Video Magnification in the Wild Using Fractional Anisotropy in Temporal Distribution

Shoichiro Takeda[1] Yasunori Akagi[2] Kazuki Okami[1] Megumi Isogai[1] Hideaki Kimata[1]
[1]NTT Media Intelligence Laboratories [2]NTT Service Evolution Laboratories
1-1 Hikarinooka, Yokosuka, Kanagawa, 239-0847 Japan
{shoichiro.takeda.us, yasunori.akagi.cu, kazuki.okami.ac, megumi.isogai.ks, hideaki.kimata.yu}@hco.ntt.co.jp

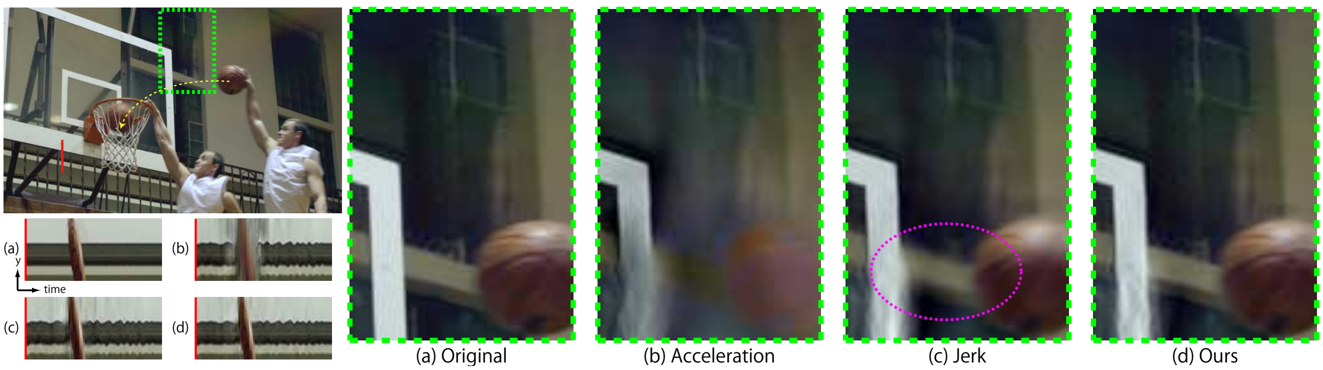

Figure 1: Top left: slam dunk video visualizing backboard deformations with ball trajectory by yellow arrow. Bottom left (a)-(d) show spatio-temporal slices along the single red line at top left. Right (a)-(d) show backgrounds in the green square at top left. (b) Acceleration method [24] produces messy artifacts due to quick ball motion. (c) Jerk method [17] magnifies meaningful subtle backboard deformations but misdetects non-meaningful subtle distortions of background window caused by photographic noise (purple circle). (d) On the contrary, our proposed method magnifies only meaningful subtle backboard deformations. See supplementary material for video results.

Abstract

Video magnification methods can magnify and reveal subtle changes invisible to the naked eye. However, in such subtle changes, meaningful ones caused by physical and natural phenomena are mixed with non-meaningful ones caused by photographic noise. Therefore, current methods often produce noisy and misleading magnification outputs due to the non-meaningful subtle changes. For detecting only meaningful subtle changes, several methods have been proposed but require human manipulations, additional resources, or input video scene limitations. In this paper, we present a novel method using fractional anisotropy (FA) to detect only meaningful subtle changes without the aforementioned requirements. FA has been used in neuroscience to evaluate anisotropic diffusion of water molecules in the body. On the basis of our observation that temporal distribution of meaningful subtle changes more clearly indicates anisotropic diffusion than that of non-meaningful ones, we used FA to design a fractional anisotropic filter that passes only meaningful subtle changes. Using the filter enables our method to obtain better and more impressive magnification results than those obtained with state-of-the-art methods.

1. Introduction

Physical and natural phenomena often cause meaningful and attractive changes in a small world. For example, muscles and skin are slightly deformed by impacts spreading through our bodies, materials deform elastically to absorb external force and thus prevent breakage and ensure safety, and strumming ukulele strings produce complicated string vibrations and result in generating wonderful sounds. However, these meaningful and attractive subtle changes are too small to see with the naked eye.

To visualize such subtle motion or color changes, video magnification methods have been proposed [22, 20, 21]. These methods are typically based on Eulerian approach which can measure subtle motion or color changes pixel by pixel. With recently developed spatio-temporal filtering [24, 17], current video magnification methods produce good results for magnifying and revealing only subtle changes under the presence of large motions of objects. However, in such subtle changes, meaningful ones caused by physical and natural phenomena are mixed with non-meaningful ones caused by noise introduced during the photographic process (i.e. low light levels, high sensor gain, short exposure time, and so on). Therefore, current methods often produce noisy magnification outputs due to the non-meaningful subtle changes; it is likely to lead users to incorrect insights and conclusions for a small world.

For detecting only meaningful subtle changes, layer-based methods have been developed [5, 7, 18]. These methods separate a target region to be magnified from the background by manual segmentation [5, 18] or depth layers [7]. They can magnify only meaningful subtle changes if they know where the changes are, but require burdensome interventions such as complicated manual operation and arranged environment for using a depth sensor.

In contrast, several methods magnify only meaningful subtle changes without the aforementioned interventions. By focusing on meaningful motion changes appearing around edges [14], edge-aware spatial smoothing methods have been proposed [20, 19]. These methods help to remove non-meaningful subtle changes in flat textured regions but have limitations in that they can not be applied to color magnification or to the removal of them around edges. Alternatively, principal component analysis (PCA) has been used for detecting only meaningful subtle changes [23] but its limitation is that meaningful subtle changes need to be larger than non-meaningful ones as the principal component in input video scenes.

This paper presents a novel video magnification method for detecting and magnifying only meaningful subtle color or motion changes under the presence of photographic noise, without additional interventions, resources, or input video scene limitations. On the basis of our observation that temporal distribution of meaningful subtle changes more clearly indicates anisotropic diffusion than that of non-meaningful ones caused by photographic noise, we considered that anisotropic diffusion in temporal distribution enables us to detect only meaningful subtle changes. Therefore, we focused on fractional anisotropy (FA), which is used in neuroscience to evaluate anisotropic diffusion of water molecules in the body for revealing the shape of tiny nerve cells [11, 2]. In developing our method, we used FA to design a novel filter, which we call a fractional anisotropic filter, that passes only meaningful subtle changes and ignores non-meaningful ones. Our method, in which the fractional anisotropic filter is applied to a state-of-the-art jerk-aware method [17], produces impressive color or motion magnification results in various input video scenes.

The main contributions of this paper are: (a) a successful application of fractional anisotropy, a popular measure in other research fields, to temporal analysis of video data, (b) a novel filter for passing only meaningful subtle color or motion changes under the presence of noise introduced during photographic process in various input video scenes, (c) a newly edge-aware regularization technique that incorporates strong normalization with hierarchical pyramid representation for refining motion information, and (d) showing of the qualitative and quantitative effects our method has on video magnification.

2. Related Work

2.1. Lagrangian Approach

Liu et al. [10] first presented the concept of video magnification with a Lagrangian approach. This approach uses optical flow to estimate the motion difference of frames. Through spatial registration of background motions, it can output a video in which subtle motion changes are magnified. However, estimating optical flow in this approach is computationally expensive and has been investigated as an unsolved problem [16, 8, 14].

2.2. Learning-based Approach

Recently, a learning-based approach that uses a deep neural network has been proposed [12]. The network takes two frames as input with an amplification factor and outputs a new frame, in which subtle changes are magnified. The implicitly learned motion representations in the network enable better noise handling to be achieved than with previous hand-crafted approaches. However, as this approach often misses subtle changes due to a strong dependence on a training dataset, the application range is still limited.

2.3. Eulerian Approach

Unlike the above approaches, our method is based on the most commonly used Eulerian approach [22, 20, 21, 24, 17]. Eulerian-based methods do not explicitly require object tracking and can detect subtle motion changes, as well as subtle color changes at a fixed position over time. They first decompose image sequences into Gaussian pyramids for color magnification [22, 24, 17] or complex-steerable pyramids (or Riesz pyramids) for motion magnification [20, 21, 24, 17], then the signals of each pixel at each pyramid are temporally filtered to detect subtle changes to be magnified. Although these methods can not distinguish subtle changes and other large object motions in pixels, recently developed spatio-temporal filtering techniques [24, 17] can isolate subtle changes from the large motions and selectively magnify them. However, in such subtle changes, meaningful ones caused by physical and natural phenomena are mixed with non-meaningful ones caused by photographic noise. Therefore, current methods often produce noisy and misleading magnification outputs due to the non-meaningful subtle changes.

For detecting only meaningful subtle changes, layer-based methods have been developed [5, 7, 18]. Elgharib et al. [5] and Verma et al. [18] require a user to select a region whose subtle changes are magnified. Kooij et al. [7] proposed a depth-weighted bilateral steerable filter for automatically selecting a region to be magnified at the same depth layer. After the magnification process, these methods synthesize the magnified region and other regions to output final results. They can magnify only meaningful subtle changes under the presence of photographic noise if a user knows where they are. However, they require complicated human interventions [5, 18] or an arranged environment suitable for a depth sensor [7]. Consequently, these methods are time consuming and error prone.

In contrast, several methods magnify meaningful subtle changes under the presence of photographic noise without the aforementioned interventions. By focusing on meaningful motion changes appearing around edges [14], Wadhwa et al. [20] applied an edge-weighted Gaussian filter (EWG) to motion changes spatially in each image pyramid, and Verma et al. [19] used a local Laplacian filter (LLP) [13] to improve pyramid decomposition in video magnification methods in terms of edges and details. These methods help to remove non-meaningful subtle changes in flat textured regions but have limitations in that they can not be applied to color magnification or to the removal of them around edges. Alternatively, Wu et al. [23] adopted PCA to video magnification as a pre-processing approach. This method can magnify only meaningful subtle changes in video sequences, but for enabling PCA to work well, it has a limitation that meaningful subtle changes need to be larger than non-meaningful ones as the principal component in input video scenes.

Our proposed method is more advanced than the aforementioned methods because it can magnify not only meaningful subtle motion changes but also color changes under the presence of photographic noise without the additional requirements or input video scene limitations.

3. Methods

We now present the details of our method. First, we explain our problem definition and why FA is a useful index to distinguish meaningful subtle changes and non-meaningful ones caused by photographic noise. Second, we describe how we designed our fractional anisotropic filter that passes only meaningful subtle changes and ignore non-meaningful ones. Finally, we show how we applied this filter to the current color or motion magnification method. We also present our edge-aware regularization in the motion magnification subsection.

3.1. Problem Definition

Given an input image signal $I(\boldsymbol{x}, t)$ at an image position $\boldsymbol{x}$ that denotes 2D pixel coordinates and a time t, video magnification methods [22, 20, 24, 17] attempt to detect subtle changes $B(\boldsymbol{x}, t)$. However, such subtle changes are often contaminated by photographic noise as

$$B(\boldsymbol{x}, t) = \hat{B}(\boldsymbol{x}, t) + \tilde{B}(\boldsymbol{x}, t), \tag{1}$$

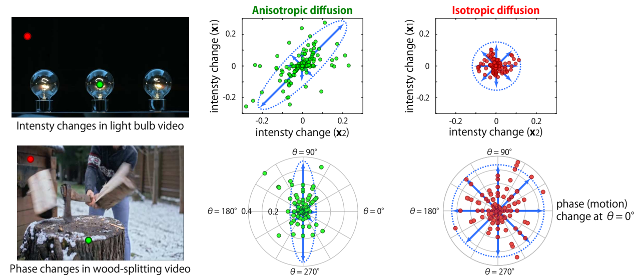

Figure 2: The temporal distributions of subtle intensity (top) and phase changes that represent motion (bottom). When the meaningful subtle intensity changes appear, they are correlated between neighboring pixels (top green), but do not when photographic noise only appears (top red). The meaningful subtle phase changes occur in a vertical direction (bottom green) but in no direction if they are not meaningful (bottom red). We noticed that temporal distribution of meaningful subtle changes more clearly indicates anisotropic diffusion than that of non-meaningful ones caused by photographic noise (blue arrow representing the trend).

where $\hat{B}(\boldsymbol{x}, t)$ is meaningful subtle changes and $\tilde{B}(\boldsymbol{x}, t)$ is non-meaningful ones caused by photographic noise. Therefore, current methods often produce noisy and misleading magnification outputs due to the $\tilde{B}(\boldsymbol{x}, t)$.

3.2. Fractional Anisotropy

Our key idea is based on our observation that temporal distribution of meaningful subtle changes more clearly indicates anisotropic diffusion than that of non-meaningful ones because they are subject to the regularity of nature (Fig.2). We considered that anisotropic diffusion in temporal distribution enables us to detect only meaningful subtle changes and focused on an index called fractional anisotropy (FA).

FA is used in neuroscience to evaluate anisotropic diffusion of water molecules in the body [11, 2], and its definition is based on the diffusion equation as

$$f(\boldsymbol{g}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{D}|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{g}^\top \boldsymbol{D}^{-1}\boldsymbol{g}\right), \tag{2}$$

where $f(\boldsymbol{g})$ is a probability distribution of water molecules in directions $\boldsymbol{g} \in \mathbb{R}^d$, and $\boldsymbol{D}$ is a positive semi-definite matrix that represents diffusion strength of the distribution $f(\boldsymbol{g})$ along or between directions $\boldsymbol{g}$. To the best of our knowledge, on the basis of $f(\boldsymbol{g})$, FA is defined in 3D case but we generalize it for multi-dimensional case as

$$FA := \sqrt{\frac{d}{d-1}} \cdot \frac{\sqrt{\sum_{i=1}^{d}(\lambda_i - \bar{\lambda})^2}}{\sqrt{\sum_{i=1}^{d}\lambda_i^2}}, \tag{3}$$

where $(\lambda_1, ..., \lambda_d)$ are eigenvalues of $\boldsymbol{D}$ and $\bar{\lambda} = \frac{1}{d}\sum_{i=1}^{d}\lambda_i$. The eigenvalues of $\boldsymbol{D}$ indicate diffusion strength to the direction of eigenvectors in original directions $\boldsymbol{g}$. $\sqrt{\frac{d}{d-1}}$ normalizes the FA value between 0 and 1.

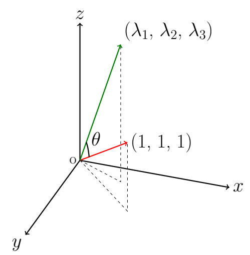

Figure 3: Intuitive interpretation of FA in 3D case. FA is proportional to sinθ, where θ is the angle between two vectors, $(\lambda_1, \lambda_2, \lambda_3)$ which are eigenvalues of $\boldsymbol{D}$, $(1, 1, 1)$. If all the eigenvalues are equal such as $(3, 3, 3)$, which means isotropic diffusion, θ is 0 and the FA value is 0. If only one eigenvalue is high such as $(5, 0, 0)$, which means anisotropic diffusion, θ is maximum and the FA value is 1.

Moreover, we found intuitive interpretation of FA as follows: let θ be the angle between two vectors, $(\lambda_1, ..., \lambda_d)$, $(1, ..., 1) \in \mathbb{R}^d$, we can rewrite the definition of FA (Eq.3) as

$$FA = \sqrt{\frac{d}{d-1}} \cdot \sin\theta. \tag{4}$$

For proof of this, see the supplementary material. This equation implies that FA purely evaluates the degree of match between the eigenvalues without depending on the magnitude of them. Since the positive semi-definite matrix of $\boldsymbol{D}$ makes all the eigenvalues positive, if only one eigenvalue is high, which means anisotropic diffusion, θ is maximum and FA value is 1, but if all the eigenvalues are equal, which means isotropic diffusion, θ is 0 and FA value is 0 (Fig.3).

In neuroscience fields, it is known that nerve axons have high FA values due to anisotropic diffusion of water molecules along their long stick structures, but if their axonal structures injury occurs due to such as a traffic accident or a neural disease, the probability distribution of water molecules in the injured area indicates isotropic diffusion and the FA value becomes lower [11, 2]. Thus, a changing of FA value sensitively assesses the loss or recovery process of the shape of nerve cells in humans or animals.

From these findings, we considered that FA values strongly respond to meaningful subtle changes compared with non-meaningful ones, due to the anisotropic diffusion in temporal distribution of meaningful ones. To visualize our hypothesis, we show FA values estimated by temporal distribution of subtle phase changes in ukulele-playing video (Fig.4). Figure 4 indicates that the FA value is higher when the meaningful subtle phase changes appear, such as hand swaying and vibrations of ukulele strings.

3.3. Fractional Anisotropic Filter

On the basis of our knowledge of FA, we designed a fractional anisotropic filter. This filter is designed using FA

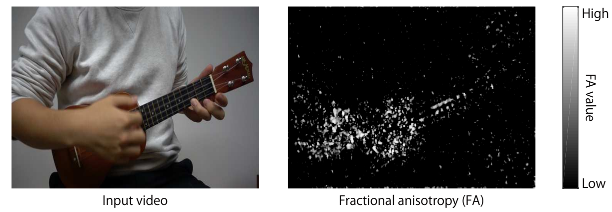

Figure 4: Visualizing fractional anisotropy (FA) values. FA values are estimated by temporal distribution of subtle phase changes and are high when the meaningful subtle phase changes appear such as hand swaying and vibrations of ukulele strings.

estimated from diffusion in temporal distribution of subtle changes so that it will pass only meaningful subtle color or motion changes, which have high FA values.

First, we get FA value $FA(\boldsymbol{x}, t)$ as follows. Given an image patch $\mathcal{P}_{\boldsymbol{x}} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{h\times w}\}$ centered at $\boldsymbol{x}$ for a time period $T = \{t_1, \ldots, t_N\}$ centered at t, let $\boldsymbol{y}_{t_j}$ be a $(h\times w)$-dimensional vector $[B(\boldsymbol{x}_1, t_j), \ldots, B(\boldsymbol{x}_{h\times w}, t_j)]^\top$ that represents subtle changes in $\mathcal{P}_{\boldsymbol{x}}$ at a time t_j. We assume that the N vectors $\boldsymbol{y}_{t_1}, \ldots, \boldsymbol{y}_{t_N}$ are i.i.d samples from a temporal distribution $f(\boldsymbol{y})$ defined as

$$f(\boldsymbol{y}) = \frac{1}{(2\pi)^{h\times w/2}|\boldsymbol{D}|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{D}^{-1}\boldsymbol{y}\right). \tag{5}$$

Using maximum likelihood estimation method, we estimate $\boldsymbol{D}$ representing diffusion strength of the temporal distribution $f(\boldsymbol{y})$ along or between the image positions in $\mathcal{P}_{\boldsymbol{x}}$ as

$$\boldsymbol{D} = \mathrm{cov}\left([\boldsymbol{y}_{t_1}, ..., \boldsymbol{y}_{t_N}]\right), \tag{6}$$

where $\mathrm{cov}(X)$ is the variance-covariance matrix of X. Then, we get $FA(\boldsymbol{x}, t)$ by using Eq.(3) for Eq.(6).

After calculating $FA(\boldsymbol{x}, t)$, we design the fractional anisotropic filter $FAF(\boldsymbol{x}, t)$ with a weight γ for adjusting the filter response as

$$FAF_{\sigma,\gamma}(\boldsymbol{x}, t) = (Norm(G_\sigma \otimes FA(\boldsymbol{x}, t)))^\gamma, \tag{7}$$

where $\otimes$ is a convolution operator, G_σ is a 2D Gaussian filter with a variance σ^2 to smooth filter responses, and $Norm(X)$ normalizes X value from 0 to 1. This filter has a high value only when anisotropic diffusion in temporal distribution of subtle changes appears, which means it can pass only meaningful subtle changes and ignores non-meaningful ones.

3.4. Video Color Magnification in the Wild

We present a novel color magnification method in which the proposed fractional anisotropic filter is applied to the jerk-aware color magnification method [17]. The jerk method decomposes input video signal $I(\boldsymbol{x}, t)$ into $I^l(\boldsymbol{x}, t)$ at pyramid level l using a Gaussian pyramid and detects subtle changes $B^l_f(\boldsymbol{x}, t)$ with a desired frequency f as

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

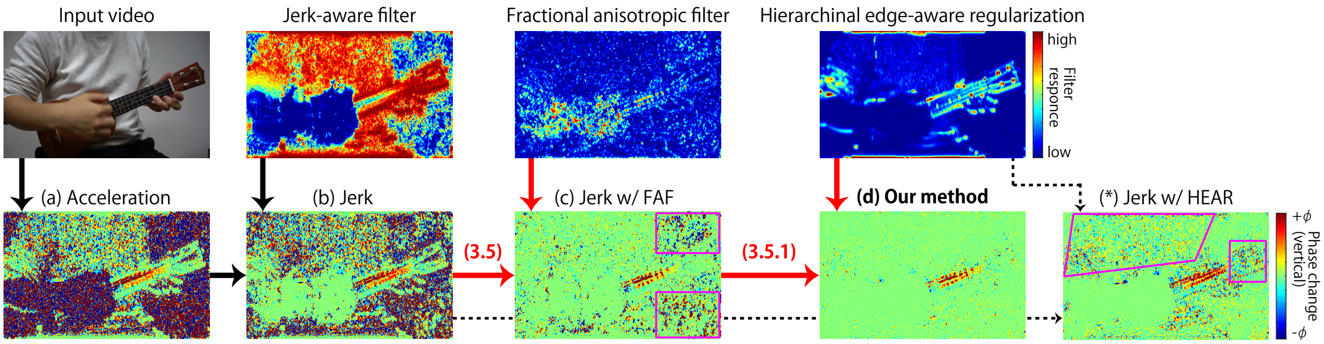

Figure 5: Our motion magnification method. (a) Acceleration method [24] misdetects quick hand strumming. (b) Jerk method [17] ignores the quick motion but misdetects non-meaningful subtle phase changes caused by photographic noise. (c) By using fractional anisotropic filter (FAF), we can detect only meaningful subtle phase changes of the ukulele strings but slightly misdetects non-meaningful ones in flat textured areas (purple quadrangles). (d) Our method further applies hierarchical edge-aware regularization (HEAR) to refine them. (*) Using only HEAR is insufficient for complex areas (purple quadrangles). These results indicate both of FAF and HEAR are needed.

where $H_f(t)$ is the temporal acceleration filter [24], $JAF^l_f(\boldsymbol{x},t)$ is the jerk-aware filter [17], and $\circ$ is an element-wise product. For details, see [17].

However, these subtle changes often include non-meaningful ones caused by photographic noise. To detect only meaningful subtle color changes, we design a fractional anisotropic filter $FAF^l_{f,\sigma,\gamma}(\boldsymbol{x},t)$ from $B^l_f(\boldsymbol{x},t)$ through Eqs.(5)-(7). After this designing, we obtain the color magnification signals $\hat{I}^l_f(\boldsymbol{x},t)$ with the amplification factor α at each pyramid level l as

$$\hat{I}^l_f(\boldsymbol{x},t) = I^l(\boldsymbol{x},t) + \alpha(FAF^l_{f,\sigma,\gamma}(\boldsymbol{x},t) \circ B^l_f(\boldsymbol{x},t)). \tag{9}$$

This process enables us to produce good color magnification results under the presence of photographic noise.

3.5. Video Motion Magnification in the Wild

For magnifying subtle motions, we used the jerk-aware phase-based method [17]. This method is based on the use of the local phase changes in video that represent local motion changes [6]. To obtain local phase information, complex steerable filters $\psi^l_{\omega,\theta}$, which are sets of filters with a motion orientation θ at each spatial scale ω and pyramid level l, are applied to $I(\boldsymbol{x},t)$ as

$$\psi^l_{\omega,\theta} \otimes I(\boldsymbol{x},t) = A^l_\omega(\boldsymbol{x},\theta,t)\mathrm{e}^{i\phi^l_\omega(\boldsymbol{x},\theta,t)}, \tag{10}$$

where $A^l_\omega(\boldsymbol{x},\theta,t)$ is amplitude and $\phi^l_\omega(\boldsymbol{x},\theta,t)$ is phase.

After obtaining the phase information, we detect subtle phase changes $C^l_{f,\omega}(\boldsymbol{x},\theta,t)$ with a desired frequency f as

$$C^l_{f,\omega}(\boldsymbol{x},\theta,t) = pJAF^l_{f,\omega}(\boldsymbol{x},\theta,t) \circ \left(H_f(t) \otimes \phi^l_\omega(\boldsymbol{x},\theta,t)\right), \tag{11}$$

where $pJAF^l_{f,\omega}(\boldsymbol{x},\theta,t)$ is the jerk-aware filter with pyramid correction [17]. For details, see [17].

However, these subtle phase changes include non-meaningful ones. To detect only the meaningful ones, we design the fractional anisotropic filter $FAF^l_{f,\omega,\sigma,\gamma}(\boldsymbol{x},t)$ about the subtle phase changes $C^l_{f,\omega}(\boldsymbol{x},\theta,t)$ as follows.

We consider an image patch $\mathcal{P}_{\boldsymbol{x}}$ for a time period T such as that described in Section 3.3 and orientations $\Theta = \{\theta_1,\dots,\theta_M\}$. Since phase changes occur along or between the orientations in Θ and the subtle phase changes $C^l_{f,\omega}(\boldsymbol{x},\theta,t)$ are similar within the image patch $\mathcal{P}_{\boldsymbol{x}}$, let $\boldsymbol{y}_{\boldsymbol{x}_i,t_j}$ be a M-dimensional vector $\left[C^l_{f,\omega}(\boldsymbol{x}_i,\theta_1,t_j),\dots,C^l_{f,\omega}(\boldsymbol{x}_i,\theta_M,t_j)\right]^\top$ that represents the subtle phase changes in a position $\boldsymbol{x}_i \in \mathcal{P}_{\boldsymbol{x}}$ at a time t_j. We assume that the $N \times h \times w$ vectors $\boldsymbol{y}_{\boldsymbol{x}_1,t_1},\dots,\boldsymbol{y}_{\boldsymbol{x}_{h\times w},t_N}$ are i.i.d samples from a temporal distribution $f(\boldsymbol{y})$ defined as Eq.(5). Then, using Eqs.(6)-(7), we design the phase-based fractional anisotropic $FAF^l_{f,\omega,\sigma,\gamma}(\boldsymbol{x},t)$.

After this designing, we can obtain the synthesis phase information in which meaningful subtle phase changes are only magnified as

$$\hat{\phi}^l_\omega(\boldsymbol{x},\theta,t) = \phi^l_\omega(\boldsymbol{x},\theta,t) + \alpha(FAF^l_{f,\omega,\sigma,\gamma}(\boldsymbol{x},t) \circ C^l_{f,\omega}(\boldsymbol{x},\theta,t)). \tag{12}$$

Figure 5 (c) shows that $FAF^l_{f,\omega,\sigma,\gamma}(\boldsymbol{x},t)$ can pass only meaningful phase subtle changes of ukulele strings and ignore non-meaningful ones caused by photographic noise. However, as the reliability of phase changes are low in flat textured areas [20, 19], $FAF^l_{f,\omega,\sigma,\gamma}(\boldsymbol{x},t)$ has errors and slightly misdetects non-meaningful subtle phase changes in the flat areas (Fig. 5 (c), purple quadrangles).

3.5.1 Hierarchical Edge-Aware Regularization

For refining the subtle phase changes, we use amplitude information A^l_ω in the same way as the previous technique [20]. However, since this technique [20] uses only amplitude information at each pyramid level l, we develop hierarchical amplitude correction via z-transform as

$$\hat{A}^l_\omega = \max_{-N_l \leqq i \leqq N_l}(\mathcal{Z}(A^l_\omega), res(\mathcal{Z}(A^{l+i}_\omega),l)), \tag{13}$$

where N_l is the number of the pyramid level l used for this, $\mathcal{Z}(A)$ converts A into z-scores that are comparable between each pyramid level l, and $res(A^{l+i}, l)$ resizes the amplitude information at the pyramid level $l+i$ to that at pyramid level l with bicubic interpolation.

Furthermore, as the previous technique [20] adopts amplitude-based smoothing but is weak in regularizing flat textured areas due to its use of smoothing alone, we propose a new strong regularization technique $HEAR_\sigma(\boldsymbol{x}, \theta, t)$ as

$$HEAR_\sigma(\boldsymbol{x}, \theta, t) = Norm(G_\sigma \otimes \hat{A}^l_\omega(\boldsymbol{x}, \theta, t)), \quad (14)$$

where $Norm(X)$ and G_σ are the same functions in Eq.(7).

By applying this regularization to Eq.(12), we can refine meaningful subtle phase changes as shown in Figure 5 (d).

4. Results

4.1. Experimental Setup

To evaluate the usefulness of our proposed method, which magnifies only meaningful subtle changes under the presence of photographic noise, we conducted experiments on real videos and synthetic ones with ground-truth magnification. We assessed the performance qualitatively for real videos and quantitatively against ground-truth for synthetic ones. We set the parameters for each experiment as listed in Table 1, but σ in Eqs.(7, 14) are equal to the spatial filter widths used to construct image pyramids. We show all the magnification results in the supplementary material.

Color Magnification. We used a Gaussian pyramid to decompose each video frame into multi-scales and magnified the green color intensity changes on the fifth pyramid level.

Motion Magnification. We performed each method in YIQ color space. To obtain amplitude and phase information from input video, we used a complex steerable pyramid [20] with half-octave bandwidth filters and 8 orientations. We set parameter N in jerk method [17] as 5 and N_l in Eq.(13) as 2. For designing a fractional anisotropic filter (Eq.(7)), we set the size of $\mathcal{P}_{\boldsymbol{x}}$ as 5x5, and T as the same sampling time used to detect subtle changes with the target frequency f.

4.2. Real Videos

We compared our proposed method with two state-of-the-art methods, acceleration [24] and jerk methods [17], both of which can perform color or motion magnification without user annotations or additional information in the same way as our method.

4.2.1 Comparison with Color Magnification

Figure 6 illustrates subtle face color changes due to blood flow through the face of a stationary man. Processing this video with acceleration [24] or jerk methods [17] succeeds in magnifying meaningful subtle face color changes on the face, but it also misdetects and magnifies non-meaningful background color fluctuations caused by photographic noise. In contrast, our proposed method magnifies only meaningful subtle face color changes.

Video	α	f	fs	β	γ	source
Slam dunk	200	2	120	1	2	[1]
Ukulele	260	40	240	1	5	[17]
Face	180	0.5	60	0.001	3	[22]
Wood	230	2	120	3	2	[1]
Gun	100	20	480	0.5	1	[17]
Tennis	180	10	600	1	1	[1]
Synthetic ball	100	10	60	1	2	-
Golf	80	2	60	0.8	2	[1]
Drone	300	2	60	1	3	[17]

Table 1: Parameters for all videos: amplification factor α in our method (this parameter in other methods was adjusted to magnify meaningful subtle changes as much as ours), target frequency f, sampling rate fs, large motions suppression parameter β in the jerk method [17], and hyper parameter γ in Eq.(7).

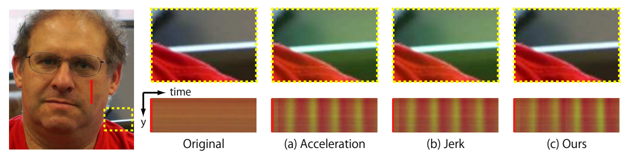

Figure 6: Color magnification at blood flow through the face of a stationary man. Our proposed method magnifies only meaningful subtle face color changes (bottom), while acceleration [24] and jerk methods [17] misdetect and magnify non-meaningful background color fluctuations caused by photographic noise (top).

4.2.2 Comparison with Motion Magnification

Figure 1 shows the motion magnification results from a basketball video, to magnify and reveal the subtle deformations of the backboard when trying to absorb the impact of a slam dunk for preventing breakage. Acceleration method [24] does not work well due to the misdetection of the quick ball motion. Jerk method [17] magnifies meaningful subtle deformation of the backboard but also misdetects non-meaningful subtle shape collapses of background window caused by photographic noise. In contrast, our proposed method magnifies only meaningful subtle deformations of the backboard without the effects of noise.

Figure 7 shows a video sequence on the ability of a wood-splitting stand to absorb the shock from a hand axe for preventing injury. Acceleration method [24] produces messy result due to the quick downswing of the hand axe. Jerk method [17] can magnify subtle deformations of the wood-splitting stand but produces pixel intensity disturbances due to non-meaningful background fluctuations

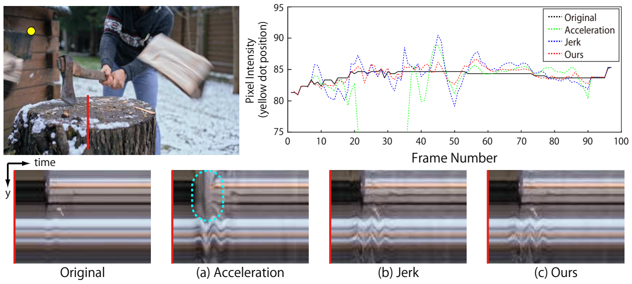

Figure 7: Wood-splitting video: visualizing deformations of a wood-splitting stand. The graph shows pixel intensity changes at yellow dot in top left. Our proposed method magnifies only meaningful subtle deformations of the wood-splitting stand, while acceleration method [24] misdetects the quick downswing of hand axe (cyan circle) and jerk method [17] produces pixel intensity disturbance due to non-meaningful background fluctuations (graph).

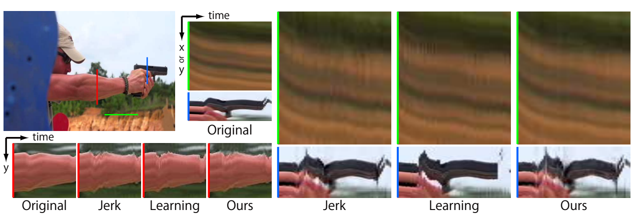

Figure 8: Gun-shooting video: visualizing gun-shooting impact spreading throughout body. Our method magnifies only meaningful subtle arm deformations (left bottom) but jerk method [17] misdetects background distortions caused by photographic noise (right top) and learning method [12] induces disappearance of the tip of the gun due to quick gun recoil motions (right bottom).

caused by photographic noise. Our method magnifies only meaningful subtle deformations of the wood-splitting stand under the presence of photographic noise.

Figure 8 shows a gun-shooting video. In this video, we also tested a state-of-the-art learning method [12] with a $5\times$ dynamic mode. Jerk method [17] misdetects distortions of background caused by photographic noise. Learning method [12] also misdetects them slightly and induces disappearance of the tip of the gun due to quick gun recoil motions. Our method magnifies only meaningful subtle deformations of muscles and skin due to the gun-shooting impact spreading throughout the body.

Figure 9 shows a ball-hitting video with magnification of impact spreading throughout a tennis racket. Acceleration method [24] produces racket shape collapse due to the quick swing motion. Jerk method [17] magnifies subtle racket deformations when the ball is hit but induces pixel intensity disturbances due to non-meaningful background fluctuations caused by photographic noise. In contrast, our method magnifies only meaningful deformations related to sport activities under the presence of photographic noise.

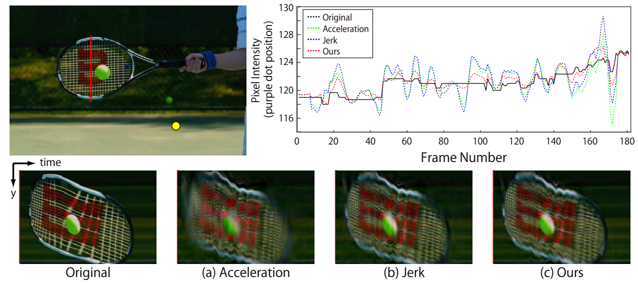

Figure 9: Tennis video: visualizing impact spreading throughout a tennis racket. Our method magnifies only meaningful subtle tennis racket deformations, but acceleration [24] and jerk methods [17] produce pixel intensity disturbance due to non-meaningful background fluctuations caused by photographic noise (graph).

4.3. Controlled Experiments

In this section, we quantitatively assess the effectiveness of our method using peak signal-to-noise ratio (PSNR) between magnified synthetic video by each magnification method and the ground-truth. Figure 10 (top left) shows a 4-second synthetic ball video with background texture from the Describable Textures Dataset [4]. The ball has vertical **meaningful** subtle motions defined as $d = 0.5\cdot\sin(2\pi\frac{f}{f_s}j)$, where j is the frame number. When j reaches 80 frames, the ball moves quickly and horizontally as $d_q = 100\cdot\sin(2\pi\frac{2}{f_s}j)$, but after 20 frames the ball movement returns to what it was before. Moreover, Gaussian noise with an average of 0 and standard deviation σ_n of 0–0.1 was added to only the background in videos as the photographic noise that causes **non-meaningful** subtle motions. To obtain the ground-truth of meaningful subtle motion magnification, we created magnification videos while changing d to $5\cdot d$.

Note that to investigate the effectiveness of our proposed method precisely, we prepared five additional methods: a jerk method with EWG proposed by [20], a jerk method with PCA, a jerk method with FAF, a jerk method with No-hierarchical edge-aware regularization as $EAR_\sigma(\boldsymbol{x},\boldsymbol{\theta},t) = \mathcal{N}(G_\sigma \otimes A^l_\omega(\boldsymbol{x},\boldsymbol{\theta},t))$, and a jerk method with HEAR.

Figure 10 right shows PSNR in each area and each background, at the real noise level $\sigma_n = 0.005$ estimated by [9]. In the ball area, Eulerian-based methods [19, 23], acceleration method [24] and learning method [12] suffer from handling quick motion and produce low PSNR, but all jerk based methods that contain our proposed method magnify only meaningful subtle motion and have high PSNR except for jerk method with PCA, which can not magnify meaningful ones due to large non-meaningful ones regarded as a principal component. On the other hand, in the noise area, jerk method produces very low PSNR due to non-meaningful ones magnified by the large amplification factor compared with acceleration method [24]. Jerk method with EWG [20], PCA, our proposed FAF, and HEAR ignore non-meaningful ones and increase PSNR compared with jerk

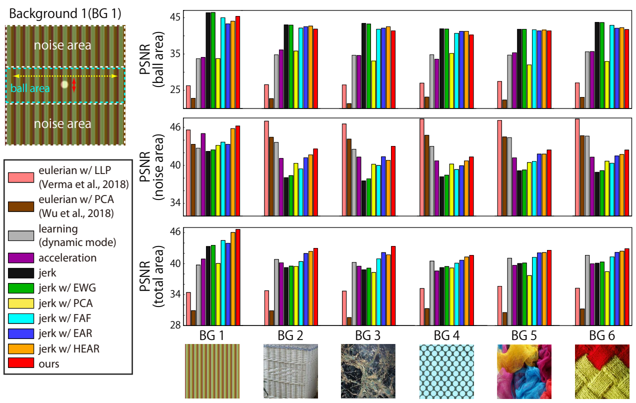

Figure 10: Left: synthetic ball video with background. The ball has **meaningful** subtle motion (red arrow) and quick motion (yellow arrow). Noise is added only to background and causes **non-meaningful** subtle motion. Right: PSNR at $\sigma_n = 0.005$. Our proposed method magnifies only meaningful subtle ball motions under the presence of noise and has the highest PSNR in the total area despite the complex background textures.

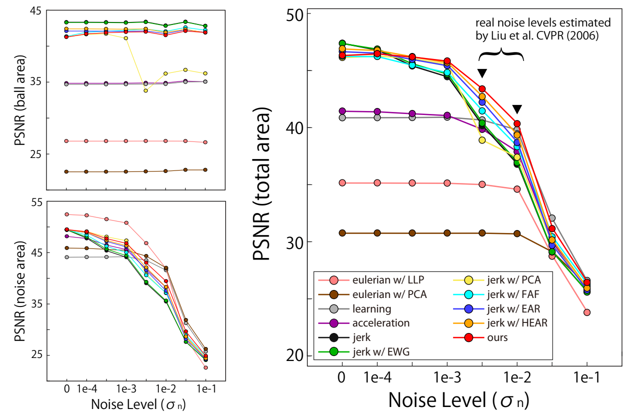

Figure 11: The effect of noise variance σ_n on PSNR for all the background videos on average. In the total area, our proposed method resists noise increase and has the highest PSNR in the real noise situations [9].

method [17] but all of these are insufficient. Our proposed method, which considers anisotropic diffusion in temporal distribution by FAF and hierarchical amplitude information by HEAR, ignores non-meaningful ones very well and has high PSNR in the noise area. After all, our proposed method magnifies only meaningful subtle ball motion under the presence of noise and has the highest PSNR in the total area despite the complex background textures.

Figure 11 shows the effect of noise variance σ_n on the average of PSNR for all the background videos. In the ball area, each magnification method maintains almost the same PSNR. However, jerk method with PCA can not do so because the principal component in video is switched from meaningful subtle motions to non-meaningful ones when $\sigma_n = 0.005$. In the noise area, PSNR in all methods gets lower in proportion to the noise increase. However, if we compare each magnification method for the total area, our proposed method resists the effect of noise increase and has the highest PSNR in the real noise situations ($\sigma_n = 0.005, 0.01$). Thus, our method produces the best meaningful and non-misleading magnification results.

5. Discussion and Limitations

Our proposed method expands the applicable range of video magnification by detecting and magnifying only meaningful subtle changes under the presence of photographic noise but has some limitations below.

Our proposed fractional anisotropic filter can detect only meaningful subtle changes, but it relies on the assumption that the temporal distribution of non-meaningful ones caused by photographic noise indicates isotropic diffusion. In real videos, such a characteristic like Gaussian distribution often occurs but other ones also need to be considered: gamma, exponential, uniform, impulse, and so on [3]. Thus, we should handle such characteristics to expand the applicable range of video magnification in future work.

If an input video size is large, our method has slow running time due to the eigen-decomposition at each position, time, and pyramid level. If one wants to precisely reveal meaningful subtle changes and show the results, our method should be used to prevent magnified non-meaningful changes that may be misleading. However, a faster algorithm for our method needs to be developed.

Moreover, empirical estimation of covariance in FA of our method is not robust to outliers under the Gaussian assumption in Eq.(2). To increase the robustness, we consider that a minimum covariance determinant approach [15] can be useful. Even so, we should develop a simple and principled approach as a substitute for using FA in future work.

6. Conclusions

We proposed a novel video magnification method for detecting and magnifying only meaningful subtle changes under the presence of noise introduced during photographic process, without the user annotations, additional information, or input video scene limitations that previous methods required [5, 18, 7, 20, 19, 23]. In developing our method, we presented a novel use of the index in neuroscience called fractional anisotropy to detect only meaningful subtle changes, and a hierarchical edge-aware regularization to refine motion representations. Our proposed method detects only meaningful subtle changes and ignores non-meaningful ones caused by photographic noise, and produces impressive magnification results exceeding those obtained with state-of-the-art methods. The results we obtained demonstrate that we succeeded in expanding the applicable range of "video magnification in the wild."

References

[1] www.videoblocks.com. 6

[2] Andrew L. Alexander, Jee Eun Lee, Mariana Lazar, and Aaron S. Field. Diffusion tensor imaging of the brain. *Neurotherapeutics*, 4(3):316–329, Jul 2007. 2, 3, 4

[3] Ajay Boyat and Brijendra Joshi. A review paper: Noise models in digital image processing. *Signal and Image Processing : An International Journal (SIPIJ)*, 2015. 8

[4] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, , and A. Vedaldi. Describing textures in the wild. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 7

[5] Mohamed A. Elgharib, Mohamed Hefeeda, Frédo Durand, and William T. Freeman. Video magnification in presence of large motions. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 2, 3, 8

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

[7] Julian FP Kooij and Jan C van Gemert. Depth-aware motion magnification. In *The European Conference on Computer Vision (ECCV)*, 2016. 2, 3, 8

[8] Till Kroeger, Radu Timofte, Dengxin Dai, and Luc Van Gool. Fast optical flow using dense inverse search. In *The European Conference on Computer Vision (ECCV)*, 2016. 2

[9] Ce Liu, William T. Freeman, Richard Szeliski, and Sing Bing Kang. Noise estimation from a single image. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006. 7, 8

[10] Ce Liu, Antonio Torralba, William T. Freeman, Frédo Durand, and Edward H. Adelson. Motion magnification. *SIGGRAPH*, 2005. 2

[11] Susumu Mori and Jiangyang Zhang. Principles of diffusion tensor imaging and its application to basic neuroscience research. *Neuron*, 51:527–539, 2006. 2, 3, 4

[12] Tae-Hyun Oh, Ronnachai Jaroensri, Changil Kim, Mohamed Elgharib, Frédo Durand, William T Freeman, and Wojciech Matusik. Learning-based video motion magnification. In *The European Conference on Computer Vision (ECCV)*, 2018. 2, 7

[13] Sylvain Paris, Samuel W. Hasinoff, and Jan Kautz. Local laplacian filters: Edge-aware image processing with a laplacian pyramid. *ACM Transactions on Graphics (ACM TOG)*. 3

[14] Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui, Cordelia Schmid, Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui, and Cordelia Schmid Epicflow Edge. Epicflow : Edge-preserving interpolation of correspondences for optical flow. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 2, 3

[15] Peter J. Rousseeuw. Least median of squares regression. *Journal of the American Statistical Association*, 1984. 8

[16] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 2

[17] Shoichiro Takeda, Kazuki Okami, Dan Mikami, Megumi Isogai, and Hideaki Kimata. Jerk-aware video acceleration magnification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1, 2, 3, 4, 5, 6, 7, 8

[18] Maninsha Verma and Shanmuganathan Raman. Interest region based motion magnification. *ICIAP*, 2017. 2, 3, 8

[19] Manisha Verma and Shanmuganathan Raman. Edge-aware spatial filtering-based motion magnification. In *Proceedings of 2nd International Conference on Computer Vision & Image Processing*, pages 117–128, Singapore, 2018. Springer Singapore. 2, 3, 5, 7, 8

[20] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T Freeman. Phase-based video motion processing. *SIGGRAPH*, 2013. 1, 2, 3, 5, 6, 7, 8

[21] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T. Freeman. Riesz pyramids for fast phase-based video magnification. In *The IEEE International Conference on Computational Photography (ICCP)*, 2014. 1, 2

[22] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. *SIGGRAPH*, 2012. 1, 2, 3, 6

[23] Xiu Wu, Xuezhi Yang, Jing Jin, and Zhao Yang. Pca-based magnification method for revealing small signals in video. *Signal, Image and Video Processing*, 12:1293–1299, 2018. 2, 3, 7, 8

[24] Yichao Zhang, Silvia L. Pintea, and Jan C. van Gemert. Video acceleration magnification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 1, 2, 3, 5, 6, 7