STRUCTURAL CONTROL AND HEALTH MONITORING
Struct. Control Health Monit. (2016)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/stc.1850

Target-free approach for vision-based structural system identification using consumer-grade cameras

Hyungchul Yoon[1,*,†], Hazem Elanwar[1,2], Hajin Choi[1], Mani Golparvar-Fard[1] and Billie F. Spencer Jr.[1]

[1]*Department of Civil and Environmental Engineering, University of Illinois at Urbana–Champaign, Urbana, IL 61802, USA*
[2]*Department of Structural Engineering, Cairo University, Cairo, Egypt*

SUMMARY

Recent reports on America's infrastructure have emphasized the importance of structural health monitoring of civil infrastructures. System identification is a key component of many structural health monitoring strategies. Current system identification methods estimate models of a structure by measuring displacements, accelerations, and strains with wired or wireless sensors. However, these methods typically involve installation of a limited number of sensors at discrete locations and require additional data acquisition devices. To overcome these limitations, computer vision-based techniques have been introduced recently that employ high-speed and high-resolution cameras. Such cameras can be quite costly and require tedious installation of targets. This paper investigates the potential of using consumer-grade cameras for structural system identification without the need to install targets. The underlying methods for target-free displacement measurements are introduced, including region of interest selection, feature detection, point tracking, and outlier removal. A set of experiments are conducted to assess the efficacy of the proposed approach by comparing the accuracy of the identified model with one obtained using a conventional wired system. Careful comparison of the results demonstrates the significant potential of the proposed approach. Copyright © 2016 John Wiley & Sons, Ltd.

Received 8 May 2015; Revised 6 January 2016; Accepted 22 January 2016

KEY WORDS: structural health monitoring; system identification; computer vision; feature tracking; consumer-grade cameras

1. INTRODUCTION

The American Society of Civil Engineers' report card gave America's civil infrastructure an overall D+ in 2013. This report indicates a total of $3.6 trillion is needed to rehabilitate existing infrastructure systems [1]. National budget priorities do not allow for this high level of investment, so many existing infrastructure systems will be left deficient. Determining prioritization of the maintenance, repair, and replacement of this infrastructure requires inspection, a process that can be costly and prone to error.

In the case of large-scale infrastructure systems such as bridges and high-rise buildings, diverse dynamic loads, including traffic, wind, and earthquakes can aggravate the condition of the structures. Structure health monitoring (SHM) has been proposed to ensure the safety of civil infrastructure. According to Kim [2], the goal of SHM can be outlined as (i) cost-effective assessment of structural performance, (ii) load estimation, (iii) detection and location of damage, and (iv) structural prognosis. A key component of SHM is system identification, the focus of which is to build an effective model of the

*Correspondence to: Hyungchul Yoon, Department of Civil and Environmental Engineering, University of Illinois at Urbana–Champaign, Urbana, IL 61802, USA.
†E-mail: yoon.illinois@gmail.com

Copyright © 2016 John Wiley & Sons, Ltd.

structural system. The identified model and system properties can be used to update analytical models and potentially identify damage.

Structural system identification requires acquisition of structural responses such as displacements, accelerations, and strains. The associated wired sensors can capture the structural behavior accurately, but require the sensors be directly attached to a structure. Furthermore, additional data acquisition (DAQ) devices and the accompanying long wires make the measurement procedure tedious [3]. Wireless sensor have been proposed to overcome cabling problems with measurement of the dynamic response of structures [4,5]. While the enhancements in both the hardware and software aspects of improved the performance of wireless sensors, these methods are still challenged by the need for manual installation of the sensors.

Recent computer vision-based techniques provide an opportunity to measure the dynamic movement of the structures with minimal effort. Compared with conventional measurement instruments, these methods do not require installation and maintenance of expensive sensor setups. Some of their early examples of these methods are provided in Nogueira *et al.* [6], Chang and Xiao [7], Khalil [8], and Lee *et al.* [9], which present a method for displacement measurement using specially designed targets. Morlier [10], Ji and Chang [11], and Caetano *et al.* [12] employ the optical flow-based method to automatically measure the structural displacement from a sequence of images. Shih and Sung [13], and Kim *et al.* [14] leverage the digital image correlation technique [15] to extract displacements of structures and show an example of characterizing vibration mode of a cantilever beam. Schumacher and Shariati [16] use virtual visual sensors and measure the rate of change in intensity of certain points to characterize dynamic motions for simple structures in both laboratory and field conditions. Fukuda *et al.* [3] and Feng *et al.* [17] used template matching technique called OCM to track points without target panel.

Despite success in preliminary experiments, most of these vision-based measurement techniques still have one or more of limitations among the next. First, these techniques either require installment of targets, which makes the process tedious, or employ a template matching algorithm, which may be computationally costly. Second, most of these techniques require relatively high-speed and high-resolution cameras, or require additional acquisition equipment and lenses, which can be prohibitively expensive. Moreover, issues such as temporal aliasing and sampling frequency variations have not been discussed. Finally, most vision-based methods have been described and verified in the purpose of the displacement measurement itself, but not for the dynamic analysis of the structures and system identification.

This paper develops a target-free approach for vision-based structural system identification using consumer-grade cameras such as smartphones or action cameras. Recent consumer-grade cameras have improved dramatically in terms of both frame rate and resolution. Now, a camera with 1920×1080 at 120fps (GoPro Hero 4 Black) is available, and today's smartphones are capable of taking video with 1280×720 at 240fps (iPhone6). The framework for using these consumer-grade cameras that are carried by most people in the everyday life is discussed while eliminating the need for installing targets on the structure. Also, a computationally effective data processing approach is proposed by tracking the intrinsic features of the object using optical flow estimation [18], a method to estimate the relative motion between the scene. Outlier detection and subpixel measurement algorithms for structures were introduced to minimize the error using such low-cost cameras. An experiment on a six-story model building is conducted in which both accelerometers and the proposed vision-based system are used to collect response data. The adequacy of any DAQ system can only be measured in terms of how the data is going to be used, and thus, the accuracy of the identified model should be used as the metric to assess the efficacy of the techniques for system identification. Therefore, to demonstrate the efficacy of the proposed approach models are identified using ERA, a popular system identification method, the accuracy of the natural frequencies, and the modes shapes of the identified a models are compared.

2. APPROACH

Figure 1 shows the overview of the proposed method. The underlying pipeline is composed of three main components: (i) camera calibration, (ii) vision-based displacement measurement, and (iii) system identification. The section presents the details of each component.

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

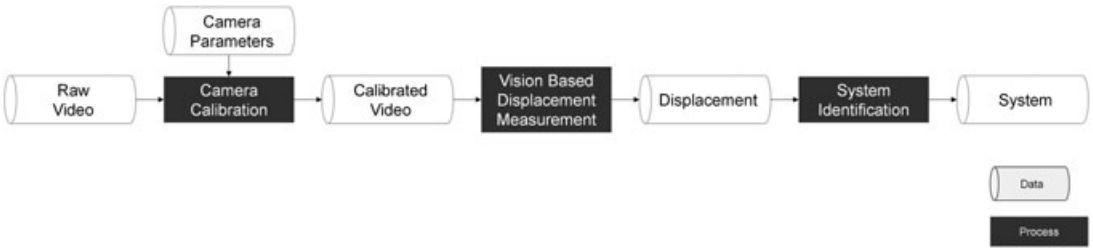

Figure 1. Overview of the target-free vision based system identification.

2.1. Camera calibration

The first step in the proposed approach is camera calibration. Modern consumer-grade camera lenses have improved dramatically in recent years. However, the recent trend toward small, lightweight, low-cost, and high-definition action cameras, such as GoPro or those used in commodity smartphones, often employ very wide angle lens. These lenses increase the field-of-view by intentionally introducing significant radial distortion. To remove this distortion and get accurate displacement measurements using consumer-grade cameras, performing camera calibration is necessary.

The camera calibration process determines the pixel scaling factor, camera focal length, lens axis offset, and the lens distortion characteristics, which then allows the uncalibrated video images to be transformed into calibrated images, from which displacement measurements can be extracted. To remove distortion effects, the camera calibration method reported by Zhang [19] was employed in this work. For completeness, the method is briefly reviewed here. The method uses a pinhole camera model and precisely estimates the focal length (f), radial distortion (k_1 and k_2), rotation matrix ($\boldsymbol{R}$), and translation vector ($\boldsymbol{T}$). The formula for projecting a 3D point $\boldsymbol{X}$ into a camera is as follows:

$$\boldsymbol{P} = \boldsymbol{R}\boldsymbol{X} + \boldsymbol{T} \tag{1}$$

where $\boldsymbol{X}$ is in global coordinates and $\boldsymbol{P}$ is the projection on the 2D image with homogenous coordinates. By dividing $\boldsymbol{P}$ by its third component P_z, the perspective is taken into account as follows:

$$\boldsymbol{p} = -\boldsymbol{P}/P_z \tag{2}$$

Finally, the radial distortion function, $r(p)$, is defined as

$$r(\boldsymbol{p}) = 1.0 + k_1 ||\boldsymbol{p}||^2 + k_2 ||\boldsymbol{p}||^4 \tag{3}$$

and used to convert $\boldsymbol{p}$ into the undistorted coordinates $\boldsymbol{p}'$

$$\boldsymbol{p}' = f \cdot r(\boldsymbol{p}) \cdot \boldsymbol{p} \tag{4}$$

The distortion function is required to be estimated only once for each camera, and can be used to undistort the image frames obtained from any video taken by the camera.

2.2. Vision-based Displacement Measurement

Once the camera is calibrated and the image distortion is removed, the dynamic response of structures is determined by analyzing the video frame-by-frame. This process involves four consecutive steps, which are shown in Figure 2: (i) selecting a region of the interest (ROI), (ii) detecting features, (iii) tracking the detected features, and (iv) removing tracking outliers.

Here, the ROI indicates the location of the object that needs to be tracked. Each ROI typically contains several features, which are used for displacement measurement purposes. The ROI is selected manually by drawing a box in the first frame of the video (more details are introduced in Section 3).

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

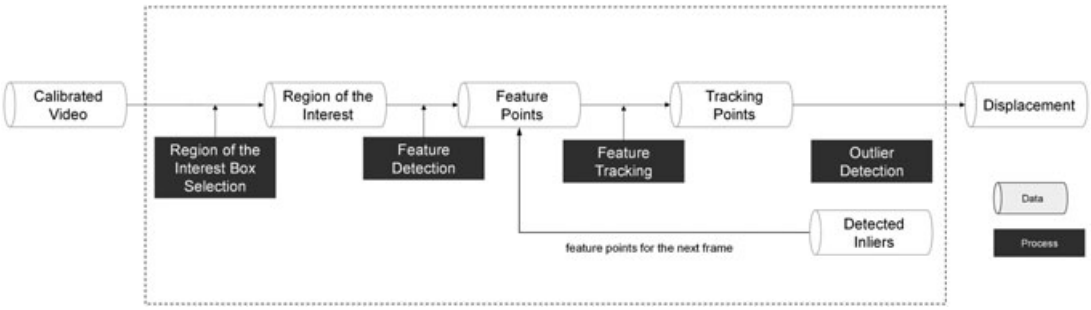

Figure 2. Data and process in the vision-based displacement measurement method.

To achieve reliable tracking, distinct (highly discriminative and salient) features need to be detected from the objects of interest. These features should be invariant to changes in illumination, scale, and pose (rotation and affine), as well as characterize the local proximity of the points of interest. There are several feature detection methods that can be used for this purpose. In this paper, the corner detection method suggested by Harris and Stephens [20] is used to extract the features within the ROI in the initial video frame. Harris corner detection works well with tracking methods such as the Kanade–Lucas–Tomasi (KLT) algorithm [21], which will be discussed in the next section. The Harris corner detection method characterizes the weighted sum of the squared differences of their intensities, assuming that the transitional shift (x, y) is small. This relationship is expressed as

$$\mathrm{E}_{\mathrm{x,y}} = \sum_{u,v} w_{u,v}\left[I_{x+u,y+v} - I_{u,v}\right]^2 = \sum_{u,v} w_{u,v}\left(\frac{\partial \mathrm{I}}{\partial \mathrm{x}}x + \frac{\partial \mathrm{I}}{\partial \mathrm{y}}y\right)^2 = (x\ y)\mathbf{M}(x\ y)^T \tag{5}$$

where $I_{u,v}$ is the initial 2D video frame and $I_{x+u,y+v}$ is the shifted frame. In this equation, **M** is the second moment matrix that characterizes whether the intensity gradients are hirozontal or vertical, and is expressed as

$$\mathbf{M} = \begin{bmatrix} \left(\frac{\partial \mathrm{I}}{\partial \mathrm{x}}\right)^2 & \left(\frac{\partial \mathrm{I}}{\partial \mathrm{x}}\right)\left(\frac{\partial \mathrm{I}}{\partial \mathrm{y}}\right) \\ \left(\frac{\partial \mathrm{I}}{\partial \mathrm{x}}\right)\left(\frac{\partial \mathrm{I}}{\partial \mathrm{y}}\right) & \left(\frac{\partial \mathrm{I}}{\partial \mathrm{y}}\right)^2 \end{bmatrix} \tag{6}$$

Considering α and β as the two eigenvalues of **M** – that is, the two directions of the fastest changes of intensity gradients – then a corner will be found only when both α and β have large positive values. To reduce the computational cost, the following response function is used calculate corner response M_c instead of directly finding the eigenvalue values.

$$M_c = \mathrm{Det}(\mathbf{M}) - k\mathrm{Tr}(\mathbf{M})^2 \tag{7}$$

where k is a tuuable parameter. M_c will have high positive values in the corner region, negative in the edge regions, and small in the lat region.

Once the feature are selected for the initial frame, the KLT algorithm [18,22] is adopted to track the point features for the entire duration of a video. The intensity of a current frame, $J(\mathbf{x})$, can be expressed by using the intensity of the previous frame, $I(\mathbf{x})$, as shown in Equation (8) by assuming a small motion for the features.

$$J(\mathbf{x}) = I(\mathbf{x} - \mathbf{d}) = I(\mathbf{x}) - \mathbf{g}\cdot\mathbf{d} \tag{8}$$

where **d** is the displacement vector between the two consecutive video frames and the gradient vector $\mathbf{g} = \left(\frac{\partial \mathrm{I}}{\partial x}, \frac{\partial \mathrm{I}}{\partial y}\right)$. The residue ϵ for a small window of pixels around the feature can be defined by the following equation:

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

$$\epsilon = \int [I(\mathbf{x} - \mathbf{d}) - J(\mathbf{x})]^2 w dA = \int [I(\mathbf{x}) - \mathbf{g}\cdot\mathbf{d} - J(\mathbf{x})]^2 w dA = \int (h - \mathbf{g}\cdot\mathbf{d})^2 w dA \tag{9}$$

where w is a weighting function and $h = I(\mathbf{x}) - J(\mathbf{x})$. To minimize the residue, the equation earlier can be differentiated with respect to $\mathbf{d}$ and set the result equal to zero as

$$\frac{d\epsilon}{d\mathbf{d}} = \int (h - \mathbf{g}\cdot\mathbf{d})\mathbf{g} w dA = 0 \tag{10}$$

Finally, the displacement vector $\mathbf{d}$ can be obtained by the following equation.

$$\mathbf{d} = \mathrm{G}^{-1}\boldsymbol{e} \tag{11}$$

where $\mathrm{G} = \int \boldsymbol{g}\boldsymbol{g}^T w dA$ and $\mathbf{e} = \int (I - J)\boldsymbol{g} w dA$. For each feature $\mathbf{x}_i$, the displacement vector $\mathbf{d}_i$ is calculated using the KLT algorithm.

In this framework, the MLESAC modeling fitting method described in [23] was used to remove the displacements that are not consistent with the dominant geometric transformation between two consecutive frames. This strategy minimizes the noise and out-of-plane errors. The motion of the ideal features $\underline{\mathrm{x}}$ (without any noise) between two consecutive frames can be expressed as the equation later.

$$\underline{\mathrm{x}}'_i = \mathbf{H}\underline{\mathrm{x}}_i \tag{12}$$

where $\underline{\mathrm{x}}_i$ is the set of the homogenous images points of the features in the first frame, $\underline{\mathrm{x}}'_i$ in the next frame, and $\mathbf{H}$ is the transformation matrix. In general, there are four types of transformation: (i) translate transformation, which only consider the translation of the two coordinates, (ii) similarity transformation, which considers the scale together with the translation, (iii) affine transformation, which adds skew of the motion, and (iv) projective transformation, which considers the full projective motion of the object. In this paper, the transformation matrix was considered as projective transformation to take account for out-of-plane motion of the structure. To estimate the projective transformation, four pairs of features are required to be tracked. However, because of the noise, the result of the transformation matrix may not be exact. Here, the residual error due to transformation is considered as the sum of the distance between the original and transformed features in the consecutive frames. The MLESAC randomly selected a minimal number of points and uses the expectation maximization algorithm together with this reprojection error distance explicitly to minimize the likelihood estimate of the residual errors. Hence, points that fit more closely to the underlying transformation model are considered more favorable than the points that are farther from the model. These inliers will then be tracked on the next video frame using the KLT algorithm. This procedure is repeated until the last frame of the video. The displacement of the object can finally be measured in pixels by tracking the center of the ROI.

2.3. System Identification

Once the dynamic displacement of the features and the ROIs on the structure are obtained, the next step involves analyzing the dynamic characteristics of the structure by using system identification. As discussed in the introduction, system identification techniques estimate a mathematical model to represent the physical structure. In this paper, the eigensystem realization algorithm (ERA) proposed by Juang and Pappa [24] is used to identify this system with known input.

Four sequential steps need to be conducted to identify the system using the ERA method (Figure 3). First, Based on the displacements determined from the tracking algorithm, the power spectral density (PSD) function can be obtained. Next, the transfer function is obtained by dividing the output PSD function with the input PSD function. Finally, the impulse response function, which will be used as the input of the ERA, can be obtained by applying inverse Fourier Transform.

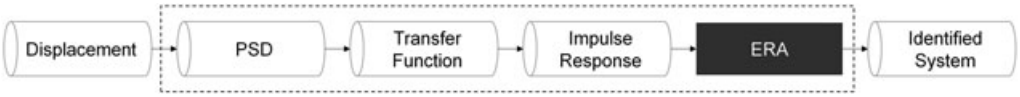

Figure 3. Flowchart for system identification.

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

Eigensystem realization algorithm uses the impulse response function to construct Hankel Matrix, which represents the data structure for the Ho-Kalman algorithm [25]. The Hankel matrix is decomposed by singular-value decomposition to determine the order of the system. Finally, the system can be obtained by finding the eigenvalues of the realized state matrix.

3. VALIDATION TEST

3.1. Test Setup

An experiment is designed to verify the proposed method. More specifically, the purpose is to verify whether the proposed method can leverage commercially available cameras to characterize the structural displacements at wide frequency ranges and use these measurements for system identification purposes. The six-story building model was selected to validate the proposed method for multi degree of freedom (DOF) structures with broad frequency ranges. To measure accuracy, the identified system will be compared with the accelerometers results as a reference solution.

Figure 4 shows the different components of the experiments, where the analyzed structure is a six-story model with equally distributed masses that are connected through elastic springs. The model is fixed on a uni-directional shaking table with maximum displacement stroke of 5 inches. The model response is recorded using two different instruments, which are the following:

1. *Cameras*: two consumer-grade cameras are used to record the displacement output of the model and the shaking table. These cameras are oriented perpendicular to the motion axis as shown in Figure 5. The first camera is a GoPro Hero3 Black edition, which can record with frame rate of 120 fps with 720p spatial resolution (1280×720 pixels in progressive or non-interlaced scanning mode in which all the rows of each videos frame are captured in sequence). The other is the camera on an LG G3 smartphone with recording rate of 30 fps and a resolution of 1080p (1920×1080 pixels, also in progressive mode). In order to improve the accuracy in detecting features necessary for displacement measurements, a fixed white background is used behind the experiment setup. The camera calibration procedure as discussed in Section 2.1 was performed to the two cameras, and radial distortions were removed prior to conducting the experiments.
2. *Reference sensors*: six accelerometers are attached to the different stories of the structure in addition to one accelerometer attached to the shaking table that measures the input accelerations. The accelerometers are connected with wires to a DAQ system called VibPilot, which, in turn, is connected to a computer to record the response. For each sensor, the sensitivity is 100 mV/g and a weight of 0.95 oz, while the sampling frequency is adjusted to be 1024 Hz with anti-aliasing filter on.

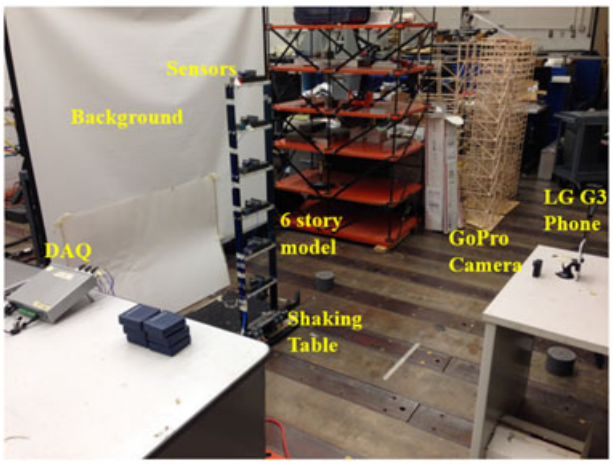

Figure 4. Different components for the validation experiment.

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

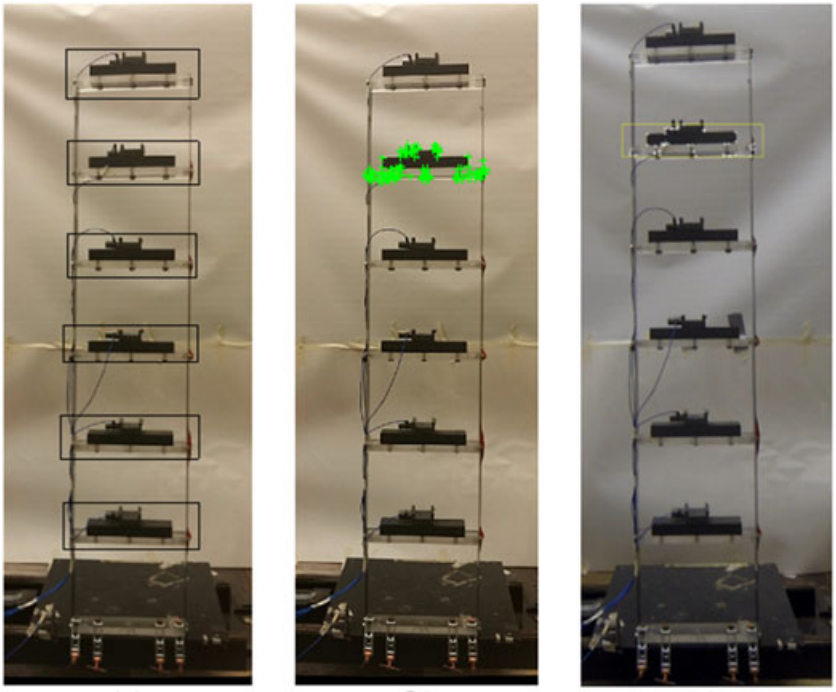

Figure 5. The computer vision based discplacement measurement procedure; (a) the region of the interest selection, (b) feature detection, and (c) tracking features within each region of the interest.

In order to excite the different structure modes of vibration, a band-limited white noise (BLWN) is adopted as an input motion in this experiment, which can be designed to cover a wide range of frequencies.

3.2. Displacement Results

Based on the described test setup, the experiment is executed, and the recorded videos are analyzed. As mentioned before, the KLT tracker is utilized to track the features from the model during the analysis. The procedure adopted to determine the displacement from the recorded video can be summarized as follows:

1. The video is calibrated to remove the optical distortions for accurate displacement measurement.
2. The user specifies the region of interest for each DOF in the model as shown in Figure 5a. Afterwards, the Harris corner detection is applied to detect the features within each of the specified regions, Figure 5b.
3. The KLT algorithm tracks the detected features and determines their pixel coordinates (motion trajectories) throughout the subsequent frames in the recorded video, Figure 5c.
4. The MLESEC algorithm is used to remove the outliers and calculate the rigid motion of the object. The determined coordinates are subtracted from the initial values to calculate the relative motion of the features.
5. A scaling factor was calculated to transform the units of the displacements from pixels to physical unit, which can be obtained by measuring a known length in the model. In this experiment, the scaling factor was 22 pixel/inch.
6. A band-pass filter is applied to eliminate any unwanted noise that might be induced during the testing process. In this experiment, the elliptic filter with band-pass width of 1 to 120 Hz was applied.

Using the proposed procedure, the displacement response are plotted against the time series for the six stories in the analyzed model as shown in Figure 6. In these results, the model was subjected to the BLWN input, and GoPro camera was used to record the video. Using the measured displacement by

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

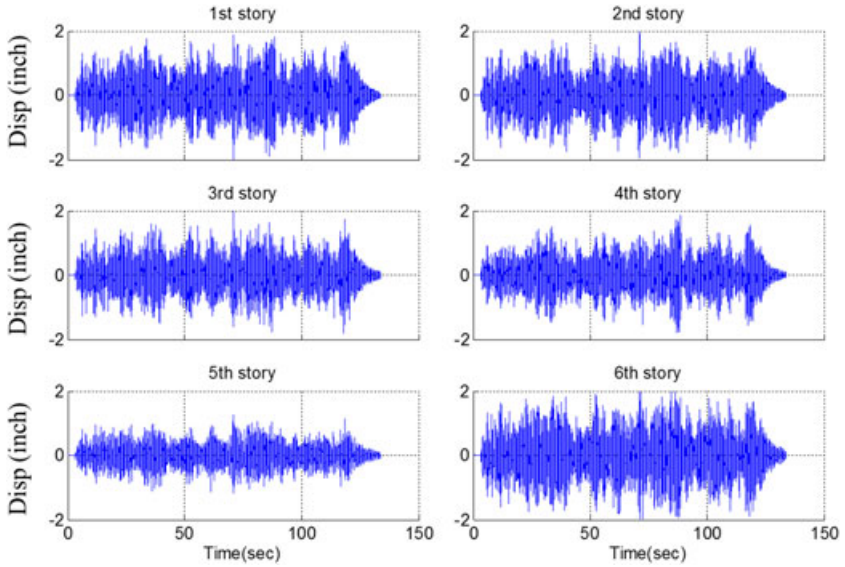

Figure 6. The displacement response by band-limited white noise for the six-story building.

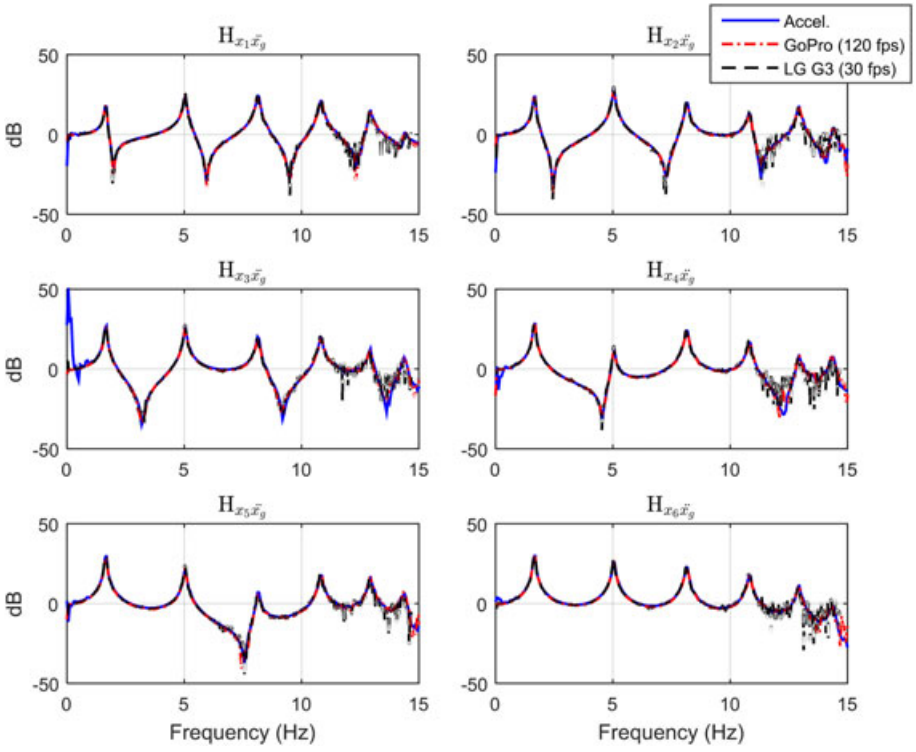

Figure 7. Comparison between the transfer functions using different measuring instruments (Figure best seen in color).

proposed method, frequency response function (FRF) was calculated as shown in Figure 7. The lower frequency ranges of the FRF obtained from the proposed vision-based system showed more accurate result compared with the measurement by reference sensor (accelerometer). On the other hand, the higher frequency ranges of the FRF obtained from the proposed system had some noises (especially for G3), because low frame rate of the cameras (which can be corresponding to sampling rate of the sensor measurement) will introduce some noises at the higher frequency ranges caused by temporal aliasing. The temporal aliasing issue will be discussed more in detail in the next section.

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

3.3. System Identification results

In this section, a system identification technique is applied to the six-story model when subject to the BLWN input motion. Both of the consumer-grade cameras (GoPro Hero3 camera 120 fps and the LG G3 smartphone camera 30 fps) were used to document this experiment. The results are then compared with the reference accelerometers attached at each DOF of the model as described in the test setup section. The accelerometers recorded the data at a rate of 1024 Hz to achieve a reliable system results relative to the both cameras. In addition, anti-aliasing filter was applied through the data acquisition system to eliminate the effect of frequencies above the Nyquist value. The state space matrix and the modal properties of the system are evaluated using the ERA method. The displacements determined from the two cameras and the accelerations measured using the accelerometers are utilized for comparison purposes.

According to the state space matrix identified for the tested model, the natural frequencies and the mode shapes can be determined. Table I shows the natural frequencies using the different instruments. In these results, the accelerometers are assumed to provide the most reliable results. The cameras were able to capture the natural frequencies within errors of about 0.2% and 0.7% for the GoPro and LG G3 cameras, respectively. It can be observed that the GoPro camera achieved more accurate results compared with the LG G3 camera for the six natural frequencies. This results is due to the higher recording frame rate of the GoPro camera, which provides more reliability to the estimated system.

Figure 8 compares the mode shapes of the structure estimated using ERA method. It can be observed from the figure that, except the sixth mode, which could not be captured accurately by LG G3 due to the low frame rate, the other mode shapes follow a similar pattern and have relatively close magnitudes for the six DOFs. A closer examination reveals that the mode shapes using LG G3 (1080p) has better accuracy (Avg. RMS error of 0.08) than the one generated with the GoPro (720p) measurements (Avg. RMS error of 0.11). This result can be justified because the higher camera resolution

Table I. The natural frequencies for the different measurement cases.

	Natural frequencies (Hz)			Error (%)	
Mode	Reference	GoPro	LG G3	GoPro	LG G3
1	1.657	1.660	1.652	0.164	−0.323
2	5.038	5.038	5.004	−0.001	−0.684
3	8.138	8.143	8.086	0.065	−0.639
4	10.833	10.834	10.759	0.004	−0.689
5	12.930	12.931	12.850	0.008	−0.623
6	14.339	14.368	14.303	0.205	−0.252

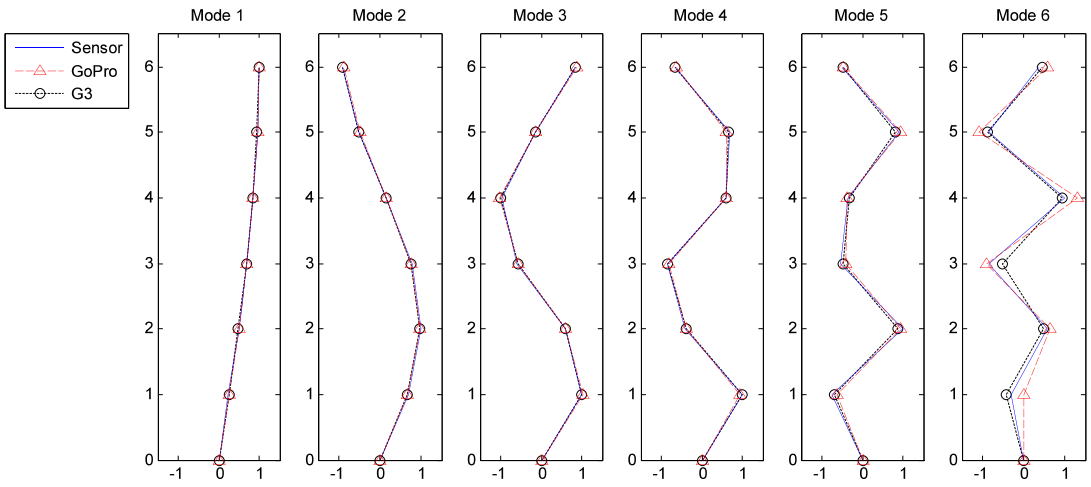

Figure 8. Comparison of the mode shapes estimated using eigensystem realization algorithm method.

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

allows the tracking algorithm to detect the vibration amplitude more precisely. Therefore, it can be concluded that the higher frame rate results in better accuracy in the temporal domain results (e.g., natural frequency), and the resolution of the pixel provides better result in the spatial domain (e.g., displacement amplitude and mode shape).

4. DISCUSSIONS

This paper presented a new computer vision-based method toward identifying the dynamic characteristic of structures, which does not require targets and can leverage commercially available cameras. Despite promising results, there are several research challenges that remain open for future investigations. These challenges include the following:

1. *Selecting the ROI Box*: Selecting the ROI is the only procedure where user input is required. However, this procedure can affect the quality of the result significantly. First, if the ROI is too large and not local (Figure 9a), then some of the features might fall outside of the boundaries of the object of the interest. If compared with the number of features in the object of the interest, an object, other than the object of the interest, has more features that are likely going to be detected and tracked, and then these features need to be removed by considering as outliers. On the other hand, if the ROI is too small (Figure 9b), then the number of the features detected inside the box will be too small, which will challenge the application of the tracking method. Finally, if the object of the interest is not rigid (Figure 9c), then the features in the object will not have a consistent motion. One thing to note is that the transformation matrix discussed in Section 2 considers projective transformation. If the object has multiple DOF, then only one mode, which has the largest numbers of features, will be tracked, and other modes will be neglected.

By considering the issues discussed earlier, the boundaries of the regions of the interest should be selected carefully by following directions. First, the size of the ROI bounding box should be selected such that the number of features in the object of interest is greater than the number of features in any other object in the ROI. Second, the size of the ROI should be selected such that at least four features (i.e., eight values) exist in the ROI considering the requirements for the projective transformation. Minimum of three set of points are required to obtain affine transformation matrix, however, considering noise and imperfection of the processes, selecting more than three points is recommended. Finally, the object of the interest should be firmly rigid.

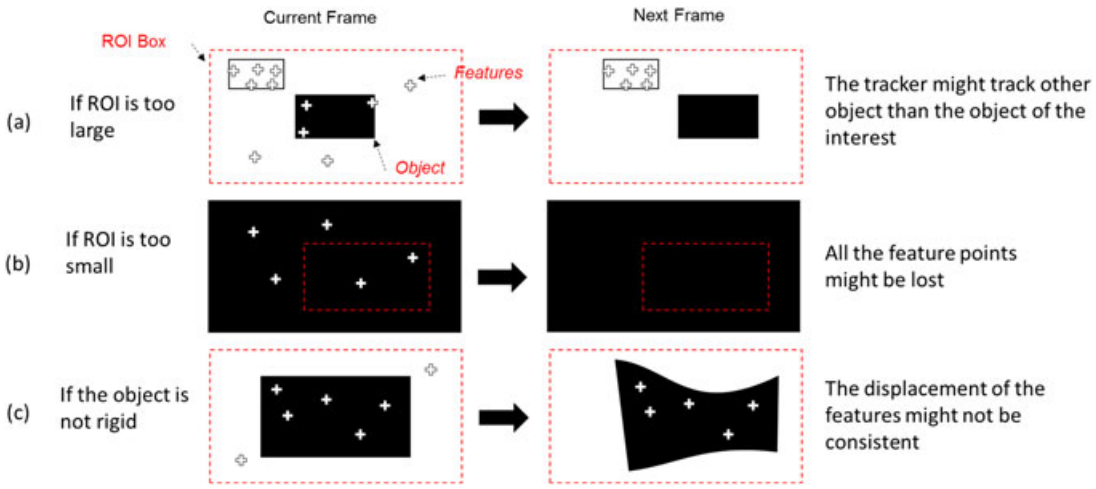

Figure 9. Practical issues associated with selecting the boundaries of the regions of the interest.

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

2 *Temporal Aliasing*: Professional high-speed cameras are currently being used for the dynamic displacement measurement necessary for structural system identification. However, these cameras are heavy and are also expensive. In contrast, consumer-grade cameras such as smartphone or GoPro cameras are inexpensive and are convenient to use. Most importantly there are widely available. While these cameras have a number of advantages, they still suffer from temporal aliasing effects. Compared with the high-speed cameras, the consumer-grade cameras usually have low frame rate (e.g., 30 or 60 fps). When structure response has frequency of higher than half of the frame rate, the measured displacement will contain aliased information from higher frequencies. The aliasing effect can be usually removed in wired/wireless sensor system by having anti-aliasing filter. However, the temporal aliasing cannot be removed in vision-based systems, because the images are already aliased when taking the digital images. Therefore, the maximum frequency of the structure response that a camera can measure will be half of the frame rate (fps or Hz). When the highest frequency exceeds the half of the frame-rate exists, a camera with higher frame will be required otherwise the measured displacement will be incorrect.

3 *Inaccurate Frame Rate*: Another issue with the frame rate in consumer-grade cameras is that not only the rate is low, but also this rate is typically inaccurate and is sometimes unreliable. The frame rate provided in camera specifications is not accurate enough to be used in the method proposed in this paper. For example, the camera on the LG G3 smartphone exhibits the frame rate of 30 fps in its specification. However, the actual/recorded frame rates in the meta-data were 29.45 fps, which reveals an offset of about 2%. The error in frame rate can introduce significant error in the measurement of the natural frequency of the structures; therefore, the frame rate should be accurately obtained by using the metadata before the proposed method is deployed.

4 *Perpendicular to the Line of Sight of the Camera*: The proposed method and the validation was performed under the assumption that the vibration of the structure is taken perpendicular to the line of sight of the camera. However, in some situations, it might not be easy to have this conditions because of obstacles and other environments. When the video is taken in such conditions, additional picture will be taken with perpendicular view, and the resulted tracked points in non-perpendicular view will be transformed into the coordinate that is perpendicular to the structure by Homographies [26].

5. CONCLUSION

This paper presented a target-free vision-based method for system identification using inexpensive consumer-grade cameras. The method is composed of three procedures: (i) camera calibration, (ii) vision-based dynamic displacement measurement, which extracts the features with Harris comer detection, tracks the features by KLT tracker, and removes the outliers by MLESAC algorithm, and (iii) system identification using the ERA. Different from the state-of-the-art vision-based system identification methods that require targets and professional cameras, the proposed method enables the dynamic characteristics of structures to be derived with consumer-grade cameras without attaching any artificial targets to the structures. To validate the proposed method, response of a six-story model in a shaking table was measured by two consumer-grade cameras together with the reference accelerometer. The experimental results show that proposed method has potential to identify the natural frequency and the mode shape with reasonable levels of accuracy. The practical considerations and limitations of the proposed system including the ROI selection, aliasing, and inaccurate frame rate were discussed. Future works include full automation process of the proposed method, utilizing the method with moving cameras and applying it to cameras mounted on moving platforms such as unmanned aerial vehicles.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site.

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

ACKNOWLEDGEMENTS

The authors are grateful for contributions to the support provided by Dr. Elnashai and Dr. Popovics. Any views, findings, conclusions, or recommendations expressed in this paper do not necessarily represent the individuals mentioned earlier.

REFERENCES

1. ASCE. Report Card for America's Infrastructure, 2013.
2. Kim J. System Identification of Civil Engineering Structures Through Wireless Structural Monitoring and Subspace System Identification Methods. Ph.D. thesis, University of Michigan: Ann Arbor, MI, 2011.
3. Fukuda Y, Feng MQ, Narita Y, Kaneko SI, Tanaka T. Vision-based displacement sensor for monitoring dynamic response using robust object search algorithm. *Sensors Journal, IEEE* 2013; **13**(12):4725–4732.
4. Spencer BF, Ruiz-Sandoval ME, Kurata N. Smart sensing technology: opportunities and challenges. *Structural Control and Health Monitoring* 2004; **11**(4):349–368.
5. Lynch JP, Loh KJ. A summary review of wireless sensors and sensor networks for structural health monitoring. *Shock and Vibration Digest* 2006; **38**(2):91–130.
6. Nogueira FMA, Barbosa FS, Barra LPS. Evaluation of structural natural frequencies using image processing. Proceedings of EVACES, 2005.
7. Chang C, Xiao X. An integrated visual-inertial technique for structural displacement and velocity measurement. *Smart Structures and Systems* 2010; **6**(9):1025–1039.
8. Khalil AM. Two-dimensional displacement measurement using static close range photogrammetry and a single fixed camera. *Alexandria Engineering Journal* 2011; **50**(3):219–227.
9. Lee J-J, Ho H-N, Lee J-H. A vision-based dynamic rotational angle measurement system for large civil structures. *Sensors* 2012; **12**(6):7326–7336.
10. Morlier J. A pedagogical image processing tool to understand structural dynamics. In Structural Dynamics, Vol. **3**. Springer: New York, 2011; 1215–1224.
11. Ji Y, Chang C. Nontarget image-based technique for small cable vibration measurement. *Journal of Bridge Engineering* 2008; **13**(1):34–42.
12. Caetano E, Silva S, Bateira J. A vision system for vibration monitoring of civil engineering structures. *Experimental Techniques* 2011; **35**(4):74–82.
13. Shih M-H, Sung W-P. Developing dynamic digital image techniques with continuous parameters to detect structural damage. *The Scientific World Journal* 2013; **2013**: Article ID 453468, 7. DOI: 10.1155/2013/453468
14. Kim S-W, Jeon B-G, Kim N-S, Park J-C. Vision-based monitoring system for evaluating cable tensile forces on a cable-stayed bridge. *Structural Health Monitoring* 2013; **12**(5-6):440–456.
15. Hild F, Roux S. Digital image correlation: from displacement measurement to identification of elastic properties–a review. *Strain* 2006; **42**(2):69–80.
16. Schumacher T, Shariati A. Monitoring of structures and mechanical systems using virtual visual sensors for video analysis: fundamental concept and proof of feasibility. *Sensors* 2013; **13**(12):16551–16564.
17. Feng MQ, Fukuda Y, Feng D, Mizuta M. Nontarget vision sensor for remote measurement of bridge dynamic response. *Journal of Bridge Engineering* 2015; **20**(12):04015023.
18. Lucas BD, Kanade T. An iterative image registration technique with an application to stereo vision. In IJCAI. Morgan Kaufmann Publishers: San Francisco, USA, 1981.
19. Zhang ZY. A flexible new technique for camera calibration. *Ieee Transactions on Pattern Analysis and Machine Intelligence* 2000; **22**(11):1330–1334.
20. Harris C, Stephens M. *A combined corner and edge detector* in *Alvey vision conference*. 1988. Manchester, UK.
21. Shi, J. and C. Tomasi. *Good features to track in computer vision and pattern recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*. 1994. IEEE.
22. Tomasi C, Kanade T. Detection and Tracking of Point Features. School of Computer Science, Carnegie Mellon Univ: Pittsburgh, 1991.
23. Torr PH, Zisserman A. MLESAC: a new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding* 2000; **78**(1):138–156.
24. Juang J-N, Pappa RS. An eigensystem realization algorithm for modal parameter identification and model reduction. *Journal of Guidance, Control, and Dynamics* 1985; **8**(5):620–627.
25. HO B, Kálmán RE. Editorial: effective construction of linear state-variable models from input/output functions. *at-Automatisierungstechnik* 1966; **14**(1-12):545–548.
26. Loop C, Zhang Z. Computing rectifying homographies for stereo vision. in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*. 1999. IEEE.

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc