Jerk-Aware Video Acceleration Magnification

Shoichiro Takeda Kazuki Okami Dan Mikami Megumi Isogai Hideaki Kimata
NTT Media Intelligence Laboratories 1-1 Hikarinooka, Yokosuka, Kanagawa, Japan
{shoichiro.takeda.us, kazuki.okami.ac, dan.mikami.vp, megumi.isogai.ks, hideaki.kimata.yu}@hco.ntt.co.jp

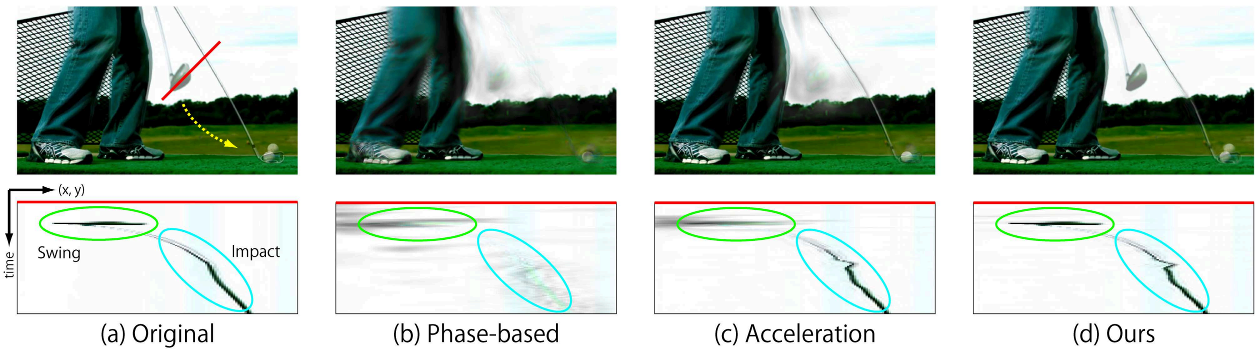

Figure 1: Sports use-case: visualizing the impact spread in the iron shaft. The yellow arrow depicts the golf swing along a trajectory. The top row shows 2 frames overlaid to indicate the swing phase and the impact phase of the ball. The bottom row shows the spatiotemporal slices along a single diagonal red line in the top of row of (a), and the green and cyan circles in them respectively indicate the swing phase and impact phase. (a) Original video. (b) Phase-based motion magnification [25]. (c) Video acceleration magnification [28]. (d) Our proposed jerk-aware video acceleration magnification. Our method only magnifies subtle deformation of the iron shaft without artifacts caused by quick swinging motions in other methods. See the supplementary material for the video results.

Abstract

Video magnification reveals subtle changes invisible to the naked eye, but such tiny yet meaningful changes are often hidden under large motions: small deformation of the muscles in doing sports, or tiny vibrations of strings in ukulele playing. For magnifying subtle changes under large motions, video acceleration magnification method has recently been proposed. This method magnifies subtle acceleration changes and ignores slow large motions. However, quick large motions severely distort this method. In this paper, we present a novel use of jerk to make the acceleration method robust to quick large motions. Jerk has been used to assess smoothness of time series data in the neuroscience and mechanical engineering fields. On the basis of our observation that subtle changes are smoother than quick large motions at temporal scale, we used jerk-based smoothness to design a jerk-aware filter that passes subtle changes only under quick large motions. By applying our filter to the acceleration method, we obtain impressive magnification results better than those obtained with state-of-the-art.

1. Introduction

The human visual system sometimes misses essential properties of dynamic objects because the clues lie within a very small world. For example, muscles and skin are slightly deformed in doing sports, drones sway to stabilize themselves in quick flight, and tiny vibrations of strings in instruments play wonderful sounds. However, these meaningful and attractive subtle changes are too small to see with the naked eye, and are often hidden under large motions.

In recent years, video magnification algorithms have been developed as a way to magnify and visualize such subtle motion changes, or color changes in video [27, 25, 26]. These methods adopt efficient displacement analysis called the Eulerian approach. This approach measures subtle image changes at a fixed position without object tracking. Eulerian video magnification achieves good results for magnifying the sway of a bridge, the breathing of an infant, and facial color changes due to blood circulation. However, as the Eulerian approach cannot distinguish between subtle changes and large ones, they perform only when objects and camera remain still.

To overcome this limitation, layer-based Eulerian video magnification techniques have been developed by Elgharib et al. [5] and Kooij et al. [13]. These techniques separate a target region of the above Eulerian magnification methods from background motions by manual segmentation [5] or depth layers [13]. They can magnify subtle changes in the presence of large motions, but they require burdensome interventions such as the need for human manipulation and preparing for an environment suitable for a depth camera. On the other hand, an impressive method called video acceleration magnification has recently been proposed by Zhang et al. [28]. This method is close to the original Eulerian approach [27], but only magnifies subtle acceleration changes of linear motions by assuming that the large motions are typically linear at temporal scale. This method can magnify subtle changes and ignores slow large motions, which can be approximated as linear ones, without these interventions. However, quick large motions severely distort this method due to their having non-linearity: golf swings, quick drone flights, and strumming in music play.

This paper presents a jerk-aware video acceleration magnification method for magnifying subtle changes in the presence of slow and quick large motions without the aforementioned interventions. Our method uses jerk to make the video acceleration technique [28] robust to quick large motions. On the basis of our observation that subtle changes are smoother than quick large motions at temporal scale, we consider that grasping the difference in the smoothness enables us to isolate subtle changes from quick large motions. For that purpose we focus on jerk, which has been used to evaluate smoothness of time series data in the neuroscience [6, 23, 19] and mechanical engineering fields [20]. In developing our method, we used jerk-based smoothness to design a novel filter (which we call a jerk-aware filter) that only passes subtle changes in the presence of quick large motions. By applying our filter to the acceleration method [28], we obtain impressive magnification results without artifacts caused by slow and quick large motions.

The contributions of this paper are below. 1) We propose a novel filter for passing subtle changes only in the presence of quick large motions. 2) We successfully apply jerk, a popular feature in other research fields, to grasp characteristics of motions and show the qualitative and quantitative effects it has on video magnification. 3) We demonstrate the practical insight our method provides and analyze the success obtained with it and its limitations.

2. Related Works

2.1. Lagrangian Approach

The concept of video magnification originates with Liu et al. [16] and their Lagrangian approach. This approach detects image changes by matching feature points between frames and estimates motions based on optical flow. They use this approach to segment background motions and motions of interest for magnification in input videos. Through spatial registration of background motions, interest motion can be re-estimated and magnified. However, the estimation of optical flow in this approach is expensive and has been researched as an unsolved problem [22, 14]. Unlike this approach, our method is based on Eulerian approaches that do not require object tracking explicitly and can magnify subtle motion changes, as well as subtle color changes.

2.2. Eulerian Approach

In comparison with the Lagrangian approach, Eulerian approaches analyze image changes at a fixed position over time without object tracking. Eulerian video magnification first decomposes image sequences into multi-scale pyramids, and then signals at each pyramid are temporally filtered to detect subtle changes to be magnified [27]. This method produces good color magnification results, but it can only support small amplification factors for motion magnification. To counter this problem, phase-based Eulerian video magnifications have been proposed [25, 26]. They build a complex-steerable pyramid [8, 21, 24] or a Riesz pyramid [26] to obtain phase variations of each pixel. The phase information corresponds to motion independent from color [7], which makes it possible to handle larger amplification factors for motion magnification. However, these Eulerian methods can only perform well when objects and cameras remain still, because they cannot distinguish subtle changes and other large motions.

For handling large motions, layer-based Eulerian video magnification techniques have been developed by Elgharib et al. [5] and Kooij et al. [13]. Elgharib et al. [5] require a user to select a region whose large motions are stabilized by homography transform. After the stabilization, subtle changes in the selected region are magnified by the above-mentioned Eulerian magnification methods. In the technique described by Kooij et al. [13], the region to be magnified is automatically selected by using a depth-weighted bilateral filter that detects all pixels at the same depth layer. However, these methods require human manipulation [5] or an environment suitable for a depth camera [13]; consequently, the methods are time consuming and error prone.

In contrast, Zhang et al. [28] tried to detect subtle changes in the presence of large motions without these additional requirements. By assuming that the large motions are approximately linear at temporal scale, they only magnify subtle acceleration changes of linear motions. This acceleration method shows good subtle motion and color magnification results in the presence of slow large motions that can be approximated as linear ones, but fails to ignore quick large motions due to their having non-linearity. Their method excessively magnifies quick large motions and produces noisy magnification results.

Using an approach similar to that described by Zhang et al. [28], we aim to magnify subtle changes in videos without the use of the aforementioned requirements. However, we had to deal with a completely different and more advanced problem: video magnification in the presence of slow and quick large motions.

3. Methods

We present jerk-aware video acceleration magnification method for magnifying subtle variations within slow and quick large motions without requiring additional resources. First, we will show that jerk is a useful index to handle quick large motions in video magnification. Second, we will describe how we designed a jerk-aware filter that passes subtle changes only under quick large motions by using jerk-based smoothness. Finally, we will show how we applied this filter to color acceleration magnification and motion acceleration magnification.

3.1. Jerk

Acceleration magnification [28] utilizes the fact that large motions are approximately linear at temporal scale, whereas our key idea is based on our observation that subtle changes depict smoother trajectories than quick large motions at temporal scale (Fig.2). We consider that grasping the difference in the smoothness better enables us to isolate subtle changes from quick large motions. Therefore, we focus on the characteristic feature called jerk.

Jerk is a third temporal derivative of displacement, and represents the rate of change in acceleration per unit of time. It is an effective index to assess steepness or smoothness of time series data. Its value becomes low during smooth changes but high during steep changes. It has been used in many research fields for assessing movements and trajectories [6, 23, 19, 20, 4]. In neuroscience, it has been used to model the trajectory of voluntary arm movements [6] and to assess the recovery of motor performance in stroke patients [23, 19]. In mechanical fields, the trajectories of the robot models with jerk restrictions make it possible to obtain smooth control [20]. Through these findings, we assume that subtle changes have a lower jerk value than quick large motions due to the smoothness. To check our hypothesis, we simply calculated the third temporal derivative of luminance intensity changes in a gun-shooting video (Fig.3). The result shows that static objects, such as the body or the arm, which may include invisible smooth subtle deformations, have lower jerk values than quick movement objects, such as gun and cartridge.

3.2. Jerk-Aware Filter

On the basis of our knowledge of jerk, we designed the jerk-aware filter. This filter is designed to have jerk-based smoothness so that it will pass subtle changes only and cut off quick large motions. To obtain the jerk-based smoothness, we first calculated the jerk value with a desired frequency component. Given input image signal $I(\boldsymbol{x}, t)$ at position $\boldsymbol{x}$ that denotes 2D pixel coordinates and time t, by referring to a previous study [28], we combined the third temporal derivative of $I(\boldsymbol{x}, t)$ and a Gaussian filter. Gaussian filter prevents spurious resolution [11] and its linearity

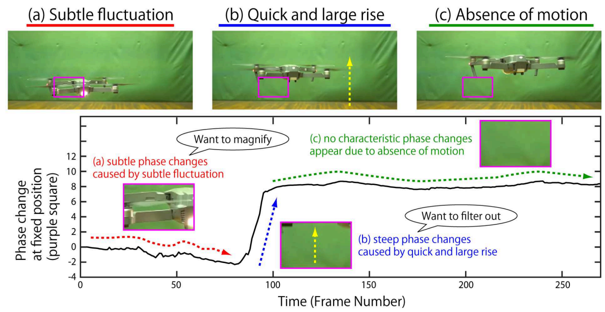

Figure 2: Our observation. Eulerian approaches [27, 25, 28] analyze image or motion changes in video as time series data of intensity or phase changes at a fixed position (purple square in (a), (b), and (c)). We observed that the time series data of subtle phase changes caused by subtle fluctuation of the drone (a) are smoother than those of steep ones caused by quick and large rise motion (b) because their magnitudes are very small.

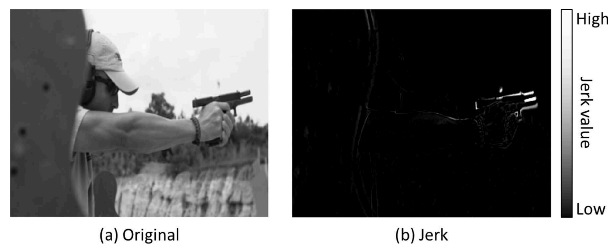

Figure 3: Gun-shooting video in luminance space (a) and jerk calculated by the temporal luminance intensity change (b). Jerk only responds to quick large motions such as gun blowback and gun cartridge release.

[12] permits its calculation of jerk with a Gaussian filter to be rewritten as:

$$\begin{aligned} Jerk_\sigma(\boldsymbol{x}, t) &= G_\sigma(t) \otimes \frac{\partial^3}{\partial t^3} I(\boldsymbol{x}, t) \\ &= \frac{\partial^3 G_\sigma(t)}{\partial t^3} \otimes I(\boldsymbol{x}, t), \end{aligned} \tag{1}$$

where $\otimes$ is convolution operator, $G_\sigma(t)$ is a Gaussian filter with variance σ^2 and $\frac{\partial^3}{\partial t^3}$ is the jerk calculation. Selecting the scale parameter σ of the Gaussian filter makes it possible to detect the jerk value with the desired frequency component [15, 18]. Therefore, we set the scale parameter σ of the Gaussian filter as: $\sigma = \frac{r}{4f\sqrt{2}}$, where r indicates the video frame rate and f is the desired frequency. Subsequently, the temporal window width of Gaussian filter is defined as $\frac{r}{4f}$.

Second, we transformed the jerk value into jerk-based smoothness that has a high value (close to 1) when smooth changes appear and a low value (close to 0) when no such changes appear as:

$$nJerk_\sigma(\boldsymbol{x}, t) = \frac{|Jerk_\sigma(\boldsymbol{x}, t)| - \min_{\boldsymbol{x},t} |Jerk_\sigma(\boldsymbol{x}, t)|}{\max_{\boldsymbol{x},t} |Jerk_\sigma(\boldsymbol{x}, t)| - \min_{\boldsymbol{x},t} |Jerk_\sigma(\boldsymbol{x}, t)|}, \tag{2}$$

$$smoothness_\sigma(\boldsymbol{x}, t) = 1 - nJerk_\sigma(\boldsymbol{x}, t). \tag{3}$$

Finally, to provide a filter capable of easily adjusting the weight of the smoothness value, we correct Eq. (3) on the basis of a hyper parameter $\beta (> 0)$ and obtain jerk-aware filter $JAF_\sigma(\boldsymbol{x}, t)$, which can selectively permeate subtle changes but cut off quick large motions:

$$JAF_\sigma(\boldsymbol{x}, t) = smoothness_\sigma(\boldsymbol{x}, t)^\beta. \tag{4}$$

3.2.1 Pyramid-based Correction

In video magnification algorithms, to reduce the unnatural appearance of synthetic videos in which subtle changes are magnified, the input image sequences are decomposed into multiple resolution pyramids at each frame before magnification processing [27, 25, 26, 5, 13, 28]. In this way, Eqs. (1) to (4) are simply transformed by taking pyramid level l into consideration. For example, Eq. (4) can be written as:

$$JAF^l_\sigma(\boldsymbol{x}, t) = smoothness^l_\sigma(\boldsymbol{x}, t)^\beta, \tag{5}$$

where l is a pyramid level.

However, given the pyramid decomposition processing, the meaning of our proposed filter $JAF^l_\sigma(\boldsymbol{x}, t)$ changes in accordance with the pyramid level. As mentioned in previous studies [17, 9], image sequences at higher level pyramids can handle large displacements, but their values decrease in proportion to the resolution at the pyramid level. This means that though our filters at higher level pyramids capture much quicker large motions, they are designed by lower jerk values.

Therefore, we define the filter correction on the basis of the down sampling factor $(0 < \lambda < 1)$ used to construct each pyramid level as:

$$JAF^l_{\sigma,\lambda}(\boldsymbol{x}, t) := JAF^l_\sigma(\boldsymbol{x}, t)^{1/\lambda}. \tag{6}$$

Furthermore, we consider that our proposed filter will need to be modified by a similar coarse-to-fine approach [17, 9]. As pyramids with higher levels observe image changes in wider space, they can accurately capture quick large motions and calculate correct jerk. On the other hand, quick large motions do not fit in the observation range at lower pyramid levels. This means that jerk at lower pyramid levels cannot reflect this essence, even if quick large motions occur. Therefore, it is necessary to propagate the information of the filter $JAF^l_{\sigma,\lambda}(\boldsymbol{x}, t)$ at a higher pyramid level to a filter at a lower pyramid level. We define this propagation correction as:

$$pJAF^l_{\sigma,\lambda}(\boldsymbol{x}, t) = \prod_{i=l}^{l+N} res(JAF^i_{\sigma,\lambda}(\boldsymbol{x}, t), l), \tag{7}$$

where N is the number of the pyramid level for this correction, and the function of $res(JAF^i_{\sigma,\lambda}(\boldsymbol{x}, t), l)$ resizes the filter size at the pyramid level i to that at pyramid level l with bicubic interpolation. Through this correction, we obtain a sophisticated jerk-aware filter $pJAF^l_{\sigma,\lambda}(\boldsymbol{x}, t)$ that can distinguish the essence of the difference between subtle and quick large changes.

3.3. Jerk-Aware Acceleration Magnification

3.3.1 Color Magnification

We present jerk-aware color acceleration magnification combining an acceleration technique and the jerk-aware filter. In the acceleration technique for color magnification [28], Gaussian pyramid used to decompose input signal $I(\boldsymbol{x}, t)$ to $I^l(\boldsymbol{x}, t)$. For detecting the subtle acceleration changes $B^l_f(\boldsymbol{x}, t)$ with a desired frequency f, the temporal acceleration filter $H_f(t)$ [28] is convolved to $I^l(\boldsymbol{x}, t)$ as:

$$B^l_f(\boldsymbol{x}, t) = H_f(t) \otimes I^l(\boldsymbol{x}, t). \tag{8}$$

After that, $B^l_f(\boldsymbol{x}, t)$ multiplied by the magnification factor α is added to $I^l(\boldsymbol{x}, t)$ for obtaining the synthesis signal $\hat{I}^l_f(\boldsymbol{x}, t)$, in which subtle color changes are magnified at each pyramid level:

$$\hat{I}^l_f(\boldsymbol{x}, t) = I^l(\boldsymbol{x}, t) + \alpha B^l_f(\boldsymbol{x}, t). \tag{9}$$

For details, see [28].

This synthesis signal includes excessive magnification of color changes due to quick large motions. To cut them off and keep subtle color changes, we apply the jerk-aware filter as below:

$$\hat{I}^l_f(\boldsymbol{x}, t) = I^l(\boldsymbol{x}, t) + \alpha(JAF^l_{\sigma,\lambda}(\boldsymbol{x}, t) \times B^l_f(\boldsymbol{x}, t)). \tag{10}$$

Note that as the color magnification method amplifies $I^l(\boldsymbol{x}, t)$ only in the third level of the pyramid [27, 28], we use the jerk-aware filter (Eq. 6), without propagation correction (Eq. 7). Through this process, our method produces good color magnification results without artifacts caused by slow and quick large motions.

3.3.2 Motion Magnification

For magnifying subtle motions, we used a phase-based acceleration technique [28]. This technique utilizes the local phase changes in video that represent local motion changes [7]. To obtain local phase information, this technique decomposes $I(\boldsymbol{x}, t)$ into a number of oriented frequency bands $R^l_{\omega,\theta}(\boldsymbol{x}, t)$ by applying complex steerable filter $\psi^l_{\omega,\theta}$, which contains a set of filters at various spatial scales ω and orientations θ at each pyramid level l. This equation can be written as:

$$\begin{aligned} R^l_{\omega,\theta}(\boldsymbol{x}, t) &= (I(\boldsymbol{x}, t) \otimes \psi^l_{\omega,\theta})(\boldsymbol{x}, t) \\ &= A^l_{\omega,\theta}(\boldsymbol{x}, t) e^{i\phi^l_{\omega,\theta}(\boldsymbol{x}, t)}. \end{aligned} \tag{11}$$

For detecting subtle acceleration phase information $C^l_{f,\omega,\theta}(\boldsymbol{x},t)$ with a desired frequency f, the temporal acceleration filter $H_f(t)$ [28] is convolved to $\phi^l_{\omega,\theta}(\boldsymbol{x},t)$ as:

$$C^l_{f,\omega,\theta}(\boldsymbol{x},t) = H_f(t) \otimes \phi^l_{\omega,\theta}(\boldsymbol{x},t). \tag{12}$$

For details, see [28].

However, this information includes quick large motions as well as subtle motion changes. To cut them off and keep subtle motion changes, we design a jerk-aware filter by using phase information. Using our method, we calculate the jerk on this local phase information as:

$$\begin{aligned} Jerk^l_{\sigma,\omega,\theta}(\boldsymbol{x},t) &= G_\sigma(t) \otimes \frac{\partial^3}{\partial t^3}\phi^l_{\omega,\theta}(\boldsymbol{x},t) \\ &= \frac{\partial^3 G_\sigma(t)}{\partial t^3} \otimes \phi^l_{\omega,\theta}(\boldsymbol{x},t). \end{aligned} \tag{13}$$

After that, we create a jerk-aware filter by using phase information with Eqs. (2)-(7); this filter applies $C^l_{f,\omega,\theta}(\boldsymbol{x},t)$ so as to only ignore quick large motions as:

$$\hat{C}^l_{f,\lambda,\omega,\theta}(\boldsymbol{x},t) = pJAF^l_{\sigma,\lambda,\omega,\theta}(\boldsymbol{x},t) \times C^l_{f,\omega,\theta}(\boldsymbol{x},t). \tag{14}$$

Finally, $\hat{C}^l_{f,\lambda,\omega,\theta}(\boldsymbol{x},t)$ multiplied by the magnification factor α is added to $\phi^l_{\omega,\theta}(\boldsymbol{x},t)$ for obtaining the synthesis phase information that subtle motion changes are magnified at each pyramid level and orientations as:

$$\hat{\phi}^l_{f,\lambda,\omega,\theta}(\boldsymbol{x},t) = \phi^l_{\omega,\theta}(\boldsymbol{x},t) + \alpha\hat{C}^l_{f,\lambda,\omega,\theta}(\boldsymbol{x},t). \tag{15}$$

Similar to [28], we use phase unwrapping [10] to correct unstable phase jumps due to the phase value being wrapped within the range of $[-\pi,\pi]$.

4. Results

4.1. Experimental Setup

To evaluate the effectiveness of our proposed method, we performed experiments on real videos as well as on synthetic ones with ground truth magnification. For real videos, we assessed the performance qualitatively. For synthetic ones, we assessed the performance quantitatively against ground truth in control experiments. We set the magnification factor α, the target frequency f to be magnified, and the hyper parameter β as given in table 1. We applied our proposed method to video sequences in YIQ color space. We show the video magnification results in the supplementary material.

Color Magnification. We used a Gaussian pyramid to decompose each video frame into multi-scales and magnified the intensity changes only on the third level of the pyramid. This approach is similar to [27, 28].

Video	α	f	fs	β	α (other)
Gun	10	20	480	0.3	8
Light bulb 1	25	10	600	0.0001	25
Golf [1]	20	2	60	0.8	12
Light bulb 2 [1]	40	2	160	20	40
Drone	25	2	30	1	18
Ukulele	25	40	240	1	18
Synthetic ball	35	10	60	0.0001-5	20
Eye [3]	40	10	500	1.5	30
Plate [2]	20	2	160	3	12

Table 1: Parameters for all videos: magnification factor α, target frequency f, sampling rate fs, and correction hyper parameter β. Gun and Light bulb 1 are from [28]; the others are new. The full sequences and results are available in the supplemental video.

Motion Magnification. To decompose each video frame into magnitude and phase information, we used a complex steerable pyramid [25] with half-octave bandwidth filters and eight orientations. We set the parameter N as 5 in propagation correction (Eq. 7) and this correction was done independently for each orientation.

4.2. Real Videos

4.2.1 Comparison with Color Magnification

We compared our jerk-aware color acceleration magnification technique with two state-of-the-art techniques, linear [27] and acceleration [28], which can perform color intensity magnification without user annotations or additional information in the same way as our technique does.

In Figure 4, we first show that our method did not have any negative effects on a video for which the acceleration method [28] produced good color magnification results in the presence of slow large motions.

Figure 5 shows light bulbs shattered by a bullet shot from a gun. The color variations in the bulb caused by electrical current changing are hardly visible without magnification (see the original in Fig. 5). Processing this video with a linear method [27] reveals color intensity changes, but creates intensity clipping artifacts. While an acceleration technique [28] succeeds in magnifying subtle color intensity changes clearly, it detects steep changes in color intensity due to quick flying of the transparent fragments of broken light bulbs and produces messy artifacts. On the other hand, our proposed method is able to magnify only subtle color intensity changes before and after the light bulbs shattering despite the quick flying of the transparent fragments.

4.2.2 Comparison with Motion Magnification

We compared our jerk-aware motion acceleration magnification technique with two state-of-the-art techniques.

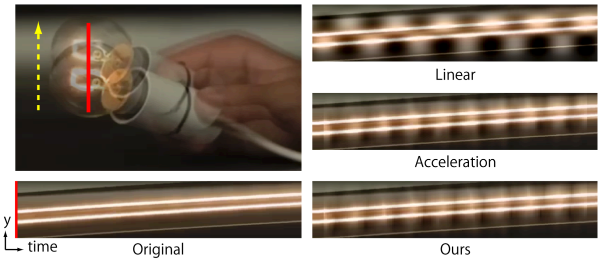

Figure 4: Color magnification in the presence of slow large motions; the light bulb moves upward slowly. Acceleration [28] and our method are well able to magnify the intensity for videos without artifacts caused by the effects of slow large motions.

These were phase-based [25] and phase-based acceleration [28] techniques that can perform motion magnification without any need for user annotations and additional information.

For Figures 1 and 6, we considered a sports use-case in which an athlete is shocked externally by hitting a ball or reacting to the recoil of gun. Our goal in these experiments was to ascertain what kind of impact they felt while they were engaged in sports activities.

Figure 1 shows the motion magnification results for a golf swing video to magnify the subtle deformation of the iron that occurs when the ball is hit. The phase-based method [25] induces large artifacts due to the quick swing. The acceleration method [28] can magnify subtle deformation of the iron that occurs when the ball is hit, but induces collapsing of the shape of the iron due to the quick large swing motion. Our proposed method can reveal this deformation by magnifying the acceleration motions [28] and ignores the effects of the quick large swing motion automatically.

Figure 6 shows a gun-shooting video with slow camera panning and quick gun recoil motion. We magnify the subtle deformation of the muscles and the skin due to the strong gun recoil. The phase-based method [25] induces large noise due to the slow camera panning and quick gun recoil motion. The acceleration method [28] can magnify subtle skin deformation of the arm in the presence of slow camera panning, but induces collapsing of the shape of the gun due to misdetected quick gun recoil motion. Our method only magnifies the skin deformation of the arm in the presence of slow camera panning and quick gun recoil motion.

Figure 7 shows an example of applying our proposed magnification to a mechanical use-case for managing the quality of mechanical stability. In this case, a drone is subtly fluctuating with various types of motions: slow parallel transition, quick rising, and 3D rotation of the body shift. Our proposed magnification is able to magnify only the subtle fluctuation of the drone without the effects of various large motions as mentioned above.

Figure 8 shows the case for a ukulele strumming video

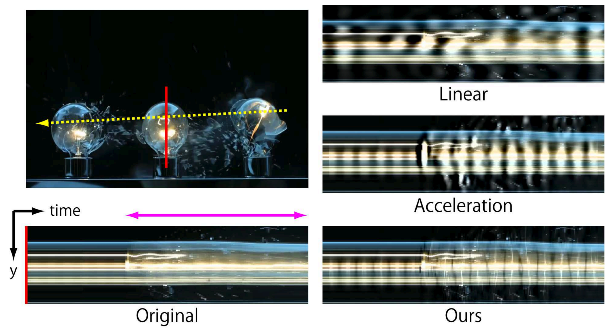

Figure 5: Color magnification in the presence of quick large motions; light bulbs shattered by a gun bullet are depicted by the yellow arrow. Our proposed method only magnifies subtle electrical current changing during quick flying of the transparent fragments of broken light bulbs (see the purple arrow time intervals).

in which quick hand motions appear many times. Our proposed magnification automatically ignores all the strumming hand motions and can magnify the subtle vibration of ukulele strings without user annotations and additional information.

4.3. Controlled Experiments

In Figure 9 (left), we show a 4-second synthetic ball video. We set the radius of the ball as 20 pixels. The ball has vertically subtle motion that is defined as $d_{subtle} = Asin\left(2\pi\frac{f}{fs}j\right)$ where $A = 0.5$ pixels, $f = 10$ cycles/frame, $f_s = 60$ frames/second, and j is the frame number. Moreover, the ball has vertical slow large motions on the screen from the top to bottom, with 0.5 pixels/frame. When the frame number j reaches 80 frames, the ball moves quickly and horizontally with $d_{quick} = A_{quick}sin\left(2\pi\frac{f_{quick}}{f_s}j\right)$ where $A_{quick} = 0 \sim 100$ pixels, $f_{quick} = 2$ cycles/frame, but after 20 frames the ball movement returns to what it was before.

To obtain the ground truth of the subtle motion magnification, we created a true magnification video while changing d_{subtle} to $d_{subtleMag} = \alpha Asin\left(2\pi\frac{f}{fs}j\right)$ where α is an amplification factor.

4.3.1 Effects of Quick Motion Magnitude

Our purpose is to assess the effectiveness of each motion magnification technique for magnifying subtle motions and ignoring quick horizontal large motions of the synthetic ball while changing the A_{quick} parameter relative to the ground truth video. We applied different magnification methods to the synthetic ball video. We fixed the parameters at $\alpha = 20$ and $\beta = 1$ for all methods except ours for which α was set to be 35.

Note that to investigate the effectiveness of our proposed pyramid-based correction, we prepared two magnification

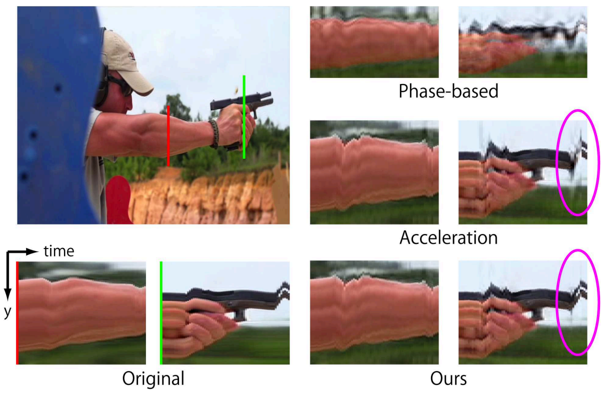

Figure 6: Sports use-case: visualizing the impact spread in an athlete's body. We show the spatiotemporal slices along a single red and green line (top-left). Our method magnifies deformations in the arm without the effect of camera panning and gun recoil motion (see the purple circles).

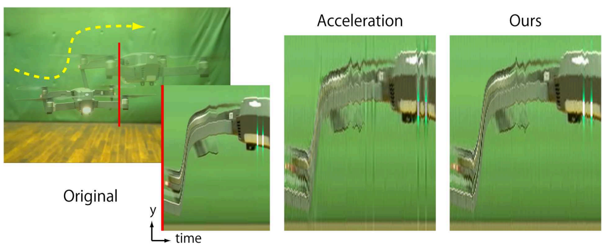

Figure 7: Mechanical use-case: analyzing the quality of mechanical stability. We show the spatiotemporal slices along a single vertical red line (left). Our proposed magnification is able to magnify and reveal the subtle fluctuation of the drone without artifacts caused by various types of large motions.

methods: a jerk method that uses a jerk-aware filter without down sampling correction (Eq. 6) or propagation correction (Eq. 7), and a jerk-down method that uses a jerk-aware filter without propagation correction (Eq. 7).

Figure 9 (right) shows the MSE (Mean Square Error) we obtained between each magnification result and the ground truth motion magnification as $A_{quick} = 100$, measured in each frame. For the phase-based method [25], we magnified the vibration in the frequency range of 9 to 11 Hz. This method incurs major errors in all frames due to slow and quick large motions. The acceleration method [28] magnifies subtle motions during the time slow large motions appear but produces artifacts during the time quick large motions appear. The jerk and jerk-down methods can cope with quick horizontal large motions fairly well, but our proposed jerk-aware method, despite its bigger amplification factor, best handles quick horizontal large motions while magnifying subtle motions that resemble the ground truth in all frames.

Figure 10 shows how a synthetic ball video behaves with different quick horizontal large motions A_{quick}. Here, at each A_{quick} we calculate the mean of MSE during

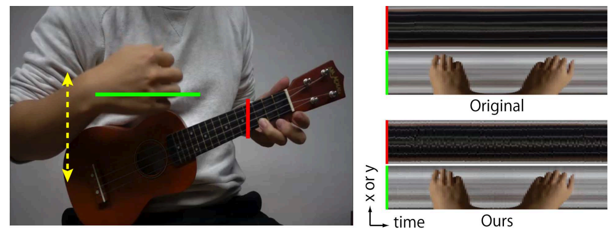

Figure 8: Music playing video: a ukulele being strummed with repetitive and quick hand motions. We show the spatiotemporal slices along a single red and green line (left). Our proposed method automatically ignores all the strumming hand motions and can magnify the subtle vibration of ukulele strings without hand manipulation or additional information.

the time subtle motions appear with slow large motions ($\mathrm{mMSE}_{\mathrm{subtle}}$) and the mean of MSE during the time quick horizontal large motions appear ($\mathrm{mMSE}_{\mathrm{quick}}$) relative to the ground truth. The phase-based method [25] takes large $\mathrm{mMSE}_{\mathrm{subtle}}$ and $\mathrm{mMSE}_{\mathrm{quick}}$ every A_{quick} values due to the effects of slow and quick large motions. Acceleration [28] and jerk methods keep $\mathrm{mMSE}_{\mathrm{subtle}}$ low, but the $\mathrm{mMSE}_{\mathrm{quick}}$ increases in proportion to A_{quick}. The jerk-down method keeps $\mathrm{mMSE}_{\mathrm{quick}}$ lower than the above three methods, but our proposed method is best for keeping $\mathrm{mMSE}_{\mathrm{subtle}}$ and $\mathrm{mMSE}_{\mathrm{quick}}$ low even if A_{quick} is increasing.

4.3.2 Effects of Pyramid-based Correction

To evaluate the effectiveness of our proposed pyramid-based correction, we applied the jerk method, the jerk-down method, and our proposed jerk-aware method to a synthetic ball video at the parameter of $A_{quick} = 100$ while changing the hyper parameter β.

In Figure 11, we show $\mathrm{mMSE}_{\mathrm{subtle}}$ and $\mathrm{mMSE}_{\mathrm{quick}}$ for every β relative to the ground truth video. Although the hyper parameter β increased in this case, the jerk method was not able to handle the quick large motions well (Fig.11 left). As we added down sampling correction to the jerk-aware filter, the jerk-down method correctly obtained the value of quick large motions in proportion to the pyramid level. Thus, this method can obtain lower $\mathrm{mMSE}_{\mathrm{subtle}}$ and $\mathrm{mMSE}_{\mathrm{quick}}$. However, it cannot completely ignore quick large motions; as can be seen from the center of Figure 11, $\mathrm{mMSE}_{\mathrm{quick}}$ does not reach 0. Our proposed method uses the jerk-aware filter with all pyramid corrections: down sampling correction and propagation correction. By integrating spatial information across the pyramid hierarchy through propagation correction, our method produces the result that the value of $\mathrm{mMSE}_{\mathrm{quick}}$ is almost 0 and $\mathrm{mMSE}_{\mathrm{subtle}}$ is kept low; this implies our proposed method best copes with quick large motions and magnifies subtle motions in the presence of slow large motions without user annotations or additional information (Fig.11 right).

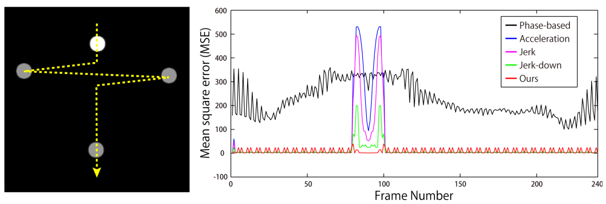

Figure 9: Left: synthetic ball video. Four frames are overlaid to indicate the ball trajectory that is depicted by the yellow arrow. The ball performs quick horizontal large motion between 80-100 frames. Right: MSE with the ground truth for each frame of the synthetic ball video. Smaller MSE is better. Our method outperforms all examined techniques.

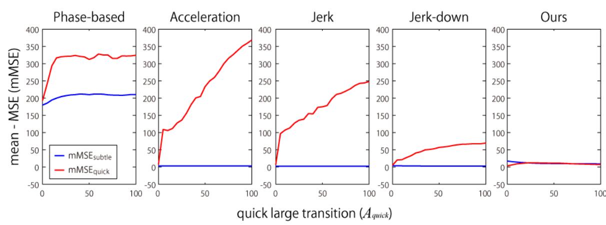

Figure 10: Mean-MSE during the appearance of subtle motions in the presence of slow large motions ($\text{mMSE}_{\text{subtle}}$) and mean-MSE during the appearance of quick large motions ($\text{mMSE}_{\text{quick}}$) with ground truth over different quick large transition A_{quick}. Our method handles quick large displacement with lower artifacts better than all the other techniques we cited.

5. Discussion and Limitations

While our proposed method expands the applicable range of video magnification by overcoming disturbance of quick large motions, it has some limitations.

Our jerk-aware filter can cut off quick large motions while permeating subtle changes. In other words, our approach is based on the assumption that subtle changes and quick large motions are spatio-temporally independent. For example, Figure 1 shows our method can magnify subtle deformations of the iron shaft that occur when the ball is hit, but cannot magnify them while the golf club is being swung. This is due to the fact that the subtle deformations mixed with quick large swing motions are regarded as being subject to removal by the jerk-aware filter. However, such motions are out of the range of our magnified targets. Even if we can magnify such subtle motions, quick large motions overwhelm these magnification results and we cannot follow them with the naked eye. We consider that a method for detecting and magnifying subtle changes mixed with quick large motions can be developed as a subject for future work.

Another limitation of our method is due to the assumption that all quick motions are large. If quick subtle changes exist, our method adversely affects them. As shown in Figure 9 to 11, our method slightly increases MSE during the time subtle changes appear. However, as we use magnification techniques for revealing subtle changes in videos, the

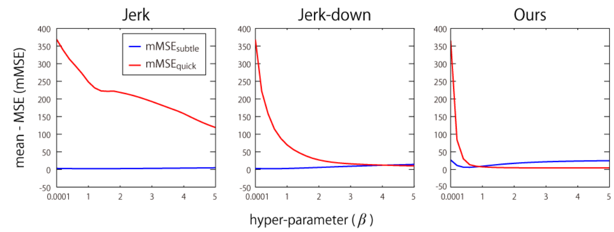

Figure 11: $\text{mMSE}_{\text{subtle}}$ and $\text{mMSE}_{\text{quick}}$ with ground truth over different hyper parameter β. Our proposed method best handles quick large motions while magnifying subtle motions; $\text{mMSE}_{\text{quick}}$ is almost 0 and $\text{mMSE}_{\text{subtle}}$ is kept low.

most important points are detecting and magnifying their amplitude. We define the amplitude of subtle changes as the distance between the top side of the crest and the bottom of the trough. Since our jerk-aware filter adopts a Gaussian filter, the surroundings of the crest and trough are smoothly connected. This indicates that our method does not affect the crest and the trough of subtle changes much and preserves their amplitude. Therefore, for example, compared with directly using the amplitude of changes to remove large motions, our method is superior in that only the quick large motions can be filtered out without worrying about the amplitude of the subtle changes being affected. However, an unresolved problem is that our proposed method slightly distorts magnified subtle changes; Figure 5 shows our color intensity magnification video slightly sharpens subtle light intensity changes compared with [28].

6. Conclusions

We present a method for magnifying subtle variations in the presence of slow and quick large motions without the need for user annotations or additional information.

Video acceleration magnification [28] has been recently proposed for magnifying subtle changes in the presence of large motions without requiring the additional resources. However, this method can only handle slow large motions. It excessively magnifies quick large motions and produces noisy magnification results.

To overcome this limitation, on the basis of our observation that subtle changes are smoother than quick large motions at temporal scale, we use jerk to evaluate the difference in smoothness between them. Using the jerk-based smoothness, we designed a jerk-aware filter that passes only subtle changes under quick large motions. Our proposed filter is able to dramatically cut off artifacts caused by quick large motions in a video acceleration method [28] and thus produces impressive magnification results.

We demonstrated our proposed method on synthetic and real videos and obtained better results than those obtained with other methods. Our method is highly applicable to sports use-cases (Golf and Gun), for mechanical stability quality control (Drone), and music entertainment (Ukulele).

References

[1] www.videoblocks.com. 5

[2] www.youtube.com/channel/uckyk8smoawb2fapgkh0kbiw. 5

[3] www.youtube.com/user/theslowmoguys. 5

[4] A. Datta, M. Shah, and N. D. V. Lobo. Person-on-person violence detection in video data. *ICPR*, 1:433–438, 2002. 3

[5] M. A. Elgharib, M. Hefeeda, F. Durand, and W. T. Freeman. Video magnification in presence of large motions. *CVPR*, 2015. 1, 2, 4

[6] T. Flash and N. Hogans. The coordination of arm movements: An experimentally confirmed mathematical model. *Journal of neuroscience*, 5(7):1688–1703, 1985. 2, 3

[7] D. J. Fleet and A. D. Jepson. Computation of component image velocity from local phase information. *International journal of computer vision*, 5(1):77–104, 1990. 2, 4

[8] W. T. Freeman, E. H. Adelson, et al. The design and use of steerable filters. *IEEE Transactions on Pattern analysis and machine intelligence*, 13(9):891–906, 1991. 2

[9] Y. Hu, R. Song, and Y. Li. Efficient coarse-to-fine patchmatch for large displacement optical flow. *CVPR*, 2016. 4

[10] D. Kitahara and I. Yamada. Algebraic phase unwrapping along the real axis: extensions and stabilizations. *Multidimensional Systems and Signal Processing*, 26(1):3–45, 2015. 5

[11] J. J. Koenderink. The structure of images. *Biological cybernetics*, 50(5):363–370, 1984. 3

[12] J. J. Koenderink and A. J. van Doorn. Generic neighborhood operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(6):597–605, 1992. 3

[13] J. F. Kooij and J. C. van Gemert. Depth-aware motion magnification. *ECCV*, 2016. 1, 2, 4

[14] T. Kroeger, R. Timofte, D. Dai, and L. Van Gool. Fast optical flow using dense inverse search. *ECCV*, 2016. 2

[15] T. Lindeberg. *Scale-Space Theory in Computer Vision*, volume 256. Springer Science & Business Media, 2013. 3

[16] C. Liu, A. Torralba, W. T. Freeman, F. Durand, and E. H. Adelson. Motion magnification. *SIGGRAPH*, 24(3), 2005. 2

[17] S. Meyer, O. Wang, H. Zimmer, M. Grosse, and A. Sorkine-Hornung. Phase-based frame interpolation for video. *CVPR*, 2015. 4

[18] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. *ICCV*, 1:525–531, 2001. 3

[19] R. Osu, K. Ota, T. Fujiwara, Y. Otaka, M. Kawato, and M. Liu. Quantifying the quality of hand movement in stroke patients through three-dimensional curvature. *Journal of NeuroEngineering and Rehabilitation*, 8(1):62, 2011. 2, 3

[20] A. Piazzi and A. Visioli. Global minimum-jerk trajectory planning of robot manipulators. *IEEE Trans. on Industrial Electronics*, 47(1):140–149, 2000. 2, 3

[21] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International journal of computer vision*, 40(1):49–70, 2000. 2

[22] J. Revaud, P. Weinzaepfel, Z. Harchaoui, C. Schmid, J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. S. E. Edge. Epicflow : Edge-preserving interpolation of correspondences for optical flow. *CVPR*, 2015. 2

[23] B. Rohrer, S. Fasoli, H. I. Krebs, R. Hughes, B. Volpe, W. R. Frontera, J. Stein, and N. Hogan. Movement smoothness changes during stroke recovery. *Journal of neuroscience*, 22(18):8297–8304, 2002. 2, 3

[24] E. P. Simoncelli and W. T. Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. *ICIP*, 3:444–447, 1995. 2

[25] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman. Phase-based video motion processing. *SIGGRAPH*, 32(4):80, 2013. 1, 2, 3, 4, 5, 6, 7

[26] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman. Riesz pyramids for fast phase-based video magnification. *ICCP*, 2014. 1, 2, 4

[27] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman. Eulerian video magnification for revealing subtle changes in the world. *SIGGRAPH*, 31(4), 2012. 1, 2, 3, 4, 5

[28] Y. Zhang, S. L. Pintea, and J. C. van Gemert. Video acceleration magnification. *CVPR*, 2017. 1, 2, 3, 4, 5, 6, 7, 8