Subtle Motion Analysis and Spotting using the Riesz Pyramid

Carlos Andres Arango, Olivier Alata, Rémi Emonet, Anne-Claire Legrand and Hubert Konik

Univ. Lyon, UJM-Saint-Etienne, CNRS, Institut d Optique Graduate School, Laboratoire Hubert Curien UMR 5516,
F-42023, Saint-Etienne, France

Abstract: Analyzing and temporally spotting motions which are almost invisible to the human eye might reveal interesting information about the world. However, detecting these events is difficult due to their short duration and low intensities. Taking inspiration from video magnification techniques, we design a workflow for analyzing and temporally spotting subtle motions based on the Riesz pyramid. In addition, we propose a filtering and masking scheme that segments motions of interest without producing undesired artifacts or delays. In order to be able to evaluate the spotting accuracy of our method, we introduce our own database containing videos of subtle motions. Experiments are carried out under different types and levels of noise. Finally, we show that our method is able to outperform other state of the art methods in this challenging task.

1 INTRODUCTION

Although the human visual system is capable of detecting a plethora of objects and phenomena that surround us it has limited spatio-temporal sensitivity. There are some movements of low-spatial amplitude that are difficult to detect by the human eye which might conceal interesting information about the world (Liu et al., 2005). For instance it has been reported that the cyclical movement of blood in the human body causes the head to move in a subtle periodic motion which has been used to calculate the heart rate in a non invasive manner (Balakrishnan et al., 2013)(Irani et al., 2014). For this reason motion magnification techniques have been developed in order to reveal almost invisible signals in videos. These techniques have been used for modal analysis (the study of the dynamic properties of structures under vibrational excitation) (Chen et al., 2015), for enhancing the motion of the blood flow in the finger veins for liveness detection (Raghavendra et al., 2015), to enhance the movement of facial expressions for anti-spoofing in a face biometric system (Bharadwaj et al., 2013), for amplifying the motion of pulsating arteries during an endoscopic surgery (McLeod et al., 2014) and for micro-expression recognition (Park and Kim, 2009).

The main problem with motion magnification methods is that they exaggerate the motion on a video rather than explicitly estimating it. However, our careful examination showed that intermediate representations produced by these methods can be used as proxies for motion. Furthermore, considering that a lot of real-life applications require to detect when an event takes place, these representations could potentially be used to temporally spot subtle events (finding the temporal locations of subtle movements from a video sequence). Thus, we propose a method to analyze and spot subtle motions in a video by analyzing the phase variations between frames based on the Riesz Pyramid. This paper is organized as follows. Section 2 serves as an introduction to the theoretical material to understand the Riesz Pyramid outputs, its quaternionic representation and filtering. Section 3 describes our proposed methodology and contributions. Section 4 describes some potential applications, our experiments, results and discussion. Finally, section 5 presents our conclusions.

2 BACKGROUND

Eulerian motion magnification is a family of techniques that amplifies subtle motion in videos. They are inspired by the *Eulerian* perspective, in reference to fluid dynamics where the properties of a voxel of fluid, such as pressure and velocity, evolve over time. The first of these techniques (Wu et al., 2012) exaggerates motion by amplifying temporal color changes at fixed positions. However, this method can signif-
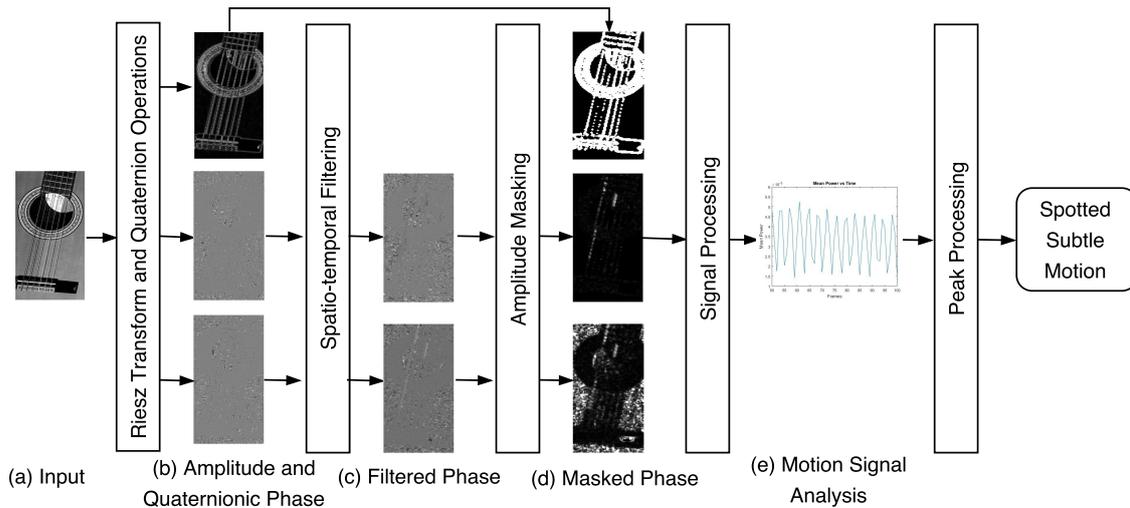
Figure 1: Subtle motion analysis framework. Each frame from an input video (a) is processed with the Riesz Pyramid to obtain the local amplitude and the quaternionic phase (b). We apply an improved spatio-temporal filter to the quaternionic phase (c). Then we use local amplitude to mask relevant areas in the quaternionic phase (d). The phase is processed into a 1-D signal (e) and we detect and classify the resulting peaks to spot the subtle motions.

icantly amplify noise when the magnification factor is increased. Another method proposed by (Wadhwa et al., 2013), amplifies the phase variation over time within each image subband using the steerable pyramid (an over-complete transform that decomposes an image according to spatial scale, orientation, and position). However, the main disadvantage of this method comes from the complex steerable pyramids which are very overcomplete and costly to construct. Later on, (Wadhwa et al., 2014a) propose a new method based on the Riesz pyramid which produced motion-magnified videos of comparable quality to the previous one, but the videos can be processed in one quarter of the time, making it more suitable for real-time or online processing applications.

Although, video magnification is a powerful tool for magnify subtle motions in videos, it doesn't indicate the moment when these motions take place. However, the filtered quaternionic phase difference obtained during the Riesz magnification approach (Wadhwa et al., 2014b) seems to be a good proxy of motion, thus it could potentially be used for analyzing and temporally segmenting subtle motions. In this section, we do an introduction to the Riesz pyramid, its quaternionic representation and quaternionic filtering.

2.1 Riesz Monogenic Signal

The Riesz pyramid is constructed by first breaking the input image into non-oriented subbands using an efficient, invertible replacement for the Laplacian pyramid, and then taking an approximate Riesz transform

of each band. The key insight into why this representation can be used for motion analysis is that the Riesz transform is a steerable Hilbert transformer and allows us to compute a quadrature pair that is 90 degrees out of phase with respect to the dominant orientation at every pixel. This allows us to phase-shift and translate image features only in the direction of the dominant orientation at every pixel.

Following (Unser et al., 2009), in two dimensions, the Riesz transform is a pair of filters with transfer functions

$$-i\frac{\omega_x}{\|\overrightarrow{\omega}\|}, -i\frac{\omega_y}{\|\overrightarrow{\omega}\|} \qquad (1)$$

with $\overrightarrow{\omega} = [\omega_x, \omega_y]$ being the signal dimensions in the frequency domain. If we filter a given image subband I using Eq. 1, the result is the pair of filter responses, $(R_1; R_2)$. The input I and Riesz transform $(R_1; R_2)$ together form a triple (the monogenic signal) that can be converted to spherical coordinates to yield the local amplitude A, local orientation θ and local phase ϕ using the equations

$$
\begin{aligned}
I &= A\cos(\phi) \\
R_1 &= A\sin(\phi)\cos(\theta) \qquad (2) \\
R_2 &= A\sin(\phi)\sin(\theta)
\end{aligned}
$$

2.2 Quaternion Representation of Riesz Pyramid

The Riesz pyramid coefficient triplet $(I; R_1; R_2)$ can be represented as a quaternion **r** with the original subband I being the real part and the two Riesz transform

components $(R_1; R_2)$ being the imaginary i and j components of the quaternion.

$$\mathbf{r} = I + iR_1 + jR_2 \tag{3}$$

The previous equation can be rewritten using (4) as:

$$\mathbf{r} = A\cos(\phi) + iA\sin(\phi)\cos(\theta) + jA\sin(\phi)\sin(\theta) \tag{4}$$

However, the decomposition proposed by (4) is not unique. That means that both (A, ϕ, θ) and $(A, -\phi, \theta + \pi)$ are possible solutions. This can be solved if we consider

$$\phi\cos(\theta), \phi\sin(\theta) \tag{5}$$

which are invariant to this sign ambiguity. If the Riesz pyramid coefficients are viewed as a quaternion, then Eq. 5 is the quaternion logarithm of the normalized coefficient[1]. Thus, the local amplitude A and quaternionic phase defined in Eq. 5 are computed:

$$A = \|\mathbf{r}\| \tag{6}$$

$$i\phi\cos(\theta) + j\phi\sin(\theta) = \log(\mathbf{r}/\|\mathbf{r}\|) \tag{7}$$

2.3 Filtering of Quaternionic Phase

In previous *Eulerian* motion amplification papers, motions of interest were isolated and denoised with temporal filters. However, the quaternionic phase cannot be naively filtered since it is a wrapped quantity (Wadhwa et al., 2014b). Therefore a technique developed in (Lee and Shin, 2002) is used to filter a sequence of unit quaternions (by first unwrapping the quaternionic phases in time and then using a linear time invariant (LTI) filter). This technique is used to filter the Riesz pyramid coefficients at each pixel in each scale in time. It is also assumed that the local orientation is roughly constant in time and space.

Suppose at a single frame n, a single pixel (x, y) in a single scale ω_r the normalized Riesz pyramid coefficients are:

$$\mathbf{r}_n = \cos(\phi_n) + i\sin(\phi_n)\cos(\theta_n) + j\sin(\phi_n)\sin(\theta_n) \tag{8}$$

In the case of ordinary complex phase unwrapping, we would take the principal value of the difference between successive terms and then do a cumulative sum to give an unwrapped sequence in which the difference between two successive terms is always in the interval $(-\pi, \pi]$. We compute the principal value of the difference between two successive coefficients by dividing them and then taking the logarithm:

$$\log(\mathbf{r}_1), \log(\mathbf{r}_2\mathbf{r}_1^{-1}), \dots, \log(\mathbf{r}_n\mathbf{r}_{n-1}^{-1}) \tag{9}$$

---

[1]An extended review of the quaternionic representation icluding complex exponentiation and logarithms can be found in (Wadhwa et al., 2014b).

If we assume that $\theta_n = \theta + \varepsilon$, that is that the local orientation is roughly constant over time at every pixel, the k term will be close to zero. More specifically,

$$\mathbf{r}_n\mathbf{r}_{n-1}^{-1} = \cos(\phi_n - \phi_{n-1}) + i\sin(\phi_n - \phi_{n-1})\cos(\theta) + j\sin(\phi_n - \phi_{n-1})\sin(\theta) + O(\varepsilon) \tag{10}$$

by ignoring the $O(\varepsilon)$ term, the logarithm is

$$i([\phi_n - \phi_{n-1}])\cos(\theta) + j([\phi_n - \phi_{n-1}])\sin(\theta) \tag{11}$$

The second step is to perform a cumulative sum of (9)

$$\phi_1\mathbf{u}, (\phi_1 + [\phi_2 - \phi_1])\mathbf{u}, \dots, \left(\phi_1 + \sum_{l=2}^{n}[\phi_n - \phi_{n-1}]\right)\mathbf{u} \tag{12}$$

where $\mathbf{u} = i\cos\theta + j\sin\theta$. If we let $\phi_n' = \phi_1 + \sum_{l=2}^{n}[\phi_n - \phi_{n-1}]$ the series can be written as:

$$i\phi_n'\cos(\theta) + j\phi_n'\sin(\theta) \tag{13}$$

Afterwards we can isolate motions of interest in the quaternionic phase signal using an LTI filter. Furthermore, the signal-to-noise ratio (SNR) of the phase signal can be increased by spatially denoising each frame with an amplitude-weighted spatial blur with Gaussian Kernel K_ρ with standard deviation ρ on the i and j components of the temporally filtered signal.

$$i\frac{A\phi'\cos(\theta) * K_\rho}{A * K_\rho} + j\frac{A\phi'\sin(\theta) * K_\rho}{A * K_\rho} \tag{14}$$

Assuming that the orientation does not change substantially in the support of K_p, then $\cos(\theta)$ and $\sin(\theta)$ can be moved outside of the convolution in Eq. 14 to get:

$$i\cos(\theta)\phi'' + j\sin(\theta)\phi'' \tag{15}$$

where

$$\phi'' = \frac{A\phi' * K_\rho}{A * K_\rho} \tag{16}$$

where (15) is the filtered quaternionic phase obtained for each pixel of each subband in each frame. This signal is then multiplied by a magnification factor and the pyramid is reconstructed to obtain the magnified video (Wadhwa et al., 2014a). In the following section, we show how this signal can be used to analyze subtle motions.

3 SUBTLE MOTION ANALYSIS

Our proposed algorithm goes as follows: first we use the Riesz Pyramid to calculate the amplitude and quaternionic phase of the images. Secondly, we implement a proper spatio-temporal filtering scheme

which can enhance motions of interest without producing delays or undesired artifacts. Thirdly, we isolate areas of potential subtle motion based on the computed amplitude. Finally, we measure the dissimilarities of quaternionic phases over time and transform them into a 1-D signal, which is used to estimate the moment when the subtle moment is taking place. A graphic representation of our framework can be seen in Figure 1.

3.1 Temporal Filtering Considerations

For an image sequence of N frames we perform the process described in Sec.2.1 and Sec.2.2 for each frame $n \in N$. However, not all levels of the pyramid are able to provide useful information about the subtle motion. Thus, after processing our video using different pyramid levels, we select the one that shows more subtle changes. We then obtain both local amplitude A_n and quaternionic phase $(\phi_n \cos(\theta), \phi_n \sin(\theta))$. We apply the process described in Sec. 2.3 to obtain ϕ'' (Eq. 16). However, since we are aiming to detect any significant quaternionic phase shifts between frames and to compensate for the cumulative sum made in (12), we calculate the difference of two consecutive filtered quaternionic phases:

$$\Delta\phi_n \mathbf{u} = \phi_n'' \mathbf{u} - \phi_{n-1}'' \mathbf{u} \tag{17}$$

where $\mathbf{u} = i\cos\theta + j\sin\theta$. The previous work in eulerian motion magnification have given their users freedom to choose any temporal filtering method available. However, since we require to pinpoint the exact moment when subtle motion is detected we cannot use traditional causal filters which may delay the signal response (Figure 2c). Therefore we propose to use a digital non-causal zero-phase finite impulse response (FIR) filter.

$$\Phi_n \mathbf{u} = b_0 \Delta\phi_n \mathbf{u} + \sum_{k=1}^{p} b_k (\Delta\phi_{n+k} \mathbf{u} + \Delta\phi_{n-k} \mathbf{u}) \tag{18}$$

where p is an even number and b_k is a coefficient of a FIR filter of length $2p+1$. One limitation of this method is that non-causal filters requires to use the previous and following p frames from the current frame (therefore for online applications there must be a delay of at least p frames). Another element to consider is that *Eulerian* amplification methods are tailored for a particular task. These methods aim to amplify subtle periodical movements (such as human breathing, the vibration of an engine, the oscillations of a guitar string, etc) by temporally band-passing some potential movements and amplify them. However, these methods do not consider subtle non periodical movements (such as blinking or facial MEs).

(a)   (b)

(c)   (d)   (e)
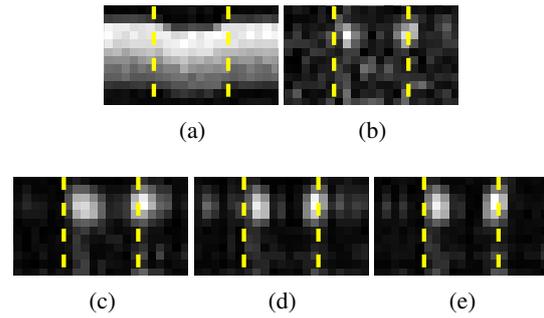
Figure 2: A comparison of different filter responses for subtle motion detection. (a) is an slice in time of an input image sequence with subtle non periodical motion (The purple dashed lines indicate when the subtle movement starts and ends). (b) is the calculated quaternionic phase shift $\Delta\phi_n$ of (a). We reduce the noise in (b) using three different filtering schemes: (c) an IIR Butterworth causal filter which delays the signal; (d) a FIR non-causal band-pass filter which does not delay the signal but it creates some artifacts before and after the motion has taken place (Gibbs phenomenon); (e) a FIR non-causal low-pass filter (our proposal).

The latter type of motion, when band-passed, creates some large oscillations near the beginning and the end of the subtle motion (Figure 2d) as stated by Gibbs phenomenon. Therefore we decided to use low-pass filtering for this type of signals (Figure 2e).

3.2 Amplitude Masking

The first step is to simplify the quaternionic phase by discarding the orientation and calculate the euclidean norm of the phase thus:

$$|\Phi_n'| = \sqrt{(\Phi_n \sin\theta)^2 + (\Phi_n \cos\theta)^2} \tag{19}$$

One thing to consider before trying to detect subtle motion is the problem of image noise. Assuming the general case of two static images corrupted with some level of additive Gaussian noise, their quaternionic phase difference would be non-zero ($|\Phi_n'| > 0$) even after the phase SNR is improved by ways of spatial and temporal filtering (Sec. 2.3). We have observed that the Φ_n' values could have a high variance in areas where local amplitude A has a relative low value regardless of the presence of motion. Considering that the motion phase-signal in regions of low amplitude is not meaningful (Wadhwa et al., 2013) we decide to isolate these areas using a threshold of validation computed from the local amplitude. However, since the scale of local amplitude might vary from subject to subject (Some videos might have objects with stronger edges compared to others) we need to normalize the local amplitude before we can
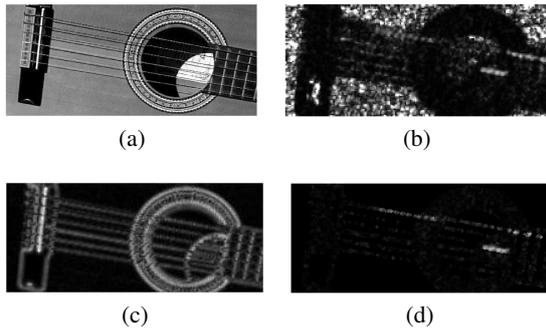
Figure 3: Quaternionic phase masked. (a) is an input video in which the first string of the guitar is subtly oscillating. (b) is the calculated quaternionic phase shift $\Delta\phi$ and (c) is its local amplitude. (d) is the result of masking (b) with (c).

threshold it.

$$M = \begin{cases} 1 & \text{if} \quad \beta \leq \frac{A_n}{A_q} \\ 0 & \text{if} \quad \beta > \frac{A_n}{A_q} \end{cases} \qquad (20)$$

where A_n is the calculated local amplitude of the image at frame n, A_q is the 95-percentile of the empirical distribution of the amplitudes along the video and β is a threshold selected by the user (See Section 4.2.3). The mask can be further refined using morphological opening. Finally, we mask the phase norm (Φ'_n) with M (as seen in Figure 3d). By masking the areas of low amplitude we have effectively selected the regions in which subtle motion can be detected.

3.3 Motion Spotting

Although the amplitude masking step aims to discard any area that could contribute with noisy data, some spurious pixels might get through this step. Thus, for each masked frame $\Delta\phi$, we select the values smaller than the 90-percentile. This is done to avoid outliers caused by noisy pixels which do not represent the subtle motion. From the selected pixels we calculate the average power:

$$P_n = \frac{1}{L} \sum_{l=0}^{L-1} |\Delta\phi_{n,l}|^2 \qquad (21)$$

where l is the index of the selected pixels and L is the total number of selected pixels. P_n is a one-dimensional signal which peaks or local maxima represents changes in the image sequence.

In order to distinguish relevant peaks (subtle motions) from local magnitude variations and background noise, we use a method to contrast the differences of P_n proposed by (Li et al., 2017). This method compares the differences of P_n within a specified interval. Since subtle motions might take more than 2 consecutive frames, we analyze micro-intervals of K

frames (an odd number bigger than 2). Then, for each current frame value, we subtract the average of the k-th frame value before the current frame and the k-th frame value after the current frame, where

$$k = \frac{1}{2}(K - 1) \qquad (22)$$

Thus, for the n-th value in the contrasted difference vector $C(\phi)$ is calculated by:

$$C_n = P_n - \frac{1}{2}(P_{n-k} + P_{n+k}) \qquad (23)$$

Finally, we select from C_n the peaks or local maxima that go over a threshold T and that are separated for at least K frames. The threshold is calculated as:

$$T = C_{median} + p \times (C_{max} - C_{median}) \qquad (24)$$

where C_{median} and C_{max} are the median and maximum value of C_k for the whole video and p is a percentage parameter in the range $[0,1]$.

4 RESULTS

Our proposed method allows an user to analyze subtle motions in videos. Although, the eulerian amplification methods are able to reveal imperceptible phenomena not previously visualized on video, our method goes further and is able to quantify this motion. Thus, in the following section, we decide to show some potential applications using the videos provided in the supplemental material of (Wadhwa et al., 2013; Wadhwa et al., 2014a).

4.1 Preliminary Evaluation

We select a video of a baby sleeping under a blanket in a cradle that, when it's magnified, amplifies the subtle movements of the baby breathing (See Figure 4a). We decided to test our method using the same filtering parameters suggested by the supplemental material of (Wadhwa et al., 2013; Wadhwa et al., 2014a). For the Riesz transform step we select the second level of the pyramid. We design a non-causal FIR bandpass FIR temporal filter of order 10 with passband between 0.04 and 0.4 Hz. We use a Gaussian Kernel K_ρ with standard deviation $\rho = 2$ for spatial filtering. For the amplitude masking step we select a threshold $\beta = 0.3$ (See Figure 4b). The segmented areas are transformed into a 1-D signal using mean power. As we can see in Figure 4c, the local maxima in the signal correspond to the moment when the breathing motion is at its peak, thus we can estimate the breathing patter of the baby.
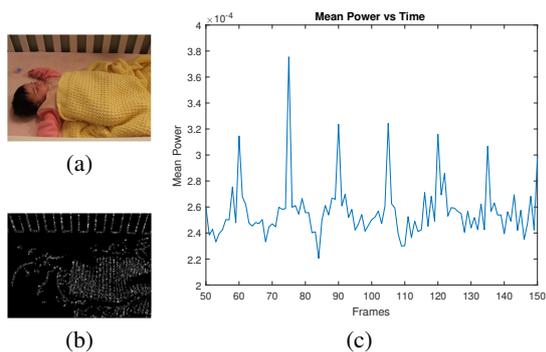
Figure 4: Subtle motion detection for breathing measurement. (a) is the a frame of a video of a baby sleeping, (b) is the masked phase norm and (c) is the obtained signal.

We also select a video of a drum, for which the motion amplification method, magnifies its vibrations (See Figure 5a). For this example we decided to design a filtering using a passband between 60 and 90 Hz instead of the narrow passband between 74 and 78 suggested by (Wadhwa et al., 2013). This video was recorded using a high speed camera at 1900 fps. For the Riesz transform step we select the third level of the pyramid. The following parameters are the ones suggested in the supplemental material of (Wadhwa et al., 2013; Wadhwa et al., 2014a). We design a non-causal FIR bandpass FIR temporal filter of order 44. with passband between 60 and 90 Hz. We use a Gaussian Kernel K_ρ with standard deviation $\rho = 2$ for spatial filtering. For the amplitude masking step we select a threshold $\beta = 0.15$ (See Figure 5b). The segmented areas are transformed into a 1-D signal using mean power. As we can see in Figure 5c, the detected vibration behaves like a combination of sinusoidal waves (which is expected since the drum is emitting acoustic waves). We can further analyze the signal spectrum using the fast Fourier transform. As we can see from Figure 5d, there are some frequency peaks outside the narrow spectrum considered by (Wadhwa et al., 2013) which could better characterize the frequency of the drum vibration.

4.2 Experiments

Although, the examples presented in the previous section show the potential of our method for subtle motion analysis, we would also like to measure its spotting accuracy and robustness. Thus, we have designed an experiment to compare the performance of our method compared with methods in the state of the art under different levels of noise.

4.2.1 Database

To the extent of our knowledge, there is not a public labeled subtle motion database available. Thus we decide to create our own database consisting of image sequences with subtle motions. We create 18 image sequences using real and artificially generated images which elements are for the most part static with the exception of one or two instances of subtle motion. We simulate the subtle motion by translating either an object in the image or the whole image by one pixel per frame during two or three frames. We also label the time when the subtle motion starts (onset) and ends (offset).

4.2.2 Method Comparison

We decide to compare our proposed method against other classical approaches for motion detection. First, the image is divided into a grid of equal-sized blocks. Then, we extract information from each block using three different feature descriptors: local binary pattern (LBP) (Ojala et al., 2002), histogram of oriented gradients (HOG) (Dalal and Triggs, 2005) and optical flow (OF) using the Lucas-Kanade method (Barron et al., 1994). Then we use the feature difference analysis method proposed by (Li et al., 2017) to compare the differences of the appearance-based features for each block within a specified interval. The difference between HOG and LBP histograms is calculated using the Chi-Squared (χ^2) distance. Then we sort the feature difference values from the blocks and calculate the mean of the greatest values that surpass the 80−percentile. Finally, we use the method to contrast the differences and the peak detection discussed in section 3.3 (Eq. 23 and 24 respectively).

4.2.3 Parameter Selection

We select the parameters for the spotting methods which will work in all videos without added noise. For our spotting method we select the second level of the Riesz pyramid. We design a FIR non-causal low-pass temporal filter with cutoff frequency of 30 Hz, corresponding to a filter of order 10. We use a Gaussian Kernel K_ρ with standard deviation $\rho = 2$ for spatial filtering. For the amplitude masking step we select a threshold $\beta = 0.25$.

For the LBP method we divide the image into a grid of 6×6 equal-sized blocks. The LBP descriptor has a radius of 3 pixels with 16 neighbors. For the HOG method we use the function *extractHOGFeatures* from Matlab. For each block produced by the function there are 2×2 cells of $[8, 8]$ pixels. For the OF method, we divide the image into a grid of 6×6 and compute the flow's amplitude[2].

---

[2]We don't extract the orientation since it doesn't provide
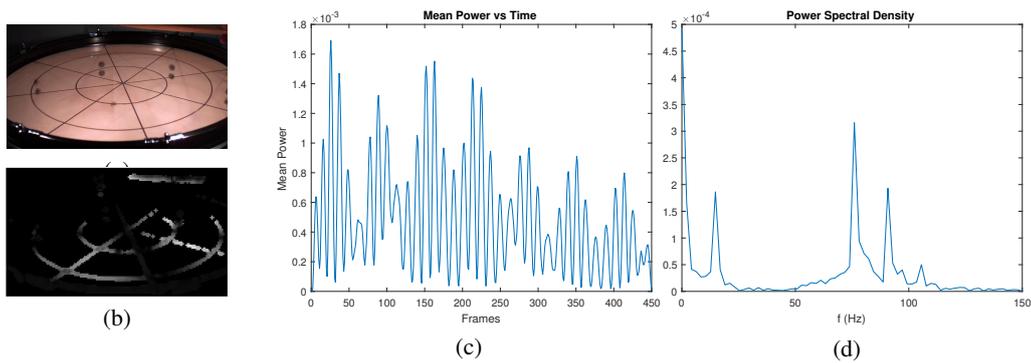
Figure 5: Subtle motion analysis for spectral analysis. (a) is the a frame of a video of a vibrating drum, (b) is the masked phase norm,(c) is the obtained sinusoidal signal and (d) is its magnitude spectrum.

4.2.4 Image Noise Measurement

Considering that subtle motions have low amplitude and, in some cases, they could be mistaken for noise, we decide to test the robustness of the motion detection methods in the presence of different levels of noise. We choose to test the videos under Gaussian additive noise and Salt and Pepper noise. In order to do a standard measurement of the noise among the different videos, we measure it using the peak signal-to-noise ratio (PSNR). PSNR is defined by the mean square error (MSE). Given a noise-free $u \times v$ monochrome image I and the image with added noise K, MSE is defined as:

$$MSE = \frac{1}{uv} \sum_{i=0}^{u-1} \sum_{j=0}^{v-1} [I(i,j) - K(i,j)]^2 \qquad (25)$$

The PSNR (in dB) is defined as:

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right) \qquad (26)$$

where MAX_I is the maximum possible pixel value of the image.

4.2.5 Evaluation Methodology

The first step is, for each video, to add a specific level of noise. Then we evaluate the methods accuracy by comparing all the detected peak frames in each method with the ground truth labels from each video in order to tell whether they are true or false positive subtle motions. Within a certain threshold level ($p = 0.75$ see Eq. 24), if one spotted peak is located within the frame range of [ONSET, OFFSET] of a labeled subtle motion video, the detected sequence is be considered as one true positive subtle motion.

_____

any important information about the instant when a subtle motion takes place.

Otherwise a penalization of a possible detected subtle motion (ψ frames) is counted as false positive. Since the noise is added randomly we repeat the test and measure the methods' accuracy 20 times.

We define the true positive rate (TPR) as the percentage of frames of correctly detected subtle motion divided by the total number of ground truth subtle motion frames in the database. The false positive rate (FPR) is calculated as the percentage of incorrectly spotted frames divided by the total number of non-subtle motion frames from all the image sequences. We evaluate the performance of the subtle motion detection methods by tracing curves with TPR and FPR as the y axis and PSNR as the x axis.

4.2.6 Experimental Results

The spotting results under different levels of Gaussian noise are presented in Figure 6. Our method has shown to have an equal or higher TPR in the presence of most levels of Gaussian noise compared to the other methods except when the PSNR is between 20 and 23 dB in which the OF method has a higher TPR. Similarly, the FPR of our method is equal or lower in most levels of Gaussian Noise except when the PSNR is between 20 and 23 dB. The LBP and HOG method had a lower TPR and a higher FPR compared to our method and the OF method.

The spotting results under different levels of density of Salt and Pepper noise are presented in Figure 7. Our method has shown to have an equal or higher TPR in the presence of most levels of Salt and Pepper noise compared to the other methods except when the PSNR is between 19 and 26 dB in which the OF method has a higher TPR. Similarly, the FPR of our method is equal or lower in most levels of Salt and Pepper Noise except when the PSNR is between 19 and 26 dB. The LBP and HOG method had a lower TPR and a higher FPR compared to our method and the OF method. However, the LBP method seems to

(a)

(b)
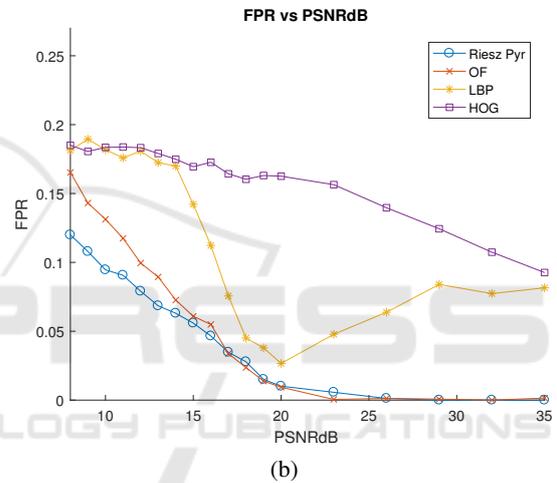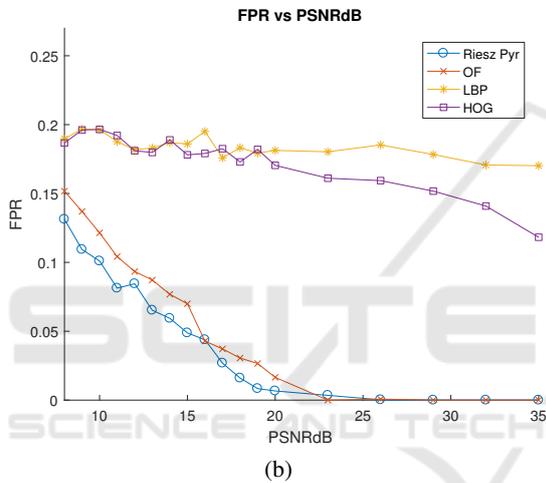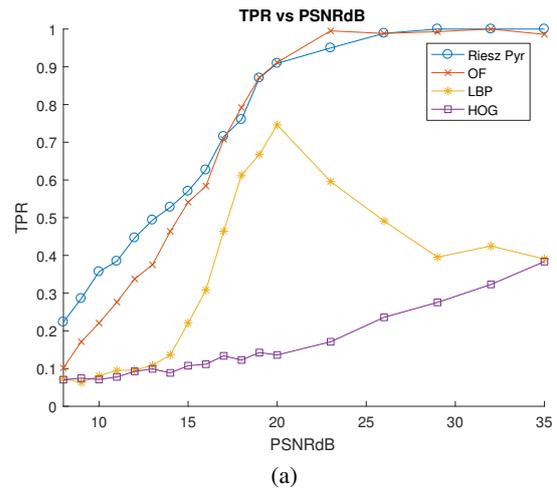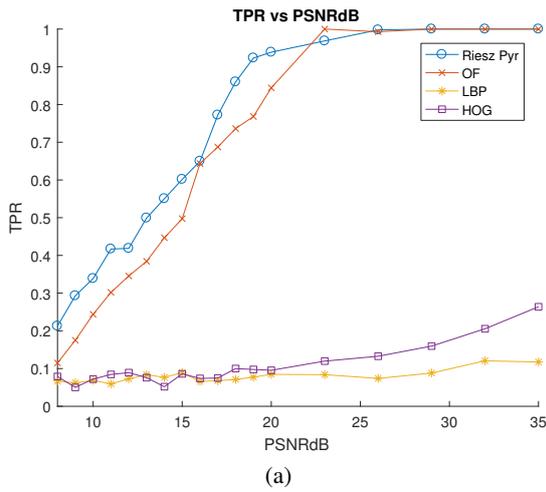
Figure 6: Performance curves of different subtle motion spotting techniques in presence of different levels of Gaussian noise.

(a)

(b)

Figure 7: Performance curves of different subtle motion spotting techniques in presence of different levels of Salt and Pepper noise.

have a better performance between 18 and 23 dB than HOG.

4.2.7 Discussion

The results in the previous section show that our method has, in general, a better spotting accuracy compared to the other methods tested. However, a closer examination of our database has shown that our method performs better in images with overall strong edges and its performance decreased in images with blurry edges. A possible explanation for this behavior is that our amplitude masking method (see section 3.2) aims to bypass areas of higher amplitude in which the phase noise is lower and discard areas of low amplitude in which the phase noise is higher. However, in images with blurry edges, the computed amplitude will be low all over the image and our system will end up bypassing areas of higher noise, thus, compromising our method's accuracy.

The performance of the OF method is comparable to our method. Since the Lucas-Kanade optical flow method was formulated under the temporal persistence assumption (motion remains small from frame to frame), it seems like an appropriate method for describing subtle motion (Kamate and Yilmazer, 2015). However, in videos with color gradients and certain levels of noise, the optical flow accuracy is decreased.

The accuracy of the HOG method was low because image gradients are sensible to noise. The accuracy of the LBP method was also low because LBP is sensitive to noise and sometimes may classify two or more different patterns falsely to the same class (Rassem and Khoo, 2014). However, the overall results suggest that our method better describes subtle motion than LBP and HOG descriptors.

5 CONCLUSIONS

We presented a subtle motion analysis and spotting method based on the Riesz pyramid. Our method adapted the quaternionic representation of the Riesz monogenic signal by proposing a new filtering scheme. We were also able to mask regions of interest where subtle motion might take place in order to reduce the effect of noise using the image amplitude. Furthermore, we illustrated the power of our subtle motion analysis method by briefly presenting a couple of potential real-life applications. After testing our method using our own database under different levels of Gaussian additive noise and salt and peper noise, we can conclude that our method surpasses other state of the art methods.

Due to the unavailability of a public labeled subtle motion database we had to test our experiments in a rather limited dataset. Further tests will require us to create or find a more complete database in order to obtain more statistically significant results.

The quaternionic representation of phase and orientation from the Riesz monogenic signal is a powerful tool that could potentially be exploited in the future for more focused applications like modal analysis, biomedical signals processing, and facial micro-expression spotting and recognition.

REFERENCES

Balakrishnan, G., Durand, F., and Guttag, J. (2013). Detecting Pulse from Head Motions in Video. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3430–3437.

Barron, J. L., Fleet, D. J., and Beauchemin, S. S. (1994). Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77.

Bharadwaj, S., Dhamecha, T. I., Vatsa, M., and Singh, R. (2013). Computationally Efficient Face Spoofing Detection with Motion Magnification. In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 105–110.

Chen, J. G., Wadhwa, N., Cha, Y.-J., Durand, F., Freeman, W. T., and Buyukozturk, O. (2015). Modal identification of simple structures with high-speed video using motion magnification. *Journal of Sound and Vibration*, 345:58–71.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893 vol. 1.

Irani, R., Nasrollahi, K., and Moeslund, T. B. (2014). Improved Pulse Detection from Head Motions Using DCT. Institute for Systems and Technologies of Information, Control and Communication.

Kamate, S. and Yilmazer, N. (2015). Application of Object Detection and Tracking Techniques for Unmanned Aerial Vehicles. *Procedia Computer Science*, 61:436–441.

Lee, J. and Shin, S. Y. (2002). General construction of time-domain filters for orientation data. *IEEE Transactions on Visualization and Computer Graphics*, 8(2):119–128.

Li, X., Hong, X., Moilanen, A., Huang, X., Pfister, T., Zhao, G., and Pietikainen, M. (2017). Towards Reading Hidden Emotions: A Comparative Study of Spontaneous Micro-expression Spotting and Recognition Methods. *IEEE Transactions on Affective Computing*, PP(99):1–1.

Liu, C., Torralba, A., Freeman, W. T., Durand, F., and Adelson, E. H. (2005). Motion Magnification. In *ACM SIGGRAPH 2005 Papers*, SIGGRAPH '05, pages 519–526, New York, NY, USA. ACM.

McLeod, A. J., Baxter, J. S. H., de Ribaupierre, S., and Peters, T. M. (2014). Motion magnification for endoscopic surgery. volume 9036, pages 90360C–90360C–8.

Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987.

Park, S. and Kim, D. (2009). Subtle facial expression recognition using motion magnification. *Pattern Recognition Letters*, 30(7):708–716.

Raghavendra, R., Avinash, M., Marcel, S., and Busch, C. (2015). Finger vein liveness detection using motion magnification. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–7.

Rassem, T. H. and Khoo, B. E. (2014). Completed local ternary pattern for rotation invariant texture classification. *The Scientific World Journal*, 2014.

Unser, M., Sage, D., and Ville, D. V. D. (2009). Multiresolution Monogenic Signal Analysis Using the Riesz-Laplace Wavelet Transform. *IEEE Transactions on Image Processing*, 18(11):2402–2418.

Wadhwa, N., Rubinstein, M., Durand, F., and Freeman, W. (2014a). Riesz pyramids for fast phase-based video magnification. In *2014 IEEE International Conference on Computational Photography (ICCP)*, pages 1–10.

Wadhwa, N., Rubinstein, M., Durand, F., and Freeman, W. T. (2013). Phase-based Video Motion Processing. *ACM Trans. Graph.*, 32(4):80:1–80:10.

Wadhwa, N., Rubinstein, M., Durand, F., and Freeman, W. T. (2014b). Quaternionic representation of the riesz pyramid for video magnification. Technical report.

Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., and Freeman, W. (2012). Eulerian Video Magnification for Revealing Subtle Changes in the World. *ACM Trans. Graph.*, 31(4):65:1–65:8.