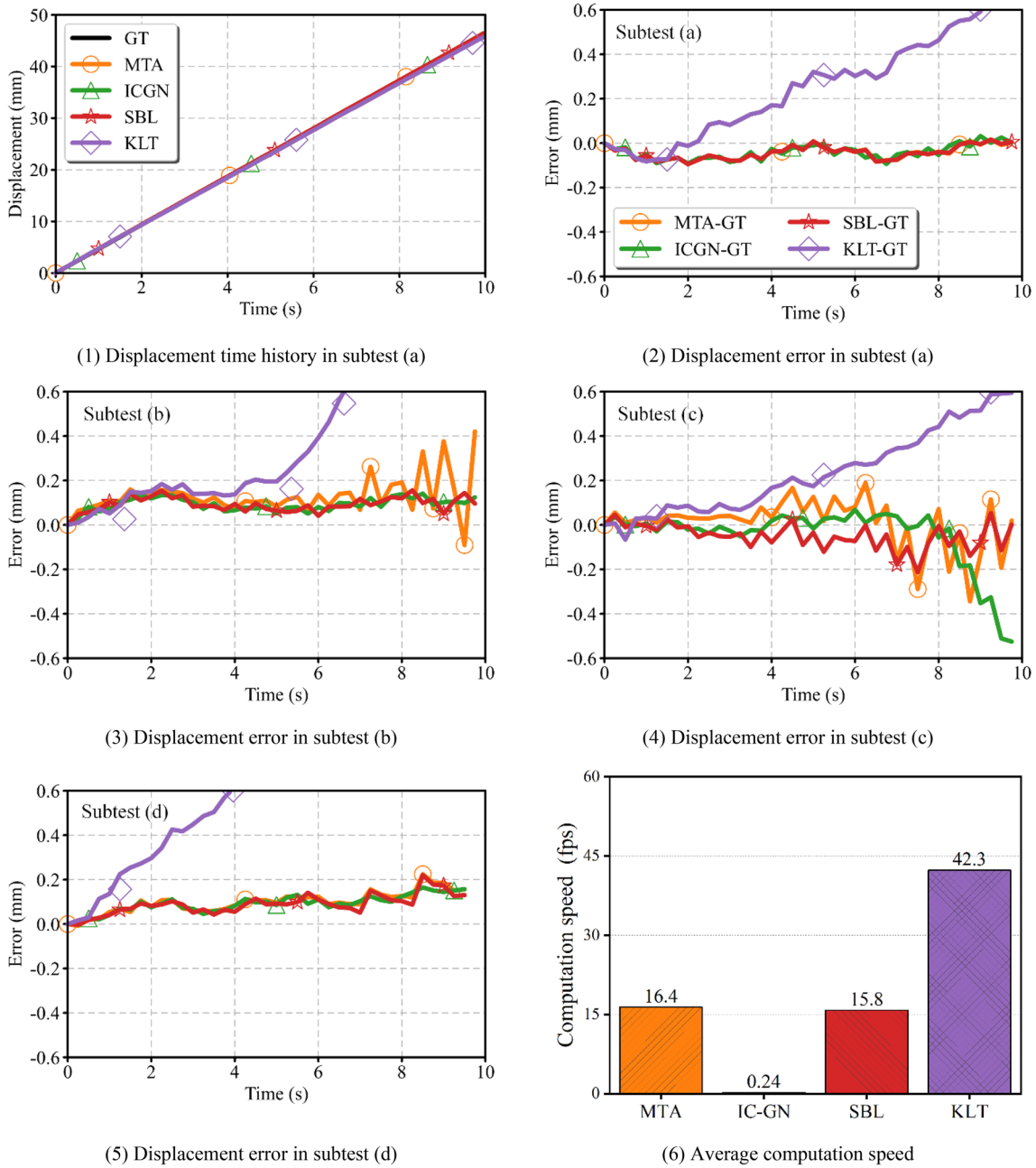


Fig. 6 Error and computational speed of each target tracking algorithm

Table 1 RMSE and computation speed of the different algorithms

Algorithm	RMSE (mm)				Average computation speed (fps)
	Subtest (a)	Subtest (b)	Subtest (c)	Subtest (d)	
SBL	0.05	0.10	0.08	0.10	15.8
ICGN	0.05	0.10	0.18	0.10	0.24
MTA	0.05	0.15	0.12	0.11	16.4
KLT	0.34	0.72	0.31	1.58	42.3

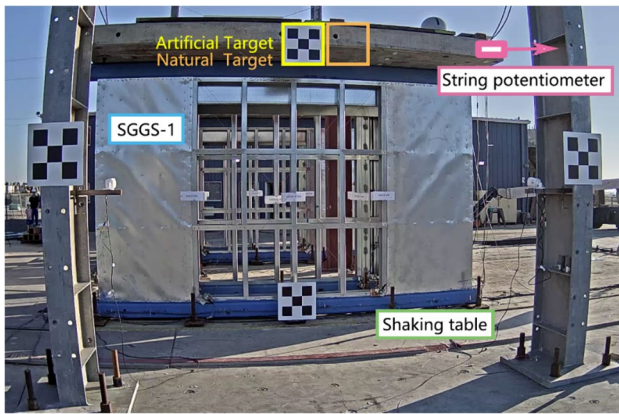


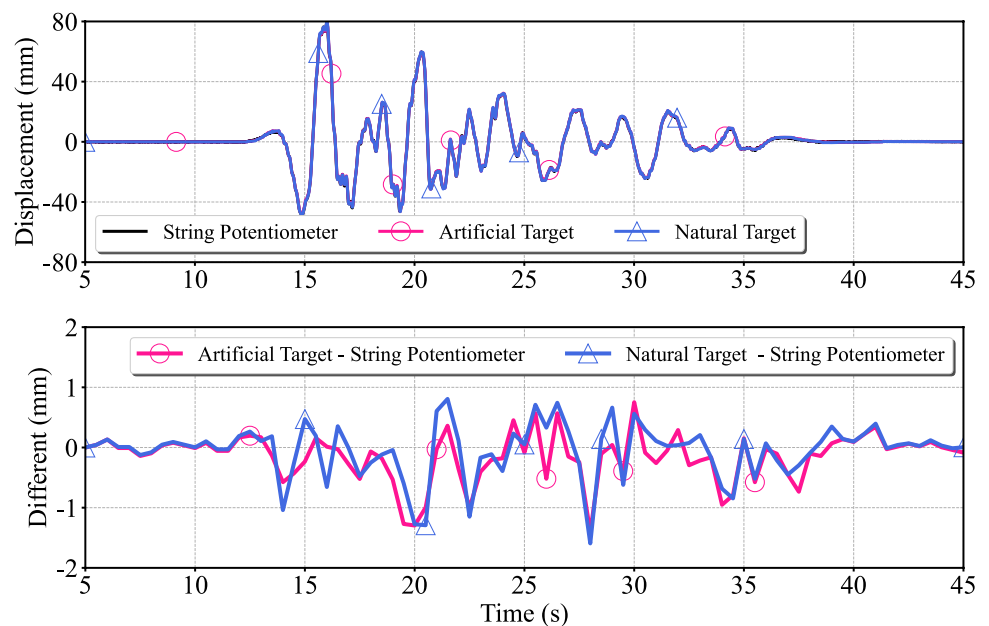
Fig. 7 The specimen SGGS-1 of Test Group 1 (Shot by the DVR)

### 3.2 Field experiment

In 2018, the University of California, San Diego conducted shake table tests to investigate the lateral response of steel sheet sheathed cold-formed steel framed in-line wall systems. The test details, reports, videos, and data can be found on DesignSafe [50]. Among the test specimens, SGGS-1 from Test Group 1 was a shear-gravity-gravity-shear wall line specimen measuring  $4.8 \times 2.7$  m, as shown in Fig. 7. To measure the wall drift, a string potentiometer was installed on the side face of the beam at the top of the specimen.

The “EQ1” test, which used the amplitude-modulated Canoga Park record component CNP196 of the 1994 Mw = 6.7 Northridge Earthquake, was selected for analysis. The specimen remained within the elastic range during the test. A fixed digital video recorder (DVR) camera placed south of the specimen recorded the dynamic test process.

Fig. 8 Displacement comparison between VDDMS and string potentiometer measurements at the top of the specimen



The captured video had a resolution of  $1920 \times 1080$  pixels, a frame rate of 30 fps, and a total of 2008 frames. VDDMS analyzed the video data to monitor the displacement of the specimen.

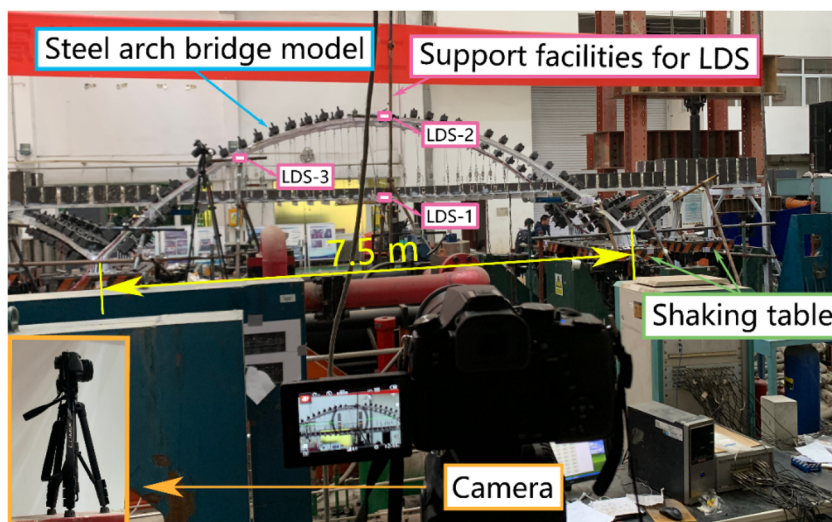
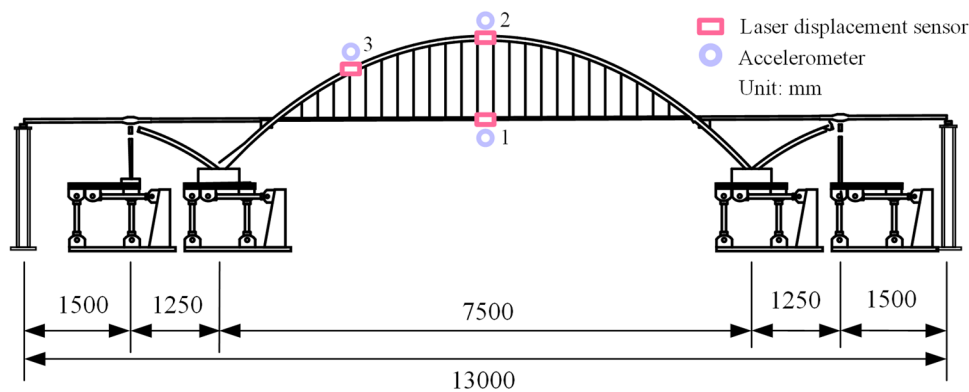
On top of the specimen, a concrete weight plate was equipped with a  $3 \times 3$  checkerboard, with each grid having a side length of 12 cm [51]. The corner coordinates of the checkerboard were extracted, and the planar homography matrix method was employed to estimate the projective transformation. Two adjacent initial templates, sized  $87 \times 87$  pixels, were selected on top of the specimen. One template contained artificial targets, while the other had natural targets.

Figure 8 illustrates the displacement of the specimen's top measured by VDDMS and the string potentiometer. The RMSE and computation speed of VDDMS are presented in Table 2, assuming that the measurement results of the string potentiometer are considered ground truth. The RMSE of the natural target measured by VDDMS was 0.37 mm, which is 2.7% higher than that of the artificial target. After applying the inverse projective transformation, the RMSE of the natural target was 1/12 pixel. For templates of the same size, VDDMS processed both natural and artificial targets at a

Table 2 RMSE and computation speed of VDDMS

Template	RMSE (mm)	RMSE (pixel)	Computation Speed (fps)
Artificial Target	0.36	0.08 (1/12)	39
Natural Target	0.37	0.08 (1/12)	39

**Fig. 9** Shaking table testing of large-scale steel arch bridge model



speed of 39 fps. Hence, the computation speed was found to be independent of the target type within the template.

## 4 Shaking table test of large-scale bridge

### 4.1 Experimental setup

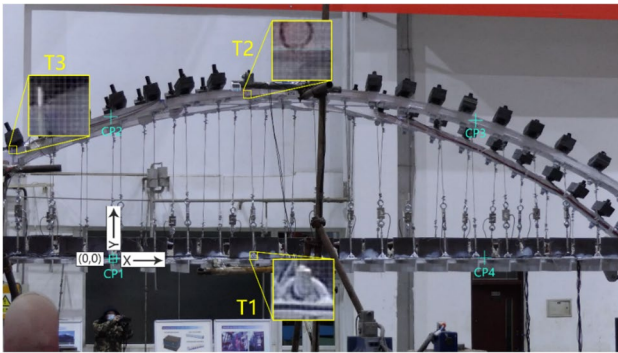
A shaking table test of a large-scale (1:40) steel box basket-handle arch bridge model was conducted in February 2022 at the Beijing University of Technology. VDDMS was used to monitor multiple points on the model bridge under various seismic excitations. As shown in Fig. 9, the bridge model features a main span of 7500 mm, two side spans of 1250 mm each, and is supported by 24 stay cables.

The shaker system consists of six small shakers with dimensions of 1 × 1 m. Table 3 provides details of the seismic excitation input scenarios used in the experiment, including the type of excitation, peak ground acceleration, and direction. Three LDS and accelerometers were installed at specific locations on the bridge model to measure longitudinal displacements. The response of

**Table 3** Seismic excitation cases

Case	Seismic excitation	Peak ground acceleration	Direction
0	White noise	0.075g	Longitudinal
1	Artificial wave 1	0.6 g	Longitudinal
2	Artificial wave 2	0.6 g	Longitudinal
3	Natural Chi-Chi wave	0.6 g	Longitudinal
4	Artificial wave 1	1.2 g	Longitudinal
5	Artificial wave 2	1.2 g	Longitudinal
6	Natural Chi-Chi wave	1.2 g	Longitudinal

the bridge model to seismic vibration was captured using the Panasonic Lumix DMC-FZ2500 camera, which is an entry-level consumer camera. The camera, positioned approximately 12 m away from the bridge model, was not precisely adjusted and had a noticeable tilt angle. The captured video had a resolution of 3840 × 2160 pixels and a frame rate of 30 fps, covering the vibration modes of most civil engineering structures.



**Fig. 10** The location of the initial templates

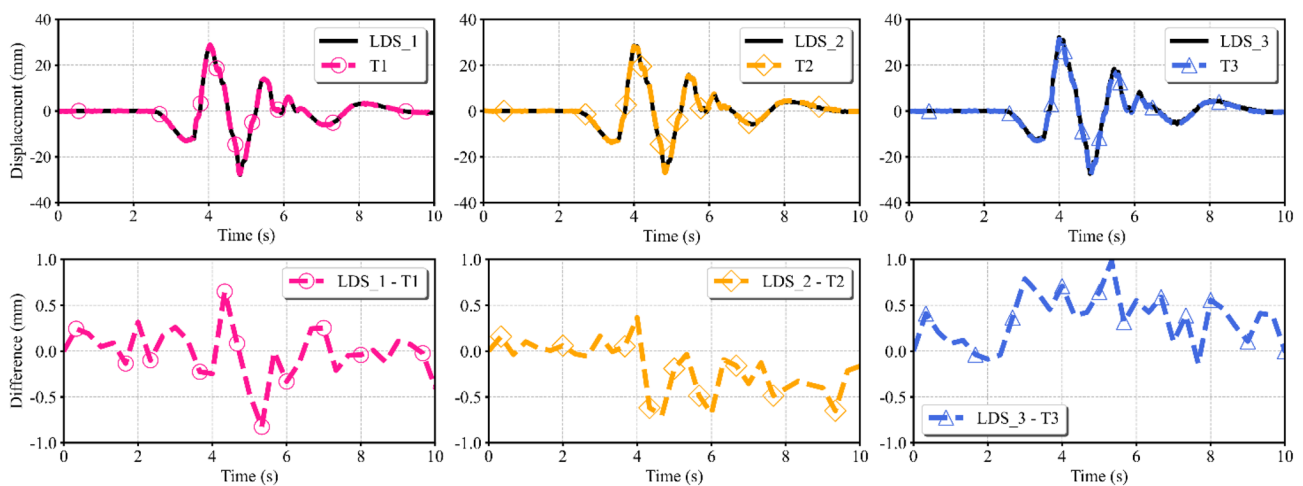
A complete camera calibration process was adopted in this experiment. The intrinsic and distortion parameters of the Panasonic Lumix DMC-FZ2500 camera were determined in advance by analyzing checkerboard images from different locations and orientations. The camera's focal length was locked, and the extrinsic parameters were determined by using four pairs of 2D-to-3D points. The distances between these control points were obtained from field measurements. A world coordinate system was established, with the X-axis aligned along the bridge span direction and the Y-axis in the vertical direction. Initial templates, sized  $51 \times 51$  pixels, were selected near the LDS measurement points (T1: mid-span of the bridge, T2: top of the arch rib, T3: 1/4 height of the arch rib), as show in Fig. 10. These templates captured the natural texture features on the structural surface. All pixel points within the templates are on the same plane and share a common motion trajectory.

## 4.2 Results of displacement

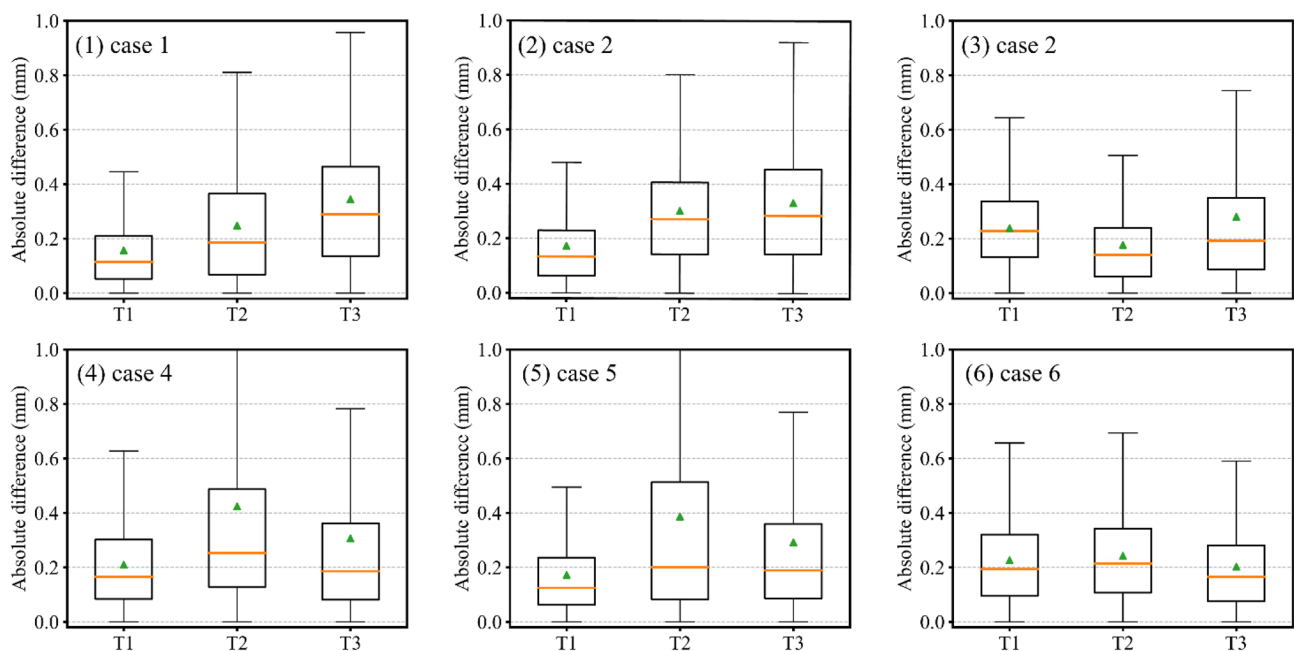
Figure 11 illustrates the displacement measurements of three targets obtained from both the VDDMS and LDS in case 1. To ensure accurate comparison, the signals from both measurement methods were aligned to a common reference time, and any minor time shifts were corrected using the maximum cross-correlation technique. The figure demonstrates a good agreement between the displacement measurements obtained from VDDMS and LDS. This indicates that the VDDMS is capable of accurately capturing the vibration displacements of the targets, comparable to the measurements obtained from the traditional LDS. Figure 12 show the distribution of absolute displacement differences measured by VDDMS in cases 1 to 6. Table 4 provides the Peak-to-peak (Pk-pk) and RMSE values of the VDDMS displacement measurements, assuming the measurement results of the LDS are considered ground truth.

One notable observation from the experiment is related to the Pk-pk displacement of different cases. Case 3 has a smaller Pk-pk displacement compared to Case 1 and Case 2. Similarly, Case 6 exhibits significantly smaller Pk-pk displacement than Case 4 and Case 5. The displacement amplitudes induced by artificial waves are larger than those caused by natural waves.

Another noteworthy observation pertains to measurement accuracy. Despite higher peak ground accelerations in the seismic excitation, resulting in faster and larger vibrations of the structure, the RMSE does not show a significant increase. Interestingly, the Normalized RMSE actually decreases noticeably. These indicates that the velocity of the target motion has a limited impact on the measurement accuracy of the VDDMS system. It can also be inferred that the primary errors in the VDDMS are fixed systematic errors, likely introduced during camera calibration.



**Fig. 11** Displacement measurements of three targets obtained from VDDMS and Laser Displacement Sensors (LDS) in Case 1



**Fig. 12** Distribution of absolute displacement differences measured by VDDMS and LDS from Case 1 to 6

**Table 4** Pk-pk and RMSE of displacement measured by the VDDMS

Case	Pk-pk (mm)			RMSE (mm)			Normalized RMSE		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
1	56	55	59	0.22	0.34	0.45	0.4%	0.6%	0.8%
2	47	48	52	0.23	0.38	0.41	0.5%	0.8%	0.8%
3	10	10	11	0.24	0.20	0.32	2.4%	2.0%	2.9%
4	109	108	114	0.26	0.64	0.43	0.2%	0.6%	0.4%
5	93	95	99	0.23	0.60	0.44	0.3%	0.6%	0.5%
6	19	19	21	0.27	0.30	0.27	1.5%	1.5%	1.3%

Note: Normalized RMSE = Pk-pk/RMSE

Furthermore, it is observed that the RMSE of target T3 is generally larger than the other two targets, indicating a larger error for the target at the edge of the image. This could be attributed to the greater degree of distortion for consumer cameras farther away from the image center. Additionally, different initial templates may have varying displacement measurement accuracies.

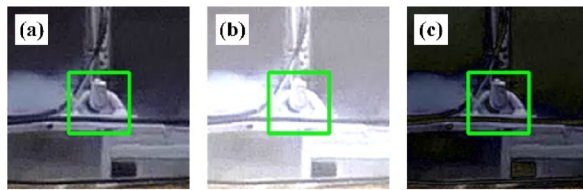
Across all measurement targets and seismic excitation conditions, the maximum and minimum values of RMSE are 0.64 mm and 0.20 mm, respectively, while the maximum and minimum values of Normalized RMSE are 2.9% and 0.2%. These results demonstrate that the VDDMS accurately monitors multiple targets within a large range on the steel arch bridge model under different seismic excitations, utilizing an entry-level consumer camera and the natural texture features of the structure's surface. Remarkably, this monitoring system does not require

precise camera position adjustments, making it practical and effective for real-world applications.

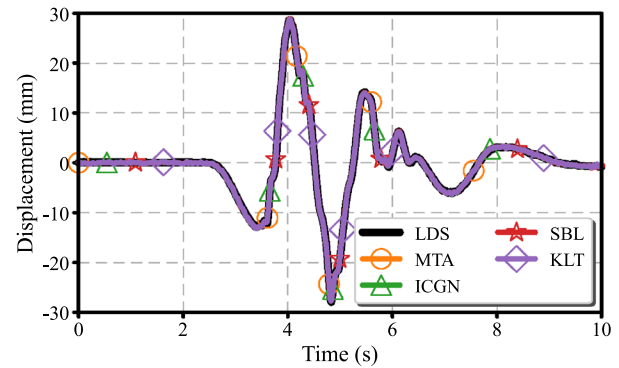
### 4.3 Illumination change robustness and computing efficiency

In this section, we evaluate the illumination robustness and computational efficiency of VDDMS, highlighting its advantages. To simulate the changing illumination during the movement of the bridge structure, we modify the grayscale values of frames in the recorded video from case 1. Three illumination conditions are considered: (a) no change, (b) brightened at 2nd seconds, and (c) dimmed at 2nd seconds. We measure the displacement of target T1 under these different illumination conditions.

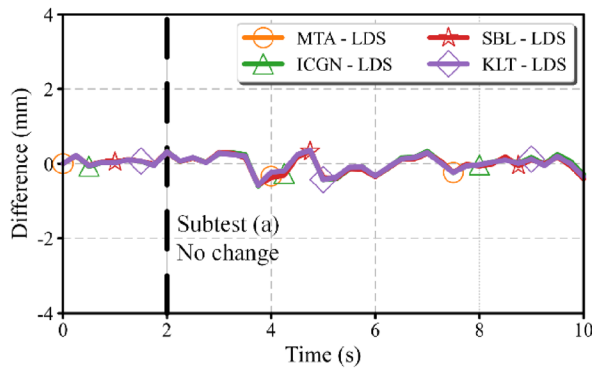
Figure 13 illustrate the difference from LDS and computation speed of VDDMS using four target tracking



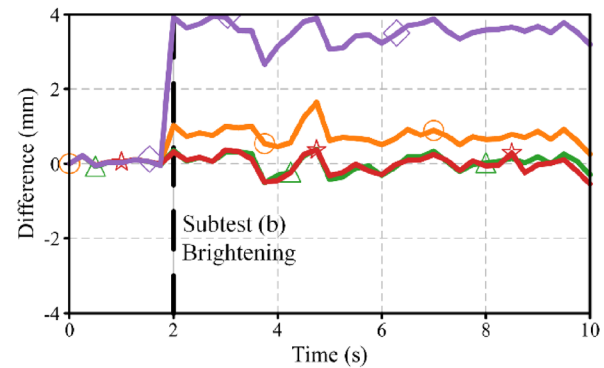
(1) Lighting conditions (a), (b), and (c)



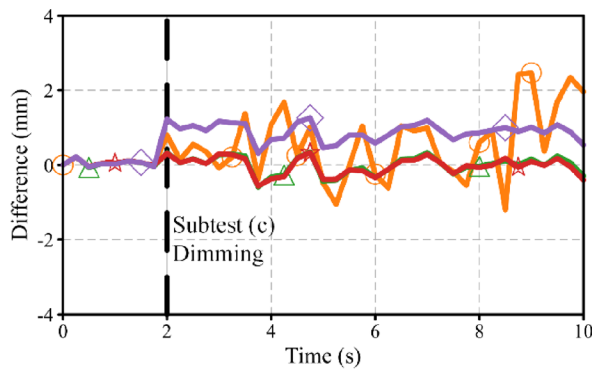
(2) Displacement time history under condition (a)



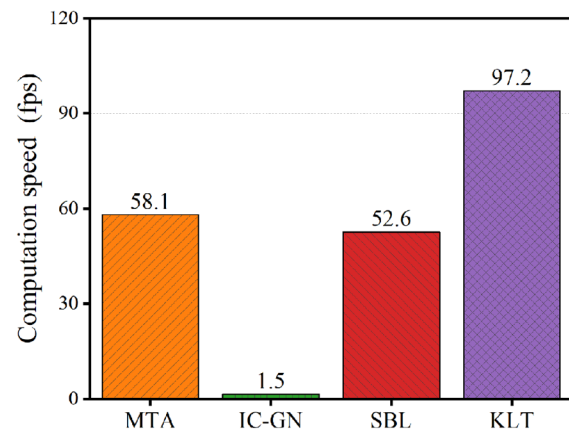
(3) Difference from LDS under condition (a)



(4) Difference from LDS under condition (b)



(5) Difference from LDS under condition (c)



(6) Average computation speed

**Fig. 13** Difference from LDS and computation speed of VDDMS using MTA, ICGN, SBL, and KLT algorithms for Target T1

algorithms: MTA, ICGN, SBL, and KLT. The corresponding RMSEs are listed in Table 5. The initial template plate size for each condition is set to  $51 \times 51$  pixels. Under illumination condition (a), the accuracy of VDDMS using all four algorithms is consistent, yielding an RMSE of 0.22 mm. However, for conditions (b) and (c), the MTA and KLT algorithms exhibit increased errors as the illumination becomes

brighter or darker, while IC-GN and SBL algorithms remain stable. The MTA algorithm calculates the displacement change between the initial frame and subsequent frames, resulting in measurement errors when there is a difference in illumination compared to the initial state. In contrast, the KLT algorithm computes displacement changes between adjacent frames, introducing errors only when there is a

**Table 5** RMSE and computation speed of VDDMS using MTA, ICGN, SBL, and KLT algorithms

Algorithm	RMSE(mm)			Average computation speed (fps)
	Condition (a)	Condition (b)	Condition (c)	
MTA	0.22	0.73	1.31	58.1
IC-GN	0.22	0.22	0.22	1.5
SBL	0.22	0.24	0.22	52.6
KLT	0.22	3.31	0.85	97.2

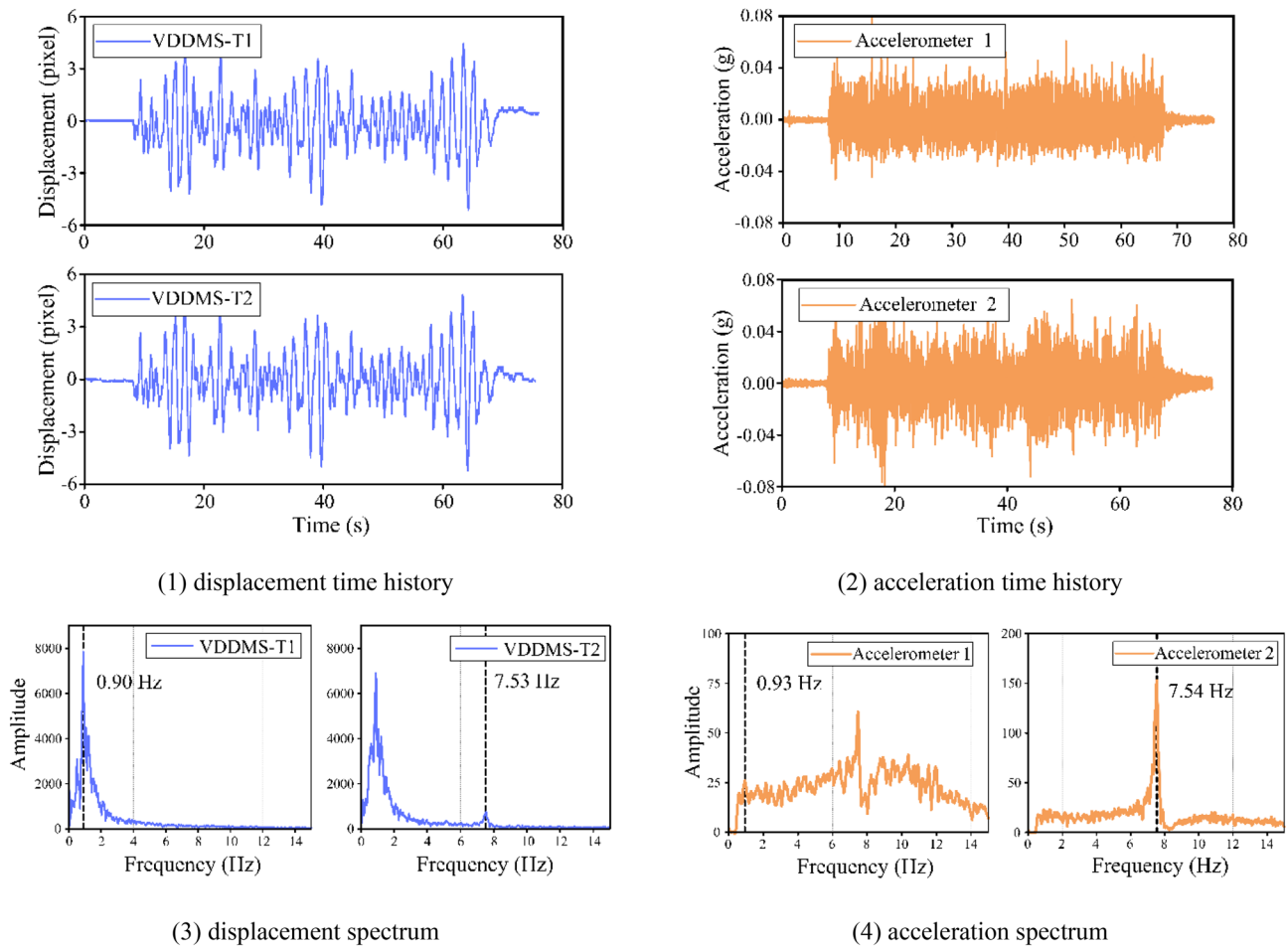
change in illumination, but these errors accumulate over time. The SBL and IC-GN algorithms exhibit higher illumination robustness.

In terms of computational speed, the IC-GN algorithm operates at a significantly slower pace, with approximately 1.5 fps. In contrast, the SBL algorithm achieves a much faster computation speed of 52.6 fps. Therefore, the proposed SBL algorithm demonstrates better illumination robustness and computational efficiency in the VDDMS system.

### 4.4 Identify structural dynamic characteristics

We conducted an analysis of the displacement spectrum of the model bridge subjected to white noise excitation. It is worth noting that the measured displacements do not require conversion into real physical displacement units. Figure 14 presents the measurements obtained from VDDMS and accelerometers under white noise excitation, along with the corresponding spectral analysis results. The first-order vibration mode of the model bridge is characterized by beam longitudinal drift, while the second-order vibration mode corresponds to arch transverse bending. Table 6 provides the frequencies of the model bridge obtained from both the accelerometer and VDDMS measurements. The frequencies measured by VDDMS closely match those obtained from the acceleration spectrum analysis, with values of 0.90 Hz and 7.53 Hz. Specifically, the VDDMS frequencies are lower by 0.01 Hz and 0.03 Hz compared to the acceleration-based results.

In conclusion, the displacement spectrum analysis demonstrates the reliability of VDDMS in accurately capturing



**Fig. 14** Spectral analysis comparison of VDDMS and accelerometer measurements under white noise excitation

**Table 6** Frequency comparison of model bridge obtained by accelerometer and VDDMS

Order	Frequency (Hz)			Measurement position
	VDDMS	Accelerometer	Difference	
1	0.90	0.93	0.03	The mid-span of the bridge (T1)
2	7.53	7.54	0.01	The top of arch ribs (T2)

the vibration frequencies of the model bridge under white noise excitation, with close agreement to the acceleration-based results.

#### 4.5 Initial template selection

The selection of the initial template is a crucial step in VDDMS, and this section investigates its influence on measurement accuracy. Five initial templates were chosen near the top of the arch ribs, and their information is presented in Table 7. The distinctiveness of template features was quantified using the Sum of the Square of Subset Intensity Gradients (SSSIG) [52] in the x and y directions, where higher SSSIG values indicate richer texture features within the template. VDDMS measured the displacement changes of five targets in case 5, and the error distribution is depicted in Fig. 15. The RMSE values for templates F1 and F3 were 0.64 mm and 0.59 mm, respectively, which were lower than templates F2 and F4. Template F5, lacking distinctive texture features, caused VDDMS to lose track of the target during the measurement process.

These findings emphasize the significance of template distinctiveness in influencing measurement accuracy. Templates with higher SSSIG values, indicative of richer texture features, exhibited lower RMSE values. Therefore, it is crucial to carefully select initial templates with distinctive texture features to ensure reliable target tracking throughout the measurement process.

#### 4.6 Error evaluation schemes

The accuracy of vision-based displacement sensing technologies is affected by various factors, including hardware devices, methods, environment, and so on [53]. In practical application, fast error evaluation is crucial. However, practical applications often lack comparable measurements from LDS, making it challenging to evaluate the reliability of vision-based displacement sensors. Hence, this study proposes two evaluation schemes for measuring the error of vision-based displacement sensors.

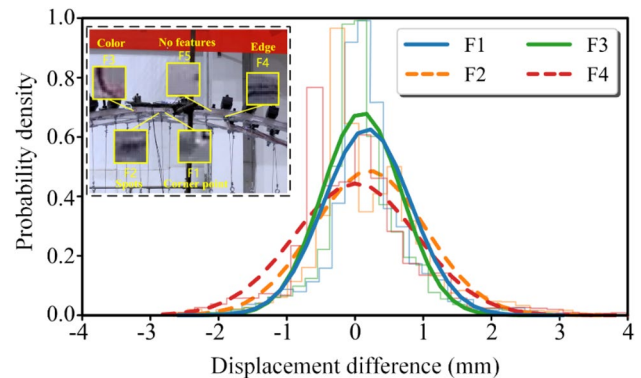
The first scheme involves extracting the displacements of measurement targets in the static state of the structure before the test, while the second scheme focuses on extracting the displacements of stationary background targets during the test. Since the actual displacement value of these targets should ideally be zero, any non-zero measurements can be

**Table 7** Information on the five initial templates

Templates	F1	F2	F3	F4	F5
Size (pixel)	21×21	21×21	21×21	21×21	21×21
Feature type	Corner	Spot	Color	Edge	Void
SSSIG	$1.3 \times 10^5$	$4.2 \times 10^3$	$1.2 \times 10^5$	$7.3 \times 10^3$	$1.5 \times 10^2$
RMSE (mm)	0.64	0.82	0.59	0.90	invalidity

considered as measurement errors. The first error evaluation scheme is designed to assess errors arising from hardware devices and algorithms, making it particularly suitable for short-term displacement measurements. On the other hand, the second scheme aims to evaluate errors attributed to environmental factors and is more applicable for long-term displacement monitoring purposes.

In this study, a video of the bridge in the static state before case 1 was captured, and the displacements of targets T1 to T3 were extracted using the first error evaluation scheme. Additionally, the displacements of the three stationary background targets BJ1 to BJ3 were extracted from the video of case 1. The results of the two error evaluation schemes are presented in Fig. 16 and Table 8. In the first scheme, the maximum RMSE is 1/7 pixels, and for the second scheme, the maximum RMSE is 1/10 pixel. It is observed that the measurement accuracy of T3 is lower than that of T1, consistent with the experimental results in the previous section. Furthermore, the second scheme indicates that the environmental factors had a minimal impact on measurement error, less than 1/10 pixel, in this experiment.

**Fig. 15** Displacement difference distribution of targets F1 to F5 as measured by VDDMS



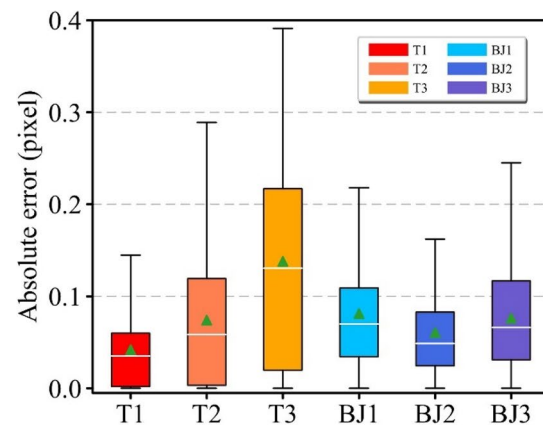
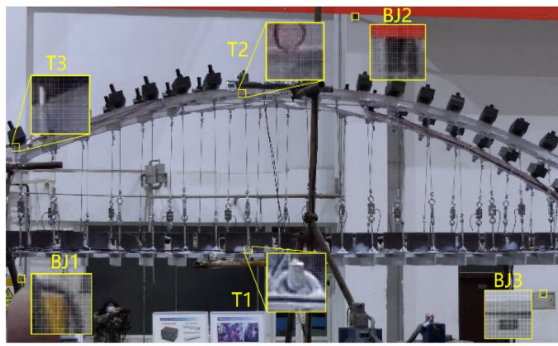


Fig. 16 Results of error evaluation schemes

Table 8 RMSE of error evaluation schemes

Targets	T1	T2	T3	B1	B2	B3
RMSE (pixel)	1/14	1/9	1/7	1/10	1/13	1/11

## 5 Conclusions

This research contributes to the advancement of vision-based sensing technologies in SHM. The robustness and versatility of VDDMS in different application scenarios are proven by a series of experiments. The key conclusions derived from this research can be summarized as follows:

1. The SBL algorithm demonstrates superior robustness in handling illumination changes compared to the KLT and MTA algorithms. It also offers a faster computational efficiency than the ICGN algorithm, approaching the high-efficiency MTA algorithm. The VDDMS accurately monitors natural targets on large-scale shear walls under outdoor conditions, achieving an RMSE of 0.37 mm, which is 3% higher than that of artificial targets.
2. By utilizing an entry-level consumer camera and leveraging the natural texture features of the structure's surface, VDDMS showcases its accurate monitoring capability for multiple targets on bridge models under various seismic excitations, eliminating the need for precise camera position adjustments. The RMSE, in comparison to laser displacement sensors, ranges from 0.2% to 2.9% of the peak-to-peak displacement. Furthermore, VDDMS effectively identifies the multi-order frequencies of the model bridge, which closely align with the results obtained from accelerometers.
3. The distinctiveness of initial template features is found to be a crucial factor influencing the accuracy of VDDMS. Furthermore, the proposed two error evaluation schemes can quickly evaluate the reliability of vision-based displacement sensing techniques, and they can be conveniently applied in field measurements.

These findings provide valuable insights for the future development of vision-based displacement sensing technologies. It is important to note that VDDMS may face challenges when natural texture features lack prominence, a common limitation in vision-based displacement sensing technologies. To overcome this limitation, future research should focus on extracting deeper target features, potentially through advanced deep learning techniques. Furthermore, the development of computer vision-based 3D displacement sensing techniques is necessary to address the challenges associated with 3D measurements in SHM applications.

**Acknowledgements** This work was supported by National Natural Science Foundation of China (No.52238011); and Beijing Municipal Education Commission (IDHT20190504). These supports are gratefully acknowledged. The results and conclusions presented in the paper are those of the authors and do not necessarily reflect the view of the sponsors.

**Data availability** Data will be made available on request.

## Declarations

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- Xu Y, Brownjohn JMW (2017) Review of machine-vision based methodologies for displacement measurement in civil structures. *J Civ Str Heal Monit* 8:91–110
- Wang M, Ao WK, Bownjohn J, Xu F (2022) A novel gradient-based matching via voting technique for vision-based structural displacement measurement. *Mech Syst Signal Proc* 171:108951
- Dong C (2019) Investigation of computer vision concepts and methods for structural health monitoring and identification applications
- Lydon D (2020) Development of a time-synchronised multi-input computer vision system for structural monitoring utilising deep learning for vehicle identification, Queen's University Belfast
- Ye XW, Dong CZ, Liu T (2016) A review of machine vision-based structural health monitoring: Methodologies and applications. *J Sensors* 2016:1–10
- Dong C, Catbas FN (2020) A review of computer vision-based structural health monitoring at local and global levels. *Str Health Monit* 20:692–743
- Wang M, Xu F, Xu Y, Brownjohn J (2023) A robust subpixel refinement technique using self-adaptive edge points matching for vision-based structural displacement measurement. *Comput Aided Civil Infrastr Eng* 38:562–579
- Wang M, Koo KY, Liu C, Xu F (2023) Development of a low-cost vision-based real-time displacement system using raspberry pi. *Eng Str* 278:115493
- Du W, Lei D, Hang Z, Ling Y, Bai P, Zhu F (2023) Short-distance and long-distance bridge displacement measurement based on template matching and feature detection methods. *J Civ Str Heal Monit* 13:343–360
- Wang Y, Brownjohn J, Jiménez Capilla JA, Dai K, Lu W, Koo KY (2021) Vibration investigation for telecom structures with smartphone camera: Case studies. *J Civil Str Health Monit* 11:757–766
- Tan D, Li J, Hao H, Nie Z (2023) Target-free vision-based approach for modal identification of a simply-supported bridge. *Eng Str* 279:115586
- Martini A, Tronci EM, Feng MQ, Leung RY (2022) A computer vision-based method for bridge model updating using displacement influence lines. *Eng Str* 259:114129
- Lado-Roigé R, Font-Moré J, Pérez MA (2023) Learning-based video motion magnification approach for vibration-based damage detection. *Measurement* 206:112218
- Peng Z, Li J, Hao H (2024) Computer vision-based displacement identification and its application to bridge condition assessment under operational conditions. *Smart Constr* 1(1). <https://doi.org/10.55092/sc20240003>
- Zhang S, Ni P, Wen J, Han Q, Du X, Fu J (2024) Intelligent identification of moving forces based on visual perception. *Mech Syst Signal Process* 214:111372
- Zhang S, Ni P, Wen J, Han Q, Du X, Xu K (2024) Non-contact impact load identification based on intelligent visual sensing technology. *Str Health Monitor* 0 14759217241227365
- Tian Y, Zhang C, Jiang S, Zhang J, Duan W (2020) Noncontact cable force estimation with unmanned aerial vehicle and computer vision. *Comput Aided Civil Infrastr Eng* 36:73–88
- Yoon H, Shin J, Spencer BF (2018) Structural displacement measurement using an unmanned aerial system. *Comput Aided Civil Infrastr Eng* 33:183–192
- Lydon D, Lydon M, Taylor S, Del Rincon JM, Hester D, Brownjohn J (2019) Development and field testing of a vision-based displacement system using a low cost wireless action camera. *Mech Syst Signal Process* 121:343–358
- Xin C, Wang C, Xu Z, Qin M, He M (2022) ker-free vision-based method for vibration measurements of rc structure under seismic vibration. *Earthquake Eng Str Dynam* 51:1773–1793
- Jana D, Nagarajaiah S (2021) Computer vision-based real-time cable tension estimation in dubrovnik cable-stayed bridge using moving handheld video camera. *Str Control Health Monitor*. <https://doi.org/10.1002/stc.2713>
- Sun C, Gu D, Zhang Y, Lu X (2022) Vision-based displacement measurement enhanced by super-resolution using generative adversarial networks. *Str Control Health Monitor*. <https://doi.org/10.1002/stc.3048>
- Cai E, Zhang Y, Quek ST (2023) Visualizing and quantifying small and nonstationary structural motions in video measurement. *Comput Aided Civil Infrastr Eng* 38:135–159
- Yang Y, Dorn C, Farrar C, Mascareñas D (2020) Blind, simultaneous identification of full-field vibration modes and large rigid-body motion of output-only structures from digital video measurements. *Eng Str* 207:110183
- Khuc T, Catbas FN (2017) Completely contactless structural health monitoring of real-life structures using cameras and computer vision. *Str Control Health Monitor* 24:e1852
- Yu S, Zhang J (2019) Fast bridge deflection monitoring through an improved feature tracing algorithm. *Comput Aided Civil Infrastr Eng* 35:292–302
- Dong C-Z, Bas S, Catbas FN (2020) Investigation of vibration serviceability of a footbridge using computer vision-based methods. *Eng Str* 224:111224
- Ye XW, Yi TH, Dong CZ, Liu T (2016) Vision-based structural displacement measurement: System performance evaluation and influence factor analysis. *Measurement* 88:372–384
- Luo L, Feng MQ (2018) Edge-enhanced matching for gradient-based computer vision displacement measurement. *Comput Aided Civil Infrastr Eng* 33:1019–1040
- Dong CZ, Ye XW, Jin T (2018) Identification of structural dynamic characteristics based on machine vision technology. *Measurement* 126:405–416
- Xiao P, Wu ZY, Christenson R, Lobo-Aguilar S (2020) Development of video analytics with template matching methods for using camera as sensor and application to highway bridge structural health monitoring. *J Civ Str Heal Monit* 10:405–424
- Brownjohn JMW, Xu Y, Hester D (2017) Vision-based bridge deformation monitoring. *Front Built Environ* 3:23
- Feng D, Feng MQ (2016) Vision-based multipoint displacement measurement for structural health monitoring. *Str Control Health Monit* 23:876–890
- Pan B, Tian L, Song X (2016) Real-time, non-contact and targetless measurement of vertical deflection of bridges using off-axis digital image correlation. *NDT E Int* 79:73–80
- Zhang D, Guo J, Lei X, Zhu C (2016) A high-speed vision-based sensor for dynamic vibration analysis using fast motion extraction algorithms. *Sensors* 16:572
- Wang M, Kei Ao W, Bownjohn J, Xu F (2022) Completely non-contact modal testing of full-scale bridge in challenging conditions using vision sensing systems. *Eng Str* 272:114994
- Zhang Z (2000) A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Mach Intell* 22:1330–1334
- Hartley RI, Zisserman A (2003) Multi-view geometry in computer vision

39. Wang J, Zhao J, Liu Y, Shan J (2021) Vision-based displacement and joint rotation tracking of frame structure using feature mix with single consumer-grade camera. *Str Control Health Monit*. <https://doi.org/10.1002/stc.2832>
40. Lee G, Kim S, Ahn S, Kim H-K, Yoon H (2022) Vision-based cable displacement measurement using side view video. *Sensors* 22:962
41. Kumar D, Chiang CH, Lin YC (2022) Experimental vibration analysis of large structures using 3d dic technique with a novel calibration method. *J Civ Str Heal Monit* 12:391–409
42. Pan B, Xie H, Wang Z (2010) Equivalence of digital image correlation criteria for pattern matching. *Appl Opt* 49:5501
43. Barron JL, Fleet DJ, Beauchemin SS (1994) Performance of optical flow techniques. *Int J Comput Vision* 12:43–77
44. Tipping ME (2001) Sparse bayesian learning and the relevance vector machine. *J Mach Learn Res* 1:211–244
45. Wang Q-A, Dai Y, Ma Z-G, Ni Y-Q, Tang J-Q, Xu X-Q, Wu Z-Y (2022) Towards probabilistic data-driven damage detection in shm using sparse bayesian learning scheme. *Str Control Health Monit* 29:e3070
46. Bishop CM, Nasrabadi NM (2006) *Pattern recognition and machine learning*, Springer
47. Prince S (2012) *Computer vision: Models, learning and inference*, Cambridge University Press
48. Schneider C, Sinnreich K (1993) Optical 3-d measurement systems for quality control in industry. *Int Arch Photogramm Remote Sens* 29:56–56
49. Dong C, Celik O, Catbas FN (2018) ker-free monitoring of the grandstand structures and modal identification using computer vision methods. *Str Health Monit* 18:1491–1509
50. Singh A, Wang X, Zhang Z, Derveni F, Castaneda H, Peterman KD, Schafer BW (2022) Hutchinson, Steel sheet sheathed cold-formed steel framed in-line wall systems. I: Impact of structural detailing. *J Str Eng* 148:04022193
51. Wang X, Lo E, De Vivo L, Hutchinson TC, Kuester F (2022) Monitoring the earthquake response of full-scale structures using uav vision-based techniques. *Str Control Health Monit* 29:e2862
52. Pan B, Xie H, Wang Z, Qian K, Wang Z (2008) Study on subset size selection in digital image correlation for speckle patterns. *Opt Express* 16:7037–7048
53. Feng D, Feng MQ (2018) Computer vision for shm of civil infrastructure: From dynamic response measurement to damage detection – a review. *Eng Str* 156:105–117

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.