Exploration of Video-Based Structural Health Monitoring Techniques

Final Report
October 2014

Ali Shariati
PhD Candidate
Center for Innovative Bridge Engineering
University of Delaware
Newark DE 19716

Thomas Schumacher (PI), PhD, PE
Assistant Professor
Center for Innovative Bridge Engineering
University of Delaware
Newark DE 19716

Nakul Ramanna (CO-PI), PhD
Research Engineer
University of Delaware
Newark DE 19716

External Project Manager
Ekin A. Senturk, PhD
HNTB Corporation

Disclaimer Statement

1. Report No. CAIT-UTC-038	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Exploration of Video-Based Structural Health Monitoring Techniques		5. Report Date October 2014	
		6. Performing Organization Code CAIT/Delaware	
7. Author(s) Ali Shariati, PhD Candidate, Thomas Schumacher, PhD, PE, Nakul Ramanna (Co-PI), PhD		8. Performing Organization Report No. CAIT-UTC-038	
9. Performing Organization, Name and Address Center for Innovative Bridge Engineering University of Delaware 301 DuPont Hall Newark, DE 19716		10. Work Unit No.	
		11. Contract or Grant No. DTRT12-G-UTC16	
12. Sponsoring Agency Name and Address Center for Advanced Infrastructure and Transportation Rutgers, The State University of New Jersey 100 Brett Road Piscataway, NJ 08854		13. Type of Report and Period Covered Final Report 8/31/13 - 9/30/2014	
		14. Sponsoring Agency Code	
15. Supplementary Notes U.S Department of Transportation/Research and Innovative Technology Administration 1200 New Jersey Avenue, SE Washington, DC 20590-0001			

16. Abstract

Structural health monitoring (SHM) has become a viable tool to provide owners with objective data for maintenance and repair. Traditionally, discrete contact sensors such as strain gages or accelerometers have been used for SHM. However, distributed remote sensors could be advantageous since they don't require cabling and can cover an area rather than a limited number of discrete points. Digital image correlation (DIC) has been used successfully but has disadvantages in terms of complexity and computational costs when applied to vides. In this research project, we investigated and proposed a different approach. The basic concept is that small changes in the intensity value of a monitored pixel with fixed (or Eulerian) coordinates caused by the vibration of structures can be captured, even if it is not invisible to the naked eye. In this research report, we explain the basic concept and mathematical theory for our proposed so-called virtual visual sensors (VVS), we present a set of laboratory experiments to demonstrate the accuracy of the approach, and provide two practical monitoring examples of in-services bridges. Finally, we discuss future work to improve the current methodology.

17. Key Words Structura health monitoring, digital videos, Eulerian coordinates, natural frequencies, structural vibrations		18 Distributional Statement	
19. Security Classification Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 40	22. Price

Form DOT F 1700.7 (8-09)

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

BACKGROUND

Introduction

The objective of this research was to evaluate a novel sensing approach for structural health monitoring (SHM) purposes which is contactless, inexpensive, and flexible in its application while addressing high computational costs of block matching algorithms such as DIC. Vibration data are important in a number of disciplines such as mechanical and structural engineering. A comprehensive review on structural health monitoring (SHM) shows the efforts put forth to estimate damage and damage location based on observed changes in natural frequencies of vibration [1]. The literature contains different resources addressing vibration-based SHM as well [2]–[8]. Finally, natural frequencies from in-service structures are often used to calibrate finite element (FE) models [9], [10].

Photogrammetry, the measurement based on visual data is a vital field subject to innovation and study among different research disciplines. Photogrammetry has also been used to measure displacements and strains in the field of structural engineering. As an example, Faghri [11] used special analog cameras, which were at the time cutting edge analog technology, to measure displacements and strains in piers and the deck of a bridge. The advent of digital video technology made the implementation of computations on the videos more convenient, enticing a wide range of innovation. To monitor bridges, Fraser et al. [12] were among the pioneers using computer vision in their integrated "decision support system" framework. In an interesting effort, Zaurin and Catbas used digital videos to measure the location as well as the approximate amount of load on a bridge [13]–[16]. They went even further by defining a unit influence line (UIL) as a measure of health of bridges [14]. Other efforts have been put forth to capture the dynamic response of bridges using video cameras. Lee et al. [17], [18] devised a system to measure displacement using digital cameras implemented with a telescopic lens and target recognition algorithms. Digital image correlation (DIC) is a block matching algorithm widely used in mechanical engineering and fluid mechanics where it is mostly known as particle image velocity (PIV) **[9]**. Both have found great interest in the civil engineering domain well [19]. Recently, the technique is finding its way to be used as an alternative to LVDT transducers [20]. Specifically, Kim et al. used DIC to measure the natural frequencies of a cable in a cable-stayed bridge to calculate the tensile force using a heuristic formula [21]. New methodologies are also being developed to make the displacement measures more accurate and feasible with reasonable computation costs for the case of structural health monitoring (SHM) of bridges [22], [23], [24],[25]. Although the application of digital videos in SHM is in its early stages, the achievements are promising. Considering the ease of use, implementation and cost efficiency of this technology, it is highly recommended to perform more research in order to develop the approach to a point where it can complement or even replace conventional sensors.

Background

The response often measured using digital videos is displacement, from which other responses and characteristics can be derived. Block matching algorithms such as digital image correlation (DIC) chose a block of pixels in one frame and search for a block with similar characteristics in the next frame. This similarity is defined using different functions which in DIC it is often cross correlation. The process is repeated by sliding the block across the entire image until the entire image has been searched. Distinctive features within a block of pixels is required for the algorithm to be able to differentiate between different regions. This is the main reason why a random speckle pattern is required on the surface in order to use DIC.

In DIC a region of interest (ROI) is identified by the user with each corner referred to as reference point as shown in Fig. 1 (a). Then the displacement is calculated at each reference point (Fig. 1 (b)) with a displacement filed associated to the rest of the image based on the calculated values at these reference points (Fig. 1 (c) and 1 (d)).
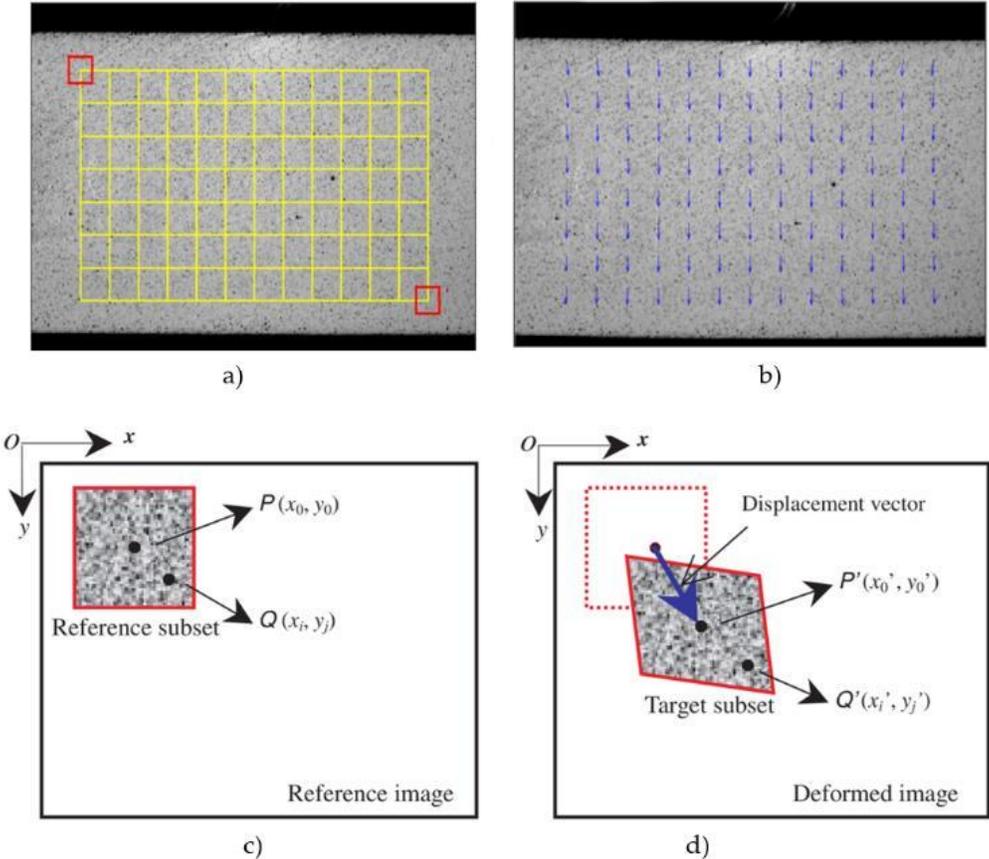
Fig 1. (a) Random speckle pattern with region of interest, (b) calculated displacements at each reference point, and (c) and (d) displacement field based on continuity assumption. Source:[26].

Another category of methods to measure displacement using camera is tracking certain features within the texture of the images and by solving the "optical flow" equation [27]In these methods, potential texture features such as a corner is extracted then the optical flow equation is solved for a patch of pixels to estimate displacement.

Using DIC, Yoneyama et al. [20] were able to measure displacement continuously along the length of a laboratory bridge girder (Fig. 2 (a)). Reportedly, they were able to quantify displacements as small as 70% of a pixel's field of view. In DIC and other block matching algorithms, sub-pixel accuracy is widely claimed but it has limitations [28]. In the application of DIC it should be noted that the computational costs of the algorithm can be quite extensive. Avoiding these costs and devising shortcuts can lead to systematic errors. For example, to achieve sub-pixel accuracy, often Newton-Raphson (NR) algorithms are used. For NR to converge, the displacement at the first reference point is calculated and associated to consecutive points. This may result in path dependent displacement fields [19]. Also, the choice of the speckle pattern is critical in making DIC continuously applicable across the object. Moreover, comparatively large displacements and rotation angles can make the detection of peaks in the correlation function difficult which will lead to implementation of coarse to fine algorithms that make the it even more computationally cumbersome [19].
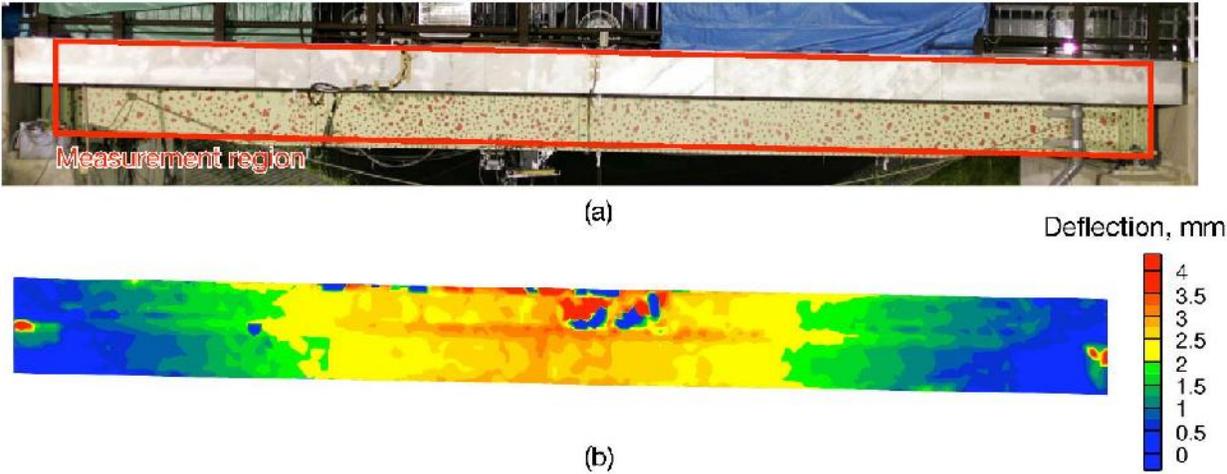
Fig 2. (a) Outline of a laboratory girder, (b) calculated displacement field. Source:.

Tracking features is another approach and has been used in monitoring of in-service bridges. Fukuda et al. [24], by defining orientation code matching (OCM), were able to measure displacement of a target point on a bridge at a distance of 300m. However, they had no physical

9

measurement to verify their displacement time history. Again, the limitations of the sub-pixel accuracy claimed in this methodology needs more theoretical and practical insight.

PROPOSED APPROACH

Fundamental Concept

We propose that every pixel in a digital video taken from a structure represents a candidate virtual visual sensor (VVS) that can be used for SHM purposes (first suggested by Patsias, S. & Staszewskiy [29]). The term "VVS" follows the terminology suggested by Song, Bowen, et al. [30]. Although the approach presented in the latter paper may appear similar, it is fundamentally different as they were employing a Lagrangian specification where a target (or feature) is tracked in space and time.
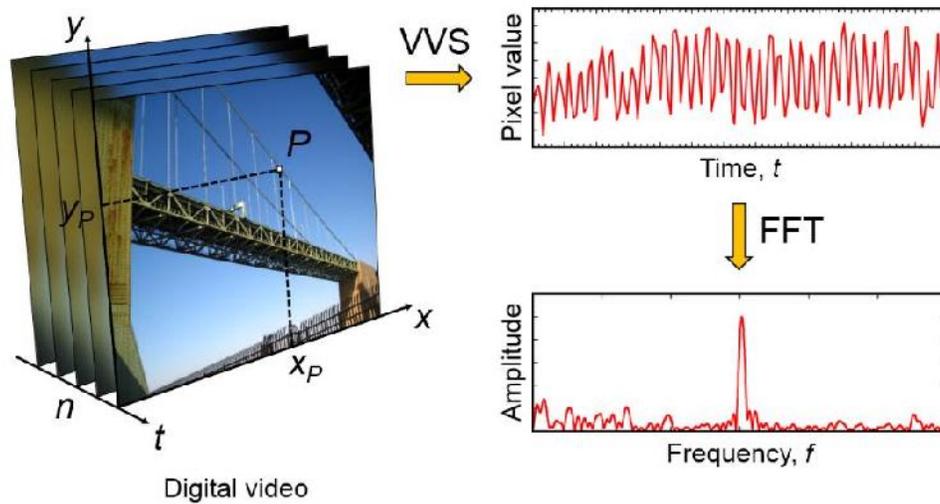
Fig 3. Proposed methodology of a virtual visual sensor (VVS) to measure structural vibrations. xp and yp represent fixed coordinates of the monitored pixel, P. "Pixel value" corresponds to the grey-scale intensity of the pixel. Source: [22]

Our proposed methodology uses an Eulerian specification where a specific pixel is selected and monitored which is illustrated in Fig. 3: the intensity of the pixel at location x_p and y_p is monitored over time and analyzed using the Fast Fourier Transform (FFT) [31], [32] to reveal the frequency of vibration. Note that the pixel value in the time domain represents gray-scale intensity and does not directly correspond to any physical entity such displacement or any of its derivatives. As a result, at this point we are not able to estimate the amplitude of vibration but only the frequency of the motion which represents a limitation.

Theoretical Basis

Digital videos are a sequence of digital images. Each digital image consists of a series of points at each x and y coordinate having three values of red, blue and green (RGB). The resolution of

digital cameras (simply the number of pixels) is an exponentially improving technology. The frame rate actually is an important key factor when measuring high frequency response. According to the Nyquist-Shannon theorem [33] the required frame-rate to measure a signal with maximum frequency content f is $2f$. This in simple terms means that if we have a camera with 30 fps, the maximum frequency we are capable of measuring is 15 Hz. This problem can be overcome by using a high frame-rate camera. As an example, apple recently announced 8 Megapixel resolution at 240 fps (= frames per second) for its iPhone 6 models. Typical frame rates of commercially available digital cameras are 25, 30, or 60 fps. In this study we used a range of cameras, including two inexpensive point-shoot cameras with 25 and 30 fps, respectively, and a new high-speed camera mainly used in the adventure sports community that can capture videos up to 240 fps. To further advance the domain of application we also tried high-speed cameras produced by Photron which will be discussed in detail later. Video frames are typically stored in RGB (red-green-blue) color mode as measured by the camera's image sensor [24]. A single grey-scale value, called intensity, I, is assigned to each pixel where 0 and 255 represent black and white, respectively. MATLAB uses a linear combination to calculate I based on RGB values that eliminates hue and saturation information while retaining the luminance [34]:

$$I = 0.2989R + 0.5870G + 0.1140B \tag{1}$$

An example of experimental data collected with a VVS and the resulting intensity curves are shown in Fig. 4. If the intensity value is smoothed using a 5-point moving average as shown in Fig. 4 (e), the quantization effects present in the raw brightness values (Fig. 4 (a), (b), and (c)) and the computed intensity value (Fig. 4 (d)) can effectively be removed to reveal a relatively harmonic motion. For this study, raw intensities (example shown in Fig. 4 (d)) were used for the computation of frequencies.
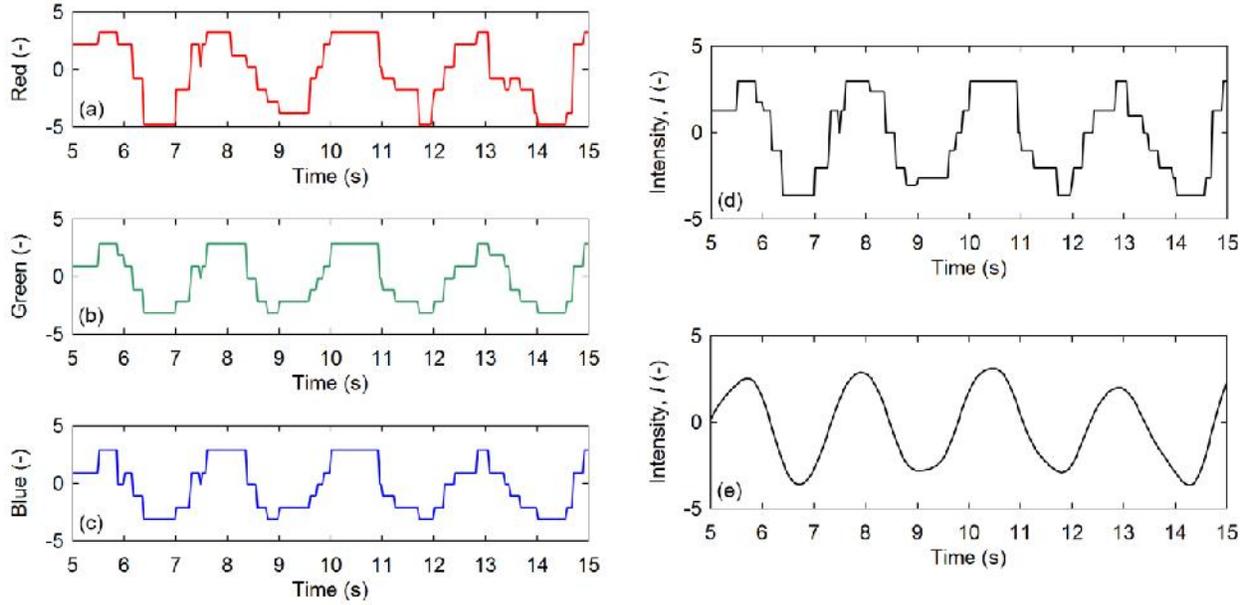
Fig 4. Example of experimental data taken with a VVS: brightness of (a) red, (b) green, and (c) blue; (d) computed intensity (used for subsequent analyses), I, and (e) smoothed intensity (for illustrative purposes). Source: [22]

Fig. 5 illustrates the factors that influence the accuracy and reliability of the proposed VVS. The dotted line represents the intensity curve $I(x)$ along a path x. For this theoretical example, the background is assumed to be light colored and the object of interest dark colored. The location and size of the monitored pixel is depicted by the grey square and denoted with $P(x,t)$. L represents the length over which the intensity changes. Fig. 5 represents a snapshot and as time t progresses the intensity curve I will oscillate horizontally (in the x-direction) and the pixel vertically, following the $I(x)$ curve.

13
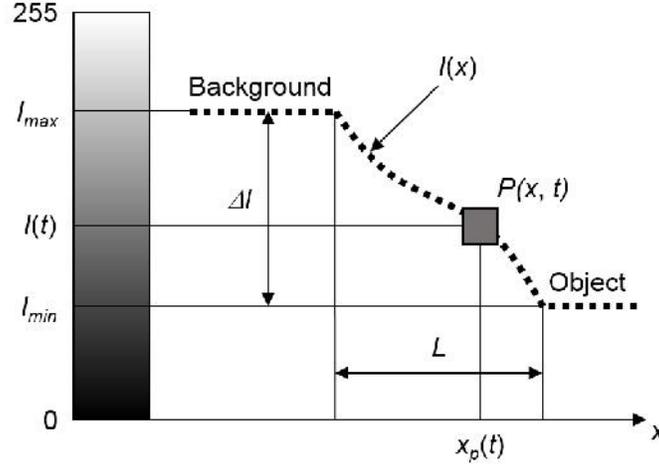
Fig 5. Illustration of the relationships between vibration of motion (one direction only for simplification), pixel location and size, and shape and location of the grey-scale intensity curve. Source: [22]

If the object is vibrating at a natural frequency ω_0 and without losing generality we can write $x_p = L\sin(\omega_0 t)$, hence the intensity value becomes $I(x_p) = I(L\sin(\omega_0 t))$. If we want to consider the effect of the intensity function $I(x)$ on our measured peak frequencies we can write:

$$F(I(x)) = \int_{-\infty}^{\infty} I(L\sin(\omega_0 t))e^{-j\omega}dt \qquad (2)$$

If $I(x)$ is a linear function by subtracting the DC term we are able to compute the exact peak frequency (neglecting any quantization noise) $I(L\sin(\omega_0 t)) = aL\sin(\omega_0 t)$ so that:

$$F(I(t)) = \frac{\pi}{j}a\left[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)\right] \qquad (3)$$

If $I(x)$ is a nonlinear function, e.g. x^n, with the identity $\sin(\omega_0 t) = \dfrac{e^{j\omega t} - e^{-j\omega t}}{2}$ we can write:

$$F(I(t)) = \int_{-\infty}^{\infty}\left(L\frac{e^{j\omega t} - e^{-j\omega t}}{2}\right)^n e^{-j\omega}dt = \left(\frac{L}{2}\right)^n \int_{-\infty}^{\infty}\sum_{k=0}^{n}(-1)^k \binom{n}{k}e^{-j(n-2k)\omega_0 t}e^{-j\omega t}dt \qquad (4)$$

Taking the sigma out of the integral we have:

14

$$F\big(I(t)\big) = 2\pi\left(\frac{L}{2}\right)^{n}\sum_{k=0}^{n}(-1)^{k}\binom{n}{k}\delta\big(\omega-(n-2k)\omega_{0}\big) \qquad (5)$$

Equation (5) shows that any nonlinearity of degree n in $I(x)$ produces peak frequencies at $(n-2k)\omega_{0}$ for $0 < k < n$. If $I(x)$ is written as a power series, the magnitude of the spurious impulses in the frequency domain can be calculated based on the above equation.
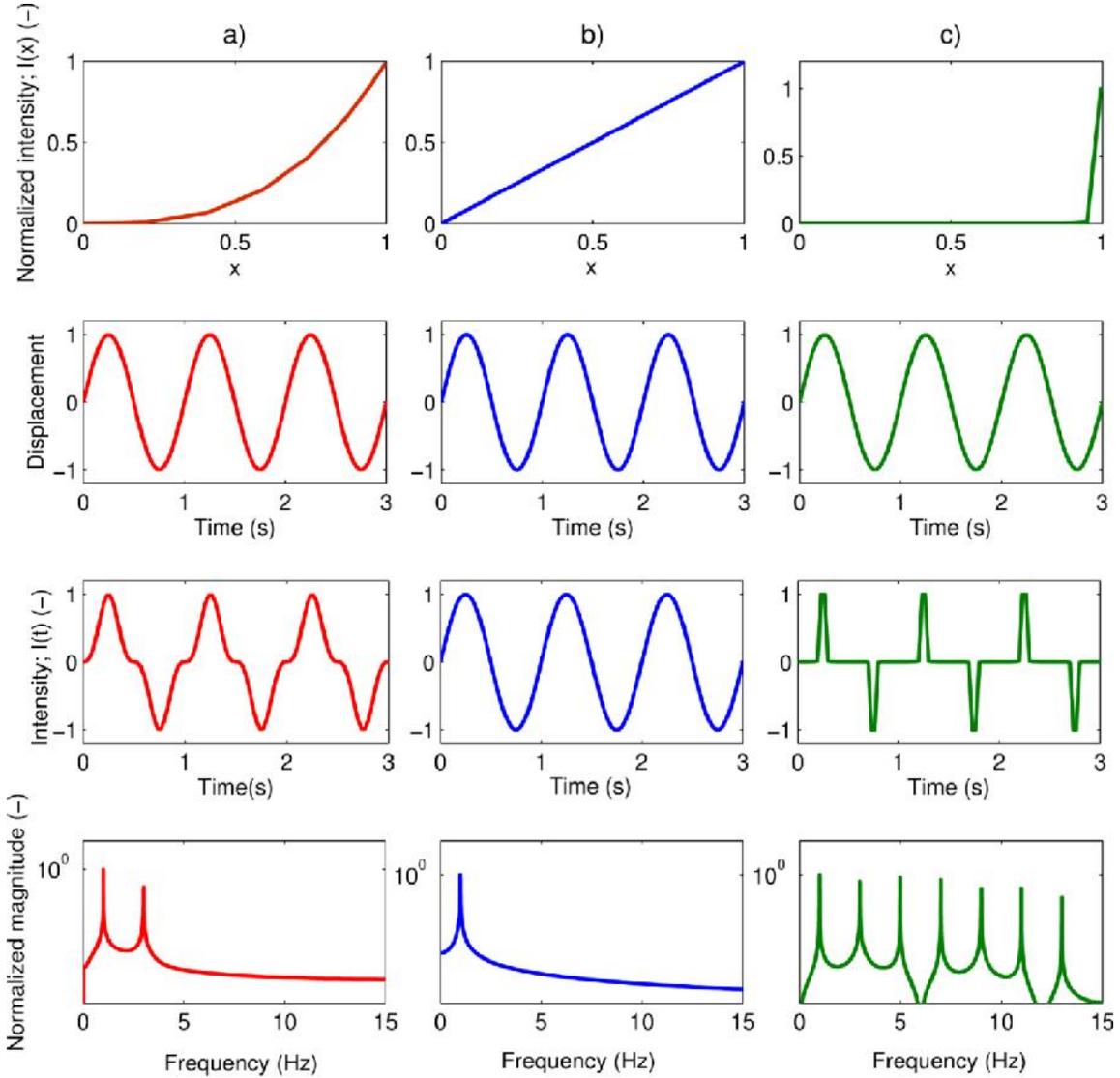
**Fig 6** The effect of non-linearity in the spatial domain. Rows one to four (top to bottom) show intensity in the spatial domain, $I(x)$, displacement, $d(x)$, observed intensity in the time domain, $I(t)$, and the Fourier transform of the former, respectively. Column a) shows the case of $I(x) = x^{3}$, b) the linear case of $I(x) = x$, and c) the case of an impulsive change of intensity modeled by $(x) = x^{99}$. Source: [35].

Fig. 6 (a) illustrates when the intensity function in the spatial domain, $I(x)$ is a third degree polynomial, $I(x) = x^3$, and the displacement follows a sinusoidal function with a frequency of 1 Hz (Fig. 6, second row) the observed intensity response in the time domain, $I(t)$ is not a sinusoidal function, as illustrated in Fig. 6 (a), third row. In the absence of noise, the Fourier transform of the intensity values have two peak frequencies at 1 and 3 Hz, as shown in the Fig. 6 (a), fourth row, which verifies Eq. 5). On the other hand, if $I(x)$ is linear, the resulting $I(t)$ is sinusoidal, as shown Fig. 6 (b). A highly non-linear case such as $I(x) = x^{99}$ results in an impulsive response of the intensity, $I(t)$ in turn leading to multiple peaks in frequency domain as shown in Fig. 6 (c). This illustrates the effect of occlusion.

From the discussion above, the following can be concluded:
- The extreme positions (peak amplitude points) of the VVS with respect to the intensity curve $I(x)$ should be located on an approximately linear portion of $I(x)$ and within L. If $I(x)$ is nonlinear, spurious frequency peaks will occur.
- The intensity range $\Delta I = (I_{max} - I_{min})$ should be maximized, i.e. a small range will increase quantization noise. This can be achieved by selecting proper background and lighting conditions.
- The number of pixels across L should be maximized which is directly related to the spatial resolution.
- The size of the VVS with respect to the length should be small to avoid averaging of measured intensity values and quantization noise.

Additionally, the following factors influence the accuracy of VVS:
- High sampling rates, i.e. large number of frames per second, decrease the quantization noise. Minimum sampling rates as given by the Shannon sampling theorem apply.
- The total signal duration T directly influences the resolution and thus the accuracy of the VVS, i.e. the resolution of a signal in the frequency domain is $\Delta f = \dfrac{1}{T}$.

Quantization Error

Analog-to-digital (A/D) conversion involves two main steps, namely: sampling in time and quantization. Errors due to quantization and their effect on the signal is a known issue addressed in the literature, e.g. in [36]. Assuming that the quantizer uniformly covers the limit values and its error is independent of the original signal, it can be deduced that the error is equivalent to an additive white noise. The white noise model, however, can be used also with high-resolution quantization, which satisfies the independence condition. In practical signal processing, in a

16

process called "dithering", some random noise within the range of quantization is added to the analog signal prior to digitization to satisfy the independence of the error from the signal [36].

In commercially available cameras, the quantization resolution to reflect the amount of absorbed energy in CCD sensors is usually 8 bit. However, as discussed previously in, this energy (the intensity value) does not correspond to any physical quantity such as displacement or any of its derivatives. In other words, higher amounts of displacement, velocity, or acceleration do not necessarily cause higher change of intensity and do not increase the quantization resolution. Assuming that the quantization error can be modeled as white noise, increasing the sampling frequency will decrease its amplitude in the frequency domain. The frame rates of commercially available cameras are in the range of 30 to 240 Hz, which is reasonably sufficient for measuring fundamental natural frequencies of major bridge components but may not be sufficient to detect all of the natural frequencies due to the high quantization error. High-frame rate cameras represent an effective yet expensive solution to this issue. In this study, we evaluate the use of gradient pattern targets mounted to the structure to significantly improve the SNR.

Patch Processing

As discussed earlier and visualized in Fig. 6, choosing one pixel in a video recorded at a comparatively low frame rate and resolution can lead to ambiguous peaks in the frequency domain which makes the detection of higher modes difficult or often impossible. To solve the problem of occlusion which produces periodic impulses in time and frequency domain, one can choose a patch of pixels and monitor their average value though time. In other words by choosing a patch of pixels, we virtually decrease the ratio of displacement to the pixel size which makes the change of intensity smoother. Patch processing can be applied to videos where no targets are used or combined with linear gradient pattern targets (LGPT) as will be discussed in more detail in the subsequent sections.

Linear Gradient Pattern Targets

In order to extend the capabilities of our proposed sensing methodology, we investigated the idea of using linear gradient pattern targets (LGPT) mounted to the structure. LGPTs with different sizes were used in the experiments as shown in Fig. 7 (a). The idea of these inexpensive and easily customizable targets, which we printed on regular white paper using a standard laser printer, is to create a well-defined, linearly-varying background to avoid non-linear behavior as discussed previously. The criteria for size is to optimize maximum amplitude of vibration, A with the length of the target, L. A typical cross section of a LGPT as it is captured and represented by the camera is shown in Fig. 7 (b). The intensity value, although designed to be linear is contaminated with

noise as shown in Fig. 7 (c), which was computed by subtracting the linear curve from the capture intensity curve. The distribution of the noise can be seen in Fig. 7 (d).
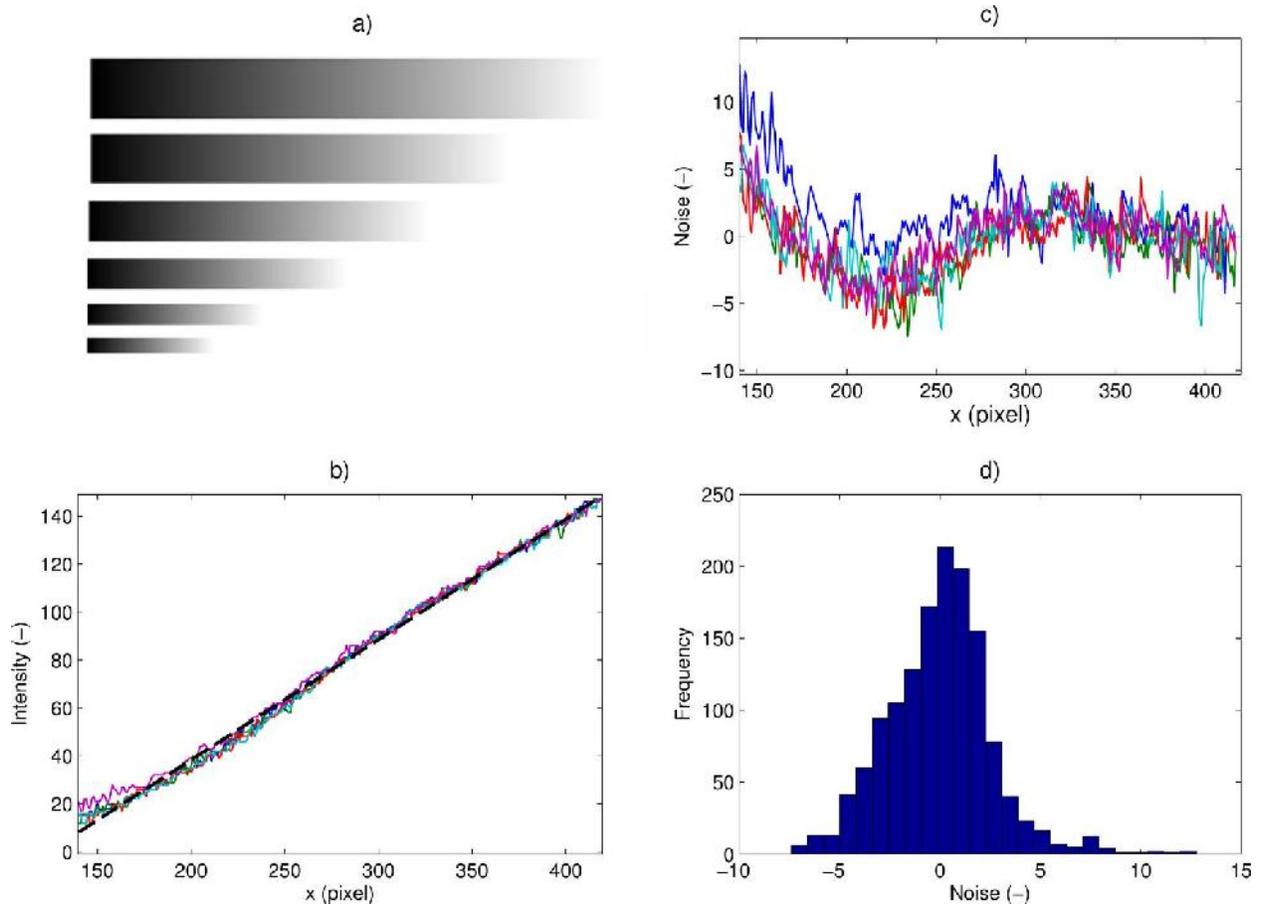
Fig 7. (a) Sample linear gradient pattern targets (LGPT), (b) intensity values captured by the camera and linear curve-fit, (c) calculated noise, and (d) the histogram of the noise. Source: [35].

Noise Reduction Strategies for LGPTs

A strategy to reduce the noise can be to average the intensities of a patch of pixels on the LGPT and average them as shown in Fig. 8 (a). From Fig. 7 (d) it can be seen that the average of the noise is close to zero so it can be deduced that by averaging the pixel values, we essentially reduce the noise. Another strategy for noise reduction can be to choose random pixels on the LGPT and fitting a linear function on them (Fig. 8 (b)). Tracking the constant part of this linear function through time can lead to a much less noisy signal (aka. smoothing). The limitation with these computationally inexpensive noise reduction techniques is that during the whole vibration phase, the chosen pixels should never leave the LGPT range, otherwise artificial non-linear behavior is introduced.
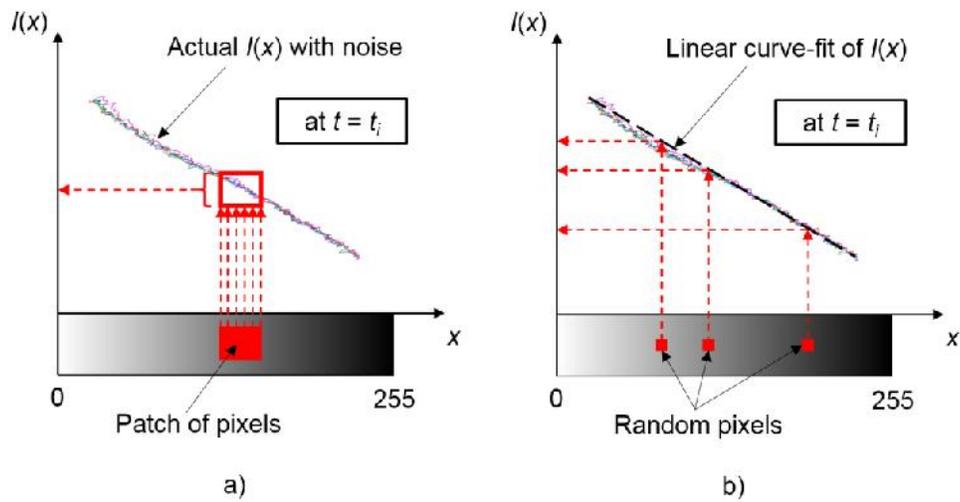
Fig 8. Noise reduction strategies for LGPTs illustrated at time instance $t = t_i$: a) Patch to average out the noise b) Linear regression approach: measured intensities are mapped onto linear curve-fit. Source: [35].

LABORATORY EXPERIMENTS

Single-Degree-of-Freedom Structure

To verify the validity and accuracy of our proposed approach, a cantilever beam with adjustable length, L_c (i.e. variable stiffness) and a concentrated constant mass on top was tested as shown in Fig. 9 (a).
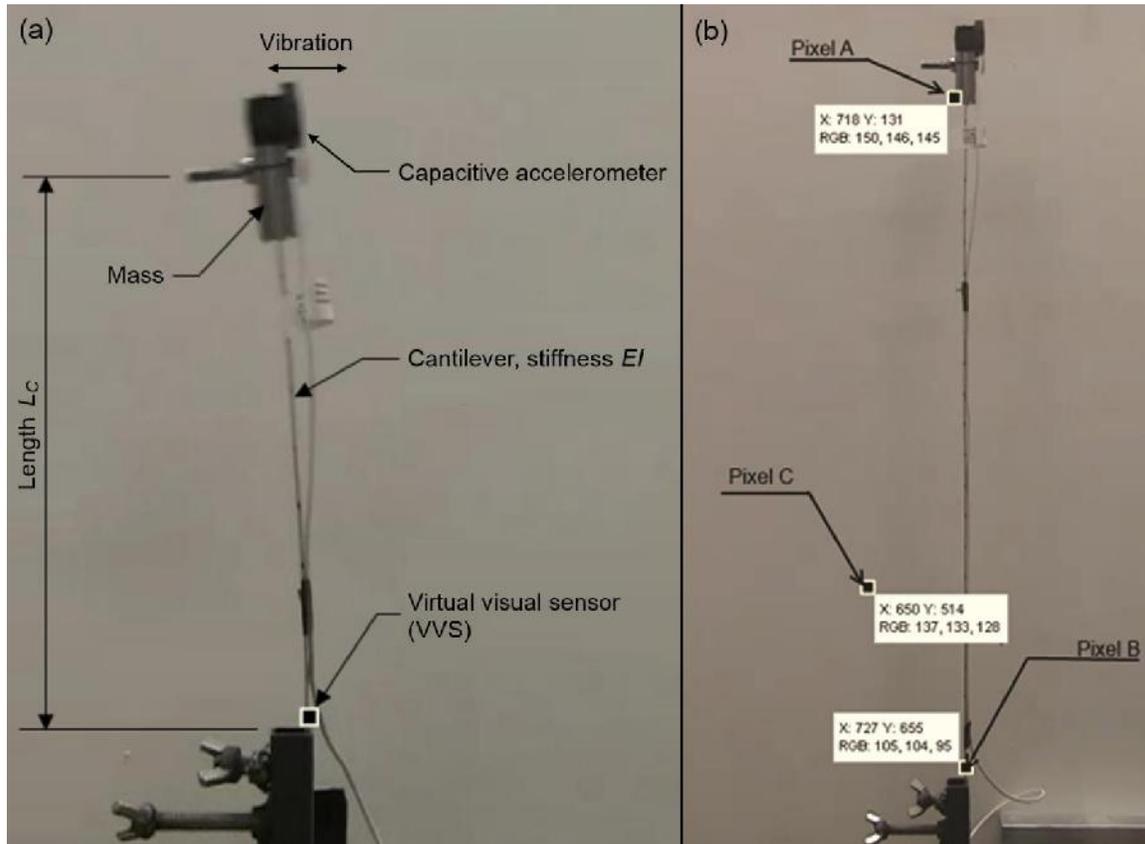
Fig 9. Snapshot of moving cantilever (a) with candidate virtual visual sensor (VVS) (b). Source: [22]

The test was initiated by creating an initial displacement (by hand) and then letting the cantilever vibrate in its natural mode of vibration. Acceleration was measured using a high-accuracy capacitive accelerometer (Model: 2260-010 by Silicon Design, sampling at 1 kHz) attached to the mass. Additionally, a digital video was taken during the test capturing the motion of the cantilever using two different cameras: a commercially available digital camcorder (Model: ViXIA HFS100 HD by Canon, recording at 30 fps) for frequencies up to 10 Hz and a relatively high-speed camera (Model: Hero 3 by GoPro, recording at 120 fps) for higher frequencies. It should be noted that, as for any digitally sampled signal, the Nyquist-Shannon [27] sampling theorem applies, i.e. the sampling rate needs to be set to at least twice the highest anticipated frequency to be distinguishable in the signal [28]. Anti-aliasing filters were set to one half of the

selected sampling frequency for the accelerometer. For the cameras, such an option is currently not available, and one of our goals was to determine whether this represents a problem.

Candidate Virtual Visual Sensors

In this section we compare and discuss the signals computed from a number of different candidate pixels. For this evaluation the cantilever length, L_c was kept constant at 25 in (635 mm). Fig. 9 (b) shows three candidate VVS: Pixels *A* and *B* are located near the top and the bottom of the cantilever where the largest and smallest displacements occur, respectively. Pixel *C* is located away from the cantilever but capturing its shadow. Intuitively one might pick pixel *A* since it is located where the largest motion takes place which should produce the best data. However, in the case of our proposed approach this does not work well as it is discussed.

Fig. 10 (a) shows example data collected with the accelerometer. Although pixel B is at the bottom of the cantilever, a place with the smallest motion which can hardly be observed by the naked eye, the change of intensity (grey scale pixel value) is represented by a relatively harmonic signal (Fig. 10 (c)). For pixel *A* located near the top of the cantilever, where the displacement is largest, the intensity value experiences periodic impulses due to the sudden occlusion of the mostly grey background by the beam. As a result, the FFT is a periodic function as well, showing pronounced harmonic peaks, as can be observed in Fig. 10 (b). Although the peak frequency is present and correct, the upper harmonics are very strong as well which makes the analysis more difficult. The peak frequency is actually the second harmonic having three times the value of the frequency of interest. Alternatively, pixel *C* is found to produce a relatively harmonic signal as well (Fig. 10 (d)). Although it is not located on the structure, it can capture the motion of its shadow. This represents an opportunity to observe vibrations indirectly, in case the actual structure is not directly observable.

21
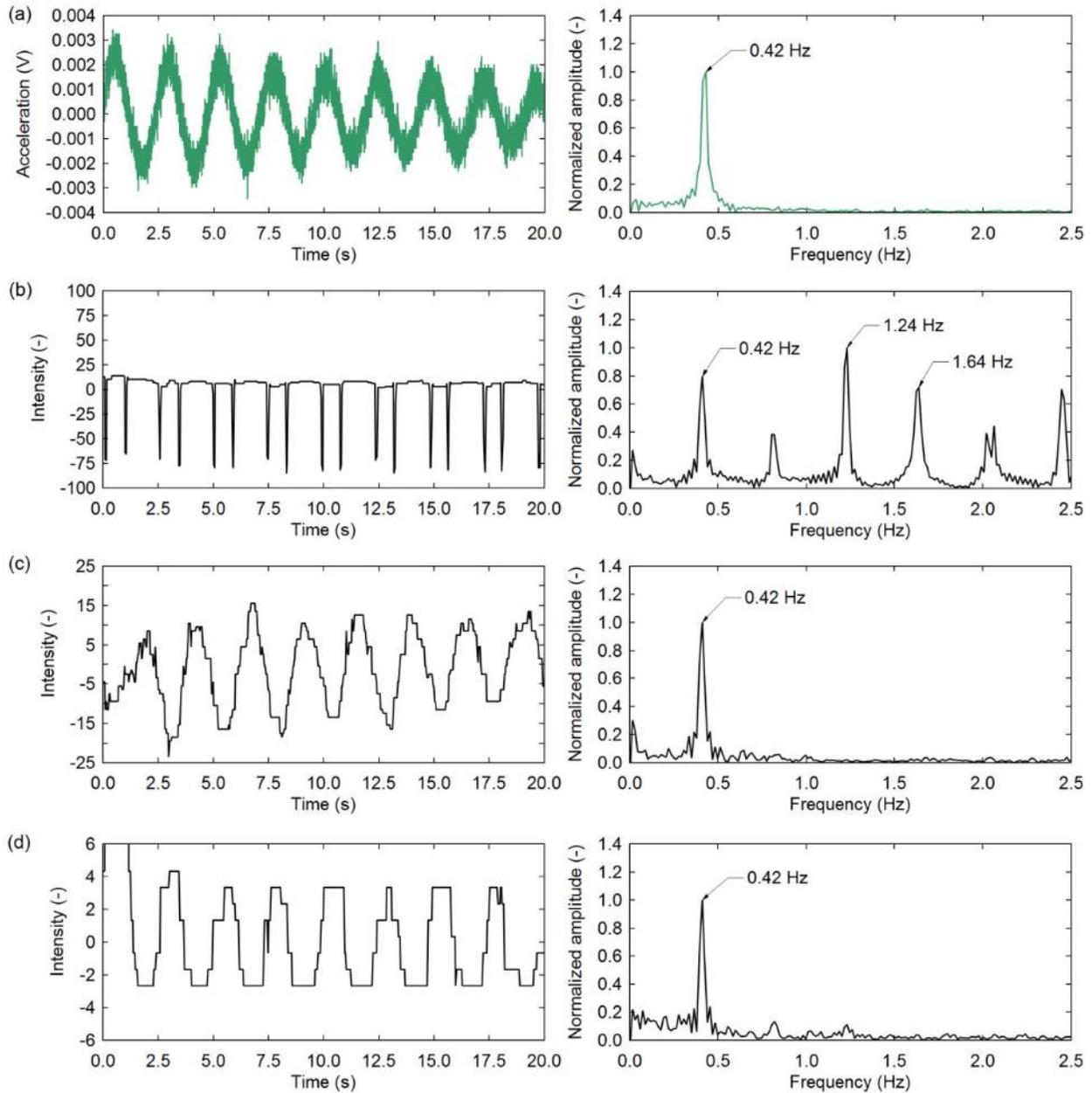
Fig 10. Time history (left column) and frequency (right column) data for (a) accelerometer, (b) pixel A, (c) pixel B, and (d) pixel C. Source: [22]

Although the total average displacement amplitude for the data shown in Figs. 10 (a) and (c) was approximately 4 in (102 mm) and 0.012 in (0.3 mm), respectively, the signal-to-noise ratios for the frequency plots are comparable. This further highlights the potential of this sensing approach to capture small vibrations.

Accuracy of Virtual Visual Sensors

In order to verify the accuracy of our proposed approach, a pixel close to the bottom of the cantilever was selected to compute the frequency as described previously and shown in Fig. 9 (a). The length, L_C was varied between 2 and 25 in (52 and 635 mm) to produce a range of natural frequencies. Fig. 11 shows the correlation between the physical accelerometer and the frequencies computed from the selected VVS. The computed frequencies listed in Fig. 11 (a) are given as $f \pm \Delta f/2$ to account for the uncertainty where $\Delta f = 1/T$ with T being the duration of the analyzed signal in seconds. Although we used zero padding to run the FFT in some cases, which will provide smoother peaks in the low frequency range, real higher accuracy is not achieved. As can be observed in Fig. 11 (b), there is excellent correlation between the frequencies computed from the two measurements. The squared correlation coefficient and standard error between accelerometer and the camera's computed frequency were found to be 99.993% and 0.0295, respectively.

(a) T	Δf	f_{ecc}	f_{VVS}
(s)	(Hz)	(Hz)	(Hz)
50.0	0.02	0.42 ±0.01	0.42 ±0.01
36.0	0.03	0.55 ±0.01	0.55 ±0.01
27.0	0.04	0.74 ±0.02	0.74 ±0.02
22.0	0.05	0.95 ±0.02	0.95 ±0.02
18.0	0.06	1.17 ±0.03	1.17 ±0.03
13.0	0.08	1.53 ±0.04	1.54 ±0.04
9.98	0.10	2.00 ±0.05	2.00 ±0.05
6.97	0.14	2.71 ±0.07	2.72 ±0.07
5.97	0.17	3.71 ±0.08	3.68 ±0.08
4.97	0.20	5.60 ±0.10	5.63 ±0.10
3.97	0.25	6.75 ±0.13	6.80 ±0.13
1.97	0.51	10.0 ±0.3	10.2 ±0.3
1.99	0.50	13.5 ±0.3	13.5 ±0.3
2.00	0.50	21.5 ±0.3	21.5 ±0.3
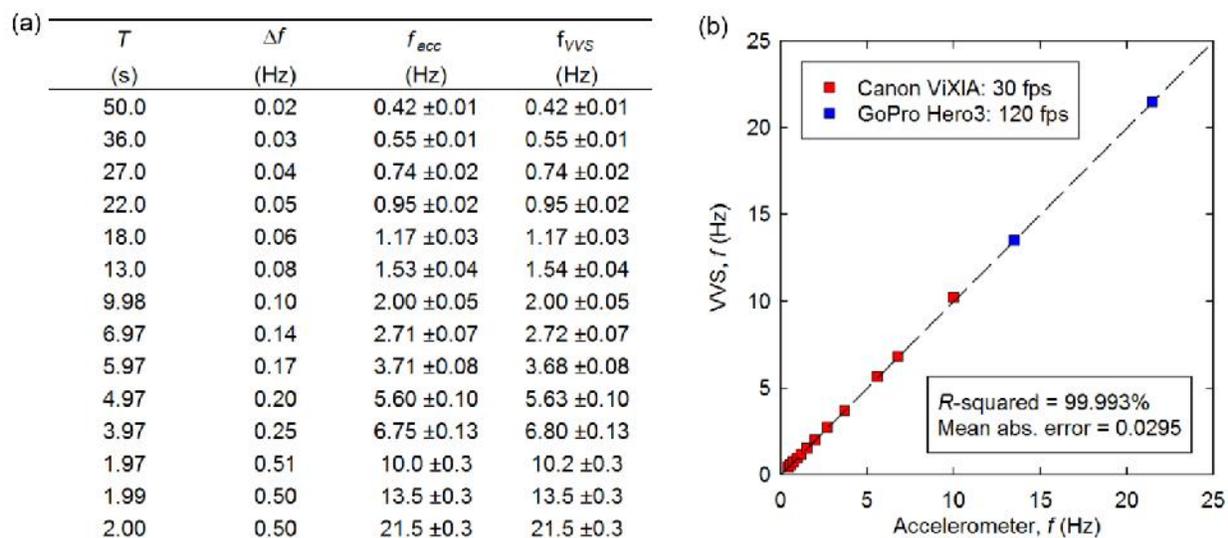
Fig 11. (a) Table and (b) plot showing correlation between physical accelerometer and virtual visual sensor (VVS). Source: [22]

Frequency Analysis over Range of Pixels

An extended approach to determine candidate VVS is to analyze a selected area of pixels in a video around the vibrating structure and then highlighting the pixels that have the same distinct peak frequency in the image as illustrated in Fig. 12. It should be noted that this only works well for small amplitudes of vibration for reasons discussed earlier. This involves the following steps:

(1)   Select range of pixels to be analyzed within video (shown as white box in Fig. 12 (b) and (c)).

(2)   Compute time history of intensity values for each of the selected pixels.

(3)   Compute the peak frequency for each pixel as described and create a histogram.

(4)   Highlight the pixels with the same peak frequency in the selected range (Fig. 12 (b)).

(5)   Normalize the color values with the magnitude of the FFT transform to reduce noise (optional, Fig. 12 (c)).

(6)   Repeat steps 1 to 5 if more than one distinct frequency peak is present in the histogram.

As can be observed from Fig. 12, the result of this analysis is essentially an image of the outline of the vibrating structure. Note that this was done for a period where the cantilever was experiencing small displacements to avoid problems as discussed in the previous section. This analysis could also be used to average peak frequencies from several measurements rather than using one measurement.
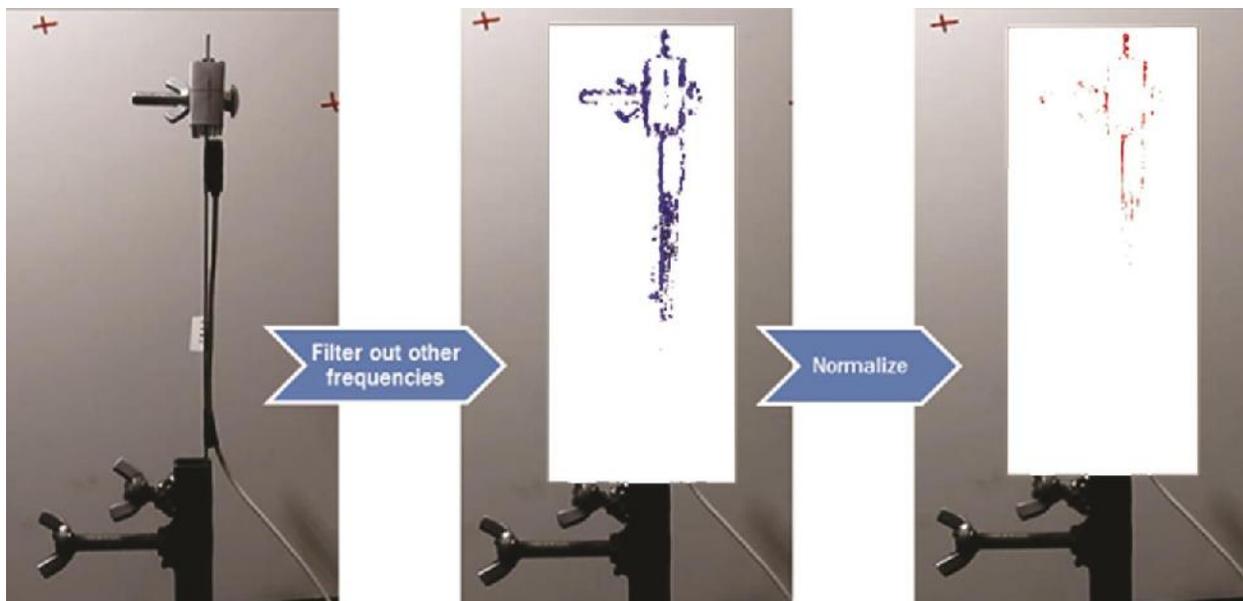
Fig 12. (a) Snapshot of cantilever, (b) snapshot with highlighted pixels of same peak frequency, and (c) snapshot with normalized highlighted pixels of same peak frequency. Source: [22]

Multi-Degree-of-Freedom Structures

Experimental Test Setup

Two experiments were conducted: Free vibration of the lab-scaled three story system as shown in Fig. 13 (a) and vibration of a simply-supported steel beam due to a hammer strike (Fig. 13 (b)). In both experiments, accelerometers were attached to the structures to give an independent measure as a comparison for the frequencies estimated from the VVS data. In the free vibration test, an initial displacement was manually imposed on the structure by hand. Following a sudden release; the system's free vibration was then recorded until it had damped out. A 142 in (3.6 m) long simply-support steel beam with a W15x87 cross-section was stroke with an instrumented

hammer to impose structural vibrations. This second test follows an NDT technique referred to as impulse response method as described in, e.g. [37].
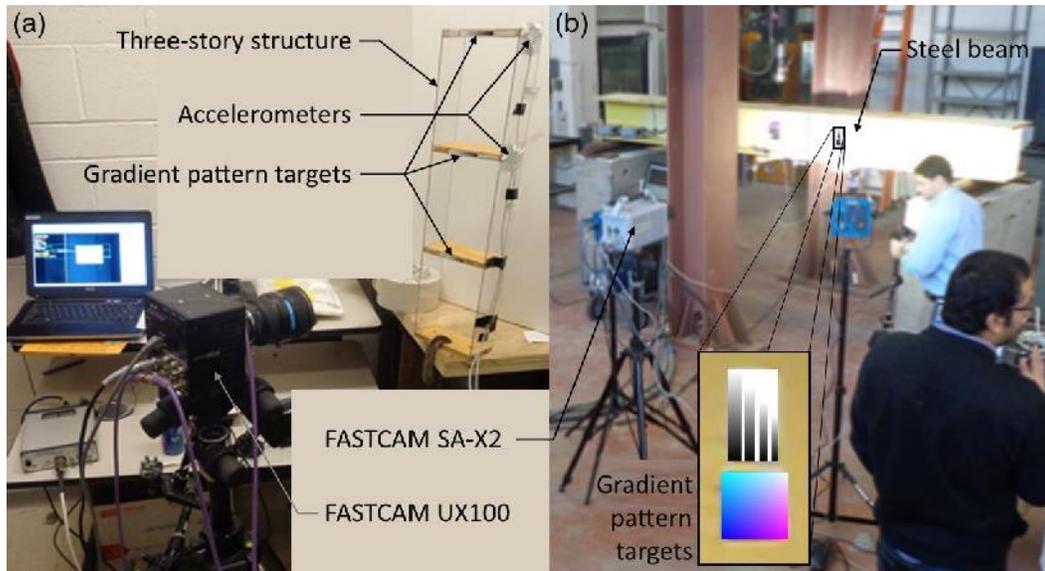
Fig 13. Experimental test setups for: (a) Three-story structure and (b) steel beam. Source: [35]

Cameras Used

In these lab experiments, three different cameras were used. For the three-story structure experiment, a GoPro Hero 3 camera and a Photron UX100 (Fig. 14 (a) and (c)) were used. The resolution of the GoPro camera was 1280 x 720 pixels and the frame rate was 120 fps. The Photron camera was used with 500 fps and its full resolution of 1280 x 1024 pixels to evaluate the ability of detecting higher-order frequencies with high-frame rate cameras. Finally, for the beam experiment, and to push to the limits of this methodology, a Photron FASTCAM SA-X2 (Fig. 14 (b)) with 5000 fps and its full resolution of 1024 x 1024 pixels was employed. It is important to note that there is a trade-off between resolution and frame rate due to the bandwidth limit of the camera hardware. Also, based on the rules of thumb and the experience of our own experiments, we found that the spatial noise in high-frame rate cameras is relatively high, i.e. higher than in regular cameras.
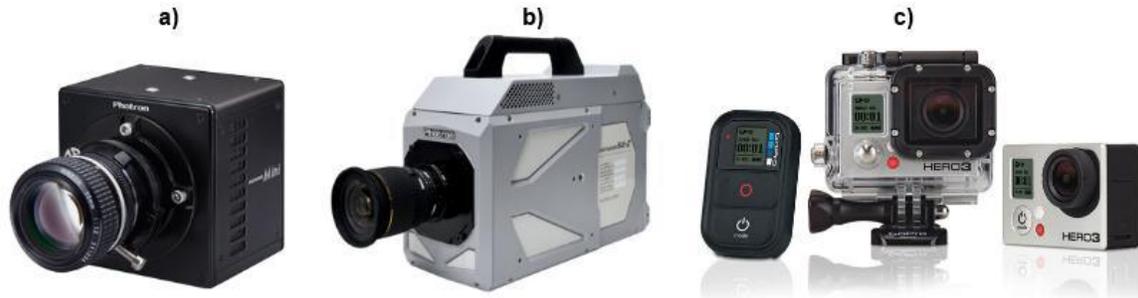
Fig 14. Cameras used for the laboratory experiments: a) Photron FASTCAM UX100, b) Photron FASTCAM SA-X2 and c) GoPro Hero 3. Source: [35]

Reference Data from Accelerometers

Three LGPTs with dimensions 0.315 x 2.36 in (8 x 60 mm) were attached to the three different masses of the three-story structure as shown in Fig. 13 (a). In addition, two high-fidelity capacitive accelerometers were attached to the side at the height of masses two and three (see Fig. 13 (a)).
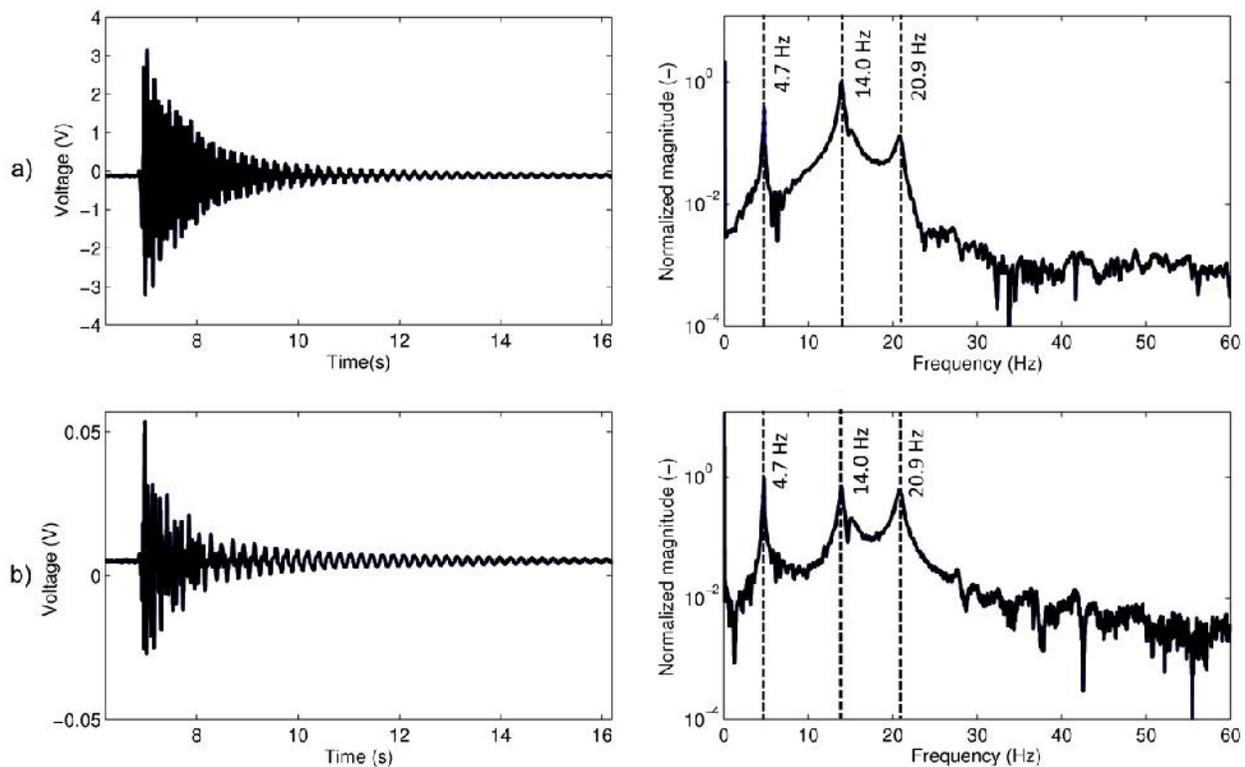
Fig 15. Sample data from the accelerometers: (a) Second story and (b) third story. Top and bottom rows show data in the time and frequency domain, respectively. Source:[35].

The natural frequencies computed from the acceleration data from the second and third story was essentially the same for all of the experiments, as shown in Fig. 15 (a) and (b), respectively. The only difference in the frequency domain was that the magnitude of the peaks slightly varied. This,

however, had no influence on the value of the peak frequency. The fundamental frequencies of vibration were found to be 4.7 Hz, 13.9 Hz and 20.8 Hz.

Results from GoPro Camera

Fig. 16 shows the results from the GoPro Camera without the use of LGPTs. In Fig. 16 (a) it can be seen that change of intensity is comparatively impulsive which led to several close peaks in the frequency domain and made the second and third peaks impossible to detect. Even a pixel at the very bottom of the three-story structure produces multiple peaks in the frequency domain (Fig. 16 (b)). Using a patch of 16 x 16 pixels it was possible to detect all of the natural frequencies. Although the magnitude of the third mode is not very large but it is still detectable (Fig. 16 (c)).
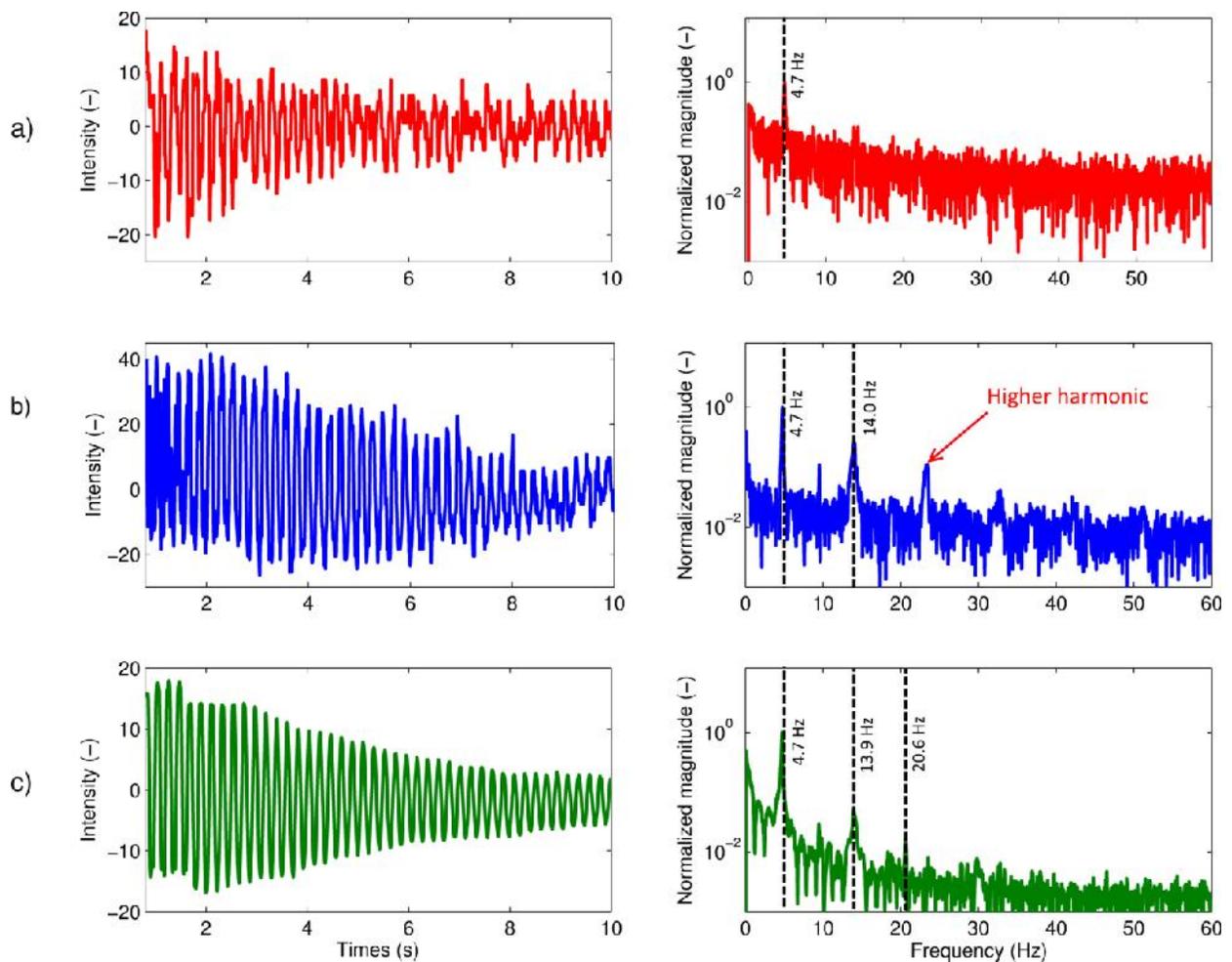
Fig 16. Data from videos of the GoPro camera without LGPTs: (a) One pixel in the middle of the first floor, (b) one pixel at the bottom, and (c) a 16 x 16 patch of pixels in the middle of the first floor. Source: [35]

As mentioned, the LGPTs should significantly improve the results in Fig. 16. By selecting a pixel on the LGPT of the first floor, all three natural frequencies could be recovered as is evident in Fig. 17 (a). Fig. 17 (b) shows the same data processed using a patch of 5 x 5 pixels on the LGPT as presented before which noticeably reduced the spatial noise. Fig. 17 (c) shows the data when the linear regression approach is employed. As can be observed, this processing step is capable of reducing the noise even better than the patch if applied for the case when LGPTs are used.
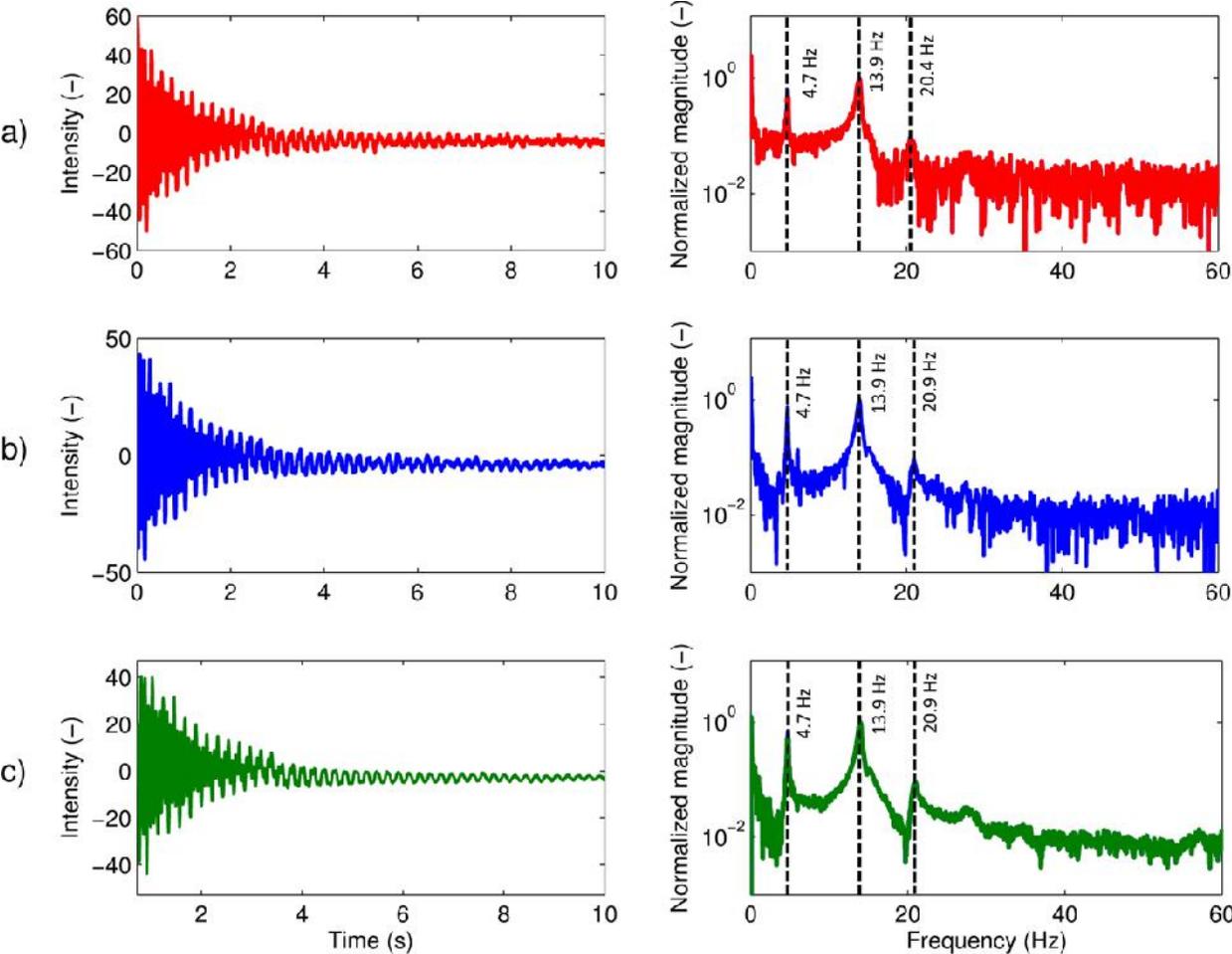
Fig 17. Data from GoPro camera using LGPTs: (a) One pixel on the LGPT, (b) a patch of 5 x 5 pixels on the LGPT, and (c) linear regression applied to 10 randomly chosen points on the LGPT. Source: [35]

It should be noted that the use of LGPTs improves the contrast and decreases quantization intervals, simultaneously. Comparing the time history part of Figs. 16 and 17 shows that the range of change in intensity values is almost twice when LGPTs are used. Also, the time history data in Fig. 17 (c) resemble an actual displacement trace with exponential decay. This, as previously

mentioned, helps reducing the effect of quantization noise, which is partially responsible for obscure signals in frequency domain.

Results from Photron FASTCAM U100 Camera

As discussed before, high speed cameras can help improve the signal-to-quantization noise ratio. For high-speed cameras, picking a pixel at the bottom of the three-story structure without any noise reduction strategy can reveal all of the natural frequencies (Fig. 18 (a)). Selecting a patch of pixels on the other hand will give a much less noisy signal and avoids presence of artificial peaks in the frequency domain as shown in Fig. 18 (b). As described before, the Photron FASTCAM U100 was used for this experiment. The problems associated with these types of cameras are their limited storage, which leads to shorter recording time, limited bandwidth, which results in sacrifice of spatial resolution with higher temporal resolution, and also higher spatial noise. Also, the higher the frame rate the brighter the medium should be in order to have high quality videos.

The use of LGPTs in conjunction with high-speed cameras can be beneficial as well: as can be seen from Fig. 19 (c), although the signal is noisy, the peaks are more pronounced. Using a patch of pixels on the LGPT (Fig. 18 (d)) reduces the noise and shows the peaks even clearer. Qualitatively, the trend of SNR here is improving from top to the bottom. Similarly to the case of the GoPro camera, the best SNR is associated with the linear regression technique (Fig. 18 (d)).
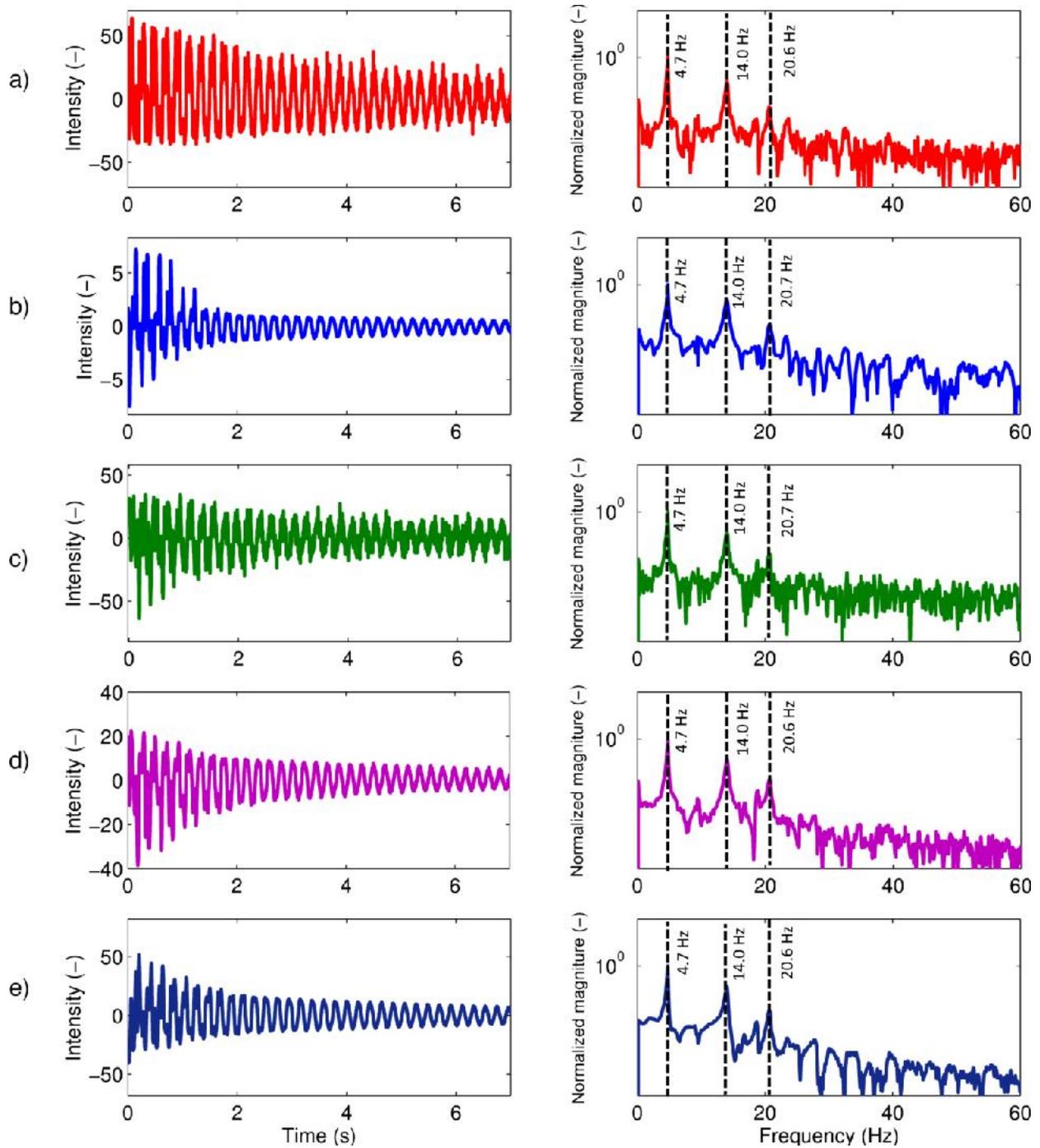
Fig 18. Data from FASTCAM UX100 camera with and without LGPTs: (a) One pixel at the very bottom, (b) a patch of pixels at the mid-level of the first floor, (c) one pixel on the LGPT, (d) a 5 x 5 pixel patch on the LGPT, and (e) linear regression applied to 10 random pixels on the LGPT. Source: [35]

Results from Photron FASTCAM SA_Z Camera

Finally, in order to push the limits of the proposed methodology, we conducted a test on a steel beam. The stimulus was a hammer impact. In this test, as it is shown in Fig. 19 (b) and (c), several peak frequencies were deducible from the accelerometer data. Monitoring a patch of pixels at the boundary of the steel beam where the gradient of the intensity is maximum (the edge), it was possible to detect several peak frequencies in congruence with the measured accelerometers' peaks.
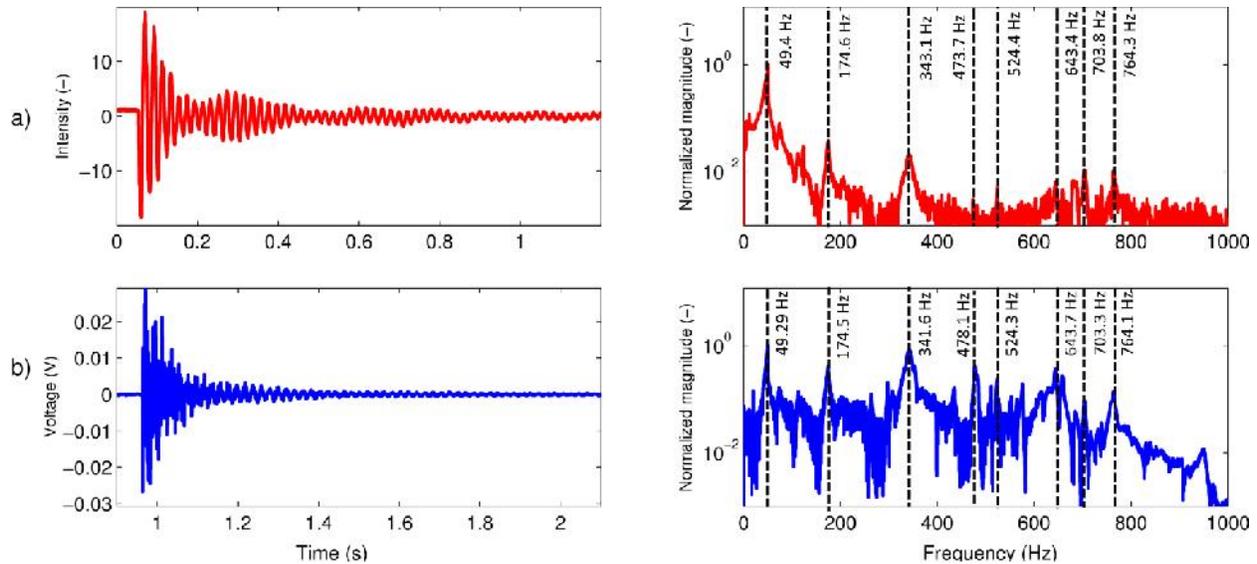
Fig 19. The steel bream test results: (a) Results from the FASTCAM SA_Z camera, and (b) the results from one accelerometer. Source: [35]

Fig. 19 (a) shows the peak frequencies detected by the camera, while Fig. 19 (b) show the data from the accelerometers for comparison. As can be seen, even without the LGPT by using a patch of pixels, several peak frequencies could be detected. The interesting point about this experiment is that it involves a continuous system where the high frequency displacements are extremely small, completely undetectable with a naked eye. However, it was possible to detect frequencies as high as 764 Hz using our proposed VVS.

IN-SERVICE MONITORING EXAMPLES

In order to evaluate the applicability of this method for practical purposes, our proposed methodology was applied to two different bridges in-service.

Example 1: Steel Truss Bridge in Oregon

A video of an existing major bridge in Oregon was evaluated. The bridge consists of a continuous steel truss and some of the vertical hangers have experienced extensive torsional vibrations due to the high transverse winds which caused vortex shedding. Concerns regarding fatigue at the connections have been raised and as a result, the Oregon Department of Transportation (ODOT) has recently retrofitted some of the susceptible members. The reason for the vibrations is the low torsional stiffness of the used I-sections. A recent research project has investigated the problem in the laboratory to make predictions on the remaining fatigue life [29].
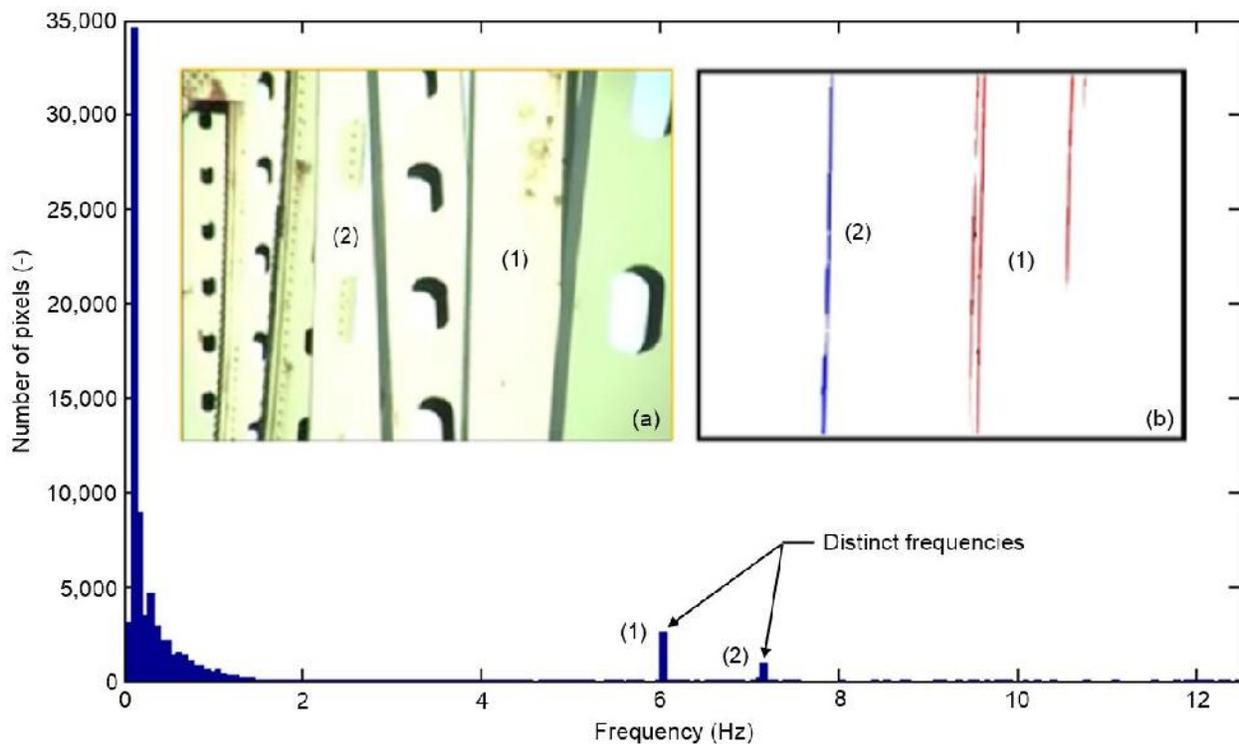
Fig 20. Histogram of peak frequencies from all pixels in the video. Insert: (a) Snapshot of video clip, (b) Colored pixels with same frequencies. Source: [22]

A four second clip from the video (made available to us by Dr. Christopher Higgins) captured severe torsional vibrations of two vertical hangers labeled (1) and (2) in Fig. 20 (a), simultaneously. For this analysis, all pixels were analyzed and their peak frequency values computed as described before. Fig. 20 shows a histogram of all computed peak frequencies. As can be observed, the majority of computed peak frequencies are close to zero which essentially

means that the majority of the pixel intensity values don't change. Frequencies between 0 and 1 can be associated with the fact that the camera was held by hand and therefore probably not completely fixed. A closer look reveals that there are two distinct frequency peaks that can be associated with the torsional motion of the two hangers.

By filtering out pixels that are not within the desired frequency range, we obtain the outline of the oscillating hangers (Fig. 20 (b)) as described in the previous section. The frequencies computed from one selected pixel for members (1) and (2) are 6.1 ±0.125 and 7.1 ±0.125 Hz, respectively. A Finite Element (FE) analysis of a hanger modeled after one of these two members [38] predicted a torsional vibration frequency of approximately 6.5 Hz. This result is not the actual measurement but proves that our computed frequencies obtained from the VVS are trustworthy. It should be noted that the movie was taken with an inexpensive point-and-shoot-type camera recording at 25 fps by hand without any mechanical stabilization. Additionally, only four seconds from the original video were usable which directly influences the resolution in the frequency domain.

Example 2: Prestressed Concrete Pedestrian Bridge
In order to evaluate the real-world performance of our proposed approach on an MDF systems, we conducted a field test on the Streicker Bridge (Fig. 21 (a)), a prestressed concrete pedestrian bridge located on Princeton University's campus in in Princeton, NJ.
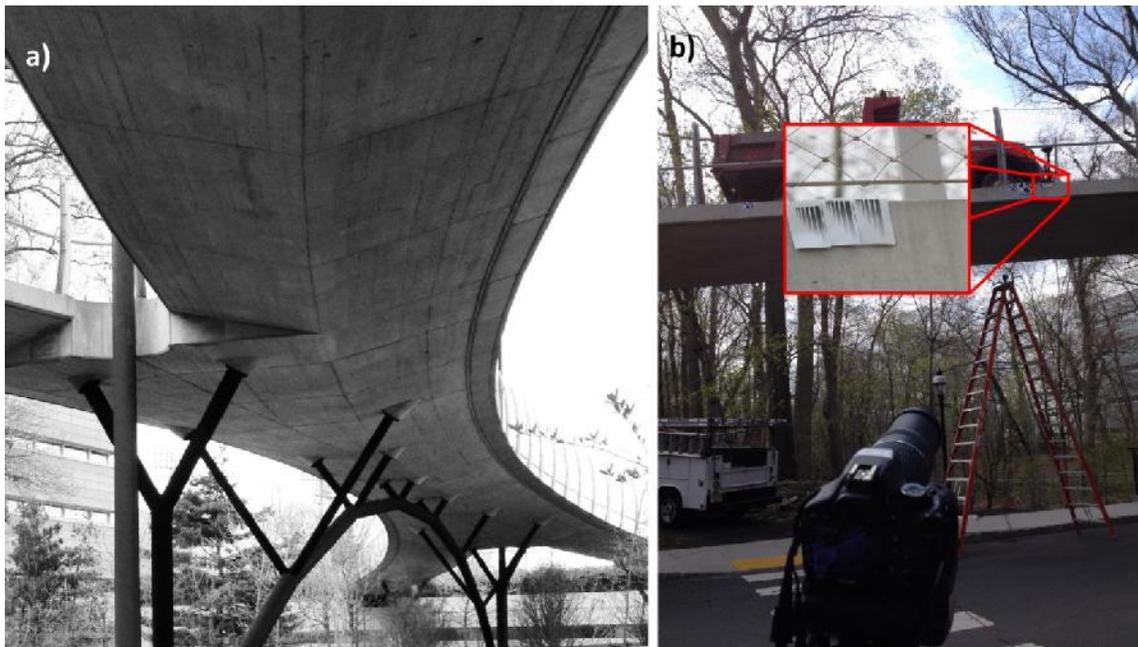
Fig 21. (a) View of the Streicker Bridge and (b) test setup, camera position and LGPTs (insert). Source: [35]

As can be seen from Fig. 21 (a), the bridge has a unique design with a main span and four horizontally curved legs. The main span consists of a deck-stiffened arch. The bridge is equipped with a structural health monitoring (SHM) system consisting of embedded fiber-optic sensors (FOS). Data from the FOS system were made available to us by Prof. Branko Glisic and allowed for a direct comparison with our measurements. The dynamic stimulus was provided in form of a group of students jumping synchronously. The test was performed on April 23, 2014, with adequate lighting conditions and some wind.

LGPTs were mounted on the inside of a curved leg to measure vertical vibrations while the cameras were on the other side of the street, approximately 26 ft (8 m) away from the LGPTs. The camera used was a Canon T4i with 60 frames per second and 128 x 730 resolution. Fig. 22 shows the results in the frequency domain for both measurements. As can be observed, the two main frequencies of vibration of the leg, namely 3.0 Hz and 3.6 Hz, were detected by both sensing approaches with comparable SNRs.
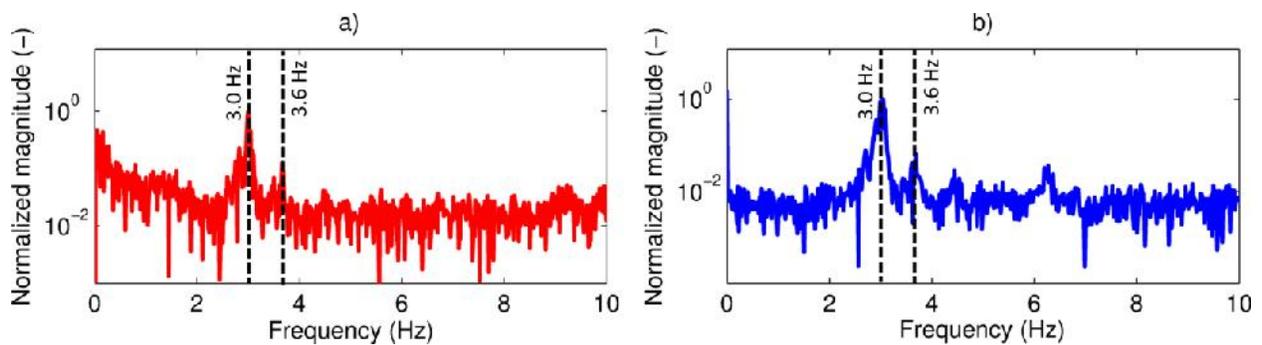
Fig 22. Frequency response of the Streicker Bridge from (a) the VVS located on a LGPT and (b) the fiber-optic sensor system data. Source: [35]

CONCLUSIONS

The concept of Eulerian-based virtual visual sensors (VVS) offers new opportunities for structural health monitoring. The following conclusions can be drawn from this research:

- The fundamental frequency of vibration of single-degree-of-freedom systems [30] can be accurately computed using the proposed methodology of VVS.
- VVS are inexpensive non-contact sensors with great application flexibility.
- Multiple independently vibrating elements in one video can be distinguished and their fundamental frequency of vibration computed.
- The accuracy and resolution of the measurements depends on a variety of factors such as sampling rate, quantization noise (function of pixel size and location with respect to intensity curve), image sensor quality and size, and lens type.
- By highlighting the pixels with a distinct frequency, the outline of the vibrating elements in a video can be recovered. This can be useful to select VVS.
- The introduction of linear gradient pattern targets (LGPT) increases the signal-to-noise-ratio (SNR) and enables detecting higher natural frequencies which is particularly helpful when ordinary cameras are used.
- Analyzing a patch of pixels rather than a single pixel can be employed when no LGPTs are used to smooth the change of intensity, i.e. minimize impulse-type response in the signal (occusion problem).
- While high-speed camera technology is still expensive and mostly used by researchers, the use of commercially-available cameras in conjunction with LGPTs allow for accurate and reliable detection of multiple natural frequencies.
- By analyzing a patch of pixels or apply a linear regression approach, the SNR of LGPTs can further be improved.
- High-speed cameras benefit from lower noise amplitude due to oversampling and are able to detect higher frequencies even without LGPTs.
- Our methodology also works in the field where we found the same peak frequencies compared to the existing structural health monitoring (SHM) system.

The findings of our studies were published in two peer-reviewed journal articles [12, 36] and one conference paper [37].

RECOMMENDATIONS

In order to further develop our proposed methodology and bring it to the point where it can be used in practice, we recommend the following additional research:

- In-depth characterization of fundamental relationships in the laboratory: resolution, frame rate, camera and lens type, distance from object, atmospheric influence, etc.
- Establishment of approach to correlate intensity with actual displacement.
- Evaluation of advanced signal processing methods to further improve the SNR and quantization noise, and adjust for variable lighting.
- Perform additional in-service tests to gain practical experience.

REFERENCES

[1] S. W. S. Doebling, C. R. C. Farrar, M. B. M. Prime, and D. W. D. Shevitz, "Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review," 1996.

[2] M. Friswell, "Damage identification using inverse methods," *Phylo Trans. R. Soc. A ...*, no. December 2006, pp. 393–410, 2007.

[3] T. Duffey and S. Doebling, "Vibration-based damage identification in structures exhibiting axial and torsional response," *Trans. ...*, 2001.

[4] K. D'Souza and B. Epureanu, "Damage detection in nonlinear systems using system augmentation and generalized minimum rank perturbation theory," *Smart Mater. Struct.*, vol. 14, pp. 989–1000, 2005.

[5] C. Fritzen, "Vibration-based structural health monitoring–concepts and applications," *Key Eng. Mater.*, vol. 294, pp. 3–18, 2005.

[6] C. Farrar and K. Worden, "An introduction to structural health monitoring," *Philos. Trans. ...*, vol. 365, no. 1851, pp. 303–315, 2007.

[7] C. Farrar and N. Lieven, "Damage prognosis: the future of structural health monitoring," *... R. ...*, vol. 365, no. 1851, pp. 623–632, 2007.

[8] D. Montalvao, N. Maia, and A. Ribeiro, "A review of vibration-based structural health monitoring with special emphasis on composite materials," *Shock Vib. Dig.*, vol. 38, no. 4, pp. 295–324, 2006.

[9] J. E. Mottershead and M. I. Friswell, "Model Updating In Structural Dynamics: A Survey," *J. Sound Vib.*, vol. 167, no. 2, pp. 347–375, Oct. 1993.

[10] D. Ribeiro, R. Calçada, R. Delgado, M. Brehm, and V. Zabel, "Finite-element model calibration of a railway vehicle based on experimental modal parameters," *Veh. Syst. Dyn.*, vol. 51, no. 6, pp. 821–856, Jun. 2013.

[11] A. Faghri, "Autostress design deformation measurements by photogrammetry and surveying," University of Washington, 1983.

[12] M. Fraser, A. Elgamal, X. He, and J. Conte, "Sensor network for structural health monitoring of a highway bridge," *J. Comput. Civ. Eng.*, no. February, pp. 11–24, 2010.

[13]   R. Zaurin and F. N. Catbas, "Integration of computer imaging and sensor data for structural health monitoring of bridges," *Smart Mater. Struct.*, vol. 19, no. 1, p. 015019, Jan. 2010.

[14]   F. N. Catbas, R. Zaurin, M. Gul, and H. B. Gokce, "Sensor Networks , Computer Imaging , and Unit Influence Lines for Structural Health Monitoring : Case Study for Bridge Load Rating," vol. 17, no. August, pp. 662–670, 2012.

[15]   F. N. C. (University of C. F.) R. Zaurín, "Computer Vision Oriented Framework for Structural Health Monitoring of Bridges," *sem-proceedings.com*.

[16]   R. Zaurin and F. Necati Catbas, "Structural health monitoring using video stream, influence lines, and statistical analysis," *Struct. Heal. Monit.*, vol. 10, no. 3, pp. 309–332, Jun. 2010.

[17]   J. Lee, Y. Fukuda, and M. Shinozuka, "Development and application of a vision-based displacement measurement system for structural health monitoring of civil structures," *Smart Struct. Syst.*, vol. 3, no. 3, pp. 373–384, 2007.

[18]   J. J. Lee and M. Shinozuka, "A vision-based system for remote sensing of bridge displacement," *NDT E Int.*, vol. 39, no. 5, pp. 425–431, Jul. 2006.

[19]   B. Pan, K. Qian, H. Xie, and A. Asundi, "Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review," *Meas. Sci. Technol.*, vol. 20, no. 6, p. 062001, Jun. 2009.

[20]   S. Yoneyama and A. Kitagawa, "Bridge deflection measurement using digital image correlation," *SEM Proceedings*, 2006. (http://sem-proceedings.com/06s/sem.org-2006-SEM-Ann-Conf-s73p02-Bridge-Deflection-Measurement-Using-Digital-Image-Correlation.pdf)

[21]   S.-W. Kim, B.-G. Jeon, N.-S. Kim, and J.-C. Park, "Vision-based monitoring system for evaluating cable tensile forces on a cable-stayed bridge," *Struct. Heal. Monit.*, vol. 12, no. 5–6, pp. 440–456, Dec. 2013.

[22]   T. Schumacher and A. Shariati, "Monitoring of structures and mechanical systems using virtual visual sensors for video analysis: fundamental concept and proof of feasibility.," *Sensors (Basel).*, vol. 13, no. 12, pp. 16551–64, Jan. 2013.

[23]   J. Chen, N. Wadhwa, and Y. Cha, "Structural modal identification through high speed camera video: Motion magnification," *Top. Modal Anal. I*, vol. 7, no. 191–197, 2014.

[24]   Y. Fukuda, M. Q. Feng, Y. Narita, and T. Tanaka, "Vision-Based Displacement Sensor for Monitoring Dynamic Response Using Robust Object Search Algorithm," *IEEE Sens. J.*, vol. 13, no. 12, pp. 4725–4732, 2013.

[25] A. Shariati and T. Schumacher, "Oversampling in Virtual Visual Sensors as a Means to Recover Higher Modes of Vibration," in *QNDE Proceeding.*Under Review

[26] P. Bing, X. Hui-min, X. Bo-qin, and D. Fu-long, "Performance of sub-pixel registration algorithms in digital image correlation," *Meas. Sci. Technol.*, vol. 17, no. 6, pp. 1615–1621, Jun. 2006.

[27] J. Shi and C. Tomasi, "Good features to track," in *Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, 1994, pp. 593–600.

[28] C. Lee, W. A. W. Take, and N. A. N. Hoult, "Optimum Accuracy of Two-Dimensional Strain Measurements Using Digital Image Correlation," *J. Comput. Civ. Eng.*, vol. 26, no. 6, pp. 795–803, Nov. 2012.

[29] S. Patsias and W. J. Staszewskiy, "Damage Detection Using Optical Measurements and Wavelets," *Struct. Heal. Monit.*, vol. 1, no. 1, pp. 5–22, Jul. 2002.

[30] Y. Song, C. Bowen, H. Kim, and A. Nassehi, "Virtual Visual Sensors and Their Application in Structural Health Monitoring," *urmassn11.iids.org*.

[31] G. Strang, *Linear Algebra and Its Applications*. Harcourth Brace Jovanovich, 1988, p. 520.

[32] G. Strang, *Introduction to Linear Algebra, Second Edition*. Wellesley-Cambridge Press, 1993, p. 472.

[33] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 86, no. 2, pp. 447–457, Feb. 1949.

[34] "Convert RGB image or colormap to grayscale - MATLAB rgb2gray." [Online]. Available: http://www.mathworks.com/help/images/ref/rgb2gray.html. [Accessed: 31-Jul-2013].

[35] A. Shariati, T. Schumacher, and N. Ramanna, "Strategies for processing data from virtual visual sensors to detect higher-order frequencies of vibration," *Meas. Sci. Technol.*,Under Review

[36] B. Widrow and I. Kollár, "Quantization noise," *Cambridge Univ. Press*, 2008.

[37] A. G. Davis, "The nondestructive impulse response test in North America: 1985–2001," *NDT E Int.*, vol. 36, no. 4, pp. 185–193, Jun. 2003.

[38] P. Keller, "Wind induced torsional fatigue behavior of truss bridge verticals," Oregon State University, 2012.