

Eulerian Video Magnification for Revealing Subtle Changes in the World

Hao-Yu Wu¹

Michael Rubinstein¹

Eugene Shih²

John Guttag¹

Frédo Durand¹

William Freeman¹

¹MIT CSAIL ²Quanta Research Cambridge, Inc.

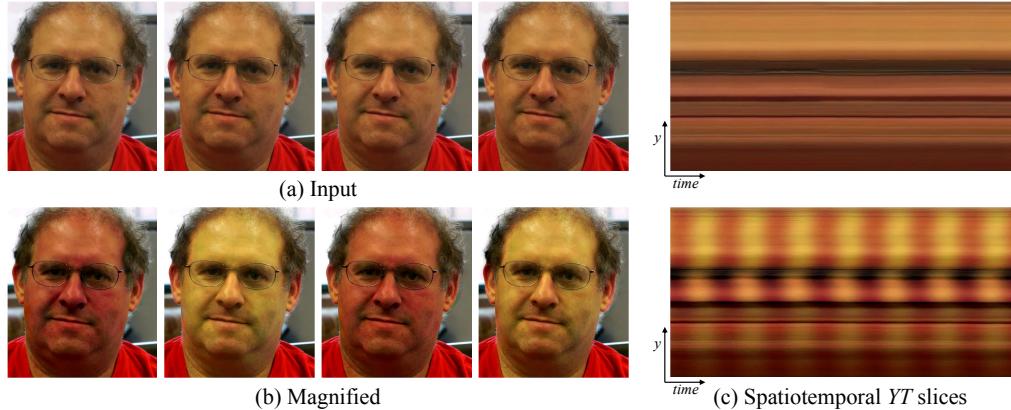


Figure 1: An example of using our Eulerian Video Magnification framework for visualizing the human pulse. (a) Four frames from the original video sequence (face). (b) The same four frames with the subject’s pulse signal amplified. (c) A vertical scan line from the input (top) and output (bottom) videos plotted over time shows how our method amplifies the periodic color variation. In the input sequence the signal is imperceptible, but in the magnified sequence the variation is clear. The complete sequence is available in the supplemental video.

Abstract

Our goal is to reveal temporal variations in videos that are difficult or impossible to see with the naked eye and display them in an indicative manner. Our method, which we call Eulerian Video Magnification, takes a standard video sequence as input, and applies spatial decomposition, followed by temporal filtering to the frames. The resulting signal is then amplified to reveal hidden information. Using our method, we are able to visualize the flow of blood as it fills the face and also to amplify and reveal small motions. Our technique can run in real time to show phenomena occurring at temporal frequencies selected by the user.

CR Categories: I.4.7 [Image Processing and Computer Vision]: Scene Analysis—Time-varying Imagery;

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

Links: DL PDF WEB

1 Introduction

The human visual system has limited spatio-temporal sensitivity, but many signals that fall below this capacity can be informative.

For example, human skin color varies slightly with blood circulation. This variation, while invisible to the naked eye, can be exploited to extract pulse rate [Verkruysse et al. 2008; Poh et al. 2010; Philips 2011]. Similarly, motion with low spatial amplitude, while hard or impossible for humans to see, can be magnified to reveal interesting mechanical behavior [Liu et al. 2005]. The success of these tools motivates the development of new techniques to reveal invisible signals in videos. In this paper, we show that a combination of spatial and temporal processing of videos can amplify subtle variations that reveal important aspects of the world around us.

Our basic approach is to consider the time series of color values at any spatial location (pixel) and amplify variation in a given temporal frequency band of interest. For example, in Figure 1 we automatically select, and then amplify, a band of temporal frequencies that includes plausible human heart rates. The amplification reveals the variation of redness as blood flows through the face. For this application, temporal filtering needs to be applied to lower spatial frequencies (spatial pooling) to allow such a subtle input signal to rise above the camera sensor and quantization noise.

Our temporal filtering approach not only amplifies color variation, but can also reveal low-amplitude motion. For example, in the supplemental video, we show that we can enhance the subtle motions around the chest of a breathing baby. We provide a mathematical analysis that explains how temporal filtering interplays with spatial motion in videos. Our analysis relies on a linear approximation related to the brightness constancy assumption used in optical flow formulations. We also derive the conditions under which this approximation holds. This leads to a multiscale approach to magnify motion without feature tracking or motion estimation.

Previous attempts have been made to unveil imperceptible motions in videos. [Liu et al. 2005] analyze and amplify subtle motions and visualize deformations that would otherwise be invisible. [Wang et al. 2006] propose using the Cartoon Animation Filter to create perceptually appealing motion exaggeration. These approaches follow a *Lagrangian* perspective, in reference to fluid dynamics where the trajectory of particles is tracked over time. As such, they rely

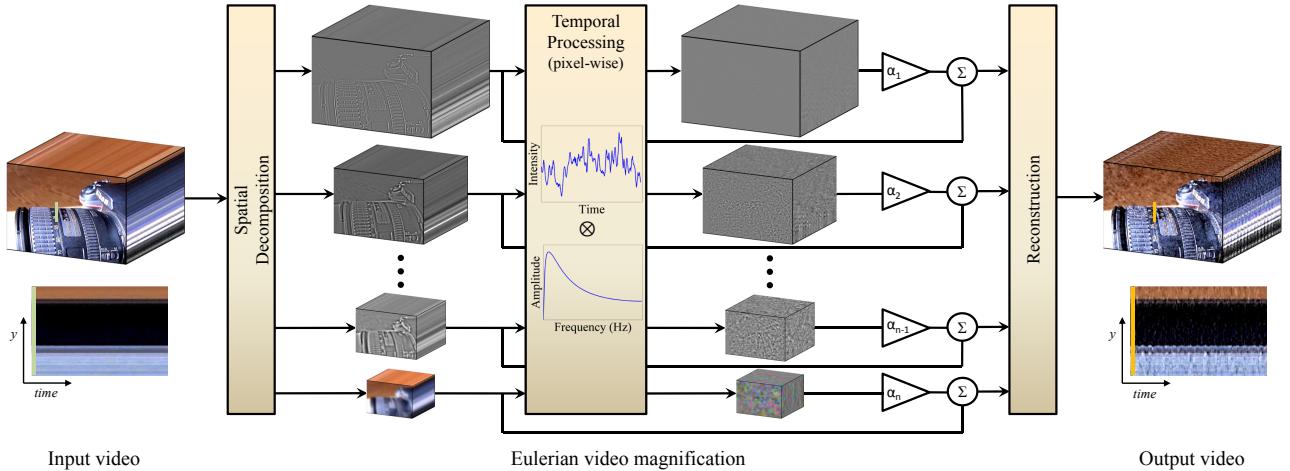


Figure 2: Overview of the Eulerian video magnification framework. The system first decomposes the input video sequence into different spatial frequency bands, and applies the same temporal filter to all bands. The filtered spatial bands are then amplified by a given factor α , added back to the original signal, and collapsed to generate the output video. The choice of temporal filter and amplification factors can be tuned to support different applications. For example, we use the system to reveal unseen motions of a Digital SLR camera, caused by the flipping mirror during a photo burst (camera; full sequences are available in the supplemental video).

on accurate motion estimation, which is computationally expensive and difficult to make artifact-free, especially at regions of occlusion boundaries and complicated motions. Moreover, Liu et al. [2005] have shown that additional techniques, including motion segmentation and image in-painting, are required to produce good quality synthesis. This increases the complexity of the algorithm further.

In contrast, we are inspired by the *Eulerian* perspective, where properties of a voxel of fluid, such as pressure and velocity, evolve over time. In our case, we study and amplify the variation of pixel values over time, in a spatially-multiscale manner. In our Eulerian approach to motion magnification, we do not explicitly estimate motion, but rather exaggerate motion by amplifying temporal color changes at fixed positions. We rely on the same differential approximations that form the basis of optical flow algorithms [Lucas and Kanade 1981; Horn and Schunck 1981].

Temporal processing has been used previously to extract invisible signals [Poh et al. 2010] and to smooth motions [Fuchs et al. 2010]. For example, Poh et al. [2010] extract a heart rate from a video of a face based on the temporal variation of the skin color, which is normally invisible to the human eye. They focus on extracting a single number, whereas we use localized spatial pooling and bandpass filtering to extract and reveal visually the signal corresponding to the pulse. This primal domain analysis allows us to amplify and visualize the pulse signal at each location on the face. This has important potential monitoring and diagnostic applications to medicine, where, for example, the asymmetry in facial blood flow can be a symptom of arterial problems.

Fuchs et al. [2010] use per-pixel temporal filters to dampen temporal aliasing of motion in videos. They also discuss the high-pass filtering of motion, but mostly for non-photorealistic effects and for large motions (Figure 11 in their paper). In contrast, our method strives to make imperceptible motions visible using a multiscale approach. We analyze our method theoretically and show that it applies only for small motions.

In this paper, we make several contributions. First, we demonstrate that nearly invisible changes in a dynamic environment can be revealed through *Eulerian* spatio-temporal processing of standard monocular video sequences. Moreover, for a range of amplification values that is suitable for various applications, explicit motion estimation is not required to amplify motion in natural videos. Our

approach is robust and runs in real time. Second, we provide an analysis of the link between temporal filtering and spatial motion and show that our method is best suited to small displacements and lower spatial frequencies. Third, we present a single framework that can be used to amplify both spatial motion and purely temporal changes, e.g., the heart pulse, and can be adjusted to amplify particular temporal frequencies—a feature which is not supported by Lagrangian methods. Finally, we analytically and empirically compare Eulerian and Lagrangian motion magnification approaches under different noisy conditions. To demonstrate our approach, we present several examples where our method makes subtle variations in a scene visible.

2 Space-time video processing

Our approach combines spatial and temporal processing to emphasize subtle temporal changes in a video. The process is illustrated in Figure 2. We first decompose the video sequence into different spatial frequency bands. These bands might be magnified differently because (a) they might exhibit different signal-to-noise ratios or (b) they might contain spatial frequencies for which the linear approximation used in our motion magnification does not hold (Sect. 3). In the latter case, we reduce the amplification for these bands to suppress artifacts. When the goal of spatial processing is simply to increase temporal signal-to-noise ratio by pooling multiple pixels, we spatially low-pass filter the frames of the video and downsample them for computational efficiency. In the general case, however, we compute a full Laplacian pyramid [Burt and Adelson 1983].

We then perform temporal processing on each spatial band. We consider the time series corresponding to the value of a pixel in a frequency band and apply a bandpass filter to extract the frequency bands of interest. For example, we might select frequencies within 0.4-4Hz, corresponding to 24-240 beats per minute, if we wish to magnify a pulse. If we are able to extract the pulse rate, we can use a narrow band around that value. The temporal processing is uniform for all spatial levels, and for all pixels within each level. We then multiply the extracted bandpassed signal by a magnification factor α . This factor can be specified by the user, and may be attenuated automatically according to guidelines in Sect. 3.2. Possible temporal filters are discussed in Sect. 4. Next, we add the magnified signal to the original and collapse the spatial pyramid to obtain

the final output. Since natural videos are spatially and temporally smooth, and since our filtering is performed uniformly over the pixels, our method implicitly maintains spatiotemporal coherency of the results.

3 Eulerian motion magnification

Our processing can amplify small motion even though we do not track motion as in Lagrangian methods [Liu et al. 2005; Wang et al. 2006]. In this section, we show how temporal processing produces motion magnification using an analysis that relies on the first-order Taylor series expansions common in optical flow analyses [Lucas and Kanade 1981; Horn and Schunck 1981].

3.1 First-order motion

To explain the relationship between temporal processing and motion magnification, we consider the simple case of a 1D signal undergoing translational motion. This analysis generalizes directly to locally-translational motion in 2D.

Let $I(x, t)$ denote the image intensity at position x and time t . Since the image undergoes translational motion, we can express the observed intensities with respect to a displacement function $\delta(t)$, such that $I(x, t) = f(x + \delta(t))$ and $I(x, 0) = f(x)$. The goal of motion magnification is to synthesize the signal

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

for some amplification factor α .

Assuming the image can be approximated by a first-order Taylor series expansion, we write the image at time t , $f(x + \delta(t))$ in a first-order Taylor expansion about x , as

$$I(x, t) \approx f(x) + \delta(t) \frac{\partial f(x)}{\partial x}. \quad (2)$$

Let $B(x, t)$ be the result of applying a broadband temporal bandpass filter to $I(x, t)$ at every position x (picking out everything except $f(x)$ in Eq. 2). For now, let us assume the motion signal, $\delta(t)$, is within the passband of the temporal bandpass filter (we will relax that assumption later). Then we have

$$B(x, t) = \delta(t) \frac{\partial f(x)}{\partial x}. \quad (3)$$

In our process, we then amplify that bandpass signal by α and add it back to $I(x, t)$, resulting in the processed signal

$$\tilde{I}(x, t) = I(x, t) + \alpha B(x, t). \quad (4)$$

Combining Eqs. 2, 3, and 4, we have

$$\tilde{I}(x, t) \approx f(x) + (1 + \alpha)\delta(t) \frac{\partial f(x)}{\partial x}. \quad (5)$$

Assuming the first-order Taylor expansion holds for the amplified larger perturbation, $(1 + \alpha)\delta(t)$, we can relate the amplification of the temporally bandpassed signal to motion magnification. The processed output is simply

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

This shows that the processing magnifies motions—the spatial displacement $\delta(t)$ of the local image $f(x)$ at time t , has been amplified to a magnitude of $(1 + \alpha)$.

This process is illustrated for a single sinusoid in Figure 3. For a low frequency cosine wave and a relatively small displacement,

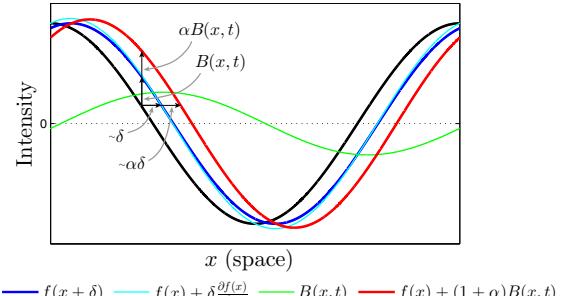


Figure 3: Temporal filtering can approximate spatial translation. This effect is demonstrated here on a 1D signal, but equally applies to 2D. The input signal is shown at two time instants: $\hat{I}(x, t) = f(x)$ at time t and $\hat{I}(x, t + 1) = f(x + \delta)$ at time $t + 1$. The first-order Taylor series expansion of $\hat{I}(x, t + 1)$ about x approximates well the translated signal. The temporal bandpass is amplified and added to the original signal to generate a larger translation. In this example $\alpha = 1$, magnifying the motion by 100%, and the temporal filter is a finite difference filter, subtracting the two curves.

$\delta(t)$, the first-order Taylor series expansion serves as a good approximation for the translated signal at time $t + 1$. When boosting the temporal signal by α and adding it back to $I(x, t)$, we approximate that wave translated by $(1 + \alpha)\delta$.

For completeness, let us return to the more general case where $\delta(t)$ is not entirely within the passband of the temporal filter. In this case, let $\delta_k(t)$, indexed by k , represent the different temporal spectral components of $\delta(t)$. Each $\delta_k(t)$ will be attenuated by the temporal filtering by a factor γ_k . This results in a bandpassed signal,

$$B(x, t) = \sum_k \gamma_k \delta_k(t) \frac{\partial f(x)}{\partial x} \quad (7)$$

(compare with Eq. 3). Because of the multiplication in Eq. 4, this temporal frequency dependent attenuation can equivalently be interpreted as a frequency-dependent motion magnification factor, $\alpha_k = \gamma_k \alpha$, resulting in a motion magnified output,

$$\tilde{I}(x, t) \approx f(x + \sum_k (1 + \alpha_k) \delta_k(t)) \quad (8)$$

The result is as would be expected for a linear analysis: the modulation of the spectral components of the motion signal becomes the modulation factor in the motion amplification factor, α_k , for each temporal subband, δ_k , of the motion signal.

3.2 Bounds

In practice, the assumptions in Sect. 3.1 hold for smooth images and small motions. For quickly changing image functions (i.e., high spatial frequencies), $f(x)$, the first-order Taylor series approximations becomes inaccurate for large values of the perturbation, $1 + \alpha\delta(t)$, which increases both with larger magnification α and motion $\delta(t)$. Figures 4 and 5 demonstrate the effect of higher frequencies, larger amplification factors and larger motions on the motion-amplified signal of a sinusoid.

As a function of spatial frequency, ω , we can derive a guide for how large the motion amplification factor, α , can be, given the observed motion $\delta(t)$. For the processed signal, $\tilde{I}(x, t)$ to be approximately equal to the true magnified motion, $\hat{I}(x, t)$, we seek the conditions under which

$$\begin{aligned} \tilde{I}(x, t) &\approx \hat{I}(x, t) \\ \Rightarrow f(x) + (1 + \alpha)\delta(t) \frac{\partial f(x)}{\partial x} &\approx f(x + (1 + \alpha)\delta(t)) \end{aligned} \quad (9)$$

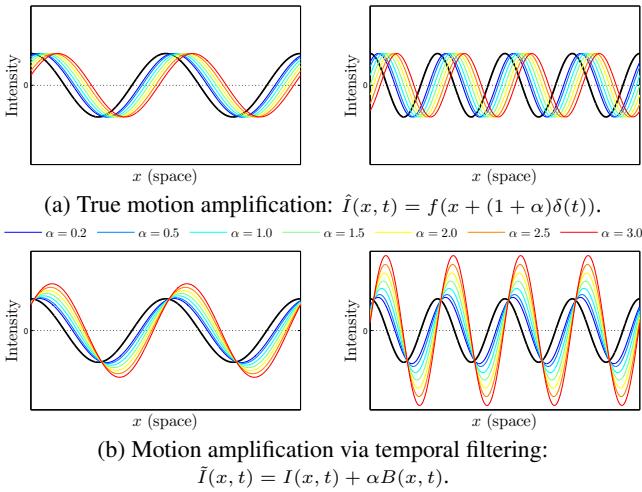


Figure 4: Illustration of motion amplification on a 1D signal for different spatial frequencies and α values. For the images on the left side, $\lambda = 2\pi$ and $\delta(1) = \frac{\pi}{8}$ is the true translation. For the images on the right side, $\lambda = \pi$ and $\delta(1) = \frac{\pi}{8}$. (a) The true displacement of $I(x, 0)$ by $(1 + \alpha)\delta(t)$ at time $t = 1$, colored from blue (small amplification factor) to red (high amplification factor). (b) The amplified displacement produced by our filter, with colors corresponding to the correctly shifted signals in (a). Referencing Eq. 14, the red (far right) curves of each plot correspond to $(1 + \alpha)\delta(t) = \frac{\lambda}{4}$ for the left plot, and $(1 + \alpha)\delta(t) = \frac{\lambda}{2}$ for the right plot, showing the mild, then severe, artifacts introduced in the motion magnification from exceeding the bound on $(1 + \alpha)$ by factors of 2 and 4, respectively.

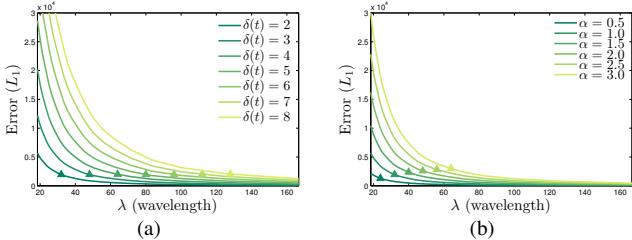


Figure 5: Motion magnification error, computed as the L_1 -norm between the true motion-amplified signal (Figure 4(a)) and the temporally-filtered result (Figure 4(b)), as function of wavelength, for different values of $\delta(t)$ (a) and α (b). In (a), we fix $\alpha = 1$, and in (b), $\delta(t) = 2$. The markers on each curve represent the derived cutoff point $(1 + \alpha)\delta(t) = \frac{\lambda}{8}$ (Eq. 14).

Let $f(x) = \cos(\omega x)$ for spatial frequency ω , and denote $\beta = 1 + \alpha$. We require that

$$\cos(\omega x) - \beta\omega\delta(t)\sin(\omega x) \approx \cos(\omega x + \beta\omega\delta(t)) \quad (10)$$

Using the addition law for cosines, we have

$$\begin{aligned} \cos(\omega x) - \beta\omega\delta(t)\sin(\omega x) &= \\ \cos(\omega x)\cos(\beta\omega\delta(t)) - \sin(\omega x)\sin(\beta\omega\delta(t)) & \end{aligned} \quad (11)$$

Hence, the following should approximately hold

$$\cos(\beta\omega\delta(t)) \approx 1 \quad (12)$$

$$\sin(\beta\omega\delta(t)) \approx \beta\delta(t)\omega \quad (13)$$

The small angle approximations of Eqs. (12) and (13) will hold to within 10% for $\beta\omega\delta(t) \leq \frac{\pi}{4}$ (the sine term is the leading ap-

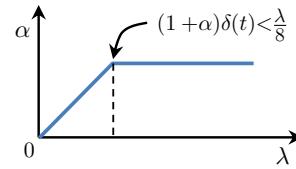


Figure 6: Amplification factor, α , as function of spatial wavelength λ , for amplifying motion. The amplification factor is fixed to α for spatial bands that are within our derived bound (Eq. 14), and is attenuated linearly for higher spatial frequencies.

proximation and we have $\sin(\frac{\pi}{4}) = 0.9\frac{\pi}{4}$. In terms of the spatial wavelength, $\lambda = \frac{2\pi}{\omega}$, of the moving signal, this gives

$$(1 + \alpha)\delta(t) < \frac{\lambda}{8}. \quad (14)$$

Eq. 14 above provides the guideline we seek, giving the largest motion amplification factor, α , compatible with accurate motion magnification of a given video motion $\delta(t)$ and image structure spatial wavelength, λ . Figure 4 (b) shows the motion magnification errors for a sinusoid when we boost α beyond the limit in Eq. 14. In some videos, violating the approximation limit can be perceptually preferred and we leave the λ cutoff as a user-modifiable parameter in the multiscale processing.

3.3 Multiscale analysis

The analysis in Sect. 3.2 suggests a *scale-varying* process: use a specified α magnification factor over some desired band of spatial frequencies, then scale back for the high spatial frequencies (found from Eq. 14 or specified by the user) where amplification would give undesirable artifacts. Figure 6 shows such a modulation scheme for α . Although areas of high spatial frequencies (sharp edges) will be generally amplified less than lower frequencies, we found the resulting videos to contain perceptually appealing magnified motion. Such effect was also exploited in the earlier work of Freeman et al. [1991] to create the illusion of motion from still images.

4 Results

The results were generated using non-optimized MATLAB code on a machine with a six-core processor and 32 GB RAM. The computation time per video was on the order of a few minutes. We used a separable binomial filter of size five to construct the video pyramids. We also built a prototype application that allows users to reveal subtle changes in real-time from live video feeds, essentially serving as a microscope for temporal variations. It is implemented in C++, is entirely CPU-based, and processes 640×480 videos at 45 frames per second on a standard laptop. It can be sped up further by utilizing GPUs. A demo of the application is available in the accompanying video. The code is available on the project webpage.

To process an input video by Eulerian video magnification, there are four steps a user needs to take: (1) select a temporal bandpass filter; (2) select an amplification factor, α ; (3) select a spatial frequency cutoff (specified by spatial wavelength, λ_c) beyond which an attenuated version of α is used; and (4) select the form of the attenuation for α —either force α to zero for all $\lambda < \lambda_c$, or linearly scale α down to zero. The frequency band of interest can be chosen automatically in some cases, but it is often important for users to be able to control the frequency band corresponding to their application. In our real-time application, the amplification factor and cutoff frequencies are all customizable by the user.

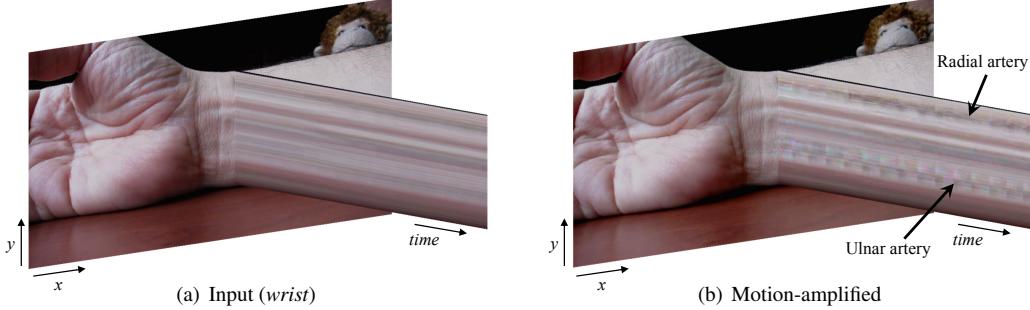


Figure 7: Eulerian video magnification used to amplify subtle motions of blood vessels arising from blood flow. For this video, we tuned the temporal filter to a frequency band that includes the heart rate—0.88 Hz (53 bpm)—and set the amplification factor to $\alpha = 10$. To reduce motion magnification of irrelevant objects, we applied a user-given mask to amplify the area near the wrist only. Movement of the radial and ulnar arteries can barely be seen in the input video (a) taken with a standard point-and-shoot camera, but is significantly more noticeable in the motion-magnified output (b). The motion of the pulsing arteries is more visible when observing a spatio-temporal YT slice of the wrist (a) and (b). The full wrist sequence can be found in the supplemental video.

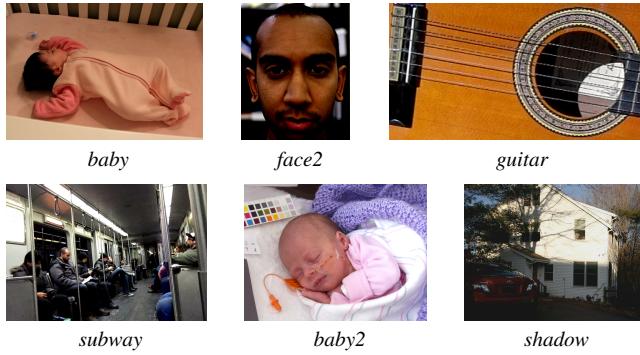


Figure 8: Representative frames from additional videos demonstrating our technique. The videos can be found in the accompanying video and on the project webpage.

We first select the temporal bandpass filter to pull out the motions or signals that we wish to be amplified (step 1 above). The choice of filter is generally application dependent. For motion magnification, a filter with a broad passband is preferred; for color amplification of blood flow, a narrow passband produces a more noise-free result. Figure 9 shows the frequency responses of some of the temporal filters used in this paper. We use ideal bandpass filters for color amplification, since they have passbands with sharp cutoff frequencies. Low-order IIR filters can be useful for both color amplification and motion magnification and are convenient for a real-time implementation. In general, we used two first-order lowpass IIR filters with cutoff frequencies ω_l and ω_h to construct an IIR bandpass filter.

Next, we select the desired magnification value, α , and spatial frequency cutoff, λ_c (steps 2 and 3). While Eq. 14 can be used as a guide, in practice, we may try various α and λ_c values to achieve a desired result. Users can select a higher α that violates the bound to exaggerate specific motions or color changes at the cost of increasing noise or introducing more artifacts. In some cases, one can account for color clipping artifacts by attenuating the chrominance components of each frame. Our approach achieves this by doing all the processing in the YIQ space. Users can attenuate the chrominance components, I and Q, before conversion to the original color space.

For human pulse color amplification, where we seek to emphasize low spatial frequency changes, we may force $\alpha = 0$ for spatial wavelengths below λ_c . For motion magnification videos, we can choose to use a linear ramp transition for α (step 4).

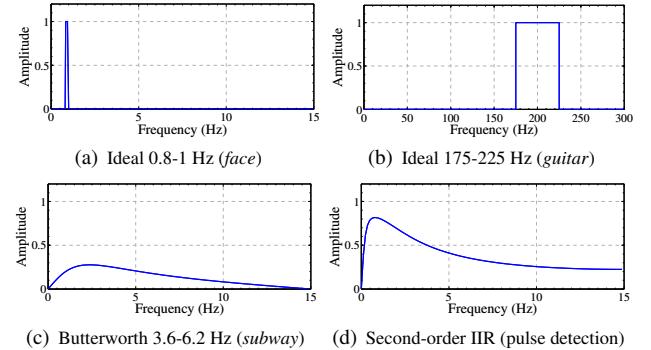


Figure 9: Temporal filters used in the paper. The ideal filters (a) and (b) are implemented using DCT. The Butterworth filter (c) is used to convert a user-specified frequency band to a second-order IIR structure and is used in our real-time application. The second-order IIR filter (d) also allows user input. These second-order filters have a broader passband than an ideal filter.

We evaluated our method for color amplification using a few videos: two videos of adults with different skin colors and one of a newborn baby. An adult subject with lighter complexion is shown in *face* (Figure 1), while an individual with darker complexion is shown in *face2* (Figure 8). In both videos, our objective was to amplify the color change as the blood flows through the face. In both *face* and *face2*, we applied a Laplacian pyramid and set α for the finest two levels to 0. Essentially, we downsampled and applied a spatial lowpass filter to each frame to reduce both quantization and noise and to boost the subtle pulse signal that we are interested in. For each video, we then passed each sequence of frames through an ideal bandpass filter with a passband of 0.83 Hz to 1 Hz (50 bpm to 60 bpm). Finally, a large value of $\alpha \approx 100$ and $\lambda_c \approx 1000$ was applied to the resulting spatially lowpass signal to emphasize the color change as much as possible. The final video was formed by adding this signal back to the original. We see periodic green to red variations at the heart rate and how blood perfuses the face.

baby2 is a video of a newborn recorded *in situ* at the Nursery Department at Winchester Hospital in Massachusetts. In addition to the video, we obtained ground truth vital signs from a hospital-grade monitor. We used this information to confirm the accuracy of our heart rate estimate and to verify that the color amplification signal extracted from our method matches the photoplethysmogram, an optically obtained measurement of the perfusion of blood to the skin, as measured by the monitor.

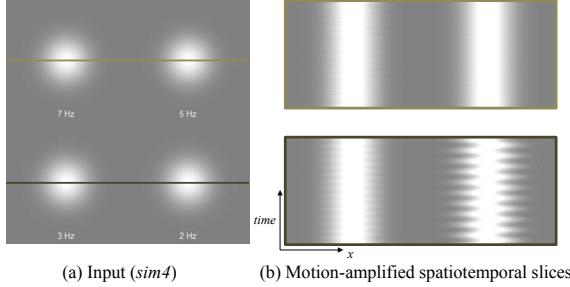


Figure 10: Selective motion amplification on a synthetic sequence (*sim4* on left). The video sequence contains blobs oscillating at different temporal frequencies as shown on the input frame. We apply our method using an ideal temporal bandpass filter of 1-3 Hz to amplify only the motions occurring within the specified passband. In (b), we show the spatio-temporal slices from the resulting video which show the different temporal frequencies and the amplified motion of the blob oscillating at 2 Hz. We note that the space-time processing is applied uniformly to all the pixels. The full sequence and result can be found in the supplemental video.

To evaluate our method for motion magnification, we used several different videos: *face* (Figure 1), *sim4* (Figure 10), *wrist* (Figure 7), *camera* (Figure 2), *face2*, *guitar*, *baby*, *subway*, *shadow*, and *baby2* (Figure 8). For all videos, we used a standard Laplacian pyramid for spatial filtering. For videos where we wanted to emphasize motions at specific temporal frequencies (e.g., in *sim4* and *guitar*), we used ideal bandpass filters. In *sim4* and *guitar*, we were able to selectively amplify the motion of a specific blob or guitar string by using a bandpass filter tuned to the oscillation frequency of the object of interest. These effects can be observed in the supplemental video. The values used for α and λ_c for all of the videos discussed in this paper are shown in Table 1.

For videos where we were interested in revealing broad, but subtle motion, we used temporal filters with a broader passband. For example, for the *face2* video, we used a second-order IIR filter with slow roll-off regions. By changing the temporal filter, we were able to magnify the motion of the head rather than amplify the change in the skin color. Accordingly, $\alpha = 20$, $\lambda_c = 80$ were chosen to magnify the motion.

By using broadband temporal filters and setting α and λ_c according to Eq. 14, our method is able to reveal subtle motions, as in the *camera* and *wrist* videos. For the *camera* video, we used a camera with a sampling rate of 300 Hz to record a Digital SLR camera vibrating while capturing photos at about one exposure per second. The vibration caused by the moving mirror in the SLR, though invisible to the naked eye, was revealed by our approach. To verify that we indeed amplified the vibrations caused by the flipping mirror, we secured a laser pointer to the camera and recorded a video of the laser light, appearing at a distance of about four meters from the source. At that distance, the laser light visibly oscillated with each exposure, with the oscillations in sync with the magnified motions.

Our method is also able to exaggerate visible, yet subtle motion, as seen in the *baby*, *face2*, and *subway* videos. In the subway example we deliberately amplified the motion beyond the derived bounds of where the first-order approximation holds in order to increase the effect and to demonstrate the algorithm’s artifacts. We note that most of the examples in our paper contain oscillatory movements because such motion generally has longer duration and smaller amplitudes. However, our method can be used to amplify non-periodic motions as well, as long as they are within the passband of the temporal bandpass filter. In *shadow*, for example, we process a video of the sun’s shadow moving linearly yet imperceptibly over 15 seconds. The magnified version makes it possible to see the change

Table 1: Table of α , λ_c , ω_l , ω_h values used to produce the various output videos. For *face2*, two different sets of parameters are used—one for amplifying pulse, another for amplifying motion. For *guitar*, different cutoff frequencies and values for (α, λ_c) are used to “select” the different oscillating guitar strings. f_s is the frame rate of the camera.

| Video | α | λ_c | ω_l (Hz) | ω_h (Hz) | f_s (Hz) |
|---------------------|----------|-------------|-----------------|-----------------|------------|
| <i>baby</i> | 10 | 16 | 0.4 | 3 | 30 |
| <i>baby2</i> | 150 | 600 | 2.33 | 2.67 | 30 |
| <i>camera</i> | 120 | 20 | 45 | 100 | 300 |
| <i>face</i> | 100 | 1000 | 0.83 | 1 | 30 |
| <i>face2 motion</i> | 20 | 80 | 0.83 | 1 | 30 |
| <i>face2 pulse</i> | 120 | 960 | 0.83 | 1 | 30 |
| <i>guitar Low E</i> | 50 | 40 | 72 | 92 | 600 |
| <i>guitar A</i> | 100 | 40 | 100 | 120 | 600 |
| <i>shadow</i> | 5 | 48 | 0.5 | 10 | 30 |
| <i>subway</i> | 60 | 90 | 3.6 | 6.2 | 30 |
| <i>wrist</i> | 10 | 80 | 0.4 | 3 | 30 |

even within this short time period.

Finally, some videos may contain regions of temporal signals that do not need amplification, or that, when amplified, are perceptually unappealing. Due to our Eulerian processing, we can easily allow the user to manually restrict magnification to particular areas by marking them on the video (this was used for *face* and *wrist*).

5 Discussion

Sensitivity to Noise. The amplitude variation of the signal of interest is often much smaller than the noise inherent in the video. In such cases direct enhancement of the pixel values will not reveal the desired signal. Spatial filtering can be used to enhance these subtle signals. However, if the spatial filter applied is not large enough, the signal of interest will not be revealed (Figure 11).

Assuming that the noise is zero-mean white and wide-sense stationary with respect to space, it can be shown that spatial low pass filtering reduces the variance of the noise according to the area of the low pass filter. In order to boost the power of a specific signal, e.g., the pulse signal in the face, we can use the spatial characteristics of the signal to estimate the spatial filter size.

Let the noise power level be σ^2 , and our prior on signal power over spatial frequencies be $S(\lambda)$. We want to find a spatial low pass filter with radius r such that the signal power is greater than the noise in the filtered frequency region. The wavelength cut off of such a filter is proportional to its radius, r , so the signal prior can be represented as $S(r)$. The noise power σ^2 can be estimated by examining pixel values in a stable region of the scene, from a gray card, or by using a technique as in [Liu et al. 2006]. Since the filtered noise power level, σ'^2 , is inversely proportional to r^2 , we can solve the following equation for r ,

$$S(r) = \sigma'^2 = k \frac{\sigma^2}{r^2} \quad (15)$$

where k is a constant that depends on the shape of the low pass filter. This equation gives an estimate for the size of the spatial filter needed to reveal the signal at a certain noise power level.

Eulerian vs. Lagrangian Processing. Because the two methods take different approaches to motion—Lagrangian approaches explicitly track motions, while our Eulerian approach does not—they can be used for complementary motion domains. Lagrangian approaches, e.g. [Liu et al. 2005], work better to enhance motions of fine point features and support larger amplification factors, while

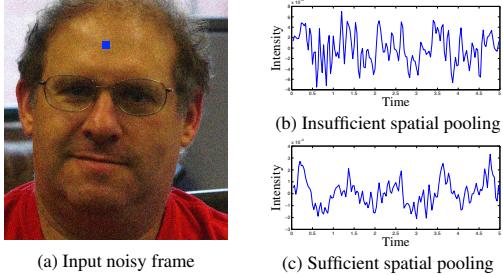


Figure 11: Proper spatial pooling is imperative for revealing the signal of interest. (a) A frame from the face video (Figure 1) with white Gaussian noise ($\sigma = 0.1$ pixel) added. On the right are intensity traces over time for the pixel marked blue on the input frame, where (b) shows the trace obtained when the (noisy) sequence is processed with the same spatial filter used to process the original face sequence, a separable binomial filter of size 20, and (c) shows the trace when using a filter tuned according to the estimated radius in Eq. 15, a binomial filter of size 80. The pulse signal is not visible in (b), as the noise level is higher than the power of the signal, while in (c) the pulse is clearly visible (the periodic peaks about one second apart in the trace).

our Eulerian method is better suited to smoother structures and small amplifications. We note that our technique does not assume particular types of motions. The first-order Taylor series analysis can hold for general small 2D motions along general paths.

In Appendix A, we further derive estimates of the accuracy of the two approaches with respect to noise. Comparing the Lagrangian error, ε_L (Eq. 29), and the Eulerian error, ε_E (Eq. 31), we see that both methods are equally sensitive to the temporal characteristics of the noise, n_t , while the Lagrangian process has additional error terms proportional to the spatial characteristics of the noise, n_x , due to the explicit estimation of motion (Eq. 27). The Eulerian error, on the other hand, grows quadratically with α , and is more sensitive to high spatial frequencies (I_{xx}). In general, this means that Eulerian magnification would be preferable over Lagrangian magnification for small amplifications and larger noise levels.

We validated this analysis on a synthetic sequence of a 2D cosine oscillating at 2 Hz temporally and 0.1 pixels spatially with additive white spatiotemporal Gaussian noise of zero mean and standard deviation σ (Figure 12). The results match the error-to-noise and error-to-amplification relationships predicted by the derivation (Figure 12(b)). The region where the Eulerian approach outperforms the Lagrangian results (Figure 12(a)-left) is also as expected. The Lagrangian method is more sensitive to increases in spatial noise, while the Eulerian error is hardly affected by it (Figure 12(c)). While different regularization schemes used for motion estimation (that are harder to analyze theoretically) may alleviate the Lagrangian error, they did not change the result significantly (Figure 12(a)-right). In general, our experiments show that for small amplifications the Eulerian approach strikes a better balance between performance and efficiency. Comparisons between the methods on natural videos are available on the project webpage.

6 Conclusion

We described a straightforward method that takes a video as input and exaggerates subtle color changes and imperceptible motions. To amplify motion, our method does not perform feature tracking or optical flow computation, but merely magnifies temporal color changes using spatio-temporal processing. This Eulerian-based method, which temporally processes pixels in a fixed spatial region, successfully reveals informative signals and amplifies small motions in real-world videos.

Acknowledgements. We would like to thank Guha Balakrishnan, Steve Lewin-Berlin and Neal Wadhwa for their helpful feedback, and the SIGGRAPH reviewers for their comments. We thank Ce Liu and Deqing Sun for helpful discussions on the Eulerian vs. Lagrangian analysis. We also thank Dr. Donna Brezinski, Dr. Karen McAlmon, and the Winchester Hospital staff for helping us collect videos of newborn babies. This work was partially supported by DARPA SCENICC program, NSF CGV-1111415, and Quanta Computer. Michael Rubinstein was partially supported by an NVIDIA Graduate Fellowship.

A Eulerian and Lagrangian Error

We derive estimates of the error in the Eulerian and Lagrangian motion magnification with respect to spatial and temporal noise. The derivation is done again in 1D for simplicity, and can be generalized to 2D. We use the same setup as in Sect. 3.1.

Both methods approximate the true motion-amplified sequence, $\hat{I}(x, t)$, as shown in (1). Let us first analyze the error in those approximations on the clean signal, $I(x, t)$.

Without noise. In the Lagrangian approach, the motion-amplified sequence, $\tilde{I}_L(x, t)$, is achieved by directly amplifying the estimated motion, $\tilde{\delta}(t)$, with respect to the reference frame, $I(x, 0)$

$$\tilde{I}_L(x, t) = I(x + (1 + \alpha)\tilde{\delta}(t), 0). \quad (16)$$

In its simplest form, we can estimate $\delta(t)$ in a point-wise manner (See Sect. 5 for discussion on spatial regularization)

$$\tilde{\delta}(t) = \frac{I_t(x, t)}{I_x(x, t)} \quad (17)$$

where $I_x(x, t) = \partial I(x, t)/\partial x$ and $I_t(x, t) = I(x, t) - I(x, 0)$. From now on, we will omit the space (x) and time (t) indices when possible for brevity.

The error in the Lagrangian solution is directly determined by the error in the estimated motion, which we take to be second-order term in the brightness constancy equation (although it is usually not paid in optical flow formulations because of Newton iterations),

$$\begin{aligned} I(x, t) &\approx I(x, 0) + \delta(t)I_x + \frac{1}{2}\delta^2(t)I_{xx} \\ \Rightarrow \frac{I_t}{I_x} &\approx \delta(t) + \frac{1}{2}\delta^2(t)I_{xx}. \end{aligned} \quad (18)$$

The estimated motion, $\tilde{\delta}(t)$, is related to the true motion, $\delta(t)$, by

$$\tilde{\delta}(t) \approx \delta(t) + \frac{1}{2}\delta^2(t)I_{xx}. \quad (19)$$

Plugging (19) in (16) and using a Taylor expansion of I about $x + (1 + \alpha)\delta(t)$, we have

$$\tilde{I}_L(x, t) \approx I(x + (1 + \alpha)\delta(t), 0) + \frac{1}{2}(1 + \alpha)\delta^2(t)I_{xx}I_x. \quad (20)$$

Subtracting (1) from (20), the error in the Lagrangian motion-magnified sequence, ε_L , is

$$\varepsilon_L \approx \left| \frac{1}{2}(1 + \alpha)\delta^2(t)I_{xx}I_x \right|. \quad (21)$$

In our Eulerian approach, the magnified sequence, $\hat{I}_E(x, t)$, is

$$\begin{aligned} \hat{I}_E(x, t) &= I(x, t) + \alpha I_t(x, t) \\ &= I(x, 0) + (1 + \alpha)I_t(x, t) \end{aligned} \quad (22)$$

similar to (4), using a two-tap temporal filter to compute I_t . Using a Taylor expansion of the true motion-magnified sequence, \hat{I} defined

in (1), about x , we have

$$\hat{I}(x, t) \approx I(x, 0) + (1 + \alpha)\delta(t)I_x + \frac{1}{2}(1 + \alpha)^2\delta^2(t)I_{xx}. \quad (23)$$

Using (18) and subtracting (1) from (23), the error in the Eulerian motion-magnified sequence, ε_E , is

$$\varepsilon_E \approx \left| \frac{1}{2}(1 + \alpha)^2\delta^2(t)I_{xx} - \frac{1}{2}(1 + \alpha)\delta^2(t)I_{xx}I_x \right|. \quad (24)$$

With noise. Let $I'(x, t)$ be the noisy signal, such that

$$I'(x, t) = I(x, t) + n(x, t) \quad (25)$$

for additive noise $n(x, t)$.

The estimated motion in the Lagrangian approach becomes

$$\tilde{\delta}(t) = \frac{I'_t}{I'_x} = \frac{I_t + n_t}{I_x + n_x} \quad (26)$$

where $n_x = \partial n / \partial x$ and $n_t = n(x, t) - n(x, 0)$.

Using a Taylor Expansion on (n_t, n_x) about $(0, 0)$ (zero noise), and using (18), we have

$$\tilde{\delta}(t) \approx \delta(t) + \frac{n_t}{I_x} - n_x \frac{I_t}{I_x^2} + \frac{1}{2}\delta^2(t)I_{xx}. \quad (27)$$

Plugging (27) into (16), and using a Taylor expansion of I about $x + (1 + \alpha)\delta(t)$, we get

$$\begin{aligned} \tilde{I}'_L(x, t) &\approx I(x + (1 + \alpha)\delta(t), 0) + \\ &(1 + \alpha)I_x \left(\frac{n_t}{I_x} - n_x \frac{I_t}{I_x^2} + \frac{1}{2}\delta^2(t)I_{xx} \right) + n. \end{aligned} \quad (28)$$

Using (19) again and subtracting (1), the Lagrangian error as a function of noise, $\varepsilon_L(n)$, is

$$\begin{aligned} \varepsilon_L(n) &\approx \left| (1 + \alpha)n_t - (1 + \alpha)n_x\delta(t) \right. \\ &\left. - \frac{1}{2}(1 + \alpha)\delta^2(t)I_{xx}n_x + \frac{1}{2}(1 + \alpha)\delta^2(t)I_{xx}I_x + n \right|. \end{aligned} \quad (29)$$

In the Eulerian approach, the noisy motion-magnified sequence becomes

$$\begin{aligned} \tilde{I}'_E(x, t) &= I'(x, 0) + (1 + \alpha)I'_t \\ &= I(x, 0) + (1 + \alpha)(I_t + n_t) + n. \end{aligned} \quad (30)$$

Using (24) and subtracting (1), the Eulerian error as a function of noise, $\varepsilon_E(n)$, is

$$\begin{aligned} \varepsilon_E(n) &\approx \left| (1 + \alpha)n_t + \frac{1}{2}(1 + \alpha)^2\delta^2(t)I_{xx} \right. \\ &\left. - \frac{1}{2}(1 + \alpha)\delta^2(t)I_{xx}I_x + n \right|. \end{aligned} \quad (31)$$

If we set the noise to zero in (29) and (31), the resulting errors correspond to those derived for the non-noisy signal as shown in (21) and (24).

References

- BURT, P., AND ADELSON, E. 1983. The laplacian pyramid as a compact image code. *IEEE Trans. Comm.* 31, 4, 532–540.
- FREEMAN, W. T., ADELSON, E. H., AND HEEGER, D. J. 1991. Motion without movement. *ACM Comp. Graph.* 25, 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.
- HORN, B., AND SCHUNCK, B. 1981. Determining optical flow. *Artificial intelligence* 17, 1-3, 185–203.
- LIU, C., TORRALBA, A., FREEMAN, W. T., DURAND, F., AND ADELSON, E. H. 2005. Motion magnification. *ACM Trans. Graph.* 24, 519–526.
- LIU, C., FREEMAN, W., SZELISKI, R., AND KANG, S. B. 2006. Noise estimation from a single image. In *IEEE CVPR*, vol. 1, 901 – 908.
- LUCAS, B. D., AND KANADE, T. 1981. An iterative image registration technique with an application to stereo vision. In *Proceedings of IJCAI*, 674–679.
- PHILIPS, 2011. Philips Vitals Signs Camera. <http://www.vitalsignscamera.com>.
- POH, M.-Z., McDUFF, D. J., AND PICARD, R. W. 2010. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express* 18, 10, 10762–10774.
- VERKRUYSSE, W., SVAASAND, L. O., AND NELSON, J. S. 2008. Remote plethysmographic imaging using ambient light. *Opt. Express* 16, 26, 21434–21445.
- WANG, J., DRUCKER, S. M., AGRAWALA, M., AND COHEN, M. F. 2006. The cartoon animation filter. *ACM Trans. Graph.* 25, 1169–1173.

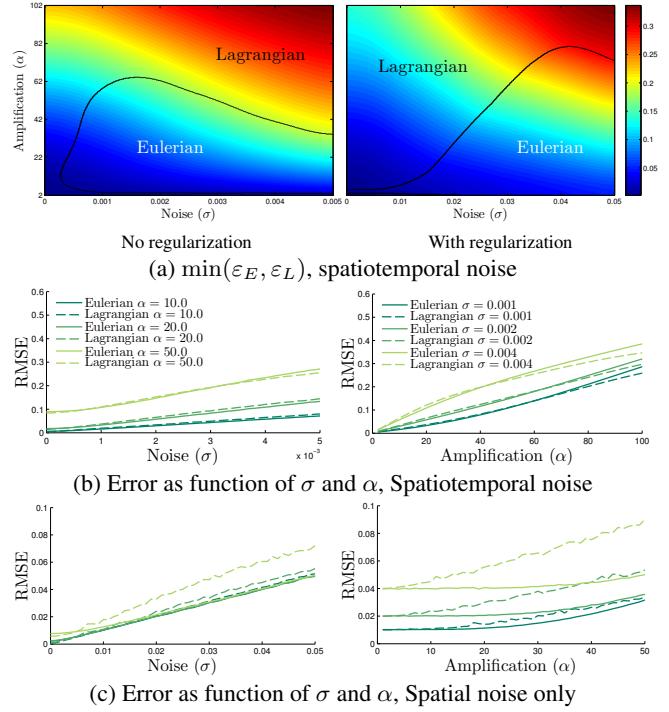


Figure 12: Comparison between Eulerian and Lagrangian motion magnification on a synthetic sequence with additive noise. (a) The minimal error, $\min(\varepsilon_E, \varepsilon_L)$, computed as the (frame-wise) RMSE between each method's result and the true motion-magnified sequence, as function of noise and amplification, colored from blue (small error) to red (large error), with (left) and without (right) spatial regularization in the Lagrangian method. The black curves mark the intersection between the error surfaces, and the overlaid text indicate the best performing method in each region. (b) RMSE as function of σ and α , Spatiotemporal noise (c) Error as function of σ and α , Spatial noise only