Q-shift dual-tree complex wavelet phase-based vibration displacement measurement for wind turbine towers

Q-Shift Dual-Tree Complex Wavelet Phase-Based Vibration Displacement Measurement for Wind Turbine Towers

Yang Yang[1,2,3]*, Yuanqing Fan[1,2,3], Zhewei Wang[1,2,3], Jie Luo[4], Giuseppe lacidogna[5], Shuai Tao[1]

[1] School of Civil Engineering, Chongqing University, Chongqing, China

[2] State Key Laboratory of Safety and Resilience of Civil Engineering in Mountain Area, Chongqing, China

[3] Institute for Liyang Smart City, Chongqing University, Liyang, China

[4] Guizhou Investment and Construction Co.,Ltd. Of China State Construction Fourth Engineering Division Corp., Ltd., Guizhou, China

[5] Department of Structural, Geotechnical and Building Engineering, Politecnico di Torino, Turin, Italy

Abstract

In the context of the global energy transition, wind turbine towers serve as the primary load-bearing structures of wind energy systems, and accurate measurement of their vibration-induced displacements is essential for structural health monitoring and safety assessment. Conventional contact-based measurement techniques suffer from complex installation procedures and limited spatial coverage. In contrast, existing vision-based vibration measurement methods are often highly dependent on surface texture and exhibit insufficient phase stability under challenging imaging conditions. To address these limitations, this study proposes a non-contact vibration displacement measurement method for wind turbine towers based on the phase information of the Q-shift dual-tree complex wavelet transform. The proposed approach exploits the multi-scale and multi-directional complex coefficients obtained from the Q-shift dual-tree complex wavelet decomposition to construct a phase representation with approximate shift invariance and improved inter-scale consistency. Displacement estimation is performed within regions exhibiting stable phase responses, such as structural edges, thereby enhancing robustness against complex backgrounds and noise. Furthermore, an adaptive downsampling strategy is introduced to alleviate

* Corresponding authors.

Email address: yangyangcqu@cqu.edu.cn (Yang Yang)

phase wrapping effects in large-scale structural vibrations, enabling reliable displacement measurement under relatively large vibration amplitudes. Numerical simulations and wind turbine tower model experiments demonstrate that the proposed method achieves sub-pixel displacement accuracy under complex environmental conditions. Compared with conventional vision-based approaches based on phase or image intensity information, the proposed method exhibits superior overall stability, providing a high-precision and robust visual vibration measurement solution for the non-contact monitoring of tall and flexible structures.

Keywords: Structural health monitoring, Q-Shift Dual-Tree Complex Wavelet Transform, Computer vision, Vibration displacement measurement, wind turbine tower

1. Introduction

In recent years, driven by the global energy transition and carbon neutrality goals, wind power has experienced rapid growth as an important component of clean and renewable energy systems [1-3]. With the continuous expansion of wind farms and the extension of their service life, wind turbine towers, as the core supporting structures of wind energy systems, are facing increasingly severe challenges [4,5]. Wind turbine towers are exposed to complex environments such as extreme weather conditions and are subjected to the combined effects of wind loads, gravity, and long-term dynamic fatigue loads, which may lead to cumulative structural damage and increased safety risks [6-8]. Therefore, structural health monitoring of wind turbine towers is of great importance to ensure operational safety and extend the service life of wind turbines [9-11].

Vibration measurement is a fundamental means of obtaining structural dynamic responses and plays a critical role in structural health monitoring [12]. By analyzing vibration data, potential damage in tower structures can be detected in a timely manner, thereby preventing catastrophic failures. Existing vibration measurement techniques can generally be classified into contact-based and non-contact approaches [13]. Traditional contact-based methods, such as accelerometers and linear variable differential transformers (LVDTs), offer high measurement accuracy and stability; however, they exhibit significant limitations in practical applications. First, sensors must be physically attached to the structure and rely on dedicated cables for signal transmission [14]. For large-scale and high-rise structures such as wind turbine towers, sensor installation involves safety risks associated with high-altitude operations, as well as complex cabling and long installation periods [15]. Moreover, contact-based methods typically provide measurements at only a limited number of discrete points, making it difficult to reconstruct the full-field vibration response of the structure [16].Non-contact measurement techniques include laser ranging, radar-based measurement, and GPS-based displacement monitoring. Laser ranging methods infer displacement through laser distance measurements and provide high accuracy and fast response, but they are highly sensitive to atmospheric conditions such as fog, rain, and dust [17]. Radar-based methods, which rely on electromagnetic wave ranging, exhibit better environmental adaptability and certain penetration capabilities; however, their measurement accuracy is relatively limited and the associated costs are high, making it difficult to meet the millimeter-level monitoring requirements of wind turbine towers [18]. GPS-based measurement methods enable large-scale three-dimensional displacement monitoring based on satellite positioning, but they suffer from signal occlusion and low sampling rates, rendering them incapable of capturing high-frequency vibrations [19]. In contrast, digital cameras offer advantages such as low cost, flexible deployment, and high spatial resolution, and when combined with computer vision techniques, they have become an active research topic in the field of non-contact vibration measurement [14].

According to their sensing mechanisms, existing vision-based vibration measurement methods can be broadly divided into image intensity–based approaches and image phase–based approaches. Image intensity–based methods estimate structural displacements by tracking temporal variations in brightness, texture, or geometric features, with representative techniques including template matching [20-23], feature point matching [24-26], optical flow analysis [27-29], and edge detection [30,31]. These methods perform well in laboratory environments with clear textures and stable illumination; however, their accuracy and reliability strongly depend on the distinctiveness of surface features and the stability of imaging conditions. For large-scale structures such as wind turbine towers, the tower surface is often smooth, low-contrast, or characterized by repetitive textures. Under long-distance outdoor monitoring conditions, these methods are highly susceptible to environmental noise and non-uniform illumination, leading to difficulties in feature extraction and degraded tracking stability, which severely limits their effectiveness in practical engineering applications.

In contrast, image phase-based methods estimate small structural displacements by analyzing temporal variations in local image phase, offering advantages such as marker-free measurement and reduced sensitivity to illumination changes. As a result, phase-based approaches have attracted increasing attention in the field of visual vibration measurement [32-35]. Existing phase-based methods typically rely on complex-valued bandpass multi-scale representations to construct local phase information, with common implementations including Gabor filters [36-38] and complex steerable pyramids [32,39]. These techniques have demonstrated high accuracy in measuring micro-vibrations of small- and medium-scale structures such as beams and plates. However, their effectiveness is constrained by the assumption that displacement amplitudes are small relative to the phase wavelength. In large-scale structural vibration measurements, these methods are prone to phase wrapping and scale mismatch issues [40]. Moreover, their scale and directional decompositions are primarily designed for general image processing tasks, offering limited adaptability to the spatial non-uniformity and dominant vibration directions characteristic of engineering structures.In recent years, some studies have introduced analytic signal representations based on the Hilbert transform to enhance phase continuity and noise robustness through short-range or local phase analysis, partially alleviating the sensitivity of conventional phase-based methods to scale selection [41,42]. Nevertheless, such approaches typically rely on local stationarity assumptions and focus on phase variations over adjacent pixels or short spatial scales. For tall and flexible structures such as wind turbine towers, which exhibit large scale spans and significant mode shape variations along the structural height, the multi-scale consistency and global phase stability of these methods remain limited, making it difficult to simultaneously achieve robust phase extraction and accurate displacement quantification under complex imaging conditions.

The dual-tree complex wavelet transform (DTCWT), constructed using pairs of approximately Hilbert-transform-related filter banks, provides a solid theoretical foundation for stable and controllable local phase representations by combining strict multi-scale analysis with approximate shift invariance and explicit directional selectivity [43]. The DTCWT and its extensions have demonstrated superior robustness compared with conventional real-valued wavelets in complex engineering signal analysis tasks, such as rotating machinery fault diagnosis, particularly under strong noise and non-Gaussian interference conditions [44]. Recent studies have further validated the effectiveness and practical feasibility of combining adaptive parameter optimization with the DTCWT for strongly non-stationary and low signal-to-noise ratio scenarios [45]. Accordingly, for tall and flexible structures under long-distance visual measurement conditions, exploiting the DTCWT to construct phase representations with multi-scale consistency and controllable directionality represents a promising research direction for achieving high-precision quantitative vibration displacement measurement.

Motivated by these considerations and the requirements of wind turbine tower structural health monitoring, this paper proposes a non-contact vibration displacement measurement method based on the phase information of the Q-shift dual-tree complex wavelet transform. The proposed method constructs a multi-scale consistent and directionally controllable phase representation via the dual-tree complex wavelet decomposition, performs displacement estimation within regions exhibiting stable phase responses such as structural edges, and incorporates phase wrapping suppression strategies to enable stable extraction and high-accuracy quantification of wind turbine tower vibration displacements under complex imaging conditions.

2. Theoretical basis

2.1 Q-Shift Dual-Tree Complex Wavelet Transform

The Dual-Tree Complex Wavelet Transform achieves multi-scale signal analysis by constructing two independent real wavelet transform trees (“real-part tree a”and “imaginary-part tree b”)[46]. While both trees employ identical wavelet bases, their filter designs remain mutually independent. The coefficients generated by the real tree represent the amplitude information of the signal, whereas those from the imaginary tree capture the phase information.

The one-dimensional DTCWT can be mathematically expressed as:

$$\psi(t)=\psi_a(t)+i\psi_b(t) \tag{1}$$

where i denotes the imaginary unit, $\psi_a(t)$ represents the real part of the complex wavelet, and $\psi_b(t)$ represents the imaginary part.

The two-dimensional DTCWT can be obtained through the tensor product of the one-dimensional DTCWT. In this way, six different directional 2D DTCWTs can be derived, corresponding to ±15°,±45°and±75°. These six directional wavelet functions can be expressed as:

$$\begin{aligned}
\psi_1(x,y)&=\phi_a(x)\psi_a(y)-\phi_b(x)\psi_b(y)+i[\phi_a(x)\psi_b(y)+\phi_b(x)\psi_a(y)]\\
\psi_2(x,y)&=\phi_a(x)\psi_a(y)+\phi_b(x)\psi_b(y)+i[-\phi_a(x)\psi_b(y)+\phi_b(x)\psi_a(y)]\\
\psi_3(x,y)&=\psi_a(x)\phi_a(y)-\psi_b(x)\phi_b(y)+i[\psi_a(x)\psi_b(y)+\psi_b(x)\phi_a(y)]\\
\psi_4(x,y)&=\psi_a(x)\phi_a(y)+\psi_b(x)\phi_b(y)+i[-\psi_a(x)\psi_b(y)+\psi_b(x)\phi_a(y)]\\
\psi_5(x,y)&=\psi_a(x)\phi_a(y)-\psi_b(x)\psi_b(y)+i[\psi_a(x)\psi_b(y)+\psi_b(x)\psi_a(y)]\\
\psi_6(x,y)&=\psi_a(x)\phi_a(y)+\psi_b(x)\psi_b(y)+i[-\psi_a(x)\psi_b(y)+\psi_b(x)\psi_a(y)]
\end{aligned} \tag{2}$$

The real-valued wavelets $\psi_a(t)$ and $\psi_b(t)$ must satisfy orthogonality or biorthogonality conditions, respectively. Furthermore, to ensure symmetry between the two trees and achieve approximate shift-invariance, $\psi_a(t)$ and $\psi_b(t)$ should form an approximate Hilbert transform pair

$$\psi_b(t)\approx H\{\psi_a(t)\} \tag{3}$$

where $H\{\cdot\}$ is the Hilbert transform operator.

The necessary and sufficient condition for these two orthogonal wavelet functions to form a Hilbert transform pair is that their corresponding low-pass filters satisfy the half-sample delay condition, which can be expressed in the frequency domain as:

$$G_0(e^{i\omega}) = e^{-0.5\omega} H_0(e^{i\omega}) \tag{4}$$

where $H_0(e^{i\omega})$ and $G_0(e^{i\omega})$ denote the frequency responses of the low-pass filters in Tree A and Tree B, respectively.

The key to DTCWT lies in designing a filter bank that must satisfy, in addition to the half-sample delay property, the following requirements: perfect reconstruction (orthogonality or biorthogonality), compact support, vanishing moment properties, and linear phase characteristics.

The Q-shift dual-tree complex wavelet transform is constructed through the Q-shift method, and its transformation framework is illustrated in Figure 1.

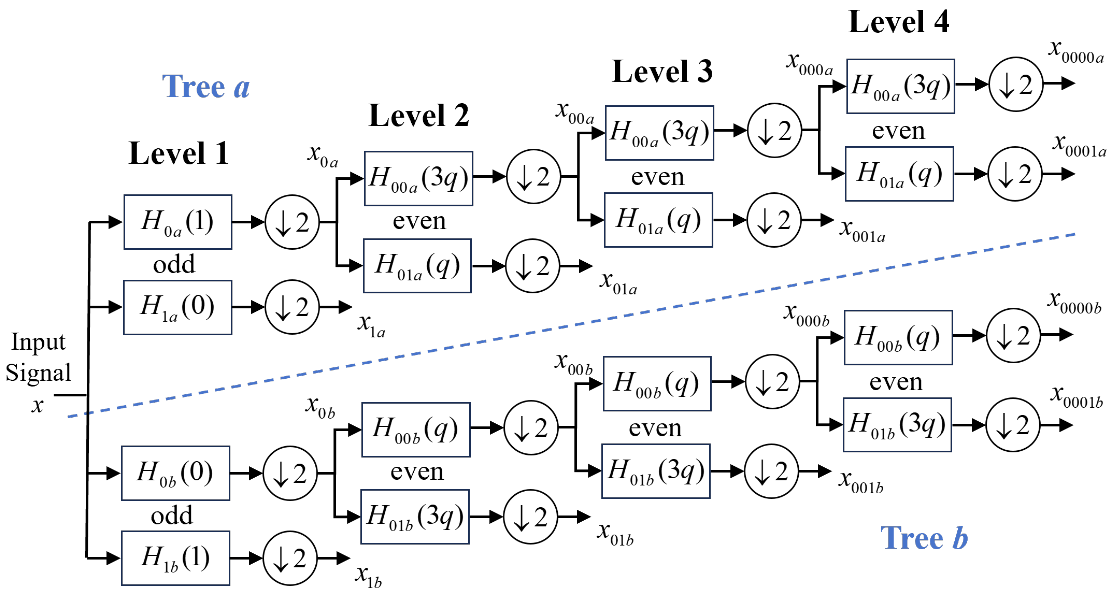


Fig. 1 Q-Shift Double Tree Complex Wavelet Transform Framework

The Q-shift(Quarter Sample Shift) DTCWT, is based on the following principles: at the first level, odd-length biorthogonal wavelets with a one-unit delay difference are employed, while for decompositions beyond the first level, even-length orthogonal wavelets with an approximate delay of a quarter-sample interval (denoted as q in the figure) are utilized. The filters used in Tree A and Tree B are designed to be temporally reversed. Since symmetry of the filter coefficients is not required, the synthesis filters ensuring perfect reconstruction can be designed to be orthogonal. All filters used at higher levels can be derived from the same orthogonal prototype sequence.

The Q-shift DTCWT offers significant advantages in image phase extraction, primarily due to its near shift-invariance, multi-directional selectivity, complex representation, and anti-aliasing properties. It can stably capture local phase information in images, reducing distortions caused by shifts or noise, while precisely extracting phase information at different resolutions through multi-scale analysis. This provides a highly robust and precise solution for phase extraction.

2.2 Phase extraction

After obtaining the wavelet functions of the two-dimensional DTCWT, the local phase and magnitude of the image at different scales and orientations can be computed by convolving the image with these wavelet functions. At each position in the image, the convolution operation calculates the response of the image at that location, thereby accurately capturing edge and texture information. The convolution of the image $I(x,y)$ with the two-dimensional Dual-Tree Complex Wavelet $\psi_i(x,y)$ can be expressed as:

$$C(x,y) = I(x,y) * \psi(x,y) \tag{5}$$

where $*$ denotes the convolution operation, $C(x,y)$ represents the complex coefficients resulting from the interaction between the image and the wavelet.

Taking the wavelet function oriented at 45° as an example, the theoretical derivation is conducted, and the wavelet function is expressed by the following equation:

$$\psi_{45^\circ}(x,y) = \mathrm{Re}_{45^\circ}(x,y) + i\,\mathrm{Im}_{45^\circ}(x,y) \tag{6}$$

where $\mathrm{Re}_{45^\circ}(x,y) = \psi_a(x)\psi_a(y) - \psi_b(x)\psi_b(y)$, $\mathrm{Im}_{45^\circ}(x,y) = \psi_a(x)\psi_b(y) + \psi_b(x)\psi_a(y)$

From equations (5) and (6), the complex response of the convolution between the image and the wavelet function can be obtained as follows:

$$C_{45^\circ}(x,y) = I(x,y) * \mathrm{Re}_{45^\circ}(x,y) + iI(x,y) * \mathrm{Im}_{45^\circ}(x,y) \tag{7}$$

By setting $R_{45^\circ}(x,y)=I(x,y)*\mathrm{Re}_{45^\circ}(x,y)$ and $I_{45^\circ}(x,y)=I(x,y)*\mathrm{Im}_{45^\circ}(x,y)$, the magnitude and phase of the complex response can be extracted from Equation (7) as:

$$A_{45^\circ}(x,y)=\left|C_{45^\circ}(x,y)\right|=\sqrt{R_{45^\circ}^2(x,y)+I_{45^\circ}^2(x,y)} \tag{8}$$

$$\varphi_{45^\circ}(x,y)=\arctan\left[\frac{I_{45^\circ}^2(x,y)}{R_{45^\circ}^2(x,y)}\right] \tag{9}$$

where $A_{45^\circ}(x,y)$ represents the local magnitude of the complex response, and $\varphi_{45^\circ}(x,y)$ represents the local phase of the complex response.

Similarly, the local phase difference between two frames of images at a specified scale and direction can be obtained.The phase difference of pixel position (x,y) between time t and time $t+\Delta t$ can be expressed as:

$$\Delta\varphi_{j,\theta}(x,y,t)=\varphi_{j,\theta}(x,y,t+\Delta t)-\varphi_{j,\theta}(x,y,t) \tag{10}$$

where j represents the scale of wavelet decomposition, and θ represents the direction of the two-dimensional dual-tree complex wavelet filter.

The phase variation is closely related to displacement. According to wave theory, the relationship between the pixel motion displacement within a plane and the phase difference can be expressed as.

$$\Delta x_{j,\theta}(x,y,t)=\frac{\lambda}{2\pi}\cdot\Delta\varphi_{j,\theta}(x,y,t) \tag{11}$$

where λ is the wavelength of the filter.

In addition, to mitigate the impact of noise in the image sequence, a local amplitude weighting method is employed to enhance the signal-to-noise ratio (SNR) [39]. The calculation formula is as follows:

$$\varphi'(x,y,t)=\frac{\left(\varphi(x,y,t)A(x,y,t)\right)\otimes h(x,y)}{A(x,y,t)\otimes h(x,y)} \tag{12}$$

$h(x,y)$ represents a two-dimensional Gaussian function, the specific expression of which is as follows:

$$h(x,y)=\exp\left[\frac{-(x^2+y^2)}{\rho}\right] \tag{13}$$

where ρ denotes the standard deviation of the Gaussian filter, which is related to the width of the spatial filter.

2.3 Optimal downsampling

In the process of estimating pixel displacement from the phase difference between two consecutive image frames, it should be noted that phase is periodic modulo 2π , Therefore, the measured phase (or phase difference) is represented within a principal interval, such as $[0, 2\pi$). When the true phase difference exceeds the boundaries of this principal interval, it is mapped back into the interval through modulo 2π , resulting in the well-known phase wrapping phenomenon. This phenomenon can cause loss or misinterpretation of phase information, thereby affecting subsequent analysis and processing[47].

The phase difference obtained from two frames of images is denoted as $\Delta\varphi_{j,\theta}^{T}(x,y,t)$, and the true phase difference can be expressed as:

$$\Delta\varphi_{j,\theta}^{R}(x,y,t) = \Delta\varphi_{j,\theta}^{T}(x,y,t) + 2n\pi \tag{14}$$

When $n \neq 0$, the phase difference $\Delta\varphi_{j,\theta}^{T}(x,y,t)$ differs from the true phase difference $\Delta\varphi_{j,\theta}^{R}(x,y,t)$ by n periods, resulting in phase wrapping. In this case, the true phase difference between the two frames of images cannot be directly obtained. When $n = 0$, $\Delta\varphi_{j,\theta}^{R}(x,y,t) = \Delta\varphi_{j,\theta}^{T}(x,y,t)$, and there is no issue of phase wrapping, allowing the true phase difference between the two frames to be directly obtained from the images.

Due to the issue of phase wrapping, the applicability of the image phase method is significantly limited. Currently, in image phase matching methods, most approaches adopt a multi-scale coarse-to-fine matching strategy to address the phase wrapping problem. This involves first performing phase matching at a coarse resolution level, then progressively refining the matching from coarse to fine levels until the finest resolution is reached. At each level, the result from the previous level is used as the initial value for the current level and is subsequently refined. However, if a phase-unstable region occurs at any level, it can lead to unrecoverable errors. Additionally, since phase estimation is required at each level, the computational cost is substantially increased.

To address the issue of phase wrapping, this section proposes a method to avoid phase wrapping between two frames by controlling the range of pixel motion in the images.

By constraining the phase difference within the range $[-\pi, \pi)$, phase wrapping can be avoided. This requires satisfying the following condition:

$$\Delta\varphi^{R}_{j,\theta}(x,y,t) < \pi \tag{15}$$

By combining equation (11), the following expression can be obtained:

$$\Delta x^{R}_{j,\theta}(x,y,t) < \frac{\lambda}{2} \tag{16}$$

Therefore, the maximum pixel displacement between two consecutive image frames cannot exceed half the wavelength ($\lambda / 2$) of the filter.

Subsequently, the maximum pixel displacement between the two frames is controlled by adjusting the image downsampling (the number of wavelet decomposition levels). The downsampling operation reduces the spatial resolution of the image, thereby compressing the size of the feature map. Typically, this process halves the width and height of the image, resulting in a reduction of the total number of pixels to one-quarter of the original. For a pair of frames in an image sequence, the reduction in spatial resolution causes the pixel displacement between the two frames to decrease by 50%. If the maximum pixel displacement between the original two frames is $\Delta_m x(x,y)$ the maximum pixel displacement after k downsampling operations can be expressed as

$$\Delta_{m,k} x(x,y) = \frac{\Delta_m x(x,y)}{2^k} \tag{17}$$

To avoid phase wrapping, the maximum pixel displacement after downsampling must not exceed half the filter''s wavelength. This allows solving for the downsampling factor k that satisfies the following equation:

$$k > \log_2\left(\frac{\Delta_m x(x,y)}{\lambda}\right) \tag{18}$$

That is, the optimal downsampling factor can be expressed as:

$$k_m = \left\lceil \log_2\left(\frac{\Delta_m x(x,y)}{\lambda}\right) \right\rceil \tag{19}$$

where $\lceil \bullet \rceil$ denotes the ceiling function.

From Equation (19), it can be inferred that when the filter is determined, the wavelength λ is known. Therefore, by determining the maximum pixel displacement $\Delta_m x(x,y)$ in the image sequence, the optimal downsampling factor k_m that avoids phase wrapping can be calculated.

To achieve this, the Canny edge detection at integer-pixel resolution is employed as an efficient preprocessing step solely for estimating the optimal downsampling factor. By computing the relative displacement between a fixed reference frame and the current frame, a coarse pixel-level displacement time history is obtained, from which the maximum inter-frame pixel displacement $\Delta_m x(x, y)$ is determined.The displacement calculation model for the image sequence is illustrated in Figure 2. This model, which utilizes a fixed reference frame, eliminates the need for frame updates during subsequent sampling processes, effectively mitigating error accumulation. It should be emphasized that the edge detection procedure is not required for the displacement measurement itself, and is introduced only as a rapid and computationally efficient means to determine the downsampling parameter. It should be emphasized that the Canny edge detection procedure is introduced solely as a preprocessing step to estimate the optimal downsampling factor. The detected edge pixels are not used for the subsequent phase-based displacement extraction.

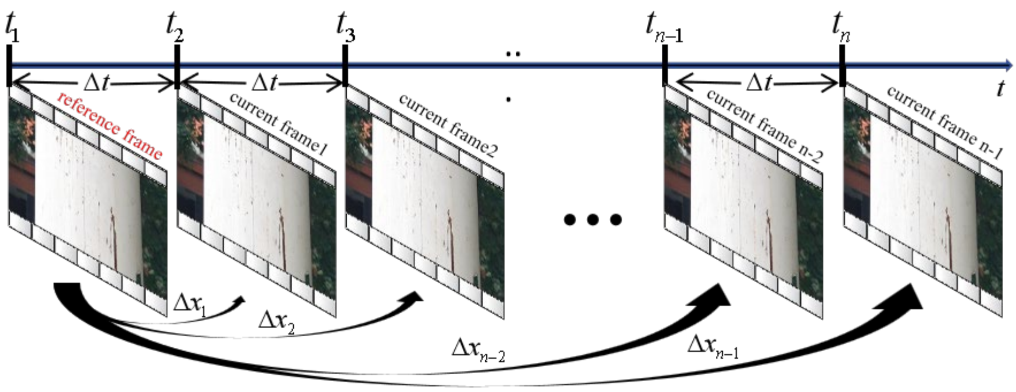


Fig. 2 Image Sequence Displacement Calculation Model

3. Proposed method

Based on the theories discussed in the previous sections, this section proposes a vibration displacement measurement method for wind turbine towers using the phase of the Q-Shift DTCWT. The implementation process is illustrated in Figure 3.

Firstly, a Region of Interest (ROI) is selected from the original video frames captured by the digital camera,. The ROI range is demarcated on the image to retain the target area required for vibration measurement. After the ROI selection is completed, the video is converted into a time-sequential series of ROI image frames. This operation significantly reduces the computational load of subsequent algorithms and minimizes the impact of background noise.

Subsequently, a rapid Canny edge detection is applied to each frame to derive a time series of edge vibration displacements with integer pixel precision. Based on the theoretical framework presented in Section 2.3, the optimal downsampling factor k_m for the Q-Shift DTCWT is computed. Here, (x_n, y_n) represents the edge coordinates detected in the n-th frame, where x_n denotes the dynamic coordinate component determined through edge detection, and y_n corresponds to the semi-length of the region of interest (ROI) in the vertical direction, This parameter remains invariant throughout the entire video sequence.

After determining the optimal downsampling factor k_m, the image sequence is decomposed in a multi-scale (with decomposition level k_m) and multi-directional manner based on the Q-Shift DTCWT to extract the local phase information of the reference frame φ_0 and the current frame φ_n in corresponding directions, thereby obtaining the phase difference $\Delta\varphi$. The spatial local weighting of the phase difference $\Delta\varphi$ is then performed using the local amplitude as the weight to enhance the signal-to-noise ratio of the phase signal.

Structural regions with pronounced intensity gradients exhibit significant complex wavelet magnitude responses and inherently provide a higher phase signal-to-noise ratio owing to the inverse-square error suppression mechanism associated with the magnitude, thereby ensuring stable phase measurements. Accordingly, these high-gradient structural regions are selected for phase-based displacement estimation, as their pronounced intensity variations and stable geometric characteristics improve the sensitivity and robustness of local motion extraction. Under practical imaging conditions, illumination variations and image noise may alter local intensity distributions and introduce local phase fluctuations. To improve robustness, multiple high-gradient structural regions are automatically identified within the ROI, and magnitude-weighted phase aggregation is subsequently performed to suppress local disturbances and ensure reliable displacement estimation under varying imaging conditions. This robustness is further validated by the parametric analyses presented in the numerical simulation section.

Given that wind turbine tower structures exhibit stable geometric boundaries and pronounced intensity gradients, the proposed method estimates the local structural response by extracting phase information from these high-gradient structural regions. Under the condition of fixing the vertical coordinate y_0 in the ROI (Region of Interest) image frames, the point (x_0, y_0) with the maximum magnitude is detected to precisely determine the extraction index for phase displacement, where the y_0 coordinate corresponds to half of the downsampled ROI’s vertical length. The magnitude-weighted phase difference is then used to extract local phase information at the indexed position (x_0, y_0). By iteratively applying this phase difference extraction between each current frame and the reference frame, the displacement signal of the image sequence is obtained.

The displacement obtained through the above procedure represents the motion component projected along the predefined measurement direction (i.e., the horizontal axis in the image coordinate system). In practical scenarios, structural responses may involve multi-degree-of-freedom motions, including vertical and rotational components. These coupled motions may introduce additional phase variations in other directions, which could influence the displacement estimation if not properly constrained. However, owing to the directional selectivity of the Q-Shift DTCWT and the orientation-specific phase projection employed in this study, the proposed method predominantly captures motion along the dominant vibration direction while suppressing contributions from orthogonal components. For wind turbine tower structures, where vibration is typically governed by a principal bending mode, the influence of multi-directional coupling on the estimated displacement is considered limited.

Finally, the acceleration time-history signal can be obtained by applying a second-order temporal differentiation to the measured displacement. However, direct numerical differentiation is highly sensitive to measurement noise and tends to significantly amplify high-frequency disturbances. To improve numerical stability and suppress noise amplification during the differentiation process, a Laplacian of Gaussian (LoG) operator is adopted, which effectively combines Gaussian smoothing with second-order differentiation[38].

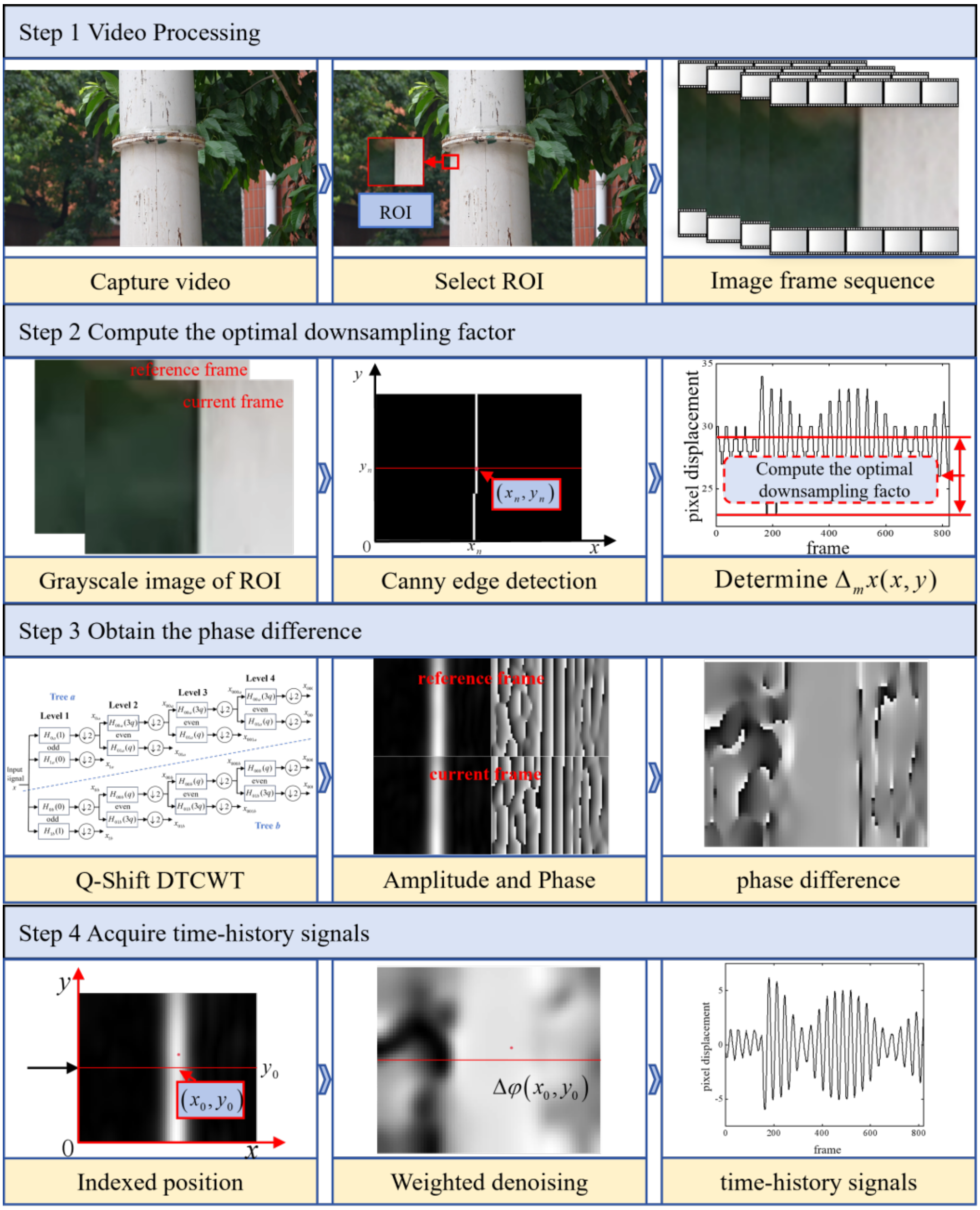


Fig. 3 Flowchart of the method

4. Numerical Simulation

4.1 Numerical Simulation Framework

To validate the effectiveness and robustness of the proposed method under controlled conditions, this section develops a fully controllable numerical simulation framework for generating synthetic vibration video sequences with known true displacements. On this basis, a comparative analysis of different displacement extraction methods is conducted.

Specifically, in the numerical simulation, the target in each frame is modeled as a two-dimensional Gaussian surface, whose grayscale distribution is expressed as

$$I(x,y,t)=A\cdot\exp\left(-\frac{\left(x-x_c(t)\right)^2}{2\sigma_x^2}-\frac{\left(y-y_c\right)^2}{2\sigma_y^2}\right) \tag{20}$$

where A is the amplitude of the Gaussian surface, σ_x and σ_y are the standard deviation parameters of the Gaussian surface in the horizontal and vertical directions, respectively, and $(x_c(t),y_c)$ denotes the center position of the Gaussian surface at time tt. At the initial moment, the center of the Gaussian surface is located at the image center (x_0,y_0), which serves as the reference frame. In subsequent frames, the center of the surface is only allowed to shift in the horizontal direction, while its vertical position remains unchanged. This setup constructs a typical vision-based vibration measurement scenario with known one-dimensional true displacement expressed in a two-dimensional image.As shown in Figure 4.

The horizontal displacement of the Gaussian surface center is given by a predefined damped harmonic function, expressed as:

$$\delta_x(t)=\sum_{k=1}^{K}A_k e^{-\varsigma_k\varpi_k t}\sin(\varpi_k t+\phi_k) \tag{21}$$

where K denotes the number of vibration frequency components. When $K=1$, it corresponds to a single-frequency vibration mode; when $K>1$, it represents a multi-frequency harmonic vibration mode. A_k is the displacement amplitude of the k-th frequency component, ϖ_n is the angular frequency, ς is the damping coefficient, and ϕ is the initial phase. Accordingly, the center position of the Gaussian surface is updated over time as:

$$x_c(t)=x_0+\delta_x(t),y_c(t)=y_0 \tag{22}$$

This approach directly maps the predefined one-dimensional temporal displacement function onto the overall translation of the two-dimensional grayscale surface, thereby avoiding the need for post-processing operations such as image-wide shifting or interpolation.

On the basis of generating the ideal vibration image sequence, this section further introduces multiple controllable perturbations in the image domain to simulate common imaging degradation factors in practical vision-based vibration measurement, thereby establishing a foundation for parametric analysis under various working conditions. First, zero-mean Gaussian white noise is superimposed on each frame to simulate sensor noise. The noise intensity is controlled by the noise standard deviation parameter, allowing the generation of video sequences under different signal-to-noise ratio conditions. Second, linear transformations are applied to the image grayscale to simulate brightness and contrast variations:

$$I'(x,y,t) = \alpha I(x,y,t) + \beta \tag{23}$$

where α is the contrast modulation coefficient and β is the brightness offset term. By varying the ranges of α and β, the effects of non-uniform illumination and dynamic range changes in grayscale on displacement extraction results can be systematically examined.

Based on the generated numerically simulated videos, the proposed method is employed to extract the displacement time-history curve. For comparison, several representative methods are introduced. Among them, the Gabor filter method [38] and the complex steerable pyramid method [39] are typical phase-based approaches, which are widely used in vision-based vibration measurement. In addition, a subpixel edge detection method based on Zernike moments [48] is included as a representative intensity-based approach. This selection enables a systematic comparison between different phase-based methods as well as between phase-based and intensity-based approaches.

Finally, to intuitively reflect the accuracy of each method, three evaluation metrics are adopted for analysis, forming a comprehensive numerical simulation framework. The first is the Pearson Correlation Coefficient (PCC), which quantifies the linear correlation between preset displacement data and data obtained by other methods, thereby assessing the consistency and similarity of the data. The second is the Root Mean Squared Error (RMSE), which measures the deviation between preset displacement data and data derived from other methods. The third is the Mean Absolute Error (MAE), which calculates the average absolute difference between the datasets.

The PCC is the quotient of the covariance between two sets of data and their standard deviations. The calculation formula is as follows:

$$P_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E\left[(X-\mu_X)(Y-\mu_Y)\right]}{\sigma_X \sigma_Y} \tag{24}$$

Here, $\mathrm{cov}(X,Y)$ represents the covariance of X and Y, σ_X and σ_Y denote the standard deviations of X and Y, E denotes the expectation, μ_X and μ_Y represent the means of X and Y, respectively.The absolute value of the PCC is always less than or equal to 1. The larger the absolute value, the stronger the correlation between the two datasets.

The formula for the RMSE can be expressed as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2} \tag{25}$$

A smaller RMSE indicates that the deviation between the predicted results and the actual values is smaller, suggesting better predictive performance. RMSE squares the errors, making it more sensitive to larger discrepancies.

The formula for the MAE can be expressed as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|X_i - Y_i\right| \tag{26}$$

MAE is similar to RMSE; a smaller value indicates a smaller deviation.

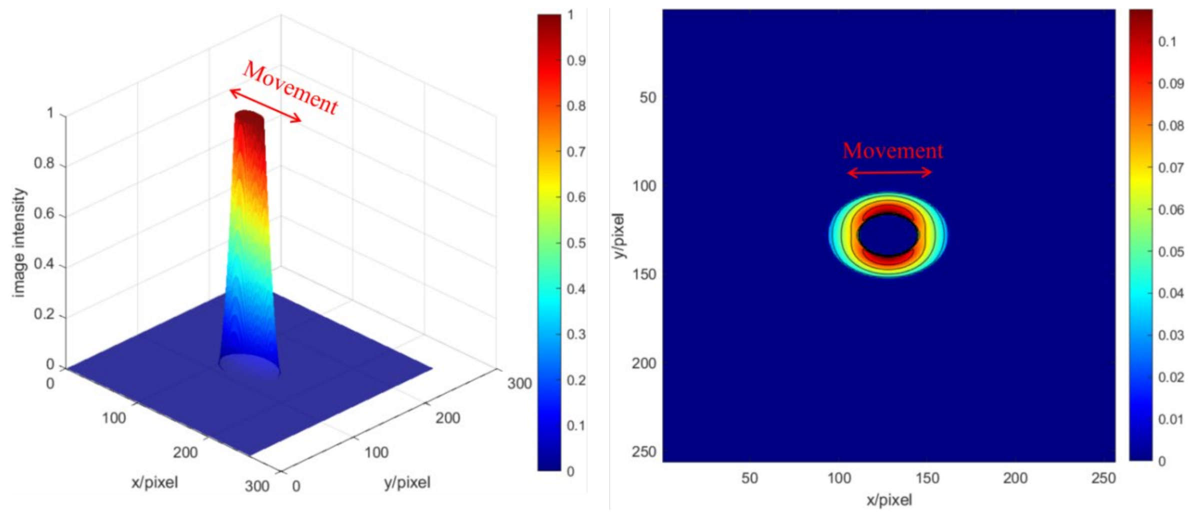


Fig. 4 The Gaussian surface

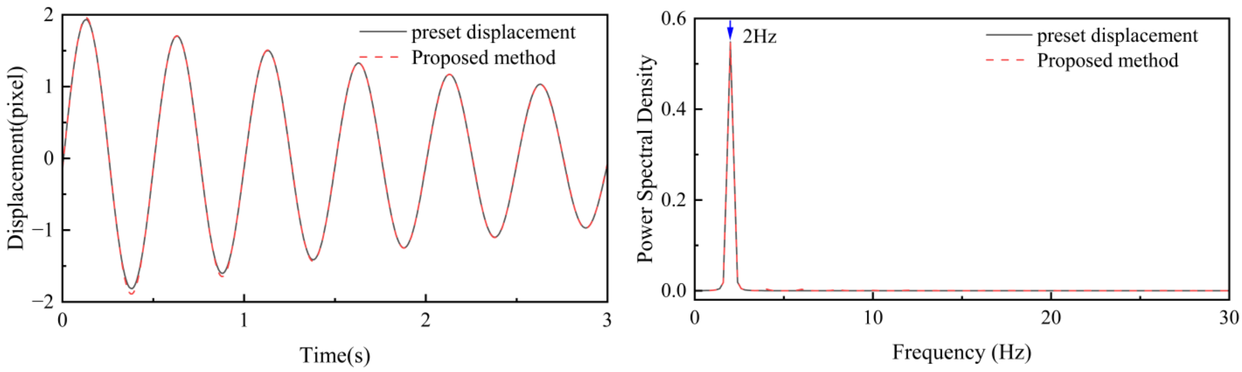
4.2 Method Validation

After establishing the complete numerical simulation framework, this section conducts numerical validation under noise-free conditions using single-frequency and multi-frequency vibration cases to verify the effectiveness and frequency resolution capability of the proposed vision-based vibration measurement method. By comparing the displacement extraction results under different vibration patterns, the accuracy of the proposed method in measuring single-frequency vibrations, as well as its capability to identify multiple frequency components under multi-frequency excitation, are systematically evaluated.

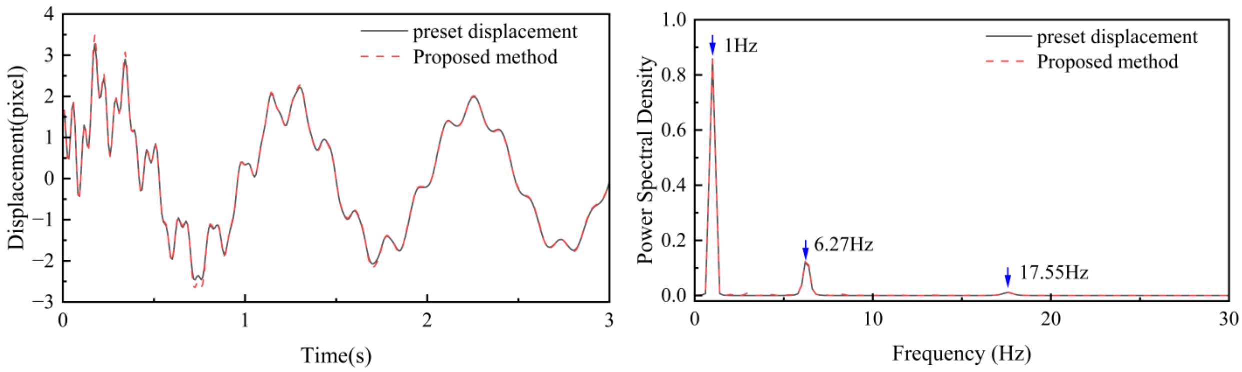
In the single-frequency validation case, a single damped harmonic displacement excitation is applied to the target structure along the horizontal direction. The vibration frequency is set to 2 Hz, with a displacement amplitude of 2.0 pixels and a damping ratio of 0.05. The corresponding video sequence is sampled at a frame rate of 60 Hz with a total duration of 3 s. This case is designed to assess the accuracy of the proposed method in extracting displacement time histories and dominant frequency components under a known single-frequency vibration condition, and it serves as a baseline for subsequent analyses under more complex vibration scenarios.

In the multi-frequency validation case, to simulate the participation of multiple structural modes in practical vibration responses, the displacement excitation is constructed as a linear superposition of multiple frequency components. The frequency ratios are selected according to the first three bending modes of a cantilever beam, corresponding to 1, 6.27, and 17.55 times the fundamental frequency. Considering that higher-order modes generally exhibit smaller vibration amplitudes in real structures, the displacement amplitudes of the frequency components are set to decrease progressively. Different initial phases and damping ratios are also introduced to increase the complexity of the vibration signal. To ensure reliable identification of higher-frequency components, the sampling frame rate of the video sequence is increased to 120 Hz, while the total duration is maintained at 3 s. This case aims to validate the capability of the proposed method to separate and identify different frequency components under multi-frequency excitation, particularly its performance in extracting higher-order vibration information. The identification results for both cases are presented in Fig. 5 and Table 1.

The results indicate that, under single-frequency excitation, the displacement time histories extracted by the proposed method are in excellent agreement with the predefined theoretical displacements, enabling highly accurate displacement measurement. Under multi-frequency vibration conditions, the displacement response contains multiple frequency components, resulting in increased signal complexity and posing higher demands on displacement extraction and frequency resolution. Nevertheless, the proposed method maintains a high level of identification accuracy and is capable of effectively capturing and separating multiple frequency components embedded in the displacement signal. The effectiveness and applicability of the proposed method are thus validated under both vibration scenarios. It should be noted that the multi-frequency simulation case includes higher-order vibration components with frequency ratios up to 17.55 times the fundamental frequency. The results demonstrate that the proposed method is capable of effectively capturing and separating multiple frequency components, including higher-order modes, under controlled conditions, indicating its capability for multi-mode vibration identification.



(a)Results for Single-Frequency Vibration



(b)Results for Multi-Frequency Vibration

Fig. 5 Method Validation Results

Table 1 Performance Metrics for Method Validation

Case	PCC	RMSE	MAE
single-frequency vibration	0.9999	0.0139	0.0096
multi-frequency vibration	0.9996	0.0350	0.0233

To evaluate the computational performance of the proposed method, a computational cost analysis is conducted under controlled numerical simulation conditions. Since the method operates on a frame-by-frame basis, the total computational time is primarily determined by the number of processed frames. Under a resolution of 256×256 and a duration of 3 s, the processing time is approximately 12.47 s for the single-frequency case (60 Hz, 180 frames), corresponding to about 0.069 s per frame. For the multi-frequency case (120 Hz, 360 frames) at the same resolution, the total time increases to 25.09 s, while the per-frame processing time remains similar at approximately 0.070 s. This indicates that the total computational time increases approximately linearly with the number of frames, while the per-frame processing time remains stable.

To further investigate the influence of image resolution, additional tests were conducted at a reduced resolution of 200×200. The results are summarized in Table 2.

Table 2 Computational time under different conditions

Case	Resolution	Frame Rate	Frames	Total Time (s)	Time per Frame(s)
single-frequency vibration	256×256	60Hz	180	12.47	0.069
multi-frequency vibration	256×256	120Hz	360	25.09	0.070
single-frequency vibration	200×200	60Hz	180	9.04	0.050
multi-frequency vibration	200×200	120Hz	360	17.89	0.050

It can be observed that the per-frame processing time remains approximately constant for a given resolution, while reducing the image size significantly improves computational efficiency. This observation highlights the importance of ROI selection in practical applications, as it directly reduces the image resolution and computational burden.

In practical applications, the wavelet decomposition level can influence the computational cost, as higher decomposition levels generally lead to increased computational complexity and processing time. However, in this study, the decomposition level is determined a priori by the optimal downsampling strategy described in Section 2 and remains fixed throughout the analysis. Therefore, it is not treated as a variable in the computational cost evaluation.

In terms of real-time performance, the proposed method achieves approximately 0.069 s per frame (about 14 fps) at 256×256 resolution and 0.050 s per frame (about 20 fps) at 200×200 resolution. Although strict real-time processing (e.g., 30 fps) is not fully achieved, the results demonstrate near real-time capability, indicating the potential applicability of the proposed method in practical structural monitoring systems.

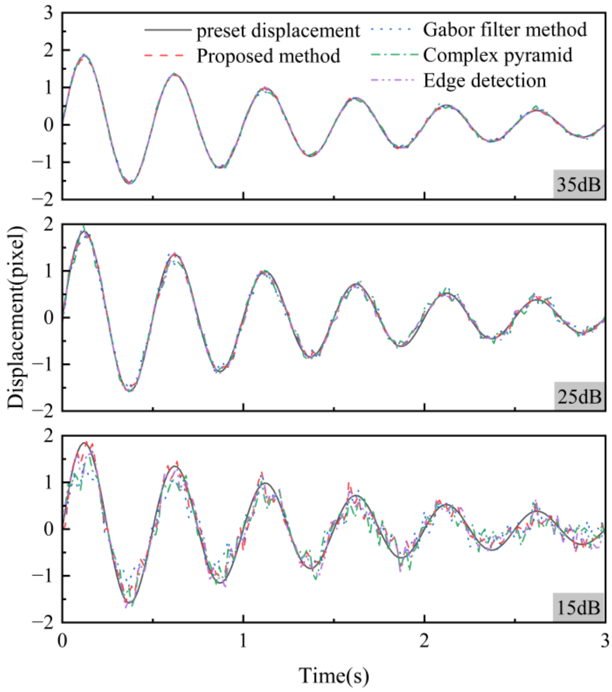
4.3 Parameter Sensitivity Analysis

In visual vibration measurement applications, imaging quality is affected by various degradation factors. Among them, imaging noise, illumination variation, and contrast degradation are the most common changes in imaging conditions, and they have a direct impact on displacement identification results. Imaging noise introduces random perturbations to the grayscale distribution and local structural information of images, thereby reducing the distinguishability of effective displacement information. Illumination variation alters the overall grayscale level of an image, which in turn affects the stability of image-based displacement extraction. In contrast, contrast degradation weakens the grayscale difference between the target and the background, leading to reduced discriminability of structural features in the image.

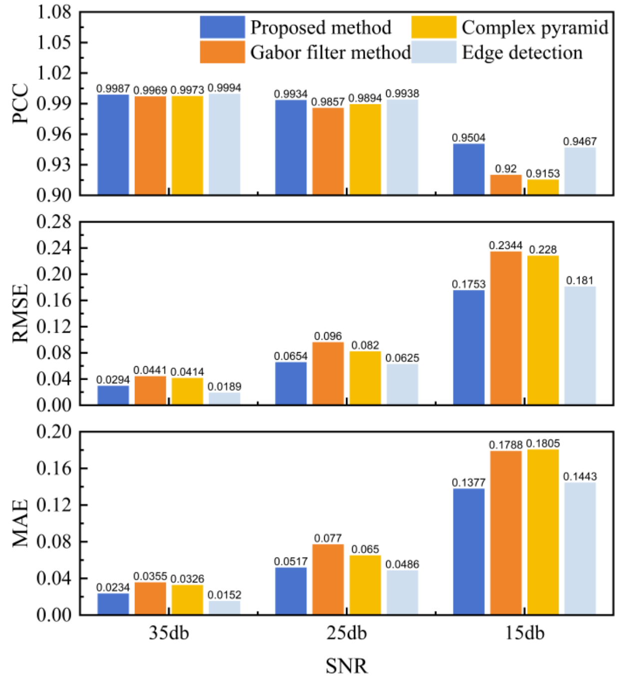
Based on the aforementioned numerical simulation framework, a total of seven parameterized working conditions are designed in this study. The first condition is taken as the baseline case. The second and third conditions simulate different noise levels by varying the signal-to-noise ratio (SNR, in dB), where 35 dB represents a typical imaging noise level, while 25 dB and 15 dB correspond to enhanced-noise and high-noise scenarios, respectively. The fourth and fifth conditions simulate different illumination environments by adjusting the normalized background grayscale level, with background brightness values of 0.3, 0.2, and 0.1 representing bright, typical, and dark imaging conditions, respectively. The sixth and seventh conditions simulate different contrast levels by varying the grayscale ratio between the target and the background, where the contrast level is quantified using the Michelson contrast formulation. The detailed parameter settings for all cases are summarized in Table 3. A systematic comparative analysis of the displacement extraction results under different imaging conditions is then conducted, and the corresponding identification results are presented in Fig. 6.

Table 3 Numerical simulation setup

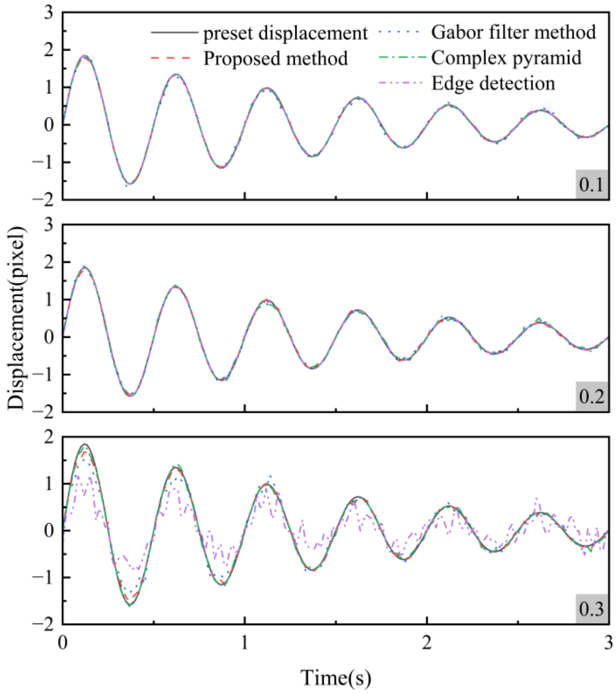
Case	SNR	Illumination level	Contrast variation
1	35	0.2	4.0
2	25	0.2	4.0
3	15	0.2	4.0
4	35	0.3	4.0
5	35	0.1	4.0
6	35	0.2	8.0
7	35	0.2	2.0

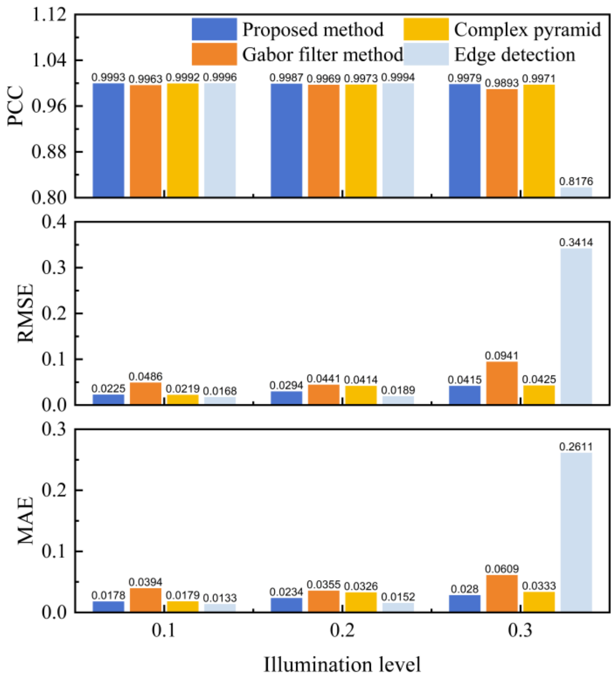


(a) Noise sensitivity analysis: displacement comparison



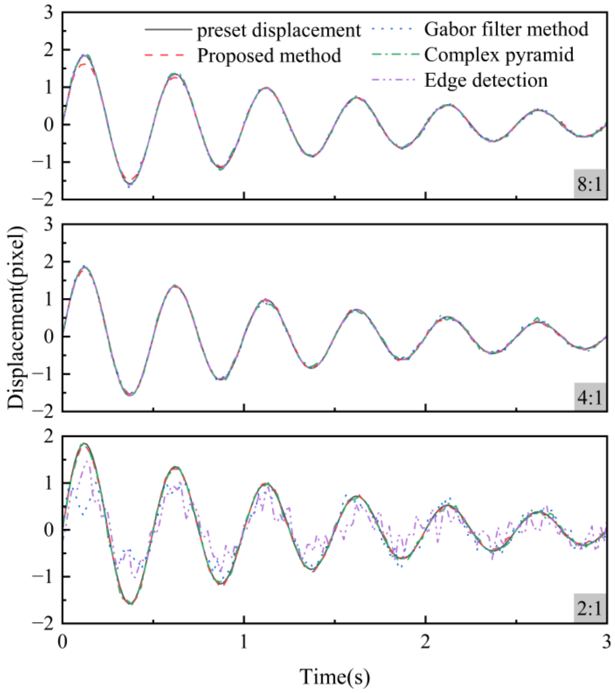
(b) Noise sensitivity analysis: metric results

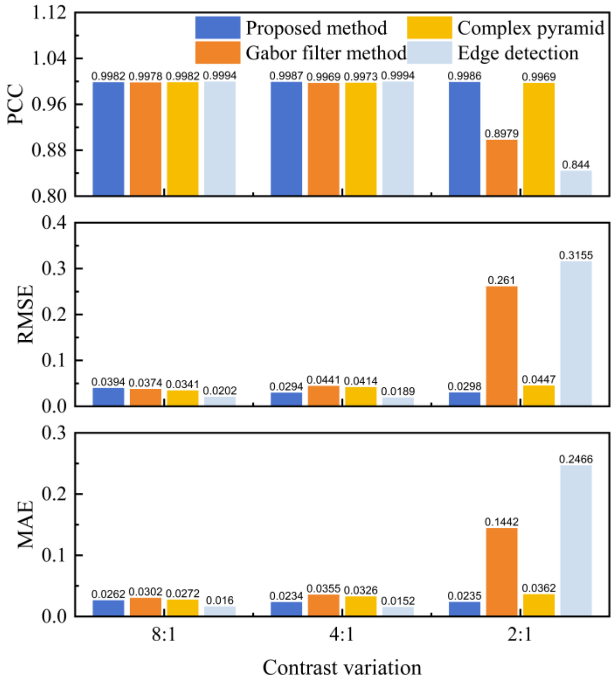




(c) Illumination sensitivity analysis: displacement comparison

(d) Illumination sensitivity analysis: metric results





(e) Contrast sensitivity analysis: displacement comparison

(f) Contrast sensitivity analysis: metric results

Fig. 6 Results of the numerical simulation

Under the above numerical simulation scenarios, the displacement identification performance of all methods exhibits a consistent degradation trend as the imaging conditions deteriorate, although the extent of performance degradation differs markedly among the methods. In the noise sensitivity analysis, when the signal-to-noise ratio is relatively high(e.g., 35 dB and 25 dB), all methods are able to achieve accurate displacement identification, and the differences among the three evaluation metrics remain small. Specifically, the PCC values of all four methods remain above 0.99, indicating a very high level of correlation with the ground truth displacement. Meanwhile, the RMSE and MAE values are both maintained at relatively low levels, demonstrating that all methods can provide reliable displacement estimation under low-noise conditions. As the noise level increases, the identification performance of all methods decreases; however, the proposed method shows the slowest performance degradation. For instance, the PCC values of the two phase-based methods decrease significantly to approximately 0.92 and 0.9153, respectively, whereas the proposed method still maintains a higher PCC of around 0.95, indicating a more robust correlation preservation capability under strong noise interference. In addition, the RMSE and MAE values of the proposed method remain noticeably lower than those of the phase-based methods, further confirming its superior accuracy and robustness. From a comparative perspective, the performance gap among the methods becomes more evident as noise intensifies. The phase-based methods using Gabor filtering and complex pyramid representations exhibit higher sensitivity to noise, leading to faster error accumulation and correlation degradation. In contrast, the proposed method and the intensity-based edge detection method demonstrate relatively stable performance, suggesting that they are less affected by noise perturbations, although the edge-based method still shows slightly higher errors than the proposed approach.

Under illumination and contrast variation scenarios, the differences in sensitivity to imaging condition changes among the methods become more pronounced. As the illumination level increases or the contrast between the target and the background decreases, the performance of the intensity-based edge detection method deteriorates significantly, while the proposed method is able to maintain relatively stable correlation and low error levels. For example, under high illumination or low contrast conditions, the PCC of the edge detection method drops sharply, accompanied by a substantial increase in RMSE and MAE, indicating its strong dependence on grayscale gradients and edge clarity. In contrast, the proposed method maintains consistently high PCC values (close to or above 0.97) and relatively low error metrics, demonstrating its robustness to intensity fluctuations. Furthermore, under extreme contrast variation (e.g., 2:1), the degradation of the edge detection method becomes particularly severe, while the proposed method still preserves acceptable accuracy, highlighting its adaptability to challenging visual conditions. This robustness can be attributed to the underlying displacement extraction mechanism, which is less dependent on absolute intensity values and more resilient to illumination and contrast distortions.

Overall, among the phase-based methods, the proposed approach demonstrates improved robustness against various imaging degradation factors, including noise, illumination variation, and contrast changes. Compared with the Gabor filter and complex pyramid methods, it exhibits more stable displacement identification performance and slower degradation under unfavorable conditions. In addition, compared with the intensity-based edge detection method, the proposed method maintains higher stability under illumination and contrast variations, where intensity-based approaches tend to suffer from significant performance degradation. These results indicate that the proposed method provides more reliable displacement estimation across different imaging conditions and achieves superior overall robustness compared with both phase-based and intensity-based approaches.

5. Experimental verification

5.1 Experimental procedure

To validate the effectiveness of the proposed method, this section presents a vibration characteristic measurement experiment conducted on a scaled steel wind turbine tower model. The experimental model was constructed using Q345B steel with the following material properties: elastic modulus of 210 GPa, Poisson's ratio of 0.3, and density of 7,850 kg/m^3.

The structural configuration consists of a 2.2 m × 2.25 m steel foundation base serving as the supporting platform. The main tower structure comprises three vertically tapered cylindrical sections with uniform wall thickness of 6 mm, featuring individual segment lengths of 1.75 m (bottom section), 1.85 m (middle section), and 1.85 m (top section). Adjacent tower segments were interconnected via high-strength flange bolted connections, with the inner diameters of successive joints progressively decreasing from base to top: 500 mm (foundation-bottom section interface), 432 mm (bottom-middle section interface), 366 mm (middle-top section interface), and 300 mm (top-wind turbine blade section interface). A simulated turbine blade inertial mass assembly was mounted atop the tower to establish an integrated tower-turbine coupled system, as illustrated in Figure 7.

This hierarchical configuration achieves both gradual stiffness gradient variations along the tower height and accurate simulation of actual wind turbine tower mechanical characteristics through parametric design of flange connections. The systematic structural design provides a reliable experimental platform for comprehensive vibration response testing, effectively replicating the dynamic behavior of full-scale wind turbine structures while maintaining geometric and mechanical similitude requirements.

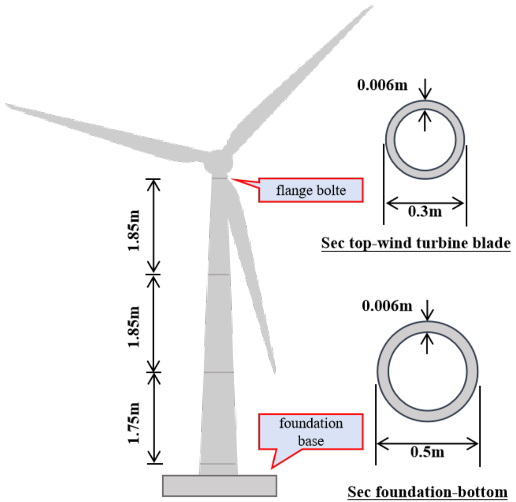


Fig.7 Wind turbine tower model

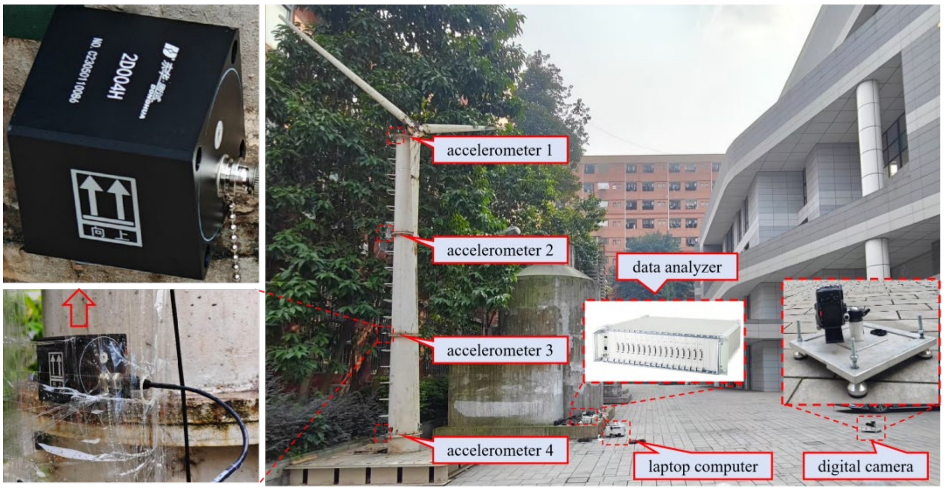


Fig.8 Experimental setup diagram

The experimental system setup is illustrated in Figure 8. The acceleration acquisition system primarily consisted of a 2D004H lateral accelerometer and a DH5922D high-performance dynamic signal test and analysis system. At each flange connection of the tower model, accelerometers were fixed using hot-melt adhesive, with their measurement axes aligned parallel to the circumferential tangent direction of the tower shell. The acceleration sampling frequency was configured at 100 Hz. The video acquisition system was centered around a Sony FX30 camera equipped with a Sony FE 24-105mm F4 lens. The digital camera was fixedly mounted along the extended line connecting the accelerometer and the tower center, positioned 10 m from the tower center, thereby satisfying the required horizontal "three-point collinear" relationship among the camera, the tower center, and the accelerometer installation point, as shown in Figure 9. The camera resolution was set to 1920×1080, with a video sampling rate of 100 Hz adhering to the Nyquist sampling theorem. The focal length was optimized based on the shooting distance, while the aperture and shutter operated in automatic mode to ensure imaging quality. Although automatic exposure adjustment may introduce slight frame-to-frame intensity variations, the influence on the proposed phase-based displacement estimation is expected to be limited, as the robustness against illumination variations has been verified through the parametric analyses presented in Section 4.3. Prior to recording, the camera was vertically aligned to minimize image distortion, and leveling bolts on the base were adjusted to ensure system stability.

All experimental equipment underwent rigorous calibration, with both power supply and data transmission systems operating normally. Horizontal excitation in the X-direction was applied using a hammer impact method. To minimize synchronization errors between the vision system and the accelerometer, the camera was wirelessly connected to a smartphone, allowing the video recording to be remotely controlled through the smartphone application, while the accelerometer data acquisition was initiated using the acquisition software on a laptop. Repeated preliminary tests indicated that the startup time difference between the two systems was approximately 0.1 s. Since manual synchronization cannot guarantee perfectly simultaneous data acquisition, the displacement signals obtained from the vision system and the accelerometer were further temporally aligned using the corresponding response peaks before calculating the quantitative evaluation metrics. Following the application of excitation, personnel promptly evacuated the test area to avoid interference from human activity.

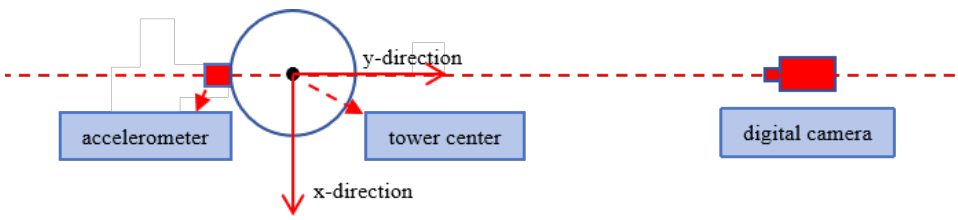


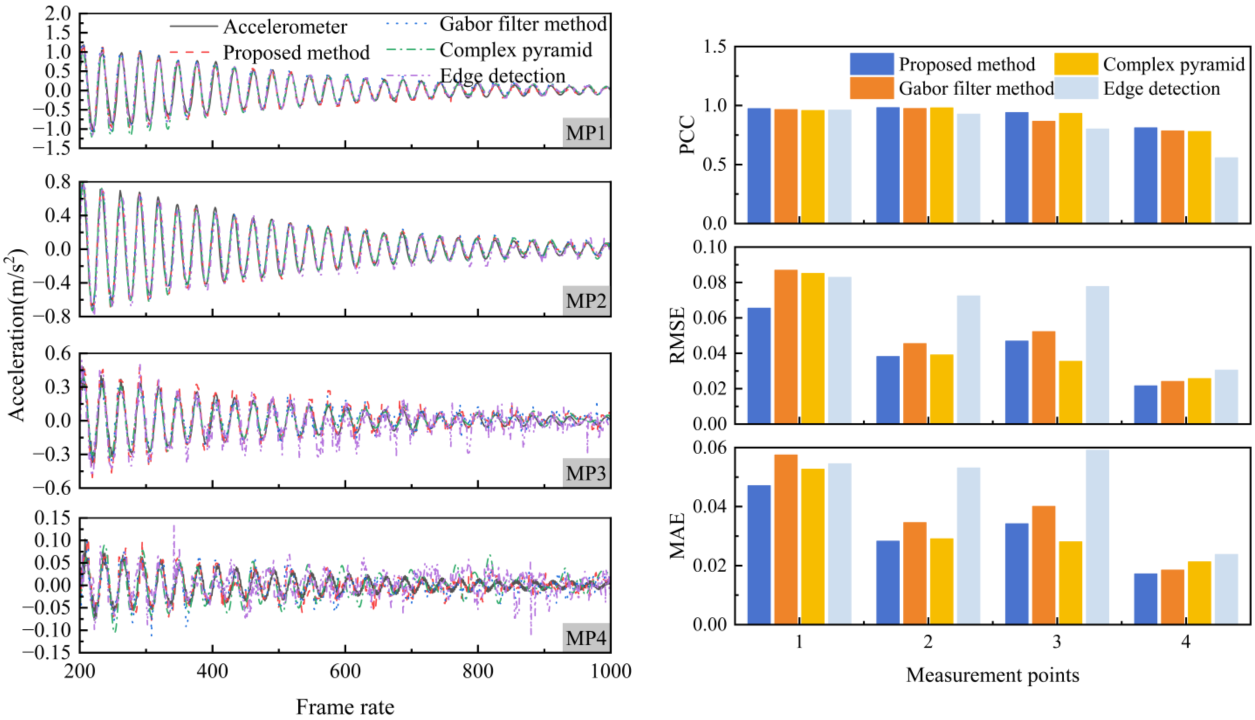
Fig.9 three-point collinear

5.2 Results analysis

Based on the measurement positions of four accelerometers, this section selected four regions of interest (MP1, MP2, MP3, and MP4) from the video captured by a digital camera and extracted the acceleration time-history curves using the proposed method for comparative analysis with the accelerometer-measured data. To validate the effectiveness of the method, three image-based vibration displacement measurement techniques were also employed for comparison. Among them, the Gabor filter method and the complex steerable pyramid method are representative phase-based approaches, while the subpixel edge detection method based on Zernike moments is a typical intensity-based approach.

The time-domain signal comparison results shown in Figure 10(a) indicate that at measurement points 1 and 2, all four methods yielded results consistent with the accelerometer data. However, at measurement points 3 and 4, the subpixel edge detection method exhibited significant deviations. In contrast, the proposed method demonstrated superior noise suppression and measurement accuracy across all measurement points, with its time-history curves showing the best agreement with the accelerometer data.

Based on the actual measurement data from the accelerometer, a comparative analysis of various evaluation metrics for the visual vibration displacement measurement methods is presented in Table 4. It can be observed that, except for the sub-pixel edge detection method, which exhibited significant errors at measurement points 3 and 4, the fundamental vibration frequency measured by all methods is 3.563 Hz, consistent with the results obtained from the accelerometer. At all four measurement points, the Pearson correlation coefficients calculated using the method proposed in this paper are higher than those of the other methods, indicating a greater degree of agreement with the accelerometer data and demonstrating a higher accuracy in vibration displacement measurement. With the exception of measurement point 3, the root mean square error and mean absolute error calculated by the proposed method are lower than those of the other methods, suggesting that the proposed method exhibits good stability and robustness during the measurement process.



(a) Comparison of Time-History Curves (b) Comparison of Performance Metrics

Fig.10 Model Test Results

In summary, the proposed method demonstrates better agreement with the accelerometer data compared with the other phase-based methods, including the Gabor filter and complex pyramid approaches. Meanwhile, the subpixel edge detection method exhibits noticeable deviations at certain measurement points, whereas the proposed method maintains more consistent performance across all locations.

It should be noted that the vibration response in the experiment is dominated by lower-order modes due to the hammer excitation, while higher-order modes receive limited energy input and therefore exhibit significantly smaller amplitudes. Under outdoor experimental conditions, environmental factors such as illumination variation and background disturbances introduce additional noise into the image sequence. Therefore, the limited identification of higher-order modes in the experiment appears to be influenced by the excitation characteristics and sampling constraints. Meanwhile, further investigation is needed to fully assess the potential limitations of the proposed method, and relevant improvements concerning experimental conditions, algorithm robustness, and higher-order mode identification will be detailed in our subsequent paper.

Table 4 Calculation Results of Evaluation Metrics

measurement points	methods	first-order natural	PCC	RMSE	MAE

		frequency			
1	Proposed method	3.563	0.9743	0.0655	0.0471
	Gabor filter method	3.563	0.9656	0.0869	0.0575
	Complex pyramid	3.563	0.9579	0.0851	0.0527
	Edge detection	3.563	0.9617	0.0829	0.0545
2	Proposed method	3.563	0.9818	0.0382	0.0283
	Gabor filter method	3.563	0.9748	0.0455	0.0346
	Complex pyramid	3.563	0.9801	0.0391	0.0291
	Edge detection	3.563	0.9275	0.0724	0.0531
3	Proposed method	3.563	0.9407	0.0469	0.0342
	Gabor filter method	3.563	0.8668	0.0522	0.0401
	Complex pyramid	3.563	0.9332	0.0355	0.0281
	Edge detection	3.500	0.8018	0.0777	0.0591
4	Proposed method	3.563	0.8117	0.0216	0.0172
	Gabor filter method	3.563	0.7851	0.0241	0.0185
	Complex pyramid	3.563	0.7799	0.0257	0.0213
	Edge detection	3.625	0.5574	0.0305	0.0238

6. Conclusion

This paper reviews the applicability of existing vibration displacement measurement techniques and proposes a non-contact vibration displacement measurement method for wind turbine towers based on the phase information of the Q-shift dual-tree complex wavelet transform. Through theoretical analysis, numerical simulations, sensitivity studies, and model-scale experimental validation, the following conclusions are drawn:

(1) The proposed Q-shift dual-tree complex wavelet phase-based method effectively exploits the approximate shift invariance, reduced aliasing, and multi-directional selectivity of the Q-shift dual-tree complex wavelet transform, enabling stable and accurate vibration displacement estimation for wind turbine tower structures. The method demonstrates clear advantages in complex imaging scenarios and multi-point synchronous vibration measurement, highlighting its potential for large-scale structural monitoring applications.

(2) To mitigate phase wrapping effects encountered in conventional phase-based measurement methods under large displacement conditions, an adaptive image downsampling strategy guided by Canny edge detection is employed. A theoretical relationship between the maximum inter-frame displacement and the minimum required downsampling ratio is established, thereby extending the applicability of phase-based

approaches to large-amplitude vibration measurement scenarios. Phase extraction is conducted within manually selected regions of interest exhibiting pronounced amplitude responses and distinct structural edges, where phase stability is enhanced. This strategy improves the reliability and practical applicability of the proposed method in engineering vibration measurement.

(3) Numerical simulations, sensitivity analyses, and model-scale wind turbine tower experiments are conducted to systematically evaluate the robustness and reliability of the proposed method. Sensitivity analyses with respect to image noise, illumination variation, and contrast degradation demonstrate that the proposed method maintains stable and accurate displacement estimation under imaging degradation conditions. Comparative experimental results further confirm the superior performance of the proposed approach over classical image phase-based vibration measurement methods in terms of measurement accuracy, noise robustness, and multi-point synchronous measurement capability.

Acknowledgments

This research was supported by the Research Funds for the Research Project of Zhejiang Provincial Department of Transportation, Intelligent Detection Method for Mechanical Indicators of Bridge Structures (No. 2025016), and the Guizhou Province Academician Workstation for Research and Application of Intelligent Equipment Technology in Civil Engineering (Qiankehe Platform KXJZ (2025) 004).

References

[1] T. Jiang, P. Lv, D. Li, A new shape reconstruction method for monitoring the large deformations of offshore wind turbine towers, Ocean Engineering 312 (2024) 119253.

[2] J. Chen, M. Zhang, H. Zhu, et al., Study on mechanical properties of wind turbine tower reinforced by ribbed interlayer, Journal of Constructional Steel Research 212 (2024) 108256.

[3] B. Moynihan, A. Mehrjoo, B. Moaveni, et al., System identification and finite element model updating of a 6 MW offshore wind turbine using vibrational response measurements, Renewable Energy 219 (2023) 119430.

[4] A. Cadoret, E. Denimal-Goy, J.M. Leroy, et al., Damage detection and localization method for wind turbine rotor based on operational modal analysis and anisotropy tracking, Mechanical Systems and Signal Processing 224 (2025) 111982.

[5] M. Serrano-Antoñanazas, J.E. Sierra-Garcia, M. Santos, et al., Identification of vibration modes in floating offshore wind turbines, Journal of Marine Science and Engineering 11(10) (2023) 101893.

[6] Y. Zhao, X. Li, X. Wang, et al., Fatigue life evaluation of wind turbine tower based on measured data, Advances in Civil Engineering (2023).

[7] L. Lin, X. Zhang, D. Zhang, et al., Damage evolution and failure analysis of the advanced transition segment behavior of wind turbine tower, Engineering Failure Analysis 152 (2023) 107527.

[8] Q. Yao, H. Bing, G. Zhu, et al., A novel stochastic process diffusion model for wind turbines condition monitoring and fault identification with multi-parameter information fusion, Mechanical Systems and Signal Processing 214 (2024) 111397.

[9] K. Lin, B. Ma, S. Yang, et al., Experimental study of wind turbine in operation under wind and seismic excitations with soil–structure interaction, Engineering Structures 308 (2024) 117927.

[10] M. Simoncelli, M. Zucca, M. Ghilardi, Structural health monitoring of an onshore steel wind turbine, Journal of Civil Structural Health Monitoring 14(6) (2024) 1423-1437.

[11] W. Jiang, J. Wu, C. Wang, et al., Health assessment of wind turbine gearbox via parallel ensemble and fuzzy derivation collaboration approach, Advanced Engineering Informatics 62 (2024) 102576.

[12] C.R. Farrar, K. Worden, An introduction to structural health monitoring, Philosophical Transactions of the Royal Society A 365(1851) (2007) 303-315.

[13] J. Baqersad, P. Poozesh, C. Niezrecki, et al., Photogrammetry and optical methods in structural dynamics - A review, Mechanical Systems and Signal Processing 86 (2017) 17-34.

[14] Y. Xu, J.M.W. Brownjohn, Review of machine-vision based methodologies for displacement measurement in civil structures, Journal of Civil Structural Health Monitoring 8 (2018) 91-110.

[15] P. Poozesh, J. Baqersad, C. Niezrecki, et al., Large-area photogrammetry based testing of wind turbine blades, Mechanical Systems and Signal Processing 86 (2017) 98-115.

[16] H. Huang, N. Baddour, M. Liang, Bearing fault diagnosis under unknown time-varying rotational speed conditions via multiple time-frequency curve extraction, Journal of Sound and Vibration 414 (2018) 43-60.

[17] P. Kohut, K. Holak, T. Uhl, et al., Monitoring of a civil structure's state based on noncontact measurements, Structural Health Monitoring 12 (2013) 411-429.

[18] C. Gentile, G. Bernardini, An interferometric radar for non-contact measurement of deflections on civil engineering structures: Laboratory and full-scale tests, Structure and Infrastructure Engineering 6 (2009) 521-534.

[19] Y. Fukuda, M.Q. Feng, M. Shinozuka, Cost-effective vision-based system for monitoring dynamic response of civil engineering structures, Structural Control and Health Monitoring 17 (2010) 918-936.

[20] L. Luo, M.Q. Feng, Z.Y. Wu, Robust vision sensor for multi-point displacement monitoring of bridges in the field, Engineering Structures 163 (2018) 255-266.

[21] P. Frankovský, I. Delyová, P. Sivák, et al., Modal analysis using digital image correlation technique, Materials 15(16) (2022) 5658.

[22] M.M. Wang, K.Y. Koo, C.Y. Liu, et al., Development of a low-cost vision-based real-time displacement system using Raspberry Pi, Engineering Structures 278 (2023) 115493.

[23] J. Zhong, Y. Shi, Q. Lin, et al., A vision-based method for displacement measurement using cross-sinusoidal composite fringe pattern, Mechanical Systems and Signal Processing 237 (2025) 113067.

[24] G. Liu, C. He, C. Zou, et al., Displacement measurement based on UAV images using SURF-enhanced camera calibration algorithm, Remote Sensing 14(23) (2022) 6008.

[25] P. Pan, C. Xing, J. Bai, et al., A remote deflection detection method for long-span bridges using adaptive mask and high-resolution camera, Measurement 201 (2022) 111774.

[26] W. Du, D. Lei, Z. Hang, et al., Short-distance and long-distance bridge displacement measurement based on template matching and feature detection methods, Journal of Civil Structural Health Monitoring 13 (2023) 343-360.

[27] C.Z. Dong, O. Celik, F.N. Catbas, Marker-free monitoring of the grandstand structures and modal identification using computer vision methods, Structural Health Monitoring 18(5-6) (2019) 1491-1509.

[28] J. Zhu, Z. Lu, C. Zhang, A marker-free method for structural dynamic displacement measurement based on optical flow, Structural and Infrastructure Engineering 18(1) (2021) 84-96.

[29] P. Li, H. Yu, R. Guo, H. Xia, et al., Full-field displacement measurements of structural vibrations using a novel two-stage neural network, Measurement 242 (2025) 115884.

[30] A. Javed, H. Lee, B. Kim, et al., Vibration measurement of a rotating cylindrical structure using subpixel-based edge detection and edge tracking, Mechanical Systems and Signal Processing 166 (2022) 108437.

[31] W. Wang, D. Cui, C. Ai, et al., Target-free recognition of cable vibration in complex backgrounds based on computer vision, Mechanical Systems and Signal Processing 197 (2023) 11039.

[32] Y. Yang, C. Dorn, M. Mancini, et al., Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification, Mechanical Systems and Signal Processing 85 (2017) 567-590.

[33] D.H. Diamond, P.S. Heyns, A.J. Oberholster, Accuracy evaluation of sub-pixel structural vibration measurements through optical flow analysis of a video sequence, Measurement 95 (2017) 166-172.

[34] S. Collier, T. Dare, Accuracy of phase-based optical flow for vibration extraction, Journal of Sound and Vibration 535 (2022) 117112.

[35] K. Luo, X. Kong, J. Li, et al., Motion magnification for video-based vibration measurement of civil structures: A review, Mechanical Systems and Signal Processing 220 (2024) 111681.

[36] Y. Miao, J.Y. Jeon, Y. Kong, et al., Phase-based displacement measurement on a straight edge using an optimal complex Gabor filter, Mechanical Systems and Signal Processing 164 (2022) 108224.

[37] Y. Miao, Y. Kong, H. Nam, et al., Phase-based vibration imaging for structural dynamics applications: Marker-free full-field displacement measurements with confidence measures, Mechanical Systems and Signal Processing 198 (2023) 110418.

[38] C. Peng, C. Zeng, Y. Wang, Camera-based micro-vibration measurement for lightweight structure using an improved phase-based motion extraction, IEEE Sensors Journal 20(5) (2019) 2590-2599.

[39] J.G. Chen, N. Wadhwa, Y.J. Cha, et al., Modal identification of simple structures with high-speed video using motion magnification, Journal of Sound and Vibration 345 (2015) 58-71.

[40] Z.X. Ma, B.R. Wang, X. Jing, et al., Computer vision-based cross-scale structural displacement estimation using amplitude–phase fusion, Mechanical Systems and Signal Processing 244 (2026) 113738.

[41] G. Liu, M.Z. Li, Z. Mao, et al., Structural motion estimation via Hilbert transform enhanced phase-based video processing, Mechanical Systems and Signal Processing 166 (2022) 108418.

[42] M. Li, G. Liu, Z. Mao, Phase-based motion estimation through short-distance Hilbert transform, Mechanical Systems and Signal Processing 211 (2024) 111219.

[43] I.W. Selesnick, R.G. Baraniuk, N.G. Kingsbury, The dual-tree complex wavelet transform, IEEE Signal Processing Magazine 22(6) (2005) 123-151.

[44] X. Wang, Y. Qin, et al., Subband averaging kurtogram with dual-tree complex wavelet packet transform for rotating machinery fault diagnosis, Mechanical Systems and Signal Processing 142 (2020) 106740.

[45] Y. Zhang, H. Li, et al., Computationally efficient UAV fault diagnosis with adaptive vibration denoising: A signal processing approach for rotorcraft systems, Mechanical Systems and Signal Processing 240 (2025) 112345.

[46] N. Kingsbury, The dual-tree complex wavelet transform: A new efficient tool for image restoration and enhancement, in: Proceedings of the 9th European Signal Processing Conference (EUSIPCO), IEEE, 1998, pp. 1-4.

[47] D.J. Fleet, A.D. Jepson, Computation of component image velocity from local phase information, International Journal of Computer Vision 5 (1990) 77-104.

[48] S. Ghosal, R. Mehrotra, Orthogonal moment operators for subpixel edge detection, Pattern Recognition 26(2) (1993) 295-306.