DOI:10.1145/3015573

Eulerian Video Magnification and Analysis

By Neal Wadhwa, Hao-Yu Wu, Abe Davis, Michael Rubinstein, Eugene Shih, Gautham J. Mysore, Justin G. Chen, Oral Buyukozturk, John V. Guttag, William T. Freeman, and Frédo Durand

Abstract

The world is filled with important, but visually subtle signals. A person's pulse, the breathing of an infant, the sag and sway of a bridge—these all create visual patterns, which are too difficult to see with the naked eye. We present Eulerian Video Magnification, a computational technique for visualizing subtle color and motion variations in ordinary videos by making the variations larger. It is a *microscope for small changes* that are hard or impossible for us to see by ourselves. In addition, these small changes can be quantitatively analyzed and used to recover sounds from vibrations in distant objects, characterize material properties, and remotely measure a person's pulse.

1. INTRODUCTION

A traditional microscope takes a slide with details too small to see and optically magnifies it to reveal a rich world of bacteria, cells, crystals, and materials. We believe there is another invisible world to be visualized: that of tiny motions and small color changes. Blood flowing through one's face makes it imperceptibly redder (Figure 1a), the wind can cause structures such as cranes to sway a small amount (Figure 1b), and the subtle pattern of a baby's breathing can be too small to see. The world is full of such tiny, yet meaningful, temporal variations. We have developed tools to visualize these temporal variations in position or color, resulting in what we call a motion, or color, microscope. These new microscopes rely on computation, rather than optics, to amplify minuscule motions and color variations in ordinary and high-speed videos. The visualization of these tiny changes has led to applications in biology, structural analysis, and mechanical engineering, and may lead to applications in health care and other fields.

We process videos that may look static to the viewer, and output modified videos where motion or color changes have been magnified to become visible. In the input videos, objects may move by only 1/100th of a pixel, while in the magnified versions, motions can be amplified to span many pixels. We can also quantitatively analyze these subtle signals to enable other applications, such as extracting a person's heart rate from video, or reconstructing sound from a distance by measuring the vibrations of an object in a high-speed video (Figure 1c).

The algorithms that make this work possible are simple, efficient, and robust. Through the processing of local color or phase changes, we can isolate and amplify signals of interest. This is in contrast with earlier work to amplify small motions[13] by computing per-pixel motion vectors and then displacing pixel values by magnified motion vectors. That technique yielded good results but it was computationally expensive, and errors in the motion analysis would generate artifacts in the motion magnified output. As we will show, the secret to the simpler processing described in this article lies in the properties of the small motions themselves.

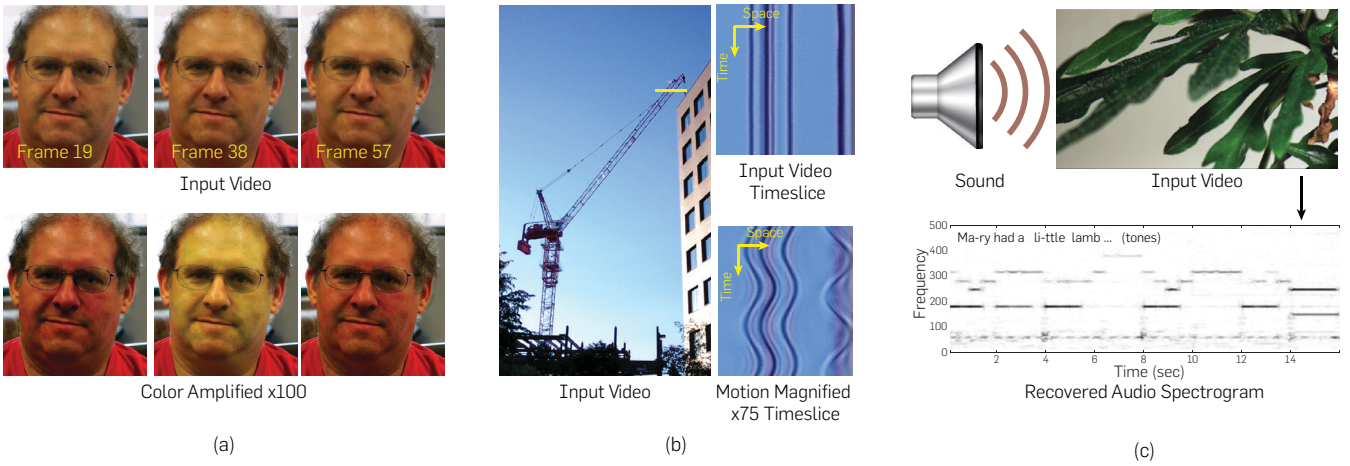
To compare our new work to the previous motion-vector work, we borrow terminology from fluid mechanics. In a *Lagrangian* perspective, the motion of fluid particles is tracked over time from the reference frame of the particles themselves, similar to observing a river flow from the moving perspective of a boat. This is the approach taken by the earlier work, tracking points in the scene and advecting pixel colors across the frame. In contrast, an *Eulerian* perspective uses a fixed reference frame and characterizes fluid properties over time at each fixed location, akin to an observer watching the water from a bridge. The new techniques we describe follow this approach by looking at temporal signals at fixed image locations.

The most basic version of our processing looks at intensity variations over time at each pixel and amplifies them. This simple processing reveals both subtle color variations *and* small motions because, for small sub-pixel motions or large structures, motion is linearly related to intensity change through a first-order Taylor series expansion (Section 2). This approach to motion magnification breaks down when the amplification factor is large and the Taylor approximation is no longer accurate. Thus, for most motion magnification applications we develop a different approach, transforming the image into a complex steerable pyramid, in which position is explicitly represented by the phase of spatially localized sinusoids. We exaggerate the phase variations observed over time, modifying the coefficients of the pyramid representation. Then, the pyramid representation is collapsed to produce the frames of a new video sequence that shows amplified versions of the small motions (Section 3). Both Eulerian approaches lead to faster processing and fewer artifacts than the previous Lagrangian approach. However, the Eulerian approaches only work well for small motions, not arbitrary ones.

Making small color changes and motions visible adds a dimension to the analysis that goes beyond simply

This Research Highlight is a high-level overview of three papers about tiny changes in videos: Eulerian Video Magnification for Revealing Subtle Changes in the World,[24] Phase-Based Video Motion Processing,[22] and The Visual Microphone: Passive Recovery of Sound from Video.[7]

Figure 1. Apparently still videos of a face, a construction crane, and a houseplant have subtle changes that can be revealed using Eulerian video magnification and analysis. Blood flow in a man's face is revealed when the color changes are amplified (a). The construction crane's motions are revealed when amplified 75× (b). A houseplant subtly vibrates in tune with a loudspeaker playing "Mary had a little lamb." The audio is recovered from a silent video of the house plant (c).



measuring color and position changes. The visualization lets a viewer interpret the small changes, and find patterns that simply measuring numbers would not reveal. It builds intuition and understanding of the motions and changes being revealed. We show results of Eulerian video magnification in a wide variety of fields, from medicine and civil engineering to analyzing subtle vibrations due to sound. Videos and all of our results are available on our project webpage (http://people.csail.mit.edu/mrub/vidmag/).

2. LINEAR VIDEO MAGNIFICATION

The core idea of Eulerian video magnification is to independently process the time series of color values at each pixel. We do this by applying standard 1D temporal signal processing to each time series to amplify a band of interesting temporal frequencies, for example, around 1 Hz (60 beats per minute) for color changes and motions related to heart-rate. The new resulting time series at each pixel yield an output video where tiny changes that were impossible to see in the input, such as the reddening of a person's face with each heart beat or the subtle breathing motion of a baby, are magnified and become clearly visible.

The idea of applying temporal signal processing to each pixels' color values is a straightforward idea, and has been explored in the past for regular videos.[10, 16] However, the results have been limited because such processing cannot handle general spatial phenomena such as large motions that involve complicated space-time behavior across pixels. When a large motion occurs, color information travels across many pixels and a Lagrangian perspective, in which motion vectors are computed, is required. One critical contribution of our work is the demonstration that in the special case of small motions, Eulerian processing can faithfully approximate their amplification. Because the motions involved are small, we can make first-order Taylor arguments to show that linear, per-pixel amplification of color variations closely approximates a larger version of the motion. We now formalize this for the special case of 1D translational motion of a diffuse object under constant lighting, but the argument applies to arbitrary phenomena such as 3D motion and shiny objects, as we discuss below.

2.1. 1D translation

Consider a translating 1D image with intensity denoted by $I(x, t)$ at position x and time t. Because it is translating, we can express the image's intensities with a displacement function $\delta(t)$, such that $I(x, t) = f(x - \delta(t))$ and $I(x, 0) = f(x)$. Figure 2 shows the image at time 0 in black and at a later time translated to the right in blue. The goal of motion magnification is to synthesize the signal

$$\hat{I}(x,t) = f(x-(1+\alpha)\delta(t)) \tag{1}$$

for some amplification factor α.

We are interested in the time series of color changes at each pixel:

$$B(x,t) := I(x,t) - I(x,0). \tag{2}$$

Under the assumption that the displacement $\delta(t)$ is small, we can approximate the first term with a first-order Taylor series expansion about x, as

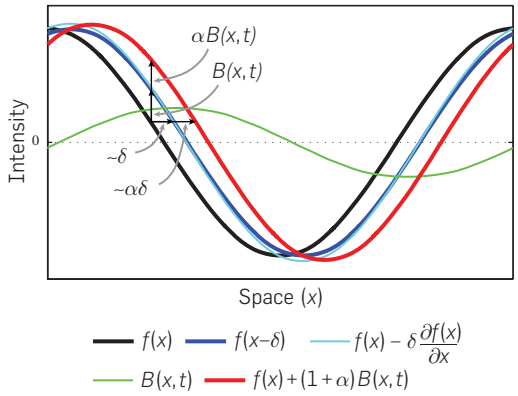
$$I(x,t) \approx f(x) - \delta(t)\frac{\partial f(x)}{\partial x}. \tag{3}$$

Because $I(x, 0) = f(x)$, the color changes at x are

$$B(x,t) \approx -\delta(t)\frac{\partial f(x)}{\partial x}. \tag{4}$$

This is the first order approximation to the well known brightness constancy equation in optical flow[12, 14]: the intensity variation at a pixel x is the negative of the product between the displacement and the spatial gradient. This

Figure 2. Amplifying per-pixel intensity variations can approximate spatial translation. The input signal is shown at two times: $I(x, 0) = f(x)$ at time 0 (black) and $I(x, t) = f(x - \delta)$ at time t (blue). The first-order Taylor series expansion of $I(x, t)$ around x approximates the translated signal (cyan). The pointwise difference between the frames (green) is amplified and added to the original signal to generate a larger translation (red). Here, the amplification factor α is 1, amplifying the motion by 100%.



can be seen as a right triangle in Figure 2, whose legs are the temporal intensity variation (vertical edge marked $B(x, t)$) and the displacement (horizontal edge marked δ) and whose hypotenuse (blue curve between the legs) has slope equal to the image's spatial derivative $\left(\frac{\partial f(x)}{\partial x}\right)$.

In our processing, we amplify the color change signal $B(x, t)$ by α and add it back to $I(x, t)$, resulting in the processed signal (red in Figure 2):

$$\tilde{I}(x,t) = I(x,t) + \alpha B(x,t). \quad (5)$$

Combining equations (3)–(5), we have

$$\tilde{I}(x,t) \approx f(x) - (1+\alpha)\delta(t)\frac{\partial f(x)}{\partial x}. \quad (6)$$

As long as $(1 + \alpha)\delta$ is small enough that a first-order Taylor expansion is valid, we can relate the previous equation to motion magnification (Eq. 1). It is simply

$$\tilde{I}(x,t) \approx f(x - (1+\alpha)\delta(t)). \quad (7)$$

This shows that this processing magnifies motions. The spatial displacement $\delta(t)$ between frames of the video at times 0 and t, has been amplified by a factor of $(1 + \alpha)$.

2.2. General case

Consider a subtle, temporal phenomenon, for example, 3D translation, rotation, or the motion of light, parameterized by a vector $\boldsymbol{\theta}$ (perhaps representing the position or orientation of objects or lights) that evolves over time as $\boldsymbol{\theta}(t)$. These parameters can be mapped to image intensities $I(\boldsymbol{x}, t)$ via a function $f(\boldsymbol{x}, \boldsymbol{\theta}(t))$ for all spatial locations $\boldsymbol{x}$. If f is a differentiable function of the parameters $\boldsymbol{\theta}$ and the changes in the parameters are small, then the video I can be approximated by its first order Taylor expansion around $\boldsymbol{\theta}(0)$

$$I(\boldsymbol{x},t) \approx f(\boldsymbol{x},\boldsymbol{\theta}(0)) + \nabla f(\boldsymbol{x},\boldsymbol{\theta}(0))^T(\boldsymbol{\theta}(t) - \boldsymbol{\theta}(0)). \quad (8)$$

That is, each pixel in the video signal is linearly related to the deviation of the parameters $\boldsymbol{\theta}$ from their initial value. If we amplify by α the difference between the image at time t and at time 0, we get

$$\tilde{I}(\boldsymbol{x},t) := f(\boldsymbol{x},\boldsymbol{\theta}(0)) + (1+\alpha)\nabla f(\boldsymbol{x},\boldsymbol{\theta})(0))^T(\boldsymbol{\theta}(t) - \boldsymbol{\theta}(0)). \quad (9)$$

By the same analysis as before, this is approximately equal to a new video in which the variations in $\boldsymbol{\theta}$ are larger by a factor $1 + \alpha$. This shows that linear Eulerian video magnification can be used to magnify many subtle, temporal phenomena. It is agnostic to the underlying imaging model and can even work in cases where brightness constancy is not true as long as the changes are small.

2.3. Limitations of the linear approach

Linear amplification relies on a first-order Taylor-expansion, which breaks down when either the amplification factor or the input motion is too large. For overly large amplification factors, the magnified video overshoots and undershoots the video's white and black levels causing clipping artifacts near edges where the second derivative $\left(\frac{\partial^2 f(x)}{\partial x^2}\right)$ is non-negligible (Figure 6a). When the input motion is too large, the initial Taylor expansion is inaccurate (Eq. 3) and the output contains ghosting artifacts instead of magnified motions.

A second limitation is that noise in the video is amplified. For example, suppose the intensity value $I(x, t)$ has an independent white Gaussian noise term $n(x, t)$ of variance σ^2 added to it. The difference between the frame at time t and at time 0 then contains the noise term, $n(x, t) - n(x, 0)$, with noise variance $2\sigma^2$. This noise term gets amplified by a factor α and the output video has noise of variance $2\alpha^2\sigma^2$, a much larger amount than in the input video (Figure 7b).

In Wu et al.,[24] noise amplification was partially mitigated by reducing the amplification of high spatial-frequency temporal variations, assuming that they are mostly noise rather than signal. This is done by constructing a Laplacian pyramid of the temporal variations and using a lower amplification factor for high spatial-frequency levels. Spatially lowpassing the temporal variations produces comparable results. A thorough noise analysis of this approach is available in the appendix of Wu et al.[24] and more information about signal-to-noise ratios is given here in Section 4.

3. PHASE-BASED MAGNIFICATION

The appeal of the Eulerian approach to video magnification is that it independently processes the time series of color values at each pixel and does not need to explicitly compute motions. However, its reliance on first-order approximations limits its scope, and its use of linear amplification increases noise power. In this section, we seek to continue using the Eulerian perspective of motion analysis—processing independent time series at fixed reference locations. But, we want to do so in a representation that better handles motions and is less prone to noise.

In the case of videos that are global translations of a frame over time, there is a representation, that is, exactly what we want: the Fourier series. Its basis functions are complex-valued sinusoids that, by the Fourier shift theorem, can be translated exactly by shifting their phase (Figure 3a,c). However, using the Fourier basis would limit us to only being able to handle the same translation across the entire frame, precluding the amplification of complicated spatially-varying motions. To handle such motions, we instead use spatially-local complex sinusoids implemented by a wavelet-like representation called the complex steerable pyramid.[19, 20] This representation decomposes images into a sum of complex wavelets corresponding to different scales, orientations, and positions. Each wavelet has a notion of local amplitude and local phase, similar to the amplitude and phase of a complex sinusoid (Figure 4a). The key to our new approach is to perform the same 1D temporal signal processing and amplification described earlier on the local phase of each wavelet, which directly corresponds to local motion as we discuss below.

3.1. Simplified global case

To provide intuition for how phase can be used to magnify motion, we work through a simplified example in which a global 1D translation of an image is magnified using the phase of Fourier basis coefficients (Figure 3).

Let image intensity $I(x, t)$ be given by $f(x - \delta(t))$ where $\delta(0) = 0$. We decompose the profile $f(x)$ into a sum of complex coefficients times sinusoids using the Fourier transform

$$f(x) = \sum_{\omega} A_{\omega} e^{i\phi_{\omega}} e^{-i\omega x}. \quad (10)$$

Because the frames of I are translations of f, their Fourier transform is given by a phase shift by $\omega\delta(t)$:

$$I(x,t) = \sum_{\omega} A_{\omega} e^{i\phi_{\omega}} e^{-i\omega(x-\delta(t))} = \sum_{\omega} A_{\omega} e^{i(\phi_{\omega}+\omega\delta(t))} e^{-i\omega x}, \quad (11)$$

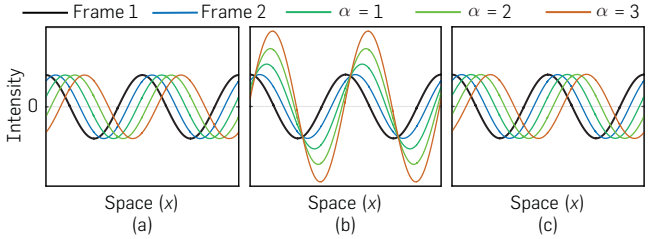
where the phase of these coefficients becomes $\phi_{\omega} + \omega\delta(t)$. If we subtract the phase at time 0 from the phase at time t, we get the phase difference

$$\omega\delta(t), \quad (12)$$

which is proportional to the translation. Amplifying this phase difference by a factor α and using it to shift the Fourier coefficients of $I(x, t)$ yields

$$\sum_{\omega} A_{\omega} e^{i\phi_{\omega}+(1+\alpha)\omega\delta(t)} e^{-i\omega x} = f(x-(1+\alpha)\delta(t)), \quad (13)$$

Figure 3. Phase-based motion magnification is perfect for Fourier basis functions (sinusoids). In these plots, the initial displacement is $\delta(t) = 1$. (a) True Amplification. (b) Linear. (c) Phase-Based.



a new image sequence in which the translations have been *exactly* magnified.

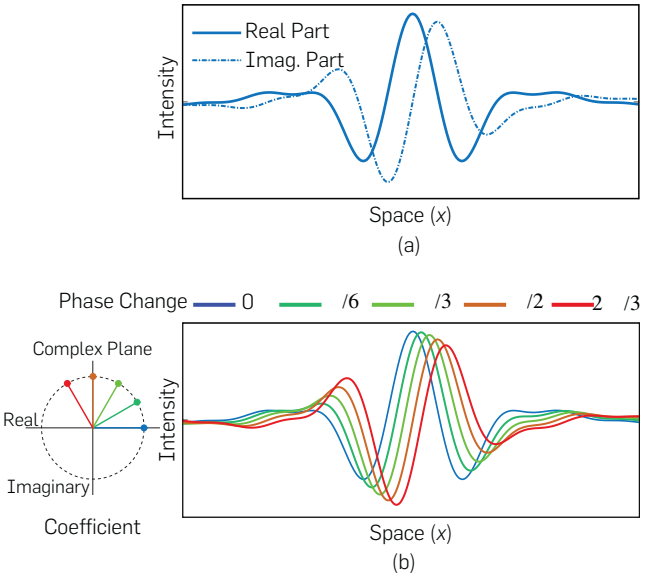
Phase-based magnification works perfectly in this case because the motions are global and because the transform breaks the image into a representation consisting of exact sinusoids (formally, the Fourier transform diagonalizes the translation operator). In most cases, however, the motions are not global, but local. This is why we break the image into local sinusoids using the complex steerable pyramid.

3.2. Complex steerable pyramid

The complex steerable pyramid[19, 20] is a complex, over-complete linear transform. It decomposes an image into a set of coefficients corresponding to basis functions that are simultaneously localized in position, spatial scale and orientation. The image is reconstructed by multiplying the coefficients by the basis functions and summing the real parts.

The transform is best-described by its self-similar basis functions. Each one is a translation, dilation, or rotation of another. So, it is sufficient to look at just one, a 1D version of which is shown in Figure 4a. It resembles an oriented complex sinusoid windowed by a Gaussian envelope. The complex sinusoid provides locality in frequency while the windowing provides locality in space. Each basis function is complex, consisting of a real, even-symmetric part (cosine) and an imaginary, odd-symmetric part (sine). This gives rise to a notion of local amplitude and local phase as opposed to the global amplitude and phase of Fourier basis functions. We use only a half-circle of orientations because basis functions at antipodal orientations $(\theta, \theta + \pi)$ yield redundant, conjugate coefficients.

Figure 4. Increasing the phase of complex steerable pyramid coefficients results in approximate local motion of the basis functions. A complex steerable pyramid basis function (a) is multiplied by several complex coefficients of constant amplitude and increasing phase to produce the real part of a new basis function, that is, approximately translating (b).



3.3. Local phase shift is local translation

The link between local phase shift and local translation has been studied before in papers about phase-based optical flow.[9, 11] Here, we demonstrate how local phase shift approximates local translation for a single basis function in a manner similar to the global phase-shift theorem of Fourier bases. We model a basis function as a Gaussian window multiplied by a complex sinusoid

$$e^{\frac{-x^2}{(2\sigma^2)}} e^{-i\omega x}, \tag{14}$$

where σ is the standard deviation of the Gaussian envelope and ω is the frequency of the complex sinusoid. In the complex steerable pyramid, the ratio between σ and ω is fixed because the basis functions are self-similar. Low frequency wavelets have larger windows.

Changing the *phase* of the basis element by multiplying it by a complex coefficient $e^{i\phi}$ results in

$$e^{\frac{-x^2}{(2\sigma^2)}} e^{-i\omega x} \times e^{i\phi} = e^{\frac{-x^2}{(2\sigma^2)}} e^{-i\omega(x-\phi/\omega)}. \tag{15}$$

The complex sinusoid under the window is translated, which is approximately a translation of the whole basis function by $\frac{\phi}{\omega}$ (Figure 4b).

Conversely, the phase difference between two translated basis elements is proportional to translation. Specifically, suppose we have a basis element and its translation by δ:

$$e^{\frac{-x^2}{(2\sigma^2)}} e^{-i\omega x}, e^{\frac{-(x-\delta)^2}{(2\sigma^2)}} e^{-i\omega(x-\delta)}. \tag{16}$$

The local phase of each element only depends on the argument to the complex exponential, and is $-\omega x$ in the first case and $-\omega(x-\delta)$ in the second. The phase difference is then $\omega\delta$, which is directly proportional to the translation. Local phase shift can be used both to analyze tiny translations and synthesize larger ones.

3.4. Our method

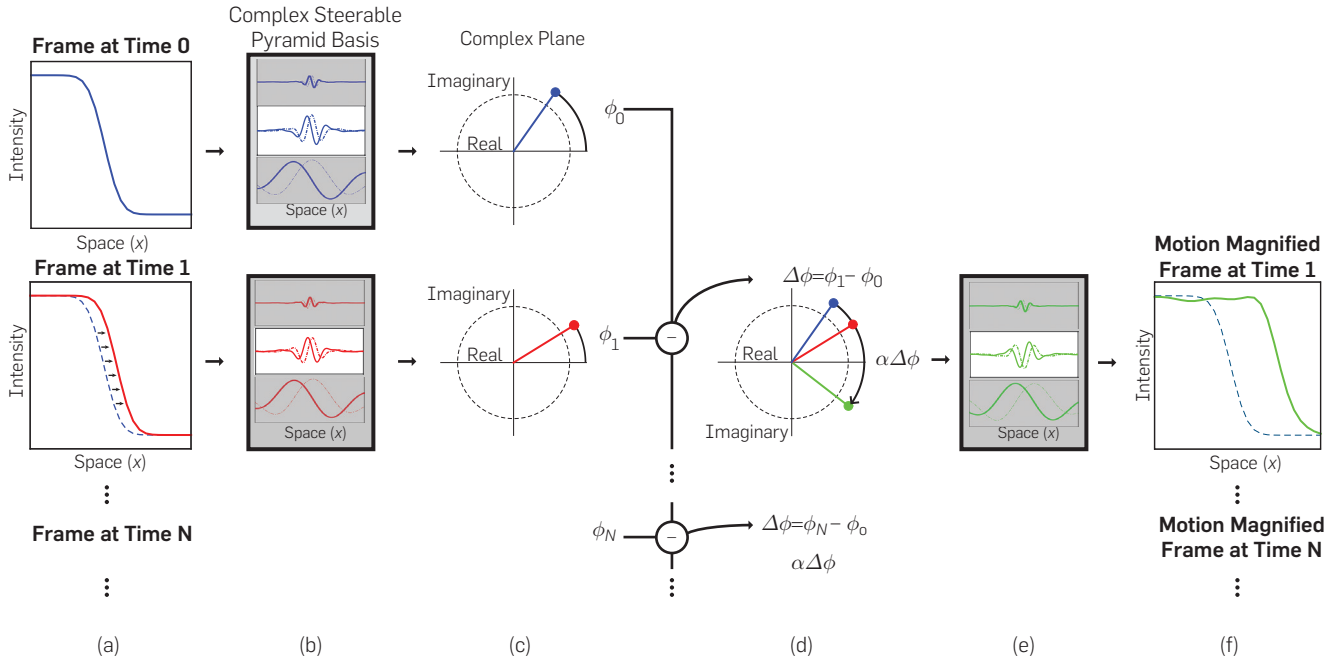
The observation that local phase differences can be used to manipulate local motions motivates our pipeline. We take an image sequence, project each frame onto the complex steerable pyramid basis and then independently amplify the phase differences between *all* corresponding basis elements. This is identical to the linear amplification pipeline except that we have changed the representation from intensities to local spatial phases.

To illustrate the pipeline, consider again an image sequence $I(x, t)$, in which the frame at time 0 is $f(x)$ and the frames at time t are translations $f(x-\delta(t))$ (Figure 5a). In our first step, we project each frame onto the complex steerable pyramid basis (Figure 5b), which results in a complex coefficient for every scale r, orientation θ and spatial location x, y, and time t. Because the coefficients are complex, they can be expressed in terms of amplitude $A_{r,\theta}$ and phase $\phi_{r,\theta}$ as

$$A_{r,\theta}(x, y, t)e^{i\phi_{r,\theta}(x,y,t)}. \tag{17}$$

In Figure 5c, we show coefficients at a specific location, scale, and orientation in the complex plane at times 0 and 1.

Figure 5. A 1D example illustrating how the local phase of complex steerable pyramid coefficients is used to amplify the motion of a subtly translating step edge. Frames (two shown) from the video (a) are transformed to the complex steerable pyramid representation by projecting onto its basis functions (b), shown in several spatial scales. The phases of the resulting complex coefficients are computed (c) and the phase differences between corresponding coefficients are amplified (d). Only a coefficient corresponding to a single location and scale is shown; this processing is done to all coefficients. The new coefficients are used to shift the basis functions (e) and a reconstructed video is produced in which the motion of the step edge is magnified (f).



Because the two frames are slight translations of each other, each coefficient has a slight phase difference. This is illustrated in Figure 5c, in which the coefficients have roughly the same amplitude but different phases. The next step of our processing is to take the phase differences between the coefficients in the video and those of a reference frame, in this case the frame at time 0:

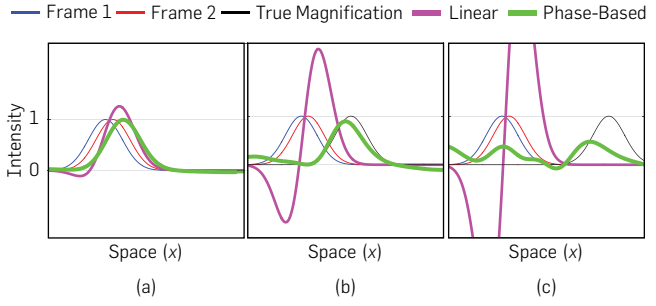
$$\Delta\phi_{r,\theta}(x, y, t) = \phi_{r,\theta}(x, y, t) - \phi_{r,\theta}(x, y, 0). \quad (18)$$

These phase differences are amplified by a factor α (Figure 5d), which yields a new set of coefficients for each frame, in which the amplitudes are the same, but the phase differences from the reference frame are larger. We reconstruct the new frames using these coefficients by multiplying them by the basis functions (Figure 5e). Then, we sum the real part to get new frames, in which the translations—and therefore the motions in the video—are amplified (Figure 5f).

Amplifying phase differences rather than pixel intensity differences has two main advantages: (a) it can support larger amplification factors, and (b) noise amplitude does not get amplified. In Figure 6, we show the two different methods being used to amplify the motions of a 1D Gaussian bump. Both methods work well for small amplification factors (Figure 6a). For larger amplification factors, amplifying raw pixel differences results in the signal overshooting the white level and undershooting the black level resulting in intensity clipping. In contrast, amplifying phase differences allows us to push the Gaussian bump much farther (Figure 6b). At very high amplification levels, the different spatial scales of the bump break apart because the high frequency components cannot be pushed as far as the lower frequency components (Figure 6c).

In Figure 7, we show the effect of both methods on a video, which consists of independent and identically distributed (iid) Gaussian noise. Unlike the linear method which increases noise power, the phase based method preserves noise power preventing objectionable artifacts in the motion magnified output. For these reasons, we found that amplifying phase differences rather than pixel differences is a better approach for magnifying small motions.

Figure 6. For non-periodic structures, both methods work for small amplification, $\alpha = 1.5$ (a). The phase-based method supports amplification factors four times as high as the linear method and does not suffer from intensity clipping artifacts, $\alpha = 6$ (b). For large amplification, different frequency bands break up because the higher frequency bands have smaller windows, $\alpha = 14$ (c).



3.5. Riesz pyramids

Using phase in the complex steerable pyramid to motion magnify videos can be slow because the representation is much larger than the input. We have developed another method that is similar in spirit and produces videos of almost the same quality. However, it is much faster and is capable of running in real-time on a laptop. More details about this can be found in this paper[23] on Riesz Pyramids.

4. AMPLIFYING THE RIGHT SIGNAL

Maximizing the signal-to-noise ratio of the temporal variations we amplify, whether local phase changes or color changes, is the key to good performance. We improve SNR by temporally and spatially filtering the variations to remove components that correspond to noise and keep those that correspond to signal. The temporal filtering also gives a way to isolate a signal of interest as different motions often occur at different temporal frequencies. A baby's squirming might be at a lower temporal frequency than her breathing.

Temporal narrowband linear filters provide a good way to improve signal-to-noise ratios for motions that occur in a narrow range of frequencies, such as respiration and vibrations. To prevent phase-wrapping issues when using these filters, we first unwrap the phases in time. The filters can also be used to isolate motions in an object that correspond to different frequencies. For example, a pipe vibrates at a preferred set of modal frequencies, each of which has a different spatial pattern of vibration. We can use video magnification to reveal these *spatial* patterns by amplifying the motions only corresponding to a range of *temporal* frequencies. A single frame from each motion magnified video is shown in Figure 8, along with the theoretically expected shape.[21]

Spatially smoothing the motion signal often improves signal-to-noise ratios. Objects tend to move coherently in local image patches and any deviation from this is likely noise. Because the phase signal is more reliable when the amplitude of the complex steerable pyramid coefficients is higher, we perform an amplitude-weighted Gaussian blur:

$$\frac{((\Delta\phi)A) * K_\rho}{A * K_\rho} \quad (19)$$

where K_ρ is a Gaussian convolution kernel given by $\exp\left(-\frac{x^2+y^2}{2\rho^2}\right)$. The indices of A and ϕ have been suppressed for readability.

Figure 7. 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) amplifies changes linearly, while (c) uses the phase-based approach.

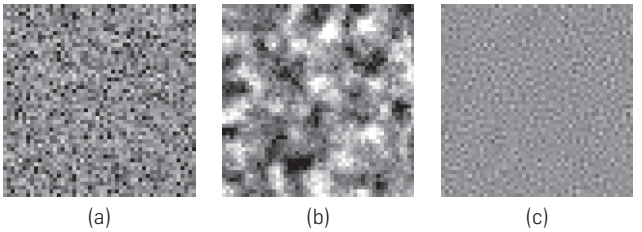
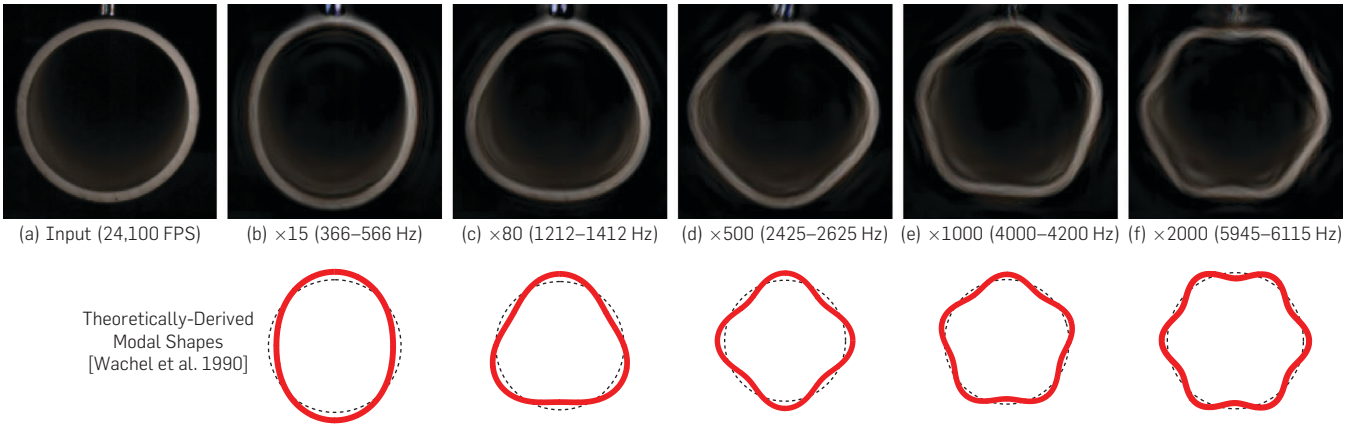


Figure 8. Isolating different types of *spatial* motions with *temporal* filtering. We took a high-speed video of a pipe being struck with a hammer. A frame from this video is shown in (a). The motions at several frequencies were magnified to isolate different modal shapes of the pipe. In (b)–(f), a frame from each of the motion magnified videos is shown. Below, the theoretically-derived modal shapes are shown in red overlayed, for comparison, over a perfect circle in dotted black.



We applied this processing to all of our motion magnification videos with ρ equal to 2 pixels in each pyramid level.

Because oversmoothing can shape even white noise into a plausible motion signal, it becomes important to verify whether the motions we are amplifying are indeed real. We have done many experiments comparing the visual motion signal with the signal recorded by accurate point-measurement devices, such as an acclerometer or laser vibrometer and the signals are always in agreement, validating that the motions are real.[2–4, 22] In addition, there are many videos where the motion is spatially coherent at a scale beyond that imposed by spatial smoothing (e.g., the pipes in Figure 8). This is unlikely to happen by chance and provides further evidence for the correctness of the amplified videos.

We can only recover motions that occur at frequencies less than the temporal Nyquist frequency of the camera. If the motions are too fast, only an aliased version of them gets amplified. In the special case that the motions occur at a single temporal frequency, aliasing can be useful. It makes such motions appear slower, which permits the visualization of fast vibrations in real-time, for example, the resonance of a wine glass.[23] However, in general we cannot recover a meaningful signal if the frames are temporally undersampled.

5. A BIG WORLD OF SMALL CHANGES

The world is full of subtle changes that are invisible to the naked eye. Video magnification allows us to reveal these changes by magnifying them. We present a selection of our magnification results and extensions of our techniques by us and other authors below.

As the heart beats, blood flows in and out of the face, changing its color slightly. This color change is very subtle, typically only half a gray-level. However, the human pulse occurs in a narrow band of temporal frequencies and is spatially smooth. For the man in Figure 1a, we can isolate signal from noise by temporally filtering the color variations in a passband of 50–60 beats per minute (0.83–1 Hz) and then spatially lowpassing them. Amplifying the result by 100× produces a color-amplified video, in which the human pulse is visible (Figure 1a). In addition to visualizing the pulse, we can plot the filtered color changes at a point on the man's forehead to get a quantitative measurement of pulse (Figure 3.8 in Ref.[17]). The beating of the human heart also produces subtle motions throughout the body. We were able to visualize the pulsing of the radial and ulnar arteries in a video of a wrist (Figure 7 in Ref.[24]). Amir-Khalili et al. and Mcleod et al. have also quantitatively analyzed subtle color and motion changes using methods inspired by the ones proposed here to identify faintly pulsing blood vessels in surgical videos,[1, 15] which may be clinically useful.

Our methods can also be used to reveal the invisible swaying of structures. We demonstrate this in a video of a crane taken by an ordinary DSLR camera (Figure 1b). The crane looks completely still in the input video. However, when the low-frequency (0.2–0.4 Hz) motions are magnified 75×, the swaying of the crane's mast and hook become apparent. In an extension to the work described here, Chen et al. quantitatively analyze structural motions in videos to non-destructively test their safety.[3] They do this by recovering modal shapes and frequencies of structures (similar to Figure 8) based on local phase changes in videos and use these as markers for structural damage.

Video magnification has also contributed to new scientific discoveries in biology. Sellon et al. magnified the subtle motions of an in-vitro mammalian tectorial membrane,[18] a thin structure in the inner ear. This helped explain this membrane's role in frequency selectivity during hearing.

6. THE VISUAL MICROPHONE

One interesting source of small motions is sound. When sound hits an object, it causes that object to vibrate. These vibrations are normally too subtle and too fast to be seen, but we can sometimes reveal them in motion magnified, high-speed videos of the object (Figure 9a). This shows that sound can produce a *visual* motion signal. Video magnification gives us a way to visualize this signal, but we can also

quantitatively analyze it to recover sound from silent videos of the objects (Figure 9b). For example, we can recover intelligible speech and music from high-speed videos of a vibrating potato chip bag or houseplant (Figures 1 and 10). We call this technique The Visual Microphone.

Figure 9. Revealing sound-induced vibrations using motion magnification and recovering audio using the visual microphone. A pure-tone version of "Mary Had a Little Lamb" is played at a chip bag and the motions corresponding to each note are magnified separately. Time slices of the resulting videos, in which the vertical dimension is space and the horizontal dimension is time, are used to produce a visual spectrogram (a), which closely matches the spectrogram of the recovered audio (b).

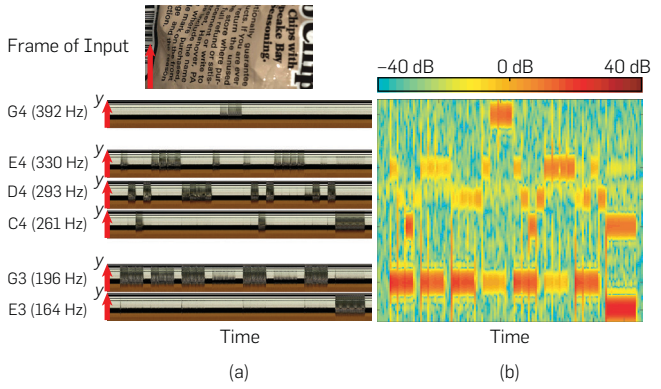
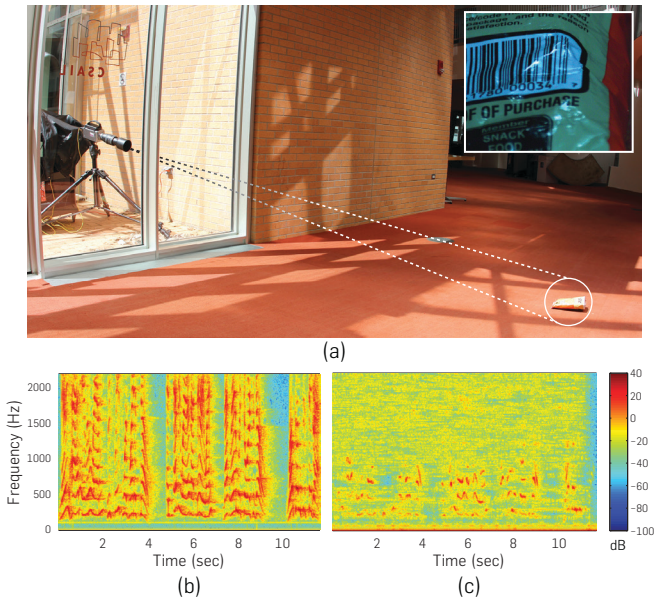


Figure 10. Speech recovered from a 4 kHz *silent* video of a bag of chips filmed through soundproof glass. The chip bag (on the bottom right in (a)) is lit by natural sunlight. The camera (on the left in (a)) is outside behind sound-proof glass. A frame from the recorded video (400 × 480 pixels) is shown in the inset. The speech "Mary had a little lamb...Welcome to SIGGRAPH!" was spoken by a person near the bag of chips. (b) The spectrogram of the source sound recorded by a standard microphone near the chip bag and (c) the spectrogram of our recovered sound. The recovered sound is noisy but comprehensible (audio clips are available on the visual microphone project webpage).



A challenge is that the vibrations are incredibly small, on the order of micrometers for 80 dB sound, and noise in the video can easily overwhelm this signal. Narrowband temporal filtering only works for narrowband sounds, not general ones that contain all frequencies. However, we want to recover a 1D audio signal, not magnify the spatial pattern of the motions. This means we can spatially combine information across the *entire* frame to attain intelligible SNR. This works because at most audible frequencies, the wavelength of the sound (10 cm–1.2 m for telephone-quality audio) is much larger than the visible portion of the object in a video. For lightweight objects, the motion caused by sound is largely coherent across each frame of the video.

Based on this observation, we seek to recover a global motion signal $R(t)$ of the object. We measure local motions by using local phase variations. We project each frame of the input video on to the complex steerable pyramid basis and then compute the time series of local phase variations $\Delta\phi_{r,\theta}(x, y, t)$ for all spatial scales r, orientations θ and positions x, y.

To place higher confidence in local motions that come from regions with localizable features, we weigh the local phase variations by the square of the amplitudes of the corresponding coefficients. We then take the average of these weighted local signals over the dimensions of the complex steerable pyramid:

$$R(t) = \sum_{r,\theta,x,y} A_{r,\theta}(x, y, t)^2 \Delta\phi_{r,\theta}(x, y, t). \qquad (20)$$

This signal measures the average motion of the object and is related to sound via an object-specific transfer function.[7] This signal can be played as intelligible sound and can be further improved with standard audio denoising techniques.

We conducted a variety of experiments to explore when, and how well, this method was able to recover sound. In each experiment, an object was filmed with a high-speed camera while being exposed to sound from either a nearby loudspeaker or a person's voice. Experiments were calibrated with a decibel meter and evaluated using perception-based metrics from the audio processing literature. We found that lightweight objects, which move readily with sound (e.g., a bag of chips, the leaves of a plant) yielded the best results. Heavier objects (e.g., bricks, an optical bench) produced much weaker results, suggesting that unintended motion, like camera shake, was not a significant factor. In our most ambitious experiment, we were able to recover human speech from a bag of chips 3–4 m away (Figure 10). More information about our experiments are in Ref.[7]

In one experiment, we played a pure-tone version (synthesized from MIDI) of "Mary had a Little Lamb" at a chip bag. Because the sound contained only pure-tones, we were able to motion magnify the chip bag in narrow temporal bands corresponding to the tones to produce six processed videos that together form a *visual spectrogram*. We show slices in time of the motion magnified videos (Figure 9a) and display them next to the recovered sound's spectrogram (Figure 9b).

Vibrations of an object can also be used to learn about its physical properties. In follow-up work, Davis and Bouman et al.[5] use the same method to analyze small vibrations in

objects and discern the material properties (stiffness, area weight, and elasticity) of fabrics from videos. Davis et al.[6] also used tiny motions to learn an image-space model of how an object vibrates and used that to perform simulations of how it would behave if stimulated.

7. LIMITATIONS

The Eulerian approach to motion magnification is robust and fast, but works primarily when the motions are small. If the motions are large, this processing can introduce artifacts. However, one can detect when this happens and suppress magnification in this case.[22] Elgharib et al.[8] also demonstrate it is possible to magnify tiny motions in the presence of large ones by first stabilizing the video. There are limits to how well spatio-temporal filtering can remove noise and amplified noise can cause image structures to move incoherently.

8. CONCLUSION

Eulerian video magnification is a set of simple and robust algorithms that can reveal and analyze tiny motions. It is a new type of microscope, not made of optics, but of software taking an ordinary video as input and producing one in which the temporal changes are larger. It reveals a new world of tiny motions and color changes showing us hidden vital signs, building movements and vibrations due to sound waves. Our visualization may have applications in a variety of fields such as healthcare, biology, mechanical engineering, and civil engineering.

Acknowledgments

We thank Quanta Computer, Shell research, QCRI, NSF CGV-1111415, the DARPA SCENICC program, the NDSEG fellowship, the Microsoft Research fellowship, and Cognex for funding this research. We also thank Dirk Smit, Steve Lewin-Berlin, Guha Balakrishnan, Ce Liu, and Deqing Sun for their helpful suggestions. We thank Michael Feng at Draper Laboratory for loaning us a laser vibrometer and Dr. Donna Brezinksi, Dr. Karen McAlmon and the Winchester Hospital staff for helping us to collect videos of newborns. C

References

1. Amir-Khalili, A., Peyrat, J.-M., Abinahed, J., Al-Alao, O., Al-Ansari, A., Hamarneh, G., Abugharbieh, R. Auto localization and segmentation of occluded vessels in robot-assisted partial nephrectomy. In *MICCAI 2014* (2014). Springer, 407–414.
2. Chen, J.G., Wadhwa, N., Cha, Y.-J., Durand, F., Freeman, W.T., Buyukozturk, O. Structural modal identification through high speed camera video: Motion magnification. In *Topics in Modal Analysis I*. Volume 7 (2014). Springer, 191–197.
3. Chen, J.G., Wadhwa, N., Cha, Y.-J., Durand, F., Freeman, W.T., Buyukozturk, O. Modal identification of simple structures with high-speed video using motion magnification. *Journal of Sound and Vibration 345* (2015), 58–71.
4. Chen, J.G., Wadhwa, N., Durand, F., Freeman, W.T., Buyukozturk, O. Developments with motion magnification for structural modal identification through camera video. In *Dynamics of Civil Structures*. Volume 2 (2015). Springer, 49–57.
5. Davis, A., Bouman, K.L., Chen, J.G., Rubinstein, M., Durand, F., Freeman, W.T. Visual vibrometry: Estimating material properties from small motions in video. In *CVPR 2015* (2015), 5335–5343.
6. Davis, A., Chen, J.G., Durand, F. Image-space modal bases for plausible manipulation of objects in video. *ACM Trans. Graph. 34*, 6 (2015), 239.
7. Davis, A., Rubinstein, M., Wadhwa, N., Mysore, G.J., Durand, F., Freeman, W.T. The visual microphone: Passive recovery of sound from video. *ACM Trans. Graph. 33*, 4 (2014), 79.
8. Elgharib, M.A., Hefeeda, M., Durand, F., Freeman, W.T. Video magnification in presence of large motions. In *CVPR 2015* (2015), 4119–4127.
9. Fleet, D.J., Jepson, A.D. Computation of component image velocity from local phase information. *Int. J. Comput. Vision* 5, 1 (1990), 77–104.
10. Fuchs, M., Chen, T., Wang, O., Raskar, R., Seidel, H.-P., Lensch, H.P. Real-time temporal shaping of high-speed video streams. *Comput. Graph. 34*, 5 (2010), 575–584.
11. Gautama, T., Van Hulle, M.M. A phase-based approach to the estimation of the optical flow field using spatial filtering. *IEEE Trans. Neural Netw. 13*, 5 (2002), 1127–1136.
12. Horn, B.K., Schunck, B.G. Determining optical flow. In *1981 Technical Symposium East* (1981). International Society for Optics and Photonics, 319–331.
13. Liu, C., Torralba, A., Freeman, W.T., Durand, F., Adelson, E.H. Motion magnification. *ACM Trans. Graph. 24*, 3 (2005), 519–526.
14. Lucas, B.D., Kanade, T., et al. An iterative image registration technique with an application to stereo vision. In *IJCAI*. Volume 81 (1981), 674–679.
15. McLeod, A.J., Baxter, J.S., de Ribaupierre, S., Peters, T.M. Motion magnification for endoscopic surgery. In *SPIE Medical Imaging* (2014), 90360C–90360C.
16. Poh, M.-Z., McDuff, D.J., Picard, R.W. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express 18*, 10 (2010), 10762–10774.
17. Rubinstein, M. Analysis and visualization of temporal variations in video. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA (2014).
18. Sellon, J.B., Farrahi, S., Ghaffari, R., Freeman, D.M. Longitudinal spread of mechanical excitation through tectorial membrane traveling waves. *Proc. Natl. Acad. Sci. 112*, 42 (2015), 12968–12973.
19. Simoncelli, E.P., Freeman, W.T. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *ICIP 1995* (1995), 3444.
20. Simoncelli, E.P., Freeman, W.T., Adelson, E.H., Heeger, D.J. Shiftable multi-scale transforms. *IEEE Trans. Info. Theory 2*, 38 (1992), 587–607.
21. Wachel, J., Morton, S.J., Atkins, K.E. Piping vibration analysis. In *Proceedings of the 19th Turbomachinery Symposium.* (1990).
22. Wadhwa, N., Rubinstein, M., Durand, F., Freeman, W.T. Phase-based video motion processing. *ACM Trans. Graph. 32*, 4 (2013), 80.
23. Wadhwa, N., Rubinstein, M., Durand, F., Freeman, W.T. Riesz pyramids for fast phase-based video magnification. In *ICCP 2014* (2014). IEEE, 1–10.
24. Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J. V., Durand, F., Freeman, W.T. Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph. 31*, 4 (2012), 65.

Neal Wadhwa, Abe Davis, John V. Guttag, William T. Freeman and Fredo Durand, MIT Computer Science and Artificial Intelligence Laboratory

Michael Rubinstein, Google Research

Eugene Shih, Cambridge Mobile Telematics

Gautham J. Mysore, Adobe Research

Justin G. Chen and Oral Buyukozturk, MIT Department of Civil and Environmental Engineering

Hao-Yu Wu, Amazon A9