WILEY | Hindawi

Research Article

An Improved Structural Displacement Monitoring Approach by Acceleration-Aided Tilt Camera Measurement

Tong Wu [ID],[1,2] **Liang Tang** [ID],[1] **Xiangyu Zhang** [ID],[1] **Yijun Liu** [ID],[1] **Xinyu Li** [ID],[1] **and Zhixiang Zhou** [ID][2]

[1]*State Key Laboratory of Mountain Bridge and Tunnel Engineering, Chongqing Jiaotong University, Chongqing 400074, China*
[2]*College of Civil and Transportation Engineering, Shenzhen University, Shenzhen 518061, China*

Correspondence should be addressed to Zhixiang Zhou; zhixiangzhou@szu.edu.cn

Computer vision is becoming one of the most popular remote-sensing techniques and has been used widely in displacement monitoring and damage identification of in-service bridges. Nevertheless, several obstacles, including limited sampling rate, insufficient resolution for remote measurement, and error induced by camera tilting, restrict the application of vision-based approaches in structural health monitoring (SHM). The combination of a traditional SHM system and a modern remote-sensing technique can significantly improve the accuracy and reliability of the monitoring system. To make full use of data collected in the traditional SHM system and computer vision technique and overcome their shortcomings, we presented an improved bridge displacement estimation approach for SHM purposes by fusing camera-based and acceleration measurements. First, we estimated the scaling factor, which transfers pixel displacement to real displacement under tilt photogrammetry, by the acceleration reconstructed and camera-based displacements in the same frequency band without the actual size of the structure or the measurement parameters. Then, we extracted the low-frequency displacement from the vision-based measurement, and we fused the high-frequency displacement that was reconstructed from the acceleration measurement to achieve high-accuracy displacement estimation. The efficiency of this method was validated through dynamic load tests on a suspension model bridge in the laboratory and field tests on a highway and subway cable-stayed bridge.

1. Introduction

Displacement response under external loads is vital for civil infrastructures such as bridges in structural health monitoring (SHM) and structural condition assessments. Traditional displacement monitoring systems are based on networks of wired or wireless contact sensors, such as linear variable differential transformers (LVDTs) and dial indicators [1]. These contact sensor-based displacement monitoring systems are inconvenient because a platform under the bridges is needed as the benchmark. Furthermore, the cost of the maintenance of these sensors is comparatively expensive because of the harsh environment they are exposed to.

In recent years, computer vision-based methods have emerged as an urgent requirement for remote and noncontact monitoring. With the advantages of remote sensing, cost-effectiveness, easy installation, and multipoint synchronous measurements, vision-based techniques have been studied widely in lab-scale experiments and field applications. In these methods, the cameras are installed far away from the target structures to track the deformation associated with the manually attached targets or the natural targets [2]. For example, Feng and Feng [3] validated the cost-effectiveness of the vision-based method according to laboratory experiments on simply supported beams and field tests of the Manhattan Bridge. Yu and Zhang [4] improved the feature tracing algorithm using a simplified fast-Hessian detector and a prepurification-based RANdom sampling consistency (RANSAC), which was applied to field testing of an arch bridge with a 100 m main girder. Luo et al. [5] developed a robust vision sensor through a gradient-based

template matching algorithm and conducted a series of experiments and field tests to validate the efficiency. Jiang et al. [6, 7] developed a robust vision-based algorithm for tracking line-like structures in a noisy background. These examples have confirmed the potential for computer vision in structural displacement monitoring.

To obtain accurate displacement, a necessary step before conducting measurement is camera calibration to identify the projection relation between three-dimensional (3D) world coordinates and two-dimensional (2D) image coordinates as well as to eliminate image distortion. The general method for camera calibration is called perspective transformation, in which the intrinsic parameters such as focal length and external parameters such as camera pose can be obtained. The most commonly used camera calibration method was proposed by Zhang [8], in which multiple images are taken from a black-and-white chessboard at various locations and orientations. Furthermore, Kim et al. [9] proposed the use of a side-view camera for cable displacement monitoring. The information related to the pose of the camera was obtained by the estimation of camera projection. In displacement measurement of civil infrastructures, the standard camera calibration method can be simplified according to the properties of the camera, lens, and motion of the target. When motion occurs in a 2D plane, the calibration can be realized by solving the homography matrix. For example, Xu et al. [10] calibrated a consumer-grade camera by solving the homography matrix and measured the displacement of a cable-stayed footbridge. Wu et al. [11] developed a vision-based displacement sensing system for 2D displacement monitoring of the experimental structures and employed the homography matrix to reconstruct the space coordinates from the image coordinates. The camera calibration procedure could be further simplified by limiting the displacement in one direction, such as the vertical displacement of bridges. A parameter called the scaling factor (SF) was introduced for images to physical displacement transformation. SF can be established based on the relationship between image size and corresponding physical size. If the optical axis of the camera is perpendicular to the object surface, SF can be estimated accurately when the focal length, measurement distance, pixel or image size, and corresponding physical size are known. If the optical axis is not perpendicular to the object surface, and the optical axis has a tilt angle to the normal line of the object plane, this angle should be considered in the estimation. A formula for angle correction was proposed by Dong et al. [12]. Feng et al. [13] discussed the influence of tilt angle on the measurement accuracy and noted that the error increased with an increase in the tilt angle, and they proposed a simplified expression. Another SF estimation method was proposed by Khuc and Catbas [14, 15], in which the checkerboard was employed to calibrate the camera before tests, and the displacement could be measured without using a physical target on structures.

The efficiency of computer vision technology has been widely validated both in experimental and field tests. Nevertheless, several practical drawbacks are still associated with applying vision-based sensors in monitoring displacement in the field [5, 16]. First, measurement accuracy depended significantly on measurement distance and light conditions. For a camera with a fixed setting, the targets of high-resolution and large field-of-view (FOV) measurements were mutually conflicting. Simultaneous spatial dense displacement monitoring of a large-scale structure with a single camera may not be practical unless the resolution of the video is decreased, which may cause inaccurate monitoring outcomes [17]. Second, the framerates of the mainstream consumer-grade cameras in the market are limited to 20-60 Hz, which might prohibit the ability to record the high-frequency dynamic motion of the civil engineering structures and could cause aliasing issues according to the sampling theorem. Finally, the optical axis of the camera should be perpendicular to the object plane; otherwise, an error would be introduced when pixel-to-physical coordinate conversion is performed. The ideal placement of a camera is difficult to identify in the field, and thus, the target structure must be monitored at a large pitch angle. Furthermore, it is inconvenient to measure the distance from the camera to the target structure as well as the structural size corresponding to the image size, both of which are essential to complete traditional SF calculations [13].

The measurement of acceleration is simpler and more convenient than displacement monitoring because no reference is required. Therefore, accelerometers are essential sensors in SHM systems. Studies about displacements that have been reconstructed by measured accelerations also have been reported [18, 19]. For example, Lee et al. [18] proposed a type of finite impulse response (FIR) filter in the time domain to reconstruct dynamic displacements from the measured accelerations. The reliability of the ultra-low-frequency response of the accelerometers is poor, however, and thus the reconstructed displacements in ultra-low-frequency are inaccurate [20, 21]. Nevertheless, for bridges, the quasistatic displacements caused by the moving load are a vital parameter for SHM. Therefore, it is generally considered that the high-frequency displacement estimated by the measured acceleration is more reliable than the low-frequency displacement. To make full use of the advantages of acceleration-based measurement, data fusion of the high-frequency displacement estimated by acceleration and low-frequency displacement measured by other sensors has been proposed to obtain an accurate displacement [22–30]. Ma et al. [22] estimated displacement using measured strains and proposed an FIR-based displacement estimation technique by fusing strain and acceleration measurements. Then, Ma et al. [23, 24] used an adaptive multirate Kalman filter to fuse an asynchronous acceleration and take computer vision measurements. Park et al. proposed the use of a pair of complementary filters and a time synchronization method to fuse vision-based displacement and acceleration [27]. Xu et al. [28, 29] fused the GPS-based and vision-based displacements with the acceleration measurements using a multirate Kalman filter, and the efficiencies were validated on a railway bridge. Wu et al. [30] estimated the full-field displacements of beam structures by data fusion of vision-based displacement and accelerations and modal expansion.

These data fusion instances as well as other vision-based displacement measurement technologies have been conducted on the condition that the optical axis of the camera is perpendicular to the object plane or that the tilt angle is small, which has restricted these applications.

In this study, to overcome the previous drawbacks and realize tilt camera-based monitoring, we proposed to introduce accelerations to improve the accuracy of vision-based displacements. First, we estimated the SF associated with pixel-physical displacement transformation under tilt photogrammetry using the camera and acceleration-based measurements under the same frequency band without the actual size of the structure or the measurement parameters. Then, the low-frequency displacement estimated from the vision-based measurements was fused with the high-frequency acceleration-based displacement using the complementary filters. Finally, the feasibility of the proposed approach was validated on a lab-scale suspension bridge model in the laboratory and a highway and subway cable-stayed bridge in the field.

2. Data Fusion of Tilt Camera-Based and Acceleration Measurements

An overview of the proposed displacement estimation approach is shown in Figure 1. The proposed approach had four main steps: first, we analyzed the raw videos acquired from the tilt camera and extracted the pixel displacements of the features. Second, we estimated the dynamic displacements of target points from the measured accelerations by double-numerical integration. Third, we estimated the SF associated with pixel-physical displacement transformation using the vision-based and acceleration-based displacements in the same frequency band. Finally, we obtained the high-fidelity displacement by data fusion of the low-frequency vision-based displacement and the high-frequency acceleration-based displacement.

2.1. Pixel Displacement Extraction from the Camera.
As shown in Figure 2, the camera was fixed outside the bridge, facing the bridge deck at a tilt angle. A natural target (i.e., holes on the concrete surface, rivets, and bolts on the steel structure surface) of the deck was involved in the camera's FOV. We used the feature tracking algorithm to extract the pixel displacement of the target. This method had four main steps.

First, we selected a region of interest (ROI) containing the target in the initial frame of the video. The location and size of the ROI were based on that of the target, and the approximate deformation range of the target should be considered as well to ensure that the size of the ROI was slightly larger than the motion range of the target. We conducted feature points detecting and tracking in the selected ROI so that the computational efficiency could be greatly reduced. We also selected ROI because the stationary points with robust features in the background could be prefiltered out to ensure that the tracked points were motional ones. We detected feature points $p_1^1, p_1^2, \ldots, p_1^M$ in the

first frame through the speed-up robust features (SURFs) [31] and described these points using the binary robust invariant scalable key point (BRISK) [32]. Then, we applied the feature matching between the first and i^{th} frames and matched the features $p_1^1, p_1^2, \ldots, p_1^M$ in the ith frames with $p_i^1, p_i^2, \ldots, p_i^M$, respectively, through the k-nearest neighbor (KNN) algorithm. The total number of M matches is shown in Figure 2. Next, we adopted RANSAC [33] to filter out the incorrect matching results. Finally, we calculated the vertical pixel displacement of the target in the ith frame through the valid matches, which can be expressed as follows [14]:

$$u_{pi}^i = \frac{1}{M} \sum_{j=1}^{M} u_j^i, \tag{1}$$

where u_{pi}^i is the vertical pixel displacement of the target in the ith frame; M is the number of valid matches; and u_j^i is the vertical pixel displacement of the jth match.

2.2. Dynamic Displacement Reconstruction from the Measured Acceleration.
Provided that the measured acceleration is available, we used the acceleration-based displacement estimation method proposed by Lee et al. [18] to solve the following optimization problem in a time window $[(k - N)\Delta t, (k + N)\Delta t]$:

$$\min_{u_a} \Pi_E(\mathbf{u}_a) = \frac{1}{2} \|\mathbf{L}\mathbf{u}_a - (\Delta t)^2 \mathbf{L}_a \mathbf{a}\|_2^2 + \frac{\mu^2}{2} \|\mathbf{u}_a\|_2^2, \tag{2}$$

where $\mathbf{a}$ and $\mathbf{u}_a$ are the vectors of the measured acceleration and the displacement estimated by the measured acceleration, respectively; $\mathbf{L} = \mathbf{L}_a \mathbf{L}_c$, where $\mathbf{L}_a$ and $\mathbf{L}_c$ are diagonal weighting matrices of order $(2k + 1)$ with all diagonal entries equal to 1 except the first and the last entries, which are equal to $1/\sqrt{2}$ and a second-order differential operator. $\|\|_2$ denotes the 2-norm of a vector; Δt is the time interval; and μ is the optimal regularization parameter to adjust the degree of the regularization in the minimization problem and can be defined as follows:

$$\mu = 46.81 (2N + 1)^{-1.95}. \tag{3}$$

The vector of displacement $\mathbf{u}_a$ estimated by the acceleration can be obtained by solving equation (2) as follows:

$$\mathbf{u}_a = \left(\mathbf{L}^{\text{T}}\mathbf{L} + \mu^2 \mathbf{I}\right)^{-1} \mathbf{L}^{\text{T}} \mathbf{L}_a \mathbf{a} (\Delta t)^2 = \mathbf{C}\mathbf{a}(\Delta t)^2, \tag{4}$$

where $\mathbf{I}$ is the identity matrix of order $(2k + 3)$ and $\mathbf{C} = (\mathbf{L}^{\text{T}}\mathbf{L} + \mu^2 \mathbf{I})^{-1} \mathbf{L}^{\text{T}} \mathbf{L}_a$ is the coefficient matrix. Because the accuracy is highest at the center data point of the time window, we applied a moving overlapping time-window technique whose length was $(2N + 1)$ and retained only the displacement estimated at the center of each time window while others were discarded.

To realize real-time reconstruction, the time window is suggested to be as short as possible. However, a short time window cannot ensure the reconstruction error has damped out sufficiently at the center point. According to Lee et al. [18], the time window length should be set as the shortest length that does not affect the reconstruction accuracy.
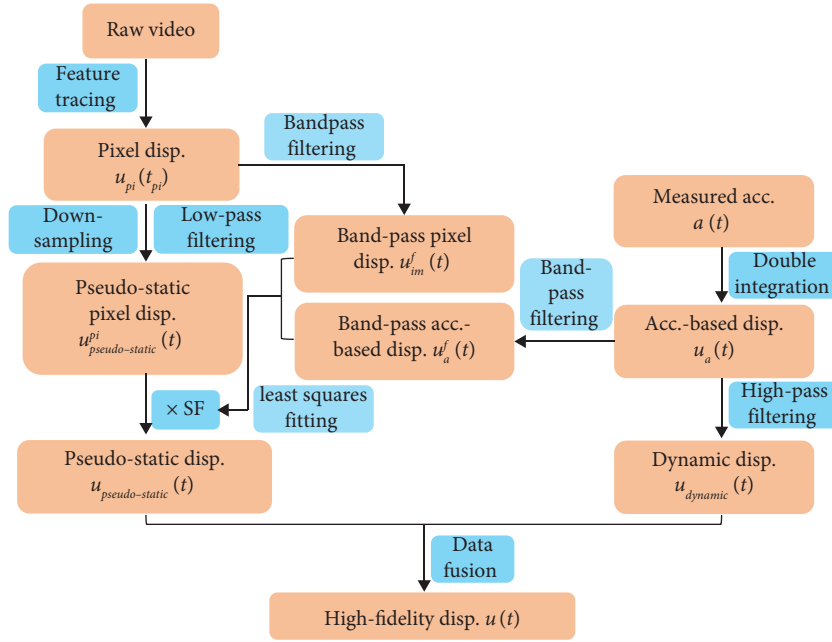
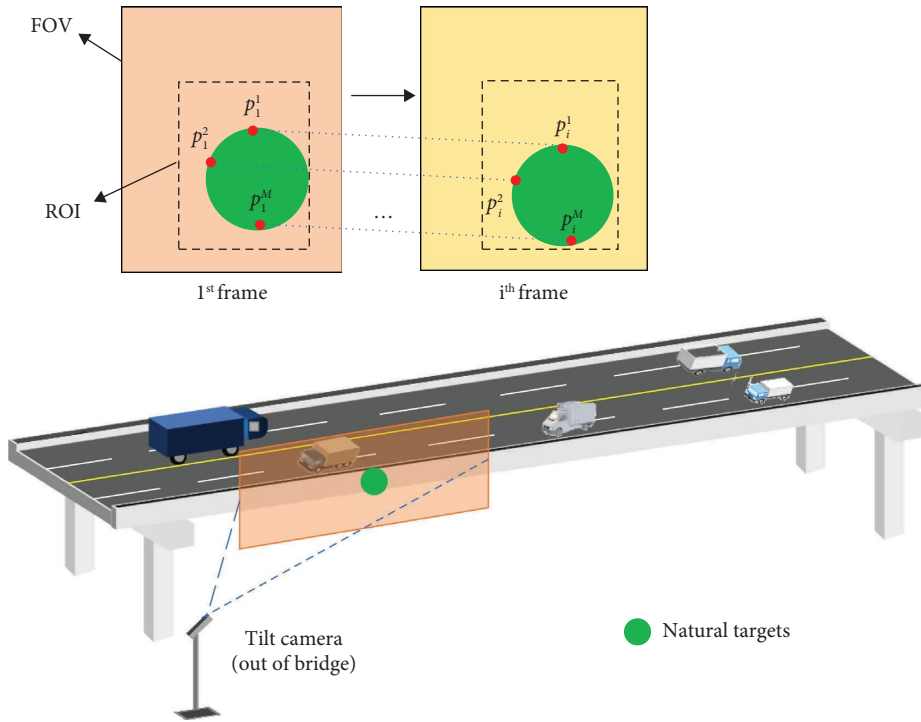FIGURE 1: The overview of the proposed displacement estimation approach.

FIGURE 2: Displacement monitoring by the tilt cameras.

Moreover, the optimal window size lies in the range of 2.22 to 3 times of the longest period of interest. To obtain the optimal window size, a detail discussion is given in Section 3 through the numerical simulations.

2.3. Scaling Factor Estimation. The extracted pixel displacement should be converted to real displacement before data fusion with the acceleration-based dynamic displacement. Because the displacement of bridges is mainly vertical, we used a simplified camera calibration method (i.e., the SF method) to realize the conversion from pixel to physical displacements.

The traditional approaches to determine the SFs are based on the establishment of the relationship between the pixel coordinate and the physical coordinate (e.g., with units of mm/pixel), as follows [12, 13]:

$$\text{SF} = \frac{D}{d}, \tag{5}$$

$$\text{SF}_1 = \frac{F}{f} d_{\text{pixel}}, \tag{6}$$

$$\text{SF}_2 = \frac{F}{f \cos^2 \theta} d_{\text{pixel}}, \tag{7}$$

where D and d are the known physical dimension on the object plane and the corresponding pixel numbers in the image plane, respectively; F and f are the distance between the camera and the object and the focal length, respectively; d_{pixel} is the pixel size; and θ is the tilt angle of the camera.

The SFs can be calculated once F, f, and θ are determined both in the perpendicular and nonperpendicular scenarios according to equations (5)–(7), and it is the same for all the pixels in the vertical direction. Nevertheless, according to Reference [13], the measurement error of vision-based displacements through the traditional estimated SF increased as the tilt angle increased. Therefore, the traditional SFs calculated by equations (6) and (7) may have introduced large errors in the displacement estimation by the tilt camera.

To explore the variability law of the SFs in the vertical direction, as shown in Figure 3, we assumed the following: tilt angle is θ; the perpendicular of the object plane passing through the camera focus, O, intersects at point P; and the camera optical axis intersects the object surface at point Q. Each edge pixel is connected with O and is extended, intersecting with the object plane at points $Q_1, Q_2, \cdots, Q_{i-1}, Q_i, \cdots, Q_{n-1}, Q_n$, respectively. Line PQ and $Q_{i-1}Q_i$ are known as $\Delta\lambda$ and λ_i, respectively, and $\angle Q_{i-1}OQ_i = \alpha_i$. Then, the following equations can be obtained according to the geometric relationship [34]:

$$\frac{(i-1)}{f} d_{\text{pixel}} = \tan(\alpha_i + \alpha_2 + \cdots + \alpha_{i-1}), \tag{8}$$

$$\frac{i}{f} d_{\text{pixel}} = \tan(\alpha_i + \alpha_2 + \cdots + \alpha_{i-1} + \alpha_i). \tag{9}$$

Subtracting the inverse function of equation (8) from the inverse function of equation (9), α_i can be obtained as follows:

$$\alpha_i = \tan^{-1}\left(\frac{i}{f} d_{\text{pixel}}\right) - \tan^{-1}\left(\frac{i-1}{f} d_{\text{pixel}}\right). \tag{10}$$

Then, according to the triangular geometric relationship of the object plane, the following equations can be obtained:

$$\tan(\theta + \alpha_i + \alpha_2 + \cdots + \alpha_{i-1})$$
$$= \frac{\Delta\lambda + \lambda_1 + \lambda_2 + \cdots + \lambda_{i-1}}{F},$$
$$\tan(\theta + \alpha_i + \alpha_2 + \cdots + \alpha_{i-1} + \alpha_i)$$
$$= \frac{\Delta\lambda + \lambda_1 + \lambda_2 + \cdots + \lambda_{i-1} + \lambda_i}{F}. \tag{11}$$

The actual size of the arbitrary pixel can be expressed as follows:

$$\lambda_i = F \cdot [\tan(\theta + \alpha_1 + \alpha_2 + \cdots \alpha_{i-1} + \alpha_i) - \tan(\theta + \alpha_1 + \alpha_2 + \cdots \alpha_{i-1})]. \tag{12}$$

When the internal and external parameters of the camera were constant, the SF of an arbitrary pixel could be calculated by equation (12). For example, assume that F and f are 500 mm and 50 mm, respectively; and d_{pixel} is determined as 0.0081 mm/pixel when the image sensor size and corresponding resolution are 1.0 inch and 1920×1080 pixels, respectively. The SFs from the center pixel to the edge pixel are shown in Figure 4. The scatter points are the corresponding SFs calculated through equation (7). As shown in Figure 4, the closer the pixel was to the edge of the object plane, the greater the SF was. Moreover, when the camera tilt angle was larger, the change of SF was more intense. The SF is calculated through equation (7), however, remained unchanged for the same tilt angle, which was approximately equal to that calculated by equation (12) at the center pixel, but the differences between the two increased with the closer to the edge pixel. Consequently, the same SF for all the targets in the vertical direction was not applicable to the case of a large camera tilt angle because it introduced large errors. Additionally, the SF calculations through equations (6) and (7) were based on the condition that the actual size of the target structure was known or that the measurement parameters were given. In in-service bridge monitoring, however, it is difficult and inaccurate to measure the structural size or the distance between the camera and the structure.

To make full use of the acceleration and vision-based displacements, we proposed a novel SF estimation method. The main procedure was as follows: first, we estimated the pixel displacement $u_{im}(t_{im})$ and the acceleration-based displacement $u_a(t)$ through the target tracing and FIR filter in the time domain, respectively. Then, we upsampled $u_{im}(t_{im})$ through the Hermite function to keep the time interval equal to $u_a(t)$, as follows:

$$u_3(t) = u_0\left(1 + 2\frac{t-t_0}{t_1-t_0}\right)\left(\frac{t-t_1}{t_0-t_1}\right)^2$$
$$+ u_1\left(1 + 2\frac{t-t_1}{t_0-t_1}\right)\left(\frac{t-t_0}{t_1-t_0}\right)^2$$
$$+ v_0(t-t_0)\left(\frac{t-t_1}{t_0-t_1}\right)^2 + v_1(t-t_1)\left(\frac{t-t_0}{t_1-t_0}\right)^2, \tag{13}$$

where u_0 and u_1 are the displacements at time t_0 and t_1, respectively; and v_0 and v_1 are the derivative displacements at time t_0 and t_1, respectively. Because the vision and accelerations were recorded by different data acquisition systems, it led to a time lag between the two measurements. Thus, we had to eliminate the time difference before SF estimation and data fusion. We realized time synchronization

FIGURE 3: Actual size of each pixel under tilt photogrammetry.

FIGURE 4: Relationship between the SFs to the distance to the center point.

by calculating the cross-correlation coefficient of the two signals. For continuous signal, the cross-correlation coefficient can be calculated as follows [35, 36]:

$$R_{xy}(\tau) = \lim_{T \longrightarrow \infty} \int_0^T x(t)y(t+\tau)\mathrm{d}\tau, \tag{14}$$

where $R_{xy}(\tau)$ is the cross-correlation coefficient of the two signals, T is the measurement time of signals $x(t)$ and $y(t)$, and τ is the time difference between $x(t)$ and $y(t)$. When $R_{xy}(\tau)$ reaches the peak value, the corresponding τ is the time difference between the two signals. For discrete signals whose time interval is Δt, the cross-correlation coefficient can be calculated as follows:

$$R_{xy}(n) = \sum_{i=1}^{N} \frac{x(i)y(i+n)}{N}, \tag{15}$$

where $R_{xy}(n)$ is the cross-correlation coefficient of the two signals, N is the half number of data points in $x(t)$ and $y(t)$, and n is the time lag numbers between $x(t)$ and $y(t)$. When $R_{xy}(n)$ reaches the peak value, the corresponding n is the time lag number between the two signals. The time difference τ can be calculated as $\tau = n \times \Delta t$.

The pixel displacement and acceleration-based displacement had the same frequency band displacement component; therefore, we applied a bandpass filter to the two signals to obtain $u_{im}^f(t)$ and $u_a^f(t)$. We set the lower cutoff frequency f_c^l to eliminate the low-frequency drift in $u_a(t)$, which is depended on the specification of the employed accelerometers. Moreover, in this study, we set the lower cutoff as the natural frequency of the tested structure. We set the upper cutoff frequency f_c^h to one-tenth of the sampling frequency of the camera to minimize the waveform distortion of the filtered vision-based displacement and obtain accurate SF estimation [23, 37]. Then, the SF can be calculated by minimizing the least squares error as follows:

$$\min_{SF} \sum_{t=1}^{n} \left\| u_a^f(t) - SF \times u_{im}^f(t) \right\|_2^2, \tag{16}$$

where $u_{im}^f(t)$ and $u_a^f(t)$ are the pixel displacement and acceleration-based displacement after bandpass filtering, respectively.

2.4. Data Fusion to Obtain High-Fidelity Displacement.

The dynamic displacements reflecting the dynamic information can be accurately reconstructed by the measured accelerations through the FIR filter in the time domain. The total displacements of the structures, however, contain the pseudo-static displacement, which is induced by a moving load on a structure without dynamic effect. The pseudo-static displacement has ultralow frequencies that cannot be measured by the accelerometer. Therefore, the computer

vision technique is considered to acquire the pseudo-static displacement owing to its advantages. Moreover, the two displacements are integrated to obtain accurate displacements. To achieve this, Park et al. [27] proposed the use of a pair of complementary filters for data fusion of the vision-based and acceleration-based displacements. The complementary filters consisted of a low-pass (H_L) and a high-pass filter (H_H) whose sum has a magnitude of 1 and phase of 0 over the defined frequency range, in which H_L is used to extract the pseudo-static from the vision-based displacement, whereas H_H is used to extract the dynamic displacement from the acceleration-based displacement. The cutoff frequency of the filters is set as f_c, so that the bandwidth of the pseudo-static and dynamic displacements are $[0, f_c]$ and $[f_c, f_{sa}/2]$, respectively, and f_{sa} is the sampling frequency of the accelerations. The displacement by data fusion from the vision-based and acceleration-based measurements can be expressed as follows:

$$u(t) = C_L u_{\text{pseudo-static}}(t) + C_H a(t), \tag{17}$$

where C_L is the filter coefficient for H_L and C_H is the filter coefficient for H_H with double integration. Moreover, they can be obtained by the inverse DFT, where $u_{\text{pseudo-static}}(t)$ and $a(t)$ are the vision-based displacement and acceleration measurements with the same time interval. The cutoff frequency of the complementary filters is selected to achieve the lowest noise level for the fused displacements and can be expressed as the intersection of the noisy power spectral density (PSD) of vision-based and acceleration reconstructed displacements as follows:

$$f_c = \frac{1}{2\pi} \sqrt{\frac{N_{\text{acc}}}{N_d}}, \tag{18}$$

where N_{acc} and N_d are the noise density of acceleration m/$(s^2\sqrt{\text{Hz}})$ and vision-based displacement (m/$\sqrt{\text{Hz}}$). Note that $u_{\text{pseudo-static}}(t)$ is converted from the pixel displacement $u_{\text{pseudo-static}}^{im}(t)$ by the proposed SF estimation and can be written as follows:

$$u_{\text{pseudo-static}}(t) = SF \times u_{\text{pseudo-static}}^{im}(t). \tag{19}$$

3. Numerical Simulations

3.1. Model Description.
As indicated in Section 2.2, the dynamic displacement reconstruction from the measured accelerations is greatly influence by the length of the time window, and according to Lee et al. [18], the optimal length of the time window lies in the range of 2.22 to 3 times of the first period of the structure. To explore the optimal length of the time window, a Euler–Bernoulli beam model with different boundary conditions was established. The total length, elastic modulus, density, and moment of inertia of the section are set as 3.84 m, 206 GPa, 7850 kg/m^3, and 8.97e4 mm^4. The description of the beam model is depicted in Figure 5. The boundary conditions, first natural frequencies, and the first periods of the beam were listed in Table 1. An impact load with a mass of 5 kg was acted on the midspan of the model. The dynamic displacements of the midspan were reconstructed with different time window sizes.

Furthermore, according to Section 2.3, the vision-based displacement is greatly influenced by the SF estimation. The selection of the cutoff frequencies would influence the SF estimation. The higher cutoff frequency is suggested to be set as less than one-tenth of the camera's framerate. For structures with large stiffness, however, the first natural frequencies are greater than one-tenth of the camera's frequency. To investigate the influence of the higher cutoff frequency of the proposed bandpass filter, a moving load with a mass of 10 kg is moved from the left to right ends of the simply supported beam at a speed of 0.6 m/s. The displacements and accelerations of the midspan were used to fit the SFs with different higher cutoff frequencies.

3.2. Influence of the Time Window Size.
The displacements and accelerations of the midspan section under impact load were calculated at a frequency of 1000 Hz. The ratio of the time window size to the first period of the beam was denoted as Nd, and the accelerations were used to reconstruct the dynamic displacements with Nd changes from 2.22 to 3. The reconstructed displacements of different boundary conditions are plotted in Figures 6–8, respectively. Overall, the reconstructed displacements were consistent with the simulated displacements at most data points. Taking the simulated displacements as references, a global index called the normalized root mean squared error (NRMSE) [22] was used to evaluate the overall error of the reconstructed displacements, which were calculated as follows:

$$\text{NRMSE} = \frac{\sqrt{(1/n)\sum_{i=1}^{n}(X_i - Y_i)^2}}{Y_{\max} - Y_{\min}} \times 100\%, \tag{20}$$

where X_i and Y_i are the ith reconstructed displacements and references, respectively. The NRMSEs of different boundaries are scattered in Figure 9. For all three conditions, the NRMSEs decreased with the increasing of Nd firstly, and reached the smallest value when Nd is around 2.74, then the NRMSEs increased with the increasing of Nd. According to the previous analysis, the time window size is suggested to be set as about 2.74 times of the first period of the structure.

3.3. Influence of the Higher Cutoff Frequency.
The displacements and accelerations of the midspan section under moving load were calculated at a frequency of 1000 Hz. The first three natural frequencies of the model can be identified through the fast Fourier transform (FFT) of the responses as $f_1 = 4.54$ Hz, $f_2 = 18.12$ Hz, and $f_3 = 40.65$ Hz. First, the displacements were low-pass filtered and resampled at 100 Hz, i.e., 1/10 of the accelerations' sampling frequency. Moreover, the vision-based displacements were low-pass filtered and resampled at 25 Hz, i.e., the same as the framerate of most network surveillance cameras. The resampled displacements were treated as vision-based displacements in the following analysis. The vision-based and acceleration reconstructed displacements are plotted in Figure 10. Then, three bandpass
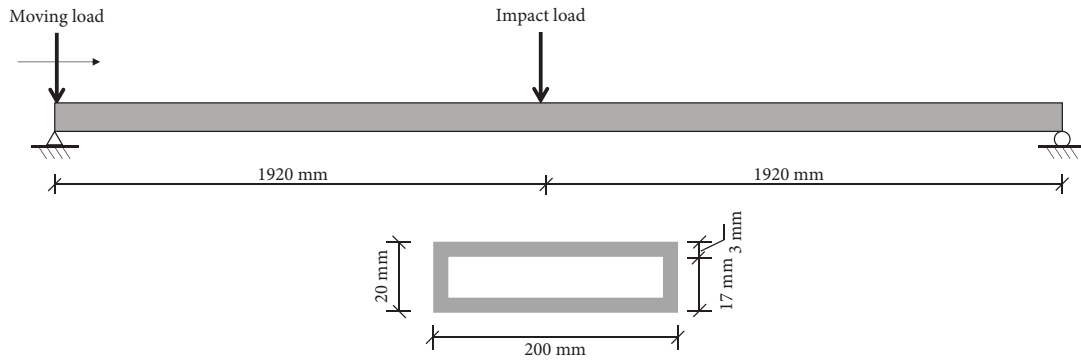
FIGURE 5: The beam model.

TABLE 1: The boundary conditions and first natural frequencies of the beam model.

Scenario	Boundary condition	First natural frequency (Hz)	Length of the first period (s)
SSB	Simply supported	4.54	0.220
SFB	Partially fixed	7.08	0.141
FFB	Completely fixed	10.26	0.097

(a)

(b)
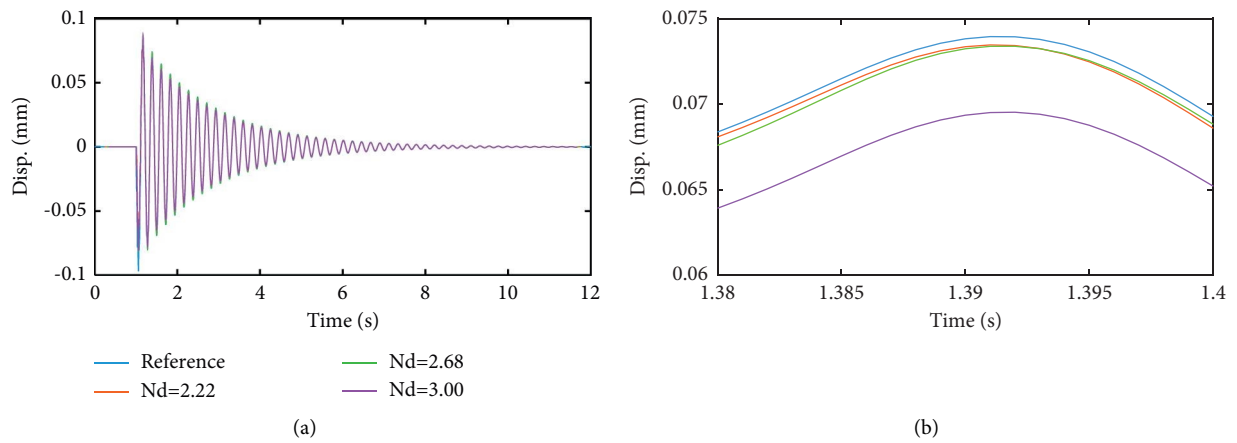
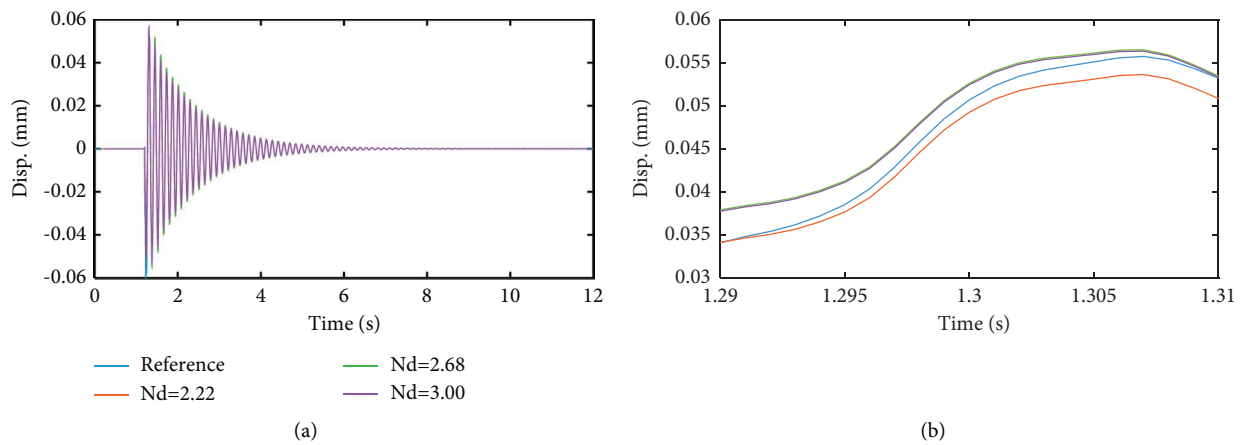FIGURE 6: Displacements comparison of the SSB: (a) overall view and (b) detail view.

(a)

(b)

FIGURE 7: Displacements comparison of the SFB: (a) overall view and (b) detail view.

(a)

(b)
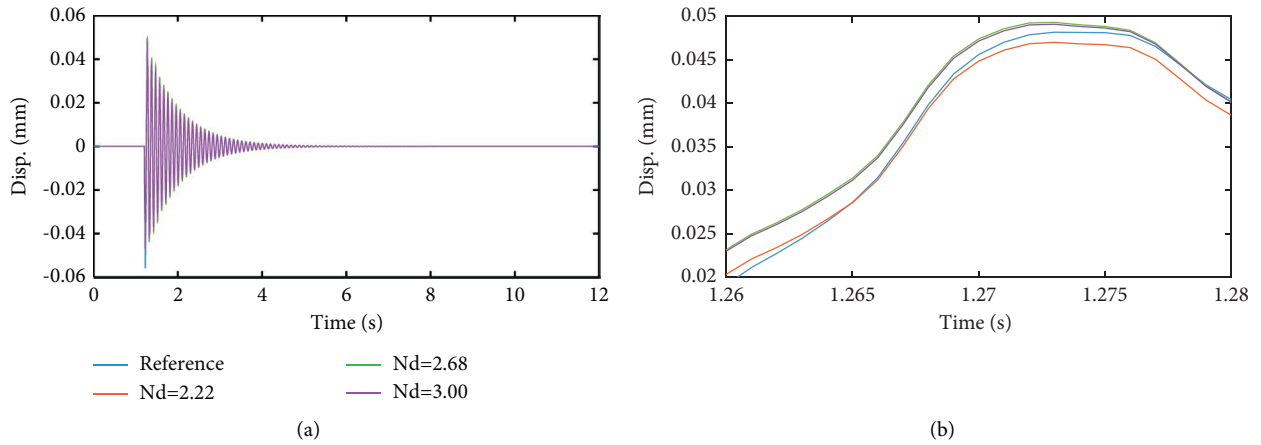
FIGURE 8: Displacements comparison of the FFB: (a) overall view and (b) detail view.
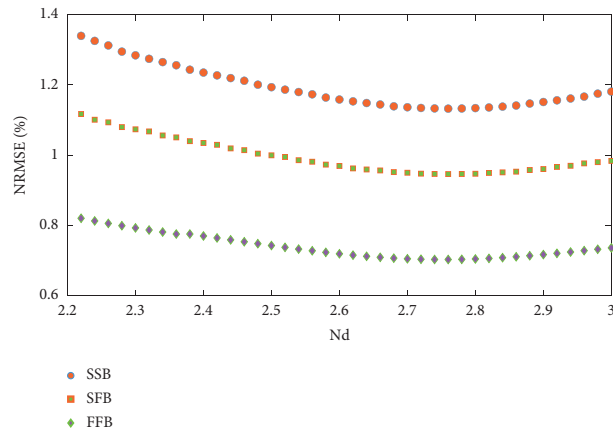
FIGURE 9: NEMSEs of the reconstructed displacements.
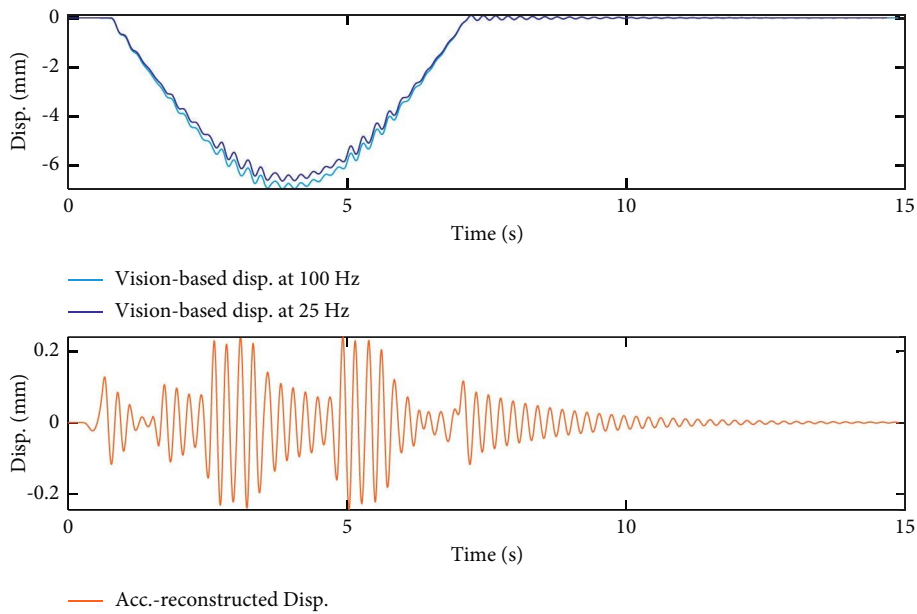
FIGURE 10: Displacements of the mid-span section.

filters with different higher cutoff frequencies were designed, and the parameters of the filters are listed in Table 2.

Then, the vision-based and acceleration reconstructed displacements were filtered by Filters 1-3, respectively. Moreover, note that the vision-based displacements sampled at 25 Hz were only bandpass filtered through Filter 3. The bandpass filtered displacements and corresponding SF estimations were plotted in Figures 11–14. The two displacements after bandpass filtered were consistent with each other at most points for all scenarios overall, and there is an obvious linear relationship between them. Because the displacements and accelerations are calculated by a numerical simulation, the true value of SF should be 1. Therefore, the proposed method can obtain SFs with relative errors of less than 3.5%, as listed in Table 2.

The SF estimation results indicate that when the higher cutoff frequency was set as less than one-tenth of the displacements sampling frequency, the SF estimation is more accurate than those of the higher cutoff frequencies that were greater than one-tenth of the displacements sampling frequency. Moreover, the accuracy is decreased with the increasing of the higher cutoff frequency. Meanwhile, when the displacement sampling frequency is less than ten times of the first natural frequency of the structure, the higher cutoff frequency should not be set as one-tenth of the displacement sampling frequency because no structural dynamic component was retained after filtering. The higher cutoff frequency is suggested to be set as greater than the first natural frequency of the structure to obtain an approximate SF estimation on this occasion.

4. Displacement Monitoring of a Self-Anchored Suspension Bridge Scale Model

4.1. Experimental Setup. We first validated the applicability of the proposed displacement monitoring approach by the experimental studies on the scale model of a self-anchored suspension bridge, as shown in Figure 15. The suspension bridge model, shown in Figure 16, had a main span of $5.3335 + 13.5330 + 5.3335 = 24.2$ m, and its width was 1.2 m. Unlike previous displacement monitoring tests on this structure [38–41], in which the cameras were perpendicular to the target region, we used a network surveillance camera, which was tilted at different angles in this study. To explore the proposed SF estimation method, we selected a total of three targets (LED lamps) located at the three-eighths of the midspan as the tracing targets, as shown in Figure 17. Note that the displacements of the three targets should have been equal because they were located at the same section in the longitudinal and transversal directions, and there was no relative displacement in the vertical direction.

The surveillance camera was fixed on a height-adjustable tripod, and the tilt angle was changed through a ball joint on the top of the tripod. To make the tilt angle as large as possible, the distance from the camera to the bridge was 0.5 m. Meanwhile, the height of the camera was changed with the tilt angle to keep the target T2 in the center of the FOV of the camera. We used a DAHUA camera, which is conventionally used as a surveillance camera in the market.

To simulate the real-world conditions, the resolution, focal length, and framerate were set as 1920×1080 pixels, 24 mm, and 25 fps, respectively, which were similar to those of ordinary surveillance cameras. To increase the discrimination between the tracking targets and the surrounding environment and improve the accuracy and robustness, we used the LED lamps as the tracking target in the lab test and adjusted the camera exposure to fit in with the brightness of the LED lamps.

In addition to the camera, we attached a uniaxial accelerometer (Model #1A202E, Donghua Testing Technology Co. Ltd., Jiangsu, China) and an LVDT (Model #LVDTC20, Donghua Testing Technology Co. Ltd., Jiangsu, China) to the bottom surface of the main girder, as shown in Figure 18. The uniaxial accelerometer measurement was used for data fusion with the vision-based response and SF estimation, and the LVDT measurement was used as the reference for the proposed displacement monitoring approach. Both the accelerometers and LVDTs were sampled at 500 Hz. To confirm the reliability of the proposed approach, we created a total of six scenarios by altering load types and weights as well as the tilt angle of the camera, as detailed in Table 3. We simulated the impact load and moving load by pedestrians jumping at the midspan and running across the bridge, respectively. Before the tests, we calibrated the camera according to Zhang's method [8] to eliminate lens distortion. During the test, we controlled the camera using the computer and changed the tilt angle only after all of the scenarios with the same tilt angle were completed to compare the results of different load types. We measured the tilt angle with a digital inclinometer, as shown in Figure 19.

4.2. Dynamic Displacement Reconstruction by Measured Accelerations. We used vision-based displacements and acceleration-reconstructed displacements as the source data for data fusion. We had to validate whether the displacements reconstructed from the measured accelerations were sufficiently accurate. In scenarios 1-1-1-3, the model bridge was impacted by a pedestrian jumping at the midspan. Because this was the only dynamic displacement, these scenarios did not experience any quasistatic displacement. Because the only difference among these three scenarios was the tilt angle of the camera, we conducted displacements comparison only for scenarios 1-2.

We took the LVDT measurements as references, plotted the comparison of the two displacements in Figure 20(a) and plotted the displacements differences in Figure 20(b). These figures indicated that the two measurements had very good consistency at most data points, and the maximum difference was approximately 0.2 mm. To quantify the errors of the proposed method, we implemented error analysis in terms of the NRMSE and the absolute peak value error (APVE), which were calculated as follows:

$$\text{APVE} = \left| \frac{|X|_{\max} - |Y|_{\max}}{|Y|_{\max}} \right| \times 100\%, \qquad (21)$$

where $|X|_{\max}$ and $|Y|_{\max}$ are the maximum absolute values of estimated displacements and references, respectively. The

TABLE 2: The parameters of the bandpass filter and corresponding SF estimations.

Scenario	The sampling frequency of vision-based displacement (Hz)	Filter order	Lower cutoff frequency (Hz)	Higher cutoff frequency (Hz)	R^2	SF	Error (%)
1	100	10	2	6	0.9929	0.9708	2.92
2	100	10	2	20	0.9921	0.9695	3.05
3	100	10	2	50	0.9914	0.9674	3.26
4	25	10	2	6	0.9919	0.9683	3.17
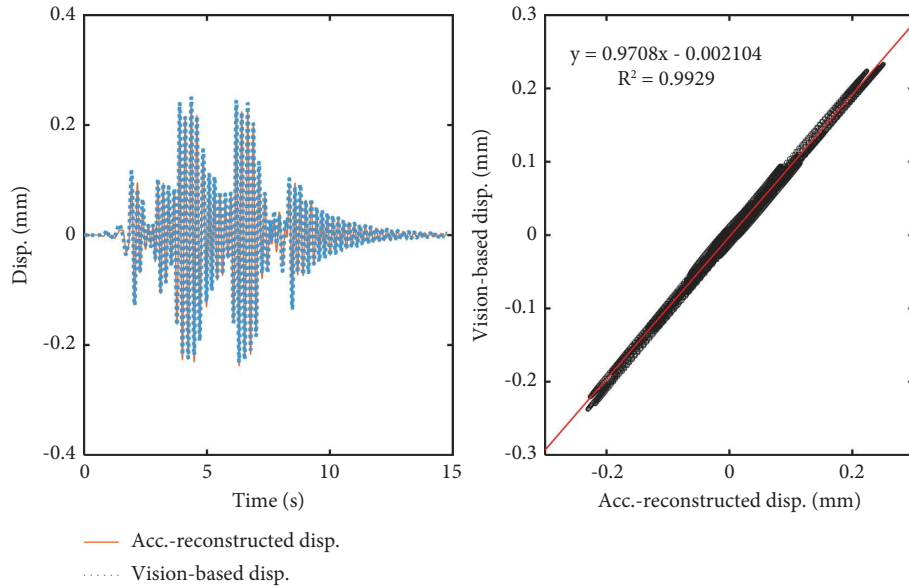
FIGURE 11: Displacements and corresponding SF of scenario 1.
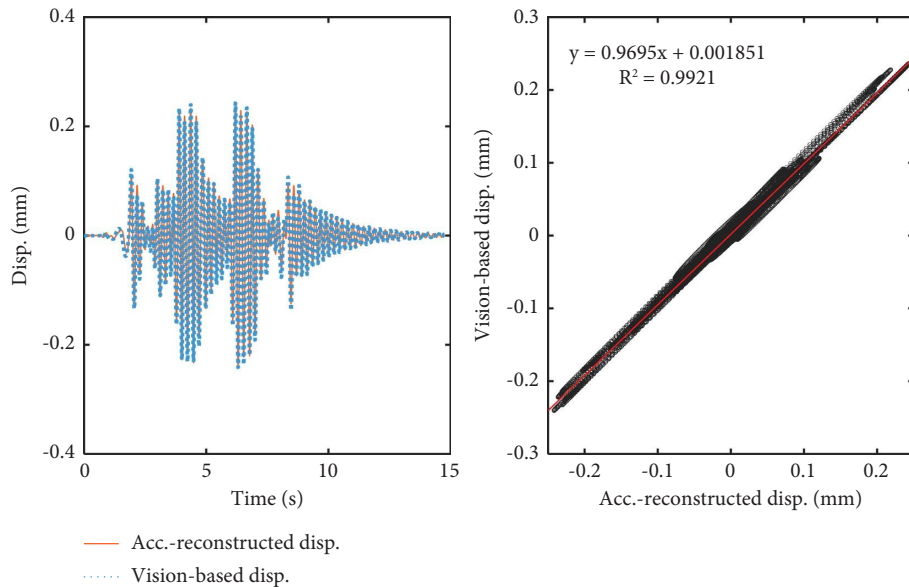
FIGURE 12: Displacements and corresponding SF of scenario 2.
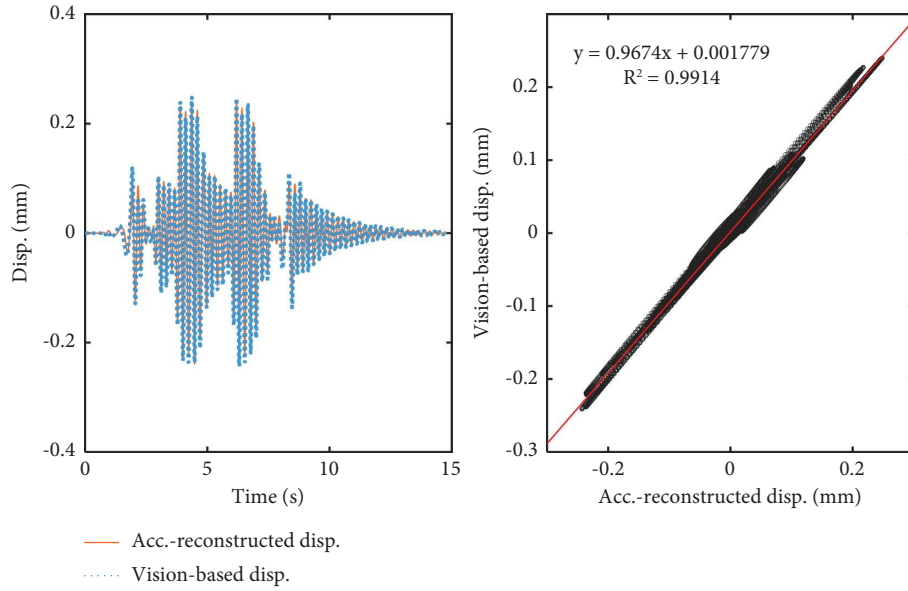
FIGURE 13: Displacements and corresponding SF of scenario 3.
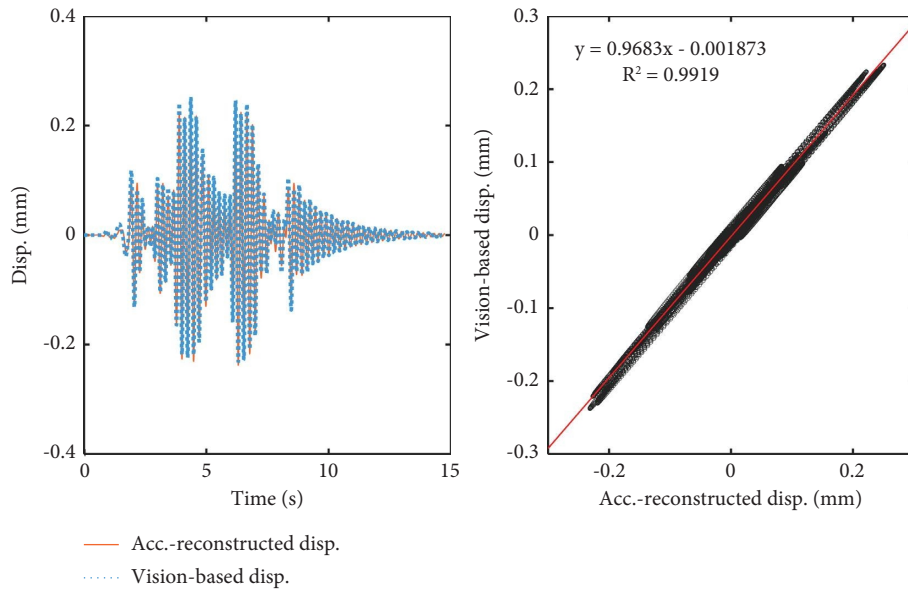
FIGURE 14: Displacements and corresponding SF of scenario 4.

NRMSE and APVE of the reconstructed displacements were 0.24% and 2.32%, respectively. Furthermore, we analyzed the power spectrum density (PSD) of the two displacements to validate reliability in the frequency domain, as shown in Figure 21. The PSDs of the two displacements were in good agreement. Then, we obtained the first five natural frequencies and corresponding damping rates by modal analysis, as listed in Table 4.

4.3. Scaling Factor Estimation Results. We calculated the vision-based displacements of all the targets by the feature tracking algorithm. Moreover, the time lag between the vision-based and acceleration-reconstructed displacements

was calculated using the previous method. The displacements of scenario 1-1 before time lag elimination are plotted, as shown in Figure 22(a). We calculated and plotted the cross-correlation coefficients between the two measurements, as shown in Figure 23. The results indicated that the time lag between the two measurements was −9.012 s. Thus, we eliminated the time lag by deleting the data points before 9.012 s and shifting the remaining data points by 9.012 s. Then, we plotted the displacements after time lag elimination, as shown in Figure 22(b). We followed the same procedures for all scenarios.

To obtain the acceleration-based and vision-based displacements at the same frequency band, we implemented bandpass filtering and set the lower and higher
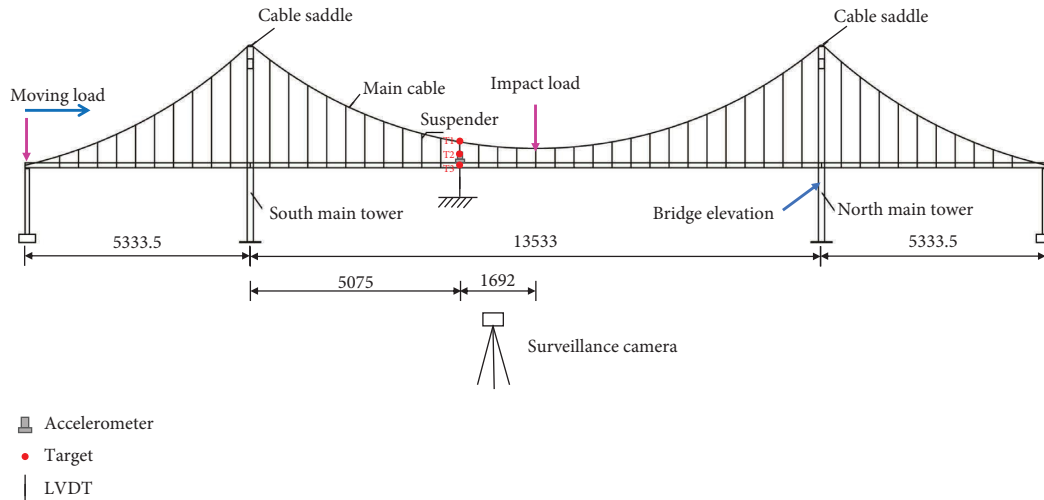
Figure 15: Experimental setup (unit: mm).

Figure 16: The bridge model.

Figure 18: Accelerometer and LVDT.
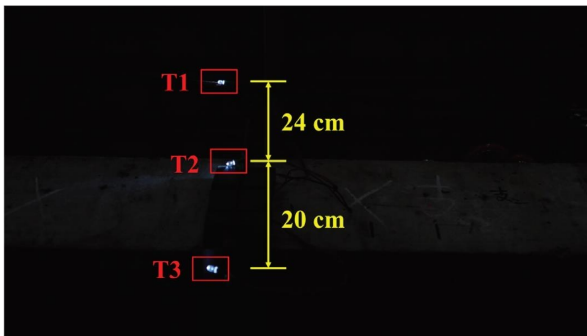
Figure 17: Three targets in the camera's FOV.

Table 3: The test specifications.

	Load type	Weight (kg)	Tilt angle (°)
Scenario 1-1	Impact load	80	0.00
Scenario 1-2	Impact load	80	45.20
Scenario 1-3	Impact load	80	70.31
Scenario 1-4	Moving load	150	0.00
Scenario 1-5	Moving load	150	45.20
Scenario 1-6	Moving load	150	70.31

Figure 19: Tilted camera and angle measurement.

cutoff frequencies as 1.7 Hz and 2.5 Hz, respectively. The displacement results after bandpass filtering of each scenario are shown in Figures 24–29, which show that with the increase in the tilt angle, the pixel displacement decreased shapely. Meanwhile, the pixel displacements of the three targets in scenarios 1-1 and 1-4, in which the tilt angles were zero, were approximately consistent, although differences appeared as the tilt angle increased. The pixel displacement of T3 was the largest, followed by T2 and T1.

(a)

(b)

FIGURE 20: The displacement comparison: (a) displacement comparison between acceleration-reconstructed displacements and LVDT measurements and (b) displacement differences.
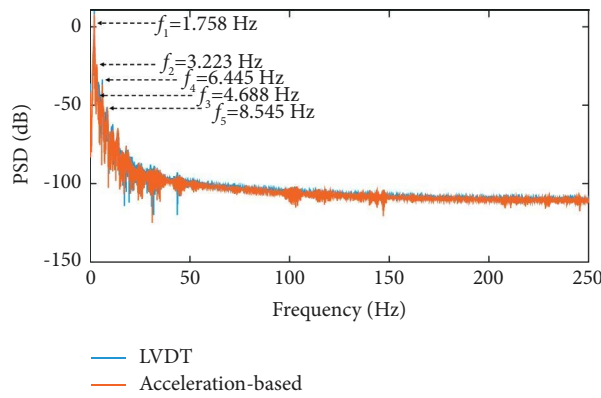
FIGURE 21: PSD comparison between the LVDT and acceleration-based displacements.

TABLE 4: Structural parameters of the model bridge.

Mode order	Frequency (Hz)	Damping rate (%)
1	1.758	2.010
2	3.223	1.170
3	4.688	1.304
4	6.445	1.910
5	8.545	0.302

(a)

(b)
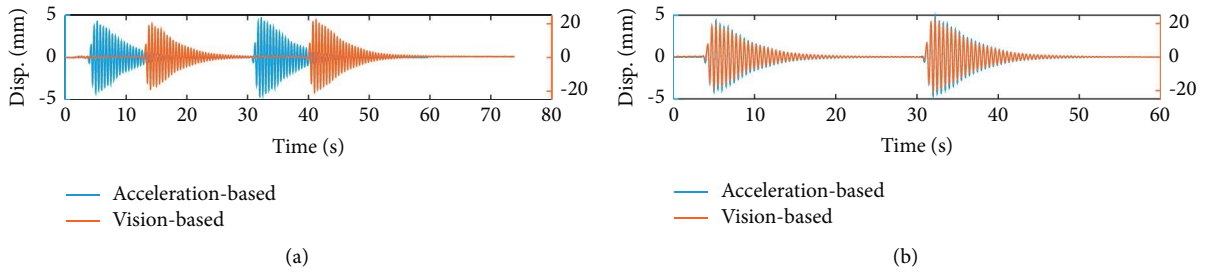
FIGURE 22: The displacements: (a) before time lag elimination and (b) after time lag elimination.
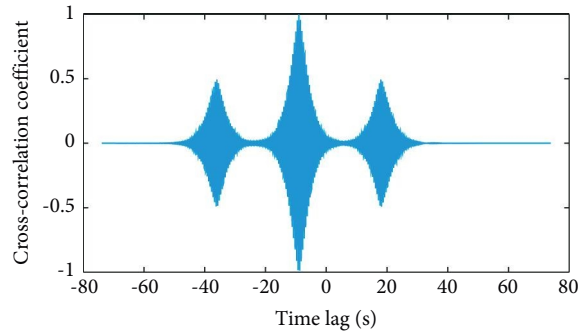
FIGURE 23: Cross-correlation coefficients of the two displacements.
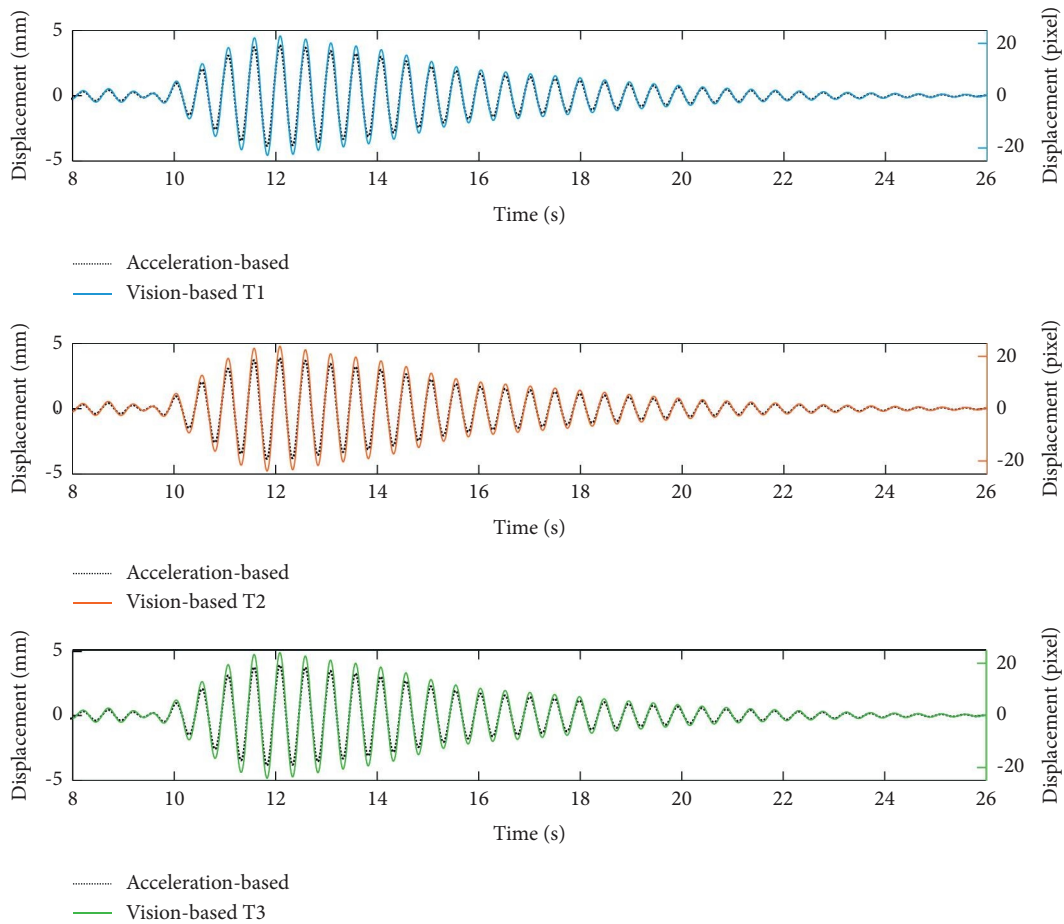
FIGURE 24: The acceleration-based and vision-based displacements after bandpass filtering of scenario 1-1.

We analyzed the relationships between the two displacements, as shown in Figures 30–35. We obtained the linear equations of the two displacements by least square fitting. The slope of the lines can be regarded as SFs. We calculated SF_1 and SF_2 by equations (6) and (7), respectively, and estimated the results according to the proposed method. According to Table 5, the SFs in scenarios 1-1 and 1-4, in which the tilt angles were zero, which was the minimum, and the SFs of all the targets were approximately equal to those calculated through the traditional method, that is, calculated by

equations (6) and (7). Then, the SFs increased as the tilt angle increased. In the large tilt angle scenarios, the SFs for three targets varied, of which T3 was the smallest, followed by T2 and T1. Furthermore, we also found some differences between the SFs estimated by the three methods, which indicated that the traditional methods for SF estimation in a large tilt angle may introduce large errors.

By comparing the impact load and moving load in the same tilt angle, we found that the linear correlation of the impact load scenarios was better than that of the moving
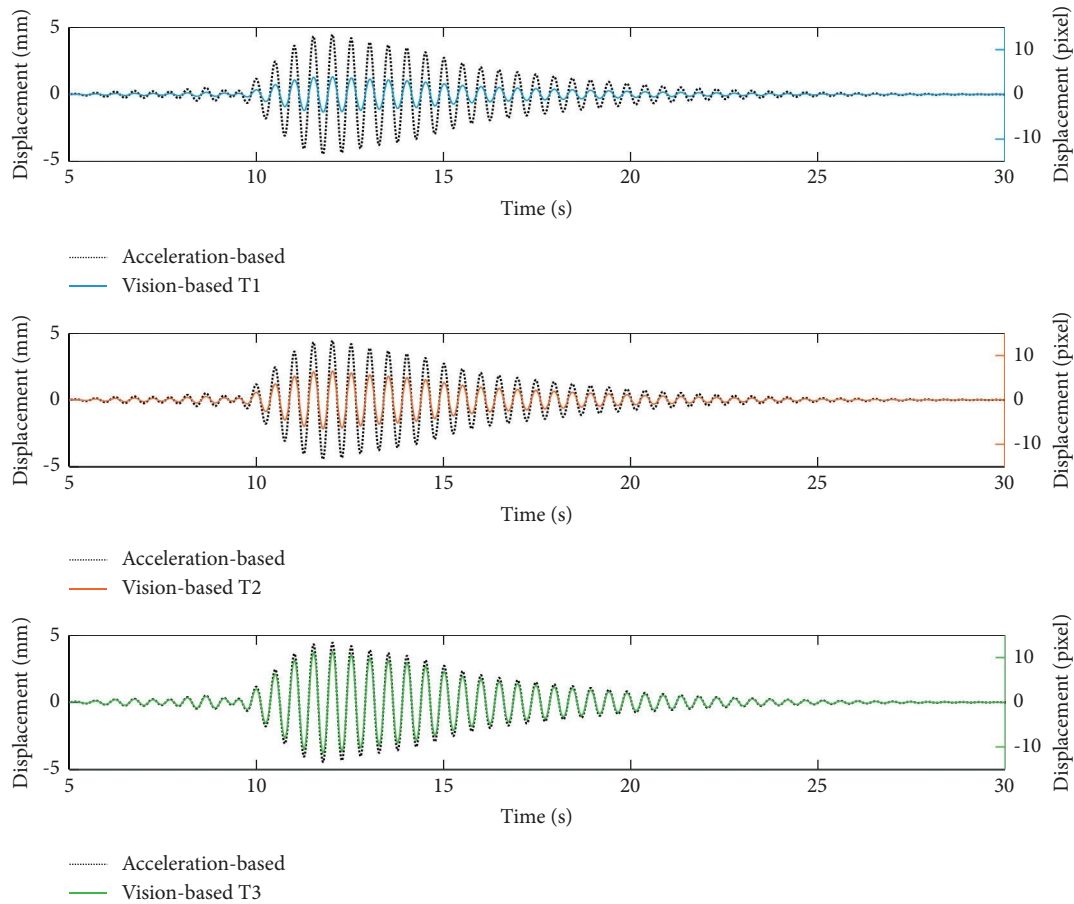
Figure 25: The acceleration-based and vision-based displacements after bandpass filtering of scenario 1-2.

load scenarios, especially for the large tilt angle scenario. Furthermore, for the same tilt angle, the SF of the impact load was a little bigger than that of the moving load. Because of the pseudostatic displacement caused by the moving load, the equilibrium positions for the dynamic displacement changed with the pseudostatic displacement, although the equilibrium position for the impact load scenarios remained unchanged. In addition, the average equilibrium position of the moving load scenario was closer to the camera than that of the impact load because of the moving load. Thus, the SFs of the moving load scenario were a little smaller than those of the impact load scenarios.

4.4. Displacement Estimation Results.

For impact load scenarios, we obtained the displacement directly by the FIR filter in the time domain and high-pass filtering of the accelerations because no pseudo-static displacement occurred. Therefore, we estimated only the displacements of the moving load scenarios. The procedures for displacement estimation are shown in Figure 1. We selected the displacements measured by LVDT as the reference and estimated the displacements using the proposed method. Meanwhile, the SFs calculated by equations (6) and (7) were used to correct the vision-based displacements and then fused with the acceleration reconstructed displacements, labelled as "Method-1" and "Method-2," respectively. The

displacements obtained by the LVDT, the traditional methods, and the proposed method are plotted in Figures 36–38. The LVDT measurements and the proposed estimations showed great consistency for the three targets in all the scenarios, and the traditional estimations showed good consistency in scenario 1-1 as well. As the tilt angle increased, however, the displacement differences between the traditional methods and LVDT measurements also increased. Because the difference between the three methods was the SF estimations, it can be inferred that the SFs estimated by the proposed method are the most accurate, especially in large tilt angle scenarios.

To quantify the errors of the proposed method, we implemented the error analysis in terms of NRMSE. The error analysis results are shown in Figures 39–41 for scenarios 1-4, 1-5, and 1-6, respectively. As shown in Figure 39, the NRMSEs of the traditional and proposed measurements were less than 2% in general, in which the SF_2-based vision measurements were the largest, followed by the SF_1-based vision and proposed measurements. The SF estimation results given in Table 5 indicated that the SFs estimated by the three methods were approximately equal to 0.162 mm/pixel for target T2 when the tilt angle was zero. Furthermore, when the tilt angle increased, the errors of the traditional estimations increased significantly because of the inaccurate SF estimations.
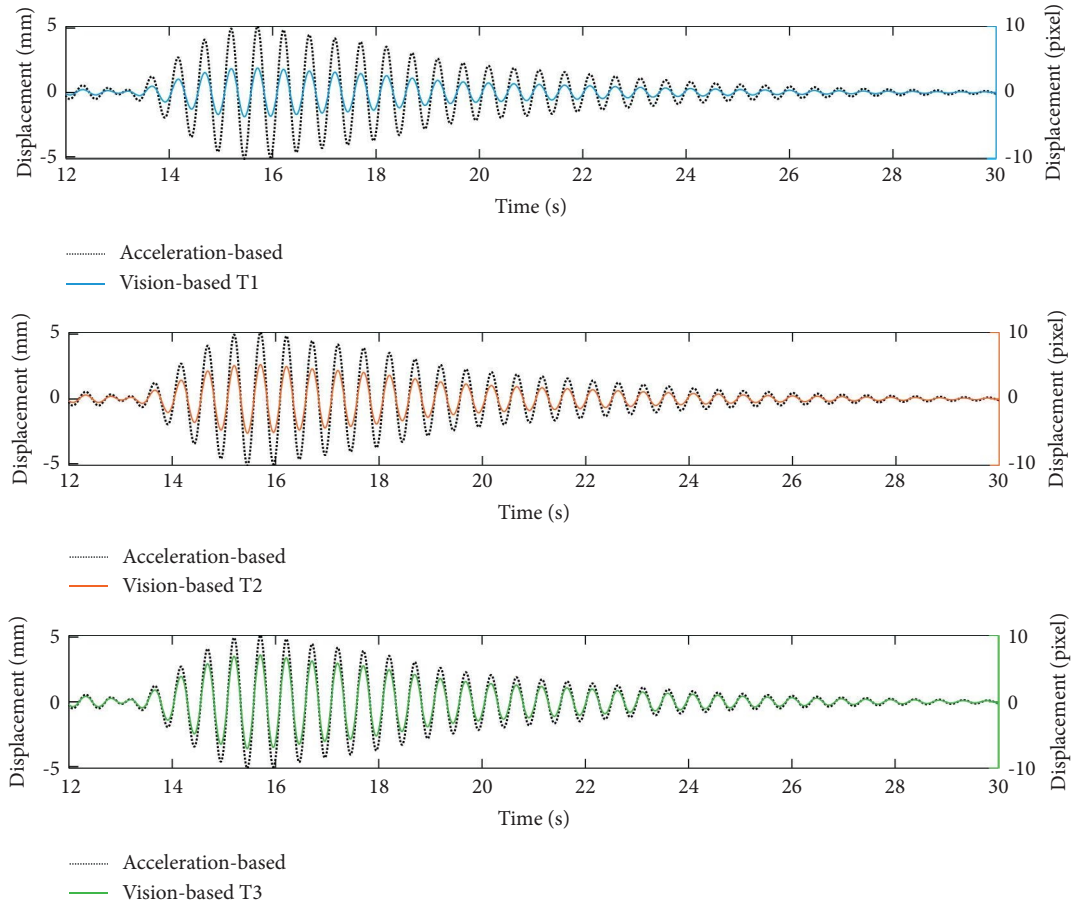
FIGURE 26: The acceleration-based and vision-based displacements after bandpass filtering of scenario 1-3.

These experimental results indicated that the SFs fitted from the vision-based and acceleration-based displacements in the same frequency band could obtain high-accuracy displacements, even in large tilt angle scenarios. The advantage of this method is that we did not have to measure the structure size or determine the camera parameters, unlike the conventional SFs calculated by equations (6) and (7).

5. Field Test on a Cable-Stayed Bridge

5.1. Experimental Setup. We further evaluated the performance of the proposed method on the Dongshuimen Yangtze River Bridge in Chongqing, China. This bridge is a two-deck steel truss girder cable-stayed bridge with a main span of $222.5 + 450 + 190.5$ m. The upper deck accommodates four-lane vehicle traffic with a design speed of 40 km/h, and the lower layer encompasses two-line subway traffic. We focused on the vertical displacement of the midspan during the subway trains passing the bridge. We attached an axial magnetoelectric accelerometer (modeled DH 2D003V, Donghua Testing Technology Co. Ltd., China) at midspan, on the ground at the sidewalk near the railing. The frequency range and sensitivity of the accelerometer were 0.125–80 Hz and $0.287 \, \text{V}/(\text{m/s}^2)$, respectively. We mounted a camera (Sony

FDR AX-700) at the riverside road under the bridge with negligible ground vibration, approximately 300 m away from the midspan of the bridge. The height difference between the camera and the bridge deck was approximately 50 m. Therefore, we applied a small elevation (approximately 15°) to keep the main girder in the center of the camera's FOV. The focal length and resolution of the camera were set as 120 mm and 3840×2160 pixels. The videos were sampled at 25 Hz, which was sufficient for this flexible long-span bridge, and the acceleration was sampled at 100 Hz. The overview of the test is shown in Figure 42.

5.2. Displacement Estimation. The data acquisition lasted approximately 30 mins, and a total of nine subway trains passed the bridge. A natural target and a stationary target (Figure 43) were tracked to monitor the displacement of the midspan section. The bridge was in normal operation during the data acquisition, and cars and pedestrians were traveling on the upper deck. The responses caused by heavy trucks were reflected in the vision-based displacement and accelerations, as shown in Figure 44. We eliminated the camera motions in the vision-based displacement response by subtracting the displacement of the stationary point in the camera's FOV from the displacement of the target [3]. The
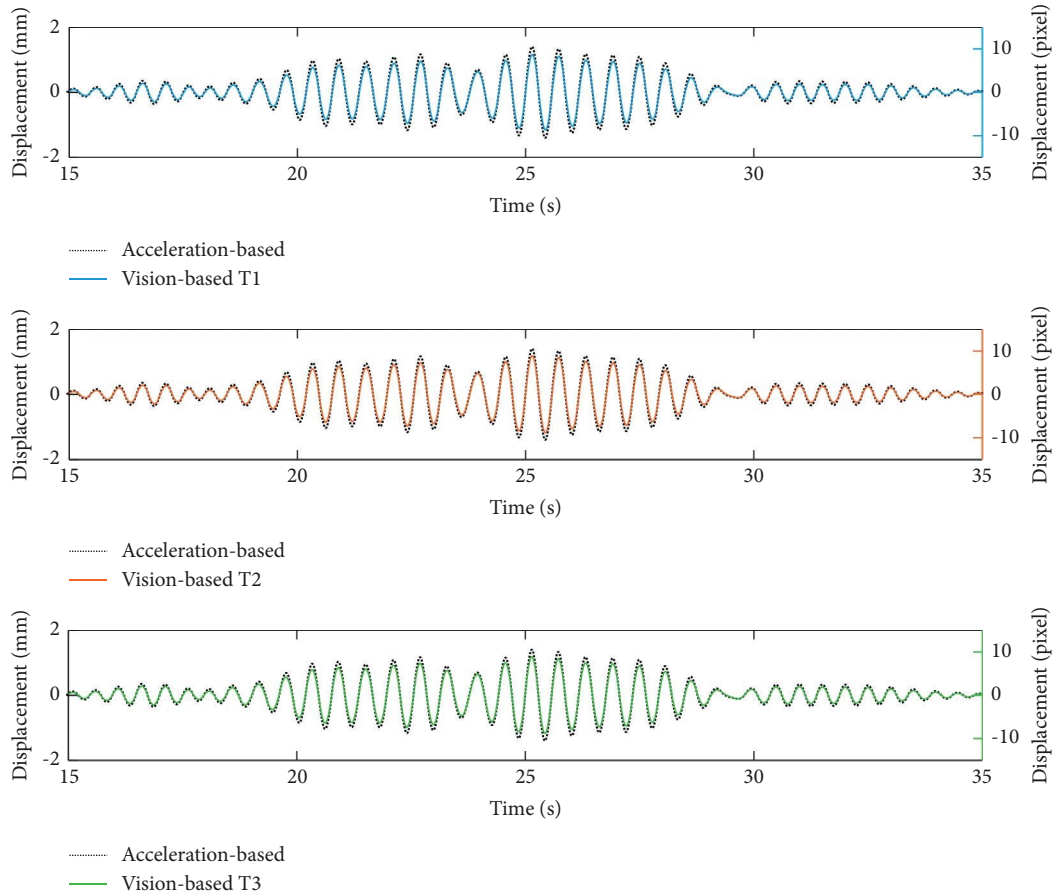
FIGURE 27: The acceleration-based and vision-based displacements after bandpass filtering of scenario 1-4.

first five natural frequencies and damping rates are listed in Table 6.

Because acceleration was more sensitive to structural vibration than vision-based displacement, we used the spectrum analysis of acceleration, as shown in Figure 45, as the reference for data filtering. The lower and higher cutoff frequencies were set at 0.35 Hz and 0.70 Hz, respectively. The vision-based and acceleration-based displacements after bandpass filtering are shown in Figure 46, and the relationships of the two signals are scattered and fitted in Figure 47. The slope of the line indicated that the SF could be determined as 9.925 mm/pixel without the measurement distance, focal length, and size of the image sensor. The operating bridges, especially such large, long-span bridges, were subjected to various external loads, and the noise from the environment had a significant influence on the vision-based displacement and acceleration, which could not be ignored. Therefore, the linear relationship between the two measurements was not as good as that of the experimental bridge.

We estimated displacement by fusing the low-pass filtered vision-based and high-pass filtered acceleration-based displacements. The data fusion results are plotted in Figure 48, and we determined that the displacement caused by subway trains was approximately 80 mm. Compared with

the original vision-based displacement shown in Figure 44(a), the fusion displacement had fewer fluctuations, especially when no trains were passing. This was because the part of the camera vibration was eliminated by low-pass filtering. It is highly difficult to use conventional contact sensors to monitor the displacement of such large-span bridges. Therefore, we did not have any references to evaluate the accuracy of the fusion displacement in the time domain. We compared the PSD of the fusion displacement with that of the acceleration-based displacement, as shown in Figure 49, which indicated that the result of data fusion was reliable.

6. Discussion

The proposed displacement estimation approach was characterized by several advantages compared with the conventional vision-based and data fusion methods. First, we used the camera with a large pitch angle for data fusion with the acceleration for high-accuracy displacement estimations. In traditional methods, it is usually suggested to fix the camera in a suitable place away from the target structure and to keep the optical axis perpendicular to the objective plane. Sometimes it is difficult to locate an ideal position, especially for bridges across valleys and rivers, and in these cases, the camera must be tilted for a better view. Second, we
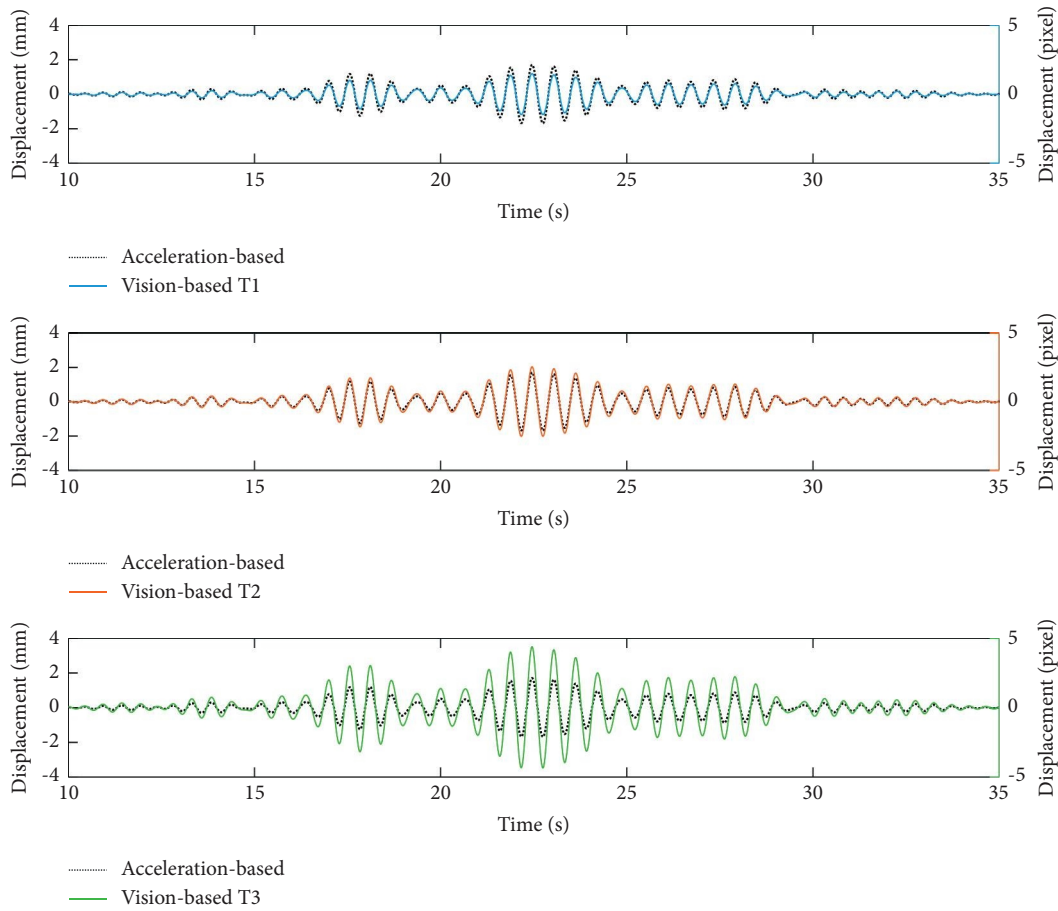
FIGURE 28: The acceleration-based and vision-based displacements after bandpass filtering of scenario 1-5.

proposed a novel SF estimation method by fitting the vision- and acceleration-based displacements in the same frequency band without any prior knowledge about the structural size or measurement distances. Furthermore, when the camera was tilted at a large angle, the SF calculated by the traditional method may have introduced an error, and this error could be reduced by the fitting SF. Third, the framerate of the mainstream consumer-grade cameras in the market is limited to 20–60 Hz, which might have prohibited recording high-frequency dynamic motion and might have caused aliasing issues. We solved these problems by data fusion with accelerations thanks to the high accuracy of the high-frequency displacement estimated by the measured acceleration.

Today, monitoring systems have been widely mounted on civil infrastructures, in which surveillance cameras can provide substantial data. The data obtained by these surveillance cameras, however, are used mostly for traffic flow surveillance and accident disposing. By employing these traffic surveillance cameras as sensors, not only the input of the structures such as moving loads can be recorded [42, 43] but also the cameras can work with other sensors to record the responses of the structures. Therefore, traffic surveillance cameras can be considered to be ready-made vision-based sensors, which can be used to achieve multiple measurements for one installation. These traffic surveillance cameras have larger tilt angles and greater vibration than the general cameras. The traffic surveillance cameras are always
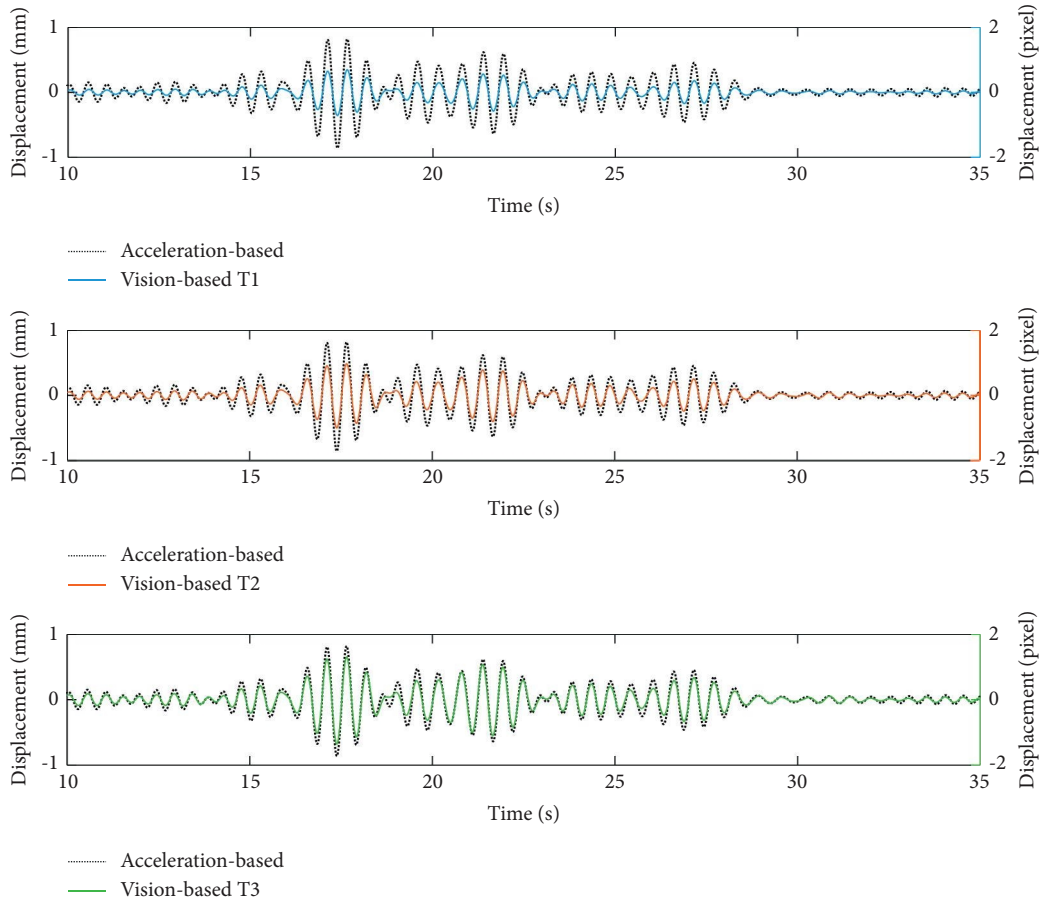
FIGURE 29: The acceleration and vision-based displacements after bandpass filtering of scenario 1-6.
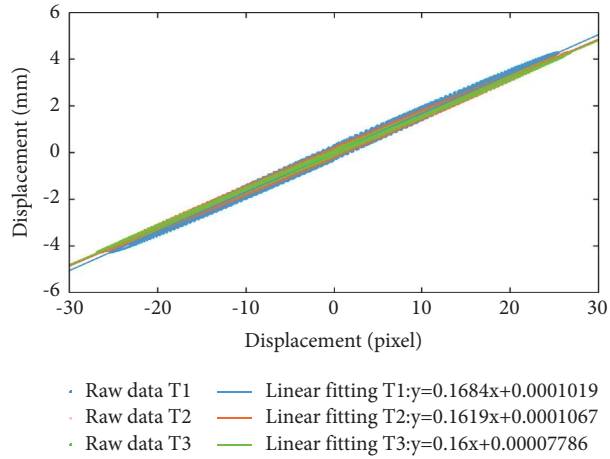
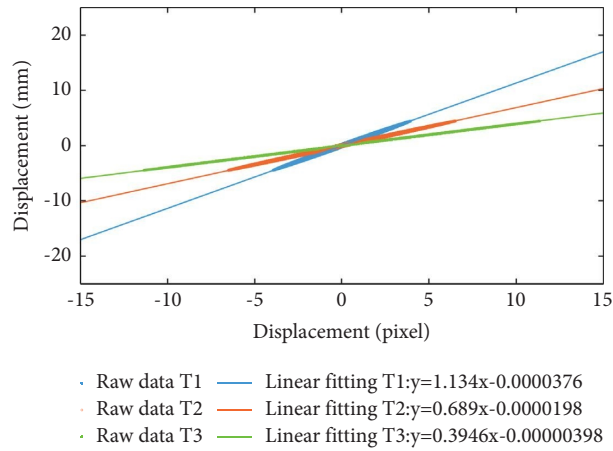FIGURE 30: Relationship between the two displacements of scenario 1-1.

FIGURE 31: Relationship between the two displacements of scenario 1-2.
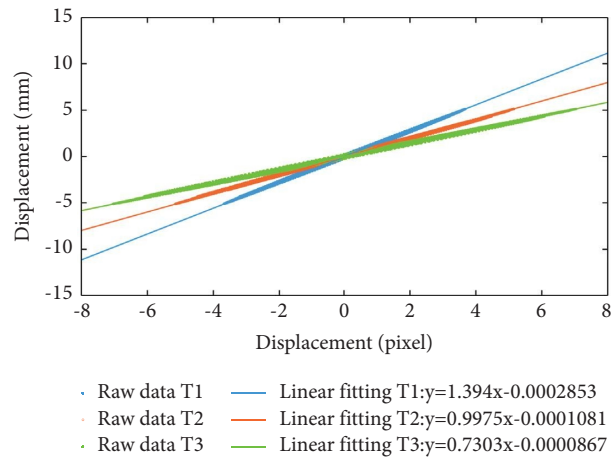

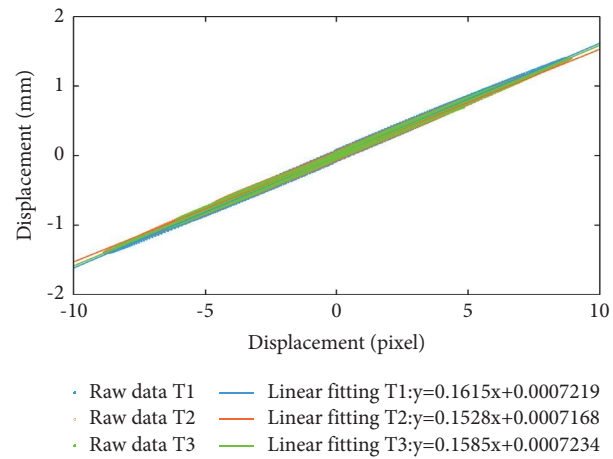FIGURE 32: Relationship between the two displacements of scenario 1-3.

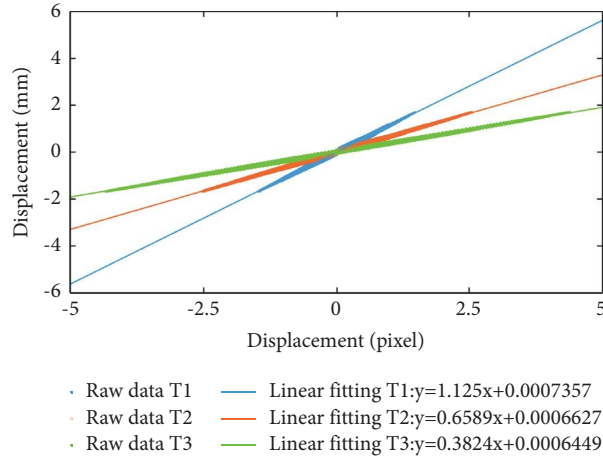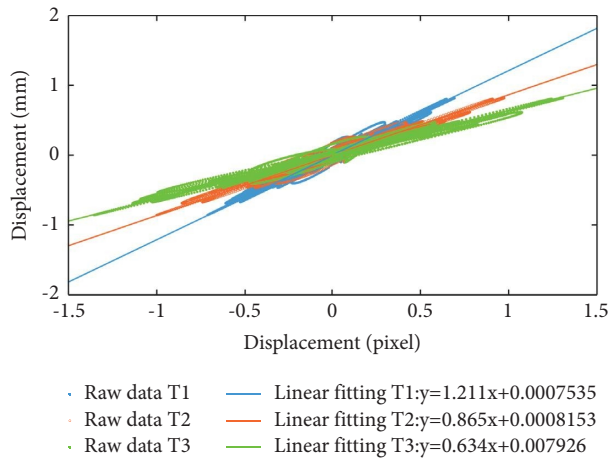FIGURE 33: Relationship between the two displacements of scenario 1-4.

Figure 34: Relationship between the two displacements of scenario 1-5.

Figure 35: Relationship between the two displacements of scenario 1-6.

Table 5: SF estimation results.

Scenarios	Tilt angle (°)	Target	SF_1 (mm/pixel)	SF_2 (mm/pixel)	SF-proposed (mm/pixel)
Scenario 1-1	0.00	1	0.162	0.163	0.168
		2			0.162
		3			0.160
Scenario 1-2	45.20	1	0.478	0.382	1.134
		2			0.689
		3			0.395
Scenario 1-3	70.31	1	1.287	1.436	1.394
		2			0.998
		3			0.730
Scenario 1-4	0.00	1	0.162	0.163	0.162
		2			0.153
		3			0.159
Scenario 1-5	45.20	1	0.487	0.382	1.125
		2			0.659
		3			0.382
Scenario 1-6	70.31	1	0.967	1.436	1.211
		2			0.865
		3			0.634
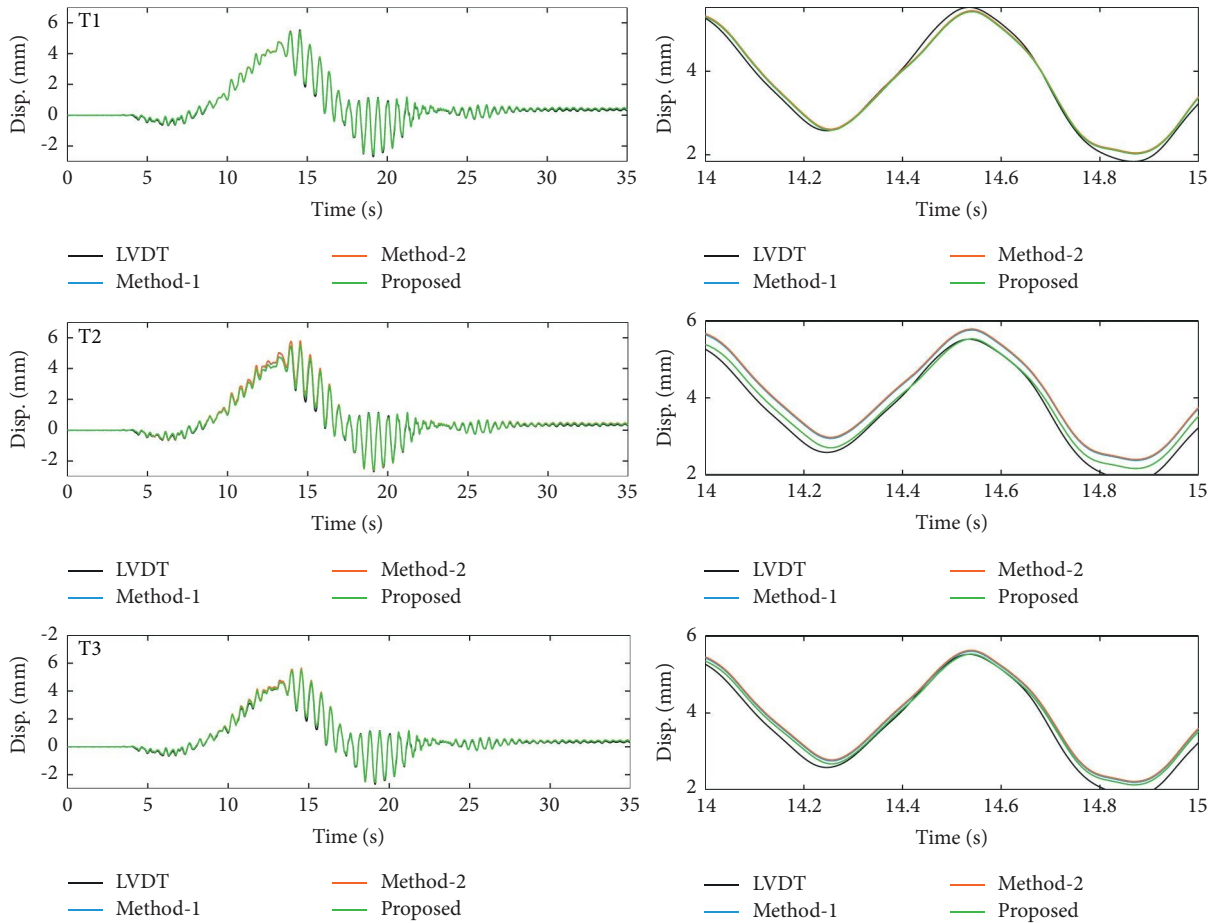
FIGURE 36: Displacement estimations of scenario 1-4.

mounted on the monitored structure and facing the structure with large tilt angles would introduce a large error if the traffic surveillance cameras were used for structural displacement monitoring. Therefore, in this study, we explored the potential for structural displacement monitoring using data fusion of a tilt camera and accelerometer measurement, which could pave the way for SHM using traffic surveillance cameras.

The proposed approach showed potential for accurate displacement estimation by data fusion of vision and acceleration measurements, and its efficiency was validated both on experimental and field tests. Furthermore, a study is needed, however, to address several limitations. First, it was difficult to eliminate the temporal aliasing in the recorded video, which was one of the major issues in the vision-based measurement system. Furthermore, the data fusion was conducted only on accelerometer locations, whereas the displacement of the acceleration unmeasured locations could not be improved. Thirdly, the necessary condition for applying the proposed SF estimation method is that the bandpass filtered displacements retain the dynamic components of the structure. For structures with large stiffness,
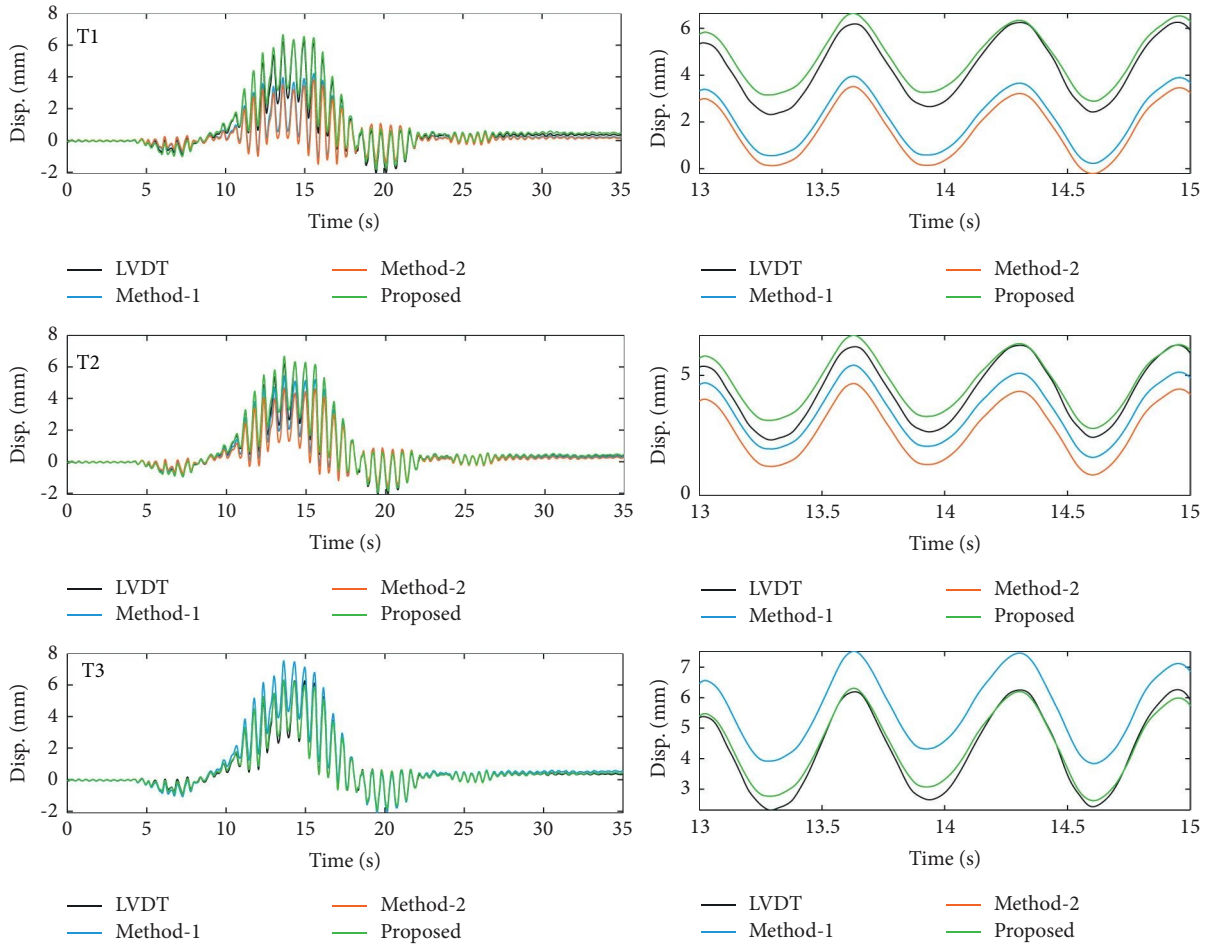
FIGURE 37: Displacement estimations of scenario 1-5.

however, a major issue is that the suggested ideal higher cutoff frequency, i.e., one-tenth of the camera framerate, would be smaller than the first natural frequency of the structure once the framerate of the camera is small. On this condition, the proposed SF estimation method might be not applicable. According to the investigation about the influence of the higher cutoff frequency on the SF estimation in Section 3, the upper cutoff frequency can be increase to greater than the first natural frequency of the structure to obtain approximate estimations, but it should not exceed 1/2 of the camera framerate.

Based on the previous limitations, our future studies should focus on sparse accelerometer-aided computer vision technique for full-field displacement monitoring of more structures. Specifically, we expected to expand the data fusion method from the accelerometer locations to other locations where no accelerometer is installed. Meanwhile, the temporal aliasing issue in vision-based measurements should also be considered. Furthermore, the proposed SF estimation method is expected to be used in more structure types and realize adaptive and accurate estimation through vision-based displacement sampled at
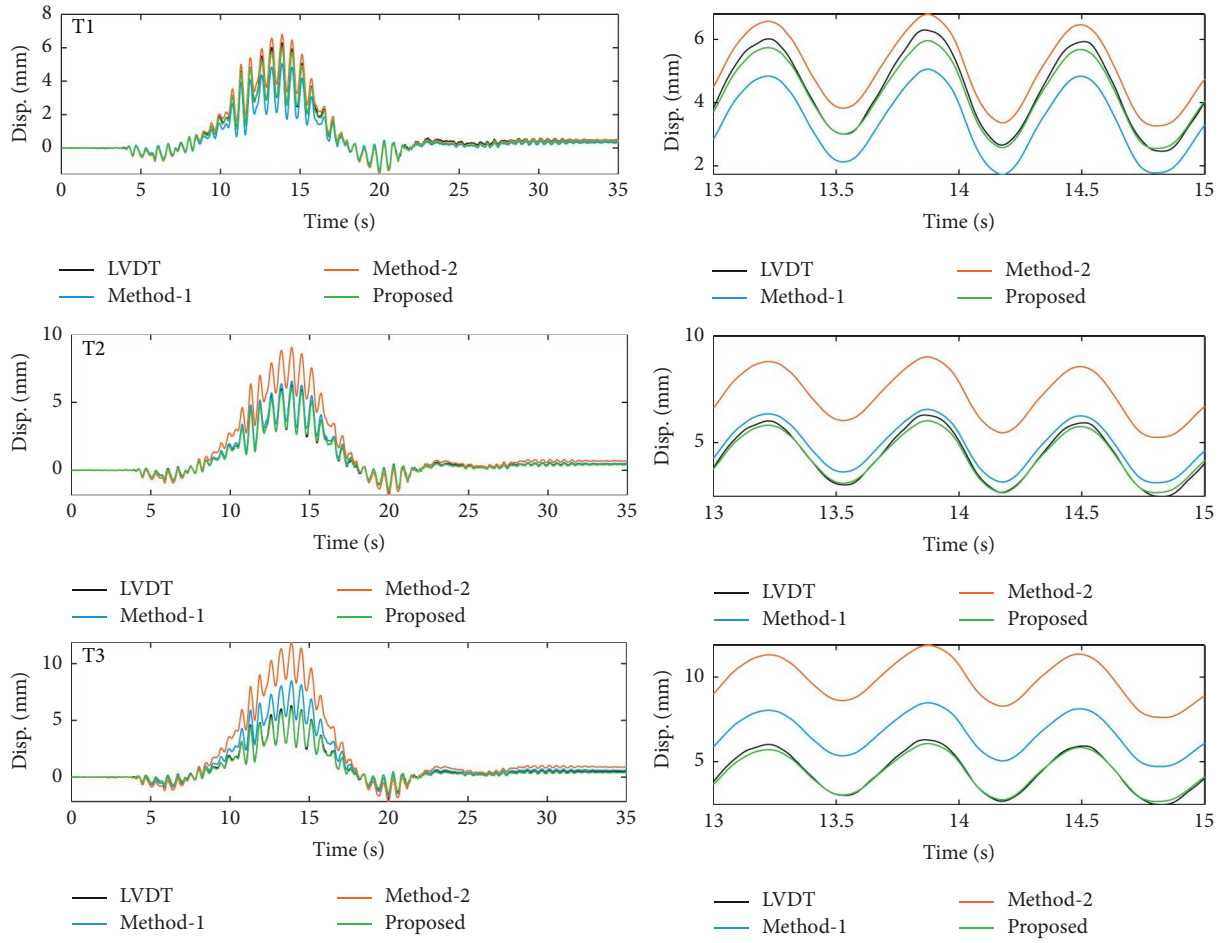
FIGURE 38: Displacement estimations of scenario 1-6.
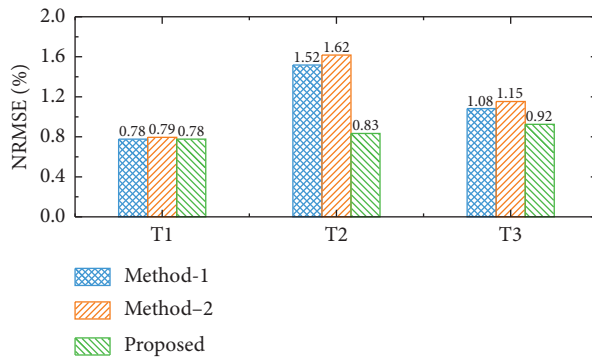

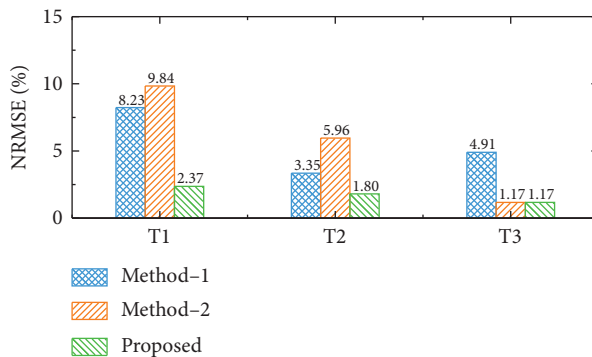FIGURE 39: Error analysis of scenario 1-4.

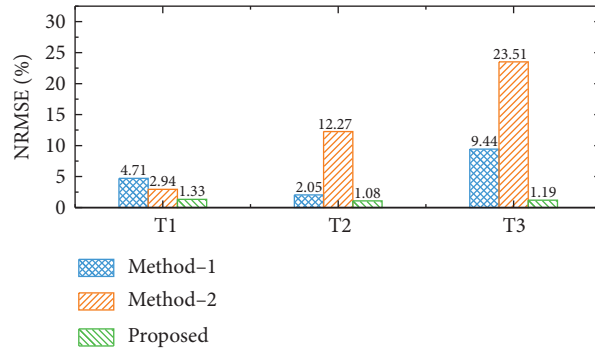FIGURE 40: Error analysis of scenario 1-5.

FIGURE 41: Error analysis of scenario 1-6.

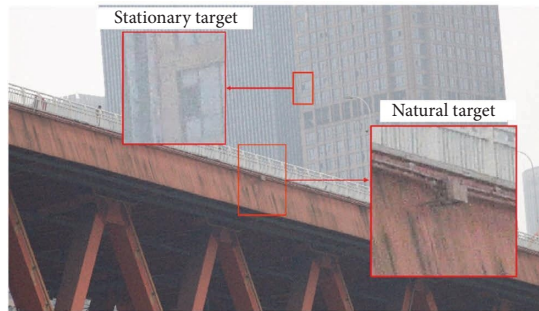FIGURE 42: Overview of the cable-stayed bridge test.

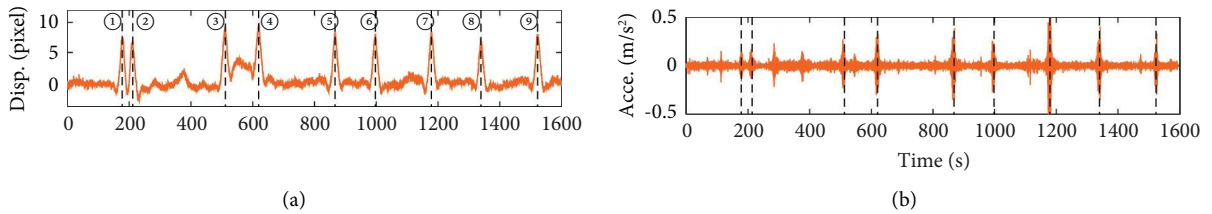FIGURE 43: The natural and stationary targets in the camera's FOV.

(a)   (b)

FIGURE 44: The original responses: (a) vision-based displacement and (b) acceleration.

TABLE 6: Structural parameters of the cable-stayed bridge.

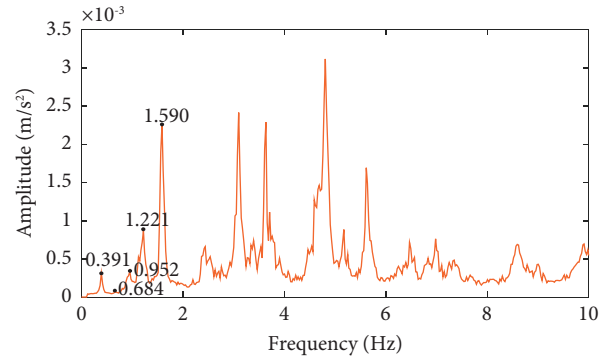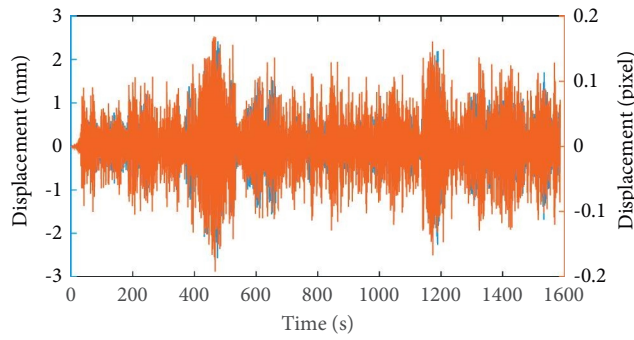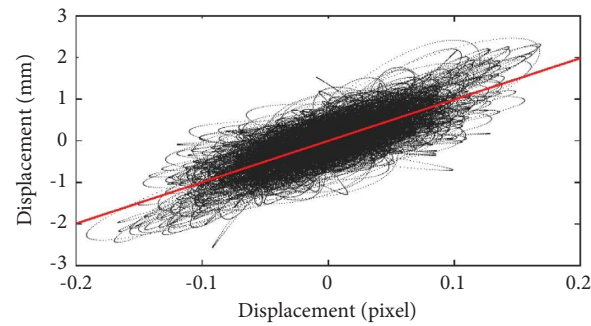| Mode order	Frequency (Hz)	Damping rate (%)
1	0.391	2.319
2	0.684	1.457
3	0.952	1.291
4	1.221	0.616
5	1.590	3.006

FIGURE 45: Acceleration spectrum.

FIGURE 46: The acceleration-based and vision-based displacements after bandpass filtering.

· Raw data
— Linear fitting: y=9.925x-0.002596
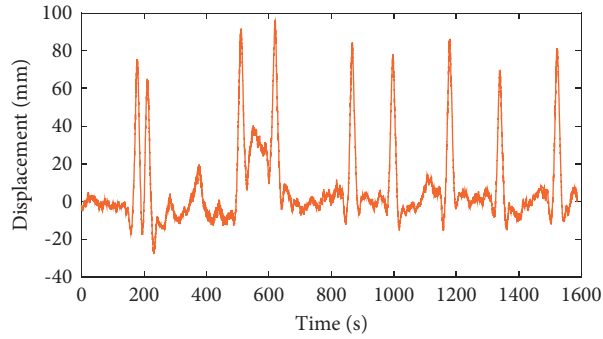
FIGURE 47: SF estimation.

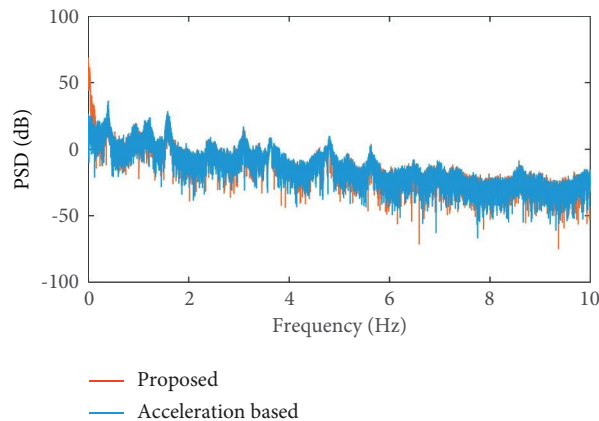FIGURE 48: Displacement estimation by data fusion.

FIGURE 49: PSD results of data fusion and acceleration-based displacements.

low frequencies and accelerations sampled at high frequencies.

7. Conclusions

Vision-based structural displacement monitoring system for SHM purposes offers advantages of noncontact, cost-efficiency, and remote sensing. In this study, we explored the potential of using tilt cameras and accelerometers to make up for the shortcomings of traditional vision-based systems to achieve accurate structural displacement monitoring. The proposed displacement monitoring method was involved in extracting the vision-based displacement from the camera, dynamic displacement reconstruction from the measured acceleration, SF estimation by the vision-based and acceleration-based displacement in the same frequency band, and data fusion of the low-frequency vision-based displacement and high-frequency acceleration-based displacements.

We validated the proposed method with both laboratory tests on a suspension bridge model and field tests on a highway and subway cable-stayed bridge. The experimental results indicated that the proposed data fusion method could obtain displacements close to the LVDT measurement compared with the traditional vision-based displacements. In the field test, we verified the performance of the proposed method by monitoring the displacement of the target bridge during the trains passing.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] P. S. Harvey and G. Elisha, "Vision-based vibration monitoring using existing cameras installed within a building," *Structural Control and Health Monitoring*, vol. 25, no. 11, p. 2235, 2018.

[2] A. Zare Hosseinzadeh, M. H. Tehrani, and P. S. Harvey, "Modal identification of building structures using vision-based measurements from multiple interior surveillance cameras," *Engineering Structures*, vol. 228, Article ID 111517, 2021.

[3] D. Feng and M. Q. Feng, "Experimental validation of cost-effective vision-based structural health monitoring," *Mechanical Systems and Signal Processing*, vol. 88, pp. 199–211, 2017.

[4] S. Yu and J. Zhang, "Fast bridge deflection monitoring through an improved feature tracing algorithm," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 3, pp. 292–302, 2019.

[5] L. Luo, M. Q. Feng, and Z. Y. Wu, "Robust vision sensor for multi-point displacement monitoring of bridges in the field," *Engineering Structures*, vol. 163, pp. 255–266, 2018.

[6] T. Jiang, G. T. Frøseth, A. Rønnquist, and E. Fagerholt, "A robust line-tracking photogrammetry method for uplift measurements of railway catenary systems in noisy backgrounds," *Mechanical Systems and Signal Processing*, vol. 144, Article ID 106888, 2020.

[7] T. Jiang, A. Rønnquist, Y. Song, G. T. Frøseth, and P. Nåvik, "A detailed investigation of uplift and damping of a railway catenary span in traffic using a vision-based line-tracking system," *Journal of Sound and Vibration*, vol. 527, Article ID 116875, 2022.

[8] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[9] S. Kim, G. Lee, S. Ahn, H. K. Kim, H. Yoon, and H. Yoon, "Vision-based cable displacement measurement using side view video," *Sensors*, vol. 22, no. 3, Article ID 22030, 2022.

[10] Y. Xu, J. Brownjohn, and D. Kong, "A non-contact vision-based system for multipoint displacement monitoring in a cable-stayed footbridge," *Structural Control and Health Monitoring*, vol. 25, no. 5, p. 2155, 2018.

[11] L.-J. Wu, F. Casciati, and S. Casciati, "Dynamic testing of a laboratory model via vision-based sensing," *Engineering Structures*, vol. 60, pp. 113–125, 2014.

[12] C.-Z. Dong, O. Celik, and F. N. Catbas, "Marker free monitoring of the grandstand structures and modal identification using computer vision methods," *Structural Health Monitoring*, vol. 18, no. 5-6, pp. 1491–1509, 2018.

[13] D. Feng, M. Q. Feng, E. Ozer, and Y. Fukuda, "A vision-based sensor for noncontact structural displacement measurement," *Sensors*, vol. 15, no. 7, pp. 16557–16575, 2015.

[14] T. Khuc and F. N. Catbas, "Computer vision-based displacement and vibration monitoring without using physical target on structures," *Structure and Infrastructure Engineering*, vol. 13, no. 4, pp. 505–516, 2016.

[15] T. Khuc and F. N. Catbas, "Completely contactless structural health monitoring of real-life structures using cameras and computer vision," *Structural Control and Health Monitoring*, vol. 24, no. 1, p. 1852, 2016.

[16] Y. Xu and J. M. W. Brownjohn, "Review of machine-vision based methodologies for displacement measurement in civil structures," *Journal of Civil Structural Health Monitoring*, vol. 8, no. 1, pp. 91–110, 2018.

[17] K. Patil, V. Srivastava, and J. Baqersad, "A multi-view optical technique to obtain mode shapes of structures," *Measurement*, vol. 122, pp. 358–367, 2018.

[18] H. S. Lee, Y. H. Hong, and H. W. Park, "Design of an FIR filter for the displacement reconstruction using measured acceleration in low-frequency dominant structures," *International Journal for Numerical Methods in Engineering*, vol. 82, no. 4, pp. 403–434, 2009.

[19] Y. H. Hong, S. G. Lee, and H. S. Lee, "Design of the FEM-FIR filter for displacement reconstruction using accelerations and displacements measured at different sampling rates,"

[20] K.-T. Park, S.-H. Kim, H.-S. Park, and K.-W. Lee, "The determination of bridge displacement using measured acceleration," *Engineering Structures*, vol. 27, no. 3, pp. 371–378, 2005.

[21] Y. K. Thong, M. S. Woolfson, J. A. Crowe, B. R. Hayes-Gill, and D. A. Jones, "Numerical double integration of acceleration measurements in noise," *Measurement*, vol. 36, no. 1, pp. 73–92, 2004.

[22] Z. Ma, J. Chung, P. Liu, and H. Sohn, "Bridge displacement estimation by fusing accelerometer and strain gauge measurements," *Structural Control and Health Monitoring*, vol. 28, no. 6, p. 7733, 2021.

[23] Z. Ma, J. Choi, and H. Sohn, "Real-time structural displacement estimation by fusing asynchronous acceleration and computer vision measurements," *Computer-Aided Civil and Infrastructure Engineering*, vol. 37, no. 6, pp. 688–703, 2021.

[24] Z. Ma, J. Choi, P. Liu, and H. Sohn, "Structural displacement estimation by fusing vision camera and accelerometer using hybrid computer vision algorithm and adaptive multi-rate Kalman filter," *Automation in Construction*, vol. 140, Article ID 104338, 2022.

[25] J. Kim, K. Kim, and H. Sohn, "Autonomous dynamic displacement estimation from data fusion of acceleration and intermittent displacement measurements," *Mechanical Systems and Signal Processing*, vol. 42, no. 1–2, pp. 194–205, 2014.

[26] K. Kim, J. Choi, G. Koo, and H. Sohn, "Dynamic displacement estimation by fusing biased high-sampling rate acceleration and low-sampling rate displacement measurements using twostage Kalman estimator," *Smart Structures and Systems*, vol. 17, no. 4, pp. 647–667, 2016.

[27] J.-W. Park, D. S. Moon, H. Yoon, F. Gomez, B. F. Spencer Jr, and J. R. Kim, Visual–inertial displacement sensing using data fusion of vision-based displacement with acceleration," *Structural Control and Health Monitoring*, vol. 25, no. 3, p. 2122, 2017.

[28] Y. Xu, J. M. W. Brownjohn, D. Hester, and K. Y. Koo, "Long-span bridges: enhanced data fusion of GPS displacement and deck accelerations," *Engineering Structures*, vol. 147, no. 9, pp. 639–651, 2017.

[29] Y. Xu, J. M. Brownjohn, and F. Huseynov, "Accurate deformation monitoring on bridge structures using a cost-effective sensing system combined with a camera and accelerometers: case study," *Journal of Bridge Engineering*, vol. 24, no. 1, Article ID 5018014, 2019.

[30] T. Wu, L. Tang, X. Li, X. Zhang, Y. Liu, and Z. Zhou, "Sparse accelerometer-aided computer vision technology for the accurate full-field displacement estimation of beam-type bridge structures," *Measurement*, vol. 25, Article ID 112532, 2023.

[31] B. Herbert, T. Tuytelaars, and L. Van Gool, "SURF: speed up robust features," *European Conference on Computer Vision*, pp. 404–417, Springer, Berlin, Germany, 2006.

[32] S. Leutenegger, M. Chli, and Y. Roland, "Siegwart. BRISK: binary robust invariant scalable key points," in *Proceedings of the International Conference on Computer Vision*, pp. 2548–2555, Barcelona, Spain, November 2011.

[33] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Readings in Computer Vision*, vol. 24, no. 6, pp. 726–740, 1987.

[34] T. Wu, L. Tang, P. Du, N. Liu, Z. Zhou, and X. Qi, "Non-contact measurement method of beam vibration with laser

stripe tracking based on tilt photography," *Measurement*, vol. 187, Article ID 110314, 2022.

[35] S. Kaneko, Y. Satoh, and S. Igarashi, "Using selective correlation coefficient for robust image registration," *Pattern Recognition*, vol. 36, no. 5, pp. 1165–1173, 2003.

[36] B. Ni, D. Xiao, and S. L. Shah, "Time delay estimation for MIMO dynamical systems – with time-frequency domain analysis," *Journal of Process Control*, vol. 20, no. 1, pp. 83–94, 2010.

[37] A. Brandt, *Noise and Vibration Analysis: Signal Analysis and Experimental Procedures*, John Wiley and Sons, Hoboken, NJ, USA, 2011.

[38] G. Deng, Z. Zhou, S. Shao, X. Chu, and C. Jian, "A novel dense full-field displacement monitoring method based on image sequences and optical flow algorithm," *Applied Sciences*, vol. 10, no. 6, Article ID 1006, 2020.

[39] G. Deng, Z. Zhou, S. Shao, X. Chu, and P. Du, "Novel approach to extract dense full-field dynamic parameters of large-scale bridges using spatial sequence video," *Journal of Civil Engineering and Management*, vol. 27, no. 8, pp. 617–636, 2021.

[40] S. Shao, Z. Zhou, G. Deng, P. Du, C. Jian, and Z. Yu, "Experiment of structural geometric morphology monitoring for bridges using holographic visual sensor," *Sensors*, vol. 20, no. 4, Article ID 2004, 2020.

[41] Z. Zhou, S. Shao, G. Deng, Y. Gao, X. Chu, and S. Wang, "Vision-based modal parameter identification for bridges using a novel holographic visual sensor," *Measurement*, vol. 179, Article ID 109551, 2021.

[42] B. Zhang, L. Zhou, and J. Zhang, "A methodology for obtaining spatiotemporal information of the vehicles on bridges based on computer vision," *Computer-Aided Civil and Infrastructure Engineering*, vol. 34, no. 6, pp. 471–487, 2019.

[43] Z. Chen, H. Li, Y. Bao, N. Li, and Y. Jin, "Identification of spatiotemporal distribution of vehicle loads on long-span bridges using computer vision technology," *Structural Control and Health Monitoring*, vol. 23, no. 3, pp. 517–534, 2015.