Measurement 159 (2020) 107769

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Swaying displacement measurement for structural monitoring using computer vision and an unmanned aerial vehicle

Tung Khuc [a,*], Tuan Anh Nguyen [b], Hieu Dao [b], F. Necati Catbas [c]

[a] Department of Bridges and Highways Engineering, National University of Civil Engineering, Hanoi, Viet Nam
[b] Department of Information Technology, National University of Civil Engineering, Hanoi, Viet Nam
[c] Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, FL, USA

ARTICLE INFO

Article history:
Received 16 May 2019
Received in revised form 29 January 2020
Accepted 17 March 2020
Available online 20 March 2020

Keywords:
Vision-based displacement measurement
UAV
Structural heath monitoring
Noncontact measurement

ABSTRACT

Data acquisition is the challenging and crucial step for any structural health monitoring (SHM) scheme, especially on numerous measurement locations that are typically at very high elevations or largely inaccessible areas, which are often linked to time-consuming, costly, and to some extent, dangerous sensor implementation and cable wiring. Noncontact vision-based measurement techniques have been recognized recently as a primarily feasible approach, although it is still characterized by some limitations. To address these constraints, the proposed study introduced an enhanced noncontact displacement measurement method that employed an unmanned aerial vehicle (UAV) and computer vision algorithms. Since UAV can carry cameras to approach any difficult-to-reach regions, the proposed system can overcome several bottlenecks of the state-of-the-art vision-based methods with regard to finding a stationary place for the camcorder and for mitigating the inaccuracy induced by the long distance between the camcorder and the measurement location. Guided by the schematic framework for the system, a camera was mounted on the UAV for filming of the measurement point, and then displacements on that point were determined by a key-point vision-based measurement method. Moreover, translations generated by the UAV were obtained by means of reference objects on the background. Additionally, an autonomous scheme based on Canny edge detection and Hough transform were introduced for calculation of scale factors between the pixel and engineering unit for every image frame to address the issue of very fluctuant distances from the UAV to the measurement location. Subsequently, the actual displacements of the measurement location were measured following the elimination of the UAV motions from the displacement data. The proposed system was verified on an experiment with a small-sized steel tower where the outcomes provided an initial confirmation of the approach's promising potential.

1. Introduction

A built structure is ascertained to undergo the processes of aging and deterioration through its estimated lifespan. Although aging is inevitable, degradation can be abated with gradual inspection, assessment, and retrofitting, among other maintenance practices, which on implied setbacks are typically costly and sometimes ineffective to the extent of expected collapse of the structure and similar accidents. As structural safety is a palpable demand in today's engineering scenarios, methods and techniques capable of monitoring actual structures have attracted great attention. There have been a number of schemes recently developed under structural health monitoring (SHM), although such are still faced with some barriers for real-life implementations, especially when it comes to the process of data acquisition process [1].

Any SHM framework begins with data acquisition from sensors. In most cases, the reliability of a specific SHM method is dependent on not only on the quality of the collected data but also on the measurement locations on the structures. For example, displacements collected from the top of a tower may carry more valuable information than the ones acquired from its base. Unfortunately, high-rise buildings, cable bridge towers, transmission steel towers, wind turbine pylons, etc., are highly characterized by locations containing rich information for monitoring studies, but often are difficult or even impossible to access. For example, mounting sensors including accelerations, strain gauges, linear variable differential transformers (LVDTs), and wiring cable on these positions are costly, time-consuming, and, to some extent, dangerous.

* Corresponding author at: Department of Bridges and Highways Engineering, National University of Civil Engineering, 55 Giai Phong Street, Hanoi, Viet Nam.
E-mail address: tungkd@nuce.edu.vn (T. Khuc).

https://doi.org/10.1016/j.measurement.2020.107769
0263-2241/

Consequently, these disadvantages also paved the way for the consideration of noncontact or wireless monitoring techniques, which have rather become appealing to the SHM community lately.

In the pool of wireless technology, the use of wireless accelerometers for modal analysis to detect dynamic properties of structures has been introduced in numerous studies for the last two decades [2–5]. Although wireless sensors are wirework unnecessary and fairly on-budget, the power supply for sensor nodes remains a critical issue, especially in long-term monitoring. Alternatively, Global Positioning Systems (GPS) can be used to obtain dynamic displacements, which are later analyzed to identify the conditions of structures [6–8]. Nevertheless, for many data processing algorithms, measurement accuracy and sampling rates of a GPS are inadequate, as, given its nature as a wireless technology, it is mainly plagued by a shared drawback of all wireless measurement techniques, of inspectors who must access the measurement locations for mounting sensor nodes or GPS receivers.

Conversely, a noncontact measurement method seems more convenient than installing a wireless sensor/GPS receiver network, as such is depicted with the conduct of remote data acquisition, thus disregarding the need for inspectors to approach measurement positions. Noncontact measurement techniques are categorized into either radar-based, laser-based, or vision-based technologies and are based on the wave reflection and transmission theory, in which vibrations from structural surfaces (measurement locations) are transmitted to the data acquisition systems via radio or light waves. On one hand, portable interferometric radar systems (IRS) have been utilized in several real structures to obtain high accuracy vibrations for detecting modal frequencies [9–11]. On the other hand, laser Doppler and scanning laser vibrometers (SLV) have been applied more extensively about a decade ago, especially in laboratories [12–14]. Even though the radar and laser-based system expense are decreasing recently, price of a potable Interferometric Radar system or a high energy Scanning Laser vibrometer that can be effectively utilized on real structures from long distance is still much more expensive than convenient sensors, for instance strain gausses, accelerometers and linear variable differential transformers (LVDTs).

With regard to vision-based measurement methods, their noncontact and cost-effective characteristics have been associated with a ballooning research interest [15]. In fact, more than a hundred SHM studies utilizing a vision-based technique have been paper-reviewed recently [16,17]. By principle, vision-based displacement measurement is based on a tracking algorithm to trace a measurement location (commonly called a target) along a series of images (extracted from video clips). Considering that the camcorder is fixed on a stationary point, the target motion is calculated by comparing the target imaging coordinates. Subsequently, the target motions are converted into actual motions of structures in an engineering unit, i.e., millimeter, by deploying camera calibration or scale factor methods.

To track the measurement locations, physical targets are employed in many SHM studies. Following this approach, pattern-matching algorithms are conducted through the use of different types of markers, such as circles, squares, or random patterns mounted on the monitored locations. A round light-emitting diode (LED) template has been employed in some studies for acquisition of structural displacements, as its eminent color can be easily segmented from the image backgrounds, for detection and tracking [18,19]. By contrast, in the last decade, implementing black dots or squares on a white background was an extensive practice among authors as the marker is easily fabricated and does not require an energy source as opposed to the LED template. An example of an early implementation for this approach was carried out by Lee and Shinozuka [20], who proposed the use of 4 black dot markers. Here the dots were detected and then tracked based on a color filtering approach, and the dimensions of the dots were used to determine the scale factors. Similar templates have been employed in the succeeding studies [21–24]. Conducting random pattern targets in vision-based measurement is less common due to the fact that such targets are irrelevant in scale factor calculation. However, these patterns were employed in several bridge monitoring studies for collecting dynamic displacements by means of digital image correlation (DIC) technique [25,26].

The more recent approach proposed by several research groups is the non-target vision-based measurement (NVM), whose main advantage is depicted in its non-reliance to a physical target, making it fully noncontact implementation. With the absence of an artificial target, monitoring positions are tracked through the use of natural markers on structure images, for instance, bolts, gussets, or any stains on structural surfaces. A group of studies introduced DIC to track regions on structure surfaces [27–29]. Here, some special regions of interest (ROIs) on either bridge structures or traffic light poles were detected and matched to estimate structural vibration. Another approach for monitoring natural features on structure images is based on image key-point detection, a crucial technique in computer vision. Several algorithms for extracting and matching image key-points were successfully adopted in some primary papers for measurement of displacements in bridge girders, stadium beams, or a small-sized steel model [30–33].

1.1. Motivations and research contributions

Although a recent introduction, NVM is still associated with a few limitations of the vision-based approach, especially during on-site implementation. Such challenges include: (i) data storage requirement for clips and images, (ii) time-consuming image processing, (iii) less precision at camera orientation in a large angle (e.g., monitoring locations under high elevations), and (iv) occlusive views and reduced accuracy attributed to the long distance between the camera and measurement locations [30]. For the purposes of this study, an enhanced NVM method is proposed through the use of a consumer-grade unmanned aerial vehicle (UAV) to measure swaying displacements of tower structures for health monitoring. In reality, the displacements of very high towers commonly acquired by GPS devices are less precise and have low sample rate. In addition, it is difficult to employ a state-of-the-art NVM method due to challenges (iii) and (iv) stated above. Because UAV can carry a camera to film from a high location top-down at a desired distance (e.g., several meters), those barriers are solved and thus, the required displacements can be obtained for further SHM analysis of tower structures, including cable bridge towers, transmission steel towers, wind turbine pylons, and base transceiver station towers. Moreover, utilizing the UAV solves the constraint of finding a stationary place to set up a camcorder on a tripod as a requirement of every conventional vision-based measurement technique.

Even though deploying the UAV has some advantages, the main challenge is in discarding the vehicular motions as such translations play a role in acquiring displacement data. An approach to determine all UAV camera translations was proposed lately employing the camera calibration algorithm [34]. Nonetheless, this technique is more suitable for a laboratory study as it requires a physical calibration pattern attached on a stationary background, i.e., chessboard. With the emergence of gimbal technology for framing cameras under UAVs, three rotations of camera motions may have less influence to the measurement results. *In this study, the swaying displacements of tower-type structures are measured by a UAV with no physical calibration pattern (or target) implementation. This is achieved via elimination of the other three translations made by UAV camera motions by means of a known-dimension part of the monitored structure, combined with some advanced computer vision*

algorithms, such as key-point detection, Hough transform, Canny edge detection, etc. As a physical calibration pattern or a target is no longer a requirement, this method is further practical for real-life SHM studies, especially on high tower structures. The proposed NVM method using a UAV is conducted on a small-sized steel tower, and displacements of the top location on the structure are determined and verified. Additionally, the structural frequencies are obtained and compared with values calculated from the acceleration data.

2. Proposed framework

The framework of the upgraded NVM system using a UAV is composed of three primary steps, as shown in Fig. 1. Initially, displacements of the measurement location were determined by processing video clips filmed by the UAV. The employed vision-based displacement measurement method was proposed in a previous study based on the image key-point detection algorithm by the same authors of the present study. Next, the obtained displacements, including the UAV motions, were analyzed to discard the influences of the UAV translations by means of a stationary object on the background. Finally, pixel displacements of the measurement location were converted into an engineering unit using scale factors, which can be automatically determined in terms of Canny edge detection and Hough transform algorithms. The main contributions of the proposed system are reflected in the last two steps.

2.1. Key-point NVM method

Motion tracking was the basis of the vision-based displacement measurement. In computer vision, tracking the motions of an object comprises three stages, namely, detection of the object, matching the object with the images, and determining the object's motions. Traditionally, a physical marker is attached on a measurement location for easier detection and matching of the object. However, such implementation is generally laborious, especially when the markers are attached in a real-life scenario. Recently, natural image features were introduced as virtual markers of measurement locations.

In this study, fundamental image features called key-points were detected and acquired from images of the measurement location. Fig. 2 describes the steps involved in the key-point NVM method. Firstly, key-points obtained from two consecutive frames were matched based on their descriptor vectors and are amended by the geometric transformation algorithm. Secondly, displacements of the measurement position on the video clips were described by the motions of the key-points based on disparities of the paired matches.

To ensure tracking of results of the monitoring locations, robust image key-points that are invariant to changes in image transformation and illumination should be achieved. With image key-points being influential features in the computer vision field, several classical techniques could be used to extract those distinctive pixels such as Scale Invariant Feature Transform (SIFT) [35], Speed-up Robust Feature (SURF) [36], Binary Robust Invariant Scalable Key-points (BRISK) [37], Fast Retina Key-points (FREAK) [38], and KAZE feature [39]. Although the SIFT technique requires highly

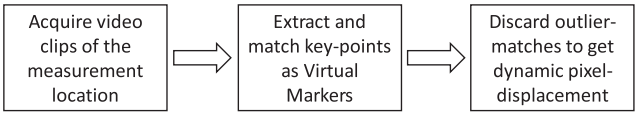

Fig. 2. Schematic framework of the key-point NVM method.

computation, it is still one of the most accurate and reliable algorithms [40]. Recently, several enhanced SIFT-based algorithms were introduced to solve the highly computation requirement of the SIFT technique [41,42], the original technique was still selected to employ in this proposed measurement framework due to two main reasons: reliability and knowledgeable implementation. Practically, the SIFT method identifies the salient points on the domain of the filtered image deviations. Fundamentally, it is based on the idea that the intensity values between two adjacent filtered images are hardly changed, except at special pixels. In implementation, the filtered images are created by convoluting numerous Gaussian kernels on the input image $I(x,y)$, mainly by preparing a pyramid of Gaussian functions corresponding to different standard deviation values of σ_i, at every scale of filtering. The filtered images $L(x,y,\sigma_i)$ were derived based on the expression:

$$L(x,y,\sigma_i) = \frac{1}{2\pi{\sigma_i}^2} e^{-\frac{x^2+y^2}{2{\sigma_i}^2}} \otimes I(x,y) \tag{1}$$

where $\otimes$ is convolution operation.

Accordingly, the filtered image deviations $D(x,y,\sigma_i)$ were formed by sequentially subtracting two adjacent filtered images (e.g., filtered images $L(x,y,\sigma_i)$ and $L(x,y,\sigma_{i+1})$). The key-point candidates were then located by detecting the local extrema values on the domain of $D(x,y,\sigma_i)$. Consequently, the robustness of the key-point candidates was verified by discarding the low-contrast and poorly located (e.g., along an edge) ones. A detailed explanation, as well as all related equations, for such implementation is found in the original paper [35]. The SIFT algorithm was employed to acquire key-points from a round plate of a steel model. The detected key-points (red circles) are illustrated in Fig. 3. Although the unreliable key-point candidates were rejected, many of these remained visible on edges and low-contrast locations.

To achieve measurement precision, only the most reliable key-points should be paired for calculating their pixel motions. Herein, two key-point sets extracted from two consecutive images were matched based on their descriptor vectors, basically formed by the gradient magnitudes and orientations of every key-point neighbor, as follows:

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \tag{2}$$

$$\theta(x,y) = tan^{-1}\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right) \tag{3}$$

where $m(x,y)$ is the gradient magnitude and $\theta(x,y)$ is the orientation of each neighboring pixel. If a neighbor size of 8 × 8 pixels was selected, there would be a total of 32 elements in the descriptor vector. Two key-points were assigned a pair when their descriptor

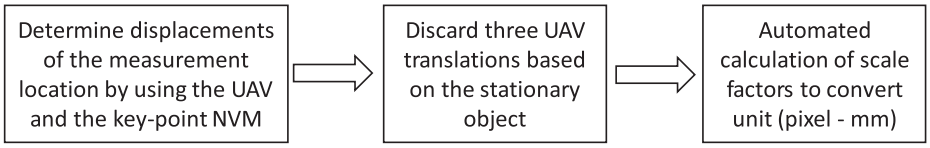

Fig. 1. Schematic framework of the NVM method using a UAV.

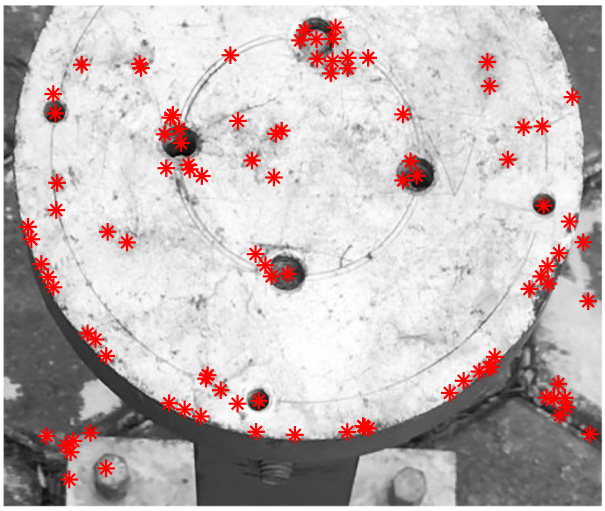

Fig. 3. Example of SIFT key-points detected on the image of the round steel plate.

vectors were the most analogous. Unexpectedly, the outcome of the matching algorithm based on the invariant descriptor vector would still provide numerous false matches, especially if the images of the measurement location were low-contrast, as shown in Fig. 4c. Due to reliable key-points would migrate together with the measurement location, and their corresponding motions must comply with the same transformation, a discarding outlier technique based on the geometric transformation approach was introduced to eliminate inaccurate matches. Assuming K_i and K_{i+1} are pixel-coordinate sets of matched key-points acquired from two consecutive images $(i)^{th}$ and $(i + 1)^{th}$, the relationship between them could be expressed in equations:

$$K_i = T \cdot K_{i+1} \tag{4}$$

$$or \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{bmatrix} \cdot \begin{bmatrix} x_{i+1} \\ y_{i+1} \\ 1 \end{bmatrix} \tag{5}$$

where T is the transformation matrix mapping $(i)^{th}$ image coordinates to $(i + 1)^{th}$ image coordinates of the measurement location. The T matrix could be determined based on the affine transformation equation. Once the matrix T was formed, the outlier relationships corresponding to inaccurate matched key-points were detected and removed. The result of this procedure is illustrated in Fig. 4d. After the erroneous matches are discarded, the motions between two set K_i and K_{i+1}, which were also the movements of the measurement position, were calculated in pixel unit.

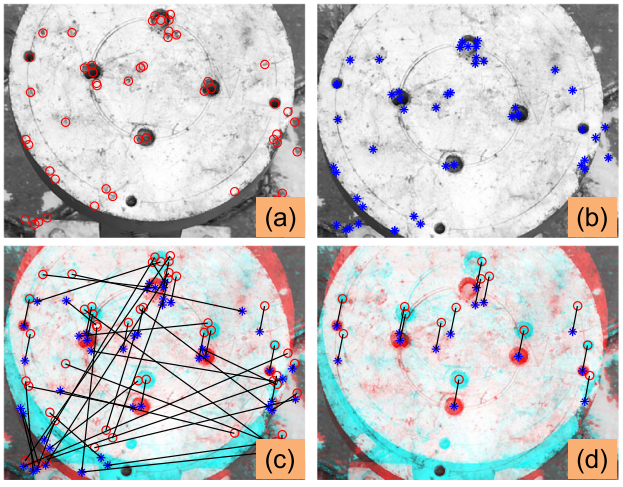

Fig. 4. Detected key-points on (a) image i^{th} and (b) image i + 1th. (c) Matches based on the invariant descriptor vector of key-point. (d) Most reliable matches after utilization of the geometric transformation algorithm.

2.2. Autonomous determination of scale factors between pixel and engineering unit

In general, there are two approaches for transforming pixel displacements obtained by vision-based methods into an engineering unit, i.e., millimeters. The first way is based on camera calibration, which commonly requires a type of calibration pattern such as a known-dimension cheeseboard, or some distances among stationary points in background. The second method is to calculate the scale factors between the pixel and the engineering unit based on a known-dimension component of the monitoring structure, i.e., bolts, gussets, sizes of shape steel elements, etc. This study utilized scale factors due to effortlessly finding a known-dimension component from the structure blueprint. Additionally, because distances from the UAV to the measurement locations were very changeable, an autonomous algorithm was introduced to track the known-dimension object, detect its edges, and measure its dimensions in pixel via Canny edge detection and Hough transform algorithms.

The scale factor is basically defined as a ratio between two dimensions of the same object or the distance in both the image and world coordinates. It can be calculated with the equation:

$$R = \frac{\Delta_{image}}{\Delta_{world}} \quad (pixel/mm) \tag{6}$$

where R is the scale factor, Δ_{image} is the dimension of the object in the image coordinate (pixel), and Δ_{world} is the dimension of the same object in the world coordinate (mm). As Δ_{world} can be measured or found from the object's blueprint, i.e., size of a bolt, width of a shape steel, thickness of a plate, etc., the Δ_{image} values are automatically determined for every image frame. First, the ROI around the known-dimension object is detected. Next, the Canny edge detection algorithm is employed on the ROI to identify the object edges. Finally, the Hough transform technique is implemented to detect lines or circles corresponding to such edges. Theoretical details of the Canny edge detection and Hough transform techniques are provided in the original papers [43,44]. Both algorithms were implemented in this study using the built-in functions of MATLAB 2018a. An example demonstrating the result of this procedure is presented in Fig. 5.

Using a sequence of images collected by the UAV, as shown in Fig. 6, the proposed autonomous process was verified to determine some dimensions of the steel model as diameters of the round plate and widths of the model base. Note that the drone altitudes were unstable during the recording period, which induced variation in the sizes among image frames, of the monitored objects. Fig. 6 also illustrates the detected circles of the round plate (highlighted in red) showing radius contrast of 51.20 pixels and 51.41 pixels, corresponding to the 1st and 150th frames. The widths of the model base (116.03 pixels and 117.02 pixels) were identified through the distances between two parallel lines (in red and in blue) obtained by the autonomous framework.

2.3. Elimination of the UAV motions

According the UAV-based NVM method, the high measurement locations (commonly on top of the tower structures) are filmed top-down, using a camera mounted on the UAV. Swaying displacements collected by the key-point NVM technique would be influ-

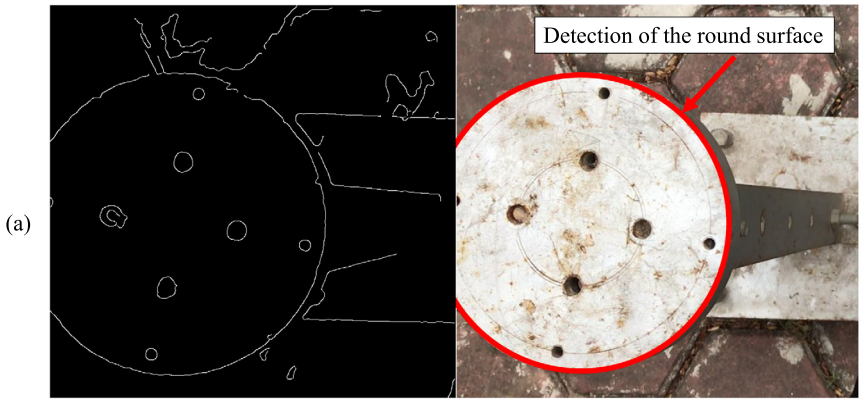

Fig. 5. (a) Canny edge detection; (b) Hough transform algorithm for detecting the steel round plate.

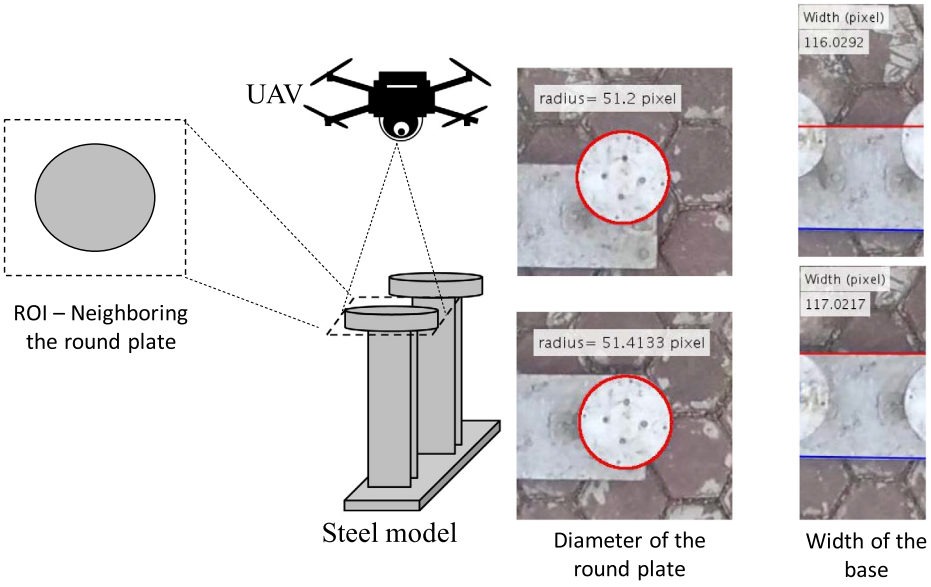

Fig. 6. Outcomes of determining round plate radius and base width at the 1st and 150th images in the UAV video clip for the autonomous scheme.

enced by the UAV motions, i.e., two horizontal translations and a vertical translation, whereas three rotations of the UAV were presumably discarded by the gimbal system. Fig. 7 illustrates the four-step method for eliminating the UAV motions, described as follows: (1) acquisition of swaying displacements of the measurement location S via the key-point NVM; (2) determination of same-direction translations D of the UAV by NVM through the use of a tracking reference object, i.e., house or a stationary structure on the ground; (3) elimination of the influence of the UAV motions, by transferring of translations D to become equivalent displacements D' at the same elevation as the measurement altitude via the perspective projection principle; and (4) determination of actual pixel displacements of the measurement location via elimination of the D' motions out of the displacement S data.

The most challenging part of this proposed procedure is transforming D into D' with fluctuant UAV elevations during monitoring. Based on the perspective projection principle, assuming that H_{UAV} and H_{mea} are the heights of the UAV and the measurement location, the values of D' can be calculated as follows:

$$D' = D \times \left(\frac{H_{UAV}}{H_{UAV} - H_{mea}}\right) \tag{7}$$

Here, H_{mea} is easily obtained, for instance, from the blueprint, but H_{UAV} is undeterminable. Thus, in this paper, H_{UAV} values were derived from those of H_{mea} in terms of two scale factors R_1 and R_2 (pixel/mm) determined at the measurement location elevation and at the altitude of the reference object on land, respectively. Due to oscillations of the UAV elevations, R_1 and R_2 should be

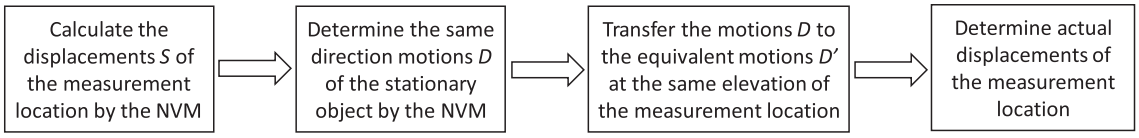

Fig. 7. Process for UAV motion elimination out of the acquired displacement data.

computed in every image frame in accordance with the autonomous process described in the previous section. In particular, H_{UAV} values were calculated as follows:

$$H_{UAV} = H_{mea} \times \frac{R_1}{(R_1 - R_2)} \quad (8)$$

Substituting Eq. (8) into Eq. (7) leads to transformed displacements $D^{'}$ of the UAV as follows:

$$D^{'} = D \times \frac{R_1}{R_2} \quad (9)$$

Note that D' motions were the equivalent horizontal pixel translations of the UAV obtained on the same elevation as the measurement point. Therefore, the influence of the UAV vertical translation changing toward the UAV horizontal motion was eliminated.

Consequently, the actual pixel displacements of the measurement location S_a were computed as follows:

$$S_a = \left(S - D^{'}\right) \quad (10)$$

Finally, S_a were transformed into an engineering unit by conversion of R_1 determined above.

$$S_a^{mm} = \frac{\left(S - D^{'}\right)}{R_1} \quad (11)$$

3. Verification experiment

The proposed vision-based displacement measurement method using a UAV was verified on a steel model at the Bridge and Tunnel Laboratory of the National University of Civil Engineering (NUCE). The steel model was a tower-type apparatus for dynamic experiments. There were two steel towers based on a rectangular steel plate of 16 mm × 200 mm × 400 mm (thickness × width × length). Each tower body was composed of two steel strips 1.2 mm thick and 50 mm wide. A round steel plate 150 mm in diameter and 19 mm thick was mounted on top of each tower to simulate a mass of 25.8 N. Due to the composition of the tower body, tower stiffness in the *X*-axis (on the long side of the model base) and *Y*-axis was very different, leading to tower movements in the *X*-axis only, as depicted in Fig. 8. Moreover, the stiffness of the tower body, as well as the mass of the round steel plate, could be altered in such a way as to obtain varying natural frequencies and other structural responses.

The sensing system used in this verification was an integration of consumer-grade UAV (DJI Phantom 3), a camcorder Canon VIXIA HF R42 attached on a tripod, and an accelerometer wired to a National Instrument data acquisition system (NI-9234 module). In this experiment, displacements of the round steel plate on the top of the tower were monitored through the UAV. The video clips were recorded top-down by a camera mounted on the DJI Phantom at the resolution of 2704 × 1520 pixels and a sample rate of 29.97 frame per second (fps). Two sorts of credentials were conducted, including displacements using the NVM system with the fixed camcorder as verified in several studies and frequencies using the accelerometer. The fixed VIXIA camcorder was placed 1.84 m away from the model to record motions of the round steel plate at the resolution of 1920 × 1080 pixels and the same rate of the UAV camera (29.97 fps). Meanwhile, the accelerometer was attached on the plate for collecting acceleration data. The specifications of the vision sensing system are shows Table 1.

3.1. Experiment results

Fig. 9 shows the displacements *S* of the round steel plate in the *X*-axis obtained by the UAV. On one hand, note the very detrending data with no oscillation being formed by the tower displacements, as opposed to that assumed beforehand. On the other hand, the same-direction UAV motions *D* were determined by tracking the reference object, which was the base of the steel model herein. The comparison mode in Fig. 9 explains that the detrending *S* data was an effect of the UAV translations. A total of 1800 images were analyzed to obtain 60 s of displacement datasets *D* and *S*. The computation time of the process is approximately 0.86 frame per second that is far below the camera frame rate of 29.97 frame per second. Thus, the processing time of the proposed measurement framework is much slower than the real-time processing. This limitation has been forecasted due to the highly computation requirement of the SIFT technique.

Following the procedure illustrated in Fig. 7, data *D* of the UAV should be necessarily transformed into *D'* by Eq. (9), before the UAV motions were discarded out of the raw displacements *S* of the round steel plate. To do so, scale factors R_1 and R_2 (pixel/mm), where R_1 was determined at the measurement location (the round steel plate) elevation and R_2 was obtained at the altitude of the stationary object (the model base), were calculated using the following equations:

$$R_1 = \frac{r_{image}}{r_{world}} \ (pixel/mm) \quad (12)$$

$$R_2 = \frac{w_{image}}{w_{world}} \ (pixel/mm) \quad (13)$$

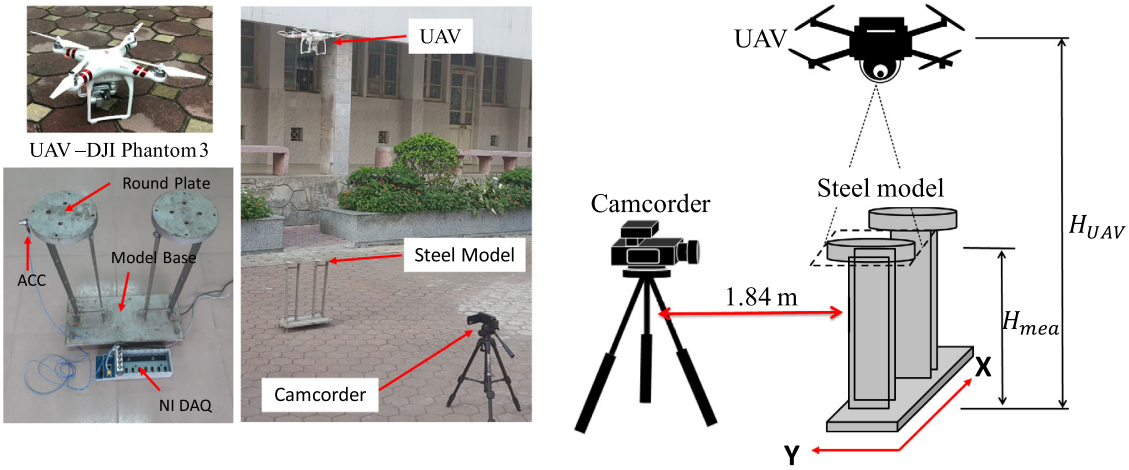

Fig. 8. Experimental setup with the steel model, the UAV, and the fixed camcorder.

Table 1
Specifications of the vision sensing system.

Item	Model	Technical Specifications	Remarks
Fixed Camcorder	Canon VIXIA HF R42	Max resolution: 1920 × 1080; Max frame rate: 59.94 fps; On board image storage; Zoom Ratio: 53× Advanced*/32× Optical/1060× Digital; Focal Length: 2.8–89.6 mm.	High resolution, with own zoom lens, suitable for close range measurement
UAV camera	Phantom 3 standard	Max resolution: 2704 × 1520 (2.7); Max frame rate: 29.97 fps; On board image storage; Zoom Ratio: fixed zoom; Focal Length: 20 mm	Very high resolution, can access targets for close range measurement by UAV
UAV	Phantom 3 standard	Max Service Ceiling Above Sea Level: 19,685 feet (6000 m); Max Flight Time: Approx. 25 min; Hover Accuracy Range on Vertical: ±0.5 m, Horizontal: ±1.5 m; Gimbal Angular Vibration Range: ±0.02°	It carries the camera to access any difficult-to-reach measurement locations, low cost.

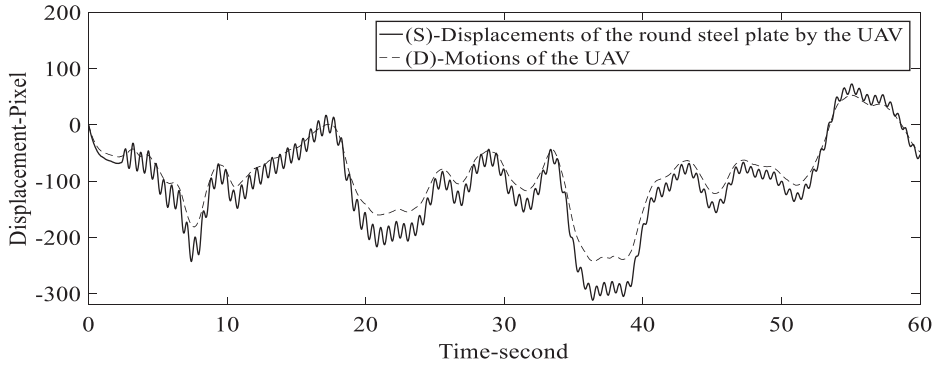

Fig. 9. Trends of displacement data of the round steel plate and motions of the UAV acquired through a stationary object (base of the model).

where r_{image} is the radius of the round steel plate estimated in each image (pixel), r_{world} is the radius of this plate in real world that is 75 mm, w_{image} is the width of the model base determined in images (pixel), and w_{world} is the real width (200 mm) of the base . Due to the up-and-down altitudes of the UAV, those scale factors were changed in every image frame and therefore were computed in the whole image sequence by an autonomous process based on Canny edge detection and Hough transform algorithms, as demonstrated in the preceding sections. Fig. 10 presents the results for R_1 and R_2. Here, although several false detections were observed, the remaining images (approx. 1800 images) were visually verified in this study through the outcome of the autonomous process. Additionally, Fig. 10 illustrates a correlation between R_1 and R_2 datasets, which should be disproportional to the altitude data of the UAV.

After the scale factors R_1 and R_2 are calculated, all the outliers are eliminated, and then the transformed displacement data D' of the UAV is determined and illustrated in Fig. 11.

Fig. 12 shows the graph of D' going through the median points of the graph of S. Such observation could be explained for the oscillation form of the round steel plate displacements after the UAV motions were discarded. Following the calculation of S of the round steel plate and the D' of the UAV, the actual displacements S_a of the steel model in the pixel unit were determined through Eq. (10). The graph of the S_a values is displayed in Fig. 12 (bottom).

Finally, the S_a data were converted to the displacements S_a^{mm} in an engineering unit based on the converting ratio R_1, as explained in Eq. (11). The result of the S_a^{mm} are presented in Fig. 13.

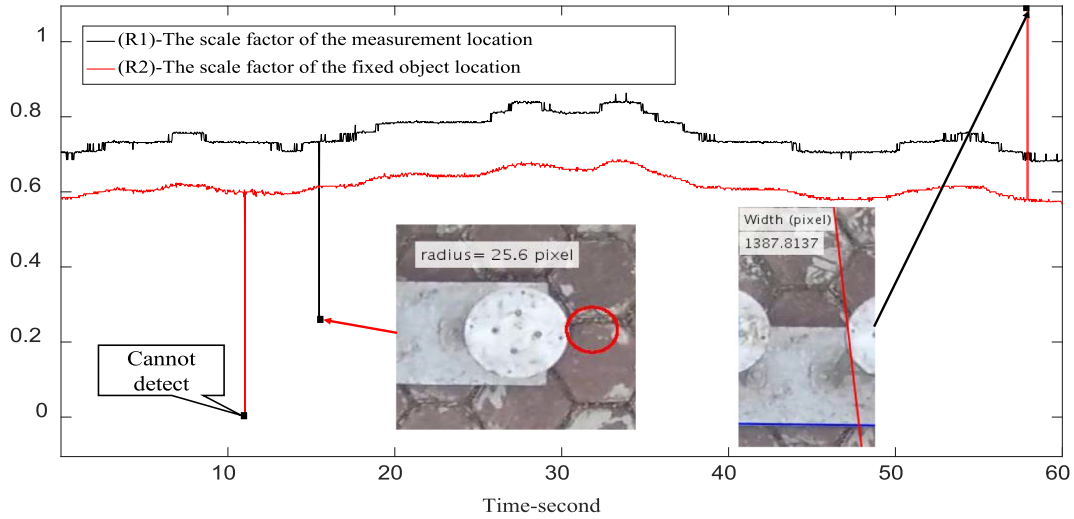

Fig. 10. Scale factors determined at different elevations for each image frame.

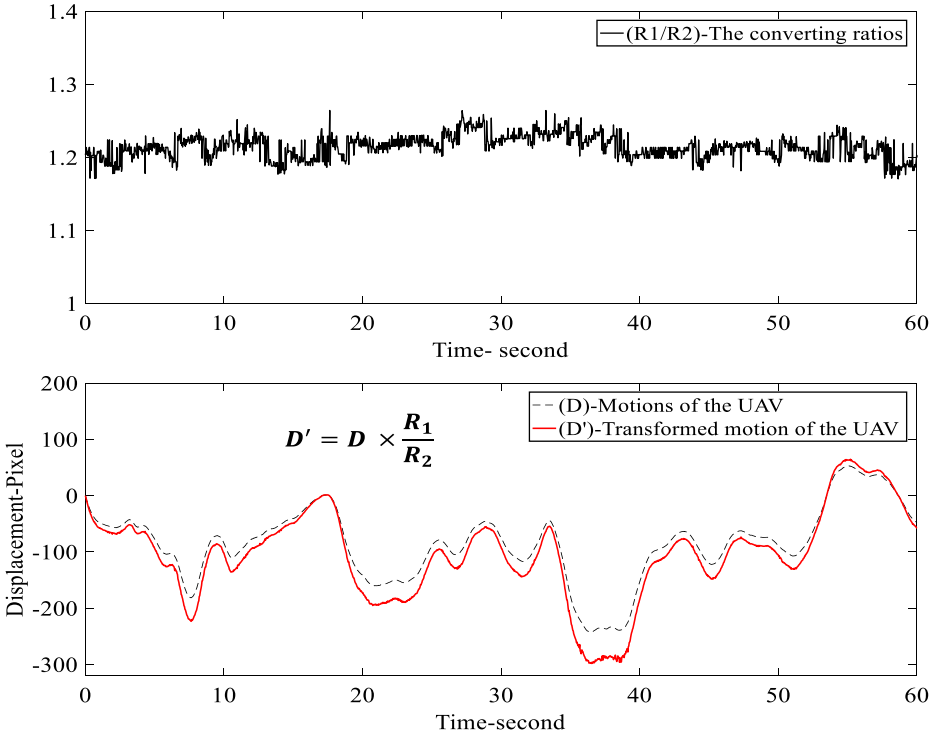

Fig. 11. Ratio between R_1 and R_2 (top) and the transformed displacement data of the UAV (bottom).

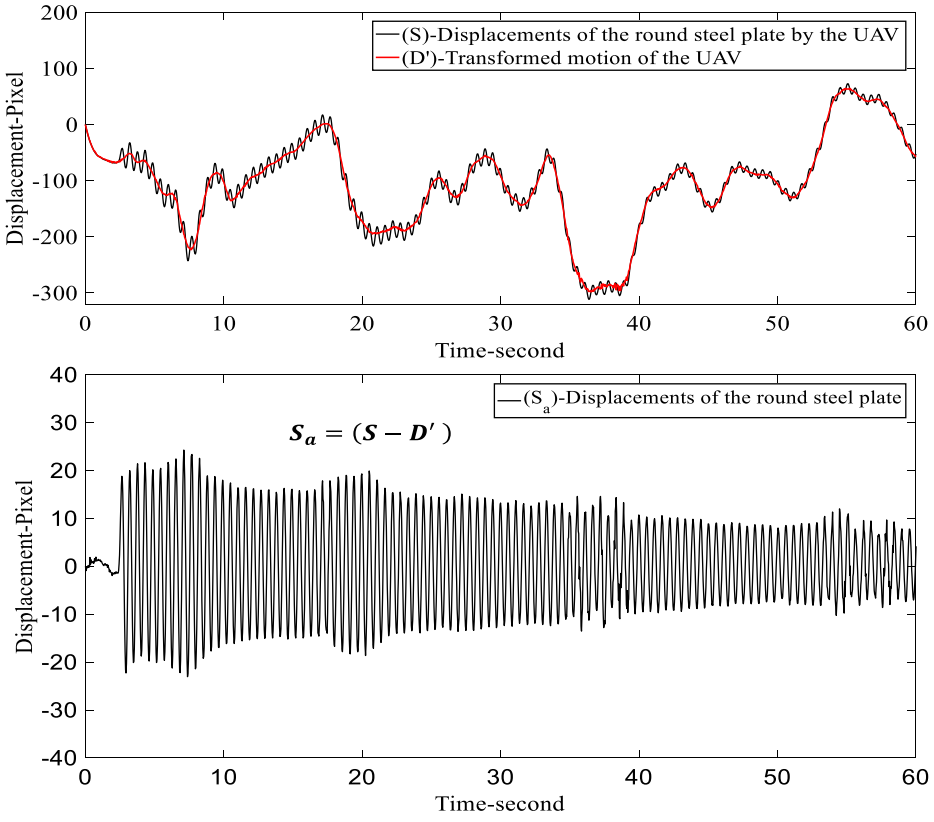

Fig. 12. Actual displacements of the round steel plate after elimination of the UAV motions.

3.2. Verification and discussions

The displacements of the round steel plate eventually determined by using a UAV were verified via comparison with data obtained by the NVM using the fixed Canon VIXIA HF R42 camcorder. This NVM method had been introduced and verified by other studies of the authors [30,32]. Based on that evaluation, the method accuracy depends on the distance from the camera

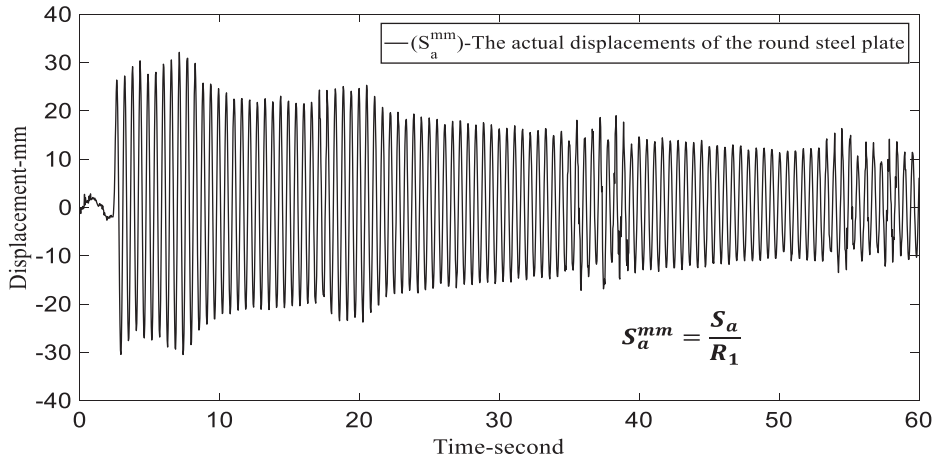

Fig. 13. Actual displacements of the round steel plate (in mm).

to measurement locations. For a particular camcorder of Canon VIXIA HF R42, the accuracy of the NVM method is ±0.01 mm for a measurement distance of 3 m; however, the accuracy becomes ±0.04 mm for a distance of 13 m [30]. Fig. 14 illustrates the displacement results obtained by both approaches in a synchronized fashion. Here, the motions of the measurement object determined by both methods highly matched each other in terms of displacement levels and vibration pattern. A zoom-in window around the 20th second presented an increase in displacements in spite of the absence of an additional force. Such problem was confirmed by visual observation of the model base movements on the clip due to instability, as captured by both the UAV camera and the fixed VIXIA camcorder. However, several mismatching patterns were observed on the time windows between 35–40 s and 52–56 s, which could have been induced by sudden turbulences and movements of the UAV under strong wind during hovering, as observed in Fig. 11.

Error analyses were performed to verify the correlation and error measurement between two datasets presented in Fig. 14. The determination coefficient *R-squared* determined using the out-

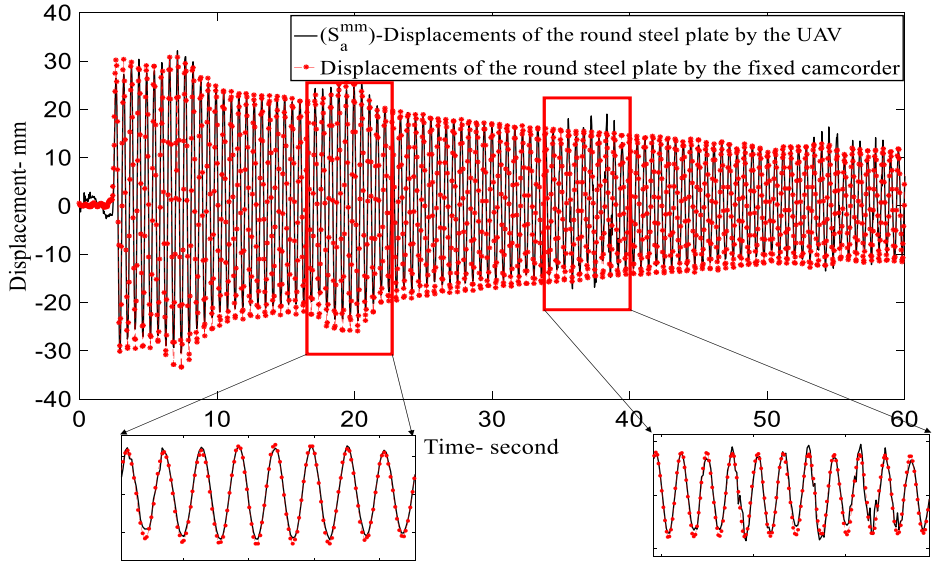

Fig. 14. Displacements of the round steel model by using a UAV and a fixed camcorder.

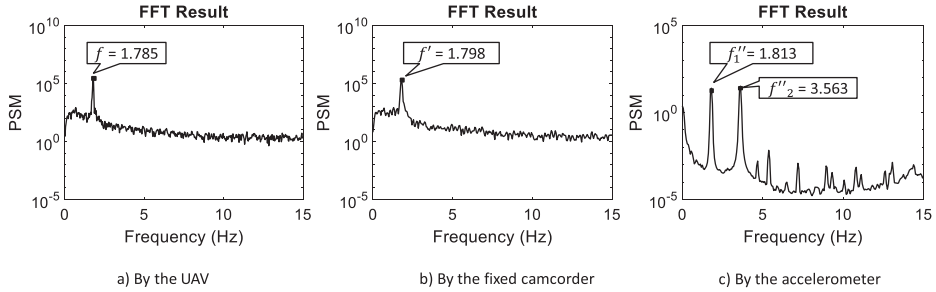

Fig. 15. Acquired natural frequencies of the steel model by a UAV, a fixed camcorder and an accelerometer.

comes from both methods is of 0.9723. It is seen that the *R-squared* value is very close to 1, which implies the good correlation and analogousness between two datasets measured by both the UAV camera and the fixed camcorder.

Another verification for the collected displacements was the structural identification. The oscillation data from the two techniques with the UAV and the fixed camcorder, along with the acceleration data from an accelerometer, were analyzed by the fast Fourier transform algorithm for finding the natural frequencies of the steel model. The frequency data provided by all methods were compared against each other to validate the ability of using a UAV-based platform for structural assessment. Fig. 15 shows the natural frequencies of the steel model identified by all three implementations. The first natural frequency identified by the UAV was $f = 1.785\ Hz$, while the one detected by the fixed camcorder was $f = 1.798\ Hz$. The first two of the model frequencies obtained by using the acceleration data were $f'_1 = 1.813\ Hz$ and $f'_2 = 3.563\ Hz$. Although the acceleration data could help detect more frequencies, the first frequencies acquired by all methods nearly matched, which implies the possibility of applying UAV for identifying modal frequencies of structures in SHM studies. Further, the two-way verification demonstrated in this section confirmed the potential of the proposed method using a UAV for displacement measurement on a tower-type structure.

4. Conclusion

The new concept of vision-based displacement measurement using a UAV proposed in this study exhibited the potential to effectively measure the dynamic displacements of tower structures in time-series, as are commonly obtained by GPS systems for SHM studies associated with some limitations as less accuracy and low sample rate data. The platform of the measurement system consisted of a low-cost UAV, a camera mounted on the UAV by a gimbal system, and computer vision. The image sequence of the measurement location filmed by the UAV was processed to obtain displacements through the key-point NVM method, which unfortunately also included UAV translations. Simultaneously, motions of the UAV were computed by means of tracking a stationary object on the background of the same image sequence. Several fundamental computer vision algorithms, i.e., Canny edge detection, Hough transform, and key-point NVM method, were introduced herein for autonomous calculation of the UAV motions.

Consequently, the UAV translations were eliminated from the displacement data of the measurement location to obtain the actual results. The proposed approach was verified at the Bridge and Tunnel Laboratory of the NUCE on an experiment of steel tower outside. Displacements obtained by the proposed vision-based technique illustrated acceptable outcomes after comparison with the other results measured by reference approaches, for example, an NVM method with a fixed camcorder and an accelerometer. The proposal enabled another way of measurement for determining the swaying displacements of tower structures, besides using conventional GPS systems. Furthermore, the introduced concept substantiated the idea of using UAVs to upgrade numerous state-of-the-art vision-based implementations in SHM. Nevertheless, a main limitation can be pinpointed related to highly computation requirement of the key-points algorithms, especially the SIFT technique. This disadvantage has limited time processing of the proposed framework far below real-time processing. Furthermore, the UAV flight time of 25 min has narrowed the time acquisition of monitoring studies. Some other challenges that needed overcoming remain, such as utilization of the method during a windy ambience, better result accuracy, UAV crashes and turbulences, etc., and should be considered for future research.

CRediT authorship contribution statement

Tung Khuc: Conceptualization, Methodology, Writing - original draft. **Tuan Anh Nguyen:** Data curation, Validation. **Hieu Dao:** Data curation, Validation. **F. Necati Catbas:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to acknowledge the supports from Vietnam Ministry of Education and Training and National University of Civil Engineering under the B2017-XDA-09 project.

References

[1] F.N. Catbas, T. Kijewski-Correa, Structural identification of constructed systems: Collective effort toward an integrated approach that reduces barriers to adoption, J. Struct. Eng. 139 (2013) 1648–1652.
[2] A. Sabato, C. Niezrecki, G. Fortino, Wireless MEMS-based accelerometer sensor boards for structural vibration monitoring: a review, IEEE Sens. J. 17 (2017) 226–235.
[3] X. Hu, B. Wang, H. Ji, A wireless sensor network-based structural health monitoring system for highway bridges, Comput.-Aided Civ. Infrastruct. Eng. 28 (2013) 193–209.
[4] J. Li, K.A. Mechitov, R.E. Kim, B.F. Spencer, Efficient time synchronization for structural health monitoring using wireless smart sensor networks, Struct. Control Health Monit. 23 (2016) 470–486.
[5] S. Cho, H. Jo, S. Jang, J. Park, H.-J. Jung, C.-B. Yun, et al., Structural health monitoring of a cable-stayed bridge using wireless smart sensor technology: data analyses, Smart Struct. Syst. 6 (2010) 461–480.
[6] P. Breuer, T. Chmielewski, P. Górski, E. Konopka, L. Tarczyński, Monitoring horizontal displacements in a vertical profile of a tall industrial chimney using Global Positioning System technology for detecting dynamic characteristics, Struct. Control Health Monit. 22 (2015) 1002–1023.
[7] F. Casciati, C. Fuggini, Engineering vibration monitoring by GPS: long duration records, Earthq. Eng. Eng. Vibrat. 8 (2009) 459–467.
[8] T.H. Yi, H.N. Li, M. Gu, Recent research and applications of GPS-based monitoring technology for high-rise structures, Struct. Control Health Monit. 20 (2013) 649–670.
[9] G. Luzi, M. Crosetto, C. Gentile, Radar interferometry as a tool for structural health monitoring: Current situation and perspectives of the technique for the next decade, in: Life-Cycle of Engineering Systems: Emphasis on Sustainable Civil Infrastructure, 2016, 145.
[10] C. Li, W. Chen, G. Liu, R. Yan, H. Xu, Y. Qi, A noncontact FMCW radar sensor for displacement measurement in structural health monitoring, Sensors 15 (2015) 7412–7433.
[11] C. Gentile, Deflection measurement on vibrating stay cables by non-contact microwave interferometer, NDT E Int. 43 (2010) 231–240.
[12] E.J. OBrien, A. Malekjafarian, A mode shape-based damage detection approach using laser measurement from a vehicle crossing a simply supported bridge, Struct. Control Health Monit., 23, 2016, pp. 1273–1286.
[13] T. Miyashita, M. Nagai, Vibration-based structural health monitoring for bridges using laser Doppler vibrometers and MEMS-based technologies, Int. J. Steel Struct 8 (2008) 325–331.
[14] W.J. Staszewski, R. Bin Jenal, A. Klepka, M. Szwedo, T. Uhl, A review of laser Doppler vibrometry for structural health monitoring applications, in: Key Engineering Materials, Trans Tech Publ, 2012, pp. 1–15.
[15] N. Catbas, C. Dong, O. Celik, T. Khuc, A vision for vision-based technologies for bridge health monitoring, in: Maintenance, Safety, Risk, Management and Life-Cycle Performance of Bridges: Proceedings of the Ninth International Conference on Bridge Maintenance, Safety and Management (IABMAS 2018), 9-13 July 2018, Melbourne, Australia, CRC Press, 2018, pp. 54.
[16] Y. Xu, J.M. Brownjohn, Review of machine-vision based methodologies for displacement measurement in civil structures, J. Civ. Struct. Health Monit. 8 (2018) 91–110.
[17] D. Feng, M.Q. Feng, Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection–A review, Eng. Struct. 156 (2018) 105–117.
[18] S. Park, H. Park, J. Kim, H. Adeli, 3D displacement measurement model for health monitoring of structures using a motion capture system, Measurement 59 (2015) 352–362.
[19] S.H. Chao, C.H. Loh, Vibration-based damage identification of reinforced concrete member using optical sensor array data, Struct. Health Monit. 12 (2013) 397–410.

[20] J.J. Lee, M. Shinozuka, Real-Time Displacement Measurement of a Flexible Bridge Using Digital Image Processing Techniques, Exp. Mech. 46 (2006) 105–114.
[21] J.M.W. Brownjohn, Y. Xu, D. Hester, Vision-Based Bridge Deformation Monitoring, Front. Built Environ. 3 (2017) 23.
[22] D. Ribeiro, R. Calçada, J. Ferreira, T. Martins, Non-contact measurement of the dynamic displacement of railway bridges using an advanced video-based system, Eng. Struct. 75 (2014) 164–180.
[23] X.W. Ye, Y.Q. Ni, T.T. Wai, K.Y. Wong, X.M. Zhang, F. Xu, A vision-based system for dynamic displacement measurement of long-span bridges: algorithm and verification, Smart Struct. Syst. 12 (2013) 363–379.
[24] Y. Xu, J. Brownjohn, Vision-based systems for structural deformation measurement: case studies, 2018.
[25] S.-W. Kim, S.-S. Lee, N.-S. Kim, D.-J. Kim, Numerical model validation for a prestressed concrete girder bridge by using image signals, KSCE J. Civ. Eng. 17 (2013) 509–517.
[26] X. Ye, T.-H. Yi, C. Dong, T. Liu, Vision-based structural displacement measurement: system performance evaluation and influence factor analysis, Measurement (2016).
[27] M.Q. Feng, Y. Fukuda, D. Feng, M. Mizuta, Nontarget vision sensor for remote measurement of bridge dynamic response, J. Bridge Eng. 20 (2015) 04015023.
[28] D.T. Bartilson, K.T. Wieghaus, S. Hurlebaus, Target-less computer vision for traffic signal structure vibration studies, Mech. Syst. Sig. Process. 60 (2015) 571–582.
[29] 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, Struct. Health Monit. 12 (2013) 440–456.
[30] T. Khuc, F.N. Catbas, Completely contactless structural health monitoring of real-life structures using cameras and computer vision, Struct. Control Health Monit. 24 (2017).
[31] H. Yoon, H. Elanwar, H. Choi, M. Golparvar-Fard, B.F. Spencer, Target-free approach for vision-based structural system identification using consumer-grade cameras, Struct. Control Health Monit. 23 (2016) 1405–1416.
[32] T. Khuc, F.N. Catbas, Computer vision-based displacement and vibration monitoring without using physical target on structures, Struct. Infrastruct. Eng. 13 (2017) 505–516.
[33] Y. Xu, J. Brownjohn, D. Kong, A non-contact vision-based system for multipoint displacement monitoring in a cable-stayed footbridge, Struct. Control Health Monit. 25 (2018) e2155.
[34] H. Yoon, J. Shin, B.F. Spencer Jr, Structural Displacement Measurement using an Unmanned Aerial System, Comput.-Aided Civ. Infrastruct. Eng. 33 (2018) 183–192.
[35] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision 60 (2004) 91–110.
[36] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, Speeded-Up Robust Features (SURF), Comput. Vis. Image Underst. 110 (2008) 346–359.
[37] S. Leutenegger, M. Chli, R.Y. Siegwart, BRISK: Binary robust invariant scalable keypoints, in: Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE, 2011, pp. 2548–2555.
[38] A. Alahi, R. Ortiz, P. Vandergheynst, Freak: Fast retina keypoint, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 510–517.
[39] P.F. Alcantarilla, A. Bartoli, A.J. Davison, KAZE features, in: European Conference on Computer Vision, Springer, 2012, pp. 214–227.
[40] S.A.K. Tareen, Z. Saleem, A comparative analysis of sift, surf, kaze, akaze, orb, and brisk, in: 2018 International conference on computing, mathematics and engineering technologies (iCoMET), IEEE, 2018, pp. 1–10.
[41] Y. He, G. Deng, Y. Wang, L. Wei, J. Yang, X. Li, et al., Optimization of SIFT algorithm for fast-image feature extraction in line-scanning ophthalmoscope, Optik 152 (2018) 21–28.
[42] J. Peng, Y. Liu, C. Lyu, Y. Li, W. Zhou, K. Fan, FPGA-based parallel hardware architecture for SIFT algorithm, in: 2016 IEEE International Conference on Real-time Computing and Robotics (RCAR), IEEE, 2016, pp. 277–282.
[43] J. Canny, A computational approach to edge detection, in: Readings in computer vision, Elsevier, 1987, pp. 184–203.
[44] R.O. Duda, P.E. Hart, Use of the Hough transformation to detect lines and curves in pictures, in: Sri International Menlo Park CA Artificial Intelligence Center, 1971.