POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Vibration measurement and modal analysis of structures based on UAV camera and video motion magnification

Laurea Magistrale in Mechanical Engineering - Ingegneria Meccanica

Author: Soheil Rasi Rezvani

Advisor: Prof. Emanuele Zappa

Academic year: 2025-2026

1. Introduction

Vibration-based Structural Monitoring relies on modal indicators such as natural frequencies, damping and mode-shape information, which are sensitive to stiffness, mass distribution and boundary conditions. Contact sensors (e.g., accelerometers) remain the reference solution for reliable vibration measurements, but dense instrumentation can be impractical in real structures due to installation time, accessibility and logistics.

Vision-based vibration measurement offers a non-contact alternative with spatially dense information. When integrated with Unmanned Aerial Vehicles (UAVs), video can be acquired from flexible and hard to access viewpoints. The main challenge is that When the structural response is subtle, even small UAV motion can dominate the signal and block modal interpretation.

This work investigates a practical workflow for UAV-based vibration for modal interpretation supported by reference accelerometer data. The final workflow combines: (i) frequency screening on accelerometer signals to identify global frequency bands; (ii) UAV video stabilization with feature tracking to suppress camera motion while preserving structural response; (iii) learning-based motion magnification in selected global frequency bands; (iv) optical flow static reporting for visualizing motions in magnified videos; and (v) verification of a selected magnified band via pattern tracking in the magnified video.

2. Literature review and methodology

Two main technical methodologies are required: (i) robust video stabilization to suppress camera motion and (ii) motion magnification focused on narrow frequency bands based on reference accelerometer evidence. The workflow is summarized below, while citing only the most relevant sources.

2.1. Stabilization

Mingkhwan and Khawsuk [3] present a practical UAV stabilization workflow based on inter-frame feature matching and geometric alignment, and they quantify stabilization quality using MSE and PSNR. In this thesis, the same concept is implemented using SURF features (Bay et al. [1]) to obtain correspondences between consecutive frames.

SURF is suitable for UAV footage because it provides repeatable features under moder-

ate changes with relatively low computational cost. Stabilization is performed by: (i) detecting SURF keypoints, (ii) matching features between frames, (iii) estimating a 2D transform, and (iv) warping the current frame onto a reference frame. As shown in Fig. 1, models range from translation and similarity up to affine and projective; more complex models can capture stronger apparent motion but may become more sensitive to mismatches and parallax.

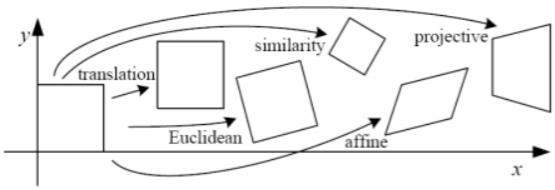

Figure 1: 2D geometric transforms used in video.

A key adaptation for vibration studies is to distinguish structural motion from camera motion. Therefore, feature extraction is constrained to the background by masking the structure. This forces the transform estimation to rely on static background, helping to preserve the structural response.

UAV videos often contain outliers. Dong et al. [2] demonstrate a practical stabilization and magnification workflow for non-contact monitoring by combining feature matching with RANSAC outlier rejection before applying motion magnification, supporting the use of robust fitting in this context. Following this, RANSAC is used here to reject mismatches when estimating the transform.

To compare candidate time windows and stabilization settings, the stabilized output is evaluated using MSE and PSNR as in [3]:

$$\mathrm{MSE} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I(i,j) - K(i,j)\right)^2, \quad (1)$$

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\right), \quad (2)$$

where I and K are two aligned frames and MAX is the maximum pixel value.

2.2. Learning-based motion magnification

After stabilization, subtle vibrations are enhanced using the learning-based motion magnification method of Oh et al. [4]. As illustrated in Fig. 2, the architecture follows an encoder–manipulator–decoder structure designed to separate appearance from motion: the encoder produces a texture representation and a shape representation, the manipulator scales the motion component α (motion magnification factor), and the decoder reconstructs a magnified frame.

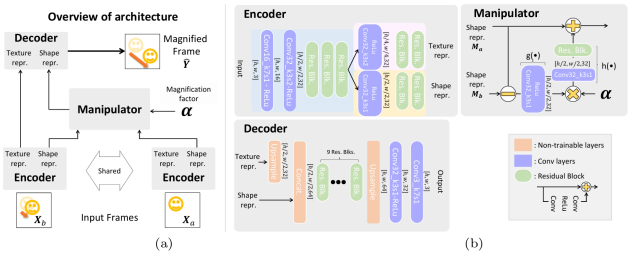

Figure 2: Learning-based video motion magnification architecture from Oh et al. [4].

For small motions, an input video can be expressed as $I(\mathbf{x},t) = f(\mathbf{x}+\delta(\mathbf{x},t))$. Ideal magnification aims to produce

$$I_e(\mathbf{x},t) = f(\mathbf{x}+(1+\alpha)\delta(\mathbf{x},t)), \quad (3)$$

where $\delta(\mathbf{x},t)$ is the displacement field and α is the magnification factor. In the learning-based method, δ is not estimated in pixels; instead, motion is manipulated in a learned latent space, which can be less sensitive to illumination flicker and high frequency texture noise than direct amplification.

In SHM, magnification is applied only within narrow frequency bands $[f_l, f_h]$ selected from accelerometer screening (Welch PSD, cross-spectrum and spectrogram). Operationally, the UAV frame rate defines the usable frequency range ($f \le f_s/2$).

The author of [4] provided access for the source and it is adopted directly in Python, while the thesis contribution focuses on frequency band selection using reference sensors, time window selection, SURF video stabilization, optical flow, and pattern matching for verification.

3. Experimental setup

The experiments were designed to validate a vision-based workflow for modal interpretation using UAV video and motion magnification, supported by reference accelerometer measurements. Two staircases at Politecnico di Milano (Lecco Campus) were selected because they are

accessible, repeatable to test, and exhibit slender components expected to respond to human excitation: (i) the external emergency staircase behind the dormitory, and (ii) the staircase connecting the main building to the laboratory level (−1).

Excitation was provided through controlled human jumping events at various locations along the staircases to generate transient responses. Reference vibration data were acquired using multiple accelerometers distributed landings for frequency analysis. A UAV also recorded 4K video with 60 fps from different views. The UAV footage was later cropped to short clips to be used as an input for the learning-based motion magnification.

Figure 3: Dormitory emergency staircase with accelerometer locations.

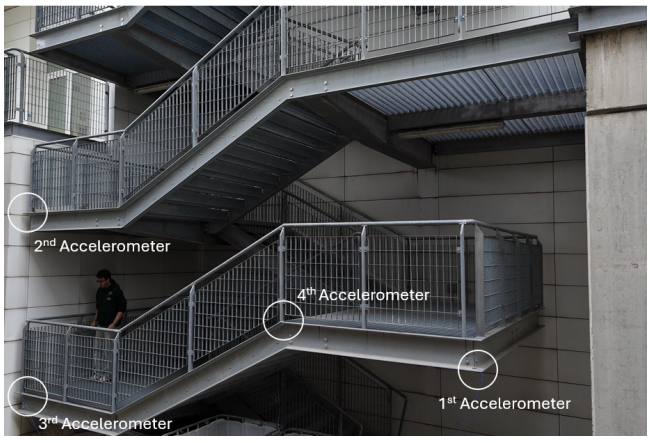

Figure 4: Laboratory-area staircase with accelerometer locations.

4. Selected Results

This section focuses on the left side view UAV recordings of the laboratory staircase for the 1st and 2nd excitation sets. The goal is to assess the integrity of the workflow when applied to different acquisitions, while keeping the same target band G3 = 12–13 Hz, and to verify the coherence between the magnified clips and the accelerometer reference data.

4.1. Reference accelerometer evidence and selected time window

The reference accelerometers are used to (a) identify repeatable candidate frequency bands and (b) select a short post-impact window for video processing, where vibration energy is high. Only the dataset of the 1st acquisition has been reported here.

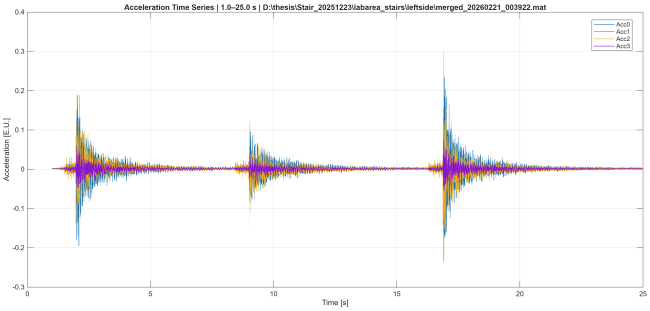

Figure 5: Left-side view (1st excitation): acceleration time series.

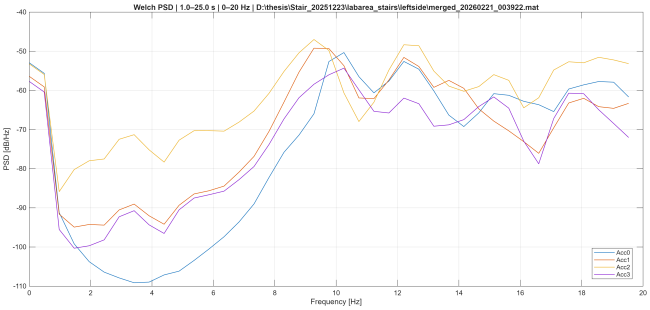

Figure 6: Left-side view (1st excitation): Welch PSD.

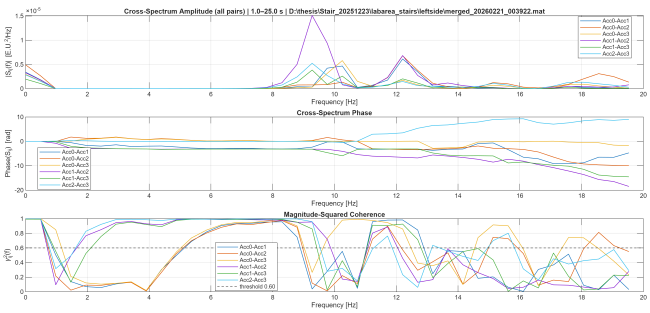

Figure 7: Left-side view (1st excitation): cross-spectrum.

If a peak appears consistently in the PSD across all accelerometer channels and is also confirmed in the cross-spectrum across multiple sensor pairs, it can be treated as a globally repeatable component and used to define a possible natural frequency band centered on that peak.

Table 1: Laboratory staircase: global frequency bands.

Band	Frequency band [Hz]
G1	9.0–9.5
G2	9.5–10.5
G3	12–13
G4	13–14
G5	17.0–19.0

spectrograms are used to find high-energy intervals. The short video is selected where energy around 12–13 Hz is most visible and stable. the interval is typically 2 s after impact.

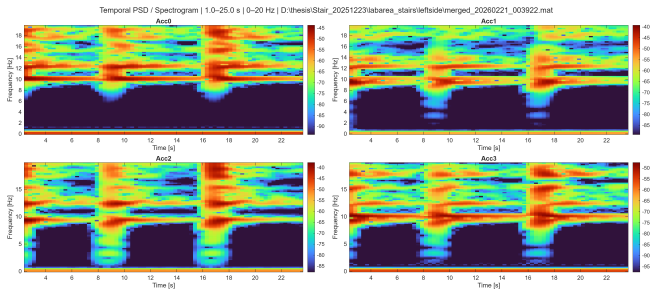

Figure 8: Left-side view (1st excitation): spectrogram (0–20 Hz).

4.2. Stabilization diagnostics

Because motion magnification amplifies residual camera motion together with structural motion, stabilization must be done. For each clip, SURF features are extracted only from the background by drawing a polygonal ROI on the structure.

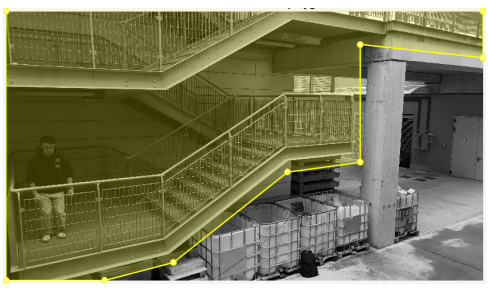

Figure 9: Left-side view: polygonal ROI.

then a robust transform is estimated with RANSAC.

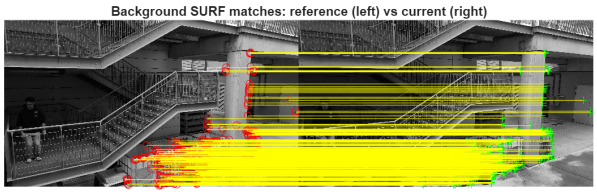

Figure 10: Left-side view: RANSAC Inliers.

In this case, the increase of PSNR and the reduction of MSE between consecutive frames is an indicator of stability improvement. The stabilized video is also visually inspected before moving to the motion magnification step.

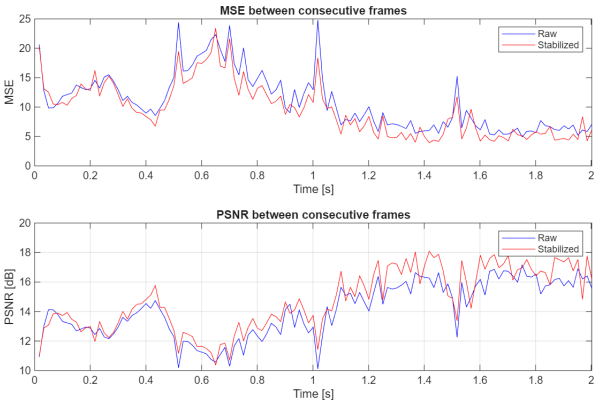

Figure 11: Left-side view: MSE and PSNR.

4.3. Motion magnification results

For the executive summary, the band G3 = 12–13 Hz is reported as the representative global mode. After stabilization and band-limited learning-based magnification ($\alpha = 30$), the optical flow magnitude maps highlight where motion repeatedly occurs during the selected window, while the vector fields provide directions.

The optical-flow summaries of the magnified clips in band G3 (12–13 Hz) exhibit a consistent spatial signature across the tested conditions. Although the excitation location, the selected post-impact time window, and the structural portion differ between the two runs, the dominant amplified motion repeatedly concentrates on the same structural components. the vectors show that railings, handlebars, and the landings experience motion in both acquisitions. Results of the proposed workflow show the repeatability of the motion localization within the same frequency band across different acquisitions.

Figure 12: Left-side view (1st excitation set), band G3 (12–13 Hz): optical-flow magnitude ($\alpha = 30$).

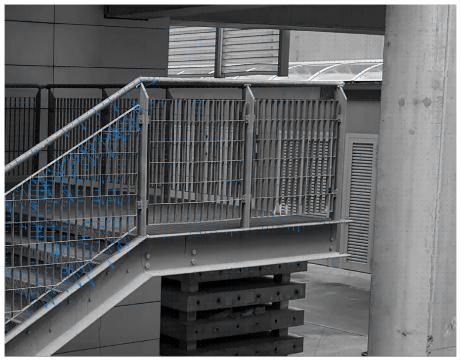

Figure 13: Left-side view (1st excitation set), band G3 (12–13 Hz): optical-flow vectors ($\alpha = 30$).

Figure 14: Left-side view (2nd excitation set), band G3 (12–13 Hz): optical-flow magnitude ($\alpha = 30$).

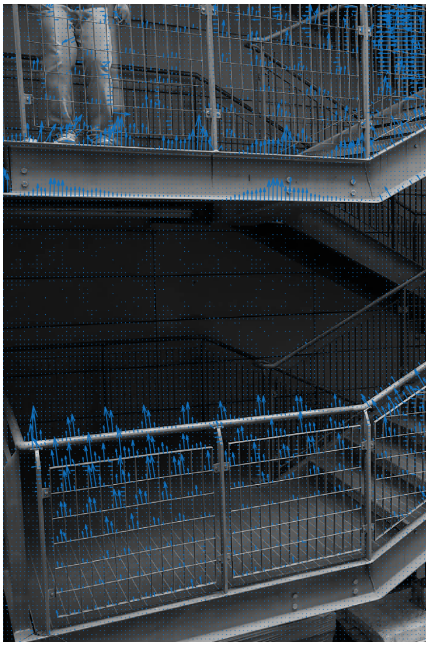

Figure 15: Left-side view (2nd excitation set), band G3 (12–13 Hz): optical-flow vectors ($\alpha = 30$).

4.4. Final verification: consistency between accelerometers and video-derived response

As a final verification step, another global natural frequency band (9.5–10.5 Hz), identified from the accelerometer screening is verified directly in the magnified video. The procedure is: (i) track a structural point in the magnified clip using pattern matching, (ii) reconstruct an acceleration signal by double differentiation, and (iii) compare its time–frequency content with the accelerometer spectrogram for the same acquisition and time window.

Figure 16: Template location for pattern matching on the landing near Acc1.

The displacement extracted from pattern matching is differentiated twice to obtain an acceleration estimate over the selected time window.

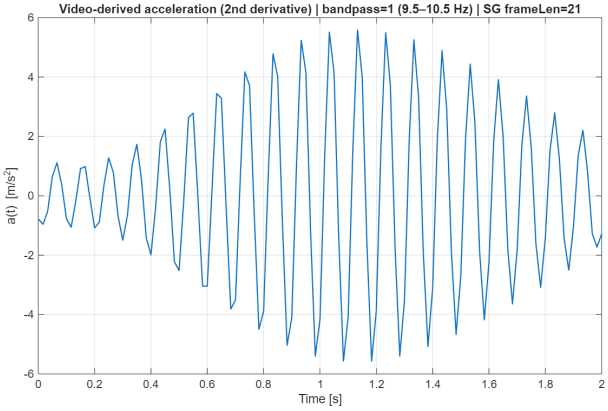

Figure 17: Acceleration reconstructed from pattern tracking.

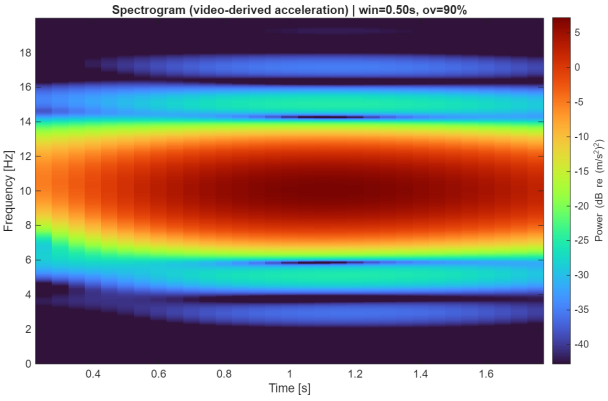

Figure 18: Spectrogram of acceleration derived from pattern matching.

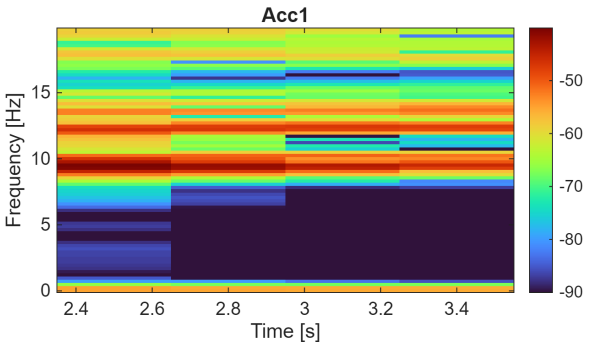

Figure 19: Accelerometer 1 spectrogram.

The verification is satisfied when the dominant part in the spectrogram of the video motion (Fig. 18) aligns with the dominant band observed in the accelerometer spectrogram (Fig. 19), within the expected bandwidth and over the same time interval.

5. Conclusions and future developments

A workflow for UAV-based modal interpretation and motion magnification was implemented and validated on two real staircases. When UAV recordings contain sufficiently stable time windows, high quality portions, and well-distributed static background features for robust stabilization, learning-based magnification produces motion patterns that are consistent with the global frequency bands derived from accelerometers. The study also identifies clear operational boundaries: close views with insufficient background and acquisitions with strong UAV motion degrade stabilization and make magnified output videos distorted.

Future work should prioritize more robust stabilization under limited background and parallax, extending verification toward more quantitative comparisons, and reducing the number of accelerometers by extracting frequency bands with pattern matching on the stabilized videos.

References

[1] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[2] Liang Dong, Lei Chen, Zhi-Cai Wu, Xing Zhang, Hou-Lin Liu, and Cui Dai. Video stabilization-based elimination of unintended jitter and vibration amplification in centrifugal pumps. *Mechanical Systems and Signal Processing*, 229:112500, 2025.

[3] Ekkaphon Mingkhwan and Weerawat Khawsuk. Digital image stabilization technique for fixed camera on small size drone. In *Proceedings of the 2017 Third Asian Conference on Defence Technology (ACDT)*, pages 12–19, Phuket, Thailand, January 2017.

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