Learning-based Axial Video Motion Magnification

Kwon Byung-Ki[1] Oh Hyun-Bin[2] Kim Jun-Seong[2]
Hyunwoo Ha[2] Tae-Hyun Oh[1,2,3]

[1] Graduate School of AI, POSTECH
[2] Department of Electrical Engineering, POSTECH
[3] Institute for Convergence Research and Education in Advanced Technology, Yonsei University
`{byungki.kwon, hyunbinoh, junseong.kim, hyunwooha, taehyun}@postech.ac.kr`

Abstract. Video motion magnification amplifies invisible small motions to be perceptible, which provides humans with a spatially dense and holistic understanding of small motions in the scene of interest. This is based on the premise that magnifying small motions enhances the legibility of motions. In the real world, however, vibrating objects often possess convoluted systems that have complex natural frequencies, modes, and directions. Existing motion magnification often fails to improve legibility since the intricate motions still retain complex characteristics even after being magnified, which likely distracts us from analyzing them. In this work, we focus on improving legibility by proposing a new concept, *axial* video motion magnification, which magnifies decomposed motions along the user-specified direction. Axial video motion magnification can be applied to various applications where motions of specific axes are critical, by providing simplified and easily readable motion information. To achieve this, we propose a novel Motion Separation Module that enables the disentangling and magnifying of motion representation along axes of interest. Furthermore, we build a new synthetic training dataset for our task that is generalized to real data. Our proposed method improves the legibility of resulting motions along certain axes by adding a new feature: user controllability. In addition, axial video motion magnification is a more generalized concept; thus, our method can be directly adapted to the *generic* motion magnification and achieves favorable performance against competing methods. The code and dataset are available on our project page: `https://axial-momag.github.io/axial-momag/`.

Keywords: Motion magnification · Motion analysis · Video processing

1 Introduction

Small motions often convey important signals in practical applications, *e.g.*, building structure health monitoring [4–8, 27], machinery fault detection [28, 32, 38], sound recovery [9], and healthcare [1, 2, 11, 16, 23]. Video motion magnification [20, 24, 39, 40, 42] is the technique to amplify subtle motions in a video, revealing details of motion that are hard to perceive with the naked eyes. This

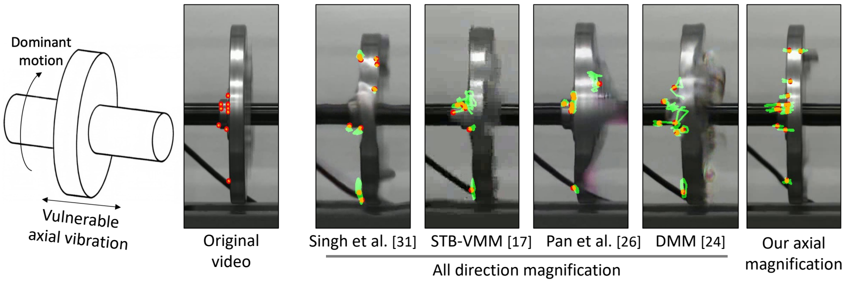


Fig. 1: Importance of axial motion magnification. When identifying faults in rotating machinery, analysis of the vulnerable axial vibration is critical [22,43]. Existing learning-based methods [17,24,26,31] amplify motions along all axes, which yield artifacts. It hinders the analyses of vulnerable axial vibration. This motivates the importance of our axial motion magnification that magnifies decomposed motions along a user-specified axis. We magnify the axial vibration only, achieving artifact-free results and the legibility of critical motions. For the visualization purpose, we overlay the sample trajectories obtained from the Kanade-Lucas-Tomasi (KLT) Tracker [21].

allows users to grasp dense and holistic behavior information of the scene of interest instantly, as long as the resulting motion is simple and easily interpretable. However, in practice, vibrating objects in the real world often possess complex systems that have complicated natural frequencies, modes, and directions [25]. Even after being magnified, the intricate movement persists, which restricts the advantages of motion magnification because the key underlying premise of its effectiveness is based on the legibility of the magnified motion in aforementioned applications, *i.e.*, effectively understanding the way objects move.

In this work, we focus on improving the legibility of magnified motion by proposing a novel concept, *axial* video motion magnification, which magnifies decomposed motions along the user-specified direction. All the existing works, *e.g.*, [17,24,26,31,39,40,42], have overlooked this key importance of legibility according to axes, although there are many practical cases where the importance of motion varies by axes. In the fault detection application of machines, vibration direction serves as the key component for the expert tree of rotating machinery's fault diagnosis [43]. As shown in Fig. 1, even small motions along the vulnerable axis are critical, while dominant rotational motions are not [22]. Likewise, many apparatus consisting of natural or artificial materials often have vulnerable axes due to the asymmetry property of microstructures, *e.g.*, fracture toughness [3,18,37]. This motivates us to separately analyze axial motions.

Specifically, we propose a novel learning-based axial video motion magnification method, where the motions in a user-specified axis are magnified. Our method can independently magnify small motions along two orthogonal orientation axes with two independent magnification factors for each axis, which facilitates the analysis of complex small motions in the lens of axes favorable to the user. To this end, we propose the Motion Separation Module (MSM), which disentangles the motion representation into two orthogonal orientations and manipulates it in the direction specified by the user. To train the proposed neural network, we develop and build a new synthetic dataset for the axial motion magnification task. Therefore, our proposed approach adds a new user control feature, which improves the legibility of resulting motions along a certain axis. This allows our axial motion magnification to become a generalization of the existing *generic* motion magnification. Thus, our method can be directly adopted to the generic motion magnification task and achieve favorable performance against competing methods. We summarize our contributions as follows:

- We propose the new concept, learning-based axial video motion magnification, which allows us to selectively amplify small motions along a specific direction.
- We propose and analyze the Motion Separation Module (MSM) for the axial motion magnification. We find that adopting MSM is effective not only in axial magnification but also in distinguishing small motions from noise.
- We propose a way to synthesize a new synthetic dataset to train the proposed axial motion magnification model and exhibit generalization to real data.

2 Related Work

Liu *et al.* [20] first pioneered the video motion magnification task, which involves estimating explicit motion trajectory via optical flow to generate magnified frames. They group and filter the motion trajectories based on motion similarity and user's intervention, and magnify them through explicit image warping, followed by video inpainting to fill holes created by the explicit warping.

Wu *et al.* [42] re-formulate the motion magnification task as an Eulerian method that represents motion by intensity changes of pixels at each fixed location without actual movement [12]. The Eulerian approach, *e.g.*, [24, 33–36, 39, 40, 42, 44], becomes standard in motion magnification due to its noise robustness, sensitivity to small motions, and simple system by avoiding challenging warp and inpaint approach for filling holes and handling occlusions. The system of the Eulerian methods typically consists of motion representation, manipulation, and reconstruction. The previous works can be categorized into two main focuses: 1) proposing motion representations or 2) motion manipulation methods.

In the first category, Wu *et al.* [42] present the motion representation motivated by the first-order Taylor expansion, which is implemented by the Laplacian pyramid as spatial decomposition. Wadhwa *et al.* [39,40] enhance the representation by modeling the motion as phase representations, which are implemented by complex steerable filters [29] in [39] and Riesz transform in [40] as spatial decomposition, respectively. These works rely on the classic signal processing theory with such hand-designed spatial filter designs, which do not model non-linear phenomenons and yield artifacts and noisy results.

To deal with, Oh *et al.* [24] first coined learning-based video motion magnification, called Deep Motion Magnification (DMM), by modeling motion representation with deep neural networks. As no real data exists for training video motion magnification, they propose a method to build motion magnification synthetic data. With the development, other learning-based variants [17,31] have been proposed, focusing on neural network architectures. These approaches demonstrate promising results by effectively handling diverse challenging scenarios such as occlusion and noisy inputs. Also, the motion magnification factors of the data-driven approaches can be controlled by the way the synthetic dataset is generated, while those of the traditional methods [39, 42] are theoretically restricted.

In the second category, when Wu *et al.* [42] present Eulerian motion magnification, they also propose to use a temporal filter on the motion representation to select the motion frequency of interest. This allows the noise to be suppressed by focusing on specific motions and increasing the legibility of magnified motion. There were attempts to extend to increase the legibility by proposing temporal filters to magnify different types of motions and deal with artifacts from large motions: acceleration [36, 44], intensity-aware temporal filter [35], velocity or all-frequency filter [24]. Our work is compatible with all these methods.

In this work, we present a new notion of motion magnification by disentangling motion axes of the user's interest. We design a neural architecture to induce disentanglement of motion in oriented axes. Also, we propose the synthetic data generation pipeline for the axial motion magnification task.

3 Learning-based Axial Motion Magnification

We first discuss preliminaries about generic motion magnification, which refers to the methods that amplify the motion regardless of direction, including the prior arts [17,24,26,31,42] (Sec. 3.1). Then, we re-frame the motion magnification problem in the view of axial motion magnification (Sec. 3.2), and elaborate on our network architecture, and synthetic data generation method (Sec. 3.3).

3.1 Preliminary – Generic Motion Magnification

Following the convention [39,42], for simplicity, we consider the 1D image intensity being shifted by the displacement function $\delta(x,t)$, which is parameterized by position x and time t. It can be generalized to local translational motion in 2D image [42]. Given an underlying intensity profile function $f(\cdot)$, the 1D image intensity $I(x,t)$ can be represented as

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

The goal of motion magnification is to synthesize the magnified image $\hat{I}(x,t)$:

$$\hat{I}(x,t) = f(x + (1+\alpha)\delta(x,t)), \tag{2}$$

where α denotes the magnification factor. The key factor of motion magnification methods lies in the extraction of the displacement function $\delta(x,t)$ from Eq. (1). If $\delta(x,t)$ can be decomposed, we can approximate $\hat{I}(x,t)$ by multiplying $\delta(x,t)$ with the magnification factor α and applying the reverse of the decomposition process. However, it is an ill-posed problem to extract exact displacements from the observed intensity images [42]. Therefore, the prior arts approximately decompose $\delta(x,t)$; for example, Wu *et al.* [42] use the first-order Taylor expansion as:

$$I(x,t) \approx f(x) + \delta(x,t)\tfrac{\partial f(x)}{\partial x}. \tag{3}$$

Learning-based methods [17,24,31] design neural networks that have intermediate representations related to $\delta(\cdot)$, called shape representation. The representations are multiplied by α, followed by reconstruction for magnification.

3.2 Axial Motion Magnification

To introduce the axial motion magnification task, we now consider the 2D spatial coordinate by slightly abusing the notations, *e.g.*, $\mathbf{x} = (x,y)$ to refer to the coordinate in the 2D image intensity $I(\mathbf{x},t)$.

Problem Definition. We can represent $I(\mathbf{x},t) = f(\mathbf{x} + \boldsymbol{\delta}(\mathbf{x},t))$ with a 2D displacement vector $\boldsymbol{\delta}(\mathbf{x},t) \in \mathbb{R}^2$. Given an angle $\phi \in \mathbb{R}$ of the user-specified direction of interest, the goal of the axial motion magnification task is to isolate and amplify the motion component corresponding to the direction angle ϕ within the displacement vector. We represent the axially magnified image $\hat{I}^{\phi}(\mathbf{x},t)$ as

$$\hat{I}^{\phi}(\mathbf{x},t) = f(\mathbf{x} + \alpha^{\phi}\boldsymbol{\delta}^{\phi}(\mathbf{x},t)), \tag{4}$$

where $\alpha^{\phi} \geq 0$ denotes the axial magnification factor and $\boldsymbol{\delta}^{\phi}(\mathbf{x},t)$ the projection of $\boldsymbol{\delta}(\mathbf{x},t)$ onto a 2D directional unit vector $\mathbf{p}^{\phi}$ with the angle ϕ, *i.e.*, the motion component. We can break down the motion component $\boldsymbol{\delta}^{\phi}(\mathbf{x},t)$ into:

$$\boldsymbol{\delta}^{\phi}(\mathbf{x},t) = \mathrm{proj}_{\mathbf{p}^{\phi}}\, \boldsymbol{\delta}(\mathbf{x},t). \tag{5}$$

Relationship with Generic Motion Magnification. If we obtain $\boldsymbol{\delta}(\mathbf{x},t)$, we can determine $\boldsymbol{\delta}^{\phi}(\mathbf{x},t)$ and $\boldsymbol{\delta}^{\phi_{\perp}}(\mathbf{x},t)$ through the projections onto $\mathbf{p}^{\phi}$ and $\mathbf{p}^{\phi_{\perp}}$. In this case, we can extend Eq. 4 to represent not only the displacement vector of an angle $\boldsymbol{\delta}^{\phi}(\mathbf{x},t)$ but also of its orthogonal direction $\boldsymbol{\delta}^{\phi_{\perp}}(\mathbf{x},t)$, as

$$\hat{I}^{\phi}(\mathbf{x},t) = f(\mathbf{x} + \alpha^{\phi}\boldsymbol{\delta}^{\phi}(\mathbf{x},t) + \alpha^{\phi_{\perp}}\boldsymbol{\delta}^{\phi_{\perp}}(\mathbf{x},t)), \tag{6}$$

where α_{ϕ}, $\alpha_{\phi_{\perp}} \geq 0$ denotes the axial magnification factors corresponding to the ϕ and $\phi_{\perp}$ directions, respectively. This formulation encompasses the various motion magnification scenarios, *e.g.*, axial and generic motion magnifications. Setting $\alpha^{\phi_{\perp}}$ to 0 leads to the formulation resulting in axial motion magnification, while setting α^{ϕ} equal to $\alpha^{\phi_{\perp}}$ results in generic motion magnification.

3.3 Neural Networks and Training

Departing from the previous learning-based methods that are confined to generic motion magnification [17, 24, 26, 31], we introduce a novel neural network architecture and a dedicated training dataset designed to learn two angle-aware

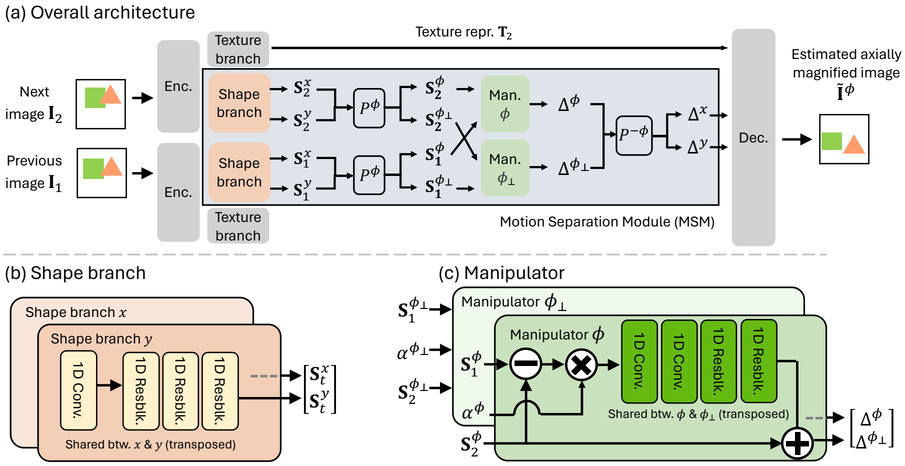


Fig. 2: Proposed architecture. (a) The *Encoder* outputs features from input images and the features are fed to the *Texture* branch and Motion Separation Module (MSM). (b) Using weight-shared 1D convolutions, the *Shape* branch extracts shape representations along the x and y-axes. These representations are fed to the projection layer P^ϕ, which generates axial shape representations, *i.e.*, $\mathbf{S}_t^\phi$ and $\mathbf{S}_t^{\phi_\perp}$. (c) the *Manipulator* amplifies them by the axial magnification factors and the inverse projection layer $P^{-\phi}$ re-project them onto the x and y-axes. Finally, the *Decoder* predicts the axially magnified image from the outputs from both the *Texture* branch and MSM.

motion representations proportional to the motion displacement $\boldsymbol{\delta}^\phi$ and $\boldsymbol{\delta}^{\phi_\perp}$, respectively. These allow our approach to unveil a distinctive feature: the magnification of motion in user-defined directions while retaining the functionality for generic motion magnification.

Network Architecture. Our whole architecture consists of *Encoder*, *Texture* & *Shape* branches, *Manipulator*, and *Decoder* similar to DMM [24] (see Fig. 2-(a)), where texture represents color and texture-related information while shape represents scene structure-related information that later leads to motion $\boldsymbol{\delta}$ [24]. To extract axial shape representations, we design Motion Separation Module (MSM) consisting of the completely re-designed and dedicated *Shape* branch and *Manipulator* as depicted in Fig. 2-(b,c). In MSM, instead of extracting a single specified direction's $\boldsymbol{\delta}^\phi$, we design to extract its orthogonal direction's $\boldsymbol{\delta}^{\phi_\perp}$ as well. This design choice is motivated by the extended axial motion magnification equation Eq. 6 and enables conducting various motion magnifications, including both axial and generic motion magnifications.

Given consecutive input video frames $\mathbf{I}_t \in \mathbb{R}^{H\times W\times 3}$ at $t = 1$ and $t = 2$ for example, texture representations $\mathbf{T}_t \in \mathbb{R}^{H/4\times W/4\times 32}$ are obtained by $\mathbf{T}_t = F(E(\mathbf{I}_t))$, where $E(\cdot)$ and $F(\cdot)$ denote the *Encoder* and the *Texture* branch, respectively. The outputs of E are fed into MSM. The same output from E is fed into the *Texture* branch and MSM, respectively.

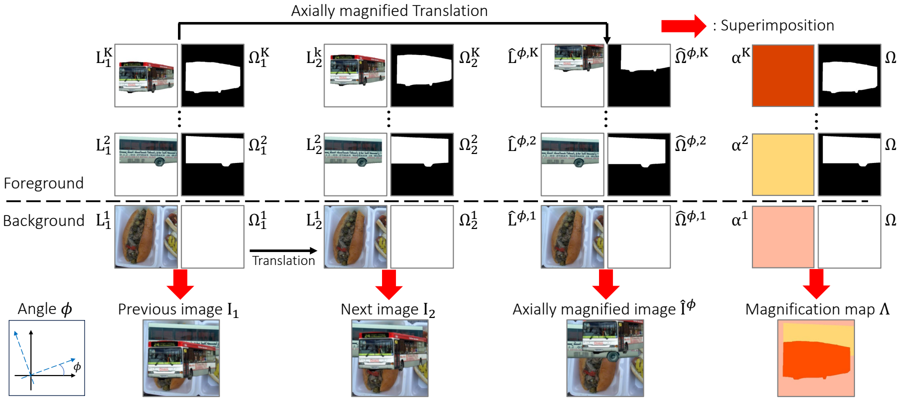


Fig. 3: Synthetic data generation pipeline for axial motion magnification. From the sampled background and foregrounds, each with their own segmentation masks, we compose the previous layer images $\{\mathbf{L}_1^k\}_{k=1}^K$ and masks $\{\boldsymbol{\Omega}_1^k\}_{k=1}^K$. To generate next layer images $\{\mathbf{L}_2^k\}_{k=1}^K$ and masks $\{\boldsymbol{\Omega}_2^k\}_{k=1}^K$, we apply the random translations to $\{\mathbf{L}_1^k\}_{k=1}^K$ and $\{\boldsymbol{\Omega}_1^k\}_{k=1}^K$. Axially magnified layer images $\{\hat{\mathbf{L}}^{\phi,k}\}_{k=1}^K$ and masks $\{\hat{\boldsymbol{\Omega}}^{\phi,k}\}_{k=1}^K$ are also synthesized by translations but with the axially magnified translation parameters. These images and masks are then superimposed into a single image to yield $\mathbf{I}_1$, $\mathbf{I}_2$, and $\hat{\mathbf{I}}^{\phi}$, respectively. The dataset also include angles ϕ and the object-wise magnification maps $\boldsymbol{\Lambda}$ generated by superimposing $\{\boldsymbol{\alpha}^k\}_{k=1}^K$ with $\{\boldsymbol{\Omega}_1^k\}_{k=1}^K$.

To extract the motion representations along two orthogonal orientations and manipulate them based on the user-defined angle, we grant the learnable parameters to learn the directionality in MSM. Our *Shape* branch $G(\cdot)$ first extracts the axial shape representations along the canonical x and y-axes by applying weight-shared 1D convolutions but with spatially transposing the convolution kernels, yielding $[\mathbf{S}_t^x, \mathbf{S}_t^y]{=}G(E(\mathbf{I}_t))$ where $\mathbf{S}_t^x, \mathbf{S}_t^y \in \mathbb{R}^{H/2\times W/2\times 32}$. Then, these are projected by the *projection* layer, which produces axial shape representations of ϕ and $\phi_\perp$ directions, *i.e.*, $\mathbf{S}_t^{\phi}$ and $\mathbf{S}_t^{\phi_\perp}$. Motivated by the steerable filters [13], where an arbitrarily rotated representation can be synthesized by a linear combination of directional representations, we design the projection layer P^{ϕ} with a linear matrix as

$$P^{\phi}\left(\begin{bmatrix} S_t^x \\ S_t^y \end{bmatrix}\right) = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} S_t^x \\ S_t^y \end{bmatrix} = \begin{bmatrix} S_t^{\phi} \\ S_t^{\phi_\perp} \end{bmatrix}. \tag{7}$$

The *Manipulator* $M(\cdot)$ computes the difference of the axial shape representations and magnifies them by multiplying the axial magnification factors α^{ϕ}. Then, these manipulated representations are fed into subsequent 1D convolutions, and added to the axial shape representation $\mathbf{S}_2^{\phi}$. For $\phi_\perp$, we use the same manipulator, of which weights are shared but spatially transposed, for applying $\alpha^{\phi_\perp}$. Note that, with this separation of ϕ and $\phi_\perp$, we can set the magnification factors α^{ϕ} and $\alpha^{\phi_\perp}$ independently, enabling broad applications of controls as another benefit. For the outputs of the *Manipulator* $\mathit{\Delta}^{\phi}, \mathit{\Delta}^{\phi_{\perp}}$, where $\mathit{\Delta}^{\phi} = M(\mathbf{S}_1^{\phi}, \mathbf{S}_2^{\phi}, \alpha^{\phi})$, we re-project them onto the canonical x and y-axes by inverse projection layer $P^{-\phi}$, obtaining $\mathit{\Delta}^{x}, \mathit{\Delta}^{y}$. Finally, the *Decoder* $D(\cdot)$ predicts the axially magnified output frame $\tilde{\mathbf{I}}^{\phi}$ as

$$\tilde{\mathbf{I}}^{\phi} = D\left(\mathbf{T}_2, \mathit{\Delta}^{x}, \mathit{\Delta}^{y}\right). \tag{8}$$

This network architecture enables the network to conduct both generic and axial motion magnification, given the user setting of the angle ϕ. The model is trained with the loss function suggested by DMM [24] with a slight modification to impose the loss separately to the x-axis and y-axis shape representations. Details of the loss function can be found in the supplementary material.

Training Data Generation. In the real world, acquiring consecutive images and magnified images at the same time is impossible. Due to this, DMM [24] proposes a synthetic training dataset for the generic motion magnification task. However, this dataset is not sufficient to induce the disentanglement of the axial property we need. Thus, we propose a new synthetic dataset specifically designed for the axial motion magnification, where the motion between $\mathbf{I}_1$ and $\hat{\mathbf{I}}^{\phi}$ is associated with the angle ϕ and axial magnification factor vector $\boldsymbol{\alpha}$=$(\alpha^{\phi}; \alpha^{\phi_{\perp}})$. Motivated by the synthetic dataset generation protocol of DMM, we synthesize the training data pairs using the widely adopted simple copy-paste method [14, 24].

Figure 3 shows the synthetic data generation pipeline. We sample one background from COCO [19] and $K-1$ number of foreground textures with segmentation masks from PASCAL VOC [10]. These elements are randomly located on image planes of resolution 384×384 to produce K previous layer images $\{\mathbf{L}_1^k\}_{k=1}^K$ and corresponding masks $\{\boldsymbol{\mathit{\Omega}}_1^k\}_{k=1}^K$. Following this, with randomly sampled K translation parameters $\{\mathbf{d}^k\}_{k=1}^K$, we generate the next layer images $\{\mathbf{L}_2^k\}_{k=1}^K$ and masks $\{\boldsymbol{\mathit{\Omega}}_2^k\}_{k=1}^K$ by translating the initial layers and masks according to $\{\mathbf{d}^k\}_{k=1}^K$. For the axially magnified layer images $\{\hat{\mathbf{L}}^{\phi,k}\}_{k=1}^K$ and their masks $\{\hat{\boldsymbol{\mathit{\Omega}}}^{\phi,k}\}_{k=1}^K$, we sample K axial magnification vectors $\{\boldsymbol{\alpha}^k\}_{k=1}^K$ and a single degree of angle ϕ. Then, we perform the same procedure as the next layers but with the axially magnified translation parameters $\{\boldsymbol{\alpha}^k(\mathrm{proj}_{\mathbf{p}^{\phi}}\, \mathbf{d}^k; \mathrm{proj}_{\mathbf{p}^{\phi_{\perp}}}\, \mathbf{d}^k)\}_{k=1}^K$. These previous, next, and axially magnified layer images and masks are then superimposed into a single image to yield $\mathbf{I}_1$, $\mathbf{I}_2$, and $\hat{\mathbf{I}}^{\phi}$, respectively. Our dataset also includes the angle ϕ and the object-wise magnification map $\boldsymbol{\mathit{\Lambda}}$ which is generated by superimposing $\{\boldsymbol{\alpha}^k\}_{k=1}^K$ segmented with $\{\boldsymbol{\mathit{\Omega}}_1^k\}_{k=1}^K$. We observe that utilizing both ϕ and $\boldsymbol{\mathit{\Lambda}}$ are useful for learning the representations distinguishing small motions from noises, which will be discussed on Sec. 4.3. Additionally, the adaptation of both ϕ and $\boldsymbol{\mathit{\Lambda}}$ enables pixel-wise axial motion magnification. We provide more details in the supplementary materials.

4 Experiments

Implementation Details. We train our learning-based axial motion magnification network on the newly proposed dataset, which contains a total of 100k

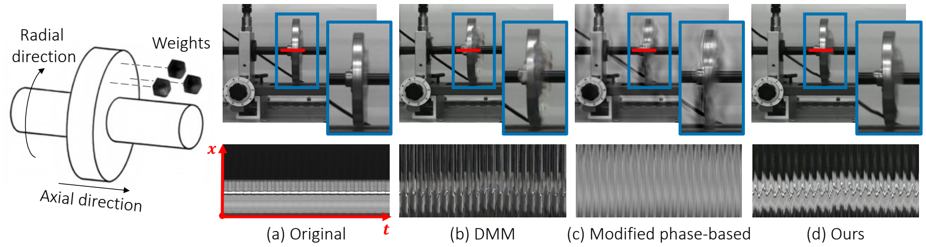


Fig. 4: [Left] Imposing an imbalance on a rotor, [Right] Qualitative results in axial motion magnification scenario. We attach weights to a rotor to impose an imbalance and acquire *rotor imbalance* sequence, which has axial vibrations. Then, we amplify only the motion of rotor's axial direction with the magnification factor $\alpha = 40$, using ours and modified phase-based method. We also show the magnified result of DMM [24] as a reference result of generic motion magnification. Our method generates magnified frames without artifacts and exhibits the x-t slice showing clearly legible axial vibrations, while modified phase-based method and DMM both suffer from severe artifacts and have unclear axial vibrations in the x-t slice.

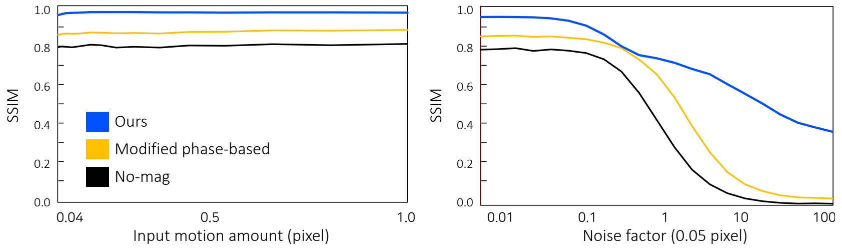
samples, for 50 epochs with a batch size of 8 and a learning rate 2×10^{-4}. For training, we use two NVIDIA TITAN RTX GPUs.

Evaluation Setup. We examine the performance of our method in axial and generic motion magnification, respectively. In generic motion magnification, we compare our method to the phase-based method [39], Singh *et al.* [31], STB-VMM [17], Pan *et al.* [26], and DMM [24]. In axial motion magnification, there is no method of handling a user-specified angle and performing axial magnification due to our novel problem setup. Therefore, we propose a new axial baseline, called *modified phase-based*, by modifying Wadhwa *et al.* [39]. Specifically, we modulate the phase-based method [39] to operate in the axial scenario by employing a half-octave bandwidth pyramid and two orientations, with one of them having its phase representation manipulated along the axis of interest. Following DMM [24], we use both the *dynamic* and *static* modes in the experiments. Additional experiments of diverse scenarios and implementation details can be found in the supplementary material and video.

4.1 Axial Motion Magnification

We evaluate our method compared to the modified phase-based method in the axial motion magnification scenario to demonstrate the effectiveness of the learning-based axial motion magnification.

Qualitative Results. We demonstrate the advantage of our method that it can amplify only the motion along the axis of interest while disentangling the motions in uninterested directions that interfere with motion analysis. To illustrate this concept concretely, consider a scenario where a shaft is rotating in the radial direction. In such cases, magnifying and examining the motion along the



(a) Subpixel test

(b) Noise test

Fig. 5: Quantitative results in axial motion magnification scenario. (a) In the subpixel test, ours shows superior performance on SSIM over the modified phase-based method across all input motion amount, ranging from 0.04 to 1.0. (b) In the noise tests when the input motion amount is 0.05 pixel, we observe a growing disparity in SSIM scores between ours and the phase-based approach, as the noise factor rises.

axial direction, which is crucial to assess the condition of the rotating machinery [22], becomes challenging due to the dominance of rotational motion over the axial component. We conduct an experiment shown in Fig. 4 by attaching weights to a rotor to impose an imbalance, which results in axial vibrations. Then, we acquire a video of the imbalanced rotor, called *rotor imbalance* sequence. We choose a horizontal-axis line in the original frame and visualize x-t slices for the magnified output frames from each method, respectively. Note that we also provide the result of DMM [24] as a reference to compare the results of axial motion magnification with generic motion magnification. As shown in Fig. 4, our method produces the magnified output frames without artifacts and exhibits the x-t slice that clearly depicts axial vibrations. In contrast, the modified phase-based method suffers from severe ringing artifacts, likely due to the overcompleteness of the complex steerable filter [29, 30], which cannot perfectly separate the phase representation into two orthogonal directions. DMM yields the magnified frames with artifacts and unclear axial vibrations in the x-t slice, since the representation of generic motion magnification method struggles to disentangle the dominant motion of the radial direction from the motion of interest, *i.e.*, axial direction's motion.

Quantitative Results. To quantitatively evaluate our learning-based axial motion magnification method, we generate an axial evaluation dataset based on the validation dataset of DMM [24]. The method of generating the dataset is almost the same as that of the training dataset. One difference is that we adjust the motion amplification factor to ensure that the amplified motion magnitude along a random axis is equal to 10. The motion amplification factor for the other axis is set to half the value. Note that we set ϕ to be 0 for this quantitative evaluation. We report the Structural Similarity Index (SSIM) [41] between the ground truth and output frames of the modified phase-based method and ours. As a reference, we provide the SSIM between ground truth and input frames. Figure 5 summarizes the results. We measure the SSIM by varying the levels of

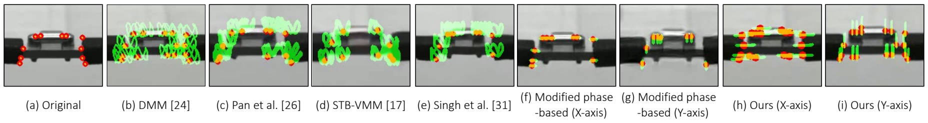

(a) Original (b) DMM [24] (c) Pan et al. [26] (d) STB-VMM [17] (e) Singh et al. [31] (f) Modified phase -based (X-axis) (g) Modified phase -based (Y-axis) (h) Ours (X-axis) (i) Ours (Y-axis)

Fig. 6: Motion legibility improvement. We visualize the 40× magnified frames of the structure, which are overlaid with the sampled trajectories from the KLT tracker. Generic motion magnification methods (b-e) produce trajectories that are more complex and hard to interpret. However, Ours (h-i) shows simplified and legible motion trajectories when magnifying along the specific axis (*i.e.*, x-axis and y-axis), while the modified phase-based methods (f-g) exhibit small and bounded amplified motions, as theoretically proven in the phase-based method [39].

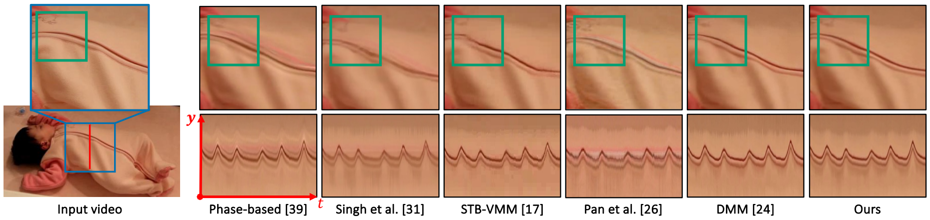
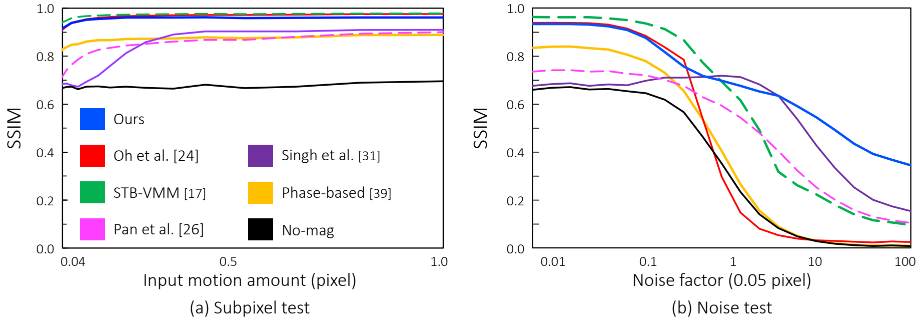


Fig. 7: Qualitative results in generic motion magnification scenario. We amplify the *baby* sequence with the magnification factor α=20, using phase-based method [39], learning-based methods [17, 24, 26, 31], and Ours. Ours and DMM favorably preserve the edges of the clothes and show no ringing artifacts in the magnified frames and the x-t slices. In contrast, the magnified output frames of the phase-based, Singh *et al.*, STB-VMM, and Pan *et al.* show ringing artifacts or blurry results.

motion (Fig. 5-(a) Subpixel test) and additive noise (Fig. 5-(b) Noise test) in the input images. The number of evaluation data samples for each level of motion and noise is $1,000$. Regardless of the input motion magnitude and noise level, our method consistently outperforms the modified phase-based approach, which indicates that our proposed network architecture and dataset are effective for learning axis-wise disentangled representations.

Motion Legibility Comparison. To demonstrate the improved legibility of magnified motions by our method, we use a structure that exhibits complex movements. We then visualize and compare the motion trajectories, tracked by the KLT tracker, of the 40× magnified video sequences of this structure using both the generic methods and the axial method (Ours). We also provide the modified phase-based for the axial method. As shown in Fig. 6, our method shows legible trajectories when magnifying along the specific axis (*i.e.*, x-axis and y-axis), while generic motion magnification method (Fig. 6-(b-e)) shows the entangled trajectories difficult to judge major motion characteristics. The modified phase-based methods (Fig. 6-(f-g)) result in bounded amplified motion, as theoretically proven in the phase-based method [39].



(a) Subpixel test

(b) Noise test

Fig. 8: Quantitative results in generic motion magnification scenario. (a) In the subpixel test, Ours outperforms phase-based method, Singh *et al.*, and Pan *et al.* and achieves favorable performance on SSIM compared to DMM and STB-VMM. (b) In the noise test, Ours shows comparable noise tolerance compared to other methods and high noise tolerance as the noise factor increases.

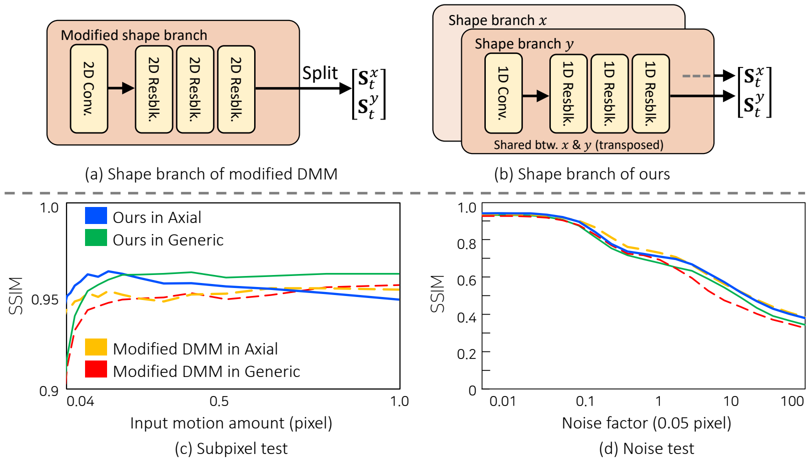
4.2 Generic Motion Magnification

Our method can be readily adapted for generic motion magnification scenarios without further training. This adaptability is achieved by simply multiplying the same magnification factors with the axis-wise shape representations. In the context of generic motion magnification, we compare our method with the phase-based method [39] and the learning-based methods [17,24,26,31].

Qualitative Results. We visualize the magnified output frames and plot the x-t slices for the *baby* sequence, comparing ours with the several motion magnification methods in the generic scenarios (see Fig. 7). Both our method and DMM [24] favorably preserve the edges of the baby's clothing and show no ringing artifacts in the magnified results of breathing motion. In contrast, the phase-based method [39], Singh *et al.* [31], STB-VMM [17], Pan *et al.* [26] and show severe ringing artifacts or blurry results[4].

Quantitative Results. To quantitatively verify the ability of our method in generic motion magnification, we synthesize a generic validation dataset. Unlike the axial case, we set the magnification factor α to be identical along the x and y axes. We report SSIM [41] between ground truth and output frames from the phase-based method [39] and the learning-based methods [17,24,26,31]. As shown in Fig. 8, for input motion ranges from 0.04 to 1.0, ours outperforms the phase-based method, Singh *et al.* [31], Pan *et al.* [26]. Compared to DMM [24] and STB-VMM [17], ours demonstrates favorable performance, which exceeds the threshold for visually acceptable SSIM scores [15]. Ours demonstrates comparable noise tolerance to other methods and exhibits high noise tolerance as noise factor increases.

[4] We reproduced all the results using the codes publicly accessible.



(a) Shape branch of modified DMM

(b) Shape branch of ours

(c) Subpixel test

(d) Noise test

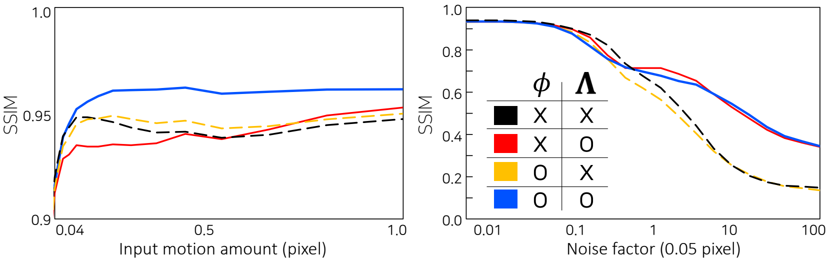
Fig. 9: [Top] Architectural difference on the *shape* branch, [Bottom] Quantitative results of ablating Motion Separation Module (MSM). (a) The modified DMM, designed for ablation study, employs 2D convolutions and splits features along channel dimensions for axial motion magnification. (c) Ours with MSM generally achieves higher SSIM in the subpixel test on generic and axial evaluation datasets. (d) In the noise test, Ours shows comparable performance to the modified DMM.

4.3 Ablation Study

In this section, we conduct ablation studies to evaluate the impact of the Motion Separation Module (MSM) and the components of the proposed synthetic training data. We carry out quantitative experiments on the evaluation dataset of both the generic case and the axial case that has random angles.

Motion Separation Module (MSM). To validate the effectiveness of MSM, we design a competitor called modified DMM, which closely resembles that of DMM [24]. As shown in the top of Fig. 9, different from our method that uses 1D convolutions, the modified DMM employs 2D convolutions in the *Shape* branch and the *Manipulator*. The axial shape representations of the modified DMM are acquired by dividing the feature map along the channel dimension. We train the networks with the same training details as Ours. The bottom of Fig. 9 shows that Ours with MSM generally achieves higher SSIM in the generic and axial subpixel tests, which shows the efficacy of the MSM in capturing small motions. In the noise test, Ours shows comparable performance to the modified DMM.

Components of Synthetic Training Data. To evaluate the impact of the angle ϕ and the object-wise motion magnification map $\boldsymbol{\Lambda}$, we generate the different types of training data varying the presence of these components. Our newly designed dataset incorporates both ϕ and $\boldsymbol{\Lambda}$, contrasting with the dataset that follows the same setup as DMM [24], which does not contain either element. In addition, we generate two more datasets that each add one of these components (*i.e.*, either ϕ or $\boldsymbol{\Lambda}$) to the base dataset that initially does not include them.



(a) Subpixel test

(b) Noise test

Fig. 10: Ablation study of the components in data generation in the generic evaluation dataset. We generate the different training data varying the presence of the angle ϕ and the object-wise motion magnification map $\boldsymbol{\Lambda}$, and evaluate the networks trained on each dataset configuration using the generic evaluation dataset. (a) Using both ϕ and $\boldsymbol{\Lambda}$ demonstrates best performance in the subpixel test. (b) In the noise test, we observe that utilizing $\boldsymbol{\Lambda}$ notably enhances noise tolerance.

Note that evaluating the networks trained on these datasets on the axial evaluation dataset is infeasible since the networks trained without ϕ cannot perform axial motion magnification. Thus, we use the generic evaluation dataset for this ablation study. Fig. 10 shows that the addition of either ϕ or $\boldsymbol{\Lambda}$ achieves no improvement in the subpixel test. The combined use of both ϕ and $\boldsymbol{\Lambda}$ yields the most significant performance improvement in the subpixel test, demonstrating that our proposed data set is beneficial in the generic motion magnification task as well. In the noise test, utilizing $\boldsymbol{\Lambda}$ notably enhances noise tolerance, while the addition of ϕ has no effect on noise tolerance.

5 Conclusion

In this work, we present a novel concept, axial video motion magnification, which improves the legibility of the motions by disentangling and magnifying the motion representations along axes specified by users. To this end, we propose an innovative learning-based approach for both axial and generic motion magnification, incorporating the Motion Separation Module (MSM) to effectively extract and magnify motion representations along two orthogonal orientations. To support this, we establish a new synthetic data generation pipeline tailored for the axial motion magnification. Our method provides user controllability and significantly enhances the legibility of the motions along chosen axes, showing favorable performance compared to competing methods, even in the generic motion magnification case. Although the axial motion magnification serves as a branch that enhances users' applicability, another branch can be the method to perform motion magnification in real-time, which is useful and beneficial for various applications. In other respects, most of existing 2D video motion magnification methods including ours assume fixed 2D viewpoints. Exploring real-time inference and moving viewpoints would be a promising direction for future research.

Acknowledgment

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant (No. RS-2022-II220290, Visual Intelligence for Space-Time Understanding and Generation based on Multi-layered Visual Common Sense; No. RS-2019-II191906, Artificial Intelligence Graduate School Program (POSTECH)) and National Research Foundation of Korea (NRF) grant (No. RS-2024-00358135, Corner Vision: Learning to Look Around the Corner through Multi-modal Signals) funded by the Korea government (MSIT).

References

1. Balakrishnan, G., Durand, F., Guttag, J.: Detecting pulse from head motions in video. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3430–3437 (2013)
2. Brattoli, B., Büchler, U., Dorkenwald, M., Reiser, P., Filli, L., Helmchen, F., Wahl, A.S., Ommer, B.: Unsupervised behaviour analysis and magnification (ubam) using deep learning. Nature Machine Intelligence **3**(6), 495–506 (2021)
3. Brodnik, N., Brach, S., Long, C., Ravichandran, G., Bourdin, B., Faber, K., Bhattacharya, K.: Fracture diodes: Directional asymmetry of fracture toughness. Physical Review Letters **126**(2), 025503 (2021)
4. Cha, Y.J., Chen, J.G., Büyüköztürk, O.: Output-only computer vision based damage detection using phase-based optical flow and unscented kalman filters. Engineering Structures **132**, 300–313 (2017)
5. Chen, J.G., Davis, A., Wadhwa, N., Durand, F., Freeman, W.T., Büyüköztürk, O.: Video camera–based vibration measurement for civil infrastructure applications. Journal of Infrastructure Systems **23**(3), B4016013 (2017)
6. Chen, J.G., Wadhwa, N., Cha, Y.J., Durand, F., Freeman, W.T., Buyukozturk, O.: Structural modal identification through high speed camera video: Motion magnification. In: Topics in Modal Analysis I, Volume 7: Proceedings of the 32nd IMAC, A Conference and Exposition on Structural Dynamics, 2014. pp. 191–197. Springer (2014)
7. Chen, J.G., Wadhwa, N., Cha, Y.J., Durand, F., Freeman, W.T., Buyukozturk, O.: Modal identification of simple structures with high-speed video using motion magnification. Journal of Sound and Vibration **345**, 58–71 (2015)
8. Chen, J.G., Wadhwa, N., Durand, F., Freeman, W.T., Buyukozturk, O.: Developments with motion magnification for structural modal identification through camera video. In: Dynamics of Civil Structures, Volume 2, pp. 49–57. Springer (2015)
9. Davis, A., Rubinstein, M., Wadhwa, N., Mysore, G.J., Durand, F., Freeman, W.T.: The visual microphone: Passive recovery of sound from video. ACM Transactions on Graphics (SIGGRAPH) (2014)
10. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International journal of computer vision **88**(2), 303–338 (2010)
11. Fan, W., Zheng, Z., Zeng, W., Chen, Y., Zeng, H.Q., Shi, H., Luo, X.: Robotically surgical vessel localization using robust hybrid video motion magnification. IEEE Robotics and Automation Letters **6**(2), 1567–1573 (2021)

12. Freeman, W.T., Adelson, E.H., Heeger, D.J.: Motion without movement. ACM SIGGRAPH **25**(4), 27–30 (1991)
13. Freeman, W.T., Adelson, E.H., et al.: The design and use of steerable filters. IEEE Transactions on Pattern analysis and machine intelligence **13**(9), 891–906 (1991)
14. Ghiasi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T.Y., Cubuk, E.D., Le, Q.V., Zoph, B.: Simple copy-paste is a strong data augmentation method for instance segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2918–2928 (2021)
15. Ha, H., Hyun-Bin, O., Jun-Seong, K., Byung-Ki, K., Sung-Bin, K., Tran, L.T., Kim, J.Y., Bae, S.H., Oh, T.H.: Revisiting learning-based video motion magnification for real-time processing (2024)
16. Janatka, M., Marcus, H.J., Dorward, N.L., Stoyanov, D.: Surgical video motion magnification with suppression of instrument artefacts. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III 23. pp. 353–363. Springer (2020)
17. Lado-Roigé, R., Pérez, M.A.: Stb-vmm: Swin transformer based video motion magnification. Knowledge-Based Systems **269**, 110493 (2023)
18. Li, B., Deng, B., Shou, W., Oh, T.H., Hu, Y., Luo, Y., Shi, L., Matusik, W.: Computational discovery of microstructured composites with optimal stiffness-toughness trade-offs. Science Advances **10**(5) (2024)
19. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European Conference on Computer Vision (ECCV) (2014)
20. Liu, C., Torralba, A., Freeman, W.T., Durand, F., Adelson, E.H.: Motion magnification. ACM transactions on graphics (TOG) **24**(3), 519–526 (2005)
21. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: IJCAI'81: 7th international joint conference on Artificial intelligence. vol. 2, pp. 674–679 (1981)
22. Luo, Y., Zhang, W., Fan, Y., Han, Y., Li, W., Acheaw, E.: Analysis of vibration characteristics of centrifugal pump mechanical seal under wear and damage degree. Shock and Vibration **2021**, 1–9 (2021)
23. Moya-Albor, E., Brieva, J., Ponce, H., Martínez-Villaseñor, L.: A non-contact heart rate estimation method using video magnification and neural networks. IEEE Instrumentation & Measurement Magazine **23**(4), 56–62 (2020)
24. Oh, T.H., Jaroensri, R., Kim, C., Elgharib, M., Durand, F., Freeman, W.T., Matusik, W.: Learning-based video motion magnification. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 633–648 (2018)
25. Oliveto, G., Santini, A., Tripodi, E.: Complex modal analysis of a flexural vibrating beam with viscous end conditions. Journal of Sound and Vibration **200**(3), 327–345 (1997)
26. Pan, Z., Geng, D., Owens, A.: Self-supervised motion magnification by backpropagating through optical flow. Advances in Neural Information Processing Systems **36** (2024)
27. Qiu, Q., Lau, D.: Defect detection in frp-bonded structural system via phase-based motion magnification technique. Structural Control and Health Monitoring **25**(12), e2259 (2018)
28. Sarrafi, A., Mao, Z., Niezrecki, C., Poozesh, P.: Vibration-based damage detection in wind turbine blades using phase-based motion estimation and motion magnification. Journal of Sound and vibration **421**, 300–318 (2018)

29. Simoncelli, E.P., Freeman, W.T.: The steerable pyramid: A flexible architecture for multi-scale derivative computation. In: Proceedings., International Conference on Image Processing. vol. 3, pp. 444–447. IEEE (1995)
30. Simoncelli, E.P., Freeman, W.T., Adelson, E.H., Heeger, D.J.: Shiftable multiscale transforms. IEEE transactions on Information Theory **38**(2), 587–607 (1992)
31. Singh, J., Murala, S., Kosuru, G.: Multi domain learning for motion magnification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13914–13923 (2023)
32. Śmieja, M., Mamala, J., Prażnowski, K., Ciepliński, T., Szumilas, Ł.: Motion magnification of vibration image in estimation of technical object condition-review. Sensors **21**(19), 6572 (2021)
33. Takeda, S., Akagi, Y., Okami, K., Isogai, M., Kimata, H.: Video magnification in the wild using fractional anisotropy in temporal distribution. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
34. Takeda, S., Isogai, M., Shimizu, S., Kimata, H.: Local riesz pyramid for faster phase-based video magnification. IEICE Transactions on Information and Systems. **103**(10), 2036–2046 (2020)
35. Takeda, S., Niwa, K., Isogawa, M., Shimizu, S., Okami, K., Aono, Y.: Bilateral video magnification filter. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
36. Takeda, S., Okami, K., Mikami, D., Isogai, M., Kimata, H.: Jerk-aware video acceleration magnification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
37. Tilbrook, M., Rozenburg, K., Steffler, E., Rutgers, L., Hoffman, M.: Crack propagation paths in layered, graded composites. Composites Part B: Engineering **37**(6), 490–498 (2006)
38. Vernekar, K., Kumar, H., Gangadharan, K.: Gear fault detection using vibration analysis and continuous wavelet transform. Procedia Materials Science **5**, 1846–1852 (2014)
39. Wadhwa, N., Rubinstein, M., Durand, F., Freeman, W.T.: Phase-based video motion processing. ACM Transactions on Graphics (TOG) **32**(4), 1–10 (2013)
40. Wadhwa, N., Rubinstein, M., Durand, F., Freeman, W.T.: Riesz pyramids for fast phase-based video magnification. In: IEEE International Conference on Computational Photography (ICCP). IEEE (2014)
41. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
42. Wu, H.Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., Freeman, W.: Eulerian video magnification for revealing subtle changes in the world. ACM transactions on graphics (TOG) **31**(4), 1–8 (2012)
43. Yang, B.S., Lim, D.S., Tan, A.C.C.: Vibex: an expert system for vibration fault diagnosis of rotating machinery using decision tree and decision table. Expert Systems with Applications **28**(4), 735–742 (2005)
44. Zhang, Y., Pintea, S.L., Van Gemert, J.C.: Video acceleration magnification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)