Received: 23 October 2024 | Revised: 8 January 2025 | Accepted: 11 February 2025 | Published online: 12 March 2025

RESEARCH ARTICLE

Archives of Advanced Engineering Science
2026, Vol. 4(1) 78–88
DOI: 10.47852/bonviewAAES52024620

BON VIEW PUBLISHING

Three-Dimensional Vibration Visual Measurement Method Based on Binocular Camera

Liangjun Luo[1], Shuo Liu[1,*], Fei Li[1], Wenhao Zou[1], Xi Wang[1], Tao Wang[1], Guoqiang Fu[1] and Caijiang Lu[1]

[1]*School of Mechanical Engineering, Southwest Jiaotong University, China*

Abstract: This paper proposes a three-dimensional micro-vibration measurement method based on amplification and tracking to achieve vibration measurement of the target. In environments with excessive lighting, which can hinder effective target tracking, the method first applies a magnification algorithm to enlarge the target image. The enlarged image is then processed using a stereo matching algorithm to align the left and right camera views. Following this, a tracking algorithm calculates the pixel displacement of the target. Finally, the target's spatial three-dimensional vibration information is derived through dimensional conversion. To address challenges posed by excessive illumination, image preprocessing is applied to extract the target before magnification. Experimental results from three sets of tests demonstrate the method's high accuracy. When measuring spatial three-dimensional vibration with an amplitude of 0.1 mm, the displacement's root mean square error is 0.0298 mm, and the normalized root mean square error is 11.57%. This new three-dimensional visual measurement method effectively and accurately measures micro-vibration amplitudes.

Keywords: visual measurement, micro-vibration, amplification, tracking, spatial vibration displacement

1. Introduction

Mechanical equipment inevitably experiences some micro-vibrations during operation. These vibrations may have negative effects on the equipment, shortening its lifespan or even posing safety risks. Therefore, eliminating these micro-vibrations is crucial [1, 2].

Currently, methods for eliminating micro-vibrations include contact and non-contact methods [3, 4]. Contact vibration measurement methods require the use of sensors such as accelerometers displacement sensors, etc. However, when using contact vibration measurement methods, sensors need to be placed on the target being detected, which can cause mass loading effects for lightweight structures, resulting in measurement errors. Additionally, when measuring vibration data of large structures, installing a large number of sensors on the surface of the structure requires a significant amount of manpower and resources for sensor placement [5–7].

Non-contact vibration measurement methods represent another approach to vibration measurement. Common non-contact methods include laser vibrometry and vision-based vibration measurement. Laser vibrometry involves using a laser vibrometer for measurement, which does not impose any loading effects on the measurement target. However, laser vibrometers can be relatively expensive, and measuring large targets with them may require significant measurement time. The vision-based vibration measurement method utilizes cameras or visual sensors to capture the vibration images of the target surface and then analyzes these images using image processing techniques to achieve vibration measurement. This method does not require direct contact with the target, thus avoiding any loading effects, and it provides high measurement accuracy.

Currently, vision-based vibration measurement methods for vibration displacement include those based on monocular cameras and those based on binocular cameras [8–10]. The monocular camera-based visual measurement method utilizes a single camera to extract feature information from images and then calculates and analyzes parameters such as the size, shape, and position of objects [11–13]. Fukuda et al. [14] proposed an image processing algorithm based on sub-pixel OCM direction coding and conducted error analysis in vibration table tests, with experimental results showing small errors and significant practical value. Building upon previous work, Caetano et al. [15] further optimized the optical flow method, applying the refined technique to vibration observation in large-scale engineering structures. Sieffer et al. [16] utilized digital image correlation technology and high-speed cameras in vibration table tests of timber frame stone infill walls, comparing the results with traditional measurement methods to find a relative error of 5%. Ansari et al. [17] compared two digital images of the structure before and after deformation using the digital image correlation method to assess the distribution of structural dislocations and surface strains. Rajaram et al. [18] demonstrated that digital image correlation methods provide precise full-field deformation information in vibration table experiments, comparing it with displacement sensor methods and finding an absolute measurement error of 0.26 mm. Jiang et al. [19] achieved vibration object recognition

*Corresponding author: Shuo Liu, School of Mechanical Engineering, Southwest Jiaotong University, China. Email: 2022ls@my.swjtu.edu.cn

and tracking in high frame rate videos by locating vibration sources and extracting high-frequency vibration information using digital filters. Zhu et al. [20] used VGG to construct a deep feature fusion network for displacement extraction from image sequences. Huang et al. [21] enhanced visual displacement measurement in harsh environments using GAN and YOLOv4. Building upon traditional image processing algorithms, Lu et al. [22] combined deep learning algorithms to improve measurement accuracy, validating the high precision of this approach through experiments. Yin et al. [23] proposed the spatial support convolutional neural network for ruler detection in switch track images.

The visual measurement method using a monocular camera is limited to measuring planar vibration displacements and cannot meet the requirements for capturing the spatial displacement of structures. In contrast, the visual measurement method based on binocular cameras can use the image sequences captured by the left and right cameras to perform stereo matching and obtain disparity, thereby recovering the three-dimensional information of the object being measured [24–26]. Yuan et al. [27] combined stereo vision with deep learning, using an intelligent detection robot with an Internet of Things data communication system to identify damage in reinforced concrete columns subjected to cyclic loading. Shao et al. [28] proposed an advanced binocular vision system for full-field three-dimensional vibration displacement measurement of structures, verifying its performance by measuring the vibration displacement of a steel cantilever beam in the laboratory. Yue et al. [29] developed an automatic deformation monitoring method using binocular vision, employing an enhanced semi-global matching algorithm for stereo matching of binocular images to accurately measure the three-dimensional displacement of the monitored target. Wang et al. [30] proposed a bidirectional correction and binocular vision positioning algorithm for the automated forging of hammer forging, with results showing a small average positioning error of 7.31 mm in the left camera coordinate system. Wu et al. [31] proposed a bridge vibration measurement method based on 3D reconstruction, using drones to capture vibration videos of bridge models and tracking the displacement of target points on the bridge model using the digital image correlation method. Yao et al. [32] processed and managed bridge vibration signals using 3D modeling and visualization algorithms, achieving an accuracy of 98.75%.

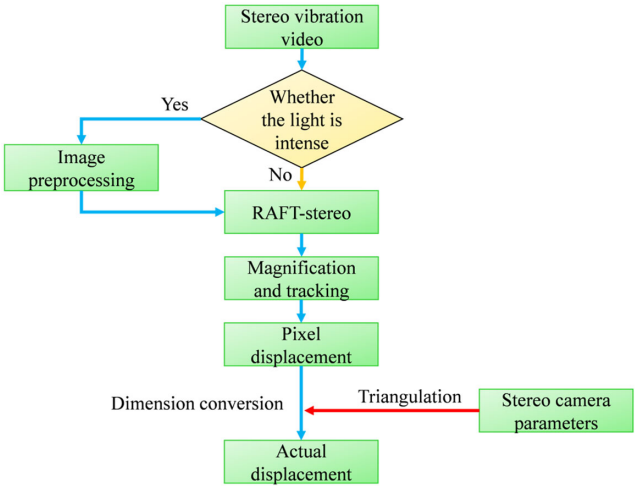
Building on existing research, a three-dimensional micro-vibration measurement method based on amplified tracking is proposed to achieve structural spatial displacement measurement. First, the overall algorithm workflow is introduced, the overall flowchart is shown in Figure 1 followed by the verification of the measurement accuracy of this method through spatial three-dimensional vibration experiments on a cantilever beam. Finally, some conclusions are drawn.

2. Three-Dimensional Vibration Measurement Algorithm

The steps of the algorithm are as follows:

Step 1: Determining if the external light intensity is too strong (judged based on whether the target tracking algorithm can track the target). If it is, image preprocessing is required to avoid interference from other objects on the vibration information. If not, proceed directly to amplification.

Figure 1
Overall flowchart



Step 2: The image preprocessing stage is undertaken for cases of excessive light intensity. This involves a number of different techniques, including region of interest processing, image smoothing filtering, color conversion and feature extraction, and morphological operations. Region of interest processing involves selecting the area of the target to be measured for ease of subsequent feature extraction. Image smoothing filtering reduces noise, removes details, and smoothens the image. Color conversion and feature extraction involve extracting specific color features after filtering based on the color of the marker, while morphological operations are used to perform region connectivity and remove outlier points on the extracted target.

Step 3: Selecting the optimal amplification factor and perform video motion amplification on the video.

Step 4: Opening the amplified video frame by frame and performing stereo matching of the left and right camera frame images using the RAFT-Stereo stereo matching algorithm.

Step 5: After that, conduct DaSiamRPN target tracking to obtain the pixel displacement of the target.

Step 6: Utilizing the obtained pixel coordinates, the calibration parameters of the binocular camera, and the principle of triangulation to obtain the spatial three-dimensional coordinates of the target.

2.1. Magnification and tracking

Oh et al. [33] proposed a learning-based motion amplification method in 2018, which possesses characteristics such as learning ability, adaptability, efficiency, generalization capability, and end-to-end optimization. These attributes enable the method to achieve better results in handling motion in videos. DaSiamRPN [34] is an efficient, accurate, and robust target tracking method that is widely used in real-time video analysis, intelligent surveillance, autonomous driving, and various other fields. The integrated framework of the amplification and tracking algorithms is depicted in Table 1 [35] and Figure 2.

Table 1
Algorithm framework

Algorithm: Magnification and tracking
Input: Two pictures displaying the frames prior to and following the video. $\{X_{a,n}\}_{n=1}^{N-1}$,$\{X_{b,n+1}\}_{n=1}^{N-1}$
For$(n=1;n++;n<N)$
Magnification: Utilizing the encoder to obtain the shape characteristics and texture attributes of the previous frame and the $\{X_{a,n}\}_{n=1}^{N-1}$ shape $\{M_{b,n+1}\}_{n=1}^{N-1}$ features $\{M_{a,n}\}_{n=1}^{N}$, then using the manipulator to obtain $\{M_{c,n+1}\}_{n=1}^{N-1}$ multiplied by a factor of α. Finally, using the decoder to fuse $\{V_{b,n+1}\}_{n=1}^{N-1}$ and $\{M_{c,n+1}\}_{n=1}^{N-1}$ to obtain output.
End
Magnification Output: the output video $\{X_{a,n}\}_{n=1}^{N-1}$ scaled by a factor of α.
For$(n=1;n++;n<N)$
Tracking: Employing the Siamese network to obtain the reference frame $\{X_{a,n}\}_{n=1}^{N-1}$ and features and detection frame $\{X_{b,n+1}\}_{n=1}^{N-1}$ features, then using the region proposal network to obtain n target. Finally, incremental learning is used to select final object with the greatest similarity from n target.
End
Tracking Output: Tracking box of the target object.

2.2. RAFT-Stereo

RAFT-Stereo algorithm [36] is a stereo matching algorithm that integrates end-to-end training and deep learning techniques of RAFlow (RAFT). By utilizing deep learning networks to learn image features and optical flow constraints, this algorithm achieves highly accurate optical flow estimation and stereo vision tasks. The overall framework of the algorithm is depicted in Table 2 and Figure 3.

2.3. Dimension conversion

2.3.1. Coordinate transformation relationship

Coordinate transformation typically involves converting coordinates from one coordinate system to another. As shown in Figure 4, this process includes transformations between the global coordinate system $O_w - X_w Y_w Z_w$, camera coordinate system $O_c - X_c Y_c Z_c$, image physical coordinate system $o - xy$, and pixel coordinate system $o_p - uv$.

The global coordinate system is a coordinate system used to describe the entire scene. It is a fixed coordinate system used to represent the position, orientation, size, and other information of all objects, denoted by (x_w, y_w, z_w).

The camera coordinate system is an internal coordinate system of the camera used to describe the spatial relationships between objects, images, and the camera during the imaging process, denoted by (x_z, y_z, z_z).

Figure 2
Overall framework

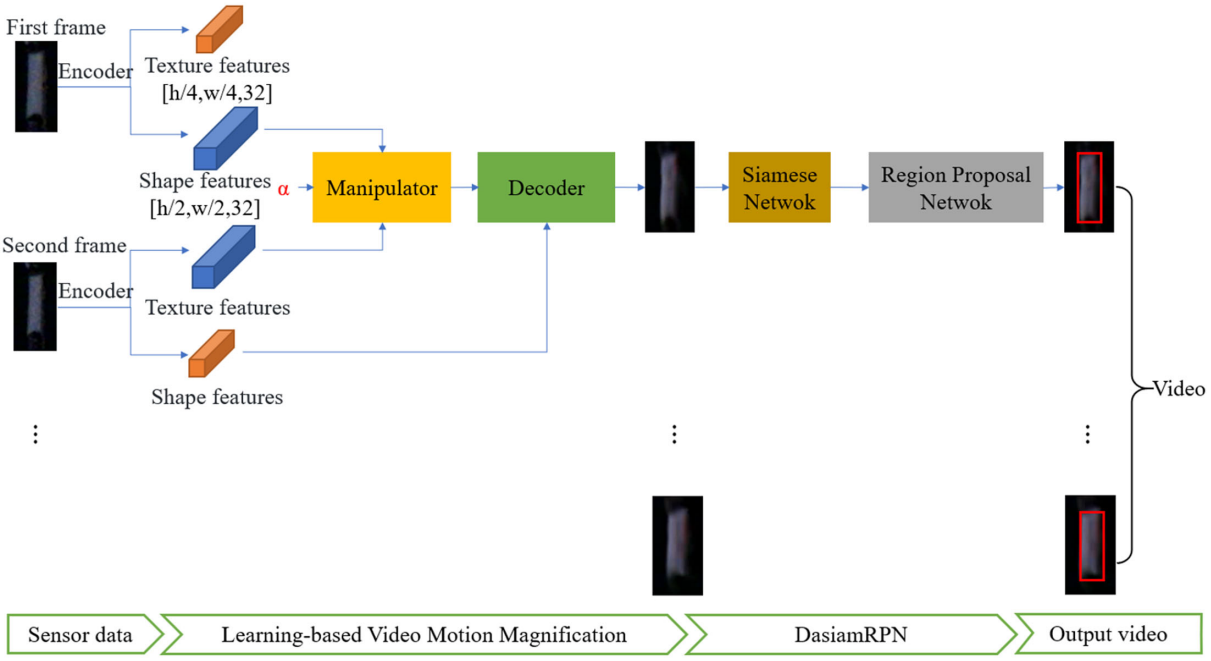


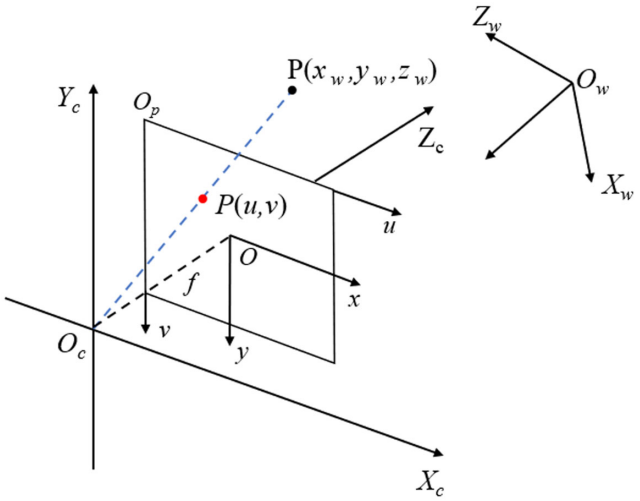
Table 2
Algorithm framework

Algorithm: RAFT-Stereo
Input: Left eye image I_L, right eye image I_R **Feature Encoder**: Composed of residual modules and downsampling layers, I_L and I_R are passed into this network structure to obtain feature maps I_L and I_R for I_F and g_R, respectively. **Context Encoder**: I_L is input to the context encoder to generate a context feature map for initializing the hidden state of the GRU module. **Correlation Pyramid**: Computing the dot product of F_L and g_R to construct a 3D correlation volume denoted as C_{ijk}. **Multi-Level Update Operator**: During the iteration process, a series of disparity maps are obtained from the initial disparity. At each iteration, a new disparity estimate d_{gt} is produced. The correlation features are obtained by looking up C_{ijk} based on the disparity estimate and then input into 2 convolutional layers. Similarly, d_{gt} is also input into 2 convolutional layers. These, along with context features, are fed into the GRU. The GRU updates the hidden state and then estimates a new disparity based on the hidden state. **Output**: Disparity map of the input left and right images.

The image coordinate system is a coordinate system used to describe the pixel positions in a digital image. In common two-dimensional images, Cartesian coordinates are typically used to represent the physical coordinates of the image, denoted by (x, y).

The pixel coordinate system refers to the coordinate system used to represent the pixel positions in an image. In two-dimensional digital images, the pixel coordinate system usually employs Cartesian coordinates, denoted by (u, v).

Figure 4
The relationships between coordinate systems



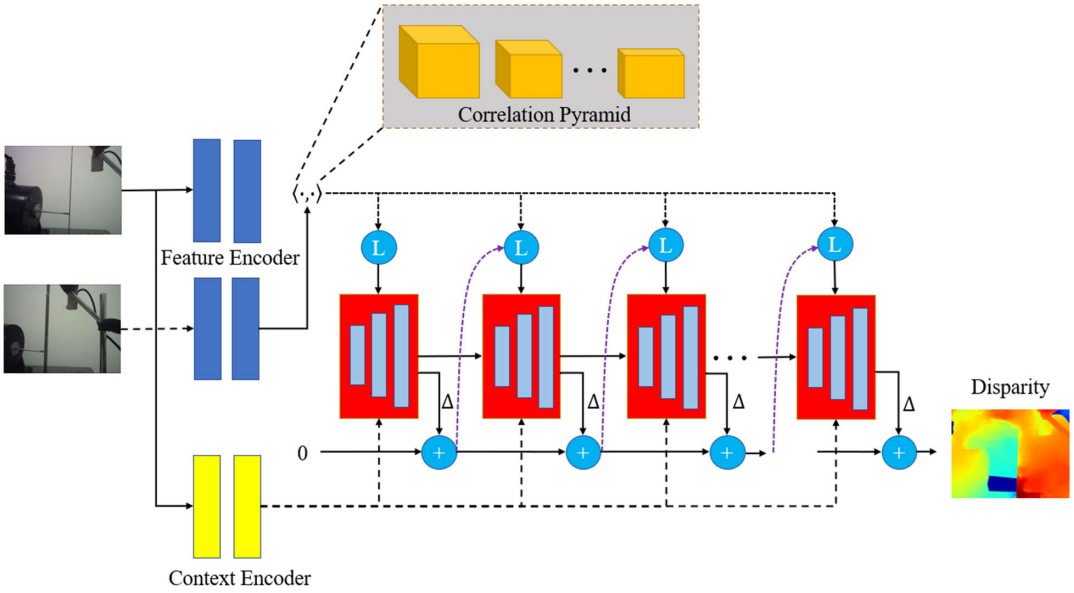
The transformation between a point in space, denoted by $P(x_w, y_w, z_w)$, and the pixel point $P(u, v)$ formed by the camera can be accomplished through the camera coordinate system, global coordinate system, and pixel coordinate system.

The formula for transforming the global coordinate system to the camera coordinate system is as follows:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} = [R \quad T] \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} \quad (1)$$

where R defines the rotation matrix that illustrates the rotational relationship between the global coordinate system and the camera coordinate system, and T represents the translation matrix describing the

Figure 3
Overall framework

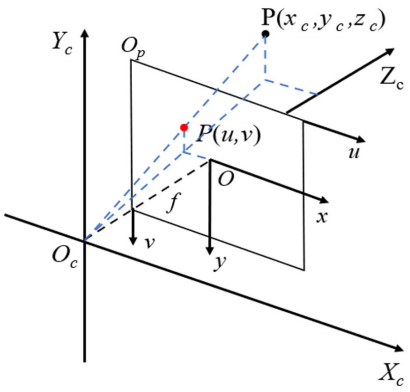


translation relationship from the global coordinate system to the camera coordinate system. Further optimization of the above formula yields:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{2}$$

Following the transformation of a spatial point from the global coordinate system to the camera coordinate system, a further conversion is required to the image coordinate system. The projection relationship between the camera coordinate system and the image coordinate system is depicted in Figure 5.

Figure 5
The projection relationship



Combining Figure 5 with the principle of camera imaging yields the following formula:

$$\frac{x}{x_c} = \frac{y}{y_c} = \frac{f}{z_c} \tag{3}$$

where f represents the distance from the lens to the focal point, which is referred to as the focal length.

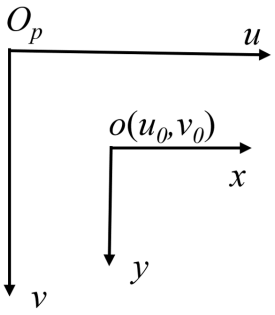
Further adjustments to the formula yield the following equation:

$$z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \tag{4}$$

Following the transformation of the spatial point from the camera coordinate system to the image coordinate system, a further transformation is required from the image coordinate system to the pixel coordinate system, as shown in Figure 6. Upon completion of this additional transformation, the entire process will result in the conversion of a spatial point into an image pixel.

Figure 6 shows if a point O_p in the image coordinate system maps to a point in the pixel coordinate system as (u_0, v_0), based on the transformation relationship, the following equation can be obtained:

Figure 6
Projection relationship from image coordinate system to pixel coordinate system



$$\begin{cases} u = u_0 + \frac{x}{dx} \\ v = v_0 + \frac{y}{dy} \end{cases} \tag{5}$$

Further manipulation of the equation yields the following expression:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{6}$$

where dx represents the pixel size in the x-direction, dy represents the pixel size in the y-direction.

Combining the coordinate transformation formulas from the aforementioned parts, we can obtain the following equation:

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r11 & r12 & r13 & t1 \\ r21 & r22 & r23 & t2 \\ r31 & r32 & r33 & t3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{7}$$

Simplifying the above equation, we get:

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M_1 M_2 \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$
$$= P \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{8}$$

where $f_x = f/dx$, f_x is the normalized focal lengths in the x-axis, $f_y = f/dy$, f_y is the normalized focal lengths in the y-axis. M_1 represents the intrinsic matrix, M_2 represents the extrinsic matrix, z_c is the scale factor, P is the projection matrix. By solving for the camera's intrinsic and extrinsic matrices, it is possible to calculate the transformation from the spatial coordinates of the target point to the image coordinates.

2.3.2. 3D global coordinate transformation

Assuming that a point in space is projected to pixel coordinates as $(u_l, v_l, 1)$ in the left camera and $(u_r, v_r, 1)$ in the right camera, with corresponding matrices P_l and P_r, based on the above information, we can determine the point X in space. The transformation principle from the global coordinate system to the pixel coordinate system yields the following equation:

$$s\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = P_l X = \begin{bmatrix} P_l^1 X \\ P_l^2 X \\ P_l^3 X \end{bmatrix} \tag{9}$$

where P_l^1 is the first row of the matrix formed in the left camera, P_l^2 is the second row of the matrix formed in the left camera, and P_l^3 is the third row of the matrix formed in the left camera.

Multiplying by s yields:

$$\begin{cases} su_l = P_l^1 X \\ sv_l = P_l^2 X \\ s = P_l^3 X \end{cases} \tag{10}$$

Further processing yields:

$$\begin{cases} u_l P_l^3 X - P_l^1 X = 0 \\ v_l P_l^3 X - P_l^2 X = 0 \end{cases} \tag{11}$$

The above equation is derived from the left camera. Similarly, the corresponding expression can be obtained for the right camera. Combining these two expressions yields the following formula:

$$\begin{cases} u_l P_l^3 X - P_l^1 X = 0 \\ v_l P_l^3 X - P_l^2 X = 0 \end{cases} \begin{cases} u_l P_l^3 X - P_l^1 X = (u_l P_l^3 - P_l^1)X \\ v_l P_l^3 X - P_l^2 X = (v_l P_l^3 - P_l^2)X \\ u_r P_r^3 X - P_r^1 X = (u_r P_r^3 - P_r^1)X \\ v_r P_r^3 X - P_r^2 X = (v_r P_r^3 - P_r^2)X \end{cases} \tag{12}$$

By substituting the calibration parameters of the stereo camera, we can obtain the three-dimensional spatial coordinates of this point in space.

3. Experimental Results

3.1. Experimental setup

The left camera of the stereo camera is placed directly in front of the cantilever beam, parallel to the cantilever beam, with the optical center of the left camera serving as the origin. The cantilever beam is excited to oscillate by means of a shaker, and the cameras register the motion of markers on the cantilever beam. After processing using the method described in this paper, the measurement results are compared and validated against those obtained from a laser displacement sensor.

Table 3
Stereo camera

Parameter	Data
Camera model	D-405
Size of the photosensitive component	3.75 um × 3.75 um
video format	MJPEG
Resolution and frame rate	Binocular 1280 × 480 90 fps Binocular 2560 × 720 60 fps Binocular 2560 × 960 30 fps
Exposure time	Automatic exposure (37 ms–1000 ms)
Image sensor	AR0135 global shutter
Lens focal length	3.4 mm

Table 4
Laser displacement sensor

Parameter	Data
Sensor model	BL-030NM
Measurement range	−5 mm~ + 5 mm
Measuring center distance	30 mm
Repetition precision	10 um
Rectilinearity	±0.1%FS
Supply voltage	12 V~24 V

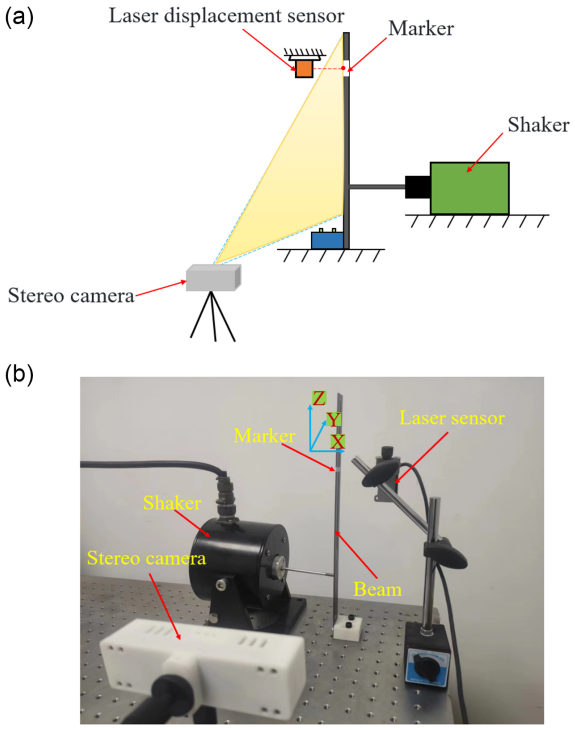
The experimental setup mainly consists of a stereo camera, a laser displacement sensor, and a shaker. The parameters of the stereo camera and the laser displacement sensor are shown in Tables 3 and 4.

3.2. Spatial 3D vibration experiment and analysis

Using the experimental platform in Figure 7, the excitation frequency of the vibrator should be adjusted to 5 Hz. Set the resolution of the stereo camera to 2560 × 720 for both eyes at 60 fps. Setting the sampling frequency of the laser displacement sensor to 60 fps and input the actual values from the laser displacement sensor. Adjust the vibration amplitude to 0.1 mm, 0.2 mm, and 0.4 mm and measure the spatial motion of the cantilever beam using the visual measurement method described in this paper. The measurement results are shown in Figure 8.

Due to the planar motion of the cantilever beam, displacement primarily occurs in the X-axis direction, with very small displacement in the Y-axis and no displacement in the Z-axis,

Figure 7
Blade modal results. (a) Experimental schematic. (b) Experimental diagram



thus ensuring accurate measurement of the spatial three-dimensional image of the cantilever beam.

To quantify the displacement measurement results, the root mean square error and normalized error are used for representation. Their main forms are

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \tag{13}$$

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}}{ymin_{max}} \tag{14}$$

where n indicates the quantity of points, x_i represents the displacement measurement acquired through the visual method at a particular instant, y_i is the displacement value of the laser displacement sensor at a particular instant, y_{max} is the peak displacement value recorded by the laser displacement sensor, y_{min} is the lowest displacement value measured by the laser displacement sensor.

Analysis of Table 5 and Figures 9 to 11 leads to the following conclusions:

Figure 8
Measurement results of 0.1 mm displacement. (a) 3D diagram of 0.1 mm amplitude in space. (b) 3D diagram of 0.2 mm amplitude in space. (c) 3D diagram of 0.4 mm amplitude in space

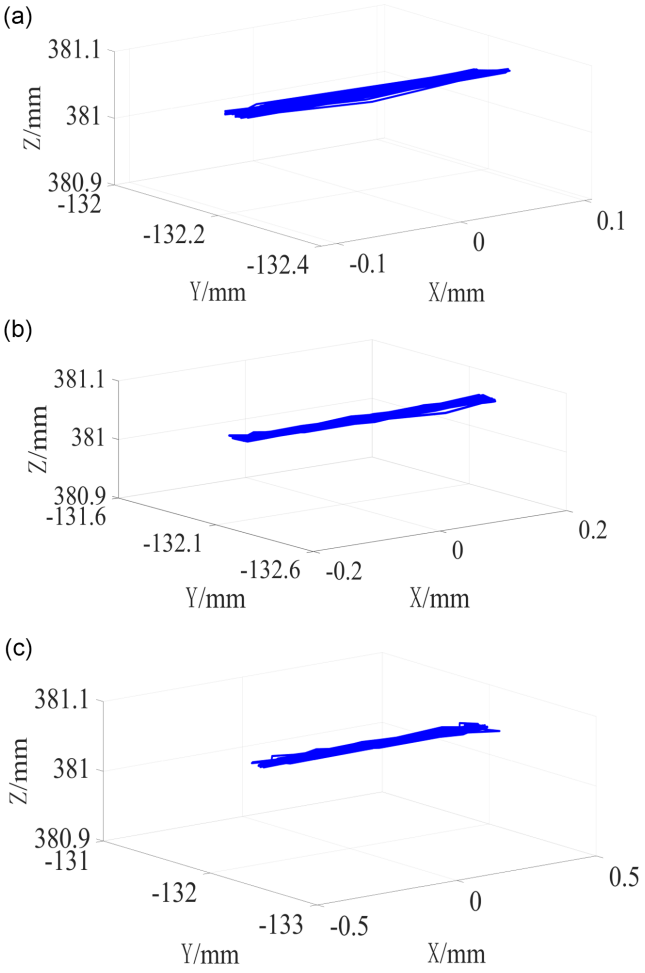


Table 5
Measurement errors of X-axis displacement at different amplitudes

Amplitude/mm	RMSE/mm	NRMSE/%
0.1	0.0298	14.57
0.2	0.0153	11.21
0.4	0.0454	7.85

Figure 9
Measurement results of 0.1 mm displacement. (a) X-axis displacement in the first second. (b) X-axis displacement in the second. (c) Error in X-axis displacement

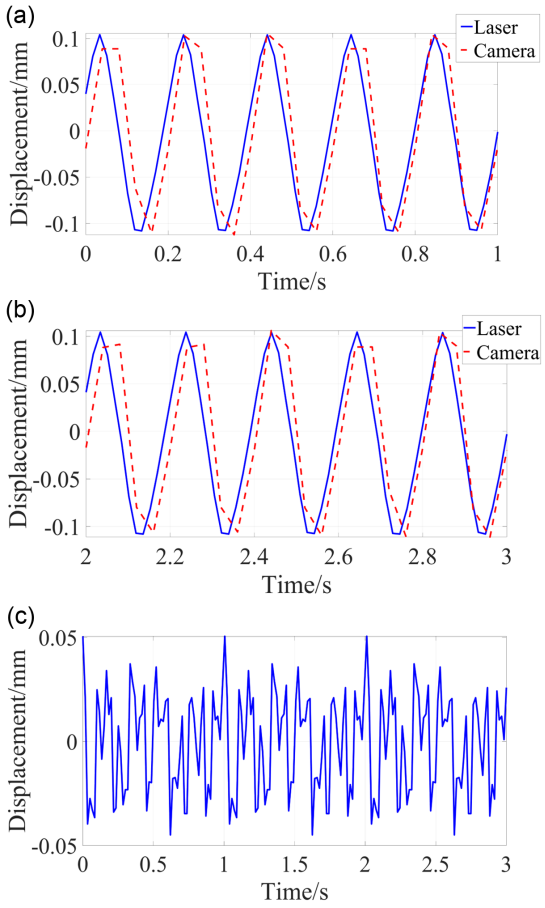
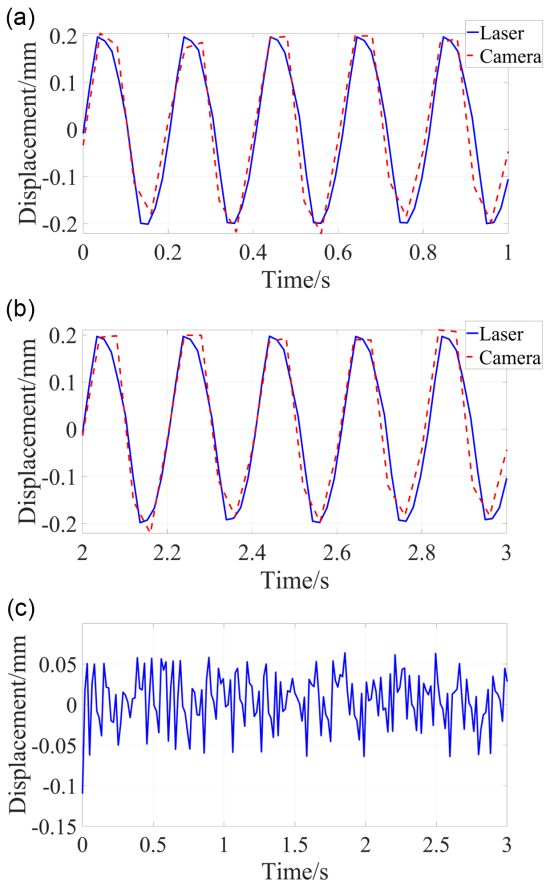


Figure 10
Measurement results of 0.2 mm displacement. (a) X-axis displacement in the first second. (b) X-axis displacement in the second. (c) Error in X-axis displacement



1) Comparing the X-axis displacement measurement results of the visual measurement method in this paper with those of the laser displacement sensor, it is found that the former method exhibits minimal errors and a high degree of correlation with the latter. The maximum amplitude error for an oscillation amplitude of 0.1 mm is 0.0153 mm, for 0.2 mm it is 0.0182 mm, and for 0.4 mm it is 0.0591 mm.
2) The visual measurement method in this paper still has some errors, which are caused by two reasons. First, the target object will deform after magnification, leading to changes in the tracking box during the tracking process, resulting in deviations in the tracking results. The second reason is that there will be a certain amount of error in stereoscopic matching and disparity map generation. Furthermore, the Y-axis and Z-axis displacements are further extracted from the spatial three-dimensional figure, as shown in Figures 12 to 14.

From Figures 12 to 14, it can be seen that due to the small displacement in the Y-axis direction, the displacement curve is chaotic. As the cantilever beam undergoes planar motion in space, the Z-axis displacement is a constant value of 381.0566 mm. The distance from the left optical center of the binocular camera to the marker measured by a ruler is 381 mm, with an error of 0.0566 mm, which is relatively small.

The displacement measurement results in the X-axis direction were transformed through Fourier analysis to obtain the spectrum plot, as shown in Figure 15.

Figure 15 shows that regardless of whether the vibration amplitude is 0.1 mm, 0.2 mm, or 0.4 mm, the frequency measured by the method is 5 Hz, and the frequency measured by the laser displacement sensor is also 5 Hz, which matches the vibration frequency set in the experiment. Therefore, the measurement results are accurate.

Taking into account the spatial three-dimensional shape of the cantilever beam, displacement measurements in the X-axis, Y-axis, and Z-axis directions, as well as the spectrum plot in the X-axis direction, it can be concluded that this method can achieve accurate measurement of three-dimensional displacement.

Figure 11
Measurement results of 0.4 mm displacement. (a) X-axis displacement in the first second. (b) X-axis displacement in the second. (c) Error in X-axis displacement

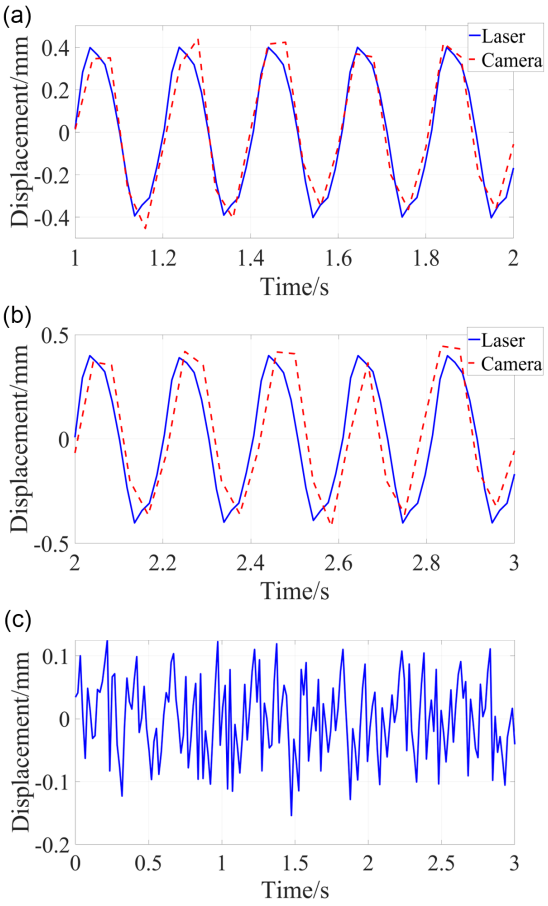


Figure 12
0.1 mm Y-axis and Z-axis displacement diagram. (a) Y-axis displacement. (b) Z-axis displacement

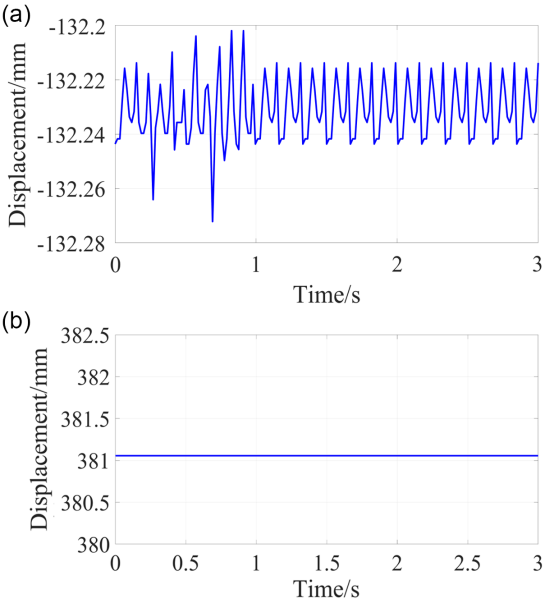


Figure 13
0.2 mm Y-axis and Z-axis displacement diagram. (a) Y-axis displacement. (b) Z-axis displacement

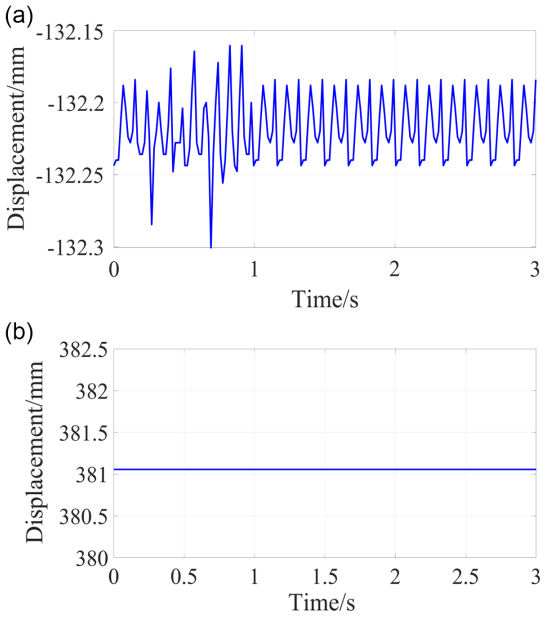


Figure 14
0.4 mm Y-axis and Z-axis displacement diagram. (a) Y-axis displacement. (b) Z-axis displacement

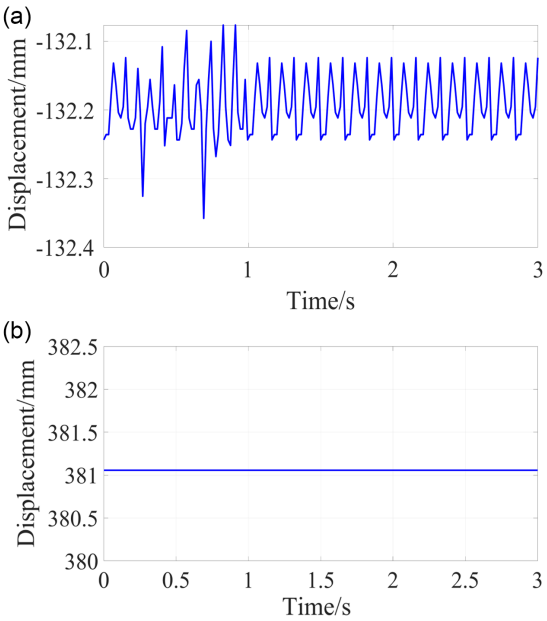
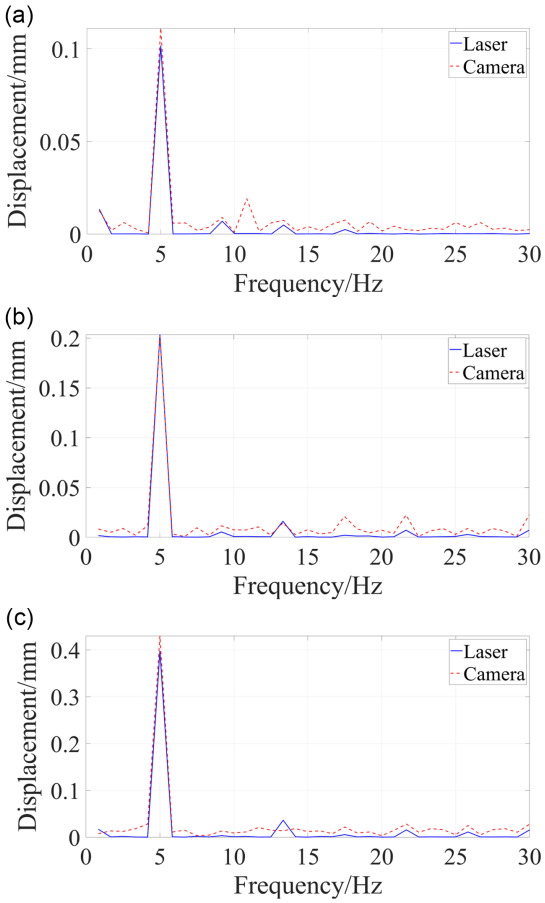


Figure 15
X-axis direction spectrum plot with different amplitudes. (a) Amplitude of 0.1 mm. (b) Amplitude of 0.2 mm. (c) Amplitude of 0.4 mm



4. Conclusion

This paper presents a method for measuring three-dimensional micro-vibration based on visual amplification. In cases where target tracking is feasible, the method utilizes a learning-based video motion amplification algorithm and the DaSiamRPN algorithm to convert captured vibration videos into pixel displacement information. Actual displacement values are then obtained through size conversion. In situations with extremely intense lighting, image preprocessing is required before motion amplification. Through spatial vibration displacement measurements of the cantilever beam, as well as measurements of displacement along the X-axis, Y-axis, and Z-axis, along with the spectrum plot in the X-axis direction, the measurement accuracy of the method was verified.

Funding Support

The authors gratefully acknowledge the support from Sichuan Science and Technology Program (2023NSFSC0858).

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Liangjun Luo: Conceptualization, Software, Validation, Investigation, Writing – original draft, Writing – review & editing. **Shuo Liu:** Methodology, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Fei Li:** Conceptualization, Methodology, Formal analysis, Data curation, Writing – review & editing. **Wenhao Zou:** Methodology, Software, Formal analysis, Data curation, Writing – review & editing, Visualization. **Xi Wang:** Conceptualization, Supervision, Funding acquisition. **Tao Wang:** Resources. **Guoqiang Fu:** Resources. **Caijiang Lu:** Resources.

References

[1] Dreier, F., Günther, P., Pfister, T., Czarske, J. W., & Fischer, A. (2013). Interferometric sensor system for blade vibration measurements in turbomachine applications. *IEEE Transactions on Instrumentation and Measurement*, *62*(8), 2297–2302. https://doi.org/10.1109/TIM.2013.2255993

[2] Peng, C., Zhu, M., Wang, K., Ren, Y., & Deng, Z. (2020). A two-stage synchronous vibration control for magnetically suspended system in the full speed range. *IEEE Transactions on Industrial Electronics*, *67*(1), 480–489. https://doi.org/10.1109/TIE.2018.2890498

[3] Haq, R., Zafar, N., & Baluch, A. (2016). Effect of structural damage on modal parameters of a stiffened composite panel using reduced order modelling technique. In *2016 13th International Bhurban Conference on Applied Sciences and Technology*, 62–68. https://doi.org/10.1109/IBCAST.2016.7429855

[4] Cakar, O., & Sanliturk, Y. (2005). Elimination of transducer mass loading effects from frequency response functions. *Mechanical Systems and Signal Processing*, *19*(1), 87–104. https://doi.org/10.1016/S0888-3270(03)00086-4

[5] Guo, L., Guo, W., Chen, D., Duan, B., & Shi, Z. (2025). Structural vibration measurement based on improved phase-based motion magnification and deep learning. *Mechanical Systems and Signal Processing*, *224*, 111945. https://doi.org/10.1016/j.ymssp.2024.111945

[6] de Figueiredo, H. V., Castillo-Zúniga, D. F., Costa, N. C., Saotome,O., & da Silva, R. G. A. (2021). Aeroelastic vibration measurement based on laser and computer vision technique. *Experimental Techniques*, *45*, 95–107. https://doi.org/10.1007/s40799-020-00399-0

[7] Dong, C. Z., Bas, S., & Catbas, F. N. (2020). Investigation of vibration serviceability of a footbridge using computer vision-based methods. *Engineering Structures*, *224*, 111224. https://doi.org/10.1016/j.engstruct.2020.111224

[8] Chen, L., Zhong, G., Han, Z., Li, Q., Wang, Y., & Pan, H. (2023). Binocular visual dimension measurement method for rectangular workpiece with a precise stereoscopic matching algorithm. *Measurement Science and Technology*, *34*(3), 035010. https://doi.org/10.1088/1361-6501/aca707

[9] Cheng, J., Jiang, H., Wang, D., Zheng, W., Shen, Y., & Wu, M. (2024). Analysis of position measurement accuracy of

boom-type roadheader based on binocular vision. *IEEE Transactions on Instrumentation and Measurement*, *73*, 1–12. https://doi.org/10.1088/1361-6501/aca707

[10] Lin, G., Lu, Z. R., Liu, J., & Wang, L. (2024). Monocular out-of-plane vibration measurement using parametric homography. *Journal of Sound and Vibration*, *570*, 118140. https://doi.org/10.1016/j.jsv.2023.118140

[11] Yang, M., Wang, Y., Liu, Z., Zuo, S., Cai, C., Yang, J., & Yang, J. (2022). A monocular vision-based decoupling measurement method for plane motion orbits. *Measurement*, *187*, 110312. https://doi.org/10.1016/j.measurement.2021.110312

[12] Wang, S., Li, X., Zhang, Y., & Xu, K. (2024). Effects of camera external parameters error on measurement accuracy in monocular vision. *Measurement*, *229*, 114413. https://doi.org/10.1016/j.measurement.2024.114413

[13] Liu, W., Cai, C., Liu, Z., Yang, M., Zuo, S., & Wang, P. (2024). Monocular vision-based dynamic calibration method for determining the sensitivities of low-frequency tri-axial vibration sensors. *Optics Express*, *32*(6), 10130–10145. https://doi.org/10.1364/OE.503013

[14] Fukuda, Y., Feng, Q., Narita, Y., Kaneko, S. I., & Tanaka, T. (2013). Vision-based displacement sensor for monitoring dynamic response using robust object search algorithm. *IEEE Sensors Journal*, *13*(12), 4725–4732. https://doi.org/10.1109/JSEN.2013.2273309

[15] Caetano, E., Silva, S., & Bateira, J. (2011). A vision system for vibration monitoring of civil engineering structures. *Experimental Techniques*, *35*(4), 74–82.

[16] Sieffert, Y., Vieux-Champagne, F., Grange, S., Garnier, P., Duccini, J. C., & Daudeville, L. (2016). Full-field measurement with a digital image correlation analysis of a shake table test on a timber-framed structure filled with stones and earth. *Engineering Structures*, *123*, 451–472.

[17] Ansari, A. I., Sheikh, N. A., & Kumar, N. (2024). Evaluation of the energy absorbing capacity of the two combinations of TPMS structure subjected to different compressive strain rates. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, *46*, 334. https://doi.org/10.1007/s40430-024-04925-8

[18] Rajaram, S., Vanniamparambil, A., Khan, F., Bolhassani, M., Koutras, A., Bartoli, I., . . . , & Kontsos, A. (2017). Full-field deformation measurements during seismic loading of masonry buildings. *Structural Control and Health Monitoring*, *24*(4), e1903. https://doi.org/10.1002/stc.1903

[19] Jiang, M., Gu, Q., Aoyama, T., Takaki, T., & Ishii, I. (2017). Real-time vibration source tracking using high-speed vision. *IEEE Sensors Journal*, *17*(5), 1513–1527. https://doi.org/10.1109/JSEN.2016.2647690

[20] Zhu, J., Zhang, C., Lu, Z., & Li, X. (2021). A multi-resolution deep feature framework for dynamic displacement measurement of bridges using vision-based tracking system. *Measurement*, *183*, 109847. https://doi.org/10.1016/j.measurement.2021.109847

[21] Huang, M., Zhang, B., Lou, W., & Kareem, A. (2021). A deep learning augmented vision-based method for measuring dynamic displacements of structures in harsh environments. *Journal of Wind Engineering and Industrial Aerodynamics*, *217*, 104758. https://doi.org/10.1016/j.jweia.2021.104758

[22] Lu, B., Bai, B., & Zhao, X. (2023). Vision-based structural displacement measurement under ambient-light changes via deep learning and digital image processing. *Measurement*, *208*, 112480. https://doi.org/10.1016/j.measurement.2023.112480

[23] Yin, H., Liu, Z., Xu, Z., & Gao, L. (2020). An automatic visual monitoring system for expansion displacement of switch rail. *IEEE Transactions on Instrumentation and Measurement*, *69*(6), 3015–3025. https://doi.org/10.1109/TIM.2019.2927547

[24] Yoon, H., Elanwar, H., Choi, H., Golparvar-Fard, M., & Spencer, Jr, B. F. (2016). Target-free approach for vision-based structural system identification using consumer-grade cameras. *Structural Control and Health Monitoring*, *23*(12), 1405–1416. https://doi.org/10.1002/stc.1850

[25] Li, H., & Zhang, B. (2021). Application of integrated binocular stereo vision measurement and wireless sensor system in athlete displacement test. *Alexandria Engineering Journal*, *60*(5), 4325–4335. https://doi.org/10.1016/j.aej.2021.02.033

[26] Zhou, Y., Li, Q., Chu, L., Ma, Y., & Zhang, J. (2020). A measurement system based on internal cooperation of cameras in binocular vision. *Measurement Science and Technology*, *31*(6), 065002. https://doi.org/10.1088/1361-6501/ab6ecd

[27] Yuan, C., Xiong, B., Li, X., Sang, X., & Kong, Q. (2022). A novel intelligent inspection robot with deep stereo vision for three-dimensional concrete damage detection and quantification. *Structural Health Monitoring*, *21*(3), 788–802. https://doi.org/10.1177/14759217211010238

[28] Shao, Y., Li, L., Li, J., An, S., & Hao, H. (2021). Computer vision based target-free 3D vibration displacement measurement of structures. *Engineering Structures*, *246*, 113040. https://doi.org/10.1016/j.engstruct.2021.113040

[29] Yue, Z., Huang, L., Lin, Y., & Lei, M. (2024). Research on image deformation monitoring algorithm based on binocular vision. *Measurement*, *228*, 114394. https://doi.org/10.1016/j.measurement.2024.114394

[30] Wang, A., Zhang, Z., & Qiao, R. (2023). A visual measurement method of the three-dimensional coordinates of the center of forgings with asymmetric flash for the intelligent hammer die forging. *Measurement*, *210*, 112581. https://doi.org/10.1016/j.measurement.2023.112581

[31] Wu, Z., Chen, G., Ding, Q., Yuan, B., & Yang, X. (2021). Three-dimensional reconstruction-based vibration measurement of bridge model using UAVs. *Applied Sciences*, *11*(11), 5111. https://doi.org/10.3390/app11115111

[32] Yao, Q., Shabaz, M., Lohani, K., Wasim Bhatt, M., Panesar, G. S., & Singh, R. K. (2021). 3D modelling and visualization for vision-based vibration signal processing and measurement. *Journal of Intelligent Systems*, *30*(1), 541–553. https://doi.org/10.1515/jisys-2020-0123

[33] Oh, H., Jaroensri, R., Kim, C., Elgharib, M., Durand, F. E., Freeman, W. T., & Matusik, W. (2018). Learning-based video motion magnification. In *Proceedings of the European Conference on Computer Vision*, 633–648.

[34] Zhu, Z., Wang, Q., Li, B., Wu, W., Yan, J., & Hu, W. (2018). Distractor-aware Siamese networks for visual object tracking. In *Proceedings of the European Conference on Computer Vision*, 101–117.

[35] Wang, X., Li, F., Du, Q., Zhang, Y., Wang, T., Fu, G., & Lu, C. (2023). Micro-amplitude vibration measurement using vision-based magnification and tracking. *Measurement*, *208*, 112464. https://doi.org/10.1016/j.measurement.2023.112464

[36] Lipson, L., Teed, Z., & Deng, J. (2021). RAFT-stereo: Multilevel recurrent field transforms for stereo matching. In *2021 International Conference on 3D Vision*, 218–227. https://doi.org/10.1109/3DV53792.2021.00032

How to Cite: Luo, L., Liu, S., Li, F., Zou, W., Wang, X., Wang, T., . . . , & Lu, C. (2026). Three-Dimensional Vibration Visual Measurement Method Based on Binocular Camera. *Archives of Advanced Engineering Science*, *4*(1), 78–88. https://doi.org/10.47852/bonviewAAES52024620