Target-Free, Vision-Based System Identification of Civil Structures Using Unmanned Aerial Vehicles

KHALID ALKADY, ACHILLES G. RASQUINHA,
JOSEF T. BRANDL, CHRISTINE E. WITTICH
and CARRICK DETWEILER

ABSTRACT

Vibration-based structural health monitoring (SHM) frameworks rely upon accurately identified natural frequencies and mode shapes of structures in the field, which is critical information for damage diagnostics and model updating. Vibration-based techniques have traditionally relied on discrete contact-based sensors. Despite the success of traditional sensing modalities, challenges and limitations remain: 1) the sensors need to be placed at discrete locations, and 2) the structure needs to be accessed for instrumentation. Advancements in the fields of computer vision and robotics have facilitated the use of remote sensing technology, such as cameras and Unmanned Aerial Vehicles (UAVs), in vibration-based SHM applications to address the limitations of traditional sensors. Although UAVs have been used to monitor the dynamic response of structures, these applications have primarily relied on targets or GPS to track the structure's motion, which is not always feasible due to the large scale of civil structures. To this end, the main objective of this study is to develop an end-to-end target-free, vision-based framework for system identification of civil structures using UAVs. The proposed framework incorporates a phase-based motion estimation approach to extract the structural vibration information from videos without placing targets in the scene. For videos collected by UAVs, a correction needs to be applied to account for the camera's rigid body motion. To compensate for this motion, the framework extracts the power spectral density (PSD) plot of a static object in the scene and subtracts it from the PSDs of the structure-of-interest. To evaluate the efficacy and robustness of the developed framework, an experimental study was conducted to monitor the free vibration response of two single-degree-of-freedom structures using three different UAVs in a controlled laboratory environment. The analysis shows strong agreement between the results extracted from UAVs equipped with high-resolution cameras with those from a stationary camera and those from accelerometers. Furthermore, the results show that camera resolution, alignment, and motion can significantly impact the accuracy of the results. This study shows the potential of successfully incorporating UAVs into target-free vision-based dynamic monitoring frameworks.

Khalid Alkady and Christine E. Wittich, Department of Civil and Environmental Engineering, University of Nebraska-Lincoln, Nebraska 68588, U.S.A.
Achilles G. Rasquinha, Josef T. Brandl, and Carrick Detweiler, Department of Computer Science and Engineering, University of Nebraska-Lincoln, Nebraska 68588, U.S.A.

INTRODUCTION

Vibration-based structural health monitoring (SHM) frameworks rely upon accurately identified natural frequencies and mode shapes of structures in the field, which is critical information for damage diagnostics, prognostics, and model updating and calibration. Vibration-based SHM methods have traditionally relied on discrete contact-based (i.e., wired or wireless) sensors for dynamic monitoring [1-4]. Despite the reliability of traditional sensing modalities, several challenges and limitations remain: 1) the sensors need to be placed at discrete locations, which reduces the spatial resolution and may result in spatial aliasing [2], and 2) the structure of interest needs to be accessed for instrumentation, which may take significant time or may not even be possible due to complex site conditions.

Advancements in the fields of computer vision and robotics have facilitated the use of remote sensing technology, such as cameras, in vibration-based SHM applications to address the limitations of traditional sensing [5-10]. Previous vision-based frameworks have shown success in estimating the dynamic response of structures by tracking artificial targets or natural features [5-10]. While the use of natural features alleviates the challenges associated with incorporating artificial targets, the results are sensitive to the illumination and surface conditions of the structure due to the reliance on image intensity [11-13]. In contrast to these methods, phase-based motion estimation extracts structural vibration from local pixel phases of video frames, which is insensitive to illumination effects and surface conditions [11-13].

While phase-based motion estimation frameworks are robust, they have primarily been tested on stationary camera platforms [11]. There is a strong need to extend targetless approaches to moving camera platforms (i.e., Unmanned Aerial Vehicles, UAVs) such that vibration-based SHM frameworks can be safely and efficiently implemented on a wider range of civil structures. As an initial step, this study aims to:

1. Develop an end-to-end phase-based motion estimation framework for system identification of civil structures from videos collected by UAVs.
2. Investigate the impacts of camera motion, image resolution, and structural natural frequency on the robustness of the proposed framework using high-sensitivity accelerometers and infrared-based motion tracking for validation.

PHASE-BASED MOTION ESTIMATION FRAMEWORK

This section discusses the three stages of the proposed framework to process the UAV's videos to estimate the structure's natural frequencies: 1) local phase extraction, 2) blind source separation, and 3) baseline correction.

Local Phase Extraction

The phase-based approach developed by [11] was followed in this study to blindly extract the vibration characteristics of structures from videos collected by stationary and moving camera devices. A video of a vibrating structure contains the motion information encoded into each video frame $I(x + \delta(x, t))$, where x is the pixel coordinate and $\delta(x, t)$ is the spatially local and temporally varying motion.

A complex-steerable pyramid was used to extract the local phases from the video frames in this project. The pyramid consists of multi-scale Gabor-based filters (i.e., transform functions) at different orientations [12]. Since the displacements monitored in this project were only in the horizontal direction, only a single filter orientation parallel to the horizontal direction at the first spatial scale $\omega = 1$ was considered in the local-phase extraction process. Figure 1a shows a summary of the framework followed for the extraction of local phases from a sequence of video frames as outlined by [11]. To extract the local phases from an image I, the chosen filter is applied to the discrete 2D Fourier transform of I, as the filtering process is done in the frequency domain.

The response of the filtering process at spatial scale ω, $R_{\omega}(x,t)$, is provided in Equation 1, where $\rho_{\omega}(x,t)$ is local amplitude, $2\pi\omega_0(x+\delta(x,t))$ is the local phase $\psi(x,t)$ and ω_0 is the spatial frequency. The motion of the structure can be approximated to the centered phase, δ' (i.e., motion matrix). δ' is recovered from $\psi(x,t)$ after removing the temporal mean $2\pi\omega_0 x$. The motion matrix $\delta' \in \mathbb{R}^{N\times T}$, where N is total number of pixels and T is number of temporal samples.

$$R_{\omega}(x,t) = \rho_{\omega}(x,t)e^{j2\pi\omega_0(x+\delta(x,t))} \tag{1}$$

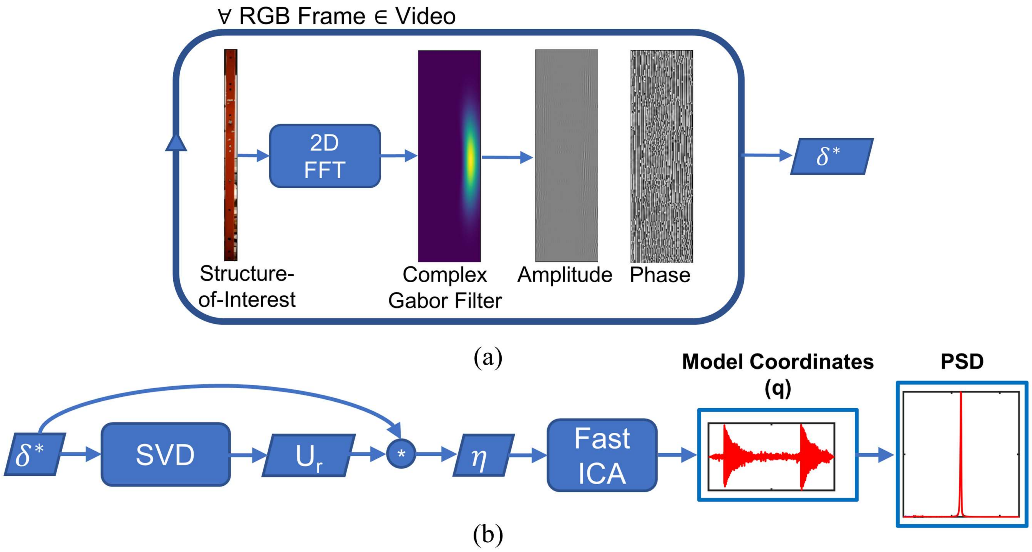
Blind Source Separation

To blindly extract the vibration characteristics from δ', a blind source separation (BSS) framework similar to [11] is followed to facilitate the output-only modal identification. BSS includes two steps: 1) dimensionality reduction of δ', and 2) modal separation. As shown in Figure 1b, singular value decomposition (SVD) is applied to extract the active components (U_r) in δ'. SVD projects most of the energy in the modal components of δ' into a small number of, r, principal components η. Equation (2) shows the computation of $\eta \in \mathbb{R}^{r\times T}$. For lightly damped structures, the number of active modes, n, can be assumed to be equal to the number of active components r. However, the components of η are linear mixtures of different modal coordinates [11]. Therefore, a FAST independent component analysis (ICA) is applied to efficiently extract de-mixed modal coordinates, q, from η, as shown in Figure 1b.

$$\eta = U_r^T * \delta' \tag{2}$$

Baseline Correction

For videos collected by a moving camera platform, a correction needs to be applied to account for the camera's rigid body motion. To compensate for this motion a novel baseline correction framework was developed, as shown in Figure 2. First, a region bounding part of the vibrating structure is cropped out, which is known as V_{str+u} . The phase-based motion extraction framework is used to generate a power spectral density (PSD) for $V_{str+uav}$, which is known as $PSD_{str+uav}$. The generated PSD plot includes the frequency information of both the UAV's motion and the structure's vibration. Therefore, to isolate the UAV's motion from the rest of the PSD, a region around a static rigid object in the vicinity of the structure-of-interest is used, which is known as V_{uav}. The phase-based framework was used to generate a PSD from V_{uav}, which is known as PSD_{uav}. Since the rigid object included in V_{uav} is assumed to remain at rest throughout the whole video, PSD_{uav} includes only the frequency information of the UAV motion, as shown in Figure 2. Hence, to isolate the camera motion, PSD_{uav} is subtracted from PSD_{str+u} , as demonstrated in Figure 2.



(a)

(b)

Figure 1. A framework for phase-based motion extraction: (a) extraction of local phases, (b) blind source separation for vibration extraction.

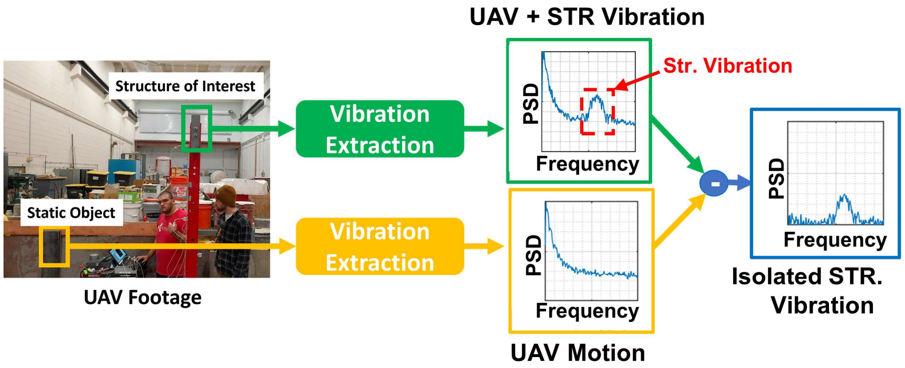
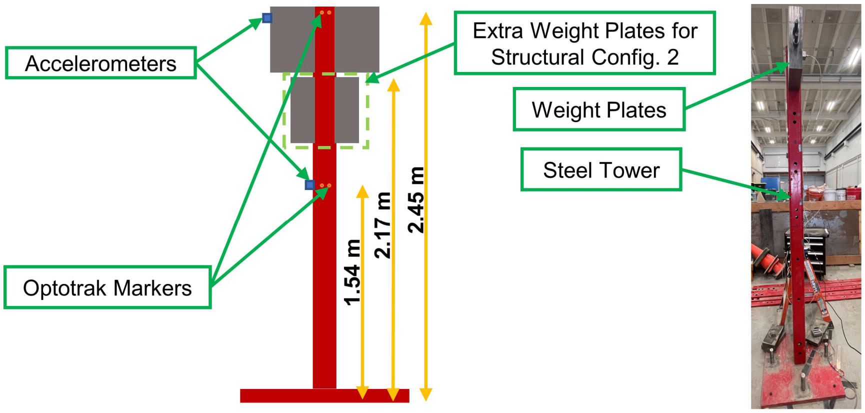


Figure 2. A framework for baseline correction

The structure-of-interest and the stationary object might not be at the same exact distance from the UAV, which might result in discrepancies between both $PSD_{str+uav}$ and PSD_{uav}. Hence, both $PSD_{str+uav}$ and PSD_{uav} are normalized and scaled before the baseline correction step to account for any discrepancies in the PSD plots.

EXPERIMENTAL SETUP AND INSTRUMENTATION

An experimental study was executed to investigate the effect of structural natural frequency, image resolution, and camera motion on the robustness of the developed phase-based vibration extraction framework. A 2.45-m tall reconfigurable HSS 101.6×101.6×9.53 mm steel tower was used as the specimen in this study, where weight plates were used to vary the dynamic properties of the steel tower during testing, as shown in Figure 3. Two structural configurations of the steel tower with unique natural frequencies were considered in the experimental program.



(a) (b)

Figure 3. Experimental Setup: (a) schematic, (b) structural configuration 1

TABLE I. IMAGERY DEVICES SPECIFICATIONS

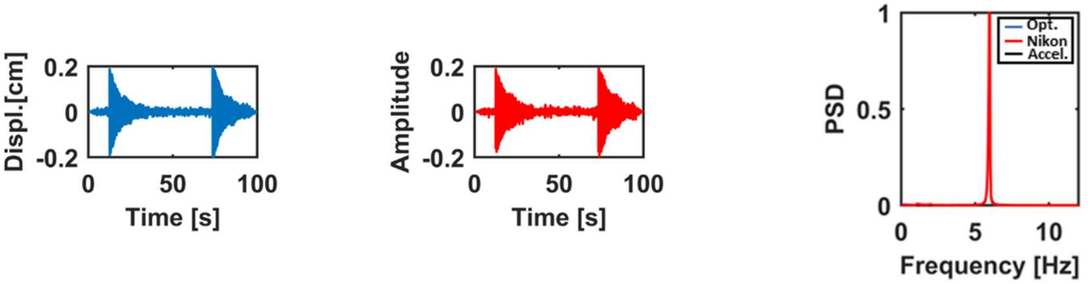
Device	Resolution [MP]	Frame Rate [Hz]	Camera Fixation
Nikon 5600 DSLR	24.2	60	Stationary
Mavic 2	20	60	UAV (gimbled)
Phantom 3	12.4	60	UAV (gimbled)
Tello	5	30	UAV (rigid)

The tower was subjected to an impulse load from a mallet at the top level approximately every 60 seconds for excitation for nearly 120 seconds. Tower motion was monitored using high-sensitivity piezoelectric accelerometers (PCB 352C34), an infrared-based motion tracking system (Optotrak Certus), a stationary camera, and cameras mounted on three commercially available UAVs. The accelerometers and infrared markers were mounted on the specimen at two levels (i.e., 1.54 and 2.45 m), and the camera and UAVs were at an approximate distance of 2.5 m from the tower. Table I shows the specifications of the three UAVs (moving camera platform) and the stationary camera. The DJI Tello UAV had a rigidly mounted camera, while DJI Phantom and Mavic had gimbled cameras. The Tello uses basic IMU stabilization, while the Phantom and Mavic use IMU and vision-based stabilization. None of the UAVs used GPS correction, as testing was indoors. For each experiment there were two tests conducted to evaluate the reliability of the results.

RESULTS AND DISCUSSION

Phase-Based Motion Estimation Validation

The videos collected by the stationary camera (Nikon DSLR) were used to validate the accuracy of the local phase extraction and blind source separation algorithms in extracting natural frequencies and vibrations. For instance, Figure 4 shows the validation results for Nikon's first test of structural configuration 1. Figures 4(a) and 4(b) show that the displacement history of the Optotrak data is in strong agreement with the displacement history of the first modal coordinate, q, extracted from the Nikon footage. Also, Figure 4(c) shows a great agreement across the PSDs of the three sensing modalities. Furthermore, Table II shows that all the natural frequencies estimated via



(a) (b) (c)

Figure 4. Phase-based motion estimation algorithm validation results: (a) Optotrak displacement, (b) Nikon's footage first modal coordinate, (c) overlaid PSDs

the Nikon DSLR are in great agreement with those of the accelerometers. Therefore, these results show the potential of the phased-based motion estimation framework to successfully characterize the dynamics properties of structures from videos.

System Identification

Table II shows summary results of the tests conducted, where each camera device was used to monitor the free response of two structural configurations. It should be noted that the test 3 was excluded as the Mavic UAV was not yet stable in the air. The results show that the developed baseline correction approach was capable of accurately eliminating the UAV's motion from the PSDs for the Mavic. However, the approach struggles to isolate the Phantom's and Tello's motion from the structure's motion. Figure 5 shows a sample of the Mavic's results before and after baseline correction for structural configuration 1, respectively.

PHANTOM

Although Table II shows that the structure's natural frequencies were not detected in any of the videos collected during the Phantom's moving tests, the natural frequencies were detected in the Phantom's stationary tests. This discrepancy can be attributed to both the UAV motion in air and misalignment of the Phantom's gimbled camera. The Phantom's camera field-of-view was not perfectly parallel to the tower's face as compared to the Mavic's and Phantom's stationary mode field-of-views. It is crucial to have the camera's field-of -view in the same plane of the structure's motion and filter orientation used in the phase-based extraction stage to yield accurate results.

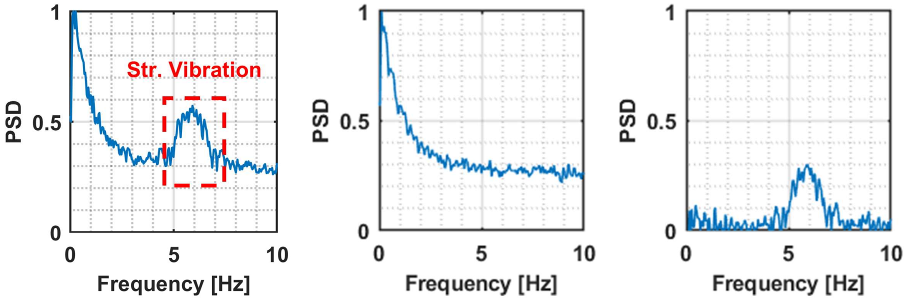
TELLO

The results show that the Tello was not able to identify the structure's natural frequencies in all the hovering and stationary tests. Although the Tello was the least stable aircraft in the air, the failure of extracting natural frequencies in the moving tests using the baseline correction approach cannot be solely attributed to the aircraft motion. The stationary test results show that the Tello's camera resolution was not sufficient to extract natural frequencies from videos collected at a distance of ~ 2.5 meters from the tower. Hence, this indicates the importance of acquiring high-resolution cameras for accurate vibration monitoring.

TABLE II. SUMMARY OF SYSTEM IDENTIFICATION RESULTS

Test No.	**Config.**	**Device**	**Test Description**	**Natural Frequency [Hz]**	
				Camera	***Accelerometer***
1	1	Nikon	Stationary	5.98	5.94
2				5.97	
3		Mavic	Hovering	ND	
4				5.95	
5		Phantom	Hovering	ND	
6				ND	
7		Tello	Hovering	ND	
8				ND	
9		Phantom	Stationary	5.97	
10				5.99	
11		Tello	Stationary	ND	
12				ND	
13	2	Nikon	Stationary	5.1	5.09
14				5.08	
15		Mavic	Hovering	5.03	
16				5.05	
17−20		Phantom/ Tello	Hovering	ND	

*ND indicates that the structure's natural frequency was not detected in the PSD plots generated



(a) (b) (c)

Figure 5. Mavic results for structural configuration 1: (a) PSD of the structural vibration and UAV motion, (b) PSD of the UAV motion, (c) PSD of the structural vibration only.

CONCLUSION

An end-to-end framework was developed to extract structural vibrations from videos using a phase-based motion extraction approach without placing targets on the structure-of-interest. A novel baseline correction approach was developed to isolate the camera motion from the structure's motion in the frequency domain. In this study, the proposed framework was used to monitor the free response of two structural configurations of a steel tower in a controlled laboratory environment considering three UAVs, a stationary camera, and an Optotrak motion capture system and high-sensitivity accelerometers for validation. The impacts of image resolution, camera motion, and structural configuration were examined in this study. The analysis shows a strong agreement between the results extracted from the stationary camera (Nikon), stationary Phantom, hovering Mavic with those of the Optotrak and accelerometers. However, the results showed that the camera motion, misalignment, and low resolution yielded inaccurate results for the Phantom and Tello UAVs, respectively. Therefore, the results indicate the potential of using the developed framework along with UAVs that are equipped with high-resolution and well-aligned cameras to conduct accurate full-field monitoring of structures.

Despite the promising results, this study was limited only to two SDOF structures of close natural frequencies. Hence, to fully evaluate the robustness of the proposed framework, complex civil structures (i.e., buildings) shall be monitored in the field. Testing structures in the field would enable the examination of the environmental conditions (i.e., wind) effects on the aircraft's motion and algorithm accuracy. Also, it shall be noted that the algorithm's current implementation requires the presence of a natural stationary object in the scene for baseline correction. Therefore, robust video stabilization techniques, which do not rely on tracking objects in the scene, can be explored to isolate the UAV's motion from the video without interfering with the structure's vibration.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of Dr. Richard L. Wood from the University of Nebraska-Lincoln for assistance with UAV flights and experiments.

REFERENCES

1. Doebling SW, Farrar CR, Prime MB. A summary review of vibration-based damage identification methods. Shock and vibration digest. 1998 Mar 1;30(2):91-105.
2. Fan W, Qiao P. Vibration-based damage identification methods: a review and comparative study. Structural health monitoring. 2011 Jan;10(1):83-111.
3. Friswell M, Mottershead JE. Finite element model updating in structural dynamics. Springer Science & Business Media; 1995.
4. Ewins DJ. Modal testing: theory, practice and application. John Wiley & Sons; 2009.
5. Spencer Jr BF, Hoskere V, Narazaki Y. Advances in computer vision-based civil infrastructure inspection and monitoring. Engineering. 2019 Apr 1;5(2):199-222.
6. Xu Y, Brownjohn JM. Review of machine-vision based methodologies for displacement measurement in civil structures. Journal of Civil Structural Health Monitoring. 2018 Jan;8:91-110.
7. Hoskere V, Park JW, Yoon H, Spencer Jr BF. Vision-based modal survey of civil infrastructure using unmanned aerial vehicles. Journal of Structural Engineering. 2019 Jul 1;145(7):04019062.
8. Yoon H, Elanwar H, Choi H, Golparvar-Fard M, Spencer Jr BF. Target-free approach for vision-based structural system identification using consumer-grade cameras. Structural Control and Health Monitoring. 2016 Dec;23(12):1405-16.
9. Sony S, Laventure S, Sadhu A. A literature review of next-generation smart sensing technology in structural health monitoring. Structural Control and Health Monitoring. 2019 Mar;26(3):e2321.
10. Feng D, Feng MQ. Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection–A review. Engineering Structures. 2018 Feb 1;156:105-17.
11. Yang Y, Dorn C, Mancini T, Talken Z, Kenyon G, Farrar C, Mascareñas D. Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification. Mechanical Systems and Signal Processing. 2017 Feb 15;85:567-90.
12. Wadhwa N, Rubinstein M, Durand F, Freeman WT. Phase-based video motion processing. ACM Transactions on Graphics (TOG). 2013 Jul 21;32(4):1-0.
13. Chen JG, Wadhwa N, Cha YJ, Durand F, Freeman WT, Buyukozturk O. Modal identification of simple structures with high-speed video using motion magnification. Journal of Sound and Vibration. 2015 Jun 9;345:58-71.