Structural Vibration Mode Identification from High-Speed Camera Footages using an Adaptive Spatial Filtering Approach

Miaoshuo Li[a], Guojin Feng[a*], Rongfeng Deng[a], Feng Gao[b], Fengshou Gu[a] and Andrew D. Ball[a]

[a] *Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield, HD1 3DH, UK*
[b] *EPSRC Future Metrology Hub, University of Huddersfield, Huddersfield, HD1 3DH, UK*

Abstract

As a non-contact and full-field testing method, high-speed camera-based modal analysis has become a feasible and acknowledged approach. However, extracting small displacements from noisy images has experienced high level of difficulty, especially in high frequency range. This paper proposes a novel adaptive spatial filtering (beamforming) algorithm to extract the displacement signals using high-speed camera. In the proposed algorithm, one pixel is considered as a sensor measuring displacement and a set of pixels are therefore taken as the elements of sensor array. Then, an adaptive spatial filtering acting on this sensor array is proposed. The proposed approach mainly includes three steps. Firstly, a set of pixels are selected to compose a sensor array according to signal to distortion and noise (SINAD). Secondly, a node/antinode searching scheme is proposed based on sinusoid-based piecewise functions, which works as an adaptive filter to match mode shape and enhance modal displacement. Finally, the output of spatial filtering is adopted as the system response for the identification of modal parameters. To validate the performance of the proposed method, simulation and experiment studies are conducted based on measuring the vibration model properties of a free-free beam, which includes a comparison with LK optical flow method and conventional accelerometer-based method. The results show that the SNR of estimated displacement and computational efficiency is significantly improved without using additional sensors. The proposed method paves a way for broadening the applications of using high-speed camera for full-field vibration measurements.

Keywords: Full-field measurement, Adaptive spatial filtering, Sinusoid-based piecewise function, High-speed camera, Modal analysis

Nomenclature

$\boldsymbol{F(x,y)}$	Reference frame
$\boldsymbol{p(x,y)}$	Predicted sharp image
$\boldsymbol{b(x,y)}$	Point spread function
$\boldsymbol{F(x)}$	One-dimension representation of reference frame
$\boldsymbol{p(x)}$	One-dimension representation of predicted sharp image
$\boldsymbol{b(x)}$	One-dimension representation of point spread function
$\boldsymbol{SINAD}$	Signal to noise and distortion
$\boldsymbol{P_s}$	Power of signal
$\boldsymbol{P_d}$	Power of distortion
$\boldsymbol{P_n}$	Power of noise
$\boldsymbol{I_A(t)}$	Intensity signal from sensor array
$\boldsymbol{F'}$	Image gradient of reference image
$\boldsymbol{u(t)}$	Sensor elements of displacement
$\boldsymbol{\varphi_i}$	Mode shape vector
$\boldsymbol{z_i(t)}$	Modal displacement
$\boldsymbol{\Phi}$	Mode shape matrix
$\boldsymbol{\eta(t)}$	Noise vector
$\boldsymbol{W}$	Weighted matrix
$\boldsymbol{\hat{z}(t)}$	Estimator of modal displacement matrix
$\boldsymbol{w_i}$	Weight vector
$\boldsymbol{\hat{z}_i(t)}$	Estimator of modal displacement
$\boldsymbol{SNR}$	Signal to noise ratio
$\boldsymbol{G_a}$	Array gain
$\boldsymbol{m}$	Number of sensor elements
$\boldsymbol{n}$	Desired order of modes
$\boldsymbol{w_{ini}}$	Initial value of weight vector
$\boldsymbol{w_{opt}}$	Optimizer of weight vector value

1 Introduction

The camera-based method or photogrammetry has been increasingly employed for modal analysis or vibration measurement due to their inherent advantages, such as non-contact and remote full-field measurement [1]. According to the application scenarios and conditions, the camera-based modal analysis can be generally classified into two main branches: operational modal analysis (OMA) and experimental modal analysis (EMA). The camera-based methods for OMA has been widely applied to the structural health monitoring of large-scale structures in civil engineering [2], [3], such as bridges [4], infrastructures [5] and wind turbines [6]–[8]. Likewise, EMA using camera-based method is also investigated intensively [9]. In contrast with OMA, the object tested in EMA is relatively small and tends to generate higher frequency components. The requirement to analyse dynamic behaviour in higher frequencies demands a broader bandwidth of measurement approach. For camera-based methods, despite the sampling frequency being sufficiently high, extracting the small displacement amplitudes of interesting modes at high frequency (generally about micrometre [10]) is also a great challenge. Thus, the capability to overcome this difficulty plays an important role in the camera-based EMA.

The most common approach for modal analysis is digital image correlation (DIC) currently [1], [9], [11]–[13], which aims at measuring the 2D or 3D strain or displacement based on maximizing the correlation coefficient of given subsets between images [14]. For example of

many applications, Helfirck et al. [9] measured the mode shape of a dryer in high frequency range [15]. DIC has been proved to be a viable option for EMA and other vibration applications [13].

The phase-based motion magnification [16], [17] or phase-based optical flow [6], [18] is another approach to estimate displacement for modal analysis, [19], [20]. The local phase and amplitude are determined by filtering an image with the complex steerable pyramids, and then the constant phase contours are used to estimate the displacement or magnify the motion [21], [22]. Chen et al. first identified the modes with phase-based motion magnification. Through cantilever beam and pipe test, modes were extracted and visualized. In addition, the phase-based motion magnification can be associated with DIC, 3D point tracking and other optical methods [23]. In [24], the modes of a 2.3 meter wind turbine blade are extracted by combining motion magnification and DIC, which shows the phase-based motion magnification can improve the signal to noise ratio (SNR) of the estimated displacement.

As a universal algorithm, the gradient-based optical flow and its variants are also widely used in image-based modal analysis [25]–[27]. Besides the typical LK-optical flow method, the simplified gradient-based optical flow is developed recently [28], which takes a subset or a pixel as an independent sensor and linearizes the relation between displacement and intensity. Afterwards, a hybrid identification method that combines a high-speed camera with accelerometers is developed [10], and recently a dynamic sub-structuring approach are applied to improve the mode shape accuracy [29].

To summarize, all approaches above can be categorized as the window-based image registration method. Because they all take advantages of a neighbourhood operation [30], which could either be a subset (window/patch/block) in DIC or gradient-based optical flow, or be a local kernel for convolution with image in phase-based methods. The inherent difficulty of these methods is the determination of window size. If the window size is too large, the displacement inside a window is aliased by averaging. Instead, if the window size is too small, the obtained result will be noisy. Furthermore, the estimation of displacement in each window is completely uncorrelated, which ignores the high correlation between responses of a dynamic system.

To avoid the deficiencies in the window-based image registration methods, this paper takes a large set of pixels to formulate a sensor array, and applies spatial filtering [31] to maximize SNR of the desired displacement signal, which is inspired by taking a single pixel as a sensor [28]. As a fundamental methodology in array signal processing [32], spatial filtering (also called beamforming) denotes an operator that perform a weighted sum to spatial samples, which has been widely applied in radar [33], wireless communications [34], acoustics [35] and many cases of structural health monitoring [36] [37] [38]. In camera-based modal analysis, when vibration signals are captured densely in space by pixels, these multipoint responses are analogous to spatial samples in the radar or microphone array case. On this basis, a novel method is developed in this paper, in which a set of pixels is taken as a sensor array and the modal displacement (displacement signal under modal coordinate) is the desired signal to extract.

The rest of paper is organized as follows. Section 2 details the methodology proposed. Section 3 introduces the framework developed, which covers the formation of sensor array based on SINAD and the design and analysis of spatial filtering. Section 4 and Section 5 conduct the simulation study and experimental validation, respectively. Finally, discussions and

conclusions are drawn in Section 6 and Section 7.

2 Methodology

In window-based image registration method, the displacement is estimated based on pixels within each window, without considering the correlation of displacement among windows. This leads to a limited estimation precision in each small window. To overcome this limitation, the spatial-filtering-based scheme is developed, which applies a set of weight vectors to linearly combine signals from sensor array (a set of selected pixels in this research). Its aim is to generate constructive interference of desired signal and suppress noise. As illustrated in Fig. 1(a), the oscillating intensity captured by pixels at different space elements are combined to enhance each modal displacement.

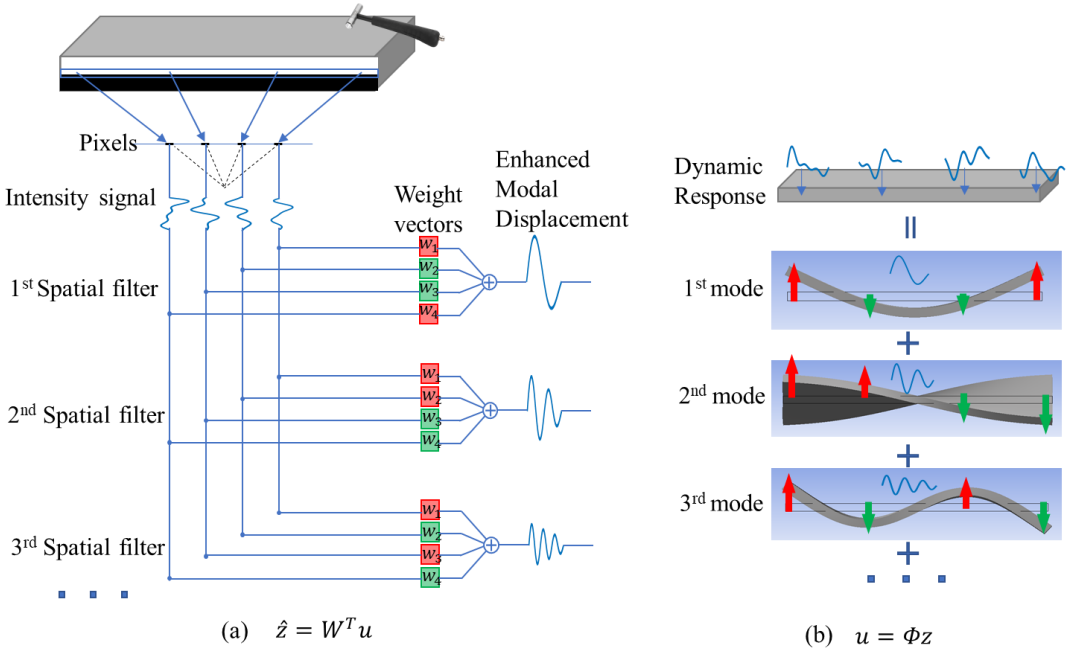


Fig. 1 Schematic diagram of the proposed method, (a) Illustration of using a pixel sensor array to enhance modal displacment signal and (b) the principle of modal superposition [32]

According to modal superposition, each dynamic response can be decomposed into several modes [32], as shown in Fig. 1(b). In each mode, all space elements are oscillating with the same frequency and in phase. As shown in Fig. 1(b), its mode shape $\boldsymbol{\Phi}$ displays a form of standing wave, in which the motion direction is opposite across nodes, as denoted by the red and green arrow, respectively. The weight vectors $\boldsymbol{W}$ are thus designed for this structure, which is set to match the mode shape, so that the constructive waves will be generated, as is illustrated in Fig. 2 (b) and the details will be discussed in Section 3.2.

The proposed spatial filtering scheme is very similar to that of other sensor arrays like microphone array but with difference in array weight formulation. As shown in Fig. 2(a), the weight vector of microphone array is designed to shift phase of signals to compensate temporal delay from a selective direction, also known as a ‘delay and sum array’. Comparatively, aiming at the modal analysis, the weight vector for the pixel sensor array is designed to change amplitude of signals along the mode shape, which can be described as a ‘matching and sum array’. Intuitively, the weight vector is similar to flip ‘green waves’ to ‘red waves’ to unify the direction of vibration before sum in Fig. 2 (b).

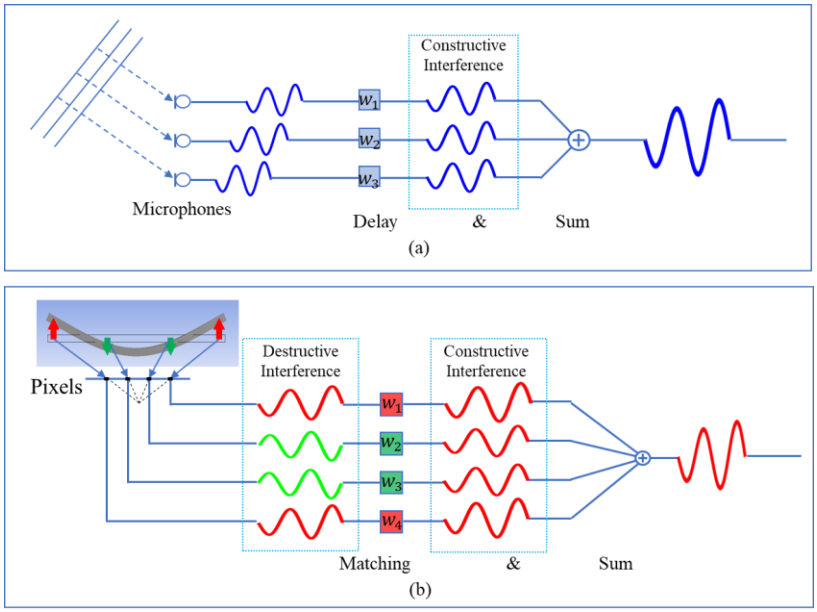


Fig. 2 Comparision of working principles between (a) conventional sensor arrary (a microphone array) and (b) a pixel sensor array used in the proposed spatial filtering method

3 Proposed Framework

Based on the spatial filtering concept introduced in Section 2, this paper proposes a framework for displacement estimation in high-speed camera-based vibration modal analysis. The proposed framework can be divided into three key modules: pixel sensor array formation, spatial filtering, and modal identification, which is illustrated in Fig. 3.

The first module, formation of a sensor array, selects pixels sensitive to displacement in a reference image [32]. By taking each pixel as a displacement sensor, the performance of the equivalent sensor measuring displacement can be used as criterion for selection of sensor elements. Concretely, SINAD is analysed because it can comprehensively measure the performance of each pixel sensor. Subsequently, a sensor array is constructed with selected pixels. The second module proposes an adaptive spatial filtering on signals collected by sensor elements according to criterion of SNR, in which a node/antinode searching scheme based on sinusoid-based piecewise function is proposed. It works as an adaptive filter to match mode shape and enhance modal displacement. The third module is modal identification in which a frequency domain method, the Least-squares rational function (LSRF) is used to identify modal parameters including the natural frequency and damping ratio.

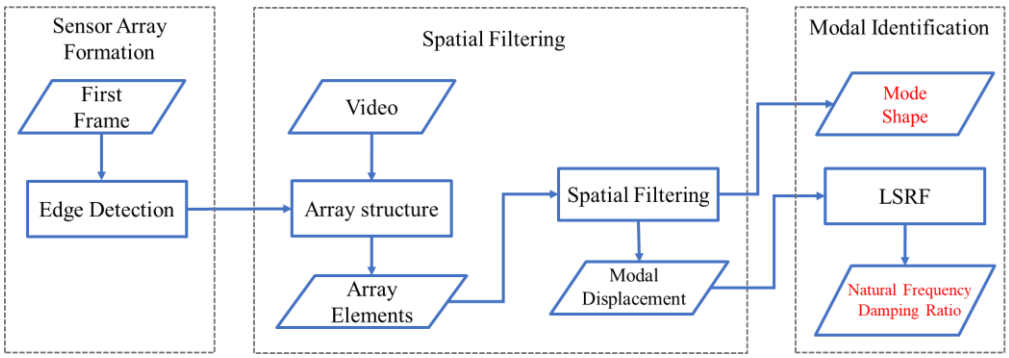


Fig. 3 Diagram of proposed framework, which mainly consists of sensor array formation, spatial filtering and modal identification.

3.1 Formation of a Sensor Array

According to the principle of array signal processing, data acquisition from each array element is independent [31]. By utilizing this condition, each pixel can be regarded as a sensor for displacement measurement, and the pixels with high performance can be selected as elements of sensor array intentionally for guaranteeing the quality of acquired raw signals.

3.1.1 Point Spread Function and Pixel Sensor

In image-based vibration measurement, the motion is often estimated based on pixel intensity. A linear relationship between intensity and motion is generally assumed for simplification [28]. However, this linear assumption cannot be met in practice because of the effect of image blurring [30]. Especially, the displacement is a pixel-level oscillation in vibration measurement, so the scale of non-linearity relative to level of micro-motion cannot be neglected indiscreetly.

Fig. 4 shows the mechanism of the image blurring resulted from defocus aberration [30]. As shown in Fig. 4 (a), the black-white stripes are painted onto the surface of target object to improve contrast of the captured image. When the edge of black-white pattern on the object projects onto image, its sharp edge will be blurred into a smooth curve. The point spread function (PSF) is often used to describe this blurring effect, Fig. 4 (b) shows the blurred image is represented by convolution of the predicted sharp image and PSF [30]:

$$F(x,y) = p(x,y) * b(x,y) \tag{1}$$

in which, $p(x,y)$ is a sharp edge described by a step function, $b(x,y)$ is the PSF kernel, * represents a convolution operation and $F(x,y)$ is the blurred image. A real image of experimental data shows the same phenomenon in Fig. 14

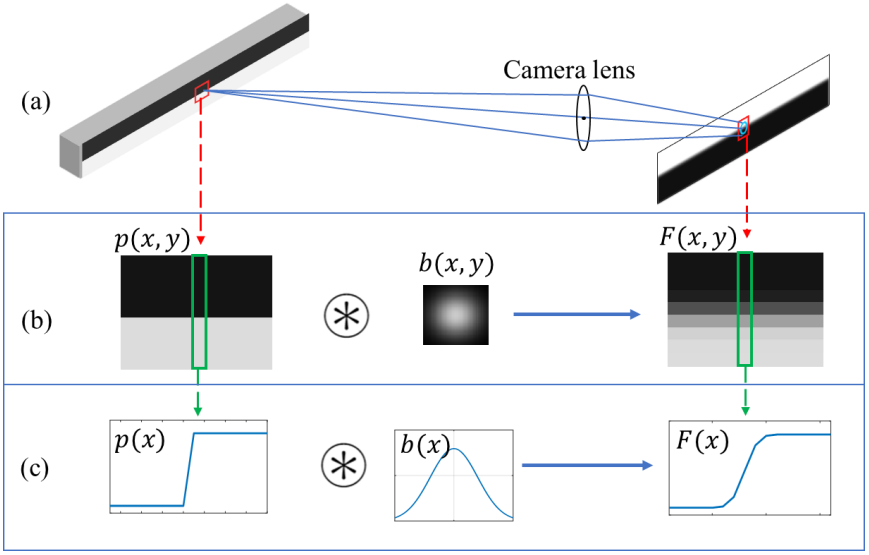


Fig. 4 Illustration of an image blurring process: left colomn for ideal image, middle colom for the kernel ofor blurring and right colomn for real image. (a) the schematic of video taking, (b) modelling 2-D blurring process and (c) modelling 1-D blurring process

To attentively investigate the transfer function between displacement and intensity, one row of pixels is picked out of an image to build a one-dimension function $F(x)$. Fig. 4(c) illustrates the one-dimensional representation of convolution: $F(x) = p(x) * b(x)$, such a profile causes the nonlinearity between intensity and displacement. Taking each pixel of this profile as a sensor for displacement measurement, the generated intensity signal of each pixel is different,

when vibration occurs. Fig. 5 demonstrates how the pixel intensity I_c changes with pixel sensor at different locations x.

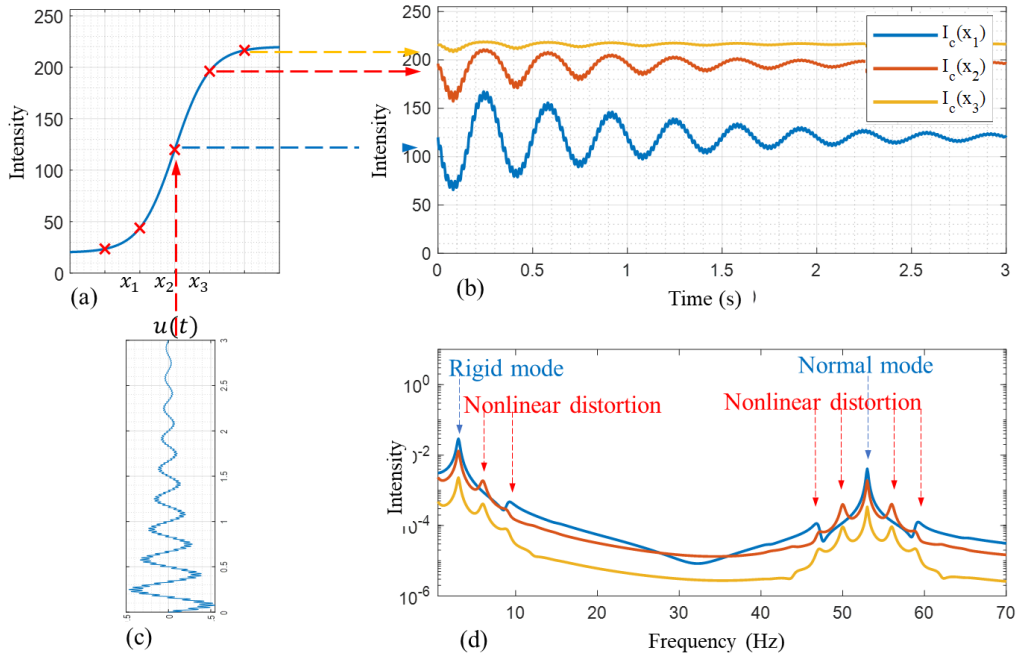


Fig. 5 Intensity signal of different pixels captured under motion (a) ROI of image, (b) response: intensity signal, (c) excitation displacement signal and (d) spectrum of the real output signal $I_C(t)$

Fig. 5(a) represents a region of image including five pixels. It shifts left and right following a displacement $u(t)$, as shown in Fig. 5 (c). The displacement is set as a damping vibration that contains a rigid mode at 3 Hz and a normal mode at 56 Hz. Fig. 5(b) shows intensity signals captured at location $x_1,\ x_2$ and x_3 with image noise. The amplitudes of three pixels differ due to their image gradient along x-axis in Fig. 5 (a). Additionally, waveforms show an obvious distortion due to the nonlinear profile, this phenomenon can also be found in experimental data (Fig. 14(c)). This nonlinear distortion produces a bunch of unwanted components in frequency domain [39], [40], as shown in Fig. 5(d). Overall, three signals show significant difference in both signal amplitude and distortion. It can be seen that pixel at x_1 (the centre of edge) obtains the highest signal amplitude with less distortions. In order to verify that this location can indeed obtain the highest signal quality, a comprehensive measure is required.

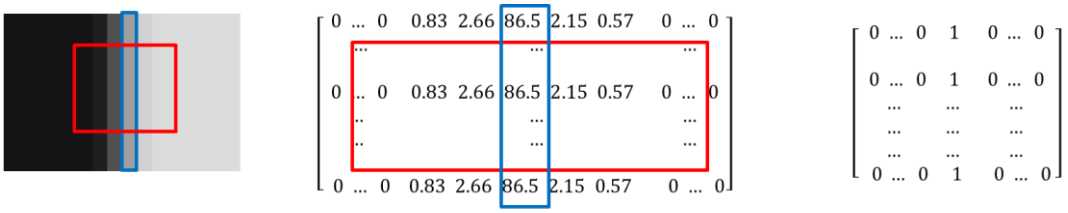
3.1.2 Array Structure based on SINAD

The signal to noise and distortion (SINAD) allows to take all the signal strength, noise influences and nonlinear distortion into account [41]. Thus, it is adopted to evaluate the performance of each pixel according to Eq. (2),

$$\boldsymbol{SINAD} = \frac{P_s}{P_d + P_n} \tag{2}$$

where P_s represents the power of signal component, P_d represents the power of distortion, and P_n represents noise power [41]. By calculating SINAD of x_1, x_2, x_3, it yields 86.5, 2.15, 0.57, respectively. Furthermore, it shows that trend of SINAD reaches the maximum at the centre of edge and gradually decreases to zero towards both sides. This verifies that pixels capturing the highest-quality signal are exactly located at the centre line of the edge.

Therefore, a SINAD matrix of entire image is built to examine SINAD of every pixel. This matrix exhibits the quality of displacement signal obtained from different pixels is uneven. In another word, only a part of pixels is sensitive to displacement, for example, if the pixel has no pattern, its SINAD will be zero or very small and obviously this pixel will have no contribution to measure displacement. Thus, only pixels on the centre line of the edge should be kept in the sensor array. To achieve this, the Sobel edge detector is used to process the reference image, which yields a binary matrix $A_s(x, y)$. It denotes the spatial coordinates of pixel sensors in the array, known as the array structure of sensor array [42]. The pixels with value one in $A_s(x, y)$ should be selected as elements of the sensor array to record the intensity signals while other pixels will be removed.



(a) $F(x, y)$ (b) $Sinad(x, y)$ (c) $A_s(x, y)$

Fig. 6 Comparision of SINAD between the proposed method and window-based method, (a) pixels selected based on two kinds of methods on image, (b) SINAD of selected pixels and (c) array structure based edge detection

Fig. 6 compares the signal quality of proposed method with the window-based registration. The blue box represents the pixels selected using $A_s(x, y)$, the red box represents a subset often used in window-based methods. Its $SINAD(x, y)$ matrix and array structure $A_s(x, y)$ is also shown. The average SINAD of the red box is far smaller than that of the blue box because it contains many pixels with small or zero SINAD, these pixel collects poor signal even no signal components. Comparably, the pixels in blue box are selective about high performance. That makes difference of efficiency in signal acquisition between the proposed method and conventional methods.

3.2 Modal Displacement Estimation via Spatial Filtering

From the relationship described in Fig. 5, the displacement can be converted from intensity signal following the Eq. (3):

$$\boldsymbol{u}(t) = \frac{\boldsymbol{I}_A(t) - F}{F'} \tag{3}$$

in which $\boldsymbol{I}_A(t)$ denotes a vector of intensity signal of pixel sensors selected by using array structure $A_s(x, y)$, F is pixel's intensity and F' is image gradient at that pixel in a reference image as shown in Fig. 5(a). Many existing methods can be used to solve the image gradient $F'(x)$, the Sobel operator is chosen in this study.

As a result of conversion, $\boldsymbol{u}(t)$ becomes a vector of displacement signals collected by the sensor array. The dynamic response $\boldsymbol{u}(t)$ can be further regarded as a linear combination of multiple modes according to the principle of mode superposition [43]. Taking into inevitable noise influences, $\boldsymbol{u}(t)$ can be expressed in a matrix form:

$$\boldsymbol{u}(t) = \boldsymbol{\Phi z}(t) + \boldsymbol{\eta}(t) \tag{4}$$

where $\boldsymbol{u}(t) \in \mathbb{R}^m$ is raw displacement signals from a sensor array with m elements. $\boldsymbol{\Phi} = [\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \dots, \boldsymbol{\varphi}_n] \in \mathbb{R}^{m \times n}$ is mode shape matrix with n modes, in which $\boldsymbol{\varphi}_i \in \mathbb{R}^m$ is a mode shape vector. $\boldsymbol{z}(t) = [z_1(t), z_2(t) \dots z_n(t)]^T \in \mathbb{R}^n$ is a vector representing modal displacement, i.e. displacement under modal coordinate, $\boldsymbol{\eta}(t) \in \mathbb{R}^m$ is a white Gaussian noise vector subject to $\boldsymbol{\eta} \sim N(0, {\sigma_\eta}^2)$. It is noted that the modal displacement $\boldsymbol{z}(t)$ is the target to be estimated in this study.

Conventional methods use measured $\boldsymbol{u}(t)$ to identify modal parameters directly, but window/subset derived $\boldsymbol{u}(t)$ usually have low SNR and such direct method often result in poor results. Considering there are a large number of channels/pixels available, an array signal processing scheme can be employed for enhancing the targeted components $\boldsymbol{z}(t)$. According to spatial filtering approach, the estimator of modal displacement $\hat{z}_i(t)$ can be obtained by applying weight vectors $\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n] \in \mathbb{R}^{m \times n}$ to the measured response $\boldsymbol{u}(t)$ [32], which yields:

$$\hat{\boldsymbol{z}}(t) = \boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{u}(t) = \boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{\Phi}\boldsymbol{z}(t) + \boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{\eta}(t) \tag{5}$$

in which $\hat{\boldsymbol{z}}(t) \in \mathbb{R}^n$ is an estimator of $\boldsymbol{z}(t)$. In section 1, Fig. 1 illustrates the structure of spatial filtering scheme. It shows that each output $\hat{z}_i(t)$ is a linear combination of all array elements. By multiplying $\boldsymbol{W}$, the displacement signals are transformed from a *m*-dimension spatial coordinate into *n*-dimension modal coordinate.

To examine the effect of $\boldsymbol{w}_i$ more conveniently on a particular estimator $\hat{z}_i(t)$, Eq. (5) is rewritten as

$$\hat{z}_i(t) = \sum_{j=1}^{n} \boldsymbol{w_i}^{\boldsymbol{T}}\boldsymbol{\varphi_j}\, z_j(t) + \boldsymbol{w_i}^{\boldsymbol{T}}\boldsymbol{\eta}(t) \tag{6}$$

It shows that $\hat{z}_i(t)$ is composed by two parts:$\boldsymbol{w}_i{}^T\boldsymbol{\varphi}_j\, z_j(t)$ due mainly to the modal contribution and noise $\boldsymbol{w}_i{}^T\boldsymbol{n}(t)$ due to noise influence. To extract the weak displacement signal, the weight vector $\mathbf{w}_i$ needs to be designed so that the ratio between signal component and noise component i.e., SNR of $\hat{z}_i(t)$ is maximized. Due to the signal components of interest in $\hat{z}_i(t)$ is only $z_i(t)$, which turns out that other modal signals should be ignored in $\hat{z}_i(t)$ and hence result in the i^{th} term $\boldsymbol{w}_i{}^T\boldsymbol{\varphi}_i z_i(t)$ alone. Therefore, the SNR of array output $\hat{z}_i(t)$ can be expressed as:

$$SNR[\hat{z}_i(t)] = \frac{E\{[\boldsymbol{w}_i{}^T\boldsymbol{\varphi}_i\, z_i(t)]^2\}}{E\{[\boldsymbol{w}_i{}^T\boldsymbol{\eta}(t)]^2\}} \tag{7}$$

where the weight vector $\boldsymbol{w}_i$ is the variable of this equation. Taking output $SNR[\hat{z}_i(t)]$ as an objective function and $\boldsymbol{w}_i$ as a variable, it enables an optimization problem to formulate as:

$$\boldsymbol{w}_i = arg\, \max_{\boldsymbol{w}_i} SNR[\hat{z}_i(t)] \tag{8}$$

In this way, a SNR-maximized $\mathbf{w}_i$ can be found. According to Cauchy–Schwarz inequality, depicted in Appendix 1, it can be shown that $SNR[\hat{z}_i(t)]$ has its upper bound as

$$SNR[\hat{z}_i(t)] = \frac{(\boldsymbol{w}_i{}^T\boldsymbol{\varphi}_i)^2}{\boldsymbol{w}_i{}^T\boldsymbol{w}_i}\,\frac{{\sigma_{z_i}}^2}{{\sigma_\eta}^2} \le (\boldsymbol{\varphi}_i{}^T\boldsymbol{\varphi}_i)\,\frac{{\sigma_{z_i}}^2}{{\sigma_\eta}^2} \tag{9}$$

in which, ${\sigma_{z_i}}^2 = E\{z_i(t)^2\}$ represent signal power and ${\sigma_\eta}^2$ is the variance of noise. It shows that if and only if $\boldsymbol{w}_i$ and $\boldsymbol{\varphi}_i$ are linearly dependent, the SNR is maximized. In another word, the upper bound of SNR can be reached, while $\boldsymbol{w}_i$ paralleling with mode shape vector $\boldsymbol{\varphi}_i$, i.e., it yields that $\widehat{\boldsymbol{\varphi}}_i = \boldsymbol{w}_{opt}$, because they are dimensionless. Moreover, Eq. (9) shows that the maximum SNR is related to that of the input signal $\boldsymbol{u}(t)$. The array gain is hence used to represent the gain of SNR before and after spatial filtering [44].

$$G_a = \frac{SNR[\hat{z}_i(t)]}{\overline{SNR}_{in}} \leq m \tag{10}$$

where $\overline{SNR}_{in}$ is defined as the average input SNR over all pixels. By excluding the same terms in Eq. (10), it shows that the upper bound of array gain is m, which is proved in Appendix 2, where m is the number of array elements. Eq. (10) shows that the maximum array gain G_a depends only on the number of array elements m, it is often expressed in dB.

$$G_{dB} \leq 10\ \log_{10}(m) \tag{11}$$

In general, increasing number of array element *m* enables a higher SNR output, which can be achieved conveniently by this video-based measurement.

3.3 Adaptive Spatial filtering and mode identification

3.3.1 Optimization Problem defined in a Constraint Form

With a unit weight vector $\boldsymbol{w}_i$, i.e. $\|\boldsymbol{w}_i\|^2 = 1$, the denominator in the middle term of Eq. (9) is eliminated, which turns Eq. (8) into a constraint optimization problem:

$$\begin{aligned} &\max_{\boldsymbol{w}_i} \quad SNR(\boldsymbol{w}_i) = ({\boldsymbol{w}_i}^T \boldsymbol{\varphi}_i)^2 \ \frac{{\sigma_{z_i}}^2}{{\sigma_\eta}^2} \\ &s.t. \qquad \|\boldsymbol{w}_i\|^2 = 1 \end{aligned} \tag{12}$$

Under this constraint condition $\|\boldsymbol{w}_i\|^2 = 1$, the noise power is not influenced by $\boldsymbol{w}_i$ as shown in Eq. (A4) in Appendix 1. Therefore, $SNR(\boldsymbol{w}_i)$ is determined by its part of signal power $P(\boldsymbol{w}_i) = ({\boldsymbol{w}_i}^T \boldsymbol{\varphi}_i)^2 {\sigma_{z_i}}^2$ only. This means that $P(\boldsymbol{w}_i)$ can substitute $SNR(\boldsymbol{w}_i)$ to serve as the objective function. In practice, $P(\boldsymbol{w}_i)$ can be easily obtained by measuring the power spectral density at frequency of interest. This objective function establishes the connection between variable parameters of filter and feedback of output signal, which is a key point in an adaptive filter[36], [45]. To solve this optimization problem, a searching scheme based on node/antinode is put forward as one example of many potential numerical solutions.

3.3.2 Approximation of Mode Shape using a Sinusoid-based Piecewise Function

In physics, mode shape of a continuous structure is often a type of normal mode presenting a form of standing wave. In addition, the location of nodes and antinodes shows where the strongest and weakest vibration occurs, which highly influences the performance and safety of the object under test. Generally, standing waves are thought to be produced by interference between waves reflected back and forth at natural frequencies, so that nodes and antinodes can be viewed as the characteristics of a standing wave.

As mode shape can be taken as a wave signal, the sinusoidal function is chosen for numerically

approximating the desired mode shape. The nodes and antinodes split the weight vector $\boldsymbol{w}$ into j pieces w_j to formulate a piecewise function and each piece can be denoted by

$$w_j(y) = A_j \cos\left(\frac{2\pi}{\lambda_j} y + \psi_j\right) \qquad 0 < y < \frac{\lambda_j}{4} \tag{13}$$

in which y is local coordinate inside each piece of function, A_j, λ_j, ψ_j is the undetermined amplitude, wavelength, and initial phase, respectively. Its derivative can be obtained as

$$\frac{dw_j}{dy} = -\frac{2\pi A_j}{\lambda_j} \sin\left(\frac{2\pi}{\lambda_j} y + \psi_j\right) \qquad 0 < y < \frac{\lambda_j}{4} \tag{14}$$

Following the physical property of wave, the mode shape and its changing rate should be continuous, therefore, the constructed function and its derivative will be continuous as well. This means the amplitudes should be identical on both sides of a node/antinode, thereafter, the boundary condition is imposed on the endpoints of every piece as following,

$$w_j(\frac{\lambda_j}{4}) = w_{j+1}(0) \tag{15}$$

Likewise, its derivative also accords with continuity,

$$\dot{w}_j(\frac{\lambda_i}{4}) = \dot{w}_{j+1}(0) \tag{16}$$

Naturally, the value of w_j at a node should be zero as well as its derivative at antinode should be zero, and then the boundary condition can be further simplified at a node. Thus, there exists a relationship for two adjacent amplitudes A_j and A_{j+1}, as shown below,

$$\frac{4A_j}{\lambda_j} = \frac{4A_{j+1}}{\lambda_{j+1}} = \tan\theta_j \tag{17}$$

This represents the parameters on both side of a node accord with this equation, namely they share the same included angle θ_j.

Similarly, at the antinode, two pieces share one amplitude, i.e.,

$$A_j = A_{j-1} \tag{18}$$

By imposing the above boundary conditions, the neighbouring pieces are coupled by the shared A_j or θ_j, and then based on the Eq. (17), A_j can be calculated by

$$A_j = \frac{1}{4}\lambda_j \tan(\theta_j) \tag{19}$$

Then, the local coordinate and local phase in each piece can be derived from the global coordinate.

$$x = y - l_j \tag{20}$$

$$\psi_j = \frac{\pi}{2}(j-1) \tag{21}$$

in which the l_j is the location of node/antinode, which is explained in detail in Appendix 3.

Correspondingly, x is under global coordinate. Then substitute all variables into the Eq. (13),

$$w(x) = \frac{tan(\theta_j)\lambda_j}{L}\cos\left(\frac{2\pi}{\lambda_j}(x - l_j) + \frac{\pi}{2}(i-1)\right) \quad l_j < x < l_{j+1} \tag{22}$$

in which $x \in (0, L)$ and L is the length of object. By using this function and considering specific boundary condition, the profile of arbitrary mode shape can be approximated numerically, if only the mode shape shows a form of standing wave.

3.3.3 Node/antinode Searching Scheme

Based on the above analysis, the location of nodes/antinodes can denote a mode shape approximately by a few parameters. By moving location of each node in the weight vector $\boldsymbol{w}_k$ along direction of increasing signal power P_k, the maximal P_k can be obtained.

In this expression, the number of variables can be reduced significantly, from number of pixels to number of nodes and antinodes, which makes the optimization problem feasible. Based on this piecewise function, the optimization process is described as follows:

1) Use the Eq. (A12) or Eq. (A13) in Appendix 3 to set $w(x)$ using node/antinode location l_i as parameter.
2) Set initial value of node/antinode location $\boldsymbol{l}_{ini}$, and obtain $\boldsymbol{w}(\boldsymbol{l}_{ini})$
3) Apply $\boldsymbol{w}(\boldsymbol{l}_{ini})$ on the input signals to obtain the output signal $\hat{\boldsymbol{z}}$, and calculate its power $P(\boldsymbol{w}_{ini})$ at a frequency range where the mode lies in.
4) Determine the iteration direction based on initial value and the first iteration, which moves the node by one-pixel distance and compares the power $P(\boldsymbol{w}_i)$ with initial one, if the power increases, it is taken as the direction for the following iterations, otherwise, the opposite direction will be used.
5) Move the node/antinode $\boldsymbol{l}_k$ and obtain its $P(\boldsymbol{w}_k)$. and repeat this step along the direction of increasing $P(\boldsymbol{w}_k)$
6) Terminate when $P(\boldsymbol{w}_k)$ starts to drop, and then determine the optimal weight vector $\boldsymbol{w}_{opt}$ which has the maximum $P(w)$.

3.3.4 Identification of Natural Frequency and Damping Ratio

In this paper, a frequency domain method, LSRF [46] is adopted to identify natural frequency and damping ratio. The modal displacement from spatial filtering is used as input for identification. The principle of identification methods is fitting the parametric model to non-parametric model. The SNR of signal will directly affect the fitting accuracy, regardless of what identification method used. As the identification results are used to verify the performance of proposed spatial filtering approach, LSRF is not irreplaceable in the framework, if any other methods that can achieve higher accuracy.

4 Simulation Study

4.1 Setting up of Synthetic Video

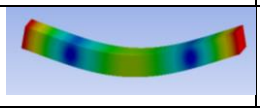
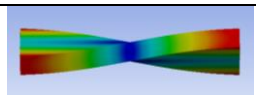
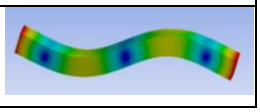
To verify the proposed method, a simulated dataset is obtained by a synthetic video footage in MATLAB. Fig. 7 shows a reference image where the black-white stripe pattern is painted on one face of a free-free beam. The image is blurred by a 5×5 PSF kernel and the region of interest has 25×1000 pixels. The colour depth is set as 8-bits and the sampling rate is set as

1,500 fps. In addition, the resulted images are further blurred by a noise floor of 5.66 dB white noise.

Fig. 7 One representative frame in synthetic video footages

The oscillation of the pixel intensity was simulated in accordance with the response of dynamic system, in which four damping signals were set to simulate the response of free-free beam under impact excitation. The characteristics are detailed in Table. 1, in which the rigid body mode at 3 Hz with amplitude of 1/2 pixel is also included to simulate real test sensoria when the beam lay on a soft base. The amplitudes of other three modes were set lower by a factor of ten, corresponding to a SNR that varies by 20 dB, which allows the performance of proposed algorithm to be evaluated sufficiently under different SNRs [47] so that it can be reliable applied to process images captured by high speed camaras.

Table. 1 Setting of dynamic characteristics.

Modes	Rigid	1st	2nd	3rd
Frequency	3 Hz	73 Hz	277 Hz	457 Hz
Amplitude	1/2 pixel	10/256 pixel	1/256 pixel	0.1/256 pixel
SNR	9.7 dB	-11.7 dB	-31.8 dB	-51.8 dB
Mode shape	Rigid-body			

An array structure $A_s(x, y)$ constructed by applying edge detection on reference image of synthetic video. As shown in Fig. 8 (a), two white rows represent one value in $A_s(x, y)$ that are 2000 selected pixel sensors of array, data matrix $\boldsymbol{u}(t)$ is constructed to be of 2000×6000. Fig. 8 (b) shows a typical signal of $\boldsymbol{u_i}(t)$ at pixel $i = 4$. As the pixel is located at the red rectangle in Fig. 8 (a), the selected pixel is likely to have the highest output as the three modes all produce the highest responses. However, the spectrum of the signal in Fig. 8 (b) shows just modal response at the 1st mode at 73Hz, but it is not possible to find the responses from two high orders of modes at 277Hz and 457Hz. This shows that the noise level of the raw signal is too high, making it difficult to resolve the two higher modes directly.

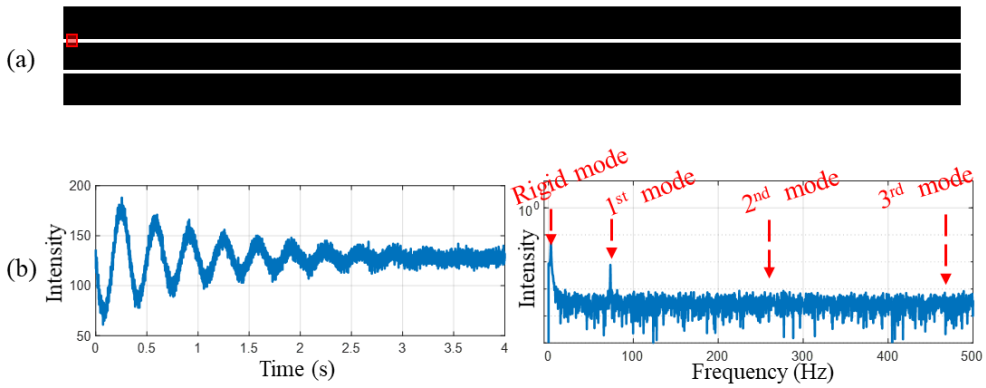


Fig. 8 Information in simualted footages, includes (a) sensor array $A_s(x, y)$ obtained by edge detection (b) intensity signals of one array element and its spectrum

4.2 Spatial Filtering and Modal Identification

Performing the procedures described in Section 3.3.1, the results can be seen from Fig. 9, every modal displacement $\hat{z}_i(t)$ is strengthened individually based on its corresponding weight vector. Furthermore, the power of modal displacement $P(w_i)$ is increasing when the weight vector is approaching to the pre-set mode shape. The convergence rate is related to the number of the nodes, and the error of mode shape is mainly affected by the noise level. For example, for the first and the second mode, the correct nodes can be always found with insignificant errors, that means the error of node location is under one pixel. For the third mode shape, the average error $Err(\boldsymbol{\varphi}_i)$ is up to 8.32% in ten sets of Monte Carlo test, but it still allows capturing the main profile features and therefore makes the amplitude at 457Hz being clearly visible. The estimated result shows smoothness is because the used sin-based function is smooth originally. The estimate error exists on the location of nodes, as well as on different curvature between estimated mode shape and real value.

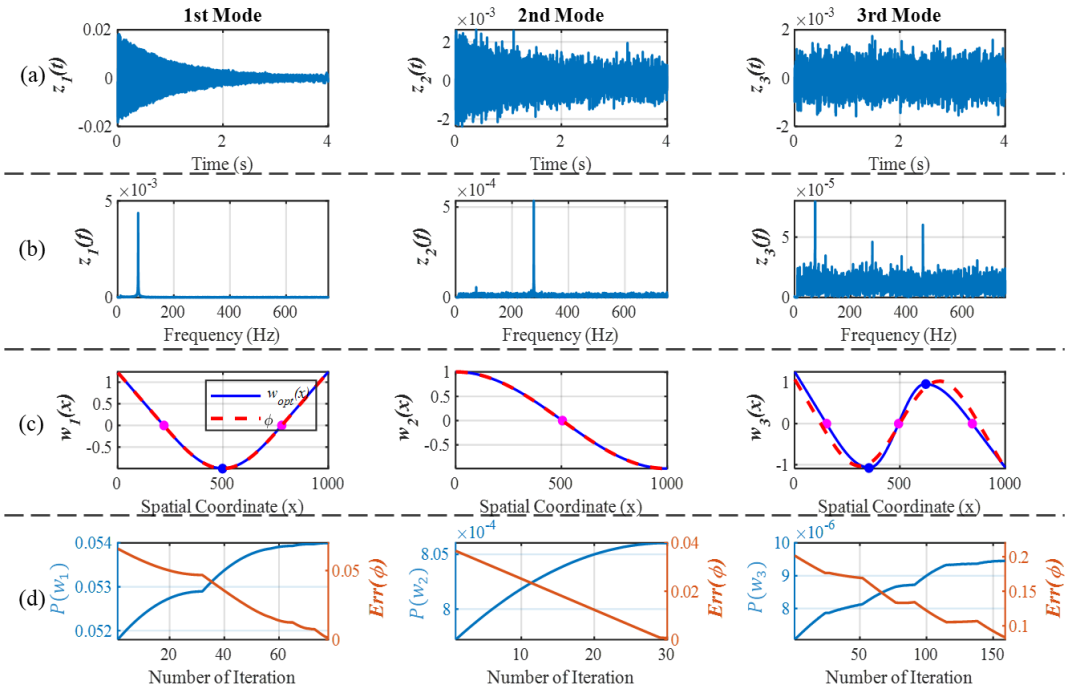


Fig. 9 Three obtained modes using spatial filtering (a) output signal (b) spectrum of output signal (c) optimum weight vector $\boldsymbol{w}_{opt}$ and pre-set mode shape $\boldsymbol{\varphi}_i$ (d) signal power $P(w_i)$ and the error of estimated mode shape $Err(\boldsymbol{\varphi}_i)$ changes with iteration

LK optical flow as a typical window-based method is compared with spatial filtering using the same video footage. 20 windows with size of 10×10 pixels are selected uniformly on image. The FRFs from spatial filtering and LK optical flow are used as the system response to identify the modal parameters using LSRF, respectively. Fig. 10 shows the stabilization diagrams, it can be observed that both methods can identify the first mode stably. However, when the amplitude decreases under 1/256 pixel in the second order, the stability of LK method declines greatly. In the third order, the LK method could not already identify the mode anymore, but spatial filtering could still get a good result.

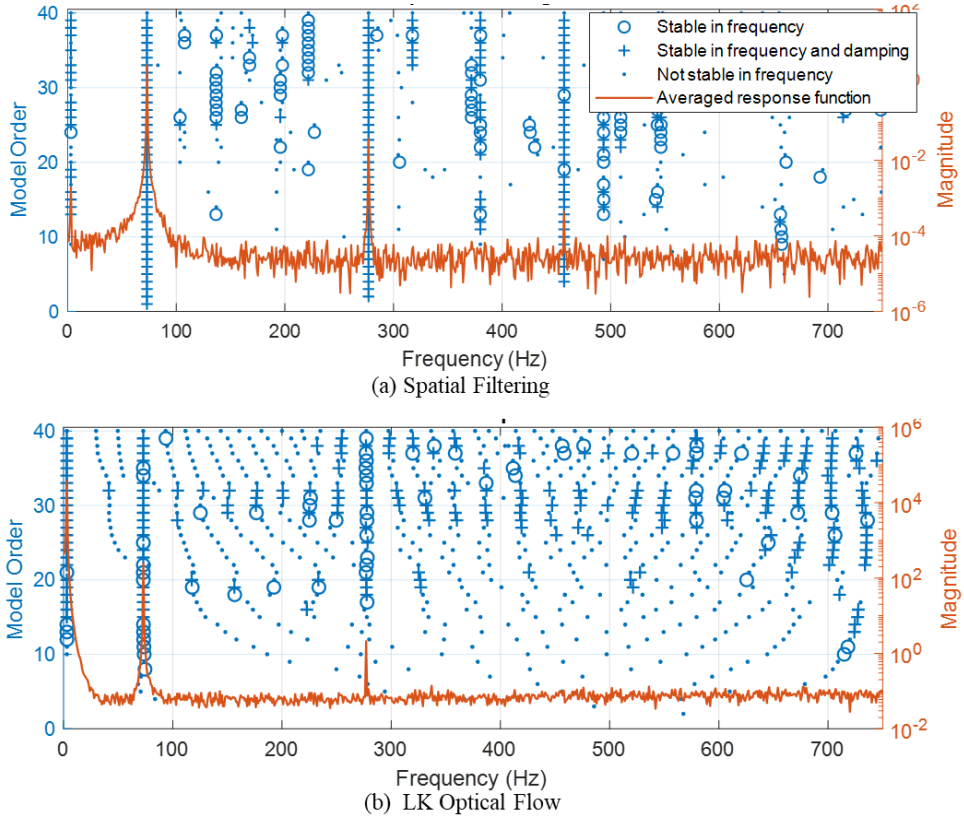


Fig. 10 Stabilization diagrams from results of (a) Spatial filtering and (b) LK optical flow, respectively

Fig 11 compares the identified natural frequencies and damping ratio. The signals from spatial filtering can successfully identify the third mode with a SNR of -51 dB, while LK optical flow cannot identify it. Moreover, the identification error of spatial filtering is smaller than that of LK optical flow method. Since video footage is the same and both results are identified by LSRF, the difference of them depends thoroughly on SNR of signals obtained by spatial filtering and LK optical flow.

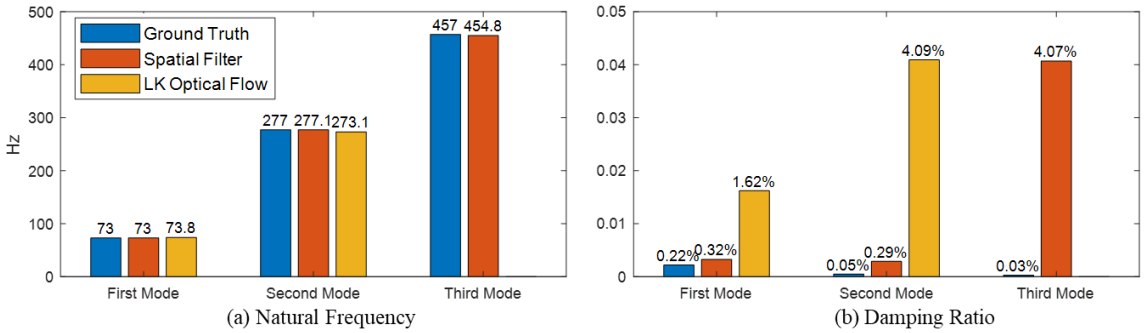


Fig 11 Natural frquency and damping ratio identified using Spatial filtering and LK optical flow

4.3 Robustness of Mode Shape Estimation

To verify the robustness of the proposed method, it was applied to match different mode shapes which were created by setting the beam with different boundary conditions with different elasticities of end supports. The initial value $\boldsymbol{w}_{ini}$ is still the same in three cases, however, the optimizer $\boldsymbol{w}_{opt}$ can approach to the different mode shapes $\boldsymbol{\varphi}$ through iterations, as shown by the first two modes in Fig 12. Although when SNR is too low in the third mode (-51.8 dB), the

error of mode shape estimate turns visible.

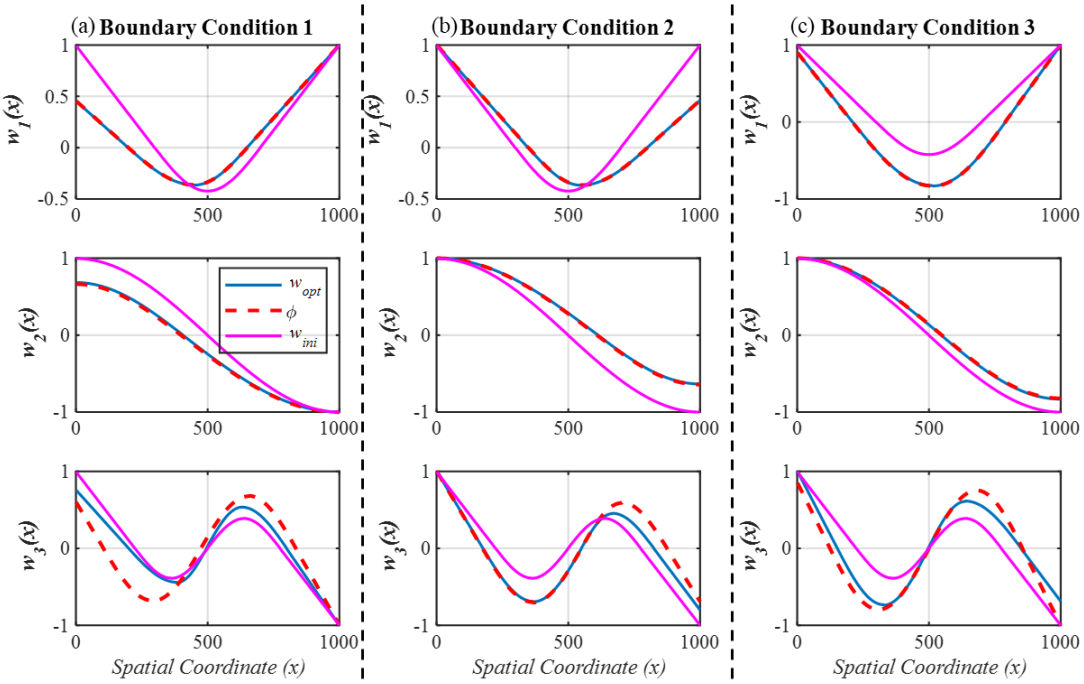


Fig 12 $\boldsymbol{w}_{opt}$ estimation under different boudary conditions (a), (b) and (c)

It demonstrates that the algorithm relies on the prior knowledge of solution forms i.e., standing waves, but the specific parameters of solutions are determined by measuring the change in signal power. However, due to the limitation of employed sinusoid function, the formulated weight vector is mainly to obtain the location of nodes. When the curvature of actual mode shape increasingly differs from sinusoid function, the error will be larger. Therefore, the robustness for more complex applications is required to improve.

4.4 Verification of Array Gain

Another simulation is implemented to examine array gain G_a of Eq. (11) under different number of pixels m. Fig. 13 shows G_a increase trend as the number of array elements increase from 400 to 4000 with a step of 400, which was obtained when the input noise is 5.6dB and other parameters remain unchanged for the synthetic datasets underlying.

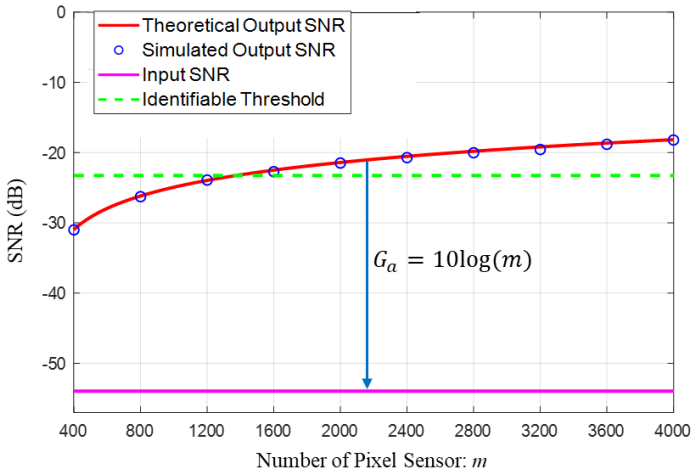


Fig. 13 SNR gain of spatial filtering with respect to number of sensor elements

According to Eq.(11), the distance between magenta and red line is exactly the array gain G_a. That shows a great consistency between theory and simulation. It demonstrates that increasing array elements is an effective way to improve output SNR. It also explains how the third mode of 1/2560 is extracted, because signal is enhanced by about 2000 times that results in its SNR excesses the identifiable threshold. In the synthetic image, ROI has a total of 25 rows of pixels, but only 2 rows at the centre of edge are utilised. In practice, without changing any other shooting conditions, just spraying more high-contrast patterns can add more array elements effectively.

5 Experiment Validation

5.1 Experimental Setup

To further verify the proposed method, an experimental study was carried out, which measures the vibration responses of a fee-free rectangular beam with the dimension of $284 \times 20 \times 75$mm and stainless-steel. The layout of the test setup is shown in Fig. 14. Two flexible porous material blocks are used to support the beam, avoiding too large amplitudes of the rigid mode. Two white strips were painted on one surface of the beam as illustrated in Fig. 14 (b). A high-speed camera was focused on the painted surface. An LED light provides a high luminous intensity and a uniform lighting. When the beam is impacted by an impact hammer, video clips were captured at a sampling rate of 14,000 fps and a resolution of 96×1024 (height×width) for multiple tests. In the meantime, the acceleration signal was also collected by a piezoelectric accelerometer for benchmarking with the results with camera-based measurements. In addition, a finite element model (FEM) was built in ANASYS software to mainly validate the mode shapes obtained by the proposed method.

5.2 Results and Analysis

Fig. 14 (c) shows the blurring effect by amplifying a small region of image in Fig. 14 (b), which validates the analysis made based the simulated image in Fig. 4. The intensity profile Fig. 14 (d) also agrees with the predicted $F(x)$ in Fig. 4. Besides, Fig. 14 (e) shows the intensity at pixel dotted in Fig. 14 (d), which is analogous to the distortion of simulated signal in Fig. 5(b). Fig. 13 (f) shows the spectrums from the outputs of three pixels, except for first mode at around 1261 Hz, all other modes are submerged in noise floor.

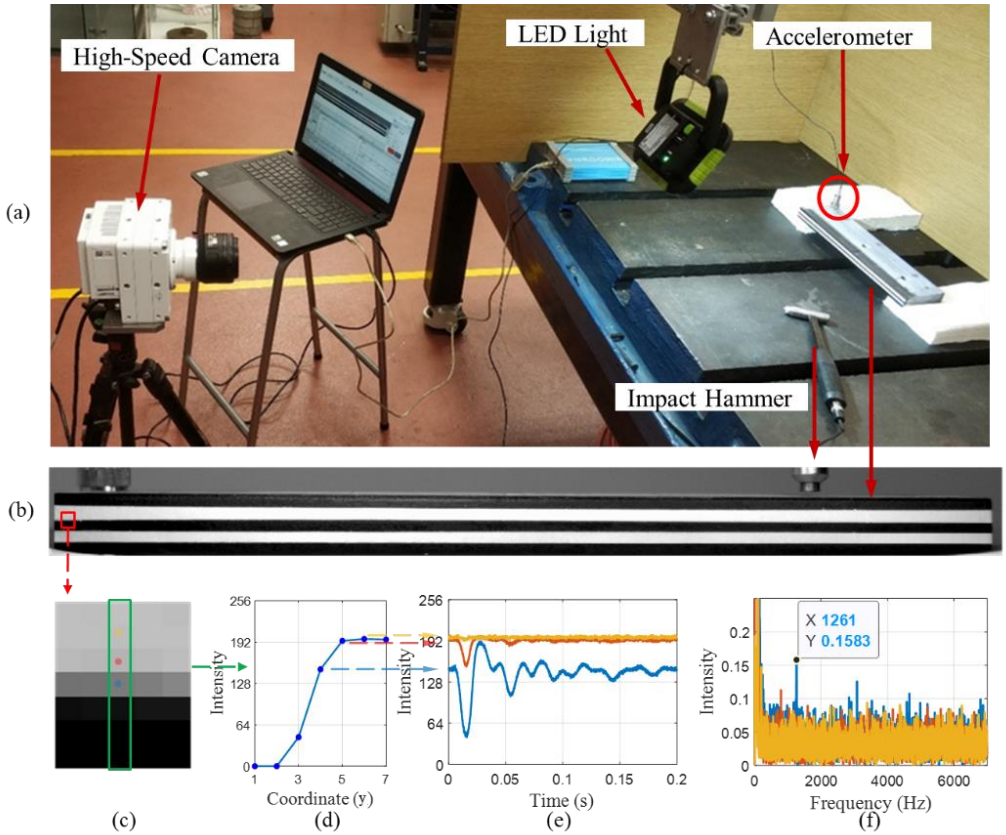
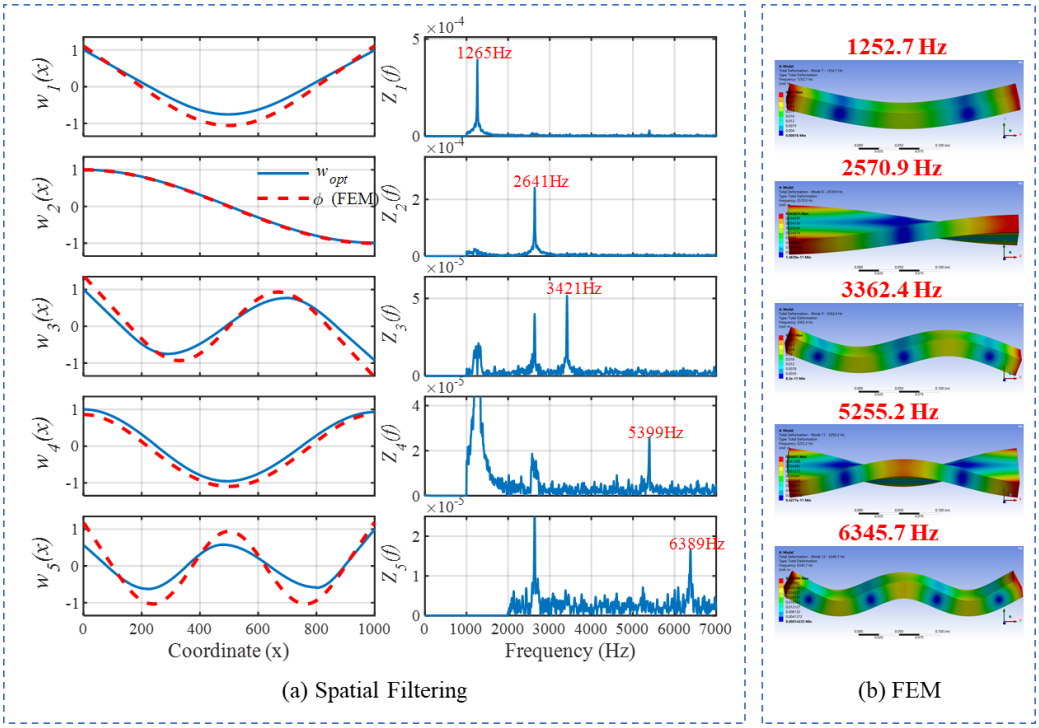


Fig. 14 Experiment data, (a) setup of experiment (b) one frame in experimental video, (c) blurred edge in red rectangle, (d) one column of blurred image, (e) intensity signals of three pixels in that column (f) spectrum

Applying the edge detection to the reference image of Fig. 14(b), according to analysis in Section 3.1.2, 4000 pixels are selected to formulate the sensor array. As a result, a data matrix of 4000×3000 is constructed as $\boldsymbol{u}(t)$, meaning that the temporal sample points are 3000 which covers the entire transient responses caused by the impact.

Comparing the proposed method with FEM, Fig. 15 shows the estimated $\boldsymbol{w}_{opt}$ and $\hat{\boldsymbol{z}}_i(t)$ generally match the prediction in FEM. The beam can show both bending modes and the twist modes when hitting on a conner of the beam. These modes can be observed by the 3D mode shapes shown in Fig. 17. However, when shooting one face (2D perspective of the beam it cannot fully separate the twisting (the 2nd and 4th) modes from bending (1st, 3rd and 5th) modes. As a result, mode mixing can occur, especially, the modes with stronger response in low frequency range will be more likely to appear in the high frequency modes. As shown in Fig. 17, the 2nd mode (twisting) appears in the 3rd and 5th mode (bending), and the 1st mode (bending) shows up in the 4th mode (twisting). However, such model mixing can be easily recognised and separated out as their frequency values are the same across different modes. Furthermore, multiple tests have shown that $\boldsymbol{w}_{opt}$ always converges in a small range, and the selection of initial weight vector does not affect the final result.



(a) Spatial Filtering (b) FEM

Fig. 15 Comparison of vibration mode identification result from (a) spatial filtering and (b) FEM

In addition, the resulted displacement spectrum is compared with the LK optical flow and the result of accelerometer in Fig. 16.The spectrum of spatial filtering is a sum of five outputs in $\widehat{\boldsymbol{Z}}(f)$. It can be seen that the peaks of all five modes are detectable in spatial filtering, however, LK optical flow cannot detect the fouth and fifth modes due to its limited noise suppression capability.

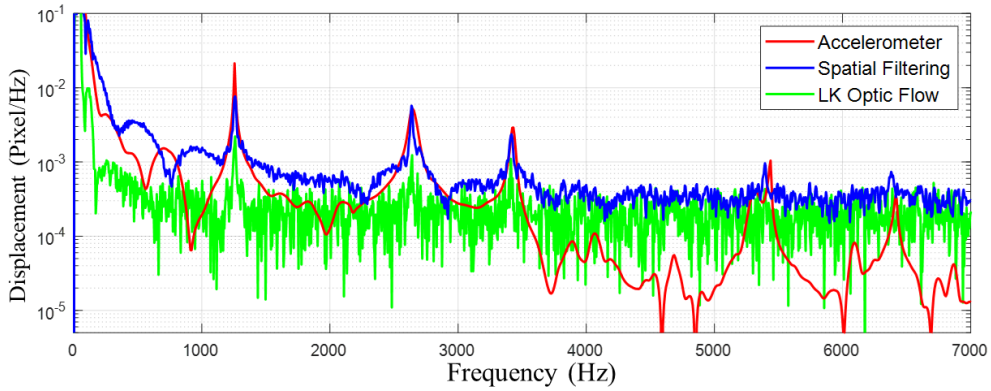


Fig. 16 Comparison of spectra obtained from accelerometer, spatial filtering and LK optical flow.

The LSRF [46] is still used to identify the modal parameters. The results from accelerometer, spatial filtering, LK optical flow and FEM are compared in Fig. 17. The LK optical flow merely detect first three modes and miss last two modes at high frequency. By contrast, the spatial filtering detected all five modes in frequency range and highly consistent with the acceleration.

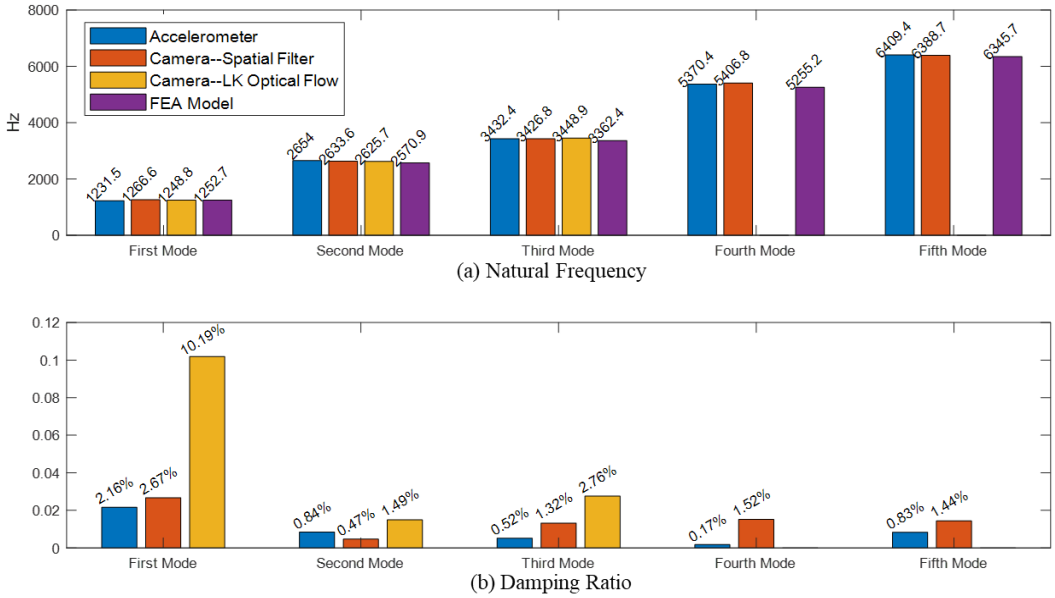


Fig. 17 Comparison of modal parameters identified in experiment: (a) the natural frequency and (b) damping ratio

The natural frequencies of acceleration, spatial filtering, optical flow, and finite element are compared together. A consistency of the five detected frequencies proves that the output of spatial filtering with high SNR can ensure that weak modes in high frequency can be identified accurately. The error of damping ratio in spatial filtering is still acceptable and it shows great advantage than LK method.

6 Discussion

The effectiveness of the proposed method has been verified by simulation and experiment. This section compares the principle, estimation precision and computational efficiency with the window-based image registration, as shown in Table. 2.

Table. 2 Comparison between the spatial filtering and window-based methods

	Proposed Spatial Filtering	**Window-based Methods**
Range of application	Mechanical vibration	No constraint of motion form
Principle	Array signal processing	Image registration
Equivalent displacement sensor	One pixel	One window (subset)
Output signal	Displacement under modal coordinate	Displacement under space coordinate
Total of pixel usage	Number of array elements: m	Total pixels in all windows
Number of outputs	Expected modal orders: n	Number of windows
Maximum gain	Number of array elements: m	Not explicit

It can be seen in Table. 2 that the primary difference between the two methods is the principles behind them. The local operator used in window-based methods is versatile for different forms of motion such as measuring strain or deformation. However, in modal analysis, all space elements oscillate around a equilibrium point at the same frequencies. Aiming at this distinctive motion, the method based on array processing can extract and processes vibration information more pertinently so as to reduce redundancy of image more efficiently.

Regarding to the computational efficiency, MATLAB is used to implement this algorithm on a laptop with Intel Core i7-7700HQ CPU @ 2.80 GHz processor and 16G GB RAM. Both the proposed method and LK optical flow are carried out to extract displacemnt signal using the same experimental dataset, the proposed method costs 29.7s totally, by contrast, LK optical flow takes 101.9 s.

7 Conclusions

This paper proposes an approach that can obtain the high-SNR modal displacement by using spatial filtering with high computational efficiency. In particular, a node/antinode searching scheme is proposed based on sinusoid-based piecewise functions, which works as an adaptive filter to match mode shape and enhance modal displacement. It shows that the proposed method can identify the weak natural frequency in the high frequency range and estimate the mode shape of continuous structures. Through theoretical analysis, the influential factors of the output SNR (weight vector and number of pixel sensors) and its upper bound (max array gain) are investigated. In simulation, the theoretical array gain is verified, and a weak mode with amplitude of 1/2560 pixel is detected through enhancing its SNR by nearly 2000 times. In experimental study, all five modes within frequency range are identified accurately, with the highest frequency being up to 6388 Hz, which cannot be identified with LK optical flow method. The fundamental reason for such an improvement can be attributed to two factors: firstly, the formation of sensor array takes pixel's sensitivity to displacement into account, so that the signal quality is guaranteed in data acquisition; secondly, spatial filtering can make modal displacements experience constructive interference by utilizing the correlation of all dynamic responses. In the proposed framework, the principle of array processing replaces image registration to process signals from image, that therefore leads to improvement. However, further development is required to improve robustness and applicability for wider applications in the future, such as structures with local modes, system with uncertain boundary condition, and complex-shape discrete system.

Appendix 1: Proof of Equation (9)

The SNR of $\hat{z}_i(t)$ is the ratio between power of signal and variance of noise:

$$SNR[\hat{z}_i(t)] = \frac{E\{[(\boldsymbol{w}_i{}^T \boldsymbol{\varphi}_i)\, z_i(t)]^2\}}{E\{[\boldsymbol{w}_i{}^T \boldsymbol{\eta}(t)]^2\}} \tag{A1}$$

Eq. (A1) can be reformed into following equation,

$$SNR[\hat{z}_i(t)] = \frac{E\{(\boldsymbol{w}_i{}^T \boldsymbol{\varphi}_i)^2 z_i(t)^2\}}{E\{\boldsymbol{w}_i{}^T \boldsymbol{\eta}(t)\boldsymbol{\eta}(t)^T \boldsymbol{w}_i\}} \tag{A2}$$

Due to the expectation is with respect to the time t, other components can be taken out of expectation. Meanwhile, $\boldsymbol{R}_\eta = \boldsymbol{\eta}(t)\boldsymbol{\eta}(t)^T$ denotes the covariance matrix of noise [32].

$$SNR[\hat{z}_i(t)] = \frac{(\boldsymbol{w}_i{}^T \boldsymbol{\varphi}_i)^2 E\{z_i(t)^2\}}{\boldsymbol{w_i}^T \boldsymbol{R}_\eta \boldsymbol{w_i}} \tag{A3}$$

in which, signal power can be presented by $\sigma_{z_i}{}^2 = E\{z_i(t)^2\}$. Then because $\boldsymbol{\eta}(t)$ is Gaussian

noise, its covariance matrix is $\boldsymbol{R}_{\eta} = {\sigma_{\eta}}^2 \boldsymbol{I}$.

$$SNR[\hat{z}_i(t)] = \frac{({\boldsymbol{w}_i}^T \boldsymbol{\varphi}_i)^2}{{\boldsymbol{w}_i}^T \boldsymbol{w}_i} \frac{{\sigma_{z_i}}^2}{{\sigma_{\eta}}^2} \tag{A4}$$

According to Cauchy–Schwarz inequality,

$$SNR[\hat{z}_i(t)] \leq \frac{({\boldsymbol{w}_i}^T \boldsymbol{w}_i)({\boldsymbol{\varphi}_i}^T \boldsymbol{\varphi}_i)}{{\boldsymbol{w}_i}^T \boldsymbol{w}_i} \frac{{\sigma_{z_i}}^2}{{\sigma_{\eta}}^2} \tag{A5}$$

By cancelling the same item of ${\boldsymbol{w}_i}^T \boldsymbol{w}_i$, the maximum SNR of the output is a value independent of the weight vector.

$$SNR[\hat{z}_i(t)] \leq ({\boldsymbol{\varphi}_i}^T \boldsymbol{\varphi}_i) \frac{{\sigma_{z_i}}^2}{{\sigma_{\eta}}^2} \tag{A6}$$

Appendix 2: Proof of Equation (10)

In order to find the array gain, the input SNR of each pixel is calculated as:

$$SNR_{in} = \frac{E\left\{\left[\varphi_{ij}\, z_i(t)\right]^2\right\}}{E\{[\eta(t)]^2\}} \tag{A7}$$

The φ_{ij} can be taken out of expectation, and then substitute ${\sigma_{z_i}}^2$ and ${\sigma_{\eta}}^2$ into Eq. (A8).

$$SNR_{in} = \frac{{\varphi_{ij}}^2 {\sigma_{z_i}}^2}{{\sigma_{\eta}}^2} \tag{A8}$$

The mean power of input signal is:

$$\overline{SNR}_{in} = \frac{1}{m} \sum_{j=1}^{m} {\varphi_{ij}}^2 \left(\frac{{\sigma_{z_i}}^2}{{\sigma_{\eta}}^2}\right) \tag{A9}$$

It can be written into a form of inner product of $\boldsymbol{\varphi_i}$.

$$\overline{SNR}_{in} = \frac{1}{m} (\boldsymbol{\varphi_i}^T \boldsymbol{\varphi_i}) \left(\frac{{\sigma_{z_i}}^2}{{\sigma_{\eta}}^2}\right) \tag{A10}$$

The array gain is defined as the gain of SNR between output and input signals.

$$G_a = \frac{SNR[\hat{z}_i(t)]}{\overline{SNR}_{in}} \tag{A11}$$

Substituting Eq. (A6) and Eq. (A10) into Eq. (A11) yields

$$G_a \leq (\boldsymbol{\varphi_i}^T \boldsymbol{\varphi_i}) \frac{{\sigma_{z_i}}^2}{{\sigma_{\eta}}^2} \Big/ \frac{1}{m} (\boldsymbol{\varphi_i}^T \boldsymbol{\varphi_i}) \frac{{\sigma_{z_i}}^2}{{\sigma_{\eta}}^2} \tag{A12}$$

By cancelling all the same terms, only m remains.

$$G_a \leq m \tag{A13}$$

In addition, the analogous relationship between array gain G_a and number of array elements m can be easily found in many kinds of sensor arrays [44].

Appendix 3: Sinusoid-based Piecewise Function for a Free-free Object

The elementary shape of each mode shape can be described by a piece of sinusoidal piecewise function with parameters of all locations l_j of nodes/antinodes

$$l_j = \{0, d_1, a_1, d_2, a_2, \dots, L\} \in (0, L) \tag{A14}$$

$j \in \mathbb{N}^n$ is number of all node/antinodes, L is the length of object, d_k and a_k denotes location of nodes and antinodes respectively. Because of the constraint in Eq. (17), these parameters are coupling as follows,

$$\frac{tan(\theta_k)}{tan(\theta_k) + tan(\theta_{k+1})} = \frac{a_k - d_k}{d_{k+1} - d_k} \tag{A15}$$

θ_k denotes included angle its $tan(\theta_k)$ is the ratio of amplitude A_{j+1} to a quarter of wavelength λ_j, which is one factor of the amplitude of each piece. Hence, there is

$$h_j = \{\tan(\theta_1), \tan(\theta_1), \tan(\theta_2), \tan(\theta_2), \dots\} \tag{A16}$$

where h_j is hence defined as a coefficient dependent of a_k, d_k and d_{k+1}. Between the adjacent node and antinode l_j and l_{j+1}., the wavelength λ_j is determined by their distance of them,

$$\lambda_j = 4(l_{j+1} - l_j) \tag{A17}$$

Then considering the free-free boundary condition, a piecewise function for twist mode shapes in one-dimension can be constituted as

$$w(x) = \frac{h_j \lambda_j}{L} \cos\left(\frac{2\pi}{\lambda_j}(x - l_j) + \frac{\pi}{2}(j-1)\right) \qquad if \quad l_j < x < l_{j+1} \tag{A18}$$

in which, $x \in (0, L)$ and $w(x)$ is the weight vector to be optimized.

Likewise, a piecewise function for bending mode shape is defined with undetermined parameters l_j to be optimized, other parameters are dependent of l_j, and coupling due to continuity and smoothness condition.

$$w(x) = \begin{cases} \dfrac{\pi h_1 \lambda_1}{2L}\left(1 - \dfrac{x}{l_2}\right) & if \quad x < l_2 \\ \dfrac{h_j \lambda_j}{L} \cos\left(\dfrac{2\pi}{\lambda_j}(x - l_j) + \dfrac{\pi}{2}(j-1)\right) & if\ l_j < x < l_{j+1} \\ (-1)^n \dfrac{\pi h_n \lambda_n}{2L}(x - l_n) & if \quad x > l_{n-1} \end{cases} \tag{A19}$$

In this numerical expression, the mode shape can be represented by variable node location.

CRediT authorship contribution statement

Miaoshuo Li: Conceptualization, Methodology, Software, Investigation, Validation, Writing, Review & editing. **Guojin Feng:** Resources, Investigation, Data analytics, Validation, review

& editing. **Rongfeng Deng:** Resources, Experiment, Data analytics, Funding acquisition, - review & editing. **Feng Gao**: Resources, review & editing. **Fengshou Gu:** Resources, Methodology improvement, Supervision, Project administration, Funding acquisition, review & editing. **Andrew D. Ball:** Supervision, Methodology improvement, Project administration, Funding acquisition.

Declaration of Competing Interest

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

Acknowledgement

This work is supported by University of Huddersfield and the China Scholarship Council. The authors wish to thank the support of National Natural Science Foundation of China under Grants No. 62076029.

References

[1] 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, doi: 10.1016/j.ymssp.2016.02.011.

[2] B. F. Spencer, V. Hoskere, and Y. Narazaki, 'Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring', *Engineering*, Mar. 2019, doi: 10.1016/j.eng.2018.11.030.

[3] D. Feng and M. Q. Feng, 'Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection – A review', *Engineering Structures*, vol. 156, pp. 105–117, Feb. 2018, doi: 10.1016/j.engstruct.2017.11.018.

[4] J. G. Chen, T. M. Adams, H. Sun, E. S. Bell, and O. Büyüköztürk, 'Camera-Based Vibration Measurement of the World War I Memorial Bridge in Portsmouth, New Hampshire', *Journal of Structural Engineering*, vol. 144, no. 11, p. 04018207, Nov. 2018, doi: 10.1061/(ASCE)ST.1943-541X.0002203.

[5] J. G. Chen, A. Davis, N. Wadhwa, F. Durand, W. T. Freeman, and O. Büyüköztürk, 'Video Camera–Based Vibration Measurement for Civil Infrastructure Applications', *Journal of Infrastructure Systems*, vol. 23, no. 3, p. B4016013, Sep. 2017, doi: 10.1061/(ASCE)IS.1943-555X.0000348.

[6] A. Sarrafi, Z. Mao, C. Niezrecki, and P. Poozesh, 'Vibration-based damage detection in wind turbine blades using Phase-based Motion Estimation and motion magnification', *Journal of Sound and Vibration*, vol. 421, pp. 300–318, May 2018, doi: 10.1016/j.jsv.2018.01.050.

[7] N. Najafi and U. S. Paulsen, 'Operational modal analysis on a VAWT in a large wind tunnel using stereo vision technique', *Energy*, vol. 125, pp. 405–416, Apr. 2017, doi: https://doi.org/10.1016/j.energy.2017.02.133.

[8] M. Ozbek, F. Meng, and D. J. Rixen, 'Challenges in testing and monitoring the in-operation vibration characteristics of wind turbines', *Mechanical Systems and Signal Processing*, vol. 41, no. 1, pp. 649–666, Dec. 2013, doi: https://doi.org/10.1016/j.ymssp.2013.07.023.

[9] M. N. Helfrick, C. Niezrecki, P. Avitabile, and T. Schmidt, '3D digital image correlation methods for full-field vibration measurement', *Mechanical Systems and Signal Processing*, vol. 25, no. 3,

pp. 917–927, 2011, doi: 10.1016/j.ymssp.2010.08.013.

[10] J. Javh, J. Slavič, and M. Boltežar, ‘High frequency modal identification on noisy high-speed camera data’, *Mechanical Systems and Signal Processing*, vol. 98, pp. 344–351, 2018, doi: 10.1016/j.ymssp.2017.05.008.

[11] D. Gorjup, J. Slavič, and M. Boltežar, ‘Frequency domain triangulation for full-field 3D operating-deflection-shape identification’, *Mechanical Systems and Signal Processing*, vol. 133, p. 106287, Nov. 2019, doi: 10.1016/j.ymssp.2019.106287.

[12] W. Wang, J. E. Mottershead, T. Siebert, and A. Pipino, ‘Frequency response functions of shape features from full-field vibration measurements using digital image correlation’, *Mechanical Systems and Signal Processing*, vol. 28, pp. 333–347, 2012, doi: 10.1016/j.ymssp.2011.11.023.

[13] P. L. Reu, D. P. Rohe, and L. D. Jacobs, ‘Comparison of DIC and LDV for practical vibration and modal measurements’, *Mechanical Systems and Signal Processing*, vol. 86, pp. 2–16, 2017, doi: 10.1016/j.ymssp.2016.02.006.

[14] B. Pan, K. Li, and W. Tong, ‘Fast, Robust and Accurate Digital Image Correlation Calculation Without Redundant Computations’, *Exp Mech*, vol. 53, no. 7, pp. 1277–1289, Sep. 2013, doi: 10.1007/s11340-013-9717-6.

[15] P. Bing, X. Hui-min, X. Bo-qin, and D. Fu-long, ‘Performance of sub-pixel registration algorithms in digital image correlation’, *Meas. Sci. Technol.*, vol. 17, no. 6, pp. 1615–1621, May 2006, doi: 10.1088/0957-0233/17/6/045.

[16] Aral Sarrafi, ‘Wind Turbine Blade Damage Detection Via 3-Dimensional Phase-Based Motion Estimation’, 2017.

[17] A. J. Molina-Viedma, L. Felipe-Sesé, E. López-Alba, and F. A. Díaz, ‘3D mode shapes characterisation using phase-based motion magnification in large structures using stereoscopic DIC’, *Mechanical Systems and Signal Processing*, vol. 108, pp. 140–155, Aug. 2018, doi: 10.1016/j.ymssp.2018.02.006.

[18] J. G. Chen, N. Wadhwa, Y.-J. Cha, F. Durand, W. T. Freeman, and O. Buyukozturk, ‘Modal identification of simple structures with high-speed video using motion magnification’, *Journal of Sound and Vibration*, vol. 345, pp. 58–71, 2015, doi: 10.1016/j.jsv.2015.01.024.

[19] A. Davis, M. Rubinstein, N. Wadhwa, G. J. Mysore, F. Durand, and W. T. Freeman, ‘The visual microphone: passive recovery of sound from video’, *ACM Transactions on Graphics*, vol. 33, no. 4, pp. 1–10, Jul. 2014, doi: 10.1145/2601097.2601119.

[20] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, ‘Phase-based video motion processing’, *ACM Transactions on Graphics*, vol. 32, no. 4, p. 1, Jul. 2013, doi: 10.1145/2461912.2461966.

[21] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, ‘Performance of optical flow techniques’, *Int J Comput Vision*, vol. 12, no. 1, pp. 43–77, Feb. 1994, doi: 10.1007/BF01420984.

[22] T. Gautama and M. A. V. 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, Sep. 2002, doi: 10.1109/TNN.2002.1031944.

[23] Y. Yang *et al.*, ‘Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification’, *Mechanical Systems and Signal Processing*, vol. 85, pp. 567–590, Feb. 2017, doi: 10.1016/j.ymssp.2016.08.041.

[24] P. Poozesh, A. Sarrafi, Z. Mao, P. Avitabile, and C. Niezrecki, ‘Feasibility of extracting operating shapes using phase-based motion magnification technique and stereo-photogrammetry’, *Journal of Sound and Vibration*, vol. 407, pp. 350–366, 2017, doi: 10.1016/j.jsv.2017.06.003.

[25] J. Javh, J. Slavič, and M. Boltežar, ‘Measuring full-field displacement spectral components using

photographs taken with a DSLR camera via an analogue Fourier integral', *Mechanical Systems and Signal Processing*, vol. 100, pp. 17–27, 2018, doi: 10.1016/j.ymssp.2017.07.024.

[26] J. Javh, J. Slavič, and M. Boltežar, 'Experimental modal analysis on full-field DSLR camera footage using spectral optical flow imaging', *Journal of Sound and Vibration*, vol. 434, pp. 213–220, Nov. 2018, doi: 10.1016/j.jsv.2018.07.046.

[27] D. Gorjup, J. Slavič, A. Babnik, and M. Boltežar, 'Still-camera multiview Spectral Optical Flow Imaging for 3D operating-deflection-shape identification', *Mechanical Systems and Signal Processing*, vol. 152, p. 107456, May 2021, doi: 10.1016/j.ymssp.2020.107456.

[28] J. Javh, J. Slavič, and M. Boltežar, 'The subpixel resolution of optical-flow-based modal analysis', *Mechanical Systems and Signal Processing*, vol. 88, pp. 89–99, 2017.

[29] T. Bregar, K. Zaletelj, G. Čepon, J. Slavič, and M. Boltežar, 'Full-field FRF estimation from noisy high-speed-camera data using a dynamic substructuring approach', *Mechanical Systems and Signal Processing*, vol. 150, p. 107263, Mar. 2021, doi: 10.1016/j.ymssp.2020.107263.

[30] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer London, 2010. [Online]. Available: https://books.google.co.uk/books?id=8_2RNQEACAAJ

[31] S. U. Pillai, *Array Signal Processing*. Springer Science & Business Media, 2012.

[32] B. D. V. Veen and K. M. Buckley, 'Beamforming: a versatile approach to spatial filtering', *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, Apr. 1988, doi: 10.1109/53.665.

[33] D. R. Fuhrmann and G. S. Antonio, 'Transmit beamforming for MIMO radar systems using signal cross-correlation', *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 1, pp. 171–186, Jan. 2008, doi: 10.1109/TAES.2008.4516997.

[34] J. Litva and T. K. Lo, *Digital Beamforming in Wireless Communications*, 1st ed. USA: Artech House, Inc., 1996.

[35] J. C. Chen, Kung Yao, and R. E. Hudson, 'Source localization and beamforming', *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 30–39, Mar. 2002, doi: 10.1109/79.985676.

[36] J. Li and P. Stoica, 'Robust Adaptive Beamforming', p. 30.

[37] A. Deraemaeker, E. Reynders, G. De Roeck, and J. Kullaa, 'Vibration-based structural health monitoring using output-only measurements under changing environment', *Mechanical Systems and Signal Processing*, vol. 22, no. 1, pp. 34–56, Jan. 2008, doi: 10.1016/j.ymssp.2007.07.004.

[38] G. Tondreau and A. Deraemaeker, 'Vibration based damage localization using multi-scale filters and large strain sensor networks', 2010.

[39] J. C. Pedro and N. B. Carvalho, *Intermodulation Distortion in Microwave and Wireless Circuits*. Artech House, 2003.

[40] M. H. Rashid, *Power Electronics Handbook*. Butterworth-Heinemann, 2017.

[41] K. M. Gharaibeh, *Nonlinear Distortion in Wireless Systems: Modeling and Simulation with MATLAB*. John Wiley & Sons, 2011.

[42] A. Sleiman and A. Manikas, 'Antenna array manifold: a simplified representation', in *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*, Jun. 2000, vol. 5, pp. 3164–3167 vol.5. doi: 10.1109/ICASSP.2000.861209.

[43] R. C. Hibbeler, *Engineering Mechanics: Dynamics*. Prentice Hall, 2010.

[44] P. S. Naidu, *Sensor array signal processing*. Boca Raton, Fla.: CRC Press, 2001.

[45] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, 'Convex Optimization-Based Beamforming', *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 62–75, May 2010, doi: 10.1109/MSP.2010.936015.

[46] A. Arda Ozdemir and S. Gumussoy, ‘Transfer Function Estimation in System Identification Toolbox via Vector Fitting’, *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 6232–6237, Jul. 2017, doi: 10.1016/j.ifacol.2017.08.1026.

[47] M. Li, G. Feng, F. Gu, and A. Ball, ‘Investigating into Minimum Detectable Displacement Signal in Image-Based Vibration Measurement’, in *Proceedings of IncoME-V & CEPE Net-2020*, Cham, 2021, pp. 882–894. doi: 10.1007/978-3-030-75793-9_82.