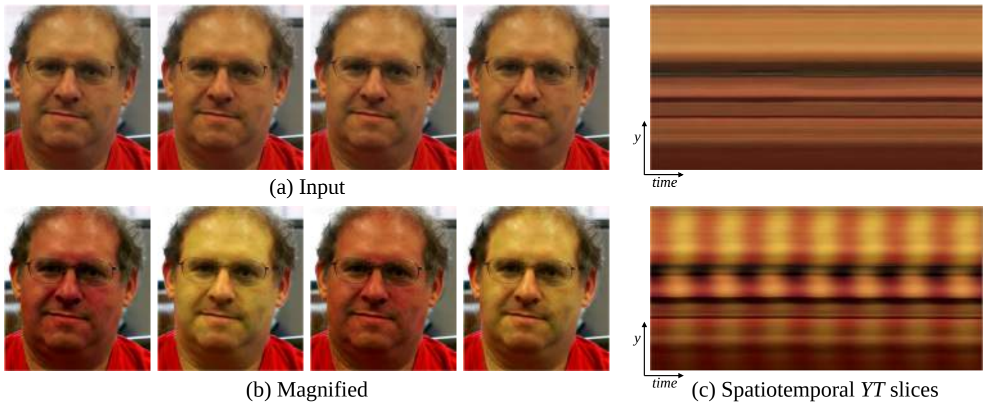
Eulerian Video Magnification for Revealing Subtle Changes in the World

Hao-Yu Wu[1] Michael Rubinstein[1] Eugene Shih[2] John Guttag[1] Frédo Durand[1] William Freeman[1]

[1]MIT CSAIL [2]Quanta Research Cambridge, Inc.



(a) Input

(b) Magnified

(c) Spatiotemporal *YT* slices

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

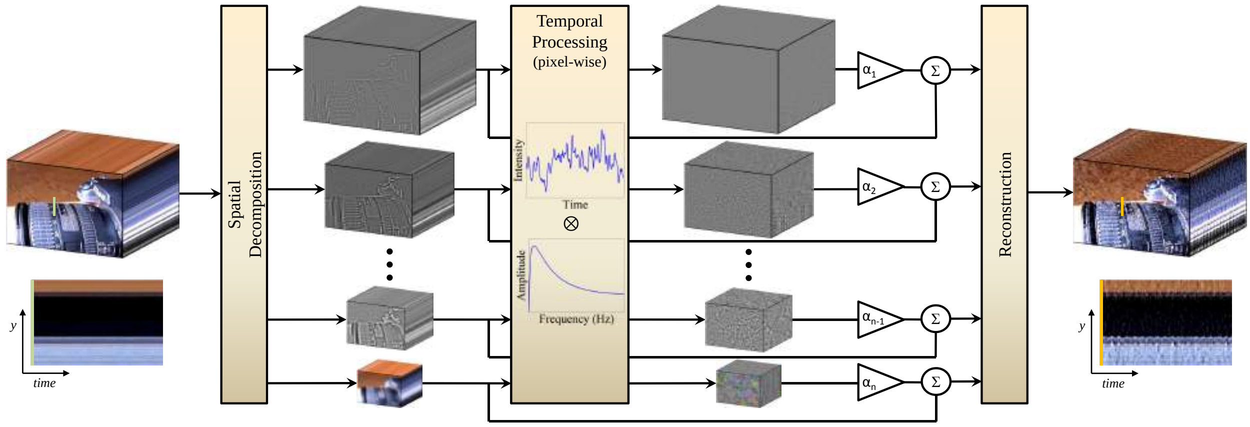


Figure 2: *Overview of the Eulerian video magnification framework. The system first decomposes the input video sequence into different spatial frequency bands, and applies the same temporal filter to all bands. The filtered spatial bands are then amplified by a given factor α, added back to the original 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

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

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

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

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

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

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

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

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

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

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

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

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,

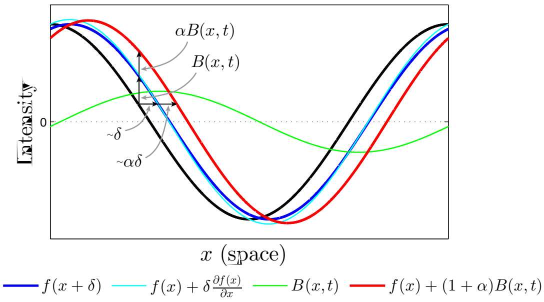


Figure 3: *Temporal filtering can approximate spatial translation. This effect is demonstrated here on a 1D signal, but equally applies to 2D. The input signal is shown at two time instants: $I(x, t) = f(x)$ at time t and $I(x, t + 1) = f(x + \delta)$ at time $t + 1$. The first-order Taylor series expansion of $I(x, t + 1)$ 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,

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

(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,

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

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

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