

Vibration Measurement With Video Processing Based on Alternating Optimization of Frequency and Phase Shifts

Jiwen Zhou[✉], Hongguang Li[✉], Li Zhang[✉], Xiaojian Wang[✉], and Yun Li[✉]

Abstract—Vision-based measurement methods have attracted growing interest over the past years because of their non-contact sensing capability for acquiring the full-field vibrations of structure. Existing vision-based measurement techniques, such as point tracking (PT) and digital image correlation (DIC), require preparation for the surface with high-contrast markers or speckle patterns before measurement. The phase-based motion magnification, enabling to extract the vibration of the structure without surface pre-processing, is a newly developed vision-based measurement method. However, it suffers from the problems, such as phase noise, phase instability, and phase unwrapping. In this article, a novel iterative algorithm called alternating optimization of frequency and phase shifts (AOFPS) is developed to measure the vibration signals from the video instead of explicitly manipulating the phase. The method of Tikhonov regularization for estimating pixel-to-pixel sinusoid patterns and the approach of the least squares for extracting frame-to-frame random phase shifts are alternately implemented to track the motion of the object. This algorithm provides robust and accurate vibration measurements without expensive computation. Simulated and laboratory experiments analyses illustrate the effectiveness and accuracy of the proposed algorithm.

Index Terms—Angular spatial frequency estimation, monocular vision, motion magnification, phase shifts, vibration measurement.

I. INTRODUCTION

VIBRATION measurement is fundamental to the modal analysis of structures [1], [2], condition monitoring and fault diagnosis of rotating machinery [3]–[5], health monitoring and damage detection of civil infrastructure [6], [7], and vital signs tracking of humans [8]. Accelerometers, usually attached to the vibrating structures, are widely used to collect vibrating data of structures. However, as a contact-based wired measurement technique, densely instrumenting the accelerometers to the large-scale structure is a laborious and time-consuming task. Besides, it is hard to apply the accelerometers

Manuscript received March 22, 2021; revised September 1, 2021; accepted September 12, 2021. Date of publication September 29, 2021; date of current version November 22, 2021. This work was supported by the National Natural Science Foundation of China under Grant 11972222. The Associate Editor coordinating the review process was Jochen Lang. (*Corresponding author: Hongguang Li*)

Jiwen Zhou, Hongguang Li, Xiaojian Wang, and Yun Li are with the State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: zhou.jiwen@sjtu.edu.cn; hgli@sjtu.edu.cn; wxj123@sjtu.edu.cn; liyun14@sjtu.edu.cn).

Li Zhang is with Shanghai Institute of Spacecraft Equipment, Shanghai 200240, China (e-mail: supzl@163.com).

Digital Object Identifier 10.1109/TIM.2021.3116322

in high-temperature environments. Particularly, the contact accelerometers will introduce extra mass to the tested structure inevitably, which will change the dynamic characteristics of the tested object, especially for the lightweight structure. This drawback sometimes cannot be tolerated in the field of modal analysis and damage detection. Furthermore, accelerometers can only achieve sparse measurements at a few discrete locations, which is far inadequate for the modal-based damage detection. To avoid these defects, non-contact vibration measurement techniques, such as eddy current sensor [9], laser vibrometry [10], and microwave interferometry [11] have been studied intensely. Though these techniques have seen many advances in vibration measurements, they still suffer from the problem of simultaneously and synchronously collecting data at multiple locations. Alternatively, the vision-based vibration measurement technique provides a new way to address these limitations of contact and non-contact sensing methods.

The vision-based vibration measurement approach [12] is a non-contact measurement method that can obtain displacement and deformation of the vibrating structure via a series of image processing techniques. The camera is the main instrument for this measurement method. Only the measurement of in-plane motion can be made by using a single camera, the out-of-plane motion measurement requires at least two cameras. As a photogrammetric technique, the point tracking (PT) method [13] is used to acquire the vibrating data by detecting and tracking the optical targets mounted on the tested structures. The accuracy and robustness of the measurement depend mainly on the quality and quantity of the optical targets. Digital image correlation (DIC) [14] is another vision-based approach to vibration measurement that calculates the displacement and strain by correlating the random speckle patterns with the referenced patterns over the region of interest. It is a continuous and full-field measurement technique. However, gray-scale variations of continuous patterns on the surface of the tested structure must be prepared before the measurement, which limits the application of this method in the case of surface treatment prohibited. Therefore, a vibration measurement technique is urgently needed without surface pre-processing.

Recently, the motion magnification technique was developed to magnify the tiny movements in videos and visualize the mode shapes of the vibrating structures [1], [15]–[19]. The motion magnification technique transforms the image sequences of the video into complex-valued image pyramids,

the phase variations of the filtering responses between adjacent frames correspond to the displacement of the vibrating structure. Combining with the method of phase-based optical flow [20], the vibration data can be easily acquired from video with the motion magnification method. Another method to measure the sub-pixel displacement of the object with the phase is depicted in [21]. However, due to the presence of phase instability and phase noise [20], [22], [23], the estimation of vibration using the phase-based optical flow method is not robust. In addition, because of the periodicity of the phase within the interval $[-\pi, \pi]$, there is a phase ambiguity that may cause artifacts in the vibration estimation. A confidence estimate was developed in [24] to correct the phase ambiguity. The phase difference greater than the given tolerance of a pixel between two scales is considered unreliable and is corrected by the shift correction method. Despite this algorithm working well in the phase-based frame interpolation [24], it is computationally intensive because the phase unwrapping process is implemented on multiple scales. An algebraic phase unwrapping method [25] is used to correct the phase ambiguity in [16], but as a path-following approach, it suffers from the issue of error propagation.

To avoid the above limitations, we proposed a novel iterative algorithm called alternating optimization of frequency and phase shifts (AOPFS) to estimate the vibration in videos inspired by Wang and Han [26] and Kayaba and Kokumai [27]. By applying the oriented spatial filters [28] to each frame of video, the filtering responses are the complex-valued sinusoids windowed by a Gaussian envelope. Instead of calculating local amplitude and local phase like that in [1], the real part that corresponds to the even-symmetric filter (cosine) and imaginary counterpart that corresponds to the odd-symmetric filter (sine) of each complex sinusoid are directly utilized to estimate the vibration within a local region. A Tikhonov regularization-based optimization algorithm is first used to estimate the angular spatial frequency of each sinusoid pattern. The angular spatial frequency in the same region of the image among each frame can be considered as a constant under the assumption of tiny motion. Then, the least squares method is utilized to estimate the temporal phase shifts frame-to-frame. The vibration displacements are derived by eliminating the angular spatial frequency from the temporal phase differences. Finally, an iterative procedure is introduced to enhance the robustness and accuracy of the algorithm. Simulated and laboratory experiments are carried to validate the effectiveness of the proposed algorithm. Compared with the work in [1], our algorithm can avoid explicitly manipulating the phase, thus it is immune to those problems induced by the phase, such as phase-instability, -derivation, and -wrapping. By using the inherent characteristics of the image, no additional digital projector is needed to reproduce fringe patterns on the surface of the structure.

The remainder of this article is organized as follows. Section II describes the theory of phase-based motion representation. Then the link is built between the vibration displacements and the video frames. In Section III, the iterative algorithm for displacement extraction based on Tikhonov regularization and the least squares method is described in detail,

and the implementation of the algorithm is also provided. In Section IV, simulated and experimental results are presented to validate the effectiveness of the AOPFS algorithm. Section V draws a conclusion.

II. THEORY OF PHASE-BASED MOTION REPRESENTATION

In this section, the principle of the displacement measurement with the phase-based optical flow method and the challenges of estimating the vibration with phase data are presented.

A. Phase-Based Motion Representation

Structure motion can be recorded by a video camera. Considering an image with its intensity $I(x, y)$ where (x, y) is the pixel coordinate, the motion-induced frame can be represented as $I(x + \delta(x, t), y + \delta(y, t))$ where $\delta(x, t)$ and $\delta(y, t)$ are the temporally varying displacements along x and y direction, respectively. For the sake of simplicity, only the one-dimensional (1-D) motion $I(x + \delta(x, t))$ is considered here. Using the Fourier Transform, the image intensity profile can be decomposed into a sum of complex sinusoids [15]

$$I(x + \delta(t)) = \sum_{\omega=-\infty}^{\infty} A_{\omega} \exp[i \omega (x + \delta(t))] \quad (1)$$

where each sub-band corresponds to a single frequency component ω .

Considering an arbitrary sub-band

$$S_{\omega}(x, t) = A_{\omega} \exp[i \omega (x + \delta(t))] \quad (2)$$

where $S_{\omega}(x, t)$ is a complex sinusoid with its local amplitude A_{ω} and local phase $\omega[x + \delta(t)]$. We can see that the local phase contains the motion $\delta(t)$. In the case of the vibration measurement, our goal is to estimate the translated displacement $\delta(t)$ from the sinusoid sub-band $S_{\omega}(x, t)$ at each pixel and time instant accurately and robustly.

For the frames of the video where vibrations are recorded, the image structure is usually non-periodic, a filter bank called complex steerable pyramid [28] is applied to transform the image frame into the sum of sinusoid-like sub-bands. The complex steerable pyramid filters decompose an image according to spatial scale, orientation, and position [15]. The details of the complex steerable pyramids can be found in [28]. Specifically, a sub-band at pixel (x, y) and time t with the fixed spatial scale ω and orientation θ can be denoted as

$$S_{\omega,\theta}(x, y, t) = \Psi_{\omega,\theta} \otimes I(x, y, t) \quad (3)$$

where \otimes is the convolution operator. $\Psi_{\omega,\theta}$ is a spatial convolution kernel and defined as

$$\Psi_{\omega,\theta}(x, y) = \exp\left(-\frac{x_{\theta}^2 + \gamma^2 y_{\theta}^2}{2\sigma^2}\right) \cdot \exp(i(\omega x_{\theta} + \psi)) \quad (4)$$

where $x_{\theta} = x \cos\theta + y \sin\theta$ and $y_{\theta} = -x \sin\theta + y \cos\theta$. ω and θ represent the orientation and scale of the filter, respectively. σ , γ , and ψ denote the standard deviation of Gaussian function, spatial aspect ratio, and phase offset, respectively. It should be noted that such a filtering procedure can also be performed in the frequency domain [28], [29]. When the

sub-band has been extracted, the motion information can be obtained from the local phase. In Section II-B, the extraction of motion displacement $\delta(t)$ will be described in detail.

B. Displacement Extraction With Phase-Based Optical Flow

According to (2), the local phase can be easily obtained as

$$\begin{aligned}\phi(x, t) &= \tan^{-1} \left(\frac{\text{Im}[S_\omega(x, t)]}{\text{Re}[S_\omega(x, t)]} \right) \\ &= \omega[x + \delta(t)]\end{aligned}\quad (5)$$

where $\text{Re}[\cdot]$ and $\text{Im}[\cdot]$ denote the real and imaginary parts of a complex exponential signal, respectively. Differentiating with respect to spatial location and time, it yields

$$\frac{\partial \phi(x, t)}{\partial x} = \omega \quad (6)$$

$$\frac{\partial \phi(x, t)}{\partial t} = \omega \frac{\partial \delta(t)}{\partial t}. \quad (7)$$

The displacement $\delta(t)$ can be obtained as

$$\delta(t) = \frac{1}{\omega} \int_0^t \frac{\partial \phi(x, t)}{\partial t} dt. \quad (8)$$

Intuitively, the implementation of this phase-based optical flow method is not difficult. However, due to the phase computed with the arctangent function is periodic within an interval $[-\pi, \pi]$, which is known as the wrapped phase, it is impossible to differentiate the wrapped phase with respect to spatial location or time directly. A continuous phase map must be prepared using phase unwrapping [30] prior to differentiating steps. Unfortunately, rapid and accurate two-dimensional (2-D) phase unwrapping is still a challenge. In addition, the phase calculated from the complex-valued coefficients which are the bandpass filtering outputs of the natural image is contaminated by the colored noise. It is difficult to cancel the colored phase noise to extract a high signal-to-noise ratio (SNR) vibration signal. Furthermore, phase singularities also affect the robustness of the estimation of displacement from the local phase. Jepson *et al.* [22], Fleet and Jepson [23] pointed out that phase will be very unstable due to the occurrence of phase singularities in particular regions, which will result in ambiguities and artifacts in the extracted signals.

To address these drawbacks, the AOPFS algorithm is proposed to estimate the vibration signal from the video by utilizing the real and imaginary parts of the complex-valued coefficients instead of directly processing the phase. The flowchart of the phase-based optical flow method and the proposed method are shown in Fig. 1. The details of the proposed iterative algorithm are presented below.

III. VIBRATION MEASUREMENTS BY THE AOPFS ALGORITHM

A. Principle of the AOPFS Algorithm for Displacement Extraction

In order to estimate the time-varying displacement $\delta(t)$, the AOPFS algorithm is introduced here. According to the phase-based motion representation, the complex exponential signal

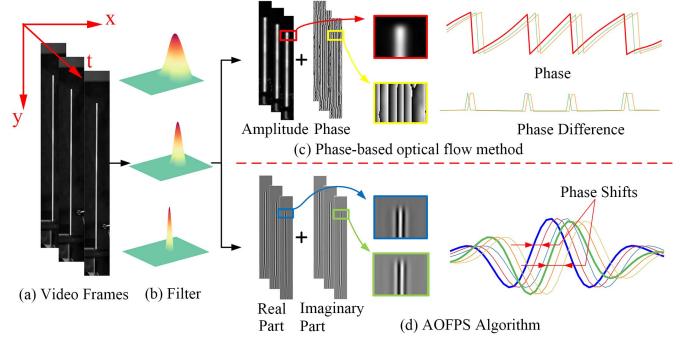


Fig. 1. Flowchart of the phase-based optical flow method and the proposed AOPFS algorithm. (a) Input video frames. (b) Bandpass filter from the complex steerable pyramids. (c) Phase-based optical method: the phase is used to derive the phase difference which corresponds to the motion, the spikes occur in phase difference because of phase wrapping. (d) Proposed algorithm: the real and imaginary parts are both utilized to find the transformed sinusoids, the phase shifts correspond to the motion.

contained the displacement $\delta(t)$ in (2) can be described as the sum of its real and imaginary parts as

$$S(x, t) = R_s(x, t) + i I_s(x, t). \quad (9)$$

With

$$\begin{aligned}R_s(x, t) &= A(x) \cos[\omega x + \omega \delta(t)] \\ I_s(x, t) &= A(x) \sin[\omega x + \omega \delta(t)]\end{aligned} \quad (10)$$

where $R_s(x, t)$ and $I_s(x, t)$ are the real and imaginary parts of $S(x, t)$, respectively. $A(x)$ is the local amplitude that varies slowly among the pixels. ω is the angular spatial frequency and can be considered as a constant within a local region. ωx is the local spatial phase distribution and $\omega \delta(t)$ is the local temporal phase shifts. If we assume that $\omega \delta(t)$ is known, (10) can be expanded as

$$\begin{aligned}R_s(x, t) &= A(x) \cos(\omega x) \cos[\omega \delta(t)] \\ &\quad - A(x) \sin(\omega x) \sin[\omega \delta(t)] \\ I_s(x, t) &= A(x) \sin(\omega x) \cos[\omega \delta(t)] \\ &\quad + A(x) \cos(\omega x) \sin[\omega \delta(t)].\end{aligned} \quad (11)$$

Let

$$\begin{aligned}\alpha(x) &= A(x) \cos(\omega x) \\ \beta(x) &= A(x) \sin(\omega x).\end{aligned} \quad (12)$$

Then (11) can be rewritten as

$$\begin{aligned}R_s(x, t) &= \alpha(x) \cos[\omega \delta(t)] - \beta(x) \sin[\omega \delta(t)] \\ I_s(x, t) &= \beta(x) \cos[\omega \delta(t)] + \alpha(x) \sin[\omega \delta(t)].\end{aligned} \quad (13)$$

Observing (13), the temporal phase shifts $\omega \delta(t)$ are assumed to be known, $R_s(x, t)$ and $I_s(x, t)$ are the real and imaginary parts of the complex sinusoid $S(x, t)$ which is obtained from the bandpass filtering of each frame. There are two equations and only two unknown variables $\alpha(x)$ and $\beta(x)$ in (13), so (13) has unique solutions.

Once $\alpha(x)$ and $\beta(x)$ are determined, the spatial phase distributions can be obtained as

$$\omega x = \tan^{-1} \left[\frac{\beta(x)}{\alpha(x)} \right]. \quad (14)$$

Differentiating with respect to the spatial location x yields

$$\omega = \frac{d(\omega x)}{dx} \Big|_{x=x_0} = \frac{\alpha(x)\beta'(x) - \alpha'(x)\beta(x)}{\alpha^2(x) + \beta^2(x)} \Big|_{x=x_0} \quad (15)$$

where $\alpha'(x)$ and $\beta'(x)$ are the first-order derivatives of $\alpha(x)$ and $\beta(x)$, respectively. ω is the local frequency at position x_0 . It should be noted that though the spatial phase distributions ωx calculated using arctangent function in (14) are wrapped in the interval $[-\pi, \pi]$, the local spatial frequency ω can be free of the phase wrapping problem by using (15).

Once the angular spatial frequency ω is determined, similar to (13), we can rearrange (11) as

$$\begin{aligned} R_s(x, t) &= a(t) \cos(\omega x) - b(t) \sin(\omega x) \\ I_s(x, t) &= a(t) \sin(\omega x) + b(t) \cos(\omega x). \end{aligned} \quad (16)$$

With

$$\begin{aligned} a(t) &= A(x) \cos[\omega\delta(t)] \\ b(t) &= A(x) \sin[\omega\delta(t)] \end{aligned} \quad (17)$$

where $R_s(x, t)$ and $I_s(x, t)$ are the same as that in (13). Note that ω has been estimated from (15), the spatial phase distributions ωx can be established with the spatial coordinate straightforwardly. As a consequence, unknown variables $a(t)$ and $b(t)$ can be obtained from (16).

Once $a(t)$ and $b(t)$ are determined, the temporal phase shifts can be obtained as

$$\omega\delta(t) = \tan^{-1} \left[\frac{b(t)}{a(t)} \right]. \quad (18)$$

It should be noted that the phase shifts $\omega\delta(t)$ calculated from the arctangent function are also wrapped within interval $[-\pi, \pi]$. Here the MATLAB built-in function `unwrap()` is used to remove the discontinuities of the phase shifts. Then the estimated displacements can be calculated via unwrapped phase shifts as

$$\delta(t) = \frac{\text{unwrap}(\omega\delta(t))}{\omega} \quad (19)$$

where the angular spatial frequency ω has been figured out in (15). Now resubstituting the estimated temporal phase shifts $\omega\delta(t)$ into (13), the new displacement estimation $\delta(t)$ can be acquired according to (13)–(19). This procedure is repeated until $\delta(t)$ satisfies the pre-defined accuracy conditions.

B. Implementation of the AOPFS Algorithm

The above derivation is only based on a single-pixel location x and a fixed time instant t . The goal of the vibration measurement is to determine all the $\delta(t)$ in discrete time $t = t_0, t_1, \dots, t_{N-1}$, expressed as a vector δ , where N is the number of frames of the recorded video. Though the discrete displacement $\delta(x, t)$ can be determined uniquely at each pixel x and each time instant t according to (13) and (16), the estimation of displacement information from a single pixel is inaccurate and less-robust due to the existence of noise. Besides, the estimation of the local frequency ω based on the single-pixel information is impossible in the discrete space, as shown in (15). In order to overcome the above

limitations, the iterative algorithm over the entire time history is implemented to improve the accuracy and robustness of vibration displacement measurement by introducing a pixels region as a region of interest (ROI). Generally, an ROI consists of M pixels which are distributed symmetrically with respect to the location x_0 . For the accuracy and convenience of calculation, it is suggested that the size of the ROI M should be greater than 2 and M takes an odd number. The indices of the ROI are defined as $L \in [-(M-1)/2, -(M-1)/2 + 1, \dots, -1, 0, 1, \dots, (M-1)/2]$. So the local frequency ω in location x_0 is equal to $\omega(0)$ estimated within the ROI. Under the assumption of intensity consistency [20], [31], the local amplitude $A(x)$ and local frequency ω in the same location remain unchanged between different frames and the temporal variation among each frame is due solely to the displacement $\delta(t)$.

Next, the temporal variation measurement in the location x_0 over the entire time history $t = t_0, t_1, \dots, t_{N-1}$ is considered. First, a new spatiotemporal matrix \mathbf{S} is built whose columns composed of all the ROIs centered in the measurement location x_0 among N frames filtered by a pre-defined bandpass filter, see (3). It yields

$$\mathbf{S} = \begin{bmatrix} S_{-\frac{M-1}{2}, t_0} & S_{-\frac{M-1}{2}, t_1} & \cdots & S_{-\frac{M-1}{2}, t_{N-1}} \\ S_{-\frac{M-1}{2}+1, t_0} & S_{-\frac{M-1}{2}+1, t_1} & \cdots & S_{-\frac{M-1}{2}+1, t_{N-1}} \\ \vdots & \vdots & \ddots & \vdots \\ S_{0, t_0} & S_{0, t_1} & \cdots & S_{0, t_{N-1}} \\ \vdots & \vdots & \ddots & \vdots \\ S_{\frac{M-1}{2}, t_0} & S_{\frac{M-1}{2}, t_1} & \cdots & S_{\frac{M-1}{2}, t_{N-1}} \end{bmatrix}. \quad (20)$$

Note that each element of \mathbf{S} is a discrete sample of the complex exponential signal $S(x, t)$ and the size of \mathbf{S} is M -by- N . According to (13), at each point, it can take two equations, there are a total of $2MN$ equations from MN samples. For convenience, these equations can be easily described in the form of the matrix-vector multiplication as

$$\mathbf{b} = \mathbf{Ax} \quad (21)$$

where

$$\mathbf{x} = \left[\alpha_{-\frac{M-1}{2}}, \beta_{-\frac{M-1}{2}}, \dots, \alpha_0, \beta_0, \dots, \alpha_{\frac{M-1}{2}}, \beta_{\frac{M-1}{2}} \right]^T. \quad (22)$$

Note that $\alpha_m = \alpha(m)$ and $\beta_m = \beta(m)$, which are defined in (12). \mathbf{A} is a block diagonal matrix and defined as

$$\mathbf{A} = \text{kron}(\mathbf{I}, \mathbf{A}_0) \quad (23)$$

where \mathbf{I} is an $M \times M$ identity matrix, $\text{kron}(\cdot)$ returns the Kronecker product of matrices \mathbf{A}_0 and \mathbf{I} . \mathbf{A}_0 is a matrix of size $(2 \times N) \times 2$ and can be expressed as

$$\mathbf{A}_0 = \begin{bmatrix} \cos[\omega\delta(t_0)] & -\sin[\omega\delta(t_0)] \\ \vdots & \vdots \\ \cos[\omega\delta(t_{N-1})] & -\sin[\omega\delta(t_{N-1})] \\ \sin[\omega\delta(t_0)] & \cos[\omega\delta(t_0)] \\ \vdots & \vdots \\ \sin[\omega\delta(t_{N-1})] & \cos[\omega\delta(t_{N-1})] \end{bmatrix} \quad (24)$$

where $\omega\delta(t)$ are the temporal phase shifts which should be initialized before the iteration. The initialization of the phase shifts will be discussed at the end of this section. The column vector \mathbf{b} is stacked by the real and imaginary parts of the complex exponential samples of $S(x, t)$, where $x = -(M-1)/2, \dots, (M-1)/2$, $t = t_0, \dots, t_{N-1}$, which are the elements of matrix \mathbf{S} and it can be given by

$$\mathbf{b} = \left[\mathbf{u}_{-\frac{M-1}{2}}, \mathbf{u}_{-\frac{M-1}{2}+1}, \dots, \mathbf{u}_0, \dots, \mathbf{u}_{\frac{M-1}{2}} \right]^T. \quad (25)$$

With

$$\mathbf{u}_m = [R_s(m, t_0), \dots, R_s(m, t_{N-1}), I_s(m, t_0), \dots, I_s(m, t_{N-1})]^T \quad (26)$$

where \mathbf{u}_m is a row vector, $R_s(m, t)$ and $I_s(m, t)$ are the real and imaginary parts of $S(m, t)$.

It is obvious that (21) can be easily solved by applying the least squares method if vector \mathbf{b} is free of noise. However, considering that the complex exponential signal $S(x, t)$ obtained by bandpass filtering of the real image inevitably contains noise, then the vector \mathbf{b} consisting of discrete samples of $S(x, t)$ is also not noise-free, which leads to an inaccurate and non-robust estimation of \mathbf{x} by using $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ directly. To overcome this limitation, an objective function is modeled in the least squares sense with a regularization term as

$$f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{Dx}\|_2^2 \quad (27)$$

where the last term imposes smoothness constraints on the phase distribution signals $\alpha(x)$ and $\beta(x)$. The regularization parameter λ allows for tuning the degree of smoothness. \mathbf{D} is a first-order difference matrix of size $(M-2) \times M$

$$\mathbf{D} = \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix}. \quad (28)$$

By setting the gradient of $f(\mathbf{x})$ to zero to minimize the objective function in (27), the phase distribution can be recovered as

$$\mathbf{x} = (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{D}^T \mathbf{D})^{-1} \mathbf{A}^T \mathbf{b}. \quad (29)$$

According to (22), vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are composed of the odd and even elements of \mathbf{x} , respectively, that are

$$\begin{aligned} \boldsymbol{\alpha} &= [\mathbf{x}(1), \mathbf{x}(3), \dots, \mathbf{x}(M)]^T \\ \boldsymbol{\beta} &= [\mathbf{x}(2), \mathbf{x}(4), \dots, \mathbf{x}(M-1)]^T \end{aligned} \quad (30)$$

where $M \in \mathbb{Z}^+$ is an odd number. Then the local frequency ω can be obtained by (15).

When ω is known, (16) can be rewritten in the form of matrix multiplication similar to (21) as

$$\mathbf{G} = \mathbf{HQ} \quad (31)$$

where

$$\mathbf{G} = \begin{bmatrix} R_{-\frac{M-1}{2}, t_0}^s & R_{-\frac{M-1}{2}, t_1}^s & \cdots & R_{-\frac{M-1}{2}, t_{N-1}}^s \\ R_{-\frac{M-1}{2}+1, t_0}^s & R_{-\frac{M-1}{2}+1, t_1}^s & \cdots & R_{-\frac{M-1}{2}+1, t_{N-1}}^s \\ \vdots & \vdots & & \vdots \\ R_{\frac{M-1}{2}, t_0}^s & R_{\frac{M-1}{2}, t_1}^s & \cdots & R_{\frac{M-1}{2}, t_{N-1}}^s \\ I_{-\frac{M-1}{2}, t_0}^s & I_{-\frac{M-1}{2}, t_1}^s & \cdots & I_{-\frac{M-1}{2}, t_{N-1}}^s \\ I_{-\frac{M-1}{2}+1, t_0}^s & I_{-\frac{M-1}{2}+1, t_1}^s & \cdots & I_{-\frac{M-1}{2}+1, t_{N-1}}^s \\ \vdots & \vdots & & \vdots \\ I_{\frac{M-1}{2}, t_0}^s & I_{\frac{M-1}{2}, t_1}^s & \cdots & I_{\frac{M-1}{2}, t_{N-1}}^s \end{bmatrix}. \quad (32)$$

Note that the matrix \mathbf{G} is created by rearranging the real and imaginary parts of the spatiotemporal matrix \mathbf{S} . The matrix \mathbf{H} , consists of matrix blocks \mathbf{h}_0 and \mathbf{h}_1 , and can be expressed as

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_0 & -\mathbf{h}_1 \\ \mathbf{h}_1 & \mathbf{h}_0 \end{bmatrix}. \quad (33)$$

With

$$\mathbf{h}_0 = \left[\cos\left(\omega\left(-\frac{M-1}{2}\right)\right) \cdots \cos\left(\omega\left(\frac{M-1}{2}\right)\right) \right]^T \quad (34)$$

$$\mathbf{h}_1 = \left[\sin\left(\omega\left(-\frac{M-1}{2}\right)\right) \cdots \sin\left(\omega\left(\frac{M-1}{2}\right)\right) \right]^T. \quad (35)$$

The matrix \mathbf{Q} with the size of $2 \times N$ is defined as

$$\mathbf{Q} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}. \quad (36)$$

With

$$\begin{aligned} \mathbf{a} &= [a(t_0) \ a(t_1) \ \cdots \ a(t_{N-1})] \\ \mathbf{b} &= [b(t_0) \ b(t_1) \ \cdots \ b(t_{N-1})] \end{aligned} \quad (37)$$

where \mathbf{a} and \mathbf{b} are the first and second rows of matrix \mathbf{Q} , respectively. According to (31), \mathbf{Q} can be calculated with the least squares method as

$$\mathbf{Q} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{G}. \quad (38)$$

Once \mathbf{Q} is determined, the vectors \mathbf{a} and \mathbf{b} can be obtained according to (37). Then the measured displacements $\boldsymbol{\delta}$ can be calculated using (18) and (19). The algorithm will iteratively update until the stopping criteria are achieved. The above algorithmic framework is summarized in Algorithm 1.

The proposed algorithm is insensitive to the initial value of ω_0 and $\boldsymbol{\delta}_0$, because the sinusoid transformed from the bandpass filter usually has only a unique spatial frequency component. For this reason, any random value can be used to initialize ω_0 . Different initial values only affect the convergence speed, but this effect is almost negligible according to our simulation and experiments. For some special cases, an alternative initialization method can be adopted as shown in (5). In our simulation and experiments, the initial value of ω_0 was all set to 1, and $\boldsymbol{\delta}_0$ is initialized to a vector of ones. The stopping criteria ε is suggested to $1 \times 10^{-7} - 1 \times 10^{-10}$ and the regularization parameter λ is set to 10–1000 based on our experience. It should be noted that a formal proof about the convergence of our algorithm is not given and the counter k is introduced to prevent divergence with its maximum value set to 100.

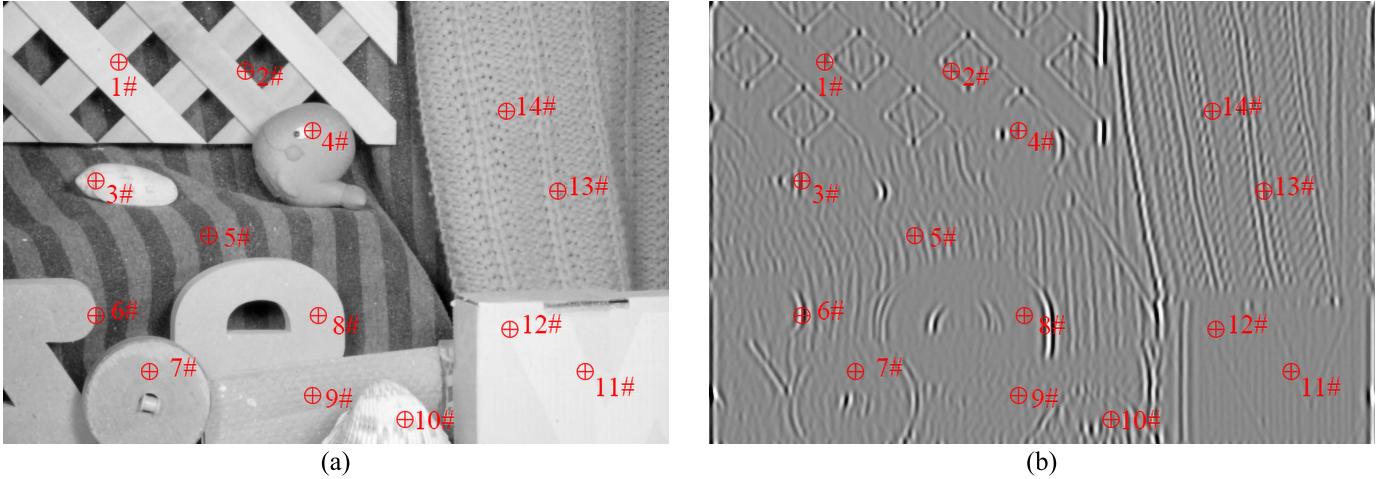


Fig. 2. (a) Original image used to generate the simulated video and locations of 14 measuring points. (b) Filtered response of the original image by the steerable pyramid filter at horizontal orientation and scale 3.

Algorithm 1 AOPPS Algorithm for Displacement Extraction

```

1: Input spatiotemporal matrix  $\mathbf{S}$ ; parameter  $\lambda > 0$ ;
stopping threshold  $\varepsilon$ 
2: Initialize  $\omega_0$ ,  $\delta_0$ , and create matrix  $\mathbf{A}_0$ . Create vector  $\mathbf{b}$ 
and matrix  $\mathbf{G}$  with  $\mathbf{S}$ .
3: Set  $k = 1$ 
4: While  $\|\delta_k - \delta_{k-1}\|_2^2 > \varepsilon$  and  $k < 100$  do
5:  $\mathbf{A} = \text{kron}(\mathbf{I}, \mathbf{A}_0)$ 
6:  $\mathbf{x} = (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{D}^T \mathbf{D})^{-1} \mathbf{A}^T \mathbf{b}$ 
7: Take the odd and even index elements of  $\mathbf{x}$  respectively
to generate vectors  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$ 
8:  $\omega_k = \frac{\alpha(x)\beta'(x) - \alpha'(x)\beta(x)}{\alpha^2(x) + \beta^2(x)}|_{x=x_0}$ 
9: Create vectors  $\mathbf{h}_0$  and  $\mathbf{h}_1$  based on (34)-(35), update
matrix  $\mathbf{H}$  based on (33)
10:  $\mathbf{Q} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{G}$ 
11: Update  $\mathbf{a}$  and  $\mathbf{b}$  with  $\mathbf{Q}$  based on (36)
12:  $\delta_k = [\frac{1}{\omega_k} \tan^{-1}(\frac{\mathbf{b}}{\mathbf{a}})]^T$ 
13: Update matrix  $\mathbf{A}_0$  with  $\omega_k$  and  $\delta_k$  based on (24)
14:  $k = k + 1$ 
15: end while
16: Output displacement signal  $\delta_k$ .
```

IV. SIMULATION AND EXPERIMENT

In order to investigate the performance of our proposed algorithm, we evaluate it on synthetic and experimental videos. It should be noted that all the motions in these videos are in-plane and along a single axis, and only a single camera is used to record the experimental videos. All the videos are processed in YIQ color space, and only the Y channel is used in our simulation and experiments. The filter used to transform the video is from the complex steerable pyramid [28]. The complex steerable pyramid with octave bandwidth has a total of n ($n \in \mathbb{Z}^+$) scales and n is determined by the size of the video image. From the lowest to the highest frequency components, the filters are marked as scales 1– n . Except for the lowest and highest frequency parts, each scale has four orientations (0° , 45° , 90° , and 135°), where 0° and 90°

TABLE I
PARAMETERS OF THE VIDEOS USED IN SIMULATION AND EXPERIMENT

Video	Size(Pixels)	FPS	Frames	Test Points	Time (Seconds)
RubberWhale	584×388	500	300	14	37
Cantilever beam	1024×128	10000	50000	706	2997
RK4 Rotor	1664×380	1000	10000	120	1407

represent the horizontal and vertical orientations respectively. In our experiments, only a single filter at a fixed scale and orientation was used to process the video, which will be detailed in the experiments. The parameters of the synthetic and experimental videos are given in Table I. All the process was done in a computer with CPU i5-4460 and 16 GB of RAM. All video frames will be processed by the preferred filter, but only the motion at some selected test points will be extracted. The number of frames and test points and the execution time are also given in Table I. Because our method involves transforming and processing every video frame, the implementation efficiency of the algorithm depends on the volume of the video. So using GPU may significantly reduce the computation time. We also provide the example videos in the supplementary material.

A. Simulation Test

To investigate the performance and effectiveness of the proposed algorithm, a real-life image is first carried out to generate the simulated video. Fig. 2(a) shows the original image which consists of a variety of textures and patterns. To simulate the motion, we use cubic-optimal maximal-order-minimal-support functions (OMOMS) interpolation [32], [33] to generate a series of video frames which are the shifted copies of the original one. To evaluate the performance of the proposed algorithm at different locations, we defined a global motion field, which means that the velocity at each pixel on the image is consistent at any moment. For simplicity, the global

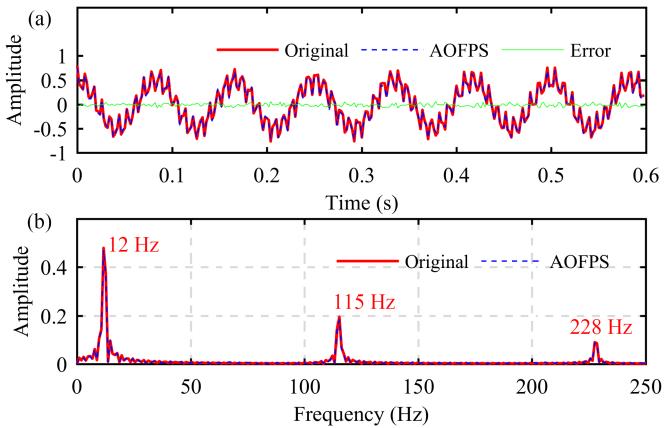


Fig. 3. (a) Comparing the original signal and the estimated result using the AOFPS algorithm at location 3. (b) FFT spectrum analysis of the estimated result and the original one.

motion field is defined as in-plane motion along the horizontal axis as follows:

$$\delta(t) = 0.5 \cos(2\pi f_1 t) + 0.2 \cos(2\pi f_2 t) + 0.1 \cos(2\pi f_3 t) \quad (39)$$

where $f_1 = 12$ Hz, $f_2 = 115$ Hz, and $f_3 = 228$ Hz. Note that the maximum displacement of $\delta(t)$ is less than 1 pixel, which guarantees the accuracy of the interpolation method with cubic-OMOMS. On the other hand, the simulated video can be used to verify the subpixel accuracy of the proposed algorithm. For this video, the complex steerable pyramid has six scales, and the filter to process the video is at scale 3 and horizontal orientation. The simulated video can also be found in the supplementary materials.

First, we use our algorithm to extract the simulated motion at location 3 marked by a symbol as shown in Fig. 2(a). The filtered response of the original image by the steerable pyramid filter is shown in Fig. 2(b). The estimated time history by the proposed algorithm and the original one are shown in Fig. 3(a). It is obvious that the estimated signal with our algorithm matches well with the ground truth in the time domain, and a very small error is induced. The Fourier spectrum in Fig. 3(b) also shows that our algorithm can retrieve all the three frequency components accurately from the simulated video. In order to further show the advantages of our method, we selected a total of 14 measuring locations from the simulated video to compare the accuracy of the proposed algorithm with the phase-based method. The 14 measuring locations are shown in Fig. 2(a). Note that at each location, the image has different structures, textures, and edges. The RMSEs calculated with our algorithm and the phase-based method at all the measuring locations are shown in Fig. 4. At locations 2, 5, and 8, the phase-based method and our algorithm obtained a result with similar accuracy. At the other locations, our algorithm outperforms the phase-based method in accuracy. Fig. 5(a)–(f) shows the comparison between the original signal with the motion signals extracted with our AOFPS algorithm and the phase-based method at locations 1, 2, 5, 6, 10, and 13, respectively. As shown in Fig. 5(b) and (c), our algorithm and the phase-based method both acquired an accurate estimation, but in Fig. 5(a) and (d)–(f), the estimation accuracy of the

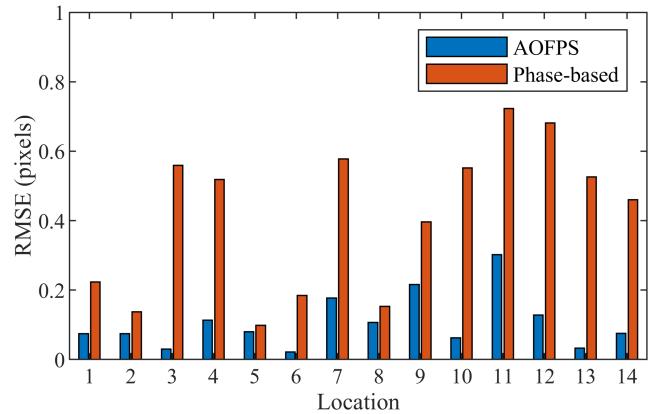


Fig. 4. RMSEs of the estimated results calculated with our AOFPS method and the phase-based method at all the 14 measuring locations.

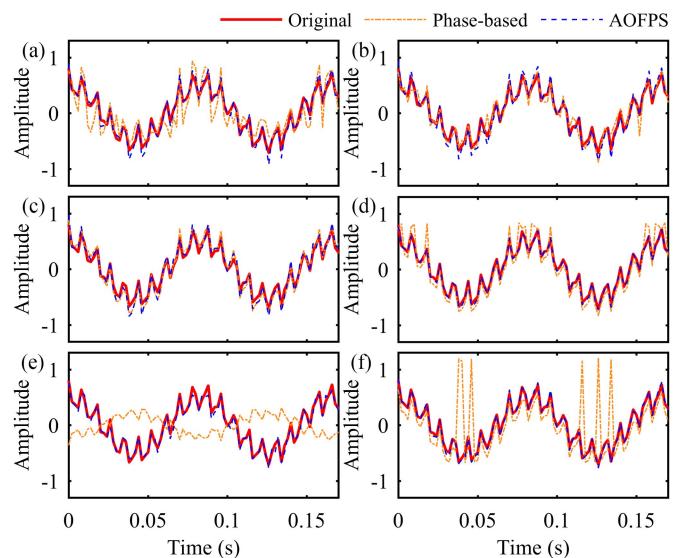


Fig. 5. Comparing the original signal with the motion signals extracted with our AOFPS algorithm and the phase-based method at locations (a) 1, (b) 2, (c) 5, (d) 6, (e) 10, and (f) 13. The abnormal peaks shown in (d) and (f) are induced by the phase wrapping problem.

phase-based method deteriorates rapidly because of the phase singularities and phase wrapping issues. As a comparison, our proposed AOFPS algorithm can still maintain the robustness and accuracy of the estimation at those locations.

Besides, we investigated the performance of our algorithm on videos with different levels of texture. The simulated video was blurred with a 2-D Gaussian filter. The standard deviation σ of the Gaussian filter is used to tune the smoothness which represents the level of texture. Namely the original video is the texture-full video, and as the value of σ increases, the video becomes smoother and the texture becomes less. In this simulation, the value of σ changes from 5 to 30 and increasing by five each time, producing six sets of videos with different levels of texture. Fig. 6 shows the blurred images filtered by Gaussian filters with different standard deviation σ , and $\sigma = 0$ represents the original video with full textures. Our algorithm was used to extract the motion signal from the varied textures videos. Fig. 7 shows the RMSEs between the results estimated

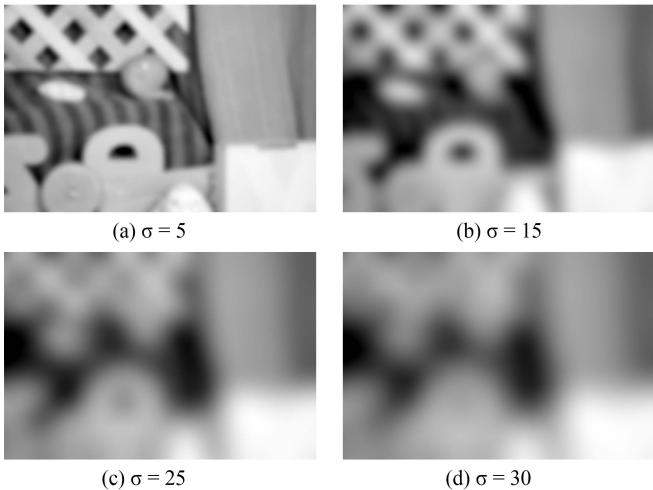


Fig. 6. Images blurred by Gaussian filters with different deviations. (a) $\sigma = 5$. (b) $\sigma = 15$. (c) $\sigma = 25$. (d) $\sigma = 30$. The level of textures is inversely proportional to the degree of smoothness.

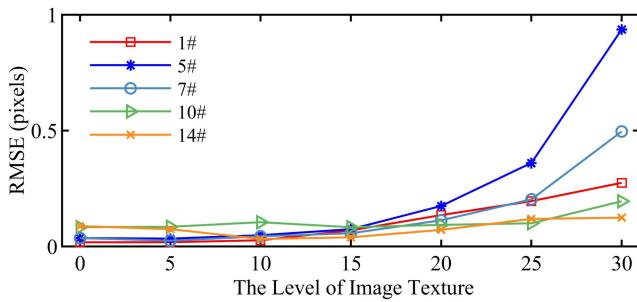


Fig. 7. RMSEs between the ground truth and the signals extracted from the varying levels of texture videos with our AOFPS algorithm at locations 1, 5, 7, 10, and 14.

with our method and the ground truth at five different locations versus the level of texture. It can be seen that our algorithm performs similarly at each measurement location when σ is less than 15. When σ exceeds 15, the performance of our algorithm drops dramatically because of the lack of valid textures. It is obvious that our method can achieve more accurate results on videos with more textures. For the videos with fewer textures, the accuracy of the estimation with our method decreases because there are fewer valid sinusoids been acquired by the complex steerable pyramid bandpass filter. Therefore, it is necessary to guarantee that there are enough textures or edges on the image to estimate the motion with our algorithm in a real application.

B. Cantilever Beam Test

In this subsection, a cantilever beam test is carried out in a laboratory environment to validate the performance of the proposed AOFPS algorithm which is applied to process the image sequences for the measurement of displacements. The experimental setup for the vibration measurement of the cantilever beam is shown in Fig. 8(a). An aluminum cantilever beam with a size of 300 mm × 40 mm × 3 mm is employed

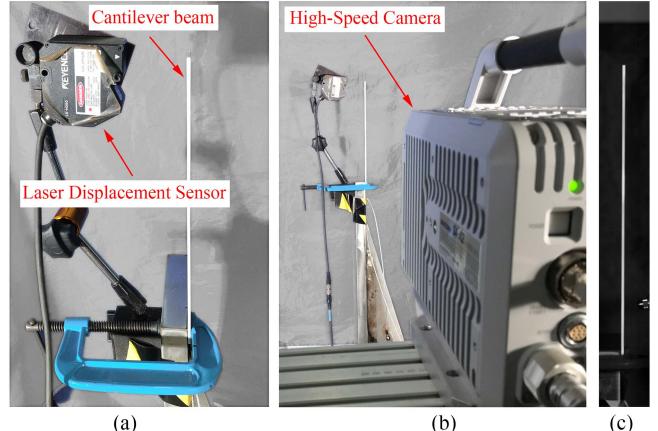


Fig. 8. Experimental setup for a cantilever beam measurement with a high-speed camera and a laser displacement sensor. (a) Laser displacement sensor. (b) High-speed camera. (c) Recorded experimental image.

as the vibrating target, and its one end is fixed on the base and the other end is free. An impact hammer is utilized to excite the beam near the base, and a high-speed camera (FASTCAM SA-X2, Photron) and a lens (ZEISS Planar F1.4/50 mm) are used to record the subsequent vibration. The resolution of the camera is 1024×128 pixels, and the frame rate is set to 10000 FPS (frames per second). In addition, a laser displacement sensor (LK-H050, Keyence) is regarded as the reference for the displacements measurement. This sensor is with the ability to achieve ultrahigh-speed and high-accuracy displacement measurements. Its measuring range and reference distance are ± 10 and 50 mm, respectively. In this experiment, the sampling frequency is set to 20.48 kHz. It should be noted that the measuring point of the laser displacement sensor is located near the free end of the cantilever beam, which can avoid the node positions of structural vibration as much as possible. For comparison, the laser displacement signal is down-sampled to 10 kHz, which is consistent with the frame rate of the high-speed camera.

The displacements derived from the high-speed camera and measured by the laser displacement sensor near the same measuring location are shown in Fig. 9(a). Enlarged view of time history from 0.15 to 0.25 s and 1.5 to 1.9 s are shown in Fig. 9(b) and (c), respectively. The displacement derived with our algorithm from the high-speed camera closely matches the one measured by the laser displacement sensor within the whole 3 s of data. Fig. 10 shows the frequency domain comparison between the displacements measured by the high-speed camera and the laser displacement sensor. Table II shows the frequencies calculated by the Abaqus software and measured by the camera and the laser displacement sensor. The laser displacement sensor data shows five resonant frequencies from 0 to 2000 Hz. The camera spectrum not only shows these five resonant frequencies but also contains two more resonance peaks at 426.9 and 1295.6 Hz, respectively, caused by the torsional vibration of the cantilever beam. Due to the position of the measuring point, the laser displacement sensor fails to measure these two torsion frequencies. This verification shows

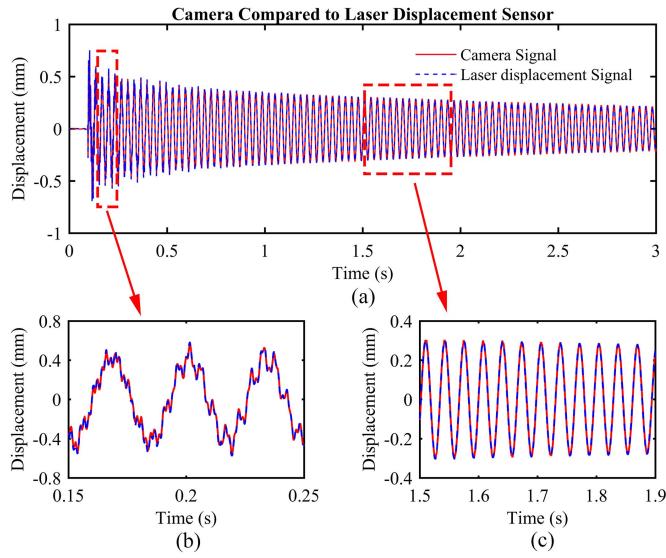


Fig. 9. (a) Comparison between displacements derived from the camera and the laser displacement sensor. (b) Enlarged view of time history marked by the left red dotted rectangle of (a). (c) Enlarged view of time history marked by the right red dotted rectangle of (a).

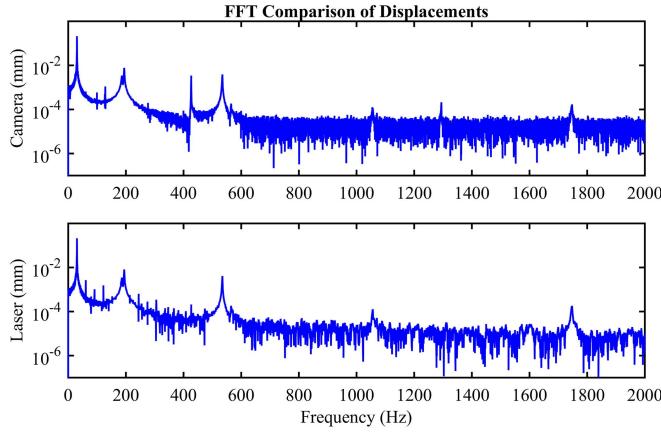


Fig. 10. Frequency domain comparison between displacements measured by the high-speed camera and the laser displacement sensor.

that our algorithm for vibration measurement using a camera is accurate and reliable.

Another advantage of the vision-based method to measure vibration is the ability to obtain the full-field motion of the structure, which makes it easier to acquire the modal shape of the structure compared with the traditional measurement methods that require multiple sensors to be placed on the surface of the structure. We obtained the vibration responses from a total of 118 measuring locations on the cantilever beam using our algorithm and then the least-squares complex exponential algorithm [34] is used to calculate the operational modal shapes of the cantilever beam as shown in Fig. 11(a). As a comparison, the motion magnification method [15] is also used to obtain the modal shapes of the first five bending vibrations of the cantilever beam as shown in Fig. 11(b). The modal shapes extracted with these two methods match well with each other. Although the natural frequency of torsional vibration can

TABLE II
COMPARISON OF THE MODAL FREQUENCIES CALCULATED BY THE ABAQUS SOFTWARE AND MEASURED BY THE CAMERA AND THE LASER DISPLACEMENT SENSOR

Mode	Frequency (Hz)		
	Abaqus	Camera	Laser
1st bending	30.9	30.5	30.5
2nd bending	193.7	194.5	194.5
1st torsion	413.3	426.9	—
3rd bending	542.4	535.1	535.1
4rd bending	1063.7	1056.8	1056.2
2nd torsion	1252.9	1295.6	—
5th bending	1759.8	1747.5	1748.1

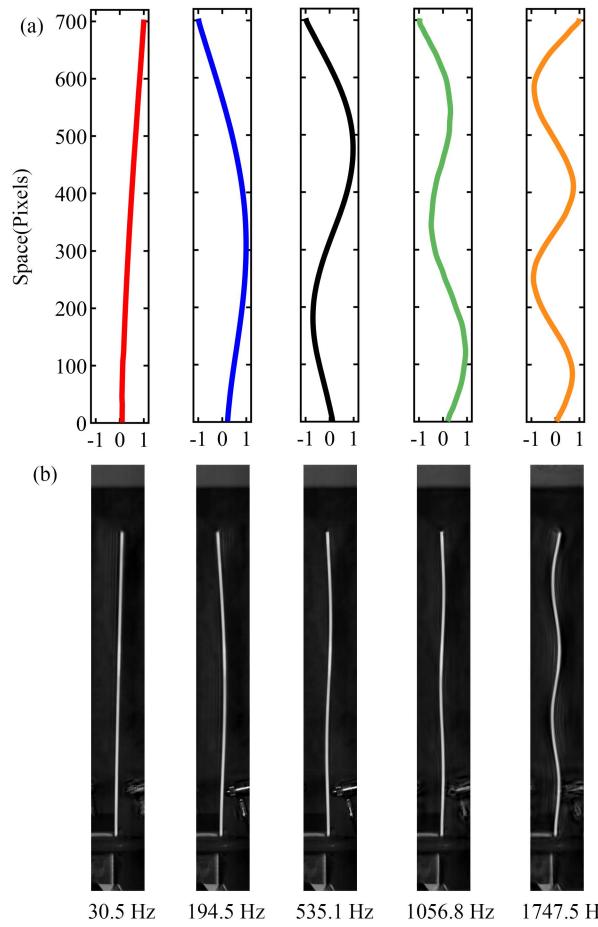


Fig. 11. First five bending vibration modal shapes of the cantilever beam extracted with (a) AOFPS algorithm and (b) motion magnification method.

be extracted with our method, it is difficult to obtain the modal shape of torsional vibration because only a single camera is used in this experiment. The acquisition of torsional vibration modal shapes will be carried out in future research.

C. Rotational Machinery Test

In this subsection, we show the application of the vision-based AOFPS algorithm in the vibration test of the rotor



Fig. 12. Rotational machinery test experiment setup. (a) Experiment scene. (b) RK4 Rotor Kit. (c) Screenshot of the rotor kit from the captured video filmed by the high-speed camera.

TABLE III
 QUANTITATIVE COMPARISON OF THE DISPLACEMENT SIGNALS MEASURED BY THE HIGH-SPEED CAMERA AND THE EDDY CURRENT SENSOR BOTH IN THE TIME DOMAIN AND THE FREQUENCY DOMAIN

	1400 rpm		1800 rpm		2200 rpm	
	L	R	L	R	L	R
Correlation	99.18%	96.79%	97.25%	99.00%	98.16%	98.58%
RMSE (mm)	0.0214	0.0081	0.0611	0.0165	0.0288	0.0218
AOFPS	Frequency (Hz)	23.32	23.32	29.91	29.91	36.74
	Amplitude (mm)	0.1324	0.0360	0.2721	0.1399	0.1479
Eddy Current Sensor	Frequency (Hz)	23.19	23.19	29.91	29.91	36.62
	Amplitude (mm)	0.1282	0.0365	0.2647	0.1405	0.1501
Approximation error		3.28%	1.37%	2.80%	0.43%	1.47%
						6.20%

system, and further investigate the performance of our algorithm in the vibration test by comparing it with the eddy current displacement sensor. Fig. 12 shows the experimental setup of the vibration test for the rotor system. The Bently Nevada RK4 Rotor Kit is used to simulate rotational machinery behavior, and a speed controller is used to adjust the speed of the rotor. Fig. 12(b) shows the configuration of the rotor kit, which consists of a motor, a rotor shaft, a balance wheel, two bearings, a coupling, and a mechanical base. Two eddy current sensors (specified by L and R , which represent the left and right eddy current sensor, respectively) are utilized to measure the vertical displacements of the rotor shaft as the reference. The sampling frequency of the eddy current sensor is set to 1000 Hz. A high-speed camera (5KF20, Revealer) and a lens (Tamron SP AF 17–50 mm F 2.8/24 mm) are used to record the video at the rotor operating condition, and a frame image captured by the high-speed camera is shown in Fig. 12(c). For this video, the complex steerable pyramid has six scales, and the filter to process the video is at scale 4 and vertical orientation. In order to be consistent with the sampling rate of the eddy current sensor, the frame rate of the high-speed camera is also set up to 1000 FPS. To ensure that the rotor kit has sufficient illumination in high-speed photography, two LED lamps are utilized to improve the lighting condition.

Fig. 13(a) and (c) show the time history comparison of the displacement signals measured by the high-speed camera and the eddy current sensor. It should be noted that the displacement signal extracted by our algorithm is measured in the vicinity of the corresponding eddy current sensor. Fig. 13(b) and (d) show the Fourier spectrum of the displacement signals. It can be seen that the displacement signals derived with our algorithm match well with the ones measured by the eddy current sensor both in the time domain and the frequency domain. Table III shows the quantitative comparison of the displacement signals measured by those two methods. In the time domain, the correlation between the displacement signals extracted with our algorithm and the eddy current sensor is above 96% at rotor speeds of 1400, 1800, and 2200 r/min. The RMSE between the displacement signals is also low enough. In the frequency domain, our algorithm acquired almost the same peak frequencies which correspond to the rotor speeds of 1400, 1800, and 2200 r/min as the eddy current sensor obtained. The approximation errors between the vibration amplitudes of the peak frequencies are also lower than 7%. This verification shows that our algorithm can be used to measure the vibration of the rotor system accurately and intuitively. The application of camera-based measurement methods for rotor vibration measurements will be further investigated in the future.

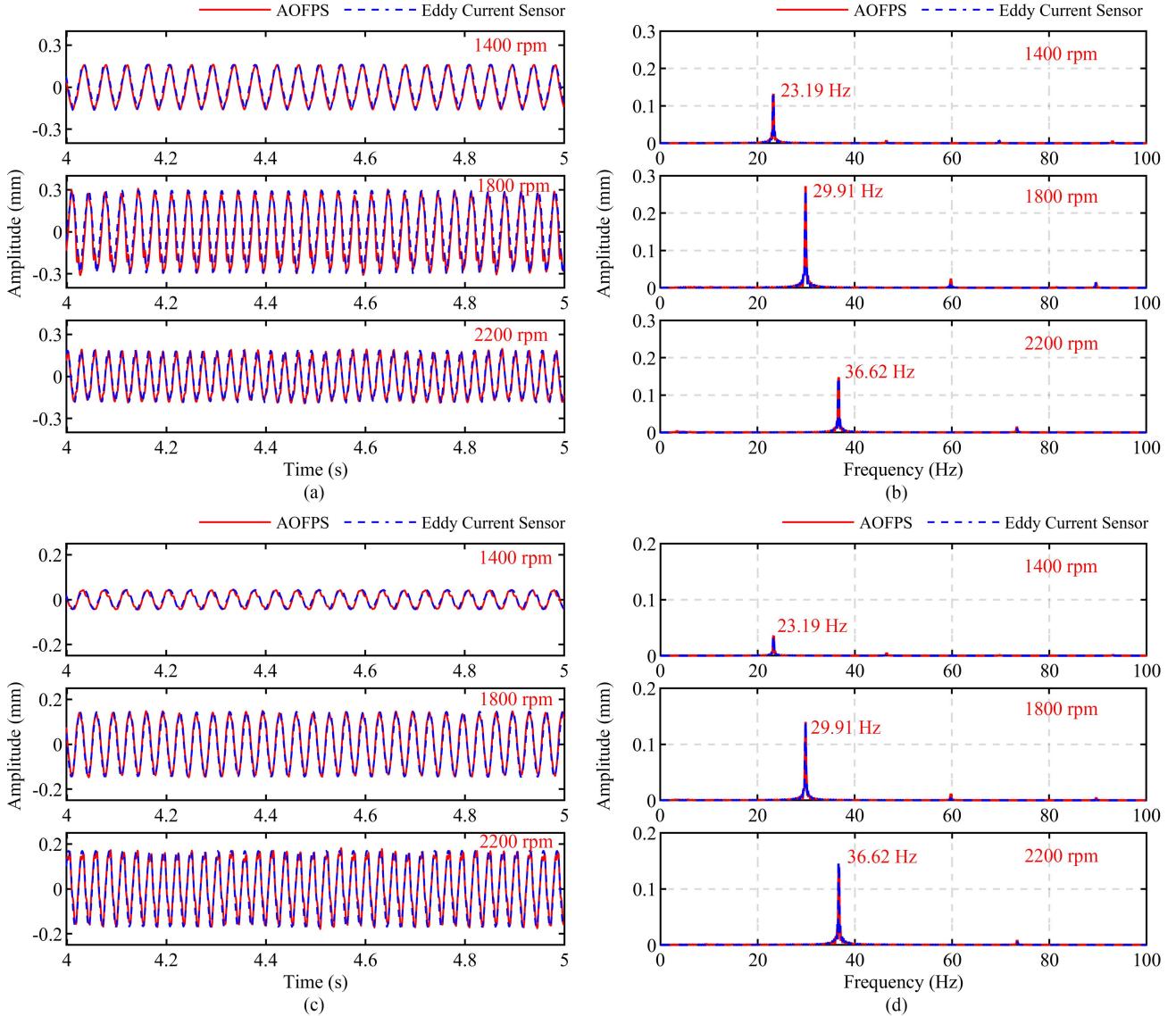


Fig. 13. Vibration test of the RK4 Rotor Kit using our AOFPS algorithm and the eddy current sensor at the rotor speed of 1400, 1800, and 2200 r/min, respectively. (a) Time history of the displacement signals measured by the camera and the left eddy current sensor, respectively. (b) Fourier spectrum of the displacement signals shown in (a). (c) Time history of the displacement signals measured by the camera and the right eddy current sensor, respectively. (d) Fourier spectrum of the displacement signals shown in (c).

In this section, we evaluated our AOFPS algorithm on synthetic videos and experimental videos respectively. Although our algorithm shows a good performance, it also has some limitations. Our algorithm has only been tested and validated in the case of in-plane motion along one direction and cannot be applied to the measurement of out-of-plane motion at this time. The measurement of out-of-plane motion requires 3-D stereo photography, for which multiple cameras must be used to capture the video, and this is the work we will do in the future. Besides, our algorithm can only measure small motions where the position of the object does not change significantly. If the object in the video has a large displacement, our algorithm does not work. We have not given a formal proof about the convergence of our algorithm although it can converge quickly according to our simulations and experiments.

V. CONCLUSION

In this study, the AOFPS algorithm for vibration measurement technique was developed to estimate the displacement of the vibrating objects which are from the recorded videos. The inherent characteristics such as textures and edges of the image were utilized to produce complex-valued fringe patterns whose phase shifts correspond to the motion of the structure among each frame. The angular spatial frequency and the temporal phase shifts of the transformed fringe patterns were alternatively calculated and optimized based on the Tikhonov regularization and least squares method. Then the displacements were derived with an iterative procedure. The proposed algorithm used the fringe patterns including both the real and imaginary parts of the complex-valued outputs of image transformation instead of the local amplitude and phase, to extract vibrations, thus it improved the robustness of estimation. A simulated

video with motion and a laboratory experiment of cantilever beam were used to demonstrate the effectiveness and accuracy of the AOPFS algorithm. Moreover, the applications of rotational machinery vibration measurement were also presented.

In future work, the extraction of operational deflection shapes and structural damage detection based on modal analysis may be considered with our vibration measurement technique.

REFERENCES

- [1] 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," *J. Sound Vib.*, vol. 345, pp. 58–71, Jun. 2015.
- [2] Y. Yang *et al.*, "Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification," *Mech. Syst. Signal Process.*, vol. 85, pp. 567–590, Feb. 2017.
- [3] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 35, nos. 1–2, pp. 108–126, Feb. 2013.
- [4] H. Li, M. Li, C. Li, F. Li, and G. Meng, "Multi-faults decoupling on turbo-expander using differential-based ensemble empirical mode decomposition," *Mech. Syst. Signal Process.*, vol. 93, pp. 267–280, Sep. 2017.
- [5] L. Ming, L. Fucai, J. Beibei, B. Huiyu, L. Hongguang, and M. Guang, "Multi-fault diagnosis of rotor system based on differential-based empirical mode decomposition," *J. Vib. Control*, vol. 21, no. 9, pp. 1821–1837, Jul. 2015.
- [6] D. Feng and M. Q. Feng, "Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection—A review," *Eng. Struct.*, vol. 156, pp. 105–117, Feb. 2018.
- [7] R. Yan and R. X. Gao, "Hilbert–Huang transform-based vibration signal analysis for machine health monitoring," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 6, pp. 2320–2329, Dec. 2006.
- [8] Y. Xiong, S. Chen, X. Dong, Z. Peng, and W. Zhang, "Accurate measurement in Doppler radar vital sign detection based on parameterized demodulation," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 11, pp. 4483–4492, Nov. 2017.
- [9] J. García-Martín, J. Gómez-Gil, and E. Vázquez-Sánchez, "Non-destructive techniques based on eddy current testing," *Sensors*, vol. 11, no. 3, pp. 2525–2565, Feb. 2011.
- [10] W. J. Staszewski, B. C. Lee, L. Mallet, and F. Scarpa, "Structural health monitoring using scanning laser vibrometry: I. Lamb wave sensing," *Smart Mater. Struct.*, vol. 13, no. 2, pp. 251–260, Apr. 2004.
- [11] Y. Xiong, Z. Peng, G. Xing, W. Zhang, and G. Meng, "Accurate and robust displacement measurement for FMCW radar vibration monitoring," *IEEE Sensors J.*, vol. 18, no. 3, pp. 1131–1139, Feb. 2018.
- [12] J. Baqersad, P. Poozesh, C. Niezrecki, and P. Avitabile, "Photogrammetry and optical methods in structural dynamics—A review," *Mech. Syst. Signal Process.*, vol. 86, pp. 17–34, Mar. 2017.
- [13] D. Kalpoe, K. Khoshelham, and B. Gorte, "Vibration measurement of a model wind turbine using high speed photogrammetry," *Proc. SPIE*, vol. 8085, Jun. 2011, Art. no. 80850J.
- [14] T. J. Beberniss and D. A. Ehrhardt, "High-speed 3D digital image correlation vibration measurement: Recent advancements and noted limitations," *Mech. Syst. Signal Process.*, vol. 86, pp. 35–48, Mar. 2017.
- [15] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, "Phase-based video motion processing," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 80:1–80:10, Jul. 2013.
- [16] Y. Zhang, S. L. Pintea, and J. C. Van Gemert, "Video acceleration magnification," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 502–510.
- [17] M. P. Popek, M. E. Danielewska, and D. R. Iskander, "Assessing the feasibility of the use of video motion magnification for measuring microdisplacements," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 9, pp. 2329–2336, Sep. 2017.
- [18] M. P. Popek and D. R. Iskander, "A new approach to the phase-based video motion magnification for measuring microdisplacements," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 2, pp. 354–361, Feb. 2020.
- [19] C. Peng, C. Zeng, and Y. Wang, "Phase-based noncontact vibration measurement of high-speed magnetically suspended rotor," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4807–4817, Jul. 2020.
- [20] T. Gautama and M. M. Van Huffel, "A phase-based approach to the estimation of the optical flow field using spatial filtering," *IEEE Trans. Neural Netw.*, vol. 13, no. 5, pp. 1127–1136, Sep. 2002.
- [21] L. Oriat and E. Lantz, "Subpixel detection of the center of an object using a spectral phase algorithm on the image," *Pattern Recognit.*, vol. 31, no. 6, pp. 761–771, Jun. 1998.
- [22] A. D. Jepson and D. J. Fleet, "Phase singularities in scale-space," *Amer. J. Phys.*, vol. 67, no. 1, pp. 55–60, 1999.
- [23] D. J. Fleet and A. D. Jepson, "Stability of Phase Information," in *Proc. IEEE Workshop Vis. Motion*, Oct. 1991, vol. 4, no. 1, p. 1.
- [24] S. Meyer, O. Wang, H. Zimmer, M. Gross, and A. Sorkine-Hornung, "Phase-based frame interpolation for video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1410–1418.
- [25] D. Kitahara and I. Yamada, "Algebraic phase unwrapping along the real axis: Extensions and stabilizations," *Multidimensional Syst. Signal Process.*, vol. 26, no. 1, pp. 3–45, Jan. 2015.
- [26] Z. Wang and B. Han, "Advanced iterative algorithm for phase extraction of randomly phase-shifted interferograms," *Opt. Lett.*, vol. 29, no. 14, p. 1671, 2004.
- [27] H. Kayaba and Y. Kokumai, "Non-contact full field vibration measurement based on phase-shifting," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2548–2556.
- [28] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3, Nov. 1995, pp. 444–447.
- [29] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 49–70, Oct. 2000.
- [30] L. Ying, *Phase Unwrapping*, vol. 24, no. 50. Hoboken, NJ, USA: Wiley, 2006.
- [31] D. J. Fleet and A. D. Jepson, "Computation of component image velocity from local phase information," *Int. J. Comput. Vis.*, vol. 5, no. 1, pp. 77–104, 1990.
- [32] T. Blu, P. Thévenaz, and M. Unser, "MOMS: Maximal-order interpolation of minimal support," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1069–1080, Jul. 2001.
- [33] C. Gilliam and T. Blu, "Local all-pass geometric deformations," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 1010–1025, Feb. 2018.
- [34] A. Brandt, *Noise and Vibration Analysis: Signal Analysis and Experimental Procedures*. Hoboken, NJ, USA: Wiley, 2011.



Jiwen Zhou received the B.Eng. degree from Xi'an Jiaotong University, Xi'an, China, in 2013. He is currently pursuing the Ph.D. degree with Shanghai Jiao Tong University, Shanghai, China, both in mechanical engineering.

He is currently with the State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University. His current research interests include digital signal processing, image processing, and vision-based vibration measurement.



Hongguang Li received the B.Eng. degree in engineering mechanics and the master's degree in computational mechanics from Dalian University of Technology, Dalian, China, in 1993 and 1996, respectively, and the Ph.D. degree in mechanical engineering from Northeastern University, Shenyang, China, in 1999.

Now he is a Professor with Shanghai Jiao Tong University, Shanghai, China. His current research interests include dynamics analysis and control and signal processing and fault diagnosis, along with reliability and life assessment techniques.

Dr. Li is a member of the Committee of Fault Diagnosis of the Chinese Society of Vibration Engineering, and the Chairperson of the Rotor Dynamics Committee of the Chinese Society of Vibration Engineering.



Li Zhang received the B.Eng. degree in refrigeration and cryogenic technology from Zhejiang University, Hangzhou, China, in 1999. He received the M.Eng. degree in mechanical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2012.

Now he is a Researcher with Shanghai Institute of Spacecraft Equipment, Shanghai. His current research interest is spacecraft ground dynamic environment test, including vibration, acoustic, modal, S/V shock separation tests, and so on.



Yun Li received the B.Eng. degree from Harbin Engineering University, Harbin, China, in 2014. She is currently pursuing the Ph.D. degree in mechanical engineering with Shanghai Jiao Tong University, Shanghai, China.

She is currently with the State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University. Her current research interests include signal processing, rotating machinery fault diagnosis, and machine learning.



Xiaojian Wang received the M.Eng. degree from Shanghai Jiao Tong University, Shanghai, China, in 2019, where he is currently pursuing the Ph.D. degree, both in mechanical engineering.

He is currently with the State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University. His current research interests include signal processing and smart structures.