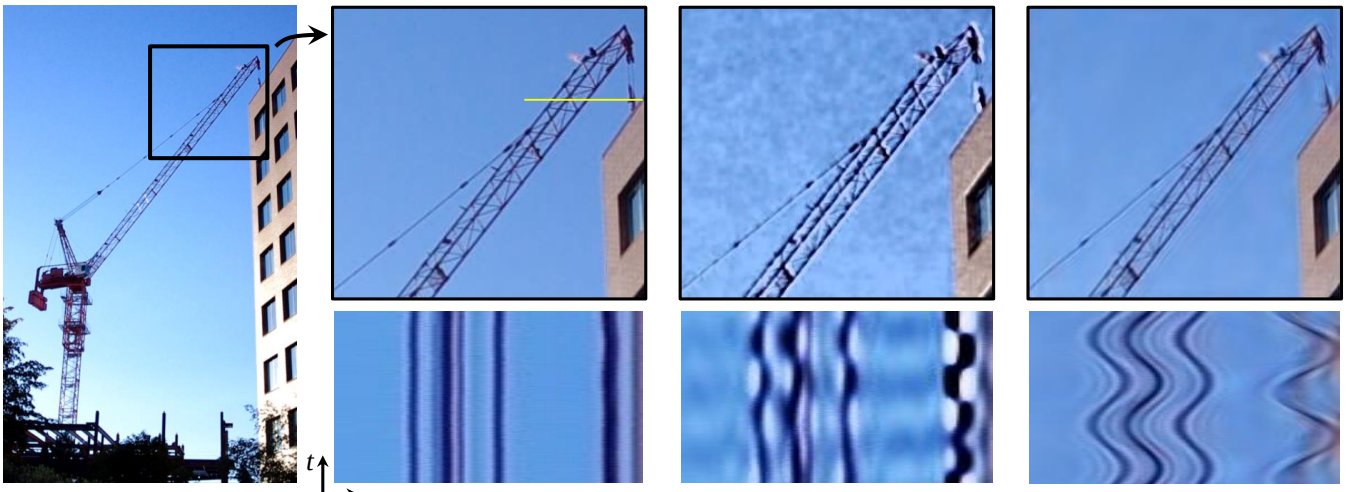
Phase-Based Video Motion Processing

Neal Wadhwa Michael Rubinstein Frédo Durand William T. Freeman

MIT Computer Science and Artificial Intelligence Lab



(a) Source (b) Linear [Wu et al. 2012] (c) Phase-based (this paper)

Figure 1: *Motion magnification of a crane imperceptibly swaying in the wind. (a) Top: a zoom-in onto a patch in the original sequence* (crane) *shown on the left. Bottom: a spatiotemporal XT slice of the video along the profile marked on the zoomed-in patch. (b-c) Linear [Wu et al. 2012] and phase-based motion magnification results, respectively, shown for the corresponding patch and spatiotemporal slice as in (a). The previous, linear method visualizes the crane's motion, but amplifies both signal and noise and introduces artifacts for higher spatial frequencies and larger motions, shown by the clipped intensities (bright pixels) in (b). In comparison, our new phase-based method supports larger magnification factors with significantly fewer artifacts and less noise (c). The full sequences are available in the supplemental video.*

Abstract

We introduce a technique to manipulate small movements in videos based on an analysis of motion in complex-valued image pyramids. Phase variations of the coefficients of a complex-valued steerable pyramid over time correspond to motion, and can be temporally processed and amplified to reveal imperceptible motions, or attenuated to remove distracting changes. This processing does not involve the computation of optical flow, and in comparison to the previous Eulerian Video Magnification method it supports larger amplification factors and is significantly less sensitive to noise. These improved capabilities broaden the set of applications for motion processing in videos. We demonstrate the advantages of this approach on synthetic and natural video sequences, and explore applications in scientific analysis, visualization and video enhancement.

CR Categories: I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Time-varying Imagery;

Keywords: video-based rendering, spatio-temporal analysis, Eulerian motion, video magnification

Links: DL PDF WEB

1 Introduction

A plethora of phenomena exhibit motions that are too small to be well perceived by the naked eye and require computational amplification to be revealed [Liu et al. 2005; Wu et al. 2012]. In *Lagrangian* approaches to motion magnification [Liu et al. 2005; Wang et al. 2006], motion is computed explicitly and the frames of the video are warped according to the magnified velocity vectors. Motion estimation, however, remains a challenging and computationally-intensive task, and errors in the estimated motions are often visible in the results.

Recently-proposed *Eulerian* approaches eliminate the need for costly flow computation, and process the video separately in space and time. Eulerian video processing was used by [Fuchs et al. 2010] to dampen temporal aliasing of motion in videos, while [Wu et al. 2012] use it to reveal small color changes and subtle motions. Unfortunately, linear Eulerian video magnification [Wu et al. 2012] supports only small magnification factors at high spatial frequencies, and can significantly amplify noise when the magnification factor is increased (Fig. 1(b)).

To counter these issues, we propose a new Eulerian approach to motion processing, based on complex-valued steerable pyramids [Simoncelli et al. 1992; Portilla and Simoncelli 2000], and inspired by phase-based optical flow [Fleet and Jepson 1990; Gautama and Van Hulle 2002] and motion without movement [Freeman et al. 1991]. Just as the phase variations of Fourier basis functions (sine waves) are related to translation via the the Fourier shift theorem, the phase variations of the complex steerable pyramid correspond to local motions in spatial subbands of an image. We compute the local phase variations to measure motion without explicit optical flow computation and perform temporal processing to amplify motion in selected temporal frequency bands, and then reconstruct the modified video.

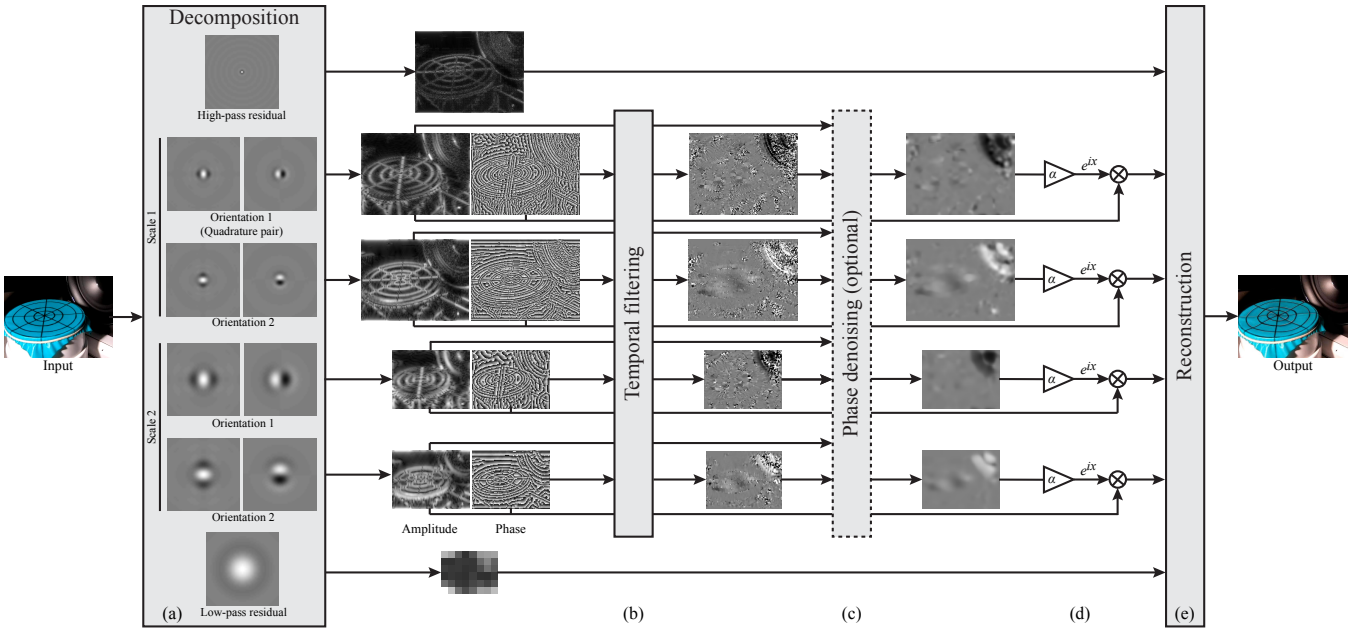


Figure 2: *Our phase-based approach manipulates motion in videos by analyzing the signals of local phase over time in different spatial scales and orientations. We use complex steerable pyramids to decompose the video and separate the amplitude of the local wavelets from their phase (a). We then temporally filter the phases independently at at each location, orientation and scale (b). Optionally, we apply amplitude-weighted spatial smoothing (c, Sect. 3.4) to increase the phase SNR, which we empirically found to improve the results. We then amplify or attenuate the temporally-bandpassed phases (d), and reconstruct the video (e). This example shows the processing pipeline for the* membrane *sequence (Sect. 4), using a pyramid of two scales and two orientations (the relative difference in size between the pyramid levels is smaller in this figure for clarity of the visualization).*

We start from the relation between motion and phase in steerable pyramids and show that by increasing the phase variations by a multiplicative factor we can amplify subtle motions. We then use this relation to analyze the limits of our method, which are set by the spatial support of the steerable basis functions. To amplify motions further, we extend the complex steerable pyramid to *sub-octave bandwidth pyramids*, comprised of filters with larger spatial support in the primal domain. While our new image representation is over-complete by a larger factor, it supports larger amplification of motions at all spatial frequencies, leading to fewer artifacts.

The phase-based method improves on the previous, linear Eulerian magnification method [Wu et al. 2012] in two important aspects (Fig. 1): the phase-based method (a) achieves larger magnifications, and (b) has substantially better noise performance. Because Wu et al. [2012] amplify temporal brightness changes, the amplitude of noise is amplified linearly. In contrast, the present method modifies phases, not amplitudes, which does not increase the magnitude of spatial noise. We demonstrate that the phase-based method can achieve larger motion magnifications with fewer artifacts, which expands the set of small-scale physical phenomena that can be visualized with motion magnification techniques.

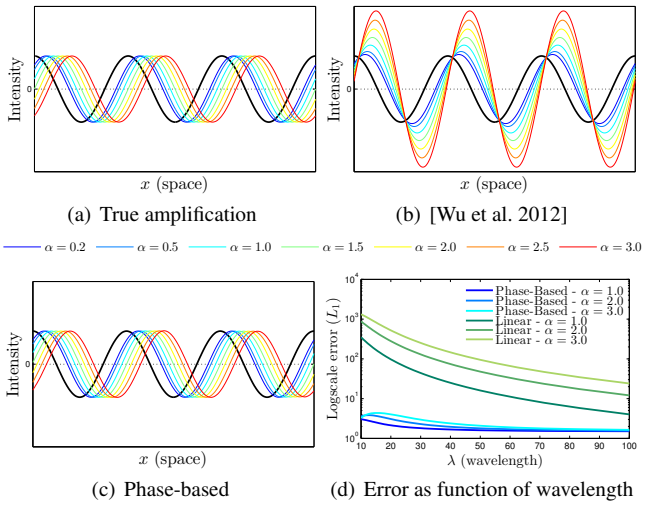
The main contributions of this paper are: (a) a novel approach for Eulerian processing of motion in videos, based on the analysis of phase variations over time in complex steerable pyramids; (b) we explore the trade-off between the compactness of the transform representation and amplitude of magnification in octave and sub-octave bandwidth pyramids; and (c) we demonstrate that the extracted low-amplitude motion signal can be refined by denoising the phase signal spatially within each image subband, improving the motion-processed results. Our new phase-based approach is able to magnify small motions further, with less noise and fewer artifacts than the previous Eulerian motion magnification method.

2 Background

Phase-based Optical Flow. Fleet and Jepson [1990] tracked constant phase contours by computing the phase gradient of a spatio-temporally bandpassed video, and showed that it provides a good approximation to the motion field, and that phase is more robust than amplitude to image changes due to contrast and scale. Gautama and Van Hulle [2002] used a similar technique in which they computed the temporal gradient of the phases of a spatially bandpassed video to estimate the motion field. We build on this link between phase and motion, but seek to avoid the explicit computation of flow vectors, and instead directly manipulate the phase variations in videos.

Complex Steerable Pyramids. The steerable pyramid [Simoncelli et al. 1992; Simoncelli and Freeman 1995] is an overcomplete transform that decomposes an image according to spatial scale, orientation, and position. The basis functions of the transform resemble Gabor wavelets, sinusoids windowed by a Gaussian envelope, and are steerable. We don't exploit the steerability of those basis functions in this work, but the transform has other properties which are important for our motion analysis: non-aliased subbands and quadrature phase filters.

We measure phase within each subband using the pairs of even and odd-phase oriented spatial filters whose outputs are the complex-valued coefficients in the steerable pyramid [Simoncelli et al. 1992], The sub-sampling scheme of the steerable pyramid avoids spatial aliasing and thus allows meaningful signal phase measurements from the coefficients of the pyramid. The real part of each coefficient represents the even-symmetric filter (cosine), while its imaginary counterpart represents an odd-symmetric filter (sine). While twice as over-complete as a real-valued pyramid, the complex-valued pyramid allows simple mea-



(a) True amplification (b) [Wu et al. 2012]

(c) Phase-based (d) Error as function of wavelength

Figure 3: *Phase-based motion magnification is perfect for sinusoidal functions. In these plots, the initial displacement is $\delta(t) = 1$. While the errors for the technique of Wu et al. [2012] are dependent on wavelength for sinusoids, there is no such dependence for the present technique and the error is uniformly small. The vertical axis in (d) is logarithmic.*

surement of local amplitude and phase, which we exploit to process motion.

The steerable pyramid has non-oriented, real-valued high and low-pass coefficients describing residual signal components not captured by the bandpass filters [Simoncelli et al. 1992]. The frequency domain transfer functions in the oriented bands of the steerable pyramid, $\Psi_{\omega,\theta}$, are scaled and rotated copies of a basic filter, indexed by scale ω and orientation θ.

The steerable pyramid is built by applying these transfer functions to the discrete Fourier transform $\tilde{I}$ of an image I to decompose it into different spatial frequency bands $S_{\omega,\theta}$ which have DFT $\tilde{S}_{\omega,\theta}(x,y) = \tilde{I}\Psi_{\omega,\theta}$. Each filter isolates a continuous region of the frequency domain and therefore has an impulse response that is localized in space (Fig. 4(Impulse Response)). The resulting spatial frequency band is localized in space, scale and orientation (see [Portilla and Simoncelli 2000] for filter design steps). The transfer functions of a complex steerable pyramid only contain the positive frequencies of the corresponding real steerable pyramid's filter. That is, the response of $2\cos(\omega x) = e^{i\omega x} + e^{-i\omega x}$ is $e^{i\omega x}$ so that there is a notion of both amplitude and phase.

In the frequency domain, the process to build and then collapse the pyramid is given by

$$\tilde{I}_R = \sum \tilde{S}_{\omega,\theta}\Psi_{\omega,\theta} = \sum \tilde{I}\Psi^2_{\omega,\theta} \tag{1}$$

where the sums are over all of the scales and orientations in the pyramid, yielding the reconstructed image, I_R. We perform filtering in the frequency domain.

3 Phase-based Motion Processing

Our processing amplifies small motions by modifying local phase variations in a complex steerable pyramid representation of the video. In this section, we describe our approach and discuss why the phase-based technique has better noise handling and maximum magnification than the linear Eulerian motion magnification technique [Wu et al. 2012]. To give intuition and to demonstrate that the phase variations correspond to motion, we show how our technique works on sinusoidal waves (Fourier basis elements). For non-periodic image structures, phase-based motion magnification is bounded by the spatial support of the complex steerable pyramid filters. We overcome this bound by using sub-octave bandwidth complex steerable pyramids that have wider spatial support.

3.1 Motion Magnification

The phase-based approach relies on complex-valued steerable pyramids because they allow us to measure and modify local motions. To give intuition for our phase-based motion processing, we first give an example using a global Fourier basis and consider the case of a 1D image intensity profile f under global translation over time, $f(x + \delta(t))$, for some displacement function $\delta(t)$ (not to be confused with a Dirac function). We wish to synthesize a sequence with modified motion, $f(x + (1 + \alpha)\delta(t))$, for some magnification factor α. We will discuss the general case at the end of this section.

Using the Fourier series decomposition, we can write the displaced image profile, $f(x + \delta(t))$, as a sum of complex sinusoids,

$$f(x + \delta(t)) = \sum_{\omega=-\infty}^{\infty} A_\omega e^{i\omega(x+\delta(t))} \tag{2}$$

in which each band corresponds to a single frequency ω.

From Eq. 2, the band for frequency ω is the complex sinusoid

$$S_\omega(x,t) = A_\omega e^{i\omega(x+\delta(t))}. \tag{3}$$

Because S_ω is a sinusoid, its phase $\omega(x + \delta(t))$ contains motion information. Like the Fourier shift theorem, we can manipulate the motion by modifying the phase.

To isolate motion in specific temporal frequencies, we temporally filter the phase $\omega(x + \delta(t))$ (Eq. 3) with a DC balanced filter. To simplify the derivation, we assume that the temporal filter has no other effect except to remove the DC component ωx. The result is

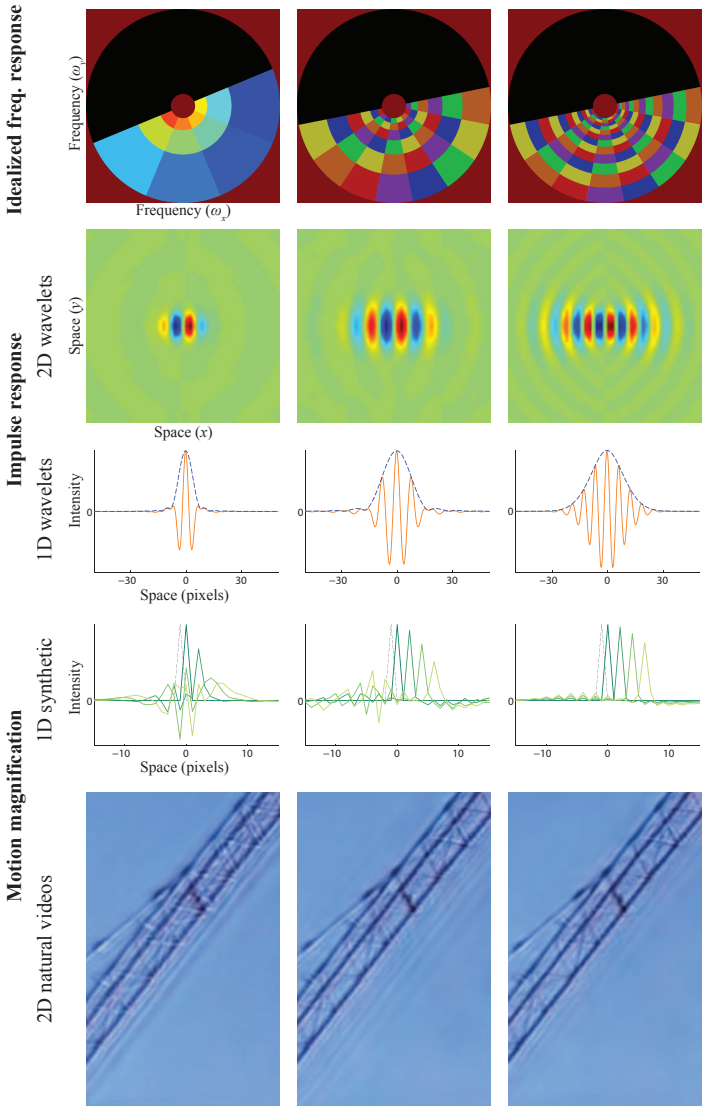
$$B_\omega(x,t) = \omega\delta(t). \tag{4}$$

We then multiply the bandpassed phase $B_\omega(x,t)$ by α and increase the phase of sub-band $S_\omega(x,t)$ by this amount to get the motion magnified sub-band

$$\hat{S}_\omega(x,t) := S_\omega(x,t)e^{i\alpha B_\omega} = A_\omega e^{i\omega(x+(1+\alpha)\delta(t))}. \tag{5}$$

The result $\hat{S}_\omega(x,y)$ is a complex sinusoid that has motions exactly $1 + \alpha$ times the input (Fig. 3). We can reconstruct the motion-magnified video by collapsing the pyramid. In this analysis, we would do this by summing all the sub-bands to get the motion magnified sequence $f(x + (1 + \alpha)\delta(t))$.

In general, motions in a video are local and $\delta(t)$ is actually $\delta(x,t)$. We use the complex steerable pyramid to deal with local motions as its filters have impulse responses with finite spatial support (Fig. 4(Impulse Response)). Specifically, our method works as follows (Fig. 2). We compute the local phase over time at every spatial scale and orientation of a steerable pyramid. Then, we temporally bandpass these phases to isolate specific temporal frequencies relevant to a given application and remove any temporal DC component. These temporally bandpassed phases correspond to motion in different spatial scales and orientations. To synthesize magnified motion, we multiply the bandpassed phases by an amplification factor α. We then use these amplified phase differences to magnify (or attenuate) the motion in the sequence by modifying the phases of each coefficient by this amount for each frame.



(a) 4 orientations, octave bandwidth, 18 Filters (12x over-complete)

(b) 8 orientations, half-octave bandwidth, 82 filters (33x over-complete)

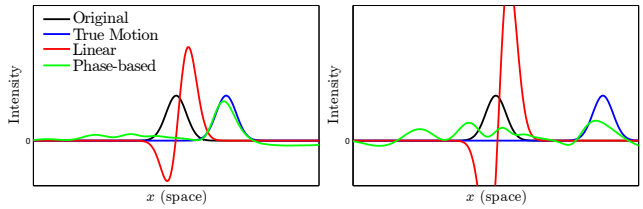
(c) 8 orientations, quarter-octave bandwidth, 162 filters (56x over-complete)

Figure 4: *A comparison between octave and sub-octave bandwidth pyramids for motion magnification. Each color in the idealized frequency response represents a different filter. (a) The original steerable pyramid of Portilla and Simoncelli [2000]. This pyramid has octave bandwidth filters and four orientations. The impulse response of the filters is narrow (rows $2-3$), which reduces the maximum magnification possible (rows $4-5$). (b-c) Pyramid representations with two and four filters per octave, respectively. These representations are more over-complete, but support larger magnification factors.*

3.2 Bounds

As we move an image feature by phase-shifting each complex pyramid filter covering that feature, we eventually reach a limit beyond which we can't move the feature because of the limited spatial support of each pyramid filter (Fig. 2(a) and Fig. 4(1D Wavelets)).

As an approximate analytic model of an image feature moved by the



(a) $\alpha = 6$ (b) $\alpha = 14$

Figure 5: *For general non-periodic structures, we achieve performance at least four times that of Wu et al. [2012], and do not suffer from clipping artifacts (a). For large amplification, the different frequency bands break up due to the higher bands having a smaller window (b).*

localized filters of the steerable pyramid, we consider the case of a single Dirac under uniform translation over time, moved by phase shifting Gabor filters, complex sinusoids modulated by a Gaussian window function. As the Dirac is phase-shifted, it is attenuated by the Gaussian window of the Gabor filters. Therefore, we bound the maximum phase shift such that the Dirac is only attenuated by a small amount.

A one dimensional Gabor filter has frequency domain transfer function

$$e^{-2\pi(\omega_x-\omega_0)^2\sigma^2}, \tag{6}$$

where ω_0 is the frequency the filter selects for and $\frac{1}{\sqrt{2\sigma}}$ is the width of Gaussian window in the frequency domain. Typically, σ depends on the frequency ω_0 (self-similar wavelets). The inverse Fourier transform gives us the following impulse response in the spatial domain (up to a constant factor):

$$S_\omega(x,0) = e^{-x^2/(2\sigma^2)}e^{2\pi i\omega_0 x}, \tag{7}$$

a complex sinusoid windowed by a Gaussian envelope. Respectively, the impulse response of a Dirac function shifted by $\delta(t)$ pixels (not to be confused with the Dirac function) at time t is

$$S_\omega(x,t) = e^{-(x-\delta(t))^2/(2\sigma^2)}e^{2\pi i\omega_0(x-\delta(t))} \tag{8}$$

Note that the spatial Gaussian envelope (the left term on the RHS of Eq. 8) does not affect the phase.

Applying a finite difference bandpass filter ($[1\ -1]$) to the phase at time 0 and time t, gives

$$B_\omega(x,t) = 2\pi\omega_0\delta(t), \tag{9}$$

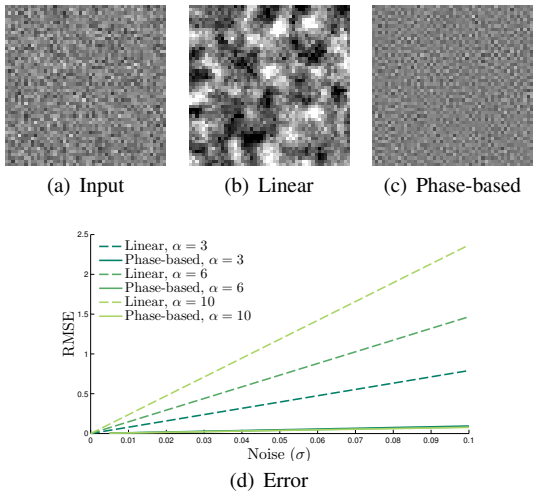
and the synthesized phase difference for modulating the motion by α is then

$$2\pi\omega_0\alpha\delta(t). \tag{10}$$

This phase difference corresponds to a shift of the Dirac by an additional $\alpha\delta(t)$ pixels. We need to bound the shift $\alpha\delta(t)$ such that the amplified shift approximates well the true shifted signal. We use one standard deviation of the Gaussian window as our bound. This maintains roughly 61% of the amplitude (Fig. 4 (1D Wavelets)), and so we have

$$\alpha\delta(t) < \sigma. \tag{11}$$

In the octave-bandwidth steerable pyramid of Portilla and Simoncelli [2000] (Fig. 4(a)), there is approximately one period of the sinusoid under the Gaussian envelope. That is, $4\sigma \approx \frac{1}{\omega_0}$, which gives the bound $\alpha\delta(t) < \sigma = \frac{1}{4\omega_0}$. By equating the spatial wavelength $\lambda = \frac{1}{\omega_0}$, we get[1]



(a) Input (b) Linear (c) Phase-based

(d) Error

Figure 6: *Comparison between linear and phase-based Eulerian motion magnification in handling noise. (a) A frame in a sequence of IID noise. In both (b) and (c), the motion is amplified by a factor of* 50*, where (b) uses the technique from Wu et al. [2012] and (c) uses the phase-based approach. (d) shows a plot of the error as function of noise for each method, using several magnification factors.*

$$\alpha\delta(t) < \frac{\lambda}{4}. \tag{13}$$

From Eq. 13, we see that the motions of the low spatial frequencies can be magnified more than those of the high spatial frequencies. Indeed, from Eq. 9, phase changes between frames will be much greater for the high frequency components than for the low frequency components. While derived for an impulse image feature moved by Gabor filters, we find the bound (and its extension below for sub-octave bandwidth pyramids) to be valid for both synthetic examples (Fig. 5) and natural videos (Fig. 1, Fig. 4, Sect. 4).

Exceeding the bound in Eq. 14 manifests as artifacts or blur, as not all image pyramid components are present in their proper ratios to reconstruct the desired translated feature. In Fig. 5(b), a Gaussian function magnified using our approach breaks up.

3.3 Sub-octave Bandwidth Pyramids

We see, therefore, that the bound in Eq. 13 is directly related to the spatial support of the filters. The smaller the filters in the frequency domain the larger their support is in the spatial domain, which allows us to shift the signals underneath their windows further. In the limit of a having a filter for every frequency band, the representation becomes equivalent to the Fourier transform and motion magnification is achieved via the shift theorem. However, we then lose the ability to measure or synthesize any spatial variation in the amount of motion. We found a good compromise between localization and magnification ability when using pyramid filters about two times as wide (in the sinusoidally varying spatial direction) as those described in Portilla and Simoncelli [2000]. They specify their steerable pyramid filters as being self-similar and having octave bandwidth (Fig. 4(a)), and we extend their representation to sub-octave bandwidth pyramids (Fig. 4(b,c)).

A simple way to accomplish this is to scale the filters in log space. This method works well for a half-octave bandwidth pyramid, while pyramids with more filters per octave need to be constructed differently, as discussed in Appendix A.

For the half octave pyramid, there are 2 periods under the Gaussian envelope of the wavelet. Thus, $4\sigma \approx \frac{2}{\omega_0}$, and the bound on the amplification (Eq. 13) becomes

$$\alpha\delta(t) < \frac{\lambda}{2}. \tag{14}$$

This bound improves over the one derived in Wu et al. [2012] using a Taylor series approximation by a factor of 4.[1]

There is a trade-off between the compactness of the representation and the amount of motion-magnification we can achieve. The 4-orientation, octave-bandwidth pyramid of Portilla and Simoncelli (Fig. 4(a)) is over-complete by a factor of 12 (each orientation contributes a real and imaginary part), and can easily support real time processing, but limits the amount of motion-magnification that can be applied. On the other hand, an 8-orientation half-octave pyramid (Fig. 4(b)) supports larger amplification, but is over-complete by a factor of 33.

3.4 Noise handling

Phase-based motion magnification has excellent noise characteristics. As the amplification factor is increased, noise is translated rather than amplified. At a particular scale and orientation band, the response for a noisy image $I + \sigma_n n$ might be

$$S_\omega = e^{i\omega(x+\delta(t))} + \sigma_n N_\omega(x,t), \tag{15}$$

where $N_\omega(x,t)$ is the response of n to the complex steerable pyramid filter indexed by ω. We assume that σ_n is much lower in magnitude than the noiseless signal, so that temporal filtering of the phase is approximately $\omega\delta(t)$ as in Eq. 4. To magnify the motion, the response in the Eq. 15 is shifted by $e^{i\alpha\omega\delta(t)}$, so that the motion magnified band is

$$\hat{S}_\omega = e^{i\omega(x+(1+\alpha)\delta(t))} + \sigma_n e^{i\alpha\omega\delta(t)} N_\omega(x,t) \tag{16}$$

The only change to the noise after processing is a phase shift. When the pyramid is collapsed, this phase shift corresponds to a translation of the noise. In contrast, the linear magnification method [Wu et al. 2012] amplifies the noise linearly in α (Fig. 6).

Still, noise in the input sequence can also cause the phase signal itself to be noisy, which can result in incorrect motions being amplified. We found that we consistently got better results when low-passing the phase signal spatially as a simple way to increase its SNR. However, as the phase-signal in regions of low amplitude is not meaningful, we use an amplitude-weighted spatial Gaussian blur on the phases. For each band i of the representation and each frame k, we have a phase signal $\phi_{i,k}$ and amplitude $A_{i,k}$. We compute a weighted Gaussian blur:

$$\frac{(\phi_{i,k}A_{i,k}) * K_\rho}{A_{i,k} * K_\rho} \tag{17}$$

where K_ρ is a Gaussian kernel given by $\exp(-\frac{x^2+y^2}{\rho^2})$. We chose ρ to be equal to that of the spatial domain filter widths. This step incurs a small computational cost that may be avoided for performance considerations, as the results without it are usually good.

[1] Notice that the bound on the phase-based method is expressed in terms of $\alpha\delta(t)$, while in [Wu et al. 2012] it is expressed in terms of $(1+\alpha)\delta(t)$. This is because in this paper, we express the motion magnified image profile at at time t is generated by modifying (phase-shifting) the shifted, but unamplified image profile at time t, whereas in the analysis in [Wu et al. 2012], the motion magnified image profile at time t is generated by modifying the unshifted image profile at time 0.

4 Results

Our algorithm allows users to see small motions without excessive noise or computational cost, as well as remove motions that may distract from an underlying phenomena of interest. We show several applications of our algorithm in this section. Please refer to the supplemental video for the full sequences and results.

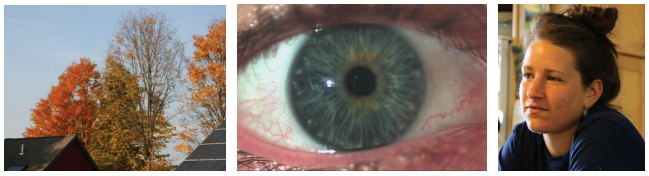
Unless mentioned otherwise, our processing was done using a complex steerable pyramid with a half-octave bandwidth filters and eight orientations. We computed the filter responses in the frequency domain. The processing was done in YIQ color space and processing was done on each channel independently. If running time is a concern (for a real-time application), amplifying only the luminance channel will give good results. Processing a 512×512 video with 300 frames took 56 seconds with an octave-bandwidth pyramid and two orientations, and 280 seconds with the aforementioned half-octave pyramid, using non-optimized MATLAB code on a laptop with 4 cores and 16GB of RAM. With an octave-bandwidth pyramid and 2 orientations, our method can be efficiently implemented to run in real time similar to Wu et al. [2012], as computing a compact steerable–rather than Laplacian–decomposition introduces a relatively minor performance overhead (about 8x slower, but still within the 30 frames per second range on 512×512 videos using an efficient C++ or GPU implementation). Also similar to Wu et al. [2012], the user has control over the amplification factor and the temporal bandpass filter.

A Big World of Small Motions The world is full of subtle and small motions that are invisible to the naked eye. Our phase-based approach allows pushing motion magnification further than before, to reveal imperceptible phenomena, not previously visualized, in clarity and detail.

In *eye* (Fig. 7), we were able to magnify subtle, involuntary, low amplitude (10-400 micron) movements in the human eye and head such as *microsaccades* [Rolfs 2009]. This video was taken with a high speed camera at 500 Hz. A one second (500 frames) sequence was processed with an ideal bandpass filter with passband between $30 - 50$ Hz and the motions were amplified 150x. A spatial mask was applied to the phase shifts to emphasize the motion around the iris. Such a detection system may have medical applications, as the frequency content of ocular microtremor was shown to have clinical significance [Bojanic et al. 2001].

Structures are design to sway in the wind, but their motion is often invisible. In *crane*, we took a video of a construction crane on a uniform background during a windy day. In the original video, the superstructure does not appear to move, however when amplifying low-frequency motions in the video within $0.2 - 0.4$ Hz 150x, the swaying of the crane's mask and undulation of its hook become apparent. For this sequence, a half-octave pyramid yields good results, however, because the crane was a solitary moving object over a uniform background, we found that we were able to further increase the motion and remove artifacts by using a quarter-octave pyramid (Fig. 4(c)).

Trees and *woman* (Fig. 7) demonstrate ordinary videos also contain changes at different frequencies over time that we cannot normally perceive. In *trees*, motions of lower temporal frequency correspond to larger structures (heavy branches), while motions of higher temporal frequency correspond to smaller structures (leaves). A simple interface allows the user to *sweep* through the frequency domain and examine temporal phenomena in a simple and intuitive manner.



trees *eye* *woman*

Figure 7: *A big world of small motions. Representative frames from videos in which we amplify imperceptible motions. The full sequences and results are available in the supplemental video.*

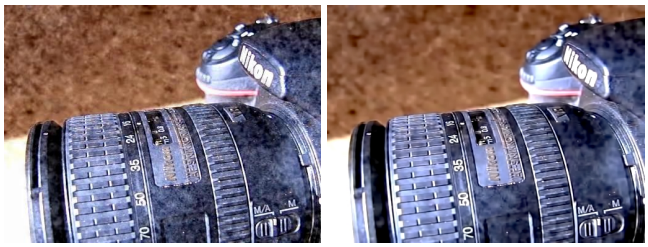
Comparison with Wu et al. [2012] The main differences between the phase-based approach and Wu et al.'s approach are summarized in Table 1. In particular, the new method supports larger amplification factors and gives a fundamentally better way of handling noise for Eulerian motion magnification. To demonstrate that, we compared the results from this work with those from Wu et al. [2012]. Several comparisons are available in Fig. 1 and the supplemental video. To illustrate that shifting phases is better than directly modifying pixel intensities, we did not spatially-smooth the phase signal in these comparisons.

On *all* the sequences we tested, we found the proposed approach to perform better. In particular, the magnified motions in the phase-based results (e.g. the respiratory motions of the baby and the vibrations of the guitar strings) appear crisper, and contain significantly fewer artifacts and noise.

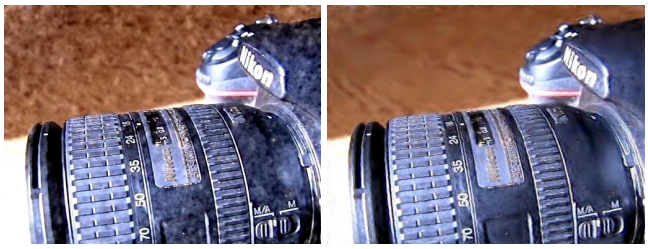
We also compared the phase-based results with noise removal processing not suggested in the Wu et al. paper: preceding and following the linear magnification processing by video denoising. We tested several denoising algorithms, namely NL-means [Buades et al. 2008], VBM3D [Dabov et al. 2007], and the recent motion-based denoising algorithm by Liu and Freeman [2010]. We tuned the denoising methods so as to produce the best result on each sequence. We achieved the overall best performance with VBM3D applied to the motion-magnified video (comparisons with all the denoising methods in pre- and post-processing are available in the supplementary material). We found that in some cases (e.g. *guitar*) denoising the video before magnification in fact kills the low-amplitude motion signal we are after. For the low-noise *baby* and *guitar* sequences, the denoised results were visually comparable to that of the phase-based method, although achieved at a higher computational cost, 17 times slower. For the higher-noise *camera* and *eye* sequences, the denoised Wu et al. result looks significantly worse than the phase-based results, as the denoising algorithms cannot do much with the medium frequency noise (Fig. 8).

	Linear [Wu et al. 2012]	**Phase-based (This paper)**
Decomposition	Laplacian pyramid	Complex steerable pyramid
Over-complete	$4/3$	$2k/(1-2^{-2/n})$
Exact for	Linear ramps	Sinusoids
Bounds	$(1+\alpha)\delta(t) < \lambda/8$	$\alpha\delta(t) < \lambda n/4$
Noise	Magnified	Translated

Table 1: *The main differences between the linear approximation of Wu et al. [2012] and our approach for motion magnification. The representation size is given as a factor of the original frame size, where k represents the number of orientation bands and n represents the number of filters per octave for each orientation.*

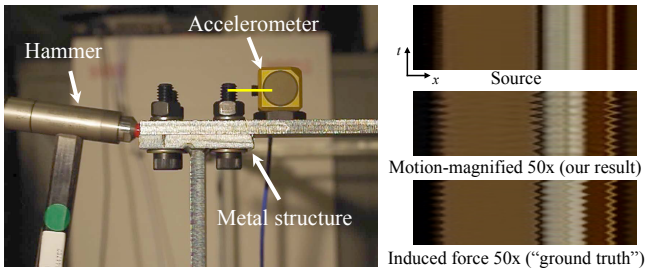


(a) Linear [Wu et al. 2012] (b) (a) denoised by [Dabov et al. 2007]

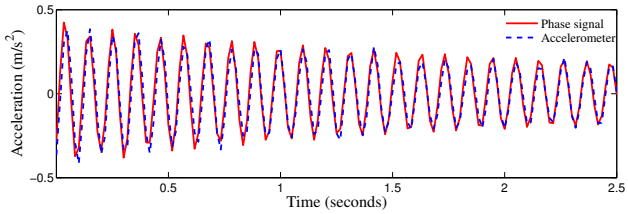


(c) (a) denoised by [Liu and Freeman 2010] (d) Phase-based

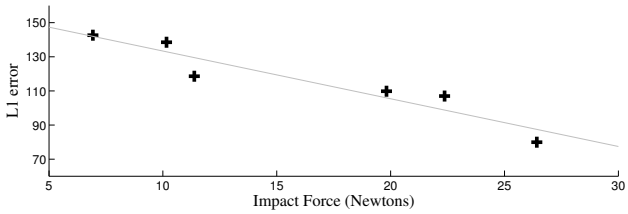
Figure 8: *Comparison of our result on the* camera *sequence (d) with the result of Wu et al. [2012] (a), denoised by two state-of-the-art video denoising algorithms: VBM3D [Dabov et al. 2007] (b) and motion-based denoising by Liu and Freeman [2010] (c). The denoising algorithms cannot deal with the medium frequency noise, and are computationally intensive. The full videos and similar comparisons on other sequences are available in the supplementary material.*



(a) Experimental setup (b) Spatiotemporal slices



(c) Acceleration comparison



(d) Error vs. force

Figure 9: *A controlled motion magnification experiment to verify our framework. (a) A hammer strikes a metal structures which then moves with a damped oscillatory motion. (b) A sequence with oscillatory motion of amplitude* 0.1 *pixels is magnified* 50 *times using our algorithm and compared to a sequence with oscillatory motion of amplitude* 5 *pixels (*50 *times the amplitude). (c) A comparison of acceleration extracted from the video with the accelerometer recording. (d) The error in the motion signal we extract from the video, measured as in (c), as function of the impact force. Our motion signal is more accurate as the motions in the scene get larger. All videos are available in the supplementary material.*

Controlled Experiments At the miniature scales of motion we are after, one might ask: are the signals we pick out and amplify real (the actual motion signals in the scene)? Would our magnified motions resemble the motions in the scene had they actually been actually larger? To answer these questions, we conducted two controlled experiments. In the first, we recorded ground truth motion data along with a (natural) video (*structure*, Fig. 9). We induced small motions in a metal structure, and affixed an accelerometer to it to capture its vibrations. To induce the motion we used an impact hammer with a sensor at its tip allowing to record the exact amount of force applied. We then recorded the structure using a standard DSLR video camera at 60 frames per second, along with the accelerometer and impact hammer data. We applied our transform to every frame and recorded the phase changes between the Nth frame and the first frame in one level of the pyramid oriented in the direction of the motion for a salient region of pixels near the accelerometer. These phase changes corresponded to displacement. To recover acceleration, we took a second derivative of Gaussian filter. Once scaled and aligned, the resulting signal matched the data from the accelerometer very closely 9(c). We also took two different sequences of the structure, one in which the amplitude of the oscillatory motion was 0.1 pixels and another in which it was 5 pixels (50x larger, from a harder hammer hit). We magnified the former 50 times and found the result to be visually comparable to the latter (Fig. 9(b)).

In a second experiment, we mount a sheet of rubber on a section of PVC pipe using a rubber band to create a tense membrane (Fig. 2). We use a loudspeaker to vibrate air in specific frequencies that in turn vibrates the membrane, and capture the result with a high speed camera. Through experimentation, we found two modes of the membrane when waveforms at 76Hz and 110Hz were sent through the loudspeaker. We then took a video of the membrane when a composite waveform of these two frequencies was sent through the loudspeaker and used our algorithm to separate and amplify these two modes. The results of this experiment are in the supplemental material.

Motion Attenuation Our phase-based formulation also lends itself naturally to attenuation of motions in videos, which allows us to remove low-amplitude, short-term motions while larger amplitude motions continue to pass through. Motion attenuation is achieved by setting the amplification factor α to a negative value in the range $[-1, 0)$, where $\alpha = -1$ zeros-out all the phase changes over time within the desired frequency band, effectively canceling out the motions within that band. The result is not the same as a constant frame as the coefficient amplitudes are still evolving over time. This is similar to motion denoising [Rubinstein et al. 2011] and video deanimation [Bai et al. 2012], but can be done efficiently in our approach (when the motions in the scene are small enough).

We apply motion attenuation for two applications: turbulence removal and color amplification (Fig. 10). Atmospheric turbulence is manifested as low-mid frequency jitters in a video of the moon as it passes through the night sky (see supplemental video). We pass a temporal window over the video (we used a window of 11 frames), transformed to our representation, and set the phases in each spa-

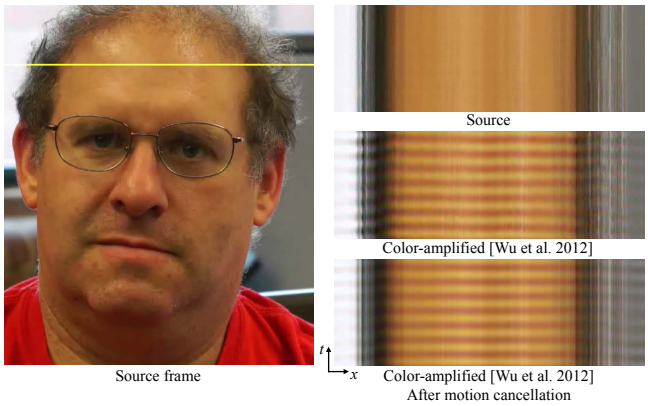


Figure 10: *Motion attenuation stabilizes unwanted head motions that would otherwise be exaggerated by color amplification. The full sequence is available in the supplemental video.*

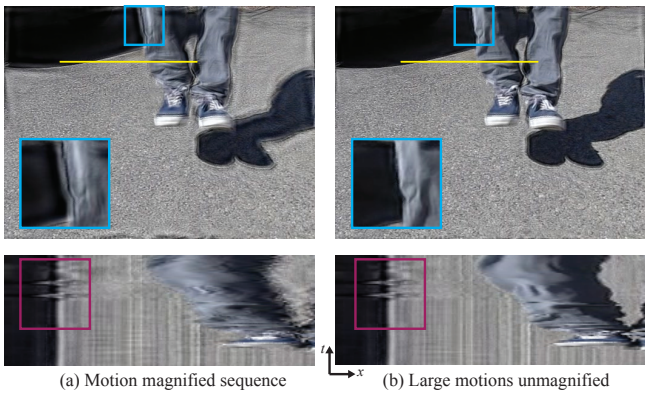


Figure 11: *Motion magnification can cause artifacts (cyan insets and spatiotemporal timeslices) in regions of large motion such as those in this sequence of a boy jumping on a platform (a). We can automatically remove such artifacts by identifying regions where the phase change exceeds our bound or a user-specified threshold (b). When the boy hits the platform, the time slice (purple highlights) shows that the subtle motions due to impact with the platform are magnified in both cases.*

tial scale and orientation of the center frame to the corresponding median phase of the transformed frames within the temporal window. This effectively *shifts* pixels in order to compensate for the turbulent motions.

Since the magnification method of Wu et al. [2012] amplifies color changes and motions *jointly*, small motions of the face become much larger, visible when amplifying the color changes corresponding to the pulse, which may not be desirable. By canceling the motions as a pre-process to their algorithm, we are able to remove those motions from their results (Fig. 10).

A similar color amplification result as that of Wu et al. [2012] can be achieved entirely with steerable pyramids. We can temporally bandpass the amplitude A_ω (Eq. 2) and the low pass residual and add a multiple of the resulting amplitude variations to the amplitude signal. This yields similar results because in both cases the same processing is applied to the low-pass residual band of an image pyramid (Laplacian pyramid in one case, steerable pyramid in the other).

5 Discussion and Limitations

Lagrangian approaches to motion magnification (e.g. [Liu et al. 2005]) are complementary to the Eulerian approach proposed in this paper. Such methods can amplify the motion in a video arbitrarily, but rely on accurate optical flow estimates, image segmentation, and inpainting. Such processing is difficult to do well and requires long computation times. In addition, Wu et al [2012] showed (Section 5 and Appendix A in their paper) that for moderate magnification and noisy inputs, the Eulerian approach performs better than Lagrangian. The phase-based method significantly reduces the sensitivity to noise of Eulerian video magnification over that of Wu et al., as well as increases its supported range of amplification, which further expands the regime where it performs better than Lagrangian approaches. Since the main contribution of this paper is in an improved Eulerian approach for motion processing, comparisons were done with the state-of-the-art Eulerian method.

While the analysis of Wu et al. [2012] is exact in the case of linear ramps, the phase-based approach is exact for sinusoidal waves (Fig. 3), since such signals contain only a single spatial frequency. However, both methods rely on spatial pyramids, where each level is band limited. We argue that such spatially bandpassed images are better approximated by sinusoidal waves than linear ramps.

Our half-octave bandwidth pyramid representation, in which the windowing function of the wavelets in the primal domain is larger, extends the magnification capability of Wu et al. [2012] by a factor of 4, and pyramids with more filters per octave may improve on it by even larger factors. While this allows us to magnify motions further, the wavelets are also more likely to span multiple motions as their support get larger, which may corrupt the phase signal and eventually lead to artifacts in the results. Currently, the user can select the desired representation based on the motions in the scene and the available computational resources.

If the input video has large motions, than the bandpassed phase (Eq. 4) will not reflect the true motion in the scene and the motion magnified video will suffer from artifacts in the regions of large motion (Fig. 11(a)). To mitigate this, we can automatically detect regions where phase exceeds our bound (or some user-specified threshold) and set the amplification to be zero in these regions. To increase robustness, we spatiotemporally lowpass the absolute value of the phase and compare the result to a threshold to determine which regions have large motions. The supplememntal video shows an example.

Finally, for sequences in which the phase signal is noisy, parts of the image in the magnified video may appear to move incoherently. Using an image or motion prior to regularize the processing may improve the results in such cases, and is a direction for future work.

6 Conclusion

We describe an efficient, Eulerian method for processing and manipulating small motions in standard videos by analyzing the local phase over time at different orientations and scales. The local phase is computed using complex steerable pyramids, which we extend to work with sub-octave bandwidth filters in order to increase the spatial support of the filters and allow us to push motion magnification further. Our method then magnifies the temporal phase differences in the corresponding bands of these pyramids to hallucinate bigger or smaller motions. We demonstrated that this phase-based technique improves the state-of-the-art in Eulerian motion processing both in theory and in practice, provides a fundamentally better way of handling noise, and produces high quality photo-realistic videos with amplified or attenuated motions for a variety of applications.

Acknowledgements We would like to thank the SIGGRAPH reviewers for their comments. We thank Justin Chen for his assistance with the controlled metal structure experiment. We acknowledge funding support from: Quanta Computer, Shell Research, the DARPA SCENICC program, NSF CGV-1111415 and a gift from Cognex. Michael Rubinstein was supported by the Microsoft Research PhD Fellowship. Neal Wadhwa was supported by the DoD through the NDSEG fellowship program.

References

BAI, J., AGARWALA, A., AGRAWALA, M., AND RAMAMOORTHI, R. 2012. Selectively de-animating video. *ACM Transactions on Graphics*.

BOJANIC, S., SIMPSON, T., AND BOLGER, C. 2001. Ocular microtremor: a tool for measuring depth of anaesthesia? *British Journal of Anaesthesia 86*, 4, 519–522.

BUADES, A., COLL, B., AND MOREL, J.-M. 2008. Nonlocal image and movie denoising. *International Journal of Computer Vision 76*, 123–139.

DABOV, K., FOI, A., AND EGIAZARIAN, K. 2007. Video denoising by sparse 3d transform-domain collaborative filtering. In *Proc. 15th European Signal Processing Conference*, vol. 1, 7.

FLEET, D. J., AND JEPSON, A. D. 1990. Computation of component image velocity from local phase information. *Int. J. Comput. Vision 5*, 1 (Sept.), 77–104.

FREEMAN, W. T., ADELSON, E. H., AND HEEGER, D. J. 1991. Motion without movement. *SIGGRAPH Comput. Graph. 25* (Jul), 27–30.

FUCHS, M., CHEN, T., WANG, O., RASKAR, R., SEIDEL, H.-P., AND LENSCH, H. P. 2010. Real-time temporal shaping of high-speed video streams. *Computers & Graphics 34*, 5, 575–584.

GAUTAMA, T., AND VAN HULLE, M. 2002. A phase-based approach to the estimation of the optical flow field using spatial filtering. *Neural Networks, IEEE Transactions on 13*, 5 (sep), 1127 – 1136.

LIU, C., AND FREEMAN, W. 2010. A high-quality video denoising algorithm based on reliable motion estimation. In *Computer Vision ECCV 2010*, K. Daniilidis, P. Maragos, and N. Paragios, Eds., vol. 6313 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 706–719.

LIU, C., TORRALBA, A., FREEMAN, W. T., DURAND, F., AND ADELSON, E. H. 2005. Motion magnification. *ACM Trans. Graph. 24* (Jul), 519–526.

PORTILLA, J., AND SIMONCELLI, E. P. 2000. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vision 40*, 1 (Oct.), 49–70.

ROLFS, M. 2009. Microsaccades: Small steps on a long way. *Vision Research 49*, 20, 2415 – 2441.

RUBINSTEIN, M., LIU, C., SAND, P., DURAND, F., AND FREEMAN, W. T. 2011. Motion denoising with application to time-lapse photography. *IEEE Computer Vision and Pattern Recognition (CVPR)* (June), 313–320.

SIMONCELLI, E. P., AND FREEMAN, W. T. 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, 3444–.

SIMONCELLI, E. P., FREEMAN, W. T., ADELSON, E. H., AND HEEGER, D. J. 1992. Shiftable multi-scale transforms. *IEEE Trans. Info. Theory 2*, 38, 587–607.

WANG, J., DRUCKER, S. M., AGRAWALA, M., AND COHEN, M. F. 2006. The cartoon animation filter. *ACM Trans. Graph. 25*, 1169–1173.

WU, H.-Y., RUBINSTEIN, M., SHIH, E., GUTTAG, J., DURAND, F., AND FREEMAN, W. 2012. Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph. (Proc. SIGGRAPH) 31* (aug).

A Improved Radial Windowing Function for Sub-octave Bandwidth Pyramids

When generalizing the complex steerable pyramid of Portilla and Simoncelli [2000] to sub-octave bandwidth pyramids, we found empirically that their windowing function was well-suited for octave and half-octave pyramids. However, at a larger number of filters per octave (≥ 3 in our experiments) this scheme produces filters which are very sharp in the frequency domain and have noticeable ringing artifacts (shown in the 1D wavelet plot of Fig. 4(b)).

They define their filters in terms of independent radial and angular windowing functions. For quarter-octave and larger pyramids, we leave the angular windowing function unchanged and propose a different radial windowing function, given by

$$\cos^6(\log_2(r))I_{[-\pi/2,\pi/2]}(\log_2(r)). \tag{18}$$

This function has two nice properties: (a) it is smoother, more similar to a Gaussian, and does not introduce ringing in the primal domain, and (b) squared copies scaled by a power of $\frac{\pi}{7}$ sum to a constant factor, so that the transform is invertible and we get perfect reconstruction (Eq. 1). An example quarter-octave pyramid generated with this windowing function in shown in Fig. 4(c) and its results for motion magnification are available in the supplemental video.