Video-Based Displacement, Velocity and Acceleration Measurement of Structural Elements Using Object-Tracking Algorithms

DHATHRI MEDA, RAKESH KATAM,
PRAFULLA KALAPATAPU
and VENKATA DILIP KUMAR PASUPULETI

ABSTRACT

Civil structures experience forces from external events, operational loads, and environmental conditions that induce vibrations and can lead to deformation that can compromise their integrity and safety. Precise computation of their displacement, velocity and acceleration aids in understanding their dynamic behaviour, detect failures early, optimise design and guarantee long-term durability. Traditional sensor-based approach, need physical contact, offer limited measurement points, are expensive, face accessibility issues, and lack scalability. These drawbacks prompt the need for non-contact, vision-based measurement techniques that offer flexibility, affordability, and comprehensive motion tracking. In this project, we propose a video-based method that utilises OpenCV based object tracking algorithms to measure kinematics such as displacement, velocity and acceleration of structural elements like steel cantilever beam, single degree of freedom pendulum and scaled portal frames. Several tracking algorithms such as Lucas-Kanade optical flow, CSRT (Channel and Spatial Reliability Tracker), MIL (Multiple Instance Learning), and KCF (Kernelized Correlation Filters) were leveraged to extract exact motion data over time by processing vibration videos of these structural elements. The videos are pre-processed to identify the structural element as the region of interest (ROI). Tracker-based approaches (CSRT, MIL, KCF) and Lucas-Kanade optical flow were utilised and displacement, velocity and acceleration were derived. The workflow involves frame extraction, grayscale conversion, and tracking algorithms, where Lucas-Kanade estimates vectors at the pixel level, while feature-based trackers record the trajectory of the beam. Plots of displacement vs time, velocity vs time and acceleration vs time depict vibration patterns and natural frequencies. The Fast Fourier Transforms (FFT) derived from both accelerometer data and video-based measurements are compared to validate the accuracy of the video-based analysis. By combining state-of-the art computer vision methods with structural analysis, this work represents a possible step toward modern, non-invasive methods for evaluating dynamic structural reactions. The comparative use of multiple tracking methods improves the approach's reliability and applicability, paving the way for broader implementation in engineering processes. This approach can be leveraged to bigger structures. This method can be used in subsequent research to remotely examine the structural health of large-scale structures like buildings and bridges. Accuracy and efficiency can be further increased by incorporating deep learning for improved feature tracking and real-time processing.

INTRODUCTION AND BACKGROUND

Structural Health Monitoring (SHM) is an essential domain concerned on assessing the integrity, performance and safety of structures throughout their operational life. It is essential for extending lifespan of infrastructure, optimizing maintenance schedules, and preventing catastrophic failures. Traditional SHM systems rely heavily on contact-based sensors such as accelerometers, strain gauges, and laser vibrometers to collect data related to structural motion, stress and strain [1]. Frequency-based approaches are a key subset of vibration-based SHM (VBSHM), though the extraction of modal parameters often requires complex algorithmic processing. This work aims to review and enhance VBSHM approaches, with an emphasis on modal parameter estimation using both traditional and modern methods.

Recent advances in sensor technologies and cloud-based computation have led to a surge in data-driven approaches for SHM. As highlighted by Rakesh et al., machine learning (ML) has emerged as a powerful tool to analyse sensor-derived data, enabling better interpretation of structural stability. Integration of ML within SHM not only increases automation but also enhances predictive capabilities. The use of vision-based and vibration-based monitoring frameworks alongside ML algorithms provides a more holistic assessment of structural integrity [2]. Additionally, Rakesh et al. emphasize that structural degradation is inevitable and depends on environmental and construction variables. Unexpected loading conditions can accelerate failure mechanisms. Their study, which involved experimental and numerical analysis of 21 cantilever beams with varying damage conditions, demonstrated the sensitivity of natural frequencies to damage severity and location. These findings confirm that vibration-based damage detection, when combined with intelligent systems, holds significant promise for improving structural resilience [3]. While effective, traditional SHM systems are often labour-intensive, time-consuming, costly and complex [2].

Vision-based monitoring techniques have gained considerable traction over the past decade. Leveraging advances in computer vision, image processing, and camera technology, these methods offer a non-contact, cost-effective, and scalable alternative to traditional sensor-based approaches [4]. Several object tracking algorithms have been employed to estimate structural response from video, including correlation filter-based methods such as CSRT, KCF, and MIL, as well as optical flow techniques like Lucas-Kanade algorithm [5-8]. These trackers work by identifying a Region of Interest (ROI) on the structure and following its motion frame-by-frame enabling the estimation of displacement, velocity and acceleration without physical contact. Ponmolar et al. employed Lucas Kanade method to track vehicles using fixed cameras, allowing for accurate monitoring of movement patterns and speed estimation [9]. Rani et al. employed KCF to track objects accurately for monitoring structural integrity in real-time allowing for timely interventions in case of anomalies [10]. Liu et al. incorporated Convolutional Neural Netowork (CNN) features so that KCF can better predict object locations and adapt to scale changes, enhancing tracking accuracy [11]. Farhodov et al. combined deep learning-based object detection with CSRT, leveraging Faster R-CNN for efficient tracking, particularly in dynamic environments [12].

One of the key analyses in SHM is the frequency spectrum, obtained using the Fast Fourier Transform (FFT). Frequency-domain representations of displacement signals help in identifying dominant frequencies, which correlate with the natural modes of the structure. These dominant frequencies are essential for detecting anomalies, shifts due to damage, and understanding vibrational behaviour [13]. In this study, object tracking was performed using algorithms such as CSRT, MIL, KCF, and Lucas-Kanade Optical Flow, and the resulting displacement data was used to extract frequency components that reflect the structural dynamics- mirroring the capability of traditional physical sensors [14]. Despite their advantages, vision-based SHM methods face limitations, such as sensitivity to lighting conditions, motion blur, occlusions, and the need for camera calibration [15]. Additionally, performance may degrade for low-amplitude vibrations or long-distance monitoring. However, recent research efforts have explored ways to overcome these challenges using deep-learning, multi-camera stereo systems, and feature-based enhancement techniques [16]. In summary, vision-based techniques represent a promising direction for SHM, particularly in environments where traditional sensor deployment is impractical or cost-prohibitive. As tracking algorithms and computer vision hardware continue to evolve, their integration into mainstream SHM workflows is becoming increasingly feasible [17].

METHODOLOGY

Experimental Setup and Video Acquisition:

The investigate the dynamic response of physical structures under free vibration, an experimental setup was devised involving three distinct types- a steel cantilever beam, a single degree of freedom pendulum and a scaled portal frame. These structures were chosen in order to provide a comparative understanding of their dynamic motion utilising vision-based methodologies by capturing a variety of vibrational features including linear, nonlinear, and multi-degree-of-freedom behaviour. Each structure was excited manually to initiate free vibration. An iPhone was utilised to capture each structure's dynamic activity in order to gather motion data. The camera was set up on a sturdy tripod or long stick and positioned at the right distance to record high-resolution videos at 30 frames per second (fps). This ensured clear view of the region of interest (ROI)-typically the tip of the beam, the centre of the pendulum or critical joints in the portal frame. The videos were recorded in an outdoor setting to reduce external interference. The camera was kept motionless for each recording session, and the structure was always kept within the frame. Without making physical contact, this arrangement offered unambiguous visual data for tracking displacement. The videos were used in the following analysis step, which involved vision-based motion tracking algorithms, and were saved in common formats (.MP4/.MOV). Figure 1 represents this experimental setup.

(a) (b) (c)

Figure 1: Experimental Setup of (a) Single Degree of Freedom Pendulum, (b) Steel Cantilever Beam and (c) Scaled Portal Frame

ROI Selection, Preprocessing, and Displacement Estimation using Tracking and Optical Flow:

The first step is analysing structural motion is analysing structural motion from a video is to choose a Region of Interest (ROI) inside a frame. This is usually the area of the structure that shows the most displacement, like the cantilever's beam tip, the pendulum's centre, or a joint in the portal frame. The ROI was manually selected in the initial frame of each video, forming the basis for subsequent motion analysis. After determining a ROI, each video frame was subjected to preprocessing techniques like histogram equalization to improve contrast, Gaussian blurring to reduce noise and grayscale conversion to simplify processing. Grayscale images are ideal for tracking and optical flow algorithms since they preserve essential structural features and eliminate redundant colour information. Figure 2 represents the Region of Interest (ROI) along with frames from the tracked videos.

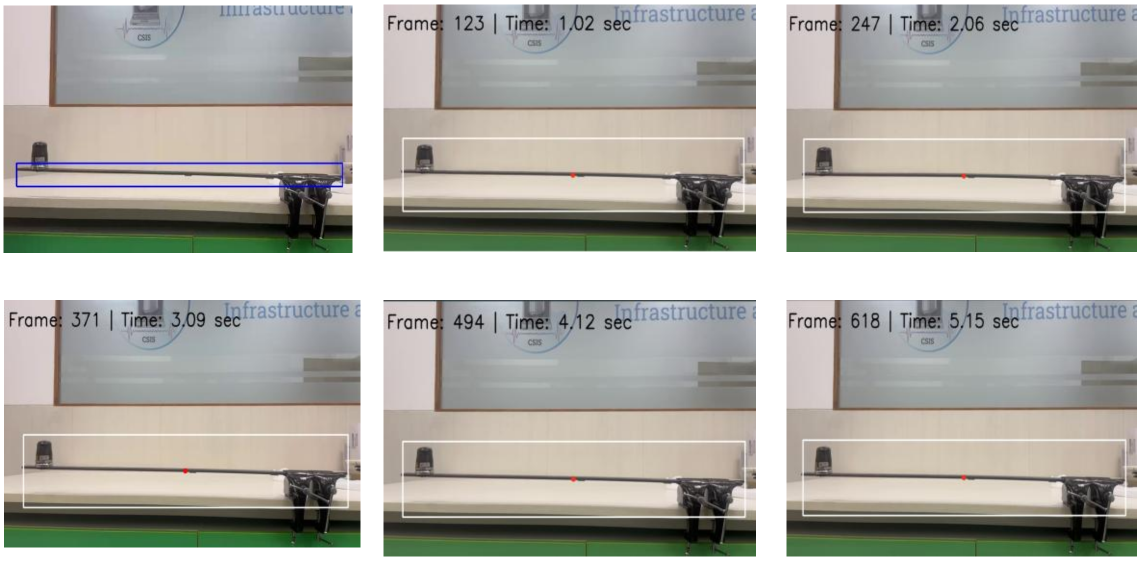

Figure 2: Region of Interest (ROI) and corresponding frames from the tracked videos

Two methods were used to estimate motion:

- Object tracking algorithms: To follow the ROI across frames, traditional trackers like CSRT, KCF, and MIL were used. These algorithms employ feature-based or correlation-based techniques to predict the object's location in the following frame. They are particularly helpful for fast and approximate displacement estimation [5][6][7].
- Optical Flow (Lucas-Kanade Method): Pixel-level motion vectors were calculated using the Lucas-Kanade optical flow for more detailed motion analysis. By examining variations in intensity between consecutive frames, this technique finds important characteristics (such as corners and edges) and monitors their movement over time. Even for tiny motions, it offers detailed and precise displacement information [8].

The Euclidean distance between consecutive tracked positions was computed to quantify the ROI's frame-to-frame displacement. Numerical differentiation techniques were then applied to estimate the velocity and acceleration of the structure. To analyse the dynamic behaviour of the system, a Fast Fourier Transform (FFT) was performed on the displacement (or acceleration) data, enabling the identification of dominant frequency components such as natural frequencies of vibration. Figure 3 represents this workflow diagram.

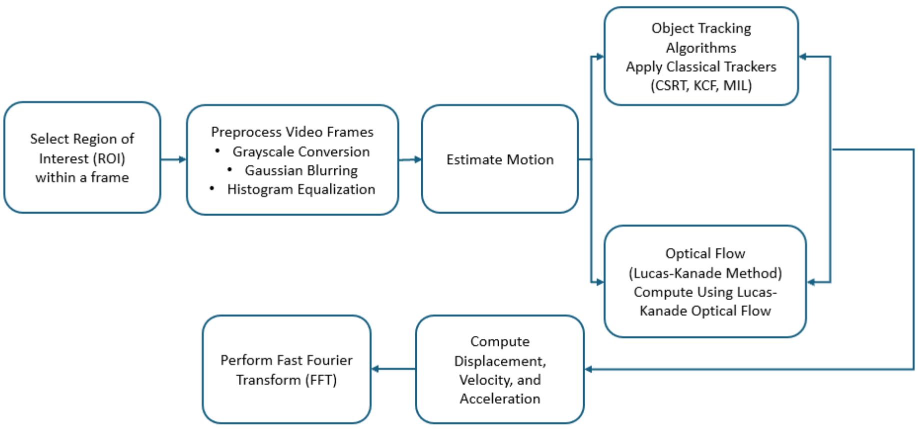

Figure 3: Workflow diagram

Validation and Analysis of Dynamic Response

The dynamic response of the structure was examined and contrasted with reference to accelerometer data in order to verify the motion predicted from video data. The analysis was based on frame-to-frame displacement of the Region of Interest (ROI), which was acquired using optical flow or object tracking. The displacement data was converted into velocity and acceleration profiles using numerical differentiation techniques.

To determine the system's dynamic characteristics, a Fast Fourier Transform (FFT) was performed on the displacement and acceleration time series. Dominant frequency components in the structural motion were identified by this frequency domain analysis. Equation (1) represents the Discrete Fourier Transform (DFT), which the Fast Fourier Transform (FFT) algorithm efficiently computes to convert time-domain signals into the frequency domain. These frequencies were then contrasted with those derived from accelerometer data, paying particular attention to the peaks that correspond to the structure's natural modes of vibration. A close match between the video-based and sensor-based dominant frequencies validated the accuracy of video processing pipeline and demonstrated its viability for non-contact structural health monitoring.

$$X(k) = \sum_{n=0}^{N-1} x(n).e^{-\frac{j2\Pi kn}{N}}, \quad k = 0, 1, 2, 3 \ldots, N-1 \tag{1}$$

Where:

- $x(n)$ is the input time-domain signal,
- $X(k)$ is the corresponding frequency component at index k,
- N is the total number of samples,
- j is the imaginary unit,
- $e^{-j2\Pi kn/N}$ represents the complex exponential basis functions.

RESULTS

The dynamic response of three different structures- the steel cantilever beam, single degree of freedom pendulum, and scaled portal frame- was extracted by utilising the most effective motion estimation algorithm for each scenario. Based on comparative performance, CSRT was employed for rubber fall providing better tracking of its shape-based movement and portal frame due to its high precision in capturing small displacements. For each case, displacement data was obtained by tracking the Region of Interest (ROI) across video frames. Velocity and acceleration were computed using numerical differentiation, and frequency analysis was performed using Fast Fourier Transform (FFT) to identify dominant vibration modes.

Figure 4 below illustrate the time series of displacement, velocity, and acceleration along with the FFT spectrum of cantilever beam, pendulum and portal frame responses. The frequency domain plot demonstrates a clear dominant peak corresponding to the natural mode of vibration.

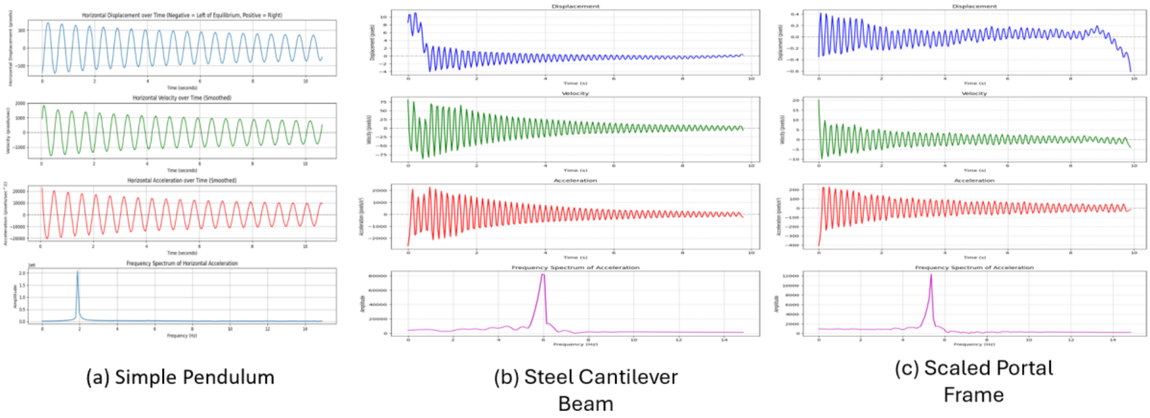

Figure 4: Time series of Displacement, Velocity and Acceleration and FFT spectrum of (a) Single Degree of Freedom Pendulum, (b) Steel Cantilever Beam and (c) Scaled Portal Frame

Figure 5 below illustrate Fast Fourier Transform (FFT) spectrums of video and accelerometer plotted together in a single graph for comparison.

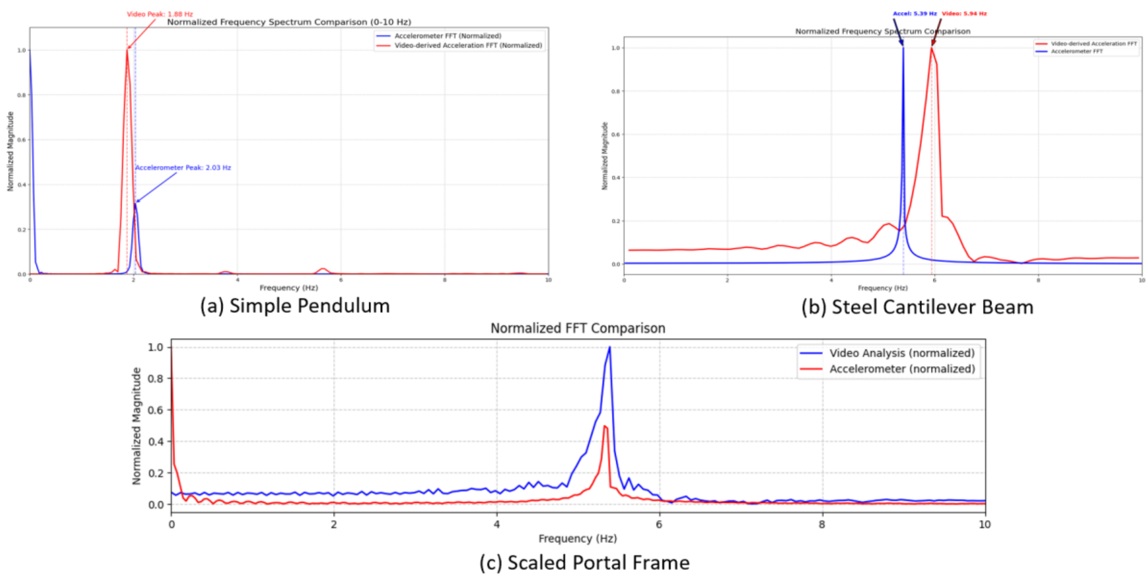

Figure 5: FFT comparison of contact and non-contact measurements for (a) Simple Pendulum, (b) Steel Cantilever Beam and (c) Scaled Portal Frame

Table 1 illustrates the comparison of contact and non-contact measurements of the three case studies that were conducted.

TABLE I. COMPARISON OF CONTACT AND NON-CONTACT MEASUREMENTS OF CASE STUDIES

Case Study	FFT from Video (Hz) (Non-contact)	FFT from Accelerometer (Hz) (Contact)	Difference (%)
Single Degree of Freedom Ball	1.88	2.03	7.39
Steel Cantilever Beam	5.94	5.39	10.20
Scaled Portal Frame	5.39	5.32	1.32

These results demonstrate minimal error across all cases. Thus, the video-based method has been shown to be successful. This further demonstrates the effectiveness of combining frequency and algorithm-specific movements.

CONCLUSION AND FUTURE WORK

This research presents a vision-based framework for tracking structural motion using object tracking algorithms and optical flow techniques. The approach effectively computes the displacement, acceleration and velocity of different physical structures under free vibration by utilizing ordinary video data. Validation against accelerometer data confirms the method's accuracy and potential for practical application. As illustrated in the comparison table, the frequency values obtained from video analysis closely match those from accelerometer data across different structural scenarios. Notably, for the Scaled Portal Frame, the difference between the FFT results from video and accelerometer is only 1.32%, indicating high precision. While slightly larger discrepancies were observed for the Single Degree of Freedom Ball and the Steel Cantilever Beam (7.98% and 10.20%, respectively), the results remain within acceptable margins for practical SHM applications.

These results indicate that with proper setup and algorithm selection, computer vision can be a reliable and affordable substitute for dynamic response monitoring, particularly in situations where the deployment of physical sensors is not feasible. Overall, this research lays the groundwork for non-contact structural health monitoring solutions that are accessible and scalable. In future work, the framework will be extended to enhance robustness and adaptability, enabling accurate motion extraction from any point of interest in the structure, regardless of scale or visibility constraints.

REFERENCES

1. Hannan, M. A., Hassan, K., & Jern, K. P. (2018). A review on sensors and systems in structural health monitoring: current issues and challenges. *Smart Structures and Systems*, *22*(5), 509–525. https://doi.org/10.12989/SSS.2018.22.5.509

2. Katam, R., Pasupuleti, V. D. K., & Kalapatapu, P. (2023). A review on structural health monitoring: past to present. *Innovative Infrastructure Solutions*, *8*(9), 248.

3. Katam, R., Kalapatapu, P., & Pasupuleti, V. D. K. (2022, June). A review on technological advancements in the field of data driven structural health monitoring. In *European workshop on structural health monitoring* (pp. 371-380). Cham: Springer International Publishing.
4. Segun, O. (2024). *Non-Invasive Measurement Techniques Using Computer Vision*. https://doi.org/10.20944/preprints202410.2006.v1
5. Hoon, J. Y. (2020). *Method and apparatus for tracking object based on correlation filter*.
6. Liu, S., Liu, D., Srivastava, G., Srivastava, G., Połap, D., & Woźniak, M. (2020). Overview and methods of correlation filter algorithms in object tracking. *Complex & Intelligent Systems*, *7*(4), 1895–1917. https://doi.org/10.1007/S40747-020-00161-4
7. Rai, S., & Mathew, R. (2019). *Adaptive Object Tracking Using Algorithms Employing Machine Learning* (pp. 381–388). Springer, Cham. https://doi.org/10.1007/978-3-030-43192-1_44
8. Zhang, J., Hu, D., & Pan, C. (2019). *A Correlation Filter Target Tracking Algorithm Combining LK Optical Flow*. https://doi.org/10.1145/3331453.3361645
9. M, P. (2023). Lucas Kanade based Optical Flow for Vehicle Motion Tracking and Velocity Estimation. *International Conference on Innovative Computing and Cloud Computing*, 1–6. https://doi.org/10.1109/ICCC57789.2023.10165227
10. Rani, A. S. N., Maik, V., & Chithravathi, B. (2017). Robust object tracking using kernalized correlation filters (KCF) and Kalman predictive estimates. *IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology*, 587–591. https://doi.org/10.1109/RTEICT.2017.8256664
11. Li, N., Wu, L., & Li, D. (2018). *Long-Term Tracking Algorithm Based on Kernelized Correlation Filter* (pp. 745–755). Springer, Cham. https://doi.org/10.1007/978-3-030-03766-6_84
12. Farhodov, X., Kwon, O.-H., Moon, K. S., Kwon, O.-J., Lee, S. H., & Kwon, K.-R. (2019). A New CSR-DCF Tracking Algorithm based on Faster RCNN Detection Model and CSRT Tracker for Drone Data. *Journal of Korea Multimedia Society*, *22*(12), 1415–1429. https://doi.org/10.9717/KMMS.2019.22.12.1415
13. Hale, E., Banerjee, P., & Ghimire, R. (2024). *Tensor Decomposition Analysis for UAV Anomaly Detection*. https://doi.org/10.2514/6.2024-4625
14. Chou, J.-Y., & Chang, C.-M. (2021). Image Motion Extraction of Structures Using Computer Vision Techniques: A Comparative Study. *Sensors*, *21*(18), 6248. https://doi.org/10.3390/S21186248
15. Dong, C.-Z. (2019). *Investigation of Computer Vision Concepts and Methods for Structural Health Monitoring and Identification Applications*. https://stars.library.ucf.edu/etd/6867/
16. Ghazvineh, S., Nouri, G., Hosseini Lavasani, S. H., Gharehbaghi, V., Noroozinejad Farsangi, E., & Noori, M. (2023). *7 Vibration-based damage detection using a novel hybrid CNN-SVM approach* (pp. 137–158). De Gruyter. https://doi.org/10.1515/9783110791426-007
17. Liu, T., Lei, Y., & Mao, Y. (2022). Computer Vision-Based Structural Displacement Monitoring and Modal Identification with Subpixel Localization Refinement. *Advances in Civil Engineering*, *2022*, 1–11. https://doi.org/10.1155/2022/5444101