WILEY | Hindawi

Research Article

KCF-Based Identification Approach for Vibration Displacement of Double-Column Bents under Various Earthquakes

Wei Cai,[1] Wen Xie [ID],[1] and Tiantao He[2]

[1]*Department of Civil Engineering, Ningbo University, Ningbo 315211, China*
[2]*Ningbo Municipal Facilities Center, Ningbo 315010, China*

Correspondence should be addressed to Wen Xie; xiewen@nbu.edu.cn

Vibration displacements are one of the most significant indicators in the health monitoring and condition assessment of bridges in the life cycle. The traditional monitoring means, such as contact sensors, have relatively high-cost and limited points for displacement measurement of bridges. This paper proposes a low-cost and non-contact monocular vision system based on the KCF algorithm to accurately and timely identify the vibration displacement of bridges. A conversion method associated with a scale ratio was established to cope with the loss of depth information in images when a monocular camera is used to monitor multiple targets in different depths of the field. A series of shaking table tests on a two-column pier with energy dissipation beams were conducted to verify the feasibility, accuracy, effectiveness, and robustness of the KCF-based identification approach. The results showed that the vibration displacements of the column identified by the monocular vision system based on the KCF algorithm are almost consistent with the measurement results obtained by the laser displacement sensors. The peak displacement discrepancies between both measurement methods are within 6% for all cases with different shaking amplitudes and earthquake waves. The RMSE of the displacement histories between both measurement methods is very low. The corresponding frequency spectra contents identified by the monocular vision system based on the KCF algorithm match well with the measurement counterparts recorded from the laser displacement sensors.

1. Introduction

The displacement of bridges is one of the most significant indicators reflecting its mechanical performance and operational status. Thus, it is crucial to use monitoring means or sensing technologies capable of timely obtaining accurate displacement of bridges. Displacement sensors mainly fall into two categories: contact displacement sensors (e.g., Linear variable displacement transducer (LVDT)) and non-contact displacement sensors (e.g., Global positioning system (GPS), Laser displacement sensor (LDS), microwave interferometric radar, and machine vision-based measurement method). The LVDT is fixed on a platform to measure the structural displacement. However, the fixed platform is hard to build when the LVDT is used to measure the displacement of large-span bridges crossing rivers and valleys

[1, 2]. The LVDT installation may cause a certain degree of damage to bridges. The vibration of the fixed platform is prone to causing measurement errors [3]. GPS measurement has low accuracy with an error of about 5 mm. It is challenging to monitor the vibration amplitudes within a few millimeters [4]. The successive vibration displacement measurements are also difficult to achieve due to the low sampling frequency of GPS [5]. Even though the LDS could overcome shortcomings of contact displacement sensors and improve measurement accuracy, its high price and installation platform limitations make it inconvenient to apply widely [6]. The microwave interferometer radar employs microwave signals for signal transmission and reception. It obtains the displacement of the bridge by analyzing the phase difference between reflected waves at a time interval. However, the temperature, humidity, and pressure in the air

may make the microwave signal time lag and bend during measurement to decrease measurement accuracy [7–10]. In addition, the installation and maintenance of conventional displacement sensors require a lot of human and financial resources [11].

Machine vision-based vibration displacement measurements are one of the research hotspots due to numerous advantages, such as higher measurement precision and sampling frequency, multi-point, non-contact, and long-distance measurements [12]. Also, machine vision-based vibration displacement measurements are not reliant on additional fixed platforms. Displacement is directly extracted from structural vibration videos recorded by cameras. Therefore, the methods are widely applied in displacement or deflection measurements [13–16] and modal recognition [17] of civil engineering structures. Machine vision-based displacement identification methods, including template matching methods [18–20], optical flow estimation methods [21–24], and correlation filtering methods [25–27], were developed in the past. Olaszek [28] established an imaging system based on the photogrammetric principle to determine the dynamic characteristic of bridges, indicating that the system reduced the influence of environmental factors on the acquisition of images. Hanssen et al. [29] developed a digital image correlation method based on the normalized cross-correlation coefficient (NCC) to measure the deflection of a three-point bending steel beam. Yu et al. [30] presented a fast and accurate machine vision-based measurement method to identify the deformation of a cantilever beam and the mid-span deflection of the Su-Tong Bridge. Yoon et al. [31] applied Kanade-Lucas-Tomasi (KLT) algorithm in measuring the displacement of a six-story-high building model and verified the algorithm by comparing the results from conventional sensors. Chen et al. [32] developed an optical flow method based on motion magnification, which was validated by the deflection of structures, cantilever beams, and pipes. Zhao et al. [33] built an approach that combined support correlation filters (SCF) and KLT to measure the vibration displacement of a cable-stayed bridge model. The results showed that the identification displacement is accurate by comparing the LDS displacement. Although the displacement identifications based on the template matching and optical flow methods achieved fruitful results, there are several shortcomings. For example, the template matching method needs high contrast or artificial targets to improve the measurement accuracy. However, the high contrast targets are often not present in actual bridges, and artificial targets may affect the structural appearance. The optical flow estimation method should meet the assumptions of constant brightness and continuous time or small motion. However, the approach cannot satisfy the abovementioned assumptions when it is used to monitor the displacement of large-span bridges. Although the optical flow method based on motion magnification could improve the identification accuracy, the efficiency of the methodology tracking and identifying the displacement of fast-moving structures reduces due to the high time complexity caused by a spatial band-pass filter.

The correlation filter (CF) could compensate for the shortcomings of the template matching and optical flow estimation methods, improve identification speed, accuracy, and robustness, and is widely applied in video tracking and identification. Bolme et al. [25] proposed the minimum output sum of squared errors filter (MOSSE) to compute the correlation response of targets by multiplication element-by-element in the Fourier domain, indicating that MOSSE significantly reduced the time complexity and could track the target with speeds of 600–700 frames per second. Henriques et al. [26] developed a circulant structure of tracking-by-detection with kernels (CSK). The CSK employs a circular matrix to obtain dense samples and the corresponding feature contents. The CSK also uses kernel functions to improve computational efficiency. Henriques et al. [27] further proposed a kernelized correlation filter (KCF) based on CSK. The KCF replaces the single-channel grayscale features in CSK with a multi-channel histogram of oriented gradients (HOG), which could enhance the expression capability of samples and the accuracy and robustness of the KCF. Du et al. [34] presented a KCF tracker integrating a three-frame-difference algorithm, which identified target displacements in the ultra-high-resolution video (3840×2160 pixels and 3600×2700 pixels). Shao et al. [35] established a velocity correlation filter (VCF) using velocity features and an inertia mechanism (IM). Yang et al. [36] embedded the KCF tracker into an unmanned aerial vehicle (UAV) tracking platform. The KCF tracker maintained high operating speed due to its low time complexity. However, its computational power was limited. Chen et al. [37] built a KCF tracking framework based on the curve fitting algorithm and evaluated the root mean square error and mean absolute deviation between the tracking displacement and theoretical displacement. Zheng and Gupta [38] proposed a multi-camera multi-target tracking system based on the KCF algorithm and the improved KCF algorithm. However, there are few publications on the KCF algorithm tracking and identifying the vibration displacement of civil structures. There is a lack of research on the KCF algorithm applied in the displacement measurement of bridge structures.

Therefore, this paper proposes a monocular vision system based on the KCF algorithm to identify the vibration displacement of bridges. First, the system establishes the relationship between physical space and pixel coordinates by camera calibration. A region of interest (ROI) containing the target is determined. The KCF algorithm is used to identify the vibration displacement of the target. The target could be artificial targets, such as 2-D codes, geometric patterns, and artificial light sources, or natural targets, such as pits, bolts, and rivets of the structural surface. Finally, the pixel coordinate displacements of the target in each frame are transformed into the physical space displacements of the target by a scale ratio. Consequently, the vibration displacement of the structure is identified. In this paper, the main contents are as follows: (1) The KCF-based identification approach is introduced, and identification processes are presented in Section 2. (2) In Section 3, the KCF-based identification approach is verified by shake table tests on

a two-column pier with energy dissipation beams. The displacement identified from the proposed KCF-based identification approach is compared with the displacement recorded from the traditional LDS. The root-mean-square errors (RMSE) and peak displacement errors between both approaches are obtained to evaluate the accuracy, feasibility, and robustness of the proposed KCF-based identification approach. (3) The critical conclusions are extracted from the analyses and discussions in Section 4.

2. Vibration Displacement Measurement Method Based on KCF Algorithm

For a machine vision-based identification approach for vibration displacement of bridge structures, a commercial high-speed camera is used to record the target movement and videos of bridge structures. Then, a proposed KCF-based identification approach is employed to identify the vibration displacement of the target from the recorded video. The specific process (see Figure 1) is as follows:

(1) Calibrating camera to calculate the scale ratio: The step aims to accurately obtain the scale ratio (φ), which could describe the relationship between the pixel coordinate displacement, d_{img}, and the physical space displacement, d_{real}, of the target.

(2) Selecting the region of interest (ROI) and extracting the sample set features: The ROI, including the target, is determined for the tth frame of the video recorded by the camera. The ROI is used to obtain the sample set by a cyclic shift operator. The sample set information is composed of the FHOG features of the sample.

(3) Training and updating the filter tracker: For the tth frame of the video recorded by the camera, the sample set information, such as FHOG features, is applied to train and update the filter tracker.

(4) Acquiring the physical space displacement of targets: For the $(t + 1)^{\text{th}}$ frame of the video recorded by the camera, the filter tracker is used to detect the target position. Then the pixel coordinate displacements of the target are converted into the physical space displacements of the target.

Each step is discussed in depth in the following subsections.

2.1. Calibrating Camera to Calculate the Scale Ratio.
Camera calibration is an indispensable step in machine vision-based displacement measurement. The purpose of camera calibration is to obtain the scale ratio between the pixel coordinate displacement, d_{img}, and the physical space displacement, d_{real}, of the target. Simultaneously, the camera calibration could eliminate the influences of the geometric aberrations caused by optical imaging through the calibration method proposed by Zhang [39]. However, only scale ratio calculation is highlighted in this section. The calculation principle of the scale ratio could be presented by

the pinhole model (see Figure 2). Namely, the physical space displacement of the target is proportional to the pixel coordinate displacement of the target, which is only related to the distance between the camera and the target when the optical axis of the camera is perpendicular to the plane of the target. The scale ratio, φ, can be determined according to the mapping relationship between pixel and physical space length.

$$\varphi = \frac{L_{\text{AB}}}{L_{\text{ab}}}, \tag{1}$$

where L_{AB} is the physical space length of the target, L_{ab} is the pixel length on the image plane projected by a physical space length.

Multi-targets in different depths of field for the bridge structures inevitably need to be simultaneously monitored by the monocular camera, resulting in the depth information of the image loss. Even if the sizes of the structural targets in physical space with various depths of field are different, the structural targets are mapped to the same pixel length on the image plane (Figure 2(b)). Therefore, various scale ratios, φ_ι, need to be calculated for structural targets in physical space at different depths of the field.

$$\varphi_\iota = \frac{L_{A_\iota B_\iota}}{L_{\text{ab}}}, \tag{2}$$

where $L_{A_\iota B_\iota}$ denotes the physical space length of the structure surface with different depths of field and ι represents the ιth depth of field plane.

2.2. Selecting the ROI and Extracting the Sample Features

2.2.1. ROI Acquisition.
An adaptive method was established to adjust the display window of the video with high resolution to satisfy the display of the computer monitor with lower resolution. The KCF-based tracking algorithms identify the structural targets within the ROI. Therefore, ROI acquisition is one of the critical steps. The ROI is a two-dimensional image containing the target. An ideal ROI could make the target tracking and identification take less processing time and improve accuracy and robustness. The key steps generating the ROI in the KCF-based tracking algorithm are as follows:

(1) Selecting and extending the ROI (selected by mouse). The selected ROI is expanded to N times (e.g., $N = 2.5$ [27]). The ROI expansion could prevent the target from being decomposed and reconstructed during cyclic shift sampling. The expansion operation could also increase the weight of the pixel in the target edge for the feature extraction because the feature extraction ignores the boundary element.

(2) Resampling the ROI resolution to obtain the appropriate size. The expanded ROI increases the number of internal pixels of the target, resulting in the operation speed of the tracking algorithm slowing down. However, the running speed of the KCF-based tracking algorithm could be improved by
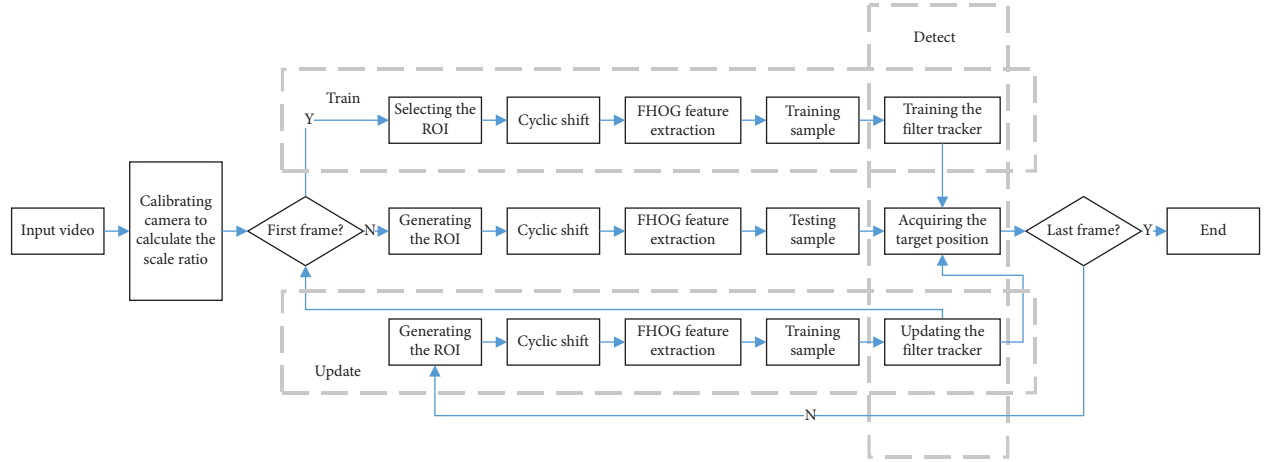
FIGURE 1: The process of the Vibration displacement measurement method based on KCF algorithm.

(a)   (b)
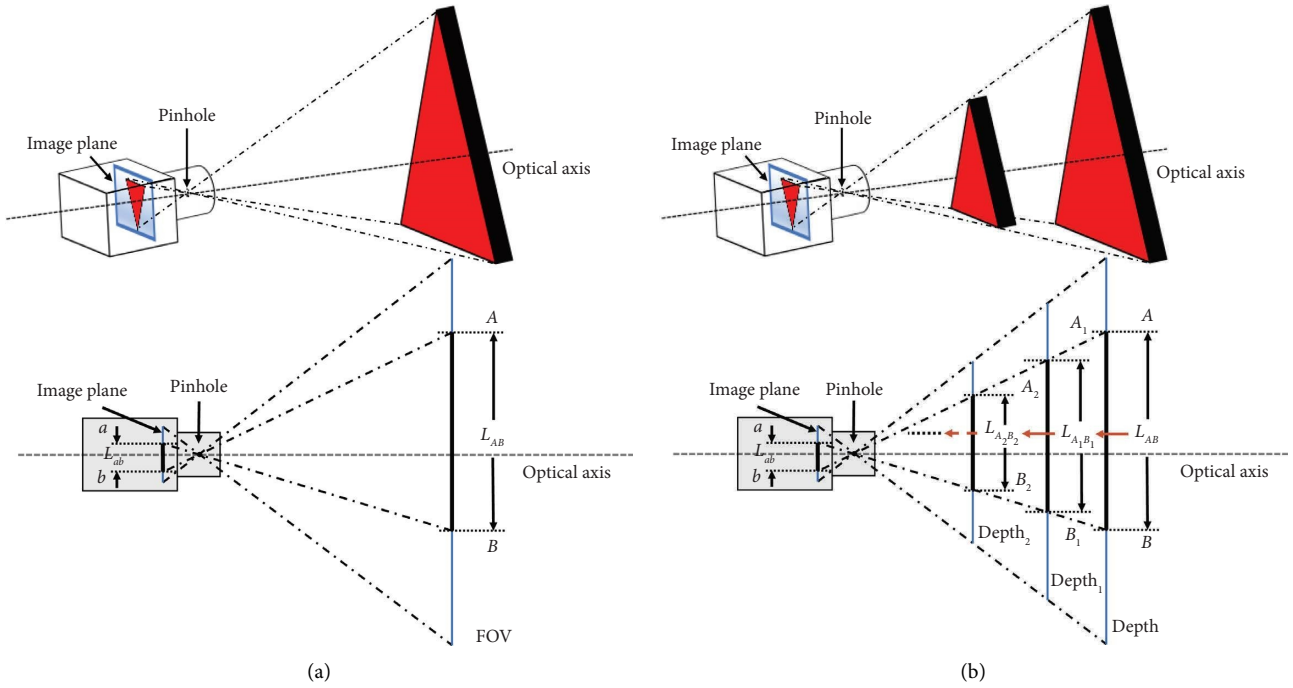
FIGURE 2: Pinhole model: (a) pinhole model for a single target; (b) pinhole model for multi targets in different depth of field.

the bilinear interpolation sampling adjusting the ROI resolution. Simultaneously, the primary feature contents of the target within the ROI could be preserved.

2.2.2. Cyclic Shift.
The ROI is sampled by a cyclic shift operator established in the KCF-based tracking algorithm to collect more sample data for training the filter tracker. The process of the cyclic shift is relatively complex for the ROI. Therefore, the computational procedures and principles of the cyclic shift operator are introduced in a one-dimensional case. Namely, a one-dimensional vector (called the base sample) is given in the equation (3). The cyclic shift operator is presented by a specific matrix to shift the one-dimensional

vector. The base sample vector, u, and cyclic shift operator matrix, R, are defined as follows.

$$u = [v_1, v_2, v_3, \ldots, v_m], \tag{3}$$

$$R = \begin{pmatrix} 0-1-0-\cdots & 0 \\ 0-0-1-\cdots & 0 \\ \vdots - \vdots - \vdots & -\vdots \\ 0-0-0-\cdots & 1 \\ 1-0-0-\cdots & 0 \end{pmatrix}. \tag{4}$$

The base sample vector, u, multiplied by the cyclic shift operator, R, equals to a negative sample, u_1, according to the equation (5). Namely, the first element of the base sample

vector is shifted to the second element of the negative sample, and the rightmost element is shifted to the leftmost. If the base sample vector separately multiplied by the cyclic shift operator of powers, R^n, a data matrix, U, is obtained, where $n = 0, 1, \ldots, m - 1$. Consequently, the data matrix, U, contains one base sample and $m - 1$ negative samples.

$$u_1 = uR^1 = [v_m, v_1, v_2, \ldots, v_{m-1}], \tag{5}$$

$$\{uR^n | n = 0, 1, \ldots, m - 1\}, \tag{6}$$

$$U = C(u) = \begin{pmatrix} v_1 - v_2 - v_3 - \cdots & v_m \\ v_m - v_1 - v_2 - \cdots & v_{m-1} \\ \vdots - \vdots - \vdots - & \vdots \\ v_3 - v_4 - v_5 - \cdots & v_2 \\ v_2 - v_3 - v_4 - \cdots & v_1 \end{pmatrix}. \tag{7}$$

Similarly, the derivative process of the one-dimensional vector could be extended to the ROI. The ROI in each frame multiplied by the cyclic shift operator is regarded as movement along the horizontal or vertical directions. The ROI is shifted using the cyclic shift operator to obtain the sample set. Figure 3 shows several typical samples. The positive and negative signs represent the shift down and up of the image sample, respectively. The number (such as 10 and 20) represents the times of shifts, and "0" represents the base sample for the image.

2.2.3. FHOG Feature Extraction. The KCF-based tracking algorithm extracts the features of the sample set to locate the target position in each frame. The features of the sample set should have invariance properties even if the sizes and posture of the target in the sample set and even environmental lighting change. However, how effectively describing the features of the sample set is a challenging task. Herein, the HOG feature [40] is used to present the features of the sample set, such as gradient features of the target. In the HOG feature extraction process, the image of the sample set is converted to a grayscale image, and the image contrast is modified by gamma correction to lessen the impact of uneven lighting. The gradients of the image are obtained from the convolution between the $[-1, 0, 1]$ and $[-1, 0, 1]^T$ operators and the image. The gradient values $G_p(p, q)$ and $G_q(p, q)$ of an image pixel in the horizontal and vertical directions are obtained according to equations (8) and (9), respectively.

$$G_p(p, q) = H(p + 1, q) - H(p - 1, q), \tag{8}$$

$$G_q(p, q) = H(p, q + 1) - H(p, q - 1), \tag{9}$$

where $H(p, q)$ denotes the gray value at the pixel coordinate (p, q).

The amplitude and direction of the gradient for a pixel is determined according to equations (10) and (11), respectively.

$$G(p, q) = \sqrt{\left(G_p(p, q)\right)^2 + \left(G_q(p, q)\right)^2}, \tag{10}$$

$$\theta(p, q) = \arctan \frac{G_p(p, q)}{G_q(p, q)}, \tag{11}$$

where $G(p, q)$ represents the amplitude of the gradient, and $\theta(p, q)$ denotes the direction of the gradient.

The image is discrete into several cells by the HOG feature. Each cell consists of 4×4 pixels, and adjacent four cells establish a block. The gradient direction of each cell is divided into unsigned and signed histograms using the HOG feature with a weighted method. The unsigned gradient direction histogram uniformly divides the gradient direction of each pixel into 9 bins in 0–180 degrees. The signed gradient direction histogram equably divides the gradient direction of each pixel into 18 bins in 0–360 degrees. The histograms in the block are normalized by four types of normalization methods, resulting in a 108-dimensional data acquisition. A higher dimensional data is obtained for the images composed of several blocks, leading to the processing with high time complexity. Therefore, it is importance to reduce the number of dimensions of the feature while retaining the primary feature contents.

The abovementioned high-dimensional data of the features could be a dimension reduction process by the FHOG method [41]. For example, the FHOG method could reduce the 108-dimensional HOG features to 31-dimensional FHOG features. The detailed dimension reduction process is shown in Figure 4.

As shown in Figure 4, the 27-dimensional features consist of the column accumulation of 4 normalization operators under 27 bins (including 9 unsigned gradient histograms and 18 signed gradient histograms). The 4-dimensional features consist of the row accumulation of 27 bins under 4 normalization operators.

2.3. Training and Updating the Filter Tracker. The core step of the KCF-based tracking algorithm is to train the filter tracker that is used to locate the target position in each frame. The computation, training, and updating of the filter tracker are taken as an example to process each frame. First, a Gaussian regression label, $y_i (y_i \subset [0, 1])$, in the Fourier domain is established by utilizing the training sample x_i. The Gaussian regression label gradually decreases as the number of cyclic shifts increases. It is worth noting that the target is itself when the Gaussian regression label equals 1. The training sample set (acquisition process as shown in Figure 5(a)) is used to train the filter tracker to obtain the equation (12) in the tth frame, which minimizes the squared error over the training sample x_i and its Gaussian regression label y_i, as presented in the equation (13).

$$f(x_i) = \omega^T x_i, \tag{12}$$

$$\min_\omega \sum_i (f(x_i) - y_i)^2 + \lambda \|\omega\|^2, \tag{13}$$

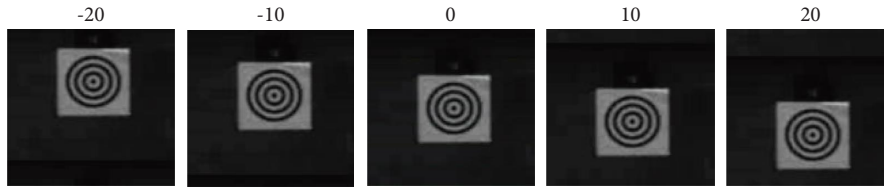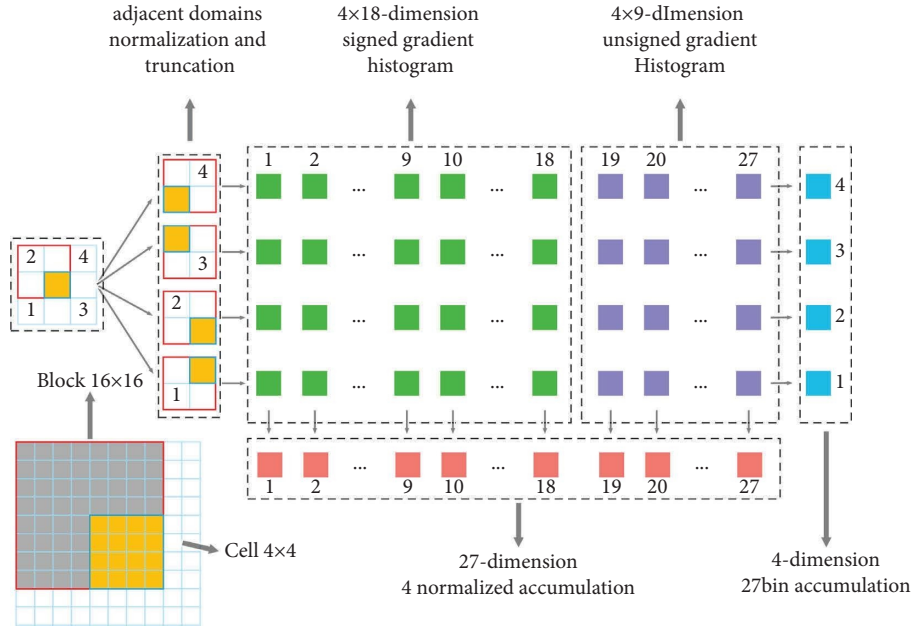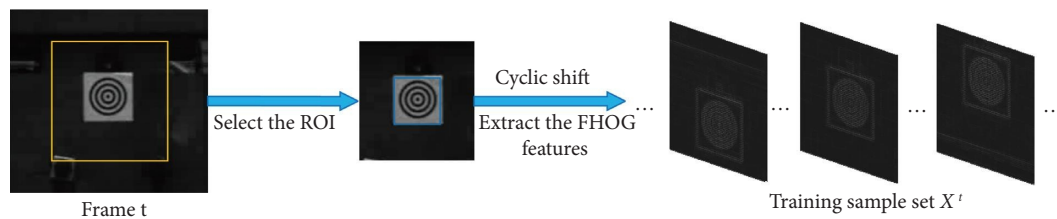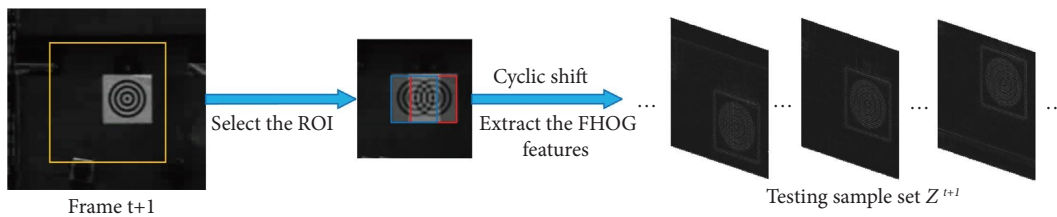Figure 3: The ROI after specific shifts along the vertical direction.

Figure 4: FHOG feature extraction process.

Frame t

☐ The ROI in each frame
☐ The target in the frame t
☐ The target in the frame t+1

(a)

Frame t+1

☐ The ROI in each frame
☐ The target in the frame t
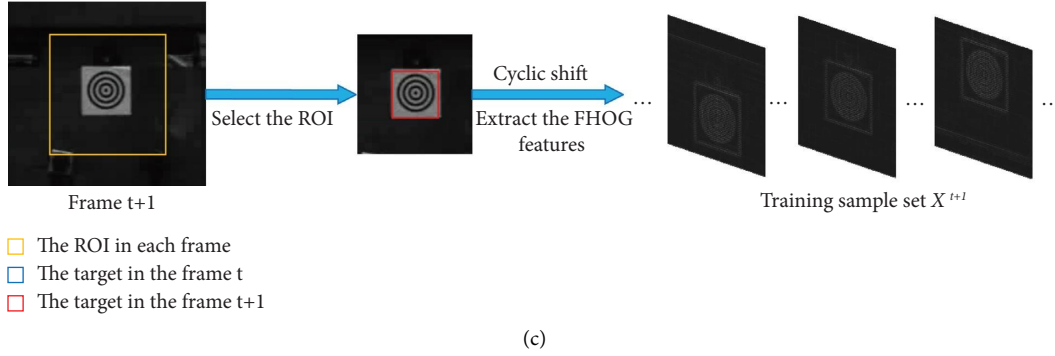☐ The target in the frame t+1

(b)

Figure 5: Continued.

(c)

FIGURE 5: Acquisition process of sample sets: (a) training sample set in the tth frame; (b) testing sample set in the $(t + 1)^{\text{th}}$ frame; (c) training sample set in the $(t + 1)^{\text{th}}$ frame.

where λ represents the regularization parameter to prevent overfitting, and ω denotes the regression coefficient.

The unique closed-form solution of ω is presented as follow.

$$\omega = \left(X^H X + \lambda I\right)^{-1} X^H y, \tag{14}$$

where X is the training sample set composed of training samples x_i, $X^H = (X^*)^T$ is the Hermitian transpose, X^* is the complex-conjugate matrix of X, I is the unit matrix, and y is the vector composed of Gaussian regression labels y_i.

To improve the classification performance of the filter tracker, a mapping ϕ is used to map the training samples x_i into Hilbert space. Simultaneously, a kernel function, as presented in the equation (15), is introduced for optimization. According to the Representer Theorem [42], the regression coefficient ω can be expressed as a linear combination of the mapped samples, as presented in the equation (16). Therefore, the equation (12) is transformed into the equation (17).

$$k\left(x_i, x_j\right) = \phi^T\left(x_i\right)\phi\left(x_j\right), \tag{15}$$

$$\omega = \sum \alpha_i \phi\left(x_i\right), \tag{16}$$

$$f\left(x_j\right) = \sum \alpha_i \phi\left(x_i\right)\phi\left(x_j\right) = \alpha^T k\left(x_i, x_j\right), \tag{17}$$

where α_i is the combination coefficient, and α is the vector composed of α_i. The solution of ω is transformed into the solution of α in the dual space.

The solution of the kernelized version of α in the dual space is presented as follows [43].

$$\alpha = \left(K^{\text{xx}} + \lambda I\right)^{-1} y, \tag{18}$$

where K^{xx} is the kernel matrix of the training sample, namely, $K_{ij}^{\text{xx}} = k\left(x_i, x_j\right)$.

The kernel matrix K^{xx} has the structure of the cyclic matrix [26], which could be further optimized in the Fourier domain by the following equation.

$$\widehat{\alpha} = \frac{\widehat{y}}{\widehat{k}^{\text{xx}} + \lambda}, \tag{19}$$

where the symbol ˆ denotes the Discrete Fourier Transform (DFT) of the variable, k^{xx} is the kernel matrix K^{xx} in the Fourier domain, and $\widehat{\alpha}$ is the filter tracker in the Fourier domain.

Due to the robustness requirements of the KCF-based tracking algorithm, the filter tracker $\widehat{\alpha}$ needs to be updated in each frame. The filter tracker is updated by the following equation

$$\widehat{\alpha}_{t+1} = (1 - 0.012) \times \widehat{\alpha}_{t-1} + 0.012 \times \widehat{\alpha}_t, \tag{20}$$

where $\widehat{\alpha}_{t+1}$ is the filter tracker for the $(t + 1)^{\text{th}}$ frame, $\widehat{\alpha}_t$ is the filter tracker for the tth current frame, and $\widehat{\alpha}_{t-1}$ is the filter tracker for the $(t - 1)^{\text{th}}$ previous frame.

2.4. Acquiring the Target Physical Space Displacement. In the $(t + 1)^{\text{th}}$ frame, the response of the testing sample set (acquisition process as shown in Figure 5(b)) is detected by the filter tracker $\widehat{\alpha}_{t+1}$ in the equation (20), as presented in the equation

$$f(z) = F^{-1}\left(\widehat{k}^{\text{xz}} \odot \widehat{\alpha}\right), \tag{21}$$

where the symbol ˆ denotes the DFT for the variable, and F^{-1} is the Fourier inverse transform matrix. k^{xz} is the kernel matrix K^{xz} in the Fourier domain, and $f(z)$ is the response set of the testing sample set Z.

The position of the target in the $(t + 1)^{\text{th}}$ frame could be located by the largest response in the equation (21). The pixel coordinate displacement d_{img} of the target is determined by difference between the target coordinate of the $(t + 1)^{\text{th}}$ frame and the target coordinate of the first frame. The physical space displacement d_{real} of the target in the $(t + 1)^{\text{th}}$ frame is calculated by the scale ratio φ, as presented in the equation

$$d_{\text{real}} = \varphi\left(d_{\text{img}}\right). \tag{22}$$

During long-term health monitoring and even shaking table tests on bridges, the identified vibration displacements may drift from the baseline due to environmental noise or other uncertain factors. Therefore, the KCF algorithm was improved to eliminate the baseline drift. Consequently, the
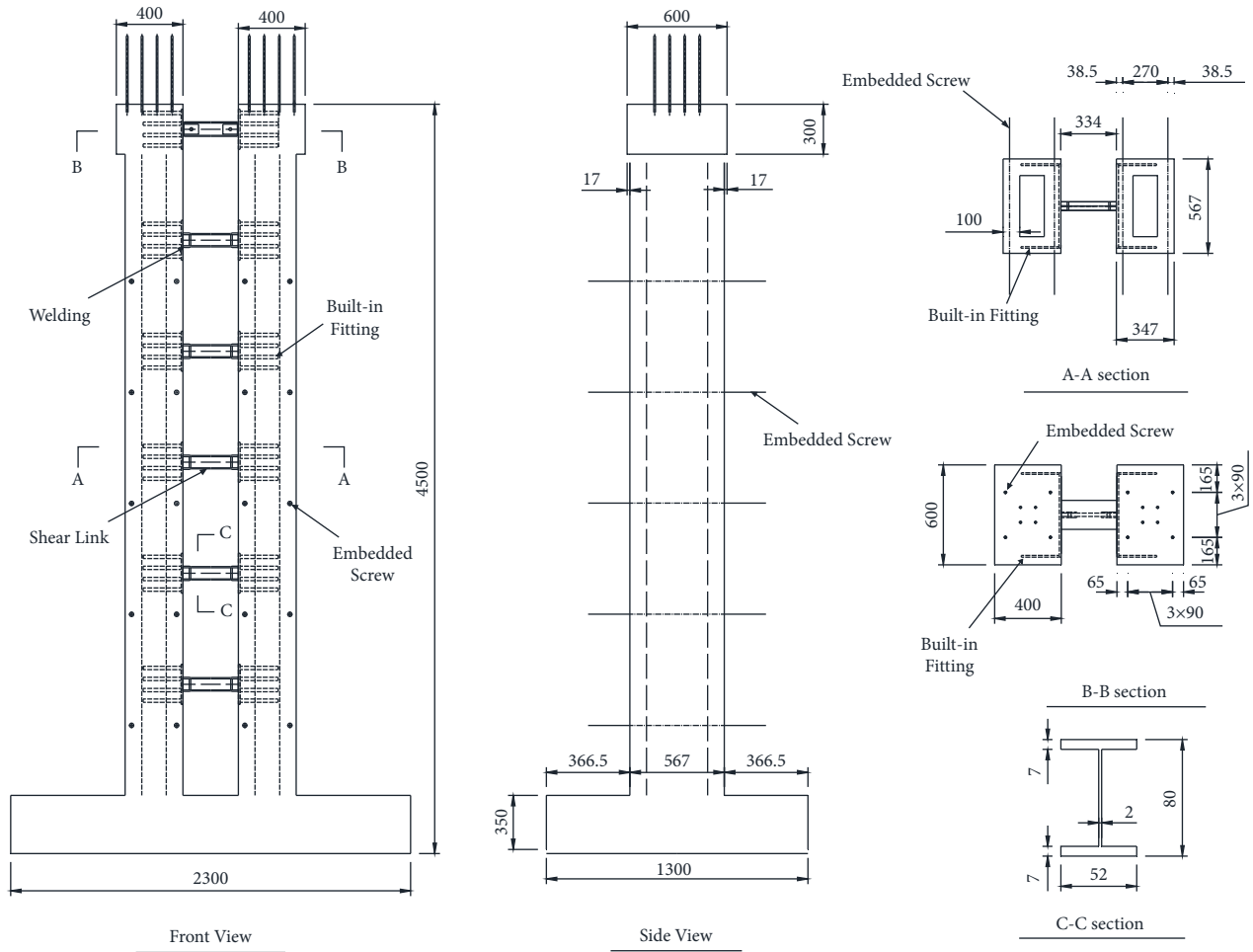
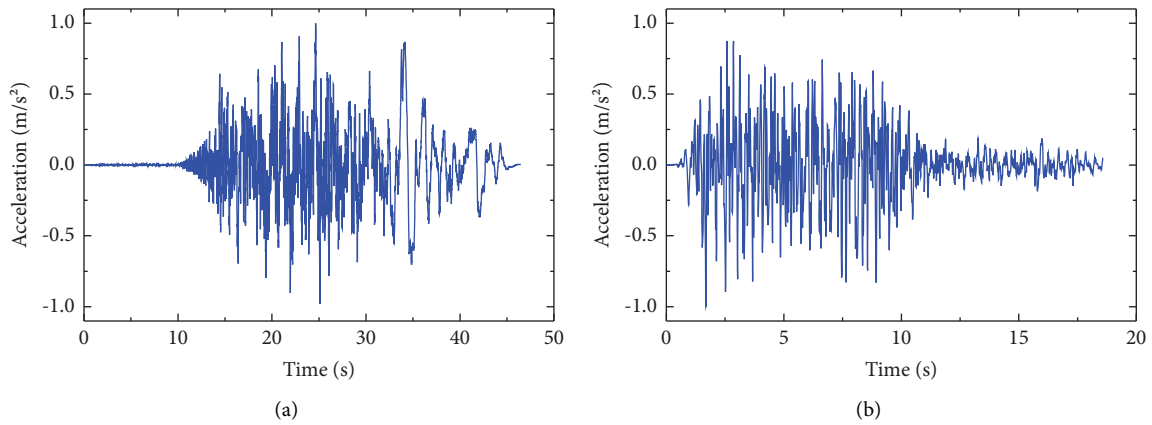FIGURE 6: Shaking table model of the two-column pier with energy dissipation beams (mm).

(a)

(b)

FIGURE 7: Continued.

(c)

(d)

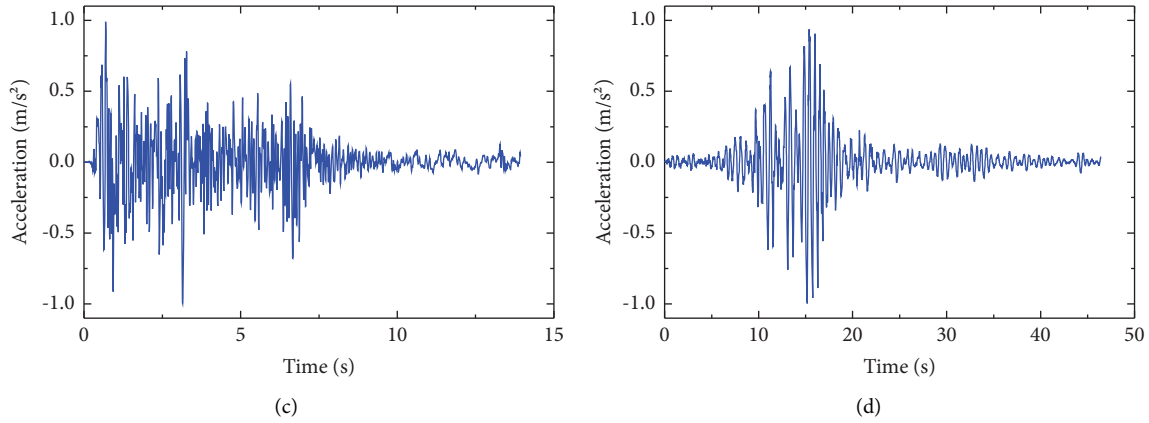Figure 7: Acceleration time histories of seismic waves. (a) Chi-Chi wave. (b) Artificial wave. (c) EI-Centro wave. (d) Mexico-City wave.
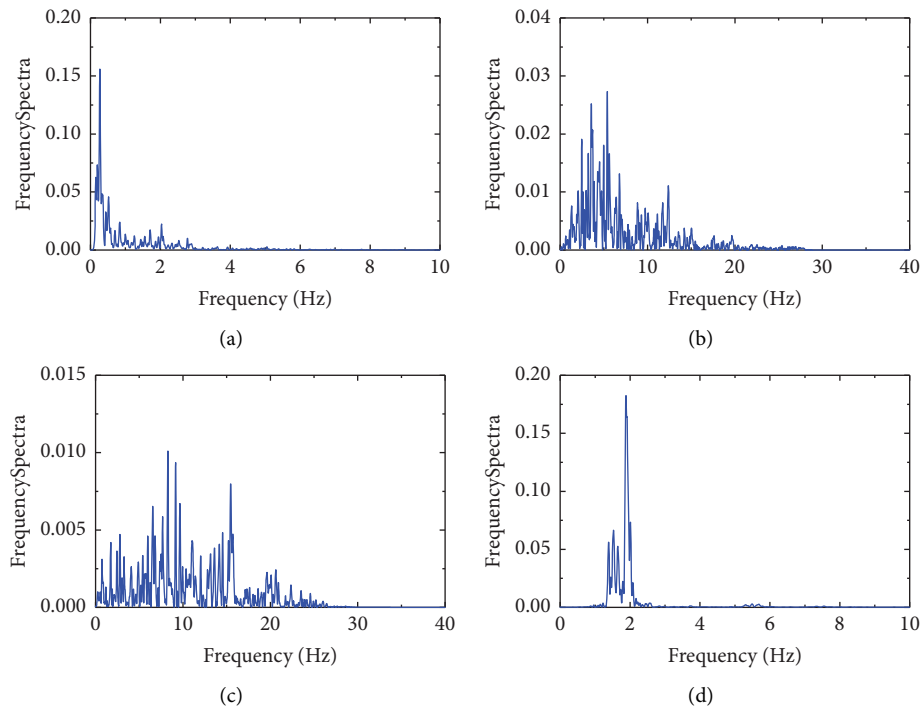

(a)

(b)

(c)

(d)

Figure 8: Acceleration frequency spectra of seismic waves. (a) Chi-Chi wave. (b) Artificial wave. (c) EI-Centro wave. (d) Mexico-City wave.

Table 1: All test cases for the test model.

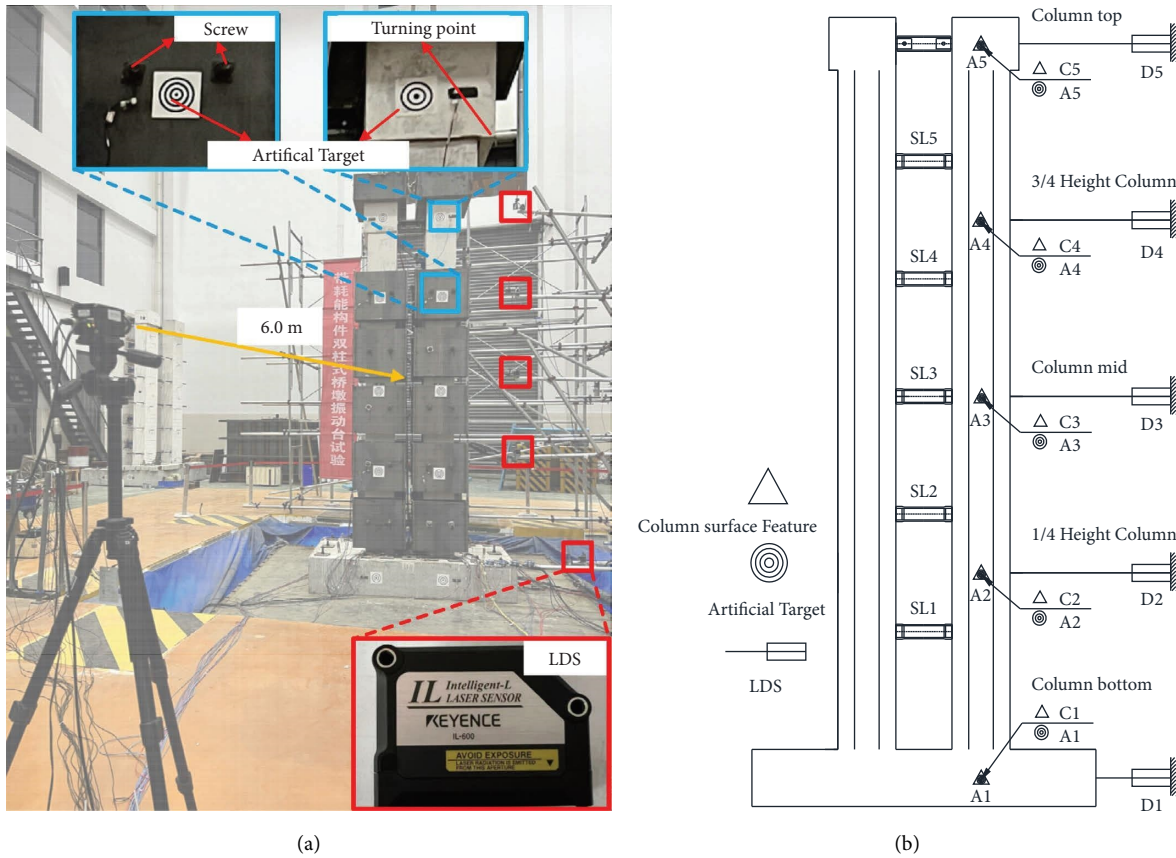| Case number	Seismic waves	Inputting direction	PGA/g
1			0.10
2			0.20
3			0.30
4			0.40
5	Chi-Chi wave	Transverse direction (Y)	0.50
6			0.60
7			0.68
8			0.80
9			0.90
10	Artificial wave	Transverse direction (Y)	0.68
11	EI-Centro wave	Transverse direction (Y)	0.68
12	Mexico-City wave	Transverse direction (Y)	0.68

(a)

(b)

FIGURE 9: Targets and laser displacement sensors layouts of the test model: (a) test site; (b) measurement positions.
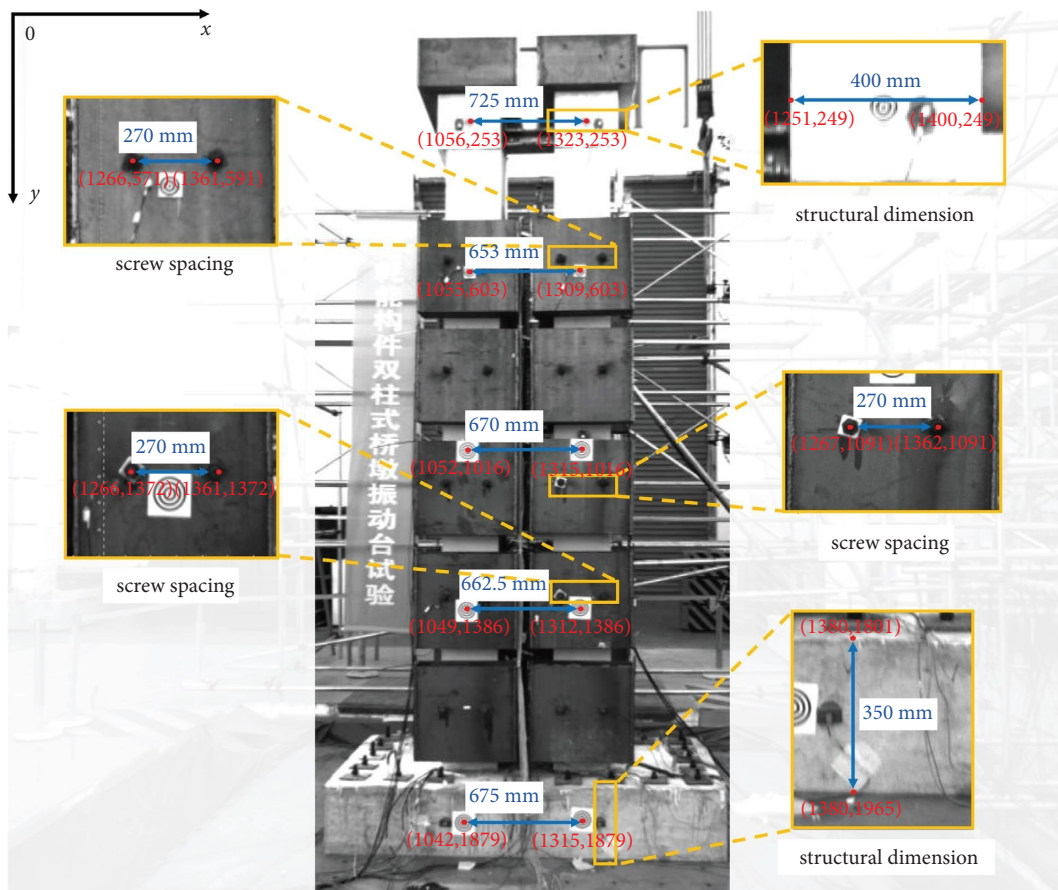

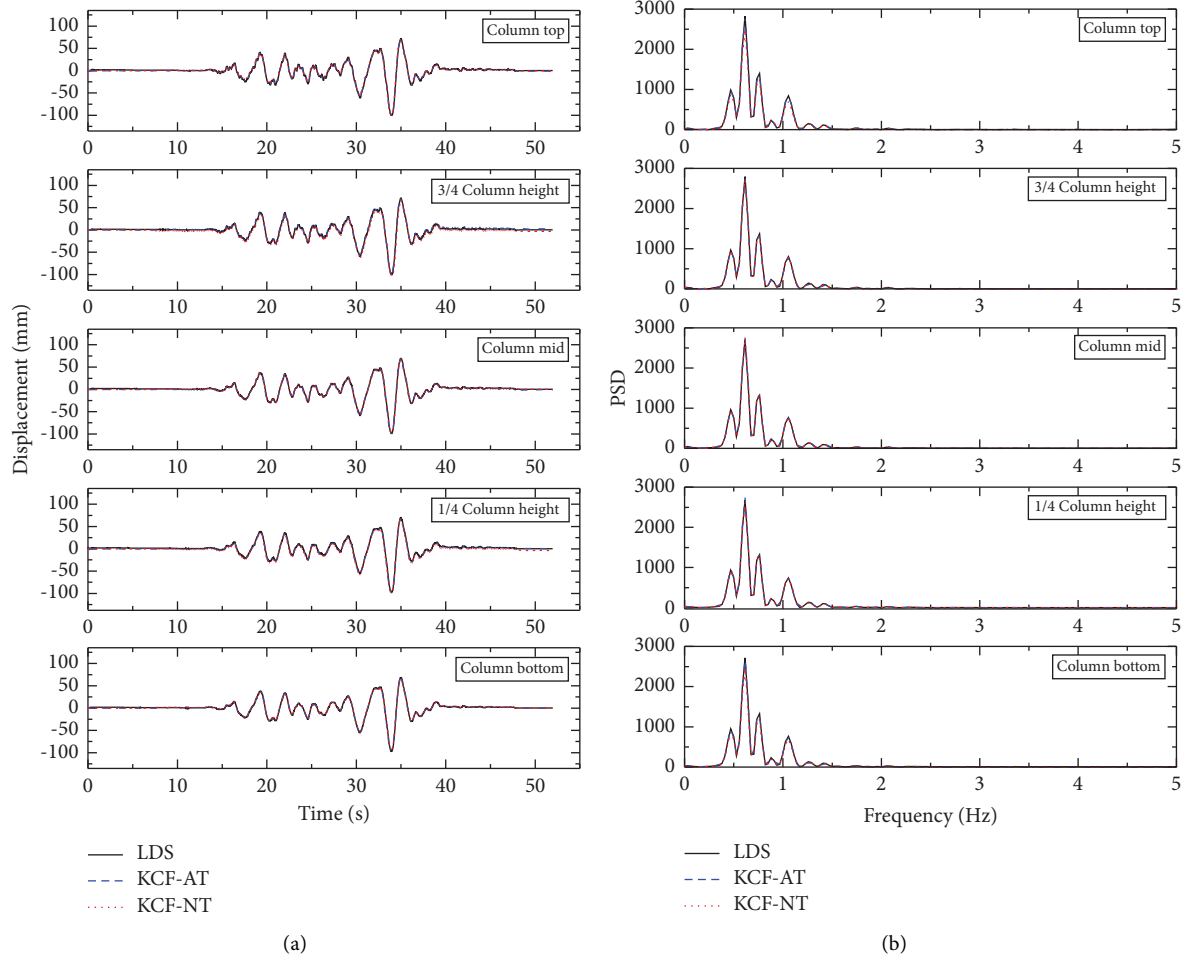FIGURE 10: Image captured by the high-speed camera.

FIGURE 11: Displacement responses and the corresponding frequency contents along the column height of Case 1. (a) Displacement responses. (b) Frequency contents of displacement.

robustness and accuracy of the KCF algorithm identifying the vibration displacements improve.

3. Experiments and Validation

3.1. Experimental Schemes. Shaking table tests on a two-column pier with energy dissipation beams were conducted to verify the feasibility, accuracy, effectiveness, and robustness of the KCF-based identification approach. The test model of the two-column pier with energy dissipation beams was designed and built according to the similarity ratios. The similarity ratios of the geometric and elasticity modulus are 1 : 15 and 0.3 : 1, respectively. The total height of the test model is 4500 mm. Each column is a box cross section. The geometric dimensions of the box cross section are 567 × 347 mm, and the wall thickness is 100 mm (see Figure 6). Five I-type energy dissipation beams made of low-yield steel were equally installed between both columns. The cross section is composed of the flange and web. The flange and web thickness are 7 mm and 2 mm, respectively. The flange wideness and web height are 52 mm and 66 mm, respectively. The test model was built using the HRB400 steel bar with 8 mm diameter and M15 cement mortar. The longitudinal reinforcement ratio is 1.526%. The additional counterweight, 9993 kg, was installed along the height of the column to satisfy the dynamic similarity requirement.

Seismic waves should be reasonably selected as vibration inputs for the shaking table tests. Therefore, a typical Chi-Chi wave was chosen as the vibration input because the wave with pulse effects may markedly influence the seismic response of the test model. The accuracy and feasibility of the KCF-based identification approach were investigated by gradually increasing the peak ground motion acceleration (PGA) of the Chi-Chi wave. Furthermore, the different frequency contents of other seismic waves, such as the Artificial wave, the El-Centro wave, and the Mexico City wave, were selected as vibration inputs to evaluate the effectiveness and robustness of the KCF-based identification approach. It is worth noting that all seismic waves must be compressed by the time similarity ratio of 0.2582 to consider
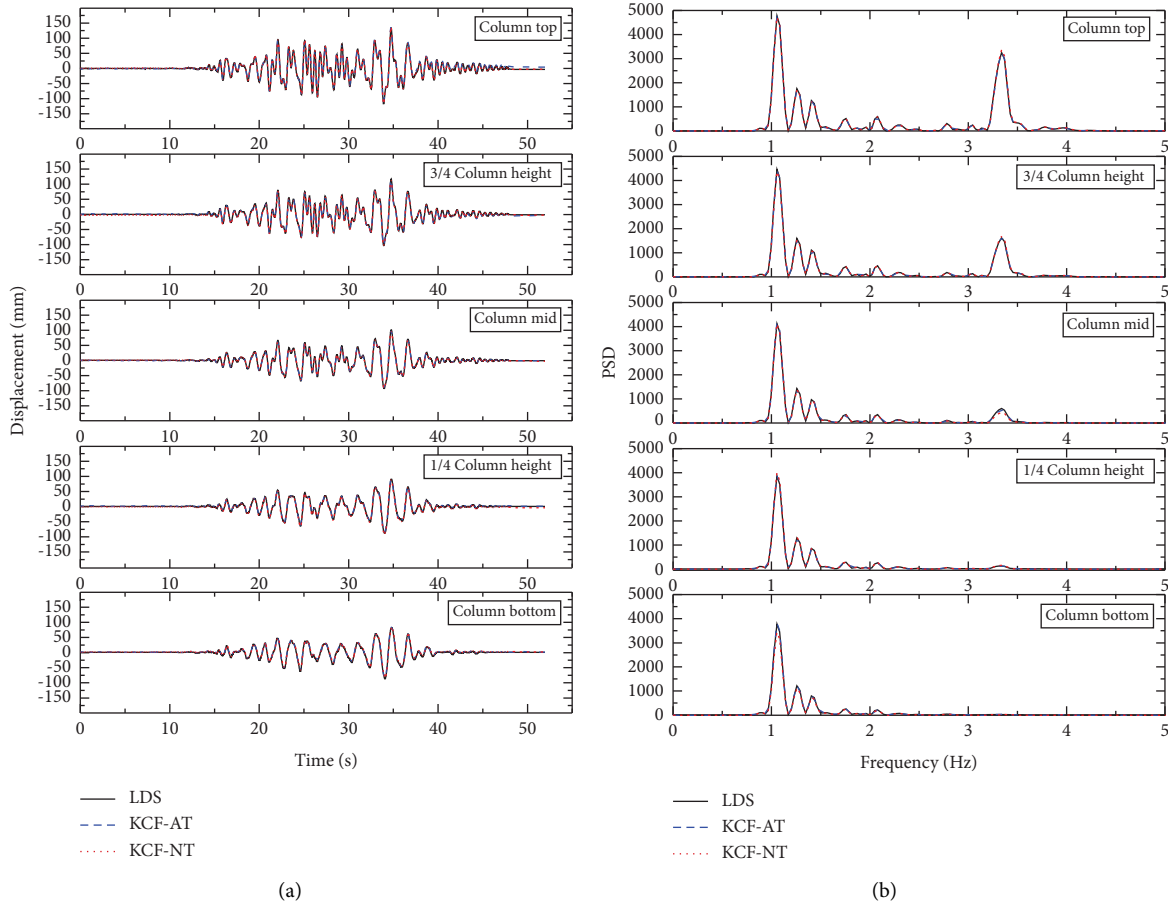
FIGURE 12: Displacement responses and the corresponding frequency contents along the column height of Case 3. (a) Displacement responses. (b) Frequency contents of displacement.

the dynamic similarity of the test model. The acceleration time histories and frequency spectra are shown in Figures 7 and 8, respectively. The seismic waves were applied to the test model in the transverse direction. All test cases are listed in Table 1.

The artificial circular targets were adhesive to the column to easily track by a high-speed camera, as shown in Figure 9(a). In addition, the natural targets of the test model, such as screws and structural corners, were selected as tracking objects because the artificial target is not easily placed for an actual structure. Signs A1–A5 and C1–C5 represent artificial and natural targets, respectively (see Figure 9(b)). In particular, signs C2, C3, and C4 present screws, and C1 and C5 represent structural corners as the natural target. The high-speed camera distance from the test model is approximately 6 m. The optical axis of the high-speed camera is perpendicular to the surface of the test model by reasonably adjusting the visual angle. The sampling frequency of the high-speed camera is 120 Hz, and its

resolution is 2448 × 2048. The traditional LDS mounted on the fixed platform was used to measure the vibration displacement of the test model, indicated by D1–D5 (see Figure 9(b)). The vibration displacement of the column measured by LDS was used to verify the feasibility, accuracy, effectiveness, and robustness of the KCF-based identification approach. It is worth noting that each traditional LDS mounted at the column height is almost identical to the targets. The sampling frequencies of the LDS and high-speed cameras are 256 Hz and 120 Hz, respectively. Moreover, both methods did not simultaneously record the vibration displacement due to the limitations of the equipment. Therefore, the peak value alignment retrieval was used to synchronize the time series of vibration displacements obtained from the high-speed camera and the LDS.

The geometric parameters of the targets were determined from the coordinates in the 2-D pixel space and the 3-D physical space of the targets, as shown in Figure 10. The red coordinates represent the target position in the pixel
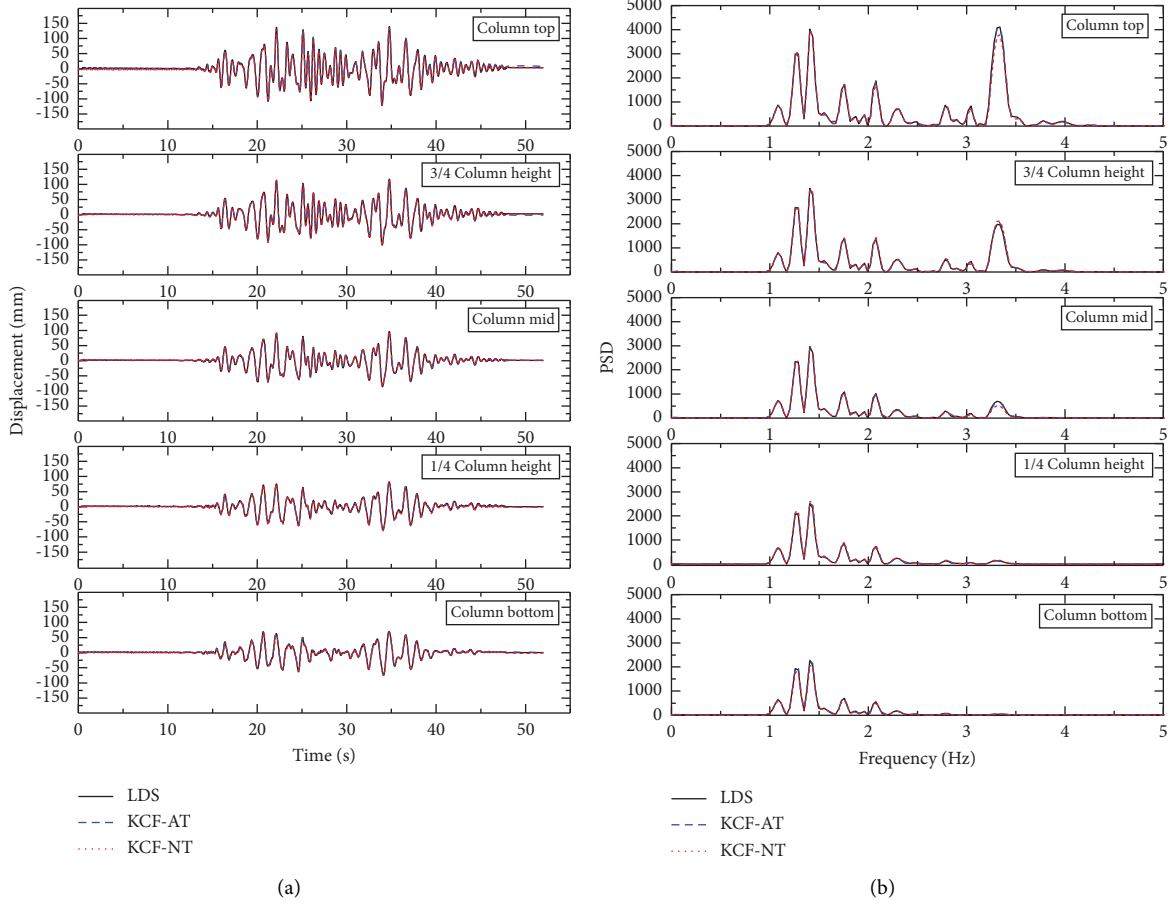
FIGURE 13: Displacement responses and the corresponding frequency contents along the column height of Case 5. (a) Displacement responses. (b) Frequency contents of displacement.

coordinate system. The blue lines and numbers represent the distance between both targets in the 3-D physical space. The scale ratios were calculated from the geometric parameters for various targets, resulting in the pixel coordinate displacements of the target transformed into the physical space displacements of the target. Similarly, the scale ratios of the natural targets were obtained from the adjacent structural dimensions or the geometric spacings between adjacent screws. In particular, the different scale ratios were calculated for various targets in different depths of the field.

3.2. Validation of the Identification Accuracy. For simplicity, only representative experimental displacements from shaking table tests were taken as comparisons to assess the feasibility, accuracy, effectiveness, and robustness of the KCF-based identification approach when the two-column pier with energy dissipation beams subjected to various seismic waves with different amplitudes and frequency contents. For comparisons, the vibration displacements identified by the KCF-based identification approach based

on artificial and natural targets were referred to as KCF-AT and KCF-NT, respectively. The vibration displacements measured with laser displacement sensors were referred to as LDS. Furthermore, the RMSE and peak displacement errors were computed to evaluate the accuracy and robustness of the KCF-based identification approach. The RMSE and peak displacement errors are calculated as follows.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{w=1}^{N} \left(d_{\text{LDS}}^{w} - d_{\text{KCF}}^{w} \right)^2}, \tag{23}$$

$$\text{Peak Error} = \left| \frac{\max \left| d_{\text{LDS}}^{w} \right| - \max \left| d_{\text{KCF}}^{w} \right|}{\max \left| d_{\text{LDS}}^{w} \right|} \right| \times 100\%, \tag{24}$$

where N represents the number of vibration displacement data, the vector, d_{LDS}, is the vibration displacement recorded by the laser displacement sensor, the vector, d_{KCF}, is the vibration displacement identified by the KCF-based identification approach, $\max|d_{\text{LDS}}^{w}|$ and $\max|d_{\text{KCF}}^{w}|$ represents the
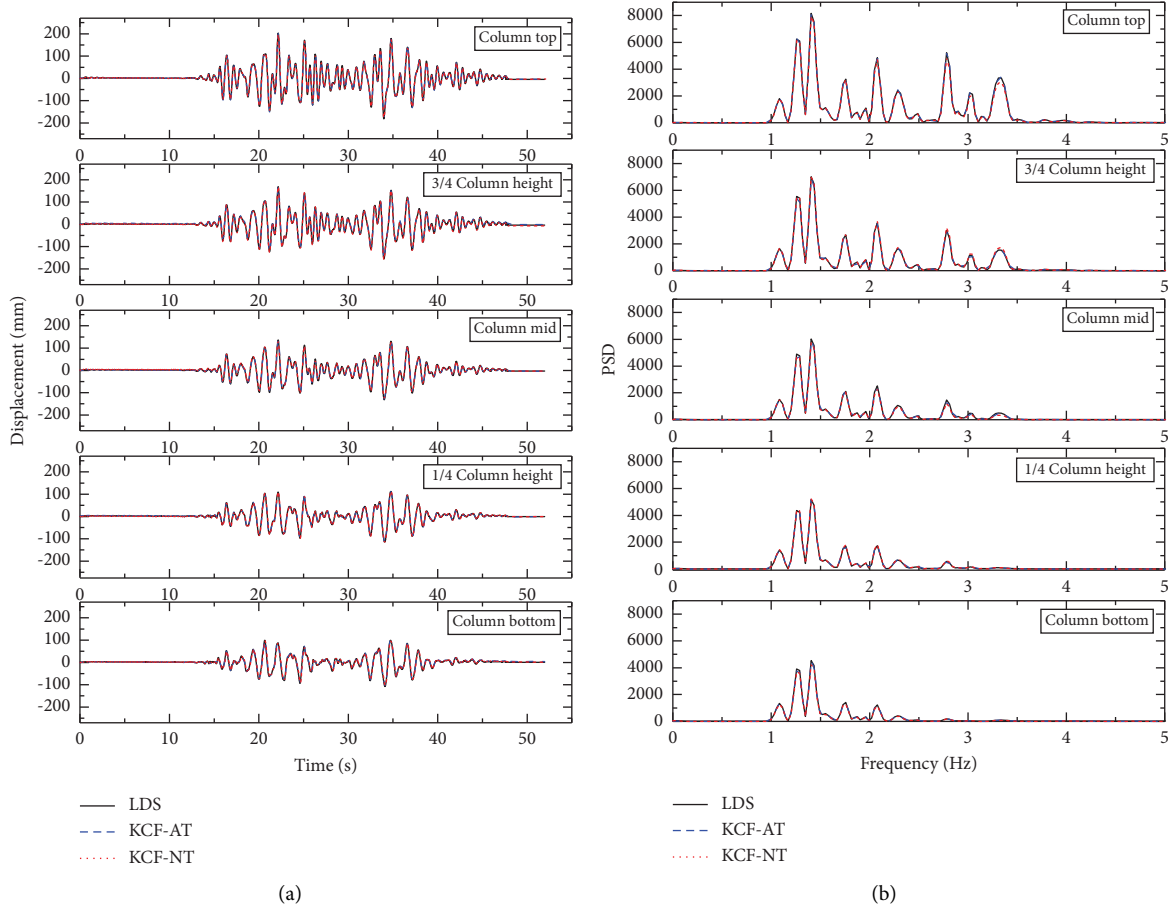
FIGURE 14: Displacement responses and the corresponding frequency contents along the column height of Case 7. (a) Displacement responses. (b) Frequency contents of displacement.

displacement with the maximum absolute values of d_{KCF} and d_{LDS}, respectively.

3.2.1. Displacement Response and the Corresponding PSD of the Column under Chi-Chi Seismic Wave. The Chi-Chi wave has a pulse effect, which significantly affects the displacement response of the test model. Therefore, the influences of shaking intensities on the accuracy of the KCF-based identification approach for the vibration displacement were studied using gradually increasing the PGAs of the Chi-Chi wave (see Table 1). However, the vibration displacements identified by the KCF-based identification approach were compared with the results recorded by the LDS for the test model under the Chi-Chi wave with typical PGAs (0.1 g, 0.3 g, 0.5 g, 0.68 g, and 0.9 g), as shown in Figures 11(a)– 15(a). The corresponding power spectrum density (PSD) of the vibration displacement is shown in Figures 11(b)–15(b), presenting the frequency spectra contents. The PSD was

calculated using the Welch average power diagram method [44].

The vibration displacements identified by the KCF-based identification approach are almost identical to the waveforms, change trends, and peak values of those recorded by the LDS for the test model under the Chi-Chi wave with various PGAs (see Figures 11(a)– 15(a)). The peak displacement errors and RMSE values between the vibration displacements measured by different methods are low (see Figure 16), indicating that the KCF-based identification approach identifying the vibration displacements of the test model has high accuracy. For example, the peak displacement errors between KCF-AT and LDS are within 4%, and the peak displacement errors between KCF-NT and LDS are less than 5%. The peak displacement errors and the RMS between the KCF-AT and KCF-NT are within 4.8%. The RMSE values between KCF-AT, KCF-NT, and LDS are within 6 mm, and the maximum value is 5.9 mm. The RMSE values between the KCF-AT and KCF-NT are within 6.6 mm
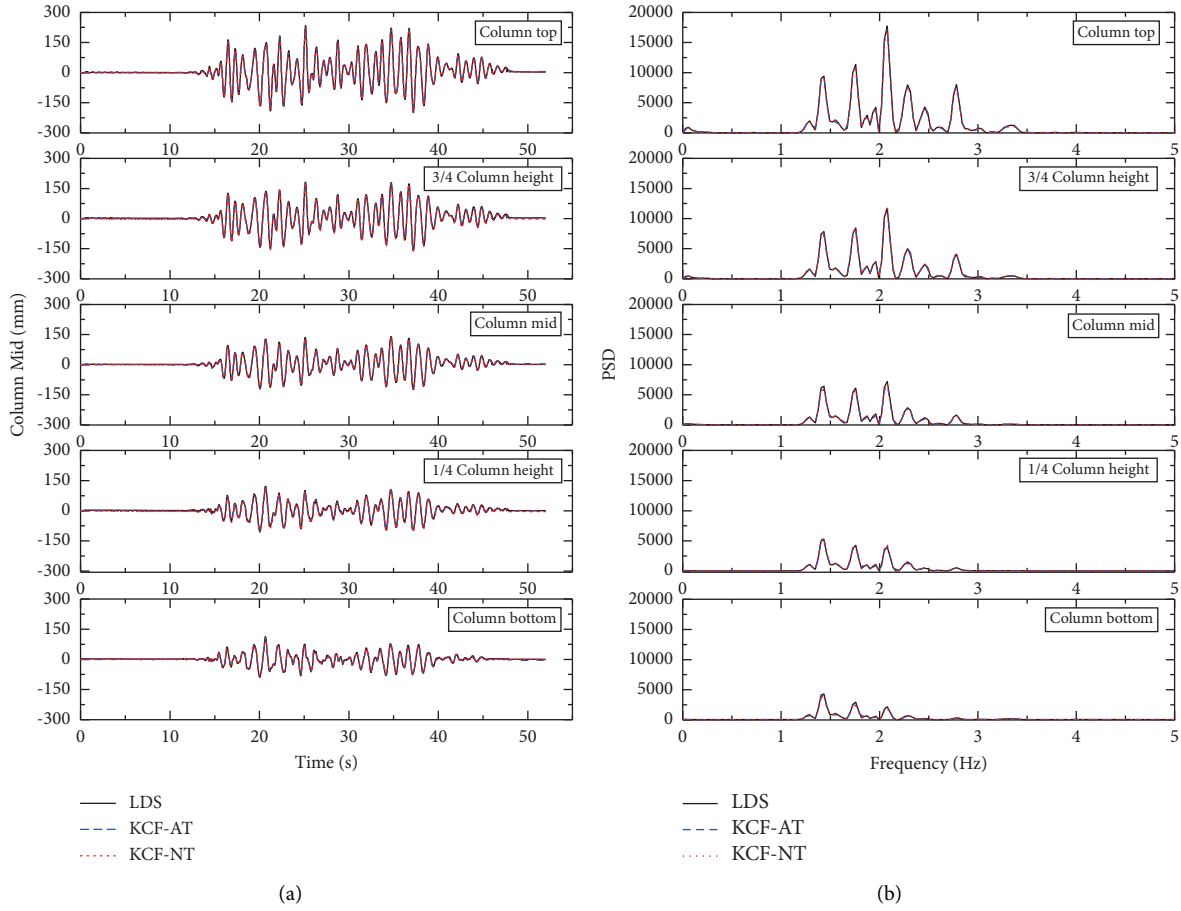
Figure 15: Displacement responses and the corresponding frequency contents along the column height of Case 9. (a) Displacement responses. (b) Frequency contents of displacement.
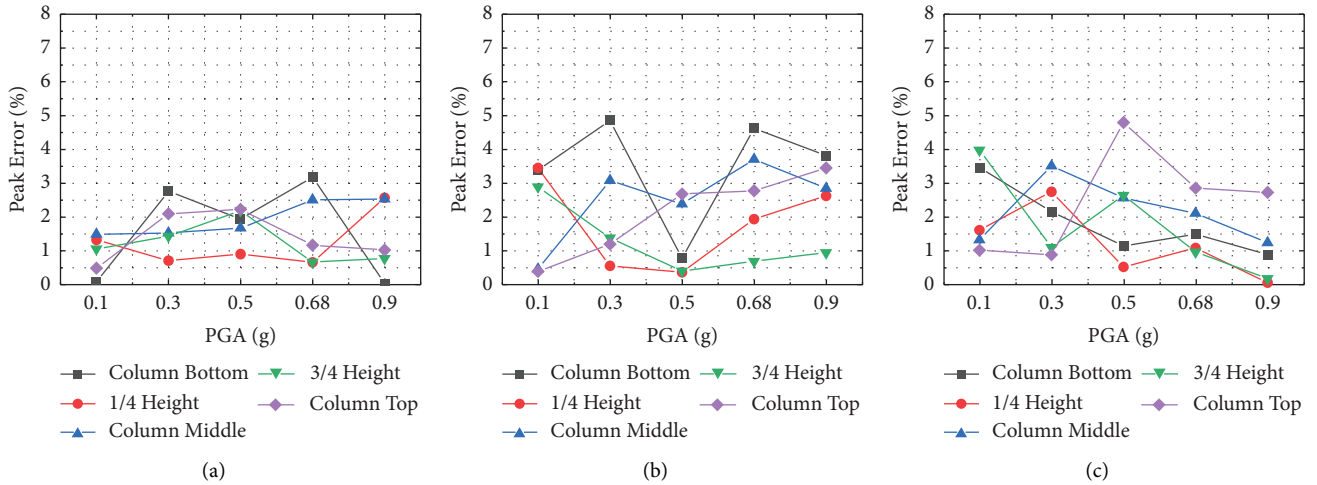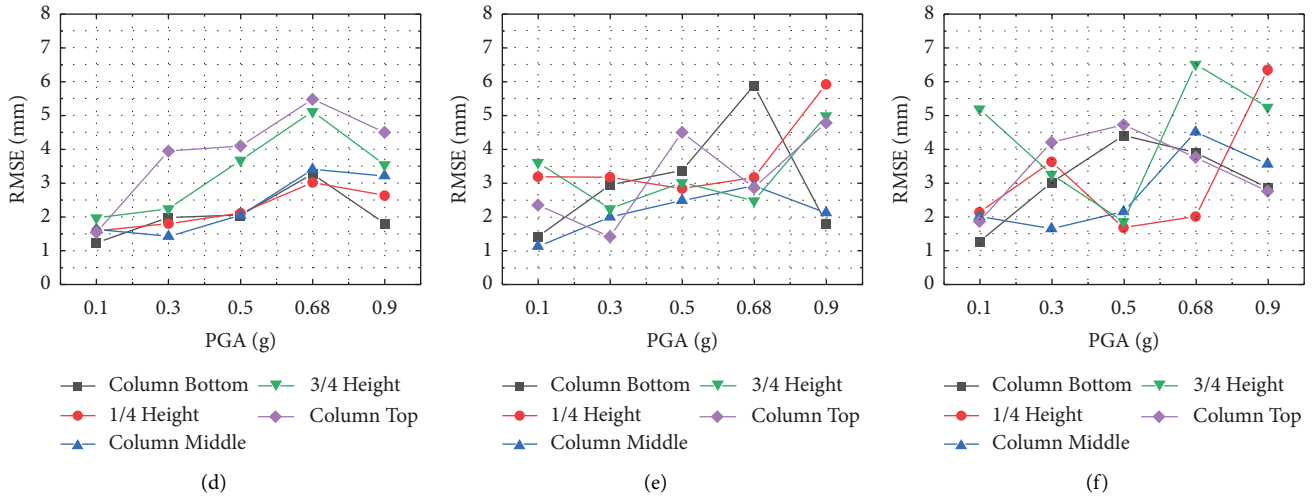

Figure 16: Continued.

FIGURE 16: Peak displacement errors and RMSE of various vibration displacement of the test model under Chi-Chi wave with various PGAs. (a) Peak displacement errors between KCF-AT vs. LDS. (b) Peak displacement errors between KCF-NT vs. LDS. (c) Peak displacement errors between KCF-NT vs. KCF-AT. (d) RMSE between KCF-AT vs. LDS. (e) RMSE between KCF-NT vs. LDS. (f) RMSE between KCF-NT vs. KCF-AT.
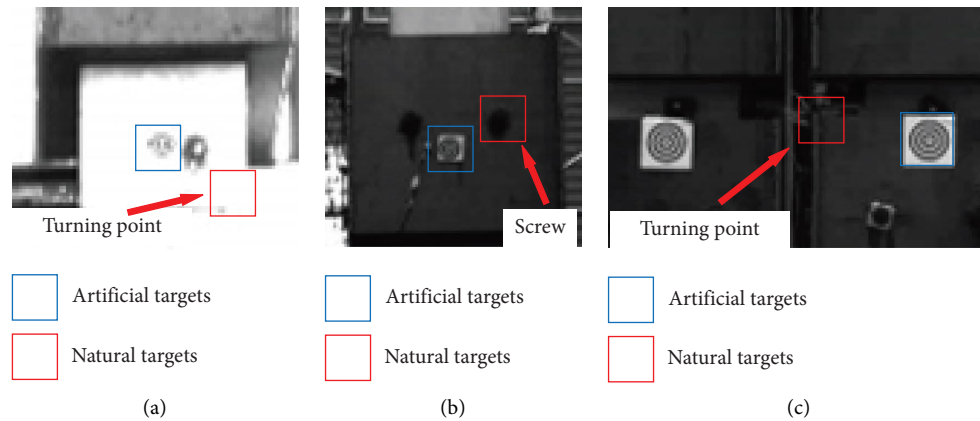

FIGURE 17: The contrast of the natural targets and the artificial targets under natural illumination: (a) column top; (b).3/4 column height; (c) column mid.

(Figure 16). However, the peak displacement errors between KCF-NT and LDS are slightly higher than those between KCF-AT and LDS. It is because the contrast of the natural targets is lower than that of the artificial targets under natural illumination. Therefore, the KCF-based identification approach according to natural targets is a higher level of difficulty. The identification accuracy is slightly lower for the KCF-NT (see Figure 17).

The frequency spectra characteristics and the wave crests of the vibration displacements identified by the KCF-AT and KCF-NT are in agreement well with those of the LDS for the test model under the Chi-Chi waves with different PGAs (see Figures 11(b)–15(b)). For example, the spectra values corresponding to the dominant frequency for the identification displacement at the column top are 4774.6, 4714.6, and 4774.5 for the KCF-AT, the KCF-NT, and LDS, respectively, when the test model subjected to the Chi-Chi wave with 0.3 g (Case 3). The corresponding errors between the KCF-AT, KCF-NT, and LDS are 0.002% and 1.26%, respectively. For the Chi-Chi wave with 0.68 g (Case 7), the spectra values at the dominant frequency for the identification displacement at the middle of the column are 5861.1, 5729.7, and 6016,4 for the KCF-AT, KCF-NT, and LDS, respectively. The corresponding differences for the KCF-AT and KCF-NT are 2.58% and 4.77%, respectively, compared to the results from the LDS. Under the Chi-Chi wave with 0.9 g (Case 9), the spectra values at the dominant frequency for the identification displacement at the column top are 17541.3, 17052.9, and 17766.7 for the KCF-AT, KCF-NT, and LDS, respectively. The corresponding errors between the KCF-AT,
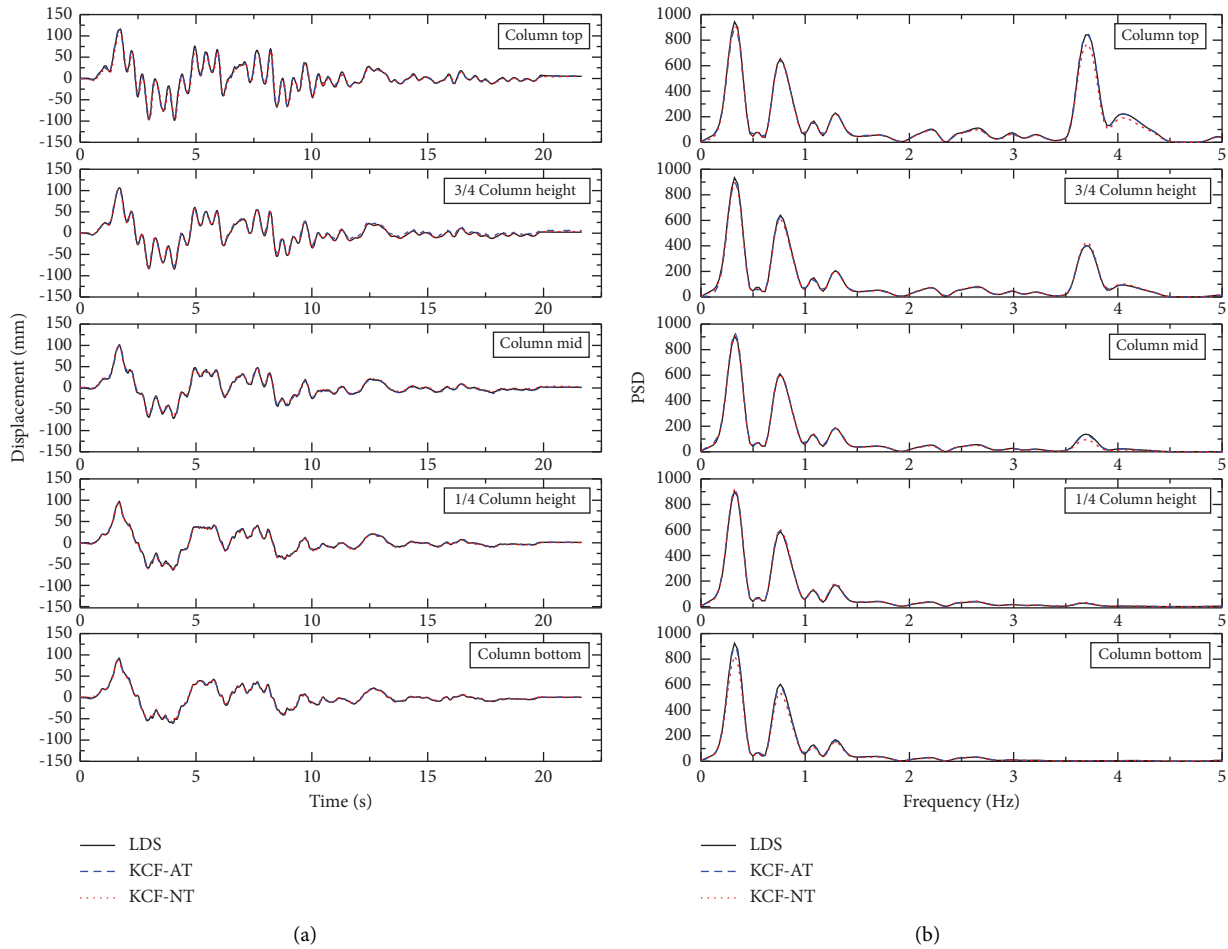
Figure 18: Displacement responses and the corresponding frequency contents along the column height of Case 10. (a) Displacement responses. (b) Frequency contents of displacement.

KCF-NT, and LDS are 1.27% and 4.02%, respectively. Consequently, the identification accuracy was verified from the frequency spectra contents of the vibration displacement recorded by the KCF-based identification approach and the LDS.

3.2.2. Displacement Response and the Corresponding PSD of the Column under Different Seismic Waves. Other seismic waves with different frequency contents are also selected as inputs to validate the robustness and effectiveness of the KCF-based identification approach. The vibration displacements identified by the KCF-AT and KCF-NT are compared with the results recorded by the LDS for the test model under four seismic waves with 0.68 g, as shown in Figures 14(a), 18(a)–20(a). The corresponding frequency spectra characteristics of the measurement displacements between different methods are illustrated in Figures 14(b), 18(b)–20(b) for the test model.

The vibration displacements identified by the KCF-AT and KCF-NT are consistent with the waveforms, variation tendencies, and peak responses of the results recorded by the LDS for the test model under the seismic waves with different frequency contents (see Figures 14(a) and 18(a)–20(a)). The peak displacement errors and RMSE values between various vibration displacements measured by the three methods (KCF-AT, KCF-NT, and LDS) are small (see Figure 21), indicating the high robustness and accuracy of the KCF-based identification approach identifying the structural vibration displacements. For example, the RMSE values of the vibration displacements between the KCF-AT, KCF-NT, and LDS are within 6 mm. The RMSE values between the KCF-AT and KCF-NT are within 7.3 mm. The peak displacement errors between the KCF-AT and LDS are lower than 4%, and the peak displacement errors between the KCF-NT and LDS are less than 6%. The peak displacement errors between the KCF-AT and KCF-NT are fewer than 7.2% (Figure 21). However, the peak
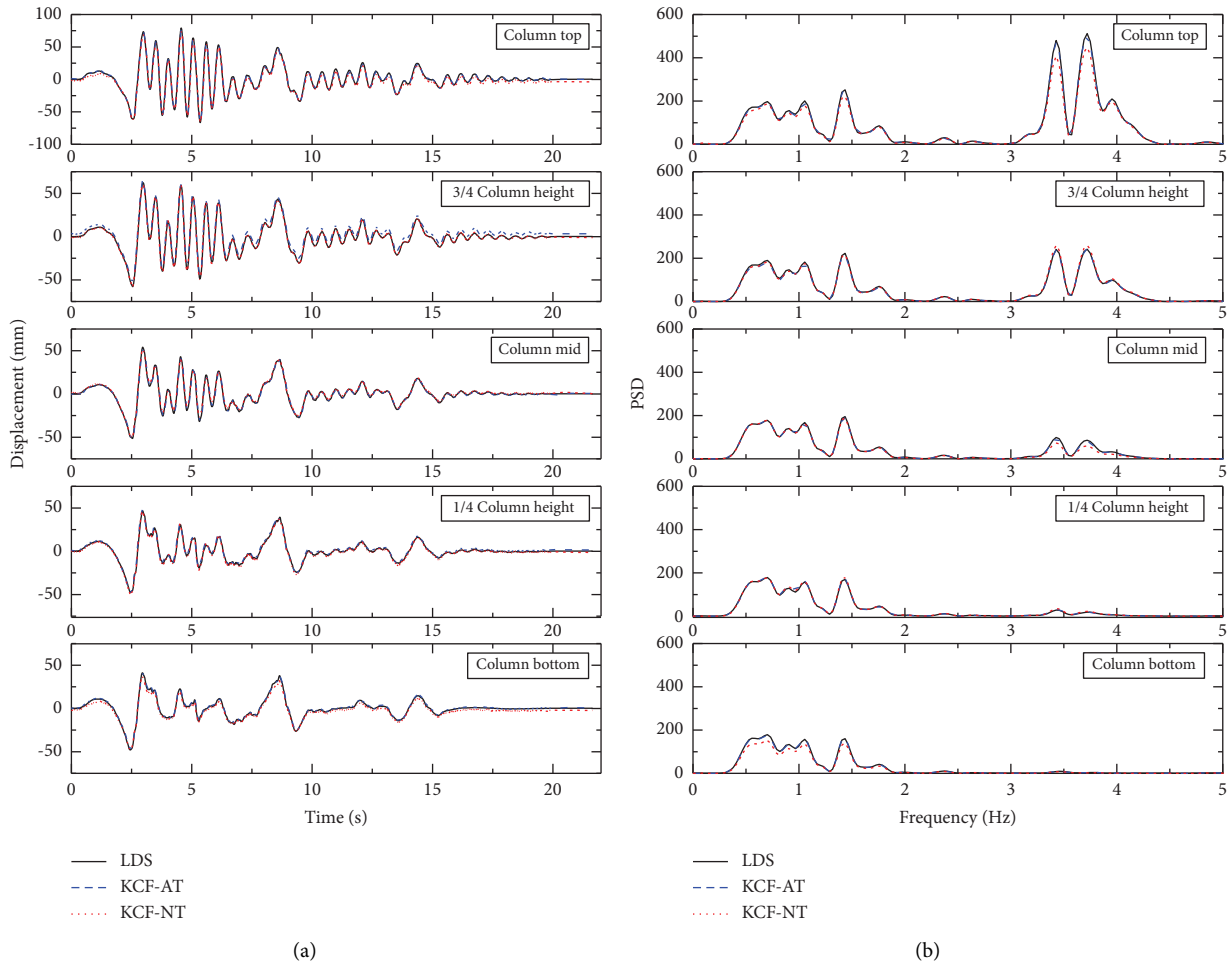
FIGURE 19: Displacement responses and the corresponding frequency contents along the column height of Case 11. (a) Displacement responses. (b) Frequency contents of displacement.

displacement errors between the KCF-NT and LDS are slightly higher than those between the KCF-AT and LDS. It is since that the natural targets showed lower contrast compared to the artificial targets under the natural illumination in the shaking table lab (see Figure 17).

The frequency spectra characteristics and corresponding wave crests of the vibration displacements measured by the KCF-AT and KCF-NT are the same as those of the vibration displacement recorded by the LDS for the test model under four seismic waves with 0.68 g (see Figures 14(b) and 18(b)–20(b)). For instance, the frequency spectra peaks at the dominant frequency of the vibration displacements at the column top from the KCF-AT, KCF-NT, and LDS are 897.6, 924.6, and 943.9, respectively, when the test model subjected to the Artificial wave with 0.68 g. The corresponding peak errors between

the KCF-AT, KCF-NT, and LDS are 4.91% and 2.04%, respectively. Under the El-Centro wave with 0.68 g, the frequency spectra peaks at the dominant frequency of the vibration displacement at the column top identified by the KCF-AT and KCF-NT are 2.30% and 14.14% less than those recorded by the LDS, respectively. For the case of the Mexico City wave with 0.68 g, the frequency spectra peaks of the dominant frequency of the vibration displacement at the column middle measured by the KCF-AT and KCF-NT are 97.30% and 94.58% compared to the results recorded by the LDS, respectively. Consequently, the KCF-based identification approach identifying the vibration displacement has high robustness and accuracy for the test model under seismic waves with different frequency contents.
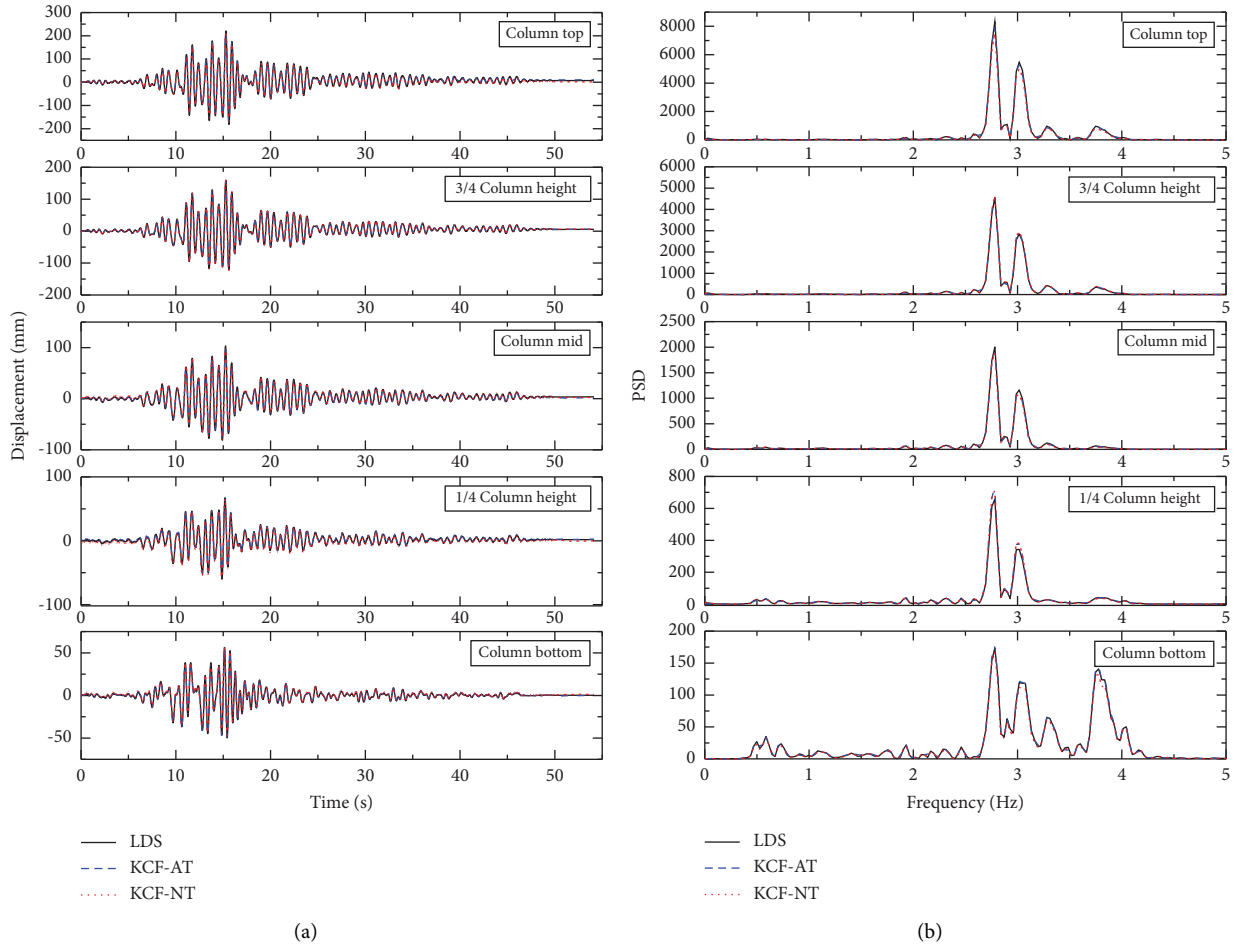
FIGURE 20: Displacement responses and the corresponding frequency contents along the column height of Case 12. (a) Displacement responses. (b) Frequency contents of displacement.
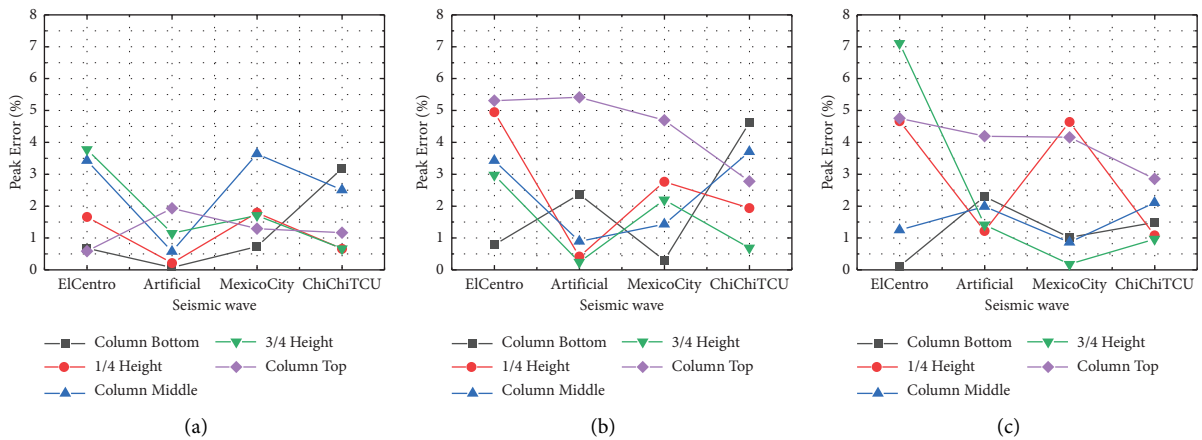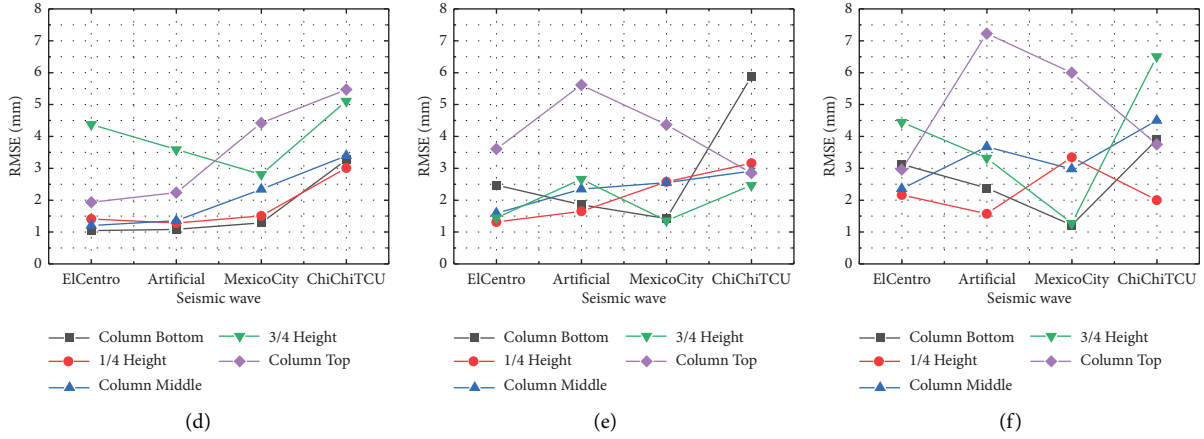

FIGURE 21: Continued.

FIGURE 21: Peak displacement errors and RMSE of various vibration displacement of the test model under seismic waves with various frequency contents. (a) Peak displacement errors between KCF-AT vs. LDS. (b) Peak displacement errors between KCF-NT vs. LDS. (c) Peak displacement errors between KCF-NT vs. KCF-AT. (d) RMSE between KCF-AT vs. LDS. (e) RMSE between KCF-NT vs. LDS. (f) RMSE between KCF-NT vs. KCF-AT.

4. Conclusions

This paper proposes a KCF-based identification approach considering various targets in different depths of the field based on monocular vision, which is employed to identify vibration displacements of bridges. A two-column pier with energy dissipation beams was designed and tested using shaking table tests under the different seismic waves. The vibration displacements of the two-column pier with energy dissipation beams were recorded by the proposed KCF-AT, KCF-NT, and conventional LDS approaches. The measurement vibration displacements between the three methods were compared to verify the feasibility, accuracy, effectiveness, and robustness of the proposed KCF-based identification approach. The crucial conclusions were summarized from the series of analyses:

(1) A conversion method associated with the scale ratio was established and adapted to various targets in different depths of the field, which is incorporated into the KCF-based tracking algorithm. The scale ratio could be used to directly achieve the physical space displacement of artificial and natural targets at different depths of the field. The vibration displacement identified by the proposed KCF-based identification approach was consistent with the results recorded by the LDS. The results show that the KCF-based identification approach has high accuracy and robustness in identifying vibration displacements.

(2) The vibration displacements and the corresponding frequency spectra contents identified by the KCF-based identification approach are almost consistent with the measurement results obtained by the laser displacement sensors for the test model under the Chi-Chi wave with different PGAs and other seismic waves with various frequency contents. The peak displacement errors and RMSE values between the

vibration displacement recorded by different methods are small. The peak displacement errors between the vibration displacement recorded by KCF-AT, KCF-NT, and LDS are less than 5% and 6%, respectively. The RMSE values between the vibration displacement recorded by the KCF-based identification approach and LDS are within 6 mm. It is indicated that the proposed KCF-based identification approach has good accuracy and robustness.

(3) The vibration displacements and the corresponding frequency spectra contents from the KCF-based identification approach according to natural targets are almost identical to the results from the KCF-based identification approach according to artificial targets. The peak displacement errors and RMSE values between the vibration displacement recorded by KCF-NT and KCF-AT are within 7.2% and 7.3 mm, respectively. It is indicated that the KCF-based identification approach based on the natural targets has the same identification accuracy and robustness. Therefore, the KCF-based identification approach based on the natural targets is more convenient in applying practical bridge engineering.

(4) The influences of complex environmental factors, such as climatic environments and low contrast natural targets, on the identification accuracy and robustness of the KCF-based identification approach in future work, especially for practical bridge engineering in more complex environments. Moreover, the KCF-based identification approach will be applied in more scenarios.

Data Availability

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

Conflicts of Interest

Acknowledgments

References

[1] 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, 2018.

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

[3] J. Lee, K. C. Lee, S. Cho, and S. H. Sim, "Computer vision-based structural displacement measurement robust to light-induced image degradation for in-service bridges," *Sensors*, vol. 17, no. 10, p. 2317, 2017.

[4] J. W. Park, J. J. Lee, H. J. Jung, and H. Myung, "Vision-based displacement measurement method for high-rise building structures using partitioning approach," *NDT & E International*, vol. 43, no. 7, pp. 642–647, 2010.

[5] P. Psimoulis, S. Stiros, and F. Moschas, "Identification of soil-structure interaction effects based on geodetic monitoring of a railway bridge," in *Proceedings of the 2nd International Conference on Bridges Innovations on Bridges and Soil-Bridge Interaction*, Athens, Greece, October 2014.

[6] 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.

[7] S. Roedelsperger, G. Laeufer, C. Gerstenecker, and M. Becker, "Monitoring of displacements with ground-based microwave interferometry: IBIS-S and IBIS-L," *Journal of Applied Geodesy*, vol. 4, no. 1, pp. 41–54, 2010.

[8] X. Zhao, H. Lan, L. Li, Y. Zhang, and C. Zhou, "A multiple-regression model considering deformation information for atmospheric phase screen compensation in ground-based SAR," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 99, pp. 1–13, 2019.

[9] B. Zhang, X. Ding, C. Werner et al., "Dynamic displacement monitoring of long-span bridges with a microwave radar interferometer," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 138, pp. 252–264, 2018.

[10] W. Zhao, G. Zhang, and J. Zhang, "Cable force estimation of a long pan cable tayed bridge with microwave interferometric radar," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 12, pp. 1491–1433, 2020.

[11] S. C. Kim, H. K. Kim, C. G. Lee, and S. B. Kim, "A vision system for identifying structural vibration in civil engineering constructions," in *Proceedings of the 2006 SICE-ICASE International Joint Conference*, Busan, Korea, October 2006.

[12] X. W. Ye and C. Z. Dong, "Review of computer vision-based structural displacement monitoring," *China Journal of Highway and Transport*, vol. 32, no. 11, pp. 21–39, 2019.

[13] C. Z. Dong, X. W. Ye, and T. Jin, "Identification of structural dynamic characteristics based on machine vision technology," *Measurement*, vol. 126, pp. 405–416, 2018.

[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] X. W. Ye, C. Z. Dong, and T. Liu, "Force monitoring of steel cables using vision-based sensing technology: methodology and experimental verification," *Smart Structures and Systems*, vol. 18, no. 3, pp. 585–599, 2016.

[16] W. Lu, Y. Cui, and J. Teng, "Structural displacement and strain monitoring based on the edge detection operator," *Advances in Structural Engineering*, vol. 20, no. 2, pp. 191–201, 2016.

[17] J. G. Chen, T. M. Adams, H. Sun, E. S. Bell, and O. Buyukozturk, "Camera-based vibration measurement of the world war I memorial bridge in portsmouth, New Hampshire," *Journal of Structural Engineering*, vol. 144, no. 11, Article ID 4018207, 2018.

[18] K. Briechle and D. H. Uwe, "Template matching using fast normalized cross correlation," in *Proceedings of the Optical Pattern Recognition XII*, Orlando, FL, USA, April 2001.

[19] G. D. Evangelidis and E. Z. Psarakis, "Parametric image alignment using enhanced correlation coefficient maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1858–1865, 2008.

[20] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Optics Letters*, vol. 33, no. 2, pp. 156–158, 2008.

[21] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

[22] M. J. Black and P. Anandan, "The robust estimation of multiple motions: parametric and piecewise-smooth flow fields," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75–104, 1996.

[23] A. Khaloo and D. Lattanzi, "Pixel-wise structural motion tracking from rectified repurposed videos," *Structural Control and Health Monitoring*, vol. 24, no. 11, 2017.

[24] T. Gautama and M. A. Van Hulle, "A phase-based approach to the estimation of the optical flow field using spatial filtering," *IEEE Transactions on Neural Networks*, vol. 13, no. 5, pp. 1127–1136, 2002.

[25] D. S. Bolme, J. Ross, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proceedings of the 2010 IEEE computer society conference on computer vision and pattern recognition*, San Francisco, CA, USA, June 2010.

[26] J. F. Henriques, C. Rui, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Proceedings of the European conference on computer vision*, Florence, Italy, October 2012.

[27] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2015.

[28] P. Olaszek, "Investigation of the dynamic characteristic of bridge structures using a computer vision method," *Measurement*, vol. 25, no. 3, pp. 227–236, 1999.

[29] L. M. Hanssen, S. Yoneyama, P. V. Farrell, A. Kitagwa, K. Kitamura, and H. Kikuta, "Deflection distribution measurement of steel structure using digital image correlation," *Optical Diagnostics*, vol. 5880, pp. 133–140, 2005.

[30] S. S. Yu, J. Zhang, and X. Y. He, "An advanced vision-based deformation measurement method and application on a long-span cable-stayed bridge," *Measurement Science and Technology*, vol. 31, no. 6, Article ID 65201, 2020.

[31] H. Yoon, H. Elanwar, H. J. Choi, M. Golparvar-Fard, and B. F. Spencer, "Target-free approach for vision-based structural system identification using consumer-grade cameras," *Structural Control and Health Monitoring*, vol. 23, no. 12, pp. 1405–1416, 2016.

[32] J. G. Chen, N. Wadhwa, Y. J. Cha, F. Durand, W. T. Freeman, and O. Buyukozturk, "Modal identification of simple structures with high-speed video using motion magnification," *Journal of Sound and Vibration*, vol. 345, pp. 58–71, 2015.

[33] J. Zhao, Y. Q. Bao, Z. G. Guan, W. M. Zuo, J. Z. Li, and H. Li, "Video-based multiscale identification approach for tower vibration of a cable-stayed bridge model under earthquake ground motions," *Structural Control and Health Monitoring*, vol. 26, no. 3, p. 2314, 2019.

[34] B. Du, Y. J. Sun, S. H. Cai, C. Wu, and Q. Du, "Object tracking in satellite videos by fusing the kernel correlation filter and the three-frame-difference algorithm," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 2, pp. 168–172, 2018.

[35] J. Shao, B. Du, C. Wu, and L. F. Zhang, "Tracking objects from satellite videos: a velocity feature based correlation filter," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 10, pp. 7860–7871, 2019.

[36] H. Yang, J. Yu, S. Wang, and X. Peng, "Design of airborne target tracking accelerator based on KCF," *Journal of Engineering*, vol. 2019, no. 23, pp. 8966–8971, 2019.

[37] X. Q. Chen, X. Q. Xu, Y. S. Yang, H. F. Wu, J. J. Tang, and J. S. Zhao, "Augmented ship tracking under occlusion conditions from maritime surveillance videos," *IEEE Access*, vol. 8, pp. 42884–42897, 2020.

[38] B. Zheng and A. Gupta, "Soccer player video target tracking based on deep learning," *Mobile Information Systems*, vol. 2022, Article ID 8090871, 6 pages, 2022.

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

[40] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, June 2005.

[41] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.

[42] B. Schölkopf and J. S. Alexander, *Learning with Kernels*, The MIT Press, Cambridge, Massachusetts, 2018.

[43] R. Ryan, G. Yeo, and T. Poggio, "Regularized least-squares classification," *Nato Science Series Sub Series III Computer and Systems Sciences*, vol. 190, pp. 131–154, 2003.

[44] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.