

Simultaneous Magnification of Subtle Motions and Color Variations in Videos Using Riesz Pyramids

Thomas V. Fontanari^a, Manuel M. Oliveira^{a,*}

^a*Instituto de Informática, UFRGS, Porto Alegre, 91509-900, Brazil*

Abstract

Videos often contain subtle motions and color variations that cannot be easily observed. Examples include, for instance, head motion and changes in skin face color due to blood flow controlled by the heart pumping rhythm. A few techniques have been developed to magnify these subtle signals. However, previous techniques were targeted specifically towards magnification of either motion or color variations. Trying to magnify both aspects applying two of these techniques in sequence does not produce good results. We present a method for magnifying subtle motions and color variations in videos, which can be performed either separately or simultaneously. Our approach is based on the Riesz pyramid, which was previously used only for motion magnification. Besides modifying the local phases of the coefficients of this pyramid, we show how altering its local amplitudes and its residue can be used to magnify intensity/color variations. We demonstrate the effectiveness of our technique in multiple videos by magnifying both subtle motion and color variations simultaneously.

Keywords: Eulerian video magnification, Riesz pyramids

1. Introduction

The world is full of subtle motions and color variations in time that tend to be invisible to the naked eye, but nevertheless carry a rich amount of information. For instance, the almost imperceptible skin color variations in one's face reveal the person's cardiac cycle, minute motion of our shoulders and chest exposes the pattern of our respiration, and small motions of the eye may be symptomatic of a neural disease. Structures undergoing pressure or bearing heavy weights may also deform causing subtle motions, and a drone will move slightly in order to maintain its stability during flight. The ability to extract and visualize subtle signals in videos, therefore, has many practical applications, ranging from the develop-

ment of tools for supporting medical diagnosis to the design of high-precision inspection equipment. Given this potential, a number of algorithms for video motion and color magnification and analysis have been recently proposed [1, 2, 3, 4, 5, 6, 7], and a variety of applications have been developed [8, 9, 10, 11, 12, 13, 14, 15, 16]. While these previous solutions have achieved considerable success in magnifying either subtle motions or color variations, none of them was designed for achieving joint magnification of both motions and intensity/color variations. Trying to simultaneously obtain both effects on a video by applying two such techniques in sequence does not produce good results.

We present a technique for simultaneous magnification of subtle motions and color variations in videos. Having a unified framework is useful in cases where both signals are of interest. Note that these signals might have different frequencies, not needing to be synchronous. For instance, consider the head and chest motion due to breathing, and

*Corresponding author.

Email addresses: tvfontanari@inf.ufrgs.br (T. V. Fontanari), oliveira@inf.ufrgs.br (M. M. Oliveira)

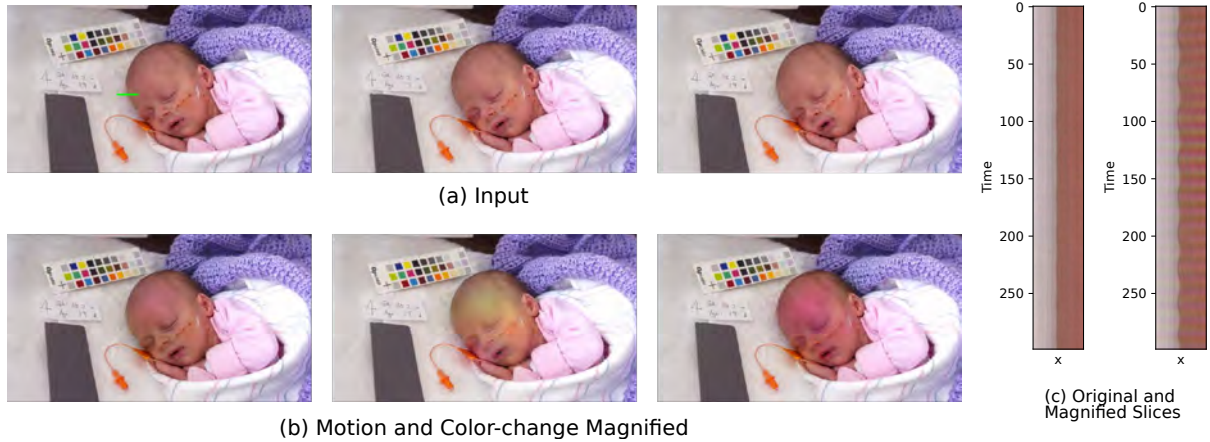


Figure 1: Magnification of subtle motion and color variations produced by our technique. (a) Three non-consecutive frames from a video of a baby (*baby2*). (b) Motion and color magnification applied to the head of the baby for the corresponding frames on top. (c) Comparison of the colors along a small strip of pixels (shown in green in (a) left) across the original (left) and magnified (right) video frames. The variations in width (motion) and colors of the portion of the segment corresponding to the baby skin result from magnification.

skin color variations due to the cardiac rhythm. Our unifying framework is flexible and can be used to magnify only motion or only color variations, as well as combinations of multiple signals at different frequencies (which can be magnified by different factors) all simultaneously. A user interface allows one to select the portions of the video to have their signals magnified in real time. Figure 1 illustrates the use of our technique to perform joint motion and color magnification to the frames of a video. Figure 1 (a) shows three (non-consecutive) frames from a video of a baby. Figure 1 (b) shows the results produced by our technique for motion and color magnification applied to the head of the baby for the corresponding frames shown on top. The variation in color is more easily noticeable. Figure 1 (c) compares the colors of a small strip of pixels along the original (left) and magnified (right) video frames. The strips of pixels are indicated by the green line segment ranging from the baby’s head to the crib bedspread, shown in Figure 1 (a) (left). Note the variations in width (motion) and color of the portion of the segment corresponding to the baby skin in Figure 1 (c) (right).

Our technique uses the Riesz pyramids introduced for subtle motion magnification in [4]. The Riesz pyramids are based on the Riesz transform [17, 18] which provides an image representation in which local phase and local

amplitude are separated from each other. We then use the local phases for motion magnification, as also done in [4], and use the local amplitudes, together with the low-pass residue of the pyramid, for magnifying color changes. Figure 2 summarizes this process. The first stage (Figure 2 (a)) performs a per-frame input video decomposition into Riesz pyramids, resulting in a *phase pyramid*, an *amplitude pyramid*, and a *residue* term. The second stage (Figure 2 (b)) applies a per-pixel temporal filter to the phase pyramid to select the (phase) frequency band ϕ_{B_M} for which motion should be magnified. An independent per-pixel temporal filter is applied to both the amplitude pyramid and to the Riesz pyramid residual. These select the (amplitude) frequency band A_{B_C} , and (residue) frequency band I_{R,B_C} for which color variations should be magnified. The final stage (Figure 2 (c)) recovers the resulting magnified video frames. For this, it scales the selected phase by α_M , and amplitude and residue frequency bands by α_C , recombines them to the original phase ϕ , amplitude A , and residue I_R values, producing the magnified video frames.

The **contributions** of this work include:

- A technique for simultaneous magnification of subtle motions and intensity/color variations in videos

(Section 4). Our technique can be applied to multiple signals defined at different frequency bands;

- A technique to magnify intensity/color variations based on the scaling of amplitude coefficients and residues of Riesz Pyramids (Section 4);
- A chrominance-based masking technique for delimiting regions of interest where signal magnification should be applied (Section 5.1).

2. Related Work

The first method for motion magnification in videos was a layer-based technique that relied on computing the trajectories of feature points obtained from a reference frame [1]. This algorithm is computationally intensive, requiring hours for processing even short video sequences with just a few seconds long. Because of its dependence on tracking features, the technique was later described as a *Lagrangian* approach. The technique is restricted to motion and cannot be used for magnification of intensity/color variations.

Later on, Wu et al. [2] introduced an approach that relied only on local spatial and temporal filtering. As such, the solution was referred to as an *Eulerian* method. The technique was based on computing the Laplacian pyramid for each frame of the video and manipulating the intensities of each pixel in the pyramid. Specifically, each pixel of the pyramid was temporally filtered (across the various frames), selecting only motions whose temporal spectrum was contained in a temporal sub-band of interest. The resulting temporally-filtered signals could then be multiplied by a magnification factor and added back to the original pyramid. Collapsing this modified pyramid produced a magnified version of the frame. Furthermore, applying this process to the residue of the Laplacian pyramid would magnify the color changes in the frame. This technique, however, produced low-quality and limited motion magnification, resulting from clipping artifacts which distorted the frames. The technique also significantly increased the noise levels when motion was magnified.

To improve the quality of the earlier linear Eulerian method and increase the amount of motion magnification, Wadhwa et al. introduced phase-based motion magnification techniques [3, 4]. Those methods were inspired by

previous works that had already demonstrated the relation between local phases of frames and movement [19, 20]. Furthermore, the phase-based approaches still have an Eulerian character, since they rely on local spatial and temporal filtering. Their basic idea is to obtain a measure of *local phase*, which is related to the local motion of a region in the video. Manipulating these phases is then equivalent to manipulating the local motion. The results obtained using these techniques were a significant improvement over the linear Eulerian method of Wu et al. [2], allowing for larger magnification factors and introducing much lower noise levels. Furthermore, Wadhwa et al. [3] suggested (they have not demonstrated it) that the image representation which they used for motion magnification, the *steerable pyramid* [21, 22, 23], could also be used for magnifying color changes in videos. In our work, we show how the Riesz pyramids, introduced in [4] where they were only used for motion magnification, can also be used to magnify intensity/color variations. Unlike Wu et al. [2] who proposed to perform color magnification by scaling the Laplacian pyramid residue, we show that high-quality intensity/color amplification can be achieved by scaling both the amplitude coefficients of the Riesz pyramid and the pyramid’s residue. As such, our work is the first to demonstrate simultaneous high-quality magnification of subtle motions and intensity/color variations in videos.

Additional techniques [24, 5, 25] were designed to specifically overcome the problems that larger motions or color changes cause to the Eulerian methods, as motions and color variations also get modified by these methods. Large motions, however, do not fit the Eulerian frameworks, as they rely on the premise that motions are local. Moreover, large color variations are generally not the ones of interest, as they are already visible, and magnifying them leads to clipping artifacts as their magnified values might get too big when trying to amplify the subtle signals. These methods, however, handle either motion or color-change amplification, but not both simultaneously. Sharing this same limitation, some deep-learning techniques have been recently proposed [26, 27]. Oh et al. [27] presented a technique for magnifying small motions, and Chen and McDuff [26] used separate models for motion and color-change magnification.

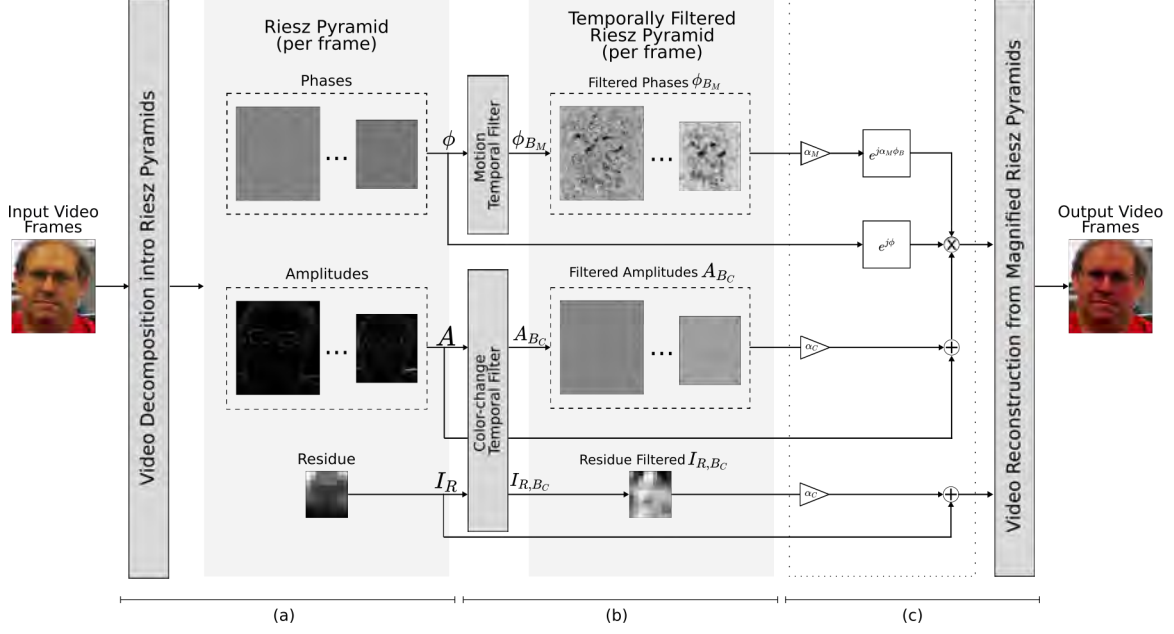


Figure 2: Our simultaneous amplification of subtle motion and intensity/color variation pipeline. (a) The first stage decomposes each frame of the input video into a Riesz pyramid: *phase*, *amplitude*, and *residue*. (b) Per-pixel temporal filter applied to the phase, amplitude, and residue of the pyramid selects the frequencies bands to be magnified: ϕ_{B_M} (for motion), and A_{B_C} and I_{R,B_C} (for intensity/color). (c) The magnified video frames are recovered after scaling the selected frequency bands by factors α_M (motion) and α_C (intensity/color) and recombining them with the original Riesz pyramid elements.

3. Phase-Based Motion Magnification

The main intuition behind phase-based motion magnification is the fact that shifting the phase of waves can be perceived as motion [3]. The relation between phase shifting and motion magnification can be understood by considering the relation between a global phase, obtained through Fourier decomposition, and the displacement. For simplicity, we present the concept in 1D and later extend it to 2D. Thus, consider a single video scanline from a video $I(x, y, t)$ characterized only by displacements, where x and y are respectively the pixel column and row coordinates, and t is time representing the individual video frames. We further simplify the scanline notation by dropping the y coordinate. Thus, a scanline can be described by a function $f: I(x, t) = f(x + \delta(t))$, where $\delta(t)$ corresponds to the displacement over time. Its targeted motion-magnified version can be written as

$\hat{I}(x, t) = f(x + (1 + \alpha)\delta(t))$. Decomposing each “frame” of this video using the complex representation of the Fourier series gives

$$f(x + \delta(t)) = \sum_{\omega=-\infty}^{\infty} A_{\omega} e^{j\omega(x+\delta(t))}, \quad (1)$$

where A_{ω} is the contribution of the frequency ω to the displaced image and $\omega(x + \delta(t))$ is its phase. An arbitrary spatial frequency ω of such a 1D frame of the video can therefore be written as

$$I_{\omega}(x, t) = A_{\omega} e^{j\omega(x+\delta(t))}. \quad (2)$$

Removing the constant term ωx from the phase $\omega(x + \delta(t))$, which can be done by subtracting from it the phase of the frame at $t = 0$, one obtains $\omega\delta(t)$. This term can then be multiplied by a (magnification) factor α and added back to the phase shift by multiplying Equation 2 by the complex

exponential $e^{\alpha\omega\delta(t)}$, producing

$$\hat{I}_\omega(x, t) = A_\omega e^{j\omega(x+(1+\alpha)\delta(t))}. \quad (3)$$

If one applies this same phase shifting process to all frequencies ω and sum over them as in Equation 1, one obtains the targeted motion magnified version of the video

$$\hat{I}(x, t) = f(x + (1 + \alpha)\delta(t)). \quad (4)$$

Hence, by shifting the phases of each pixel x , one also obtains a motion magnified version of the video. However, using the global phases obtained through a Fourier decomposition would not work in general, since each wave covers the whole space. Motion, on the other hand, is in general local. Therefore, in order to magnify motions, it is necessary to obtain a *local phase*.

3.1. Motion Magnification Using Riesz Pyramids

Wadhwa et al. [4] introduced an efficient approach for phase-based motion magnification which uses the local phases obtained from a *Riesz pyramid*. The Riesz pyramid is based on the two-dimensional extension of the Hilbert transform. The transfer function $H_{\mathcal{H}}(\omega)$ of the Hilbert transform $\mathcal{H}$ is given by

$$H_{\mathcal{H}}(\omega) = -j \operatorname{sign}(\omega) = -j \frac{\omega}{\|\omega\|}, \quad (5)$$

where $\operatorname{sign}(\omega)$ is the *sign* function. The Hilbert transform phase shifts each component (i.e., each $\sin$ and $\cos$ function associated to each frequency ω) of the input signal by 90° . Thus, each $\cos$ becomes a $\sin$ and each $\sin$ becomes a $-\cos$. This allows the computation of the quadrature pair [28]

$$f(x) + j \mathcal{H}\{f(x)\} = A(x)e^{j\phi(x)}, \quad (6)$$

where $f(x)$ is the input signal being transformed, and $A(x)$ and $\phi(x)$ are respectively the *local amplitude* and *local phase* of $f(x)$.

As an example, we can consider the quadrature pair obtained when $f(x) = \cos(\omega x)$. Its transform is simply $\sin(\omega x)$, which results in the quadrature pair $\cos(\omega x) + j \sin(\omega x) = e^{j\omega x}$. In general, however, a function will not be defined by a single frequency. Hence, when analysing more complex signals, it is necessary to use a multi-scale approach by first decomposing the original signal into

multiple sub-bands and treat them independently. The local amplitude and phase for each sub-band can then be obtained from the quadrature pair corresponding to the sub-band. The obtained local phases can be used to estimate and magnify the motion at each pixel x . However, before one can apply this to videos, it is necessary to have a generalization of the Hilbert transform to 2D.

The Riesz transform $\mathcal{R}$ generalizes the Hilbert transform to two or more dimensions, and in 2D can be defined by the pair of transfer functions

$$H_{\mathcal{R}}(\vec{\omega}) = \left(\frac{-j\omega_x}{\|\vec{\omega}\|}, \frac{-j\omega_y}{\|\vec{\omega}\|} \right), \quad (7)$$

where ω_x and ω_y are the frequency coordinates in the frequency domain and $\vec{\omega} = (\omega_x, \omega_y)$, and $\|\cdot\|$ is the L2-norm operator. As was done with the one-dimensional case, one can write this pair together with the original function. Since this results in a three component entity, we cannot write it as a complex number and instead write the vector

$$I_R = (I(x, y), R_1(x, y), R_2(x, y)), \quad (8)$$

where $I(x, y)$ is the two-dimensional original function (e.g., an input video frame or a sub-band of a frame), and $R_1(x, y)$ and $R_2(x, y)$ are the results of applying the first and second components of the Riesz transform to $I(x, y)$. This vector can be written in spherical coordinates (Figure 3)

$$I(x, y) = A(x, y) \cos(\phi(x, y)), \quad (9)$$

$$R_1(x, y) = A(x, y) \sin(\phi(x, y)) \cos(\theta(x, y)), \quad (10)$$

$$R_2(x, y) = A(x, y) \sin(\phi(x, y)) \sin(\theta(x, y)), \quad (11)$$

where $A(x, y)$, $\phi(x, y)$ and $\theta(x, y)$ are respectively the *local amplitude*, *local phase*, and *local orientation* of the pixel at position (x, y) . The local phase in this equation has a similar meaning to that of the phase obtained for the one-dimensional case with the Hilbert transform. Furthermore, the local phase is associated to the wave that points towards the local orientation θ , which is in fact the dominant orientation (i.e., the gradient) at pixel (x, y) . The value of the local phase (ϕ) can be obtained from the following quadrature pair, which is analogous to the one produced by the Hilbert transform:

$$I + jQ = Ae^{j\phi}, \quad (12)$$

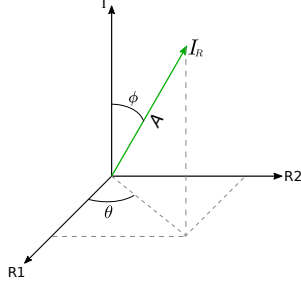


Figure 3: Vector representation of I_R , showing the local phase ϕ , local orientation θ and local amplitude A .

where

$$A = \sqrt{I^2 + R_1^2 + R_2^2}, \quad (13)$$

and

$$Q = \sqrt{R_1^2 + R_2^2} = A \sin \phi. \quad (14)$$

The reason why this local phase can be used for motion magnification is illustrated by Wadhwa et al. [4] using a 2D sinusoidal wave being translated horizontally. This is described by the equation $I(x, y, t) = \cos(\omega_x(x - \delta(t)) + \omega_y y)$, where $\delta(t)$ corresponds to the displacement over time. The Riesz transform of this signal is

$$\mathcal{R}\{I(x, y, t)\} = \frac{(\omega_x, \omega_y)}{\sqrt{\omega_x^2 + \omega_y^2}} \sin(\omega_x(x - \delta(t)) + \omega_y y). \quad (15)$$

The quadrature pair $I + jQ$ can then be computed using $Q = \sin(\omega_x(x - \delta(t)) + \omega_y y)$ (from Equation 14). The local phase at each pixel is therefore

$$\phi(x, y) = \omega_x x + \omega_y y - \omega_x \delta(t). \quad (16)$$

Subtracting from Equation 16 the phase of the first frame (where $\delta(t) = 0$), the factor $-\omega_x \delta(t)$ can be isolated and amplified by multiplication with a constant α_M to obtain the value $-\alpha_M \omega_x \delta(t)$. Finally, the phase of the quadrature pair $I + jQ$ is updated accordingly by multiplication by the complex exponential $e^{-j\alpha_M \omega_x \delta(t)}$. The real part of the resulting pair corresponds to the motion amplified video

$$\hat{I}(x, y, t) = \cos(\omega_x(x - (1 + \alpha)\delta(t)) + \omega_y y). \quad (17)$$

As in the one-dimensional case, the image needs to be decomposed into multiple non-oriented sub-bands before

a phase analysis can be performed. This can be done using a Laplacian pyramid. The Riesz transform is then applied to each of the levels of the Laplacian pyramid, resulting in the Riesz pyramid. The local phases of each level of the Riesz pyramid can then be modified through the process just described, and the real parts of each level is treated as a level of a Laplacian pyramid, which is then collapsed in order to produce the magnified frame.

Furthermore, the local phases can also be spatially and temporally filtered before being used to modify the pyramid. Similarly to Wu et al. [2], Wadhwa et al. [4] temporally filter the spatial local phases (i.e., the local phases in each video frame) in order to select only the motions whose temporal frequencies fall in band of interest. However, as they show, filtering the local phases directly leads to discontinuity problems. Therefore, they choose to instead filter the quantities $\phi \cos \theta$ and $\phi \sin \theta$, where ϕ and θ stand for the local phase and local orientation, respectively. After filtering these quantities, they add them to obtain the filtered local phase. The rest of the process proceeds as described earlier, that is, by multiplying the filtered phases by a magnification factor α and shifting the phases of the pixels of the Riesz pyramid by multiplication with a complex exponential. Further details on the reasoning behind these equations and on the motion magnification using the Riesz pyramids can be found in Appendix A and in [4].

4. Magnification of Color Variations with Riesz Pyramids

Intensity/color variations can be magnified by modifying the residue and the local amplitudes given by the coefficients of the Riesz pyramids. The simplest way to do it is to magnify only the residue. Since the Riesz pyramid is constructed based on the Laplacian pyramid, the result produced is similar to the color-change magnification in the linear Eulerian method developed by Wu et al. [2]. In order to magnify the residue, we temporally filter directly the intensities of its pixels in order to select the color variations of interest. The resulting band is then multiplied by a magnification factor and added back to the residue (see the bottom-most part of Figure 2 (c)). When the Riesz pyramid of the frame is collapsed at the end of the process, the result is a frame with intensity/color variations magnified.

Besides its residue, the amplitude of the Riesz pyramid coefficients also contains information on the intensity/color variations. Furthermore, since the amplitude is separated from the local phase, it is less affected by the motion of the pixel. This is in contrast to altering the intensity of the pixels or of the coefficients of a Laplacian pyramid directly. This process can be better understood by first considering its one-dimensional version, which uses the quadrature pair obtained through the Hilbert transform. Thus, consider a video given by a sinusoid undergoing a small amount of translation (phase shift, $\delta_M(t)$) and amplitude scaling ($\delta_A(t)$): $I(x, t) = \delta_A(t)A \cos \omega_0(x - \delta_M(t))$, for which one wishes to construct its amplitude-only magnified version $\hat{I}(x, t) = (1 + \alpha)\delta_A(t)A \cos \omega_0(x - \delta_M(t))$. Figure 4 (a) illustrates this situation, where the blue line represents the signal at time $t = 0$ and the orange line represents the signal at time $t = 1$.

The quadrature pair of $I(x, t)$ is given by

$$I_{QP}(x, t) = \delta_A(t)Ae^{j\omega_0(x - \delta_M(t))}, \quad (18)$$

where $\delta_A(t)$ and $\delta_M(t)$ correspond respectively to the scaling and motion over time. Isolating the magnitude (Figure 4 (b)), one obtains $\delta_A(t)A$. This quantity can be multiplied by a magnification factor α_C and the resulting value is added back to the original sub-band by adding to the sub-band the complex number (see the mid portion of Figure 2 (c))

$$\alpha_C \delta_A(t)Ae^{j\omega_0(x - \delta_M(t))}. \quad (19)$$

That is, the variation in magnitude is isolated and modified while the phase is kept unaltered. The real part of the resulting complex number is the intensity-magnified video, while the imaginary part is the intensity-magnified version of the Hilbert transform of the input video $I(x, t)$. This is illustrated in Figure 4 (c). As is done for the phases (in the case of motion amplification), one can also use a band-pass filter to select only the temporal frequencies of the amplitude signal which relate to the intensity/color variations that one wishes to magnify.

A similar process can also be applied to video frames using the Riesz transform. In that case, one writes

$$I_{QP}(t) = I(t) + jQ(t) = A(t)e^{j\phi(t)} \quad (20)$$

where the quantities $I(t)$ and $Q(t)$ refer to an arbitrary coefficient of the Riesz pyramids at coordinates (x, y) . The

amplitude $A(t)$ is given by Equation 13. Temporally filtering it and multiplying by the magnification factor α , we find similarly as before $A(t) = \sum_k \alpha_k A_k(t)$, where A_k are the frequency components of $A(t)$. This can be added back to the original image by first multiplying it with $e^{j\phi(t)}$, resulting in the magnified signal

$$\hat{I}_A(t) = \sum_k (1 + \alpha_k) A_k(t) e^{j\phi(t)} \quad (21)$$

The magnified frame is then given by the real part of Equation 21. As mentioned previously, images in general are composed of multiple spatial frequency components. In order to perform the previously described process, it is necessary to first decompose the image into multiple non-oriented sub-bands, such as by using a Laplacian pyramid. The magnification steps are then applied to each of the levels of the pyramid. It is important to notice, however, that color variations are usually very weak and can be indistinguishable from noise in the initial (more detailed) levels of the pyramid. For that reason, we have in general only magnified the amplitudes of the highest (less detailed) levels of the Riesz pyramid, besides the residue, when magnifying color variations.

The effects of magnifying the amplitudes of the Riesz Pyramid are illustrated in Figures 5 and 6. In this example, the video is built using only a single period of a cosine and clipped outside of the center of the image. The function moves periodically a small amount ($\pm 1px$) along the x axis and its amplitude is also periodically scaled 0.01x. For comparison, we also show the result of magnifying the intensity of the pixels directly instead of the amplitudes of the quadrature pairs.

5. Simultaneous Motion and Intensity Magnification

Since the amplitude and the phase in $I + jQ$ are independent of each other, the amplitude of the Riesz pyramids can be modified without affecting the motion and vice-versa. We note that this would not be the case if instead of modifying the amplitudes of the pyramid we had chosen to amplify the intensity I of the pixels directly, as this would modify the phases of the quadrature pairs $I + jQ$ (see Figure 5). Furthermore, since the residue of the pyramid constitutes an additional spatial sub-band to those whose phases are modified when magnifying motion, we can also modify the residue without

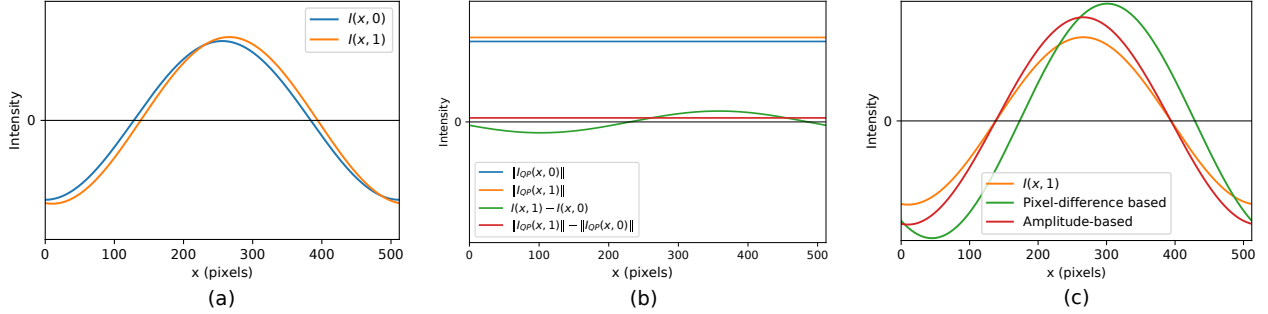


Figure 4: The Hilbert transform can be used to magnify intensity variations in one-dimensional signals. (a) Sub-band of 1-D video at time steps $t = 0$ and $t = 1$. The first frame is both shifted to the right by 10 pixels and scaled up $0.05\times$. (b) The difference between the absolute values of the quadrature pairs at times 0 and 1 is shown by the red line. Since this is an ideal scenario, the magnitudes of the sub-bands capture perfectly only its intensity scaling, while the more straightforward linear difference approach, shown in green, is also influenced by the motion. (c) The magnification of the color changes between time steps using the difference of the magnitudes of the quadrature pair is shown in red. Since in this case motion and scaling can be separated perfectly, the process using the magnitudes is not affected by the motion of the band, while the magnification based on the direct computation of the differences of the pixels (in green) also magnifies motion.

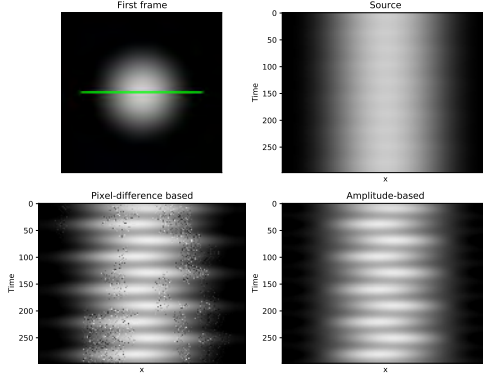


Figure 5: Intensity magnification by scaling the amplitudes of the quadrature pairs. (top left) First frame of a video sequence of a blob oscillating horizontally. The green segment corresponds to the portion of the frame analyzed over time in the other images. (top right) Image obtained by stacking a line of the oscillating blob (under the green segment) over time. (bottom left) Directly amplifying the intensity of the pixels also magnifies motion. (bottom right) The amplitude-based method more closely magnifies only the intensity changes.

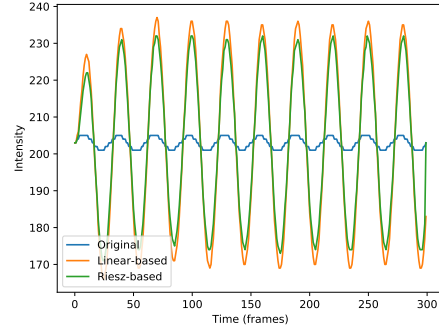


Figure 6: The value of the pixel at the center of the frame (Figure 5) is shown over time for the source and magnified videos. Both methods magnify the intensity changes.

causing problems to the phase magnification. Although some level of motion magnification is unavoidable, the residue contains only lower spatial frequencies and therefore magnifying its intensities will not produce motion of edges. Hence, to simultaneously magnify motion and intensity/color, we scale the local phases (of the coefficients

of the pyramid) for motion magnification, and the corresponding amplitudes and residue for intensity/color magnification.

The asymptotic cost of our technique is similar to the ones used by the previous phase-based motion magnification methods of Wadhwa et al. [3, 4]. But unlike these methods, ours can perform joint magnification of both subtle motions and color variations. The cost of performing a per-frame Riesz pyramid decomposition is $O(n \log n)$ on the number of pixels of the frame. Obtaining a temporally-filtered Riesz pyramid has cost $O(n \log n)$.

Obtaining a motion-color-magnified Riesz pyramid also has cost $O(n \log n)$. Reconstructing the video from the magnified pyramid has cost $O(n \log n)$. Therefore, the total cost of our method is $O(n \log n)$ on the number of pixels of the input video.

5.1. Chrominance-based Mask

The magnification methods described are applied to the entire video frames. This may introduce distracting artifacts in regions where no magnification is intended. Furthermore, one might also be interested in magnifying the subtle signals only in some regions and not in others. Because of the Eulerian character of the motion magnification methods, one can select some frame regions to be magnified while preserving the original content of the remaining ones. Elgharib et al. [5], for instance, used a user-provided alpha matte to select a region of interest. This, unfortunately, requires creating a separate matte for each region of interest in each video. We instead introduce a much simpler and natural way of specifying regions of interest based on chrominance masking. The user can select a region from a video frame by simply pointing at a pixel p contained in the region. The algorithm then performs contextual segmentation and only magnifies pixels in the connected component defined by having chrominance values similar to p 's. The similarity between the chrominance channels of two pixels is computed using the Euclidean distance between their chrominances. This is illustrated in Figure 7.

6. Results

We have implemented the described techniques in Python and used them to magnify both motion and color variations on a large number of videos. We implemented a user interface that allows one to specify the frequency bands (in Hz) for the signals of interest, after which a Riesz pyramid decomposition is obtained and the amplitudes and phases of the pyramid coefficients are computed. Such preprocessing steps take approximately 21 seconds for a 592×528 RGB video with 300 frames. All performance numbers reported in the paper were measured on a notebook with a i7-10510U @1.80 GHz CPU and 16 GB of RAM memory. After the preprocessing, the user can interactively change the magnification factors for

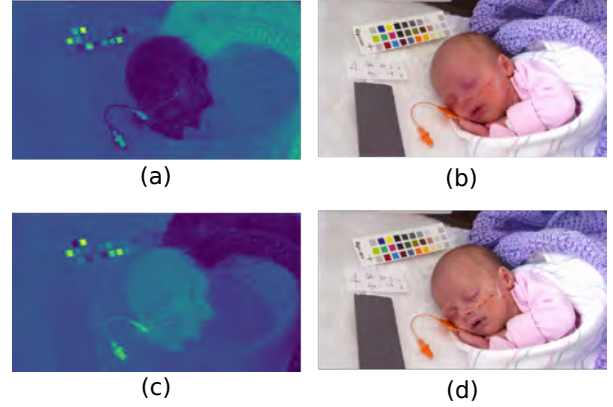


Figure 7: Region selection through chrominance masking. (a) chrominance distances between a user-selected pixel in the baby's face and other pixels in the image are color coded, where darker colors represent smaller distances. In (c), the same process is shown for a pixel selected on the purple blanket. (b) and (d) show color magnification results for the regions corresponding to the darker pixels in the masks shown in (a) and (c), respectively.

the selected bands, as well as the frame regions for which the different magnifications should be applied to, receiving instant feedback. The accompanying video shows various examples of simultaneous magnification of subtle motion and color variations produced with our technique. It also illustrates the interactive selection of independent parameters and regions of interest for the magnification of these two kinds of signals.

In the results shown here the frames were processed in the YIQ color space, with the Riesz pyramids containing 7 layers plus the residual. We magnify motion using only in the luminance (Y) channel and used both luminance and chrominance channels for magnifying color variations (in this case, additional Riesz pyramids for the I and Q channels are created and processed). It is also possible to use only the luminance channel for color magnification, but the inclusion of chrominance channels produce better results. Other color spaces can be used instead of YIQ, as long as there is a separation between luminance-like and chrominance channels. When magnifying color changes, besides the residue, we also magnify the amplitude coefficients from levels 5 up to 7 the Riesz pyramids. We have not used levels 1 to 4 since color variations are generally very weak, being indistinguishable from noise in the

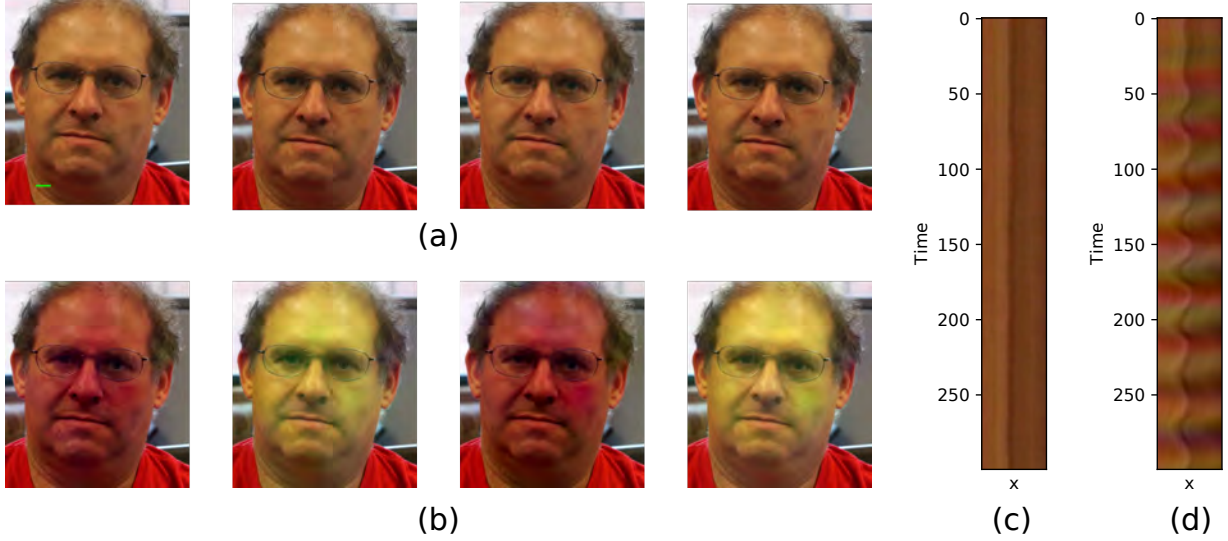


Figure 8: In (a), the frames from the input video *face* are shown. Their motion and color-change magnified version showing the pulse of the man are shown in (b). (c) and (d) show the time slices over his neck (small green line segment on (a) left) in the original and magnified videos, revealing the color changes and motion associated with the pulse in his coronary artery. The original video is from Wu et al. [2].

lowest levels of the pyramids. Moreover, we have used a Laplacian pyramid as the basis for building the Riesz pyramid using the approximate Riesz transform from [3]. Table 1 summarizes the magnification parameters used in the experiments presented in the paper, while the frame dimensions, number of frames, and full processing time for each video is shown in Table 2. *Kernel Sigma* is the standard deviation of a Gaussian blur kernel for the amplitude-weighted blur used to avoid phase discontinuities during phase-based motion magnification (see Appendix A).

Figure 8 illustrates the magnification of both the color variations and motion associated to a man's blood flow. The time slices reveal specifically how his neck moves as the blood passes through an artery. Furthermore, the heart beats also cause the head to move slightly. This is illustrated in Figure 9.

One can also choose to magnify different signals for motion and color changes by using a band-pass filter for the phases and another one for the amplitudes and the residue. In Figure 10 we use a band-pass filter with lower cut-off frequencies for the phases in order to magnify the man's head motion due to respiration together with the

Table 1: Summary of the magnification parameters used for each video.

Video	α_M	α_C	Motion Temporal Bands (Hz)	Color Temporal Bands (Hz)	Kernel Sigma
face heartbeat	30	150	0.83 - 1.10	0.83 - 1.10	4
face2	25	122	0.83 - 1.10	0.83 - 1.10	4
face heartbeat and respiration	12	70	0.20 - 0.33	0.83 - 1.10	4
baby2	10	150	0.61 - 1.91	2.33 - 2.67	4
eye	50	70	30.0 - 40.0	0.83 - 2.00	2
violin	100	-	340 - 370	-	2
drum	5	-	74.0 - 78.0	-	2
guitar	25	-	72.0 - 92.0	-	2
baby	10	-	0.25 - 3.00	-	4
plants	6	-	0.2 - 7.75	-	4

Table 2: Processing time for each video.

Video	Frame Dimensions	# of Frames	Processing Time (s)	Frames per Second
face	592 x 528 x 3	300	29.99	10.04
face2	718 x 570 x 3	300	36.50	8.22
baby2	352 x 640 x 3	500	37.47	13.34
eye	358 x 460 x 3	500	25.26	19.87
violin	360 x 480 x 3	300	10.86	27.63
drum	360 x 640 x 3	450	21.37	21.06
guitar	192 x 432 x 3	300	5.57	53.88
baby	544 x 960 x 3	300	31.25	9.63
plants	700 x 750 x 3	145	15.36	9.44

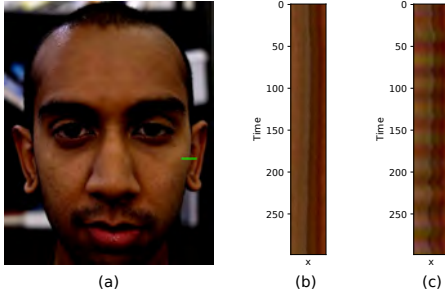


Figure 9: In (a), a reference frame from the source video *face2* is shown. The time slice from the original frame corresponding to the small green line segment is shown in (b) and the time slice of the magnified video is shown in (c). Here what is shown is the color-changes and the head movement which is associated with the pulse rate of the man. The original video is from Wu et al. [2].

color changes caused by his pulse cycle.

We can also select signals with different sampling rates and magnify distant frequency bands for motion and color changes. In Figure 11, a video captured with a sampling rate of 500 Hz is magnified. The temporal filter for the color changes selects frequencies in the range of the heart rate (between 0.83 and 2 Hz), while the temporal filter for motion selects the band from 30 to 40 Hz, corresponding to microsaccades of the eye. Since our technique extends the Riesz motion magnification [29] to also support color, we can also use it for phase-based motion magnification only. In *violin* (Figure 12), a video with a sampling rate of 5,600 Hz has the temporal bands from 340 and 370 magnified, revealing the motion of the bow that plays the strings.

Figure 13 (*drum*) illustrates the magnification of mo-

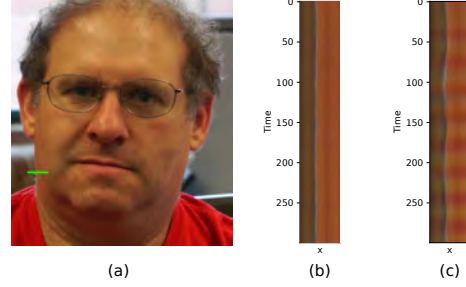


Figure 10: In (a), a reference frame from the source video *face* is shown. The time slice from the original video for the small green line segment is shown in (b) and the time slice from the magnified version is shown in (c), where motion associated to the respiration was magnified together with color changes associated with the pulse. The original video is from Wu et al. [2].

tions from 74 to 78 Hz, revealing the vibrations of the skin of a drum. Figure 14 (*guitar*) shows the recovery of small vibrations from a specific string of a guitar. Finally, Figure 15 (*baby*) presents an example of magnification of the periodic motion of a baby’s chest during respiration.

Our technique also produces better results than a sequential application of a motion magnification method followed by a color-change magnification method, or vice-versa. This is illustrated in Figure 16. When one method is applied after the other, the result of the first magnification algorithm introduces variations in the video (in order to make the signal visible), which are then processed by the next algorithm. Such cascading effect introduces clipping artifacts in the case of motion followed by intensity magnification (Figure 16 (a)), and causes spurious motion magnifications in the case of intensity followed by motion magnification (Figure 16 (b)). Our method does not suffer from these problems, since the phases of the Riesz pyramids are independent of the amplitudes and the residue. Figure 17 shows another example exhibiting similar artifacts. The supplementary material provides side-by-side video comparisons among the results of our technique and the sequential application of color and motion magnification in both orders, highlighting the advantages of our method.

6.1. Discussion and Limitations

Our technique extends the Riesz motion magnification framework to also support magnification of color varia-

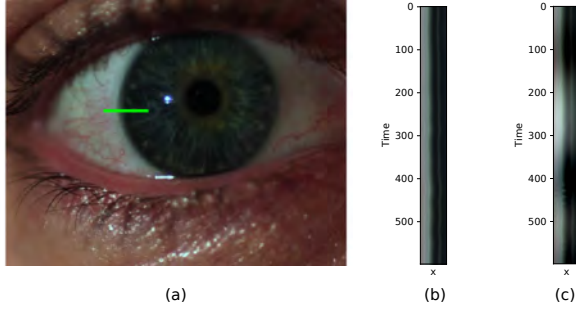


Figure 11: In (a), a reference frame from the source video *eye* is shown. Figures (b) and (c) show the time slices from the border of the eye in the original and magnified videos, where it is possible to see both the intensity change associated with the heart rate and the microsaccades. The original video is from Wadhwa et al. [3].

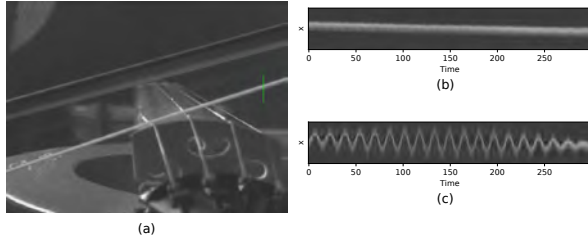


Figure 12: In (a), a reference frame from the source video *violin* is shown. Figures (b) and (c) show original and magnified time slices of the bow playing the violin. The original video is from Wadhwa et al. [3]

tions. Thus, one can obtain phase-based motion magnification and color-change magnification within the same framework. Because the motion magnification component of our method is based on the Riesz motion magnification technique described by Wadhwa et al. [4], both have the same strengths and limitations. Specifically, the Riesz motion magnification method depends on the frames having a single dominant direction and show problems when this is not the case. However, we have not found this to be a problem on real videos. Furthermore, the Riesz framework uses an approximate Riesz transform for efficiency purposes, which in principle leads to worse estimations of the local phase, causing artifacts. In practice, however, we have not observed any artifacts when compared with the use of the ideal Riesz transform. Our technique can be used by any application that re-

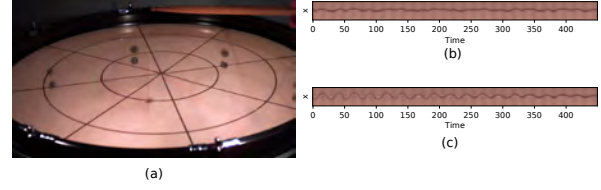


Figure 13: In (a), a reference frame from the source video *drum* is shown. Figures (b) and (c) show original and magnified time slices of the skin of the drum. The original video is from Wadhwa et al. [3]

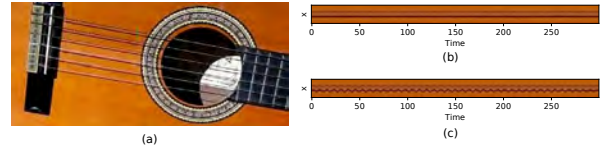


Figure 14: In (a), a reference frame from the source video *drum* is shown. Figures (b) and (c) show original and magnified time slices of the vibrating string. The original video is from Wadhwa et al. [3]

quires subtle motion and/or color-variation magnification, including [8, 9, 10, 11, 12, 13, 14, 15, 16].

To minimize the impact of noise, our technique for color magnification avoids using the finest levels of the spatial decomposition. Ideally, one should magnify all spatial bands for visualizing color changes. However, besides noise issues, this actually also magnifies subtle motions [2]. On the other hand, since we are using the amplitudes of Riesz pyramids and they separate phase changes from amplitude changes – which in an ideal scenario with a single sub-band means that intensity variations and motion are completely separated – our color magnification strategy should not affect motion so much.

Another relevant question is whether the magnified signals indeed correspond to the actual signals that should be magnified. In order to verify this, for instance, Wu et al. [2] attempted to recover the heart rate of a baby from a video and compared it to the true value measured by an electrocardiogram. We performed the same experiment, as illustrated in Figure 18. In our case, we used the magnified video produced by our technique to analyze the frequency response of the temporal values of the pixels in a region of the baby’s face. After removing the DC component, we selected the frequency with the highest magnitude as the estimated heart rate. The ground truth value

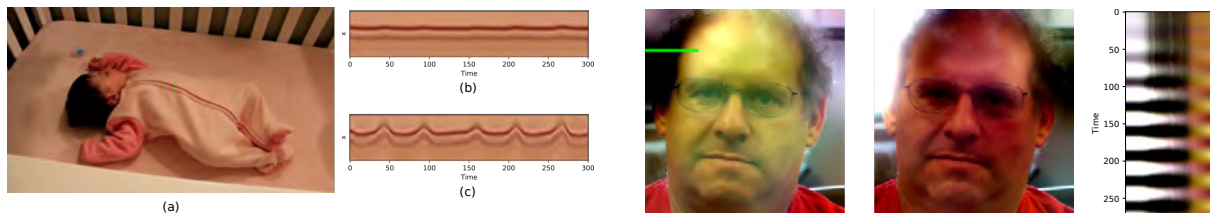


Figure 15: In (a), a reference frame from the source video *baby* is shown. Figures (b) and (c) show original and magnified time slices of the chest of the baby. The original video is from Wadhwa et al. [3]

obtained from an electrocardiogram monitor is 151 bpm (Figure 18 (a) right). The value estimated from the amplified signal using our technique was 150 bpm. This value was obtained by multiplying the detected frequency with highest magnitude (2.5 Hz) by 60 seconds to obtain the number of beats per minute (Figure 18 (c)). Such an experiment shows that our technique does magnify the intended signals.

Finally, we did not address the problem caused by the presence of larger motions in videos. However, since our method is similar to the previous Eulerian techniques, one can apply any of the approaches for video magnification in the presence of larger motions such as the ones by Elgharib et al. [5] and Zhang et al. [24].

7. Conclusion

We presented a method for simultaneous magnification of subtle motions and intensity/color variations in videos. Our technique uses Riesz pyramids, previously used only for motion magnification. We showed how local amplitude of the coefficients of the pyramid and its residue can be used for magnifying intensity/color variations, thus obtaining a method for magnifying motions and intensity/colors variations simultaneously. We demonstrated the effectiveness of our technique on multiple videos for which previously only a single signal had been observed at once.

Possible directions for future exploration include improving our technique by adding to it the methods developed for dealing with larger motions [24, 5, 25]. Furthermore, the orientation of the Riesz coefficients could be used as a parameter for creating directional selective

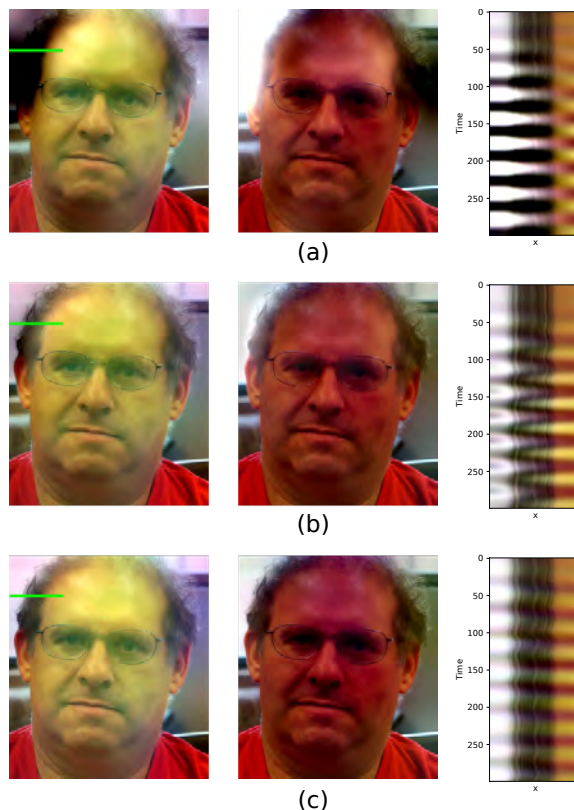


Figure 16: Sequentially applying motion and color-change magnification algorithms introduces artifacts in the resulting videos. The images in the third column were obtained by stacking the pixels corresponding to the green segment on the left for all frames in the video sequence. (a) Applying motion magnification followed by color-change magnification introduces clipping artifacts (intensity values above 255 or below 0, which are clipped). These artifacts can be observed in the alternating black and white stripes on the image on the right. (b) Applying color-change magnification followed by motion magnification mistakenly causes the color variations to appear to also be moving. Thus, in the rightmost column, note the alternating protrusions from the bright background and from the forehead color into the dark region corresponding to the man's hair, making it to appear wiggling. (c) Our technique does not suffer from these issues, as the Riesz pyramid representation provides independent measures of phase and amplitude.

masks. In this case, it would be possible to only magnify motions oriented along a group of intended directions.

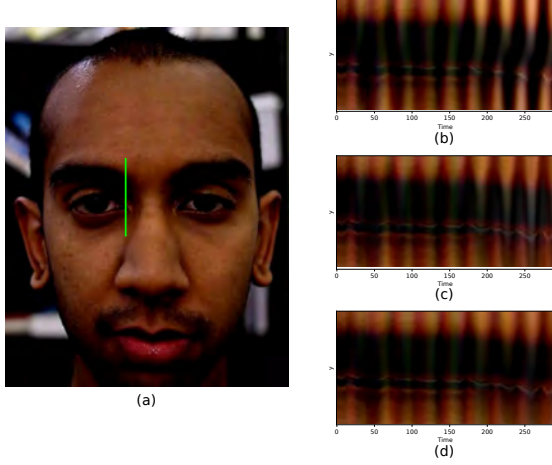


Figure 17: Artifacts resulting from the sequential application of motion and color-change magnification. The images in the right column were obtained by stacking the pixels corresponding to the green segment in (a). (b) Clipping artifacts caused by magnifying intensity variations resulting from motion magnification in a previous step. (c) Applying color-change magnification followed by motion magnification causes the man’s skin color to extend over his eyebrows. (d) Our technique does not suffer from these artifacts (see accompanying video).

Appendix A. Quaternionic Phase Filtering

Instead of temporally filtering the local phases ϕ directly, Wadhwa et al. [4] chose instead to filter the quantities $\phi \cos \theta$ and $\phi \sin \theta$. One reason for this comes from the fact that filtering the local phases directly leads to discontinuity problems. For instance, if the local orientation θ is limited to the interval $[0, \pi]$ while the local phase goes from $[-\pi, \pi]$, a small change in the orientation between two adjacent pixels (spatially or temporally) can lead to a sudden phase change from ϕ to $-\phi$ or vice-versa. The same discontinuity does not exist when the aforementioned quantities are filtered.

After the applying the temporal filter, Wadhwa et al. [29] also filter the phases spatially through the use of an amplitude weighted blur. That is, the phases are convolved with a Gaussian kernel K and weighted by the value of the amplitude of the Riesz pyramid coefficient at the pixel. This spatial filtering reduces noise from the computation of the phases and also reduces errors from approximations used when constructing the Riesz pyra-

mid. Finally, the filtered quantities are recombined. The equation below summarizes these steps

$$\cos \theta \frac{A \cos(\theta) \phi * K}{A * K} + \sin \theta \frac{A \sin(\theta) \phi * K}{A * K}, \quad (\text{A.1})$$

where $*$ represents convolution. The rest of the process proceeds as described earlier, that is, by multiplying the filtered phases by a magnification factor α and shifting the phases of the pixels of the Riesz pyramid by multiplication with a complex exponential.

The same quantities $\phi \cos \theta$ and $\phi \sin \theta$, however, can also be deduced naturally by using a quaternionic representation of the Riesz pyramid coefficients, as shown in a later work by Wadhwa et al. [29]. Instead of using the vector representation of the Riesz pyramid coefficients in Equation 8, one can write them in a quaternionic form as

$$\mathbf{r} = I + iR_1 + jR_2, \quad (\text{A.2})$$

where $\mathbf{r}$ is an arbitrary coefficient of the Riesz pyramid and I , R_1 and R_2 are defined as in Section 3. Similarly as before, this can also be expressed in terms of the local phase $\phi(x, y)$, local orientation $\theta(x, y)$, and local amplitude $A(x, y)$, resulting in

$$\mathbf{r} = A \cos(\phi) + iA \sin(\phi) \cos(\theta) + jA \sin(\phi) \sin(\theta). \quad (\text{A.3})$$

When working with complex numbers, its phase is given by the logarithm of the normalized complex number. Here, similarly, Wadhwa et al. [29] compute the *quaternionic phase* using the logarithm of the normalized quaternion. The norm $\|\mathbf{r}\|$ is given by A and the quaternionic phase is therefore

$$\log \left(\frac{\mathbf{r}}{\|\mathbf{r}\|} \right) = \frac{\mathbf{v}}{\|\mathbf{v}\|} \arccos(q_1) = i\phi \cos(\theta) + j\phi \sin(\theta), \quad (\text{A.4})$$

where $\mathbf{v} = i\phi \cos(\theta) + j\phi \sin(\theta)$ and q_1 is the real part of $\mathbf{r}$. This is the same quantity that was used in Equation A.1 for filtering in place of the phase and it has no discontinuities in the phase arising from whether orientation is represented by θ or by $\theta + \pi$.

Acknowledgments

This work was funded by CNPq-Brazil (fellowships and grants 312975/2018-0 and 423673/2016-5), and CAPES Finance Code 001.

References

- [1] Liu, C, Torralba, A, Freeman, WT, Durand, F, Adelson, EH. Motion magnification. *ACM transactions on graphics (TOG)* 2005;24(3):519–526.
- [2] Wu, HY, Rubinstein, M, Shih, E, Gutttag, J, Durand, F, Freeman, W. Eulerian video magnification for revealing subtle changes in the world. *ACM transactions on graphics (TOG)* 2012;31(4):1–8.
- [3] Wadhwa, N, Rubinstein, M, Durand, F, Freeman, WT. Phase-based video motion processing. *ACM Transactions on Graphics (TOG)* 2013;32(4):1–10.
- [4] Wadhwa, N, Rubinstein, M, Durand, F, Freeman, WT. Riesz pyramids for fast phase-based video magnification. In: 2014 IEEE International Conference on Computational Photography (ICCP). IEEE; 2014, p. 1–10.
- [5] Elgharib, M, Hefeeda, M, Durand, F, Freeman, WT. Video magnification in presence of large motions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, p. 4119–4127.
- [6] Wu, X, Yang, X, Jin, J, Yang, Z. Pca-based magnification method for revealing small signals in video. *Signal, Image and Video Processing* 2018;12(7):1293–1299.
- [7] Liu, L, Lu, L, Luo, J, Zhang, J, Chen, X. Enhanced Eulerian video magnification. In: 2014 7th International Congress on Image and Signal Processing. IEEE; 2014, p. 50–54.
- [8] Arango, C, Alata, O, Emonet, R, Legrand, AC, Konik, H. Subtle motion analysis and spotting using the riesz pyramid. 2018,.
- [9] Lauridsen, H, Gonzales, S, Hedwig, D, Perrin, KL, Williams, CJ, Wrege, PH, et al. Extracting physiological information in experimental biology via eulerian video magnification. *BMC biology* 2019;17(1):1–26.
- [10] Wadhwa, N. Revealing and analyzing imperceptible deviations in images and videos. Ph.D. thesis; Massachusetts Institute of Technology; 2016.
- [11] Le Ngo, AC, Johnston, A, Phan, RC, See, J. Micro-expression motion magnification: Global Lagrangian vs. local Eulerian approaches. In: 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018). 2018, p. 650–656. doi:10.1109/FG.2018.00102.
- [12] Chen, W, Picard, RW. Eliminating physiological information from facial videos. In: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). IEEE; 2017, p. 48–55.
- [13] Perrot, V, Salles, S, Vray, D, Liebgott, H. Video magnification applied in ultrasound. *IEEE Transactions on Biomedical Engineering* 2018;66(1):283–288.
- [14] Davis, A, Bouman, KL, Chen, JG, Rubinstein, M, Durand, F, Freeman, WT. Visual vibrometry: Estimating material properties from small motion in video. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, p. 5335–5343.
- [15] Balakrishnan, G, Durand, F, Gutttag, J. Detecting pulse from head motions in video. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2013, p. 3430–3437.
- [16] Belaid, A, Boukerroui, D, Maingourd, Y, Lerallut, JF. Phase-based level set segmentation of ultrasound images. *IEEE Transactions on Information Technology in Biomedicine* 2010;15(1):138–147.
- [17] Felsberg, M, Sommer, G. The monogenic signal. *IEEE transactions on signal processing* 2001;49(12):3136–3144.
- [18] Unser, M, Sage, D, Van De Ville, D. Multiresolution monogenic signal analysis using the riesz-laplace wavelet transform. *IEEE Transactions on Image Processing* 2009;18(11):2402–2418.
- [19] Freeman, WT, Adelson, EH, Heeger, DJ. Motion without movement. *ACM Siggraph Computer Graphics* 1991;25(4):27–30.
- [20] Gautama, T, Van Hulle, MA. A phase-based approach to the estimation of the optical flow field

using spatial filtering. IEEE Transactions on Neural Networks 2002;13(5):1127–1136. doi:10.1109/TNN.2002.1031944.

[21] Simoncelli, EP, Freeman, WT. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In: Proceedings., International Conference on Image Processing; vol. 3. IEEE; 1995, p. 444–447.

[22] Freeman, WT, Adelson, EH, et al. The design and use of steerable filters. IEEE Transactions on Pattern analysis and machine intelligence 1991;13(9):891–906.

[23] Simoncelli, EP, Freeman, WT, Adelson, EH, Heeger, DJ. Shiftable multiscale transforms. IEEE transactions on Information Theory 1992;38(2):587–607.

[24] Zhang, Y, Pinteá, SL, van Gemert, JC. Video acceleration magnification. In: Computer Vision and Pattern Recognition. 2017,.

[25] Wu, X, Yang, X, Jin, J, Yang, Z. Amplitude-based filtering for video magnification in presence of large motion. Sensors 2018;18(7):2312.

[26] Chen, W, McDuff, D. Deepmag: Source specific motion magnification using gradient ascent. arXiv preprint arXiv:180803338 2018;.

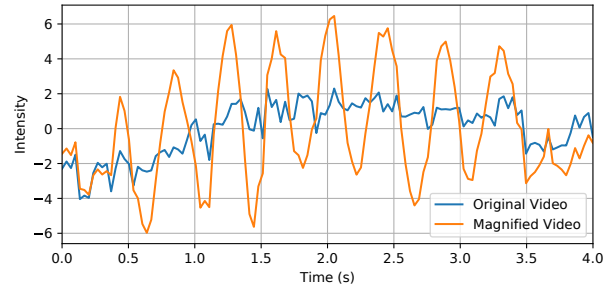
[27] Oh, TH, Jaroensri, R, Kim, C, Elgharib, M, Durand, F, Freeman, WT, et al. Learning-based video motion magnification. In: Proceedings of the European Conference on Computer Vision (ECCV). 2018, p. 633–648.

[28] Knutsson, H. Texture analysis using two-dimensional quadrature filters. In: IEEE Comput. Soc. Workshop Comp. Architecture Patt. Anal. Image Database Mgmt. 1983, p. 206–213.

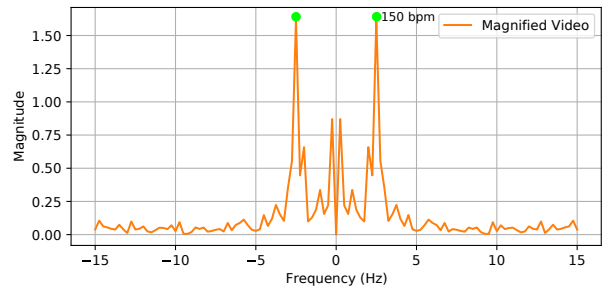
[29] Wadhwa, N, Rubinstein, M, Durand, F, Freeman, WT. Quaternionic representation of the riesz pyramid for video magnification. 2014.



(a)



(b)



(c)

Figure 18: The estimated heart rate of the baby from a frequency analysis matches the actual value obtained with an electrocardiogram. (a) A representative frame of the *baby2* video with a picture of the electrocardiogram monitor used by Wu et al. [2] to measure the actual heart rate. The electrocardiogram measures 151 bpm. (b) Values over time of the pixels in the region marked in green in (a) for the original and magnified videos. (c) Corresponding frequency spectrum. The DC component was removed before generating the plot to facilitate visualization. The value of 150 bpm was obtained by multiplying the 2.5 Hz by 60 seconds.