Comparison of operational modal analysis methods from displacement estimation by video processing

Cédric Marinel, Olivier Losson, Benjamin Mathon, Jean Le Besnerais, Ludovic Macaire

Comparison of operational modal analysis methods from displacement estimation by video processing

C. Marinel [1,2], **O. Losson** [1], **B. Mathon** [1], **J. Le Besnerais** [2], **L. Macaire** [1]
[1] Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL,
Lille, France

[2] EOMYS Engineering,
Lille, France
e-mail: **cedric.marinel@eomys.com**

Abstract
Operational modal analysis is generally performed using contact sensors that may be time-consuming to setup and may introduce mass loading. Taking advantage of works about motion estimation by video analysis, several video modal analysis methods have emerged in the last decade. These new methods make it possible to perform modal analysis with a camera by estimating displacement from a video instead of contact motion measurements. Thereby, each pixel may be used as a contactless sensor. This study compares different strategies to perform video-based modal analysis. Two sub-pixel displacement estimation methods based on the phase of frame multi-scale decomposition are compared. In addition, two operational modal analysis methods using displacement estimations are studied. The methods are validated and compared on synthetic videos of a vibrating vertical cantilever beam. Different videos are generated to assess the robustness of these methods against motion amplitude, white noise, blurring, and gray level quantization.

1 Introduction

Monitoring properties of civil structures is important to detect failures at an early stage. The goal of operational modal analysis (OMA) is to identify the modal properties of a structure from local displacements, velocity, and/or acceleration measurements [1]. Traditionally, these measurements are obtained by contact sensors such as accelerometers or linear variable differential transformers. However, placing sensors on the structure can be tough and time-consuming. Furthermore, these sensors are generally expensive. During the last decade, video-based modal analysis methods have emerged thanks to high-speed camera improvements [2–5]. By considering each pixel as a sensor, one performs contactless modal analysis at low cost by estimating small displacements in video. Phase-based approaches estimate sub-pixel displacement with no need of any speckle pattern projected on the structure [5]. Two methods can be followed to estimate displacement from a multi-scale pyramid decomposition of each frame. Wadhwa et al. [6] analyze the multi-scale pyramid to estimate the displacement at each pixel, whereas Yang et al. [7] perform displacement estimation using a single scale. Moreover, OMA can be performed either by a combination of principal component analysis for data size reduction and complexity pursuit for blind source separation [8], or by covariance-driven stochastic subspace identification [9]. Few works compare the OMA performances reached by video-based strategies [5]. Because no study focuses on phase-based methods, we propose to compare their performances using synthetic videos that represent a vibrating vertical cantilever beam.

Section 2 describes how displacement is estimated by multi-scale and single-scale phase-based methods, and Sec. 3 how modal analyses are performed with different model orders to build stabilization diagrams. These graphs are then automatically processed to compare the results with the theoretical modal basis. In Sec. 4, we generate videos with different motion amplitudes to study sub-pixel efficiency. We also study the robustness of the methods against additive noise, blurring, and gray level quantization.

2 Displacement estimation by video analysis

2.1 Phase-based displacement estimation

Let $I(x, y; k)$ be the intensity at spatial coordinates (x, y) in frame $k \in [\![0, \mathcal{N}_k - 1]\!]$ and $\boldsymbol{\delta}$ be the displacement field along horizontal and vertical directions at k:

$$\boldsymbol{\delta}(x, y; k) = \begin{pmatrix} \delta^h(x, y; k) \\ \delta^v(x, y; k) \end{pmatrix} \in \mathbb{R}^2. \tag{1}$$

Assuming illumination is spatially and spectrally constant over time, the intensity associated to a given surface element can be considered as constant:

$$I(x, y; 0) \approx I(x + \delta^h(x, y; k), y + \delta^v(x, y; k); k). \tag{2}$$

To estimate the displacement field $\boldsymbol{\delta}$, each frame is decomposed into a complex steerable pyramid (CSP). To do so, spatial frequencies are transformed as $(\omega^h, \omega^v) = (\omega_r \cos(\theta), \omega_r \sin(\theta))$ into polar coordinates corresponding to different scales $r = 1, \ldots, \mathcal{N}_r$ and orientations $\theta = 0, \ldots, (\mathcal{N}_\theta - 1)\pi/\mathcal{N}_\theta$, where $\mathcal{N}_r$ and $\mathcal{N}_\theta$ are the number of scales and orientations. The CSP is then built by convolving each frame with a set of quadrature complex filters that split it into spatial frequency sub-bands. Each filter $G_{r,\theta}$ provides a complex response $S_{r,\theta} = G_{r,\theta} * I$ with magnitude $\rho_{r,\theta} = |\mathcal{S}_{r,\theta}|$ and phase $\varphi_{r,\theta} = \arctan\left(\mathfrak{Im}(\mathcal{S}_{r,\theta})/\mathfrak{Re}\left(\mathcal{S}_{r,\theta}\right)\right)$.

Using the constant illumination assumption, a filter response for frame 0 can be expressed from its response and the displacement for frame k:

$$S_{r,\theta}(x, y; 0) = G_{r,\theta} * I(x, y; 0) \overset{(2)}{\approx} S_{r,\theta}(x + \delta^h(x, y; k), y + \delta^v(x, y; k); k). \tag{3}$$

For each sub-band, the filter response phase thus verifies:

$$\varphi_{r,\theta}(x, y; 0) \approx \varphi_{r,\theta}(x + \delta^h(x, y; k), y + \delta^v(x, y; k); k). \tag{4}$$

Assuming that $\varphi_{r,\theta} \in \mathcal{C}^1$ for all r, θ, and k, a first-order Taylor expansion of Eq. (4) yields:

$$\varphi_{r,\theta}(x, y; 0) - \varphi_{r,\theta}(x, y; k) \approx \nabla\varphi_{r,\theta}(x, y; k) \cdot \boldsymbol{\delta}(x, y; k). \tag{5}$$

Because the phase gradient $\nabla\varphi_{r,\theta}$ is approximately equal to the filter central spatial frequencies [10], displacement can be estimated by replacing $\nabla\varphi_{r,\theta}$ by (ω^h, ω^v) in Eq. (5):

$$\varphi_{r,\theta}(x, y; 0) - \varphi_{r,\theta}(x, y; k) \approx (\omega^h, \omega^v) \cdot \boldsymbol{\delta}(x, y; k). \tag{6}$$

Let us use the Dirac comb to sample continuous space quantities I, $\rho_{r,\theta}$, and $\varphi_{r,\theta}$, and denote them in discrete space as $I[x, y; k]$, $\rho_{r,\theta}[x, y; k]$, $\varphi_{r,\theta}[x, y; k]$, and $\delta[x, y; k]$, with $[x, y; k] \in [\![1, \mathcal{N}_x]\!] \times [\![1, \mathcal{N}_y]\!] \times [\![0, \mathcal{N}_k - 1]\!]$, where $\mathcal{N}_x$, $\mathcal{N}_y$, and $\mathcal{N}_k$ are the number of pixel columns, pixel rows, and frames.

2.2 Multi-scale displacement estimation

To decompose each frame into a CSP, Wadhwa *et al.* [6] use Simoncelli and Freeman frequency filters, whose supports are shown in Fig. 1. Then they solve a weighted least square (WLS) problem to estimate displacement by fusing sub-band phases:

$$\begin{aligned} \hat{\boldsymbol{\delta}}[x, y; k] = \underset{\boldsymbol{\delta}[x,y;k]}{\arg\min} \sum_r \sum_\theta \sum_{c=-9}^{9} \sum_{\ell=-9}^{9} & \mathcal{G}[c, \ell] \cdot \rho_{r,\theta}^2[x + c, y + \ell; k] \\ & \cdot \Big[(\omega^h, \omega^v) \cdot \boldsymbol{\delta}[x, y; k] - \big(\varphi_{r,\theta}[x + c, y + \ell; 0] - \varphi_{r,\theta}[x + c, y + \ell, k]\big)\Big]^2. \end{aligned} \tag{7}$$

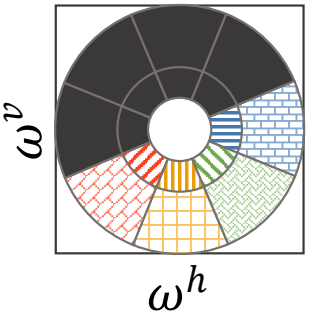

Figure 1: Ideal Simoncelli and Freeman [11] frequency filters $G_{r,\theta}$ supports of a pyramid with $\mathcal{N}_r = 2$ scales and $\mathcal{N}_\theta = 4$ orientations.

Weights are based on the squared filter response magnitude $\rho^2_{r,\theta}$ derived from sub-band decomposition. Indeed, the phase in a sub-band is meaningful only if the associated magnitude is high. The authors also assume that displacement is locally constant and add a spatial consistency constraint via a Gaussian kernel $\mathcal{G}$ (with 3 px standard deviation and 19×19 px support). Furthermore, phase $\varphi_{r,\theta}[x, y; k]$ is wrapped in $(-\pi, \pi]$. Before solving Eq. (7), phase is temporally unwrapped to compare phase shift between frame k and 0. In Eq. (7), $\varphi_{r,\theta}$ and $\rho_{r,\theta}$ are upsampled by bicubic interpolation for $r > 1$ to get the same spatial resolution as $\varphi_{1,\theta}$ and $\rho_{1,\theta}$.

2.3 Single-scale horizontal displacement estimation

Yang *et al.* [7] also use the frequency filters of Fig. 1 for CSP frame decomposition. Besides, they assume that the vertical displacement in their vertical cantilever beam videos can be neglected (i.e., $\delta^v(x, y; k) \approx 0$), which gives from Eq. (4):

$$\varphi_{r,\theta}(x, y; 0) \approx \varphi_{r,\theta}(x + \delta^h(x, y; k), y; k). \tag{8}$$

Using a Taylor expansion and the phase partial derivative approximation [10], Eq. (8) becomes:

$$\varphi_{r,\theta}(x, y; 0) = \varphi_{r,\theta}(x, y; k) + \omega^h \delta^h(x, y; k). \tag{9}$$

The authors only use the response of horizontal filters ($\theta = 0$) and estimate horizontal displacement at scale r by:

$$\hat{\delta}^h_r[x, y; k] = \frac{\varphi_{r,0}[x, y; 0] - \varphi_{r,0}[x, y; k]}{\omega_r}. \tag{10}$$

Phase is also temporally unwrapped before displacement estimation. $\hat{\delta}^h_1$ is computed at frame resolution, whereas for $r > 1$, $\hat{\delta}^h_r$ is first computed with subsampled phase $\varphi_{r,0}$, then upsampled by bicubic interpolation to get the full frame spatial resolution.

3 Modal analysis

As our experiments focus on a vertical cantilever beam, we neglect the vertical displacement of the multi-scale displacement estimator described in Sec. 2.2. The single-scale operator (see Sec. 2.3) provides a horizontal displacement estimation at a given scale. To perform modal analysis, the horizontal displacement estimated in either case is sampled at $\mathcal{N}_p$ pixels of interest (PoI), and reorganized as a matrix $\hat{\boldsymbol{\delta}}^h \in \mathbb{R}^{\mathcal{N}_p \times \mathcal{N}_k}$.

3.1 Complexity pursuit after principal component analysis

The objective of principal component analysis (PCA) and complexity pursuit (CP) [7] is to decompose the displacement matrix $\hat{\boldsymbol{\delta}}^h$ as:

$$\hat{\boldsymbol{\delta}}^h = \boldsymbol{\Phi}\boldsymbol{q}, \tag{11}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{\mathcal{N}_p \times N}$ and $\boldsymbol{q} \in \mathbb{R}^{N \times \mathcal{N}_k}$ are the mode shape and modal coordinate matrices. The method takes a parameter N called the model order and is performed in two successive steps: *i*) a model order reduction using PCA to reduce the number of displacement matrix rows from $\mathcal{N}_p$ to N, and *ii*) an estimation of modal coordinates using blind source separation by CP algorithm.

i) When $\mathcal{N}_p$ is high, video-based modal analysis may examine a high-dimensional displacement matrix. To reduce its size, it is factorized by singular value decomposition (SVD):

$$\hat{\boldsymbol{\delta}}^h = \begin{bmatrix} \boldsymbol{U} \, \overline{\boldsymbol{U}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \overline{\boldsymbol{\Sigma}} \end{bmatrix} \begin{bmatrix} \boldsymbol{V} \\ \overline{\boldsymbol{V}} \end{bmatrix}. \tag{12}$$

The displacement matrix is projected by PCA upon the first N left-singular vectors gathered in $\boldsymbol{U} \in \mathbb{R}^{\mathcal{N}_p \times N}$, whose conjugate transpose provides the reduced displacement matrix $\boldsymbol{\eta}$ that is defined as:

$$\boldsymbol{\eta} = \boldsymbol{U}^\star \hat{\boldsymbol{\delta}}^h \in \mathbb{R}^{N \times \mathcal{N}_k}. \tag{13}$$

ii) Assuming that $\boldsymbol{\eta}$ can be decoupled into modal coordinates $\boldsymbol{q}$:

$$\boldsymbol{q} = \boldsymbol{W}\boldsymbol{\eta}\,, \tag{14}$$

where $\boldsymbol{W} \in \mathbb{R}^{N \times N}$ is the demixing matrix, this blind source separation problem is solved using CP [8]. The method estimates each row $\boldsymbol{w}_i$ of $\boldsymbol{W}$ so that $\boldsymbol{q}_i = \boldsymbol{w}_i\boldsymbol{\eta} \in \mathbb{R}^{\mathcal{N}_k}$, $i \in [\![1,\ N]\!]$, has the highest temporal predictability defined as:

$$P(\boldsymbol{q}_i) = \log\left(\frac{\sum_{k=0}^{\mathcal{N}_k-1}(\bar{q}_i[k] - q_i[k])^2}{\sum_{k=0}^{\mathcal{N}_k-1}(\check{q}_i[k] - q_i[k])^2}\right), \tag{15}$$

where $\bar{q}_i[k]$ and $\check{q}_i[k]$ are long and short exponential moving average of $\boldsymbol{q}_i$, respectively.

For a given model order N, natural frequencies are estimated as the frequencies that maximize the discrete Fourier transform of modal coordinates:

$$f_i^N = \arg\max_f \left|\mathrm{DFT}\{\boldsymbol{q}_i\}[f]\right|. \tag{16}$$

Each damping ratio ζ_i^N is estimated from q_i using the logarithmic decrement method. Each mode shape $\boldsymbol{\phi}_i^N$ is a column of the mode shape matrix $\boldsymbol{\Phi}^N$ that is computed as:

$$\boldsymbol{\Phi}^N = \boldsymbol{U}\boldsymbol{W}^{-1}. \tag{17}$$

3.2 Covariance-driven stochastic subspace identification

The free motion equation for a system with N degrees of freedom (N corresponding to the model order) and viscous damping is:

$$\boldsymbol{M}\ddot{\boldsymbol{u}}(t) + \boldsymbol{D}\dot{\boldsymbol{u}}(t) + \boldsymbol{K}\boldsymbol{u}(t) = 0\,, \tag{18}$$

where $\boldsymbol{M}$, $\boldsymbol{D}$, and $\boldsymbol{K} \in \mathbb{R}^{N \times N}$ are the mass, damping, and stiffness matrices, and $\boldsymbol{u}(t) \in \mathbb{R}^N$ is the time-varying displacement vector. This problem can be recast into a discrete-time state space form:

$$\begin{aligned} \boldsymbol{x}[k+1] &= \boldsymbol{A}\boldsymbol{x}[k] + \boldsymbol{w}[k] \\ \boldsymbol{y}[k] &= \boldsymbol{C}\boldsymbol{x}[k] + \boldsymbol{v}[k]\,, \end{aligned} \tag{19}$$

where $\boldsymbol{x}[k] = \begin{pmatrix} \boldsymbol{u}(k\Delta t) \\ \dot{\boldsymbol{u}}(k\Delta t) \end{pmatrix} \in \mathbb{R}^{2N}$ is the state vector, Δt the time step between two successive frames, $\boldsymbol{y}[k] \in \mathbb{R}^{\mathcal{N}_p}$ the observation vector, $\boldsymbol{A} \in \mathbb{R}^{2N \times 2N}$ the state-space matrix, $\boldsymbol{C} \in \mathbb{R}^{\mathcal{N}_p \times 2N}$ the observation matrix, and $\boldsymbol{w}[k] \in \mathbb{R}^{2N}$ and $\boldsymbol{v}[k] \in \mathbb{R}^{\mathcal{N}_p}$ are the observation and input noise vectors.

The objectives of stochastic subspace identification (SSI) [9] is to get estimates $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{C}}$ of these matrices only from observations $\{\boldsymbol{y}[k]\}_{k=0}^{\mathcal{N}_k-1}$ (columns of $\hat{\boldsymbol{\delta}}^h$) to obtain the modes. The covariance-driven method takes a parameter $R < \mathcal{N}_k/2$ and considers a set of covariance matrices $\{\boldsymbol{\Lambda}_j\}_{j=1}^{2R-1}$ between time-shifted observations:

$$\boldsymbol{\Lambda}_j = \frac{1}{\mathcal{N}_k - 2R} \sum_{k=0}^{\mathcal{N}_k-2R-1} \boldsymbol{y}[k+j]\boldsymbol{y}[k]^\intercal \,. \tag{20}$$

These matrices are used to form a block Toeplitz matrix of R block rows as:

$$\boldsymbol{T}_{1:R} = \begin{pmatrix} \boldsymbol{\Lambda}_R & \boldsymbol{\Lambda}_{R-1} & \cdots & \boldsymbol{\Lambda}_1 \\ \boldsymbol{\Lambda}_{R+1} & \boldsymbol{\Lambda}_R & \cdots & \boldsymbol{\Lambda}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Lambda}_{2R-1} & \boldsymbol{\Lambda}_{2R-2} & \cdots & \boldsymbol{\Lambda}_R \end{pmatrix} \in \mathbb{R}^{R\mathcal{N}_p \times R\mathcal{N}_p} \,. \tag{21}$$

The SVD of $\boldsymbol{T}_{1:R}$ provides estimates $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{C}}$ [9]. Then, from the eigenvalue decomposition $\hat{\boldsymbol{A}} = \boldsymbol{\Psi}\boldsymbol{S}\boldsymbol{\Psi}^{-1}$ with $\boldsymbol{S} = \mathrm{diag}(\lambda_i), i \in [\![1, 2N]\!]$, natural frequencies, damping ratios, and mode shapes can be computed as:

$$f_i^N = \frac{1}{2\pi}\left|\frac{\log \lambda_i}{\Delta t}\right|, \tag{22}$$

$$\zeta_i^N = \frac{\mathfrak{Re}(\lambda_i)}{|\lambda_i|}, \tag{23}$$

$$\boldsymbol{\Phi}^N = \hat{\boldsymbol{C}}\boldsymbol{\Psi}. \tag{24}$$

The $2N$ modes come as N complex conjugate pairs and only N positive frequencies are kept.

3.3 Fast covariance-driven stochastic subspace identification

The size $R\mathcal{N}_p \times R\mathcal{N}_p$ of the Toeplitz matrix may be huge for video-based modal analysis. To reduce the number of observations, we use a similar projection as in Eq. (13). The displacement matrix is first decomposed by SVD, then projected on the first P left-singular vectors to obtain a reduced observation matrix $\check{\boldsymbol{y}}$:

$$\hat{\boldsymbol{\delta}}^h = \left[\boldsymbol{U}_P\, \overline{\boldsymbol{U}_P}\right] \begin{bmatrix} \boldsymbol{\Sigma}_P & \boldsymbol{0} \\ \boldsymbol{0} & \overline{\boldsymbol{\Sigma}_P} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_P \\ \overline{\boldsymbol{V}_P} \end{bmatrix}, \tag{25}$$

$$\check{\boldsymbol{y}} = \boldsymbol{U}_P^\star \hat{\boldsymbol{\delta}}^h \in \mathbb{R}^{P \times \mathcal{N}_k}. \tag{26}$$

SSI is performed on $\check{\boldsymbol{y}}$ with N and R as parameters. The Toeplitz matrix $\check{\boldsymbol{T}}_{1:R} \in \mathbb{R}^{RP \times RP}$ is constructed to obtain the natural frequencies $\{f_i^N\}_{i=1}^N$ and damping ratios $\{\zeta_i^N\}_{i=1}^N$, and a reduced mode shape matrix $\check{\boldsymbol{\Phi}}^N$. The mode shape matrix $\boldsymbol{\Phi}^N$ on the $\mathcal{N}_p$ pixels can then be computed as:

$$\boldsymbol{\Phi}^N = U_P \check{\boldsymbol{\Phi}}^N. \tag{27}$$

We denote this method as FSSI.

3.4 Stabilization diagram

In practice, the number of modes N is not known. Therefore, a stabilization diagram is used to plot the poles obtained from a modal analysis method for different model orders [12]. Irrelevant models produce spurious modes that can be discarded by a stability analysis. Indeed, physical poles tend to be stable while spurious ones tend to be unstable (in frequency, damping, and/or mode shape).

Let p_i^N, $i \in [\![1, N]\!]$, be a pole of order N with f_i^N, ζ_i^N, and $\boldsymbol{\phi}_i^N$ its natural frequency, damping ratio, and mode shape. We consider it stable if the following predicate holds:

$$S\left(p_i^N\right) = \exists p_j^{N-1} \left[\left(\frac{|f_i^N - f_j^{N-1}|}{f_j^{N-1}} < 0.01 \right) \wedge \left(\frac{|\zeta_i^N - \zeta_j^{N-1}|}{\zeta_j^{N-1}} < 0.05 \right) \wedge \left(\text{MAC}(\boldsymbol{\phi}_i^N, \boldsymbol{\phi}_j^{N-1}) > 0.98 \right) \right], \quad (28)$$

where MAC is the modal assurance criterion defined for any two mode shapes $\boldsymbol{\phi}$ and $\tilde{\boldsymbol{\phi}}$ by:

$$\text{MAC}(\boldsymbol{\phi}, \tilde{\boldsymbol{\phi}}) = \frac{(\boldsymbol{\phi}^\intercal \tilde{\boldsymbol{\phi}})^2}{(\boldsymbol{\phi}^\intercal \boldsymbol{\phi})(\tilde{\boldsymbol{\phi}}^\intercal \tilde{\boldsymbol{\phi}})} \, . \quad (29)$$

The key idea is that spurious modes occur randomly and are not stable for two consecutive model orders.

In addition to poles, complex mode indication functions (CMIFs) are also represented on the diagram [13]. These functions are the squared eigenvalues of the estimated frequency response function matrix, sorted in descending order for each frequency. The first CMIF is the most important, and each of its peaks indicates the presence of a mode at the associated frequency.

4 Experiments

4.1 Experimental setup

To compare the methods considered in Sec. 3, we generate synthetic videos of a vertical cantilever beam using the Euler-Bernoulli beam model. This model requires adjusting the following physical beam parameters: length L (m), Young modulus E (Pa), moment of inertia J (m^4), and mass per unit length μ ($\text{kg}\cdot\text{m}^{-1}$). The center line of the vertical beam is defined in the scene coordinate system by the point set:

$$\{(g(z,t), z; t) \in \mathbb{R} \times [0, L] \times [0, (\mathcal{N}_k - 1)\Delta t]\}, \quad (30)$$

where $g(z,t) = \sum_{m=1}^{\mathcal{N}_m} \phi_m(z)\, q_m(t)$ with, for all $m \in [\![1, \mathcal{N}_m]\!]$, ϕ_m and q_m solutions of:

$$\begin{cases} \dfrac{\partial^4 \phi_m}{\partial z^4}(z) - \dfrac{\mu(2\pi f_m)^2}{EJ}\phi_m(z) = 0 & (31) \\ \phi_m(0) = 0, \quad \dfrac{\partial \phi_m}{\partial z}(0) = 0 & (32) \\ \dfrac{\partial^2 \phi_m}{\partial z^2}(L) = 0, \quad \dfrac{\partial^3 \phi_m}{\partial z^3}(L) = 0 & (33) \\ (2\pi f_m)^2 q_m(t) + 4\pi f_m \zeta_m \dfrac{\partial q_m}{\partial t}(t) + \dfrac{\partial^2 q_m}{\partial t^2}(t) = \dfrac{1}{\mu}\displaystyle\int_0^L \phi_m(z)\gamma(z,t)dz & (34) \end{cases}$$

In this experiment, the input force γ (N) is represented by a time and space Dirac function to simulate a horizontal hammer impact at the free end of the beam ($z = L$) at $t = 0$. To simulate the behavior of our experimental beam, we set its volume to $900 \times 30 \times 6\,\text{mm}^3$ with $L = 900$ mm, its mass to $1.413\,\text{kg}$, and its Young modulus to $E = 210 \cdot 10^9\,\text{Pa}$. We set the number of modes to $\mathcal{N}_m = 4$. Theoretical natural frequencies $\{f_m\}_{m=1}^4$ are computed from Eqs. (31)–(33), and damping ratios are set from the results of an experimental modal analysis of our beam. Their values are listed in Table 1.

Table 1: Beam theoretical natural frequencies and damping ratios.

m	1	2	3	4
f_m (Hz)	6.19	38.79	108.60	212.82
ζ_m (%)	0.11	1.13	0.29	0.13

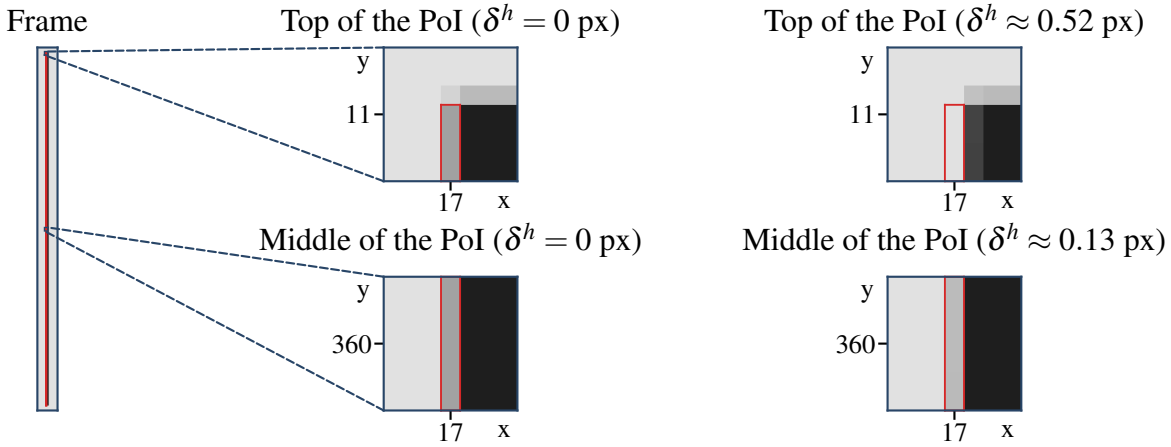

Figure 2: Edge pixels of interest for displacement estimation, with PoI in red.

Frame resolution is set to 720×40 px and since the beam covers 97% of the frame height, pixel resolution is $1.289\,\text{mm}\cdot\text{px}^{-1}$. Videos last 4 s at 436 fps (frame rate of TIS DMK 33UX287 camera). Each pixel intensity value is computed in proportion to the area of the intersection between the pixel in the image plane and the projected beam using a pinhole camera model without optical distortion. Values are then scaled between 30 and 225 to encode the gray level of each pixel on $\mathcal{N}_b = 8$ bits. These values ensure a high contrast between the beam and the background, and avoid saturation after Gaussian noise addition (see Sec. 4.3).

Our pixels of interest (PoI) are the 699 pixels on the beam left edge, as displayed in Fig. 2. As the beam is vertical, we consider that its vertical displacement is negligible and focus on the horizontal one. Table 2 shows the amplitude of the true (model-based) horizontal displacement δ^h computed at the top edge pixel according to the input force γ.

To estimate displacement $\boldsymbol{\delta}^h$ at the PoI, each video is analyzed by the multi-scale estimator using Eq. (7) to compute $\hat{\boldsymbol{\delta}}^h$, and by the single-scale estimator using Eq. (10) for $\hat{\boldsymbol{\delta}}^h_1$ and $\hat{\boldsymbol{\delta}}^h_2$. We adapt the Gaussian kernel $\mathcal{G}$ used in the multi-scale estimator (1 px standard deviation and 7×7 px support) to account for the beam thickness (5 px). Displacements are estimated with respect to the first frame, hence are not necessarily centered on the cantilever beam equilibrium. In practice, we remove the temporal mean of the displacements estimated at each pixel to analyze centered vibrations. Figure 3 shows all the 699 theoretical and estimated displacements with each estimator for two input forces of 0.08 and 1.31 N. We can observe that for $\gamma = 1.31$ N, estimations are close to the theoretical displacements. The single-scale estimator $\hat{\delta}^h_2$ overestimates displacement, while the single-scale estimator $\hat{\delta}^h_1$ and multi-scale one $\hat{\delta}^h$ overestimate it. For $\gamma = 0.08$ N, estimated displacements follow the main frequency of the theoretical one, but the low displacement amplitudes introduce a quantization effect because gray levels have too few different values.

The estimated displacements of the 699 pixels are extracted and used as input data of the three modal analysis methods described in Sec. 3: CP after PCA (PCA + CP), SSI, and FSSI. For PCA + CP, we set the model order N from 2 to 25. For SSI, the number of block rows in the Toeplitz matrix is set to $R = 20$ and the

Table 2: Amplitude of true horizontal displacement δ^h at the top of the beam vs. input force γ.

Force γ (N)	0.08	0.16	0.33	0.65	1.31	2.62	5.24	10.47
$\max(\delta^h) - \min(\delta^h)$ (px)	0.03	0.06	0.12	0.25	0.5	1	2	4

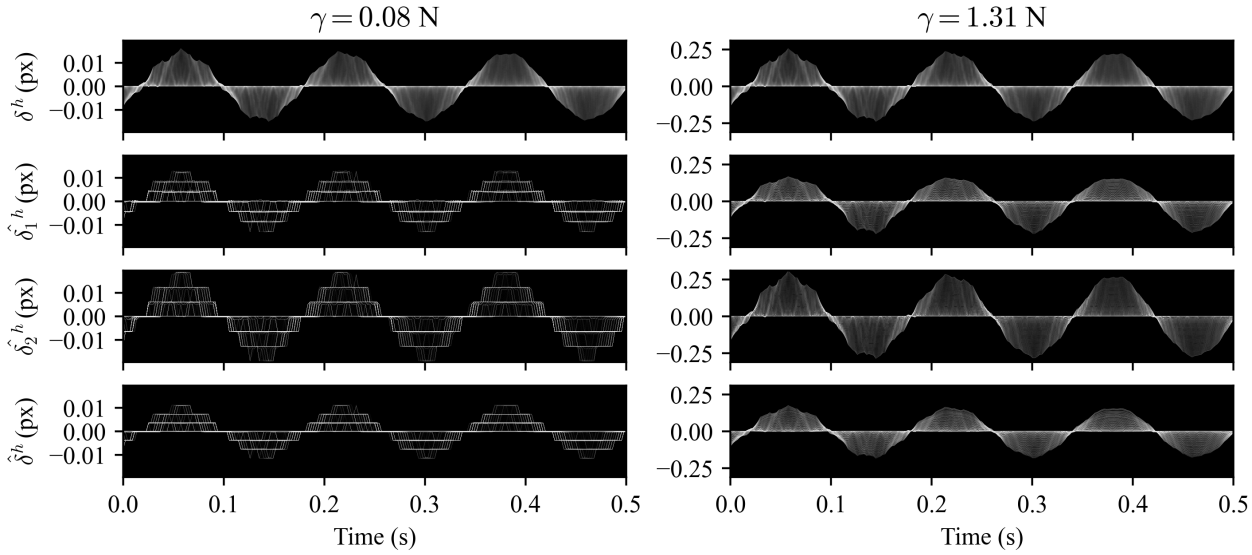

Figure 3: Theoretical and estimated displacements with input forces $\gamma = 0.08$ and 1.31 N.

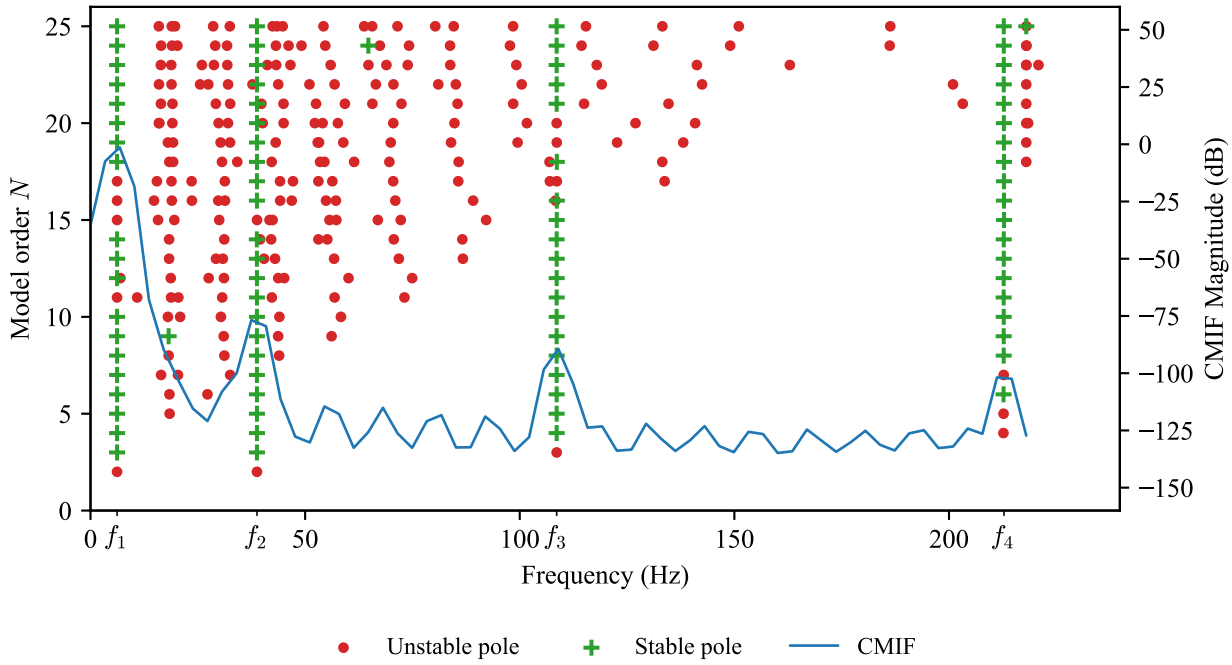

Figure 4: Stabilization diagram of SSI with $\hat{\delta}_2^h$ for the video with input force $\gamma = 1.31$ N.

model order N also varies from 2 to 25. For FSSI, we first keep the first $P = 50$ principal components from the displacement matrix $\hat{\boldsymbol{\delta}}^{\boldsymbol{h}}$ to obtain the reduced observation matrix, then set the same SSI parameters.

According to the displacement estimator and the modal analysis method, we then obtain nine stabilization diagrams per video, of which Fig. 4 provides an example. To compare the approaches, we assume that spurious poles are successfully removed with the stabilization diagram and the CMIF. For a given theoretical frequency, we count how many among the 24 poles of different orders are stable according to Eq. (28) provided that their estimated natural frequencies differ by no more than 1% relative to the theoretical one. We consider that the number of stable pole is a quality indicator of the method efficiency, and that a mode is successfully retrieved if at least five stable poles are retrieved at the mode frequency.

Four sensitivity studies are then performed to test the robustness of the methods against the following variations: *i*) input force γ, *ii*) Gaussian white noise standard deviation in the video, *iii*) Gaussian blur standard

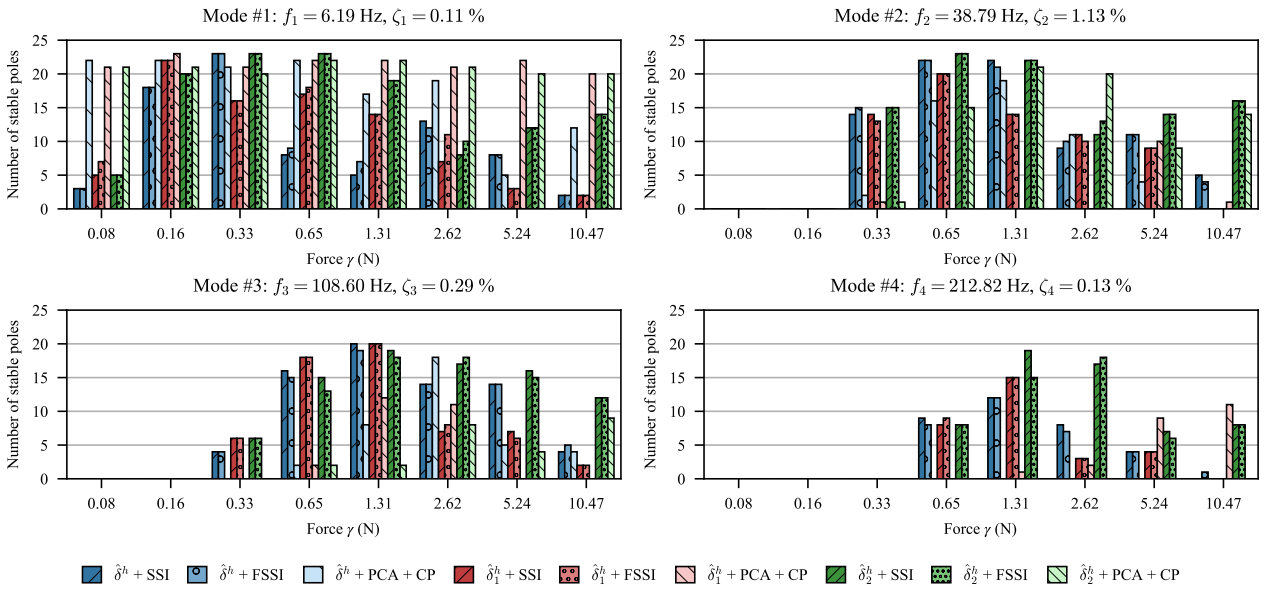

Figure 5: Number of stable poles around theoretical natural frequencies against input force γ.

deviation in the video, and *iv*) quantization bits of the video.

4.2 Robustness against input force

We perform a sensitivity study of the methods against the input force γ to inspect their robustness to displacement amplitude. The input force γ varies from 0.08 N to 10.47 N according to Table 2. For each video, the three horizontal displacements estimations $\hat{\boldsymbol{\delta}}^h$, $\hat{\boldsymbol{\delta}}^h_1$, and $\hat{\boldsymbol{\delta}}^h_2$ are computed, followed by the three modal analyses on each of them. Each stabilization diagram is processed as described in Sec. 4.1. The results gathered in Fig. 5 show that the estimator $\hat{\delta}^h_2$ tends to give better results than $\hat{\delta}^h$ and $\hat{\delta}^h_1$ for $\gamma \geq 2.62\,\text{N}$. PCA + CP generates more stable poles for the first mode, but is much less efficient than SSI and FSSI for modes at f_3 and f_4. FSSI is more efficient than SSI: the reduced observations does not deteriorate the results and FSSI is 100 times faster than SSI. For videos with small motion amplitudes (i.e., $\gamma \leq 0.16\,\text{N}$), only the first mode is retrieved. This may be due to a quantization of the estimated displacement, caused by a low variation of the intensity of each pixel, as shown in Fig. 3 for $\gamma = 0.08\,\text{N}$.

4.3 Robustness against noise

In real acquisition conditions, high-speed videos may be corrupted by noise, especially in low-light conditions. To check the robustness of the methods against noise, Gaussian noise with standard deviation from $\sigma_n = 0$ to $\sigma_n = 4$ is added to the video with input force $\gamma = 1.31\,\text{N}$. Figure 6 shows the number of stable poles near every theoretical frequencies according to Gaussian noise standard deviation. Modes with natural frequencies f_3 and f_4 are less tolerant to noise. The multi-scale displacement estimator $\hat{\delta}^h$ is less robust than $\hat{\delta}^h_1$ and $\hat{\delta}^h_2$ for modes 2, 3, and 4. FSSI still gives similar results as SSI.

4.4 Robustness against blur

According to the depth of field of the camera, parts of the acquired scene (that are not in the same plane) may be blurred in real experimental conditions. To study the effect of blur on the considered methods, Gaussian blur is applied to each frame of the video synthesized for $\gamma = 1.31\,\text{N}$ before displacement estimation and modal analysis. Results for different values of the Gaussian blur standard deviation σ_b are shown in Fig. 7. The study shows that results of the first-scale displacement estimator $\hat{\delta}^h_1$ are improved by blur. Indeed, modal

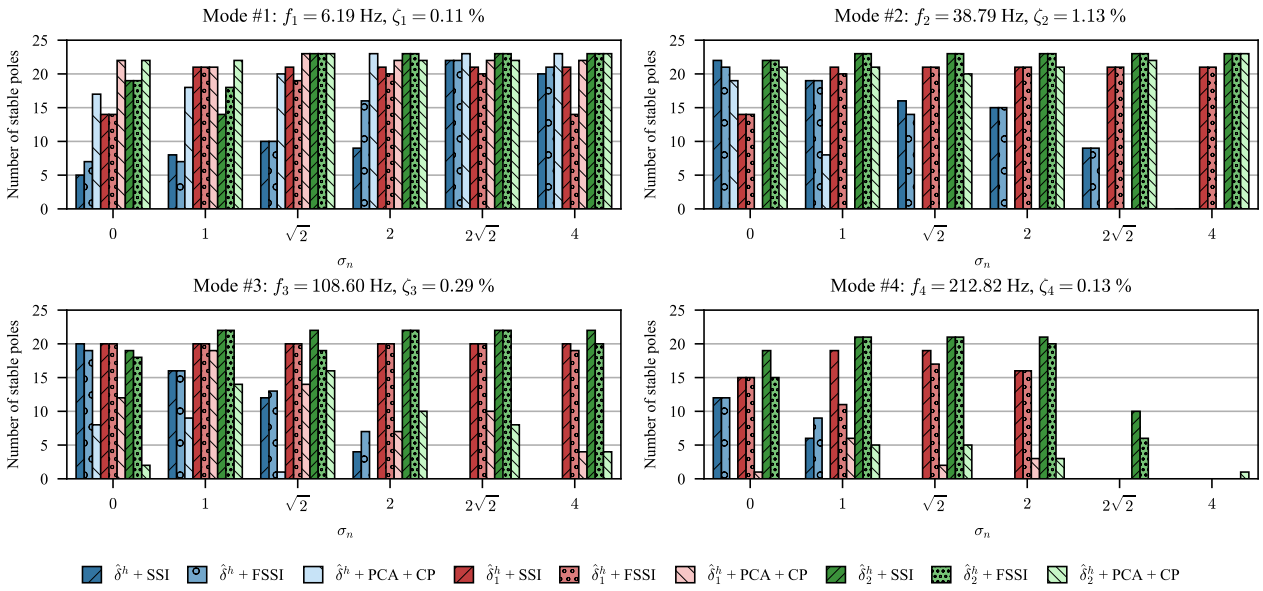

Figure 6: Number of stable poles around theoretical natural frequencies against Gaussian noise standard deviation.

analyses with $\hat{\delta}^h_1$ tend to have more stable poles with $\sigma_b = 2.0$ px than with $\sigma_b = 0$ px. In particular, PCA + CP do not retrieve the second mode from $\hat{\delta}^h_1$ with $\sigma_b = 0$ px whereas all modes are retrieved with $\sigma_b = 2.0$ px. This may be because blur modifies high spatial frequencies of each frame and that neighboring pixel values are also taken into account. SSI and FSSI methods are robust against blur, since they both correctly estimate the four modes with any displacement estimator. PCA + CP with the single-scale estimator $\hat{\delta}^h_2$ does not retrieve the third mode, and the number of stable modes around the last mode decreases as blur increases. In conclusion, each method is barely affected by blur.

4.5 Robustness against quantization

When acquisitions take place outside, illumination is uncontrolled and contrast between the structure and background may vary. Low contrast can be simulated by reducing the bit depth of intensity at each pixel. Moreover, decreasing the bit depth may be desirable to increase the acquisition frame rate. Therefore, we quantize the gray levels of the video with $\gamma = 1.31$ N onto a number of bits from 5 to 8 before motion estimation and modal analysis. The number of stable poles per method are represented in Fig. 8. The first mode is retrieved by all methods. PCA + CP needs at least 7 quantization bits to estimate the second mode, while SSI and FSSI can retrieve it with 6 quantization bits. For the third and fourth modes, the first-scale estimator $\hat{\delta}^h_1$ is less robust than the multi-scale estimator $\hat{\delta}^h$ and second-scale estimator $\hat{\delta}^h_2$ against quantization. For each displacement estimation, FSSI is as reliable as SSI.

4.6 Synthesis

To synthesize the results of each sensitivity study, for each approach we count the percentage of modes retrieved over all values of force, noise and blur standard deviations, and quantization bits. We consider that a mode is retrieved if there are at least five stable poles near its natural frequency. These percentages are gathered in Table 3. We can see that the methods that give the best results whatever the study are SSI and FSSI. As FSSI is 100 times faster thanks to dimension reduction, we conclude that FSSI is the most efficient method.

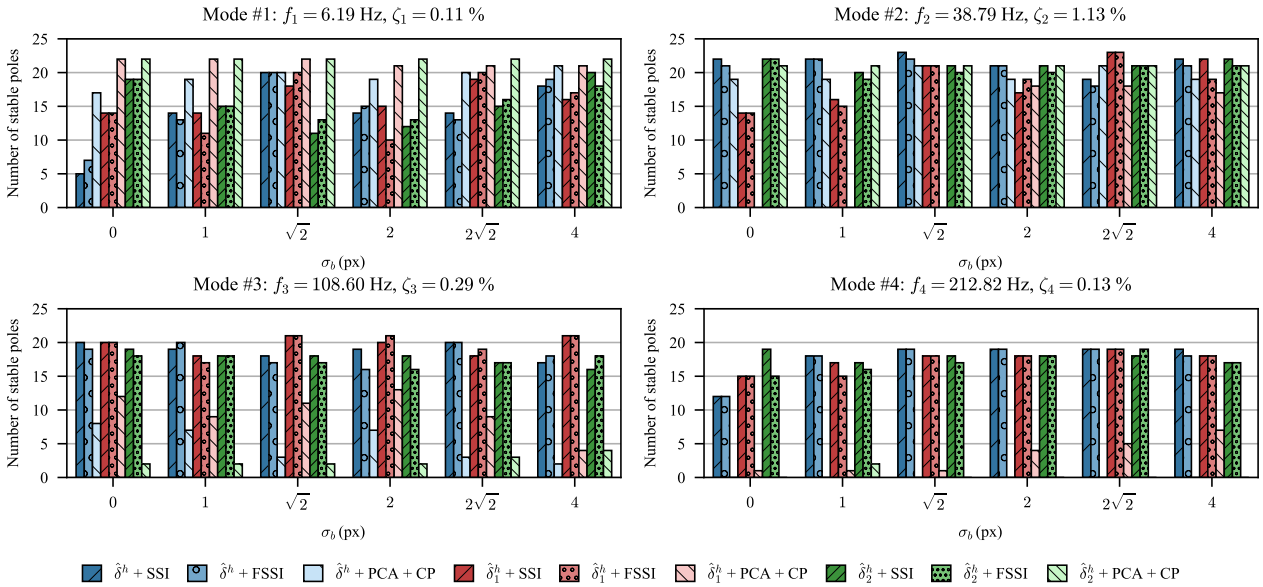

Figure 7: Number of stable poles around theoretical natural frequencies against Gaussian blur standard deviation.

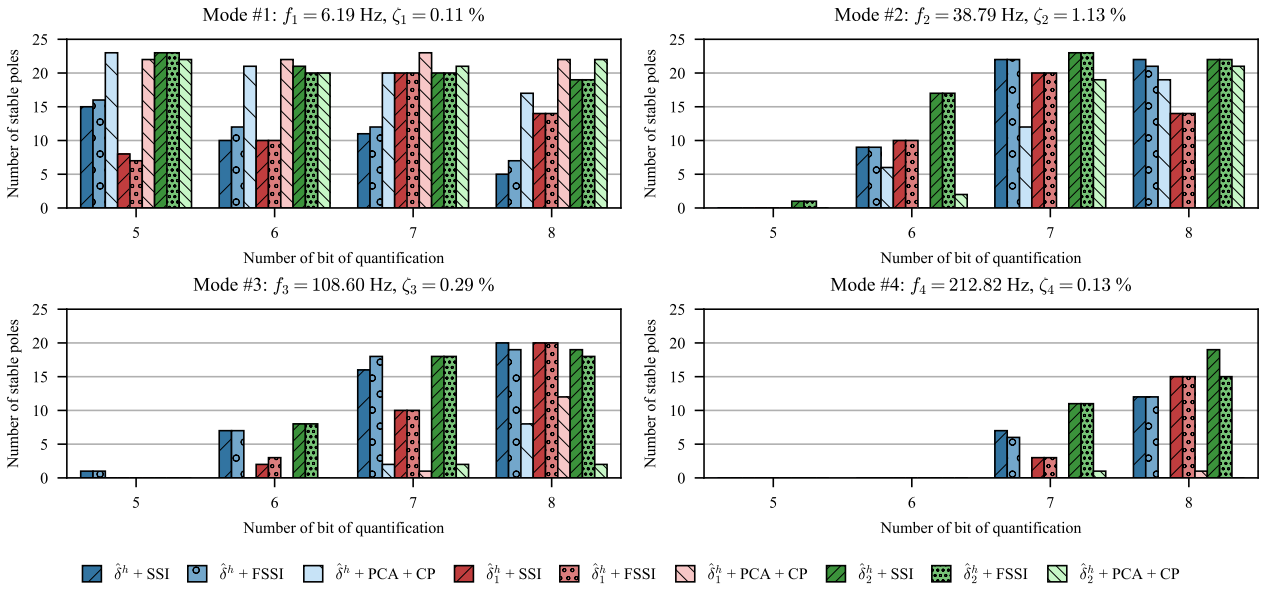

Figure 8: Number of stable poles around each theoretical natural frequencies against the number of quantization bits.

Table 3: Percentage of modes retrieved per method for each sensitivity study. Bold indicates the best result.

Displacement estimator OMA method	$\hat{\delta}^h$ SSI	$\hat{\delta}^h$ FSSI	$\hat{\delta}^h$ PCA + CP	$\hat{\delta}_1^h$ SSI	$\hat{\delta}_1^h$ FSSI	$\hat{\delta}_1^h$ PCA + CP	$\hat{\delta}_2^h$ SSI	$\hat{\delta}_2^h$ FSSI	$\hat{\delta}_2^h$ PCA + CP
Input force (γ)	59%	59%	41%	56%	56%	38%	**78%**	**78%**	47%
Noise (σ_n)	67%	71%	42%	92%	92%	50%	**96%**	**96%**	75%
Blur (σ_b)	**100%**	**100%**	62%	**100%**	**100%**	67%	**100%**	**100%**	50%
Quantization	**75%**	**75%**	50%	63%	63%	31%	**75%**	**75%**	38%

5 Conclusion

In this paper, we compare different methods to perform modal analysis from a video. These methods are split into two successive steps. First, the displacement is estimated at each frame pixel using responses of spatial-frequency filter phases. The displacement is then used as input of a modal analysis method. Three displacement estimators are compared: a multi-scale one that merges phases from the first two scales using a WLS estimator, and two single-scale estimators using the phase at first or second scale. Three modal analysis methods are also compared in combination with each displacement estimator. The first one uses a principal component analysis to reduce the dimension of the displacement matrix, and a blind source separation algorithm called complexity pursuit to estimate the modal coordinates. The second one is the classical covariance-driven stochastic subspace identification. The last one is the same as the previous, but preceded by a principal component analysis to speed up computations. The nine combinations of displacement estimator and modal analysis method are compared using synthetic videos of a vertical cantilever beam whose model is based on Euler-Bernoulli beam theory.The stabilization diagram is constructed for each method by iterating on the model order and is then automatically processed by counting stable poles around the four theoretical natural frequencies.

Different forces are applied to generate videos with different displacement amplitudes, and the comparison is achieved for different standard deviations of Gaussian noise and blur, and different numbers of quantization bits. The results show that the best method is the stochastic subspace identification on displacements estimated by the single-scale (at scale 2) estimator. However, the multi-scale estimator may be useful for videos that represent objects at different depths in the scene. Furthermore, the principal component analysis between motion estimation and stochastic subspace identification gives similar results with faster computation. However, these methods do not succeed in estimating all the modes when input force $\gamma \leq 0.33\,\mathrm{N}$ (i.e., displacement amplitude $\delta^h \leq 0.12\,\mathrm{px}$). This is due to the estimated displacement quantization, caused by low variations of the pixel intensities. Principal component analysis followed by complexity pursuit gives the worst results. The blind source separation method does not provide a good estimator of modal coordinates, so that estimated natural frequencies and damping ratios are not reliable.

Future tests on experimental videos should be performed to confirm the results of this work on synthetic videos. Moreover, these experiments focus on horizontal displacements. Extension of this work should make no assumption on the displacement direction.

References

[1] A. Brandt, *Noise and vibration analysis: signal analysis and experimental procedures*. Chichester: Wiley, 2011.

[2] S.-W. Kim and N.-S. Kim, "Multi-point displacement response measurement of civil infrastructures using digital image processing," *Procedia Engineering*, vol. 14, pp. 195–203, Oct. 2011.

[3] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Transactions on Graphics*, vol. 31, no. 4, pp. 1–8, Aug. 2012.

[4] J. Javh, J. Slavič, and M. Boltežar, "The subpixel resolution of optical-flow-based modal analysis," *Mechanical Systems and Signal Processing*, vol. 88, pp. 89–99, May 2017.

[5] J.-Y. Chou and C.-M. Chang, "Image motion extraction of structures using computer vision techniques: A comparative study," *Sensors*, vol. 21, no. 18, Sep. 2021, paper 6248.

[6] N. Wadhwa, J. G. Chen, J. B. Sellon, D. Wei, M. Rubinstein, R. Ghaffari, D. M. Freeman, O. Büyüköztürk, P. Wang, S. Sun, S. H. Kang, K. Bertoldi, F. Durand, and W. T. Freeman, "Motion microscopy for visualizing and quantifying small motions," *Proceedings of the National Academy of Sciences*, vol. 114, no. 44, pp. 11 639–11 644, Oct. 2017.

[7] Y. Yang, C. Dorn, T. Mancini, Z. Talken, G. Kenyon, C. Farrar, and D. Mascareñas, "Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification," *Mechanical Systems and Signal Processing*, vol. 85, pp. 567–590, Feb. 2017.

[8] Y. Yang and S. Nagarajaiah, "Blind modal identification of output-only structures in time-domain based on complexity pursuit," *Earthquake Engineering & Structural Dynamics*, vol. 42, no. 13, pp. 1885–1905, May 2013.

[9] B. Peeters and G. de Roeck, "Reference-based stochastic subspace identification for output-only modal analysis," *Mechanical Systems and Signal Processing*, vol. 13, no. 6, pp. 855–878, Nov. 1999.

[10] D. J. Fleet and A. D. Jepson, "Computation of component image velocity from local phase information," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 77–104, Aug. 1990.

[11] E. Simoncelli and W. Freeman, "The steerable pyramid: a flexible architecture for multi-scale derivative computation," in *Proceedings of the 2nd International Conference on Image Processing (ICIP'95)*, vol. 3, Washington, DC, USA, Oct. 1995, pp. 444–447.

[12] H. Van der Auweraer and B. Peeters, "Discriminating physical poles from mathematical poles in high order systems: use and automation of the stabilization diagram," in *Proceedings of the 21st IEEE Instrumentation and Measurement Technology Conference (IMTC'04)*, vol. 3, Como, Italy, May 2004, pp. 2193–2198.

[13] C. Shih, Y. Tsuei, R. Allemang, and D. Brown, "Complex mode indication function and its applications to spatial domain parameter estimation," *Mechanical Systems and Signal Processing*, vol. 2, no. 4, pp. 367–377, Oct. 1988.

Appendix

A Notations

$\boldsymbol{A}$	Matrix
$f(x)$	Continuous function
$f[x]$	Discrete function
$\hat{f}$	Estimator of f
t	Continuous time
k	Discrete time