Developments with Motion Magnification for Structural Modal Identification through Camera Video

Justin G. Chen*†, Neal Wadhwa†,**
Frédo Durand‡, William T. Freeman**‡, Oral Buyukozturk*‡**

*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology,
77 Massachusetts Avenue, Cambridge, MA 02139, USA
**Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology,
77 Massachusetts Avenue, Cambridge, MA 02139, USA
†Ph.D Candidate, ‡Professor

ABSTRACT

Non-contact measurement of the response of vibrating structures may be achieved using several different methods including the use of video cameras that offer flexibility in use and advantage in terms of cost. Videos can provide valuable qualitative information to an informed person, but quantitative measurements obtained using computer vision techniques are essential for structural assessment. Motion Magnification in videos refers to a collection of techniques that amplify small motions in videos in specified bands of frequencies for visualization, which can also be used to determine displacements of distinct edges of structures being measured. We will present recent developments in motion magnification for the modal identification of structures. A new algorithm based on the Riesz transform has been developed allowing for real-time application of motion magnification to normal-speed videos with similar quality to the previous computationally intensive phase-based algorithm. Displacement signals are extracted from strong edges in the video as a basis for the data necessary for modal identification. Methodologies for output-only modal analysis applicable to the large number of signals and short length signals are demonstrated on example videos of vibrating structures.

Keywords: Computer Vision, Non-contact, Modal Identification, Mode Shape, High Speed Video

1 INTRODUCTION

Non-contact techniques for measuring structures have only recently been available as a serious option. Laser vibrometers provide an accurate way of measuring the surface velocity of an object and are now widely used as the gold standard for non-contact vibration measurement [1]. For large structures, scanning laser vibrometers have the capability of measuring a large number of points in succession, however for phased measurements they require another stationary laser vibrometer to serve as a reference, or they must be continuously scanning systems [2]. Laser vibrometer arrays consisting of multiple laser vibrometers measuring simultaneously are also an option [3]. However in general, laser vibrometers may be prohibitively expensive for many applications.

Video cameras collect light from a scene of interest onto a sensor consisting of an array of pixels and can obtain data with high spatial density. Interpretation of this data can be complex, especially to translate a video into quantitative measurement of the vibration and displacements of a structure. Computer vision techniques and digital image correlation have been successfully used to measure displacements of structures and quantify vibrations in structures [4−8].

This paper presents developments in structural modal identification in videos using the family of algorithms called motion magnification [9−11]. These algorithms can visualize operational deflections shapes (ODS) in a vibrating structure by magnifying motions in a video in a particular frequency band [12]. The most recent development of interest is the development of a Riesz transform based method that allows for fast or real-time processing for phase-based Motion Magnification [13]. This will be demonstrated by visualizing the ODS of a cantilever beam by using a subsampling technique in real time. The other development is expanding the use of the previously developed displacement extraction algorithms to obtain a displacement at many pixels in a video of interest at any strong edges of a structure [12]. Two different methods of identifying ODS from the many signals extracted from the video are demonstrated: peak picking and frequency domain decomposition. For these two developments, real-time Motion Magnification and ODS identification with numerous signals from video, some theory on the underlying methodologies is presented, followed by experimental results, and a conclusion offering directions for future work.

2 DERIVATION

2.1 Real-time Motion Magnification

Recently, Wadhwa et al. [13] showed that Motion Magnification can be done in real-time. They decompose frames of a video into spatially-bandpassed subbands that correspond to different spatial scales. They then use the Riesz transform to compute an oriented local phase that approximately corresponds to the motion in the direction of the dominant orientation at every point, in every scale. This motion signal can be temporally filtered and amplified. More details about this technique can be found in their paper, but some of the key ideas are summarized here.

Subbands of an image are spatially bandpassed versions of that image. For example, the Fourier transform can be viewed as a way of decomposing an image $I(x)$ into a linear combination of complex sinusoids, each of which is a subband of that image. The phase of these subbands is linked to the global motion of the image $I(x)$ via the Fourier shift theorem

$$I(x) = \sum_{n=0}^{N} A_n e^{inx} \Rightarrow I(x-\delta) = \sum_{n=0}^{N} A_n e^{inx} e^{-in\delta}. \tag{1}$$

However, the Fourier transform is only able to decompose the image into global subbands that can only characterize global motions, while most motions in a video are local. Therefore, it is better to decompose the image into localized subbands using one of a variety of transforms like the Laplacian or steerable pyramids [14, 15]. Rather than decompose the image using filters that correspond to a single sinusoid like the Fourier transform, these transforms have filters that correspond to a wider range in the frequency domain making them more local in space.

In case of the Laplacian pyramid and the steerable pyramid, the produced subbands are real-valued signals, which are missing the corresponding imaginary part. In one dimension, the Hilbert transform, specified by transfer function

$$-i\frac{\omega}{|\omega|}, \tag{2}$$

can be applied to a real subband to yield a 90 degree phase-shifted version of it, which corresponds to the imaginary part. For example, the Hilbert transform of $\cos(x)$ is $\sin(x)$. These two signals can be combined to form a complex signal, which can be phase-shifted arbitrarily

$$\text{Real}((\cos(x) + i\sin(x))e^{i\delta}) = \cos(x-\delta) \tag{3}$$

In two dimensions, the Hilbert transform must be applied along a specific orientation. This is problematic as images do not have a preferred orientation. One solution used in Wadhwa et al. 2013 [11] was to first decompose the image into multiple scales and then further decompose each scale into several orientations using a complex steerable pyramid. This results in a large number of subbands, with 21x the total number of pixels of the original image. In addition, the filters required to divide the image into many orientations are often non-separable making their evaluation inefficient. All of this makes the processing unable to run in real-time. A second solution, allowing for real-time Motion Magnification, proposed in Wadhwa et al. 2014 was to instead use the Riesz transform, the two dimensional generalization of the Hilbert transform [13, 16, 17].

The Riesz transform is defined by a pair of filters with transfer functions

$$-i\frac{\omega_x}{\|\vec{\omega}\|}, -i\frac{\omega_y}{\|\vec{\omega}\|}. \tag{4}$$

The direction invariance follows from the fact that the Riesz transform can be steered to an arbitrary orientation. For example, the Riesz transform of each frame of the image sequence $I(x, y, t) := \cos(\omega_x x + \omega_y y + \delta(t))$ is

$$R_1 := \sin(\omega_x x + \omega_y y + \delta(t))\frac{\omega_x}{\sqrt{\omega_x^2 + \omega_y^2}} \text{ and } R_2 := \sin(\omega_x x + \omega_y y + \delta(t))\frac{\omega_y}{\sqrt{\omega_x^2 + \omega_y^2}} \tag{5}$$

This can be steered to an arbitrary direction θ via the matrix multiplication

$$\begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 \end{pmatrix}. \tag{6}$$

giving a Hilbert transform in the specified direction. Of particular interest is if the Riesz transform is steered to the dominant orientation θ_0, given by $\tan^{-1}(R_2/R_1)$. The result of the matrix multiplication in Eq. 6 is a vector with only one non-zero entry, which in our case is

$$\sin(\omega_x x + \omega_y y + \delta(t)), \tag{7}$$

the Hilbert transform along the dominant orientation of the input image sequence. Like in 1D, this can now be combined with the original image sequence to form a complex signal which has phase

$$\omega_x x + \omega_y y + \delta(t). \tag{8}$$

This can be temporally filtered to remove the DC component $\omega_x x + \omega_y y$. The remainder $\delta(t)$ can be amplified and used to phase shift the complex signal, whose real part will be a motion magnified version of the original image sequence $I(x, y, t)$. That is,

$$\text{Real}((\cos(\omega_x x + \omega_y y + \delta(t)) + i\sin(\omega_x x + \omega_y y + \delta(t)))e^{i\alpha\delta(t)}) = \cos(\omega_x x + \omega_y - (1 + \alpha)\delta(t)). \tag{9}$$

More details about this processing can be found in Wadhwa et al. 2014[13]. In practice, on a standard laptop, this processing runs at about 35 frames per second.

So far, we have described processing that amplifies small motions in videos. However, many ODS occur at temporal frequencies that are too fast to see at 35 frames per second. To handle these faster vibrations, we stroboscopically sample the scene by choosing a sampling rate that will alias the ODS frequencies to approximately 2.5Hz. In particular, for an ODS frequency F, we choose a frame rate f, such that

$$2.5\text{Hz} \approx |\text{mod}(F + f/2, f) - f/2|. \tag{10}$$

We automatically choose f by finding the f between 17Hz and 25Hz that minimizes the difference between the right and left sides of Eq. 10.

2.2 Operational Deflection Shape Extraction Methods Adapted to Camera Data

In previous work displacements were extracted from a cropped region of a video and processed to give a displacement signal that was representative of that region [12]. The details are given in [12] but the general procedure is to downsample the video to average out noise, use a quadrature pair of filters to extract the local phase information in the video, determine the locations of the strong edges, obtain displacement signals from the local phase information, and then use a weighted average to obtain a single displacement signal for the whole video. In this procedure, once the displacement signals are obtained from the local phase, each pixel in the video at the location of a strong edge has a displacement signal associated with it. These displacement signals may be somewhat noisier than the averaged result, however the benefit is high spatial sampling over the strong edges of structures in the video. We can extract high spatial resolution ODS from the displacement signals obtained from the videos. For a typical video this may result in 500 - 3000 displacement signals, which is many more than the typical accelerometer-based measurement. This presents a special challenge when performing mode shape identification, as some methods may not scale well with increasing numbers of signals. Two methods were implemented in an output only manner with use of a broadband excitation, peak picking from fast Fourier transform (FFT) spectra and frequency domain decomposition.

Peak picking to obtain ODS from a collection of signals involves taking the FFT of the displacement time signals and picking out peaks in the frequency domain that correspond to potential resonant modes. Cross power spectral densities (PSD) are taken with reference to a single signal to determine the phase difference between the displacement signals at different frequencies. This information combined with the normalized magnitudes of the FFT at the picked frequencies create the ODS. The downside to this method is that it does not work well with closely spaced modes and it is difficult to identify non-physical harmonics in the response [18].

Frequency domain decomposition (FDD) is a method used for modal identification of output-only systems which addresses some of the drawbacks of the peak picking method [18]. It involves first calculating the spectral matrix, a matrix containing the cross PSD between all the output signals. The singular value decomposition (SVD) of the spectral matrix at each frequency is taken to decompose the spectral matrix into a set of auto spectral density functions as singular values, where strong resonant peaks in the singular values correspond to possible resonant modes and the singular vectors are the corresponding ODS. The benefit is that this algorithm performs than peak picking when signals are noisy and there is better discrimination for closely spaced resonant modes and the possible identification of non-physical harmonics [18, 19].

3 RESULTS

3.1 Real-time Motion Magnification

The real-time motion processing in Wadhwa et al. [13] can run at 35 frames per second for a 640x400 image. As mentioned in Section 2.1, vibrations occurring over the Nyquist frequency of 17.5Hz can be visualized by subsampling the video at a rate to alias the modal frequencies to approximately 2.5Hz. This sampling requires low exposure times to reduce motion blur. We use a bright light to provide sufficient illumination for an exposure time of 0.3 milliseconds. The camera is a color Point Grey Grasshopper 3 (GS3-U3-23S6M-C) that is connected to a laptop with a quad-core 2.7GHz Intel i7-3820QM processor. The processing only uses a single core of this processor.

We use the processing to visualize two of the ODS of a damaged cantilevered beam when it is struck by a hammer near the support in the flexible direction. The results are shown in Fig. 1. The camera is set to sampling rates of 21.2Hz for the one node ODS and 24.5Hz for the two node ODS. The modal frequencies are 87Hz and 248Hz respectively and they get aliased to 2.5Hz and 2.9Hz. However, the spatial shape of the motions at these frequencies still correspond to the original operational deflection shape.

(a) Experimental Setup (b) Raw Frame from Camera (c) One node ODS (d) Two node ODS

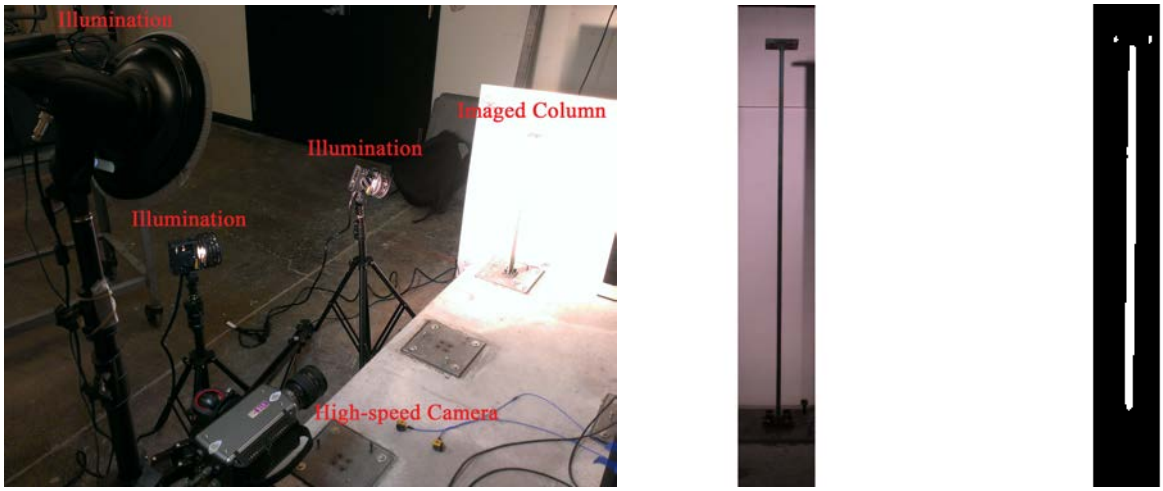
Fig. 1 Real-time motion magnification is used to visualize the second and third bending operational deflection shapes of a cantilevered beam. The experimental setup is shown in (a). An ordinary laptop processes data from the camera in real-time. An unprocessed frame from the camera is shown (b) and two frames from processed videos are shown in (c) and (d). In (b,c,d), the frame has been rotated 90 degrees counter-clockwise and the contrast has been enhanced.

3.2 Quantitative ODS Extraction from Camera Video

The experimental setup used to test the ODS identification methods involved a high-speed camera, cantilever beam, and sources of illumination as shown in Fig. 2(a). The column was struck with an impact hammer on the bottom third to induce vibrations in as many bending modes as possible. Video was recorded at 2000 frames per second with a full frame resolution of 1248×200 using the high speed camera with a frame shown in Fig. 2(b) . The video was downsampled by a factor of two such that after the processing, the valid frame was 304×42 (12768) pixels in the video and from that, edges were strong enough to obtain signals for 1234 pixels in the image. The mask shown in Fig. 2(c) has in white the pixels where displacement signals were obtained.

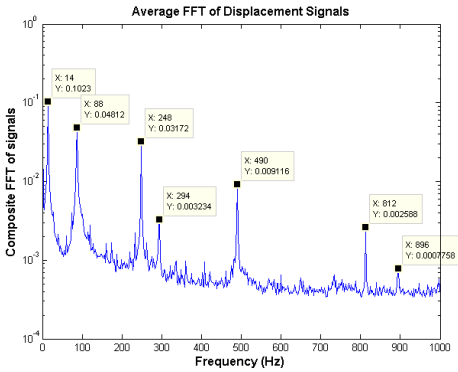
From the displacements the ODS were identified using the two methods described in the derivation section, peak picking from FFT spectra, and frequency domain decomposition. The spectra obtained from the displacement signals are shown in Fig. 3 with the one used for the peak picking technique in (a) and the singular values given in (b) and (c). The frequencies determined to be candidates for resonant modes are the same for both methods, which makes sense as the FDD method uses the same FFTs of the displacement signals to generate the singular values. The results of the ODS identification are shown in Fig. 4, with 2D plots showing the normalized magnitude of the ODS at each identified frequency and the phase relationships between the signals.

The resulting identified ODS for both cases end up being the first five bending modes at 14 Hz, 88 Hz, 248 Hz, 490 Hz, and 812 Hz and the first two torsional modes at 294 Hz and 896 Hz. Comparing the two methods the results are almost identical for the five bending modes. They are also similar for the torsional modes, albeit with the phase information more varied in the case of the peak picking method. A more detailed comparison of the identified ODS is given in Fig. 5 where the ODS magnitudes are averaged in horizontal rows to give a single value for each vertical sample which results in a more visually interpretable 1D mode shapes for the column. The ODS are shown in blue for the result from the FFT peak picking method and red for FDD. For the bending modes, the results are almost identical, but for the torsional modes the differences are more clear, especially with the second torsional mode where the peak picking ODS is much noisier than the FDD ODS. The MAC values comparing the identified ODS are given in Table 1 with the bending and torsional modes denoted by either B or T. The MAC values quantitatively match the qualitative observations previously described.

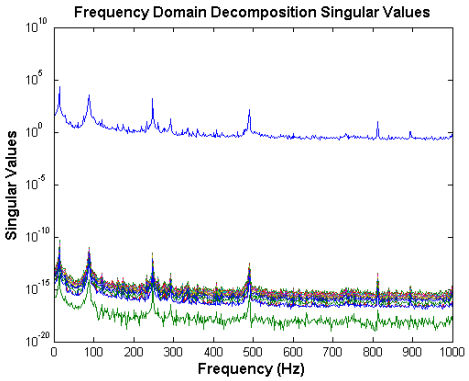


(a) Experimental Setup (b) Raw Frame from Camera (c) Pixel Mask

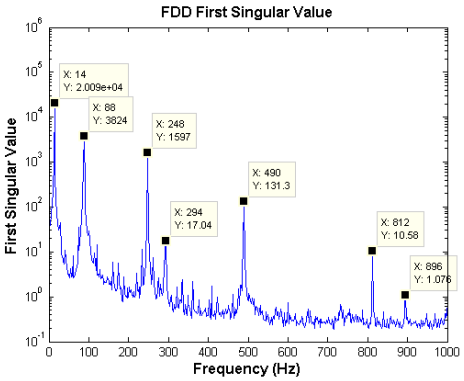
Fig. 2 Setup for ODS identification of a cantilever beam: (a) a picture of the experimental setup, (b) a frame from the camera video, and (c) the mask showing pixels with extracted displacements



(a) Average of FFT of all displacement signals

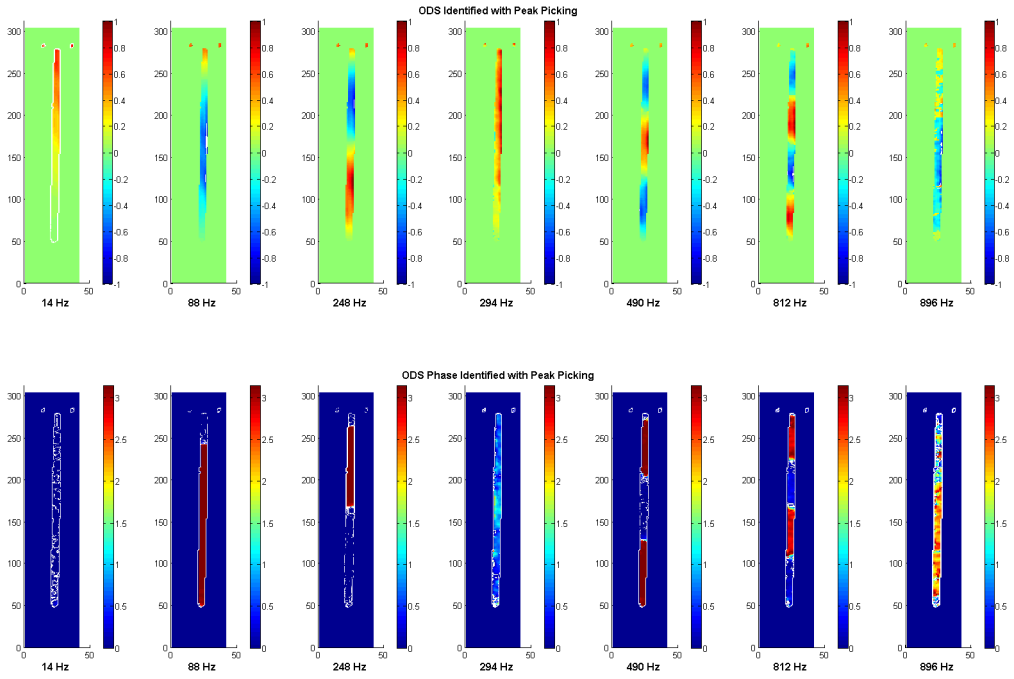


(b) Singular values from FDD

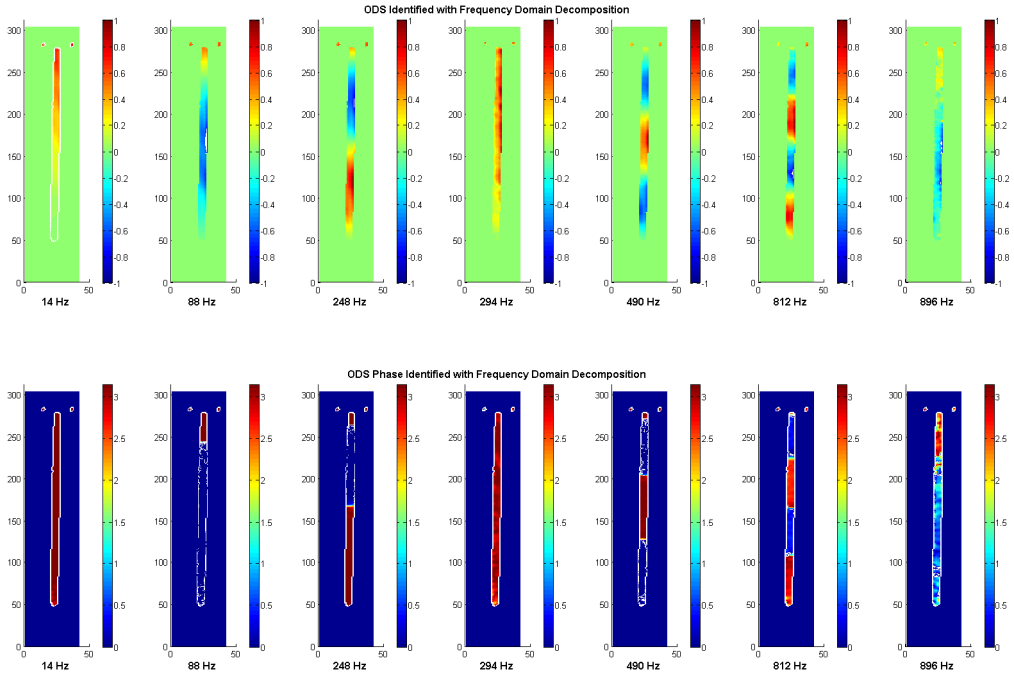


(c) First singular value from FDD

Fig. 3 Spectra used to determine potential resonant modes of interest, (a) average FFT of displacement signals and selected frequencies of interest, (b) singular values as a result of FDD, and (c) first singular value and selected frequencies of interest



(a) Normalized ODS and phases identified with peak picking



(b) Normalized ODS and phases identified with FDD

Fig. 4 Identified ODS using the (a) peak picking method and the (b) frequency domain decomposition method

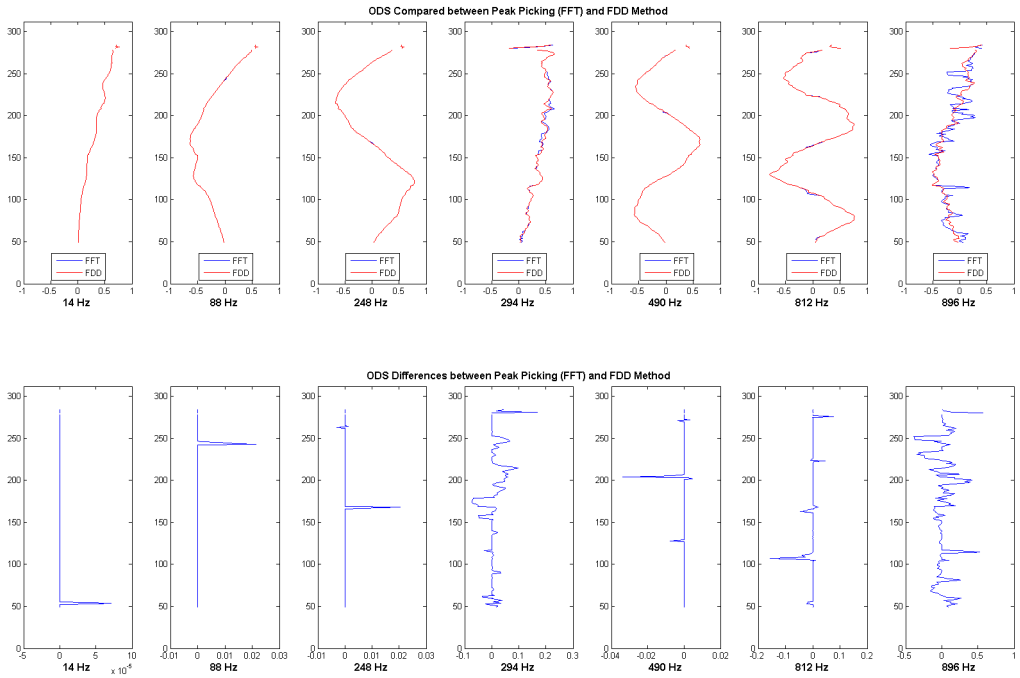


Fig. 5 Comparison of video identified ODS

Table 1 MAC values comparing ODS identified with FDD with peak picking (FFT) techniques for the various bending (B) and torsional (T) modes

ODS	FFT B1	FFT B2	FFT B3	FFT B4	FFT B5	FFT T1	FFT T2
FDD B1	100.00%	7.75%	4.38%	1.17%	0.00%	84.50%	0.41%
FDD B2	7.78%	100.00%	1.56%	2.53%	1.19%	33.61%	63.21%
FDD B3	4.38%	1.56%	100.00%	0.19%	5.07%	1.38%	20.33%
FDD B4	1.15%	2.57%	0.19%	100.00%	0.09%	1.96%	1.03%
FDD B5	0.00%	1.30%	4.85%	0.12%	99.91%	0.18%	0.54%
FDD T1	83.34%	34.91%	1.07%	1.58%	0.09%	99.52%	13.48%
FDD T2	1.08%	78.09%	25.88%	3.34%	0.10%	14.37%	76.41%

4 CONCLUSION

In this paper we have presented two developments in the use of motion magnification algorithms with video for structural model identification. The first development is the demonstration of real-time Motion Magnification of a cantilever beam using a new technique involving Riesz pyramids for video processing. Using a sub sampling technique, the second and third vibrational mode are visualized in real-time. The second development is the application of modal identification techniques to the large number of displacements extractable from video of a cantilever beam. Upwards of 1000 displacement signals are extracted from video of a cantilever beam and they are processed with both the peak picking and frequency domain decomposition techniques to identify the first five bending modes and first two torsional modes. The techniques provide comparable results except in the case of the second torsional mode which has less signal than the other modes. The hope is that these developments bring closer the reality of using cameras for modal inspection of structures.

ACKNOWLEDGEMENTS

The authors acknowledge the support provided by Royal Dutch Shell through the MIT Energy Initiative, and thank chief scientists Dr. Dirk Smit, Dr. Sergio Kapusta, project managers Dr. Keng Yap and Dr. Yile Li, and Shell-MIT liaison Dr. Jonathan Kane for their oversight of this work. We also acknowledge Dr. Michael Feng and Draper Laboratory for providing experimental equipment. At the time of this work, Neal Wadhwa was supported by the MIT Department of Mathematics and the NSF Graduate Research Fellowship Program under Grant No. 1122374. Special thanks are due to Reza Mohammadi Ghazi, James Long, and Young-Jin Cha for their help with experimental collection of the data.

REFERENCES

[1] **Castellini, P., Martarelli, M.** and **Tomasini, E.**, *Laser Doppler Vibrometry: Development of advanced solutions answering to technology's needs*, Mechanical Systems and Signal Processing, Vol. 20, No. 6, pp. 1265--1285, 2006.

[2] **Stanbridge, A.** and **Ewins, D.**, *Modal testing using a scanning laser Doppler vibrometer*, Mechanical Systems and Signal Processing, Vol. 13, No. 2, pp. 255--270, 1999.

[3] **Aranchuk, V., Lal, A., Sabatier, J. M.** and **Hess, C.**, *Multi-beam laser Doppler vibrometer for landmine detection*, Optical Engineering, Vol. 45, No. 10, pp. 104302--104302, 2006.

[4] **Caetano, E., Silva, S.** and **Bateira, J.**, *A vision system for vibration monitoring of civil engineering structures*, Experimental Techniques, Vol. 35, No. 4, pp. 74--82, 2011.

[5] **Lee, J. J.** and **Shinozuka, M.**, *A vision-based system for remote sensing of bridge displacement*, Ndt & E International, Vol. 39, No. 5, pp. 425--431, 2006.

[6] **Wahbeh, A. M., Caffrey, J. P.** and **Masri, S. F.**, *A vision-based approach for the direct measurement of displacements in vibrating systems*, Smart Materials and Structures, Vol. 12, No. 5, pp. 785, 2003.

[7] **Patsias, S.** and **Staszewskiy, W.**, *Damage detection using optical measurements and wavelets*, Structural Health Monitoring, Vol. 1, No. 1, pp. 5--22, 2002.

[8] **Ferrer, B., Espinosa, J., Roig, A. B., Perez, J.** and **Mas, D.**, *Vibration frequency measurement using a local multithreshold technique*, Optics express, Vol. 21, No. 22, pp. 26198--26208, 2013.

[9] **Liu, C., Torralba, A., Freeman, W. T., Durand, F.** and **Adelson, E. H.**, *Motion magnification*, ACM Trans. Graph., Vol. 24, pp. 519--526, Jul 2005.

[10] **Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F.** and **Freeman, W.**, *Eulerian Video Magnification for Revealing Subtle Changes in the World*, ACM Trans. Graph. (Proc. SIGGRAPH), Vol. 31, aug 2012.

[11] **Wadhwa, N., Rubinstein, M., Durand, F.** and **Freeman, W. T.**, *Phase-Based Video Motion Processing*, ACM Trans. Graph. (Proceedings SIGGRAPH 2013), Vol. 32, No. 4, 2013.

[12] **Chen, J. G., Wadhwa, N., Cha, Y.-J., Durand, F., Freeman, W. T.** and **Buyukozturk, O.**, *Structural modal identification through high speed camera video: Motion magnification*, *Topics in Modal Analysis I, Volume 7*, pp. 191--197, Springer, 2014.

[13] **Wadhwa, N., Rubinstein, M., Durand, F.** and **Freeman, W. T.**, *Riesz Pyramids for Fast Phase-Based Video Magnification*, 2014.

[14] **Burt, P.** and **Adelson, E.**, *The Laplacian pyramid as a compact image code*, IEEE Trans. Commun., Vol. 31, No. 4, pp. 532--540, 1983.

[15] **Simoncelli, E. P.** and **Freeman, W. T.**, *The steerable pyramid: a flexible architecture for multi-scale derivative computation*, *Proceedings of the 1995 International Conference on Image Processing (Vol. 3)-Volume 3 - Volume 3*, ICIP '95, pp. 3444--, IEEE Computer Society, Washington, DC, USA, 1995.

[16] **Felsberg, M.** and **Sommer, G.**, *The monogenic signal*, IEEE Trans. Signal Process., Vol. 49, No. 12, pp. 3136--3144, 2001.

[17] **Unser, M., Sage, D.** and **Van De Ville, D.**, *Multiresolution monogenic signal analysis using the Riesz--Laplace wavelet transform*, IEEE Trans. Image Process., Vol. 18, No. 11, pp. 2402--2418, 2009.

[18] **Brincker, R., Zhang, L.** and **Andersen, P.**, *Modal identification of output-only systems using frequency domain decomposition*, Smart materials and structures, Vol. 10, No. 3, pp. 441, 2001.

[19] **Brincker, R., Andersen, P.** and **Møller, N.**, *An indicator for separation of structural and harmonic modes in output-only modal testing*, *International Modal Analysis Conference (IMAC XVIII), San Antonio, TX, Society for Engineering Mechanics, Bethel, CT*, pp. 1649--1654, 2000.