Title: **Modal Imaging of Portsmouth, New Hampshire Bridge for Video-Based Structural Health Monitoring**

Authors : Abe Davis
Justin G. Chen
Travis M. Adams
Hao Sun
Erin S. Bell
Frédo Durand
Oral Büyüköztürk

ABSTRACT

We compare the results of modal imaging [1], a method for measuring operational deflection mode shapes and frequencies in video, to the predictions of a finite element simulation, as well as accelerometer measurements, of a Portsmouth, New Hampshire bridge. The finite element simulation is based on schematics obtained from the New Hampshire Department of Transportation. We also compare the results of modal imaging on the bridge to previous visual analysis from [2], and discuss some of the algorithmic differences that appear to improve results.

INTRODUCTION

Structural health monitoring depends on reliable, interpretable data to inform engineers about damage and other properties of structures. The right data equips stakeholders with the ability to make better decisions about maintenance and repairs, saving money, and, in some cases, lives. However, data collection that takes structures out of service carries a high cost and makes persistent monitoring impossible. Over the last couple of decades, consumer digital cameras have become ubiquitous, low cost, and better performing devices due to competition driven by consumer market forces. Thus the opportunity to use consumer devices as inexpensive sensors has only recently become tractable.

Recent work in computer vision and graphics has demonstrated the ability to quantitatively measure small sub-pixel motion in video. Some of these works focus on visualizing small displacements using a technique called motion magnification [3–6].

Abe Davis, Cornell University, Department of Computer Science, Cornell University, Ithaca, NY 14850, U.S.A..

Abe Davis, Fredo Durand. CSAIL, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, U.S.A.

Justin G. Chen, Oral Buyukozturk, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, U.S.A.

Travis M. Adams, Erin S. Bell, Dept. of Civil and Environmental Engineering, Univ. of New Hampshire, Durham, NH 03824

Hao Sun, Department of Civil and Eivronmental Engineering, Northeastern University

a) b) c)

Figure 1. (Left) Wide angle view of the bridge (Middle) View of capture setup–including camera and tripod–with bridge seen in the background. (Right) Satellite image showing the camera's location relative to the bridge. More details on the original data can be found in [2].

Others focus more specifically on the analysis of mechanical vibration, applying this analysis to applications including the recovery of sound from vibrating objects in silent video [7], characterizing material properties based on the vibrations of objects measured from video [8, 9], and realistic simulation of object motion based on object dynamics recovered from video [10]. This and similar work has spawned research using vision based measurement systems for structural health monitoring.

Vision based measurement systems have specific advantages over standard contact sensors. They can be less expensive per ”sensor” as cameras collect information from many pixels forming virtual sensors. As non contact sensors, they are also less labor intensive to use to instrument a structure as the setup of several cameras is far easier than having to physically attach accelerometers or strain gauges to a structure, and can more easily reach physically difficult to access locations. With other recent improvements in computer vision, vision based systems have become quite popular in structural health monitoring applications [11]. Work related to the previously mentioned papers [4, 7] has demonstrated the use of video cameras and motion magnification for the measurement and characterization of civil infrastructure for the purposes of structural health monitoring [12–20]. The work in this paper compares the results of [2] to a new analysis performed with the method described in [1, 10, 14] (a slightly refined version of the method in [7–9]).

EXPERIMENTAL SETUP

Capture:

The data we analyze in this paper was originally captured by Chen et al. in [2]. We summarize their capture setup and analysis here, referring to the original publication for additional details.

The bridge was filmed from a distance of approximately 80 meters (260 feet) using a Point Grey camera mounted on a heavy tripod (see Figure 1). An accelerometer was also

Figure 2. Originally Fig 4 of [2], reproduced here with permission. (Left) An image of the camera during capture. Plastic cover is used to protect from rain. (Right) an example frame recorded from the camera, with labels indicating the placement of accelerometers and strain gauges on the bridge.

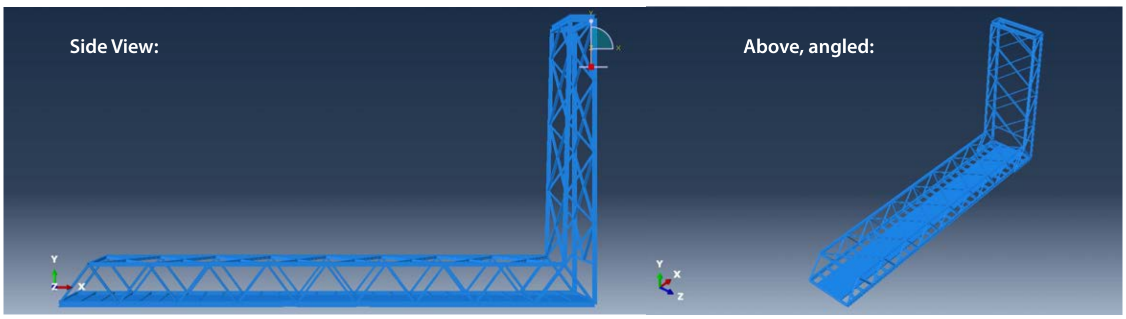


Figure 3. Finite element model built in ABAQUS based on plans provided by the New Hampshire Department of Transportation.

used to measure the camera's vibrations during capture. The bridge was instrumented with two additional accelerometers and two strain gauges prior to capture, providing contact measurements for comparison. Figure 2 (taken with permission from Chen et al. [2]) shows the camera setup and an example frame of captured video, as well as the locations of contact sensors on the bridge.

Finite-Element Model:

Chen et al. also constructed a finite-element (FE) model for the bridge based on plans provided by the New Hampshire Department of Transportation (NHDOT). Figure 3 shows rendered images of the model, which consists of $1,790$ linear beam elements, with $1,673$ nodes, and a total of $9,776$ unrestrained degrees of freedom. This model is used to predict mode shapes and frequencies based on the bridge's original design, serving as a rough proxy for its predicted behavior prior to construction, and a geometric point of comparison for operational mode shapes.

The original finite element model is no longer available, but we were able to obtain additional data and images from the original work to facilitate new comparisons.

PREVIOUS ANALYSIS

Figure 4. a) Screenshot of downsampled video frame. b) Subject mask used for modal imaging in this paper. Note that the background, a pillar sitting in front of the bridge, and a square portion of the bridge that contained a moving person, have all been excluded in the mask. c) Derived horizontal mask used when combining local signals in [2]. d) The corresponding derived vertical mask from [2].

Chen et al. analyzed mode shapes and frequencies based on a 33 second long video of the bridge containing vibrations caused by the impact of the center part of the bridge returning to its supports. The video was recorded in a single color channel at a resolution of 800x600 pixels and a sampling rate of 30 frames per second. Their analysis is based largely on the algorithm described in [13].

ALGORITHMIC COMPARISON WITH MODAL IMAGING

Here we describe some of the algorithmic differences between the method used in [2] and modal imaging [1].

Subject Masking:

Modal imaging leverages the observation that masking can serve two distinct purposes: the first is to indicate what part of the scene should be included in analysis, and the second is to weigh higher-confidence observations more heavily in that analysis. The first purpose is critical for video that contains moving objects other than the structure of interest. For example, the input bridge video shows a second, separate bridge in the distant background. In order to exclude this bridge from our analysis we need some indication that it is independent of our subject. In some of their experiments, Chen et al. [2] accomplish this by cropping the input video to some smaller region of interest. This strategy is able to remove the background bridge, but discards a great deal of useful data along with it. In modal imaging, we instead use a per-pixel mask to indicate what should be included in our calculations (see Figure 4b). This lets us precisely remove content that we do not want considered.

Chen et al. [2] do use a pixel mask in their processing of the video. However, this mask does not distinguish between subject and background content, and is instead used to exclude low-confidence local motion measurements, as we discuss below.

Weighing Local Signals

The reliability of local observations is strongly affected by the strength of local image gradients, which determine how well we can localize motion in different orientations. This relates to the well-known *aperture problem* in computer vision. As image gradients vary with orientation–even within a single spatial location–it is useful to decouple the weights of different orientations in our analysis. Algorithms based on the complex steerable pyramid [21] generally use local pyramid amplitudes to estimate confidence in different orientations. These amplitudes are very similar to oriented image gradients at different scales and are well-correlated with the noise of local displacement estimates.

The masking used in Chen et al. [2] is implemented as a threshold on local amplitudes. This effectively discards local information that falls below some confidence threshold, placing equal weight on anything above said threshold. If we cast the combination of local signals as a weighted sum of measurements and assume that noise is zero-mean, this approximately corresponds to weights of zero for every measurement with variance above a threshold, and uniform weights for all other measurements. In this light, it is easy to see how an appropriate choice of threshold could reduce the variance of our sum: the theoretically optimal weight for each local measurement should be inversely proportional to its variance, so we can think of thresholding as a highly quantized version of this optimal strategy. The weighting used in our implementation of modal imaging can be interpreted as a slightly closer approximation to optimal weights implemented locally in space. The motion signal for a given image location is a weighted combination of nearby observations with weights proportional to our confidence estimates and a Gaussian falloff with image distance (for details, see [1]). This produces non-binary weights that offer a closer approximation to inverse-variance weighting.

Combining Local Signals:

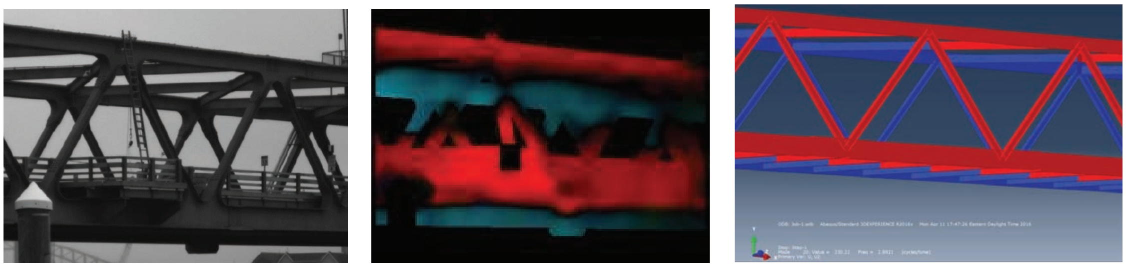
Another key difference between modal imaging and the method in [2] has to do with how spatially local information is aggregated to compute a frequency response curve. At a high level, the main difference is in what values are being averaged. Chen et al. sums local signals in the time domain, while our analysis computes the per-frequency sum over local power spectra. The advantage of summing over power spectra is that it lets local motions with opposite phase (as are common with modal vibration) add constructively. Normally, this comes at the cost of also causing zero-mean noise to add constructively. However, any noise that is zero-mean within the scale used for filtering local signals (as described in the previous section) should be greatly attenuated before local signals are combined.

Mode Visualization:

The local filtering and calculation of frequency responses in modal imaging lets us use a spatial slice of Fourier coefficients—called a *modal image*—to visualize mode shapes. These shapes are a useful tool for reasoning about the behavior of different modes, and could provide valuable insights for SHM. These modal images are especially useful for relating observed modes to those predicted by a finite element model, as observed frequencies may vary much more significantly than shapes.

RESULTS AND COMPARISON

Figure 6 shows a comparison of frequency response functions recovered by Chen et al. to those produced by our implementation of modal imaging. Figure 7 shows



Single frame of video | Recovered Vibration Mode Shape at 2.54Hz | Predicted Vibration Mode at 2.89Hz

Figure 5. Comparison of vertical mode shape recovered by modal imaging from video at 2.54 Hz with the predicted torsional vibration mode at 2.89 Hz obtained from a finite element model of the bridge originally reported in [2]

recovered shapes for sevaral modes. Figure 8 shows a vibration mode shape that was not discovered in [2], but became visible with modal imaging. We also show what we believe to be the corresponding predicted mode shape from the FE model.

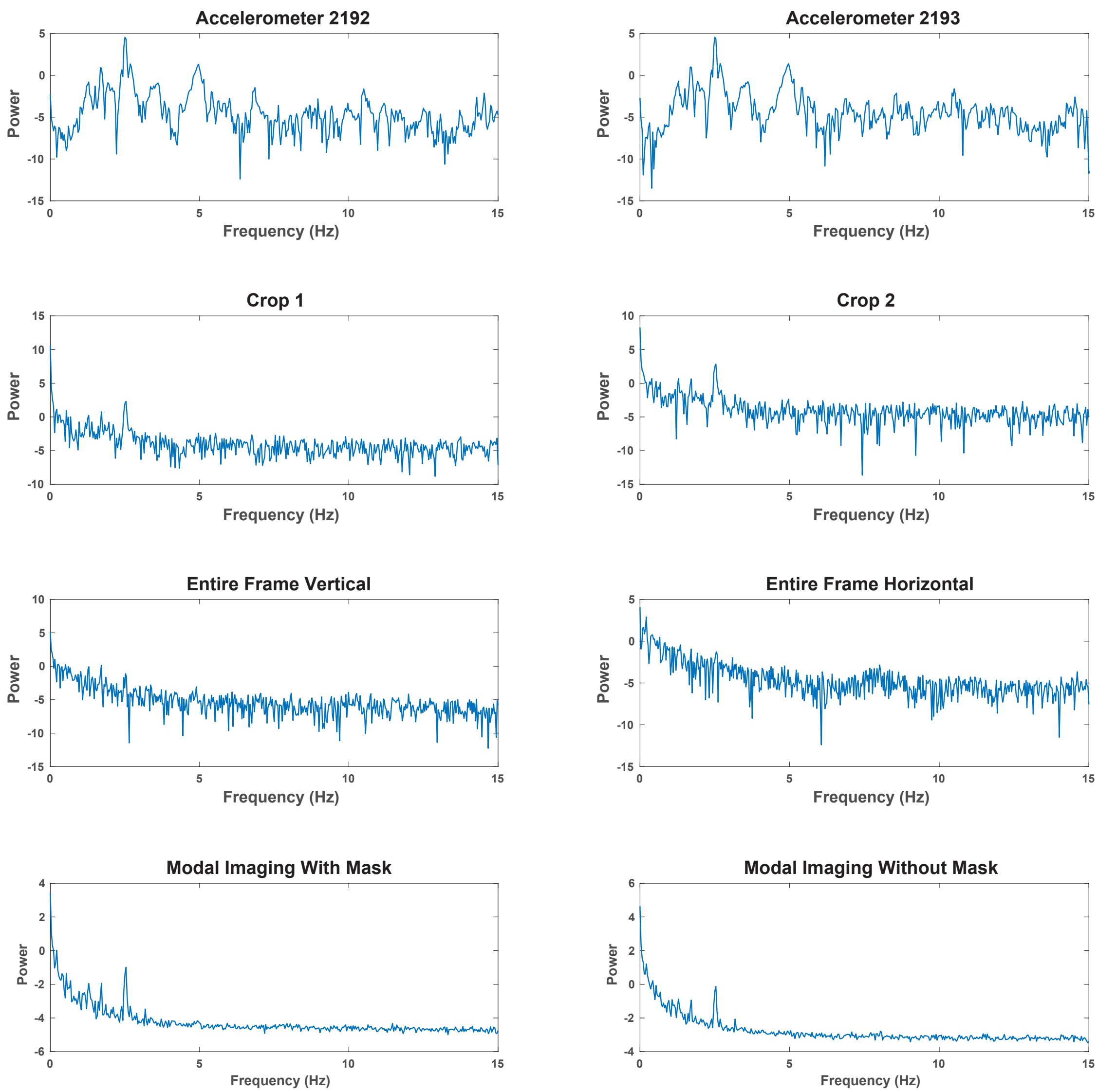


Figure 6. A comparison of the signals extracted from the (top) accelerometer sensors on the bridge, with the (top center) camera measurement from [2] of video crops from around those sensors, (bottom center) the signals extracted from the full video frame as reported in [2] and (bottom) the modal imaging technique processing the full frame video as recorded by the camera. Note that for these experiments, we cropped the original video to its first 28 seconds, as the end of the video contained significant undesirable motion. This appeared to improved the results for both methods.

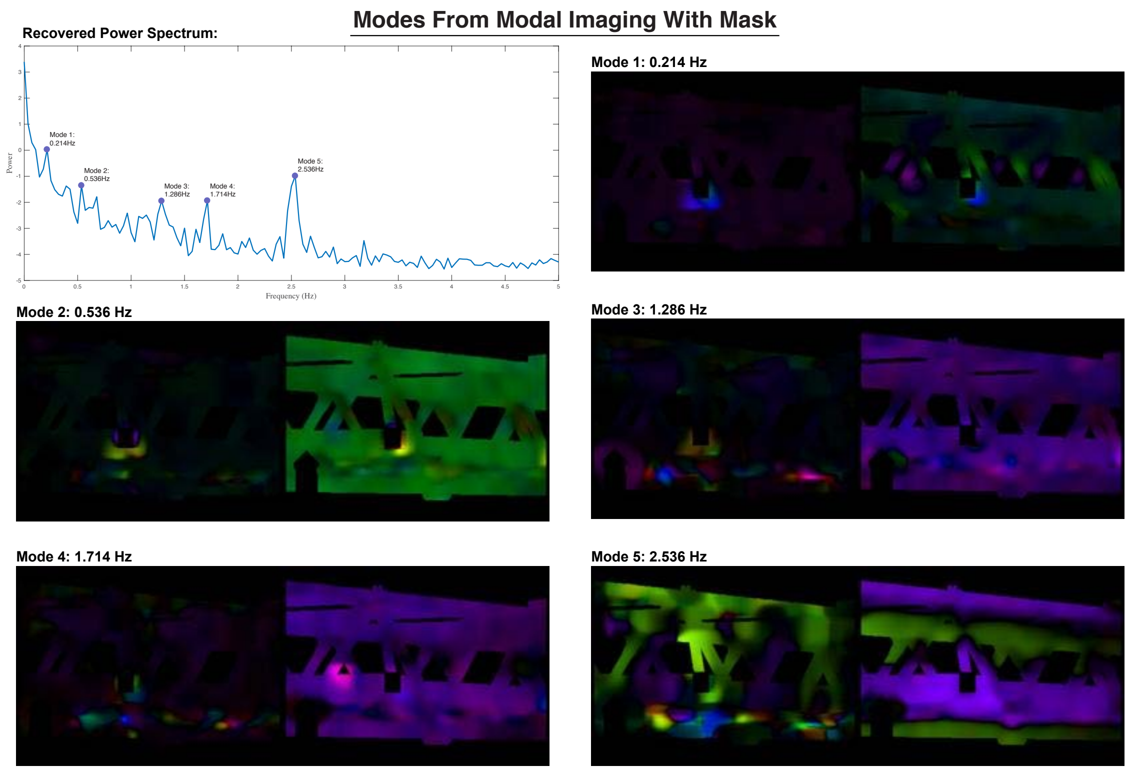


Figure 7. (Top Left) Power spectra extracted from modal imaging processing and other figures showing various extracted modes of the Portsmouth bridge.

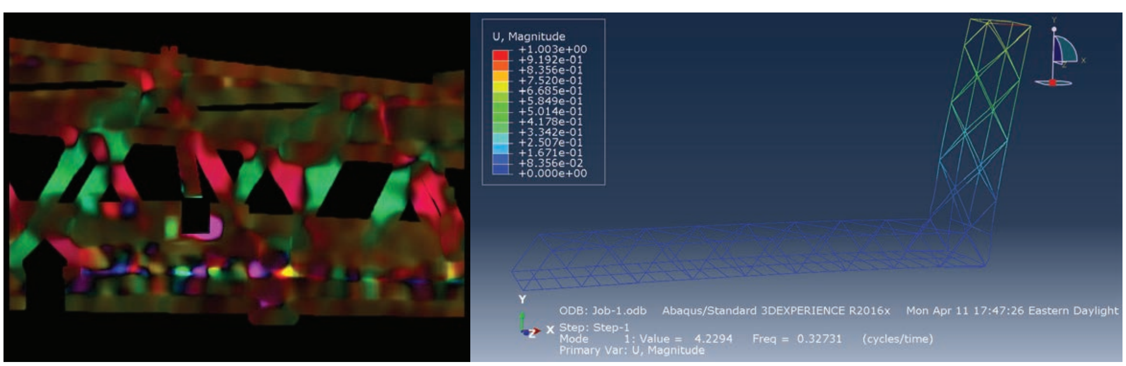


Figure 8. Using modal imaging, we detected a mode at approximately $0.21Hz$ that had not shown up in the analysis of [2]. The corresponding vertical mode shape (a brightened version of the one shown in Figure 7)is shown brightened on the left. On the right is what we believe to be a corresponding mode shape predicted by the finite element model. The opposite phase relationships shown in the modal image could be caused by shearing in response to motion in the vertical segment of the bridge (off camera).

REFERENCES

1. Davis, A. 2016. *Visual Vibration Analysis*, Ph.D. thesis, Massachusetts Institute of Technology.
2. Chen, J. G., T. M. Adams, H. Sun, E. S. Bell, and O. Bykztrk. 2018. “Camera-Based Vibration Measurement of the World War I Memorial Bridge in Portsmouth, New Hampshire,” *Journal of Structural Engineering*, 144(11):04018207, doi:10.1061/(ASCE)ST.1943-541X.0002203.
3. Wu, H.-Y., M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman. 2012. “Eulerian video magnification for revealing subtle changes in the world,” *ACM Transactions on Graphics (TOG)*, 31(4):65.
4. Wadhwa, N., M. Rubinstein, F. Durand, and W. T. Freeman. 2013. “Phase-based video motion processing,” *ACM Transactions on Graphics (TOG)*, 32(4):80.
5. Wadhwa, N., M. Rubinstein, F. Durand, and W. T. Freeman. 2014. “Riesz Pyramid for Fast Phase-Based Video Magnification,” in *Computational Photography (ICCP), 2014 IEEE International Conference on*, IEEE.
6. Wadhwa, N., J. G. Chen, J. B. Sellon, D. Wei, M. Rubinstein, R. Ghaffari, D. M. Freeman, O. Büyüköztürk, P. Wang, S. Sun, et al. 2017. “Motion microscopy for visualizing and quantifying small motions,” *Proceedings of the National Academy of Sciences*, 114(44):11639–11644.
7. Davis, A., M. Rubinstein, N. Wadhwa, G. J. Mysore, F. Durand, and W. T. Freeman. 2014. “The Visual Microphone: Passive Recovery of Sound from Video,” *ACM Trans. Graph.*, 33(4):79:1–79:10, ISSN 0730-0301, doi:10.1145/2601097.2601119.
8. Davis, A., K. L. Bouman, J. G. Chen, M. Rubinstein, F. Durand, and W. T. Freeman. 2015. “Visual Vibrometry: Estimating Material Properties From Small Motion in Video,” .
9. Davis*, A., K. L. Bouman*, J. G. Chen, M. Rubinstein, O. Bykztrk, F. Durand, and W. T. Freeman. 2017. “Visual Vibrometry: Estimating Material Properties from Small Motions in Video,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):732–745, ISSN 0162-8828, doi:10.1109/TPAMI.2016.2622271.
10. Davis, A., J. G. Chen, and F. Durand. 2015. “Image-space Modal Bases for Plausible Manipulation of Objects in Video,” *ACM Trans. Graph.*, 34(6):239:1–239:7, ISSN 0730-0301, doi:10.1145/2816795.2818095.
11. Spencer Jr, B. F., V. Hoskere, and Y. Narazaki. 2019. “Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring,” *Engineering*.
12. Chen, J. G., N. Wadhwa, Y.-J. Cha, F. Durand, W. T. Freeman, and O. Buyukozturk. 2015. “Modal identification of simple structures with high-speed video using motion magnification,” *Journal of Sound and Vibration*, 345:58–71.
13. Chen, J. G., A. Davis, N. Wadhwa, F. Durand, W. T. Freeman, and O. Büyüköztürk. 2017. “Video Camera–Based Vibration Measurement for Civil Infrastructure Applications,” *Journal of Infrastructure Systems*, 23(3).
14. Davis, A., J. G. Chen, O. Buyukozturk, F. Durand, and D. L. James. 2017. “Structural Health Monitoring from the Window Seat of a Passenger Airplane,” *10th International Workshop on Structural Health Monitoring (IWSHM 2015)*, doi:10.12783/shm2017/14234.
15. Cha, Y.-J., J. Chen, and O. Büyüköztürk. 2017. “Output-only computer vision based damage detection using phase-based optical flow and unscented Kalman filters,” *Engineering Structures*, 132:300–313.
16. Yang, Y., C. Dorn, T. Mancini, Z. Talken, G. Kenyon, C. Farrar, and D. Mascareñas. 2017. “Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification,” *Mechanical Systems and Signal Processing*, 85:567–

590.
17. Bao, Y., P. Seshadri, and S. Mahadevan. 2017. “Motion Magnication For Quantifying Aeroelastic Modes From High-Speed Videos,” in *58th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, p. 0869.
18. Sarrafi, A., P. Poozesh, C. Niezrecki, and Z. Mao. 2017. “Mode extraction on wind turbine blades via phase-based video motion estimation,” in *SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring*, International Society for Optics and Photonics, pp. 101710E–101710E.
19. Sarrafi, A. and Z. Mao. 2019. “Mapping Motion-Magnified Videos to Operating Deflection Shape Vectors Using Particle Filters,” in *Rotating Machinery, Optical Methods & Scanning LDV Methods, Volume 6*, Springer, pp. 75–83.
20. Yang, Y., L. Sanchez, H. Zhang, A. Roeder, J. Bowlan, J. Crochet, C. Farrar, and D. Mascareñas. 2019. “Estimation of full-field, full-order experimental modal model of cable vibration from digital video measurements with physics-guided unsupervised machine learning and computer vision,” *Structural Control and Health Monitoring*:e2358.
21. Simoncelli, E. P. and W. T. Freeman. 1995. “The steerable pyramid: a flexible architecture for multi-scale derivative computation,” in *Proceedings of the 1995 International Conference on Image Processing (Vol. 3)-Volume 3 - Volume 3*, IEEE Computer Society, Washington, DC, USA, ICIP ’95, ISBN 0-8186-7310-9, pp. 3444–.

DEStech Publications, Inc.

CONTRIBUTING AUTHOR COPYRIGHT RELEASE FORM

As author of the chapter/contribution titled “Modal Imaging of Portsmouth, New Hampshire Bridge for Video-Based Structural Health Monitoring”, to appear in the *Proceedings of Structural Health Monitoring 2019*, I hereby agree to the following:

1. To grant to DEStech Publications, Inc., 439 North Duke Street, Lancaster, PA, 17602, copyright of the above named chapter/contribution (for U.S. Government employees to the extent transferable), in print, electronic, and online formats. However, the undersigned reserve the following:
a. All proprietary rights other than copyright, such as patent rights.
b. The right to use all or part of this article in future works.

DEStech Publications thereby retains full and exclusive right to publish, market, and sell this material in any and all editions, in the English language or otherwise.

1 I warrant to DEStech Publications, Inc., that I am the (an) author of the above-named chapter/contribution and that I am the (a) copyright holder of the above-named chapter/contribution granted to DEStech Publications, Inc.

2 I warrant that, where necessary and required, I have obtained written permission for the use of any and all copyrighted materials used in the above-named chapter/contribution. I understand that I am responsible for all costs of gaining written permission for use of copyrighted materials.

3 I agree to assume full liability to DEStech Publications, Inc. and its licensee, and to hold DEStech Publications, Inc. harmless for any claim or suit filed against DEStech Publications, Inc. for violation of copyrighted material used in the above-named contribution.

Please sign and date this form and retain a copy for your records. Please include original form with your chapter/paper.

Thank you for your cooperation.

Please print name: ____Abe Davis______________________________________

Signed: ______________________________ Dated: __July 8, 2019_________

439 NORTH DUKE STREET • LANCASTER, PENNSYLVANIA 17602-4967, U.S.A. Toll Free: (866) 401-4337 • Tel: (717) 290-1660 • Fax: (717) 509-6100 E-mail: info@destechpub.com • Internet address: www.destechpub.com