Vision-Based Modal Analysis of Built Environment Structures with Multiple Drones*

Michele Bolognini[a,*], Giovanni Izzo[a], Daniele Marchisotti[b], Lorenzo Fagiano[a], Maria Pina Limongelli[c], Emanuele Zappa[b]

[a]*Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Via Giuseppe Ponzio 34/5, 20133 Milan, Italy*
[b]*Dipartimento di Ingegneria Meccanica, Politecnico di Milano, via La Masa 1, 20156 Milan, Italy*
[c]*Dipartimento di Architettura, Ingegneria delle Costruzioni e Ambiente Costruito, Politecnico di Milano, Via Giuseppe Ponzio 31, 20133 Milan, Italy*

Abstract

Unmanned Aerial Vehicles are employed for vision-based modal analysis of civil infrastructure, as they overcome range limitations of fixed cameras and measure the oscillations of a structure up close. Nevertheless, their potential is not fully exploited: they are often piloted manually and one at a time, though one drone is unable to capture high resolution displacement of a whole structure. An approach is explored here, employing multiple drones simultaneously to estimate natural frequencies and modal shapes of a structure, by synchronizing their measurement. The ability of the method to detect modal parameter variations is assessed, such that it can identify anomalies in the structure. Procedures are applied to a test structure, yielding maximum natural frequency estimation errors of 0.2% with respect to accelerometers. The results suggest the accuracy of the approach is high enough to warrant further development and support autonomous, multi-drone applications to the inspection of the built environment.

Keywords: Unmanned Aerial Vehicle, Building inspection, Vision-based inspection, Modal analysis, Signal cross-correlation, Signal synchronization.

1. Introduction

Unmanned Aerial Vehicles (UAVs) are increasingly being adopted in many sectors, such as agriculture [1], military [2], natural asset management [3, 4] and construction [5, 6]. Besides the recent decreasing trend of their cost, the reasons for their success are multiple. Their ease of use makes previously cumbersome tasks much simpler. In fact, they can reach remote places, thus extending the reach of human operators significantly. Furthermore, they are versatile, as they can bear a diverse array of payloads, hosting

*This research has been supported by the Italian Ministry of University and Research (MIUR) under the PRIN 2017 grant n. 201732RS94 "Systems of Tethered Multicopters".
*Corresponding author
Email address: michele.bolognini@polimi.it (Michele Bolognini)

for example on-board sensors like RGB and thermal cameras for visual inspections, LiDARs and depth cameras for mapping and 3D scanning, or gas sensing systems [7]. Thanks to these features, the use of UAVs in public infrastructure monitoring is promising. This is especially important in countries with an aging infrastructure, which poses serious safety concerns and ever-increasing monitoring and maintenance cost. In such regions, large investments for renewal are in general not planned in the foreseeable future, therefore management and maintenance are even more important. In fact, infrastructure expenditure still accounts for a sizable portion of national budgets; it is estimated that the countries of the European Union combined spend close to a hundred billion dollars in road infrastructure maintenance alone each year, and the USA bear a comparable cost [8]. As a consequence, any tool that could help lower at least some of the costs is worth investigating.

Within the maintenance and inspection field, we focus on vibration-based Structural Health Monitoring (SHM), which is commonly performed with accelerometers. These sensors measure accelerations in three dimensions (3D), which can be used to estimate modal parameters such as natural frequencies and modal shapes [9]. Nevertheless, a wide network of synchronized sensors is necessary to identify modal shapes with a high spacial resolution, and accelerometers are expensive to maintain. Additionally, most existing structures were built without an embedded sensor network, thus performing the analysis would entail installing accelerometers ad-hoc, a cumbersome and expensive task. Vision-based modal analysis shines in these scenarios: cameras can capture videos of relatively large sections of a structure, thus collecting more data than a single accelerometer [10]. Furthermore, they measure displacement rather than acceleration, although they do so more accurately along the two directions orthogonal to the camera axis than along the parallel one [11]. This typically results in a more accurate displacement measurement than the one obtained by integrating acceleration twice, due to the possible noise-induced drift the integration process. In this work, we investigate the potential of vision-based SHM with multiple drones and provide the following contributions. First, we introduce a mechanism through which multiple drone-mounted cameras are synchronized and data can be gathered about large sections of a structure at a time; second, we validate the new approach experimentally on a small cantilever structure and evaluate the accuracy with respect to both a fixed camera and accelerometers, as well as the capability to detect and isolate alterations, such as added mass or a faulty connection to the supporting structure. In the process, we show that markers are not necessary to produce an accurate measurement, since the approach can exploit visual features of the structure, such as a junction, a bolt or a welding spot.

The remainder of this paper is organized as follows. After a brief overview of the literature, we describe the problem in section 2 and present our method in section 3. Section 4 details the experimental set-up, and the following one presents the results of our experiments. The final section contains our conclusions and a discussion of the limitations of the approach in its current state, and it points to future development paths.

1.1. Literature Overview

Vibration modal analysis has been developed as a tool for SHM decades ago [12, 13]. It is defined as the process of determining the dynamic characteristics of a mechanical system in terms of natural frequencies, modal shapes and damping factors. These can

be used to formulate a modal model of the structure, and their variation over time can be tracked as an indicator of its health. Modal analysis is traditionally performed with accelerometers, but the decreasing costs of cameras and computational power has pushed the development of computer-vision methods [14]. It has been shown that, by measuring displacement with a fixed camera, it is possible to estimate the first modal frequencies and recreate the modal shapes of a beam that is excited by an external input in a real world scenario. In [15], the displacement of a structure across time is measured by identifying features in the first frames of the recorded video and tracking their motion in the following frames with the Lucas-Kanade [16] algorithm, based on optical flow calculation. One advantage of such a technique is that there is no need to install markers on the structure. Similarly, some studies [17, 18] attained sub-pixel accuracy in displacement measurement using markers on the target structure, thanks to pattern matching techniques adopted in image processing, both in laboratory and in realistic conditions. This fact suggests that markers could be exploited where accuracy is critical, such as in scenarios where the camera is far away from the structure or the oscillations are small enough to be hardly distinguishable from measurement noise. Furthermore, the authors of [19] have shown that natural frequencies can be estimated reliably with inexpensive consumer grade cameras, at least in laboratory conditions. While applications of these techniques in realistic scenarios are possible, further data processing is often necessary to detect oscillations of low amplitude [20, 21]. Most of the real buildings where these techniques are tested are bridges or slender structures [18, 22, 21, 23], which tend to exhibit wider oscillations than other structures. The same techniques can be applied to measurement of frequencies and mode shapes of cables [24], with applications to cable-stayed bridges and stadium roofs, among others. Natural frequencies, in particular, can be used in the estimation of cable tension, provided that its length, mass and elasticity are known [25, 26, 27]. Photogrammetry and other optical methods involving multiple viewpoints have been employed as well for structural dynamics measurement [28], especially with Digital Image Correlation (DIC). Such techniques can achieve very high precision 3D measurements and support stitching of multiple views together [29], but they often rely on the presence of a custom-painted speckled pattern on the surface of the target structure [30].

Drones too have been studied and developed as tools for SHM. Recent projects deploy them as platforms for digitalization and Building Information Modeling [31], and even mobile contact sensors for non-destructive testing [32, 33]. Some technologies have been developed for the contact-based inspection of reinforced concrete [33]. In this scenario, a UAV is fitted with electrical resistance sensors, in order to detect cracks and corrosion in steel under the surface. Other applications include inspection of oil and gas pipes [34]. The main application of UAVs, though, is in visual inspection, where they excel due to their freedom of movement along the three spacial dimensions [35], the ability to hover in place and the variety of cameras they can host. When equipped with thermal cameras, they can be used for thermography, offering for example extremely quick surveys of the status of solar panel fields [36]. Regular RGB cameras are exploited for photogrammetry [37] and acquiring photographs that are subsequently processed through Machine Learning (ML) techniques [38], aiming at automatically detecting surface cracks, concrete spalling, rust, humidity and other surface-level issues.

Several video-based techniques for SHM using regular cameras have been proposed,

and the issue of multiple camera synchronization is well known in literature. The fixed nature of regular cameras makes it possible to use wired connections, thus frame synchronization is achieved with trigger signals sent to all sensors at the same time. This is possible through simple configurations where all cameras are connected to a central clock [39] and also in more complex networks relying on the Ethernet standard [40]. Regardless of the technology, the error is as low as a few microseconds. Such a low error enables approaches like DIC for vibration measurement [41]. Unfortunately, due to their reliance on a physical connection between the cameras, these solutions cannot be applied to drone-mounted sensors.

Still, some drone-mounted video-based techniques have been proposed to monitor the condition of buildings. The authors of [42] deployed a commercially available drone to measure the displacement of a structure, both in laboratory and on a real railway bridge. They performed natural feature tracking to estimate the displacement of the targets, and compensated egomotion by estimating it with respect to background features. Results were compared to measurements taken with a fixed camera + LED system and found to be of good quality, though some simplifying hypotheses were introduced in the design phase of the experiment. Furthermore, the authors noted that their technique of egomotion compensation is not robust. The authors of [43] aimed at estimating modes of vibration of a scale model of a wind turbine, obtaining promising results through DIC. The same technique was exploited in [44] to estimate two-dimensional strain. In [45], a single commercially available drone was used to measure the oscillating motion of both a six-story building model and a real pedestrian bridge. Displacement was measured through marker detection and KLT (Kanade-Lucas-Tomasi) tracking across video frames. Measurement quality was found to be remarkable even in the more realistic case on the pedestrian bridge, whose motion was forced by people jumping on the deck. Some limitations of the approach are mentioned, namely: the low number of acquisitions that can be made with a single battery, the issue of detecting natural frequencies within the bandwidth of the motion of the drone, the issue of positional stability with respect to the target in GPS-denied environments, and the issue of applying and maintaining a significant forcing input to a larger, more rigid structure. At least for the first modes of vibration, the comparatively low Nyquist frequency introduced by cameras with respect to accelerometers is usually not an issue. In fact commonly available cameras easily reach at least 30 frames per second (FPS), whereas the first natural frequencies appear in the $0 - 10$ Hz bandwidth.

1.2. Scope

In the state of the art, we note the absence of drone-enabled vision-based vibration modal analysis techniques that can capture the response of large spans of structure at once with high spacial accuracy: [42] achieves high accuracy but focuses on a single area of the building, whereas [45] calculates the whole modal shape of a pedestrian bridge, but it does so by sequentially repeating the displacement measurement with a single drone-mounted camera on several sections while an external force is continuously applied to the deck. This trade-off between accuracy, acquisition time and completeness is due to the fact that when a single aircraft is deployed, it must hover rather close to the structure to measure its displacement with sufficient accuracy through the video analysis, thus losing the possibility of capturing the whole building at once. By using multiple

drones, instead, it is possible to perform in parallel accurate measurements of larger stretches at the same time, thus also overcoming time limitations due to battery charge and removing the need to maintain a continuous external forcing input for a prolonged amount of time. Furthermore, to the best of the authors' knowledge no solutions exist in literature to the problem of drone-mounted camera synchronization. As previously noted, the non-stationary nature of drones makes it impractical to physically connect them (even though some research in this direction is being carried out, see [46, 47]), therefore a fundamentally different solution must be found.

In this context, we present an experimental feasibility study of a novel system to perform video-based modal analysis through multiple cameras mounted on collaborative drones, that measure natural frequencies and estimate modal shapes. Our approach exploits the idea of overlapping the target areas measured by two different cameras to perform *a posteriori* signal synchronization. With respect to [45], we show experimentally that markers on the structure are not necessary to measure displacement with sufficient accuracy to extract modal shapes.

Using multiple UAVs at the same time can provide a more complete view of the response of a structure to an arbitrary input excitation with limited error propagation, which is hardly possible when using a single device, addressing the mentioned research gap. Deploying several drones at a time can also help to reconstruct the 3D motion of the structure by capturing it from several different points of view, thus overcoming another limitation of a single camera, *i.e.* the reduced accuracy at which displacement is measured along the direction of the camera's axis. In this regard, using several cameras with different orientations with respect to the structure can effectively perform the same type of measurement an accelerometer network offers.

We show promising preliminary results in a test case where two drones are employed. Using commercially available UAVs similar to those we adopted, this procedure can already be applied to small-size structures without incurring in complexity scaling issues. For example, with three aircraft hovering 2.5 m away from the target surface and recording footage with 1920×1080 resolution, it is already possible to capture the oscillations of structures less than 10 m long, such as barriers, antennas, parapets, small metallic pedestrian bridges or portions of larger structures.

2. Problem Description

The problem at hand is to obtain the modal parameters of a structure, namely natural frequencies ω_i and modal shapes $\mathbf{\Phi}_i$, where $i = 1, ..., n$ is an index that identifies the different modes, up to the n-th. This is to be done through vision only, without relying on contact. Our main objectives are three:

- Identifying natural frequencies and modal shapes through displacements measured using cameras mounted on drones;
- Establishing if our method is sufficiently accurate to detect changes in modal parameters due to damage of the structure;
- Testing the viability of multiple drones to obtain and fuse information collected across the whole structure.

3. Proposed Method

The procedure for estimating the modal parameters of the structure is summarized in these steps:

1. Recording the behaviour of the structure and identifying features in the collected frames;
2. Measuring the structure's displacement by tracking the identified features across frames;
3. Estimating natural frequencies from the measured displacement;
4. Estimating the modal shape for each identified natural frequency.

We adopt the classic pinhole camera model to describe our visual sensors, hence we can write the following relationship between the 3D position of objects in the real world $\mathbf{P}$ with respect to a global reference frame and their positions in the image plane $\mathbf{p}$:

$$s\mathbf{p} = K\,[R|v]\,\mathbf{P}. \tag{1}$$

Here s is the scale factor, R and v are the rotation matrix and translation vector that operate the projection onto the 2D plane and K represents the intrinsic parameters of the camera. The lens distortion is corrected after camera calibration.

3.1. Recording and Feature Detection

In this study, the structure is excited with a kinetic hammer and its oscillation is recorded by the cameras. In order to track the displacement of an object, *i.e.* the movement of the targets across pixels in consecutive frames, we first apply the Harris corner detection algorithm [48] to find features within the gray-scale version of the first frame. Some of those features are manually selected for tracking: automation of this step will be the object of further study. This procedure is based on an estimation of the bi-dimensional gradient of the color in each frame and can reliably distinguish homogeneous areas, edges, and corners. Such an algorithm is compatible with markers featuring many corners such as ArUco [49], like the ones in Figure 1, but also works without them. We initially chose to exploit the markers to ease the feature selection procedure, maximize robustness of the measurements and streamline the presentation of the algorithm. Nevertheless, we also repeated the measurement by tracking different features, belonging to the surface of the structure. We show in section 5 that markers are not strictly necessary to track structural displacement. In fact, tracking naturally occurring features on the surface of the target yields equally accurate results, as visible in Figure 2.

3.2. Displacement measurement

The second step of motion tracking is estimating how the selected features move across frames. To make the measurement robust, the chosen feature is the mean position of the four corners. Tracking is performed through the KLT algorithm [16], which solves an optimization problem to estimate the displacement d of single pixels in consecutive pictures as optical flow. Optical flow is defined as the pattern of motion that objects perform in a scene caused by the relative movement between the camera

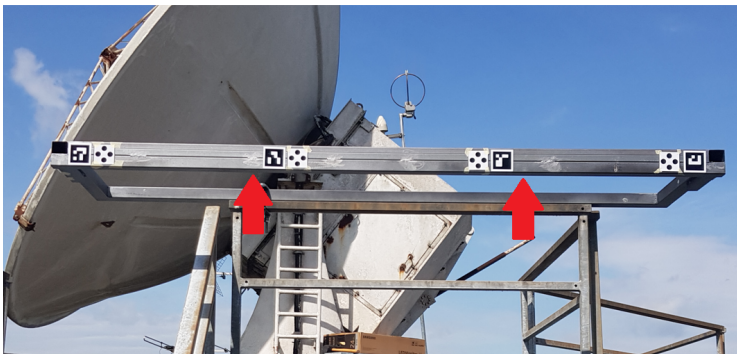

Fig. 1. The structure used for vibration tests. Spot markers are present alongside ArUco ones, but they were not exploited in these procedures. The arrows indicate where the accelerometers (not visible in this picture) were placed.

and the scene itself. This method works reliably under the hypothesis that out-of-plane motion is negligible with respect to the mean distance between the camera and the target. Furthermore, if it is necessary to estimate the vibration amplitude in addition to the frequencies, the optical axis of the camera must also be orthogonal to the plane of motion of the target. This introduces limitations both on the amplitude and speed of the measurable motion with respect to the frame rate, which are not an issue in our specific case, where the amplitude is limited to a few pixels. Finally, a scale factor must be computed to express the amplitude of motion in physical units. It is necessary to rely on a known physical distance between two features to establish a relationship between lengths in pixels and in millimeters. To this end we exploit the corners of the markers, knowing their dimensions. For a fixed camera, this scale factor can be computed on the first frame and used on all of them, but the same is not true for flying cameras. Due to their non-null movement in the out-of-plane direction, the scale factor may change across time. Nevertheless the distance between the features is fixed, hence the scale factor can be computed at every frame as:

$$s(t) = \frac{|\boldsymbol{P}_1 - \boldsymbol{P}_2|}{|\boldsymbol{p}_1(t) - \boldsymbol{p}_2(t)|}, \tag{2}$$

where t is the discrete-time index, $|\boldsymbol{P}_1 - \boldsymbol{P}_2|$ expresses the physical distance between two features, which is constant across time, and $|\boldsymbol{p}_1(t) - \boldsymbol{p}_2(t)|$ describes the same distance in pixels, in the image plane. The scale factor not only converts the camera plane displacement, measured in pixels, into an absolute displacement, measured for example in millimeters; it is also exploited to compensate the out of plane motion of the camera, since the physical distance between the features is constant across time. In the absence of markers, there are at least two ways to solve the scale problem:

- exploiting the known dimensions of some component of the structure, such as the size of a joint or bolt. This requires some previous knowledge of the structure or additional measurement to be carried out;
- calculating the physical distance between two features appearing in the image

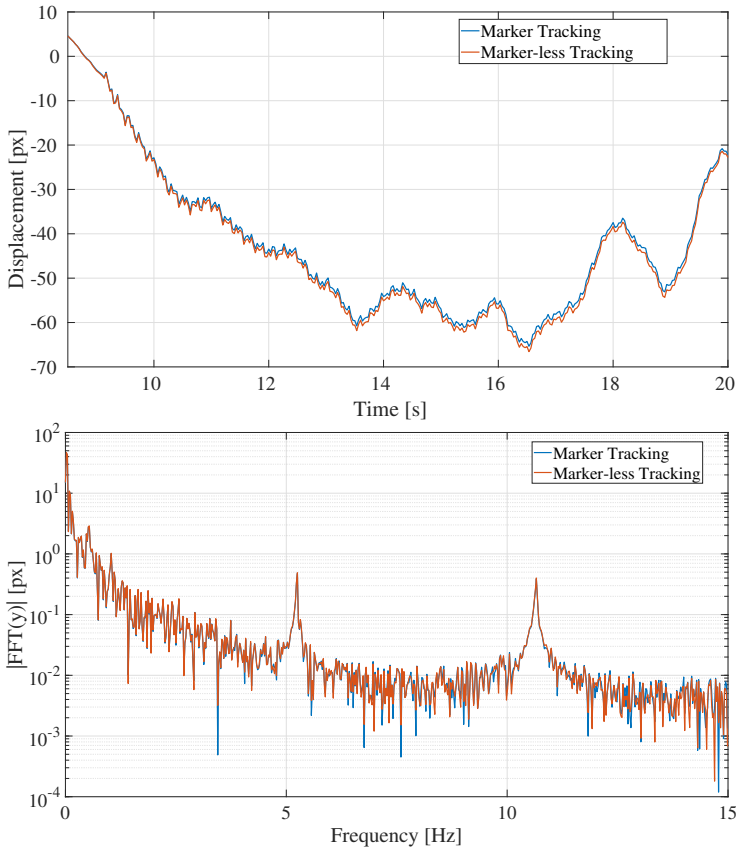

Fig. 2. Non-filtered time history and frequency spectrum of the averaged vertical displacement of features on a marker and naturally occurring on the structure itself, such as a junction. The plots overlap to the point of being almost indistinguishable. The time history is dominated by the relative motion between drone and structure, but the frequency components are well separated.

plane with trigonometry. This requires measuring the distance between the camera and the surface in real time during the recording, which is possible through LiDAR or ultrasonic sensors, for example. Such sensors are often present on UAVs for collision avoidance.

3.3. Estimating Natural Frequencies

We apply a Fast Fourier Transform (FFT) to obtain the spectrum of frequency components of the signal. In the case of drone-mounted cameras, noise originated by the movement of the aircraft and the action of the motors must be accounted for. In commercially available drones, ours included, cameras are mounted on active gimbals, capable of compensating some of these disturbances. Furthermore, we verify experimentally that the noise does not contain frequency components that could corrupt the signal (see Figure 3). The effect of the rotating propellers is not visible because of its

high frequency: according to the manufacturer's claims, the motor rotation frequency of small drones while hovering is higher than 80 Hz. On the other hand, the drone motion relative to the ground presents frequency components that are below the natural ones of the structure. As a consequence, we can apply a high-pass filter to separate the natural frequencies of the structure from the frequency components introduced by the motion of the camera. A reasonable choice in our case is to select a cutoff frequency of 2 Hz, as it is lower than the first natural frequency of the structure at hand, but higher than the components related to the camera movement. We exploit here a fundamental assumption, which will be recalled multiple times throughout the paper, and is verified for the considered structure (see section 4.2): the natural frequencies of the target do not overlap with the main spectral components of camera displacement. If this were not the case, the relative displacement can still be compensated, for example by accurately measuring the motion of the camera with respect to the still background and subtracting it from the one of the structure relative to the camera [42]. Figure 4 exemplifies the effect of filtering on the time history of relative displacement between the structure and the camera: the spectral components of camera displacement are canceled and the resulting signal is a time-history of the structure's vibration alone. Furthermore, from the obtained FFT, the natural frequencies ω_i of the target structure can be measured. Figure 3 displays the spectrum of the signal that is obtained if the structure is not excited, further confirming that the peaks visible in tests with excitation represent the behavior of the structure only.

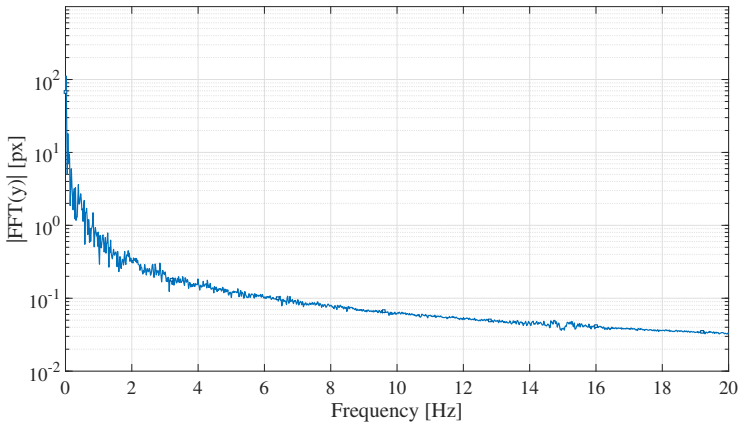

Fig. 3. Spectrum of the measurement of a fixed feature, representing the frequency components of measurement noise only. By comparison with Figure 4, we conclude that the peaks at 5.06 Hz and 11.16 Hz represent the oscillating behavior of the structure in response to excitation. This is also consistent with an estimated rotational speed of the propellers during hovering of 5000 revolutions per minute, equivalent to more than 80 Hz.

3.4. Estimating Modal Shapes

In order to obtain the modal shapes, we need to know the pair-wise phase delays between features across the length of the structure, therefore, we take the signals two at a time and estimate the transfer function between them. Evaluating the phase of such

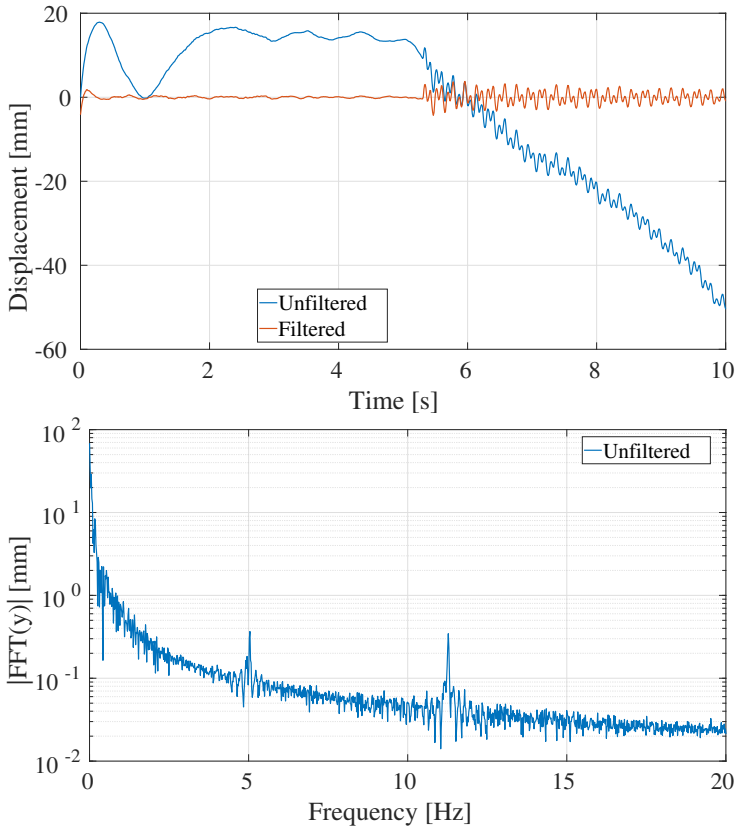

Fig. 4. Vertical displacement (top) of a feature as measured by a flying camera, in the first ten seconds of a test. The hammer impacts around the 5.5 s mark. Notice how the unfiltered measurement is dominated by the relative movement of the camera. In the frequency domain, these effects are distinguishable, as conveyed by the FFT (bottom). The first two natural frequencies of the structure (5.06 Hz and 11.16 Hz) are clearly visible.

transfer function at the natural frequencies yields the phase delay between each feature pair:

$$\Delta\phi_{i,a-b} = \arg\left(G_{a-b}(j\omega_i)\right), \tag{3}$$

where j is the imaginary unit, $G_{a-b}(j\omega_i)$ is the value of the continuous-time transfer function between the signals describing the displacement of features a and b, evaluated at point $j\omega_i$.

To calculate pair-wise phase delays between features appearing in different video sources, we first synchronize them. This is achieved through cross-correlation of the high-pass filtered displacement measurements of a feature or marker appearing in both videos, in the overlapping region. This step is necessary for data collected with commercially available drones, for which hardware synchronization of the cameras is not possible.

We then estimate the transfer function between signals extracted from the two different but synchronized recordings. To better describe such a method, let us consider two cameras, without loss of generality. In a generic scenario, not only the initial time instant of the two recordings might be different, but also their frame rate. If however at least one feature is recorded by both cameras, we can artificially over-sample the two time-histories recorded at the same feature to the least common multiple of their frame rates, then apply a high-pass filter to isolate the oscillation of the structure from the relative motion of the cameras. Then, the cross-correlation between the two processed signals is calculated to estimate their relative delay in time, indicated with $\hat{\tau}$. We expect this delay to be always well defined, because the two signals correspond to the displacement of the same feature. Finally, a correction is applied to one of the two original signals to compensate the estimated time delay $\hat{\tau}$, to synchronize the two video sources. Then, the relative displacements among all features, including those not appearing in the overlapping regions of the two cameras, can be estimated. Thus, this method makes it possible to directly estimate phase delays also between features that are recorded by different cameras. Figure 5 summarizes the proposed procedure.

4. Experimental set-up

To determine the accuracy of the estimates, at least for portions of structures of limited size, we compare those obtained from flying cameras with those taken with accelerometers and with a fixed, high frame-rate camera. In order to test the capability of the proposed approach to provide information about possible alterations of the structure or faults, we then repeat and compare the measurements in different conditions. First, we identify the modal parameters in the reference situation, corresponding to an an undamaged state. Then we identify the same parameters in modified scenarios obtained by gradually adding more weights to the structure. Finally we simulate a faulty scenario, by restoring the initial weight and loosening one of the fixings.

4.1. Instrumentation

The accelerometers are PCB 333B30 monoaxial sensors with full scale range of ±50g, acquired with NI 9234 board at 2048 Hz. The bandwidth of these sensors and the acquisition system is $1 - 700$ Hz, while the range of interest for this test is $1 - 100$ Hz. Motivated by their high accuracy, we take accelerometers acquisitions as ground truth to evaluate the results obtained with the cameras. Accelerometers were placed on top of the structure to measure the vertical displacement, as indicated by the red arrows in Figure 1.

The fixed camera is a FLIR Blackfly S BFS-U3-16S2M, a monochrome camera, with a resolution of 1440x1080 px and the framerate used for this experiment is 200 FPS. The camera was placed on a tripod at a distance of about 5 m to capture the entire structure. The focal length is 12 mm, to have a Field Of View (FOV) large enough for the entire structure.

The first UAV is a DJI Mavic Pro equipped with a 12 MP camera, capable of shooting full HD video (1920×1080) at 96 FPS and HD video (1280×720) at 120 FPS. The second is a DJI Mavic 2 Enterprise Dual, whose camera captures 4K Ultra HD

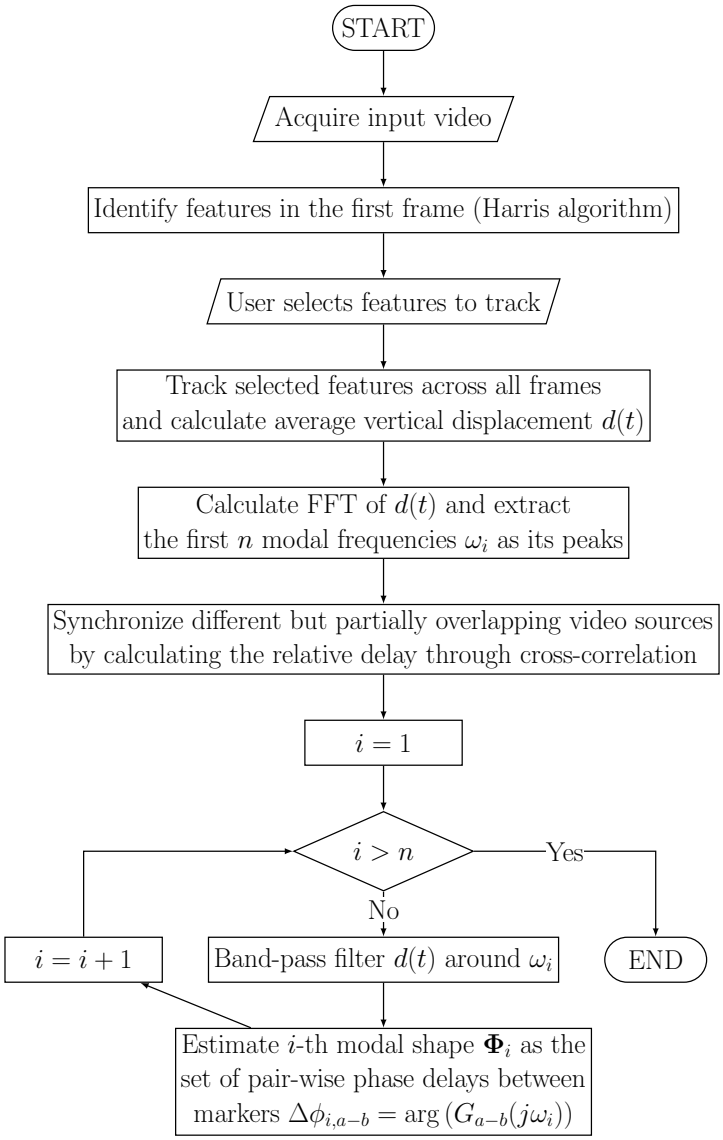

Fig. 5. Layout of the proposed methodology.

footage (3840×2160) at 30 FPS. Both drones are commercial, and they are equipped with a stabilizing gimbal for the camera. This device is controlled by the on board computer and actively compensates part of the disturbance induced by the aircraft motion, by decoupling its attitude with respect to that of the camera.

The resolutions of the cameras are different, which influences the minimum structural displacement they can record. To ensure that every camera can capture the

oscillation of the target in every test, they are at a suitable distance from the structure. The resulting scale factors, for all cameras and all tests, were within the $0.5-1.8\ 10^{-3}\ \frac{m}{px}$ (millimeters per pixel) range, which is enough to capture oscillations with amplitude of a few millimeters (see Figure 4). By the Nyquist-Shannon sampling theorem, frame-rates also introduce an upper limit of $\frac{FPS}{2}$ on the measurable frequencies. In our case, the lowest frame-rate is 30 FPS, corresponding to the Mavic 2 drone, therefore we limit our analysis to natural frequencies below 15 Hz, which pertain to the first two modes.

4.2. Test structure and procedure

The structure for our experiments has two main components: a railing, welded on a metal base fixed to the ground, and a cantilever section on top (see Figure 1). The two parts are made of steel and attached with two clamps. Both are made of hollow tubes with square cross-section. The cantilever was designed to have suitable natural frequencies. In particular, lower bounds on the frequency values were introduced by the frequency content of drone movement, in the $0-1$ Hz range. The Nyquist rates of the two flying cameras, on the other hand, posed an upper bound. As a consequence, the cantilever was designed to have frequencies within the $3-12$ Hz bandwidth. The cantilever was welded into a rectangular, 2000×550 mm shape with a mass of 20.5 kg. A finite element model of the structure was developed to simulate its behavior and determine the values of the natural frequencies for several different mass distributions.

Four square, 55×55 mm markers are attached on the cantilever, two near the corners and two near the center, equally spaced along the structure (see Figure 1). All the tests are performed with the same procedure: an impulse is generated with a rubber hammer on the top to excite all frequencies. The impact point is at the center of the protruding end of the cantilever, and the hammer impacts the structure vertically. The fixed camera was placed 5 m away from the structure, facing it directly, whereas drones were hovered at approximately 1 to 2 m from the cantilever, to compensate for their wider field of view.

A total of 13 tests were carried out in the span of two hours, with consistent lighting conditions. The duration of each test is on average of 30 s, which is enough to include the whole response of the structure, up to the moment when it stops vibrating. Consistency was kept by keeping the static camera and the accelerometers fixed between tests and by manually piloting the drones during the tests so as to keep them as stationary as possible.

5. Experimental Results

5.1. Measuring Modal Parameters

The first objective of the procedure is to accurately measure the natural frequencies of the structure, and to evaluate the accuracy obtained with the drones as compared with the accelerometers, which we take as ground truth. Hence, a test was designed where the accelerometers were fixed onto the structure as shown in Figure 1, the fixed camera was placed in front of the target at a distance of 5 m and the DJI Mavic Pro was flown at a 1.5 m distance, hovering in place. The drone was positioned slightly above the structure in order not to obstruct the visual of the fixed camera. The comparison of the

Mode	Accel.	Fixed Camera (err.)	Flying Camera (err.)
$i = 1$	5.06 Hz	5.01 Hz (1.0%)	5.05 Hz (0.2%)
$i = 2$	11.16 Hz	11.14 Hz (0.2%)	11.17 Hz (0.1%)

Table 1. Estimated natural frequencies

Mode	ω_i	$\Delta\phi_{i,1-2}$	$\Delta\phi_{i,1-3}$	$\Delta\phi_{i,1-4}$
$i = 1$	5.05 Hz	$-1.10°$	$0.61°$	$-0.24°$
$i = 2$	11.17 Hz	$0.75°$	$-0.22°$	$0.56°$

Table 2. Phase delays estimated with a single camera

FFTs of the accelerometers, fixed camera and flying camera in the relevant bandwidth is reported in Figures 6 (fixed and flying cameras) and 7 (accelerometers) and in Table 1. Error is calculated as relative difference with respect to the accelerometer value:

$$e_i = 100\frac{\left|\omega_{i,accelerometer} - \omega_{i,camera}\right|}{\omega_{i,accelerometer}}. \tag{4}$$

The two measured modes are within the bounds specified in the design phase; in particular, their values are sufficiently close to those obtained in finite-element simulations: 4.88 Hz and 10.66 Hz. Furthermore, the error with respect to the accelerometers is at most 0.2% for the flying camera, and slightly higher for the fixed one, notwithstanding its higher frame rate and the absence of motion.

It is possible to extract modal shapes from the data, comparing the movement of the four different markers and their relative pairwise phase delays $\Delta\phi_{i,a-b}$. The phase delay provided by the two accelerometers placed on the protruding end of the cantilever is approximately zero for both investigated frequencies (see Figure 7).

The pair-wise transfer functions and phase delays between the four measurements recorded by the DJI Mavic Pro were estimated as described in Section 3. Table 2 details the calculated delays, where each column represents the estimated phase delay between a marker and the next one. The phase delays are all very close to zero, which indicates, in accordance with the information provided by the accelerometers, that both modes correspond to a vertical rigid displacement of the front side of the cantilever. This behavior is again compatible with the one observed in the finite-element simulations. The two are likely due to the railing oscillating with respect to the ground and the cantilever vibrating with respect to the railing itself. The cantilever does not exhibit bending modes at such low frequencies.

We verified experimentally that measurements taken by tracking naturally occurring features on the surface of the structure are as accurate as those obtained by exploiting markers. In order to do so, we selected as features the edges of the soldering points (see Figure 8 for an example). These results are summarized in Figure 2, which shows that the measured displacement and its frequency spectrum are nearly identical in the two cases.

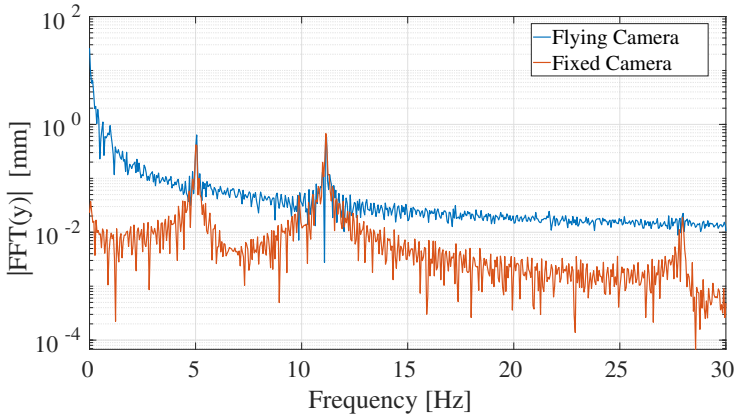

Fig. 6. FFT of the displacement measurement, comparing flying camera and fixed camera. The main difference between the sensors is not the accuracy of the frequency peaks, but the baseline noise. This is the reason why the third natural frequency is not identifiable with the flying camera.

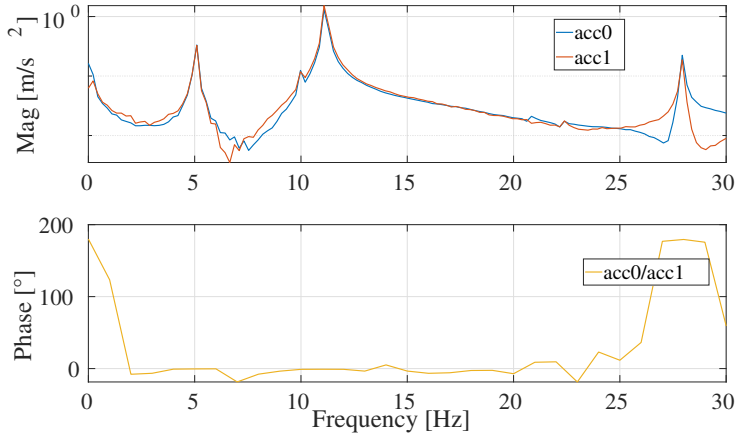

Fig. 7. The magnitude of the FFT of the accelerometers placed to measure the vertical displacement (top) and their relative phase delay (bottom), obtained from the transfer function estimated with Welch's averaged periodogram method.

5.2. Synchronization of Two Drones

We next investigate the acquisition of the use of data collected by two drones, each one with a partial view of the whole structure and with an overlapping region with respect to the other aerial vehicle. For this reason a second test was carried out, where only the two UAVs described in Section 4.1 were used to record the displacement. The structure was still excited with a rubber hammer and the drones were hovering in front of the structure at a 1 m distance (see Figure 9). The first aircraft recorded at 96 FPS, the second at 30 FPS. After calculating and correcting the delay between the two videos, through oversampling and cross-correlation of the displacement of the overlapping marker (as described in Section 3), the pairwise relative phase delays between markers

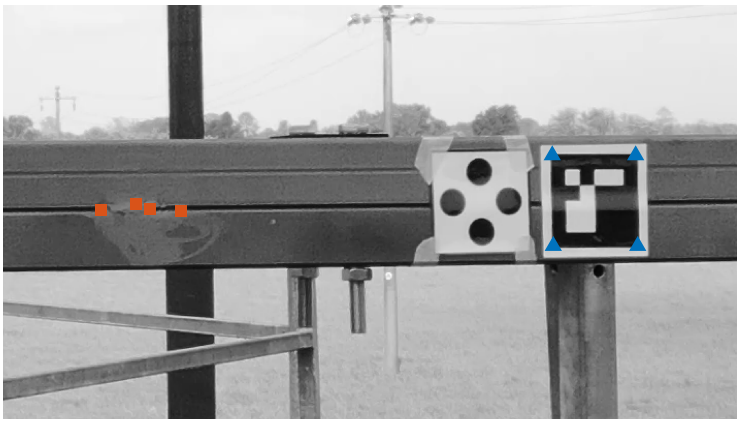

Fig. 8. Comparison between using markers and features that are inherent in the structure.. Triangles are the corners of the marker, while squares belong to a soldering spot on the structure.

Fig. 9. Setup of the second test. Drones are positioned such that each one is unable to capture the whole structure.

can be computed. Synchronized displacements were treated as if they were recorded by the same camera, and yielded the results shown in Table 3. The error is calculated as the absolute value of the difference with respect to the phase delays estimated with a single camera. Phase delays between non-consecutive markers can be computed as

$$\Delta\phi_{i,a-c} = \Delta\phi_{i,a-b} + \Delta\phi_{i,b-c}, \tag{5}$$

and they are comparable to the ones obtained with a single flying camera.

5.3. Detection of Overload and Tampering

In order to ascertain the ability of this procedure to detect variations in modal parameters, a third series of tests was carried out. The same excitation technique was adopted,

ω_i	$\Delta\phi_{i,1-2}$ (err.)	$\Delta\phi_{i,1-3}$ (err.)	$\Delta\phi_{i,1-4}$ (err.)
5.05 Hz	$-1.17°$ ($0.07°$)	$-1.87°$ ($2.48°$)	$-2.73°$ ($2.49°$)
11.17 Hz	$-3.02°$ ($3.77°$)	$-2.21°$ ($1.99°$)	$-2.23°$ ($2.79°$)

Table 3. Phase delays estimated with two drone-mounted cameras and synchronization

while recording and measuring with a single drone, as in the first scenario. Three tests were carried out after gradually adding masses to the protruding end of the cantilever. In particular, we added three 4.0 kg masses, one at a time, and repeated the measurement. While such a significant modification of the modal parameters may be unrealistic, we show our procedure still correctly identifies the change and the measurement of the parameters is commensurate with that obtained with accelerometers.

Another test was conducted after removing the additional masses and loosening one of the clamps that held together the structure. Results were compared to the base case to find out if the measured modal parameters differ significantly, and they are collected in Table 4. Figure 10 shows the effect of gradually adding mass to the structure. As

Mode	ω_i	$\Delta\phi_{i,1-2}$	$\Delta\phi_{i,2-3}$	$\Delta\phi_{i,3-4}$
$i = 1$	4.96 Hz	$3.51°$	$0.50°$	$-2.75°$
$i = 2$	8.19 Hz	$4.32°$	$0.81°$	$-4.95°$

Table 4. Estimated phase delays after tampering

expected, the natural frequencies decrease, since they are, in general, proportional to the square root of the ratio between the modal stiffness and mass. It should also be noted that while the two frequencies decrease, they maintain the same ratio. No significant differences emerge in the modal shape analysis: the main oscillations are still related to vertical rigid displacement of the protruding end of the cantilever and no bending takes place.

The effect of a tampered clamp, on the other hand, is noticeably different (see Figure 10). Here the natural frequencies are still impacted, but the second one drops to 8.19 Hz, while the first is 4.96 Hz, thus the ratio is not maintained. This decrease in frequency is likely due to an overall reduction of the stiffness of the structure. The modal shape is still the same, though some phase delays increase with respect to the base scenario, indicating that bending of the cantilever beam is probably starting to happen due to the loosened clamp, with a modal shape that is not symmetric anymore.

6. Conclusion, Limitations and Future Developments

We described and tested a method to conduct vision-based modal analysis of small structures or portions of structures through the use of measurements acquired simultaneously by multiple airborne cameras. We measured the natural frequencies and calculated the modes of vibration of a metal structure and showed that accurate displacement measurements can be carried out based on features occurring on the surface

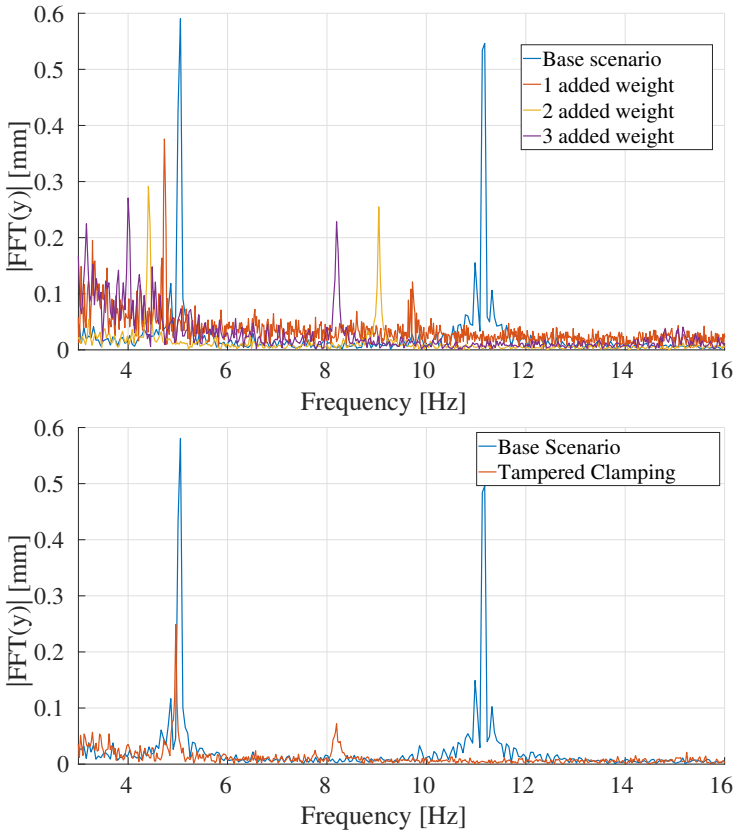

Fig. 10. Effect of increasing mass (top) and of loosening a clamp (bottom) on natural frequencies.

of the target structure, without markers. We also tested the ability of such method to detect variations in modal parameters. The results are encouraging, despite the limited bandwidth due to rather low FPS cameras, with respect to accelerometers. We deem this solution feasible for structures of limited size and convenient when accelerometers are not already present and their installation is difficult due to the structure being precarious or hard to reach. Even though the method, in its current state, is not yet suited to large structures which would need a large number of UAVs making the solution complex, we think that it can be developed further to extend both its range of applicability and the accuracy of the estimates. For example, the range of frequencies that it is possible to study can be extended towards the lower end of the spectrum, where camera movement noise is significant, by devising a more complex way of filtering ego-motion, such as measuring camera displacement with respect to a still background or estimating it based on data collected from the drone's Inertial Measurement Units (IMUs). Furthermore, higher autonomy can be reached both in feature selection, through specifically designed algorithms, and in aircraft deployment, by employing autonomous UAVs. The technology for achieving this is already available while the regulatory framework

is not sufficiently developed yet. A technical limit to further scaling is the relationship between camera resolution, field of view and oscillation amplitude, which however is favorable for slender structures and cables. Motion blur due to the camera hovering can in principle impact measurement accuracy, but we verified that its contribution in our tests is negligible, as displacements with amplitude below 1 px are identified. Finally, in case the surface of the target structure does not offer a sufficient number of features, it is possible to spray-paint suitable patterns from a distance, also exploiting drones [50]. As long as environmental conditions are favorable (limited wind, suitable lighting and absence of reflections), the procedure can be applied as it stands to some real world tasks, including modal analysis of small, flexible structures such as barriers, antennas, parapets and small metallic bridges. This method can be suitable also for vision-based estimation of tension force in cables, such as those in stadiums, cable-ways, power transmission systems and cable-stayed bridges. There is ongoing research within our group in this field with positive results.

7. Data Availability Statement

All data and code that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] M. Kulbacki, J. Segen, W. Knieć, R. Klempous, K. Kluwak, J. Nikodem, J. Kulbacka, A. Serester, Survey of drones for agriculture automation from planting to harvest, in: 2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES), IEEE, 2018, pp. 000353–000358. doi:10.1109/INES.2018.8523943.

[2] S. K. Chaturvedi, R. Sekhar, S. Banerjee, H. Kamal, Comparative review study of military and civilian unmanned aerial vehicles (uavs), Bulletin of the National Institute for Aerospace Research" Elie Carafoli" 11 (3) (2019) 183–198. doi:10.13111/2066-8201.2019.11.3.16.

[3] M. J. Sousa, A. Moutinho, M. Almeida, Thermal infrared sensing for near real-time data-driven fire detection and monitoring systems, Sensors 20 (23) (2020) 6803. doi:10.3390/s20236803.

[4] Y. Jiang, Y. Bai, S. Han, Determining ground elevations covered by vegetation on construction sites using drone-based orthoimage and convolutional neural network, Journal of Computing in Civil Engineering 34 (6) (2020) 04020049. doi:10.1061/(ASCE)CP.1943-5487.0000930.

[5] R. Ali, D. Kang, G. Suh, Y.-J. Cha, Real-time multiple damage mapping using autonomous uav and deep faster region-based neural networks for gps-denied structures, Automation in Construction 130 (2021) 103831. doi:10.1016/j.autcon.2021.103831.

[6] Y. Tan, S. Li, H. Liu, P. Chen, Z. Zhou, Automatic inspection data collection of building surface based on bim and uav, Automation in Construction 131 (2021) 103881. doi:10.1016/j.autcon.2021.103881.

[7] M. Rossi, D. Brunelli, A. Adami, L. Lorenzelli, F. Menna, F. Remondino, Gas-drone: Portable gas sensing system on uavs for gas leakage localization, in: SENSORS, 2014 IEEE, IEEE, 2014, pp. 1431–1434. doi:10.1109/ICSENS.2014.6985282.

[8] I. T. Forum, Transport infrastructure investment and maintenance, accessed: 2022-06-17. doi:10.1787/g2g55573-en.

[9] A. Sabato, C. Niezrecki, G. Fortino, Wireless mems-based accelerometer sensor boards for structural vibration monitoring: a review, IEEE Sensors Journal 17 (2) (2016) 226–235. doi:10.1109/JSEN.2016.2630008.

[10] Y. Xu, J. M. Brownjohn, F. Huseynov, Accurate deformation monitoring on bridge structures using a cost-effective sensing system combined with a camera and accelerometers: Case study, Journal of Bridge Engineering 24 (1) (2019) 05018014. doi:10.1061/(ASCE)BE.1943-5592.0001330.

[11] M. A. Sutton, J. H. Yan, V. Tiwari, H. Schreier, J.-J. Orteu, The effect of out-of-plane motion on 2d and 3d digital image correlation measurements, Optics and Lasers in Engineering 46 (10) (2008) 746–757. doi:10.1016/j.optlaseng.2008.05.005.

[12] L. M. Khoo, P. R. Mantena, P. Jadhav, Structural damage assessment using vibration modal analysis, Structural Health Monitoring 3 (2) (2004) 177–194. doi:10.1177/1475921704042680.

[13] Z.-F. Fu, J. He, Modal analysis, Elsevier, 2001, ISBN 978-0750650793.

[14] C.-Z. Dong, F. N. Catbas, A review of computer vision-based structural health monitoring at local and global levels, Structural Health Monitoring 20 (2) (2021) 692–743. doi:10.1177/1475921720935585.

[15] C.-Z. Dong, O. Celik, F. N. Catbas, Marker-free monitoring of the grandstand structures and modal identification using computer vision methods, Structural Health Monitoring 18 (5-6) (2019) 1491–1509. doi:10.1177/1475921718806895.

[16] C. Tomasi, T. Kanade, Detection and tracking of point, International Journal of Computer Vision 9 (1991) 137–154.

[17] D. Feng, M. Q. Feng, Experimental validation of cost-effective vision-based structural health monitoring, Mechanical Systems and Signal Processing 88 (2017) 199–211. doi:10.1016/j.ymssp.2016.11.021.

[18] D. Feng, M. Q. Feng, Vision-based multipoint displacement measurement for structural health monitoring, Structural Control and Health Monitoring 23 (5) (2016) 876–890. doi:10.1002/stc.1819.

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

[20] Y. Yang, C. Dorn, T. Mancini, Z. Talken, G. Kenyon, C. Farrar, D. Mascareñas, Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification, Mechanical Systems and Signal Processing 85 (2017) 567–590. doi:10.1016/j.ymssp.2016.08.041.

[21] V. Fioriti, I. Roselli, A. Tatì, R. Romano, G. De Canio, Motion magnification analysis for structural monitoring of ancient constructions, Measurement 129 (2018) 375–380. doi:10.1016/j.measurement.2018.07.055.

[22] J. G. Chen, T. M. Adams, H. Sun, E. S. Bell, O. Büyüköztürk, Camera-based vibration measurement of the world war i memorial bridge in portsmouth, new hampshire, Journal of Structural Engineering 144 (11) (2018) 04018207. doi:10.1061/(ASCE)ST.1943-541X.0002203.

[23] J. J. Lin, A. Ibrahim, S. Sarwade, M. Golparvar-Fard, Bridge inspection with aerial robots: Automating the entire pipeline of visual data capture, 3d mapping, defect detection, analysis, and reporting, Journal of Computing in Civil Engineering 35 (2) (2021) 04020064. doi:10.1061/(ASCE)CP.1943-5487.0000954.

[24] Y. Ji, C. Chang, Nontarget stereo vision technique for spatiotemporal response measurement of line-like structures, Journal of engineering mechanics 134 (6) (2008) 466–474. doi:10.1061/(ASCE)0733-9399(2008)134:6(466).

[25] S.-W. Kim, B.-G. Jeon, N.-S. Kim, J.-C. Park, Vision-based monitoring system for evaluating cable tensile forces on a cable-stayed bridge, Structural Health Monitoring 12 (5-6) (2013) 440–456. doi:10.1177/1475921713500513.

[26] X. Zhao, K. Ri, N. Wang, Experimental verification for cable force estimation using handheld shooting of smartphones, Journal of Sensors 2017 (2017). doi:10.1155/2017/5625396.

[27] D. Feng, T. Scarangello, M. Q. Feng, Q. Ye, Cable tension force estimate using novel noncontact vision-based sensor, Measurement 99 (2017) 44–52. doi:10.1016/j.measurement.2016.12.020.

[28] J. Baqersad, P. Poozesh, C. Niezrecki, P. Avitabile, Photogrammetry and optical methods in structural dynamics – a review, Mechanical Systems and Signal Processing 86 (2017) 17–34, full-field, non-contact vibration measurement methods: comparisons and applications. doi:https://doi.org/10.1016/j.ymssp.2016.02.011.

[29] F. Chen, X. Chen, X. Xie, X. Feng, L. Yang, Full-field 3d measurement using multi-camera digital image correlation system, Optics and Lasers in Engineering 51 (9) (2013) 1044–1052. doi:https://doi.org/10.1016/j.optlaseng.2013.03.001.

[30] K. Patil, V. Srivastava, J. Baqersad, A multi-view optical technique to obtain mode shapes of structures, Measurement 122 (2018) 358–367. doi:https://doi.org/10.1016/j.measurement.2018.02.059.

[31] D. Liu, X. Xia, J. Chen, S. Li, Integrating building information model and augmented reality for drone-based building inspection, Journal of Computing in Civil Engineering 35 (2) (2021) 04020073. doi:10.1061/(ASCE)CP.1943-5487.0000958.

[32] P. J. Sanchez-Cuevas, A. Gonzalez-Morgado, N. Cortes, D. B. Gayango, A. E. Jimenez-Cano, A. Ollero, G. Heredia, Fully-actuated aerial manipulator for infrastructure contact inspection: Design, modeling, localization, and control, Sensors 20 (17) (2020) 4708. doi:10.3390/s20174708.

[33] P. Pfändler, K. Bodie, U. Angst, R. Siegwart, Flying corrosion inspection robot for corrosion monitoring of civil structures–first results, in: SMAR 2019-Fifth Conference on Smart Monitoring, Assessment and Rehabilitation of Civil Structures-Program, SMAR, 2019, pp. We–4. doi:10.3929/ethz-b-000365572.

[34] M. Á. Trujillo, J. R. Martínez-de Dios, C. Martín, A. Viguria, A. Ollero, Novel aerial manipulator for accurate and robust industrial ndt contact inspection: A new tool for the oil and gas inspection industry, Sensors 19 (6) (2019) 1305. doi:10.3390/s19061305.

[35] L. Duque, J. Seo, J. Wacker, Synthesis of unmanned aerial vehicle applications for infrastructures, Journal of Performance of Constructed Facilities 32 (4) (2018) 04018046. doi:10.1061/(ASCE)CF.1943-5509.0001185.

[36] P. Zhang, L. Zhang, T. Wu, H. Zhang, X. Sun, Detection and location of fouling on photovoltaic panels using a drone-mounted infrared thermography system, Journal of Applied Remote Sensing 11 (1) (2017) 016026. doi:10.1117/1.JRS.11.016026.

[37] Y. S. Lee, D. G. Lee, Y. G. Yu, H. J. Lee, Application of drone photogrammetry for current state analysis of damage in forest damage areas, Journal of Korean Society for Geospatial Information System 24 (3) (2016) 49–58. doi:10.7319/kogsis.2016.24.3.049.

[38] B. F. Spencer Jr, V. Hoskere, Y. Narazaki, Advances in computer vision-based civil infrastructure inspection and monitoring, Engineering 5 (2) (2019) 199–222. doi:10.1016/j.eng.2018.11.030.

[39] P. Hu, X. Hao, J. Li, C. Cheng, A. Wang, Design and implementation of binocular vision system with an adjustable baseline and high synchronization, in: 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), IEEE, 2018, pp. 566–570. doi:10.1109/ICIVC.2018.8492907.

[40] A. Noda, Y. Yamakawa, M. Ishikawa, Frame synchronization for networked high-speed vision systems, in: SENSORS, 2014 IEEE, IEEE, 2014, pp. 269–272. doi:10.1109/ICSENS.2014.6984985.

[41] M. N. Helfrick, C. Niezrecki, P. Avitabile, T. Schmidt, 3d digital image correlation methods for full-field vibration measurement, Mechanical systems and signal processing 25 (3) (2011) 917–927. doi:10.1016/j.ymssp.2010.08.013.

[42] H. Yoon, J. Shin, B. F. Spencer Jr, Structural displacement measurement using an unmanned aerial system, Computer-Aided Civil and Infrastructure Engineering 33 (3) (2018) 183–192. doi:10.1111/mice.12338.

[43] A. Khadka, B. Fick, A. Afshar, M. Tavakoli, J. Baqersad, Non-contact vibration monitoring of rotating wind turbines using a semi-autonomous uav, Mechanical Systems and Signal Processing 138 (2020) 106446. doi:10.1016/j.ymssp.2019.106446.

[44] C. Lee, W. A. Take, N. A. Hoult, Optimum accuracy of two-dimensional strain measurements using digital image correlation, Journal of Computing in Civil Engineering 26 (6) (2012) 795–803. doi:10.1061/(ASCE)CP.1943-5487.0000182.

[45] V. Hoskere, J.-W. Park, H. Yoon, B. F. Spencer Jr, Vision-based modal survey of civil infrastructure using unmanned aerial vehicles, Journal of Structural Engineering 145 (7) (2019) 04019062. doi:10.1061/(ASCE)ST.1943-541X.0002321.

[46] L. Fagiano, Systems of tethered multicopters: modeling and control design, IFAC-PapersOnLine 50 (1) (2017) 4610–4615. doi:10.1016/j.ifacol.2017.08.653.

[47] M. Bolognini, L. Fagiano, Lidar-based navigation of tethered drone formations in an unknown environment, IFAC-PapersOnLine 53 (2) (2020) 9426–9431. doi:10.1016/j.ifacol.2020.12.2413.

[48] K. G. Derpanis, The harris corner detector, York University 2 (2004).

[49] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, M. J. Marín-Jiménez, Automatic generation and detection of highly reliable fiducial markers under occlusion, Pattern Recognition 47 (6) (2014) 2280–2292. doi:10.1016/j.patcog.2014.01.005.

[50] A. Uryasheva, M. Kulbeda, N. Rodichenko, D. Tsetserukou, Dronegraffiti: autonomous multi-uav spray painting, in: ACM SIGGRAPH 2019 Studio, 2019, pp. 1–2. doi:10.1145/3306306.3328000.