Video Motion Magnification for Vibration Measurement in Hydropower Applications

Florian Fritzsche[1], Alexander Jung[2], Alexander Rubbert[3], Elisa Sanchez[4], and Axel Busboom[5]

[1,4,5] *Munich University of Applied Sciences, Institute for Sustainable Energy Systems, Munich, Germany*
{*fritzsche.florian; elisa.sanchez; axel.busboom*}*@hm.edu*

[2,3] *Voith Hydro, Heidenheim, Germany*
{*alexander.jung; alexander.rubbert*}*@voith.com*

ABSTRACT

Ensuring the mechanical integrity of hydropower plants requires robust structural health monitoring to detect issues like rotor imbalance and cavitation. Video motion magnification offers a promising non-contact alternative for vibration measurement. This paper presents an experimental comparison of three state-of-the-art algorithms (phase-based, learning-based, and Swin Transformer-based) for quantitative vibration measurement. Rather than evaluating only the final output, a novel framework analyses motion signals across multiple stages of the algorithms' processing pipelines to identify optimal extraction points. The frequency detection capabilities of these algorithms are then evaluated using both industrial and consumer-grade cameras. The focus is on comparing their ability to accurately measure vibrations with different input data quality. The results demonstrate the importance of the quality of the input data on the performance of the algorithm, as the compressed videos from the consumer-grade camera performed significantly worse than the uncompressed videos from the industrial camera. The learning-based method demonstrated the best overall performance, particularly with high-quality video data. This enabled the oscillation frequency to be measured at amplitudes over 70 times smaller than a pixel.

1. INTRODUCTION

Hydropower plays a central role in the global energy mix. As the world's largest source of renewable electricity, it provides approximately half of all renewable generation capacity, serving as a reliable and flexible foundation for integrating other intermittent energy sources, such as wind and solar power (International Energy Agency, 2024). Therefore, maintaining the mechanical integrity and operational stability of hydropower plants is of significant economic and environmental importance.

Structural health monitoring is essential for ensuring stability. Of the many physical quantities that can be observed, vibration measurements are particularly informative. They enable the early detection of issues such as rotor imbalance, bearing wear, shaft misalignment, cavitation in turbines and resonance phenomena in penstocks or generator housings (Mohanta et al., 2017). Identifying these issues at an early stage enables predictive maintenance, reduces costly downtime and improves the overall safety of plant operation.

Recent advances in video-based motion analysis offer an interesting alternative to the contact-based methods of measuring vibration. Techniques such as motion magnification enable small vibrations to be visualised across the entire camera view. If these vibrations can be accurately quantified, particularly using low-end, consumer-grade hardware, the cost of vibration measurement hardware could be reduced and diagnostics simplified. This method would also serve as a means of mobile ad hoc measurement, in contrast to the use of permanently installed systems, which are associated with the costs of cabling and integration.

This paper presents an experimental comparison of three state-of-the-art video motion magnification algorithms for quantitative vibration measurement in hydropower applications. While previous literature focuses on qualitative visualization, this work introduces a novel comparison framework that evaluates motion signals at multiple, distinct stages of the algorithms' processing pipelines. This multi-stage analysis allows for the identification of the optimal extraction point for accurate subpixel vibration measurement. Finally, the viability of these algorithms is tested using an industrial camera and consumer-grade hardware to ensure practical applicability.

The paper is structured as follows: Section 2 provides back-

ground information on vibrations in hydropower plants and an overview of video motion magnification techniques. Section 3 describes the algorithms used and their adaptation to vibration measurement. Section 4 outlines the experimental setup and evaluation metrics used to assess algorithm performance. Finally, Section 5 presents the results of the comparison and discusses their implications for vibration measurement in hydropower applications, with Section 6 concluding the study.

2. BACKGROUND AND RELATED WORK

2.1. Vibrations in Hydropower Plants

Hydropower plants convert the potential and kinetic energy of water into mechanical energy via turbines and subsequently into electricity using generators (Sun et al., 2021) (Yildiz & Vrugt, 2019). The selection of hydropower plant layout and turbine type depends on site-specific hydraulic and topographical conditions. Francis and Kaplan turbines dominate medium- and low-head applications due to their robust design and wide operating range (Giesecke & Heimerl, 2014).

The primary reason why vibration measurement and monitoring are so important for hydropower units is that they are rotating machines. Beyond these fundamental rotor-dynamic conditions, vibration behavior is subsequently influenced by hydraulic, mechanical, and structural interactions. These secondary sources include cavitation, pressure pulsations in turbines and draft tubes, and dynamic loads on shafts, bearings, and generator components (L. Zhang et al., 2019) (Shrestha et al., 2022) (X. Zhang et al., 2021). One of the most common types of oscillations in hydropower plants occurs at the blade passing frequency (BPF). The BPF is defined as the product of the number of runner blades and the rotational frequency of the turbine shaft.

The BPF is specific to each hydropower plant, as its value depends on the characteristics of the power plant and the turbine used. Values can range from 4 Hz with Kaplan turbines to significantly higher values of up to 200 Hz. For large-scale hydropower plants, the BPF typically does not exceed 70 Hz.

Vibration severity is commonly assessed using ISO 20816-5, which specifies root mean square (RMS) vibration velocity v_{RMS} limits for safe operation and intervention (International Organization for Standardization, 2018). The velocity value of 0.5 mm/s RMS serves as the strictest baseline limit, establishing the most conservative threshold for the 'safe for long-term operation' category. The highest 'intervention necessary' value is 3.0 mm/s RMS (International Organization for Standardization, 2018). Fault conditions such as imbalance or bearing damage lead to increased broadband vibration levels and pronounced harmonics (Nässelqvist et al., 2013), motivating reliable vibration monitoring solutions. Table 1 outlines the key vibration parameters relevant to hydropower applications, adapted from ISO 20816-5. Specifically, it illustrates the peak-to-peak amplitude d_{pp} and RMS acceleration a_{RMS} across representative frequencies for the established RMS vibration velocity v_{RMS} thresholds.

Table 1. Relevant vibrations characteristics for hydropower applications adapted from ISO 20816-5

v_{RMS} [$\text{mm}\,\text{s}^{-1}$]	f [Hz]	d_{pp} [µm]	a_{RMS} [$\text{m}\,\text{s}^{-2}$]
0.5	10	22.5	0.03
	50	4.5	0.16
	100	2.3	0.31
3.0	10	225.00	0.19
	50	135.05	0.94
	100	67.53	1.88

Conventional vibration monitoring in hydropower plants predominantly utilizes a combination of contact-based instrumentation, such as accelerometers, and non-contact sensors, like inductive displacement and velocity transducers (Romanssini et al., 2023). These sensors are hardwired to critical components of the machinery, such as turbine guide bearings and generator stators, and continuously collect time-waveform and spectral data for condition assessment. Vibration analysis typically relies on two principal measurement techniques: absolute vibration, quantified as root-mean-square (RMS) velocity, and relative vibration, expressed as peak-to-peak displacement. While these traditional approaches are highly reliable and have been established as the industry standard, they require complex cabling infrastructure and present challenges regarding physical installation and maintenance in harsh operating environments.

2.2. Video Motion Magnification

In recent years, a novel approach to visualising structural vibrations has emerged. The visualisation of slightest structural movements in videos can be facilitated by implementing Eulerian Video Motion Magnification (VMM) methodologies. These techniques were first described by Wu et al. (2012).

The Eulerian approach to motion detection involves examining fixed locations in an image to see how visual signals, such as pixel intensity or colour, change over time. This method essentially observes temporal variations at each pixel in order to reveal subtle movements or vibrations without explicitly tracking individual objects. In contrast, the Lagrangian approach involves identifying features or particles and tracing their trajectories through successive frames in order to measure motion.

2.2.1. Eulerian Approaches

The initial iteration of Eulerian methods used manually designed filters to extract temporal variations. The concept introduced by Wu et al. (2012) enhances small variations in colour and motion through bandpass filtering of pixel intensity

signals. Building on this idea, Wadhwa et al. (2013) proposed a phase-based motion processing approach that leverages local phase variations in a complex steerable pyramid, offering improved robustness in motion representation. Wadhwa et al. (2014) refined their methodology using Riesz pyramids, enabling more efficient, real-time motion magnification.

2.2.2. Learning-Based Approaches

The use of hand-crafted filters limited the adaptability of early methods because incorrect filter settings reduced the effectiveness of the approaches. To address this issue, deep learning frameworks were introduced to automate filter learning. Oh et al. (2018) presented a learning-based VMM framework that demonstrated the potential of neural networks for end-to-end motion extraction. Since its introduction, subsequent works have further refined this architecture. For instance, Ha et al. (2024) optimized the learning-based approach for real-time processing to reduce computational overhead, while Byung-Ki et al. (2025) introduced an axial decomposition technique to further isolate and enhance specific directional motions.

More recently, Lado-Roigé & Pérez (2023) presented a Swin Transformer-based Video Motion Magnification (STB-VMM) method that claimed to have achieved state-of-the-art results in visual motion magnification.

3. ALGORITHMS AND MOTION SIGNALS

Originally designed for visualisation purposes, the algorithms were modified to output per-frame motion tensors, effectively utilizing the latent space as the core motion signal. To ensure a fair evaluation, this latent space signal is extracted at multiple stages of the processing pipeline. Evaluating these various candidate signals determines which specific representation within the latent space is best suited for quantitative vibration measurement.

3.1. Phase-based Method

In the phase-based method (Wadhwa et al., 2013), the motion signal was first extracted from the phase variations in the complex steerable pyramid representation of the video frames (M_{PB1}). The algorithm then applies a temporal filter after which the motion signal was stored a second time (M_{PB2}). Finally, the motion signal is amplified, generating the final magnified video from which the motion signal was extracted a third time (M_{PB3}). To facilitate the extraction of these motion tensors and subsequent processing, we adapted a PyTorch implementation[1] rather than using the original MATLAB codebase. This was done to simplify the subsequent processing of motion tensors and to improve computational performance by using hardware acceleration.

[1] https://github.com/itberrios/phase_based

3.2. Learning-based Method

The learning-based method (Oh et al., 2018) uses an encoder network to separate each image into two components: texture, which captures appearance; and shape, which represents geometry. There are two variants of magnification: a two-frame approach and a temporal filtering approach.

In the two-frame magnification approach, the algorithm processes pairs of consecutive frames. Both frames are encoded to obtain their shape representations, and the difference between these shapes indicates motion. This motion signal is then multiplied by an amplification factor and added to the shape of the later frame. A decoder then combines the amplified shape with the texture from the later frame to reconstruct a new frame. Repeating this process sequentially over all pairs of frames produces a motion-magnified video.

In the temporal filtering magnification approach, each video frame is first encoded to extract its shape representation. A temporal band-pass filter is then applied to the sequence of shape representations to isolate motions within a desired frequency range, such as subtle periodic movements. The filtered shape signal is then amplified and combined with the original texture of each frame. The decoder then reconstructs the final frames, resulting in a video that highlights specific temporal motions more effectively.

The motion signal was extracted at two stages. First, the shape difference between consecutive frames was saved before amplification (M_{LB1}). This was identical for both methods. Secondly, the amplified shape signal was extracted for the two-frame (M_{LB2}) and temporal filtering (M_{LB2t}) approaches.

3.3. Swin Transformer-based Method

The Swin Transformer-based method (Lado-Roigé & Pérez, 2023) processes two input frames through a multi-stage pipeline to produce a motion-magnified output.

First, both frames are passed through an initial convolutional layer that downsamples them by a factor of eight in both spatial dimensions and extracts shallow, low-level features. These features are then fed into a deep feature extraction stage comprising of Residual Swin Transformer Blocks, which generate high-level representations for each frame.

Motion magnification is achieved by first computing the difference between the deep features of the two frames and interpreting this as motion, which was extracted as M_{STB1}. This motion is then passed through convolutional blocks and scaled by a specified amplification factor, which was saved as M_{STB2}. Afterwards it is added back to the deep features of the second frame. The resulting motion-enhanced features are then refined further using a Mixed Magnified Transformer Block, which integrates the amplified motion and helps to suppress artefacts.

Finally, the refined features are upsampled to the original resolution using transposed convolution, after which a final convolution layer reconstructs the magnified output frame.

The downsampling used in this approach reduces the resolution of the motion tensors by a factor of 64. This is irrelevant for video output, as the tensors are combined with input frames to reconstruct the full resolution. However, it significantly reduces the spatial resolution of the tensors used for vibration frequency analysis.

4. EXPERIMENTAL SETUP

As the algorithms were originally developed for visualisation purposes, an experimental comparison was conducted to evaluate their performance when measuring vibrations relevant to hydropower applications. The experiment involved recording the vibrations of a calibration device under controlled conditions using both an industry camera and a common smartphone camera.

The industry camera was used to serve as the reference, when using uncompressed videos. The smartphone was used to evaluate the performance of the algorithms with videos from consumer-grade hardware.

The recorded videos were then processed using phase-based (Wadhwa et al., 2013), Learning-based (Oh et al., 2018) and Swin Transformer-based (Lado-Roigé & Pérez, 2023) methods, adapted to export motion data for vibration measurement extraction. While the recent advancements of these methods, which are mentioned in Section 2.2, offer promising improvements in efficiency and capability, evaluating them falls outside the scope of this comparative study, which focuses on the foundational architectures.

These measurements were then compared to the set vibration frequency of the calibration device.

4.1. Video Acquisition

The experimental setup consisted of a vibration calibration device (MMF VC21), which can generate precise, adjustable sinusoidal oscillations in terms of frequency and acceleration. To ensure consistent and high contrast, the calibration device was set up in front of a white background. These oscillations were recorded using two cameras: a Teledyne FLIR ORYX-10GS-32S4M-C industrial camera and an Apple iPhone 15, which represented consumer-grade hardware.

Different camera-object distances (COD) were selected to achieve a similar resolution at the target for both cameras, making the recordings comparable. The resulting resolutions are listed in Table 2. The industry camera recorded videos at 240 and 400 frames per second with a focal length of 25 mm and CODs of 150 cm, 250 cm, and 400 cm. The iPhone recorded at 240 fps with a 26 mm focal length and CODs of 25 cm, 50 cm and 75 cm. A battery-powered LED light was used for all recordings to maintain consistent lighting conditions and prevent mains-induced flicker.

Table 2. Camera-Object Distance and Resolution

Device	COD [cm]	Resolution [$\mu\mathrm{m}/\mathrm{px}$]
FLIR Oryx	150	201.61
	250	342.46
	400	568.18
iPhone 15	25	164.47
	50	362.31
	75	568.18

4.2. Dataset

The dataset used for the experiments consisted of videos recorded with both cameras at various vibration settings. According to the Nyquist theorem, the analysable frequency range is limited by the frame rates of the cameras. The industry camera videos covered frequencies of up to 160 Hz, whereas the maximum frame rate of the iPhone meant it could only record vibrations of up to 80 Hz. Table 3 summarises the calibration device settings and oscillation characteristics of the videos in the dataset. The frequency and acceleration steps are the result of limitations in the calibration device.

The videos from the industry camera were saved as uncompressed AVI files. In contrast, the iPhone videos were recorded using the non-native camera app "Final Cut Camera"[2] and saved using H.265 compression in the MOV format with a bitrate of approximately 50.5 Mbps. The non-native camera app was used because it outputs recordings with a higher bitrate than the native camera app.

All videos were processed by the algorithms with an amplification factor of 20, to facilitate comparability between different tests. Each test was conducted on one second of video footage. As the algorithms are computationally expensive, all videos were cropped to the same region of interest around the vibrating target. The cropping was performed without altering the video signal in the area, to avoid influencing the subsequent processing and the accuracy of the algorithms.

4.3. Metrics

The performance of the algorithms was assessed by comparing the extracted motion signals to the expected sinusoidal vibration set on the calibration device. This study focuses solely on the ability of the algorithms to detect the frequency of the vibrations. To achieve this, the measured signal from each frame was analysed using a fast Fourier transform (FFT). The frequency corresponding to the highest peak in the FFT

[2] https://apps.apple.com/us/app/final-cut-camera/id6469552837

Table 3. Oscillation characteristics and subpixel-ratio of the dataset videos

f	a_{RMS}	Frame Rate	y_{RMS}	d_{pp}	Oryx SPxR (COD in cm)			iPhone SPxR (COD in cm)		
[Hz]	[m s^{-2}]	[fps]	[mm s^{-1}]	[µm]	150	250	400	25	50	75
15.92	1	240	10.0	282.68	0.71	1.21	2.01	0.58	1.28	2.01
15.92	2	240	20.0	565.37	0.36	0.61	1.00	0.29	0.64	1.00
40.00	1	240	4.0	44.78	4.50	7.65	12.69	3.67	8.09	12.69
40.00	2	240	8.0	89.56	2.25	3.82	6.34	1.84	4.05	6.34
40.00	5	240	19.9	223.89	0.90	1.53	2.54	0.73	1.62	2.54
80.00	1	240	2.0	11.19	18.01	30.59	50.76	14.69	32.36	50.76
80.00	2	240	4.0	22.39	9.00	15.30	25.38	7.35	16.18	25.38
80.00	5	240	9.9	55.97	3.60	6.12	10.15	2.94	6.47	10.15
80.00	10	240	19.9	111.95	1.80	3.06	5.08	1.47	3.24	5.08
159.2	1	400	1.0	2.83	71.32	121.15	201.00	–	–	–
159.2	2	400	2.0	5.65	35.66	60.57	100.50	–	–	–
159.2	5	400	5.0	14.13	14.26	24.23	40.20	–	–	–
159.2	10	400	10.0	28.27	7.13	12.11	20.10	–	–	–

spectrum was identified as the vibration frequency f_{meas}. For the identification of f_{meas}, we only considered frequencies between 5 Hz and the Nyquist frequency limit of the respective video. Frequencies below 5 Hz were excluded as signal drift would affect some of the algorithms, making subsequent evaluation impossible as the low frequencies would dominate the FFT. To avoid relying on frequency filters, which could affect other parts of the signal, parts of the FFT signal outside the set interval were simply ignored.

4.3.1. Signal-to-Noise Ratio

The most important metric for enabling a fair comparison between the algorithms is the signal-to-noise ratio (SNR). This metric quantifies the dominance of the main frequency component in relation to the background noise power. It is defined as the ratio of the peak signal power to the average noise power

$$\text{SNR} = \frac{P_{\text{S}}}{P_{\text{N}}} = \frac{\max_{k\in K}(P(k))}{\frac{1}{|K|-1}\left(\sum_{k\in K} P(k) - \max_{k\in K}(P(k))\right)}, \quad (1)$$

where $P(k)$ is the power spectrum derived from the fast Fourier transform of the motion magnitude, and K is the set of evaluated frequency bins above the specified cut-off. Here, P_{S} represents the peak power at the dominant frequency component, P_{N} is the average power of all remaining frequency components in K, and $|K|$ is the total number of evaluated frequency bins. A higher SNR indicates a clearer and more distinct vibration signal, as it means that the main frequency component stands out significantly from the background noise. The result is then converted to decibels using

$$\text{SNR}_{\text{dB}} = 10 \cdot \log_{10}\left(\frac{P_{\text{S}}}{P_{\text{N}} + \epsilon}\right) \text{[dB]}, \quad (2)$$

where ϵ is a small constant added to the denominator to ensure numerical stability and prevent division by zero.

Although a high SNR indicates a strong dominant frequency, this does not necessarily mean that the frequency matches the expected vibration frequency. Therefore, when evaluating algorithm performance, the SNR should be considered alongside the accuracy of the detected frequency.

In order to mitigate the effects of low-frequency noise and motion signal offsets, the frequency cut-off of 5 Hz is applied to the power spectrum prior to calculating the SNR. These offsets may occur because the algorithms are set to use the first frame of the video as the reference for the motion analysis, and this frame is not necessarily taken in the neutral position of the oscillation. By restricting the evaluated frequency set K to bins above this cut-off index, any DC component that could skew the results is effectively removed.

In order to facilitate comparability between the different tests, the FFT is limited to a window of 200 frames. This ensures that the SNR values are comparable, regardless of the original video length or frame rate.

4.3.2. Subpixel-Ratio

An important property of video motion magnification is the ability to measure vibrations that are smaller than a pixel. To quantify this capability, the subpixel-ratio (SPxR) metric is introduced. This metric indicates how many times smaller than a pixel the measured vibration amplitude is. It is defined as

$$\mathrm{SPxR} = \frac{d_{\mathrm{px}}}{d_{\mathrm{pp}}}, \tag{3}$$

where d_{pp} is the peak-to-peak displacement amplitude of the set oscillation in µm and d_{px} is the resolution at the target quantified as distance across a pixel in µm. A higher achieved SPxR indicates that the algorithm can measure smaller vibrations relative to the pixel size, demonstrating its effectiveness in capturing subtle motion details. The resolution was determined by measuring the pixel width of the target in the video frames and dividing the actual width of the target by it. Table 3 lists the SPxR values for all videos of the dataset.

5. RESULTS

To evaluate the algorithms, the videos recorded with both cameras were processed. First, the extracted motion signals M at the different stages were compared for each algorithm, based on a qualitative visual assessment of the extracted motion signals. The best signal was then selected for each algorithm and the signals were compared between the algorithms to identify the best-performing signal for each method.

Afterwards, the selected best signals were compared between the algorithms to determine which method provides the most accurate vibration measurements for the industry camera and iPhone videos.

5.1. Algorithm-Internal Performance Analysis

The comparison of the extracted motion signals at different extraction stages reveals significant differences in performance among the algorithms.

5.1.1. Phase-based Method

For the phase-based method, the signal extracted after the amplification stage (M_{PB3}) had the most pixels with a high SNR on the target. Unfortunately, it detected the wrong frequency, which was either half or double the ground truth frequency. The signal extracted after temporal filtering (M_{PB2}) showed almost no visible signal and had the same issue as M_{PB3}. The first motion signal, M_{PB1}, was the only one that contained the correct frequency and was therefore selected as the best-performing signal for the phase-based method, despite the fact that the SNR values on the target were lower and occurred less frequently, as with M_{PB3}.

5.1.2. Learning-based Method

This study does not consider the temporal filtering approach, as this method was not originally developed for temporal filtering. This was only added by Oh et al. (2018) because it produced comparable results to the non-filtered approach, for which the deep convolutional neural networks that sets the filters was trained. Therefore, only two signals (M_{LB1} and M_{LB2}) were compared for the learning-based method. Fig. 1 shows an example of a comparison of motion signals from a video taken with the industry camera at $f = 40$ Hz and $a_{\mathrm{RMS}} = 1\ \mathrm{m\,s^{-2}}$.

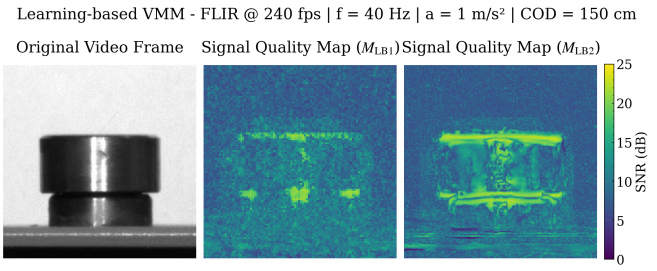


Figure 1. Comparison of extracted motion signals for the Learning-based method.

Both signals contained the correct frequency. However, the second signal (M_{LB2}) had a higher SNR over a larger portion of the target and was more distinct. It was therefore selected as the best-performing signal for the learning-based method.

5.1.3. Swin Transformer-based Method

Similar to the learning-based method, the two signals M_{STB1} and M_{STB2} were compared for the Swin Transformer-based method. Fig. 2 shows an example comparison for the motion signals of a video from the industry camera at $f = 40$ Hz and $a_{\mathrm{RMS}} = 1\ \mathrm{m\,s^{-2}}$.

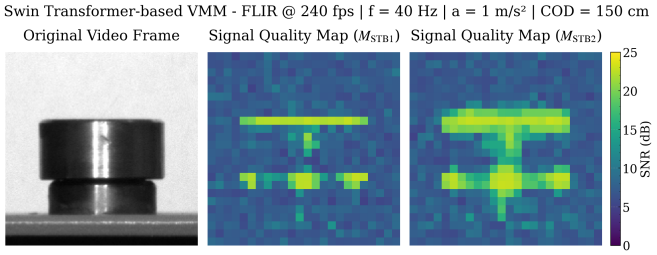
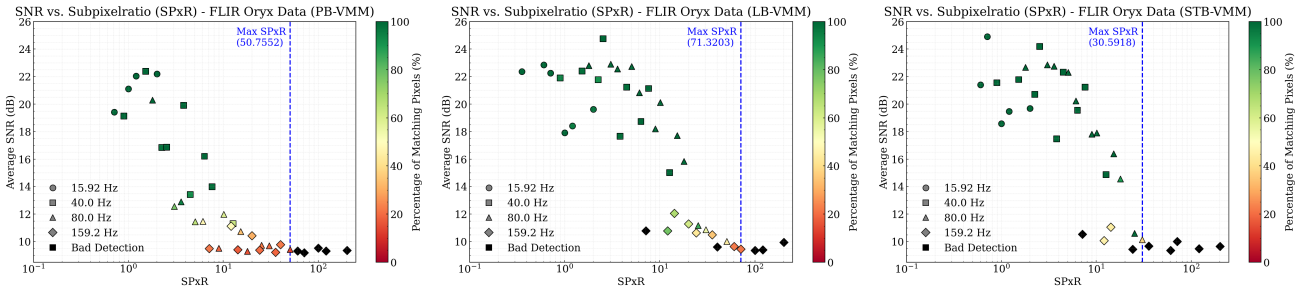


Figure 2. Comparison of extracted motion signals for the Swin Transformer-based method.

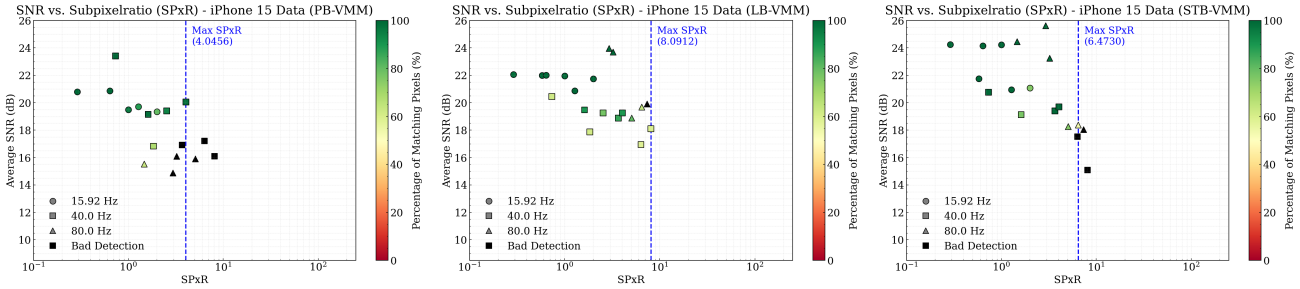
As with the learning-based method, both signals contained the correct frequency, with the second signal (M_{STB2}) having a higher SNR over a larger portion of the target. Consequently, the post-magnification signal was chosen as the optimal signal for the Swin Transformer-based approach.

5.2. Cross-Algorithm Performance Comparison

While the previous section identified the best-performing signals for each algorithm, this section compares these signals to determine the most accurate vibration measurement algorithm for the industry camera and iPhone. As saving all motion signals at all points adds non-equally distributed computational overhead to each algorithm, runtime and memory usage are not considered in this comparison.



(a) PB-VMM Algorithm – FLIR Oryx (b) LB-VMM Algorithm – FLIR Oryx (c) STB-VMM Algorithm – FLIR Oryx



(d) PB-VMM Algorithm – iPhone 15 (e) LB-VMM Algorithm – iPhone 15 (f) STB-VMM Algorithm – iPhone 15

Figure 3. Comparison of the algorithms outputs for the industry camera and iPhone videos.

To evaluate the performance of the algorithms, the entire dataset was processed with each algorithm, extracting the best-performing signal for the respective method. The time-motion series for each pixel was then processed with a FFT to identify the dominant frequency and calculate the SNR. To ensure a fair and rigorous comparison of the algorithms, the evaluation was restricted to the 95th percentile of pixels exhibiting the highest SNR values for each video. This threshold was chosen because the target does not fill the entire field of view. Furthermore, the algorithms predominantly detect motion along the top and bottom faces of the target because these are the regions with the highest contrast, generating the most significant intensity gradients during motion. By isolating the top 5%, the evaluation effectively filters out irrelevant background data and uniform target regions, focusing the comparative analysis exclusively on pixels containing a meaningful signal. The dynamic sample size was chosen to maintain consistency, compensating for output resolutions that vary with the video recording distance. This adjustment is particularly necessary as the Swin Transformer-based method works with significant downsampling, which reduces the output resolution by a factor of 64.

If the dominant frequency of these pixels matched the set vibration frequency within a tolerance of $\pm 10\%$, the SNR values are averaged for all matching pixels to obtain a single average SNR value for the output in question. The SPxR is also calculated for each output, based on the target resolution and the set vibration amplitude, d_{pp}.

The results for these analyses are visualised in Fig. 3. The x-axis represents the SPxR, which indicates how many times smaller than a pixel the measured vibration amplitude is. The y-axis represents the SNR in decibels, which quantifies the clarity of the detected vibration signal. Each point in the plots corresponds to a video from the dataset, with its position determined by the SPxR and SNR values calculated for the corresponding output. The points are colour-coded based on the percentage of pixels that match the actual vibration frequency f for that video. Outputs with no correct frequency detections are not included in the graphics. The shape of each data point represents the actual vibration frequency f, enabling an analysis of the impact of frequency on algorithm performance.

The detection threshold is the SPxR value up to which an algorithm was able to detect the correct frequency of a vibration. It was determined by letting the top five per cent of pixels vote for their dominant frequency. Next, it was checked whether the frequency with the most votes also matches the target frequency within a $\pm 10\%$ interval. If the detected frequency was incorrect, the corresponding data point in the diagram is coloured black. The highest SPxR value with correct detection was selected as the 'detection limit'. We chose this limit, which at first glance appears rather lenient, so as not to lower it unnecessarily, given that detecting the 159.2 Hz oscillation proved very difficult. The threshold is indicated by the blue, dashed vertical line in the plots of Fig. 3. The results for the best-performing signals of each algorithm are summarised in Table 4, which lists the detection thresholds and video characteristics for these signals.

Table 4. Detection thresholds and video characteristics for the best-performing signals of each algorithm.

	Device	f [Hz]	d_{pp} [µm]	COD [cm]	SPxR	SNR [dB]
M_{PB1}	Oryx	80.00	11.19	400	50.76	9.42
$M_{\textbf{LB2}}$		**159.2**	**2.83**	**150**	**71.32**	**9.44**
M_{STB2}		80.00	11.19	250	30.59	10.14
M_{PB1}	iPhone	40.00	89.56	50	4.05	20.05
$M_{\textbf{LB2}}$		**40.00**	**44.78**	**50**	**8.09**	**18.10**
M_{STB2}		80.00	55.97	50	6.47	18.37

5.2.1. Industry Camera Videos

For the industry camera videos revealed that the learning-based method achieved the highest maximum SPxR (71.32), followed by the phase-based method (50.76). The STB-VMM method achieved a maximum SPxR of 30.59.

In general, the learning-based method achieved the best signal quality and matching rate, which remained relatively constant up to an SPxR of 20. After this point, the SNR and matching rate began to decrease significantly. The Swin Transformer-based method generally produced only slightly worse results. However, it struggled to achieve good detection sooner than the learning-based method. The phase-based method exhibited more consistent behaviour, with a more gradual decrease in SNR and matching rate as the SPxR increased. Its maximum SPxR was only slightly lower than that of the learning-based method. Nevertheless, the SNR values were generally lower than those of the other algorithms and the matching rate was poorer at higher SPxR values. It should be noted that the phase-based method performed better than the other methods at detecting the 159.2 Hz oscillations at lower SPxR values.

5.2.2. iPhone Videos

When evaluating the iPhone videos, all algorithms showed significantly reduced performance compared to the industry camera videos. Whilst the detection rate and SNR at low SPxR values were equivalent and even outperforming the industry camera videos with the learning-based approach, the performance of all algorithms was significantly degraded at higher SPxR values. The learning-based method slightly outperformed the Swin Transformer-based method, achieving a maximum SPxR of 8.09 (6.47). Meanwhile, the phase-based method fell behind, achieving a maximum SPxR of 4.05. The significantly lower SPxR values for the iPhone videos compared to the industry camera videos suggest that the video quality substantially impacts the ability of the algorithms to accurately measure subpixel vibrations.

5.2.3. Compressed Industry Camera Videos

As iPhone videos are compressed, it is hypothesised that compression is responsible for the reduced performance of all algorithms. To test this theory, industry camera videos were compressed using H.265 with a bitrate of around 50.5 Mbps, which was increased proportionally to 84 Mbps for videos at 400 fps, in order to match the bitrate of the iPhone videos.

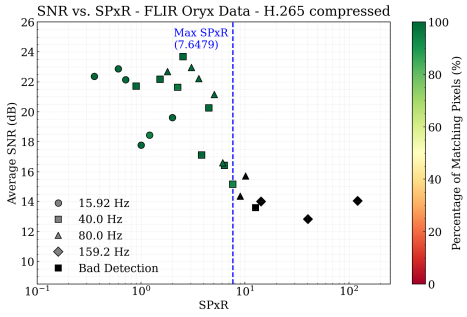


Figure 4. Performance of the learning-based method for compressed industry camera videos.

The test was limited to the learning-based method as this was the most effective for both the industry camera and the iPhone videos. As can be seen in Fig. 4, the results show a significant reduction in performance, with a maximum SPxR of 7.65, which is much closer to the results from the iPhone videos. This supports the hypothesis that compression is responsible for the reduced performance of all algorithms on iPhone videos, as the compression used appears to significantly reduce the accuracy of frequency measurements.

5.2.4. Detection of Frequencies close to the Nyquist Limit

As the Nyquist frequency is the upper limit for accurate frequency detection, it is expected that the algorithms will struggle to detect frequencies close to this limit. The results for the industry camera videos show that almost all of the incorrect detections (black points in Fig. 3) occured at the set frequency of 159.2 Hz for all algorithms. The observed behaviour as the set frequency approaches the 200 Hz Nyquist limit is consistent with the hypothesis that detection algorithms perform poorly near this threshold. In contrast, this trend is less pronounced in iPhone video data because the maximum set frequency (80 Hz) is sufficiently far from the 120 Hz Nyquist limit of these recordings.

5.3. Applicability and Practical Recommendations

When the results from the videos taken with the industry camera are applied to the relevant hydropower application vibrations mentioned in Table 1, the learning-based method should be able to measure even the smallest vibrations with the resolutions used in the experiment. The higher, critical vibrations are well within the measurement limits of both algorithms. This makes it possible to analyse higher harmonics when recording at the required frame rates.

Based on the comparative evaluation, several practical recommendations can be made regarding data acquisition, process-

ing, and model selection for deploying video motion magnification in hydropower plants:

- **Data Acquisition:** Prioritize industrial cameras capturing uncompressed video, as consumer compression (e.g., H.265) severely degrades frequency accuracy and limits the maximum SPxR. Additionally, ensure the frame rate keeps the target blade passing frequency safely below the Nyquist limit to prevent accuracy drops.
- **Data Processing:** To mitigate high computational costs, crop input videos to a tight region of interest around the target prior to processing. Furthermore, pipeline extraction points are critical. For example, the learning-based method performs optimally using the amplified shape signal rather than the pre-amplification difference.
- **Model Selection:** The learning-based method is recommended for high-precision structural health monitoring. It offers the best overall performance, achieves the highest SPxR, and avoids the severe 64x spatial downsampling resolution loss seen in the Swin Transformer method. While the phase-based method maintains a better signal-to-noise ratio at high SPxRs, it produces harmonic artifacts that hinder automated frequency detection without further filtering.

6. CONCLUSION

This study provides a novel experimental comparison of three video motion magnification algorithms for quantitative vibration measurement in hydropower environments. By moving beyond qualitative visualization, this work introduced a comprehensive evaluation framework that analyzed motion signals across multiple stages of the processing pipelines to identify the most accurate extraction points. Ultimately, the results prove the high precision of these techniques, with the learning-based algorithm notably demonstrating the ability to accurately capture oscillation amplitudes up to 71 times smaller than a single pixel using uncompressed video data.

In our evaluation, the phase-based method struggled to measure the correct frequency after the temporal filtering and amplification stages, frequently detecting the second harmonic or a subharmonic instead. Recent work by Shen et al. (2025) addresses this limitation, specifically targeting the phase decomposition errors that might be responsible for the artifacts. They demonstrate that applying a Distribution-Aware Fractional Anisotropic Filter (DFAF) can effectively suppress harmonic disturbances. While implementing DFAF is beyond the scope of this comparative study, their findings suggest that integrating frame interval selection could resolve the frequency detection anomalies observed in our tests.

The Swin Transformer-based method performed worst for the industry camera videos, with a maximum SPxR of 30.59. It had a much sharper fall-off in SNR and detection at higher SPxR values compared to the other algorithms.

When using compressed videos from the iPhone, the accuracy of the amplitude measurements was limited to a maximum of approximately an eighth of a pixel. This emphasises the importance of video quality for successful vibration measurement using video motion magnification techniques, since video processing on the iPhone has been shown to significantly reduce the frequency measurement accuracy. Although the measurements taken with consumer-grade hardware are less accurate than those taken with an industry camera, an SPxR of eight should enable critical vibration amplitudes to be measured when recording at an adequate COD.

Overall, the learning-based method demonstrated the best performance, outperforming the other algorithms when processing industry camera videos and achieving comparable results when processing iPhone videos. While beyond the scope of this paper, the learning-based method also has a significantly higher output resolution than the Swin Transformer-based method. The latter relies on 8x downsampling in both spatial dimensions, resulting in a 64x reduction in output resolution. The learning-based method achieves this higher output resolution using an amount of computational resources similar to the Swin Transformer-based method.

While the current evaluation focused on VMM algorithms, the established framework is inherently extensible to a wider class of dense optical flow methodologies.

6.1. Limitations

The experimental comparison presented in this paper has several limitations that must be taken into account when interpreting the results. Firstly, the evaluation focused solely on the ability of the algorithms to measure vibration frequency without considering amplitude accuracy. It also neglects the impact of different video compression levels, lighting conditions and camera angles on the performance of the algorithms. Furthermore, the study was conducted using a specific calibration device and may not accurately reflect real-world hydropower plant conditions. Additionally, the computational efficiency and memory usage of the algorithms were not assessed, which are important factors for practical deployment.

6.2. Future Research

In addition to addressing the above limitations, future research could pursue several approaches to further advance the field of video-based vibration measurement in hydropower applications. Future studies could expand the evaluation to encompass a wider variety of vibration types and conditions, including non-sinusoidal vibrations, transient events and different environmental factors. This would provide a more comprehensive understanding of the robustness and adaptability of the algorithms in real-world scenarios. As all algorithms can run in

either static or dynamic mode, setting the reference frame for motion extraction to the first or previous frame, it would helpful to compare the two modes to highlight their respective advantages and limitations. Finally, future research could investigate the precision of the frequency measurements, which were not considered in this study.

ACKNOWLEDGMENT

This work was partially funded by the European Union's Horizon Europe research and innovation programme under grant agreement No 101147310.

REFERENCES

Byung-Ki, K., Hyun-Bin, O., Jun-Seong, K., Ha, H., & Oh, T.-H. (2025). Learning-based axial video motion magnification. In *Computer vision - eccv 2024* (pp. 179–195). Cham: Springer Nature Switzerland.

Giesecke, J., & Heimerl, S. (2014). *Wasserkraftanlagen - Planung, Bau und Betrieb* (6th ed.). Springer Vieweg Berlin. doi: https://doi.org/10.1007/978-3-642-53871-1

Ha, H., Hyun-Bin, O., Jun-Seong, K., Byung-Ki, K., Sung-Bin, K., Tran, L.-T., ... Oh, T.-H. (2024). *Revisiting learning-based video motion magnification for real-time processing.* Retrieved from https://arxiv.org/abs/2403.01898

International Energy Agency. (2024). *Renewables 2024* (Tech. Rep.). Paris: International Energy Agency. https://www.iea.org/reports/renewables-2024.

International Organization for Standardization. (2018). *Iso 20816-5: Mechanical vibration — measurement and evaluation of machine vibration* (Tech. Rep.). Geneva: International Organization for Standardization.

Lado-Roigé, R., & Pérez, M. (2023). STB-VMM: Swin transformer based video motion magnification. *Knowledge-Based Systems*, *269*. doi: https://doi.org/10.1016/j.knosys.2023.110493

Mohanta, R. K., Chelliah, T. R., Allamsetty, S., Akula, A., & Ghosh, R. (2017). Sources of vibration and their treatment in hydro power stations-a review. *Engineering Science and Technology, an International Journal*, *20*(2), 637–648. doi: https://doi.org/10.1016/j.jestch.2016.11.004

Nässelqvist, M., Gustavsson, R., & Aidanpää, J. O. (2013). A methodology for protective vibration monitoring of hydropower units based on the mechanical properties. *Journal of Dynamic Systems, Measurement, and Control*, *135*(4). doi: https://doi.org/10.1115/1.4023668

Oh, T.-H., Jaroensri, R., Kim, C., Elgharib, M., Durand, F., Freeman, W. T., & Matusik, W. (2018). Learning-based video motion magnification. In *Computer vision - eccv 2018* (pp. 663–679). Springer International Publishing.

Romanssini, M., de Aguirre, P. C. C., Compassi-Severo, L., & Girardi, A. G. (2023). A review on vibration monitoring techniques for predictive maintenance of rotating machinery. *Eng*, *4*(3), 1797–1817. doi: https://doi.org/10.3390/eng4030102

Shen, J., Yang, X., & Cheng, D. (2025). Distribution-aware fractional anisotropic filtering for vibration displacement field measurement. *IEEE Transactions on Instrumentation and Measurement*, *74*, 1–17. doi: https://doi.org/10.1109/TIM.2024.3502742

Shrestha, R., Pradhan, S. S., Gurung, P., Ghimire, A., & Chitrakar, S. (2022). A review on erosion and erosion induced vibrations in francis turbine. *IOP Conference Series: Earth and Environmental Science*, *1037*(1), 12–28. doi: https://doi.org/10.1088/1755-1315/1037/1/012028

Sun, Y., Yang, Z., & Zhou, Z. (2021). Hydroelectric power plants: Current design principles, impacts and development prospects. In *Proceedings of the 2021 5th international conference on e-business and internet* (pp. 46–55). Association for Computing Machinery. doi: https://doi.org/10.1145/3497701.3497710

Wadhwa, N., Rubinstein, M., Durand, F., & Freeman, W. T. (2013). Phase-based video motion processing. *ACM Transactions on Graphics*, *32*(4). doi: https://doi.org/10.1145/2461912.2461966

Wadhwa, N., Rubinstein, M., Durand, F., & Freeman, W. T. (2014). *Riesz pyramids for fast phase-based video magnification.* https://people.csail.mit.edu/nwadhwa/riesz-pyramid/.

Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., & Freeman, W. T. (2012). Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics*, *31*(4). doi: https://doi.org/10.1145/2185520.2185561

Yildiz, V., & Vrugt, J. A. (2019). A toolbox for the optimal design of run-of-river hydropower plants. *Environmental Modelling and Software*, *111*, 134–152. doi: https://doi.org/10.1016/j.envsoft.2018.08.018

Zhang, L., Wu, Q., Ma, Z., & Wang, X. (2019). Transient vibration analysis of unit-plant structure for hydropower station in sudden load increasing process. *Mechanical Systems and Signal Processing*, *120*, 486–504. doi: https://doi.org/10.1016/j.ymssp.2018.10.037

Zhang, X., Zeng, J., Wu, B., & Gu, J. (2021). Study on the dynamic response of the powerhouse under the vibration load of the hydropower station. In *2021 7th international conference on hydraulic and civil engineering and smart water conservancy and intelligent disaster reduction forum, ichce and swidr 2021* (pp. 1667–1671). doi: https://doi.org/10.1109/ICHCESWIDR54323.2021.9656465