# Chapter 7
# Regime Sorting for Multiscale Vibrations and Phase-Based Motion Extraction

**Sean Collier and Tyler Dare**

**Abstract** While phase-based motion magnification has shown great success in enhancing sub-pixel motion, super-pixel motion has proven more difficult to work with. As "small motion" is usually a major assumption in the derivation of optical flow, large-scale super-pixel motion has been shown to cause artifacts and limit the usability of the method in both magnification and extraction applications. Viable methods do exist and have shown promise for scenarios where unimportant large motions are present in an otherwise ideal video. Likewise, current rules of thumb suggest utilizing a pyramid approach, where the images are downsampled until the motion becomes sub-pixel and thus avoids the issue entirely. However, the approaches for effectively removing these large-scale, super-pixel motions neglect their potential importance for objects which exhibit *both* super- and sub-pixel motions of interest. Further, for such objects, downsampling could degrade and even obliterate the relevant sub-pixel motion. Therefore, an approach is conceived and developed exploiting this degradation in conjunction with complexity pursuit (a blind source separation technique) to allow for more effective and purposeful processing using the complex steerable pyramid. With greater information in the beginning, rules of thumb and temporal bandpassing can be utilized more effectively to extract viable motion measurements in *both* regimes of pixel motion.

**Keywords** Motion extraction · Sorting · Super-pixel · Sub-pixel · Phase-based

## 7.1 Introduction

Phase-based motion magnification (PBMM) and phase-based motion extraction (PBME) are notoriously plagued by large motions that inevitably break the small-motion assumptions of the underlying theory. Being an issue across much of optical flow, a few common techniques do exist to handle these in an effective manner. The first and seemingly less used is a Newton's method approach, wherein the motion is estimated from the raw video under the assumption that the motions are small/linear; the images are shifted by the extracted linear portion and the residual video is processed again in an iterative fashion until the error converges. The more commonly used rule of thumb for dealing with large motions is to do a coarse-to-fine (pyramid) approach. This renders super-pixel motion sub-pixel and within the bounds of the method; however, initially sub-pixel motion becomes even smaller, and is eventually pushed beyond the lower bound of the method, unintentionally leading to signal degradation and (more extremely) signal obliteration.

Across the literature, many of the test articles examined using PBMM/PBME have been structures exhibiting small motions in an otherwise ideal scene and therefore provide ideal cases for the processing methods. In general, and especially outside of a lab setting, objects tend to exhibit motion across both regimes of pixel motion—that is, super- and sub-pixel—simultaneously. For example, small vibrations of a structure undergoing rigid body motion are obscured when processed as is through PBMM. The Computer Science and Artificial Intelligence Laboratory (CSAIL) Group at MIT was aware of this problem when they published [1], with the suggestion of handling this motion by choosing not to magnify frequencies where the motion exceeds some threshold set by the user. This works if the large, super-pixel motion and sub-pixel motions are frequency independent and the large motions are otherwise unimportant. For the instances where frequencies may overlap, a different solution was required, and many potential methods have been published. Dynamic video magnification (DVMAG) [2] was the first, and relies on scene segmentation similar to the original Lagrangian, intensity-based motion magnification results from 2005 [3]. While effective, the method is exceptionally complicated.

S. Collier (✉) · T. Dare
Graduate Program in Acoustics College of Engineering, The Pennsylvania State University, State College, PA, USA
e-mail: smc604@psu.edu; tpd10@psu.edu



D. D. Maio, J. Baqersad (eds.), *Rotating Machinery, Optical Methods & Scanning LDV Methods, Volume 6*, Conference Proceedings of the Society for Experimental Mechanics Series, https://doi.org/10.1007/978-3-031-04098-6_7

Another group of solutions rely on a convenient operator property and assuming large motions are roughly linear between time steps (a fact especially true for very fast sampling), thereby allowing for removal entirely by using higher-order derivatives of the motion. Hence, methods such as video acceleration magnification [4] and video jerk magnification [5] were introduced to accomplish magnification of small motions while simultaneously avoiding the magnification of larger ones. Again, these methods are effective but limited in scope by the order-of-motion assumptions associated with them. Other methods use frequency-domain representations instead of a time-domain perspective, where one example uses amplitude-based filters to remove motions above a certain threshold, much like the original CSAIL suggestion [6]. Likewise, [7] assumes background motions are complex in nature with varying frequency components that are inconsistent over time; thus, only stable peaks are kept and magnified, though this doesn't fix the large motion issue if it is also stable.

All of these methods have effectively demonstrated the ability to magnify sub-pixel motions while in the presence of larger, super-pixel motions. Importantly, many of these solutions exist on bases that conclude the large motions are otherwise unimportant and can be removed without concern, as evidenced in [13]. This dismissal of super-pixel motion is problematic in situations such as experimental modal analysis, where objects are entirely capable of exhibiting super- and sub-pixel motions simultaneously. With this in mind, the above methods for handling large motions are inappropriate and could lead to removal of important, often fundamental behavior in a vibrating structure. Unfortunately, the disadvantage of downsampling the video to render the large motions sub-pixel creates a similar problem, where now the small motions are removed, albeit inadvertently. Therefore, a method was conceived and developed to effectively sort these motions into their respective regimes to then be processed in a context more appropriate for the motion size.

The paper proceeds as follows: Theory briefly introduces the overall processing method, Implementation describes what to expect from the approach when motion exists in different regimes and how to interpret results, and Examples follow for synthetic and experimental cases, with tips for more Complicated Scenes and then Conclusions.

## 7.2 Theory

This method is based on the phase-based processing algorithm introduced by CSAIL in 2014 [1]. The fundamental aspect of PBMM/PBME is the complex steerable pyramid (CSP)—an image pyramid consisting of multiple complex steered filters at varying scales. A diagram can be seen in Fig. 7.1.

This method is based on a small-motion assumption, and is considered valid for motions spanning $2-3$ px frame to frame [8]. In part, the pyramid approach helps to handle the large motions by representing them across reduced resolution, such

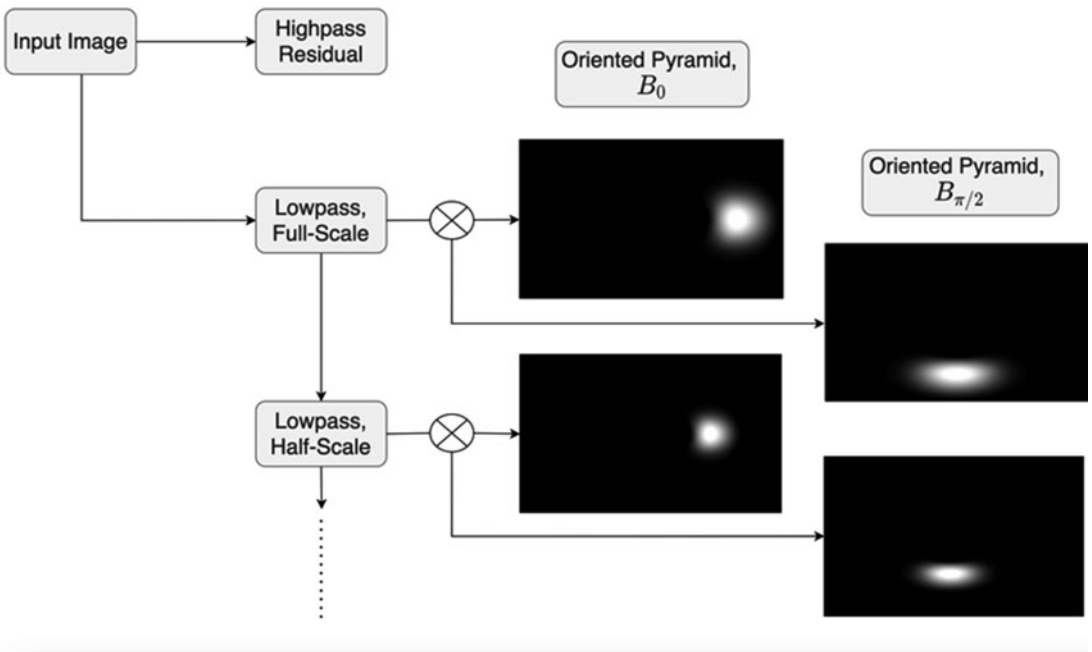


**Fig. 7.1** Schematic of a two-layer, two-orientation ($\theta = 0, \pi/2$) complex steerable pyramid

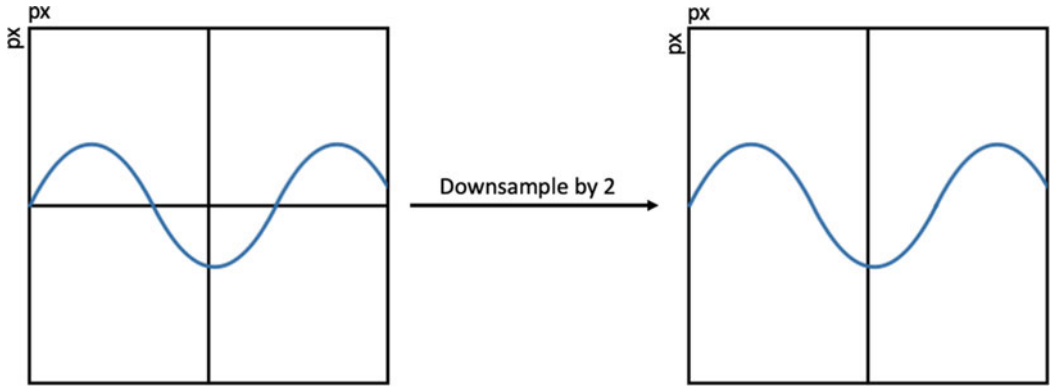


**Fig. 7.2** Example of how downsampling reduces pixel motion, as the same physical scene is captured with fewer pixels

that motion which originally spanned multiple pixels is represented with fewer spatial points, effectively making it smaller. For example, motion that originally spanned 4 pixels will span 2 when downsampled by a factor of 2. This is represented in Fig. 7.2 and helps illustrate the utility of such an approach.

While effective in helping to handle the larger motions, the originally sub-pixel motion becomes further reduced and suppressed into the noise floor of the image. This means the sub-pixel motions are sometimes reduced in signal quality and may even be suppressed so much that they are no longer detectable in the signal. Initially, this is an attribute that is potentially misleading and could lead to faulted or uninformed conclusions; however, this degradation can be exploited as an effective preprocessing measure. It will be referred to as the CSP detriment.

To help mitigate this effect while keeping the processing intuitive and familiar, the motion data gathered from the CSP is processed through complexity pursuit (CP; a blind source separation technique), with an implementation introduced by John Stone in 2001 [9]. In short, complexity pursuit is an addendum to the familiar decomposition methods commonly used in the field, those including the singular value decomposition (SVD), proper orthogonal decomposition (POD), and principal component analysis (PCA). While effective in separating modal data into basis functions that usually represent mode shapes, generally these methods separate the data into a *convenient* basis, which Yang et al. points out only coincides with the modes when there exists uniform mass distribution [11].

Given the phase information (displacements) from processing a video of a vibrating structure through the CSP, let $\delta(x,t)$ be expressed as a superposition of the response from individual modes, as

$$\delta(x,t) = \Phi(x)q(t) = \sum_{i=1}^{n} \phi_i(x) q_i(t) \tag{7.1}$$

where $n$ is the mode number and $\Phi$ the mode shape matrix with $\phi_i(x)$ being the $i$th mode and $q_i(t)$ being the corresponding time history. Applying the singular value decomposition (SVD) method to $\delta(x,t)$ will yield:

$$\delta(x,t) = U\Sigma V^T, \tag{7.2}$$

a dimension reduction that ideally results in the mode shapes of the vibrating system. Here, the eigenvectors contained within $U$ (and $V$) contain a set of bases for the vector space represented by the matrix and should correspond with the mode shapes, with the singular values contained in $\Sigma$; however, as already mentioned this may not necessarily be true and may otherwise produce a basis of convenience. Thus, consider the results of the SVD to generally still be a mixture of modes such that

$$\eta(t) = Yq(t) = \sum_{i=1}^{n} \upsilon_i q_i(t) \tag{7.3}$$

just as above, but for $\eta$ being the resultant of $\eta(t) = U^T\delta$. Then it follows that

$$\sum_{i=1}^{n} \phi_i q(t) = \delta = U\eta = U\mathrm{Y}q(t) = U\sum_{i=1}^{n} \upsilon_i q_i(t) = \sum_{i=1}^{n} (U\upsilon_i)\, q_i(t) \tag{7.4}$$

which, by comparing the left- and right-hand side of Eq. (7.4), implies $\phi_i = U\upsilon_i$, forcing the results of the SVD to align with actual modes of the system. Likewise, it is necessary to decouple the temporal component as well. This follows as $q(t) = W\eta(t)$, with $W$ (the de-mixing matrix) being estimated by Stone's implementation of complexity pursuit. Again, by comparison it follows that $Y = W^{-1}$, thereby allowing for complete decoupling of modes and temporal components. Typically, $U$ is reduced to only the first $r$ eigenvectors; in this study, a value of $r$ between 5 and 10 was used, as this was the number of modes clearly pulled from the camera data when processing a few preliminary pixels without further measures to increase SNR. This value can easily be changed, though this may result in a messier histogram if $r$ is quite large and may hinder the sorting in some ways; similarly, a smaller value used for a complicated scene may also result in weaker modes being overlooked by more prominent motions and not being adequately represented in the CP output, as orientation and pyramid scale will heavily influence what motions are picked out by the algorithm.

A more detailed description of the implementation of complexity pursuit can be found in [10] and Stone's algorithm in [9]. Likewise, this fact has already been discussed in great detail through the work of Yongchao Yang et al. [10–12], where CP is used to blindly choose ranges for the temporal bandpassing of the CSP. Instead of using as a mid-processing step as done before, CP is used as a pre-processing step to supplement a better-informed handling of algorithm parameters such as layer selection for motion extraction.

## 7.3 Implementation

For the proposed algorithm, signal degradation paired with CP (CSP + CP) is exploited to sort the motions into their respective regimes. The large (super-pixel) motions will present in (at least) all layers of the proper oriented pyramid, and its corresponding frequency component will be observed by the CP algorithm, but may also appear in both oriented pyramids as the motion is so large. Likewise, due to potential squaring-off of the sinusoids as a by-product of crossing pixels, substantial harmonic representation from a strong fundamental is also a sign that the motion is well-beyond a single pixel. These motions will survive downsampling and will benefit from it, as they will better satisfy the small-motion assumption and the SNR will increase due to the spatial smoothing used in the CSP. Small (sub-pixel) motions will not be present in every layer, as their eventual degradation and/or suppression into the noise floor will inhibit CP from accurately resolving the frequency. Therefore, motions which are observed in every layer will be processed using deeper levels of the pyramid to render them sub-pixel, whereas those observed in only a few of the CSP layers will only be processed using the first and/or second layer to preserve signal quality. This sorting results in (at most) the same number of bandpasses as if the video had been processed through the CSP otherwise with traditional filtering around each mode. However, now PBMM/PBME can be done in a more effective manner where each scale of pixel motion is given the best chance to be accurately reproduced from the video while also reducing computation time and memory usage overall. Once things have been sorted and processed, the extracted displacements—and/or motion-magnified frames—from different regimes can be recombined for the full, as-is displacement signal if desired, being much more akin to an output expected from a more traditional sensor like an accelerometer or laser Doppler vibrometer.

A diagram showing the flow and steps involved in this algorithm is provided in Fig. 7.3, and the authors give credit to CSAIL for use of the phase-based motion magnification code provided and J. Stone's code for CP.

## 7.4 Verification

Verification is done on both a synthetic example and an experimental beam to show varying expectations for simple and complex scenes.

### *7.4.1 Synthetic*

For the synthetic case, data sets were created to allow for control of individual parameters and to provide precise comparisons for frequency extractions. To do this, a cantilever beam was constructed in Blender, a 3-D computer graphics software capable of creating animations with predefined motion paths. The beam underwent prescribed motion according to Euler-Bernoulli beam theory:

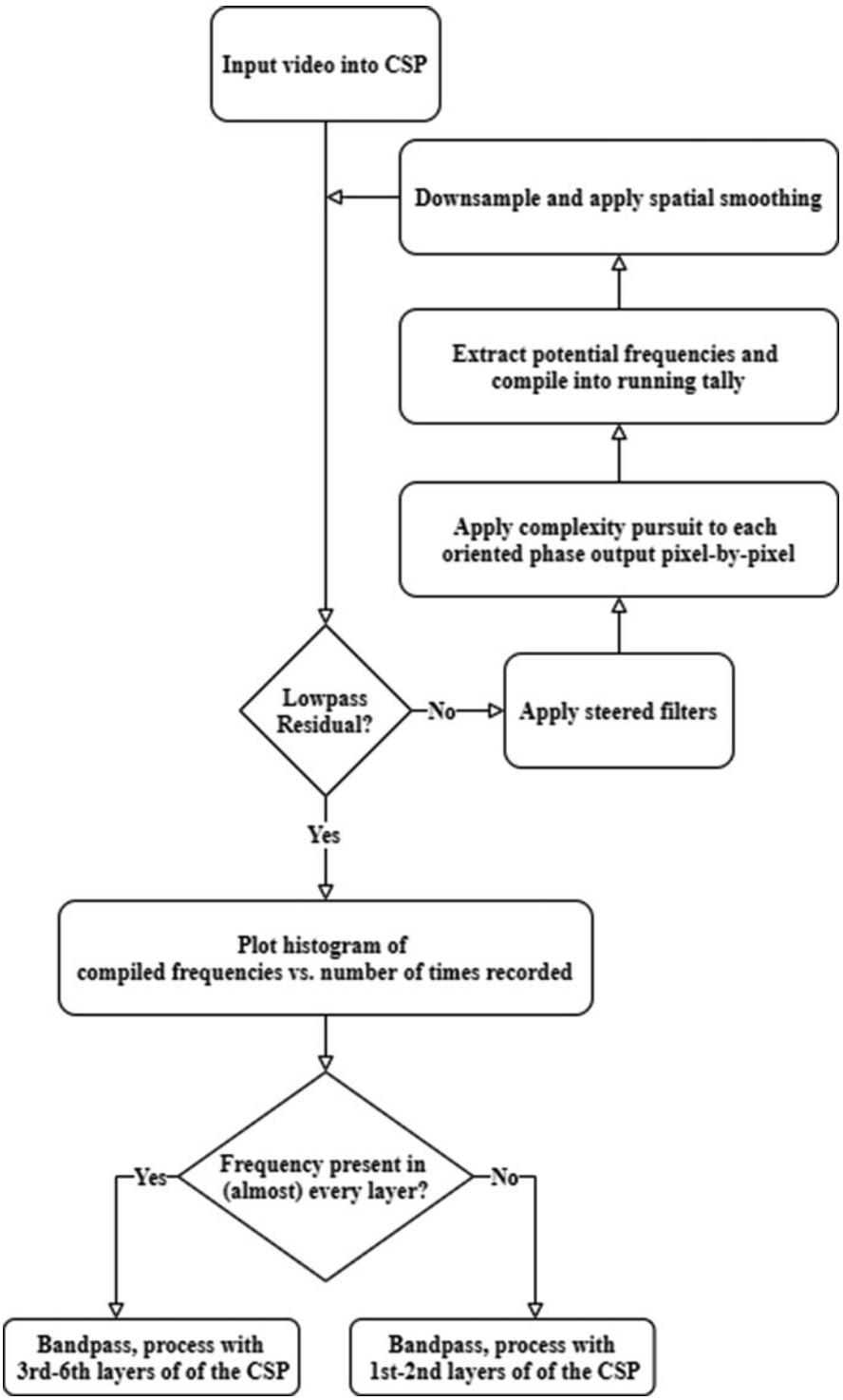


**Fig. 7.3** Flow diagram for the CSP+CP regime-sorting approach. This is done for each oriented pyramid within the CSP

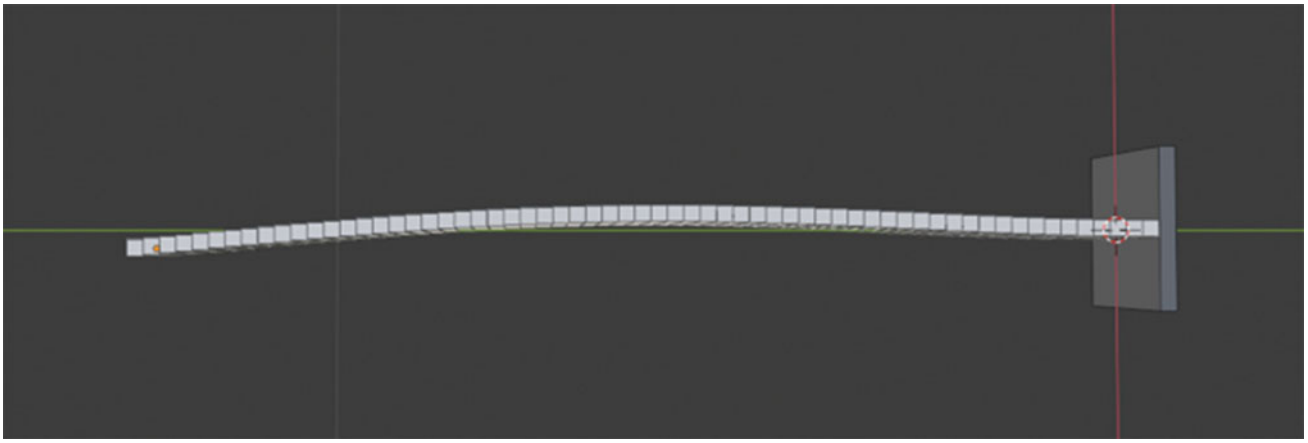

**Fig. 7.4** Blender beam, as shown in Blender

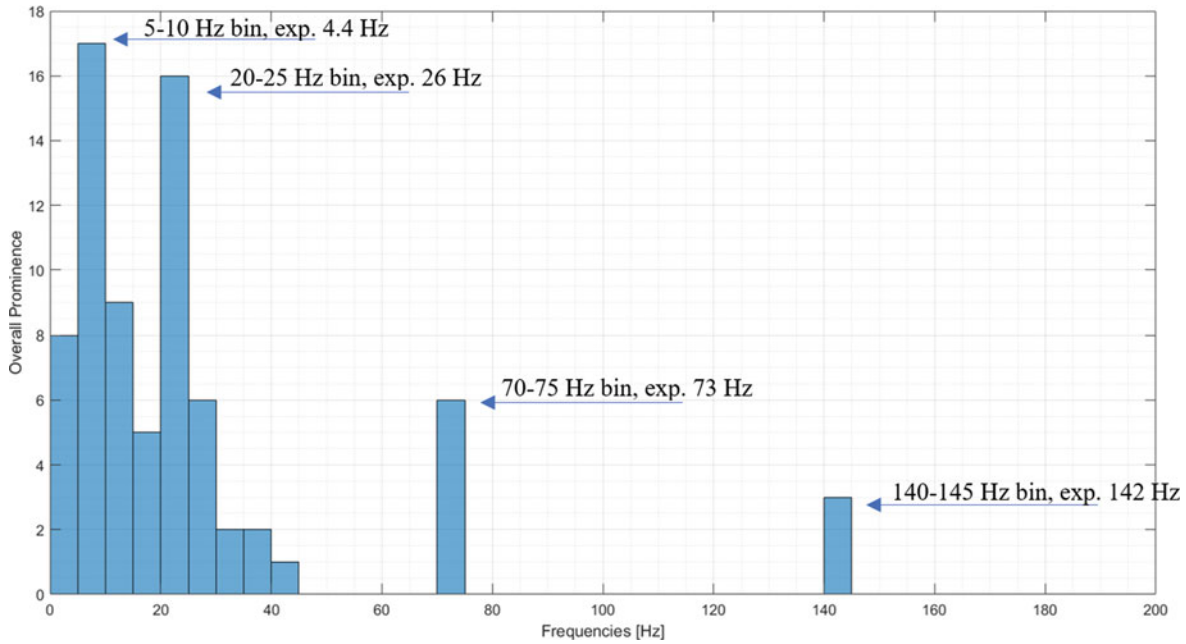


**Fig. 7.5** Output from CP algorithm, note the strong peaks between 0 and 45 Hz and the weaker peaks above. Notice that artificial harmonics and CP error/bin width have led to an extra count of the 4.4 Hz component—17 counts instead of 16

$$\xi\left(x,t\right)=|\delta|\cdot\Big[\cosh(x)\left(\frac{\omega x}{v}\right)-\cos\left(\frac{\omega x}{v}\right)$$
$$-\left(\cosh(x)\left(\frac{\omega L}{v}\right)+\cos\left(\frac{\omega L}{v}\right)/\sinh(x)\left(\frac{\omega L}{v}\right)+\sin\left(\frac{\omega L}{v}\right)\right)\cdot\left(\sinh(x)\left(\frac{\omega x}{v}\right)-\sin\left(\frac{\omega x}{v}\right)\right)\Big]$$
$$\cdot\sin\left(\omega t\right) \tag{7.5}$$

where $(x, t)$ is the spatiotemporal point, $\omega$ the temporal frequency, $L$ the length of the beam, and $v$ the dispersive wave speed. Blender's representation of the beam can be seen in Fig. 7.4. Frequencies were 4.4, 26, 73, and 142 Hz with respective amplitudes, $|\delta|$, of 3, 1, 0.1, and 0.05 px. The video was rendered at 16 bit with a resolution of 540 × 960, no noise, and no compression. A frame rate of 800 frames per second was used to ensure proper amplitude measurement. The beam is shown in Fig. 7.4. The pyramid had two orientations—horizontal and vertical—eight layers each.

After rendering the video, it was processed through the proposed algorithm of CSP+CP. The frequency vs. prominence histogram was plotted and is shown in Fig. 7.5. In this case, the pyramid had a total of 16 displacement extractions in terms of orientation and layer. It is immediately clear from the histogram that the CP algorithm has accurately predicted the frequencies prescribed in Blender. Further, the behavior described in the theory section is observed, where modes with smaller amplitudes are not present in every layer, but the larger displacements persist throughout all of the pyramid(s). From

the histogram, it can be gathered that the strong peaks present at 0–45 Hz will survive in the lower levels of the pyramid with greater amounts of downsampling while the smaller peaks at higher frequencies will instead benefit from the first few layers, as the signal degrades beyond recognition rather quickly. In the absence of aliasing, only two sets of bandpassing need be done at minimum if the interest is to capture to full displacement signal (not extract particular modes): one for the range encompassing all of the prominent—typically lower—frequencies and one for the weaker set. When aliasing is present, more bandpassing will be required to properly sort the motion spectra. With this, the frequency resolution (here 5 Hz) can be increased or decreased depending on the application, but this seems like a nice balance as exact frequency extraction is not the goal here, but rather general regime sorting.

### *7.4.2 Experimental*

For experimental data sets, an aluminum cantilever beam was initially attached to a shaker vibrating with known voltage given from a laser Doppler vibrometer (LDV) at 102400 samples per second. Estimated pixel displacement was then calculated from calibration factors and recorded for comparison later. A Phantom v1212 high-speed camera was used to record video data at a resolution of 800 × 1280 with 1 k or 5 k frames per second. The setup can be seen in Fig. 7.6 (top) and a screenshot from the video in Fig. 7.6 (bottom). Constant lighting was implemented, and a black background was used to promote good contrast at the edge for motion extraction. Extracted displacements for the first five modes were 0.34, 0.038, 0.028, 0.005, and 0.002 px. The results from CSP+CP are shown in Fig. 7.7.

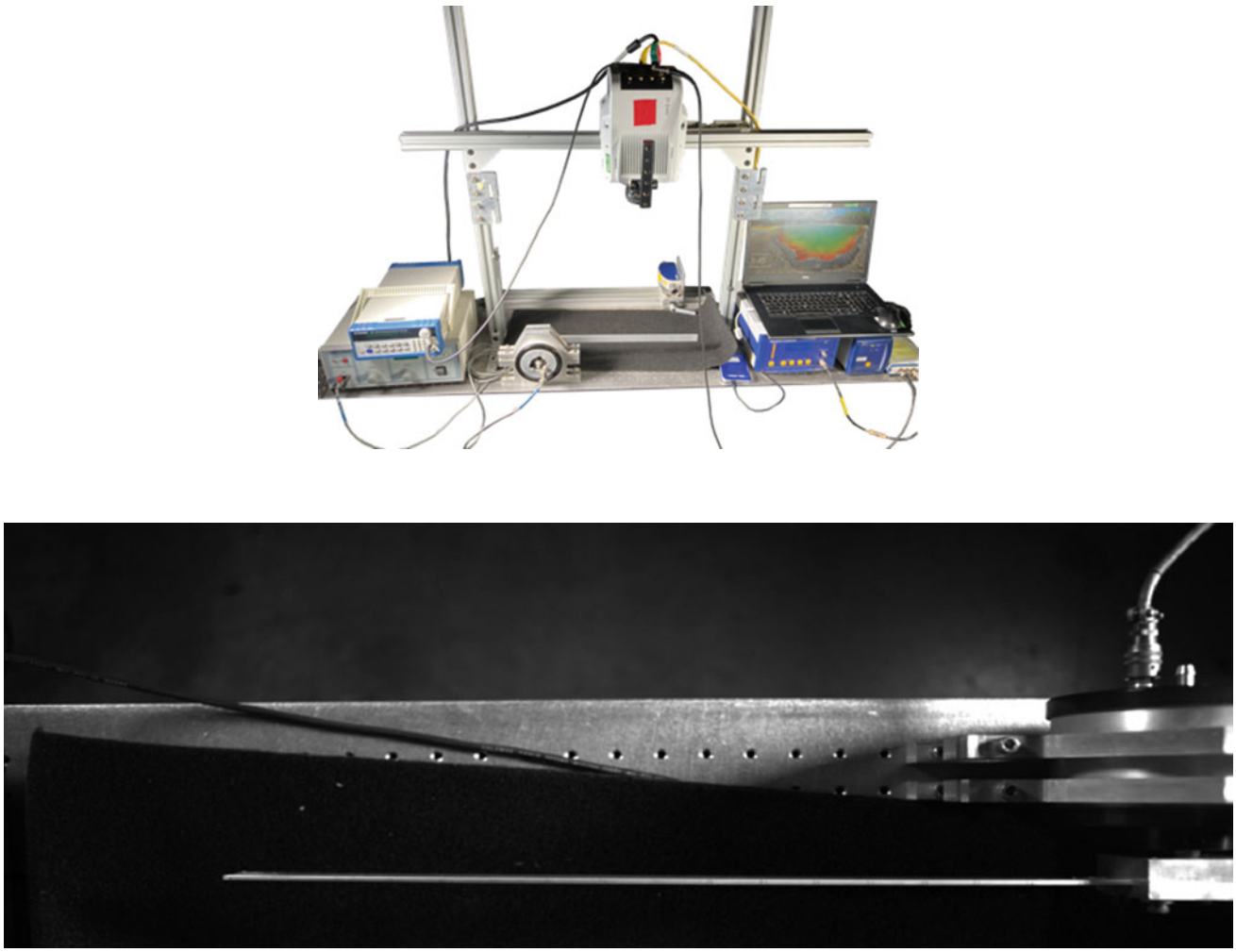

**Fig. 7.6** Experimental setup (top) and a still frame from the video recording (bottom)

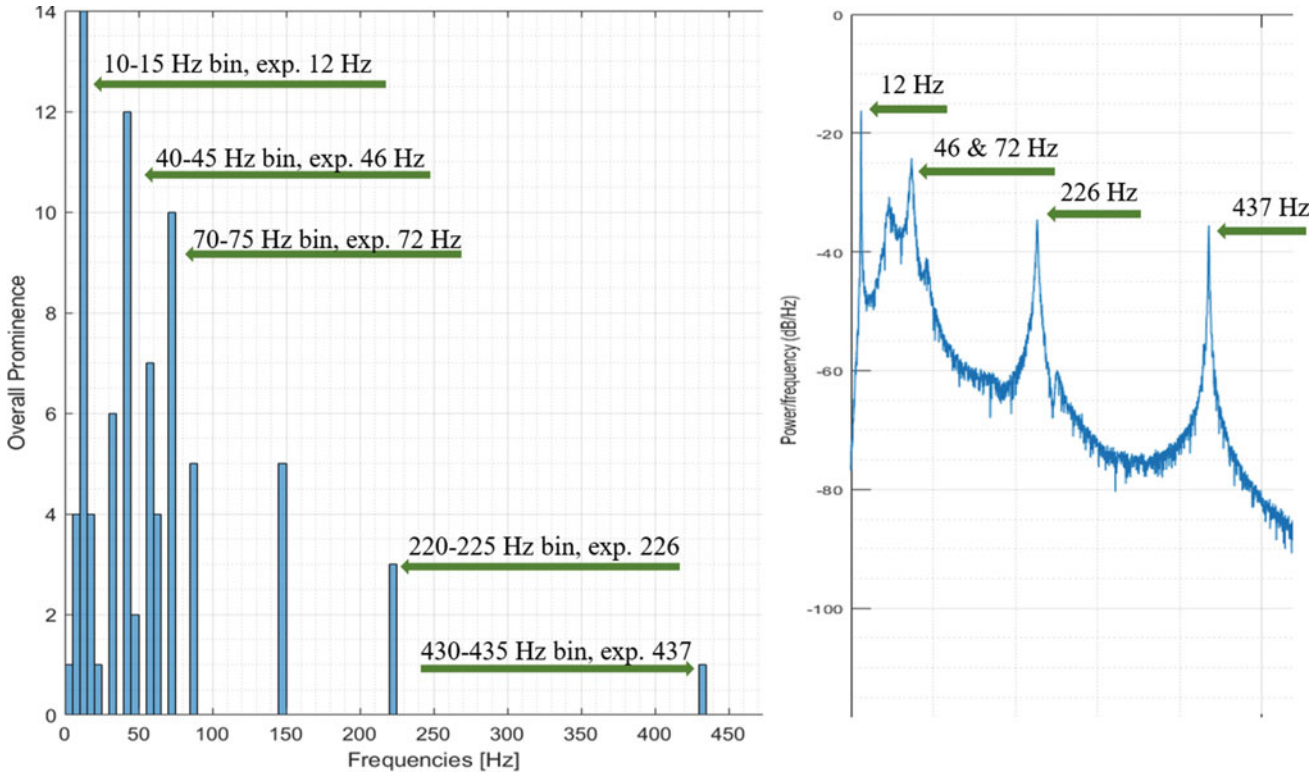


**Fig. 7.7** Output from CP algorithm for cropped region compared to laser PSD. Note the strong peaks from 0 to 100 Hz and weaker peaks from 150 to 500 Hz, all matching well to the reference laser and arrowed in green

Again, strong peaks present in all layers of the oriented pyramid (here six layers) are gathered toward the lower frequencies, where smaller peaks are mostly contained in the higher-frequency region and not omnipresent. Although none of the individual modal amplitudes are super-pixel, the superposition had a maximum of 1.2 px peak to peak as modes came in phase. Artificial harmonic content as a result of this pixel crossing is again visible and indicating the total motion is super-pixel. In this case, a low-pass filter for 0–100 Hz would capture the super-pixel motion that can then be processed using sufficient downsampling; similarly, a high pass can be used for the 150–500 Hz bandwidth, processing with the first or second layer to preserve quality. In addition to sufficient sorting, there is good agreement between frequencies extracted using the algorithm and those from the LDV power spectral density (PSD). The 145–150 Hz bin presence can be attributed to aliasing.

The beam was then excited with a modal impact hammer, which is another common experimental technique easily capable of producing super-pixel motions. Extracted displacements for the first five participating modes were 5, 0.08, 0.05, 0.04, and 0.01 px, with the impact being roughly at the node of the second bending mode. In this case, each oriented pyramid contained five layers. The CP output in Fig. 7.8 reveals abundant artificial harmonic content, resulting in extensive double counting within the running tally—hence the presence in more than 20 layers when only 10 are produced in the CSP. This content covers a wide bandwidth and is a good indicator that this frequency range will need (and survive) downsampling to process correctly. The higher-frequency modes are again present in only a few levels of the respective oriented pyramid and will benefit from minimal downsampling to preserve signal quality. As with the white noise input, the extracted frequency components given by the algorithm match well to the corresponding LDV PSD.

Both the synthetic and experimental cases have demonstrated the utility and effectiveness of this method in practice. While the CSP detriment and the squaring-off of pixel motion are nonideal factors, when considered together with the CP output they provide a complementary visual tool. Not only does this tool help to blindly identify frequencies of interest, but it allows for quick qualification of relative pixel motion size. With this information, motions can be processed in a more nuanced and appropriate way.

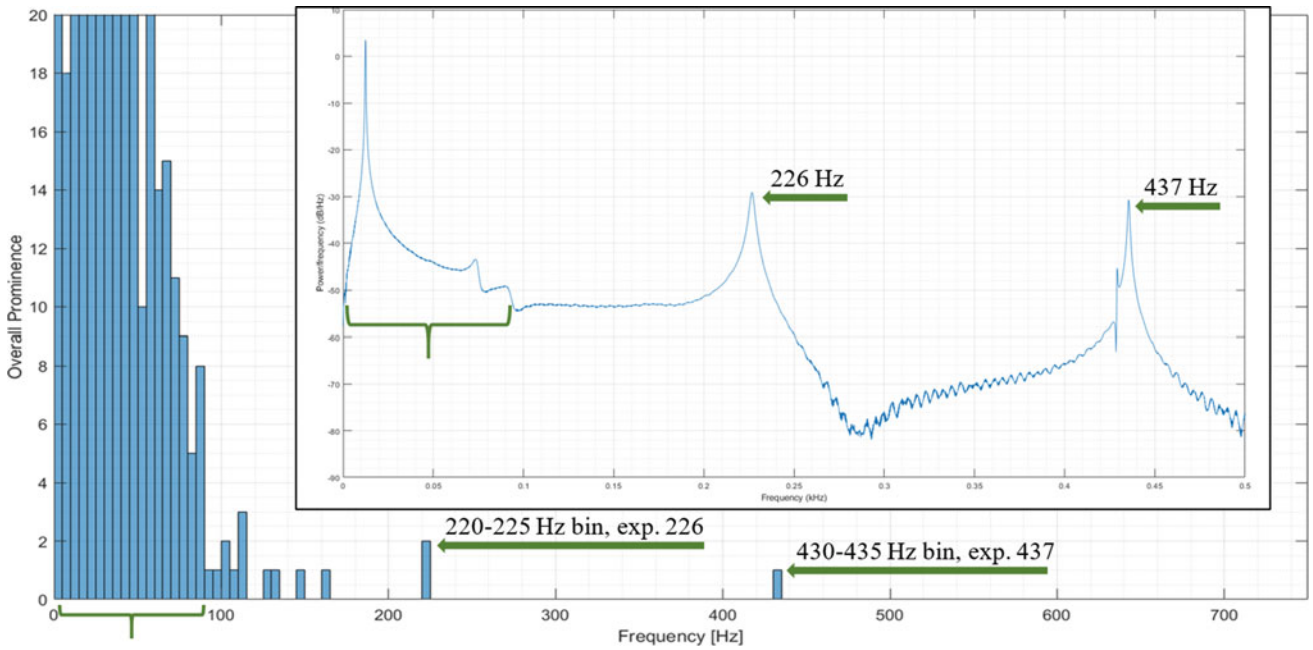


**Fig. 7.8** Output from CP algorithm for cropped region and hammer excitation. Note the abundant artificial harmonic content resulting from low-frequency, super-pixel motion

## 7.5 Complicated Scenes

In videos with only the object of interest (OOI) in the scene, the above processing will yield something akin to an FRF, where smaller motions will be at higher frequencies, and so very wide bandpassing to simply sort the small from the large is sufficient, and results in two further (but faster because fewer layers) runs of the CSP. If there are multiple OOI, like in the experimental beam, then this simple segmentation will likely need to be adjusted. For example, when the CP output (Fig. 7.9) is compared to the laser PSD (Fig. 7.10), there is okay agreement in the lower frequencies, but there exists a questionable resurgence of strong peaks in the higher frequency range which does not agree with the laser data. These are arrowed in purple. This is a result of the shaker, visible in the lower right-hand side of the frame. With this region cropped out, there is much better agreement with the PSD (Fig. 7.7). If cropping is not an option, then multiple, more precise bandpassing must to be used to adequately sort the motions.

## 7.6 Conclusions

A detriment of image pyramids is exploited using complexity pursuit to sort motions into super- and sub-pixel regimes. The proposed algorithm provides an intuitive, visual tool that allows the experimentalist to quickly identify and sort large and small motions of interest in video. This sorting allows for tailored processing, wherein initially sub-pixel motions are processed minimally to preserve quality and super-pixel motions can be downsampled sufficiently to render them sub-pixel. This facilitates the applicability of phase-based optical flow to the super-pixel regime and allows for more accurate motion extraction or magnification overall. Now that important, super-pixel motions can be handled effectively, future work might include combining this approach with one of the other described methods. This would allow for complete handling of large motions in a scene (both valued and otherwise), resulting in a more robust phase-based processing method previously plagued by such degrees of displacement.

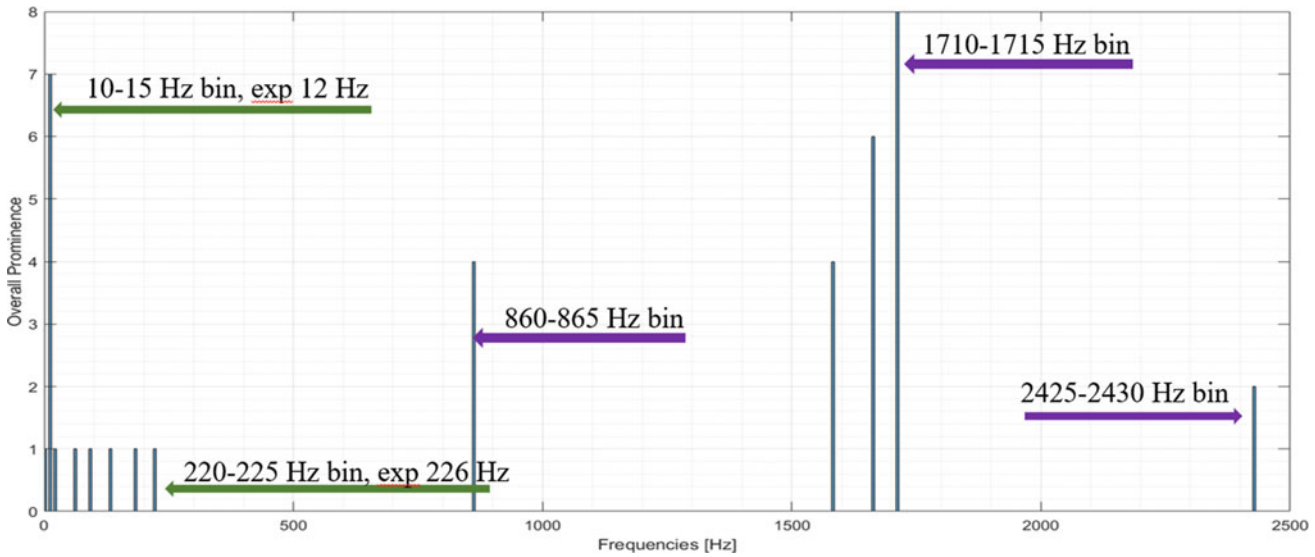


**Fig. 7.9** Output from CP algorithm for full image. Note the mix of strong and weak bins over the frequency range as a result of the shaker's presence in the frame. Bins agreeing with the PSD are arrowed in green. Those not present are in purple

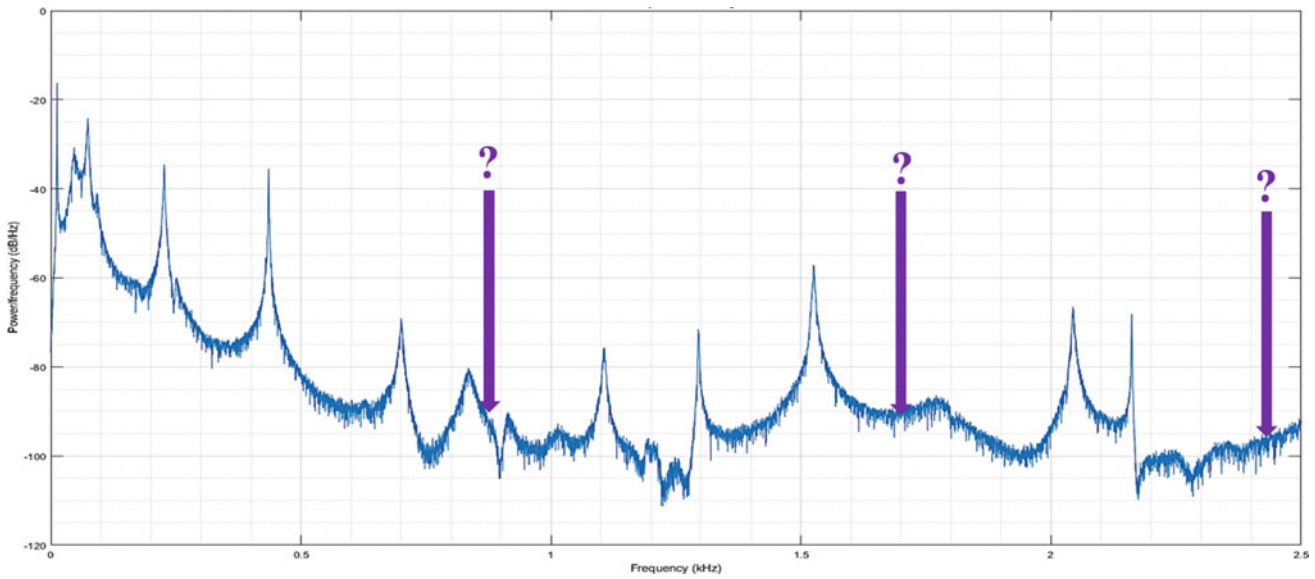


**Fig. 7.10** Full LDV PSD from the beam excited with the shaker, with frequencies not agreeing with CP arrowed in purple

**Acknowledgments** We would like to thank the ARL Walker Assistantship program for supporting this work.

## References

1. Wadhwa, N., Rubinstein, M., Durand, F., Freeman, W.T.: Phase-based video motion processing. ACM Trans. Graph. **32**(4), 1–10 (2013). https://doi.org/10.1145/2461912.2461966
2. Elgharib, M.A., Hefeeda, M., Durand, F., Freeman, W.T.: Video magnification in presence of large motions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4119–4127 (2015). https://doi.org/10.1109/CVPR.2015.7299039
3. Liu, C., Torralba, A., Freeman, W.T., Durand, F., Adelson, E.H.: Motion magnification. In: ACM SIGGRAPH 2005 Papers, pp. 519–526. Association for Computing Machinery, New York (2005). https://doi.org/10.1145/1186822.1073223
4. Zhang, Y., Pintea, S.L., van Gemert, J.C.: Video acceleration magnification. In: Computer Vision and Pattern Recognition, pp. 502–510 (2017)
5. Takeda, S., Okami, K., Mikami, D., Isogai, M., Kimata, H.: Jerk-aware video acceleration magnification. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 1769–1777 (2018). https://doi.org/10.1109/CVPR.2018.00190

6. Wu, X., Yang, X., Jin, J., Yang, Z.: Amplitude-based filtering for video magnification in presence of large motion. Sensors. **18** (2018). https://doi.org/10.3390/s18072312
7. Zhang, L., Yang, X.: Video magnification under the presence of complex background motions. IEICE Trans. Inf. Syst. **104**, 909–914 (2021). https://doi.org/10.1587/transinf.2020EDL8134
8. Wadhwa, N.: Revealing and analyzing imperceptible deviations in images and videos. DSpace. (2016)
9. Stone, J.V.: Blind source separation using temporal predictability. Neural Comput. **13**, 1559–1574 (2001)
10. Yang, Y., Nagarajaiah, S.: Blind modal identification of output-only structures in time-domain based on complexity pursuit. Earthquake Eng. Struct. Dyn. **42**, 1885–1905 (2013)
11. Yang, Y., Dorn, C., Mancini, T., Talken, Z., Kenyon, G., Farrar, C., Mascareñas, D.: Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification. Mech. Syst. Signal Process. **85**, 567–590 (2017). https://doi.org/10.1016/j.ymssp.2016.08.041
12. Yang, Y., Dorn, C., Mancini, T., Talken, Z., Nagarajaiah, S., Kenyon, G., Farrar, C., Mascareñas, D.: Blind identification of full-field vibration modes of output-only structures from uniformly sampled, possibly temporally aliased (sub-Nyquist), video measurements. J. Sound Vib. **390**, 232–256 (2017). https://doi.org/10.1016/j.jsv.2016.11.034
13. Janatka, M., Sridhar, A., Kelly, J., Stoyanov, D.: Higher order of motion magnification for vessel localisation in surgical video. In: Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) Medical Image Computing and Computer Assisted Intervention—MICCAI 2018. Lecture Notes in Computer Science, vol. 11073, (2018)