

Article

Label-Free Computer Vision Method for Measuring the Natural Frequencies of Tall Structures

Duo Chen ¹, Ziqi Quan ¹, Yonghong Zhang ¹, Qiang Gao ¹, Bo Jin ¹, Zhen Zhang ², Zexuan Li ³ and Qing Sun ^{3,*}

¹ Ordos Power Supply Company, Inner Mongolia Electric Power (Group) Co., Ltd., Ordos 017000, China; chenduogod@163.com (D.C.); m15598825512@163.com (Z.Q.); 19104779575@163.com (Y.Z.); hanghaidadao115@163.com (Q.G.); ning_xi_82@163.com (B.J.)

² Inner Mongolia Power (Group) Co., Ltd., Hohhot 010010, China; 15149007590@163.com

³ Department of Civil Engineering, Xi'an Jiaotong University, Xi'an 710049, China; robin.harden@stu.xjtu.edu.cn

* Correspondence: sunq@mail.xjtu.edu.cn

Abstract

Recent years have seen significant advancements in the field of structural health monitoring (SHM) using computer vision, which has steadily developed into a practical and effective technique for measuring the dynamic properties of structures. With the long-range, non-contact, and easy-to-use features of this measurement technology, some difficulties associated with utilizing conventional techniques to detect the natural frequencies of tall structures can be reduced. Because it is label-free, the method also avoids attaching physical targets or sensors to energized, difficult-to-access structures. This paper proposes a label-free computer vision measurement method, which uses a label-free detection system to determine the tracking feature points, and then tracks the feature points of the structure based on the KLT optical flow method. Finally, the natural frequencies of the structure are obtained by frequency domain analysis of the vibration signal of the feature points. Experimental studies on indoor transmission tower models and outdoor high street lamps were conducted to confirm the viability of the above technology. The test findings were compared with the measurements from a 941B accelerometer, and the impact of several aspects, such as frame rate, resolution, and measuring distance, on the accuracy of the results was examined. The test findings show that the label-free system identifies the first several natural frequencies, including the fundamental, with the modal-frequency estimates agreeing with the 941B accelerometer to within roughly 0.02–0.04 Hz across the first three modes; the larger percentage error at the fundamental reflects its low frequency rather than reduced accuracy, since the absolute discrepancies are comparable across all modes. The study's findings may serve as a guide for developing software that will be used in the future to assess structural health monitoring using computer vision.



Academic Editor: Tiago Filipe da Silva Miranda

Received: 6 June 2026

Revised: 5 July 2026

Accepted: 10 July 2026

Published: 14 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: computer vision; structural health monitoring; label-free; tall structures; non-contact measurement

1. Introduction

Strong winds, earthquakes, snowstorms, precipitation, snowfall, and natural aging over time can all harm civil engineering structures, particularly tall ones. Tall buildings, including skyscrapers, TV towers, transmission towers, and other structures, are vulnerable to wear and tear from a variety of natural disasters during their service lives and may even collapse under extremely high loads [1]. The safety of structures is becoming more and

more important as the number of tall buildings in the world rises. In the discipline of civil engineering, structural health monitoring (SHM) is one of the subjects that has garnered a lot of attention and is particularly significant. The primary goals of SHM are to measure, identify, and analyze structures under load in order to assess their safety and performance condition, as well as to identify possible hazards and damage in building structures early on [2,3].

In SHM, monitoring a structure's dynamic properties requires knowing its natural frequencies [4,5]. In order to do this, researchers have created a number of methods; however, there are still barriers and restrictions with regard to data collection in practice [6]. Numerous techniques for gathering data have surfaced in recent years; the two primary ones are contact measurement and non-contact measurement. In order to collect data for traditional contact measurement, physical sensors such as accelerometers, strain sensors, and stress sensors are typically used. However, installing these sensors can be labor- and time-intensive, particularly for tall buildings. It takes a lot of work to install sensors and wires [7]. Using high-voltage transmission towers as an example, these large, tall buildings present a number of challenges, including difficult deployment, inconvenient operation, and high costs associated with placing sensors throughout the tower body. On the other hand, transmission towers are typically powered by high-voltage electricity while in use, and their measurement window is frequently restricted year-round.

Consequently, non-contact measurement-based computer vision technology, such as lidar, global navigation satellite system (GNSS), and laser displacement sensor (LDS) [8–10], has been widely applied in civil engineering, particularly in structural health monitoring [11]. Specifically, SHM of engineered structures makes extensive use of GPS [12]. The aforementioned techniques do, however, still have certain drawbacks, such as the need for costly machinery and intricate procedures. In the field of structural health monitoring, vision-based approaches are widely used because of their ease of operation, higher accuracy, and strong cost-effectiveness when compared to other non-contact measuring techniques [13]. Measurements of object displacement have been made with a lot of utility in recent years using computer vision-based techniques [14]. Conceptually, the three main components of this method are feature recognition and matching, tracking algorithm establishment, and camera parameter calibration. When shooting video, the lens's position during the shooting process is simultaneously obtained, and the measurement position—also known as the target point—is identified and tracked frame by frame in a video sequence. The displacement signal is corrected, and finally, the obtained data is subjected to denoising and then conversion by the proportional factor method to obtain the displacement time history curve of the structure. The development of several widely used key point detection algorithms, such as Harris Corner Features [15], Scale Invariant Feature Transform (SIFT) [16], Fast Retina Key points (FREAK) [17], Accelerated Robust Features (SURF) [18], etc., has made it possible to apply the aforementioned methods in practice. These algorithms can be used for feature point detection of various structures and can also perform feature matching frame by frame to obtain displacement information; On the other hand, for tracking algorithms, researchers mostly use structural motion tracking algorithms based on optical flow. This algorithm can calculate the motion vector within the monitoring area based on the principles of optical flow and motion field. Currently, there are also many computational models based on optical flow in this field [19], which are widely used. For example, Yoon et al. [20] used a video of a steel frame shaking table test recorded by a consumer-grade camera and performed system identification using the KLT optical flow tracking algorithm [21] and the Shi–Tomasi corner detection method [22] to obtain the displacement signal of the steel frame. Dong et al. [23] calculated the full-field optical flow based on FlowNet2 [24] and combined it with kernel functions to monitor multi-

point displacements; Ji et al. [25] proposed a method for measuring the two-dimensional displacement of small cables without targets using an optical flow algorithm; Morliler J et al. [26] introduced a practical vision-based two-dimensional vibration measurement system with a KLT tracker for estimating the displacements of simple Oberst beams and the first two main modes of helicopter blades; Kuddus et al. [27] used the KLT algorithm to systematically identify RC columns through shaking table tests. In the above studies, researchers have proved through a large number of indoor experiments that the use of computer vision methods for systematic identification of structures is feasible and have also created application space for the application of visual methods to tall structures.

Furthermore, modal recognition of structures based on deep learning and computer vision has been extensively studied. While some researchers establish sparse and dense point cloud models using binocular vision and geometric and optical principles to obtain the three-dimensional coordinates of the structure and use them to model [28], others introduce edge detection algorithms to improve the accuracy of the results [29]. The majority of them use prefabricated neural networks to automatically extract and detect features from images [30–32]. This certainly expands computer vision's potential applications. Nonetheless, there has been a lack of research on the application of different computer vision techniques in high-rise buildings. The majority of research is restricted to validating experiments conducted indoors, such as steel frames, simply supported beams, and steel beams; it ignores high-rise structures, which are mostly exposed to wind loads and other random environmental excitations outside. Furthermore, in SHM, synthetic physical targets are frequently positioned on the structure to improve the feature point identification algorithm's accuracy. However, there are many inconveniences in arranging artificial targets for high-rise structures such as skyscrapers, high-voltage transmission towers, and signal towers.

In this research, a label-free computer vision method for determining the natural frequencies of tall structures is proposed. With this approach, the structure is photographed for a predetermined amount of time using a consumer-level camera. The visual system is then used to identify, extract, and track the structural feature points. Frame per frame, the target point coordinates in the image are collected and transformed into the matching dynamic displacement (acceleration) time history curve. This curve is then transformed using the fast Fourier transform (FFT) into a spectrum diagram for frequency domain analysis. A verification test of an indoor model transmission tower and an outdoor high-street lamp wind load simulation test were set up to simulate the tall structure's service process under environmental stimulation in order to confirm the efficacy, accuracy, and feasibility of the suggested method. The test content primarily consists of a comparison between the visual system's measurement results and those of the conventional accelerometer 941B. It also investigates the impact of various factors on the accuracy of the results, including the measuring distance between the camera and the structure, the pixel or frame rate of the instrument itself, etc.

2. Materials and Methods

The acquisition of the natural frequencies of the structure using the label-free computer vision method is mainly divided into four steps. As shown in Figure 1, the camera parameters are first calibrated to correct the distortion caused by the lens, and then the tracking target is determined using the label-free detection system. The KLT optical flow tracking method is used to track the target point and the reference point, and the position coordinates of each frame are determined. Error correction is performed, and pixel coordinates are converted to real coordinates to obtain the displacement time history curve of the target point. Finally, data post-processing and fast Fourier (FFT) transformation are

performed to achieve the measurement of the natural frequencies of the structure. Although the first natural frequency is the primary quantity of interest, the same procedure yields the higher modal frequencies up to the Nyquist limit set by the camera frame rate.

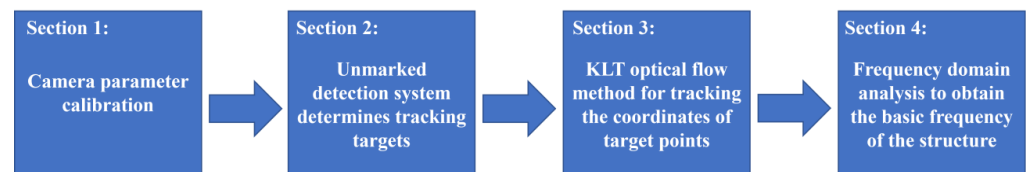


Figure 1. Natural frequency frame diagram obtained using a label-free visual method.

2.1. Camera Parameter Calibration

The structure is filmed with a consumer-level camera. Since wide-angle lenses are mostly used, the video is affected by lens distortion, including mirror distortion and tangential distortion. In order to improve the accuracy and dependability of upcoming computer vision jobs, it is required to calibrate the camera and lens in order to improve the shape and position of the item in the image.

Under ideal conditions, the principle of imaging an object in a camera is the pinhole imaging principle, and the imaging model is shown in Figure 2, which converts three-dimensional coordinates in reality into the two-dimensional coordinates of the imaging plane. Since this mathematical model is an ideal “pinhole imaging” model, the influence of lens distortion is not considered. When calibrating the camera, it is also necessary to obtain the lens distortion coefficient and then correct the original image.

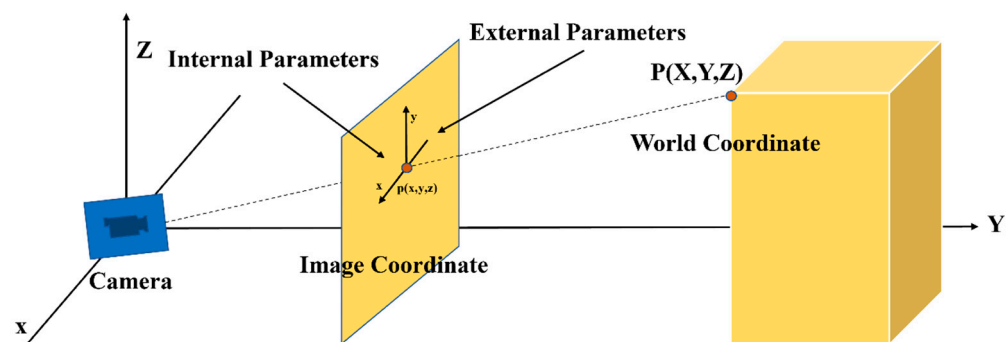


Figure 2. Pinhole imaging principle diagram.

The conversion equation is shown in Formula (1).

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = A_{3 \times 3} \begin{bmatrix} R_{3 \times 3} & T_{3 \times 1} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

$$A_{3 \times 3} = \begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

In the above formula, s is the scaling factor, $A_{3 \times 3}$ is the intrinsic parameter matrix, f_x and f_y are the focal lengths of the camera in the horizontal and vertical directions, c_x and c_y are the offsets of the optical axis in the horizontal and vertical directions, and γ is the tilt factor. R and T are both extrinsic parameter matrices, representing the rotation and translation from the 3D real coordinate system to the 3D camera coordinate system. Both internal and external parameters will affect the imaging results.

The checkerboard approach for camera parameter calibration proposed by Zhang [33] is used in this research. As seen in Figure 3, the camera parameter calibration is quickly finished to obtain the parameter matrices in a stable lighting environment after obtaining a set of images of a black and white chessboard with a known size and a specific number of clearly visible corners at different camera positions. The calibration plate image set's reprojection error is displayed in Figure 4.

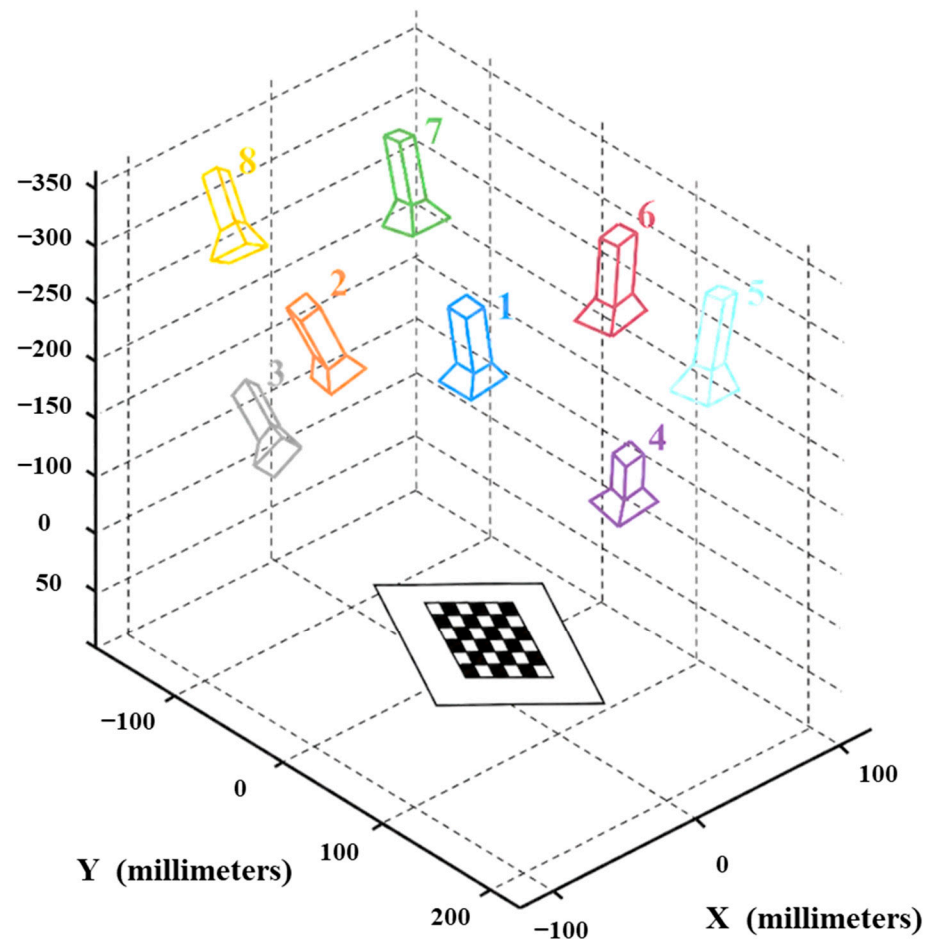


Figure 3. Collecting calibration plate image set. The numbers 1–8 indicate the eight different camera poses used to acquire the calibration plate images.

2.2. Label-Free Detection System

The primary features of the label-free detection system are feature detection and image processing. The two most significant variables influencing measurement results in computer vision measurements are typically illumination and noise. Noise can obscure image details or sharpen nearby pixels that are misinterpreted as significant features, leading to inaccurate feature extraction and matching. The recognition and processing of details might be hampered by uneven brightness and color distortion caused by illumination, which can make certain portions of the image overly bright or too dark. To improve the overall visual impression of the image, increase the details and edges of the image, and improve the accuracy of feature extraction and pattern recognition, the image should be grayed, denoised, and contrast-adjusted before employing feature detection and extraction.

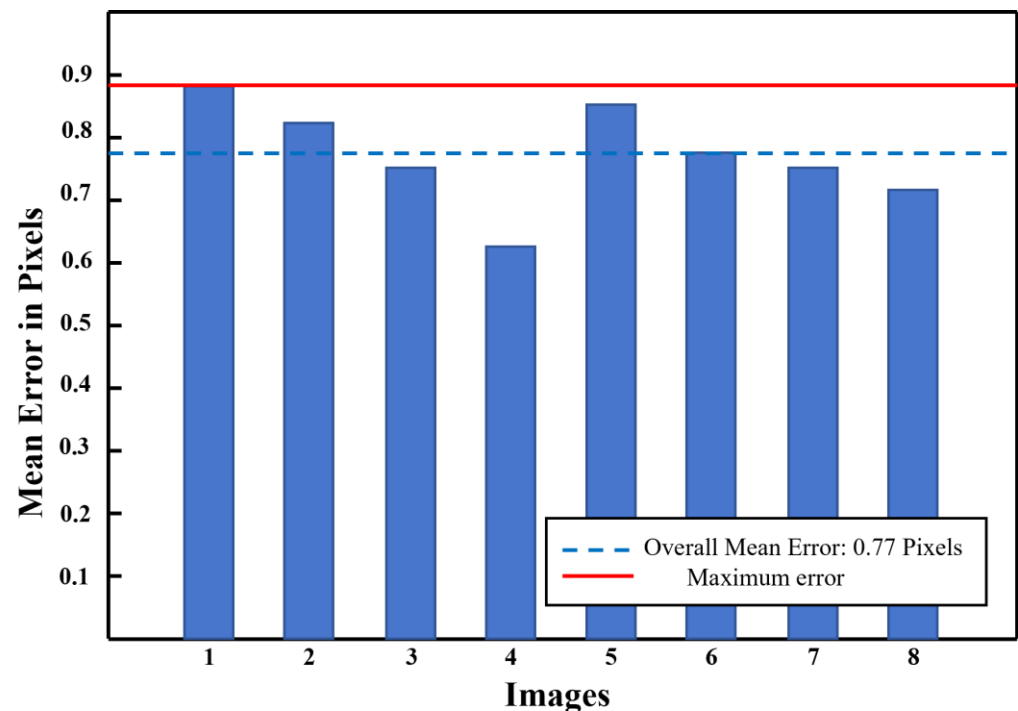


Figure 4. Image set reprojection error.

The label-free detection technique in this study primarily uses the corner detection approach, which selects points with notable properties like pixel values or curvature values as feature points, based on the multi-angular characteristics of the building structure. When using this system, choose the right detectors and algorithms for the various sorts of structures, as well as the feature points that need to be tracked. The results of this paper's evaluation of the most widely utilized corner detectors are displayed in Table 1. In this study the Shi–Tomasi detector was used together with the KLT tracker, since it offers the highest accuracy and stability among the detectors in Table 1 and supplies the minimum-eigenvalue features on which the KLT formulation is based.

Table 1. Evaluation results for commonly used corner detectors.

Detector Name	Detection Speed	Accuracy	Stability	Disadvantage
Harris	Faster	Higher	better	Sensitive to noise, dependent on threshold selection
Shi–Tomasi	Faster	Highest	best	
SIFT	Fastest	Highest	best	High image processing requirements
FAST	Fastest	Higher	better	Sensitive to noise, missed detection
SUSAN	Fastest	Higher	better	Sensitive to noise, High error

2.3. Displacement Time History Curve of the Structural Target Point

The camera records motion as the frame-to-frame pixel displacement of the tracked points. This pixel motion is converted, through the reference-point correction and the scale factor, into a physical displacement time history, and the natural frequencies are then obtained from the dominant peaks of the spectrum of that signal. The displacement time history is therefore the bridge between the raw video and the frequency domain modal information, which is why it is determined before the frequency analysis.

2.3.1. KLT Optical Flow Method

Optical flow is used to measure the instantaneous velocity of a point on a spatial structure moving on an imaging plane. Horn BKP et al. [34] linked the two-dimensional

velocity field with the brightness field, introduced the optical flow constraint equation, and proposed the basic method for optical flow calculation.

$$f_x u + f_y v + f_t = 0 \quad (3)$$

In the above formula, f_x and f_y represent the spatial rate of change of the brightness of the image point (x, y) in the x direction and the y direction; u and v represent the components of the optical flow of the point in the x direction and the y direction; and f_t represents the temporal rate of change of the brightness of the point. The relevant principle is to compare the changes in the pixels of the area around the marked point in the image sequence in the spatiotemporal domain, and then calculate the motion information of the object between adjacent frames. The KLT target tracking algorithm used in this paper is a type of optical flow method. The core idea of the algorithm is to perform a quadratic derivative operation on the grayscale image to obtain the matrix and its two eigenvalues, determine whether the pre-selected window $w(x, y)$ is suitable as a feature point, and then solve the translation of the window pixel point. The solution formula of the z matrix is

$$z = \begin{Bmatrix} I_x I_x & I_x I_y \\ I_y I_x & I_y I_y \end{Bmatrix} \quad (4)$$

In the formula, I_x represents the first-order derivative of the image in the x direction of the feature window w , and I_y represents the first-order derivative in the y direction. If the eigenvalue of the solution matrix is greater than the preset threshold, the point can be used as a feature point.

Then the next step is to match the feature points to solve the translation D of the window pixel points. In this paper, the sum of the squares of the grayscale differences in w in SSD video frames is used as the measurement method. In the grayscale image, the KLT algorithm assumes that the position of the feature window w at the image frame $I(x, y, t)$ at t time and the image frame $I(x, y, t + \tau)$ at $t + \tau$ time is

$$I(x, y, t) = I(x - \Delta x, y - \Delta y, t) \quad (5)$$

$$D = (\Delta x, \Delta y) \quad (6)$$

In w , each pixel point in $I(x, y, t + \tau)$ can be obtained by translating the corresponding window pixel point D in $I(x, y, t)$.

Assume that the feature window at $t + \tau$ time is $B(x) = B(x, y, t)$, where $X = (x, y)$ is the window coordinate. The feature window at t time is

$$A(X - d) = A(x - \Delta x, y - \Delta y, t) \quad (7)$$

For the general case, it is

$$B(X) = A(X - d) + n(X) \quad (8)$$

In the formula, $n(X)$ represents the noise. The noise $n(X)$ needs to be squared first, and the entire window is integrated X to calculate the sum of squares of the grayscale difference in the window image:

$$\varepsilon = \iint_v [A(X - d) - B(X)]^2 \omega(X) dX \quad (9)$$

When the error ε is minimized, the feature points of each feature matching window w can achieve better-matching results.

According to the basic assumption of the KLT optical flow method that “objects are accompanied by tiny movements in a very short time”, it can be considered that d relative to X is negligible. By Taylor expanding $A(X - d)$ and removing the high-order terms in the formula, we receive

$$A(X - d) = A(X) - g \times d \quad (10)$$

Substitute into Equation (9), take the derivative with respect to z , set both sides of the equation equal to 0, and solve for the ε minimum value of

$$\frac{\partial \varepsilon}{\partial d} = \iint_v [A(X) - B(X) - g^T d] \times gw(X) d(X) = 0 \quad (11)$$

Combined with the above, when the two eigenvalues of the z matrix both satisfy $\lambda > \lambda_{\max}$, the feature window has a good tracking effect, and the following can be obtained by substituting it into Formula (11):

$$\begin{bmatrix} G_{xx} & G_{xy} \\ G_{yx} & G_{yy} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} E_x \\ E_y \end{bmatrix} \quad (12)$$

$$\Delta x = \frac{E_x - \frac{G_{xy}}{G_{yy}} E_y}{G_{xx} - \frac{G_{xy}^2}{G_{yy}}}, \Delta y = \frac{E_x \frac{G_{xy}}{G_{yy}} E_y}{G_{xx} - \frac{G_{xy}^2}{G_{yy}}} \quad (13)$$

In this way, the translation D of the window pixel points and the coordinate set of the tracking points can be obtained.

2.3.2. Displacement Time History Curve

After using the KLT optical flow method to obtain the coordinate sets of the target point T_1 and the reference point T_0 , the displacement of the target point is corrected to correct the error caused by the camera's self-vibration during the shooting process:

$$X = X_1 + X_2 \quad (14)$$

In Formula (14), X is the absolute displacement of the target point, X_1 is the relative displacement of the target point with respect to the moving image frame, and X_2 is the apparent displacement induced by camera self-vibration. The reference point T_0 is used to estimate the camera-induced component, and the corrected structural displacement is obtained by subtracting the trajectory of T_0 from that of the target point.

Finally, the displacement is processed with the scale factor R , and the final displacement–time history curve can be obtained. The scale factor R is the ratio of the real coordinate difference in the target point T_1 and the reference point T_0 to the pixel coordinate difference:

$$R = \frac{(T_1 - T_0)}{(T'_1 - T'_0)} \quad (15)$$

In addition, a displacement change threshold is set in the visual system used in this study. In this study, this threshold was set to 5 pixels. If the displacement change in adjacent frames is greater than the threshold, it is considered a tracking failure and needs to be reinitialized to reduce the error of the current frame displacement and enhance the adaptability of the visual system. The specific flow chart is shown in Figure 5.

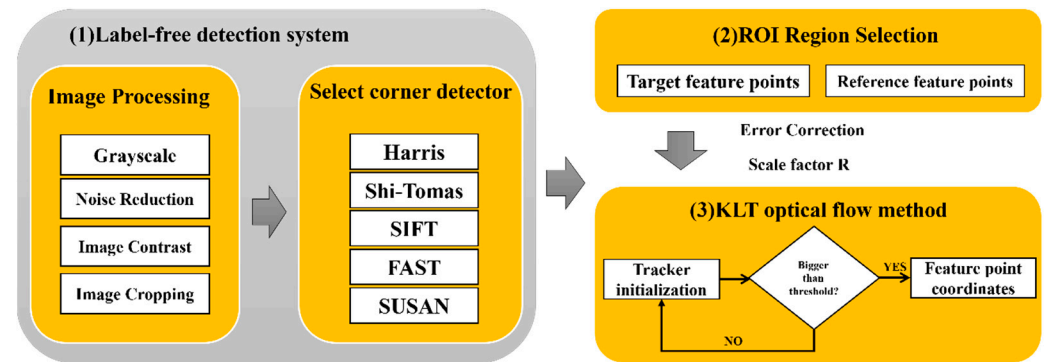


Figure 5. Determining and tracking feature points.

2.3.3. Data Post-Processing

The acquired signal must be post-processed, which mostly entails actions such as trend elimination, fast Fourier transform (FFT), and bandpass filtering, in order to determine the structure's natural frequencies. After that, by doing a frequency domain analysis on the signal, the natural frequencies of the structure can be found.

The spectra reported in this study were obtained from 60 s records using an FFT. The camera sampling frequency was equal to the frame rate, namely 60 Hz for the 60 FPS videos and 30 Hz for the 30 FPS comparison case. The 941B accelerometer signal was sampled at 50 Hz. Therefore, the FFT sample numbers were 3600, 1800, and 3000 for the 60 FPS video, 30 FPS video, and accelerometer records, respectively. The resulting frequency resolution was $\Delta f = f_s/N = 1/T = 0.0167$ Hz. No statistical averaging across repeated trials was performed; therefore, the spectra should be interpreted as single-record estimates.

3. Experimental Verification

3.1. Experimental Setup and Data Acquisition

Verification tests on indoor model transmission towers and outdoor high street lamp wind load simulation tests were set up to confirm the viability and applicability of the proposed method for obtaining the natural frequency of tall structures. The influence of various factors on the accuracy of the results in real-world applications was also discussed. As far as possible, the test conditions are designed to have stable lighting and little noise interference because the test results will be impacted by both external noise and noise produced by the equipment's internal sensors. An iPhone, a cell phone holder, a computer, a 941B accelerometer, an acquisition device, and some software are the primary components of the verification procedure. In the experiments, the camera recorded video at 60 frames per second, and the 941B accelerometer signal was sampled at 50 Hz. The two systems recorded the structural vibration response under the same test conditions. The frequency comparison was based on the dominant peaks obtained from FFT analysis, rather than on time-domain phase matching; therefore, the video and acceleration signals were analyzed using their respective sampling frequencies.

3.2. Indoor Model Tower Experiment

The camera is positioned at a specific distance from the tower to record a video of its free vibration along the ground in the indoor model tower experiment, as depicted in Figure 6.

Four ROI (regions of interest) areas in total are configured to choose feature points, as seen in Figure 7. T0 is the symmetrical point of T3, which serves as a point of reference. T1, T2, and T3 are points on the tower body at various heights. A consumer-level camera with a 60 frames per second frame rate and 1080p resolution was used for the test. Experience has

shown that the vibration frequencies of civil engineering structures are typically modest. In the frequency domain diagram, the fundamental frequencies of importance to engineers can already be covered by the frame rate of 60 frames per second. Figure 8 displays the feature points that the label-free detection algorithm chose for each of the ROI sections. Instead of depending on hand annotations, it integrates the structure's features. No load was applied to the structure during this experiment in order to replicate the actual ambient excitation conditions. Measuring was limited to the tower's signal vibrating freely with the ground. There were also other sizable machinery near the test location that were always in operation, which produced a lot of noise.

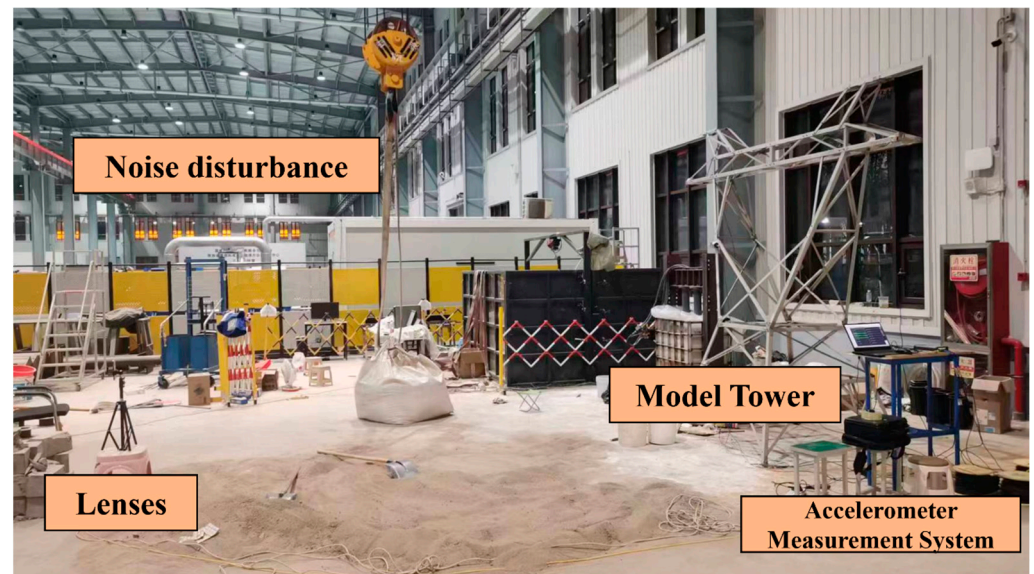


Figure 6. Indoor testing equipment.

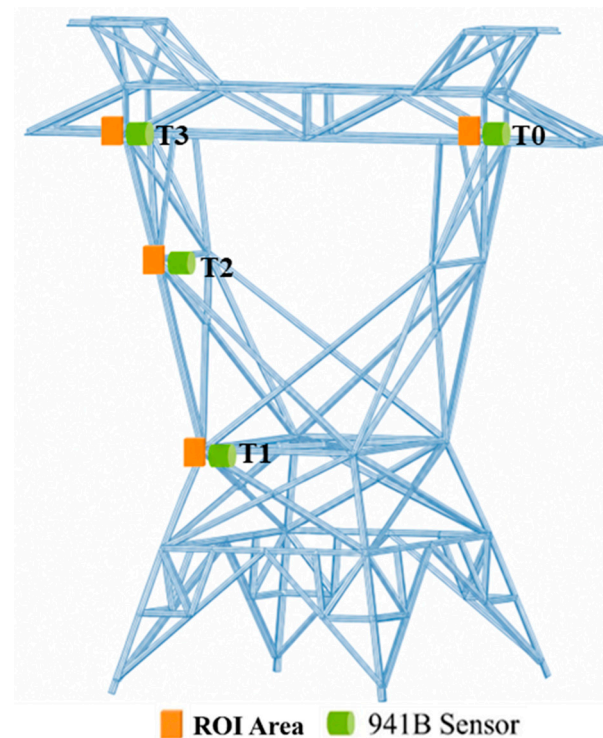


Figure 7. Model tower test diagram (T1, T2, and T3 are measurement points on the tower body, while T0 is the reference point used for camera-motion correction).

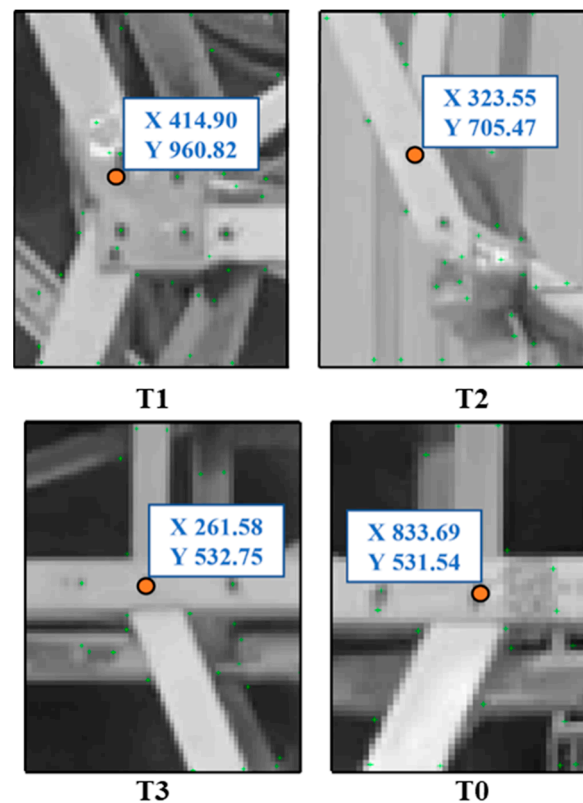


Figure 8. Select tracking point.

Figure 9 displays the test findings. The first three frequency peaks identified from the vibration signals at measurement points T1 and T2 are discussed. As shown in Figure 9, the dominant frequency peaks obtained by the vision system are generally consistent with those measured by the 941B accelerometer, indicating that the proposed label-free vision method can capture the main vibration-frequency components of the model tower. However, the spectral amplitudes and noise levels of the two methods are not identical. These differences may be related to image noise, illumination variation, feature-tracking uncertainty, camera-motion correction error, and environmental vibration from nearby machinery. Therefore, the comparison focuses mainly on the identified peak frequencies rather than on exact spectral amplitudes.

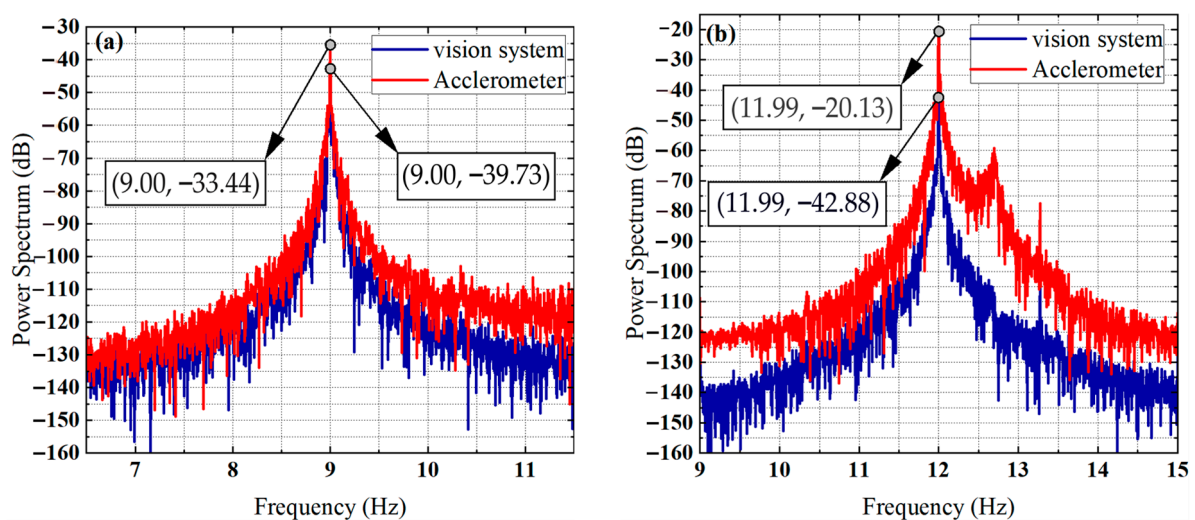


Figure 9. Cont.

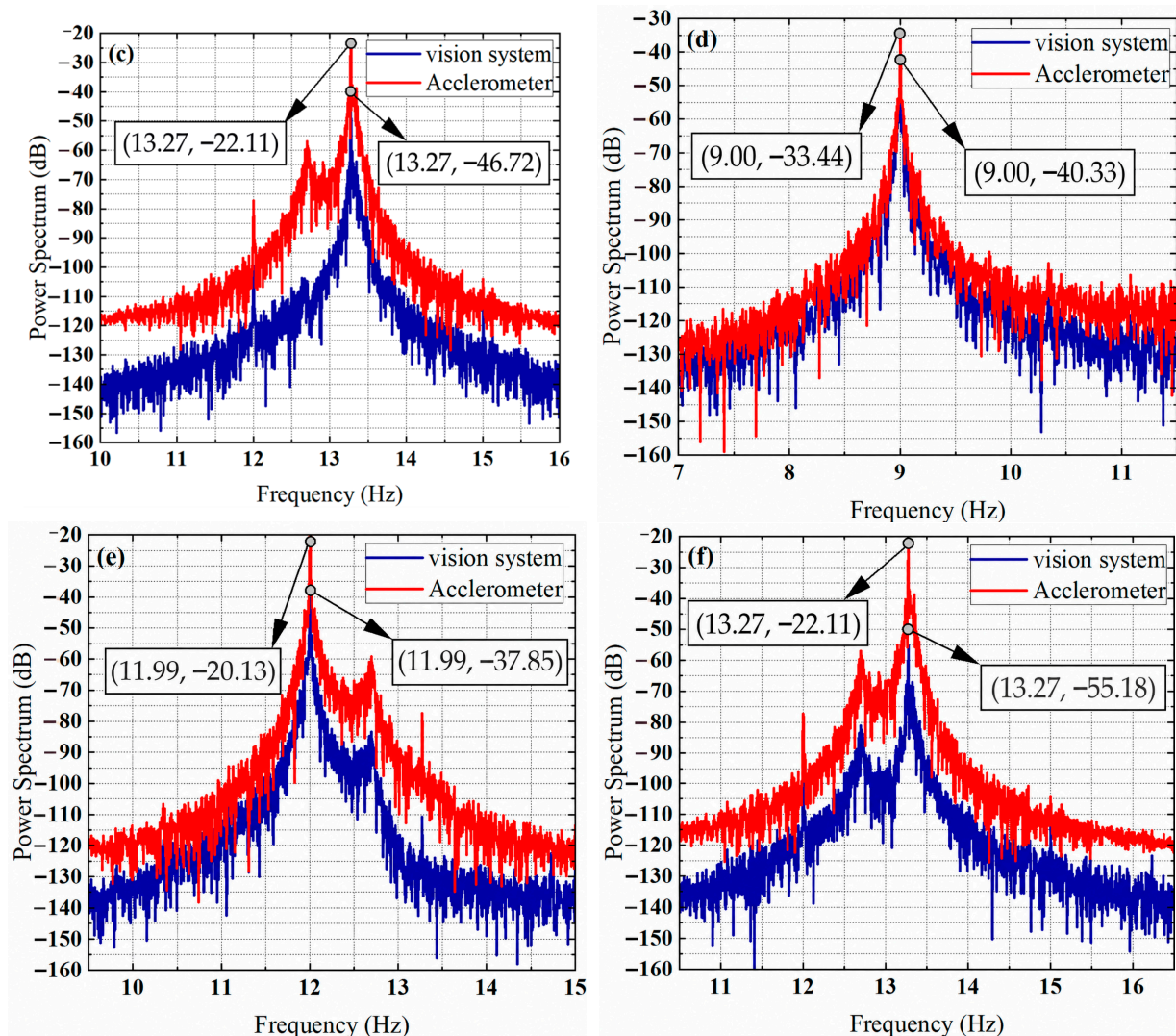


Figure 9. Vision system and accelerometer measurement results in indoor model tower experiment: (a–c) measurement results of feature point T1; (d–f) measurement results of feature point T2.

The feasibility of the visual system in this study is demonstrated through indoor testing. But unlike the accelerometer, which is directly connected to the vibration of the tower, the visual system is impacted by things like lighting, noise, and parameter adjustments. The frequency domain analysis may reveal energy augmentation or loss, which would prevent the signal from fully overlapping.

3.3. Outdoor High Street Lamp Test

3.3.1. Experiment Overview

As seen in Figure 10, the high street lamp has three measurement points (T1, T2, and T0) installed. The base-mounted point T0 serves as a reference point for determining the camera posture of the high street lamp during vibration. The high street lamp experiences free vibration and is mostly exposed to environmental stimulation from natural wind load during the measuring process. Throughout this procedure, the 941B accelerometer and vision system will measure the high street lamp's vibration signal and carry out frequency domain analysis. For the outdoor lamp test, a 60 s record was visually inspected to avoid obvious drift, abrupt disturbances, or tracking loss, and no spectral averaging was applied before peak picking.

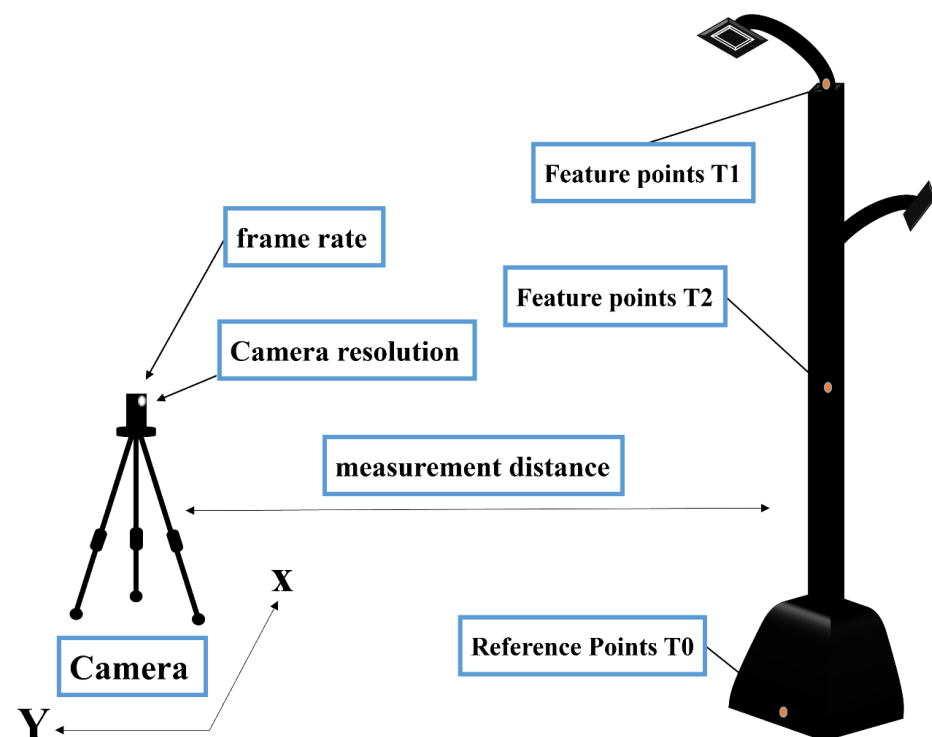


Figure 10. Outdoor testing diagram.

Simultaneously, five test conditions were considered to investigate the impact of camera resolution, frame rate, and measurement distance on the accuracy of visual system results, as indicated in Table 2.

Table 2. Outdoor experimental conditions.

Case	Camera Resolution	Frame Rate (FPS)	Measuring Distance (M)
1	1080P	60	9
2	1080P	60	12
3	1080P	60	15
4	1080P	30	9
5	4K	60	9

3.3.2. Comparison of Visual System and 941B Measurement Results

Taking case 1 as an example, as shown in Figure 11, a comparison diagram of the vibration signals of two characteristic points of the structure in the visual system and the accelerometer 941B is constructed.

After comparing the frequency spectrum of the structure, an error analysis was performed, as shown in Table 3. The results show that the vision-based system is suitable for measuring the natural frequency of tall structures and has the ability to analyze the dynamic characteristics of the structure.

Table 3. Comparison results of vision system and accelerometer experiments.

Measuring Point Methods and Errors	T1			T2		
	First	Second	Third	First	Second	Third
Frequency	1.022	9.285	10.793	0.982	9.282	10.797
Vision system	1.000	9.317	10.775	1.000	9.317	10.775
Accelerometer	2.21%	0.32%	0.17%	1.81%	0.38%	0.20%
Error						

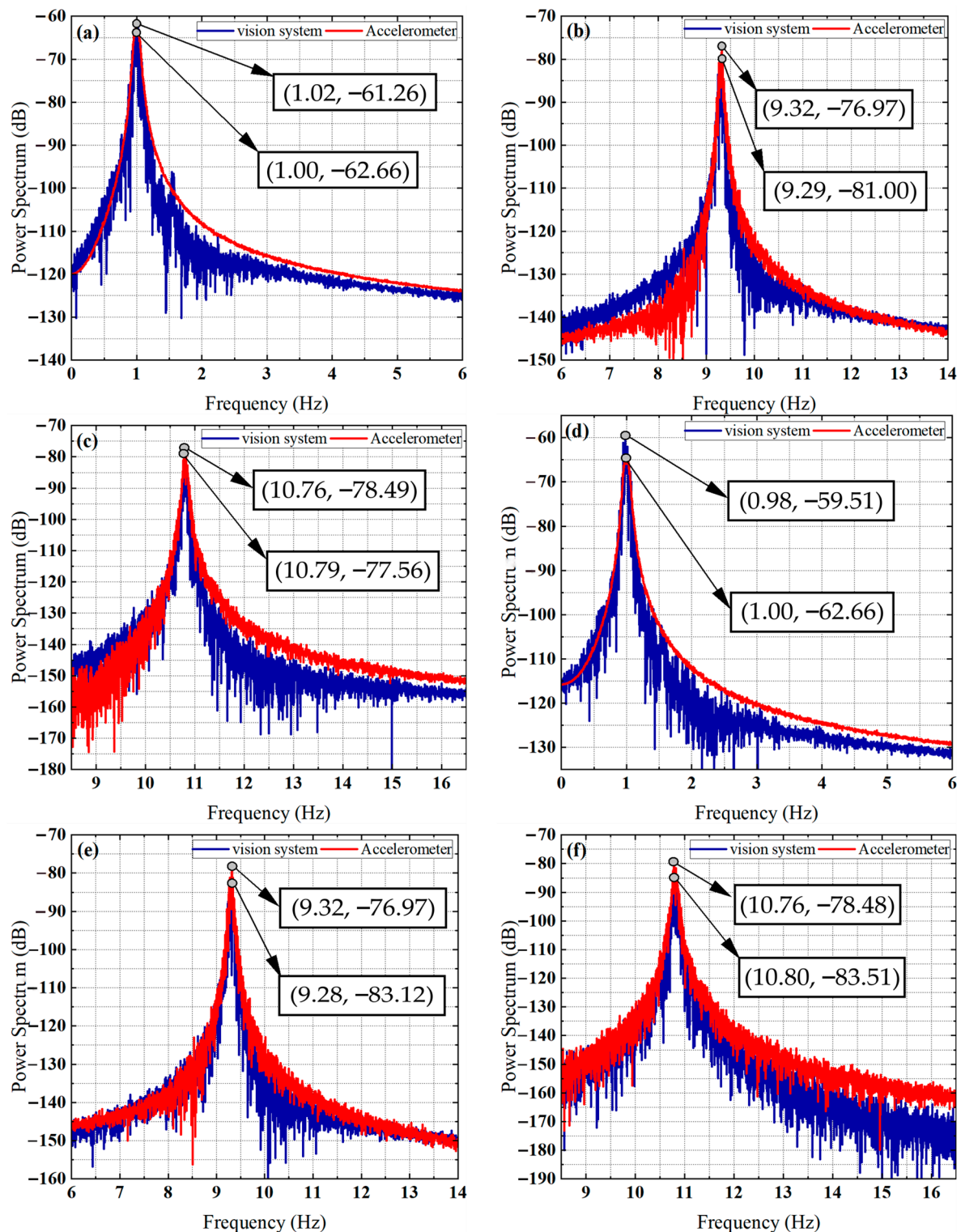


Figure 11. Measurement results of the visual system and accelerometer: (a–c) measurement results of feature point T1; (d–f) measurement results of feature point T2.

3.3.3. The Impact of Measurement Distance, Camera Frame Rate, and Resolution on the Accuracy

This study used three measurement distances, two frame rates, and two resolutions in a series of experiments to investigate whether frame rate, resolution, and measurement distance affect the accuracy of the results.

Using feature point T1 as an example, Figure 12 shows how measurement distance affects the accuracy of the result.

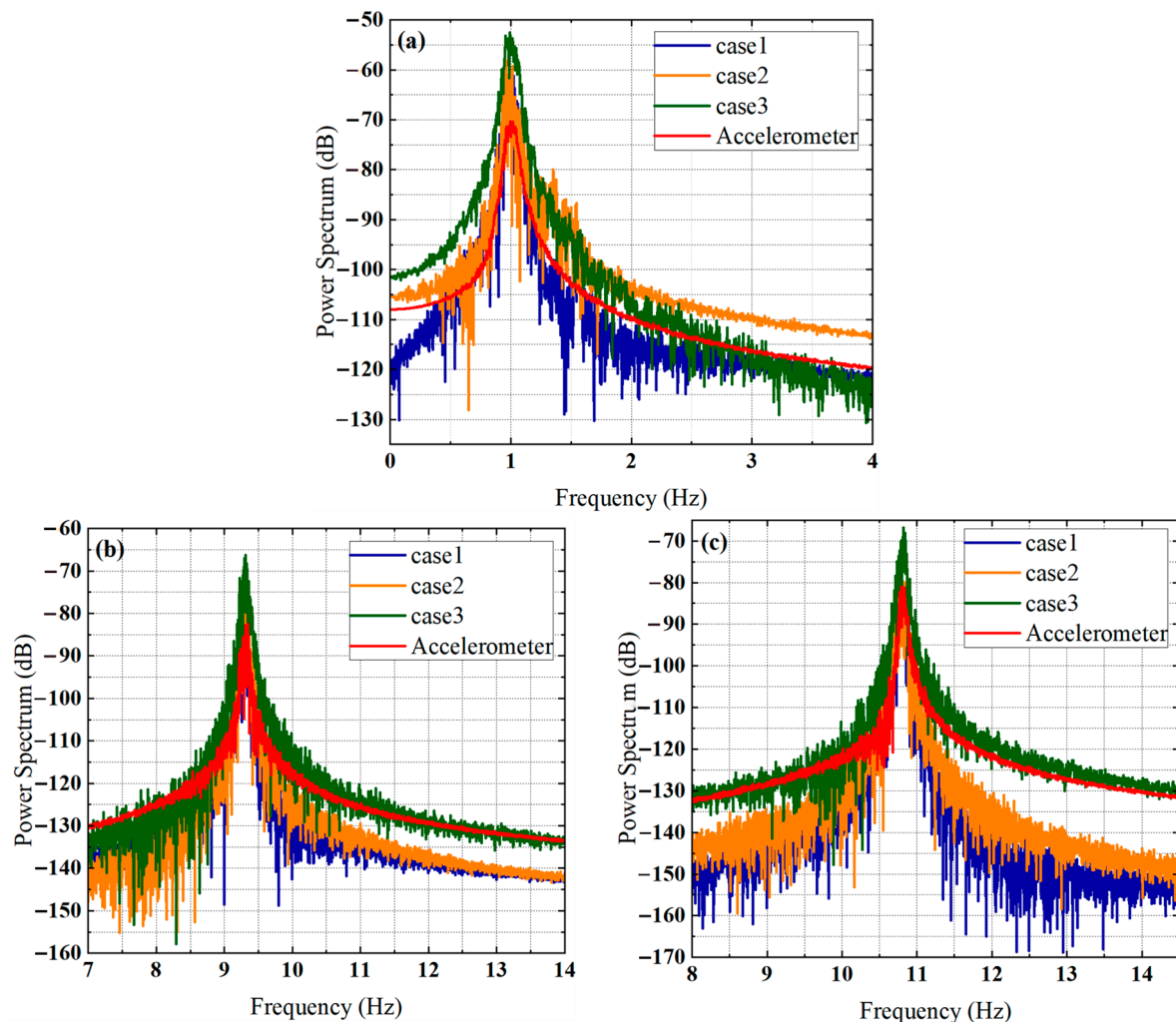


Figure 12. Effect of measurement distance on results. (a) first natural frequency; (b) second natural frequency; and (c) third natural frequency.

The results indicate that the influence of the measuring distance is mode-dependent within the tested range of 9–15 m.

As shown in Table 4, the first-frequency estimate is more sensitive to the measurement distance. The first-frequency error increases from 2.21% at 9 m to 3.91% at 12 m and 4.21% at 15 m. However, the second-frequency error remains nearly unchanged, ranging from 0.31% to 0.42%, and the third-frequency error shows a small but non-monotonic change from 0.17% to 0.13% and then to 0.34%. Therefore, increasing the measuring distance should not be interpreted as causing a uniform monotonic decrease in accuracy for all identified frequencies.

A possible explanation is that, as the measurement distance increases, the same physical displacement corresponds to a smaller pixel displacement in the image, which reduces displacement resolution and tracking signal-to-noise ratio. This influence is more evident for the low-frequency first mode, while the second- and third-frequency peaks remain relatively stable under the present test conditions.

In addition, this study compared the effects of camera frame rate and image resolution on the identified frequencies. The corresponding error analyses are shown in Tables 5 and 6.

Table 4. Experimental results at different measurement distances.

Measuring Distance (m) Natural Frequency	9				12			15		
First	1.022				0.961			0.960		
Second	9.285				9.278			9.278		
Third	10.793				10.761			10.812		
Error	2.21%	0.31%	0.17%	3.91%	0.42%	0.13%	4.21%	0.42%	0.34%	

Table 5. Experimental results at different frame rates.

Frame Rate (FPS) Natural Frequency	60				30		
First	1.022				1.027		
Second	9.285				9.293		
Third	10.793				10.780		
Error	2.21%	0.31%	0.17%		2.72%	0.25%	0.05%

Table 6. Experimental results at different resolutions.

Camera Resolutions Natural Frequency	1080p				4K		
First	1.022				1.008		
Second	9.285				9.282		
Third	10.793				10.797		
Error	2.21%	0.31%	0.17%		0.81%	0.37%	0.20%

As shown in Tables 5 and 6, the differences caused by frame rate and image resolution are relatively small under the present test conditions. Since repeated independent trials were not conducted for all test conditions, the results should be interpreted as feasibility observations rather than statistically confirmed trends; further repeated measurements are needed to quantify repeatability.

4. Conclusions

This study presents a label-free computer vision method for measuring the natural frequencies of tall structures, which is validated on an indoor model transmission tower and an outdoor street lamp. The key contributions and findings can be summarized as follows:

(1) The method was validated on an indoor model transmission tower and an outdoor street lamp using a 941B accelerometer as a reference. The vision-based estimates of the first three frequencies agreed with the accelerometer results, with absolute discrepancies of approximately 0.02–0.04 Hz. The first frequency showed a larger percentage error than the higher frequencies, mainly because its frequency value is smaller; the absolute errors were of the same order.

(2) The parameter study showed that measurement distance mainly affected the first-frequency estimate within the tested range, whereas the second- and third-frequency estimates did not show a clear monotonic trend. The effects of frame rate and image resolution were relatively small under the present conditions. Overall, the results indicate that the proposed method is feasible for non-contact frequency measurement of tower- and pole-type structures under suitable imaging conditions.

(3) The main limitations are the limited number of records, the lack of quantified run-to-run repeatability, sensitivity to lighting and noise, dependence on parameter selection,

and the need for an effectively stationary reference point. Future work will focus on repeated tests, uncertainty quantification, numerical or analytical comparison, and full-scale field validation.

Overall, this approach provides a useful methodological reference for applying computer vision to the modal-frequency measurement of tall structures and for developing related structural-health-monitoring software. Improving measurement accuracy, repeatability, robustness, and automation will be the focus of future research.

Author Contributions: Conceptualization, Q.S. and D.C.; methodology, Q.S. and Y.Z.; validation, D.C., Z.L. and B.J.; formal analysis, Z.L. and Q.G.; investigation, D.C., Z.Q. and Y.Z.; resources, B.J., Z.Q. and D.C.; data curation, Z.L. and Z.Z.; writing—original draft preparation, Z.Z. and B.J.; writing—review and editing, Q.S., Z.Z. and D.C.; visualization, Z.L. and Z.Z.; supervision, Q.S.; project administration, B.J. and Z.Q.; funding acquisition, B.J. and Q.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Inner Mongolia Electric Power Group Co., Ltd. under project number Nei Dian Sheng (2022) 6.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: Authors Duo Chen, Ziqi Quan, Yonghong Zhang, Qiang Gao and Bo Jin were employed by the Ordos Power Supply Company, Inner Mongolia Electric Power (Group) Co., Ltd. Author Zhen Zhang was employed by the Inner Mongolia Power (Group) Co., Ltd., Inner Mongolia Autonomous Region. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Li, C.; Pan, H.Y.; Tian, L.; Bi, W. Lifetime multi-hazard fragility analysis of transmission towers under earthquake and wind considering wind-induced fatigue effect. *Struct. Saf.* **2022**, *99*, 102266.
2. Bao, Y.Q.; Chen, Z.C.; Wei, S.Y.; Xu, Y.; Tang, Z.; Li, H. The State of the Art of Data Science and Engineering in Structural Health Monitoring. *Engineering* **2019**, *5*, 234–242. [[CrossRef](#)]
3. Yang, Y.; Zhang, Y.; Tan, X.K. Review on Vibration-Based Structural Health Monitoring Techniques and Technical Codes. *Symmetry* **2021**, *13*, 1998. [[CrossRef](#)]
4. Xu, Y.; Brownjohn, J.M.W. Review of machine-vision based methodologies for displacement measurement in civil structures. *J. Civ. Struct. Health Monit.* **2018**, *8*, 91–110.
5. Gul, M.; Catbas, F.N.; Hattori, H. Image-Based Monitoring of Open Gears of Movable Bridges for Condition Assessment and Maintenance Decision Making. *J. Comput. Civ. Eng.* **2015**, *29*, 04014034. [[CrossRef](#)]
6. Catbas, F.N.; Kijewski-Correa, T. Structural Identification of Constructed Systems: Collective Effort toward an Integrated Approach That Reduces Barriers to Adoption. *J. Struct. Eng.* **2013**, *139*, 1648–1652. [[CrossRef](#)]
7. Fukuda, Y.; Feng, M.Q.; Shinozuka, M. Cost-effective vision-based system for monitoring dynamic response of civil engineering structures. *Struct. Control Health Monit.* **2010**, *17*, 918–936. [[CrossRef](#)]
8. Feng, D.M.; Feng, M.Q. Experimental validation of cost-effective vision-based structural health monitoring. *Mech. Syst. Signal Process.* **2017**, *88*, 199–211. [[CrossRef](#)]
9. Abdelbarr, M.; Chen, Y.L.; Jahanshahi, M.R.; Masri, S.F.; Shen, W.-M.; A Qidwai, U. 3D dynamic displacement-field measurement for structural health monitoring using inexpensive RGB-D based sensor. *Smart Mater. Struct.* **2017**, *26*, 125016. [[CrossRef](#)]
10. Sony, S.; Laventure, S.; Sadhu, A. A literature review of next-generation smart sensing technology in structural health monitoring. *Struct. Control Health Monit.* **2019**, *26*, e2321. [[CrossRef](#)]
11. Bhowmick, S.; Nagarajaiah, S.; Lai, Z.L. Measurement of full-field displacement time history of a vibrating continuous edge from video. *Mech. Syst. Signal Process.* **2020**, *144*, 106847. [[CrossRef](#)]
12. Yi, T.H.; Li, H.N.; Gu, M. Experimental assessment of high-rate GPS receivers for deformation monitoring of bridge. *Measurement* **2013**, *46*, 420–432. [[CrossRef](#)]

13. Catbas, N.; Dong, C.Z.; Celik, O.; Khuc, T. A Vision for Vision-based Technologies for Bridge Health Monitoring. In Proceedings of the 9th International Conference on Bridge Maintenance, Safety and Management (IBAMAS), Swinburne University Technology, Melbourne, Australia, 9–13 July 2018; pp. 54–62.
14. Feng, D.M.; Feng, M.Q. Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection—A review. *Eng. Struct.* **2018**, *156*, 105–117. [[CrossRef](#)]
15. Harris, C.G.; Stephens, M. A combined corner and edge detector. *Alvey Vis. Conf.* **1988**, *15*, 10–5244.
16. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
17. Alahi, A.; Ortiz, R.; Vandergheynst, P. FREAK: Fast Retina Keypoint. In *IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2012; pp. 510–517.
18. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [[CrossRef](#)]
19. Beauchemin, S.S.; Barron, J.L. The computation of optical flow. *ACM Comput. Surv.* **1995**, *27*, 433–466. [[CrossRef](#)]
20. Yoon, H.; Elanwar, H.; Choi, H.J.; Golparvar-Fard, M.; Spencer, B.F., Jr. Target-free approach for vision-based structural system identification using consumer-grade cameras. *Struct. Control Health Monit.* **2016**, *23*, 1405–1416. [[CrossRef](#)]
21. Tomasi Kanade, T. *Detection and Tracking of Point*; Technical Report CMU-CS-91-132; Carnegie Mellon University: Pittsburgh, PA, USA, 1991.
22. Gundogdu, E.; Alatan, A.A. Good Features to Correlate for Visual Tracking. *IEEE Trans. Image Process.* **2018**, *27*, 2526–2540. [[CrossRef](#)]
23. Dong, C.Z.; Celik, O.; Catbas, F.N.; O'Brien, E.J.; Taylor, S. Structural displacement monitoring using deep learning-based full field optical flow methods. *Struct. Infrastruct. Eng.* **2020**, *16*, 51–71.
24. Ilg, E.; Mayer, N.; Saikia, T.; Keuper, M.; Dosovitskiy, A.; Brox, T. FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks. In Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2462–2470.
25. Ji, Y.F.; Chang, C.C. Nontarget image-based technique for small cable vibration measurement. *J. Bridge Eng.* **2008**, *13*, 34–42. [[CrossRef](#)]
26. Morlier, J.; Michon, G. Virtual Vibration Measurement Using KLT Motion Tracking Algorithm. *J. Dyn. Syst. Meas. Control-Trans. ASME* **2010**, *132*, 011003.
27. Kuddus, M.A.; Li, J.; Hao, H.; Li, C.; Bi, K. Target-free vision-based technique for vibration measurements of structures subjected to out-of-plane movements. *Eng. Struct.* **2019**, *190*, 210–222. [[CrossRef](#)]
28. Qu, C.X.; Jiang, J.Z.; Yi, T.H.; Li, H.-N. Computer vision-based 3D coordinate acquisition of surface feature points of building structures. *Eng. Struct.* **2024**, *300*, 117212. [[CrossRef](#)]
29. Bai, X.; Yang, M.J.; Ajmera, B. An Advanced Edge-Detection Method for Noncontact Structural Displacement Monitoring. *Sensors* **2020**, *20*, 4941. [[CrossRef](#)] [[PubMed](#)]
30. Shao, Y.D.; Li, L.; Li, J.; An, S.; Hao, H. Computer vision based target-free 3D vibration displacement measurement of structures. *Eng. Struct.* **2021**, *246*, 113040. [[CrossRef](#)]
31. Ono, Y.; Trulls, E.; Fua, P.; Yi, K.M. LF-Net: Learning Local Features from Images. In Proceedings of the 32nd Conference on Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 2–8 December 2018; p. 31.
32. Dusmanu, M.; Rocco, I.; Pajdla, T.; Pollefeys, M.; Sivic, J.; Torii, A.; Sattler, T. D2-Net: A Trainable CNN for Joint Description and Detection of Local Features. In Proceedings of the 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–21 June 2019; pp. 8084–8093.
33. Zhang, Z.Y. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [[CrossRef](#)]
34. Horn, B.K.P.; Schunck, B.G. Determining optical-flow. *Artif. Intell.* **1981**, *17*, 185–203. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.