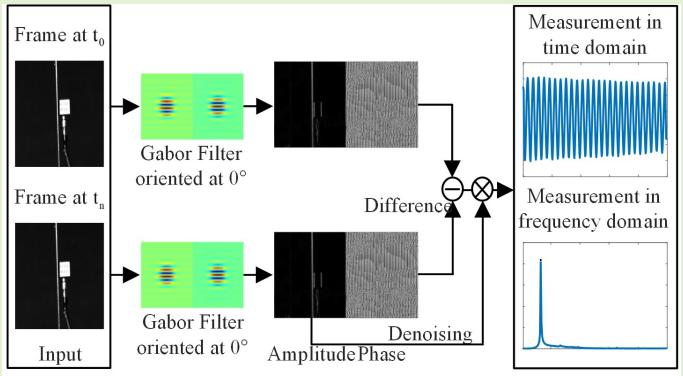
Camera-based Micro-vibration Measurement for Lightweight Structure using an Improved Phase-based Motion Extraction

Cong Peng, *Member*, *IEEE*, Cong Zeng, and Yangang Wang, *Member*, *IEEE*

***Abstract*—Video cameras are capable to collect high-density spatial data as each pixel is a sensor. An issue is that the video data is difficult to convert into the motion information, such as displacement. This paper presents a novel camera-based vibration measurement methodology for the lightweight structure. First, the phase-based motion extraction method is improved by directly transforming the phase variations into the displacement without the computation of phase gradient. The displacement is then converted into acceleration information to describe the vibration. Moreover, the proposed vibration measurement is investigated for a lightweight structure without attached sensors to change the physical properties. The measurements of the acceleration signals from camera and accelerometer are compared for verification. The experimental results for the effect of mass-loading, different camera parameters and down-sampling on the measurement validate the effectiveness of the proposed method.**



***Index Terms*—Vibration measurement, non-contract measurement, video camera, phase-based optical flow, micro-vibration.**

I. Introduction

Vibration is an important factor affecting the working condition and service life of the equipment. Vibration measurement of equipment can identify system parameters, monitor the working condition of equipment and diagnose the fault of industrial equipment. Therefore, vibration testing analysis occupies an increasingly important position in the field of engineering [1][2]. However, it is still a great challenge to develop a vibration measurement system with strong adaptability, high automation and high accuracy.

From the existing vibration measurement technology, there are two kinds of measurement methods, the contact measurement and the non-contact measurement [3]. The former one requires physical-attached sensors, such as accelerometers and LVDT displacement sensors, etc. The sensor is arranged on the measuring object according to certain rules and connected with the upper system. The sensor is used to collect vibration data, and the upper system is used to process data and display them in the form of reports and curves. However, the contact measurement causes mass-loading effects, and can only measure the corresponding signal of a single point without providing full-field spatial resolution [4]. In addition, for some large structures, handling all the wiring and meters is time-consuming and labor-intensive.

Another non-contact measurement method usually depends on electromagnetic radiation to transmit information. This method is different from the traditional contact measurement and can obtain different dimension information without changing the configuration and installation of the existing equipment. For example, the laser vibrometers used for the vibration measurement do not need sensors to install on the structure, thus avoiding the mass-loading effects [5]-[7]. As an alternative non-contact method, digital cameras in conjunction with computer vision techniques, has been utilized to determine the displacements of structures [8]-[12]. Compared to laser vibrometers, the digital cameras have the advantages of low cost, high automation and suitable for simultaneous full-field measurement.

There are various camera vision-based measurement method. The 3D digital image correlation method (3D DIC) [13]-[17] is widely used in previous work. It needs to set speckle patterns or high contrast markers on the surface of the structure. The digital

Manuscript received Oct 2019. This work was supported in part by the National Key R&D Program of China (No. 2018YFB1403901), the National Natural Science Foundation of China (No. 61703203 and 61806054), the Fundamental Research Funds for the Central Universities under Grant NS2019020, and the Natural Science Foundation of Jiangsu Province (No. BK20170812 and BK20180355). This work was done when Cong Peng was a visiting scholar at Southeast University, Nanjing, China. (*Corresponding author: Yangang Wang*)

Cong Peng and Cong Zeng are with the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106, China. (e-mail: pengcong.2006@163.com; zengcong718@nuaa.edu.cn)

Yangang Wang is with the School of Automation, Southeast University, Nanjing, 210096, China.(e-mail: ygwangthu@gmail.com)

speckle image before and after surface deformation can be acquired by image equipment, and the 3D full-field vibration information of the object in a large range can be measured. Two high-speed cameras were used in [13] to acquire vibration speckle images of the drying cabinet panel by using 3D digital correlation method, and the 3D geometry and displacement of the structure surface were obtained. Two high-definition, high-frequency cameras were adopted in [16] to obtain 3D vibration speckle image sequence by 3D digital correlation method, and the measurement of non-contact, full-field vibration modal parameters and strain of the structure was realized. In [15], an iPhone's camera is combined with digital image correlation technique to track a circular target, the arbitrary 3-D structural displacement is monitored. Although 3D digital correlation method provides high-precision smooth pattern shapes, it needs two high-definition cameras to collect data, and the vibration information needs to be obtained by a large number of matching calculations. Moreover, the rigorous condition is required for computer systems and cameras, and surface preparation is necessary, hence it is not practical. Another vision-based measurement method, optical flow method can extract motion information according to the natural structure of the image without any surface preparation. The very first attempts to estimate motion fields in videos can be traced back to Lucas-Kanade (LK) [18] and Horn-Schunck (HS) [19] methods. These optical flow methods are based on image intensity information. The intensity-based optical flow has been widely used in vibration measurement of structures [20]-[23]. However, the intensity-based optical flow method is sensitive to noise and disturbunces. When the illumination of the scene changes abruptly, it leads to errors in the estimated motion field. For the sub-pixel motion, the accuracy measured by the intensity-based optical flow method is relatively low.

More recent methods make use of the phase-based optical flow method introduced by Fleet *et al.* [24] to estimate the motion according to the phase field of the image. Phase-based optical flow can avoid the above-mentioned surface preparation with less computational burden. Compared with the intensity information, the phase is more robust with respect to the variations of image brightness and contrast [25]. Furthermore, for low amplitude sub-pixel motion, the effect of signal extraction can be significantly improved by denoising the phase signal. Hence the phase-based optical flow outperforms 3D DIC and other intensity-based optical flow techniques, especially in micro-vibration measurement. Chen *et al.* [26] combined the phase-based optical flow method with the motion amplification algorithm to extract resonance frequencies and motion deformation modes of several laboratory-scale reference structures. Yang *et al.* [27] combined phase-based optical flow method with motion amplification theory and blind source separation to generate unsupervised learning methods for modal analysis. Sarrafi *et al.* [28] combined the phase-based visual principle to carry out structural health monitoring of wind turbine blades based on vibration. However, the phase based optical flow needs the computation of flow vectors and phase gradient, which aggravates the time complexity and noise sensitivity. Moreover, it is noted that the image capture quality is important for the camera-based measurement. The references mentioned above did not specify the image quality requirements of phase-based optical flow method.

In this work, we propose a novel camera-based vibration measurement methodology, which based on an improved phase-based displacement extraction. The structural displacement is extracted directly according to the denoised local phase variations without calculating the optical flow vectors and phase gradient. The extracted displacement is used to compute the vibration signals in the time domain and frequency domain. Finally, a series of measurement experiments are carried out to verify the accuracy and applicable conditions of the proposed camera-based vibration measurement method. Owing to no need for gradient calculation, not only the computation cost is reduced, but also the robustness of the motion computation are improved.

II. Phase-based Displacement Extraction

In this section, we suggest an improve phase-based motion extraction method, which is based on spatial filtering the images using 2D Gabor filter. The local phase and local amplitude information of the images are extracted. Local phase gives a way to compute local motion. Local amplitude is used to denoise the phase signal. Different from the conventional phase-based optical flow method, the displacement of each pixel is computed directly according the local phase variations without the computation of flow vectors and phase gradient.

A. 2D Gabor Filter for Image Spatial Filtering

The proposed phase-based motion extraction method is inspired by the Fourier transform shift theorem, which indicates that motion in spatial domain leads to the phase change in frequency domain. However, Fourier transform only reflects the overall characteristics of the signal. The phase and the amplitude of Fourier transform correspond to the position and intensity of global motion. Therefore, Fourier transform is used to estimate global motion.

However, localization analysis is impossible for the Fourier transform, then the Gabor transform is introduced. The Gabor transform is generated based on the simulation of the human visual system. By simulating the human visual system, retinal imaging can be decomposed into a set of filtered images, each of which can reflect intensity variations in frequency and direction in a local range. The fundamental of the Gabor transform is the design of the Gabor filter. In the spatial domain, a 2D Gabor filter is a sinusoid function windowed by a Gaussian envelope. The sinusoid function is the tuning function while the Gaussian envelope is the window function. The 2D Gabor function can be expressed as,

$$g\left(x,y;\lambda,\theta,\psi,\sigma,\gamma\right)=\exp\left(-\frac{x_\theta{}^2+\gamma^2 y_\theta{}^2}{2\sigma^2}\right)\exp\left(i\left(2\pi\frac{x_\theta}{\lambda}+\psi\right)\right) \tag{1}$$

where λ is the wavelength of the sinusoids and ψ is the phase offset of tuning function. γ is the spatial aspect ratio, which determines the shape of the Gabor function. σ represents the standard deviation of the Gaussian function, which determines the size of the acceptable region of the Gabor filter kernel. Moreover, the spatial variables x_θ and y_θ also contain directional information, and can be presented as,

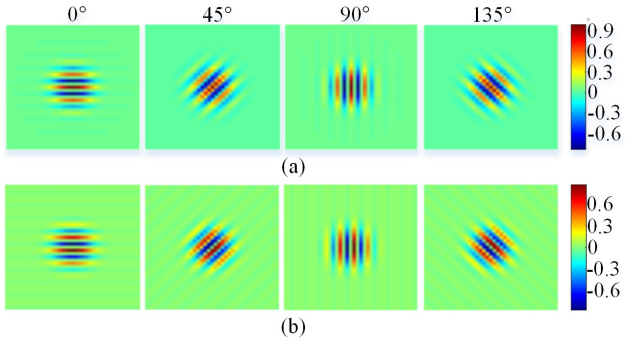


Fig. 1. Gabor filter in four different orientations. (a) The real parts. (b) The imaginary parts.

$$x_\theta = x\cos\theta + y\sin\theta \tag{2}$$

$$y_\theta = -x\sin\theta + y\cos\theta \tag{3}$$

where x and y denote spatial location. $\theta \in (0°, 360°)$ represents the direction of the parallel stripe in the Gabor filtering kernel.

The 2D Gabor function can also be expressed as a simpler form of the real part G_θ and the imaginary part H_θ,

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = G_\theta + iH_\theta \tag{4}$$

where

$$\begin{cases} G_\theta = \exp\left(-\dfrac{x_\theta^{\,2} + \gamma^2 y_\theta^{\,2}}{2\sigma^2}\right)\cos\left(2\pi\dfrac{x_\theta}{\lambda} + \psi\right) \\ H_\theta = \exp\left(-\dfrac{x_\theta^{\,2} + \gamma^2 y_\theta^{\,2}}{2\sigma^2}\right)\sin\left(2\pi\dfrac{x_\theta}{\lambda} + \psi\right) \end{cases} \tag{5}$$

Fig. 1 shows the real and imaginary pairs of 2D Gabor filter oriented at $0°$, $45°$, $90°$ and $135°$.

A video consists of a sequence of image frames with the image intensity $I(x, y, t_0)$ at the specific time t_0. Then a 2D Gabor filter is convolved with the image intensity, and specific frequency representation of each frame in a particular direction θ can be computed. This can be expressed with a simple form,

$$A_\theta(x, y, t_0)\exp\left(i\phi_\theta(x, y, t_0)\right) = I(x, y, t_0) \otimes (G_\theta + iH_\theta) \tag{6}$$

where G_θ and H_θ are the real and imaginary parts of the filter, respectively. $A_\theta(x, y, t_0)$ denotes the spatial local amplitude and $\phi_\theta(x, y, t_0)$ denotes the spatial local phase.

Using the features of direction and frequency selectivity of Gabor filter, the local amplitude and local phase in different directions can be extracted. Then the displacement signals are obtained according to the relationship between local phase and displacement.

B. Displacement Extraction and Noise Handling

Fig. 2 shows the whole workflow diagram of phase-based displacement extraction. Taking the vibration video of steel rod as an example, the 2D Gabor filter oriented at $0°$ is used to extract the horizontal amplitude and phase information of steel rod. The amplitude is used to perform a weighted spatial Gaussian blur to compute the displacement signal. The spatial Gaussian blur is on the phase difference between the frame at t_0 and the frame t_n.

The displacement extraction is then presented in details.

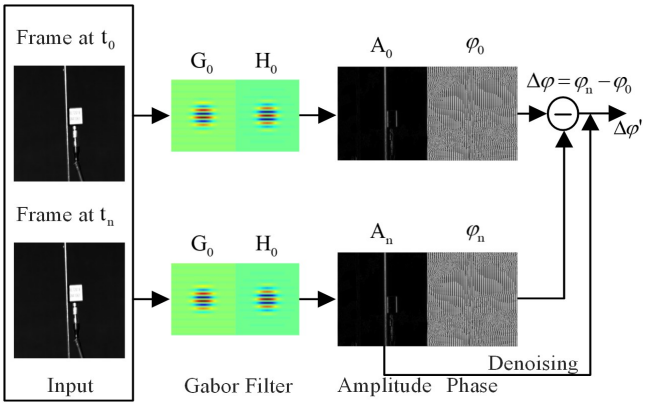


Fig. 2. The workflow diagram of the phase-based displacement extraction.

Assuming that the time interval between the two consecutive frames of the video is Δt, and local motion $(\Delta x, \Delta y)$ occurs at the spatial location (x, y). Define the first frame as the reference frame with the image intensity $I(x, y, t_0)$ and the second as the motion frame with the image intensity $I(x + \Delta x, y + \Delta y, t_0 + \Delta t)$. By convolution operation with 2D Gabor function, image intensity $I(x, y, t)$ is transformed into frequency domain $F(x, y, t)$,

$$F(x, y, t) = I(x, y, t) \otimes g(x, y; \lambda, \theta, \psi, \sigma, \gamma) \tag{7}$$

It also can be calculated by the integral,

$$F(x, y, t) = \int\int_{-\infty}^{+\infty} I(u, v, t)\, g(x - u, y - v; \lambda, \theta, \psi, \sigma, \gamma) du dv \tag{8}$$

Take the extraction of horizontal motion as an example. Select $\theta = 0^\circ$, according to (2) and (3), the spatial variables can be expressed as $x_\theta = x$ and $y_\theta = y$. So the reference frame can be transformed as,

$$F(x, y, t) = \int\int_{-\infty}^{+\infty} I(u, v, t) e^{\left(-\frac{(x-u)^2 + \gamma^2 (y-v)^2}{2\sigma^2}\right)} e^{i\left(2\pi\frac{x-u}{\lambda} + \psi\right)} du dv \tag{9}$$

And the motion frame can be transformed as,

$$F(x, y, t_0 + \Delta t) = \int\int_{-\infty}^{+\infty} I(u + \Delta x, v + \Delta y, t + \Delta t) e^{\left(-\frac{(x-u-\Delta x)^2 + \gamma^2 (y-v-\Delta y)^2}{2\sigma^2}\right)} \cdot e^{i\left(2\pi\frac{x-u-\Delta x}{\lambda} + \psi\right)} d(u + \Delta x) d(v + \Delta y) \tag{10}$$

Equations (9) and (10) are rearranged by putting the phase term variables which are independent of integration variables outside the integral and simplify the equation as,

$$F(x, y, t_0) = e^{i\left(-2\pi\frac{x}{\lambda}\right)} \int\int_{-\infty}^{+\infty} I(u, v, t) e^{\left(-\frac{(x-u)^2 + \gamma^2 (y-v)^2}{2\sigma^2}\right)} e^{i\left(2\pi\frac{u}{\lambda} + \psi\right)} du dv \tag{11}$$

$$F(x, y, t_0 + \Delta t) = e^{i\left(-2\pi\frac{x + \Delta x}{\lambda}\right)} \int\int_{-\infty}^{+\infty} I(u + \Delta x, v + \Delta y, t + \Delta t) \cdot e^{\left(-\frac{(x-u-\Delta x)^2 + \gamma^2 (y-v-\Delta y)^2}{2\sigma^2}\right)} e^{i\left(2\pi\frac{u}{\lambda} + \psi\right)} du dv \tag{12}$$

Now the phase terms being integrated of (11) and (12) are both $e^{i\left(2\pi\frac{u}{\lambda}+\psi\right)}$. Then, compute the phase angle of (11) and (12) by the final definite integral φ',

$$\varphi\left(F\left(x,y,t_0\right)\right)=-2\pi\frac{x}{\lambda}+\varphi' \tag{13}$$

$$\varphi\left(F\left(x,y,t_0+\Delta t\right)\right)=-2\pi\frac{x+\Delta x}{\lambda}+\varphi' \tag{14}$$

According to (13) and (14), the phase difference is presented as,

$$\varphi\left(F\left(x,y,t_0+\Delta t\right)\right)-\varphi\left(F\left(x,y,t_0\right)\right)=2\pi\frac{\Delta x}{\lambda} \tag{15}$$

It is easy to infer from (15) that the horizontal motion Δx is proportional to the phase difference.

However, the conventional phase-based optical flow computes the phase gradient to estimate the motion field. Assuming $\Delta y=0$ by only considering the horizontal direction, the phase-based optical flow is,

$$\varphi\left(F\left(x,y,t_0+\Delta t\right)\right)-\varphi\left(F\left(x,y,t_0\right)\right)=\frac{\partial\varphi}{\partial x}\Delta x \tag{16}$$

Comparing (15) and (16), it is noted that $\partial\varphi/\partial x$ in the existing phase-based optical-flow is replaced by a constant value $2\pi/\lambda$. There is no need to compute the phase gradient in the whole image. The horizontal motion Δx is directly obtained via the proportion to the temporal phase differences. Moreover, the motion in the vertical directions can also related to phase differences via changing the orientation of Gabor filter to $\theta=\pi/2$.

However, the noise of the input image sequence can interfere the phase signal, and affect the final displacement extraction performance. As noise is always low amplitude phase signals, to reduce these meaningless signals, local amplitude is used to perform a weighted spatial Gaussian blur on the phases. For the Nth frame, the weighted phase signal φ'_N can be computed as,

$$\varphi'_N=\left(\varphi_N A_N\right)\otimes h\left(x,y\right)/\,A_N\otimes h\left(x,y\right) \tag{17}$$

where A_N and φ_N represent the amplitude and phase signal of the Nth frame, respectively. $h\left(x,y\right)$ is the 2D Gauss function, which can be expressed as,

$$h\left(x,y\right)=\exp\left[-\left(x^2+y^2\right)/\rho^2\right] \tag{18}$$

The standard deviation of Gaussian filter ρ represents the spatial domain filter widths. The larger the standard deviation, the wider 2D Gauss image and the better filtering effect is obtained. This step incurs a small computational cost, but improves the signal-to-noise ratio (SNR) and lowers the noise floor, which can better reflect the real signal.

The acceleration signal is recovered by imposing 1D Laplacian of the Gaussian (LOG) operator on the displacement signal. The 1D LOG operator is the combination of gaussian function and Laplacian function. The kernel function of the 1D LOG operator is as follows,

$$LOG=\frac{\partial^2}{\partial x^2}G_\sigma\left(x\right)=\frac{x^2-2\sigma^2}{\sigma^4}e^{-\frac{x^2}{2\sigma^2}} \tag{19}$$

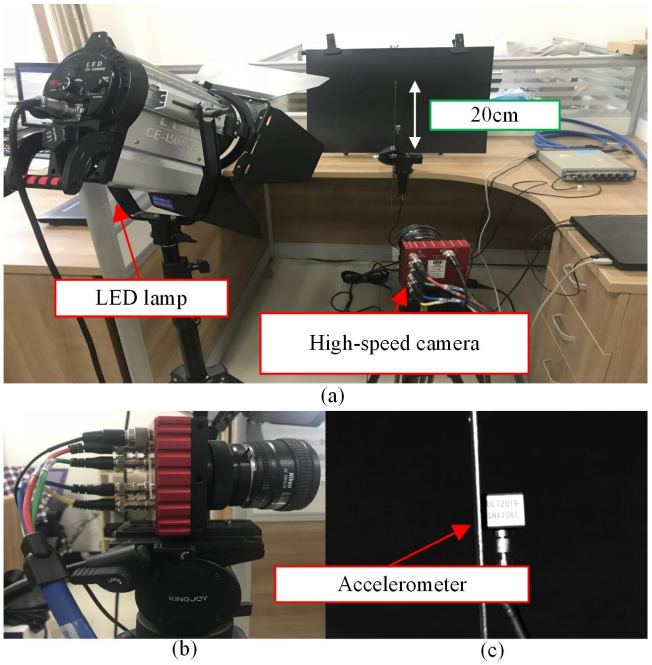


Fig. 3. Experimental setup. (a) The side view of the experimental scene. (b) The high-speed camera with 20mm lens. (c) The screenshot of the camera.

where σ represents the standard deviation of the Gaussian function, x represents the displacement.

The displacement signal is denoised by gaussian function $G_\sigma\left(x\right)$, and the Laplacian function takes the second derivative of the denoised displacement signal. The second derivative of the displacement signal is the acceleration signal.

III. Experimental Setup

The vison measurement system is made of a high-speed camera (FLARE 12M180MCX), a single lens reflex camera (SLR) camera (Canon EOS 80D), an impact hammer, an accelerometer, light sources and a computer for data storage.

Fig. 3 shows the experimental setup. The high-speed camera is fixed by a tripod and adjusted to the appropriate viewing position. Meanwhile, the LED lamp is used to illuminate the structure to provide sufficient brightness conditions and improve the quality of the captured images. The steel rod is clamped on the table to a length of 20cm long with an accelerometer mounted for measurement. Moreover, it is noted that the addition of sensors and other devices on the equipment, the physical properties of vibrating objects would be changed. This work is mainly to verify the accuracy and applicable conditions of the proposed camera-based vibration measurement method. Therefore, a light-weight structure experimental device is used to verify the experiment.

Fig. 3(b) shows the enlarged side view of high-speed camera. The high-speed camera is mounted with a high-quality 20-mm optical lens (manufacturer Nikon). It can adjust the size of any pixel range below the highest resolution 4096×3076. When the

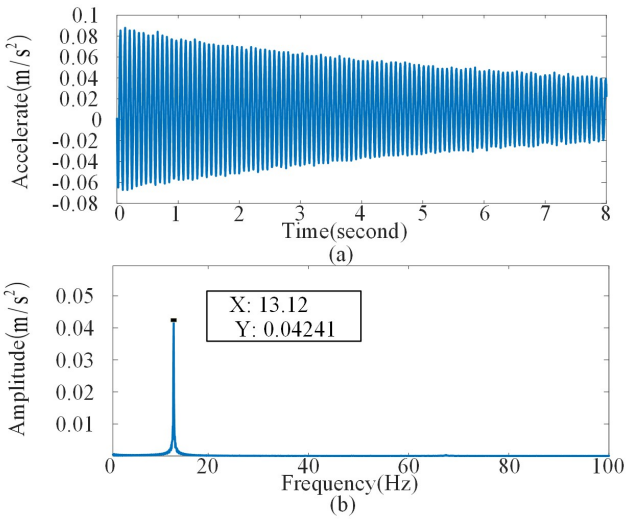


Fig. 4. Vibration measurement results from the accelerometer. (a) The acceleration signal in time domain. (b) The acceleration signal in frequency domain.

output format and the resolution are reduced, the minimum frame period of the camera (maximum frame rate) can be increased and the range for the exposure time is also re-calculated. The frame rate of high-speed cameras can reach up to 3000 frames per second.

In order to facilitate the acquisition, we develop a video acquisition and storage system for industrial equipment based on high-speed camera and general PC platform. The system is capable of high-speed real-time image acquisition above 2500fps. The camera parameters and acquisition card parameters can be quickly set up at the same time, which significantly reduces the cost of data acquisition and improves the configuration flexibility.

Fig. 3(c) shows one of the sample screenshots of the high-speed camera. All the videos are captured at the resolution of 1280×720 and the frame rate of 200fps (frames per second). For comparison, several videos with varying focal length and aperture are taken.

IV. Experimental Results

A series of controlled experiments are carried out to verify the accuracy and applicable conditions of the proposed camera-based vibration measurement method.

A. Comparison with Acceleration Measurement

Firstly, an experiment is formulated to compare the results to the traditional sensors, in order to validate the performance of the proposed camera-based measurement method. The motion of the structure is simultaneously measured by the accelerometer and the high-speed camera. Then a fast Fourier transform (FFT) is performed to transform the signal into the frequency domain. The sampling frequency of the accelerometer is 2000 Hz and the recording time is 8 seconds.

Fig. 4 presents the vibration signal for the steel rod measured by the conventional accelerometer. The acceleration in time domain is shown in Fig. 4(a) and the acceleration signal in the frequency domain is presented in Fig. 4(b). As can be observed from Fig. 4 (b), the peak frequency of the steel rod is 13.12Hz,

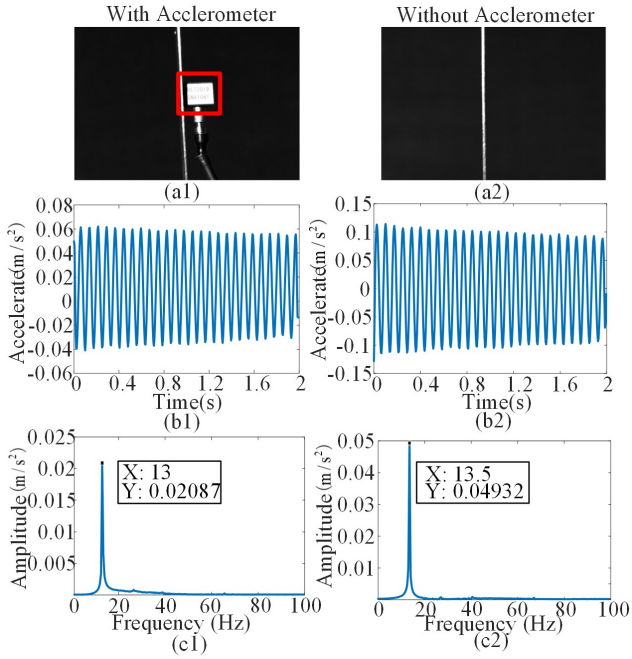


Fig. 5. The vibration signal measured by the proposed camera-based measurement for the steel rod with/without accelerometer. (a1) A video screenshot for the steel rod equipped with the accelerometer. (b1) The acceleration signal in time domain for the steel rod equipped with the accelerometer. (c1) The acceleration signal in frequency domain for the steel rod equipped with the accelerometer. (a2) A video screenshot for the steel rod without the accelerometer. (b2) The acceleration signal in time domain for the steel rod without the accelerometer. (c2) The acceleration signal in frequency domain for the steel rod without the accelerometer.

and the amplitude of the peak frequency is $0.04241\ \mathrm{m/s^2}$. It is noted that the first 100 Hz part of the spectrum is shown, and there is no peak frequency in the latter part.

Fig. 5 shows the vibration signals measured by the proposed camera based non-contact method with and without the accelerometer. The resolution of the camera is set up to 1280×720. The frame rate is 200 frames per second, the focal length is 20mm and the aperture is 2.8. The sequence of the images is measured up to 2 s (400 frames). The motion amplitude is very small and invisible to the naked eye, which is equivalent to micro-motion. Fig. 5 (b1) show the time series data of the marker's vertical position captured by the camera system. A pixel near the accelerometer is selected as the video marker. The phase information is extracted by the proposed phase-based vibration measurement method and transformed into displacement information in units of pixels. In the frames, the length of the accelerometer plane is 120 pixels and the width of the accelerometer plane is 132 pixels, with 1.6 millimeters per pixel. Then the displacement information is converted into acceleration information in order to describe the vibration by using the Laplacian of the Gaussian (LOG) operator. The displacement in units of pixels can be converted to units of millimeters by multiplying 1.6. A peak frequency of about 13 Hz was detected from the results of both Fig. 4(b) and Fig. 5(c1); this confirms that the peak frequency of the lightweight structure model can be detected from measurement results of the vision system just as well as from the data of

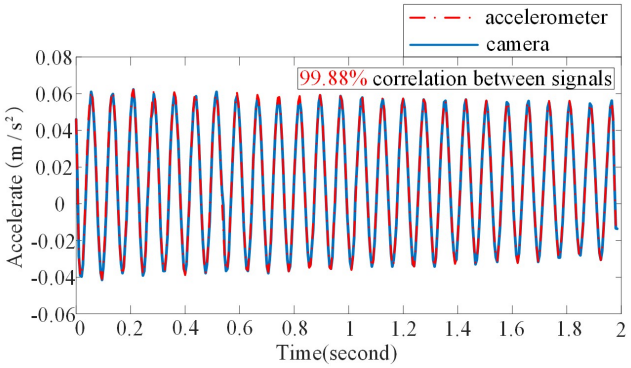


Fig. 6. Comparison between the aligned acceleration signals in time domain derived from the camera and the accelerometer.

conventional accelerometer.

Fig. 5 (b2) presents the time series data of the same marker position with that of Fig. 5(b1). As for the lightweight structure, the installation of accelerometer can definitely affect the inherent frequency. Then the proposed vision system is implemented for the lightweight structure without installing accelerometer. The FFT analysis for time series data in Fig. 5(b2) is shown in Fig. 5 (c2). As can be seen from Fig. 5 (c2), the peak frequency of the FFT result is 13.5 Hz, which is 3.8% greater than that with accelerometer in Fig. 5 (c1). It can be verified that instrumentation of the lightweight structures with the physical-attached sensors induces mass-loading effects and reduces the natural frequency of the structure.

Furthermore, in order to compare the measured acceleration signals in time domain, we manually align the accelerometer data with the camera data. Fig. 6 presents the aligned acceleration signals in time domain derived from the camera and the accelerometer. As can be observed from Fig. 6, the aligned acceleration signal in time domain from the camera matches that from the accelerometer very closely for the whole time range. For a quantitative comparison, the correlation between the camera and accelerometer acceleration signals is 99.88%. The correlation is a statistical indicator that reflects the degree of correlation between signals. This verifies that the acceleration signal extracted from video is in accordance with that from the conventional accelerometer measurement.

B. Comparison with Different Camera Parameters

Different camera parameters setting also have some effect on the measurement results. Here, the influence of resolution, focal length and aperture of the camera are mainly discussed on the measurement results.

Fig. 7 shows the screenshots from videos with different focal length. Four cases are carried out with different focal length including 20mm, 35mm, 50mm and 85mm. Other conditions remain unchanged (resolution, 1280×720; frame rate, 200fps; aperture, 2.8). Different from the accelerometer method only for single point measurement, the proposed vision-based non-contact method is a multi-point measurement. That is, the pixels of the image can be selected as measurement points. In this work, Point 1, Point 2 and Point 3 are selected as the measurement points. Point 1 is at the top of the edge of the ruler, while Point 2 and Point 3 are at the middle and bottom positions, respectively. Moreover, Point 2 is the connection point between

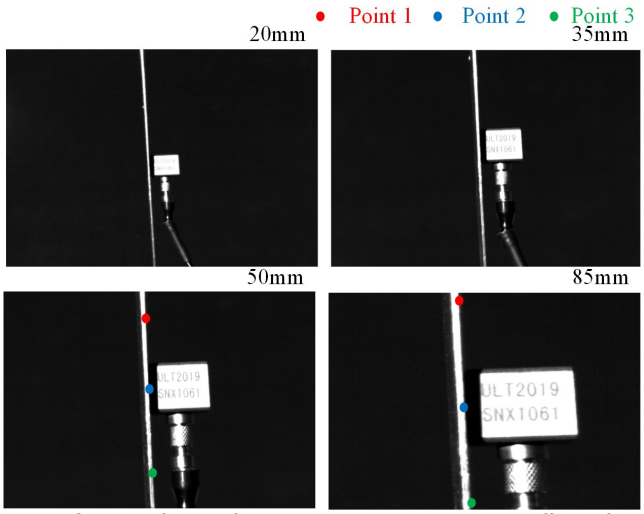


Fig. 7. Cropped frames from the corresponding videos differ in focal length.

TABLE I
THE PROPOSED CAMERA-BASED MEASUREMENT RESULTS OF DIFFERENT POINTS WITH VARYING FOCAL LENGTH.

Point	Focal length	Aperture	Amplitude	Frequency
Point 1	20mm	2.8	0.11587	13Hz
	35mm	2.8	0.04451	13Hz
	50mm	2.8	0.08498	13Hz
	85mm	2.8	0.05371	13Hz
Point 2	20mm	2.8	0.05515	13Hz
	35mm	2.8	0.02080	13Hz
	50mm	2.8	0.04244	13Hz
	85mm	2.8	0.02988	13Hz
Point 3	20mm	2.8	0.03152	13Hz
	35mm	2.8	0.01441	13Hz
	50mm	2.8	0.02982	13Hz
	85mm	2.8	0.02335	13Hz

accelerometer and the lightweight structure.

Table I summarizes the camera-based measurement results of three measurement points with varying focal length. For every point, the frequency and amplitude of the vibration FFT analysis are calculated. It is obvious that the peak frequencies measured at different points are all 13 Hz. Take the video for focal length of 20mm as an example, the measured amplitudes of Point 1, Point 2 and Point 3 are 0.11587, 0.05515 and 0.03152 respectively. The actual amplitude of the measurement points depends on the excitation magnitude and the position of the point. Since the impact hammer strikes the top of the lightweight structure, it is actually supposed that the amplitude of Point 1 is the largest and the amplitude of Point 3 is the smallest, which is consistent with the camera-based measurement. The amplitude measurement results for the video at other focal lengths obtain the similar performance.

Moreover, the signal correlation between the accelerometer and the vision system is deduced for Point 2 position. According to Table I, the correlation for focal length 20mm and 35mm is 99.81% and 99.88%, respectively. While for the greater focal length, the similar conclusion can be included. That is, the correlation between the accelerometer and the videos with focal length 50mm and 85mm are 99.90% and 99.94%, respectively. The increasing of the focal length can

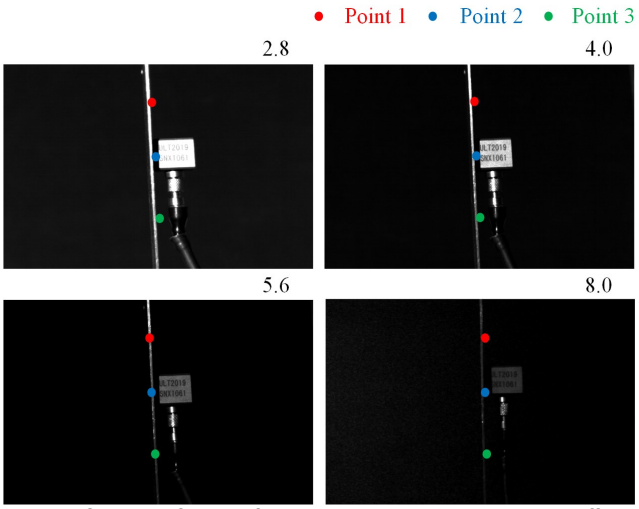


Fig. 8. Cropped frames from the corresponding videos differ in aperture.

TABLE II
THE PROPOSED CAMERA-BASED MEASUREMENT RESULTS OF DIFFERENT POINTS WITH VARYING APERTURE.

Point	Focal length	Aperture	Amplitude	Frequency
Point 1	35mm	2.8	0.04451	13Hz
	35mm	4.0	0.17328	13Hz
	35mm	5.6	0.12216	13Hz
	35mm	8.0	/	/
Point 2	35mm	2.8	0.02080	13Hz
	35mm	4.0	0.08210	13Hz
	35mm	5.6	0.06458	13Hz
	35mm	8.0	/	/
Point 3	35mm	2.8	0.01441	13Hz
	35mm	4.0	0.04864	13Hz
	35mm	5.6	0.03842	13Hz
	35mm	8.0	/	/

improve the accuracy of visual measurement.

Then the influence of aperture is discussed. Fig. 8 shows the screenshots from videos with different aperture parameter. The aperture of the camera is adjusted from 2.8 to 4.0, 5.6 and 8.0. Other conditions remain unchanged (resolution, 1280 × 720; frame rate, 200fps; focal length, 35mm). Similarly, Point 1, Point 2 and Point 3 on the rod are selected for measurement. Since the relative position of the camera and the rod remains unchanged, the same pixel point is selected.

Table II resents the camera-based measurement results for Point 1, Point 2 and Point 3 from videos with varying aperture. Similar to previous measurements with different focal lengths, the peak frequency of the vibration signal in FFT analysis is 13 Hz. For the peak frequency, the amplitudes of Point 1, Point 2 and Point 3 with the aperture of 2.8 are 0.04451, 0.02080 and 0.01441 respectively. The peak amplitude of Point 1 is about twice that of Point 2 and about 3 times that of Point 3. The amplitude proportional relationship among the three points under other apertures is also similar. It is noted that the amplitude and frequency results for aperture 8.0 are not displayed in Table II. In this case, the accuracy of camera-based measurement is relatively low, and the spectrums obtained do not have reference meaning and value.

The correlation for acceleration signals between the accelerometer and the videos with varying aperture (2.8, 4.0, 5.6, 8.0) are 99.88%, 99.90%, 96.61% and 53.82%, respectively. With the increase of aperture value, the measurement accuracy decreases accordingly. When the aperture value is adjusted to 8.0, the overall effect of the image is dim and the contrast between foreground and background is not obvious enough. In this case, the phase information cannot be extracted correctly. Hence, a certain brightness and contrast is necessary for the image capture.

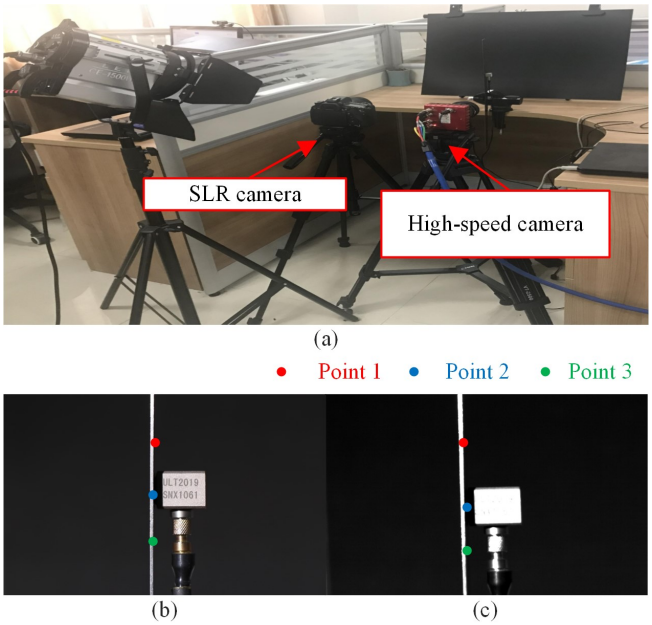


Fig. 9. (a) The side view of the experimental scene. (b) A sample frame from the SLR camera. (c) A sample frame from the high-speed camera.

Furthermore, expensive high-speed camera is not required for some low-frequency (frequency less than 15 Hz) motions. Theoretically, using common single lens reflex camera (SLR) for data acquisition is sufficient for the needs of measurement. Fig. 9 shows the vision-based measurement system by the SLR and high-speed camera. As shown in Fig. 9, videos are captured simultaneously by SLR camera (Canon EOS 80D, 1280× 720 pixel resolution, RGB) and high-speed camera (FLARE 12M180MCX, 1280× 720 pixel resolution, grayscale) at the same frame rate (50 fps). Fig. 9 (b) and Fig. 9 (c) presents the sample frames from the SLR camera and the high-speed camera, respectively. Point 1, Point 2 and Point 3 on the rod are selected as the measurement points.

Fig. 10 shows the acceleration time series and FFT results of Point 1 from high-speed camera and SLR camera. As can be seen from Fig. 10, both the time-domain and the frequency-domain signals from different cameras are close. The correlation between the time-domain signals is 99.92%.

Table III summarizes the measurement results for Point 1, Point 2 and Point 3 from high-speed camera and SLR camera. The peak frequency of vibration signal of high-speed camera and SLR camera is exactly the same, 13Hz. For peak frequency, the amplitude of point 1, point 2 and point 3 measured by high-speed camera is 0.2956, 0.1492 and 0.0944 respectively. While the amplitudes of point 1, point 2 and point 3 measured by SLR camera are 0.2998, 0.01508 and 0.1053 respectively. The peak amplitude of Point 1, Point 2 and Point 3 measured from different cameras are also very close. For a quantitative

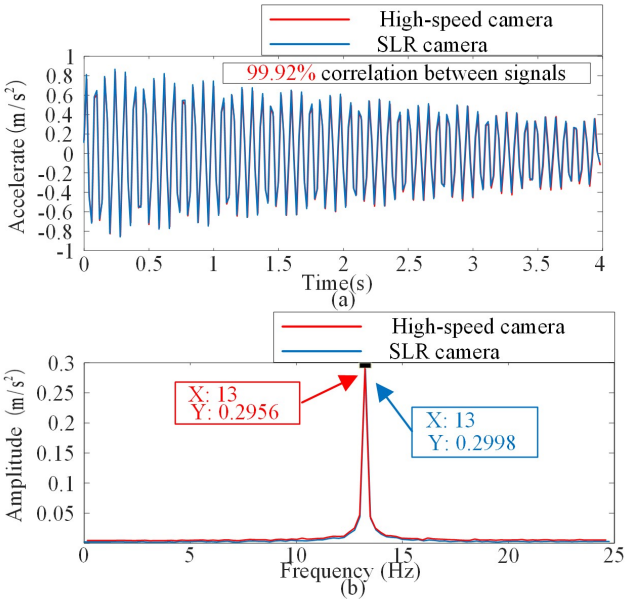


Fig. 10. Vibration signal measurement from the high-speed camera and SLR camera. (a) The acceleration signal in time domain. (b) The acceleration signal in frequency domain.

TABLE III
THE PROPOSED CAMERA-BASED MEASUREMENT RESULTS OF DIFFERENT POINTS WITH VARYING APERTURE.

Point	Camera	Amplitude	Frequency	Correlation
Point 1	High-speed camera	0.2956	13Hz	99.92%
	SLR camera	0.2998	13Hz	
Point 2	High-speed camera	0.1492	13Hz	99.94%
	SLR camera	0.1508	13Hz	
Point 3	High-speed camera	0.0944	13Hz	99.92%
	SLR camera	0.1053	13Hz	

comparison, the correlation for acceleration signals between the high-speed camera and SLR camera are calculated. The correlation of Point 1, Point 2 and Point 3 are 99.92%, 99.94% and 99.92%, respectively. Therefore, the high-speed camera can be replaced by the SLR camera, especially for the low-frequency vibration measurement.

C. Comparison for Video Downsampling

In some cases, the vibration signal contains multi-frequency components. Downsampling is ultilized to extract the principal component and remove the frequency component with small and insignificant amplitude.

As shown in Fig. 11(a), a rubber is mounted on the steel rod to generate noise component. Fig. 11 shows the measured acceleration signals in the time and frequency domain. As can be seen in Fig. 11, both time-domain signal and frequency-domain signal do not agree with the predicted results. The measured frequency spectrum contains several noises.

Then the video is downsampled by a factor of two in each dimension to a size of 640 × 360 using binomial filter. Fig. 12(a) shows the measured acceleration signal in time domain and Fig. 12(b) shows the measured acceleration signal in frequency domain. The acceleration signal in the time domain is an approximate attenuated sinusoidal signal. Fig. 12(b) shows the acceleration signal in frequency domain, which shows a peak

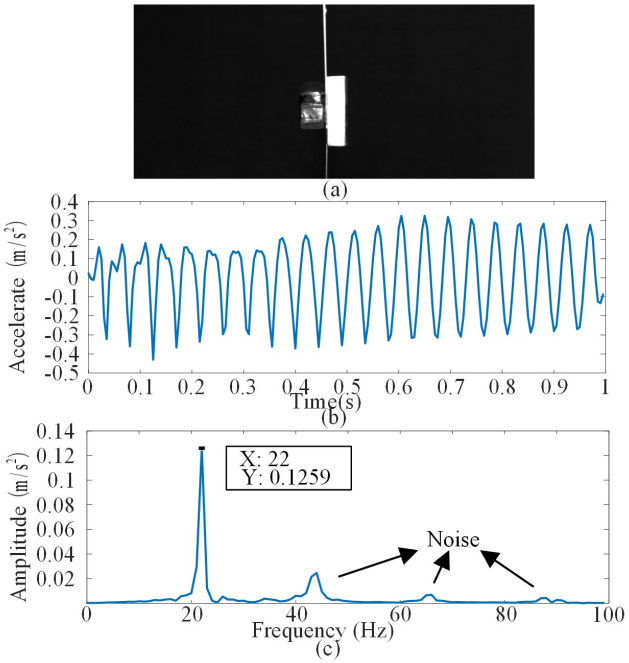


Fig. 11. Vibration signal measurement before video downsampling. (a) The acceleration signal in time domain. (b) The acceleration signal in frequency domain.

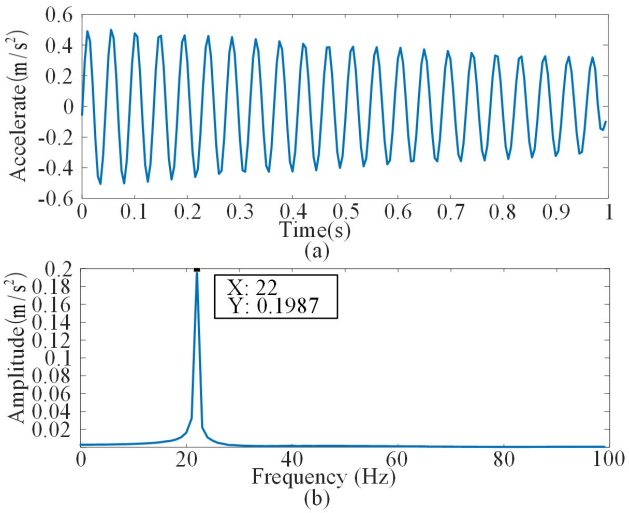


Fig. 12. Vibration signal measurement after video downsampling. (a) The acceleration signal in time domain. (b) The acceleration signal in frequency domain.

TABLE IV
SUMMARY OF COMPARISON OF DIFFERENT LEVELS OF DOWNSAMPLING

Downsampling Level	Resolution	Frame Length	Processing Time(s)	Correlation
1	1280 × 720	400	49.50	99.88%
2	640 × 360	400	23.93	99.91%
4	320 × 180	400	17.69	99.94%
8	160 × 90	400	9.81	99.94%

frequency at 22Hz without noise component. This experiment shows that the principal frequency components can be extracted well by downsampling the video.

For comparison, downsampling each dimension of the video (resolution, 1280 × 720; focal length, 35mm; aperture, 2.8) by a

factor of 1, 2, 4 and 8. The correlation between the measured time-domain acceleration signals from camera and accelerometer at different spatial resolutions is calculated. Considering the high accuracy of accelerometer measurement, the correlation is related to accuracy. Table IV shows the comparison for different levels of downsampling. From comparison, more spatial downsampling not only improves the measurement accuracy but also reduces the processing time as there are fewer pixels to process. However, this is at the expense of less spatial resolution and fewer measurement points. Therefore, the downsampling level should be selected based on the trade-off between spatial resolution and accuracy. In addition, we utilized the phase-based optical flow and the proposed method for the vibration measurement of the same video (resolution, 1280×720; focal length, 35mm; aperture, 2.8) respectively. The processing time of the phase-based optical flow is 58.01s, while the processing time of the proposed method is 49.49s. This verifies that the proposed method reduces the computation time.

V. Conclusion

This paper proposed a novel camera-based vibration measurement methodology by applying the phase-based optical flow. A series of controlled experiments on the lightweight structure were implemented. The peak frequency of 13 Hz measured by the camera agreed within 0.9% of the frequency measured by the accelerometer from close range. The correlation between the camera and accelerometer acceleration signals was 99.88%. It verifies that the acceleration signal extracted from video was in accordance with that from the accelerometer. Compared to the conventional measurement with accelerometer, the proposed camera-based method effectively avoids the influence of the load accelerometer, especially for the lightweight structure. Furthermore, the experiments for micro-vibration measurement with the influence of camera parameters, shooting conditions, and data processing was also discussed. The results showed that the proposed method can definitely improve the measurement accuracy.

References

[1] F. Dreier, P. Günther, T. Pfister, J.W. Czarske, and A. Fischer, "Interferometric sensor system for blade vibration measurements in turbomachine applications," *IEEE Trans. Instrum. Meas*., vol. 62, no. 8, pp. 2297-2302, Aug. 2013.

[2] C. Peng, M. Zhu, K. Wang, Y. Ren, and Z. Deng, "A Two-stage synchronous vibration control for magnetically suspended rotor system in the full speed range," *IEEE Trans. Ind. Electron.,* vol. 67, no. 1, pp. 480-489, Jan. 2020.

[3] J. Baqersad, P. Poozesh, C. Niezrecki, and P. Avitable, "Photogrammetry and optical methods in structural dynamics-a review," *Mech. Syst. Signal Pr*., vol. 86, pp. 17-34, 2017.

[4] H. Huang, N. Baddour, and M. Liang, "Bearing fault diagnosis under unknown time-varying rotational speed conditions via multiple time-frequency curve extraction," *J. Sound Vib*., vol. 414, pp. 43-60, 2018.

[5] R.Li, T. Wang, Z. Zhu, and W. Xiao, "Vibration characteristics of various surfaces using an LDV for long range voice acquisition," *IEEE Sens. J*., vol. 11, no. 6, pp. 1415-1422, June 2011.

[6] A. Hati and C. W. Nelson, "W-band vibrometer for non-contact thermoacoustic imaging," *IEEE Trans. Ultrason., Ferr*., published on line, 2019.

[7] S. Dhiraj and X. Omkar, "Object drop detection on railway track through rayleigh wave sensing using laser vibrometer," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9158-9172, Oct. 2018.

[8] H. Sakaino, "Camera-vision-based water level estimation," *IEEE Sensors J*., vol. 16, no. 21, pp. 7564-7565, Nov. 2016.

[9] F. Zhong and C. Quan, "A single color camera stereo vision system," *IEEE Sens. J.*, vol.18, no. 4, pp. 1474-1482, Dec. 2017.

[10] L. Alberto and Z. Emanuele, "Modeling uncertainty for a vision system applied to vibration measurements," *IEEE Trans. Instrum. Meas*., vol. 65, no. 8, pp. 1818-1826, Aug. 2016.

[11] D. Feng and M. Feng, "Experimental validation of cost-effective vision-based structural health monitoring," *Mech. Syst. Signal Pr*., vol. 88, pp. 199-211, May. 2017.

[12] D. Lydon, M. Lydon, J. M. del. Rincon, S. E. Taylor, D. Robinson, E. O'Brien, "Development and field testing of a time-sychronized system for multi-point displacement calculation using low-cost wireless vision-based sensors," *IEEE Sens. J*., vol. 18, no. 23, pp. 9744-9754, Dec 2018.

[13] M. N. Helfrick, C. Niezrecki, P. Avitabile and T. Schmidt, "3D digital image correlation methods for full-field vibration measurement," *Mech. Syst. Signal Pr*., vol. 25, no. 3, pp. 917-927, Apr. 2011.

[14] F. Q. Zhong, X. X. Shao, and C. Quan, "A comparative study of 3D reconstruction methods in stereo digital image correlation," *Opt. Laser Eng*., vol. 122, pp. 142-150, 2019.

[15] M. Jiang, Q. Gu, T. Aoyama, T. Takaki, I. Ishii, "Real-time vibration source tracking using high-speed vision," *IEEE Sens. J.*, vol. 17, no. 5, pp. 1513-1527, Mar. 2017.

[16] 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," *Mech. Syst. Signal Pr*., vol. 28, no. 28, pp. 333-347, Apr. 2012.

[17] N. Wang, K. Ri, H. Liu, X. Zhao, "Structural displacement monitoring using smartphone camera and digital image coorelation," *IEEE Sens. J.* vol. 18, no. 11, pp. 4664-4672, Jun. 2018.

[18] S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying Framework," *Int. J. Comput. Vision*, vol. 56, no. 3, pp. 221-255, Feb. 2004.

[19] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell*., vol. 17, no. 1-3, pp. 185-203, Aug. 1981.

[20] D. Feng and M. Feng, "Vision-based multipoint displacement measurement for structural health monitoring," *Struct. Control. Health Monit.*, vol. 19, no. 4, pp. 2363-2374, May. 2016.

[21] T. Khuc and F. Catbas, "Computer vision-based displacement and vibration monitoring without using physical target on structures," *Struct. Infrastruct. Eng.*, vol. 13, no. 4, pp. 505-516, Jun. 2017.

[22] Y. Yang, X. Sang, S. Yang, X. Hou, Y. Huang, "High-precision vision sensors method for dam surface displacement measurement," *IEEE Sens. J*., published on line, 2019.

[23] J. Javh, J. Slavic, and M. Boltezar, "High frequency modal identification on noisy high-speed camera data," *Mech. Syst. Signal Proc.*, vol. 98, pp. 344-351, Jan. 2018.

[24] D. J. Fleet and Y. Weiss, "Optical Flow Estimation," in Handbook of Mathematical Models in Computer Vision, New York: Springer, 2006.

[25] A. Davis, K. L. Bouman, J. G. Chen, M. Rubinstein, F. Durand, and W. T. Freeman, "Visual vibrometry: estimating material properties from small motions in video," IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015.

[26] 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.

[27] Y. Yang, C. Dom, T. Mancini, Z. Talken, and G. Kenyon, "Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification," *Mech. Syst. Signal Pr*., vol. 85, pp. 567-590, Feb. 2017.

[28] A. Sarrafi, M Zhu, C. Niezrecki, and P. Poozesh, "Vibration-based damage detection in wind turbine blades using phase-based motion estimation and motion magnification," *J. Sound Vib*., vol. 421, pp. 300-318, May. 2018.