

Received 31 October 2022, accepted 18 November 2022, date of publication 24 November 2022,  
date of current version 30 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3224601

## RESEARCH ARTICLE

# Phase-Based Motion Magnification for Structural Vibration Monitoring at a Video Streaming Rate

JAE YOUNG AN AND SOO IL LEE<sup>✉</sup>

Department of Mechanical and Information Engineering, University of Seoul, Seoul 02504, South Korea

Corresponding author: Soo Il Lee (leesooil@uos.ac.kr)

This work was supported by the 2019 Research Fund of the University of Seoul under Grant 201904301017.

**ABSTRACT** Here we propose a novel approach to image magnification that enables frame-by-frame motion magnification application at a video streaming rate of around 30 fps. This novel approach can be used instead of batch processing an image file in conventional phase-based magnification (PBM). This new PBM method can instantly show a magnified video streaming on display at 30 fps, which is helpful for vibration measurement and for monitoring tasks where magnified images need to be viewed simultaneously. To accomplish video streaming-rate magnification, the proposed PBM uses time-domain convolution in temporal bandpass filtering for frame-by-frame operation whereas conventional PBM employs a frequency-domain filter that is applied to the entire image file at once. An experiment was conducted to monitor the vibration of a cantilever using a webcam streaming at the same frame rate, and data were collected simultaneously using a laser Doppler vibrometer for comparison. The experiment confirmed that the proposed PBM approach is more effective than the conventional magnification method and it also analyzed the system performance for vibration measurement. Additionally, the proper orthogonal mode could be found through the singular value decomposition from vibration displacement data of the cantilever that was collected instantly from the magnified image frames. Furthermore, the dominant mode could be effectively extracted from excitation at the resonance frequency. Because of magnification factors, the vibration displacements from the proposed method were estimated using linear regression and the accuracy of the estimated displacements was within the permissible error bound. Since the proposed PBM is a frame-by-frame operation, instantaneous adjustment of magnification parameters is available, even while the magnification is being processed. In addition, the proposed PBM is independently adjustable to the number of image frames for temporal FIR filter order.

**INDEX TERMS** Phase-based magnification, time-domain convolution, video streaming rate, vibration monitoring.

## I. INTRODUCTION

To circumvent the limitations of commonly used vibration sensors, such as accelerometers or laser vibrometers, computer vision-based techniques utilizing high-speed or conventional digital cameras have been investigated [1], [2], [3]. The camera-based measurement method is unaffected by loading and temperature errors and has the advantage of obtaining a full-field image [4], [5], [6], [7]. Nevertheless, data acquisition from video frames is challenging when the vibration motion tracked in video images is as tiny as subpixel; however, small vibrations can be amplified using motion

magnification algorithms [8], [9], [10], [11], [12], [13] and can be applied to vibration measurements [14], [15], [16].

Liu et al. [8] developed an approach to visually show tiny motions in an image by employing the Lagrangian description used in fluid mechanics. The Lagrangian approach is a method for tracking individual particles with high precision, and it has a high computational cost. On the other hand, Wu et al. [9] demonstrated an effective technique called Eulerian video magnification (EVM) that is based on the Eulerian description. EVM works similarly to optical flow in that it amplifies the pixel intensities that change at a fixed pixel position. While this Eulerian magnification considerably reduces the amount of computing required, a high magnification factor cannot be applied to high-frequency

The associate editor coordinating the review of this manuscript and approving it for publication was Joewono Widjaja<sup>✉</sup>.

motion and noise increases linearly as the magnification factor increases [10]. To address this issue, Wadhwa et al. [10] suggested phase-based magnification (PBM) using a complex steerable pyramid (CSP) in the Eulerian framework. [17] Magnification factors greater than those used in conventional Eulerian magnification can be used, providing a magnified image that is more robust to artifacts and noise.

Despite the fact that CSP has a long computation time due to overcompleteness and a computational disadvantage [11], [17], it has substantial advantages in terms of obtaining accurate magnified motion and in detecting motion in many directions. Therefore, a method for reducing computing time by employing the Riesz pyramid, instead of CSP, has been proposed [11] for calculating faster using the local Riesz pyramid approach [13]. However, compared to the CSP, the Riesz pyramid has a disadvantage in detecting motion in various directions, and the local Riesz pyramid has a limitation in that motion is weakly magnified. Oh, et al. [12] proposed a learning-based motion magnification model based on convolutional neural network architecture. Their learning-based method demonstrated superior performance in terms of noise and accuracy without a temporal filter. This strategy, however, has disadvantages of requiring a large amount of training data and time to construct a trained model. Despite these alternatives, many vision-based vibration analysis studies still use the steerable pyramid-based motion magnification [14], [15], [16].

In contrast, the conventional phase-based motion magnification approach can examine the outputs after all frames of a finite-length video clip have been processed, but many applications require video streaming-rate processing. In conventional PBM, all video frames are processed simultaneously, and all video data are allocated and processed in memory. The processing time is proportional to the image resolution and frame number. However, PBM applications, such as structural health monitoring, require instant results, and real-time level performance for applying motion magnification would be beneficial.

This paper introduces a novel approach that allows frame-by-frame processing instead of batch processing of image files in conventional PBM, allowing motion magnification to be applied at the video streaming rate. In addition, vibration measurement for a simple cantilever example was conducted by applying magnification from real-time streaming video. We propose an algorithm for frame-by-frame evaluation of the magnification output by converting it from a bandpass filter in the frequency domain to a convolution in the time domain for video streaming processing. This streaming-rate magnification is implemented on a small portable laptop equipped with a GPU at a processing speed of 30 fps as the webcam's frame rate. An experiment was conducted to monitor the vibration of a cantilever using a webcam streaming at the same frame rates, and data were collected simultaneously using a laser Doppler vibrometer (LDV) for reference. The experiment confirmed that the proposed approach

is as effective as the conventional magnification method in many circumstances, and it also evaluated the system validity for vibration measurement in displacement, frequency, and vibration modes.

## II. PHASE-BASED MOTION MAGNIFICATION FOR VIBRATION MEASUREMENT

### A. PHASE-BASED MOTION MAGNIFICATION

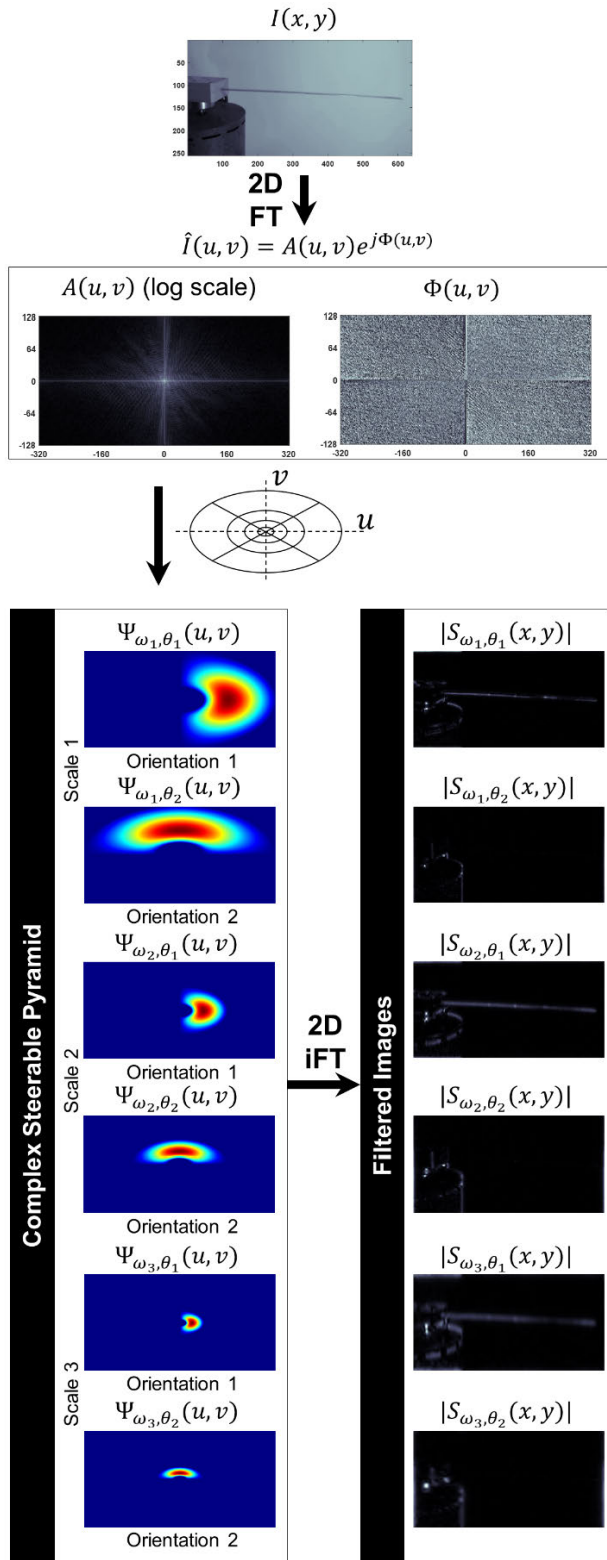
The core idea of phase-based motion magnification is a CSP that decomposes an image into complex-valued data scales. As a result, phase-based motion magnification can handle image motion data expressed as a phase angle difference. The CSP [10] consists of multi-level complex steerable filters according to multiple orientations and spatial frequency bands. As illustrated in Fig. 1, original image  $I(x, y)$  is transformed by 2D Fourier transform into amplitude  $A(u, v)$  and phase  $\Phi(u, v)$  regarded as complex-valued transformed image data  $\hat{I}(u, v) = A(u, v)e^{j\Phi(u, v)}$ , where  $(x, y)$  is geometry coordinates and  $(u, v)$  is image frequency coordinates. Then, steerable filters  $\Psi_{\omega, \theta}(u, v)$  are applied.  $\Psi_{\omega, \theta}(u, v)$  is a filter component of the CSP that corresponds to the spatial frequency bands with scale levels  $\omega$  and orientation  $\theta$  in the frequency domain. This description is also a polar coordinate scheme of the spatial frequency domain  $(u, v)$  because  $\omega$  and  $\theta$  represent the distance and the direction from the origin, respectively, as shown in Fig. 1.

After applying CSP, the results were show the magnitude of inverse Fourier transformed image (of image geometry domain), which displays multi-orientation and multi-scale frequency components  $|S_{\omega, \theta}(x, y)|$  in Fig. 1. The outer region corresponds to the high frequency in the spatial frequency domain, which correlates with a sharp change in pixel intensity value or patterns with a short wavelength. On the other hand, the inner region correlates with a low frequency, implying smooth variations in pixel intensity or long-wavelength patterns in the image. The direction from the origin to a specific steerable filter region in frequency coordinates is the orientation that the filter extracts, as shown in Fig. 1. For example, the center point of filter  $\Psi_{\omega_3, \theta_1}(u, v)$  at Scale 3 in Fig. 1 is horizontally aligned with the origin. Orientation  $\theta_1$ , therefore, represents the horizontal features of the image so that the shape of the cantilever is well shown as  $|S_{\omega_3, \theta_1}(x, y)|$ .

The image reconstruction is a sequence of each spatial frequency component, demonstrating mathematically how magnification proceeds. The complex steerable filter is comprised of windowing and sinusoids in the spatial domain. With spatial domain  $(x, y)$  defined as vector  $\mathbf{x}$ , the spatial frequency of a decomposed image,  $S_{\omega}(\mathbf{x})$ , applied with a filter of windowing,  $W(\mathbf{x})$ , and spatial frequency  $\omega$  can be described as follows:

$$S_{\omega}(\mathbf{x}) = I(\mathbf{x}) \otimes W(\mathbf{x})e^{j\omega\mathbf{x}} = |S_{\omega}(\mathbf{x})| e^{j\omega\mathbf{x}}, \quad (1)$$

where the  $\otimes$ (circled times) represents the spatial-domain convolution and  $W(\mathbf{x})e^{j\omega\mathbf{x}}$  is the inverse Fourier transform of steerable filter  $\Psi_{\omega}(\mathbf{u})$ , where  $\mathbf{u}$  is a vector expression of  $(u, v)$ .



**FIGURE 1.** A complex steerable pyramid (CSP) decomposes an original image (top) into scale and orientation components (bottom left). After inverse Fourier transform through CSP, the filtered images match each scale and orientation (right).

As the purpose of reconstructing an image is to collect all frequency image components,  $S_{\omega}(\mathbf{x})$ , the image notation can

be represented as an infinite series

$$I(\mathbf{x}) = \sum_{\omega=-\infty}^{\infty} |S_{\omega}(\mathbf{x})| e^{j\omega\mathbf{x}}. \quad (2)$$

If  $I(\mathbf{x})$  is an initial state and it moves some position over time  $t$ , the value yields

$$I(\mathbf{x} + \delta_{\mathbf{x}}(t)) = \sum_{\omega=-\infty}^{\infty} |S_{\omega}(\mathbf{x})| e^{j\omega(\mathbf{x} + \delta_{\mathbf{x}}(t))}, \quad (3)$$

where  $\delta_{\mathbf{x}}(t)$  is the subtle motion of the image. The phase difference between the first and current image frames is calculated using the phase angle in (2) and (3).

$$\begin{aligned} \Delta\Phi(\mathbf{x}) &= \arg \left[ |S_{\omega}(\mathbf{x})| e^{j\omega(\mathbf{x} + \delta_{\mathbf{x}}(t))} \right] - \arg \left[ |S_{\omega}(\mathbf{x})| e^{j\omega\mathbf{x}} \right] \\ &= \omega \delta_{\mathbf{x}}(t), \end{aligned} \quad (4)$$

where  $\Delta\Phi$  denotes the phase difference between the two images. As can be seen in the equation, phase difference  $\Delta\Phi$  encompasses the phase difference  $\delta_{\mathbf{x}}(t)$ . To amplify the phase difference, we apply a magnification factor  $\alpha$  to  $\Delta\Phi$ , which is then  $\exp(j\alpha\Delta\Phi)$  multiplied by (3) to yield the magnified image as

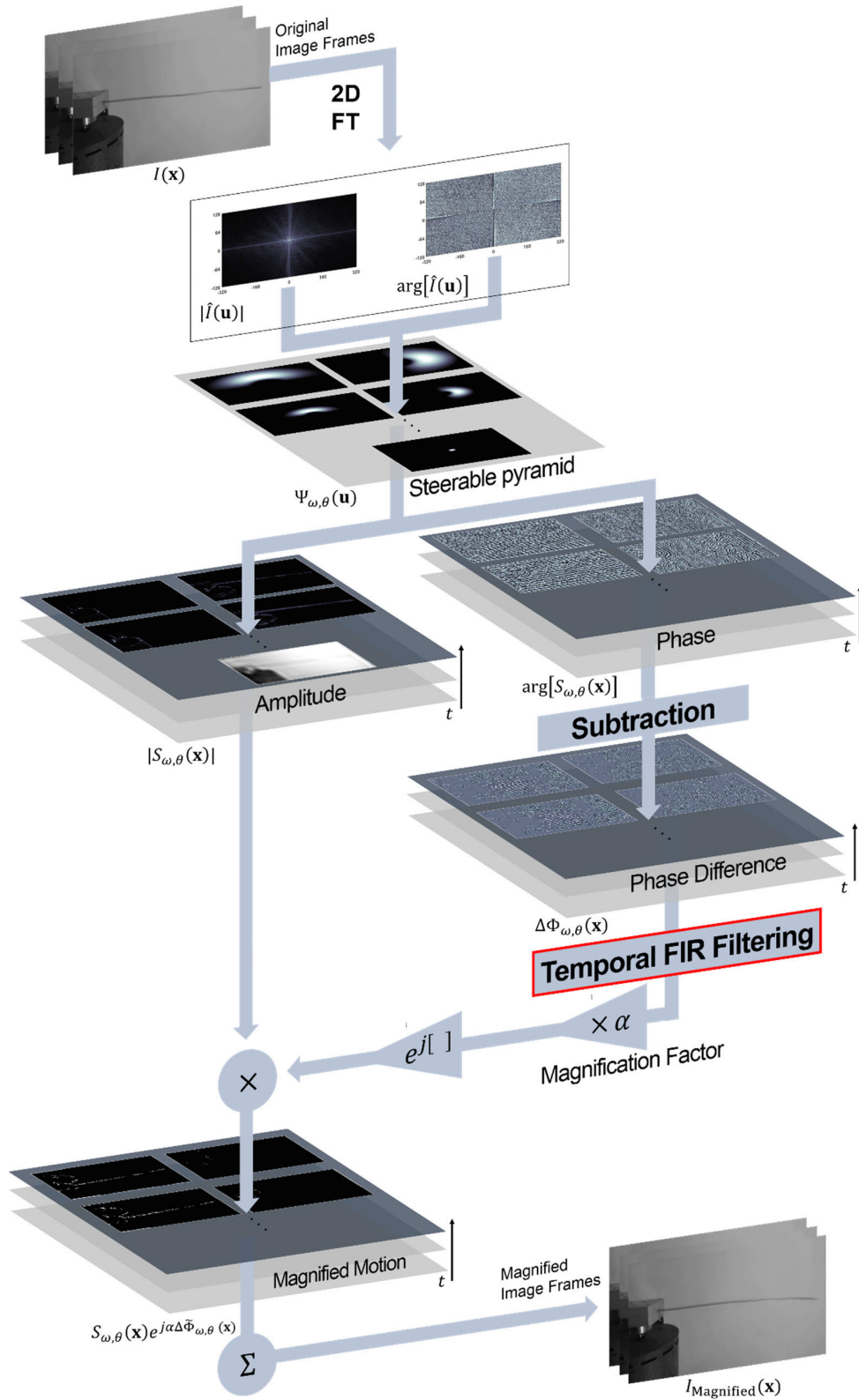
$$\sum_{\omega=-\infty}^{\infty} |S_{\omega}(\mathbf{x})| e^{j\omega(\mathbf{x} + \delta_{\mathbf{x}}(t))} e^{j\omega\alpha\Delta\Phi(\mathbf{x})} = I(\mathbf{x} + (1 + \alpha)\delta_{\mathbf{x}}(t)). \quad (5)$$

Therefore, motion-magnified images can be acquired in (5).

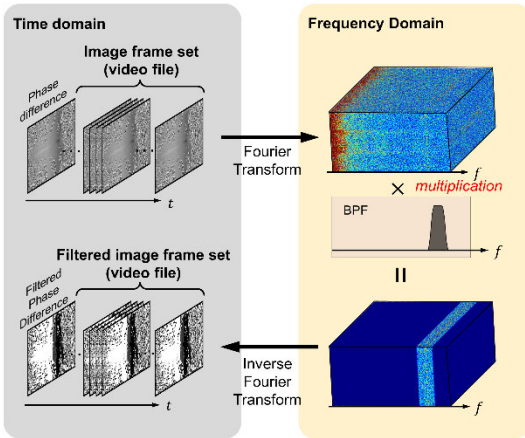
Practical phase-based video motion processing procedures are illustrated in Fig. 2 using PBM. Although not explicitly stated above,  $\Delta\Phi$  must be subjected to temporal band-pass filtering prior to being inserted in order to focus on motion and eliminate the DC component [10]. Consequently, Fig. 2 depicts the filtered phase difference,  $\Delta\tilde{\Phi}$ , following temporal filtering. Additionally, while the orientation of the steerable filter is not depicted individually in Fig. 2 for clarity of explanation, the actual process of image decomposition must take the orientation into account.

### B. TEMPORAL FILTERING AND CONVOLUTION

When contemplating frequency band filters to magnify the motion of a particular frequency component, either a finite impulse response (FIR) or an infinite impulse response (IIR) filter might be considered. As shown in Fig. 2 as ‘‘Temporal FIR Filtering,’’ located between the phase difference frame stack and the magnification factor, the FIR filter is used due to its implementation simplicity and frame-by-frame processing capability in video processing [18]. The response of the FIR filter has a linear phase and is always stable, making it excellent for video processing, even during online streaming. However, IIR filters have a non-linear phase response and they cannot be guaranteed to operate in a stable manner [18]. In addition, because FIR filters may be designed to behave similarly to IIR filters, even with lower filter orders, this high-speed attribute contributes to the processing time reduction.



**FIGURE 2.** The general concept of phase-based motion magnification (PBM). 1) In the frequency domain, a CSP with scales and orientations is applied to the image. After decomposition and filtering, the images contain complex values. 2) The phase angle of the decomposed image and the phase angle difference reflecting the motion is calculated. 3) Temporal filtering is applied to the phase difference in order to extract the motion of the specified frequency component. 4) The filtered phase difference is amplified by the magnification factor and then combined with the original frame. 5) After completing the preceding procedures on each CSP-decomposed local image, the reconstructed image exhibits a magnified motion.



**FIGURE 3.** Schematic illustrating how conventional filtering employs frequency-domain temporal filtering on all frames in a batch.

Conventional PBM [10] employs a Fourier transform to perform bandpass filtering of an entire set of phase frames in batch processing at once, as shown in Fig. 3. Let  $b(t)$  be the FIR filter coefficients in the time domain prepared using the windowing method [19]. Then, temporal filtering of phase difference under the conventional method is represented as

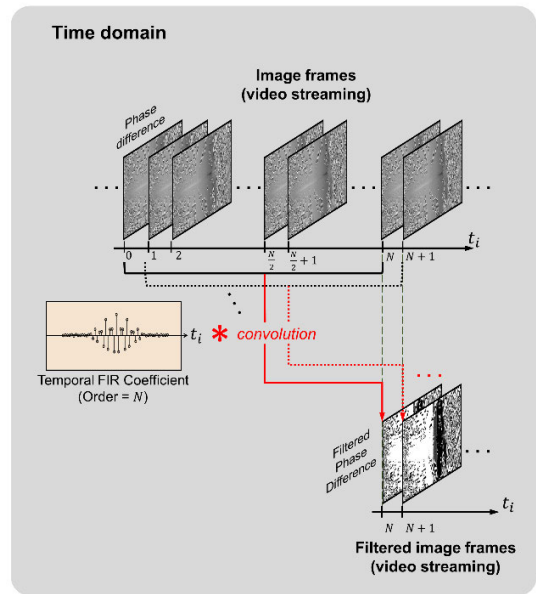
$$\Delta \tilde{\Phi}(t) = \mathcal{F}^{-1} [\hat{b}(f) \times \Delta \hat{\Phi}(f)], \quad (6)$$

where  $\Delta \tilde{\Phi}(t)$  is the filtered phase difference in the time domain. Moreover,  $\hat{b}(f)$  and  $\Delta \hat{\Phi}(f)$  are the Fourier transform of filter coefficients  $b(t)$  and phase difference  $\Delta \Phi(t)$ , respectively. Since the discrete Fourier transform returns a spectral array with the same length as the input signal array, the FIR filter coefficients will be of the same size as the filter length. This temporal filtering is a multiplication operation in the frequency domain, as shown in Fig. 3. Since the two arrays must have the same size, the length of the FIR filter must be the same as the length (number of frames) of the image sequence. This frequency-domain multiplication requires a higher-order filter, leading to much computation but not improvement in filtering performance as the filter order increases. Also, it is impossible to ascertain whether the filtering band frequency and magnification factor are correctly set in processing until the entire frame is processed.

The proposed PBM conducts temporal bandpass filtering using a convolution directly in the time domain, rather than via frequency-domain operation, as denoted by

$$\Delta \tilde{\Phi}(t) = b(t) * \Delta \Phi(t), \quad (7)$$

where the \* (asterisk) represents the time-domain convolution. As illustrated in Fig. 4, in actual time-domain convolution, the digital filter operation is accomplished with a presenting number of phase image frames by matching the discrete FIR coefficients to the filter length. Considering  $N = 2n$  as an even-numbered discrete filter order, the FIR filter coefficient of order  $N$  is  $b_N$ , which is also the impulse response of the filter. Then, the filter length of  $b_N$  is  $N + 1$  and the center frame is on  $n$ . From (7), the discrete version of the



**FIGURE 4.** Schematic illustrating how the proposed phase-based magnification technique uses time-domain convolution to perform temporal filtering on a single frame sequence. The proposed PBM supports frame-by-frame execution via image frame streaming.

filtered phase difference and the time-domain convolution are described as

$$\Delta \tilde{\Phi} [t_{i-n}] = \sum_{k=0}^N b_N [k] \times \Delta \Phi [t_i - k], \quad (8)$$

where  $t_i$  represents  $i$ th discrete instance of the temporal frame. As the filter has a group delay of  $n$  image frames, which is half of the filter order [19], the result of the convolution corresponds to the phase difference at  $t_{i-n}$ , as shown in (8). The discretized time convolution (8) is employed instead of (7) in the actual code implementation of temporal filtering.

This filtering approach produces a single frame-by-frame result for each processing iteration, allowing instantaneous adjustment of the frequency bands and magnification factors. Additionally, the magnified image can be monitored in real-time, which is advantageous for online video streaming applications. Compared to conventional PBM, the proposed approach is not order-dependent, hence computation is minimized. As a result, the user can control the order of the filters to improve processing time and image quality.

### III. IMPLEMENTATION AND EXPERIEMENTS

#### A. IMAGE MAGNIFICATION SOFTWARE

The proposed PBM technique, which enables frame-by-frame magnification processing, was implemented in Python, executed on a laptop PC equipped with a GPU, and obtained a video streaming rate of 30 fps. However, there are intensive operations for multi-dimensional arrays in the motion magnification image, so an ordinary CPU-only PC is insufficient. A GPU is required for faster processing, even without optimized Python code. The proposed PBM code was tested

```

proposedPBM_main.py
1   while(1):
2       frame = ReadImage()
3       decomposedFrames = ComplexSteerablePyramidDecomposition(frame)
4
5       for level in range(0, pyramidLevel):
6           # step 1: calculate the phase angle difference of the decomposed images
7           phaseDifference = PhaseSubtraction(decomposedFrame[level])
8
9           # step 2: take temporal filtering on phase difference
10          filteredPhaseDifference = TemporalFiltering(phaseDifference)
11
12          # step 3: magnify the filtered phase difference and combine
13          #         decomposed images
14          magnifiedPhaseDifference = alpha*filteredPhaseDifference
15          magnifiedDecomposedFrames[level] = \
16              decomposedFrame[level]*exp(1j*magnifiedPhaseDifference)
17
18          # step 4: reconstruct by gathering all decomposed images
19          magnifiedFrame = ComplexSteerablePyramidReonstrcution(
20              magnifiedDecomposedFrames
21          )
22
23          # step 5: save the motion-magnified image
24          WriteImage(magnifiedFrame)

```

**FIGURE 5.** The main routine of the proposed phase-based magnification code implemented in Python (pseudo-code).

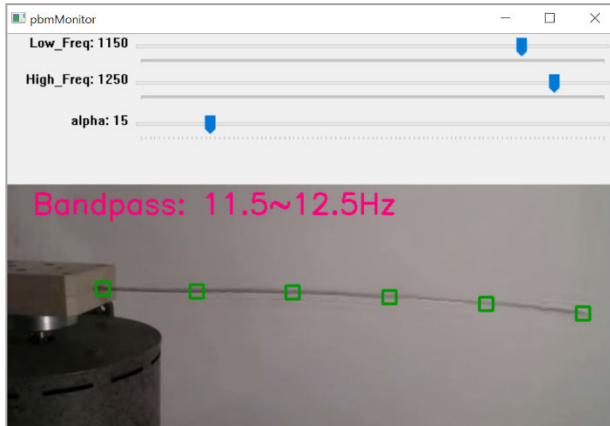
on a small laptop equipped with an Intel Core i7 mobile CPU and an NVIDIA GTX1650 GPU to process webcam streaming ( $640 \times 256$ -pixel resolutions at 30 fps) at the same processing capabilities (at least 1/30-s each frame). The PBM software was developed in Python (version 3.7) with image processing and temporal FIR filtering implemented using OpenCV (version 4.5.3) and SciPy (version 1.7) packages. While the NumPy (version 1.21.1) module is the de facto standard for array operations in Python programs, the CuPy (version 9.2 with CUDA 10.1) module was primarily used here for working with arrays utilizing the GPU. NumPy was only used to convert the processed data array back to an image at the end. Fig. 5 shows the pseudo-code of the core routine used to implement the proposed PBM algorithm with the temporal convolution illustrated in Fig. 4. Within the main loop, the camera reads the frame and the program returns a result when processing is complete; the program then obtains the next frame and repeats the subsequent calculation, providing real-time processing via frame-by-frame computation.

The following algorithm was used. (1) After filtering the image with the steerable pyramid, the phase angle difference

is extracted by subtraction. (2) Direct temporal filtering with convolution is employed to filter the object motion equal to the subtracted phase whereas the conventional method employs a Fourier and inverse Fourier transform of the time domain before and after bandpass filtering. (3) After multiplying the filtered motion by the magnification factor, it is merged into the existing frame. (4) Images from steps (1) to (3) are re-merged for each level, along with the residual image from the low-pass filter. (5) The magnified motion image is saved. If checking is required, showing of the magnified images, as illustrated in Fig. 6, can be added. While the image is being streamed, steps (1) to (5) can be repeated indefinitely.

## B. EXPERIMENT

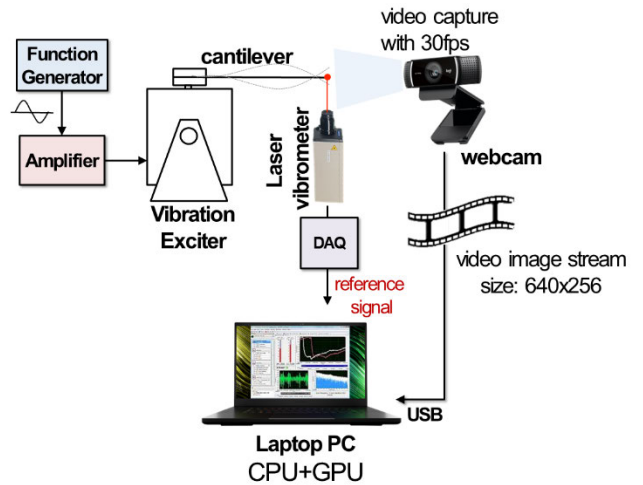
To verify the performance of the proposed PBM method, which operates at a video rate, an experiment to measure cantilever beam vibration was designed and compared with batch processing of entire frames in the conventional PBM method. The key parameters affecting the processing speed of the proposed real-time method are the size (resolution) of the image and the filter order chosen for temporal convolution,



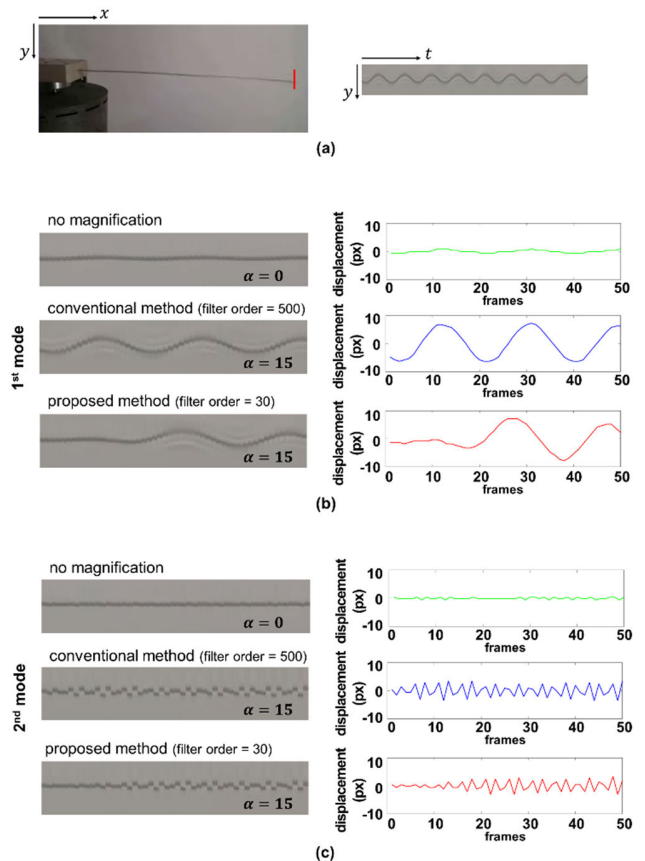
**FIGURE 6.** Screenshot executing the proposed phase-based magnification code with slide bars for adjusting the frequency band and the magnification factor.

which is dependent on the number of incorporated image frames. The processing time for the proposed PBM method on an image size of  $640 \times 256$  can be achieved by 1/30-s per frame when the FIR filter order is 30. Therefore, we set the temporal FIR filter order to 30 to achieve a video streaming rate of 30 fps. In addition, the total length of the video file is limited to 500 frames for convenience though the proposed method can be applied to a continuous online video sequence. We also used a conventional PBM of the same length as the recorded video. As mentioned in Section II.B, the filter length of the conventional PBM method should be set the same as the frame length; therefore, the filter order is set to 499 as the filter length is 500. Moreover, the measurable frequency at 30-fps video is up to 15 Hz according to sampling theory.

As shown in Fig. 7, a thin metal cantilever was mounted to an electromagnetic exciter and a function generator was used to generate an excitation signal at a specific frequency. An experiment was conducted to magnify images captured through a webcam connected to a laptop PC. The actual amplitude and frequency were determined using a reference LDV. The experimental cantilever was constructed of stainless steel of 20.8-g and was 0.5-mm-thick, 15-mm-wide, and 400-mm-long with uniform cross-section. It was established that the independently measured natural frequencies of the cantilever were 1.53 and 12.2 Hz in the first and second resonance modes, respectively. The applied frequencies of the first two cantilever modes were less than 15 Hz, which is the Nyquist limit in the experiment. When the base of the cantilever was excited with a tiny excitation force according to the natural frequency of each mode,  $640 \times 256$ -pixel images at 30 fps were captured and processed using the conventional and the proposed methods. The benchmark time was recorded for each program step in each method. The vibration displacement was extracted from the image using the centroid tracking method [20] and compared with reference displacement data acquired simultaneously with the LDV.



**FIGURE 7.** Experimental setup for cantilever vibration measurement with webcam image streaming and reference LDV.



**FIGURE 8.** Transverse vibration displacement of the cantilever tip captured from the magnified image with the centroid tracking method followed by a red line segment, as shown in (a). (b) and (c) show extracted data in the case of the 1st and 2nd mode frequencies, respectively. The magnification factors were set at  $\alpha = 15$ .

### C. RESULT

We obtained the free-tip displacements of the cantilever from the magnified video using the centroid tracking method [20]. Fig. 8(a) shows the frame stack process for extracting the

TABLE 1. Estimated displacement from the intercept of the fitted curve.

	RMS Amplitude of Displacement (mm)			
	LDV	No Magnification	Conventional Method	Proposed Method
	Measurement	(calibrated)	(calibrated & rescaled)	(calibrated & rescaled)
Mode 1	0.271	0.345	0.299	0.224
Mode 2	0.228	0.307	0.253	0.236

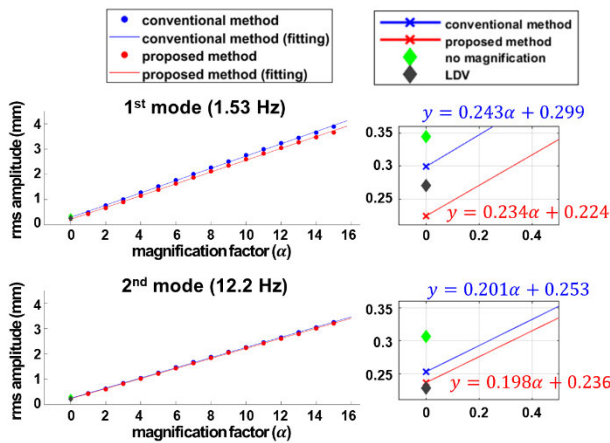


FIGURE 9. RMS amplitude of vibration displacement increases linearly as the magnification factor,  $\alpha$ , increases and can be expressed as a linear regression. The y-intercept value of the regression equation is the estimated RMS amplitude expected without magnification ( $\alpha = 0$ ). The estimated RMS amplitudes for both PBM methods are compared with those of the no-magnification (original) image and the LDV measurement.

tip displacement of the cantilever from the magnified image. First, the pixel displacement data were extracted from the original (unmagnified), the conventional PBM, and the proposed PBM videos, respectively, for the case of  $\alpha = 15$ , as shown in Fig. 8(b) and (c). Then, the pixel values were conventional and proposed methods exhibit identical magnified displacement waveforms, as shown in Fig. 8(b) and (c), the proposed method only provides the magnification after specific frames of the transient response due to the group delay aspect of the FIR filter design [19]. Indeed, steady-state magnified displacement may be obtained at the video rate after the initial transient delay, even with the proposed method. After an initial delay of as many frames as the half order of the temporal filter, magnification processing is carried out at the frame rate.

In theory, since both conventional and proposed PBMs use a linear magnification mechanism, the displacement is linearly magnified by magnification factor  $\alpha$ . The linear relationship between magnified root-mean squared (RMS) displacement amplitudes and magnification factors ranging from  $\alpha = 1$  to 15 is evident in Fig. 9, as is the linear regression with excellent correlation. The amplitude value at the amplitude-axis intercept of the regression line seen in Fig. 9 (when  $\alpha = 0$ ) can be regarded as the de-scaled value from the magnified amplitude and it is approximated as the actual

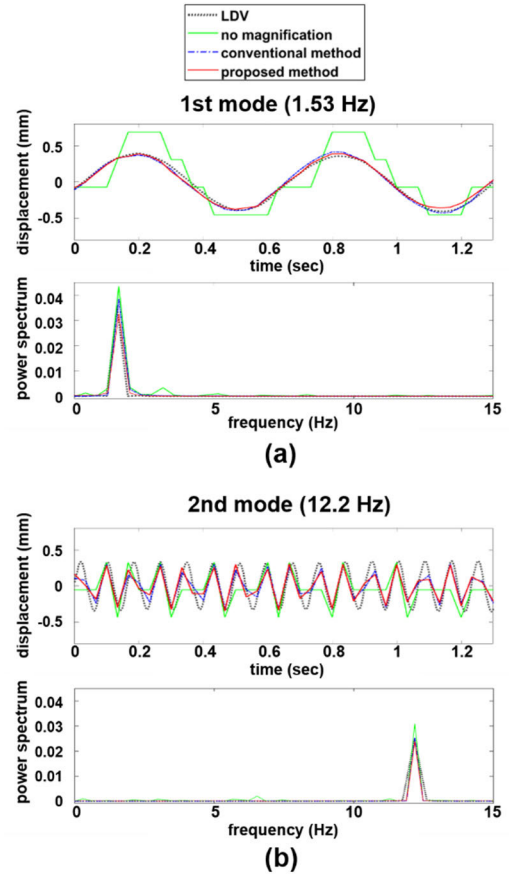


FIGURE 10. Data obtained through displacement extraction from the  $\alpha = 15$  video, rescaled by the pixel-to-metric conversion, and then compared with the LDV reference data. (a) and (b) show the 1st and 2nd mode frequency cases, respectively.

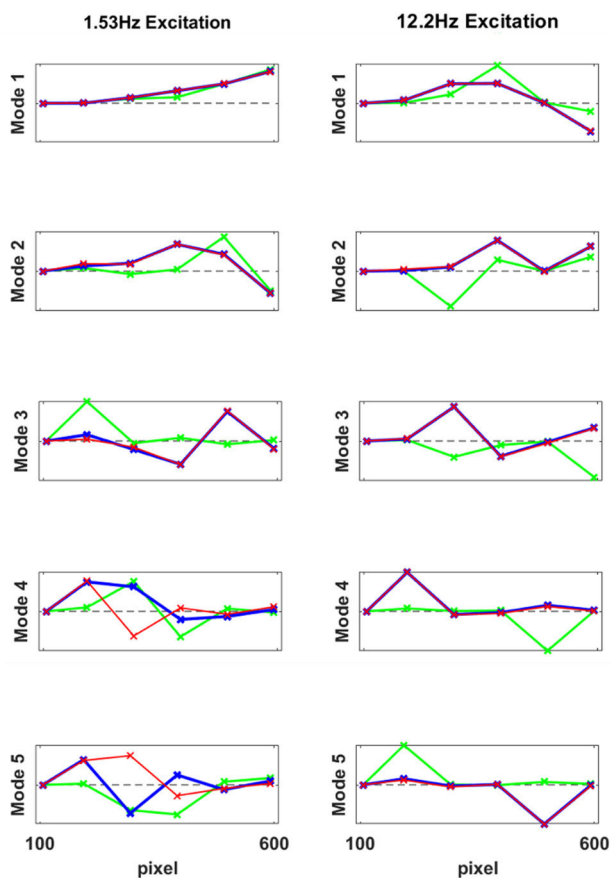
unmagnified amplitude of the vibration. The estimated displacement amplitude differs slightly from the displacement amplitude extracted from the unmagnified image. Instead, the estimated displacement amplitude from the magnified image is much closer to the LDV reference amplitude. The calibrated and estimated displacement RMS amplitudes for each case and for the LDV measurement data listed in Table 1.

As shown in Table 1 and Fig. 10, the data obtained from the original unmagnified image, which imposes subpixel displacement, is overestimated in comparison to the actual LDV reference data, even after calibration. It is known that calibrated data obtained by rescaling a magnified image yield



**TABLE 2.** Relative proportion of singular values of a vibrating cantilever with base excitation using SVD.

POM	1.53-Hz excitation (1 <sup>st</sup> modal resonance)			12.2-Hz excitation (2 <sup>nd</sup> modal resonance)		
	No Magnification $\sigma$ (%)	Conventional PBM $\sigma$ (%)	Proposed PBM $\sigma$ (%)	No Magnification $\sigma$ (%)	Conventional PBM $\sigma$ (%)	Proposed PBM $\sigma$ (%)
Mode 1	8.98 (50.24)	135.20 (81.22)	126.49 (76.46)	13.69 (45.85)	115.60 (89.27)	116.65 (88.33)
Mode 2	3.08 (17.23)	15.87 (9.53)	17.53 (10.60)	8.98 (30.06)	5.91 (4.57)	6.43 (4.87)
Mode 3	2.65 (14.83)	7.46 (4.48)	9.45 (5.71)	4.71 (15.77)	3.70 (2.85)	4.39 (3.32)
Mode 4	1.64 (9.19)	4.07 (2.44)	6.58 (3.98)	1.54 (5.16)	2.36 (1.82)	2.46 (1.86)
Mode 5	1.52 (8.51)	3.88 (2.33)	5.37 (3.25)	0.94 (3.16)	1.92 (1.49)	2.13 (1.61)
Optimal SVHT	3.81	10.72	13.58	6.77	5.32	6.31



**FIGURE 11.** POMs of a vibrating cantilever with base excitation using SVD. The mode extracted from the magnified image is closer to the resonance mode at the excitation frequency than the original unmagnified image. (a) and (b) show the 1st and 2nd mode frequency case, respectively.

a more precise result. In Fig. 10, each frequency can also be estimated from the power spectrum transformed from the temporal displacements. It can be seen that the estimated frequency obviously coincides well with the excitation frequency of the corresponding mode. The excitation frequencies are 1.53 and 12.2 Hz.

We set regions of interest of six points along the cantilever, including the base as shown in the screenshot in Fig. 6, and

simultaneously extracted dynamic displacement data of six points from the original and magnified images using the centroid tracking method. From the data extracted using singular value decomposition (SVD) as a proper orthogonal decomposition, the operational deflection shapes or proper orthogonal modes (POMs) representing the dynamic signature of the structure were obtained. These POMs represent the mode in forced vibration [22]. The forced vibration modes obtained in each case are shown in Fig. 11. The singular mode (i.e., POM) extracted from the magnified image is closer to the normal mode at the resonance frequency than the original unmagnified image. Table 2 shows the relative proportion of singular values of a vibrating cantilever with base excitation using SVD. The first POM has the largest singular value and the first singular value from the magnified image at each resonance frequency is larger than the singular value from the original (unmagnified) image. This supports extracting the vibration mode associated with the corresponding frequency by image magnification from the vibrating structure. Fig. 12 shows the modal assurance criteria (MAC) [23], which indicates how similar the POM is, according to the magnification for each excitation frequency, to the analytical resonance mode of the corresponding frequency. The first POM in 1.53-Hz excitation (1st resonance) has the largest MAC value with the first eigenmode, whereas the first POM in 12.2-Hz excitation (2nd resonance) has the largest MAC value with the second eigenmode. In particular, vibration mode extraction through SVD in the proposed PBM is as reliable as that for the conventional PBM. In this case, if important modes are selected among the measured POM modes by calculating the optimal hard threshold [24] in SVD, the values are Mode 1 or Mode 2 above the dotted line in Table 2. In each POM of Fig. 11, not only is the singular value relatively very small, but it also has a different shape from the normal mode due to the influence of noise in the case of Mode 3 or higher.

**IV. DISCUSSION**

**A. PROCESSING TIME**

The processing time of each algorithm routine was measured to evaluate whether the proposed PBM approach can sustain a processing speed of lower than 30 fps. Since the

TABLE 3. Total processing time and equivalent FPS.

Process	Step 1 Phase Subtraction	Step 2 Temporal Filtering	Step 3 Magnification	Step 4 Low-pass Residual	Step 5 Video Save	Total	Equivalent FPS
<b>Time (ms)</b>							
<b>Conventional PBM</b>	6.908	0.468	13.560	0.276	17.892	39.108	25.6
<b>Proposed PBM</b>	8.281	2.195	7.257	0.061	14.903	32.697	30.6

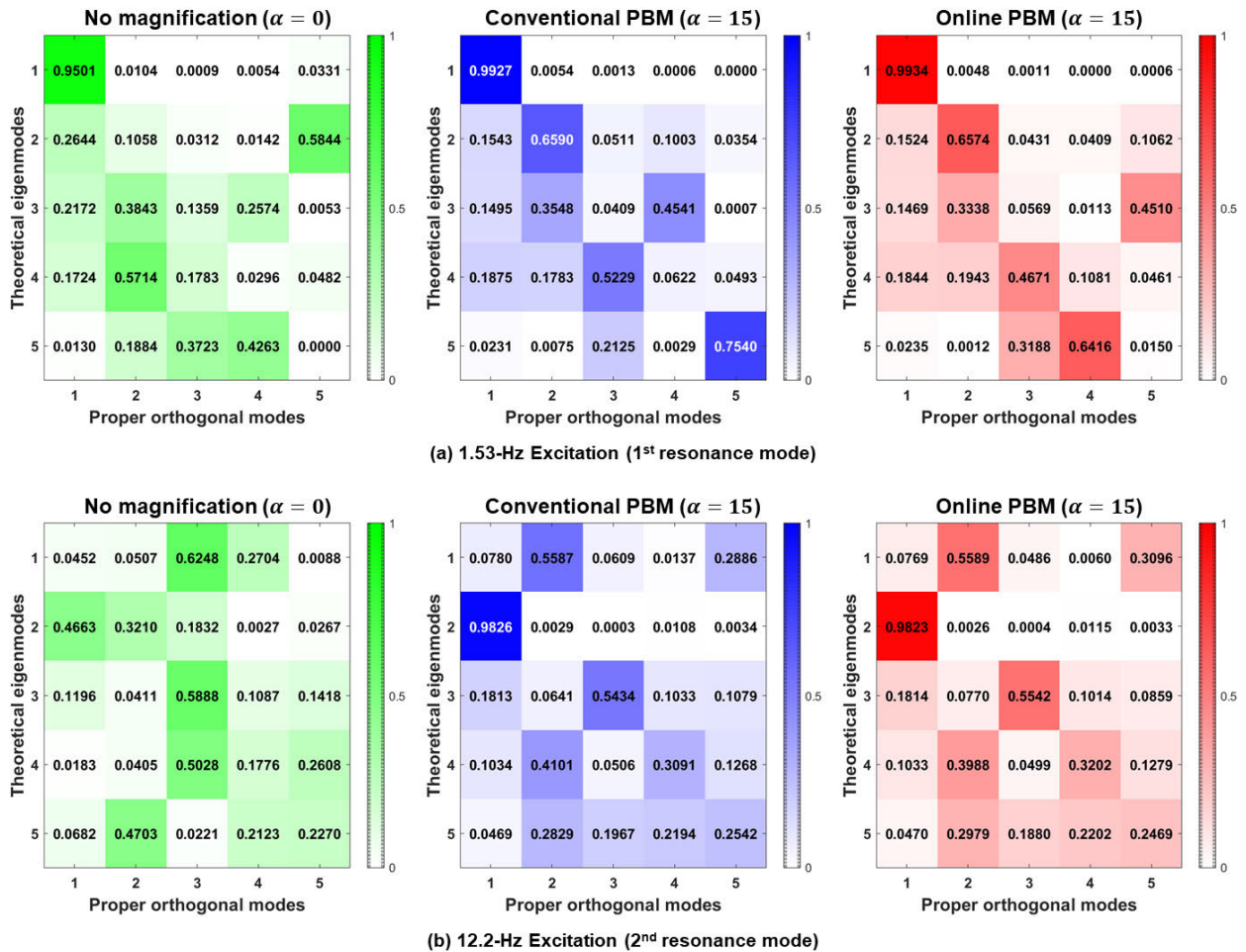


FIGURE 12. Modal assurance criteria (MAC) correlation between POMs and theoretical eigenmodes.

conventional method processes the entire set of image frames from a previously saved video file into memory, the processing time was measured at five different process steps (Table 3), ranging from applying complex steerable pyramids to saving the image frames, except for the image loading stage. The processing time is dependent on the hardware and the results reported in Table 3 are based on the laptop PC equipped with a GPU used in the experiment. The proposed method can check every single frame result in real-time, so the processing time for each iteration was measured and

averaged over 500 frames as seconds-per-frame. Since the processing time of the conventional method can be measured only as entire frames, it is measured in a total processing time and then divided by the number of frames (500 frames in this study) to evaluate the seconds-per-frame. Finally, the processing times of both methods are transformed into equivalent frame-per-seconds (fps) and compared. According to Table 3, the conventional PBM method processed each frame in 39.108 ms, whereas the proposed method processed each frame in 32.697 ms. If the frames-per-second rate exceeds

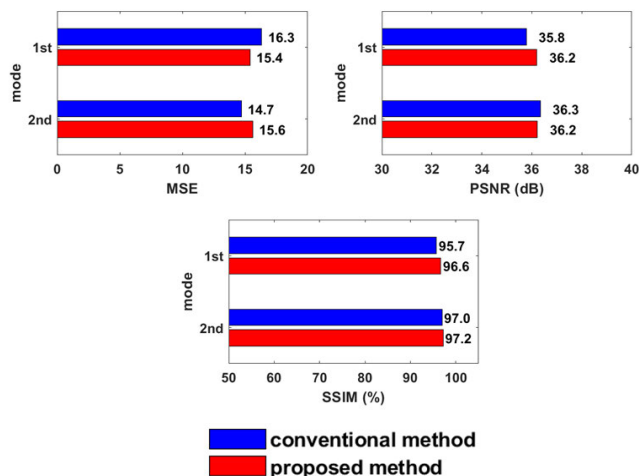


FIGURE 13. Image quality metrics applied on magnified videos with  $\alpha = 10$ .

30 fps, the images streamed from the webcam can be processed at a real-time video streaming rate. Therefore, the proposed PBM could achieve real-time performance.

The proposed method performs about 20% faster processing than the conventional method because it needs only direct temporal convolutions with the filter order from temporal filtering to the magnification process, which takes advantage of the frame-by-frame processing. However, the conventional method requires Fourier and inverse transforms and performs temporal bandpass filtering in the frequency domain via filter multiplication. The (tensor) multiplication is a computationally intensive operation for data arrays with dimensions of  $640 \times 256 \times 500$ , which correspond to the image size and total frame count in this experiment. While all transforming and filtering operations are expedited by the GPU, managing such large amounts of data concurrently may be slower due to computational and hardware constraints.

The overall processing time is not significantly different between the two methods; the proposed method provides the essential advantage of frame-by-frame processing and enables real-time magnification of online streaming video at around 30 fps. Additionally, the proposed method allows instantaneous adjustment of magnification parameters, such as bandpass frequencies and magnification factors. Moreover, users can specify an FIR filter with lower order, which is less vulnerable to system restrictions. Because the proposed method processes a fixed number of image frames required for the FIR filter order for convolution, regardless of the total number of frames, it is efficient for image memory access in the program. On the other hand, the conventional method should specify the FIR filter order in terms of the total number of frames. Therefore, the proposed magnification approach efficiently controls resources and processes them faster with acceptable image quality and data extraction accuracy. Also, as shown in Fig. 8, initial delay occurs as often as the number of image frames equals the order of the applied FIR filter

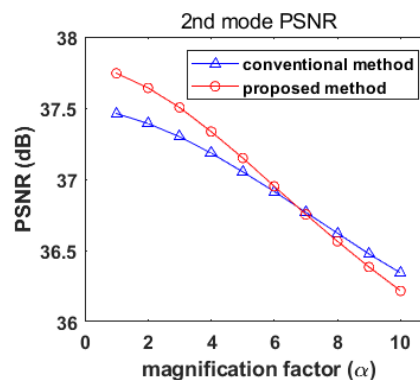


FIGURE 14. Peak signal-to-noise ratio (PSNR) scores according to  $\alpha$  for 2nd mode excitation.

in the proposed method. Therefore, magnification usually occurs after an incomplete magnification (delay) time from the initial stage of the image.

### B. IMAGE QUALITY ASSESSMENT

The mean squared error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM) were used to determine the degree of distortion relative to the original image [21]. MSE is a fundamental metric that calculates the average squared difference between all pixel intensity values. PSNR is the decibel-scale ratio of noise power to the maximum possible power of an image, calculated using MSE. When the MSE equals zero, PSNR approaches infinity and, as MSE increases, PSNR decreases in value. SSIM is a metric that quantifies the similarity of the structure of an object without quantitatively evaluating the error and evaluates the quality based on image similarity [25]. Both conventional and proposed PBMs exhibit a distorted object shape as the  $\alpha$  is increased excessively, which affects subsequent operations. Visual noise becomes discernible only when the cantilever image is magnified at around  $\alpha = 10$ . Therefore, the results were compared with the original video according to both PBM methods when  $\alpha = 10$ .

As can be seen, there is no noticeable difference in image quality between the two methods and the images obtained using the proposed method are as acceptable as those produced using the conventional magnification method. Nevertheless, as shown in Fig. 13, the proposed magnification method is slightly superior in the first-mode frequency while the conventional method is slightly superior in the second-mode MSE and PSNR values. As a result, it is difficult to assign consistent superiority in image quality metrics comparing both magnification techniques. Fig. 14 shows the PSNR score for each method at the second-mode frequency as a function of  $\alpha$ . As such, the proposed method tends to be better with lower  $\alpha$  values (i.e.,  $\alpha < 6$ ), but as  $\alpha$  increases, the quality score of the conventional magnification method takes precedence. Therefore, the appropriate  $\alpha$  value for better image quality varies from case to case. In particular, when both PSNRs are greater than 30 dB, the image quality

of the two methods is comparable at the human-perceived level [26].

### C. DATA ACCURACY AND APPLICATION

Both magnification approaches estimated the displacement amplitude closer to the reference LDV value than the extracted amplitude without magnification. The proposed method shows slightly underestimated displacements than conventional magnification in Table 1. From Fig. 9 and in the same viewpoint as Table 1, the conventional method tends to be magnified more for the same value of  $\alpha$ , whereas the proposed method tends to be less magnified. In particular, the displacement (green) in the case of no magnification for the low-frequency first mode (Fig. 10(a)) is very overestimated and more noticeably undersampled than the scaled displacement data extracted from the magnified image (blue and red) and the reference LDV displacement (black dotted). This overestimation in displacement is because of geometric discretization according to the pixel resolution when converting the displacement of the vibration image into the pixel intensity values. In the second-mode frequency (Fig. 10(b)), the amplitude difference of displacements extracted from magnified and unmagnified images is insignificant, except for the time sampling effect, as they are at the first-mode frequency. As illustrated in Fig. 9, the RMS response in the second mode is set to be less than that in the first mode in the experiment, but by chance the actual amplitude matches the image unit pixel level, limiting overestimation as in the first mode. Instead, the displacement may not be extracted due to unpacking by the pixel resolution of the image.

By integrating the proposed magnification technique with the current system specifications in this study, real-time processing at a 30-fps rate is conceivable. As a result, real-time monitoring of structures and machinery will be possible by acquiring monitoring video streams and magnifying subtle motions and vibrations in real-time. The primary issue with vibration measurement via images is the difficulty in accurately quantifying tiny vibrations at the subpixel level. As a result, magnification methods, particularly PBM, are beneficial. The vibration can be observed and measured using a conventional magnification algorithm. Nonetheless, it is challenging to apply motion magnification to vibration monitoring by post-processing image files. However, qualitative real-time and online vibration monitoring might be easily achieved by applying the proposed frame-by-frame magnification technique.

### V. CONCLUSION

The purpose of this study was to offer a novel approach for effectively implementing a PBM algorithm that leverages direct time-domain convolution at the temporal filtering stage. The developed real-time PBM algorithm produced a frame rate of up to 30 fps in dynamic displacement measurements based on online vibration monitoring using a nominal webcam. To verify the performance of the proposed PBM, the

magnification processing time, the magnified image quality, and the vibration displacement extraction for a simple cantilever structure vibrating less than 15 Hz were compared to the conventional PBM. Because of the magnification factors, the vibration displacements from the proposed method were estimated using linear regression and the accuracy of the estimated displacements was within the permissible error bound. Finally, the vibration characteristics were compared with the frequency spectrum of the extracted displacement.

The following advantages of the proposed video streaming-rate PBM can be summarized on the basis of the discussion:

- 1) While conventional batch processing works at a speed of 25.6 fps in a typical laptop PC environment, the proposed frame-by-frame processing technique can achieve a faster speed of 30.6 fps, which is 19.5% improvement.
- 2) The proposed PBM has the advantage of not being independent of the number of image frames for the temporal FIR filter order. A low-order filter provides rationale for fast processing without significantly affecting overall image quality and accuracy of the estimated vibration displacement.
- 3) Since the proposed PBM is frame-by-frame processing, instant adjustment of magnification parameters is always available, even while the magnification is being processed. This enables implementation of online vibration monitoring.

### REFERENCES

- [1] B. H. Kim, "Extracting modal parameters of a cable on shaky motion pictures," *Mech. Syst. Signal Process.*, vol. 49, nos. 1–2, pp. 3–12, Dec. 2014.
- [2] J. G. Chen, N. Wadhwa, Y.-J. Cha, F. Durand, W. T. Freeman, and O. Buyukozturk, "Modal identification of simple structures with high-speed video using motion magnification," *J. Sound Vib.*, vol. 345, pp. 58–71, Jun. 2015.
- [3] A. Havaran, M. Mahmoudi, and R. Ebrahimpour, "Extraction of the structural mode shapes utilizing image processing method and data fusion," *Mech. Syst. Signal Process.*, vol. 151, Apr. 2021, Art. no. 107380.
- [4] M. N. Helfrick, C. Niezrecki, P. Avitabile, and T. Schmidt, "3D digital image correlation methods for full-field vibration measurement," *Mech. Syst. Signal Process.*, vol. 25, no. 3, pp. 917–927, 2011.
- [5] S. Yoneyama and H. Ueda, "Bridge deflection measurement using digital image correlation with camera movement correction," *Mater. Trans.*, vol. 53, no. 2, pp. 285–290, 2012.
- [6] J. Baqersad, C. Niezrecki, and P. Avitabile, "Extracting full-field dynamic strain on a wind turbine rotor subjected to arbitrary excitations using 3D point tracking and a modal expansion technique," *J. Sound Vib.*, vol. 352, pp. 16–29, Sep. 2015.
- [7] L. Yu and B. Pan, "Single-camera high-speed stereo-digital image correlation for full-field vibration measurement," *Mech. Syst. Signal Process.*, vol. 94, pp. 374–383, Sep. 2017.
- [8] C. Liu, A. Torralba, W. T. Freeman, F. Durand, and E. H. Adelson, "Motion magnification," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 519–526, Jul. 2005.
- [9] H.-Y. Wu, M. Rubinstein, E. Shih, J. Gutttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 1–8, Aug. 2012.
- [10] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, "Phase-based video motion processing," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 1–10, Jul. 2013.
- [11] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, "Riesz pyramids for fast phase-based video magnification," in *Proc. IEEE Int. Conf. Comput. Photogr. (ICCP)*, May 2014, pp. 1–10.

- [12] T. H. Oh, R. Jaroensri, C. Kim, M. Elgharib, F. Durand, W. T. Freeman, and W. Matusik, "Learning-based video motion magnification," *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 663–679.
- [13] S. Takeda, M. Isogai, S. Shimizu, and H. Kimata, "Local Riesz pyramid for faster phase-based video magnification," *IEICE Trans. Inf. Syst.*, vol. E103.D, no. 10, pp. 2036–2046, 2020.
- [14] A. J. Molina-Viedma, L. Felipe-Sesé, E. López-Alba, and F. Díaz, "High frequency mode shapes characterisation using digital image correlation and phase-based motion magnification," *Mech. Syst. Signal Process.*, vol. 102, pp. 245–261, Mar. 2018.
- [15] M. Eitner, B. Miller, J. Sirohi, and C. Tinney, "Effect of broad-band phase-based motion magnification on modal parameter estimation," *Mech. Syst. Signal Process.*, vol. 146, Jan. 2021, Art. no. 106995.
- [16] H. Kim, Y. Chung, J. Jin, and J. Park, "Manifestation of flexural vibration modes of rails by the phase-based magnification method," *IEEE Access*, vol. 9, pp. 98121–98131, 2021.
- [17] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. Int. Conf. Image Process.*, 1995, pp. 444–447.
- [18] L. Litwin, "FIR and IIR digital filters," *IEEE Potentials*, vol. 19, no. 4, pp. 28–31, Oct./Nov. 2000.
- [19] R. J. Schilling and S. L. Harris, *Digital Signal Processing Using MATLAB*, 3rd ed. Boston, MA, USA: Cengage, 2016.
- [20] G. Fu, W. Bo, X. Pang, Z. Wang, L. Chen, Y. Song, Z. Zhang, J. Li, and R. Wu, "Mapping shape quantitative trait loci using a radius-centroid-contour model," *Heredity*, vol. 110, no. 6, pp. 511–519, Jun. 2013.
- [21] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.
- [22] G. Kerschen and J. C. Golinval, "Physical interpretation of the proper orthogonal modes using the singular value decomposition," *J. Sound Vib.*, vol. 249, no. 5, pp. 849–865, Jan. 2002, doi: [10.1006/jsvi.2001.3930](https://doi.org/10.1006/jsvi.2001.3930).
- [23] M. Pastor, M. Binda, and T. Harčarik, "Modal assurance criterion," *Proc. Eng.*, vol. 48, pp. 543–548, Jan. 2012, doi: [10.1016/j.proeng.2012.09.551](https://doi.org/10.1016/j.proeng.2012.09.551).
- [24] M. Gavish, S. Member, and D. L. Donoho, "The optimal hard threshold for singular values is  $4/\sqrt{3}$ ," *IEEE Trans. Image Process.*, vol. 60, no. 8, pp. 5040–5053, Aug. 2014, doi: [10.1109/TIT.2014.2323359](https://doi.org/10.1109/TIT.2014.2323359).
- [25] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [26] N. Ponomarenko, V. Lukin, K. Egiazarian, J. Astola, M. Carli, and F. Battisti, "Color image database for evaluation of image quality metrics," in *Proc. IEEE 10th Workshop Multimedia Signal Process.*, Oct. 2008, pp. 403–408, doi: [10.1109/MMSP.2008.4665112](https://doi.org/10.1109/MMSP.2008.4665112).



**JAЕ YOUNG AN** received the B.S. degree in mechanical and information engineering from the University of Seoul, in 2021, where he is currently pursuing the M.S. degree in mechanical and information engineering. His research interests include image processing, AI technique, and vibration analysis for mechanical systems.



**SOO IL LEE** received the B.S., M.S., and Ph.D. degrees from the Department of Mechanical Design and Production Engineering, Seoul National University, South Korea, in 1991, 1993, and 1997, respectively. He worked at Samsung Electronics Company Ltd., from 1997 to 2000. He was a Research Associate of mechanical engineering at Purdue University, West Lafayette, IN, USA, from 2001 to 2003. He is currently a Professor with the Department of Mechanical and Information Engineering, University of Seoul, South Korea. His research interests include nonlinear dynamic analysis of multiscale systems, and image and signal processing for vibration issues in mechanical systems.

...