Hindawi
Structural Control and Health Monitoring
Volume 2023, Article ID 7904198, 25 pages
https://doi.org/10.1155/2023/7904198

Research Article

Video Motion Magnification and Subpixel Edge Detection-Based Full-Field Dynamic Displacement Measurement

Da-You Duan,[1] K. S. C. Kuang,[2] Zuo-Cai Wang,[1,3,4] and Xiao-Tong Sun[1]

[1]*School of Civil Engineering, Hefei University of Technology, Hefei 230009, Anhui, China*
[2]*Department of Civil and Environmental Engineering, National University of Singapore, Singapore*
[3]*Anhui Engineering Laboratory for Infrastructural Safety Inspection and Monitoring, Hefei 230009, Anhui, China*
[4]*Engineering Research Center of Safety-Critical Industrial Measurement and Control Technology of the Ministry of Education, Hefei 230009, Anhui, China*

Correspondence should be addressed to Zuo-Cai Wang; wangzuocai@hfut.edu.cn

Received 8 March 2023; Revised 15 July 2023; Accepted 31 August 2023; Published 8 September 2023

Academic Editor: Zoran Rakicevic

Noncontact measurement techniques in structural dynamics field have progressed significantly in the past few decades. Vision-based measurement techniques are unique in that they have the ability to achieve full-field measurement and possess the typical advantages associated with noncontact measurement techniques. Recently, vision-based techniques have also been applied to streaming of videos for structural dynamic displacement measurement. The most recent trends in vision-based measurements include target tracing, digital image correlation, and target-less approaches. There are, however, some shortcomings of the vision-based techniques such as susceptibilities to image noise, prevailing light conditions, and limit in measurement resolution. To reduce these shortcomings, a method known as video motion magnification (MM) can be used to amplify small structural motions. Using the phase-based motion magnification (PBMM) and subpixel edge detection methods, the full-field dynamic displacements of the structure can be obtained. The deep convolutional long short-term memory (ConvLSTM) network is applied to aid in the selection of the frequency band for magnification in the PBMM algorithm. To achieve higher measurement accuracy, the displacement results with and without MM are combined with the finite impulse response (FIR) filter which can reduce the error caused by the PBMM procedure. In the tests, plastic optical fiber (POF) displacement sensors are introduced and used as reference measurements to compare the dynamic displacement results from the proposed vision-based method. Compared with the measured displacements with POF sensors, the proposed method offers high level of accuracy for full-field displacement measurement.

1. Introduction

Dynamic displacement is one of the most important indices for structural safety assessment. Although accelerometers, strain gages, and linear variable displacement transducers can be used to obtain the structural dynamic displacement, there are some limitations that make these contact methods unattractive and impractical in many applications [1]. As an alternative to contact measurement approaches, noncontact measurement methods are more convenient to deploy and have no effect on the structural modal response. Laser Doppler vibrometers, microwave interferometer, Global Navigation Satellite System (GNSS), and other noncontact measurement methods have drawn much attention recently in the field of structural dynamic displacement monitoring [2]. Continuously scanning laser Doppler vibrometer system has been used to measure the structural dynamic displacement [3]. Microwave interferometer has also been used for remote sensing of structural displacements in buildings [4]. The accuracy of GNSS systems, however, depends on the satellite signal strength which can be influenced by the satellite geometry, environmental factors, and atmospheric condition [5]. There are still some challenges encountered when using the above methods in addition to the high cost of the devices involved.

With advances in technological innovation, computer vision sensing and monitoring technology is gaining

significant attention from many workers in the field [6]. The sampling rate and spatial resolution achievable for these methods have also been demonstrated to meet the technical requirements for dynamic displacement measurement projects [7, 8]. Some of the most frequently used approaches in vision-based structural displacement data extraction techniques are target tracing [9, 10], digital image correlation (DIC) [11], and target-less approaches. Xu et al. [12] have achieved multipoint displacement monitoring by tracking deck and cable targets in a cable-stayed footbridge. Jeong and Jo [13] applied a Siamese network-based visual tracker as a real-time tracking tool for dynamic monitoring. Jiao et al. [14] proposed a tracking algorithm that combines random sample consensus algorithm and efficient second-order minimization technique, which refines the final estimates to subpixel accuracy and avoids tracking drift and nonsmoothness effectively. Except for target-tracing methods, the DIC and target-less approaches have the ability to observe the full-field structural displacement [15]. Edge detection, optical flow [16], and pattern matching algorithms are popular in target-less measurement. Bhowmick et al. [17, 18] have made achievements in full-field displacement measurement using video image sequences. They applied Gaussian filter on the spatiotemporal video signal to convert the edge information into zero crossing signal and used the optical flow to compute the Lagrangian displacement of each pixel of the edge for every frame in the video with subpixel accuracy. Dong et al. [19] proposed a structural displacement monitoring method using deep learning-based full-field optical flow.

Due to the limitation in video resolution, obtaining the small vibratory motion of the structure in the video is one of the challenges in vision-based dynamic displacement measurement. To achieve high accuracy, cameras with high resolution are required in vision-based measurement methods, and where necessary, zoom lenses are used to focus in on a region of interest (ROI) which may reduce the advantage of a full-field measurement. Although many types of subpixel methods have been proposed [20] for improving the edge accuracy, a subpixel edge detection method can only use the information in a single image frame of the video captured, while the time series of the video frames contains many more information within the image sequence. The developments in computer vision have generated a method known as motion magnification, a technique which can amplify the small motions of the vibrating structure as captured by the video recording. The first attempt to magnify barely visible structural motions in a video was published in 2005 by Liu et al. [21]. In 2012, Wu et al. [22] first proposed the Eulerian video magnification (EVM) method which can also linearly amplify noise while amplifying motion. Wadhwa et al. [23] developed the phase-based motion magnification (PBMM) technique. Using this method, microvibration of an object that was originally barely observable can be significantly amplified within the frequency band of interest. Nowadays the PBMM has been applied to structural health monitoring purposes. Civera et al. [24] used the PBMM for instantaneous damage localization and tracking the damage growth over time. Besides, the PBMM can return displacement time histories as accurate as physically attached sensors under appropriate conditions [25].

The PBMM processing can generate unwanted artifacts if the parameters chosen are inappropriate. The blurring effect and other artifacts in the video can significantly reduce the accuracy of the dynamic displacement measurement. The key parameters in question are related to the structural vibration frequency and amplitude. To produce clear motion magnification results, nonnegative matrix factorization [26] and singular value decomposition [27] have been used to identify independent modal components. In view of its potential, machine learning methods have recently drawn much attention in field of vibration monitoring and dynamic analysis [1, 28]. Deep neural networks such as CNN [29] and LSTM [30] have shown data modeling capabilities in many engineering applications. As for video-based structural vibration monitoring, CNN-LSTM-based computer vision architectures have been used for modal frequency extraction [31]. It has been shown that the vibration frequency spectrum peaks can be first identified using convolutional long short-term memory (ConvLSTM) networks [32, 33]. The parameters of the PBMM are then selected using the frequency information to achieve a better motion magnification effect.

In this research, the PBMM algorithm was used to efficiently magnify the small motions of the vibrating structure captured in the video stream with the help of a ConvLSTM neural network. Following that, the subpixel edge detection algorithm based on the partial area effect [34] was implemented to locate the boundaries of the structure. As the PBMM algorithm does not work well with static and abrupt displacements in the video, the displacements at low frequencies can be measured from the original video with the subpixel edge detection method. Furthermore, due to the noisy and complex background, the edge detection results are not found to be satisfactory to obtain the structural deformation, and hence polynomial fitting is adapted to reconstruct the displacement curve of the beam or shape of the cable. From the experiments conducted, the proposed method was found to be capable of achieving high accuracy in capturing the vibration response. The vibration response of the structure was also captured using an array of plastic optical fiber (POF) displacement sensors as a reference measurement to compare with the video-based results.

With the help of the ConvLSTM network and subpixel edge detection, the proposed method accomplishes to obtain full-field displacement response of vibrating structural components. The main contributions of this paper are as follows. (1) Novel ConvLSTM-based computer vision architecture for modal frequency extraction is proposed. The outlined architecture is entirely autonomous when it comes to processing vibration videos and extracting modal frequencies. Additionally, the developed approach extrapolates and performs well on unfamiliar data. (2) With the frequency estimated by the proposed ConvLSTM network, the PBMM results have been improved and the side effect like blur has been alleviated. (3) A methodology to obtain full-field displacement response of vibrating structural

components has been proposed using PBMM, convolutional LSTM network, and subpixel edge detection. The displacement monitoring accuracy has been improved compared to extracting the displacement directly from the original video.

2. Methodology

2.1. Phase-Based Video Magnification. Several approached such as the Lagrangian, linear-Eulerian, and phase-based techniques have been proposed for the motion magnification technology. In Lagrangian approaches, the errors are found to be large in the magnified video. The basic methodology of the linear EVM involves the use of the time series of pixel brightness (grayscale) and amplifying any variation in a specified temporal frequency band of interest. For the PBMM method, improvements over the former methods are achieved in two important aspects: (a) larger magnification scales and (b) better performance under noisy conditions [23]. The processes of PBMM and EVM technology have many similarities, but their corresponding spatial filters are very different. In the spatial-domain decomposition part of the PBMM algorithm, the Gauss or Laplace pyramids of the Eulerian linear method are replaced with complex steerable pyramid. The algorithm takes the phase difference in the wavelet to amplify the video motion. These pyramids result in an efficient and accurate linear decomposition for video frames in the scale and orientation subbands.

Based on the phase-based optical flow method [35], PBMM algorithm indirectly represents the micromotion in the video using the phase information from the Euler perspective and magnifies the motion in the video using the phase information. The basis is to convert the position information of the pixels in the video image in the spatial domain into the phase information in the frequency domain by using the time-shift property of Fourier transform [36]. The complex steerable pyramid [37] is used to establish the relationship between the local motion information in the video image with the local phase information, in order to achieve the amplification of the local motion of the video through the operation of the local phase.

By using transfer function $\Psi_{\omega,\theta}$, the discrete Fourier transform (DFT) $\widetilde{I}$ of an image I is decomposed into different spatial frequency bands $S_{\omega,\theta}$. Each spatial frequency band has $\mathrm{DFT}\widetilde{S}_{\omega,\theta}(x,y) = \widetilde{I}\Psi_{\omega,\theta}$, in which the spatial scale is$\omega$and orientation is θ. The complex steerable pyramid decomposes an image into different spatial frequency bands, each of which is localized in space, scale, and orientation. The transfer functions of a complex steerable pyramid contain only the positive frequencies of the corresponding real steerable pyramid's filter, allowing for a representation of both amplitude and phase.

In the frequency domain, the process of building and collapsing the steerable pyramid is given by equation (1). The image $\widetilde{I}_R$ is reconstructed by the sums in the equation over all the scales and orientations in the pyramid.

$$\widetilde{I}_R = \sum \widetilde{S}_{\omega,\theta}\Psi_{\omega,\theta} = \sum \widetilde{I}\Psi^2_{\omega,\theta}. \tag{1}$$

The phase-based motion magnification approach uses complex-valued steerable pyramids to measure and modify local motions. As an example, consider a 1D image intensity profile undergoing global translation over time, denoted as $f(x+\delta(t))$, where $\delta(t)$ is the displacement function. The goal is to synthesize a sequence with modified motion $f(x+(1+\alpha)\delta(t))$, for some magnification factor α. Using a Fourier series decomposition, the displaced image $f(x+\delta(t))$ can be expressed as

$$f(x+\delta(t)) = \sum_{\omega=-\infty}^{\infty} \mathrm{Amp}_\omega e^{i\omega(x+\delta(t))}, \tag{2}$$

where Amp is the spatial amplitude, and each frequency band in this sum corresponds to a single frequency ω.

Based on equation (2), the band for frequency ω is the complex sinusoid.

$$S_\omega(x,t) = \mathrm{Amp}_\omega e^{i\omega(x+\delta(t))}. \tag{3}$$

Because S_ω is a sinusoid, its phase $\omega(x+\delta(t))$ contains motion information. Like the Fourier shift theorem, the motion can be manipulated by modifying the phase.

To isolate motion in specific temporal frequencies, the phase $\omega(x+\delta(t))$ is temporally filtered with a direct current balanced filter. The result is

$$B_\omega(x,t) = \omega\delta(t). \tag{4}$$

The band-passed phase $B_\omega(x,t)$ is then multiplied by α and the phase of subband $S_\omega(x,t)$ is increased by this amount to get the motion magnified subband.

$$\widehat{S}_\omega(x,t) \coloneqq S_\omega(x,t)e^{i\alpha B_\omega} = \mathrm{Amp}_\omega e^{i\omega(x+(1+\alpha)\delta(t))}. \tag{5}$$

The result $\widehat{S}_\omega(x,t)$ is a complex sinusoid that has motions exactly $1+\alpha$ times the input. The motion magnified video can be reconstructed by collapsing the pyramid. In this analysis, the motion magnified sequence $f(x+(1+\alpha)\delta(t))$ is obtained by summing all the subbands.

The whole process contains four parts. (1) Each frame of the input video is decomposed by a complex steerable pyramid to obtain the local amplitude spectrum and the local phase spectrum of the video image. (2) The phase difference signal in the frequency band of interest is extracted by time-domain band-pass filtering such as linear phase FIR (fine impulse response) band-pass filter and IIR (infinite impulse response) band-pass filter. (3) The selected phase difference signal of interest is multiplied by the set amplification factor to obtain the linear amplification result of the small phase difference signal. (4) The amplified data are reconstructed by complex steerable pyramid, and the amplified output video is reconstructed by combining the high-pass residual and low-pass residual of the input video.

2.2. Convolutional LSTM Network. ConvLSTM is a type of recurrent neural network that is designed for spatiotemporal prediction. It has a convolutional structure in both the input-to-state and state-to-state transitions, which means that it uses convolutional filters to process the input data and

update the internal state of the network. The ConvLSTM model is composed of a grid of cells, where each cell processes a small region of the input data and maintains its own internal state. The internal state of a cell at time t is determined by the inputs and past states of its local neighbors, as well as its own previous state. This allows the ConvLSTM to capture the dependencies between different spatial locations over time, making it well suited for tasks such as video prediction or anomaly detection.

One of the key differences between the ConvLSTM and other types of LSTM networks is that the ConvLSTM uses convolutions directly as part of reading the input into the LSTM units themselves. This allows the ConvLSTM to eliminate matrix multiplication in the LSTM, which can make the model more efficient and easier to train [38]. This is in contrast to the traditional LSTM, which reads the data indirectly in order to calculate the internal state and state transitions, and the CNN-LSTM, which interprets the output from CNN models [39]. The key equations of ConvLSTM are shown below, and the symbols $*$ and $\odot$ denote the convolution operator and the Hadamard product, respectively:

$$\begin{aligned} i_t &= \sigma\left(W_{\mathrm{xi}} * X_t + W_{\mathrm{hi}} * H_{t-1} + W_{\mathrm{ci}} \odot \mathscr{C}_{t-1} + b_i\right), \\ f_t &= \sigma\left(W_{\mathrm{xf}} * X_t + W_{\mathrm{hf}} * H_{t-1} + W_{\mathrm{cf}} \odot \mathscr{C}_{t-1} + b_f\right), \\ \mathscr{C}_t &= f_t \odot \mathscr{C}_{t-1} + i_t \odot \tanh\left(W_{\mathrm{xc}} * X_t + W_{\mathrm{hc}} * \mathscr{H}_{t-1} + b_c\right), \\ o_t &= \sigma\left(W_{\mathrm{xo}} * X_t + W_{\mathrm{ho}} * \mathscr{H}_{t-1} + W_{\mathrm{co}} \odot \mathscr{C}_t + b_o\right), \\ \mathscr{H}_t &= o_t \odot \tanh\left(C_t\right). \end{aligned} \tag{6}$$

As mentioned in [40, 41], the structure of a ConvLSTM cell is drawn in Figure 1. The ConvLSTM cell consists of input gate (i_t), forget gate (f_t), and output gate (o_t) which are used to update the hidden state based on the input at time t (X_t). The main structure of the ConvLSTM cell is similar to that of a traditional LSTM cell, but it also includes additional components that allow it to incorporate convolutional operations. These additional components are used to process the input data using convolutional filters and update the cell output and hidden state using convolutions. The cell output ($\mathscr{C}_t$) and hidden state (H_t) are used to store and propagate information over time, allowing the ConvLSTM cell to capture dependencies between inputs at different times.

2.3. Subpixel Edge Detection. In the field of vibration measurement, ordinary pixel level edge detection methods such as Prewitt, Sobel, Laplacian, and Canny edge detection are not sufficiently accurate for the present application. A subpixel level edge detection method is required instead to obtain accurate displacement data of the vibrating structure. To estimate the subpixel edge position, the subpixel edge detection based on partial area effect is used in this paper.

The subpixel edge detection is based on the hypothesis that pixel values of the image are proportional to the intensities and areas at both sides of the edge. For the edge of a straight line, the edge equation can be represented as $y = a + \mathrm{bx}$. Considering an ideal image with a straight edge

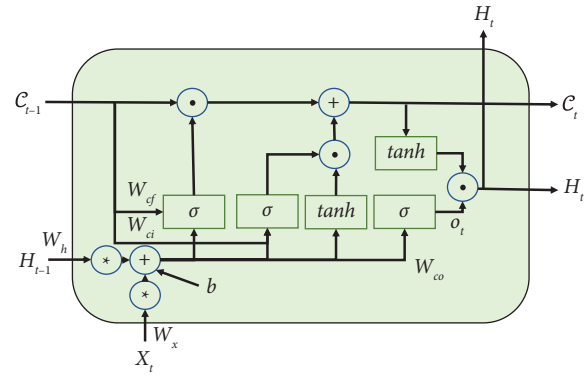


Figure 1: ConvLSTM cell.

like Figure 2(a), the edge divides the image plane into two regions of different gray intensities, A and B. Every pixel in the image is a $h \times h$ square. Assume that S_A and S_B are the areas in one pixel covered by A and B, respectively ($h^2 = S_A + S_B$). When an edge crosses over that pixel, the intensity of that pixel is

$$\text{Intensity} = \frac{\mathrm{AS}_A + \mathrm{BS}_B}{h^2}. \tag{7}$$

To determine the subpixel position of the edge, the vertical distance from the center of a pixel to the edge can be calculated and represented by the parameter "a." A 5×3 window is centered on the target pixel as shown in Figure 2(b). The areas of left, middle, and right column below the edge line can be expressed as $\mathrm{ah} - \mathrm{bh}^2 + (5/2)h^2$, $\mathrm{ah} + (5/2)h^2$, and $\mathrm{ah} + \mathrm{bh}^2 + (5/2)h^2$, respectively. Then, the intensity of the left, middle, and right column pixels has the following expressions:

$$\begin{aligned} L &= 5B + \frac{A-B}{h^2}\left(\mathrm{ah} - \mathrm{bh}^2 + \frac{5}{2}h^2\right), \\ M &= 5B + \frac{A-B}{h^2}\left(\mathrm{ah} + \frac{5}{2}h^2\right), \\ R &= 5B + \frac{A-B}{h^2}\left(\mathrm{ah} + \mathrm{bh}^2 + \frac{5}{2}h^2\right). \end{aligned} \tag{8}$$

With the expressions, the coefficients a and b of the edge line can be derived as

$$\begin{aligned} a &= \frac{2M - 5(A+B)}{2(A-B)}h, \\ b &= \frac{R - \mathrm{L}}{2(A-B)}. \end{aligned} \tag{9}$$

The normal vector to the edge can be calculated using the expression

$$N = \frac{A-B}{\sqrt{1+b^2}}[b, -1]. \tag{10}$$

The magnitude of the normal vector to the edge in an image represents the change in intensity between the two regions on either side of the edge. The edge location and the normal vector can therefore be calculated by this method.

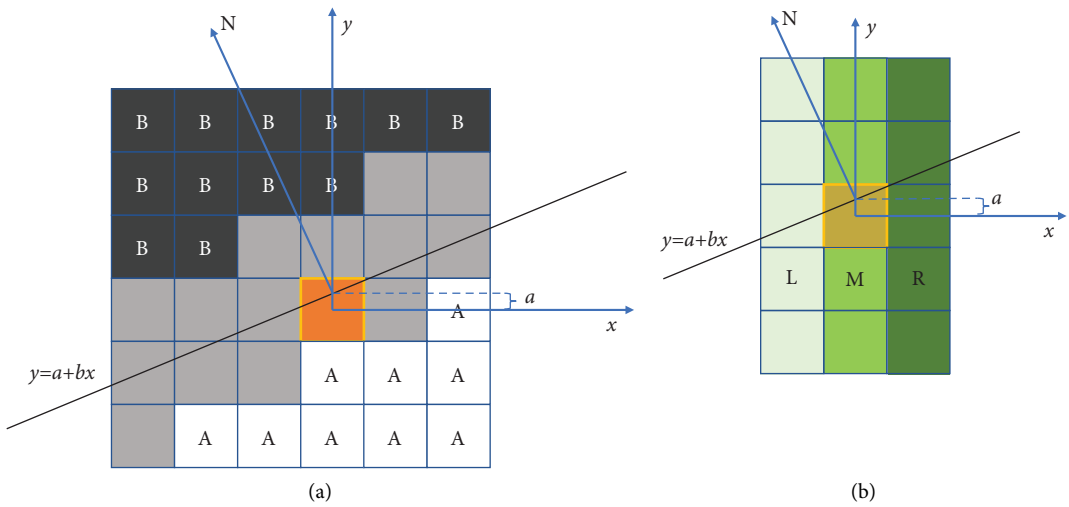


Figure 2: Edge features obtained for the highlighted pixel. (a) Edge line between areas A and B. (b) A 5×3 window centered on the target pixel.

2.4. Workflow of the Proposed Method. Though the PBMM algorithm can magnify the small motions in the video, the process may cause distortion at locations with relatively large displacements with low-frequency vibrations. For example, it is typical that the measured displacement results are normalized with the magnification factor. Following the normalization process, the large motions will end up smaller than the ground truth. Concurrently, the subpixel edge detection can measure the relatively low-frequency large motions directly from the original video. In this paper, the proposed method aims to combine the measurement results from the original and the magnified video to maintain the accuracy of the vibration measurement.

The workflow of the proposed method is illustrated in Figure 3. Firstly, the spectrum peaks are estimated by the deep ConvLSTM neural network. The cutoff frequency is identified as the first peak of the estimation. Adopting the selected frequency band from the first spectrum peak to the Nyquist frequency, the video motion can be magnified by the PBMM algorithm. The displacements of the object boundaries of both the original video and the processed video are extracted by the subpixel edge detection. With the use of the FIR filter, data from the low-frequency part of the original video and the high-frequency part of the magnified video results are combined. In every frame, the bending curve is smoothened through polynomial fitting, and the final full-field dynamic displacement can then be obtained.

3. ConvLSTM Network Training and Video Magnification Parameter Optimization Based on Steel Strand Experiment

In the proposed processes of the full-field dynamic displacement measurement, the video motion magnification needs the frequency parameter to work and the ConvLSTM network needs to be pretrained before the measurement. Thus, a cable test was used to provide the datasets for ConvLSTM network training and the videos of the vibrating cable with different tension and modal frequencies were recorded. Once the network is pretrained, it can be fine-tuned on different frequency estimation cases. This approach enables the network to learn the spatio-temporal patterns and relationships between the frequency and the input data, leading to accurate frequency predictions and further helping to measure the dynamic full-field displacement.

3.1. Data Collection and Training of the ConvLSTM Network. A 7-wire steel strand which is 5 mm in diameter is tested in this study. The experimental set is shown in Figure 4. The anchorages at the end plates provide fixed constraints at both ends. The steel strand used has the following property and dimension: $\overline{m} = 0.123$ kg/m and the free length $L = 3$ m. In order to vibrate the strand, a random hammering excitation is provided manually. The accelerometer is attached at the end of the strand, and the acceleration is captured at 200 Hz sampling rate. The video is recorded with a Hikrobot area scan camera set at 1440×1080 pixel resolution and 100 fps. Different tension forces are applied to the strand to change its natural modal frequencies. Videos of the steel strand vibrating at different natural modal frequencies are recorded during the tests. Due to factors such as ambient lighting condition and camera sensor, noise in the captured video is expected.

The schematic drawing and the selected rectangle ROI of video frames are shown in Figure 5. The peaks observed in the frequency spectra correspond to the modal frequencies of the strands. The first three modes of natural frequencies of the strands with different tension are shown in Table 1. It can be seen from Table 1 that the frequency range of this study is from 7.813 Hz to 38.086 Hz. The range of the frequency can be extended or narrowed by changing the cable tensile force according to the actual requirement. Finally, the proposed deep learning network is trained with the video frame streams and the corresponding modal frequencies obtained from the displacement spectra.

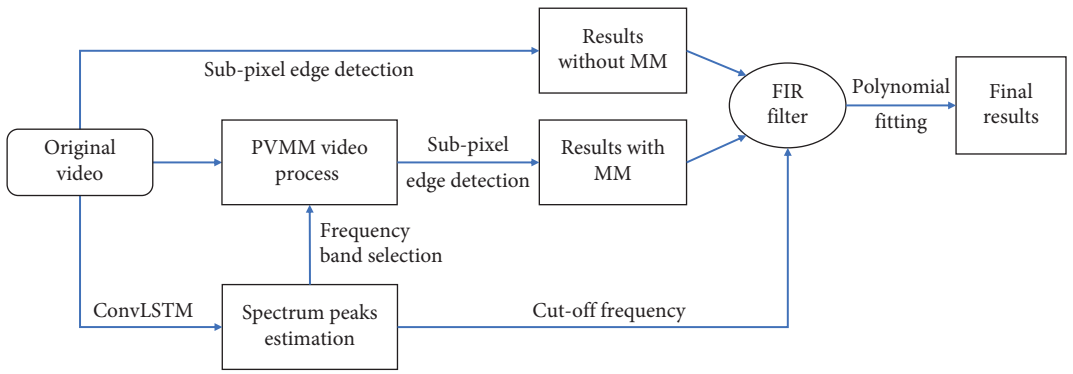


Figure 3: The whole processing of the full-field dynamic displacement measurement.

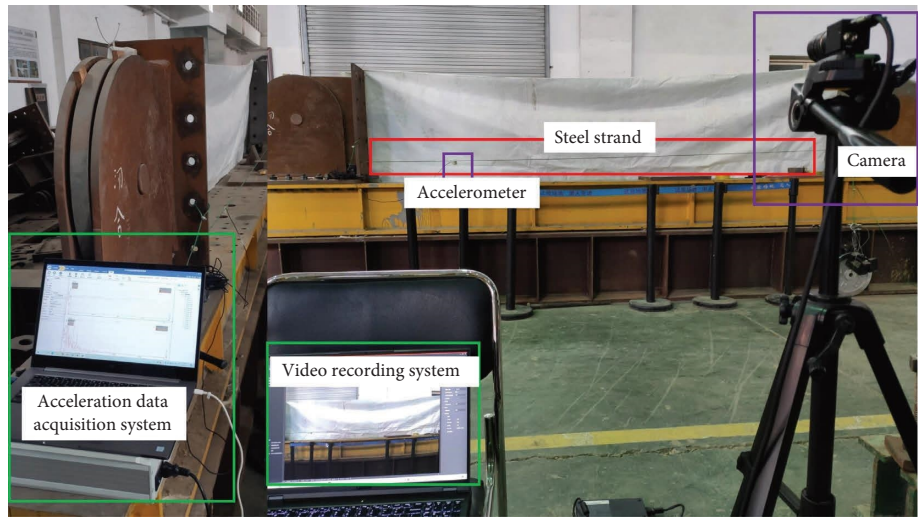


Figure 4: The steel strand experimental set.

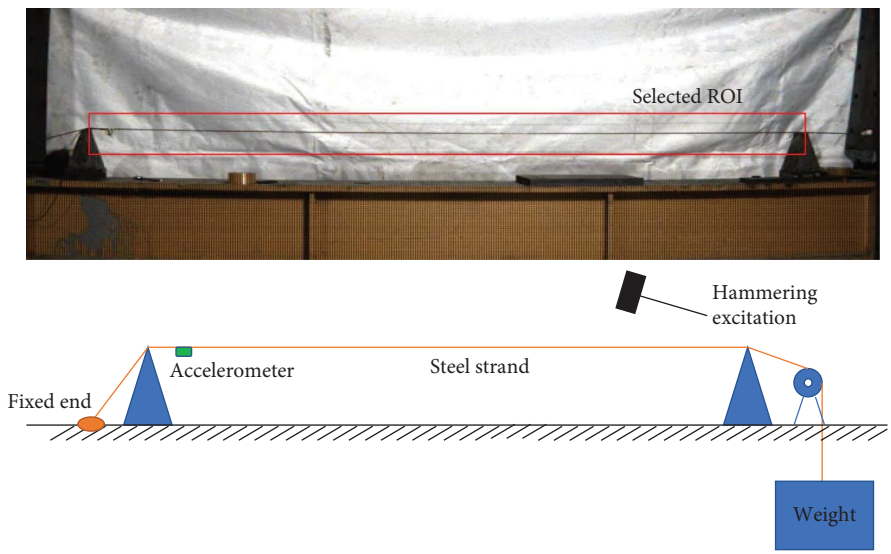


Figure 5: The select rectangle ROI and schematic drawing of the steel strand with supports.

Table 1: The natural frequency (Hz) of the strands.

No.	The first-order frequency	The second-order frequency	The third-order frequency
1	7.813	15.234	22.852
2	9.571	18.756	27.734
3	10.156	19.727	28.906
4	11.523	22.852	33.594
5	12.305	24.609	36.133
6	13.086	25.977	38.086

Each raw frame has the shape of $1440 \times 1080 \times 3$, and the central section of the strand in the video is selected for the deep learning processing. The raw frames were resized into $64 \times 64 \times 3$ section as the input data for the ConvLSTM model. A total of 70400 video frames are collected for the dataset, and each of the set of 200 frames was assigned as one kind of sequential input data for the natural frequency regression task. On the whole, there are 240 training samples, 48 validation samples, and 64 test samples. The frames in training and test dataset are from different videos.

The process of predicting the frequency is shown in Figure 6. The mean absolute error (MAE) method is used to evaluate the accuracy of the proposed ConvLSTM approach, which is also adopted as the loss function for the training of the proposed deep learning architectures. MAE is a measure of the absolute difference between the prediction and the ground truth values. It is given by the following equation:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} \left| \widehat{y}_i - y_i \right|}{n}, \tag{11}$$

where n is the sample size and $\widehat{y}$ and y represent the predicted value and ground truth value, respectively.

Three ConvLSTM networks with different depths are tested in this research. The ConvLSTM models contain different number of ConvLSTM layers from 3 to 1. The Maxpooling and Flatten layers are used to reshape the data inside the model, and the dense layers help to output the results. In Table 2, the hyperparameter configurations of the three networks are listed and the first one of these has the best performance among them.

Hyperparameter tuning involves finding the optimal set of hyperparameters for a particular learning algorithm that minimizes a predefined loss function on a given independent dataset, resulting in a high-performing model [42]. The number of layers, layer size, filter size, learning rate, and batch size are the typical hyperparameters. In the training phase, two components comprising L2 regularization and early stopping algorithms are adopted. By incorporating penalties in the loss function on the layer parameters, it helps to solve the problem of overfitting during optimization. L2 defines the regularization term as the sum of the squares of all the feature weights. Early stopping is another regularization algorithm that can provide guidance on the number of iterations required before the model begins to encounter the problem of overfitting. The learning rate is set as 0.001; batch size is 10; activation function is Relu; and the training epoch is 100.

Besides, the CNN-LSTM can also estimate the vibration frequency with the video input [31]. One CNN-LSTM model made by stacking three convolutional layers and one LSTM layer is given for the comparison. The learning rate, batch size, and kernel size are the same as ConvLSTM model 1. The MAE values are calculated based on the models' performance of the validation samples. The MAE values over the validation dataset are shown in Figure 7, and the model with three ConvLSTM deep layers has the best performance.

Extrapolability is the measure of a model's estimation capability on a dataset outside its training domain range. ConvLSTM model 1 and the CNN-LSTM model are trained on data from any 5 given cases in Table 1 and then deployed for predicting the natural frequency of the remaining one case. In total, there are six experiments. Each experiment comprises training on five cases, followed by testing on the remaining one.

The MAE values of the ConvLSTM and CNN-LSTM models are shown in Figures 8 and 9. Although the MAEs are higher when using unfamiliar datasets for verification, the results of ConvLSTM model are still close to the accelerometer and more accurate than that of CNN-LSTM model. The ConvLSTM model's ability of spatiotemporal information extraction helps it obtain the information from the video frame sequences. Overall, the proposed ConvLSTM model has better performance than the CNN-LSTM model.

3.2. Full-Field Displacement Extraction of the Steel Strand. After the frequency estimation, the PBMM processing is provided to amplify the strand vibration in the video. It is not necessary to process the entire image frame if the ROI has been selected. The ROI selection can reduce the workload and time consumption of feature extraction and matching calculation. The ROI is adjusted to include both the upper edge and the lower edge of the steel strand. The small motion in the selected ROI of the video stream is magnified by the PBMM algorithm. In the motion magnification process, three parameters (frame rate, magnification band, and magnification factor) are required. The natural frequencies can be estimated using the proposed ConvLSTM network, and subsequently the magnification band width can then be chosen. The frequency prediction of the first three modes of natural frequency is 7.868 Hz, 16.156 Hz, and 23.433 Hz. According to the Nyquist–Shannon sampling theorem, the magnification band width should not be over the Nyquist frequency which is half that of the video frame rate. To ensure the accuracy in the motion magnification,

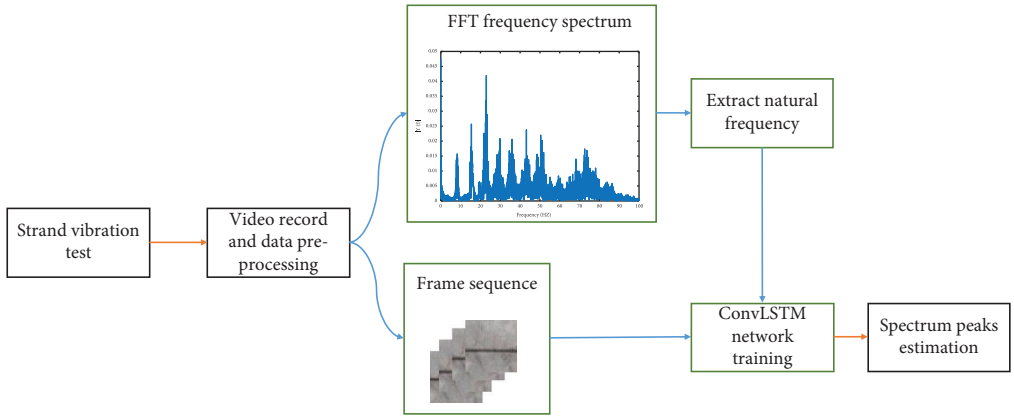


Figure 6: The processing of the frequency estimation.

Table 2: ConvLSTM model structure.

Configuration	Model 1	Model 2	Model 3
ConvLSTM2D	Filters = 32, kernel size = (3, 3)	Filters = 32, kernel size = (3, 3)	Filters = 32, kernel size = (3, 3)
Batch normalization			
ConvLSTM2D	Filters = 32, kernel size = (3, 3)	Filters = 32, kernel size = (3, 3)	—
Batch normalization			
ConvLSTM2D	Filters = 32, kernel size = (3, 3)	—	—
Batch normalization			
Conv2D	Filters = 32, kernel size = (3, 3)	Filters = 32, kernel size = (3, 3)	Filters = 32, kernel size = (3, 3)
MaxPooling2D	Pool size = (4, 4)	Pool size = (4, 4)	Pool size = (4, 4)
Flatten	—	—	—
Dense	512	512	512
Dropout	0.5	0.5	0.5
Dense	512	512	512
Dense	128	128	128
Dense	3	3	3

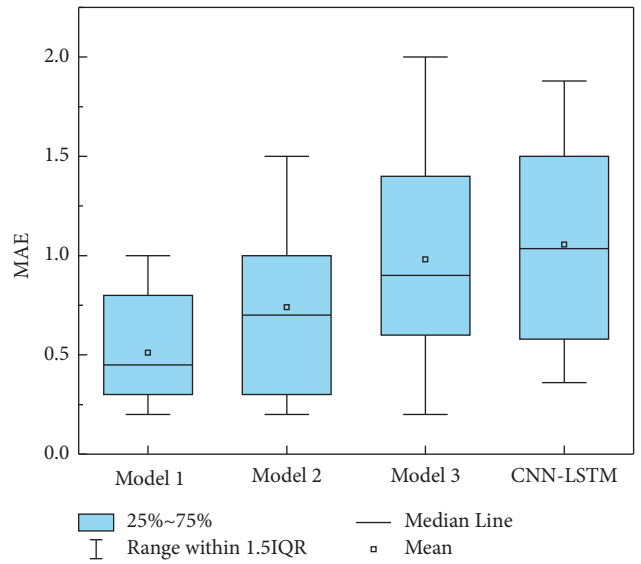


Figure 7: The MAE values for the ConvLSTM and CNN-LSTM models.

broad-band magnification [43] is ideal. But one of the limitations of this technique is that it can cause blur when there are large low-frequency motions in the video [23]. In addition, if the results are normalized according to the magnification factor, the displacement measurement outside the selected band will lose its accuracy. In view of these factors, it is challenging to maintain the accuracy of the results.

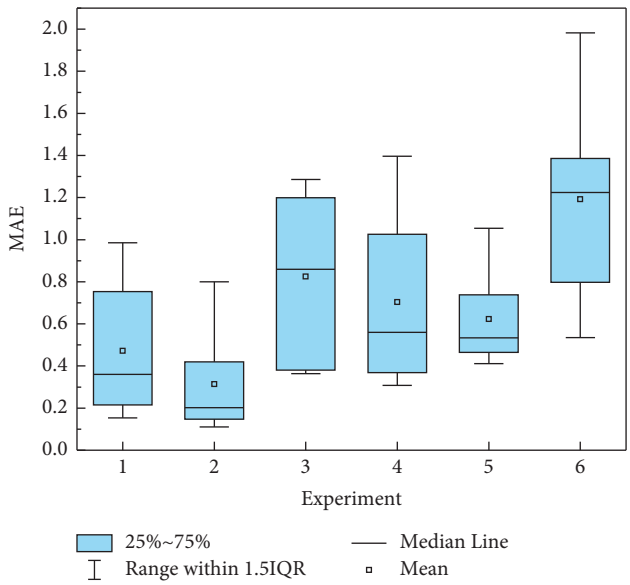


Figure 8: Estimation error of ConvLSTM network under unfamiliar frequencies.

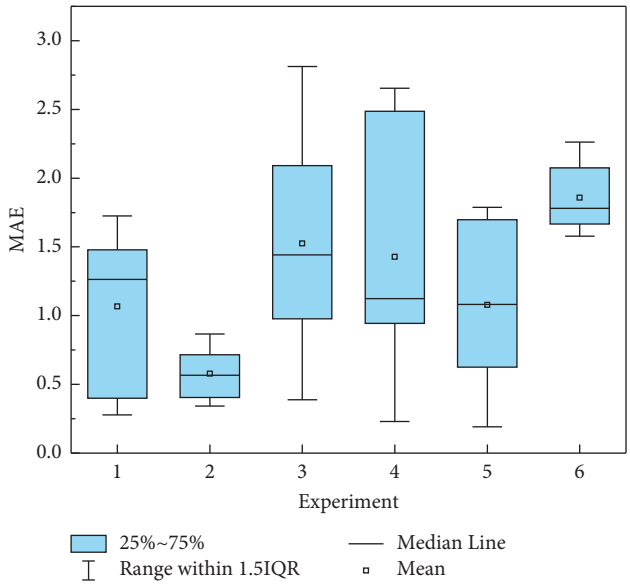


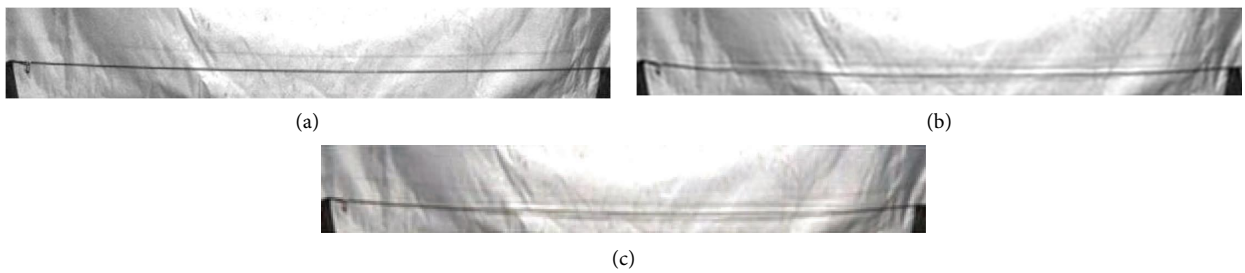
Figure 9: Estimation error of CNN-LSTM network under unfamiliar frequencies.

After trials, the frequency band between the first-order frequency and the Nyquist frequency is selected in the PBMM algorithm. With the subpixel edge detection algorithm, the vibration is firstly measured from the original video, and then the vibration within the selected frequency band is magnified by PBMM and extracted. The FIR filter is used to extract the low-frequency part of the vibration without MM and high-frequency part of the vibration with PBMM. The cutoff frequency is the estimated first spectrum peak. With the superposition of the two parts, the accuracy of the measurement could be improved. The motion magnification results can be significantly affected by the magnification factor. To prevent blurring of the image in the motion magnification process, the magnification factor should be optimized. The PBMM effect of the whole strand is presented in Figures 10(a)–10(c). It can be seen from Figure 10 that it is difficult to locate the strand due to the motion blur when the amplification factor is set to a value of 10. After several trials, the magnification factor of 5 was chosen to produce sufficient level of magnification while minimizing image distortion.

The vibration of the test steel strand under random hammer excitation is amplified by a factor of five using the PBMM algorithm before the subpixel edge detection is applied in this experiment. In Figure 11(a), the vibration amplitude of the steel strand is magnified, and the vibrations of red line sampled region are drawn in y-t slice view. The significant magnification effect is further shown by comparing Figures 11(b) and 11(c). To measure the full-field dynamic displacement of the strand, the edges in every frame need to be detected. The subpixel edge detection of one video frame is shown in Figure 12. The displacements between the first frame and every other frame can be obtained. The distance between the upper and lower edge of the strand is known as the diameter of the strand, which can be used to calibrate the displacement in the images. The edge values of all the pixels along the ROI length are collected. The mean value of the strand upper edge and lower edge is used to describe the strand deformation. Besides, the strand deflection is always in a smooth continuous curve shape, and polynomial fitting can be used to optimize the results. In Figure 13, the subpixel displacement results of the strand in one frame are shown and the predicted displacement of the curve estimated through the use of fourth-degree polynomial fitting is shown. In the original video, the edge curve obtained by the subpixel edge detection is not found to be sufficiently smooth due to the background and the lighting condition as shown in Figures 13(a) and 13(b). After the PBMM process, the vibration amplitude in the video is amplified and the edge result is much less noisy.

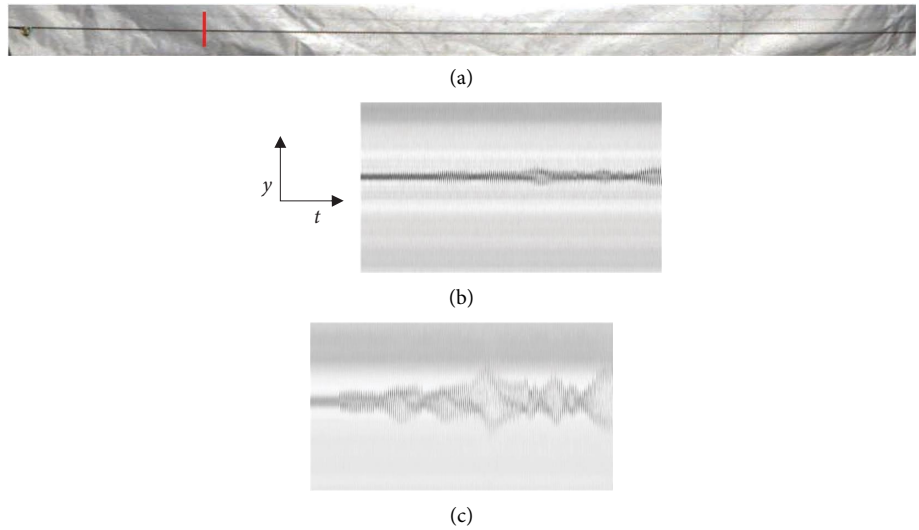
Furthermore, the full-field deformation of the strand during the measurement time is shown in Figure 14; the results are smoothened via fourth-degree polynomial fitting. After smoothening the deformation curves at every moment, the result shows the dynamic displacements of the entire steel strand over time. Following the polynomial fitting, the displacement time history for a given point on the steel strand could be obtained directly from Figure 14. The obtained displacement is further presented in Figure 15. In the original video, slight vibrations with small amplitude and high frequency are not clearly visible in the image, which results in low accuracy and some noise in the subpixel edge detection results. By superpositioning the original video and magnified video with an FIR filter, the fidelity of the high frequency segment is maintained. In addition, the noise from the original video data is eliminated. Following the superpositioning process, the proposed method is found to be able to measure the dynamic displacement accurately.

The comparison of the displacement and acceleration frequency spectra is further shown in Figure 16. The peaks in the displacement spectrum match the acceleration peaks in the first three modes of natural frequencies. The FFT



(a) (b) (c)

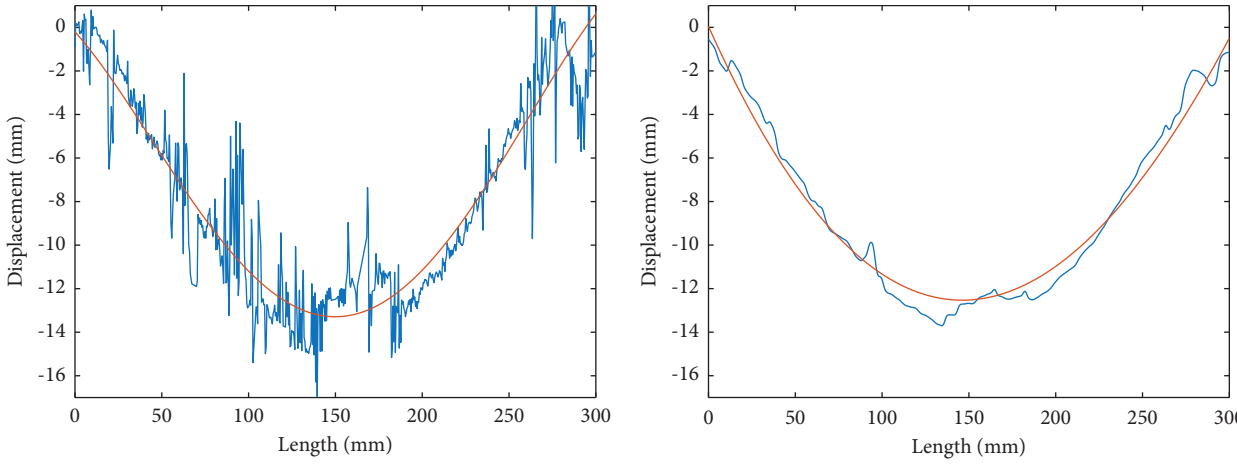
Figure 10: The PBMM effect of the whole strand. (a) Original. (b) Motion magnified 5x. (c) Motion magnified 10x.



(a) (b) (c)

Figure 11: The motion magnification results. (a) One video frame. (b) Original. (c) Motion magnified 5x.

Figure 12: The subpixel edge of the steel strand.



(a) (b)

Figure 13: Fourth-degree polynomial curve fitting of strand deformation at frame number 1800. (a) Deformation curve without motion magnification. (b) Deformation curve with motion magnification.

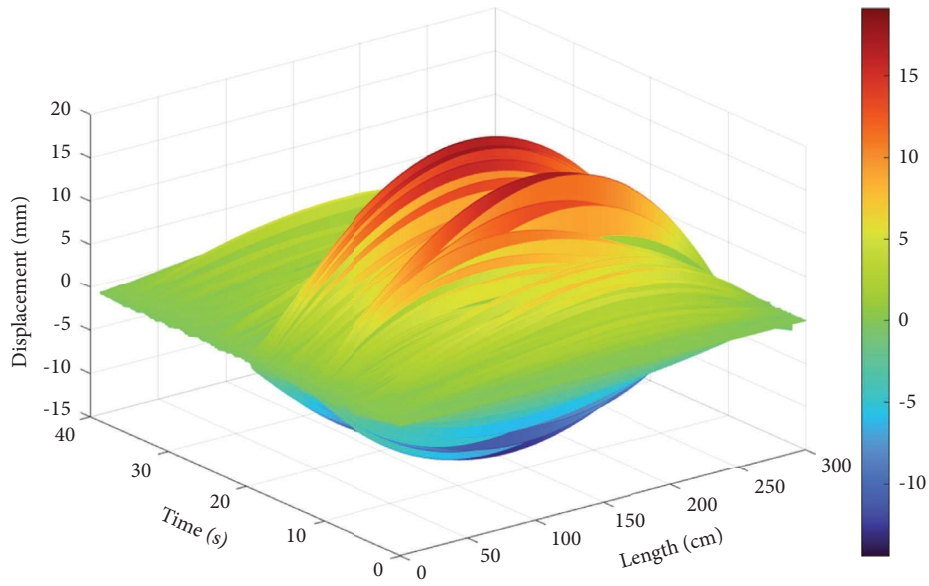


Figure 14: Full-field deformation of the steel strand along time.

amplitude of the acceleration in frequencies lower than 15 Hz does not match the vision methods, and the reason is that the accelerometer is not attached completely along the vibration direction due to the round cross-section of the steel strand. The higher modes of natural frequency are not considered in the analysis. The inaccuracies of the higher modes of vibration are due to the video frame imaging time and synchronization. The performance of the video capture device, the complexity of the video signal, and the efficiency of the video processing can affect the video frame synchronization which can cause inaccuracy in the high frequency. Besides, it is hard to extract the higher order modal frequency peaks above the first-order frequency using the original raw video.

The accuracy of the displacement measurement in the frequency domain can be improved using the proposed method. By adopting the proposed method, the displacement spectrum peaks are clearer than the results of original video. Table 3 shows the first three modes of natural frequency estimated using the accelerometer, deep ConvLSTM, original video extraction, and proposed method, and the relative errors of different methods are calculated, respectively, comparing the accelerometer. Here, it is clear that the frequency peaks obtained via the proposed method with PBMM match well with the acceleration peaks.

4. Beam Full-Field Displacement Measurement Results and Comparison

In this section, an aluminum beam is tested to verify the accuracy of the proposed method. The dynamic displacement of the beam is measured with both the camera and POF sensor. POF sensors are versatile and robust sensors that can measure various physical parameters such as strain and displacement. For dynamic displacement measurements, POF sensors use the principle of intensity modulation, where the variation in the intensity of light transmitted through the fiber is related to the target displacement. By comparing the results obtained from both techniques, the accuracy of the proposed method can be assessed and validated. Each POF sensor can only measure the displacement of one point, and the results at specific points of the beam are compared.

4.1. Comparison of Results to Study the Effect of Varying Vibration Amplitudes. To further verify the proposed measurement method, an aluminum beam with dimensions of $800 \times 12.7 \times 4$ mm is tested in the experiment. Young's modulus, Poisson's ratio, and the density of the beam are 70 GPa, 0.33, and 2700 kg/m^3, respectively. The beam is clipped at both ends which are considered as fixed supports at these locations. The video stream is captured by a Sony a7m3 video camera using the 1080P@100fps format. The setups of the camera, data logger, and aluminum are shown in Figure 17.

To obtain the actual displacements for verification, the displacement collected by plastic optical fiber [44] displacement sensors is used and set as the ground truth for comparison of results. Three POF displacement sensors are placed at the 1/4, 1/2, and 7/8 span of the test beam, measured 200 mm, 400 mm, and 700 mm, respectively, from the left support as shown in Figure 18. The POF sensor system comprises a Keyence FS-N11MN fiber amplifier and a Hioki LR8400 data acquisition unit to capture the vibration of the host beam. This sensor sends a light signal through the POF and receives a reflected light containing information of the vibration of the beam. The reflected light intensity is converted into raw voltage values and is continuously logged via the data acquisition unit at 100 Hz sampling rate. The raw voltage signal is converted to displacement data (in mm) via a calibration constant obtained earlier from a calibration study.

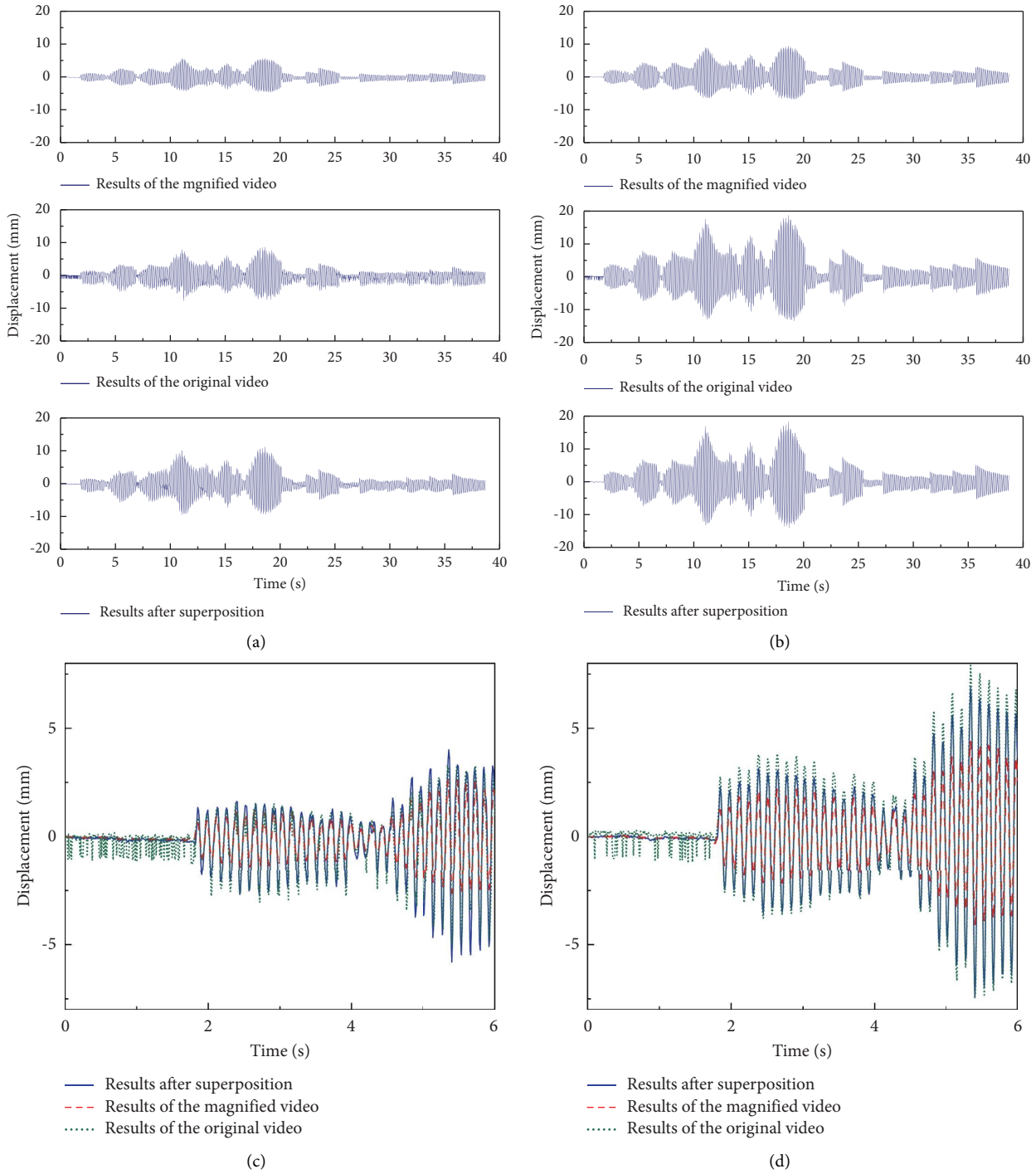


FIGURE 15: Measured displacement time history of one point: (a) at length = 75 cm, (b) at length = 150 cm, (c) zoom-in view of the displacement time history at length = 75 cm, and (d) zoom-in view of the displacement time history at length = 150 cm.

Firstly, the ConvLSTM network predicts the peaks reflected in the vibration spectrum as outlined in Section 3. The deep ConvLSTM network used refers to the pretrained three-layer ConvLSTM summarized in Section 3.1. In the first test, the first three peaks predicted were 3.879 Hz, 11.265 Hz, and 22.571 Hz, respectively. The process was the same as that described in the previous section. The displacement information was extracted with and without applying the MM algorithm. Following that, the low-frequency part of the displacement without MM and the high-frequency part of the displacement with MM were combined.

The video was recorded at 100 frames per second, and correspondingly the Nyquist frequency was 50 Hz. Based on the spectrum peak estimation, the magnification band selected ranged from 3.8 Hz to 50 Hz. A magnification factor of

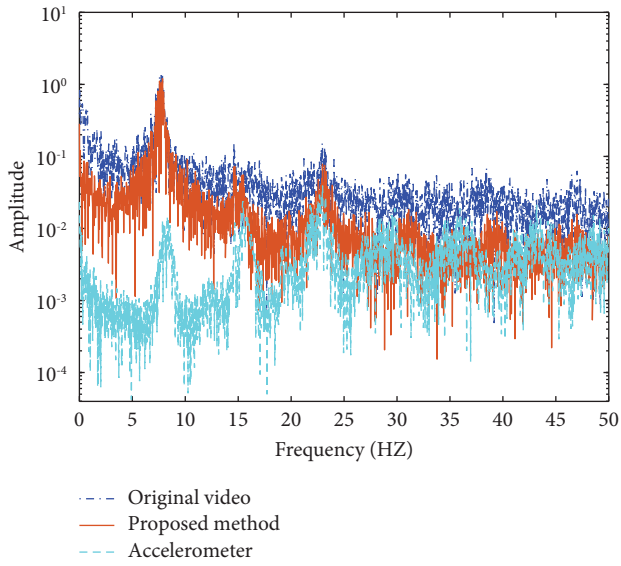


Figure 16: FFT of the strand vibration.

Table 3: The natural frequency (Hz) estimation and the relative error.

Method	Accelerometer	ConvLSTM		Original video extraction		Proposed method	
	Value	Value	Error (%)	Value	Error (%)	Value	Error (%)
The first-order frequency	7.813	7.868	0.7	7.709	1.3	7.792	0.3
The second-order frequency	15.234	16.156	6.1	15.821	3.9	15.726	3.2
The third-order frequency	22.852	23.433	2.5	23.317	2.0	23.179	1.4

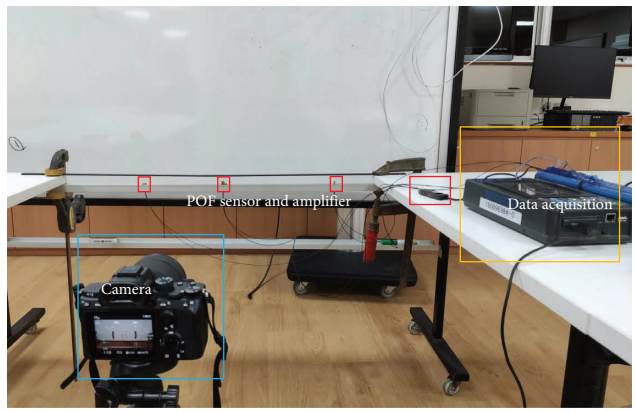


Figure 17: The beam experimental set.

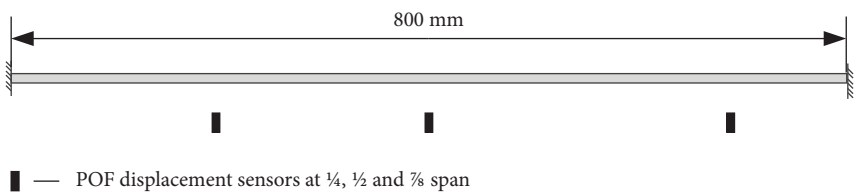


Figure 18: The arrangement of the POF sensors.

5 was adopted. With the subpixel edge detection method, the full-field dynamic deformation can be acquired and calibrated. With the FIR filter, displacement data extracted from the original video vibrating above 3.8 Hz were filtered out and combined to the magnified video displacement data in the frequency band from 3.8 Hz to 50 Hz to obtain the final results.

In this study, different test cases were conducted to verify the accuracy of the proposed method by varying the vibration amplitude of the aluminum beam. In a series of three cases, the free-response vibration of the beam at different initial displacements was captured and processed using the proposed method. Here, the POF displacement sensors were located at these positions along the beam, i.e., at lengths 200 mm, 400 mm, and 700 mm from the left support, respectively. A mass was tied to the beam with a thread at length 400 mm (middle point of the beam) to provide an initial displacement to the beam. The thread was cut, and the beam was left to vibrate freely.

For Case 1, the mass was attached to the beam at position 400 mm where a 10 mm local initial displacement was introduced due to the weight of the mass. The thread was cut to trigger the free vibration of the beam. The vibration of the beam was recorded via a video camera and the POF displacement sensors. The full-field dynamic displacement is shown in Figure 19. The displacement time histories obtained by the proposed method and POF sensors are presented in Figure 20. The root-mean-square error (RMSE) values of the measured displacement by using the proposed method and the method with original video are illustrated in Figure 21. The RMSE is defined as follows:

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(\widehat{y}_i - y_i)^2}{n}}, \tag{12}$$

where $\widehat{y}$ is the predicted value, y is the ground truth value, and n is the sample size.

In Case 2, the mass was applied on the middle span (i.e., 400 mm) and the initial displacement was 5 mm. As before, free vibration of the beam was initiated after the mass was detached and the vibration response was captured. The full-field displacement over time is shown in Figure 22. Comparison of the measurement results from the three POF displacement sensors and the video (original and after processing) is shown in Figure 23. The RMSEs are calculated in Figure 24.

In Case 3, the mass was applied to the same position of the beam as before but here the initial displacement was set to 1 mm. The free vibration response of the beam was recorded as before after detaching the mass. The results are shown in Figures 25–27.

In three cases with different displacement amplitudes, the initial displacement is applied by the hanging mass at the middle span. It can be noticed that RMSEs of the displacement measurements extracted by the proposed method were shown to be nearly 50% lesser than those based on the original video. The maximum RMSE occurs in Case 1 among the three cases due to the biggest displacement amplitude. The RMSEs are 0.88%, 1.64%, and 2.68% of the displacement amplitudes at length 400 mm in Cases 1, 2, and 3, respectively. Though the RMSEs increased with the decrease in amplitude, the proposed method significantly improved the accuracy in the cases with different amplitudes.

4.2. Comparison of Results to Study the Effect of Varying Mass Weights. To further test the robustness of the proposed method, another test series of three cases were conducted where the weight of the mass was varied for different cases. For the first case in this series, Case 4, the initial weight was positioned at 200 mm along the beam from the left support, and the initial deflection at middle point of the beam (i.e., at 400 mm) was 2 mm. The full-field dynamic displacement for Case 4 is shown in Figure 28. The vibration responses of the beam captured via the video camera at specific points along the length of the beam were compared to the POF sensor. The result comparisons are shown in Figures 29 and 30.

In Case 5, the beam structure subjected to abrupt initial displacement was tested to evaluate the accuracy of the proposed method. Three 100 g masses were attached at the middle of the span, and then, each of them was sequentially released to generate the free vibration. Comprising three masses, each weighing 100 g, they were placed at specific position in each test (corresponding to the positions of the POF displacement sensor installed at 200 mm, 400 mm, and 700 mm along the beam, respectively). By cutting the threads holding the masses sequentially, the beam was subjected to different free vibrations under different initial deflection conditions as illustrated in the displacement plots in Figures 31–33.

In Case 6, random excitations using a hammer were introduced to the test beam. The dynamic full-field displacements are captured as before, and the results are shown in Figures 34–36.

In Cases 4 to 6, with varying mass weights, the proposed method can obtain the abrupt and random impact displacements. Like Cases 1, 2, and 3, the RMSE of the proposed method is almost half of that based on original videos in Cases 4, 5, and 6. The RMSEs are 1.26%, 0.64%, and 2.63% of the maximum amplitude in Cases 4, 5, and 6, respectively. In Case 6, the RMSEs of vibration under the random excitation are larger than those in other cases, which may due to the smaller vibration amplitude of the test beam and the distortion caused by the motion magnification.

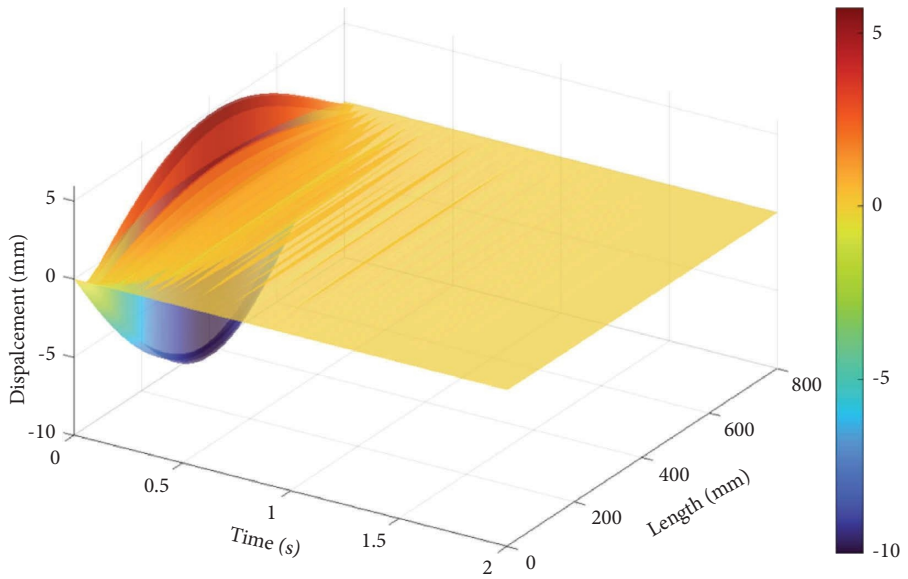


FIGURE 19: Illustration showing the full-field displacement of the beam in Case 1.

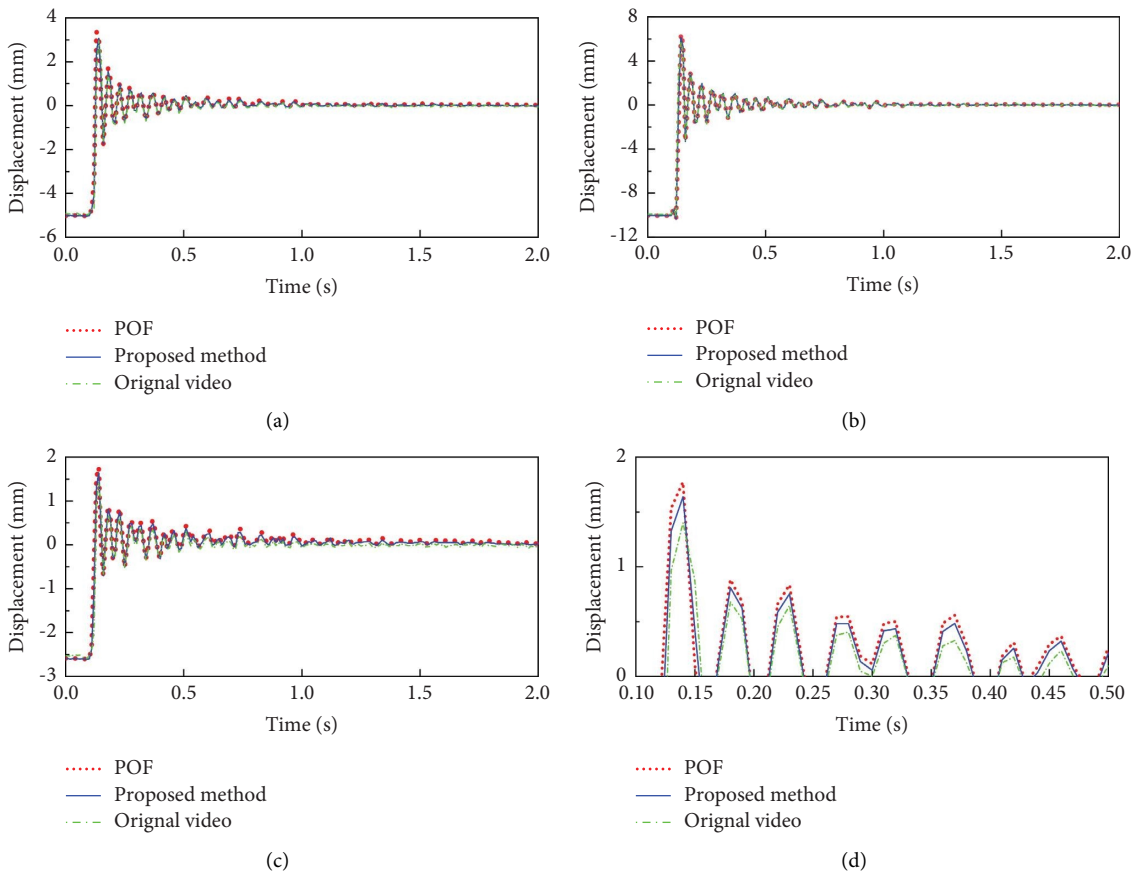


FIGURE 20: Plots showing a comparison of the displacement time history results of POF sensor and those from the video capture (original and processed) at specific points along the beam for Case 1. (a) Length = 200 mm. (b) Length = 400 mm. (c) Length = 700 mm. (d) Length = 700 mm, zoomed in.

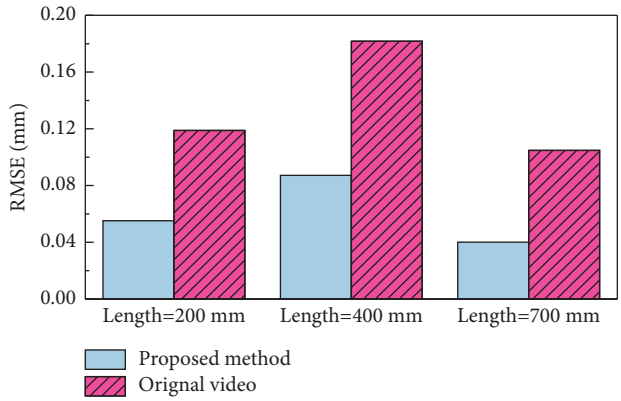


Figure 21: RMSE comparison for Case 1.

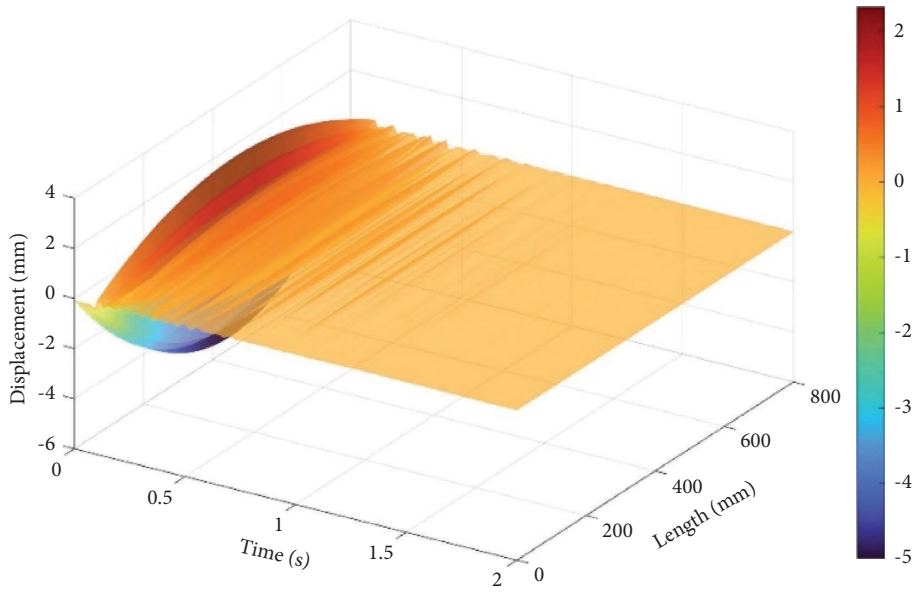


Figure 22: Illustration showing the full-field displacement of the beam in Case 2.

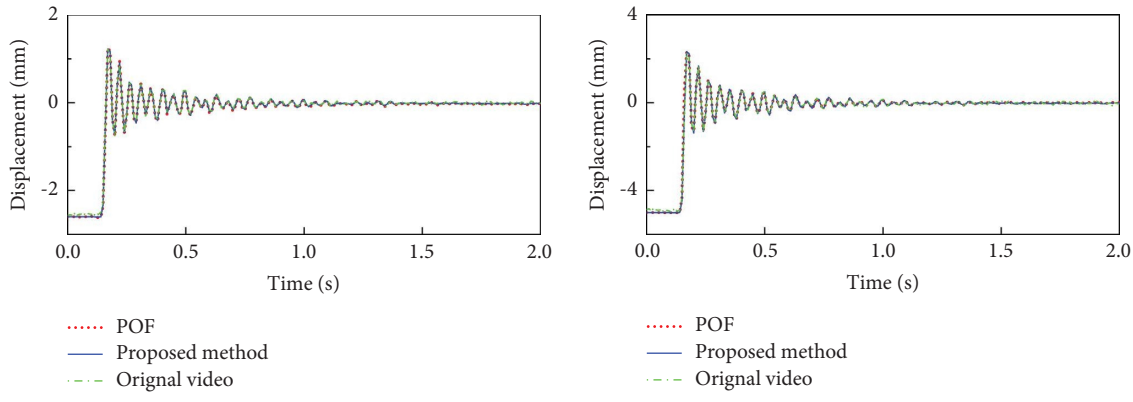


Figure 23: Continued.

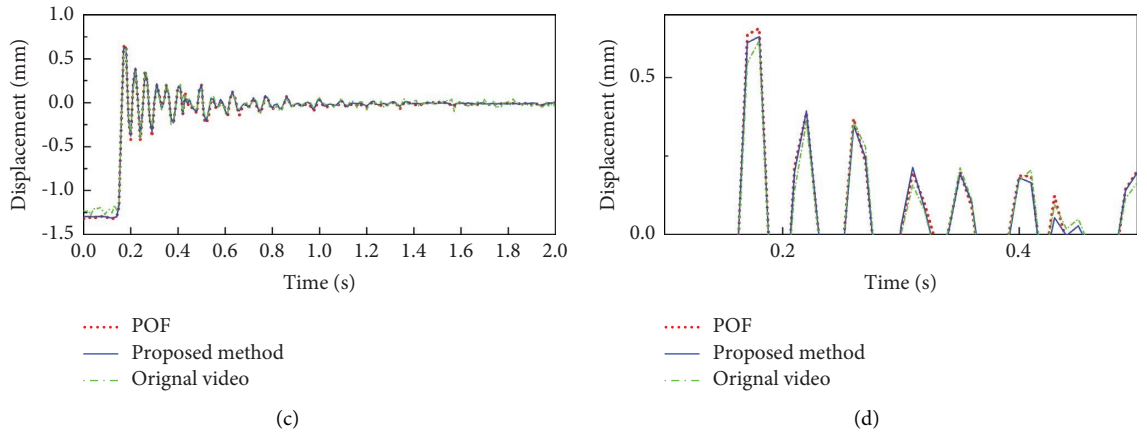


Figure 23: Plots showing a comparison of the displacement time history results of POF sensor and that from the video capture (original and processed) at specific points along the beam in Case 2. (a) Length = 200 mm. (b) Length = 400 mm. (c) Length = 700 mm. (d) Length = 700 mm, zoomed in.

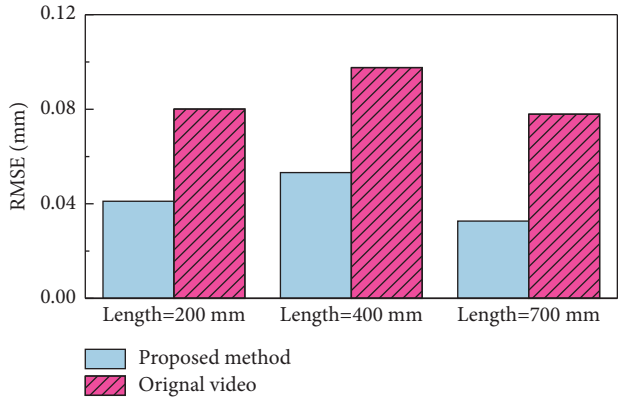


Figure 24: RMSE comparison for Case 2.

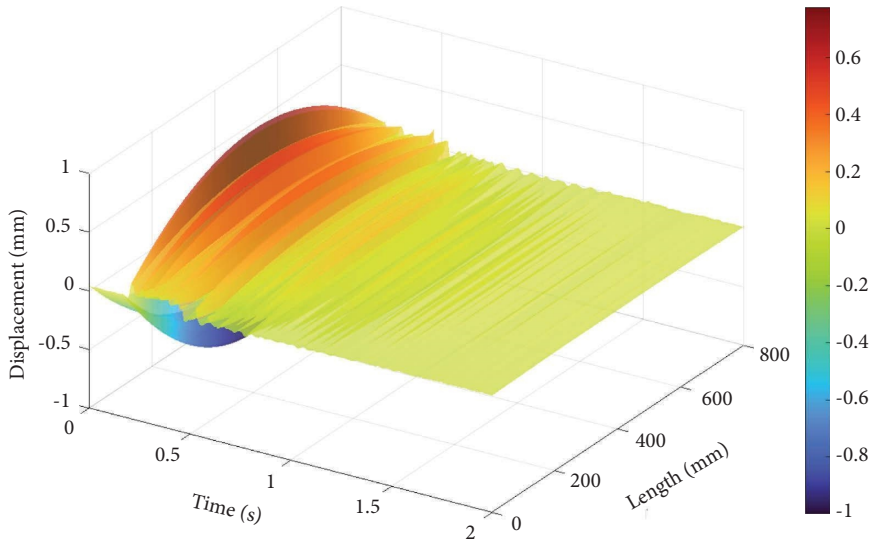


Figure 25: Illustration showing the full-field displacement of the beam in Case 3.

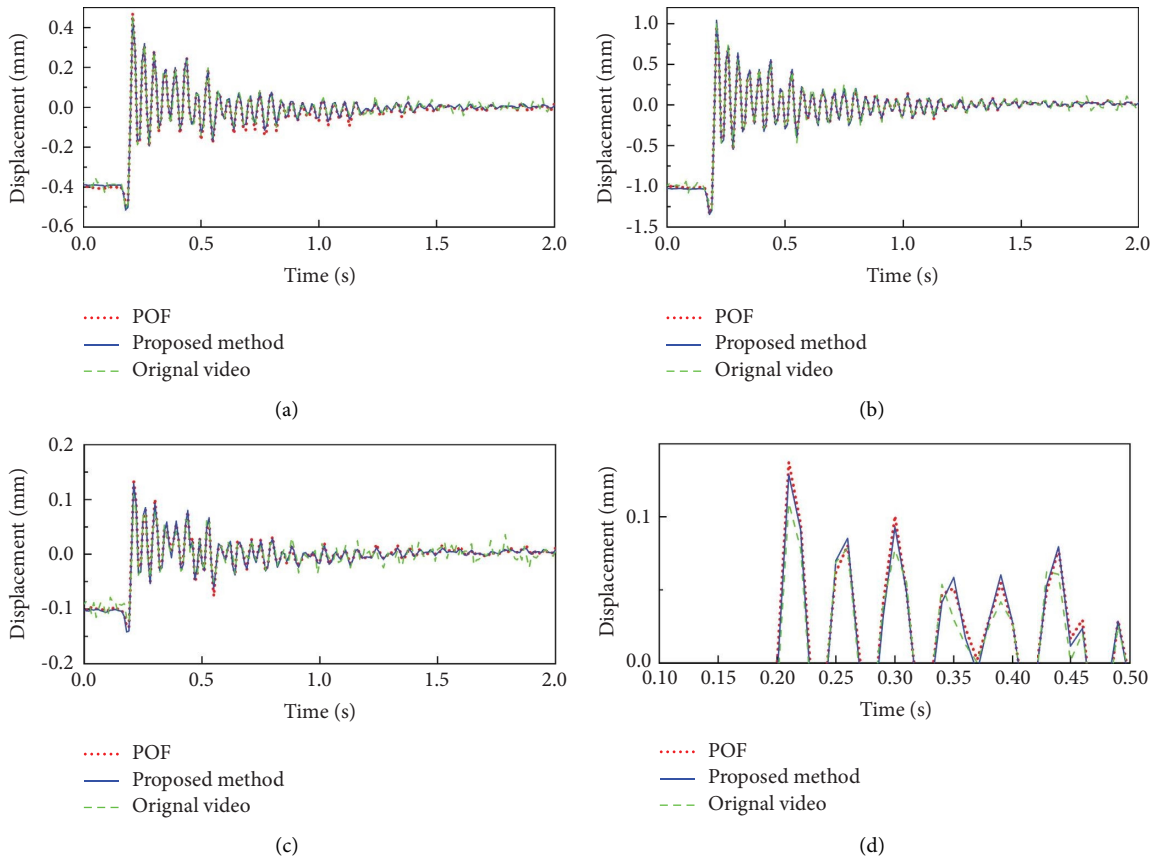


Figure 26: Plots showing a comparison of the displacement time history results of POF sensor and that from the video capture (original and processed) at specific points along the beam for Case 3. (a) Length = 200 mm. (b) Length = 400 mm. (c) Length = 700 mm. (d) Length = 700 mm, zoomed in.

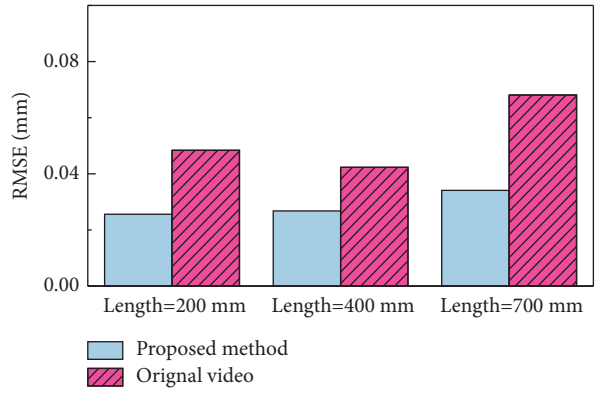


Figure 27: RMSE comparison for Case 3.

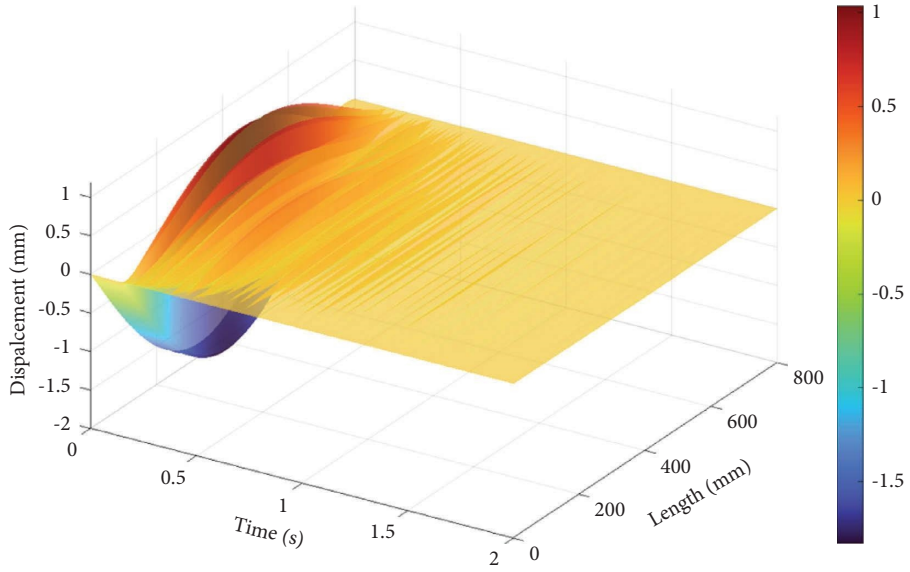


FIGURE 28: Illustration showing the full-field displacement of the beam in Case 4.

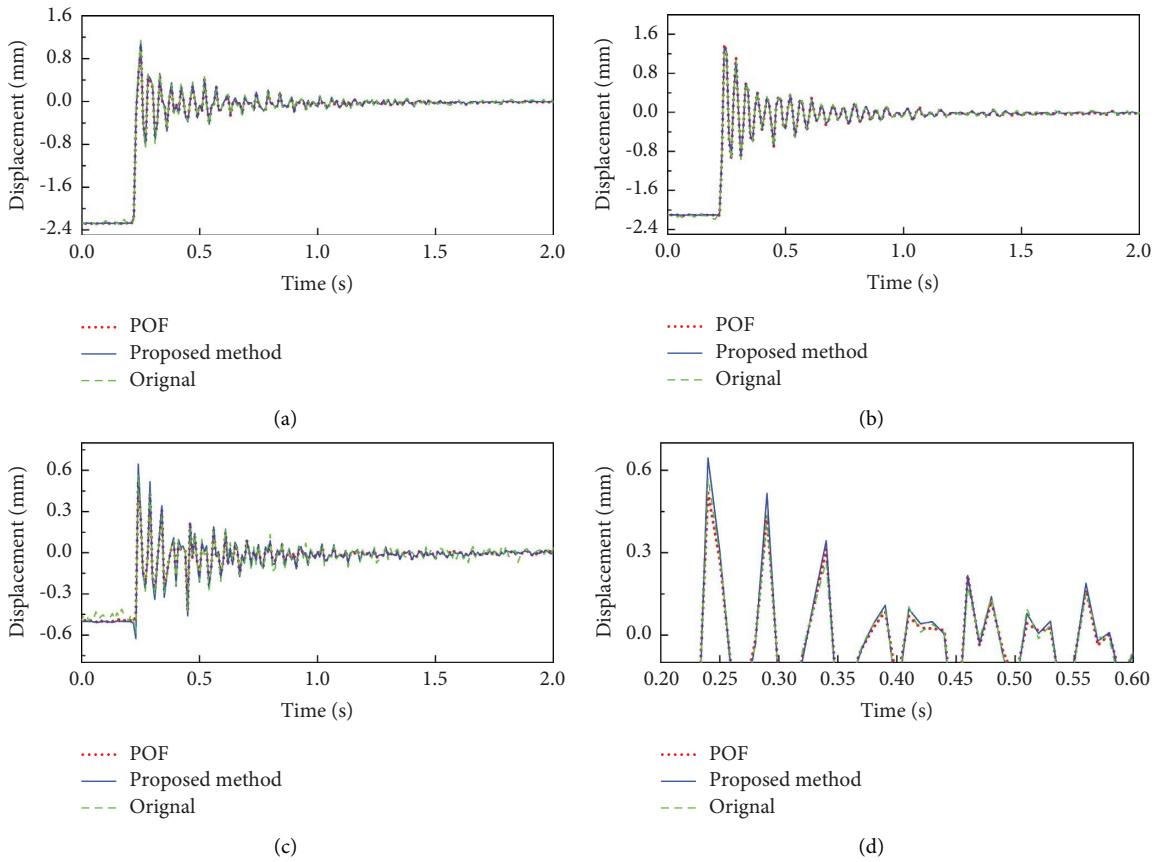


FIGURE 29: Plots showing a comparison of the displacement time history results of POF sensor and that from the video capture (original and processed) at specific points along the beam for Case 4. (a) Length = 200 mm. (b) Length = 400 mm. (c) Length = 700 mm. (d) Length = 700 mm, zoomed in.

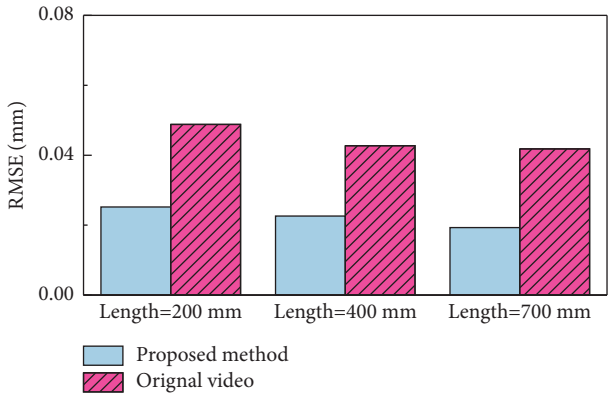


Figure 30: RMSE comparison for Case 4.

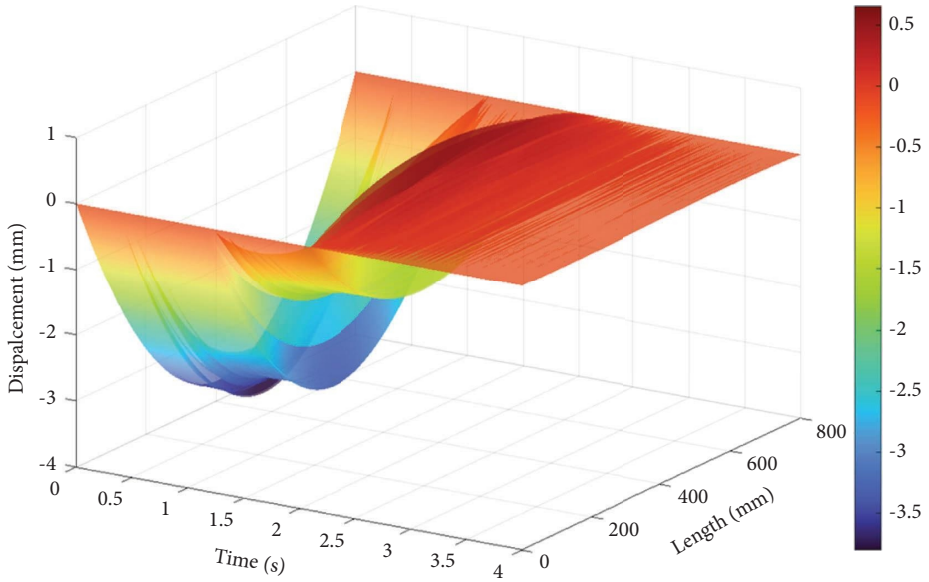


Figure 31: Illustration showing the full-field displacement of the beam in Case 5.

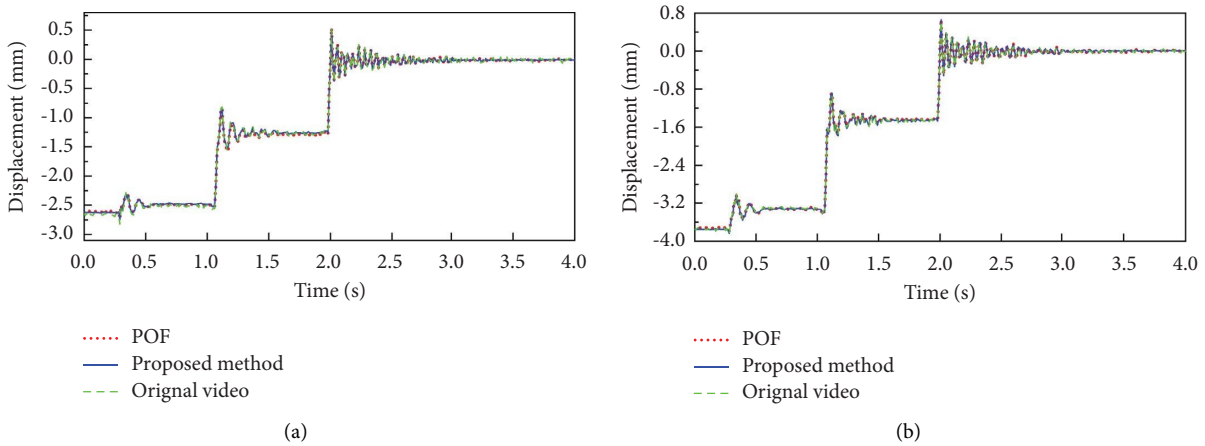


Figure 32: Continued.

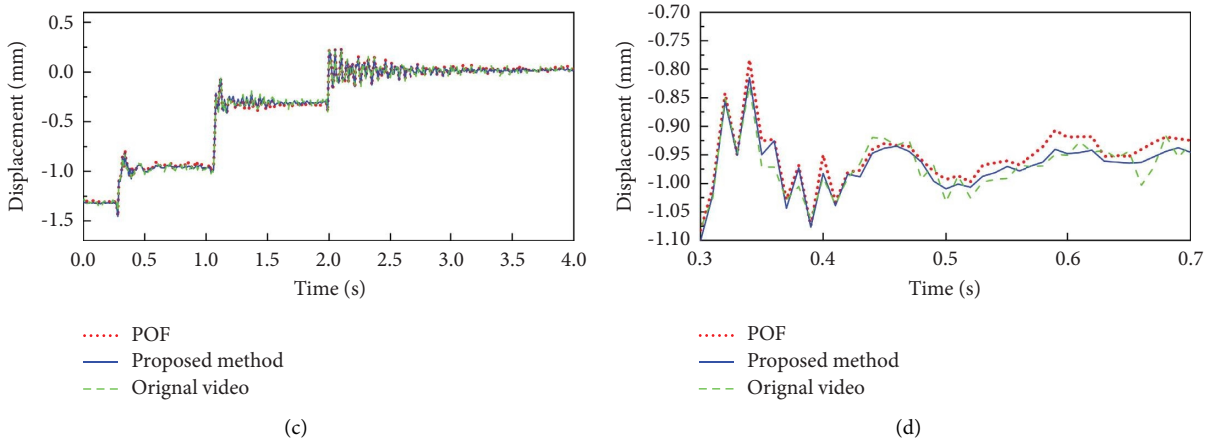


FIGURE 32: Plots showing a comparison of the displacement time history results of POF sensor and that from the video capture (original and processed) at specific points along the beam for Case 5. (a) Length = 200 mm. (b) Length = 400 mm. (c) Length = 700 mm. (d) Length = 700 mm, zoomed in.

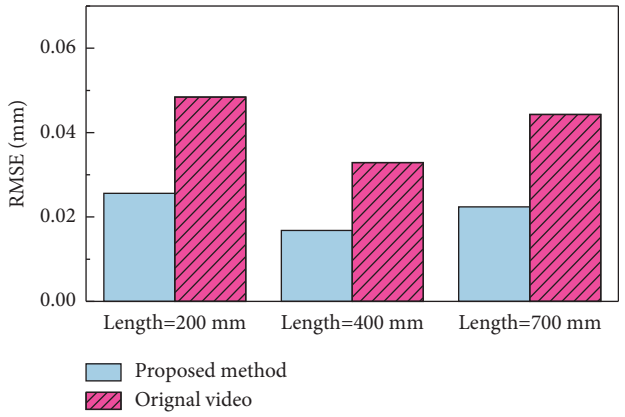


FIGURE 33: RMSE comparison for Case 5.

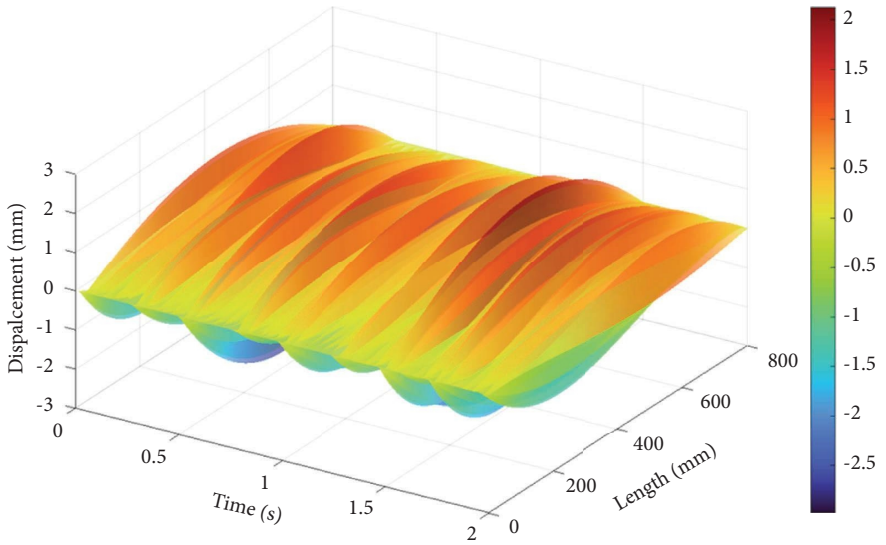


FIGURE 34: Illustration showing the full-field displacement of the beam in Case 6.

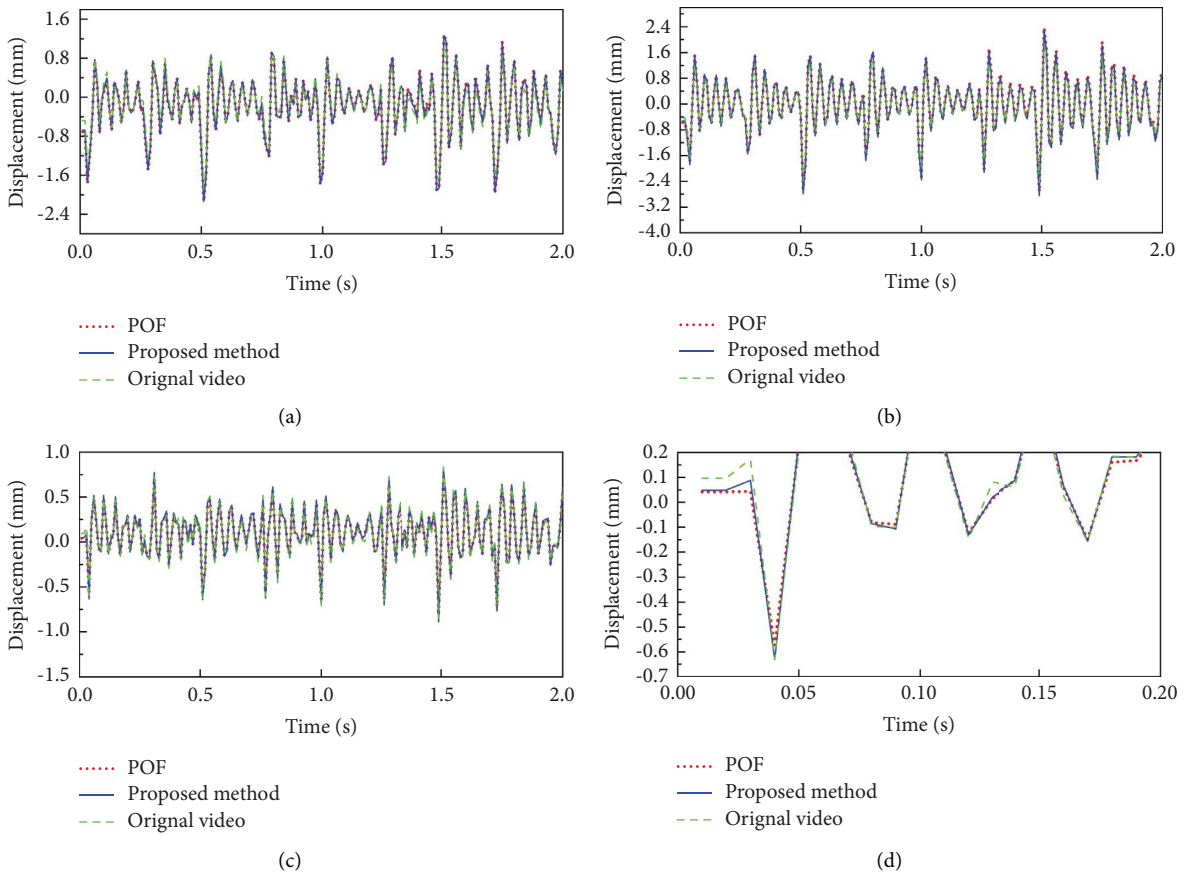


Figure 35: Plots showing a comparison of the displacement time history results of POF sensor and that from the video capture (original and processed) at specific points along the beam for Case 6. (a) Length = 200 mm. (b) Length = 400 mm. (c) Length = 700 mm. (d) Length = 700 mm, zoomed in.

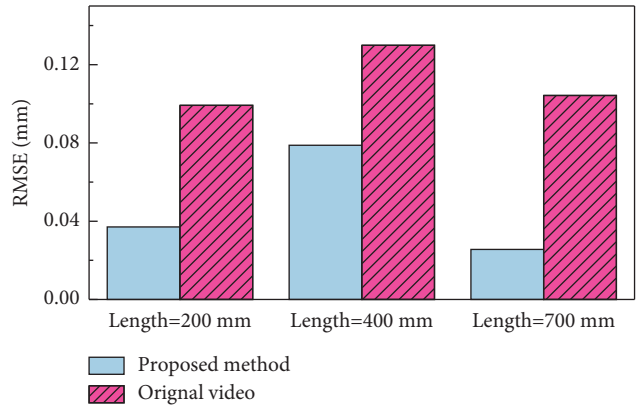


Figure 36: RMSE comparison for Case 6.

5. Conclusions

Though PBMM can significantly magnify the low and high-frequency vibrations, it is hard for the PBMM algorithm to deal with the abrupt and relatively large motions in the video. By using FIR filters, the proposed method combines the displacements extracted from the original video and the video that had undergone the PBMM process, and the accuracy in both low-frequency and high-frequency parts of the full-field displacement could be maintained. The ConvLSTM deep network was used to predict the structure vibration spectrum peaks to help identify the magnification band prior to the PBMM process. Using a magnification factor of 5x, this was found to be sufficiently effective in improving the accuracy of displacement measurement after the superposition of the two kinds of displacement information from the original video and the magnified video, respectively. The full-field dynamic displacements are not smooth along the length at every frame due to the measuring noises in the edge detection and result combination process. In contrast, structures like beams, cables, and tubes tend to exhibit smooth and continuous deformation shapes in the real world. To improve the smoothness of the displacement curves, fourth-degree polynomial fitting is used to optimize the results of the subpixel edge detection process.

(1) The deep ConvLSTM network can efficiently predict the modal frequencies of the vibrating objects in the video images. The accuracy is satisfactory, and the three-layer deep ConvLSTM network has the best accuracy compared to ConvLSTM networks with two or one layer.

(2) In all cases of the aluminum beam test, the POF results are set as ground truth, and the proposed method could achieve the full-field dynamic displacement measurement accurately. The RMSEs based on the raw original video recordings and the processed results based on the proposed method are computed, and the results showed that the proposed method yielded significantly better displacement measurement accuracy compared to the raw unprocessed video recordings. The fusion of the data from the original and the motion magnified videos eliminated the shortcomings of the PBMM algorithm leading to higher displacement measurement accuracy. The proposed method was able to reduce almost 50% of the RMSE in all cases compared to the displacement data extracted from the raw original videos.

(3) The proposed method improved the accuracy of the full-field dynamic displacement measurement using the processes described in the paper. In the cable test, the proposed method measured the full-field dynamic displacements and the vibration spectrum showing clearer modal peaks. The results also illustrated that the proposed method based on subpixel edge detection used in conjunction with the PBMM can produce a time history of the beam displacement information. Compared to methods which only trace limited finite number of points on the structure, the proposed method is superior in achieving high degree of accuracy in capturing the full-field dynamic displacement measurement.

Data Availability

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

Disclosure

The results and opinions expressed in this paper are those of the authors, and they do not necessarily represent those of the sponsors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Financial support to complete this study was provided in part by the National Natural Science Foundation of China under grant no. 52278301 and by the China Scholarship Council Scholarship under grant no. 202106690025.

References

[1] O. Avci, O. Abdeljaber, S. Kiranyaz, M. Hussein, M. Gabbouj, and D. J. Inman, "A review of vibration-based damage detection in civil structures: from traditional methods to Machine Learning and Deep Learning applications," *Mechanical Systems and Signal Processing*, vol. 147, Article ID 107077, 2021.

[2] S. Sony, S. Laventure, and A. Sadhu, "A literature review of next-generation smart sensing technology in structural health monitoring," *Structural Control and Health Monitoring*, vol. 26, no. 3, Article ID e2321, 2019.

[3] Y. Xu, D.-M. Chen, and W. Zhu, "Modal parameter estimation using free response measured by a continuously scanning laser Doppler vibrometer system with application to structural damage identification," *Journal of Sound and Vibration*, vol. 485, Article ID 115536, 2020.

[4] M. Pieraccini, G. Luzi, D. Mecatti et al., "Remote sensing of building structural displacements using a microwave interferometer with imaging capability," *NDT & E International*, vol. 37, no. 7, pp. 545–550, 2004.

[5] J. Yu, X. Meng, B. Yan, B. Xu, Q. Fan, and Y. Xie, "Global Navigation Satellite System-based positioning technology for structural health monitoring: a review," *Structural Control and Health Monitoring*, vol. 27, no. 1, Article ID e2467, 2020.

[6] X.-W. Ye, C.-Z. Dong, and T. Liu, "A review of machine vision-based structural health monitoring: methodologies and applications," *Journal of Sensors*, vol. 2016, Article ID 7103039, 10 pages, 2016.

[7] Y. Yang, C. Dorn, C. Farrar, and D. Mascareñas, "Blind, simultaneous identification of full-field vibration modes and large rigid-body motion of output-only structures from digital video measurements," *Engineering Structures*, vol. 207, Article ID 110183, 2020.

[8] 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, Article ID e1852, 2017.

[9] X. Ye, C. Dong, and T. Liu, "Image-based structural dynamic displacement measurement using different multi-object tracking algorithms," *Smart Structures and Systems*, vol. 17, no. 6, pp. 935–956, 2016.

[10] X. Ye, T.-H. Yi, C. Dong, and T. Liu, "Vision-based structural displacement measurement: system performance evaluation and influence factor analysis," *Measurement*, vol. 88, pp. 372–384, 2016.

[11] M. A. Sutton, J. J. Orteu, and H. Schreier, *Image Correlation for Shape, Motion and Deformation Measurements: Basic Concepts, Theory and Applications*, Springer Science & Business Media, Singapore, 2009.

[12] 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, Article ID e2155, 2018.

[13] J. H. Jeong and H. Jo, "Real-time generic target tracking for structural displacement monitoring under environmental uncertainties via deep learning," *Structural Control and Health Monitoring*, vol. 29, no. 3, Article ID e2902, 2022.

[14] J. Jiao, J. Guo, K. Fujita, and I. Takewaki, "Displacement measurement and nonlinear structural system identification: a vision-based approach with camera motion correction using planar structures," *Structural Control and Health Monitoring*, vol. 28, no. 8, p. e2761, 2021.

[15] J. Baqersad, P. Poozesh, C. Niezrecki, and P. Avitabile, "Photogrammetry and optical methods in structural dynamics–A review," *Mechanical Systems and Signal Processing*, vol. 86, pp. 17–34, 2017.

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

[17] S. Bhowmick and S. Nagarajaiah, "Identification of full-field dynamic modes using continuous displacement response estimated from vibrating edge video," *Journal of Sound and Vibration*, vol. 489, Article ID 115657, 2020.

[18] S. Bhowmick, S. Nagarajaiah, and Z. Lai, "Measurement of full-field displacement time history of a vibrating continuous edge from video," *Mechanical Systems and Signal Processing*, vol. 144, Article ID 106847, 2020.

[19] C.-Z. Dong, O. Celik, F. N. Catbas, E. J. O'Brien, and S. Taylor, "Structural displacement monitoring using deep learning-based full field optical flow methods," *Structure and Infrastructure Engineering*, vol. 16, no. 1, pp. 51–71, 2019.

[20] J. Ye, G. Fu, and U. P. Poudel, "High-accuracy edge detection with blurred edge model," *Image and Vision Computing*, vol. 23, no. 5, pp. 453–467, 2005.

[21] C. Liu, A. Torralba, W. T. Freeman, F. Durand, and E. H. Adelson, "Motion magnification," *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 519–526, 2005.

[22] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Transactions on Graphics*, vol. 31, no. 4, pp. 1–8, 2012.

[23] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, "Phase-based video motion processing," *ACM Transactions on Graphics*, vol. 32, no. 4, pp. 1–10, 2013.

[24] M. Civera, L. Zanotti Fragonara, and C. Surace, "An experimental study of the feasibility of phase-based video magnification for damage detection and localisation in operational deflection shapes," *Strain*, vol. 56, no. 1, Article ID e12336, 2020.

[25] M. Civera, L. Zanotti Fragonara, P. Antonaci, G. Anglani, and C. Surace, "An experimental validation of phase-based motion magnification for structures with developing cracks and time-varying configurations," *Shock and Vibration*, vol. 2021, Article ID 5518163, 16 pages, 2021.

[26] M. Silva, B. Martinez, E. Figueiredo, J. C. W. A. Costa, Y. Yang, and D. Mascareñas, "Nonnegative matrix factorization-based blind source separation for full-field and high-resolution modal identification from video," *Journal of Sound and Vibration*, vol. 487, Article ID 115586, 2020.

[27] S. Wangchuk, D. M. Siringoringo, and Y. Fujino, "Modal analysis and tension estimation of stay cables using non-contact vision-based motion magnification method," *Structural Control and Health Monitoring*, vol. 29, no. 7, Article ID e2957, 2022.

[28] Y. Shao, L. Li, J. Li, S. An, and H. Hao, "Computer vision based target-free 3D vibration displacement measurement of structures," *Engineering Structures*, vol. 246, Article ID 113040, 2021.

[29] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[31] R. Yang, S. K. Singh, M. Tavakkoli et al., "CNN-LSTM deep learning architecture for computer vision-based modal frequency detection," *Mechanical Systems and Signal Processing*, vol. 144, Article ID 106885, 2020.

[32] M. S. Dizaji, Z. Mao, and M. Haile, "A hybrid-attention-ConvLSTM-based deep learning architecture to extract modal frequencies from limited data using transfer learning," *Mechanical Systems and Signal Processing*, vol. 187, Article ID 109949, 2023.

[33] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[34] A. Trujillo-Pino, "Accurate subpixel edge location based on partial area effect," *Image and Vision Computing*, vol. 31, no. 1, pp. 72–90, 2013.

[35] T. Gautama and M. 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.

[36] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Pearson Education, London, UK, 2014.

[37] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 49–70, 2000.

[38] Z. Sun and M. Zhao, "Short-term wind power forecasting based on VMD decomposition, ConvLSTM networks and error analysis," *IEEE Access*, vol. 8, pp. 134422–134434, 2020.

[39] J. Brownlee, *Deep Learning For Time Series Forecasting: Predict The Future With Mlps, CNNS and LSTMS In Python*, Machine Learning Mastery, Vermont, Australia, 2018.

[40] C. Shi, Z. Zhang, W. Zhang, C. Zhang, and Q. Xu, "Learning multiscale temporal–spatial–spectral features via a multipath convolutional LSTM neural network for change detection with hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

[41] R. Vrskova, R. Hudec, P. Kamencay, and P. Sykora, "A new approach for abnormal human activities recognition based on ConvLSTM architecture," *Sensors*, vol. 22, no. 8, Article ID 2946, 2022.
[42] M. Claesen and B. De Moor, "Hyperparameter search in machine learning," 2015, https://arxiv.org/abs/1502.02127.
[43] M. Eitner, B. Miller, J. Sirohi, and C. Tinney, "Effect of broad-band phase-based motion magnification on modal parameter estimation," *Mechanical Systems and Signal Processing*, vol. 146, Article ID 106995, 2021.
[44] K. Kuang and W. Cantwell, "The use of plastic optical fibre sensors for monitoring the dynamic response of fibre composite beams," *Measurement Science and Technology*, vol. 14, no. 6, pp. 736–745, 2003.