# Video Camera–Based Vibration Measurement for Civil Infrastructure Applications

Justin G. Chen[1]; Abe Davis[2]; Neal Wadhwa[3]; Frédo Durand[4]
William T. Freeman[5]; and Oral Büyüköztürk[6]

**Abstract:** Visual testing, as one of the oldest methods for nondestructive testing (NDT), plays a large role in the inspection of civil infrastructure. As NDT has evolved, more quantitative techniques have emerged such as vibration analysis. New computer vision techniques for analyzing the small motions in videos, collectively called motion magnification, have been recently developed, allowing quantitative measurement of the vibration behavior of structures from videos. Video cameras offer the benefit of long range measurement and can collect a large amount of data at once because each pixel is effectively a sensor. This paper presents a video camera-based vibration measurement methodology for civil infrastructure. As a proof of concept, measurements are made of an antenna tower on top of the Green Building on the campus of the Massachusetts Institute of Technology (MIT) from a distance of over 175 m, and the resonant frequency of the antenna tower on the roof is identified with an amplitude of 0.21 mm, which was less than 1/170th of a pixel. Methods for improving the noise floor of the measurement are discussed, especially for motion compensation and the effects of video downsampling, and suggestions are given for implementing the methodology into a structural health monitoring (SHM) scheme for existing and new structures. **DOI: 10.1061/(ASCE)IS.1943-555X.0000348.** *© 2016 American Society of Civil Engineers.*

**Author keywords:** Video camera; Computer vision; Vibration analysis; Motion magnification; Condition assessment.

## Introduction

Noncontact measurement methods for the condition assessment of civil infrastructure can have advantages over traditional measurement methods. Standard wired accelerometers used for structural health monitoring (SHM) are labor-intensive to instrument a structure with, as wiring issues and physical placement can be cumbersome. However, systems are usually meant to last for long-term monitoring where the initial time investment pays off. In the case of nondestructive testing (NDT), it is time consuming to inspect a large structure with ultrasonic testing or other similar methods (although much more detailed information is obtained). Noncontact methods have distinct advantages because of these issues, and typically some form of electromagnetic radiation is measured from the structure. An example of this is synthetic aperture radar (SAR), which is an interferometric radar technique that can be used for monitoring bridges or other civil infrastructure (Farrar et al. 1999; Pieraccini et al. 2006), and laser vibrometry used for modal testing of a building (Valla et al. 2014). Another noncontact method would be the use of video cameras, which generally measure visible light and can be easily set up and measure a large scene of interest as every pixel collects a time series. However, the tradeoff is less-precise data compared to contact techniques.

Video cameras have been used to measure civil infrastructure on a variety of different structures in previous work. Many camera measurements have been made of bridges and the cables on suspension bridges using small lights or paper targets. However, this nullifies the easy-to-instrument advantage of the camera as the targets need to be placed on the structure (Wahbeh et al. 2003; Lee and Shinozuka 2006; Kim and Kim 2011; Cigada et al. 2014). Target-less measurement, a much more ideal use which does not require any preparation other than the camera itself where virtual sensors are imposed over the video (Schumacher and Shariati 2013), have also been made of bridges and cables (Kim and Kim 2013; Cigada et al. 2014) and a traffic structure (Bartilson et al. 2015). Particularly impressive is the Caetano et al. (2011) study where frequencies of cables in a cable-stayed bridge were identified from a distance of 850 m with the smallest estimated amplitude being approximately 1/3 of a pixel. In this paper the smallest identified resonant frequency has an amplitude of less than 1/100 of a pixel using newly developed computer vision techniques.

Recently, researchers have been able to use computer vision techniques to analyze small motions in videos. This work is called motion magnification, which amplifies small imperceptible motions in specified frequency bands, effectively producing a visualization of an object's operational deflection shapes (Wu et al. 2012; Wadhwa et al. 2013). Similarly, by analyzing small motions in video, Davis et al. (2014) were able to recover sound from a

[1] Research Assistant, Dept. of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139. E-mail: ju21743@mit.edu

[2] Research Assistant, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139. E-mail: abedavis@mit.edu

[3] Research Assistant, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139. E-mail: nwadhwa@mit.edu

[4] Professor, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139. E-mail: fredo@mit.edu

[5] Thomas and Gerd Perkins Professor, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139. E-mail: billf@mit.edu

[6] Professor, Dept. of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139 (corresponding author). E-mail: obuyuk@mit.edu

Note. This manuscript was submitted on September 1, 2015; approved on September 14, 2016; published online on December 8, 2016. Discussion period open until May 8, 2017; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Infrastructure Systems***, © ASCE, ISSN 1076-0342.

Downloaded from ascelibrary.org by Massachusetts Institute of Technology (MIT) on 12/08/16. Copyright ASCE. For personal use only; all rights reserved.

Downloaded from ascelibrary.org by Massachusetts Institute of Technology (MIT) on 12/08/16. Copyright ASCE. For personal use only; all rights reserved.

loudspeaker using only video. In related work using similar video processing, a camera was used to measure the frequency response spectra of rods and fabrics and use them to detect changes in their material properties (Davis et al. 2015). Previous work using the same processing method presented in this paper was able to identify the mode shapes of cantilever beams and pipes in a laboratory setting (Chen et al. 2015b, c). This study expands that previous work to outdoor, uncontrolled environments and much longer distances, and is an expanded version of conference proceedings presented at the 2015 International Symposium Nondestructive Testing in Civil Engineering (NDT-CE) (Chen et al. 2015a).

The purpose of this paper is to demonstrate a method for video camera-based vibration measurement that is able to see smaller subpixel displacements than previous work with camera-based measurements. A measurement was made from a distance of over 175 m of the antenna tower on top of the Green Building at the Massachusetts Institute of Technology (MIT). This measurement was made in an outdoor setting with uncontrolled lighting and without any experimenter controlled active excitation of the structure. Methods for improving the noise floor of the measurement by correcting for camera motion are discussed, and a study on the effect of video downsampling is presented. Suggestions are given for implementing this methodology for the condition assessment of civil infrastructure

## Calculating Displacement from Videos

The procedure for extracting a displacement signal from a video is inspired by motion magnification (Wadhwa et al. 2013) and described in detail in the Chen et al. (2015b, c) papers. Individual images in the video, with intensity values $I(x,y,t_0)$ are decomposed into a local spatial amplitude $A_\theta(x,y,t_0)$ and phase $\phi_\theta(x,y,t_0)$ representation using quadrature pair basis filters, represented by $(G_2^\theta + \mathrm{i}H_2^\theta)$ for some orientation $\theta$ in [Eq. (1)], specified in Freeman and Adelson (1991). This representation is analogous to the Fourier coefficients of amplitude and phase, but are local to a small kernel around the pixel

$$A_\theta(x,y,t_0)e^{i\phi_\theta(x,y,t_0)} = \left(G^\theta + iH^\theta\right) \otimes I(x,y,t_0) \tag{1}$$

The displacement signal is obtained from the motion of constant contours of local phase in time (Fleet and Jepson 1990; Gautama and Van Hulle 2002), which can be expressed as

$$\phi_\theta(x,y,t) = c \tag{2}$$

for some constant $c$. The displacement signal in a single direction then comes from the distance the local phase contours move between the first frame and the current frame. Considering only the horizontal direction, this is obtained by differentiating Eq. (2) with respect to time:

$$\left[\frac{\partial\phi_0(x,y,t)}{\partial x}, \frac{\partial\phi_0(x,y,t)}{\partial y}, \frac{\partial\phi_0(x,y,t)}{\partial t}\right] \cdot (u,v,1) = 0 \tag{3}$$

Since $\partial\phi_0(x,y,t)/\partial y \approx 0$, this gives us an expression for the horizontal velocity in units of pixels:

$$u = -\left[\frac{\partial\phi_0(x,y,t)}{\partial x}\right]^{-1} \frac{\partial\phi_0(x,y,t)}{\partial t} \tag{4}$$

The velocity $u$ is integrated in time to obtain the displacement over time in units of pixels.

The displacement in units of pixels can be converted to units of millimeters by a scale factor determined by the physical size and pixel width of an object at the same depth in the video frame. Chen et al. (2015b) proved that this procedure gives accurate measurements by comparing the displacement signal measured by a camera with signals measured by a laser vibrometer and an accelerometer in a laboratory setting.

### *Assumptions*

A major assumption is that the illumination in the scene remains unchanged, otherwise there might be apparent motion that doesn't exist. Changing lighting or background conditions due to clouds either passing over the sun or behind the structure could introduce erroneous apparent motion signals onto the object. Care must be taken to avoid video sequences with too many changes in these conditions.

Displacement signals are least noisy in areas with greater texture or contrast (e.g., edges), while textureless regions have ambiguous

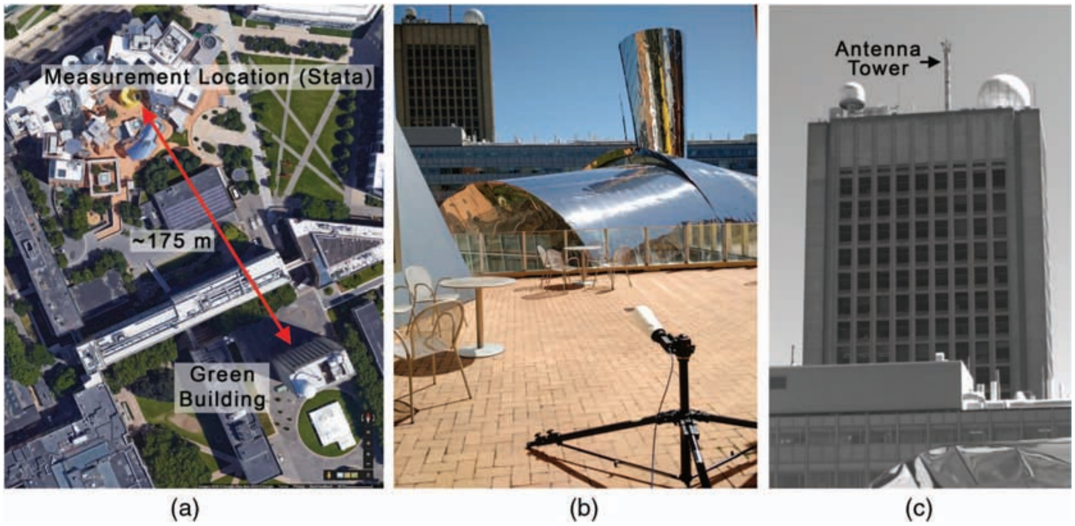


**Fig. 1.** (Color) Experimental setup for measurement of MIT's Green Building showing: (a) the satellite view of the measurement location relative to the Green Building (imagery ©2015 Google, map data ©2015 Google); (b) the view from the measurement location and the camera setup (image by Justin G. Chen); (c) a screenshot from the recorded video with the antenna tower labeled (image by Justin G. Chen).

Downloaded from ascelibrary.org by Massachusetts Institute of Technology (MIT) on 12/08/16. Copyright ASCE. For personal use only; all rights reserved.

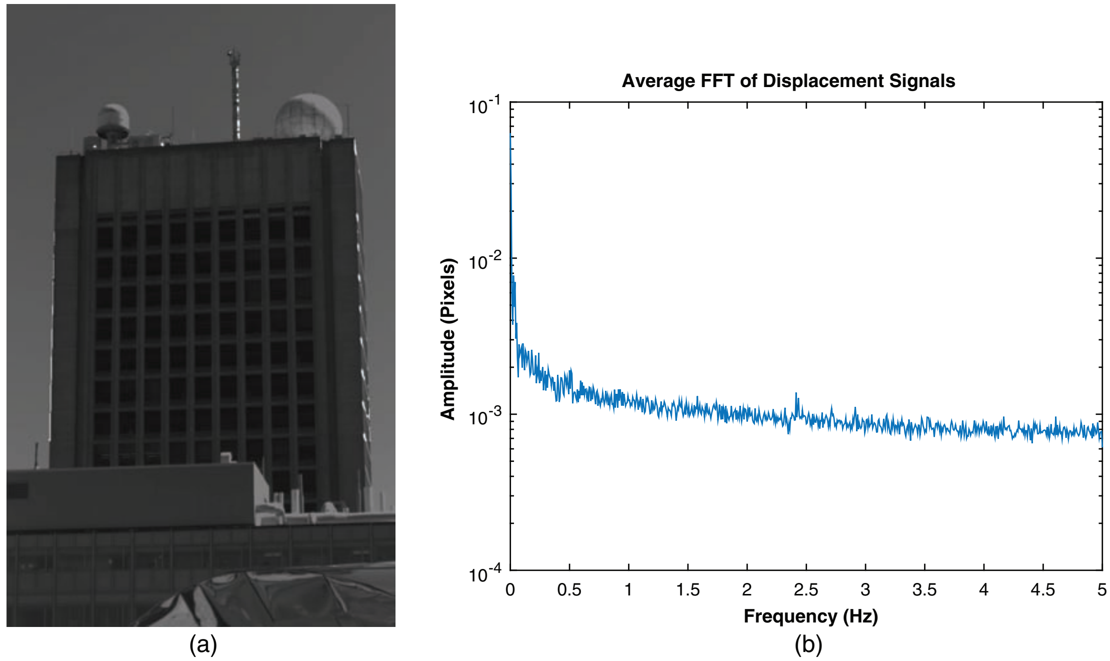


**Fig. 2.** (Color) Initial data processing results showing (a) screenshot from the video with pixel mask overlaid, bright pixels being those with extracted displacements (image by Justin G. Chen); (b) average frequency spectrum of displacement signals extracted from the video

motion due to the aperture problem. In order to determine which pixels have good texture or contrast, a minimum threshold on the amplitude coefficient which corresponds to the contrast is chosen. The threshold used is half of the median of the 30 pixels with the largest amplitude, but a different threshold can be chosen.

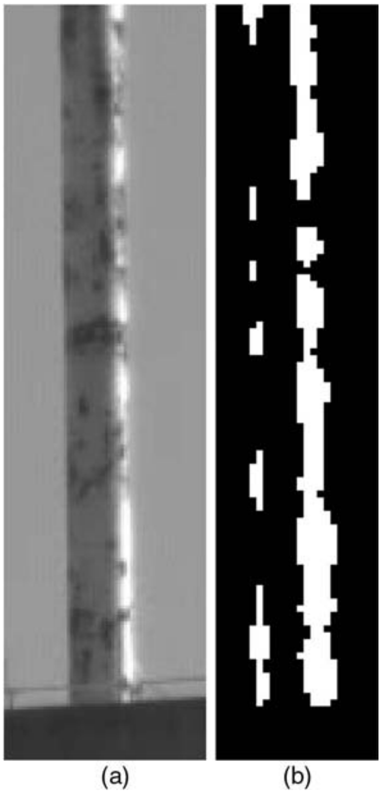


**Fig. 3.** (a) Screenshot (image by Justin G. Chen); (b) pixel mask for the video cropped to the antenna tower

### Video Preprocessing

The video may be downsampled prior to processing to change the scale on which the basis filters are acting and increase the signal-to-noise ratio (SNR) for each individual pixel. Intensity noise is the main contributor of noise for each pixel, so averaging neighboring pixels together will reduce the amount of noise, lowering the noise floor of the measurement. This downsampling is done by factors of two along both dimensions in a video using binomial filters, which also reduces the spatial resolution of the video. A study of the effects of video downsampling on the SNR and the noise floor of measured data is shown in the "Results" section.

### Frequency Identification

After extracting the displacement signals, there are typically too many signals to inspect by hand (i.e., in the hundreds or thousands). To get a general sense of the structure in the video they are averaged, then a fast Fourier transform (FFT) is used to transform the average signal into the frequency domain to obtain a frequency spectrum of the average displacement signal. Conversely, the displacement signals may undergo the FFT first, then become averaged in the frequency domain to obtain an average frequency spectrum for the signals. Examining these two average frequency spectra give a good idea of whether or not the measurement shows appreciable signal.

For a more in-depth analysis of the displacement signals, standard frequency-domain modal analysis methods such as peak picking or frequency domain decomposition (FDD) can be used (Brincker et al. 2001). A comparison between the two methods as applied for video camera-measured signals is summarized here, and was previously described in Chen et al. (2015c). Peak picking is computationally quick. However, if the resonant frequency signal is relatively weak or only belongs to part of the structure, it will often not be seen in the average frequency spectra and will not produce any useful results. FDD is able to pick out resonant peaks with lower SNR or local resonances. However, it takes much more

Downloaded from ascelibrary.org by Massachusetts Institute of Technology (MIT) on 12/08/16. Copyright ASCE. For personal use only; all rights reserved.

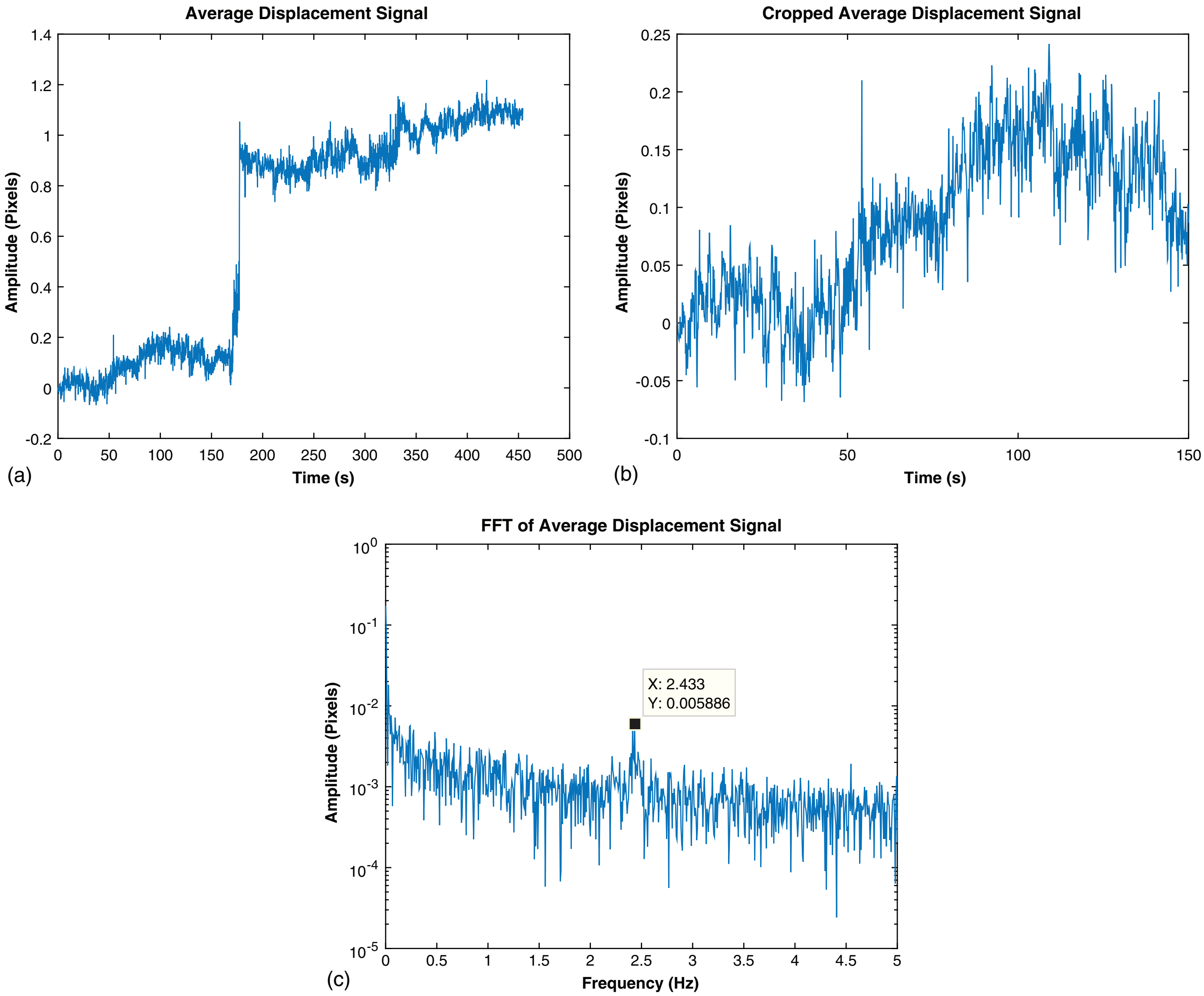


**Fig. 4.** (Color) Average displacement signals measured from cropped video of Green Building antenna tower: (a) the full 454-s time series; (b) the cropped first 150 s of the time series; (c) FFT of the cropped time series signal

time to run, especially as the signal count grows as it depends on the calculation of the spectral matrix and a singular value decomposition. Both methods result in potential resonant frequencies and operational mode shapes for the structure, but FDD may be more useful as it is a nonparametric method. A more in-depth comparison of operational modal analysis methods in general is described in Reynders (2012). Any local vibration modes that are found usually warrant more in-depth processing with only the signals from that local structure.

## Experimental Setup

The video measurement was made of the Green Building on MIT's campus, a 21-story, 90-m tall reinforced concrete building. The camera was located on a terrace of the Stata Center, another building at MIT a distance of approximately 175 m away from the Green Building, as shown in a satellite view in Fig. 1(a). The view from the measurement location is shown in Fig. 1(b) also showing the experimental setup. A Point Grey Grasshopper3 camera was used with a 24–85 mm zoom lens set to 30 mm to capture the whole building and the antenna tower on the roof, as seen in a screenshot from the recorded video in Fig. 1(c). The resolution of the video was 1,200 × 1,920, resulting in a scale factor of 3.65 cm per pixel at the depth of the structure, determined from the 36-m width of the structure. The video was recorded in greyscale in a raw 8-bit pixel resolution format, which disables any onboard image processing. A 454-second long video at 10 frames per second (fps) was captured. The weather was clear during the measurement itself with a temperature of 18.9°C and 4.6 m/s winds (The Weather Company, LLC 2016).

## Results

### *Green Building Initial Processing*

The measured video of the Green Building was processed to determine if any frequency peaks indicative of a possible resonance of the building or other structures were captured in

Downloaded from ascelibrary.org by Massachusetts Institute of Technology (MIT) on 12/08/16. Copyright ASCE. For personal use only; all rights reserved.

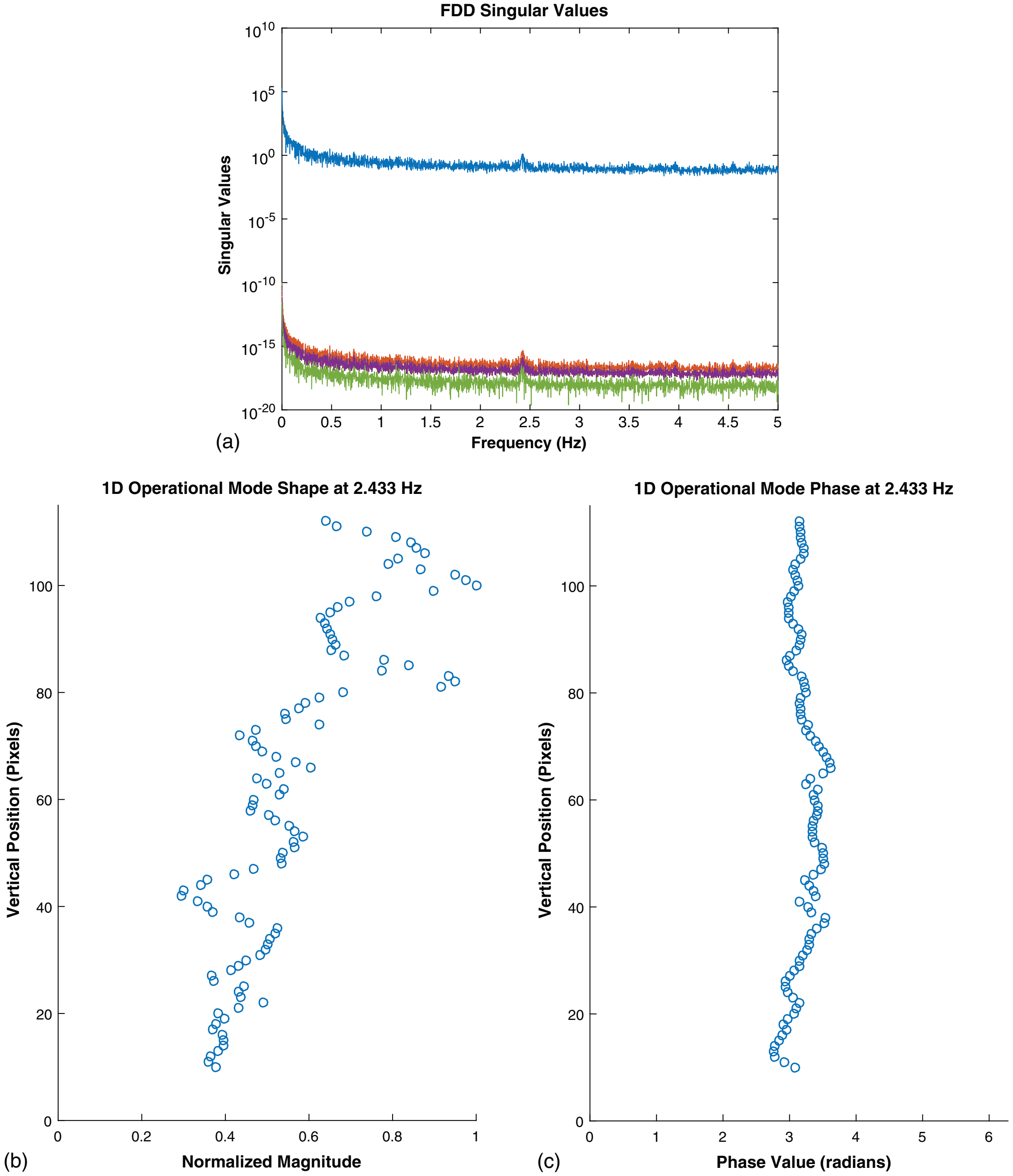


**Fig. 5.** (Color) Results of FDD on the Green Building antenna tower horizontal displacement signals showing (a) singular values and at 2.433 Hz; (b) operation mode shape; (c) phase

the measurement. The video was downsampled by a factor of two in each dimension to a size of 600 × 960. Out of a possible 563,584 pixels, slightly lower than the total number due to the size of the filter kernel with valid convolution results, 1,191 pixels with displacements were extracted as shown in white, overlaid over the a video screenshot in Fig. 2(a). Of the 454 s data collect, the first 150 s is without much camera motion, so the signals were cropped to the first 150 s for analysis. After using an FFT to obtain the frequency spectra, they were averaged to obtain an average frequency spectrum for all the signals shown in Fig. 2(b). The most prominent resonant peak was at 2.413 Hz, and it was found that the pixels in the video that contributed to this peak corresponded to the antenna tower located on the roof of the building.

### *Antenna Tower Processing*

To better analyze the motion of the antenna tower, the original video of the whole Green Building was cropped to a video containing only the tower, with a resolution of 64 × 240, shown in Fig. 3(a). Before processing, the video was downsampled by a factor of two in each dimension to a size of 32 × 120, and with the filter kernel size a possible 2,688 pixels with displacements. 441 out of those 2,688 pixels were high-contrast pixels with

Downloaded from ascelibrary.org by Massachusetts Institute of Technology (MIT) on 12/08/16. Copyright ASCE. For personal use only; all rights reserved.

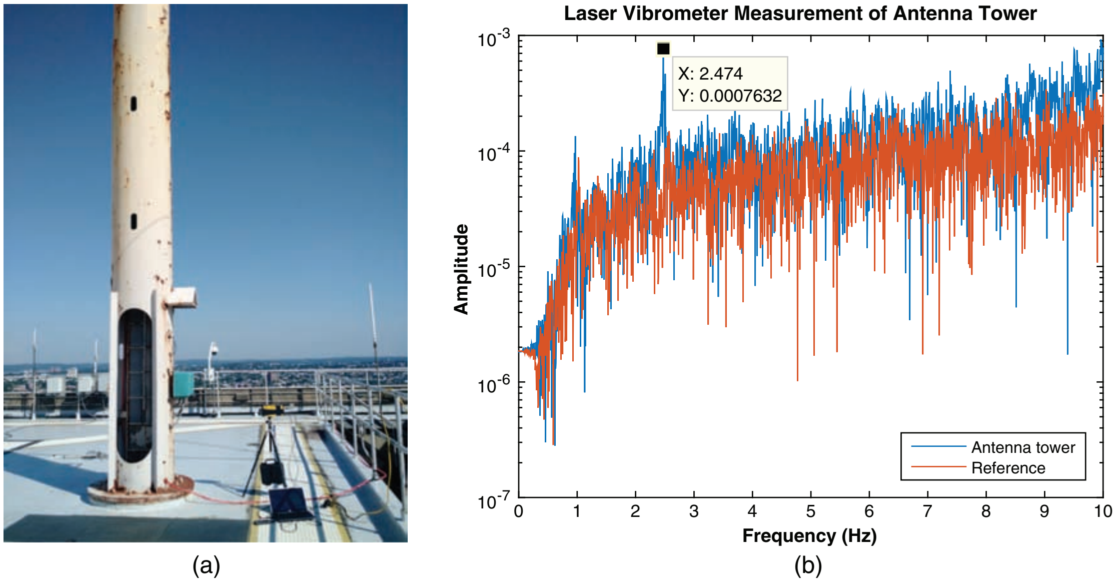


**Fig. 6.** (Color) Laser vibrometer verification of MIT Green Building antenna tower resonant frequency: (a) measurement setup (image by Justin G. Chen); (b) result with static reference for comparison

extracted displacements, shown in white in Fig. 3(b). The resulting average displacement time series is shown in Fig. 4(a). As with the Green Building measurement video, due to camera motion the time series was cropped to the first 150 s of the video, which is shown in Fig. 4(b).

The FFT of the average displacement signal is shown in Fig. 4(c) which shows a resonant peak at 2.433 Hz. A one-dimensional representation of the structure's horizontal motions was generated by averaging the motions of pixels at the same vertical height. Frequency domain decomposition was used to determine the operational deflection shape of the structure at 2.433 Hz and the resulting singular values are shown in Fig. 5(a). The resulting shape is shown in Fig. 5(b) which shows something similar to a first-bending mode shape of a cantilever beam, and the phase values are shown in Fig. 5(c). Even though the operational shape at 2.433 Hz is quite noisy, the trend of the measured shape is plausibly a mode shape and phase relationships are consistent. The amplitude of the peak is 0.00589 pixels, much smaller than similar previous work has been able to detect. Using the scaling factor of 3.65 cm per pixel, the amplitude of the resonant peak is calculated to be 0.215 mm, and given the noise floor of 0.07 mm, the SNR of this measurement is 3.

## Verification

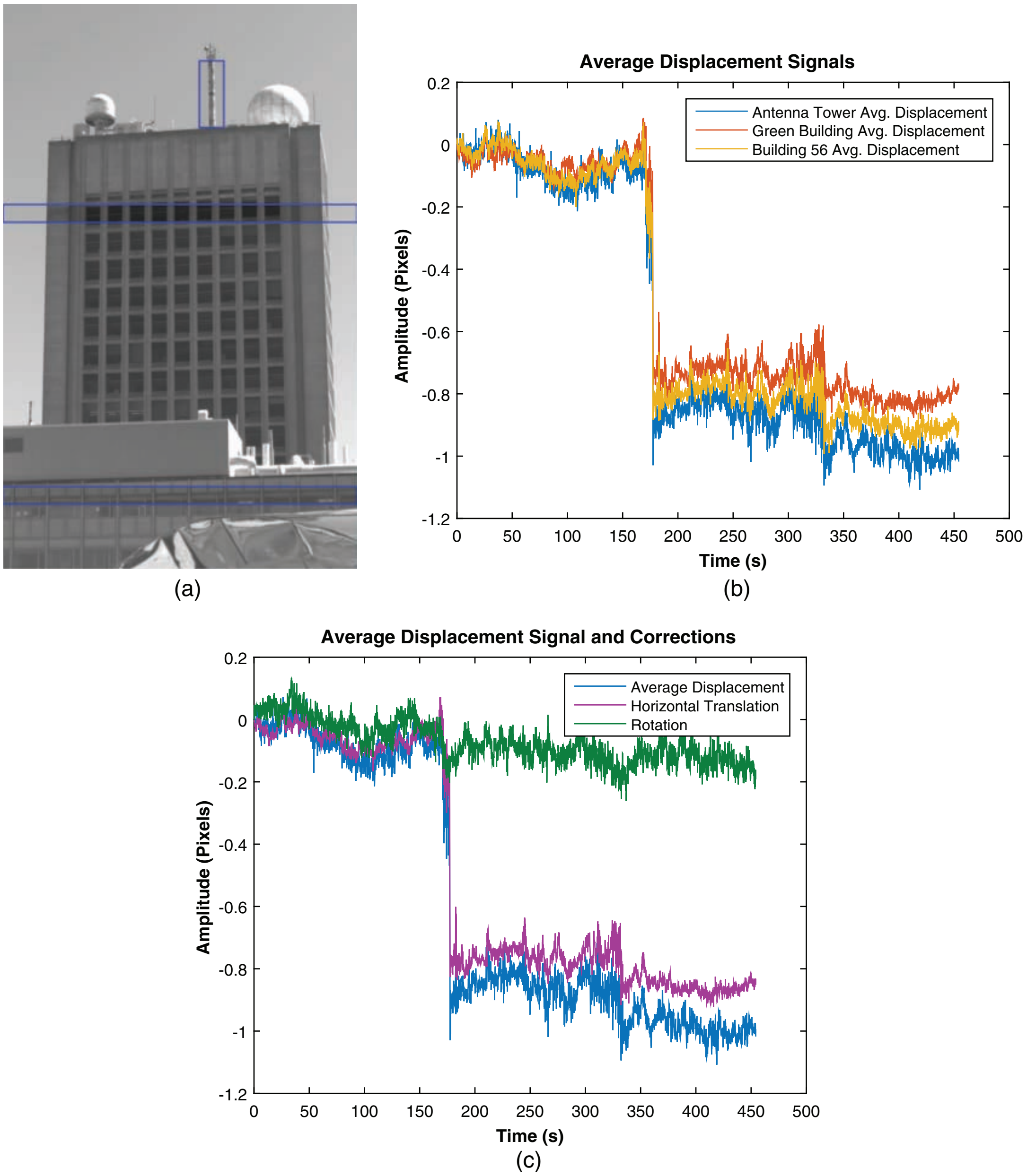
A laser vibrometer was used at close range to measure the frequency response of the antenna tower during a different day with similar weather conditions, clear with a temperature of 28.3°C and 5.7 m/s winds (The Weather Company, LLC 2016), to determine the accuracy of the resonant frequency as measured by the video camera. The measurement setup on the roof of the building is shown in Fig. 6(a) with the laser vibrometer located about one meter away from the base of the tower, measuring a point approximately one meter high. To discount any potential resonances of the laser vibrometer and tripod system itself, a measurement was also made of the ground next to the antenna tower as a reference. The resulting laser vibrometer measurement from the antenna tower is shown in blue in Fig. 6(b) with the reference signal shown in red. The resonant peak that stands out in the antenna tower measurement front the reference spectrum is a peak at 2.474 Hz, which is similar to the 2.433 Hz measured by the camera. The percent difference between the measurements is only 1.7%, and is within the potential variation in resonant frequencies due to temperature variations (Clinton et al. 2006; Sohn 2007), 28.3°C during the laser vibrometer measurement and 18.9°C during the camera measurement.

## Camera Motion Compensation

The full 454-s video was not used in the case of the antenna tower measurement because of camera motion that introduced motion in the measurement displacement signal. To correct for camera motion, reference objects in the frame of the video that are assumed to be stationary can be used to determine the apparent camera motion signal. This camera motion signal can then be subtracted from the displacement signal of the structure of interest, i.e., the antenna tower. Given that there was no measurable response from the Green Building itself, it can be used as a reference to measure the camera motion from the video. A shorter building in the foreground, Building 56, is also used as a stationary reference. Fig. 7(a) shows the regions of interest around reference objects used to calculate displacement signals to accomplish this motion compensation. The signals are shown in Fig. 7(b) and they include the antenna tower in blue, a slice from the Green Building in red, and a slice from a structure in front of the Green Building, i.e., Building 56, in yellow. All three signals look quite similar, demonstrating that there is significant camera motion to be corrected.

Both horizontal translation and rotation of the camera were corrected. Translational motion is calculated from the average of the Building 56 and the Green Building displacement signals, shown in purple in Fig. 7(c), while rotational motion of the camera is calculated from the difference of the two signals and shown in green. The resulting corrected displacement signal of the antenna tower is shown in red in Fig. 8(a), as compared to the original displacement signal shown in blue. Much of the camera motion, especially the large jump at 175 s into the signal, is removed from the displacement signal. The resulting effects in the Fourier domain are shown in Fig. 8(b). Most of the difference seems to be in

Downloaded from ascelibrary.org by Massachusetts Institute of Technology (MIT) on 12/08/16. Copyright ASCE. For personal use only; all rights reserved.



(a)

(b)

(c)

**Fig. 7.** (Color) Motion compensation signals and corrections: (a) video regions of interest for motion compensation (image by Justin G. Chen); (b) average displacement signals of regions of interest; (c) resulting correction signals

the frequencies below 0.2 Hz, which is shown in detail in Fig. 8(c), where the noise floor is reduced by a factor of two. Additional improvements in the noise floor can be gained by sacrificing the field of view and zooming in on the structure to have a lower physical distance per pixel scaling factor.

### *Video Downsampling Comparison*

A comparison was conducted to determine the effect of spatial downsampling on the noise floor and SNR of the resulting measured frequency spectrum. The previous processing was all done with the video spatially downsampled by one level, or reduced by a factor of two in each dimension; a comparison is made to no downsampling called level zero, or downsampling twice, or by a factor of four in each dimension of the video. Fig. 9(a) shows the mean noise floor for individual pixels between 3 and 5 Hz versus the edge strength of the pixel for different levels of downsampling for all pixels with valid displacements in the video. More spatial downsampling lowers the noise floor of each individual pixel, but there are fewer pixels in total. This also shows that the noise floor gets much worse for pixels with insufficient edge strength, which is the rationale behind using a threshold to select pixels good texture or edges.

Fig. 9(b) shows the differences in the measured average frequency spectrum, and downsampling noticeably reduces the noise floor of the measurement while preserving the amplitude of the vibration peak. Table 1 summarizes the results for the different levels of downsampling. In this case, the SNR is calculated from the average frequency spectrum with pixels above the previously specified threshold for edge strength. With two levels of downsampling, not only is the processing time reduced by half as there are far fewer pixels to process, but the SNR is improved as well. However, this is at the expense of less spatial resolution as the frame size ends up being approximately 1/16 of the size of the original video.

Downloaded from ascelibrary.org by Massachusetts Institute of Technology (MIT) on 12/08/16. For personal use only; all rights reserved.

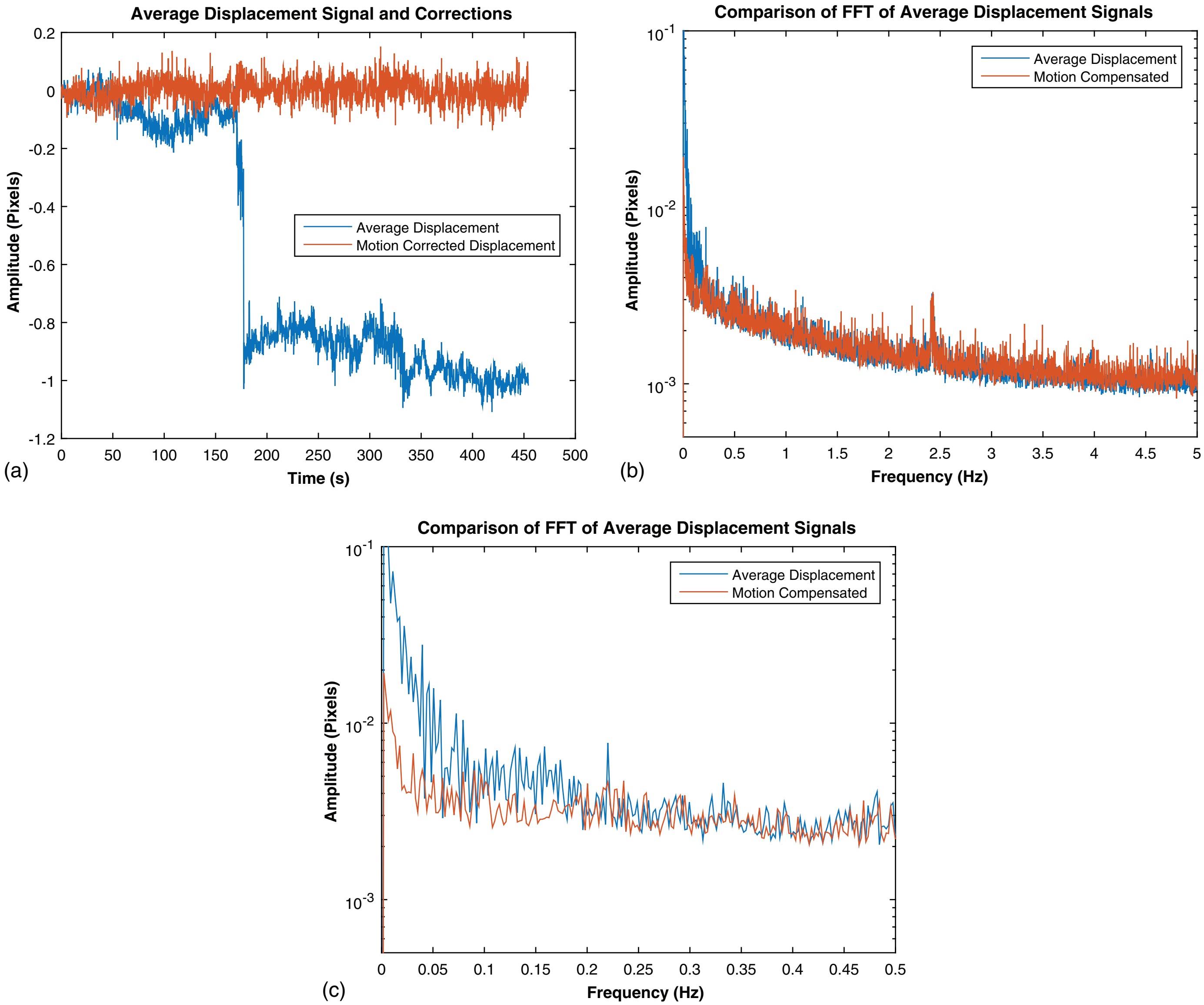


**Fig. 8.** (Color) (a) Motion compensated antenna tower signal comparison; (b) frequency spectrum comparison; (c) 0–0.2 Hz range of the frequency spectrum comparison in detail

## Discussion

The results present the various processing steps of the measured Green Building video from the initial processing of the full video to specific processing of the antenna tower and compensation for camera motion. Each step is necessary to narrow down the focus of what is actually being measured by the camera. In the initial measurement the camera measures the relative motion between the camera itself and any objects in the frame. The antenna tower was determined to be an object of interest, and as camera motion was quite significant other portions of the video were used as references to subtract the camera motions from the displacement signal from the antenna tower. The resonant frequency of the antenna tower measured by the camera was successfully verified against the laser vibrometer measurement made from the roof.

This current study has only identified the resonant frequency from the antenna tower. If the system were to be used for structural health monitoring possible methods for monitoring changes in the displacement signal include statistical pattern recognition techniques (Sohn et al. 2001), one-class machine learning methods (Long and Buyukozturk 2014), analysis of nonlinear features (Mohammadi Ghazi and Büyüköztürk 2016), or other damage detection algorithms. Without the need for physical access to instrument a structure, cameras can more easily collect data from structures that might otherwise be difficult or time consuming to instrument. Continuous monitoring of structures with cameras seems tractable in the near future as the camera hardware involved is relatively inexpensive since civil infrastructure tends to vibrate at relatively low frequencies accessible to normal consumer cameras.

### *Limitations*

The main limitations of the methodology are the higher noise floor of the measurement compared to traditional sensors, the requirement of a mostly stationary camera, and potential problems with flickering lighting or changing background conditions. The noise floor of the camera methodology is still higher when compared to contact accelerometers by up to several orders of magnitude depending on the frequency range. This is illustrated in the direct laboratory measurement comparison between the camera system,

Downloaded from ascelibrary.org by Massachusetts Institute of Technology (MIT) on 12/08/16. Copyright ASCE. For personal use only; all rights reserved.

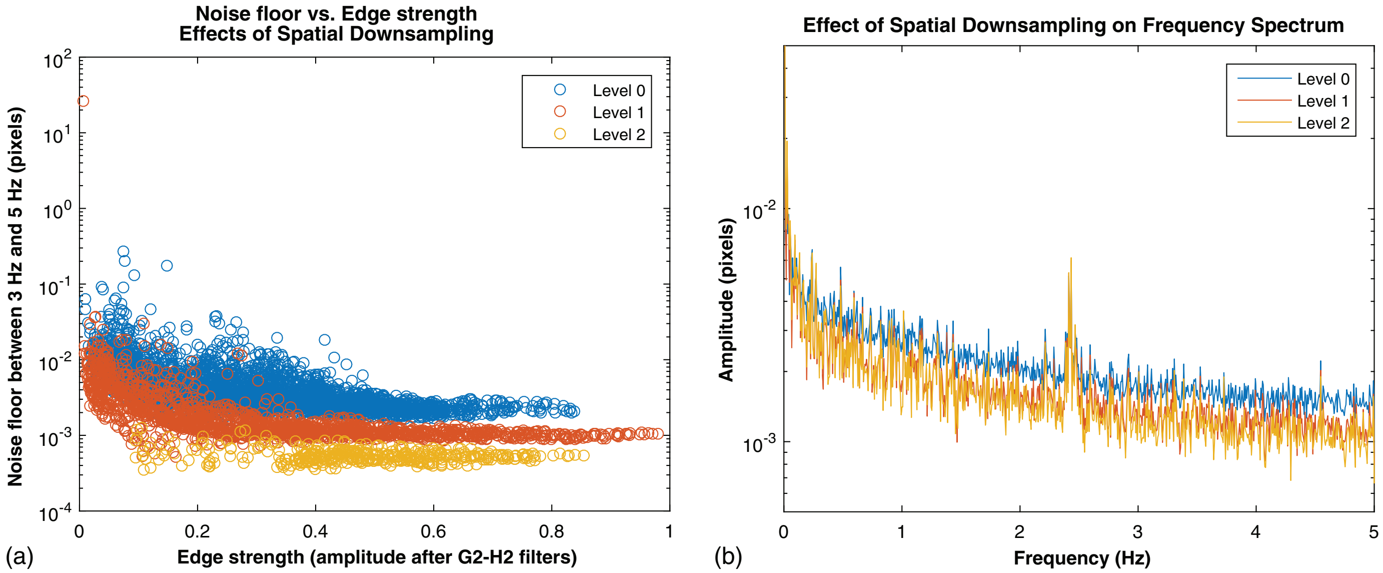


**Fig. 9.** (Color) Comparison of the effects of video spatial downsampling on: (a) the noise floor of individual pixels; (b) the measured frequency spectrum

**Table 1.** Summary of Comparison of Different Spatial Downsampling Levels

| Level | Processing time (s) | Frame size | Valid pixels | SNR |
|---|---|---|---|---|
| 0 | 64.5 | 56 × 232 | 3,908 | 3.03 |
| 1 | 40.7 | 24 × 112 | 1,729 | 3.98 |
| 2 | 28.2 | 8 × 52 | 416 | 4.54 |

Note: SNR = signal-to-noise ratio.

laser vibrometer, and an accelerometer (Chen et al. 2015b). At higher frequencies structures will vibrate with smaller amplitudes, which are more difficult to measure as displacements compared to typical sensors such as accelerometers or strain gauges which can be extremely sensitive. Another limitation is that the camera needs to be somewhat stationary so that only small motions that are less than a couple of pixels are present in the video. Not only do camera motions introduce apparent motion that need to be corrected for, but if the motions are too large the processing procedure can break down. Changing lighting and background conditions can also introduce apparent motions into a video when there are none. This could happen due to clouds passing over the sun or moving behind a building of interest.

### Camera Noise Floor versus Expected Building Motion

To be able to obtain meaningful displacement signals from many different types of civil infrastructure the main improvement that needs to be made is a lower noise floor. Much of the previous work with camera-based measurements require multiple pixel displacements, and only a few methods are able to access the $1/5$ of a pixel range of subpixel measurements. This technique gives a noise floor of less than $1/500$ of a pixel or 0.07 mm, which in this measurement is sufficient to measure the motion of the antenna tower but insufficient to measure the motion of the Green Building itself. The signal from the antenna tower should contain the motion of the Green Building as the tower is directly and rigidly mounted to the roof. However, there is no evidence of the Green Building resonant frequencies which are likely under the noise floor of the current measurement as the resonant frequencies of the Green Building in the East-West direction, 0.68 Hz for the first bending mode and 2.49 Hz for the second, are not present in the displacement signals (Çelebi et al. 2014).

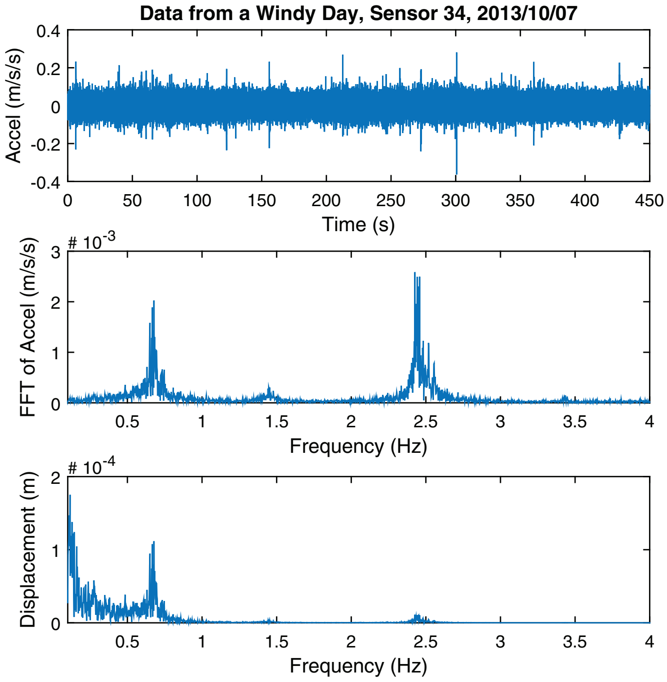


**Fig. 10.** (Color) Green building accelerometer data from an especially windy day

The Çelebi et al. (2014) study found that ambient motion of the Green Building had accelerations with an amplitude of $0.02\ \mathrm{cm/s^2}$ in the east-west direction, which at the first bending mode corresponds to displacements of 0.01 mm, which are below the noise floor of our measurement. Fig. 10 shows accelerometer data from the roof of the Green Building, in the east-west direction during a wind advisory and severe thunderstorm watch with winds of $8.9\ \mathrm{m/s}$ from the south-southwest on October 7, 2013. The building motion on this particularly windy day is larger at $1 \times 10^{-4}$ meters or 0.1 mm, which would be borderline measurable by the camera. It is unlikely that there are techniques that can be used to reduce the noise floor of the camera system low enough to sufficiently measure the ambient vibrations of the Green Building under normal conditions. However, during more extreme weather events and with a moderate improvement in the measurement noise floor, the building may move enough to be measurable by the camera system.

There are several possible options to achieve a lower noise floor. A more powerful zoom lens would give a lower physical distance-per-pixel ratio and lower the physical scale of the noise floor at the expense of a smaller field of view. Longer measurement windows could also improve the SNR, achievable as a part of a continuous monitoring scheme where a video camera might be able to collect valid data during any daylight hours. Because structural motions may not persist throughout the full measurement and coherent averaging may average out the vibration signal, incoherent averaging over a sliding window for a very long measurement time may be necessary. Another significant contributor to the noise floor as seen in this paper is camera motion. For low frequencies that are less than 1 Hz and long-term measurements, it is possible that background ground motions in an urban environment may be a source of seismic noise that causes camera motion (Groos and Ritter 2009) that cannot be averaged out. Camera motion due to intrinsic ground motion could be reduced by correcting the collected data with accurate external measurements of the camera's motion. There is work in computer vision that uses gyroscopes and accelerometers to correct for camera shake in deblurring images (Joshi et al. 2010), which suggests that similar methods could also be used for videos while preserving the subtle motions present in objects in the video.

## Conclusions

A camera-based vibration measurement methodology for civil infrastructure was demonstrated by measuring the ambient vibration response of the antenna tower atop MIT's Green Building from a distance of over 175 m. The resonant frequency of 2.433 Hz measured by the camera agreed within 1.7% of the frequency measured by a laser vibrometer from close range. The amplitude of the motion detected was 0.21 mm or 1/170 of a pixel, and the noise floor of the measurement was 0.07 mm or 1/500 of a pixel for a SNR of 3. This represents an improvement in the scale of the displacements measurable by the camera from long distance in terms of fractions of a pixel. However, this noise floor was too high and the camera system was unable to measure the ambient vibration of the Green Building itself. Camera motion was compensated for by using reference objects in the same video scene, and noise in the range of 0–0.2 Hz was reduced by a factor of two. The effects of video downsampling were shown in that it improved the SNR at the expense of spatial resolution.

## Acknowledgments

The authors acknowledge the support provided by Royal Dutch Shell through the MIT Energy Initiative, and thank chief scientists Dr. Dirk Smit and Dr. Sergio Kapusta, project managers Dr. Keng Yap and Dr. Lorna Ortiz-Soto, and Shell-MIT liaison Dr. Jonathan Kane for their oversight of this work. Special thanks are also given to Scott Wade for access to the roof of the Green Building to make the verification measurement of the antenna tower, and the United States Geological Service (USGS) for the accelerometer data.

## References

Bartilson, D. T., Wieghaus, K. T., and Hurlebaus, S. (2015). "Target-less computer vision for traffic signal structure vibration studies." *Mech. Syst. Sig. Process.*, 60, 571–582.

Brincker, R., Zhang, L., and Andersen, P. (2001). "Modal identification of output-only systems using frequency domain decomposition." *Smart Mater. Struct.*, 10(3), 441–445.

Caetano, E., Silva, S., and Bateira, J. (2011). "A vision system for vibration monitoring of civil engineering structures." *Exp. Tech.*, 35(4), 74–82.

Çelebi, M., Toksöz, N., and Büyüköztürk, O. (2014). "Rocking behavior of an instrumented unique building on the mit campus identified from ambient shaking data." *Earthquake Spectra*, 30(2), 705–720.

Chen, J. G., Davis, A., Wadhwa, N., Durand, F., Freeman, W. T., and Buyukozturk, O. (2015a). "Video camera-based vibration measurement for condition assessment of civil infrastructure." *Int. Symp. Non-Destructive Testing in Civil Engineering (NDT-CE)*, BAM Federal Institute for Materials Research and Testing, Berlin.

Chen, J. G., Wadhwa, N., Cha, Y. J., Durand, F., Freeman, W. T., and Buyukozturk, O. (2015b). "Modal identification of simple structures with high-speed video using motion magnification." *J. Sound Vib.*, 345, 58–71.

Chen, J. G., Wadhwa, N., Durand, F., Freeman, W. T., and Buyukozturk, O. (2015c). "Developments with motion magnification for structural modal identification through camera video." *Dynamics of Civil Structures.*, Vol. 2, Springer, Cham, Switzerland, 49–57.

Cigada, A., Mazzoleni, P., Zappa, E., et al. (2014). "Vibration monitoring of multiple bridge points by means of a unique vision-based measuring system." *Exp. Mech.*, 54(2), 255–271.

Clinton, J. F., Bradford, S. C., Heaton, T. H., and Favela, J. (2006). "The observed wander of the natural frequencies in a structure." *Bull. Seismol. Soc. Am.*, 96(1), 237–257.

Davis, A., Bouman, K. L., Chen, J. G., Rubinstein, M., Durand, F., and Freeman, W. T. (2015). "Visual vibrometry: Estimating material properties from small motions in video." *Proc., IEEE Conf. on Computer Vision and Pattern Recognition*, Computer Vision Foundation, 5335–5343.

Davis, A., Rubinstein, M., Wadhwa, N., Mysore, G. J., Durand, F., and Freeman, W. T. (2014). "The visual microphone: Passive recovery of sound from video." *ACM Trans. Graphics*, 33(4), 1–10.

Farrar, C. R., Darling, T. W., Migliori, A., and Baker, W. E. (1999). "Microwave interferometers for non-contact vibration measurements on large structures." *Mech. Syst. Sig. Process.*, 13(2), 241–253.

Fleet, D. J., and Jepson, A. D. (1990). "Computation of component image velocity from local phase information." *Int. J. Comput. Vision*, 5(1), 77–104.

Freeman, W. T., and Adelson, E. H. (1991). "The design and use of steerable filters." *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(9), 891–906.

Gautama, T., and Van Hulle, M. (2002). "A phase-based approach to the estimation of the optical flow field using spatial filtering." *IEEE Trans. Neural Networks*, 13(5), 1127–1136.

Groos, J., and Ritter, J. (2009). "Time domain classification and quantification of seismic noise in an urban environment." *Geophys. J. Int.*, 179(2), 1213–1231.

Joshi, N., Kang, S. B., Zitnick, C. L., and Szeliski, R. (2010). "Image deblurring using inertial measurement sensors." *ACM Trans. Graphics (TOG)*, 29(30), 1.

Kim, S. W., and Kim, N. S. (2011). "Multi-point displacement response measurement of civil infrastructures using digital image processing." *Procedia Eng.*, 14, 195–203.

Kim, S. W., and Kim, N. S. (2013). "Dynamic characteristics of suspension bridge hanger cables using digital image processing." *NDT and E Int.*, 59, 25–33.

Lee, J. J., and Shinozuka, M. (2006). "Real-time displacement measurement of a flexible bridge using digital image processing techniques." *Exp. Mech.*, 46(1), 105–114.

Long, J., and Buyukozturk, O. (2014). "Automated structural damage detection using one-class machine learning." *Dynamics of Civil*

Downloaded from ascelibrary.org by Massachusetts Institute of Technology (MIT) on 12/08/16. Copyright ASCE. For personal use only; all rights reserved.

*Structures*, F. N. Catbas, ed., *Conf. Proc., Society for Experimental Mechanics Series*, Vol. 4, Springer, Cham, Switzerland, 117–128.

Mohammadi Ghazi, R., and Büyüköztürk, O. (2016). "Damage detection with small data set using energy-based nonlinear features." *Struct. Control Health Monit.*, 23(2), 333–348.

Pieraccini, M., Fratini, M., Parrini, F., and Atzeni, C. (2006). "Dynamic monitoring of bridges using a high-speed coherent radar." *IEEE Trans. Geosci. Remote Sens.*, 44(11), 3284–3288.

Reynders, E. (2012). "System identification methods for (operational) modal analysis: Review and comparison." *Arch. Comput. Methods Eng.*, 19(1), 51–124.

Schumacher, T., and Shariati, A. (2013). "Monitoring of structures and mechanical systems using virtual visual sensors for video analysis: Fundamental concept and proof of feasibility." *Sensors*, 13(12), 16551–16564.

Sohn, H. (2007). "Effects of environmental and operational variability on structural health monitoring." *Philos. Trans. R. Soc., London, Ser. A*, 365(1851), 539–560.

Sohn, H., Farrar, C. R., Hunter, N. F., and Worden, K. (2001). "Structural health monitoring using statistical pattern recognition techniques." *J. dyn. Syst. Meas. Contr.*, 123(4), 706–711.

The Weather Company, LLC. (2016). "Weather history and data archive weather underground." ⟨http://www.wunderground.com/history/⟩ (Feb. 3, 2016).

Valla, M., Gueguen, P., Augère, B., Goular, D., and Perrault, M. (2014). "Remote modal study of reinforced concrete buildings using a multipath lidar vibrometer." *J. Struct. Eng.*, 141(1), 1–10.

Wadhwa, N., Rubinstein, M., Durand, F., and Freeman, W. T. (2013). "Phase-based video motion processing." *ACM Trans. Graphics*, 32(4), 1.

Wahbeh, A. M., Caffrey, J. P., and Masri, S. F. (2003). "A vision-based approach for the direct measurement of displacements in vibrating systems." *Smart Mater. Struct.*, 12(5), 785–794.

Wu, H. Y., Rubinstein, M., Shih, E., Guttag, J. V., Durand, F., and Freeman, W. T. (2012). "Eulerian video magnification for revealing subtle changes in the world." *ACM Trans. Graphics*, 31(4), 1–8.

Downloaded from ascelibrary.org by Massachusetts Institute of Technology (MIT) on 12/08/16. Copyright ASCE. For personal use only; all rights reserved.