UNIVERSITÉ DE LILLE

THÈSE

pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ DE LILLE

dans la spécialité
TRAITEMENT DU SIGNAL ET DES IMAGES

par
Cédric Marinel

Modal characterization of mechanical structures by automatic motion analysis from videos

Caractérisation modale de structures mécaniques par analyse automatique du mouvement à partir de vidéos

Thèse soutenue le 29 novembre 2023 devant le jury composé de :

Mme	CATHERINE ACHARD	Professeure, ISIR, Sorbonne Université	(Présidente)
M.	OLIVIER ALATA	Professeur, Lab. Hubert Curien, Université Jean Monnet	(Rapporteur)
M.	FABRICE MERIAUDEAU	Professeur, ICMUB, Université de Bourgogne	(Rapporteur)
M.	OLIVIER LOSSON	Maître de Conférences, CRIStAL, Université de Lille	(Examinateur)
M.	BENJAMIN MATHON	Maître de Conférences, CRIStAL, Université de Lille	(Examinateur)
M.	JEAN LE BESNERAIS	Docteur, EOMYS Engineering	(Examinateur)
M.	LUDOVIC MACAIRE	Professeur, CRIStAL, Université de Lille	(Directeur)

CRIStAL UMR 9189
BÂTIMENT ESPRIT AVENUE HENRI POINCARÉ, 59655 VILLENEUVE D'ASCQ
FRANCE

Abstract

Vibration monitoring plays a crucial role in ensuring the safety, reliability, and performance of mechanical structures such as bridges and wind turbines. It involves the continuous or periodic measurement and analysis of vibrations to assess the structure state, detect anomalies, and trigger maintenance decisions. Vibration monitoring systems are generally composed of vibration sensors such as accelerometers, a data acquisition device, and a software for data analysis. Because classical contact vibration sensors require complicated setup, this thesis focuses on video-based vibration analysis. Video-based methods perform remote measurements and provide vibration data at each pixel of the video frames. The operational modal analysis from these data determines the structure mechanical properties.

In this doctoral work conducted with EOMYS Engineering company, we investigate video motion estimation methods. We especially focus on phase-based methods that rely on the analysis of the space–frequency decomposition of video frames into complex subbands. These methods provide a dense motion estimation of the scene. A multi-subband approach, based on the phase fusion of the full space-frequency decomposition, is proposed and compared with the state of the art.

As motion can be estimated at every pixel of each frame, the amount of data is not suited to classical operational modal analysis. Thus, a new video-based operational modal analysis, based on a data reduction technique, is proposed and compared with a state-of-the-art video-based method.

The experimental comparisons are first conducted on synthetic videos of a vibrating cantilever beam, with different motion characteristics. Two experimental setups of straight and bent cantilever beams are finally studied to assess the performances of the methods on real video data acquired by a high-speed camera in controlled conditions.

RÉSUMÉ

La surveillance vibratoire joue un rôle crucial pour garantir la sécurité, la fiabilité et les performances de structures mécaniques. Elle implique la mesure et l'analyse, continue ou périodique, des vibrations afin d'évaluer l'état des structures, de détecter leurs anomalies et d'éclairer les décisions de maintenance. Les systèmes de surveillance des vibrations sont généralement composés de capteurs de vibration tels que les accéléromètres, d'un dispositif d'acquisition de données et d'un logiciel pour l'analyse des données. Comme les capteurs de vibration par contact peuvent être complexes à mettre en place, cette thèse se concentre sur l'estimation des vibrations basée sur une analyse de vidéos de la scène. Ces méthodes par vidéo permettent une mesure sans contact et fournissent des données de vibration en chaque pixel de l'image. L'analyse modale opérationnelle de ces données fournit des informations sur les propriétés mécaniques de la structure.

Ce travail doctoral, en collaboration avec l'entreprise EOMYS Engineering, porte sur l'étude des méthodes d'estimation du mouvement par analyse de vidéos. Elle met l'accent sur les méthodes basées sur la phase issue de filtres complexes. Ces méthodes s'appuient sur l'analyse de la décomposition spatio-fréquencielle des images de la vidéo en sous-bandes complexes. Elles fournissent une estimation dense du mouvement de la scène. Une méthode multi-échelle, basée sur une décomposition spatio-fréquencielle des images en sous-bandes, est mise en œuvre. Les performances de la méthode sont comparées à celles de l'état de l'art.

Comme l'estimation du mouvement peut être réalisée en chaque pixel de l'image, la quantité de données n'est pas adaptée aux méthodes classiques d'analyse modale opérationnelle. Une nouvelle méthode basée sur la vidéo, réduisant les données d'entrée de l'analyse modale, est donc proposée et comparée à une méthode de l'état de l'art utilisant la même approche.

Les comparaisons sont d'abord effectuées sur des vidéos synthétiques d'une poutre en porte-à-faux vibrant avec différentes caractéristiques de mouvement. Les analyses opérationnelles de deux poutres en porte-à-faux, l'une droite et l'autre pliée, sont finalement réalisées à l'aide de vidéos acquises par une caméra haute cadence qui opère en conditions contrôlées, pour évaluer les performances des méthodes.

Remerciements

Je tiens tout d'abord à remercier Jean pour m'avoir fait confiance tout au long de ce projet. La vie d'une petite entreprise est souvent riche et nos échanges m'ont permis de garder le cap durant ces trois années.

Je remercie vivement Ludovic, Olivier et Benjamin pour vos conseils, votre écoute et votre bienveillance tout au long de la thèse. Merci pour votre disponibilité et votre souci du détail, j'ai énormément appris grâce à vous et travailler à vos côtés fut un réel plaisir.

Je remercie ensuite tous les membres du jury de thèse. Merci aux Pr. Olivier Alata et Pr. Fabrice Meriaudeau pour vos rapports et vos questions lors de la soutenance. Merci également à la Pr. Catherine Achard pour avoir présidé la soutenance ainsi que pour vos retours et vos questions lors de celle-ci.

Je remercie aussi tous les membres de l'équipe d'EOMYS. Bien qu'ayant un sujet à la marge des activités de l'entreprise, je me suis toujours senti à ma place grâce à nos échanges, et c'est avec plaisir que je continue l'aventure. Spéciale dédicace à Martin pour avoir répondu à mes questions tout au long de ces trois années et pour la démonstration d'analyse modale à toute l'équipe du laboratoire.

Mes remerciements vont ensuite aux autres membres du laboratoire CRIStAL, avec qui j'ai pû échanger. Merci à Anis, Félix ainsi qu'à toute l'équipe 3D SAM pour ces discussions. Merci également aux membres du conseil de l'école doctorale, auquel j'ai pris part avec plaisir durant deux années.

Pour finir, je remercie ma famille, en commencant par mes parents pour m'avoir accompagné dans toutes mes tergiversations académiques et pour leur présence lors de la soutenance. Et enfin, ma compagne, pour son soutien indéfectible durant l'entiereté de mes études.

Nomenclature

Multidimensional variables are indicated in bold font.

$\boldsymbol{p}$	Pixel coordinates
x	Horizontal pixel coordinate
y	Vertical pixel coordinate
z	Beam curvilinear coordinate in scene or frame
k	Frame number
t	Time
f	Frequency
ω	Angular frequency
g	Euler–Bernoulli cantilever beam function
h	Impulse response function
j	Imaginary unit
s	Laplace complex frequency variable
f_i	Natural frequency of the i^{th} mode
ζ_i	Damping ratio of the i^{th} mode
$\boldsymbol{\phi}_i$	Mode shape of the i^{th} mode
r	Spatial filter scale
θ	Spatial filter orientation
N_x	Number of horizontal pixels
N_y	Number of vertical pixels

$N_{\boldsymbol{p}}$	Number of pixels
N_k	Number of frames
N_r	Number of spatial filter scales
N_θ	Number of spatial filter orientations
N_{PoI}	Number of pixels of interest
N_m	Number of physical modes in the frequency range
$I[\boldsymbol{p}; k]$	Gray level of pixel $\boldsymbol{p}$ in frame k
$G_{r,\theta}$	Spatial Gabor filter associated with subband (r, θ)
$\boldsymbol{\omega}_{r,\theta} = \left(\omega^h_{r,\theta}, \omega^v_{r,\theta}\right)^\mathsf{T}$	Central spatial frequency of the filter
$S_{r,\theta}$	Complex response of the filtered frame
$\rho_{r,\theta}$	Complex response amplitude of the filtered frame
$\varphi_{r,\theta}$	Complex response phase of the filtered frame
$\tilde{\boldsymbol{\omega}}_{r,\theta}$	Local spatial frequency of the complex response
$\boldsymbol{\delta}[\boldsymbol{p}; k]$	Displacement at pixel $\boldsymbol{p}$ in frame k
$\delta^h[\boldsymbol{p}; k]$	Horizontal displacement at pixel $\boldsymbol{p}$ in frame k
$\delta^v[\boldsymbol{p}; k]$	Vertical displacement at pixel $\boldsymbol{p}$ in frame k
$\hat{\boldsymbol{\delta}}[\boldsymbol{p}; k]$	Estimated displacement at pixel $\boldsymbol{p}$ in frame k
$\hat{\boldsymbol{\delta}}[z; k]$	Estimated displacement at pixel of interest z in frame k
$\mathfrak{Re}$	Real part
$\mathfrak{Im}$	Imaginary part
$\lfloor . \rceil$	Round operator
$\lceil . \rceil$	Ceiling function
$[\![a, b]\!]$	Interval of all integers between a and b included
$\boldsymbol{\nabla}$	Spatial gradient operator

Contents

Introduction

Vibration monitoring

Mechanical structures can be subjected to various types of vibrations, including ambient vibrations from environmental factors (e.g., wind and traffic), operational vibrations from equipment and machinery (e.g., rotor of wind turbine), and dynamic vibrations during seismic events.

Vibration monitoring plays a crucial role in ensuring the safety, reliability, and performance of mechanical structures such as bridges and wind turbines. It involves the continuous or periodic measurement and the analysis of structural vibrations to assess their state and detect anomalies. Thus, it is a core component of structural health monitoring (SHM), which is the process of continuously assessing the integrity of mechanical structures.

SHM systems use several sensors to collect data on structural vibrations, which are analyzed to characterize the structure state. By continuously monitoring vibrations, engineers can detect state changes in the structure. These changes can be indicative of damages or deterioration, such as cracks, corrosion, or fatigue [29]. It contributes to assess the resilience of the structure and can trigger safety alerts.

Vibration monitoring systems are typically composed of vibration sensors, a data acquisition device, and a software for data analysis. The sensors are strategically placed on or within the structure to collect relevant data. The acquired vibration data are processed to extract meaningful information. They are used by experimental modal analysis, when forces applied to the structure are measured, or by operational modal analysis (OMA), when the structure that runs in operational conditions, is subject to unknown forces.

Generally, contact sensors such as accelerometers are attached to the structure. However, they can alter the system's behavior or require invasive installation. Therefore, remote sensors such as laser vibrometers, have been used to acquire vibration data.

In the last decade, video-based modal analysis methods have emerged. Indeed, cameras can capture vibrations without any physical contact with the structure. This is beneficial for monitoring structures in hard-to-reach environments, such as industrial facilities. Furthermore, cameras offer full-field measurements, providing a global view of how vibrations propagate within the structure. However, as the vibration frequencies are high, videos should be acquired with a high frame rate, which often limits spatial resolution. Therefore, as vibration produces small pseudo-periodic motion whose amplitude is lower than the spatial pixel resolution, it is represented by motion at the subpixel scale.

Different state-of-the-art strategies have been proposed to perform video-based modal analysis. They combine video vibration estimation with modal analysis. As a vibration can be considered as a pseudo-periodic motion at different scales, frequency representation is exploited thanks to complex filtering. Phased-based motion estimation (PME) methods provide dense subpixel estimation of motion from the video.

Addressed challenges

Although there is a subsequent number of research works about video-based operational modal analysis [51], many challenges remain open. This industrial doctoral work, conducted at EOMYS Engineering company[1] and CRIStAL laboratory[2], addresses two key challenges.

The first one is the determination of the space–frequency subbands to estimate the subpixel motion. Indeed, PME methods rely on the phase analysis of a space–frequency subband decomposition. Different strategies exist to estimate motion, from the processing of a single subband phase to the fusion of the full space–frequency decomposition. However, no study has been conducted to compare these strategies or to study the influence of the decomposition. Our first contribution is the development of a new two-dimensional single-subband phase-based motion estimator [31]. For each pixel, this method selects the subband with the highest amplitude in a reference frame, and thus adapts the method to the local properties of the frame.

Furthermore, we show that multi-subband methods are more robust against gray level noise than single-subband ones [32]. Our second contribution is the development of a new two-dimensional multi-subband phase-based motion estimator. This method is based on the phase fusion of the full space–frequency decomposition, and provides estimations that are robust against additive intensity noise.

The second challenge concerns the high number of observations for modal analysis. Indeed, modal analysis is classically conducted with a few sensors. However, dense video-based methods estimate motion for every pixel in each frame and provide a huge number of observations. The third contribution of this work consists in a new video-based OMA method that combines a dimension reduction step and a state-of-the-art OMA method [30].

---

[1] `www.eomys.com`

[2] `www.cristal.univ-lille.fr`

Thesis structure

The thesis is divided into four chapters.

Chapter 1 deals with the monitoring of mechanical structures. The aim of this chapter is to provide an overview of the classical vibration monitoring systems. For this purpose, the theoretical basis of modal analysis is detailed. The different motion sensors are presented and an overview about motion estimation from videos is given.

Chapter 2 presents the main phase-based subpixel motion estimation methods. The concept of frame space–frequency decomposition into subbands is introduced. Then, state-of-the-art PME methods are detailed, and two new methods are presented. The different PME methods are experimentally compared on synthetic videos of a cantilever beam, with different motion characteristics.

Chapter 3 details the modal basis identification. Different classical OMA method are presented and a state-of-the-art video-based OMA method is detailed. A new OMA procedure suited to videos is developed, and its performances are compared with those of a state-of-the-art method on synthetic videos. Different PME methods are tested to study their influence on modal basis estimation. A preprocessing step is proposed to improve the OMA performances.

Chapter 4 assesses the performances reached by the developed video-based OMA method on experimental videos, acquired with a high-speed camera, of cantilever beams that vibrate according to subpixel motion amplitude. A first study is conducted on a straight beam that is equipped with accelerometers. The video-based OMA is then compared with accelerometer-based OMA. Then, a bent beam is observed by the camera to determine vibrations along several directions.

1 Mechanical structure monitoring

1.1 Introduction

Vibration monitoring plays a critical role in ensuring the safety and reliability of mechanical structures. For this purpose, their dynamic behavior should be studied. Modal analysis is a scientific field that studies the dynamic characteristics of a mechanical structure, including its natural frequencies, damping ratios, and mode shapes. These characteristics can be estimated using finite element analysis (FEA) of numerical models. However, models are based on ideal hypotheses and may be inaccurate for real systems. Indeed, measurement-based estimations are often needed to fit the FEA model.

Different modal testing methods have been developed to measure the behavior of mechanical structures. Experimental modal analysis (EMA) is an advanced technique to extract the dynamic characteristics of mechanical structures based on measured vibration data. In experimental conditions, the input forces applied to the structure are also measured.

Unlike traditional modal testing, operational modal analysis (OMA) does not require a controlled input to excite the structure. Instead, it relies on the operating vibrations of the structure, induced by external and uncontrolled sources such as wind, traffic, or other environmental factors. OMA can be performed while the structure runs in its normal operating conditions, which makes it well suited to structural health monitoring [34]. It is widely used in civil engineering to monitor bridges, buildings, and other structures [24, 28].

In this chapter, we first introduce the theoretical bases of EMA and OMA. Secondly, the theoretical characterization of systems with single and multiple degrees of freedom are detailed. Then we introduce different contact and remote sensors used to perform modal testing. A focus is finally given on the different video-based methods, including available commercial solutions.

1.2 Modal analysis

1.2.1 Introduction

The displacements of a mechanical system excited by a force do not only depend on the magnitude and direction of the force, but also on its temporal frequency. Thus, it is important to estimate the vibratory responses of the system. To do so, a mechanical system can be discretized in a given number of masses linked by springs and dampers [11]. A given force applied to a given mass influences the whole structure. This influence is characterized by frequency response functions (FRFs), generically denoted as $H(\omega)$, with ω the angular frequency. An FRF is a representation of the relationship between the input and output of a system at different frequencies (see Fig. 1.1). It describes how a system responds to sinusoidal inputs of varying frequencies, and is typically represented in terms of amplitude and phase. The amplitude component indicates how much the input signal is amplified or attenuated by the system. The phase component represents the time shift between the input force and output response signals.

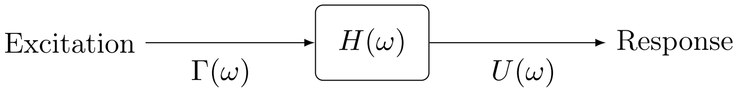

Figure 1.1: Frequency response function $H(\omega) = \dfrac{U(\omega)}{\Gamma(\omega)}$.

EMA is a technique to estimate FRFs of a system using measurements of its responses according to excitation forces. Sensors are fixed on the structure to measure the system response when a controlled excitation force is applied to a point of the structure using an impact hammer or a shaker. Multiple measurements are recorded for a single point force, and measurement noise is reduced by averaging results. Then forces are applied to other points to measure other FRFs. Fig 1.2 illustrates the EMA of a cantilever beam whose FRFs are measured at three points, denoted as 1, 2, and 3. For instance, a force is applied to a point a using an impact hammer, and the response of the structure is measured at point 3 to estimate the FRF (denoted as H_{3a}). As the relation is symmetric between two points, applying the force at a given point a and measuring the response at a point b provides the same result when a and b are switched. Once FRFs have been estimated, the modes that form the modal basis can be estimated by a modal analysis method. Each mode i is defined by a natural frequency f_i in Hz (or the equivalent natural angular frequency $\omega_i = 2\pi \cdot f_i$ in $\mathrm{rad} \cdot \mathrm{s}^{-1}$), a damping ratio ζ_i, and an associated shape $\boldsymbol{\phi}_i$. The natural frequency f_i corresponds to the frequency at which the mode resonates. The damping ratio ζ_i characterizes how the mode contribution vanishes through time when the excitation stops. The shape $\boldsymbol{\phi}_i$ is the structure deflection shape induced by the mode.

In Fig. 1.2, three modes are captured and each peak in an FRF amplitude highlights the contribution of a mode. The second mode is not visible on the FRF H_{32} because point 2 is on a static node of the mode shape ϕ_2. The system complex response can then be decomposed into a combination of three simple mode responses (see Fig 1.3).

OMA only analyzes output response measurements to estimate FRFs. To do so, the input spectrum is commonly assumed to be a white noise whose spectrum is constant, so that the response spectrum is approximately the FRF. As the excitation amplitudes are not measured, FRFs are only estimated relatively to each other, without any knowledge of their actual amplitude. Fig. 1.4 shows an example of such spectra with a white noise as input.

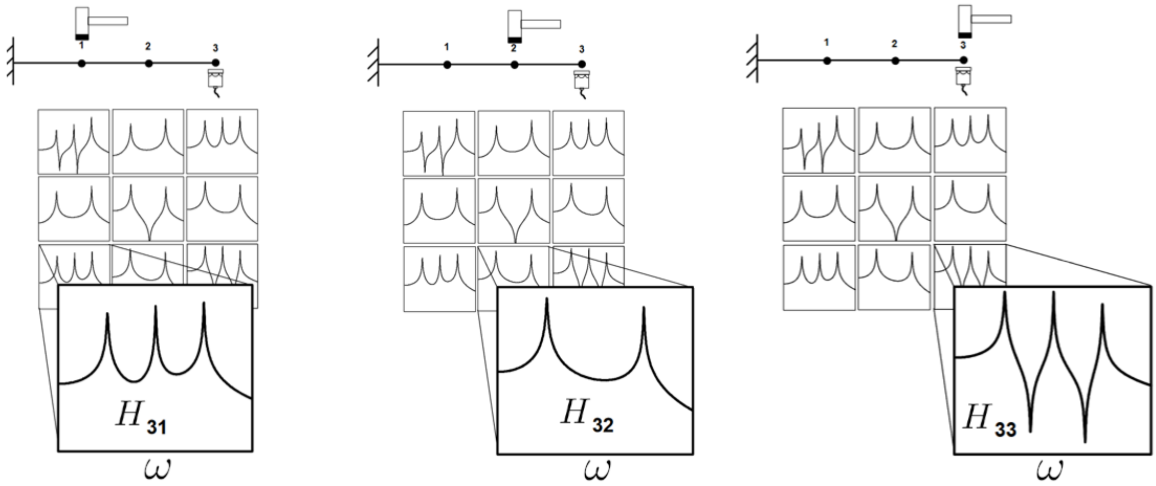

Figure 1.2: FRF measurements of a cantilever beam discretized with three points [1].

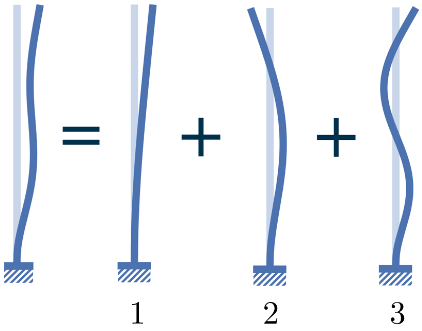

Figure 1.3: First three mode shapes of a cantilever beam.

1.2.2 Single degree of freedom

The equation of motion of a mass-spring-damper system (see Fig. 1.5), which is a single degree of freedom (SDoF) system, is:

$$m_1 \cdot \ddot{u}(t) + d_1 \cdot \dot{u}(t) + k_1 \cdot u(t) = \gamma(t), \tag{1.1}$$

with $u(t)$ the displacement response (m) and $\gamma(t)$ the input force (N). The physical parameters are the mass m_1 (kg), the viscous damping d_1 ($\mathrm{kg \cdot s^{-1}}$), and the spring stiffness k_1 ($\mathrm{N \cdot m^{-1}}$) [11].

Assuming $u(t) = U \cdot e^{s \cdot t}$, with $s \in \mathbb{C}$, and $\gamma(t) = 0$, Eq. (1.1) becomes:

$$\left(m_1 \cdot s^2 + d_1 \cdot s + k_1\right) \cdot U = 0, \tag{1.2}$$

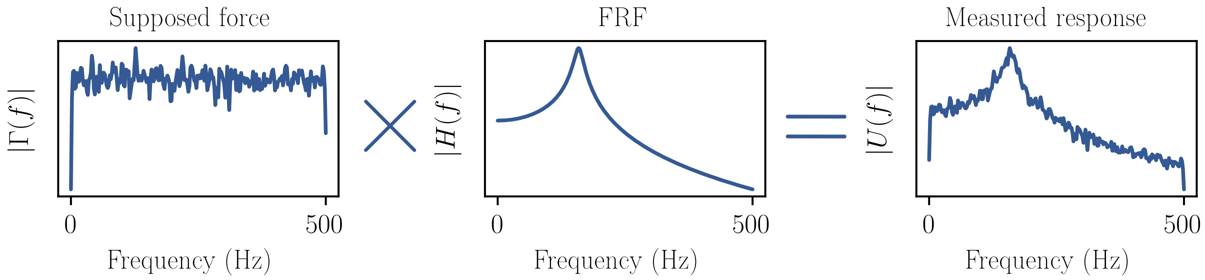

Figure 1.4: Spectrum example of an OMA with white noise input excitation.

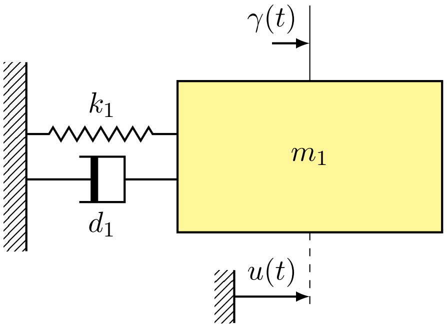

Figure 1.5: Single DoF mechanical system.

which implies:

$$m_1 \cdot s^2 + d_1 \cdot s + k_1 = 0 \;\Leftrightarrow\; s^2 + \frac{d_1}{m_1} s + \frac{k_1}{m_1} = 0 \;\Leftrightarrow\; s^2 + 2 \cdot \zeta \cdot \omega_1 \cdot s + \omega_1^2 = 0, \tag{1.3}$$

with $\omega_1 = \sqrt{\frac{k_1}{m_1}}$ the undamped natural frequency ($\mathrm{rad} \cdot \mathrm{s}^{-1}$) and $\zeta = \frac{d_1}{2\sqrt{k_1 \cdot m_1}} < 1$ the relative damping (unitless). The form of the general solution $u(t)$ is:

$$u(t) = U \cdot e^{-\omega_1 \cdot \zeta \cdot t} \cdot e^{j \cdot \left(\omega_1 \cdot \sqrt{1-\zeta^2}\right) \cdot t}. \tag{1.4}$$

1.2.2.1 Impulse response function

When the system is excited by a continuous input force $\gamma(t)$, the system response $u(t)$ is defined as:

$$u(t) := \int_0^{+\infty} h(\tau) \cdot \gamma(t - \tau) \, d\tau, \tag{1.5}$$

where $h(t)$ is the impulse response function (IRF) that characterizes how a system behaves when it is excited by a Dirac function at time $t = 0$. The IRF of a single mass-spring-damper system is expressed as:

$$h(t) = \frac{1}{m_1 \cdot \omega_1 \cdot \sqrt{1-\zeta^2}} \cdot e^{-\omega_1 \cdot \zeta \cdot t} \cdot \sin\left(\omega_1 \cdot \sqrt{1-\zeta^2} \cdot t\right). \tag{1.6}$$

The sinusoid frequency $\omega_d = \omega_1 \cdot \sqrt{1-\zeta^2}$ is called the damped frequency. An example of IRF is represented in Fig. 1.6 (a).

The so-called transfer function is the Laplace transform of the impulse response function.

1.2.2.2 Frequency response function

The frequency response function (FRF), denoted as $H(f)$, characterizes how a system responds to a given input force with temporal frequency f (Hz). The FRF can also be defined with respect to the angular frequency ω $(\mathrm{rad} \cdot \mathrm{s}^{-1})$ and is then denoted $H(\omega)$ for the sake of simplicity. It is the transfer function $H(s)$ processed at the imaginary values $s = j \cdot \omega$, with $\omega = 2\pi \cdot f$ the angular frequency. The transfer function is computed by taking the Laplace transform of Eq. (1.1):

$$\left(m_1 \cdot s^2 + d_1 \cdot s + k_1\right) \cdot U(s) = \Gamma(s), \tag{1.7}$$

$$H(s) = \frac{U(s)}{\Gamma(s)} \tag{1.8}$$

$$= \frac{1}{m_1 \cdot s^2 + d_1 \cdot s + k_1}. \tag{1.9}$$

$$\tag{1.10}$$

The FRF is then computed as:

$$H(\omega) = \frac{U(\omega)}{\Gamma(\omega)} \tag{1.11}$$

$$= \frac{1}{m_1\left(-\omega^2 + \frac{d_1}{m_1} j \cdot \omega + \frac{k_1}{m_1}\right)} \tag{1.12}$$

$$= \frac{1}{m_1\left(-\omega^2 + 2j \cdot \omega \cdot \zeta \cdot \omega_1 + \omega_1^2\right)} \tag{1.13}$$

$$= \frac{1}{k_1\left(-\left(\frac{\omega}{\omega_1}\right)^2 + 2j \cdot \zeta \cdot \frac{\omega}{\omega_1} + 1\right)}, \tag{1.14}$$

$$H(f) = \frac{1}{k_1\left(-\left(\frac{f}{f_1}\right)^2 + 2j \cdot \zeta \cdot \frac{f}{f_1} + 1\right)}, \tag{1.15}$$

with $f_1 = \dfrac{\omega_1}{2\pi}$. Examples of FRFs for different values of the damping ratio ζ are represented in Fig. 1.6 (b). The lower the damping ratio is, the larger the FRF amplitude is. Furthermore, as the FRF amplitude peak is located at the frequency $f_{\max} = f_1\sqrt{1-2\zeta^2}$, the peak locations of the represented FRF amplitudes depend on the damping ratio ζ and are thus slightly different.

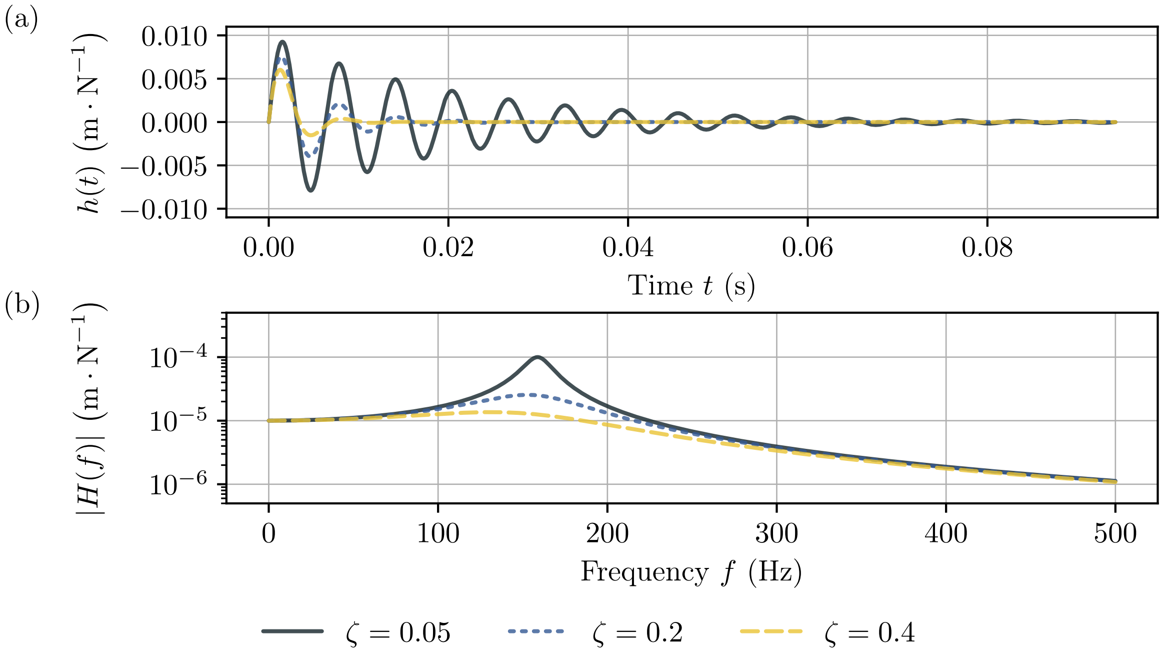

Figure 1.6: Response function of a single DoF system. (a) Impulse response function, (b) Frequency response function amplitude.

1.2.3 Multiple degrees of freedom

1.2.3.1 Equation of motion

The equation of motion of a N mass-spring-damper system, that form a multiple DoF (MDoF) system, is:

$$\boldsymbol{M}\cdot\ddot{\boldsymbol{u}}(t) + \boldsymbol{D}\cdot\dot{\boldsymbol{u}}(t) + \boldsymbol{K}\cdot\boldsymbol{u}(t) = \boldsymbol{\gamma}(t), \tag{1.16}$$

with $\boldsymbol{u}(t) \in \mathbb{R}^N$ the displacement responses of the N masses, $\boldsymbol{\gamma}(t) \in \mathbb{R}^N$ the forces applied to the N masses, $\boldsymbol{M} \in \mathbb{R}^{N\times N}$ the mass matrix, $\boldsymbol{D} \in \mathbb{R}^{N\times N}$ the damping matrix, and $\boldsymbol{K} \in \mathbb{R}^{N\times N}$

the stiffness matrix.

Figure 1.7 shows an example with $N = 2$ masses and the following equation of motion:

$$\begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix} \begin{pmatrix} \ddot{u}_1(t) \\ \ddot{u}_2(t) \end{pmatrix} + \begin{pmatrix} d_1 + d_2 & -d_2 \\ -d_2 & d_2 + d_3 \end{pmatrix} \begin{pmatrix} \dot{u}_1(t) \\ \dot{u}_2(t) \end{pmatrix} + \begin{pmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_2 + k_3 \end{pmatrix} \begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix} = \begin{pmatrix} \gamma_1(t) \\ \gamma_2(t) \end{pmatrix}. \tag{1.17}$$

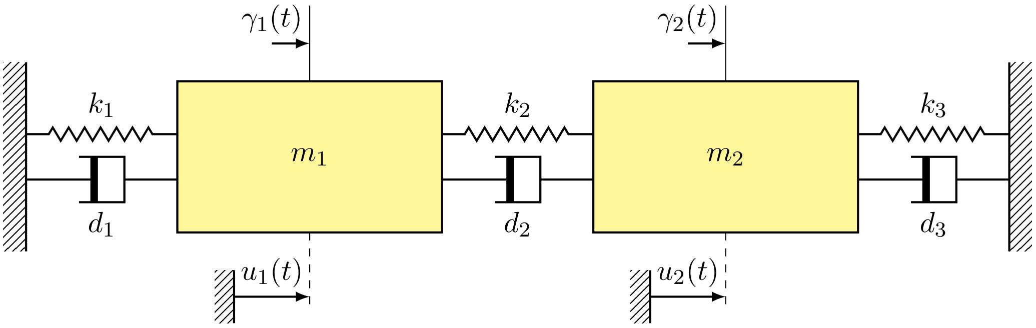

Figure 1.7: Multiple mass-spring-damper system.

Let us first consider zero excitation, *i.e.*, $\boldsymbol{\gamma}(t) = \mathbf{0}$, and assume a solution with the following form:

$$\boldsymbol{u}(t) = \boldsymbol{U} \cdot e^{s \cdot t}. \tag{1.18}$$

Substituting Eq. (1.18) into (1.16) gives:

$$\left(s^2 \cdot \boldsymbol{M} + s \cdot \boldsymbol{D} + \boldsymbol{K}\right) \cdot \boldsymbol{U} = 0. \tag{1.19}$$

This system has $2N$ solutions $\{s_i, \boldsymbol{\phi}_i\}$ that arise in complex conjugate pairs, with s_i an eigenvalue of the coefficient matrix of the system described by Eq. (1.19) and $\boldsymbol{\phi}_i$ the associated eigenvector, that is a mode shape. Denoting $\boldsymbol{\phi}_i^H$ the conjugate transpose (a.k.a. the Hermitian transpose) of a given vector $\boldsymbol{\phi}_i$, the form of an eigenvalue s_i is:

$$s_i = -\omega_i \cdot \zeta_i + j \cdot \omega_i \cdot \sqrt{1 - \zeta_i^2}, \tag{1.20}$$

where:

$$\omega_i = \sqrt{\frac{\tilde{k_i}}{\tilde{m_i}}}, \tag{1.21}$$

$$\zeta_i = \frac{\tilde{d_i}}{2\sqrt{\tilde{k_i} \cdot \tilde{m_i}}}, \tag{1.22}$$

$$\tilde{m_i} = \boldsymbol{\phi}_i^H \cdot \boldsymbol{M} \cdot \boldsymbol{\phi}_i, \tag{1.23}$$

$$\tilde{d_i} = \boldsymbol{\phi}_i^H \cdot \boldsymbol{D} \cdot \boldsymbol{\phi}_i, \tag{1.24}$$

$$\tilde{k_i} = \boldsymbol{\phi}_i^H \cdot \boldsymbol{K} \cdot \boldsymbol{\phi}_i. \tag{1.25}$$

The displacement response $\boldsymbol{u}(t)$ is a superposition of single DoF solutions (see Eq. (1.4)):

$$\boldsymbol{u}(t) = \sum_{i=1}^{N} \left(\left(e^{-\omega_i \cdot \zeta_i \cdot t} \cdot e^{j \cdot \omega_i \cdot \sqrt{1-\zeta_i^2} \cdot t} \right) \cdot \boldsymbol{\phi}_i + \left(e^{-\omega_i \cdot \zeta_i \cdot t} \cdot e^{-j \cdot \omega_i \cdot \sqrt{1-\zeta_i^2} \cdot t} \right) \cdot \boldsymbol{\phi}_i^* \right), \tag{1.26}$$

with $\boldsymbol{\phi}_i^*$ the conjugate of $\boldsymbol{\phi}_i$. This superposition principle is important because it states that the mechanical behavior of a complex structure can be expressed as that of a sum of simple mass-spring-damper systems.

1.2.3.2 Continuous state-space formulation

Each element of a multiple DoF FRF matrix expresses the relationship between the force applied at one DoF and the response of another DoF. It is a key of vibration analysis and is often estimated before the modal basis. To formulate it, Eq. (1.16) must be recast into a continuous state-space problem:

$$\underbrace{\begin{pmatrix} \boldsymbol{D} & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{0}_N \end{pmatrix}}_{\boldsymbol{A}} \underbrace{\begin{pmatrix} \dot{\boldsymbol{u}}(t) \\ \ddot{\boldsymbol{u}}(t) \end{pmatrix}}_{\dot{\boldsymbol{x}}(t)} + \underbrace{\begin{pmatrix} \boldsymbol{K} & \boldsymbol{0}_N \\ \boldsymbol{0}_N & -\boldsymbol{M} \end{pmatrix}}_{\boldsymbol{B}} \underbrace{\begin{pmatrix} \boldsymbol{u}(t) \\ \dot{\boldsymbol{u}}(t) \end{pmatrix}}_{\boldsymbol{x}(t)} = \underbrace{\begin{pmatrix} \gamma(t) \\ (\boldsymbol{0}) \end{pmatrix}}_{\boldsymbol{\gamma}(t)}, \tag{1.27}$$

with $\boldsymbol{x}(t) \in \mathbb{R}^{2N}$ the state vector, $\boldsymbol{A}$ and $\boldsymbol{B} \in \mathbb{R}^{2N \times 2N}$, $\boldsymbol{0}_N \in \mathbb{R}^{N \times N}$ a matrix of zeros, and $(\boldsymbol{0}) \in \mathbb{R}^N$ a vector of zeros.

Each eigenvector of the coefficient matrix of the system described by Eq. (1.27) without input force are of the form $\boldsymbol{\psi}_i = \begin{pmatrix} \boldsymbol{\phi}_i \\ s_i \cdot \boldsymbol{\phi}_i \end{pmatrix}$, $i = 1, \ldots, 2N$, with s_i the associated eigenvalue of the same form as in Eq. (1.20). The solutions arise as N pairs of complex conjugates. The matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ respect the following orthogonal properties:

$$\boldsymbol{\psi}_i^\intercal \boldsymbol{A} \boldsymbol{\psi}_j := \boldsymbol{M}^{\boldsymbol{A}}_{ij} = \begin{cases} \alpha_i := \boldsymbol{\phi}_i^\intercal (\boldsymbol{D} + 2 s_i \cdot \boldsymbol{M}) \boldsymbol{\phi}_i & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases} \tag{1.28}$$

$$\boldsymbol{\psi}_i^\intercal \boldsymbol{B} \boldsymbol{\psi}_j := \boldsymbol{M}^{\boldsymbol{B}}_{ij} = \begin{cases} \beta_i := \boldsymbol{\phi}_i^\intercal \left(\boldsymbol{K} - s_i^2 \cdot \boldsymbol{M}\right) \boldsymbol{\phi}_i & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases} \tag{1.29}$$

with $\boldsymbol{M}^{\boldsymbol{A}}$ called the modal $\boldsymbol{A}$ matrix, $\boldsymbol{M}^{\boldsymbol{B}}$ the modal $\boldsymbol{B}$ matrix, and $\boldsymbol{M}^{\boldsymbol{A}}_{ij}$ being the element of $\boldsymbol{M}^{\boldsymbol{A}}$ on row i and column j.
Furthermore, the $2N$ eigenvalues are $s_i = -\dfrac{\beta_i}{\alpha_i}$.

1.2.3.3 Frequency response function

Assuming $\boldsymbol{x}(0) = \boldsymbol{0}$, the Laplace transform of the state-space formulation given by Eq. (1.27) is:

$$(s \cdot \boldsymbol{A} + \boldsymbol{B}) \cdot \boldsymbol{X}(s) = \Gamma(s), \tag{1.30}$$

which can be used to compute the state-space FRF $\boldsymbol{G}(s)$ as:

$$\boldsymbol{G}(s)^{-1} = (s\boldsymbol{A} + \boldsymbol{B}) \in \mathbb{R}^{2N \times 2N} \tag{1.31}$$

$$\boldsymbol{\Psi}^\intercal \boldsymbol{G}(s)^{-1} \boldsymbol{\Psi} = \boldsymbol{\Psi}^\intercal (s\boldsymbol{A} + \boldsymbol{B}) \boldsymbol{\Psi} \tag{1.32}$$

$$\boldsymbol{\Psi}^\intercal \boldsymbol{G}(s)^{-1} \boldsymbol{\Psi} = \left(s\boldsymbol{M}^{\boldsymbol{A}} + \boldsymbol{M}^{\boldsymbol{B}}\right) \tag{1.33}$$

$$\boldsymbol{G}(s) = \boldsymbol{\Psi} \underbrace{\left(s\boldsymbol{M}^{\boldsymbol{A}} + \boldsymbol{M}^{\boldsymbol{B}}\right)^{-1}}_{\boldsymbol{\Xi}} \boldsymbol{\Psi}^\intercal, \tag{1.34}$$

where $\boldsymbol{\Psi}$ is the state-space mode shape matrix, whose i^{th} column corresponds to $\boldsymbol{\psi}_i$, and $\boldsymbol{\Xi}$ is the diagonal matrix defined by:

$$\operatorname{diag}(\boldsymbol{\Xi}) = \left\{ \frac{1}{s \cdot \alpha_1 + \beta_1}, \dots, \frac{1}{s \cdot \alpha_{2N} + \beta_{2N}} \right\}. \tag{1.35}$$

The FRF $\boldsymbol{H}(\omega)$ is the block matrix composed of the N first rows and columns of the state-space FRF $\boldsymbol{G}(s)$ evaluated at the imaginary value $s = j \cdot \omega$:

$$\boldsymbol{H}_{pq}(\omega) = \sum_{i=1}^{N} \left(\frac{\boldsymbol{\Phi}_{pi}\boldsymbol{\Phi}_{qi}}{\alpha_i \cdot \left(\zeta_i \cdot \omega_i + j\left(\omega - \omega_i\sqrt{1-\zeta_i^2}\right)\right)} + \frac{\boldsymbol{\Phi}^*_{pi}\boldsymbol{\Phi}^*_{qi}}{\alpha^*_i \cdot \left(\zeta_i \cdot \omega_i + j\left(\omega + \omega_i\sqrt{1-\zeta_i^2}\right)\right)} \right), \tag{1.36}$$

with $\boldsymbol{\Phi}$ the mode shape matrix, whose i^{th} column corresponds to mode shape $\boldsymbol{\phi}_i$. $\boldsymbol{\Phi}_{pi}$ indicates the value of DoF p in the mode shape $\boldsymbol{\phi}_i$ and $\boldsymbol{H}_{pq}(\omega)$ corresponds to the FRF matrix element on row p and column q. An example of FRF of a system with $N = 2$ DoFs is given in Fig. 1.8. The FRF matrix $\boldsymbol{H}$ is symmetric, *i.e.*, $\boldsymbol{H}_{pq}(\omega) = \boldsymbol{H}_{qp}(\omega)$. Its peaks characterize resonances. Around $\omega = 12\,\text{rad}{\cdot}\text{s}^{-1}$, the FRF amplitude $|\boldsymbol{H}_{11}|$, that characterizes the response of the first mass to a force applied on it, is really low. This is called an antiresonance. When the first mass is excited by a force with the antiresonance frequency, this mass is static and only the other moves. We can see that the antiresonance frequency of the second mass is different from that of the first one.

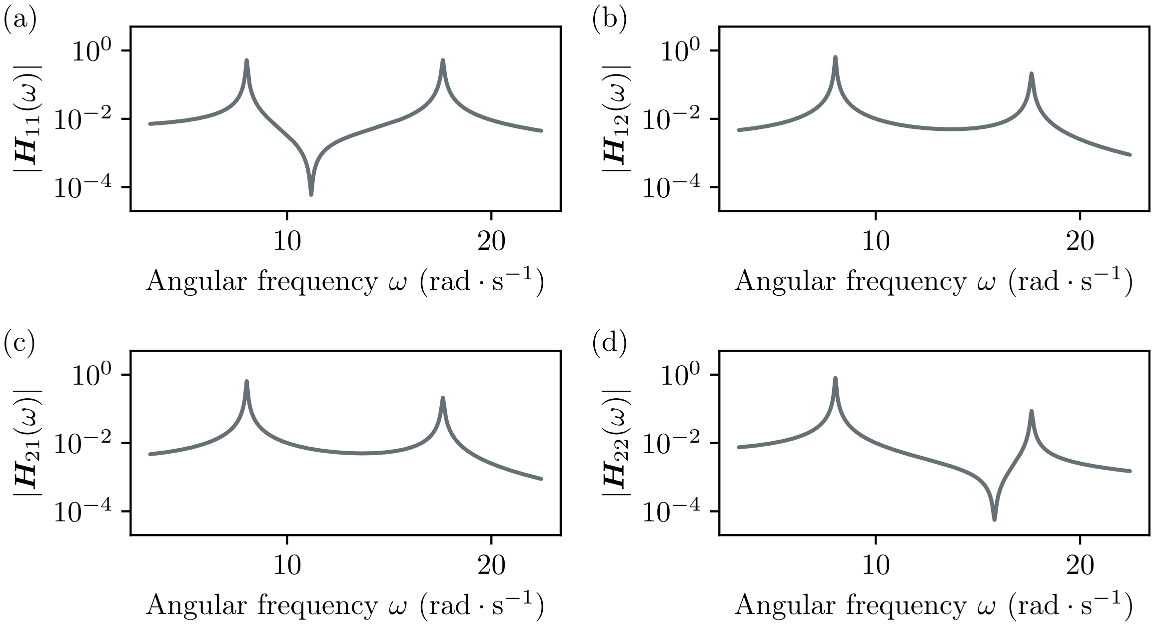

Figure 1.8: Frequency response function amplitudes of a 2-DoF system.

1.2.3.4 Operating deflection shape

Operating deflection shape (ODS) analysis is a simpler analysis than OMA. It consists in displaying the operational response of the studied system at a certain frequency or in a frequency range [39]. This quick analysis is useful to analyze the frequency behavior of a system without identifying the modal basis. For this purpose, a given DoF r is selected as that where the sinusoidal force $\boldsymbol{\gamma}_r(\omega)$ of angular frequency ω is applied. The N ODS of DoF $q = 1, \ldots, N$ are computed using inverse fast Fourier transform (FFT) as:

$$\mathrm{ODS}_q(t) = \mathrm{FFT}^{-1}\{\boldsymbol{H}_{qr}(\omega) \cdot \boldsymbol{\gamma}_r(\omega)\}. \tag{1.37}$$

As the modal basis is not identified, the ODS is a combination of the mode shapes. When natural frequencies of modes are distant from each other, the ODS displayed at an FRF peak frequency will be close to the corresponding mode shape. However, when natural frequencies are close from each other, the ODS will represent a mix of mode shapes.

In Fig. 1.9, the two modal contributions are represented together with the FRF. As the natural frequencies are distant from each other, the contribution of the second mode at the first natural frequencies is negligible compared to the first mode contribution. The ODS at this frequency will be close to the first mode. One can also note that antiresonance in H_{11} and H_{22} occur at frequencies where modal contribution have the same amplitude. This means that their directions are opposite, *i.e.*, their phases differ by π.

Figure 1.10 represents the FRFs of the third DoF of a 3-DoF system. The natural frequencies of two modes are close to 18.5 rad/s and the amplitude of first and second modal contributions are close. Thus, ODS at this frequency will represent a mix of the first two modes. In that case, the individual mode contributions and the associated mode shapes can be only identified by modal analysis.

1.2.4 Discrete state-space formulation

In experimental conditions, measurements are discrete in time and space. Furthermore, the number N of modes that can be extracted is unknown. Assuming N_p observations and white noise input, the problem described by Eq. (1.16) can be recast in a continuous state-space form:

$$\begin{pmatrix} \dot{\boldsymbol{u}}(t) \\ \ddot{\boldsymbol{u}}(t) \end{pmatrix} = \begin{pmatrix} \boldsymbol{0}_N & \boldsymbol{I}_N \\ -\boldsymbol{M}^{-1}\boldsymbol{K} & -\boldsymbol{M}^{-1}\boldsymbol{D} \end{pmatrix} \begin{pmatrix} \boldsymbol{u}(t) \\ \dot{\boldsymbol{u}}(t) \end{pmatrix} + \boldsymbol{w}(t), \tag{1.38}$$

$$\boldsymbol{y}(t) = \boldsymbol{C} \begin{pmatrix} \boldsymbol{u}(t) \\ \dot{\boldsymbol{u}}(t) \end{pmatrix} + \boldsymbol{v}(t), \tag{1.39}$$

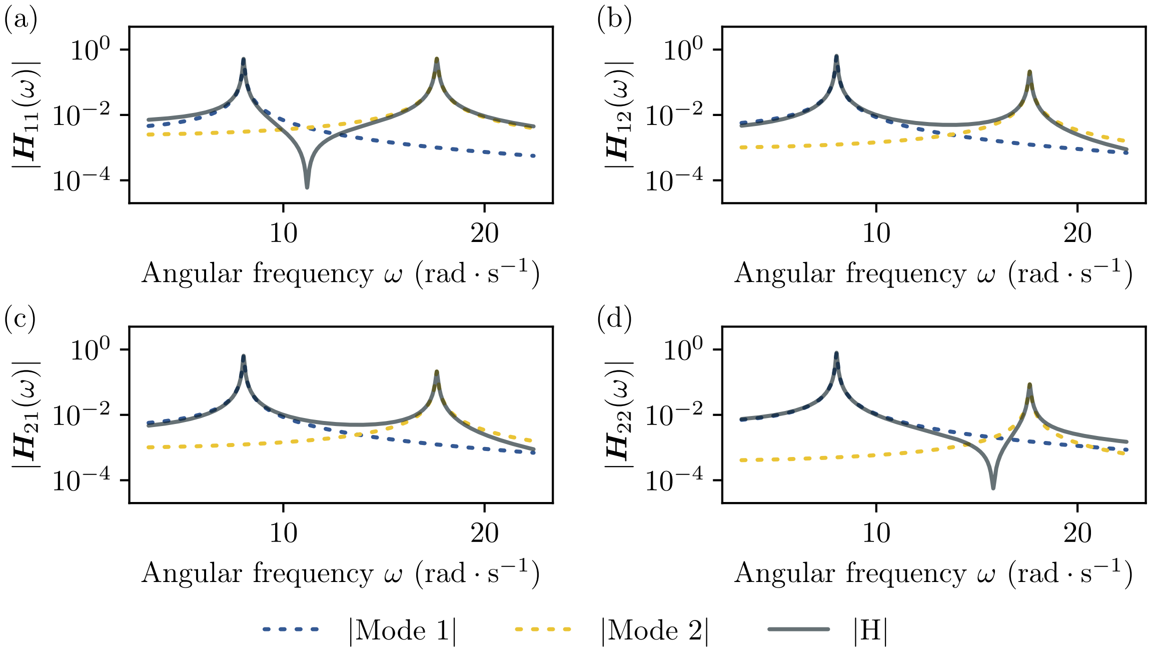

Figure 1.9: FRF amplitudes of the same 2-DoF system as in Fig. 1.8, with amplitude contributions of each mode.

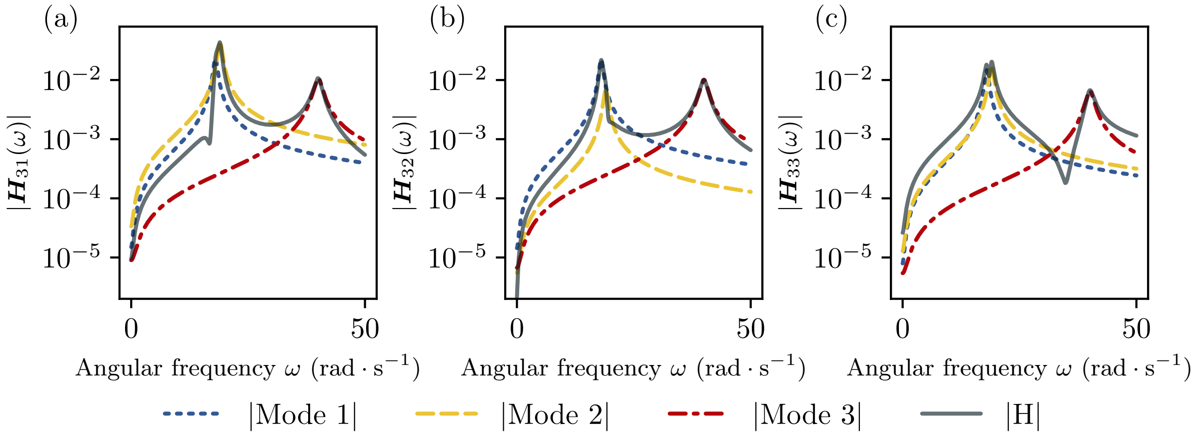

Figure 1.10: FRF amplitudes and modal contributions of a 3-DoF system with close modes.

with $\boldsymbol{C} \in \mathbb{R}^{N_p \times 2N}$ the output matrix, $\boldsymbol{y}(t) \in \mathbb{R}^{N_p}$ the observation vector, $\boldsymbol{w}(t) \in \mathbb{R}^{2N}$ and $\boldsymbol{v}(t) \in \mathbb{R}^{N_p}$ the input and observation white noise vectors. Denoting the state vector

$\boldsymbol{x}(t) = \begin{pmatrix} \boldsymbol{u}(t) \\ \dot{\boldsymbol{u}}(t) \end{pmatrix} \in \mathbb{R}^{2N}$, and the continuous state-space matrix $\boldsymbol{A}_c \in \mathbb{R}^{2N\times 2N}$, the state-space formulation becomes:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}_c\, \boldsymbol{x}(t) + \boldsymbol{w}(t), \tag{1.40}$$

$$\boldsymbol{y}(t) = \boldsymbol{C}\, \boldsymbol{x}(t) + \boldsymbol{v}(t). \tag{1.41}$$

The continuous-time state-space system described by Eqs. (1.40) and (1.41) can be discretized as:

$$\boldsymbol{x}[k+1] = \boldsymbol{A}_d\, \boldsymbol{x}[k] + \boldsymbol{w}[k], \tag{1.42}$$

$$\boldsymbol{y}[k] = \boldsymbol{C}\, \boldsymbol{x}[k] + \boldsymbol{v}[k], \tag{1.43}$$

with $\boldsymbol{x}[k] := \boldsymbol{x}(k\Delta t)$ the discrete state vector, Δt the time step between two successive measurements, $\boldsymbol{y}[k] := \boldsymbol{y}(k\Delta t)$ the discrete observation vector, $\boldsymbol{A}_d := e^{\boldsymbol{A}_c \Delta t}$ the discrete state-space matrix, and $\boldsymbol{w}[k] := \boldsymbol{w}(k\Delta t)$ and $\boldsymbol{v}[k] := \boldsymbol{v}(k\Delta t)$ the input and observation noise vectors [23].

For given observations $\boldsymbol{y}[k]$ and model order N, discrete state-space $\boldsymbol{A}_d$ and observation $\boldsymbol{C}$ matrices are estimated by the methods that will be detailed in Sec. 3.3.

Then, from the eigenvalue decomposition $\boldsymbol{A}_d = \boldsymbol{\chi} \boldsymbol{S} \boldsymbol{\chi}^{\boldsymbol{-1}}$ with $\boldsymbol{S} = \operatorname{diag}(\lambda_i), i \in [\![1, 2N]\!]$, natural frequencies f_i^N, damping ratios ζ_i^N, and mode shape matrix $\boldsymbol{\Phi}^N$ can be computed as:

$$f_i^N = \frac{1}{2\pi}\left|\frac{\log \lambda_i}{\Delta t}\right|, \tag{1.44}$$

$$\zeta_i^N = \frac{\mathfrak{Re}(\lambda_i)}{|\lambda_i|}, \tag{1.45}$$

$$\boldsymbol{\Phi}^N = \boldsymbol{C}\boldsymbol{\chi}. \tag{1.46}$$

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

Measurements in $\boldsymbol{y}$ can refer to displacement, velocity and/or acceleration. They are provided by sensors that can be split into two categories: contact and remote sensors. A brief summary of existing sensors is given in the next two sections.

1.3 Contact vibration sensing

Vibration analysis is classically performed thanks to contact sensors, with a high accuracy for frequency analysis up to kHz orders. Contact sensors specifically measure either acceleration, velocity, or displacement. As velocity can be obtained by integration of acceleration with respect to time or by derivation of displacement, sensors of different types can be used simultaneously to perform vibration analysis in acceleration, velocity, and/or displacement.

1.3.1 Accelerometer

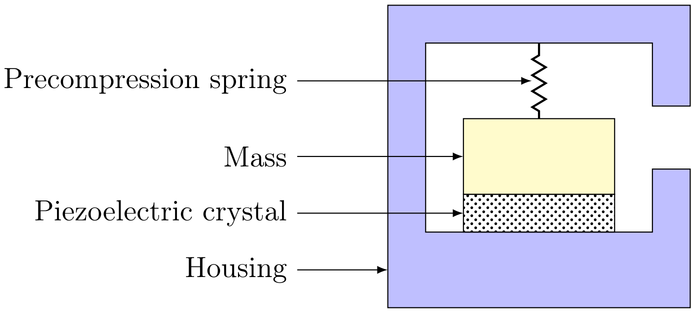

Figure 1.11: Schematic of a standard accelerometer.

The accelerometer is among of the most popular vibration sensors. It is a contact transducer that measures acceleration. It consists of a mass attached to a piezoelectric crystal, such as a quartz, which generates an electric charge proportional to the acceleration of the mass (see Fig. 1.11). The accelerometer is fixed to a structure using either adhesive, magnet, or probe tip. It delivers analog signals to a data acquisition system that converts them into digital signals. Those output signals are then processed using Fourier transform or power spectral density (PSD) to estimate the system FRF.

Depending on the application, the temporal acquisition frequency ranges from 0.1 Hz to 20 kHz. For this range, the Nyquist theorem states that the frequency analysis may be up to 10 kHz.

Accelerometers can be monoaxial, biaxial, or triaxial to measure acceleration in one, two, or three directions. Each axis of an accelerometer generates a signal and can be considered as a DoF for the modal analysis.

1.3.2 Linear variable differential transformer

A linear variable differential transformer (LVDT) is a transducer that allows for the measurement of displacement along one direction. It is composed of a body and a ferromagnetic shaft. The body has a primary coil and a pair of secondary coils on either side (see Fig 1.12). As the LVDT shaft is fixed to the structure, its displacement generates an induced current in the secondary coils. This analog signal is then sampled and stored by a data acquisition system. Its typical temporal frequency ranges from 1 Hz to 10 kHz.

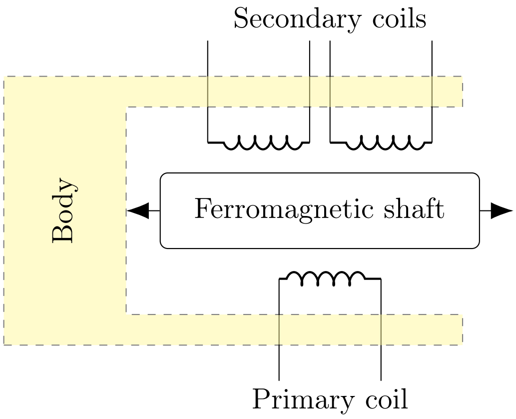

Figure 1.12: Linear variable differential transformer.

1.3.3 Contact sensor drawbacks

As contact sensors must be positioned on the structure, setting them up in operational conditions may be complicated. Maintaining such data acquisition devices for long-term monitoring is also time-consuming. Moreover, because of the mass they add to that of the structure, the estimated modal basis may be different from the structure's modal basis without mass loading.

Furthermore, each aforementioned sensor is a single-point measurement device. Experimental setup may therefore require many such sensors to get spatially dense mode shapes, which may increase the measure complexity and cost.

Different remote sensors have emerged to overcome contact sensor weaknesses.

1.4 Remote vibration sensing

1.4.1 Laser Doppler vibrometer

A basic laser Doppler vibrometer (LDV) is an axial contactless sensor that measures velocity. A laser reference beam with frequency f_0 is split to be directed both onto a photodetector and an element of the surface of interest (see Fig. 1.13). The scattered light reflected by the surface, whose Doppler frequency shift f_d is induced by the out-of-plane surface velocity, is projected onto the photodetector. The frequency shift is thus computed with the photodetector interfered signal to deduce the point out-of-plane velocity. The typical frequency working range of this device is $(0\,\text{Hz}, 20{+}\,\text{MHz}]$, with stand-off distances up to $200\,\text{m}$.

As the basic LDV measures the velocity of a single point of the surface, scanning laser Doppler vibrometers (SLDVs) have been developed to measure that of multiple points [4]. A camera is used to display the scene, so that the user defines different points of interest in the image plane. The laser measures the velocity point by point.

In-plane laser vibrometers have also been developed to measure velocity within the plane normal to the laser. Furthermore, systems combining three SLDVs allow for the measurement of velocity along three orthogonal directions.

The point-by-point asynchronous measurement is a downside of such sensors. In experimental conditions, an automatic impact hammer is able to ensure that the same force is applied for each point measurement. However, input forces are not controlled in operational conditions. Thus, an additional fixed sensor is needed as a reference for OMA.

Moreover, the measurement distance is limited by the structure reflection properties of the signal that goes back to the LDV.

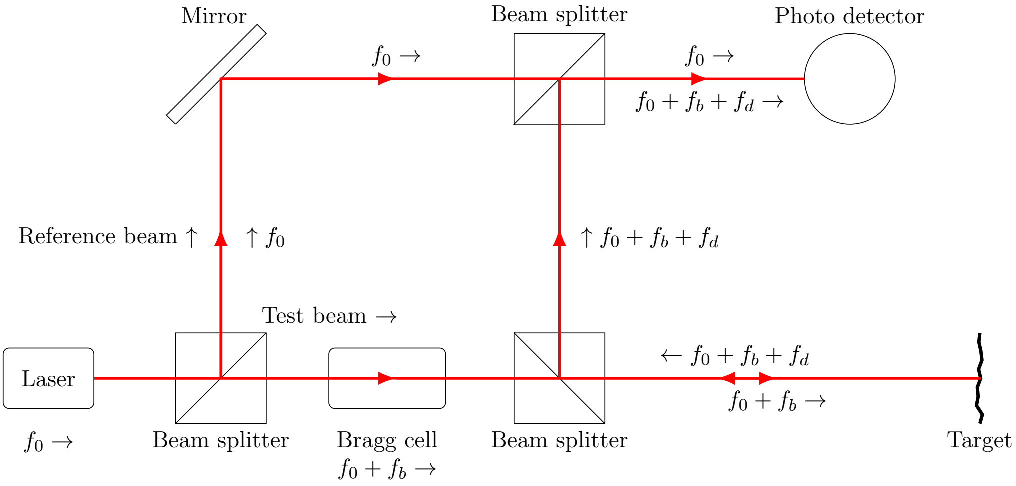

Figure 1.13: Laser Doppler vibrometer schematic[1].

1.4.2 Digital video camera

Alternatively, a digital camera acquires video frames of a scene with potential high spatial resolution. Each pixel can then be considered as a remote sensor to provide dense 2D displacement measurements in the plane orthogonal to the optical axis.

Any video-based experiment must respect several conditions to measure the displacement of a remote structure. First, the camera must be fixed without occlusion of the structure of interest. Secondly, the structure must be adequately illuminated with constant brightness to avoid external changes of brightness that would be interpreted as displacements. Without scene depth information, the structure surface of interest must also be in the plane orthogonal to the camera optical axis to correctly estimate relative displacement amplitudes.

The working temporal frequency range is limited by the camera frame rate. Indeed, the highest structure vibration frequency must be at most half of the camera frame rate to avoid spectral

---

[1] Basic schematic of a laser Doppler vibrometer: `https://en.wikipedia.org/wiki/File:LDV_Schematic.png`

aliasing. However, with multiple acquisitions at different frame rates, this frequency range can be improved [53]. Standard industrial cameras can record videos with a spatial resolution of 720×540 px and a frame rate of 539 fps (frame per second)[2]. Furthermore, high-speed cameras can increase the acquisition frame rate up to 2 457 000 fps at a resolution of 640×32 px[3]. Due to the bitrate of the interface between camera and computer, the designer chooses a relevant trade-off between frame rate and frame size. As frame rate is generally high to study vibrations with a high temporal frequency, small motions can be accurately estimated by subpixel approaches.

Moreover, an additional camera that is properly calibrated enables to extend the measurements to three dimensions using stereo vision processing [36].

1.5 Video-based vibration measurements

Structure vibrations can be measured by the analysis of N_k successive frames in a video. Let $I[\boldsymbol{p}; k]$ be the intensity at pixel $\boldsymbol{p}$ with spatial coordinates (x, y) in frame $k \in [\![0, N_k - 1]\!]$, and $\boldsymbol{\delta}$ be the displacement field, specified at a pixel $\boldsymbol{p}$ within frame k by its coordinates along horizontal and vertical directions:

$$\boldsymbol{\delta}[\boldsymbol{p}; k] = \begin{pmatrix} \delta^h[\boldsymbol{p}; k] \\ \delta^v[\boldsymbol{p}; k] \end{pmatrix} \in \mathbb{R}^2. \tag{1.47}$$

The displacement $\boldsymbol{\delta}$ is defined relatively to frame 0 and represents the structure displacement projected in the image plane. The intensity associated to a specific surface element, represented by a pixel $\boldsymbol{p}$ and that moves according to displacement $\boldsymbol{\delta}[\boldsymbol{p}; k]$, is assumed to be constant. Furthermore, when the scene is assumed to be free from any occlusion and illuminated with a spatially uniform and spectrally constant over time lighting, the intensity at pixel $\boldsymbol{p}$ in frame 0 is the same as that of pixel $\boldsymbol{p} + \boldsymbol{\delta}[\boldsymbol{p}; k]$ in frame k:

$$I[\boldsymbol{p}; 0] = I[\boldsymbol{p} + \boldsymbol{\delta}[\boldsymbol{p}; k]; k]. \tag{1.48}$$

This property is exploited to estimate displacement.

1.5.1 Image matching

Image matching techniques detect keypoints in frames using descriptors and try to match them using a specific distance. One of the most used methods is the scale invariant feature transform (SIFT) [26]. This method detects keypoints in each frame and compute descriptors using the following steps:

---

[2] TIS DMK 33UX287: `https://www.theimagingsource.com/en-fr/product/industrial/33u/dmk33ux287/`

[3] Photron Model E9 80S : `https://photron.com/pharsighted/`

1. Decompose the frame into different scales using difference of Gaussians,
2. Detect scale-space extrema pixels by comparing a pixel level with those of its neighbors in the previous, current, and next scales,
3. Eliminate low-contrast scale-space extrema by setting a threshold on maxima value of the keypoint refined by quadratic interpolation,
4. Eliminate edge responses by setting a threshold on the ratio of curvature along and perpendicular to the edge,
5. Assign an orientation to each keypoint using the orientation histogram in its neighborhood,
6. Create an invariant descriptor vector for each keypoint by computing orientation histograms in sub-neighborhoods of the keypoint relative to its orientation.

Then, the keypoints extracted from the frames are matched using the nearest neighbor rule based on the Euclidean distance between their descriptors.

There are other popular keypoint descriptors based on the same principle, such as speeded-up robust features (SURF) [2] and oriented FAST and rotated BRIEF (ORB) [40]. SURF is faster, but cannot estimate subpixel displacements.

As this family of methods automatically detects keypoints, high-contrast patterns must be located on the mechanical structure to ensure that features are adequately detected at desired places to compute the mode shape. Thus, this method can be difficult to apply when the structure is not easily accessible by an operator.

1.5.2 Image correlation

Image correlation considers a square template $\mathcal{T}_{\boldsymbol{p}}$ centered at each pixel $\boldsymbol{p}$. It determines the spatial offset between the template in frames 0 and k. For a given pixel $\boldsymbol{p}$, the template matching method searches for the best integer displacement $\boldsymbol{\Delta p} = (\Delta x, \Delta y)$ that maximizes a correlation measure between the two templates, such as the zero-mean normalized cross-correlation:

$$\hat{\boldsymbol{\delta}}[\boldsymbol{p};k] = \arg\max_{\boldsymbol{\Delta p}} \frac{\sum\limits_{\boldsymbol{q}\in\mathcal{T}_{\boldsymbol{p}}} \big(I[\boldsymbol{p}+\boldsymbol{\Delta p}+\boldsymbol{q};k] - \bar{I}[\boldsymbol{p}+\boldsymbol{\Delta p};k]\big)\cdot\big(I[\boldsymbol{p}+\boldsymbol{q};0] - \bar{I}[\boldsymbol{p};0]\big)}{\sqrt{\sum\limits_{\boldsymbol{q}\in\mathcal{T}_{\boldsymbol{p}}} \big(I[\boldsymbol{p}+\boldsymbol{\Delta p}+\boldsymbol{q};k] - \bar{I}[\boldsymbol{p}+\boldsymbol{\Delta p};k]\big)^2 \cdot \sum\limits_{\boldsymbol{q}\in\mathcal{T}_{\boldsymbol{p}}} \big(I[\boldsymbol{p}+\boldsymbol{q};0] - \bar{I}[\boldsymbol{p};0]\big)^2}}, \tag{1.49}$$

where $\bar{I}[\boldsymbol{p};k]$ is the mean level of the template $\mathcal{T}_{\boldsymbol{p}}$ [41].

A subpixel refinement can be performed by finding the maximum of the interpolated cross-correlation with quadratic surface fitting [56].

These methods are efficient when targets are fixed on the observed mechanical structure [25]. To obtain dense vibration estimations, these methods require that a speckle pattern is projected onto the structure of interest [8]. Thus, it may be difficult to apply them for monitoring of large outside structures.

1.5.3 Subpixel optical flow

Assuming constant brightness, the optical flow equation states that intensities are translated from one frame to another according to:

$$I(\boldsymbol{p}; t) = I(\boldsymbol{p} + \Delta \boldsymbol{p}; t + \Delta t). \tag{1.50}$$

In this section, equations are formulated in continuous time-space to define spatial and temporal derivatives. Those derivatives are estimated by finite difference scheme or derivative filters. The Taylor expansion at order 1 of $I(\boldsymbol{p} + \boldsymbol{\Delta p}; t + \Delta t)$ is:

$$I(\boldsymbol{p} + \boldsymbol{\Delta p}; t + \Delta t) = I(\boldsymbol{p}; t) + \boldsymbol{\nabla} I(\boldsymbol{p}; t) \cdot \boldsymbol{\Delta p} + \partial_t I(\boldsymbol{p}; t) \cdot \Delta t + O\big(\|(\boldsymbol{\Delta p}, \Delta t)\|^2\big), \tag{1.51}$$

with $\partial_t I$ the intensity time derivative and $\boldsymbol{\nabla} I$ the intensity spatial gradient. Truncating Eq. (1.50) at order 1 and injecting it into Eq. (1.51) gives:

$$\boldsymbol{\nabla} I(\boldsymbol{p}; t) \cdot \boldsymbol{\Delta p} + \partial_t I(\boldsymbol{p}; t) \cdot \Delta t = 0. \tag{1.52}$$

Denoting the velocity vector as $\boldsymbol{V}[\boldsymbol{p}] = \big(V^h[\boldsymbol{p}], V^v[\boldsymbol{p}]\big)^\intercal$, Eq. (1.52) can be reformulated in terms of velocity by dividing by Δt:

$$\boldsymbol{\nabla} I(\boldsymbol{p}; t) \cdot \boldsymbol{V}[\boldsymbol{p}] = -\partial_t I(\boldsymbol{p}; t). \tag{1.53}$$

1.5.3.1 Lucas-Kanade

The Lucas and Kanade method [27] proposes to solve a weighted least-square problem to compute the velocity $\boldsymbol{V}$ at pixel $\boldsymbol{p}$. The authors assume that the velocity is constant in a local neighborhood $\mathcal{N}_\sigma[\boldsymbol{p}]$ defined by the standard deviation σ, such that for all $\boldsymbol{q} \in \mathcal{N}_\sigma[\boldsymbol{p}]$,

$$\boldsymbol{\nabla} I(\boldsymbol{q}; t) \cdot \boldsymbol{V}[\boldsymbol{p}] = -\frac{\partial I(\boldsymbol{q}; t)}{\partial t}. \tag{1.54}$$

Each equation in this set is weighted by a Gaussian window $\mathcal{G}_\sigma$ so that the weight decreases with respect to the spatial distance between $\boldsymbol{q}$ and $\boldsymbol{p}$. Solving this problem in the least-square sense provides the velocity vector:

$$\begin{pmatrix} V^h[\boldsymbol{p}] \\ V^v[\boldsymbol{p}] \end{pmatrix} = \begin{pmatrix} \sum\limits_{\boldsymbol{q}} \mathcal{G}_\sigma[\boldsymbol{q}] \cdot \partial_x I[\boldsymbol{q}]^2 & \sum\limits_{\boldsymbol{q}} \mathcal{G}_\sigma[\boldsymbol{q}] \cdot \partial_x I[\boldsymbol{q}] \cdot \partial_y I[\boldsymbol{q}] \\ \sum\limits_{\boldsymbol{q}} \mathcal{G}_\sigma[\boldsymbol{q}] \cdot \partial_x I[\boldsymbol{q}] \cdot \partial_y I[\boldsymbol{q}] & \sum\limits_{\boldsymbol{q}} \mathcal{G}_\sigma[\boldsymbol{q}] \cdot \partial_y I[\boldsymbol{q}]^2 \end{pmatrix}^{-1} \begin{pmatrix} -\sum\limits_{\boldsymbol{q}} \mathcal{G}_\sigma[\boldsymbol{q}] \cdot \partial_x I[\boldsymbol{q}] \cdot \partial_t I[\boldsymbol{q}] \\ -\sum\limits_{\boldsymbol{q}} \mathcal{G}_\sigma[\boldsymbol{q}] \cdot \partial_y I[\boldsymbol{q}] \cdot \partial_t I[\boldsymbol{q}] \end{pmatrix}. \tag{1.55}$$

1.5.3.2 Gradient-based optical flow

The gradient-based optical flow method assumes that the displacement is along the direction of the spatial gradient [21]. Temporal derivative is approximated by the following first order forward finite difference scheme:

$$\partial_t I(\boldsymbol{p}; t) = \frac{I(\boldsymbol{p}; t + \Delta t) - I(\boldsymbol{p}; t)}{\Delta t}. \tag{1.56}$$

Injecting this approximation into Eq. (1.52) gives:

$$\boldsymbol{\nabla} I(\boldsymbol{p}; t) \cdot \boldsymbol{\Delta p} = I(\boldsymbol{p}; t) - I(\boldsymbol{p}; t + \Delta t). \tag{1.57}$$

Assuming that motion $\boldsymbol{\Delta p}$ is along the spatial gradient direction, the equation becomes:

$$|\boldsymbol{\nabla} I(\boldsymbol{p}; t)| \cdot |\boldsymbol{\Delta p}| = I(\boldsymbol{p}; t) - I(\boldsymbol{p}; t + \Delta t), \tag{1.58}$$

$$|\boldsymbol{\Delta p}| = \frac{I(\boldsymbol{p}; t) - I(\boldsymbol{p}; t + \Delta t)}{|\boldsymbol{\nabla} I(\boldsymbol{p}; t)|}. \tag{1.59}$$

Assuming small motion in the whole video, the spatial gradient is only computed within the first frame, which provides its amplitude $|\boldsymbol{\nabla} I[\boldsymbol{p}; 0]|$ and direction $\angle_{\boldsymbol{\nabla} I[\boldsymbol{p};0]}$. Furthermore, instead of computing the displacement in frame k relative to the previous frame $(k-1)$, it can be computed relatively to frame 0. The motion at pixel $\boldsymbol{p}$ in frame k is thus computed from Eq. (1.59) as:

$$\boldsymbol{\delta}[\boldsymbol{p}; k] = \frac{I[\boldsymbol{p}; 0] - I[\boldsymbol{p}; k]}{|\boldsymbol{\nabla} I[\boldsymbol{p}; 0]|} \begin{pmatrix} \cos(\angle_{\boldsymbol{\nabla} I[\boldsymbol{p};0]}) \\ \sin(\angle_{\boldsymbol{\nabla} I[\boldsymbol{p};0]}) \end{pmatrix}. \tag{1.60}$$

1.5.4 Phase-based subpixel motion estimation

Phase-based methods are optical flow methods that operate on the phase φ of a complex filter S. For this purpose, the scene brightness is supposed to be constant, and the intensity I of a given surface element is assumed to be constant through time and space. Furthermore, the intensity is assumed to be a sparse local combination of cosines. To illustrate the principle of the method, Fig. 1.14 (a) shows the intensity level of a pixel row that follows the function $I(x,t) = 100 + 100 \cdot \cos(0.04 \cdot 2\pi \cdot (x - 2 \cdot t))$. Frames 1 ($t = 1$) and 4 ($t = 4$) are translated versions of frame 0 ($t = 0$) with a global displacement to the right of 2 and 8 px respectively. This global displacement is encoded by the frame Fourier transform phase. In this example, the signal is composed of a single frequency 0.04 px^{-1}. The frame Fourier transform amplitude is equal to 0 at every frequency except 0 px^{-1}, where its value is the signal mean, and 0.04 px^{-1} (see Fig. 1.14 (b)). As Fourier transform amplitudes are the same for the two frames, their analysis cannot determine the displacement. However, the phases are different and only meaningful at the signal frequency (see Fig. 1.14 (c)). Furthermore, the phase shifts between frames 0 and 1 and frames 0 and 4 are proportional to the displacements in frames 1 and 4. Thus, the displacement is encoded by the phase shift at this frequency. Moreover, reconstructing a new frame with amplified phase shift allows for small motion amplification.

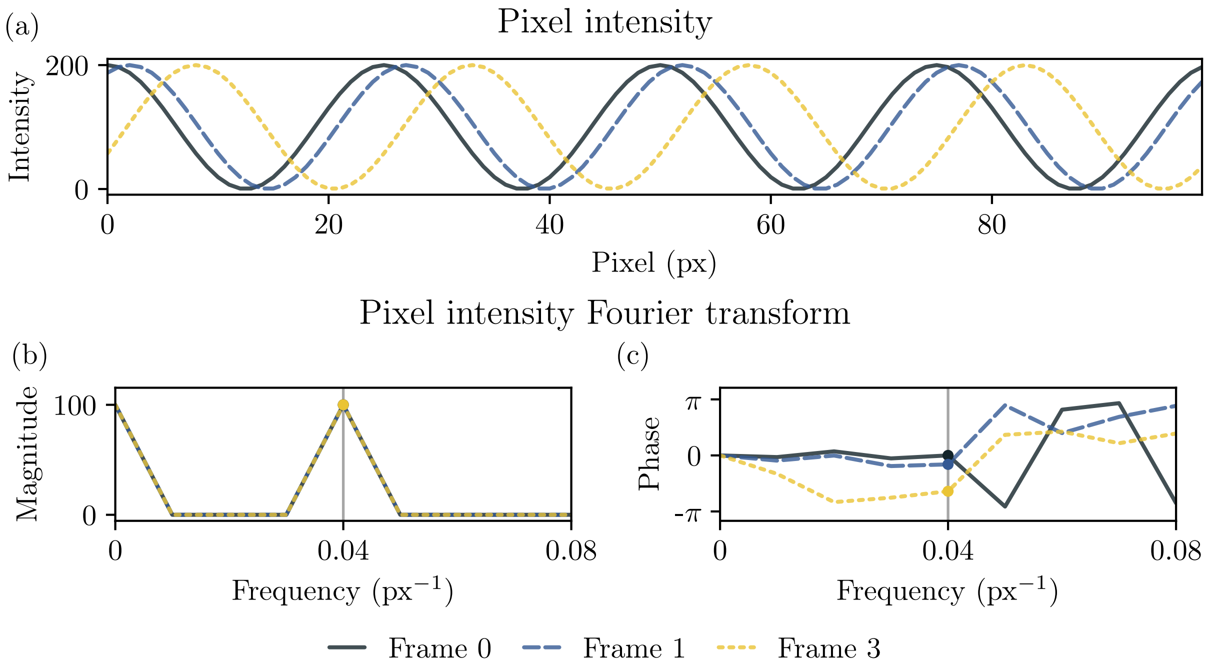

Figure 1.14: Pixel intensity phase as a motion encoder, on a single-row image.

This simple example exposes the relationship between motion and phase. For complex signals composed of different frequencies with local motion, local spatial frequency should be analyzed. For this purpose, each frame is decomposed into different oriented frequency bands, and local phase shift is computed to estimate the displacement.

Several state-of-the-art methods, that will be detailed in chapter 2, analyze the spatial local frequency. These methods amplify small motions in the video by raising the phase shift between the current frame and the unwrapped phase temporal mean [49]. ODS can thus be directly displayed in the video by temporal frequency band-pass filtering. This method of motion estimation is the ground of several commercial software products that perform video-based modal analysis.

1.5.5 Video-based industrial products

Different commercial solutions exist to achieve video-based modal analysis. They are generally composed of an acquisition system and a software that amplifies motion in videos at certain frequencies to display ODS. As far as we know, the commercial offer is composed of the products detailed in Tab 1.1.

Company	Product	Motion estimation
RDI Technologies[4]	*RDI Acquisition*	Phase-based optical flow
Mechanical Solutions, Inc.[5]	*VibVue*	Phase-based optical flow
Erbessd Instruments[6]	*DragonVision*	Image matching
MotionScope[7]	*MotionScope Tracker*	Intensity-based optical flow
Vibrant Technology[8]	*MEscope ODS Videos*	Intensity-based optical flow
Gfai tech GmbH[9]	*WaveCam*	Intensity-based optical flow

Table 1.1: Commercial video-based vibration analysis solutions.

In addition to standard digital cameras, event cameras have been used to perform vibration analysis [10]. Event cameras are composed of pixels that are independently activated when a time variation of brightness is perceived. This sparse information provides the ability to study features only where motion occurs, with an equivalent frame rate above 10 kHz.

Such methods are out of the scope of this work that focuses on classical, and widespread digital cameras.

---

[4] https://rditechnologies.com/

[5] https://www.mechsol.com/products/vibvue

[6] https://www.erbessd-instruments.com/dragon-vision/

[7] https://www.motion-scope.com/visual-modal-analysis

[8] https://www.vibetech.com/

[9] https://www.gfaitech.com/products/structural-dynamics/vibration-analysis-with-wavecam

1.6 Conclusion

In this chapter, we explain how the dynamics of complex mechanical systems can be expressed in temporal and frequency domains. The roots of modal basis estimation are developed. The discrete state-space formulation is expressed to link the system modal basis with motion measurements, which can be displacement, velocity, and/or acceleration. The different types of contact and remote motion sensors are detailed with their respective drawbacks. For dense remote motion measurements, video-based analysis achieves a trade-off between frame rate and spatial definition. The main video-based motion measurement approaches are detailed, and, as vibration produces pseudo-periodic small motion whose amplitude is lower than the spatial pixel resolution, the phase-based subpixel methods are introduced.

In chapter 2, we provide details about the state-of-the-art phase-based methods, and propose two ways to improve such methods. Performances are then compared on synthetic videos of a cantilever beam.

2 Phase-based subpixel motion estimation

2.1 Introduction

To determine motion from a video, phase-based motion estimation (PME) methods are optical flow ones that analyze the phase and amplitude of complex filter responses. Such methods provide a dense subpixel motion estimation. This chapter is dedicated to the different PME methods.

First, the frame representation model is developed in Sec. 2.2. Secondly, different decomposition strategies of spatial frequency are introduced in Sec. 2.3. Then, state-of-the-art single-subband PME methods are detailed in Sec. 2.4, and a new method is proposed. Existing multi-subband approaches are presented in Sec. 2.5, and a new method is derived to combine existing strategies by the fusion of several subbands.

To compare the performances of the studied methods, synthetic videos of a cantilever beam are generated, as detailed in Sec. 2.6. Finally, Sec. 2.7 provides a comparison of the performances reached by the methods for different motion amplitudes and additive gray level noises.

2.2 Frame representation

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

$$\boldsymbol{\delta}[\boldsymbol{p}; k] = \begin{pmatrix} \delta^h[\boldsymbol{p}; k] \\ \delta^v[\boldsymbol{p}; k] \end{pmatrix} \in \mathbb{R}^2. \tag{2.1}$$

The link between the global motion of a sinusoidal intensity and its discrete Fourier transform (DFT) phase has been illustrated in Fig. 1.14. For a single sinusoidal component, such motion is only encoded by the phase at the signal spatial frequency. However, real video frames are two-dimensional signals composed of several components with local motions. To analyze the global motion, 2D DFT phase at relevant spatial frequencies could be used. Yet, industrial applications require identifying different vibrations in local frame areas. To estimate local motion, a spectral analysis is performed within the neighborhood of each pixel.

For this purpose, the scene is supposed to be under uniform and constant lighting without occlusion. Therefore, the intensity at pixel $\boldsymbol{p}$ associated to a specific surface element that moves according to a displacement $\boldsymbol{\delta}[\boldsymbol{p}; k]$, is assumed to be constant:

$$I[\boldsymbol{p}; 0] = I[\boldsymbol{p} + \boldsymbol{\delta}[\boldsymbol{p}; k]; k]. \tag{2.2}$$

To estimate the local displacement field $\boldsymbol{\delta}$, each frame is decomposed into a spatial-frequency representation. The spatial frequency domain is divided into subbands centered at central frequency $\boldsymbol{\omega}_{r,\theta} = (\omega^h_{r,\theta}, \omega^v_{r,\theta})^\intercal = (\omega_r \cos(\theta), \omega_r \sin(\theta))^\intercal$. These subbands correspond to different scales $r = 1, \dots, N_r$ and orientations $\theta = 0, \dots, (N_\theta - 1)\pi/N_\theta$, where N_r and N_θ are the numbers of scales and orientations. An example of such decomposition is represented in Fig. 2.1.

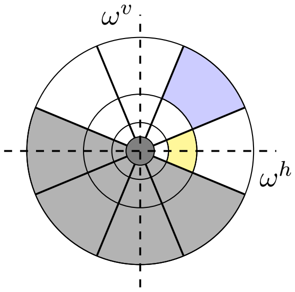

Figure 2.1: Ideal spatial frequency plane decomposition, with $N_r = 3$ scales and $N_\theta = 4$ orientations. The yellow and blue parts respectively correspond to the complex filters $\hat{G}_{2,0}$ and $\hat{G}_{1,\frac{\pi}{4}}$ amplitude supports. The bottom light gray part is redundant with the upper part and is thus not processed. The dark gray part is the low-frequency residual.

Each subband (r, θ) is characterized by a complex response $S_{r,\theta}$, that is computed using either a spatial filter $G_{r,\theta}$ or its equivalent $\hat{G}_{r,\theta}$ in the frequency domain. Such a filter is a band-pass filter characterized by a central frequency $\boldsymbol{\omega}_{r,\theta}$ and a standard deviation $\sigma_{r,\theta}$.

Additionally, it is assumed that the intensity $I[\boldsymbol{p}; k]$ at pixel $\boldsymbol{p}$ in frame k is a combination of local cosines with sparse local spatial frequencies $\tilde{\boldsymbol{\omega}}_{r,\theta}[\boldsymbol{p}]$:

$$I[\boldsymbol{p} + \boldsymbol{\delta}[\boldsymbol{p}; k]; k] = \sum_{r,\theta} \rho_{\tilde{\boldsymbol{\omega}}_{r,\theta}}[\boldsymbol{p}; k] \cdot \cos\left(\tilde{\boldsymbol{\omega}}_{r,\theta}[\boldsymbol{p}]^\intercal \cdot (\boldsymbol{p} - \boldsymbol{\delta}[\boldsymbol{p}; k])\right), \tag{2.3}$$

where $\tilde{\boldsymbol{\omega}}_{r,\theta}[\boldsymbol{p}]$ is in the subband centered at $\boldsymbol{\omega}_{r,\theta}$, and $\rho_{\tilde{\boldsymbol{\omega}}_{r,\theta}}[\boldsymbol{p}; k]$ is the amplitude of the corresponding cosine.

Hence, the complex response $S_{r,\theta} = G_{r,\theta} * I$ whose magnitude at $\boldsymbol{p}$ in frame k is $\rho_{r,\theta}[\boldsymbol{p}; k] = |S_{r,\theta}[\boldsymbol{p}; k]|$ and phase is $\varphi_{r,\theta}[\boldsymbol{p}; k] = \arctan(\mathfrak{Im}(S_{r,\theta}[\boldsymbol{p}; k])/\mathfrak{Re}(S_{r,\theta}[\boldsymbol{p}; k]))$, is expressed by:

$$\begin{aligned} S_{r,\theta}[\boldsymbol{p}; k] &= \rho_{r,\theta}[\boldsymbol{p}; k] \cdot \exp(j \cdot \varphi_{r,\theta}[\boldsymbol{p}; k]) && (2.4) \\ &= \rho_{r,\theta}[\boldsymbol{p}; k] \cdot \exp(j \cdot (\tilde{\boldsymbol{\omega}}_{r,\theta}[\boldsymbol{p}]^\intercal \cdot (\boldsymbol{p} - \boldsymbol{\delta}[\boldsymbol{p}; k]))), && (2.5) \end{aligned}$$

with $\rho_{r,\theta}[\boldsymbol{p}; k] \approx \rho_{\tilde{\boldsymbol{\omega}}_{r,\theta}}[\boldsymbol{p}; k]$ and $|\tilde{\boldsymbol{\omega}}_{r,\theta}[\boldsymbol{p}] - \boldsymbol{\omega}_{r,\theta}[p]| < \sigma_{r,\theta}$. The complex response phase $\varphi_{r,\theta}$ is thus:

$$\varphi_{r,\theta}[\boldsymbol{p}; k] = \tilde{\boldsymbol{\omega}}_{r,\theta}[\boldsymbol{p}]^\intercal \cdot (\boldsymbol{p} - \boldsymbol{\delta}[\boldsymbol{p}; k]). \tag{2.6}$$

Assuming that there is no motion in the first frame (*i.e.*, $\boldsymbol{\delta}[\boldsymbol{p}; 0] = 0$ whatever $\boldsymbol{p}$), Eq. (2.6) yields the following for each subband (r, θ), pixel $\boldsymbol{p}$, and frame k:

$$\varphi_{r,\theta}[\boldsymbol{p}; 0] - \varphi_{r,\theta}[\boldsymbol{p}; k] = \tilde{\boldsymbol{\omega}}_{r,\theta}[\boldsymbol{p}]^\intercal \cdot \boldsymbol{\delta}[\boldsymbol{p}; k]. \tag{2.7}$$

This fundamental equation links the displacement $\boldsymbol{\delta}$ to the local spatial frequency $\tilde{\boldsymbol{\omega}}_{r,\theta}$ and the phase $\varphi_{r,\theta}$.

The phase at each pixel in each frame is computed from the complex response $S_{r,\theta}$. As phase is wrapped within $(-\pi, \pi]$, it is first temporally unwrapped, such that any successive phase difference is never greater than π, to avoid that a phase jump gets interpreted as a displacement in the wrong direction. Therefore, Eq. (2.7) is rewritten:

$$\varphi_{r,\theta}[\boldsymbol{p};0] - \text{unwrap}(\varphi_{r,\theta}[\boldsymbol{p};k]) = \tilde{\boldsymbol{\omega}}_{r,\theta}[\boldsymbol{p}]^\intercal \cdot \boldsymbol{\delta}[\boldsymbol{p};k]. \tag{2.8}$$

As displacement and local frequency are assumed to be locally constant, their spatial gradients are equal to 0:

$$\left|\boldsymbol{\nabla}\delta^h[\boldsymbol{p};k]\right| = \left|\boldsymbol{\nabla}\delta^v[\boldsymbol{p};k]\right| = \left|\boldsymbol{\nabla}\tilde{\omega}^h_{r,\theta}[\boldsymbol{p}]\right| = \left|\boldsymbol{\nabla}\tilde{\omega}^v_{r,\theta}[\boldsymbol{p}]\right| = 0. \tag{2.9}$$

By deriving Eq. (2.6) with respect to $\boldsymbol{p}$, local spatial frequency is thus equal to the phase spatial gradient:

$$\begin{aligned}\boldsymbol{\nabla}\varphi_{r,\theta}[\boldsymbol{p};k] &= (x - \delta^h[\boldsymbol{p};k]) \cdot \boldsymbol{\nabla}\tilde{\omega}^h_{r,\theta}[\boldsymbol{p}] + (y - \delta^v[\boldsymbol{p};k]) \cdot \boldsymbol{\nabla}\tilde{\omega}^v_{r,\theta}[\boldsymbol{p}] \\ &\quad + \tilde{\boldsymbol{\omega}}_{r,\theta}[\boldsymbol{p}] - \tilde{\omega}^h_{r,\theta}[\boldsymbol{p}] \cdot \boldsymbol{\nabla}\delta^h[\boldsymbol{p};k] - \tilde{\omega}^v_{r,\theta}[\boldsymbol{p}] \cdot \boldsymbol{\nabla}\delta^v[\boldsymbol{p};k] \qquad (2.10) \\ &\overset{(2.9)}{=} \tilde{\boldsymbol{\omega}}_{r,\theta}[\boldsymbol{p}]. \qquad (2.11)\end{aligned}$$

Hence, from Eq. (2.8) and Eq. (2.11), motion $\boldsymbol{\delta}[\boldsymbol{p};k]$ is linked to $\varphi_{r,\theta}$ according to:

$$\varphi_{r,\theta}[\boldsymbol{p};0] - \text{unwrap}(\varphi_{r,\theta}[\boldsymbol{p};k]) = \boldsymbol{\nabla}\varphi_{r,\theta}[\boldsymbol{p};k]^\intercal \cdot \boldsymbol{\delta}[\boldsymbol{p};k]. \tag{2.12}$$

To avoid phase wrapping and discontinuity problems, the phase spatial gradient can be computed from Eq. (2.4) using Fleet and Jepson's formulation [12]:

$$\begin{aligned}\varphi_{r,\theta}[\boldsymbol{p};k] &= \mathfrak{Im}(\log(S_{r,\theta}[\boldsymbol{p};k])), \qquad (2.13) \\ \boldsymbol{\nabla}\varphi_{r,\theta}[\boldsymbol{p};k] &= \boldsymbol{\nabla}\mathfrak{Im}(\log(S_{r,\theta}[\boldsymbol{p};k])) \qquad (2.14) \\ &= \mathfrak{Im}(\boldsymbol{\nabla}\log(S_{r,\theta}[\boldsymbol{p};k])) \qquad (2.15) \\ &= \mathfrak{Im}\left(\frac{\boldsymbol{\nabla}S_{r,\theta}[\boldsymbol{p};k]}{S_{r,\theta}[\boldsymbol{p};k]}\right) \qquad (2.16) \\ &= \frac{\mathfrak{Im}\left(S^*_{r,\theta}[\boldsymbol{p};k] \cdot \boldsymbol{\nabla}S_{r,\theta}[\boldsymbol{p};k]\right)}{\left|\rho_{r,\theta}[\boldsymbol{p};k]\right|^2}, \qquad (2.17)\end{aligned}$$

with $S^*_{r,\theta}$ the complex conjugate of the filter response.

2.3 Spatial frequency decomposition

In Eq. (2.12), the local motion $\boldsymbol{\delta}[\boldsymbol{p};k]$ only depends on the local phase $\varphi_{r,\theta}[\boldsymbol{p};k]$ computed from the spatial frequency decomposition. However, spatial frequency decomposition cannot reach high resolution in both space and frequency, a phenomenon known as the uncertainty relation. Indeed, on the one hand, when a local Fourier transform is computed with a large support, frequency resolution is high, but localization of these frequencies is poor. On the other hand, when a short support is used to compute the local Fourier transform, localization is good, but frequency resolution is poor.

2.3.1 Gabor filters

A one-dimensional Gabor filter is a quadrature complex filter composed of a sinusoid, with a given central frequency, weighted by a Gaussian window. Such a filter reaches the lower bound of the uncertainty relation. This lower bound can be formulated in terms of standard deviation of the power distribution of the signal [12]. As an example, for a 1D signal $g(x) \in L^2(\mathbb{R})$ such that $\|g\|_2 = 1$ and its Fourier transform $\hat{g}(f)$, the uncertainty relation states that $\sigma_x \cdot \sigma_f \geq \frac{1}{2}$ where $\sigma_x^2 = \frac{\int x^2 \cdot |g(x)|^2 \, dx}{\int |g(x)|^2 \, dx}$ and $\sigma_f^2 = \frac{\int f^2 \cdot |\hat{g}(f)|^2 \, df}{\int |\hat{g}(f)|^2 \, df}$.
For a two-dimensional Gabor filter, a direction θ is specified for the sinusoid and Gaussian window, and another Gaussian window is applied along the orthogonal direction. The Gabor filter expression is given by:

$$G_{r,\theta}(\boldsymbol{p};\boldsymbol{\omega}_r,\theta,\sigma,\gamma) = \exp\left(-\frac{x'^2 + \gamma^2 \cdot y'^2}{2\sigma^2}\right) \cdot \exp\left(j \cdot \boldsymbol{\omega}_r \cdot x'\right), \tag{2.18}$$

with $\boldsymbol{\omega}_r$ the central spatial frequency radius, θ the filter main direction, σ the standard deviation along θ, γ the aspect ratio between the main and orthogonal directions, and $(x', y') = \Big(\boldsymbol{p}^\mathsf{T} \cdot \big(\cos(\theta), \sin(\theta)\big)^\mathsf{T}, \boldsymbol{p}^\mathsf{T} \cdot \big(-\sin(\theta), \cos(\theta)\big)^\mathsf{T}\Big)$ the pixel coordinates rotated by angle θ. An example of 2D filters is provided in Fig. 2.2 with central frequency radius $\boldsymbol{\omega}_r = \dfrac{\pi}{2^r}$, standard deviation $\sigma = \dfrac{3}{\boldsymbol{\omega}_r}$, and aspect ratio $\gamma = 1$. The support of such a filter, with size $(6\sigma + 1) \times (6\sigma + 1)$, increases with respect to the scale r. As the standard deviation σ is inversely proportional to the central frequency, the convolution can thus be time-consuming for low frequencies.

2.3.2 Complex steerable pyramid

The complex steerable pyramid (CSP) is an architecture that performs a fast space–frequency decomposition [37, 43]. The frequency spectrum is divided into octave pieces. To do so, the frame is first split into two parts corresponding to high and low frequencies. The first scale subbands are calculated by convolving the low-frequency part with band-pass filters (*e.g.*, $G_{1,\theta}$

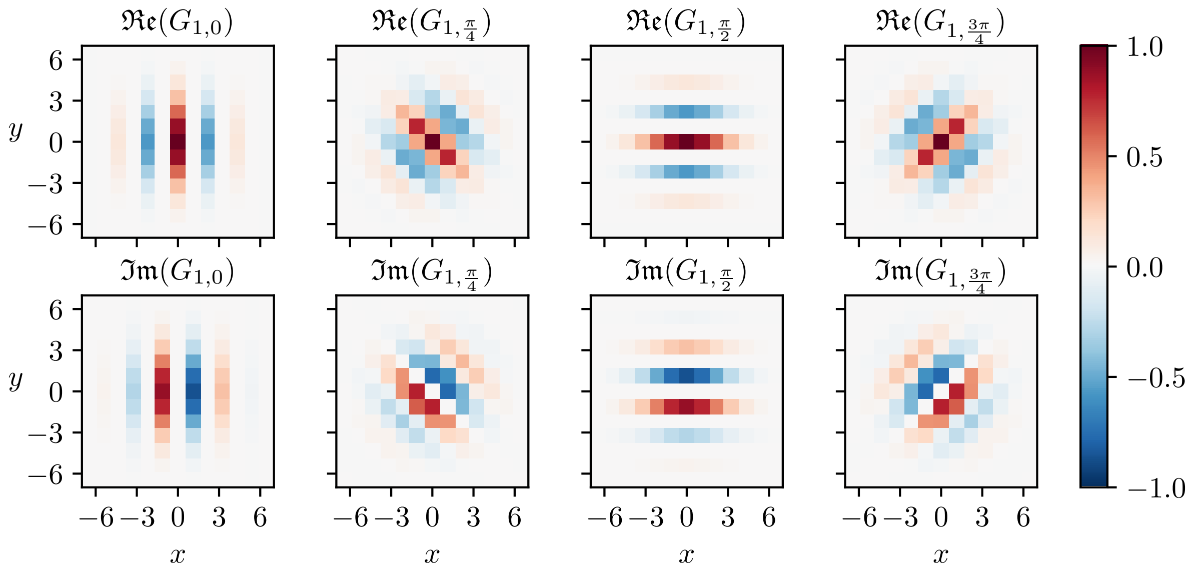

Figure 2.2: Gabor filters at scale $r = 1$, for $N_\theta = 4$ orientations.

in Fig. 2.2) in each direction θ. Then the low-frequency part is low-pass filtered another time and subsampled in each direction by a factor 2. The frame representation at next scale is computed using the same band-pass filters on the subsampled low-pass filtered frame. The last two operations are repeated recursively to represent the frame at subsequent scales. Finally, the last low-frequency part is also low-pass filtered and subsampled to get the low-pass residual. The initial frame can be reconstructed with low error by recursively upsampling the low-pass residual and summing with the band-passed subbands. A schematic of the iterative process is represented in Fig. 2.3, with H, B, and L being high-pass, band-pass, and low-pass filters respectively.

2.4 Single-subband motion estimation

On each subband (r, θ), motion can be estimated from $S_{r,\theta}$ using Eq. (2.12). In this section, we present different simple single-subband strategies that use a given scale r or subband (r, θ) to estimate the displacement field $\boldsymbol{\delta}$.

2.4.1 Given scale associated to local frequency (GSLF)

Chen et al. [6] assume that constant phase over time corresponds to the displacement field:

$$\varphi_{r,\theta}[\boldsymbol{p} + \boldsymbol{\delta}[\boldsymbol{p}; k]; k] = \varphi_{r,\theta}[\boldsymbol{p}; 0]. \tag{2.19}$$

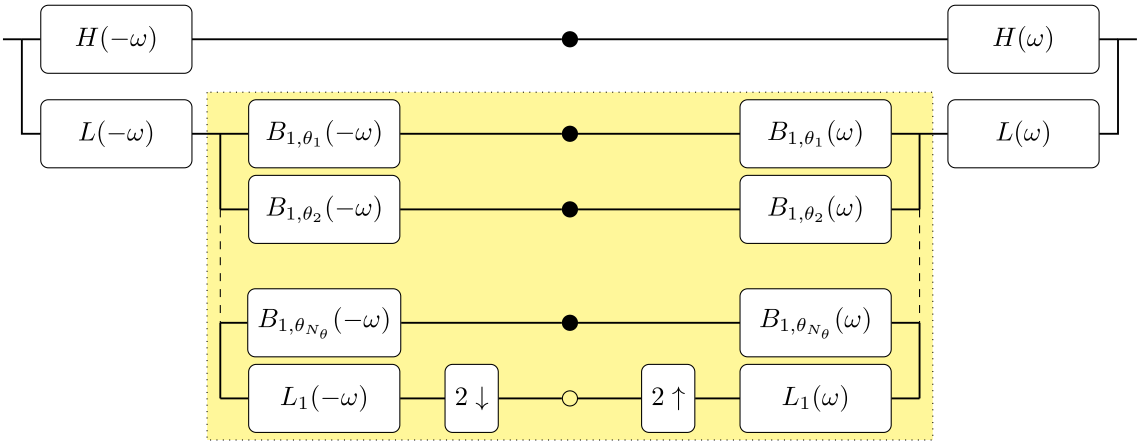

Figure 2.3: Complex steerable pyramid decomposition. The yellow rectangle is recursively applied to the bottom circle.

Furthermore, assuming small displacements in frame k, second and higher orders of the Taylor expansion can be omitted, which gives:

$$\varphi_{r,\theta}[\boldsymbol{p}+\boldsymbol{\delta}[\boldsymbol{p};k];k] = \varphi_{r,\theta}[\boldsymbol{p};k] + \boldsymbol{\nabla}\varphi_{r,\theta}[\boldsymbol{p};k]^\intercal \cdot \boldsymbol{\delta}[\boldsymbol{p};k]. \tag{2.20}$$

Plugging Eq. (2.20) into (2.19) yields:

$$\boldsymbol{\nabla}\varphi_{r,\theta}[\boldsymbol{p};k]^\intercal \cdot \boldsymbol{\delta}[\boldsymbol{p};k] = \varphi_{r,\theta}[\boldsymbol{p};0] - \varphi_{r,\theta}[\boldsymbol{p};k]. \tag{2.21}$$

Both hypotheses on constant intensity (see Eq. (2.2)) and phase (see Eq. (2.19)) lead to the same equation.

The phase is again temporally unwrapped to handle displacements larger than $\dfrac{\pi}{|\boldsymbol{\nabla}\varphi_{r,\theta}[\boldsymbol{p};k]|}$.

This method independently estimates horizontal and vertical components of motion. Indeed, when $\theta = 0$, the central spatial frequency of $G_{r,0}$ is horizontal with $\omega^v_{r,0} = 0$. Therefore, assuming that the local frequency is close to the central frequency, the phase vertical partial derivative $\dfrac{\partial\varphi_{r,0}}{\partial y}$ is negligible, so that the left-hand side of Eq. (2.21) only depends on δ^h and $\dfrac{\partial\varphi_{r,0}}{\partial x}$. A similar assumption holds for $\theta = \dfrac{\pi}{2}$, so that $\dfrac{\partial\varphi_{r,\frac{\pi}{2}}}{\partial x} \approx 0$. The following estimators are then obtained from Eq. (2.21):

$$\tilde{\delta}^h[\boldsymbol{p};k] = \left(\frac{\partial\varphi_{r,0}}{\partial x}[\boldsymbol{p};k]\right)^{-1} \cdot \left(\varphi_{r,0}[\boldsymbol{p};0] - \operatorname{unwrap}(\varphi_{r,0}[\boldsymbol{p};k])\right), \tag{2.22}$$

$$\tilde{\delta}^v[\boldsymbol{p};k] = \left(\frac{\partial\varphi_{r,\frac{\pi}{2}}}{\partial y}[\boldsymbol{p};k]\right)^{-1} \cdot \left(\varphi_{r,\frac{\pi}{2}}[\boldsymbol{p};0] - \operatorname{unwrap}(\varphi_{r,\frac{\pi}{2}}[\boldsymbol{p};k])\right). \tag{2.23}$$

Computing the 2D motion $\boldsymbol{\delta}$ from these equations requires only two subbands, computed at a single scale $\tilde{r}$ and two orientations ($\theta = 0$ and $\theta = \frac{\pi}{2}$). The scale is set arbitrarily (*e.g.*, $\tilde{r} = 4$), each frame being low-pass filtered and subsampled $\tilde{r} - 1$ times. A high scale focuses on frame content with low spatial frequency. The higher the scale is, the coarser the motion field estimation is because of the successive subsamplings. Chen et al. use the scale $\tilde{r} = 4$ to increase the signal-to-noise ratio [6]. The two orientation subbands are computed using separable quadrature band-pass complex filters (see Fig. 2.4) on the low-passed and subsampled frames.

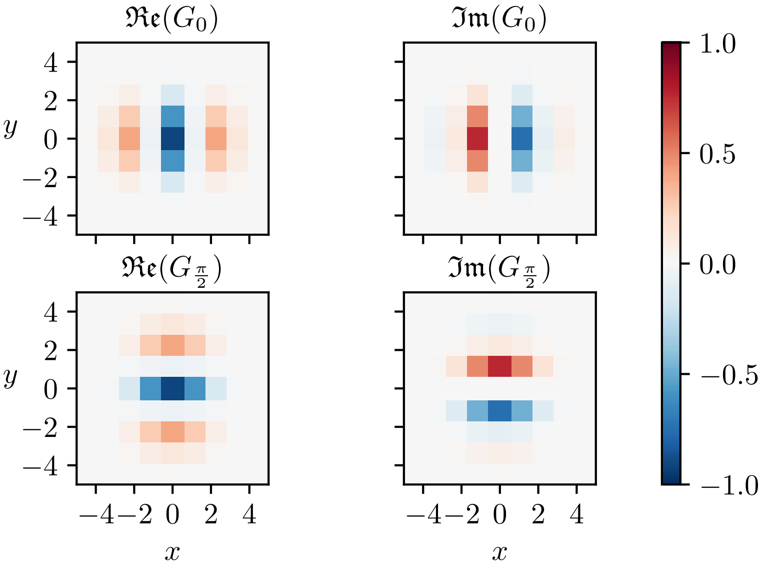

Figure 2.4: Spatial quadrature convolution filter pair.

Pixels where spatial derivative $\dfrac{\partial \varphi_{\tilde{r},0}}{\partial x}[\boldsymbol{p};k]$ or $\dfrac{\partial \varphi_{\tilde{r},\frac{\pi}{2}}}{\partial y}[\boldsymbol{p};k]$ is close to zero are removed to avoid irrelevant estimated motion [7].

In Eqs. (2.22) and (2.23), displacement $\tilde{\boldsymbol{\delta}}$ is estimated relatively to the first frame. To reduce the influence of the latter, displacement can be computed relatively to an equilibrium by removing the estimated displacement mean over all frames:

$$\hat{\delta}^h[\boldsymbol{p};k] = \tilde{\delta}^h[\boldsymbol{p};k] - \frac{1}{N_k}\sum_{i=0}^{N_k-1} \tilde{\delta}^h[\boldsymbol{p};i]. \tag{2.24}$$

Assuming constant phase partial derivative, $\frac{\partial \varphi_{\tilde{r},0}}{\partial x}[\boldsymbol{p};i] \approx \frac{\partial \varphi_{\tilde{r},0}}{\partial x}[\boldsymbol{p};k]$ for all $i \in [\![0, N_k - 1]\!]$, displacement is computed relatively to an equilibrium as:

$$\begin{aligned}\hat{\delta}^h[\boldsymbol{p};k] \overset{(2.22)}{=} & \left(\frac{\partial \varphi_{\tilde{r},0}}{\partial x}[\boldsymbol{p};k]\right)^{-1} \cdot \Big(\varphi_{\tilde{r},0}[\boldsymbol{p};0] - \operatorname{unwrap}(\varphi_{\tilde{r},0}[\boldsymbol{p};k]) \\ & - \frac{1}{N_k}\sum_{i=0}^{N_k-1}(\varphi_{\tilde{r},0}[\boldsymbol{p};0] - \operatorname{unwrap}(\varphi_{\tilde{r},0}[\boldsymbol{p};i]))\Big),\end{aligned} \tag{2.25}$$

$$= \left(\frac{\partial \varphi_{\tilde{r},0}}{\partial x}[\boldsymbol{p};k]\right)^{-1} \cdot \Bigg(\underbrace{\frac{1}{N_k}\sum_{i=0}^{N_k-1}\operatorname{unwrap}(\varphi_{\tilde{r},0}[\boldsymbol{p};i])}_{\bar{\varphi}_{\tilde{r},0}[\boldsymbol{p}]} - \operatorname{unwrap}(\varphi_{\tilde{r},0}[\boldsymbol{p};k])\Bigg), \tag{2.26}$$

with $\bar{\varphi}_{\tilde{r},0}[\boldsymbol{p}]$ the unwrapped phase temporal mean. Vertical displacement can be computed relatively to an equilibrium by following the same decomposition with $\frac{\partial \varphi_{\tilde{r},\frac{\pi}{2}}}{\partial y}[\boldsymbol{p};i]$.

2.4.2 Given scale associated to central frequency (GSCF)

Yang et al. [52] study the motion of a vertical cantilever beam and assume that vertical displacement $\delta^v[\boldsymbol{p};k]$ is negligible. The authors estimate horizontal displacement using horizontal response $S_{\tilde{r},0}$. To do so, the phase spatial gradient, which is equal to local frequency $\tilde{\boldsymbol{\omega}}_{\tilde{r},0}$ (see Eq. (2.11)), is approximated by the filter central frequency $\boldsymbol{\omega}_{\tilde{r},0}$ in Eq. (2.22). Indeed, filters $G_{\tilde{r},0}$ are spatial band-pass filters with central frequency $\boldsymbol{\omega}_{\tilde{r},0} = (\omega^h_{\tilde{r},0}, 0)^\intercal$, and local frequency given by Eqs. (2.11) and (2.17) is assumed to be in the filter frequency extent. Displacement is computed relatively to the estimated equilibrium as follows:

$$\hat{\delta}^h_{\tilde{r}}[\boldsymbol{p};k] = \frac{\bar{\varphi}_{\tilde{r},0}[\boldsymbol{p}] - \operatorname{unwrap}(\varphi_{\tilde{r},0}[\boldsymbol{p};k])}{\omega^h_{\tilde{r},0}}. \tag{2.27}$$

In Yang et al.'s work, vertical displacement is close to zero and thus not computed. However, it can be computed using the vertical filter response:

$$\hat{\delta}^v_{\tilde{r}}[\boldsymbol{p};k] = \frac{\bar{\varphi}_{\tilde{r},\frac{\pi}{2}}[\boldsymbol{p}] - \operatorname{unwrap}(\varphi_{\tilde{r},\frac{\pi}{2}}[\boldsymbol{p};k])}{\omega^v_{\tilde{r},\frac{\pi}{2}}}. \tag{2.28}$$

Although displacement can be estimated using Eq. (2.27), only phase differences are computed in the article. Indeed, OMA only needs relative displacement measurements.

2.4.3 Maximum amplitude subband (MAS)

To avoid that scale is given by the analyst (see Secs. 2.4.1 and 2.4.2), we propose a new method based on pixel-wise subband selection [31]. The first frame is decomposed into a space–frequency representation using Gabor filters. For each pixel $\boldsymbol{p}$, the retained subband $(r_{\boldsymbol{p}}, \theta_{\boldsymbol{p}})$ is that with maximum amplitude within the first frame subbands $\{S_{r,\theta}[\boldsymbol{p};0]\}_{r=0,\theta=0}^{N_r,(N_\theta-1)\cdot\pi/N_\theta}$:

$$(r_{\boldsymbol{p}}, \theta_{\boldsymbol{p}}) = \underset{r,\theta}{\operatorname{argmax}}\ \rho_{r,\theta}[\boldsymbol{p};0]. \tag{2.29}$$

Decomposing local frequency in polar coordinates $\tilde{\boldsymbol{\omega}}_{r_{\boldsymbol{p}},\theta_{\boldsymbol{p}}}[\boldsymbol{p}] = |\tilde{\boldsymbol{\omega}}_{r_{\boldsymbol{p}},\theta_{\boldsymbol{p}}}[\boldsymbol{p}]| \cdot \begin{pmatrix} \cos\left(\angle_{\tilde{\boldsymbol{\omega}}_{r_{\boldsymbol{p}},\theta_{\boldsymbol{p}}}[\boldsymbol{p}]}\right) \\ \sin\left(\angle_{\tilde{\boldsymbol{\omega}}_{r_{\boldsymbol{p}},\theta_{\boldsymbol{p}}}[\boldsymbol{p}]}\right) \end{pmatrix}$, and assuming that motion $\boldsymbol{\delta}[\boldsymbol{p};k]$ is collinear with $\tilde{\boldsymbol{\omega}}_{r_{\boldsymbol{p}},\theta_{\boldsymbol{p}}}[\boldsymbol{p}]$, *i.e.*, $\boldsymbol{\delta}[\boldsymbol{p};k] = |\boldsymbol{\delta}[\boldsymbol{p};k]| \cdot \begin{pmatrix} \cos\left(\angle_{\tilde{\boldsymbol{\omega}}_{r_{\boldsymbol{p}},\theta_{\boldsymbol{p}}}[\boldsymbol{p}]}\right) \\ \sin\left(\angle_{\tilde{\boldsymbol{\omega}}_{r_{\boldsymbol{p}},\theta_{\boldsymbol{p}}}[\boldsymbol{p}]}\right) \end{pmatrix}$, Eq. (2.8) becomes:

$$\varphi_{r_{\boldsymbol{p}},\theta_{\boldsymbol{p}}}[\boldsymbol{p};0] - \operatorname{unwrap}(\varphi_{r_{\boldsymbol{p}},\theta_{\boldsymbol{p}}}[\boldsymbol{p};k]) = |\tilde{\boldsymbol{\omega}}_{r_{\boldsymbol{p}},\theta_{\boldsymbol{p}}}[\boldsymbol{p}]| \cdot |\boldsymbol{\delta}[\boldsymbol{p};k]|. \tag{2.30}$$

Displacement can thus be estimated using the selected subband $(r_{\boldsymbol{p}}, \theta_{\boldsymbol{p}})$ either relatively to the first frame as:

$$\tilde{\boldsymbol{\delta}}[\boldsymbol{p};k] = \frac{\varphi_{r_{\boldsymbol{p}},\theta_{\boldsymbol{p}}}[\boldsymbol{p};0] - \operatorname{unwrap}(\varphi_{r_{\boldsymbol{p}},\theta_{\boldsymbol{p}}}[\boldsymbol{p};k])}{|\tilde{\boldsymbol{\omega}}_{r_{\boldsymbol{p}},\theta_{\boldsymbol{p}}}[\boldsymbol{p}]|} \cdot \begin{pmatrix} \cos\left(\angle_{\tilde{\boldsymbol{\omega}}_{r_{\boldsymbol{p}},\theta_{\boldsymbol{p}}}[\boldsymbol{p}]}\right) \\ \sin\left(\angle_{\tilde{\boldsymbol{\omega}}_{r_{\boldsymbol{p}},\theta_{\boldsymbol{p}}}[\boldsymbol{p}]}\right) \end{pmatrix}, \tag{2.31}$$

or to the equilibrium by replacing $\varphi_{r_{\boldsymbol{p}},\theta_{\boldsymbol{p}}}[\boldsymbol{p};0]$ in the above equation by $\bar{\varphi}_{r_{\boldsymbol{p}},\theta_{\boldsymbol{p}}}[\boldsymbol{p}]$.

At each pixel $\boldsymbol{p}$, the local spatial frequency within the selected subband $(r_{\boldsymbol{p}}, \theta_{\boldsymbol{p}})$ must be determined in order to estimate motion according to Eq. (2.31). For this purpose, the local frequency $\tilde{\boldsymbol{\omega}}_{r,\theta}$ is assumed to be temporally constant and is computed with Eqs. (2.11) and (2.17) using the first frame or a temporal mean.

2.5 Multi-subband motion estimation

In opposition to single-subband strategies, multi-subband methods estimate motion using all subbands deduced from the space–frequency decomposition. These methods consider the phase associated with each subband as an estimator. However, phase is inconsistent for subbands with low energy. The following methods present different strategies to fuse relevant subband estimators.

2.5.1 Weighted least-square (WLS)

The first method, introduced by Wadhwa et al. [48], uses a weighted least square estimator to merge subband phases, where each subband phase $\varphi_{r,\theta}$ is weighted by its squared amplitude $\rho^2_{r,\theta}$ and a local spatial Gaussian weight $\mathcal{G}$:

$$\begin{aligned}\hat{\boldsymbol{\delta}}[x,y;k] = \underset{\boldsymbol{\delta}[x,y;k]}{\arg\min} \sum_r \sum_\theta \sum_{c=-9}^{9} \sum_{\ell=-9}^{9} \mathcal{G}[c,\ell] \cdot \rho^2_{r,\theta}[x+c,y+\ell;k] \\ \cdot \Big[\boldsymbol{\omega}^{\intercal}_{r,\theta} \cdot \boldsymbol{\delta}[x,y;k] - \big(\varphi_{r,\theta}[x+c,y+\ell;0] - \mathrm{unwrap}(\varphi_{r,\theta}[x+c,y+\ell;k])\big)\Big]^2.\end{aligned} \tag{2.32}$$

In this formulation, the local spatial frequency $\tilde{\boldsymbol{\omega}}_{r,\theta}$ is approximated by the central spatial frequency $\boldsymbol{\omega}_{r,\theta}$.

The squared amplitude weight reduces the influence of inconsistent phases. Indeed, the phase is meaningful only when the local spatial frequency is a component of the signal. For instance, the phase in the 1D global motion example in Fig. 1.14 is only meaningful at frequency $0.04\,\mathrm{px}^{-1}$, where the DFT amplitude is high.

The local spatial Gaussian weight $\mathcal{G}$ considers the phase values in a local neighborhood of $\boldsymbol{p} = (x,y)^{\intercal}$, thus smoothes the estimator spatially. Its support is defined by a standard deviation σ, which is a parameter of the method. In their work, Wadhwa et al. set $\sigma = 3$ and compute phases from a CSP with $N_r = 2$ scales and $N_\theta = 4$ orientations.

To solve the weighted least square problem of Eq. (2.32), the following system is constructed for each pixel $(x,y)^{\intercal}$:

$$\boldsymbol{X} \cdot \delta[x,y;k] = \boldsymbol{Y}, \tag{2.33}$$

where

$$\boldsymbol{X} = \begin{pmatrix} \sum\limits_{r,\theta,c,\ell} \left(\omega^h_{r,\theta}\right)^2 \cdot \mathcal{G}[c,\ell] \cdot \rho^2_{r,\theta}[x+c,y+\ell;k] & \sum\limits_{r,\theta,c,\ell} \omega^h_{r,\theta} \cdot \omega^v_{r,\theta} \cdot \mathcal{G}[c,\ell] \cdot \rho^2_{r,\theta}[x+c,y+\ell;k] \\ \sum\limits_{r,\theta,c,\ell} \omega^h_{r,\theta} \cdot \omega^v_{r,\theta} \cdot \mathcal{G}[c,\ell] \cdot \rho^2_{r,\theta}[x+c,y+\ell;k] & \sum\limits_{r,\theta,c,\ell} \left(\omega^v_{r,\theta}\right)^2 \cdot \mathcal{G}[c,\ell] \cdot \rho^2_{r,\theta}[x+c,y+\ell;k] \end{pmatrix} \tag{2.34}$$

and

$$\boldsymbol{Y} = \begin{pmatrix} \sum\limits_{r,\theta,c,\ell} \omega^h_{r,\theta} \cdot \mathcal{G}[c,\ell] \cdot \rho^2_{r,\theta}[x+c,y+\ell;k] \cdot \big(\varphi_{r,\theta}[x+c,y+\ell;0] - \varphi_{r,\theta}[x+c,y+\ell;k]\big) \\ \sum\limits_{r,\theta,c,\ell} \omega^v_{r,\theta} \cdot \mathcal{G}[c,\ell] \cdot \rho^2_{r,\theta}[x+c,y+\ell;k] \cdot \big(\varphi_{r,\theta}[x+c,y+\ell;0] - \varphi_{r,\theta}[x+c,y+\ell;k]\big) \end{pmatrix}. \tag{2.35}$$

2.5.2 Phase confidence based estimator (PC)

As phases of low-energy subbands can be inconsistent, large phase shifts in such subbands may significantly increase the estimation error, even with squared amplitude weight.

2.5.2.1 Phase instability of a one-dimensional signal

To remove these low-energy subbands, Zhou et al. [57] propose to compute a weighted mean estimator, with a weight composed of the squared amplitude and a mask based on a phase confidence indicator. The phase confidence indicator is derived from Jepson and Fleet's work on phase stability [22]. These authors notice that phase is unstable when amplitude is low and when local frequency is far from the central frequency of interest. To come to this conclusion, they study a one-dimensional signal and define two criteria to exclude the subband thanks to a given threshold τ_r:

$$|\tilde{\omega}_r - \omega_r| > \tau_r, \tag{2.36}$$

$$\left|\frac{d_x \rho_r[x]}{\rho_r[x]}\right| > \tau_r, \tag{2.37}$$

with ω_r the central frequency, $\tilde{\omega}_r$ the local frequency, $\rho_r[x]$ the amplitude, and $d_x \rho_r[x]$ its derivative. The first criterion thresholds the absolute difference between local and central spatial frequencies. The second one states that the response amplitude ρ_r must be high and spatially constant (*i.e.*, $d_x \rho_r \approx 0$). It is thus based on the ratio between the response amplitude spatial derivative and the response amplitude. The threshold is normalized using the filter spatial standard deviation for both criteria.

An example is provided in Fig. 2.5, the signal composed of several frequencies being represented in Fig. 2.5 (a). A subband is computed from a Gabor filter with central frequency $\omega_r = 0.314\,\mathrm{rad} \cdot \mathrm{px}^{-1}$ whose amplitude and temporally unwrapped phase are shown in Figs. 2.5 (b) and 2.5 (c), respectively. For the considered pixel coordinates, the amplitude drops to zero at four locations. The unwrapped phase increases linearly, except at the same four locations, where some jumps can be observed. The two criteria are displayed in Figs. 2.5 (d) and 2.5 (e), with a threshold $\tau_r = \frac{5}{\sigma_r}$ normalized by the filter spatial standard deviation σ_r. The two criteria are above the threshold for the subband with $\omega_r = 0.314\,\mathrm{rad} \cdot \mathrm{px}^{-1}$ over the four locations, which highlights subband phase instability. Finally, the signal phase is represented in the space–frequency plane. The two criteria bounds are represented and correspond to locations with unstable phase.

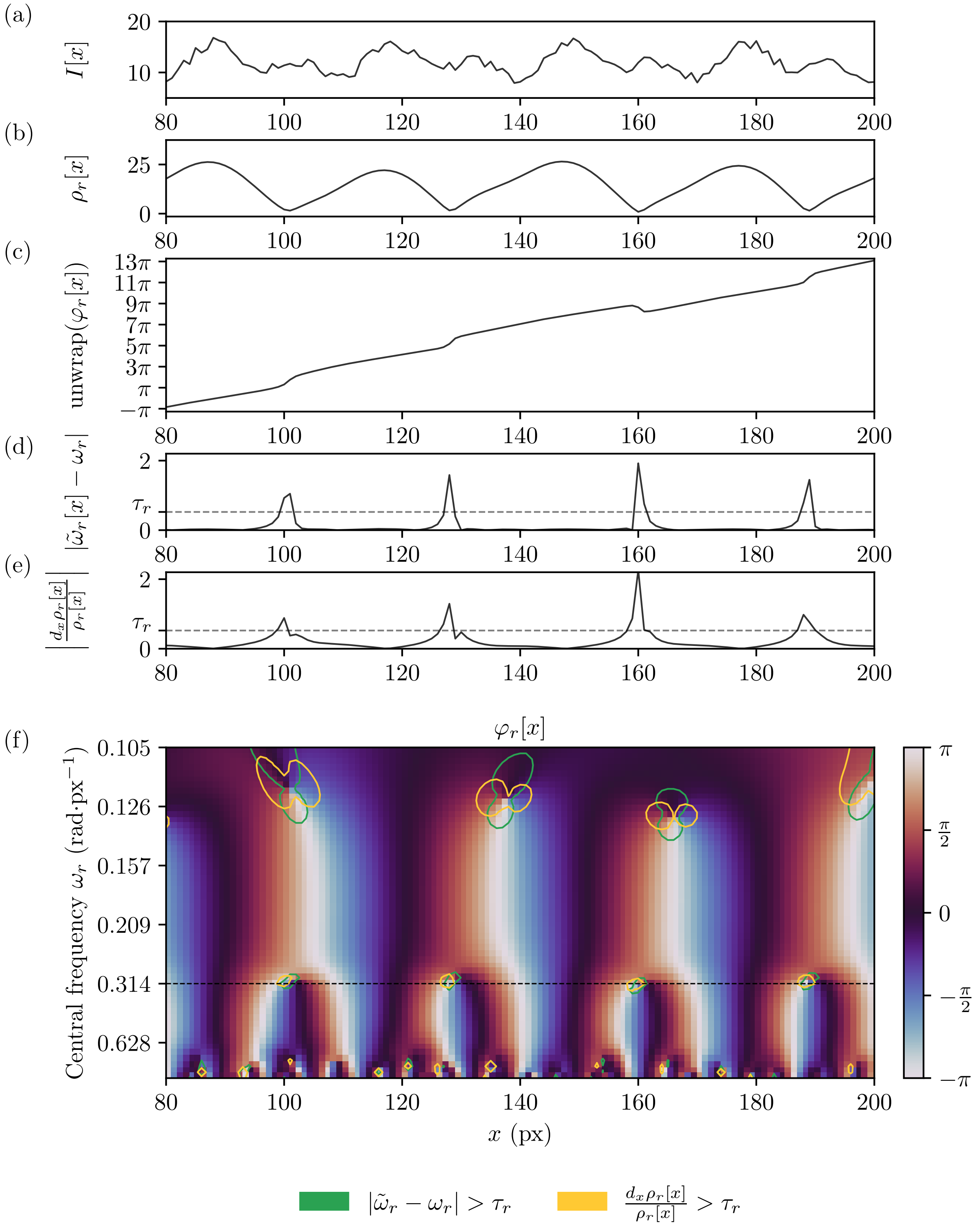

Figure 2.5: Example of phase instability, $\tau_r = \frac{5}{\sigma_r}$. (a) Intensity profile along a row, (b) amplitude at central frequency $\omega_r = 0.314\,\mathrm{rad} \cdot \mathrm{px}^{-1}$, (c) phase at central frequency ω_r, (d) criterion on local frequency at central frequency ω_r, (d) criterion on amplitude at central frequency ω_r, (e) phase frequency–space domain, with $\omega_r = 0.314\,\mathrm{rad} \cdot \mathrm{px}^{-1}$ as dashed line.

In their work, Zhou et al. focus on horizontal motion and use a single criterion with the same two key ideas as Jepson and Fleet about local frequency and amplitude ratio. The phase of horizontal subband is considered as unstable if:

$$\left| \frac{\partial_x \rho_{r,0}[\boldsymbol{p};k]}{\rho_{r,0}[\boldsymbol{p};k]} - \jmath \frac{|\tilde{\omega}^h_{r,0}[\boldsymbol{p};k] - \omega^h_{r,0}|}{\omega^h_{r,0}} \right| > \tau, \tag{2.38}$$

with $\tilde{\omega}^h_{r,0} = \partial_x \varphi_{r,0}[\boldsymbol{p};k]$ the local frequency. The threshold τ does not take the filter standard deviation $\sigma_{r,0}$ into account, but has the same given value for every scale. From this confidence measurement indicator, a binary mask $\mathcal{M}_r$ is defined for the given threshold τ:

$$\mathcal{M}^\tau_r[\boldsymbol{p};k] = \left\lfloor \left(1 + \exp\left[-\alpha\left(\tau - \left| \frac{\partial_x \rho_{r,0}[\boldsymbol{p};k]}{\rho_{r,0}[\boldsymbol{p};k]} - \jmath \frac{|\tilde{\omega}^h_{r,0}[\boldsymbol{p};k] - \omega^h_{r,0}|}{\omega^h_{r,0}} \right|\right)\right]\right)^{-1} \right\rceil, \tag{2.39}$$

with α the steepness of the logistic function. Note that parameter α can be discarded because the logistic function is rounded. This mask is efficient to remove phase singularities.

2.5.2.2 Confidence-based horizontal displacement estimator

The horizontal displacement is then estimated using a mean over scales, weighted by the squared response amplitude and the mask described previously:

$$\hat{\delta}^h[\boldsymbol{p};k] = \frac{\sum_r \rho^2_{r,0}[\boldsymbol{p};k] \cdot \mathcal{M}^\tau_r[\boldsymbol{p};k] \cdot \hat{\delta}^h_{r,0}[\boldsymbol{p};k]}{\sum_r \rho^2_{r,0}[\boldsymbol{p};k] \cdot \mathcal{M}^\tau_r[\boldsymbol{p};k]}, \tag{2.40}$$

the horizontal motion being estimated at each scale r as:

$$\hat{\delta}^h_{r,0}[\boldsymbol{p};k] = \frac{\varphi_{r,0}[\boldsymbol{p};0] - \mathrm{unwrap}(\varphi_{r,0}[\boldsymbol{p};k])}{\tilde{\omega}^h_{r,0}}. \tag{2.41}$$

2.5.2.3 Extension to 2D estimator

In this part, we propose an extension of Zhou et al.'s method to both horizontal and vertical directions. The same fusion strategy as for the sole horizontal direction can be applied to estimate 2D displacement as:

$$\hat{\boldsymbol{\delta}}[\boldsymbol{p};k] = \frac{\sum_{r,\theta} \rho^2_{r,\theta}[\boldsymbol{p};k] \cdot \mathcal{M}^\tau_{r,\theta}[\boldsymbol{p};k] \cdot \hat{\boldsymbol{\delta}}_{r,\theta}[\boldsymbol{p};k]}{\sum_{r,\theta} \rho^2_{r,\theta}[\boldsymbol{p};k] \cdot \mathcal{M}^\tau_{r,\theta}[\boldsymbol{p};k]}, \tag{2.42}$$

with $\hat{\boldsymbol{\delta}}_{r,\theta}$ the motion estimated on each subband (r,θ). For this purpose, we assume that motion is along the phase spatial gradient direction $\angle_{\boldsymbol{\nabla}\varphi_{r,\theta}[\boldsymbol{p};k]}$:

$$\hat{\boldsymbol{\delta}}_{r,\theta}[\boldsymbol{p};k] = \left|\hat{\boldsymbol{\delta}}_{r,\theta}[\boldsymbol{p};k]\right| \cdot \begin{pmatrix} \cos\left(\angle_{\boldsymbol{\nabla}\varphi_{r,\theta}[\boldsymbol{p};k]}\right) \\ \sin\left(\angle_{\boldsymbol{\nabla}\varphi_{r,\theta}[\boldsymbol{p};k]}\right) \end{pmatrix}, \tag{2.43}$$

with $\boldsymbol{\nabla}\varphi_{r,\theta}[\boldsymbol{p};k]$ estimated using Eq. (2.17). The amplitude can be then deduced from Eq. (2.12):

$$\left|\hat{\boldsymbol{\delta}}_{r,\theta}[\boldsymbol{p};k]\right| = \frac{\varphi_{r,\theta}[\boldsymbol{p};0] - \mathrm{unwrap}(\varphi_{r,\theta}[\boldsymbol{p};k])}{|\boldsymbol{\nabla}\varphi_{r,\theta}[\boldsymbol{p};k]|}. \tag{2.44}$$

Finally, we assume that phase gradient is approximately temporally constant so that:

$$\hat{\boldsymbol{\delta}}_{r,\theta}[\boldsymbol{p};k] = \frac{\varphi_{r,\theta}[\boldsymbol{p};0] - \mathrm{unwrap}(\varphi_{r,\theta}[\boldsymbol{p};k])}{|\boldsymbol{\nabla}\varphi_{r,\theta}[\boldsymbol{p};0]|} \cdot \begin{pmatrix} \cos\left(\angle_{\boldsymbol{\nabla}\varphi_{r,\theta}[\boldsymbol{p};0]}\right) \\ \sin\left(\angle_{\boldsymbol{\nabla}\varphi_{r,\theta}[\boldsymbol{p};0]}\right) \end{pmatrix}. \tag{2.45}$$

To extend the phase confidence mask to the two directions, we assume that the spatial gradient of the response amplitude $\rho_{r,\theta}$ must be small along the direction of the local frequency $\tilde{\boldsymbol{\omega}}_{r,\theta}$. The criterion on the phase linearity thus becomes:

$$\left|\frac{\boldsymbol{\nabla}\rho_{r,\theta}[\boldsymbol{p};k]^{\intercal} \cdot \tilde{\boldsymbol{\omega}}_{r,\theta}[\boldsymbol{p};k]}{\rho_{r,\theta}[\boldsymbol{p};k] \cdot |\tilde{\boldsymbol{\omega}}_{r,\theta}[\boldsymbol{p};k]|}\right| > \tau. \tag{2.46}$$

The criterion about local frequency is kept as it, such that the phase confidence mask $M^{\tau}_{r,\theta}$ in subband (r,θ) is defined as:

$$\mathcal{M}^{\tau}_{r,\theta}[\boldsymbol{p};k] = \left\lfloor \left(1 + \exp\left[-\left(\tau - \left|\frac{\boldsymbol{\nabla}\rho_{r,\theta}[\boldsymbol{p};k]^{\intercal} \cdot \tilde{\boldsymbol{\omega}}_{r,\theta}[\boldsymbol{p};k]}{\rho_{r,\theta}[\boldsymbol{p};k] \cdot |\tilde{\boldsymbol{\omega}}_{r,\theta}[\boldsymbol{p};k]|} - j\frac{|\tilde{\boldsymbol{\omega}}_{r,\theta}[\boldsymbol{p};k] - \boldsymbol{\omega}_{r,\theta}|}{|\boldsymbol{\omega}_{r,\theta}|}\right|\right)\right]\right)^{-1} \right\rceil. \tag{2.47}$$

2.5.3 Weighted spatial mean based on phase confidence (SMPC)

The estimator in Eq. (2.42) only accounts for the local phase at $\boldsymbol{p}$. However, motion of mechanical structures can be considered as spatially continuous. Therefore, we assume that the estimated motion should be spatially smooth. In order to account for the neighborhood of $\boldsymbol{p}$, we thus propose to introduce a spatial Gaussian kernel $\mathcal{G}$ with standard deviation σ and support $\mathcal{N}_\sigma$ in the estimator of Eq. (2.42):

$$\hat{\boldsymbol{\delta}}[\boldsymbol{p};k] = \frac{\sum\limits_{r,\theta}\sum\limits_{\boldsymbol{q}\in\mathcal{N}_\sigma} \mathcal{G}[\boldsymbol{q}] \cdot \rho^2_{r,\theta}[\boldsymbol{p}+\boldsymbol{q};k] \cdot \mathcal{M}^{\tau}_{r,\theta}[\boldsymbol{p}+\boldsymbol{q};k] \cdot \hat{\boldsymbol{\delta}}_{r,\theta}[\boldsymbol{p}+\boldsymbol{q};k]}{\sum\limits_{r,\theta}\sum\limits_{\boldsymbol{q}\in\mathcal{N}_\sigma} \mathcal{G}[\boldsymbol{q}] \cdot \rho^2_{r,\theta}[\boldsymbol{p}+\boldsymbol{q};k] \cdot \mathcal{M}^{\tau}_{r,\theta}[\boldsymbol{p}+\boldsymbol{q};k]}. \tag{2.48}$$

2.6 Synthetic videos of a cantilever beam

To test the performances of the motion estimation methods presented above, we generate synthetic videos of a cantilever beam. These videos are based on the Euler–Bernoulli beam equation [44], which models the deflection of a cantilever beam according to its physical properties and the force applied to it. This equation models the center line of a 1D beam. The borders of the beam are computed using the normal to the center line at each point. Then each frame is generated using a pinhole camera model.

2.6.1 Euler–Bernoulli cantilever beam model

The physical parameters of the beam are its length $L\,(\mathrm{m})$, Young modulus $E\,(\mathrm{Pa})$, cross-section area $A\,(\mathrm{m}^2)$, moment of inertia $J\,(\mathrm{m}^4)$, and volumic mass density $\rho\,(\mathrm{kg{\cdot}m^{-3}})$. First, the model is considered without force. The center line of the beam is defined by a function $g(z,t)$, $z \in [0,L]$, $t \geq 0$, as:

$$\frac{\partial^2}{\partial z^2}\left(EJ\frac{\partial^2 g}{\partial z^2}(z,t)\right) + \rho A\frac{\partial^2 g}{\partial t^2}(z,t) = 0, \tag{2.49}$$

$$\underbrace{\frac{EJ}{\rho A}}_{c^2}\frac{\partial^4 g}{\partial z^4}(z,t) + \frac{\partial^2 g}{\partial t^2}(z,t) = 0. \tag{2.50}$$

To solve this equation, the variables of the function g are separated by assuming that $g(z,t) = \phi(x)\cdot q(t)$, where ϕ encodes spatial information and q encodes temporal information. Multiple solutions ϕ_i exist whose forms are:

$$\phi_i(z) = \cosh(\beta_i z) - \cos(\beta_i z) + \frac{\sinh(\beta_i L) - \sinh(\beta_i L)}{\cos(\beta_i L) + \cosh(\beta_i L)}\cdot\left(-\sinh(\beta_i z) + \sin(\beta_i z)\right)\,, \tag{2.51}$$

with β_i solution of $\cos(\beta_i L)\cdot\cosh(\beta_i L) + 1 = 0$, which can be solved numerically. The first four mode shapes are represented in Fig. 2.6.

The modal contribution q_i associated to ϕ_i is solution of:

$$\frac{\partial^2 q_i(t)}{\partial t^2} + \omega_i^2 q_i(t) = \frac{1}{\rho A}\int_0^L \phi_i(z)\gamma(z,t)\,dz\,, \tag{2.52}$$

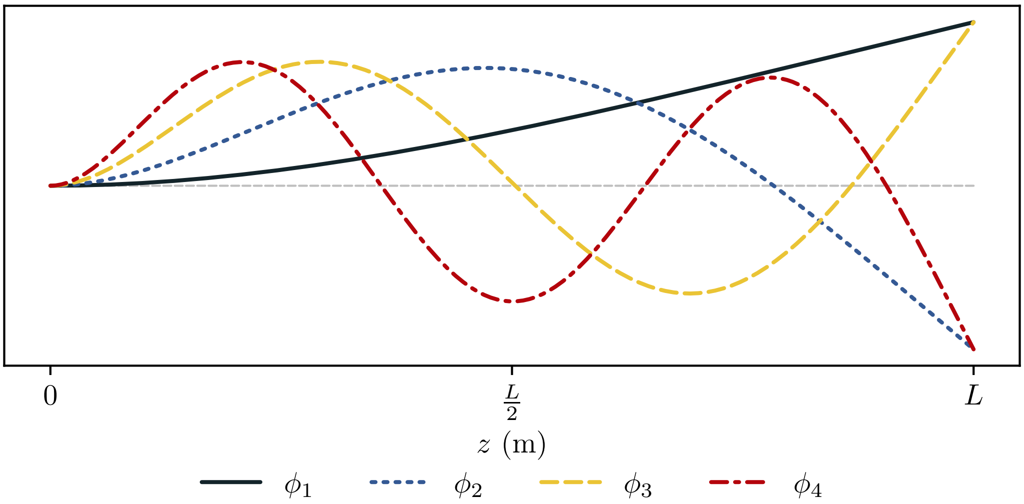

Figure 2.6: First four mode shapes of an Euler–Bernoulli cantilever beam clamped to the left end.

with $\omega_i = \beta_i^2 \cdot c$, and $\gamma(z,t)$ the input force applied to the beam. For a given input force, the central line coordinate is thus computed by summing each contribution:

$$g(z,t) = \sum_{i=1}^{\infty} \phi_i(z) \cdot q_i(t). \tag{2.53}$$

More details about how to solve Eq. (2.50) are provided in Appendix A.

2.6.2 Frame generation

To generate each frame, the central line position is computed using Euler–Bernoulli beam model. The beam borders are computed as a 2D polygon by taking the normal to the central line at each point. This polygon is then projected on the image plane using a simple pinhole camera model. To do so, we assume that the beam moves in a plane parallel to the image plane so that its size in the frame is only defined by the ratio between its distance to the camera lens and the focal length. Finally, each pixel value is computed as the proportion of its surface occupied by the beam, and quantized on $N_b = 8$ bits. An example of generated frame is provided in Fig. 2.7 for a beam with the following parameters: $L = 0.9\,\mathrm{m}$, $E = 210 \cdot 10^9\,\mathrm{Pa}$, $A = 1.8 \cdot 10^{-4}\,\mathrm{m}^2$, $J = 5.4 \cdot 10^{-10}\,\mathrm{m}^4$, and $\rho = 7850\,\mathrm{kg{\cdot}m^{-3}}$.

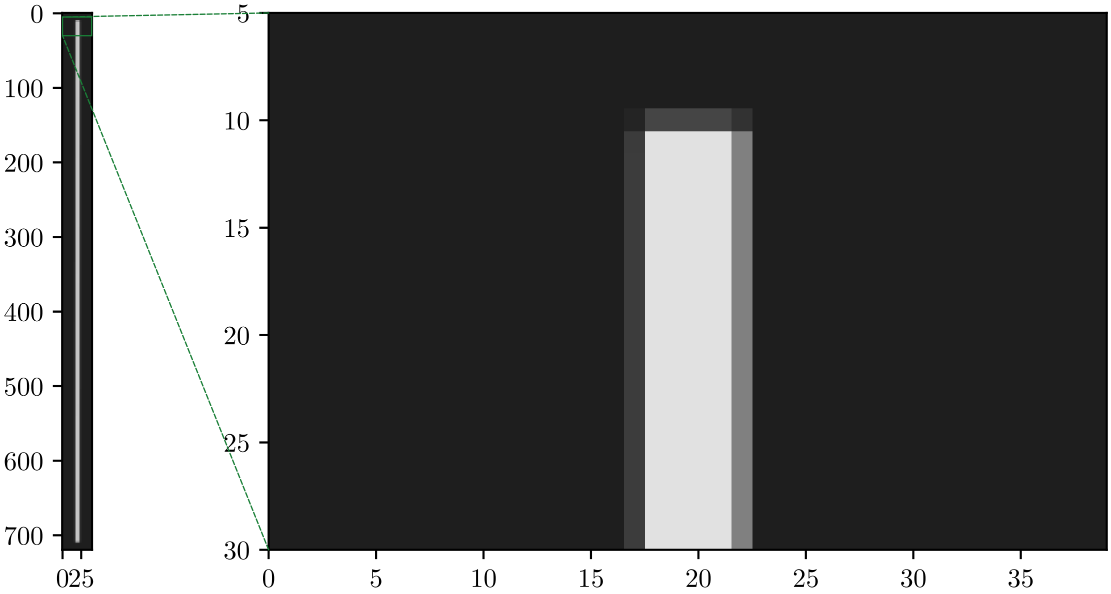

Figure 2.7: Frame of a synthetic video of a cantilever beam clamped at its bottom end: full frame (left) and top detail (right).

2.7 Estimated motion from synthetic videos

To compare the methods defined in Secs. 2.4 and 2.5, we use a synthetic video of a cantilever beam generated using the method described in Sec. 2.6. We simulate a hammer impact at the free end of the beam by setting the input force $\gamma(z,t)$ as a Dirac function in time and space:

$$\gamma(z,t) = \begin{cases} 0.64\,\mathrm{N} & \text{if } z = L \text{ and } t = 0, \\ 0\,\mathrm{N} & \text{otherwise.} \end{cases} \tag{2.54}$$

Only the first four modes are taken into account for the beam displacement, with the theoretical natural frequencies and damping ratios detailed in Tab. 2.1.

Mode m	1	2	3	4
f_m (Hz)	6.19	38.79	108.60	212.82
ζ_m ($\times 10^2$)	0.11	1.13	0.29	0.13

Table 2.1: Beam theoretical natural frequencies and damping ratios.

The video frame rate is set to 436 fps to be twice larger than the highest natural frequency (212 Hz), and thus avoid temporal frequency aliasing. The video lasts 1 s so that it is composed of 436 frames.

As the video is generated with a vertical cantilever beam, vertical displacement is close to zero. The video is rotated with an angle of 78.75° so that both vertical and horizontal displacements are studied. Each frame is thus a square of $1000 \times 1000\,\mathrm{px}$ where the beam clamped end is located in the bottom right corner. A region of interest is defined on a pixel line at the center of the beam, as represented in Fig. 2.8. Motion is estimated in each frame and the horizontal and vertical components of each pixel of interest are extracted to get $\hat{\boldsymbol{\delta}}[z;k]$, with z the curvilinear coordinate of pixels of interest along the beam in the frame. Note that the pixel at $z = 1$ represents the beam clamped end.

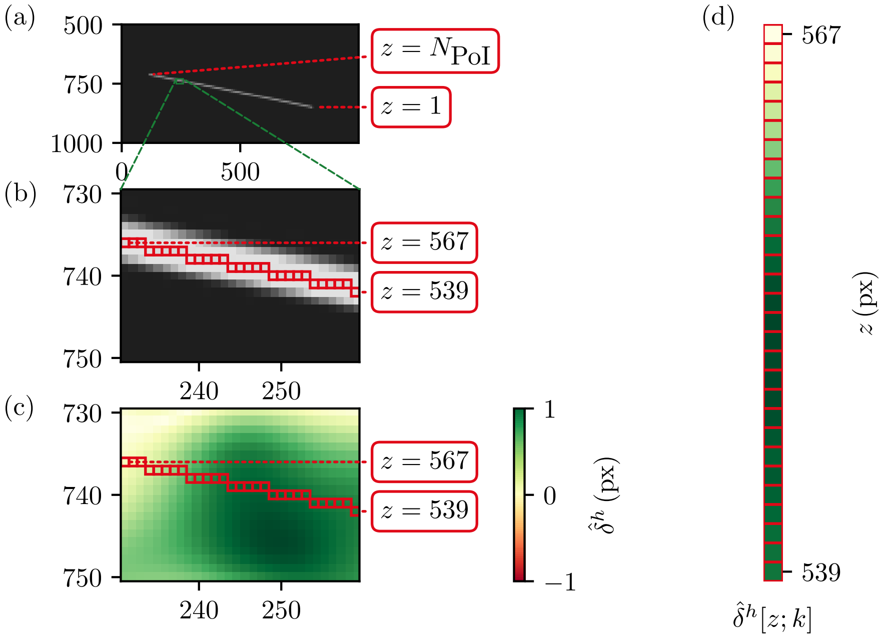

Figure 2.8: Motion extraction process of a cantilever beam. (a) Frame of a synthetic video. (b) Focus on pixels of interest in red with $z \in [\![539, 567]\!]$. (c) Estimated motion $\hat{\delta}^h[z;k]$ in the focus view. (d) Estimated motion $\hat{\delta}^h[z;k]$ at the pixels of interest.

We denote each of the six estimators as $\hat{\boldsymbol{\delta}}_i$, i corresponding to one of the following methods:

1. Given scale with local frequency (GSLF) [6] described in Sec. 2.4.1,
2. Given scale with central frequency (GSCF) [52] described in Sec. 2.4.2,
3. Our maximum amplitude subband (MAS) described in Sec. 2.4.3,
4. Weighted least-square (WLS) [48] described in Sec. 2.5.1,
5. Phase confidence (PC) [57] described in Sec. 2.5.2,

6. Our weighted spatial mean based on phase confidence (SMPC) described in Sec. 2.5.3.

To compare the performances reached by the methods, each frame is decomposed using the Gabor filters defined in Sec. 2.3.1 with $N_r = 5$ scales and $N_\theta = 8$ orientations. For GSLF ($\hat{\boldsymbol{\delta}}_1$) and GSCF ($\hat{\boldsymbol{\delta}}_2$) methods, the third scale is used. The standard deviation of the Gaussian kernel of WLS ($\hat{\boldsymbol{\delta}}_4$) and our SMPC ($\hat{\boldsymbol{\delta}}_6$) methods is set to 3 px. The phase confidence threshold τ is set to 0.5 for PC ($\hat{\boldsymbol{\delta}}_5$) and SMPC ($\hat{\boldsymbol{\delta}}_6$) estimators.

The methods are applied to the 436 frames of the video. The values obtained at the $N_{\mathrm{PoI}} = 686$ pixels of interest are extracted and compared with the theoretical displacement of the Euler–Bernoulli model. To do so, the theoretical displacements are converted from the scene coordinates to the frame coordinates to match the pixels of interest. The methods are assessed qualitatively by visually comparing the obtained estimations with the theoretical displacements. In Fig. 2.9, horizontal and vertical components of theoretical and estimated displacements at the middle of the beam are represented along the first 101 frames. Regarding vertical displacement, the six estimators seem to be consistent with the theoretical displacement. Their amplitudes are quite different, but the temporal frequency content seems to be correctly captured. $\hat{\delta}^v_2$ provides the best estimation of the motion amplitude. $\hat{\delta}^v_3$, $\hat{\delta}^v_5$, and $\hat{\delta}^v_6$ overestimate the displacement while $\hat{\delta}^v_1$ and $\hat{\delta}^v_4$ underestimate it. Regarding horizontal displacement, the single-subband estimators $\hat{\delta}^h_1$ and $\hat{\delta}^h_2$ are not consistent with the theoretical displacement. This is due to the fact that the frequency content of the horizontal subband is poor because the beam is almost horizontal. The other estimators seem to be consistent and slightly overestimate or underestimate the amplitude in the same way as for vertical displacement.

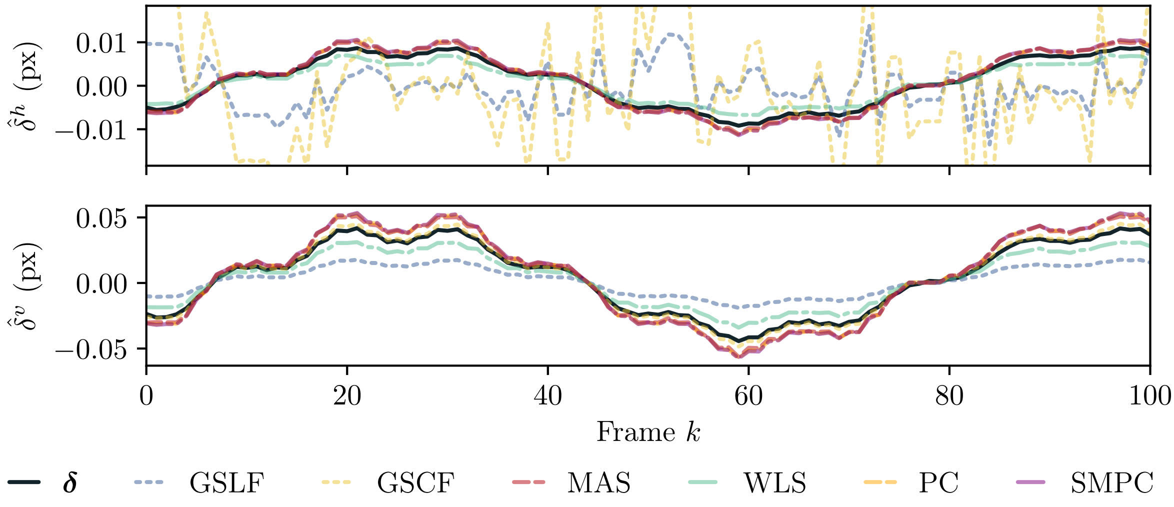

Figure 2.9: Comparison of theoretical and estimated displacements at the middle of the beam along the first 101 frames.

As this figure only analyzes the displacement estimations at one pixel, we propose to represent the relative absolute error (RAE) in the space–time domain (see Fig. 2.10). To do so, horizontal and vertical estimations are studied separately, and the RAE of a given estimator $\hat{\boldsymbol{\delta}}_i$ is computed relatively to the maximum theoretical displacement as:

$$\mathrm{RAE}^h(\hat{\boldsymbol{\delta}}_i)[z;k] = 100 \cdot \frac{\left|\hat{\delta}_i^h[z;k] - \delta^h[z;k]\right|}{\max\limits_{\tilde{z},\tilde{k}} |\delta^h[\tilde{z};\tilde{k}]|} \text{ and } \mathrm{RAE}^v(\hat{\boldsymbol{\delta}}_i)[z;k] = 100 \cdot \frac{\left|\hat{\delta}_i^v[z;k] - \delta^v[z;k]\right|}{\max\limits_{\tilde{z},\tilde{k}} |\delta^v[\tilde{z};\tilde{k}]|}, \tag{2.55}$$

with z the pixel of interest. The values above $100\,\%$ are set to 100 to avoid that large errors hide smaller ones.

Figure 2.10 shows the RAEs of the six estimators, revealing the same behaviors as for a single pixel (see Fig. 2.9). The two single-subband estimators $\hat{\boldsymbol{\delta}}_1$ and $\hat{\boldsymbol{\delta}}_2$ fail to estimate horizontal displacement, which confirms that these methods require an assumption about the main direction of the structure in the video. The four other methods provide relatively low errors, with vertical patterns that highlight amplitude error. At the free end of the beam, the pixelwise subband estimator $\hat{\boldsymbol{\delta}}_3$ has a higher error than multi-subband methods. This may be due to the fact that the highest frequency content at this location is horizontal, and that the hypothesis on the displacement in the direction of the local frequency is not correct in this region. Regarding horizontal displacement estimation, the weighted least-square method $\hat{\boldsymbol{\delta}}_4$ error seems to be less smooth than other methods as z varies. This suggests that errors are not only due to the amplitude.

To further compare the methods quantitatively, we compute the cosine similarity (CS) that measures the similarity between theoretical and estimated displacements:

$$\mathrm{CS}(\hat{\boldsymbol{\delta}}, \boldsymbol{\delta}) = \frac{\sum\limits_z \sum\limits_k (\hat{\delta}^h[z;k] \cdot \delta^h[z;k] + \hat{\delta}^v[z;k] \cdot \delta^v[z;k])}{\sqrt{\sum\limits_z \sum\limits_k (\hat{\delta}^h[z;k]^2 + \hat{\delta}^v[z;k]^2)} \sqrt{\sum\limits_z \sum\limits_k (\delta^h[z;k]^2 + \delta^v[z;k]^2)}}. \tag{2.56}$$

This measure ranges from -1 to 1 and does not account for the signal amplitudes (CS between two proportional signals is 1). This is adequate because for vibration analysis (and OMA more specifically), displacements have only to be estimated relatively to each others. This means that if a method overestimates or underestimates displacement in the same way throughout the video, the results will not be affected.

2.7.1 Sensitivity study against motion amplitude

As subpixel motion can be small in videos, we assess the methods with different motion amplitudes. To do so, we generate three other videos with input force $\gamma(L,0) \in \{0.08, 0.16, 0.32\}$ (N),

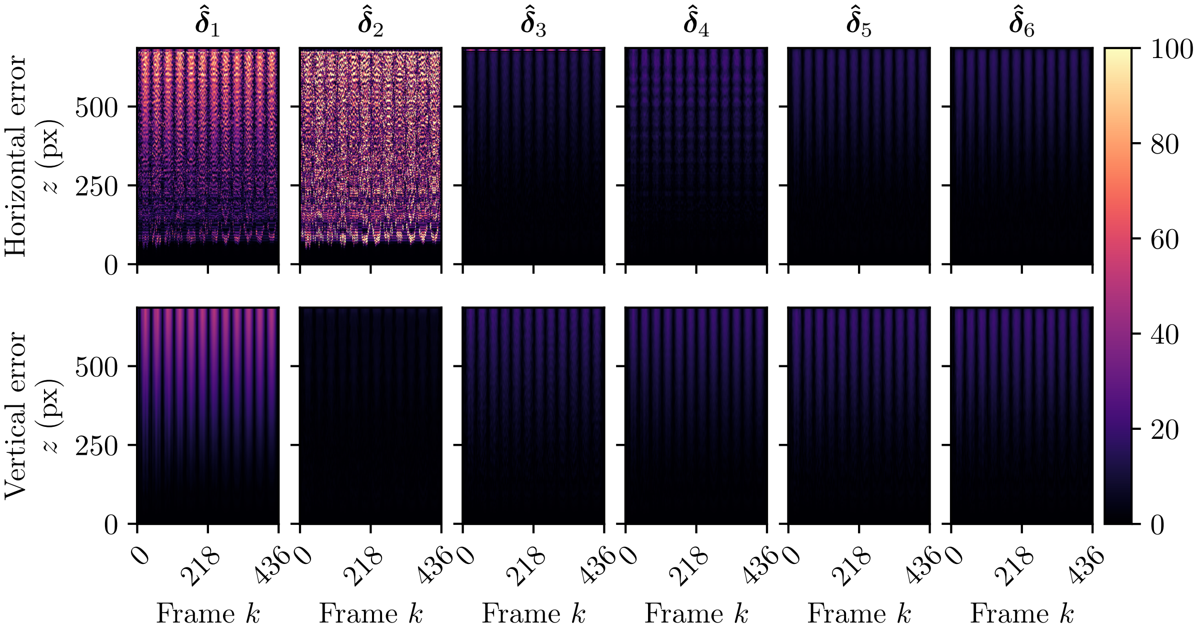

Figure 2.10: Horizontal (top row) and vertical (bottom row) RAEs of the beam (%) for the six tested estimators.

then we follow the same process as before to estimate displacement. Finally, we compute the CS of the six methods for each video. From the results shown in Tab. 2.2, we can see that CS decreases with respect to the input force (hence the motion amplitude) for all methods. Furthermore, our SMPC estimator $\hat{\boldsymbol{\delta}}_6$ provides slightly better results than $\hat{\boldsymbol{\delta}}_3$, $\hat{\boldsymbol{\delta}}_4$, and $\hat{\boldsymbol{\delta}}_5$.

Method / Force $\gamma(L,0)$	GSLF $\hat{\boldsymbol{\delta}}_1$	GSCF $\hat{\boldsymbol{\delta}}_2$	MAS $\hat{\boldsymbol{\delta}}_3$	WSL $\hat{\boldsymbol{\delta}}_4$	PC $\hat{\boldsymbol{\delta}}_5$	SMPC $\hat{\boldsymbol{\delta}}_6$
0.64 N ($\delta_{\max} = 0.26$ px)	0.940	0.940	**0.999**	**0.999**	**0.999**	**0.999**
0.32 N ($\delta_{\max} = 0.13$ px)	0.857	0.857	**0.998**	**0.998**	**0.998**	**0.998**
0.16 N ($\delta_{\max} = 0.06$ px)	0.713	0.713	0.992	0.992	0.992	**0.993**
0.08 N ($\delta_{\max} = 0.03$ px)	0.542	0.542	0.971	0.971	0.972	**0.973**

Table 2.2: Cosine similarity of the methods for different input forces $\gamma(L,0)$. Bold indicates the best result in each row.

Figure 2.11 compares the displacements estimated at the middle of the beam for the video generated with input force $\gamma(L,0) = 0.08$ N (about 0.01 px amplitude). Irrelevant horizontal values outside the interval $[-0.002, 0.002]$ px are truncated. We can see that the estimated displacements are quantized with a few numbers of levels. This suggests that the methods could give better results with more quantization bits N_b in the video.

To confirm this hypothesis, the same video has been regenerated with $N_b = 12$ quantization bits. The displacements obtained at the middle of the beam represented in Fig. 2.12 are more

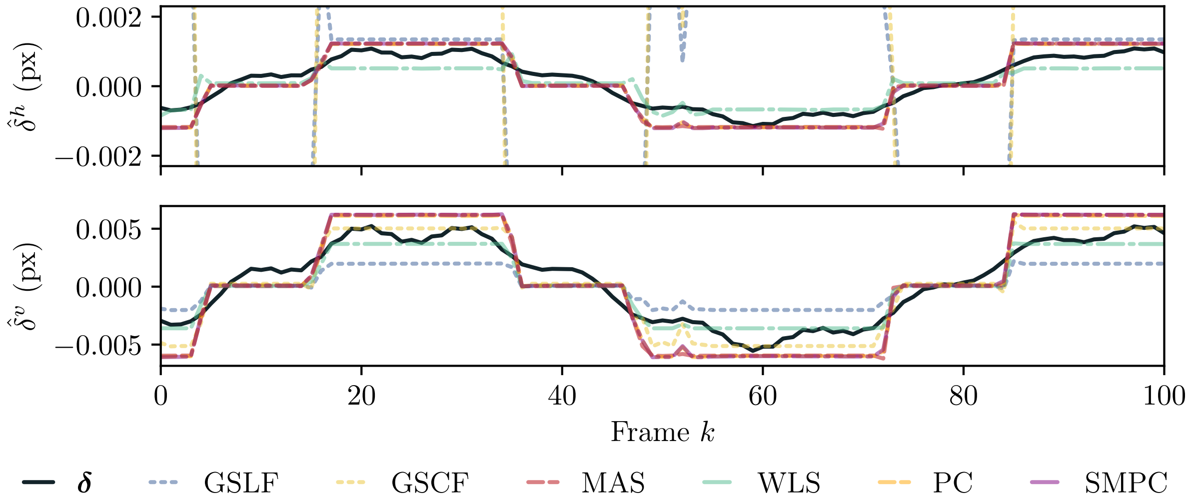

Figure 2.11: Comparison of theoretical and estimated displacement at the middle of the beam in the first 101 frames of the video with input force $\gamma(L, 0) = 0.08\,\mathrm{N}$.

consistent than that determined from the video with $N_b = 8$ quantization bits. The CS for the two videos is listed in Tab. 2.3 and confirms that more quantization bits improves estimation of small subpixel motion. It also shows that the SMPC method ($\hat{\boldsymbol{\delta}}_6$) provides the best results in both cases, even if several methods reach the same first rank when $N_b = 12$ bits.

Method / N_b	GSLF $\hat{\boldsymbol{\delta}}_1$	GSCF $\hat{\boldsymbol{\delta}}_2$	MAS $\hat{\boldsymbol{\delta}}_3$	WSL $\hat{\boldsymbol{\delta}}_4$	PC $\hat{\boldsymbol{\delta}}_5$	SMPC $\hat{\boldsymbol{\delta}}_6$
8 bits	0.542	0.542	0.971	0.971	0.972	**0.973**
12 bits	0.938	0.938	**0.998**	**0.998**	**0.998**	**0.998**

Table 2.3: Cosine similarity of the methods for different video quantization bits. Bold indicates the best result in each row.

2.7.2 Sensitivity study against gray level noise

To assess the robustness of the methods against gray level noise, we add synthetic noise to the frames before estimating motion from the video with $\gamma(L, 0) = 0.64\,\mathrm{N}$. Once the (already quantized) video file is loaded, a centered Gaussian noise with standard deviation σ_{noise} is added, then the real-valued video is re-quantized on 8 bits. Note that this method is fast but suboptimal because noise should have been added before the analog–digital conversion step to accurately represent a noise caused by sensors. Displacement are then estimated using the six methods, and CS values with theoretical displacement are computed and gathered in Tab. 2.4. We can see that the CS decreases when noise increases for all estimators, particularly for $\hat{\boldsymbol{\delta}}_1$ and $\hat{\boldsymbol{\delta}}_2$. This

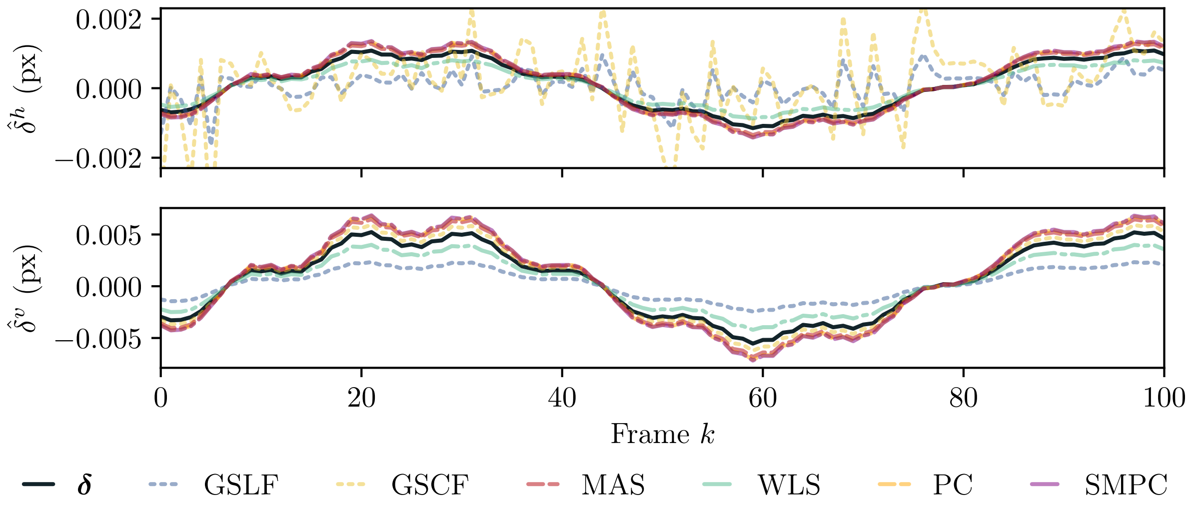

Figure 2.12: Comparison of theoretical and estimated displacement at the middle of the beam in the first 101 frames of the video with 12 quantization bits and input force $\gamma(L, 0) = 0.08\,\mathrm{N}$.

lack of robustness of $\hat{\boldsymbol{\delta}}_1$ and $\hat{\boldsymbol{\delta}}_2$ against noise is confirmed by a study on horizontal motion in [32], where we compare $\hat{\delta}_1^h$, $\hat{\delta}_2^h$, and $\hat{\delta}_4^h$ on synthetic videos of a vertical cantilever beam. As for robustness against input force, our SMPC method ($\hat{\boldsymbol{\delta}}_6$) provides the best results, even for strong noise.

Method σ_{noise}	GSLF $\hat{\boldsymbol{\delta}}_1$	GSCF $\hat{\boldsymbol{\delta}}_2$	MAS $\hat{\boldsymbol{\delta}}_3$	WSL $\hat{\boldsymbol{\delta}}_4$	PC $\hat{\boldsymbol{\delta}}_5$	SMPC $\hat{\boldsymbol{\delta}}_6$
0 px	0.940	0.940	**0.999**	**0.999**	**0.999**	**0.999**
1 px	0.415	0.415	0.998	0.996	0.998	**0.999**
$\sqrt{2}$ px	0.314	0.314	0.997	0.993	**0.998**	**0.998**
2 px	0.091	0.091	0.994	0.986	**0.997**	**0.997**
$2\sqrt{2}$ px	0.072	0.072	0.989	0.973	**0.995**	**0.995**
4 px	0.030	0.030	0.979	0.946	0.990	**0.991**

Table 2.4: Cosine similarity of the methods for different noise standard deviation σ_{noise}.

Figure 2.13 shows the displacements at the middle of the beam obtained for the six methods. All the estimators capture the main frequency of vertical motion, but only $\hat{\delta}_3^h$, $\hat{\delta}_5^h$, and $\hat{\delta}_6^h$ are close to the theoretical horizontal displacement.

2.8 Conclusion

In this chapter, we give details about different phase-based subpixel motion estimation methods. Three single-subband and three multi-subband methods are derived.

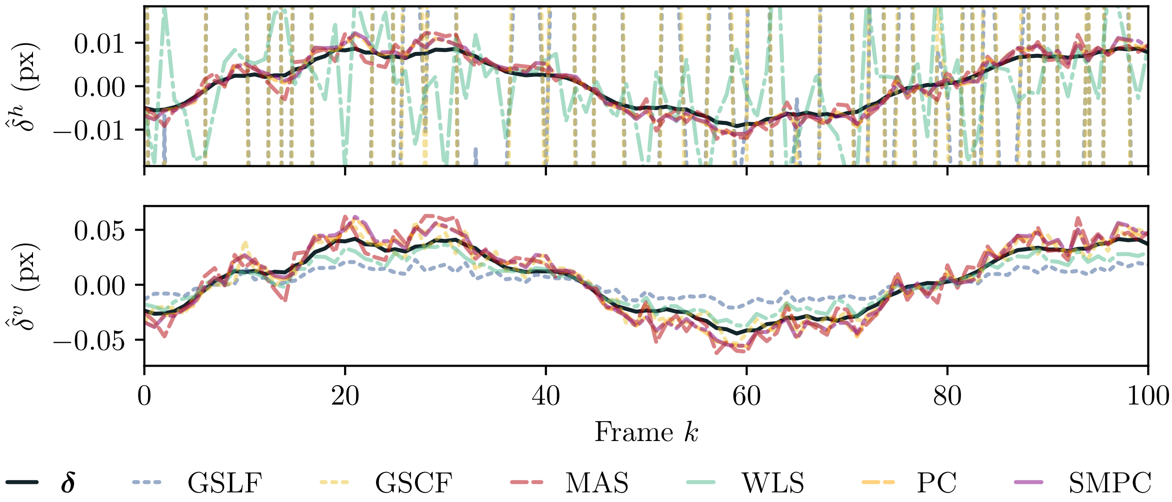

Figure 2.13: Comparison of theoretical and estimated displacement at the middle of the beam in the first 101 frames with Gaussian noise standard deviation $\sigma_{\text{noise}} = 4\,\text{px}$.

To compare the performances reached by these methods, we generate synthetic videos of a cantilever beam based on Euler–Bernoulli beam equation and a pinhole camera model. This model is very simple and does not take any point spread function or optical distortion into account. Furthermore, as the cantilever beam displacement is in a plane parallel to the image plane, the displacement is linearly projected. It could be rendered in 3D thanks to a computer graphics software such as Blender[1] to study the impact of the camera position.

Finally, the estimated motions are compared with theoretical displacements for different motion and noise amplitudes using cosine similarity. Our SMPC method based on a weighted spatial mean with phase confidence mask provides the best results in every video.

In this chapter, the focus is given on subpixel motion estimation from video analysis. Chapter 3 details the state-of-the-art and video-specific modal identification algorithms. The methods are then compared on the synthetic videos by evaluating the quality of the estimated modal basis.

---

[1] https://www.blender.org/

3 Video based operational modal analysis

3.1 Introduction

Operational modal analysis (OMA) estimates the modal basis of a structure from vibration measurements. The modal basis is composed of a set of vibration modes. Each estimated mode is defined by its natural frequency, damping ratio and shape. The theoretical foundations of OMA are presented in Sec. 1.2 of the first chapter. Many OMA methods exist and are split into two categories, temporal and frequency methods. Temporal methods estimate the modal basis directly from the vibration measurements. Frequency methods analyze the temporal-frequency space built from frequency response functions (FRFs).

In this chapter, we focus on the modal basis estimation methods. First, the estimation of FRFs is described in Sec. 3.2. Secondly, different state-of-the-art OMA methods are detailed in Sec. 3.3. Since the phase-based analysis provides a dense motion estimation, two video-based modal analysis methods are introduced in Sec. 3.4. The video-based methods are then compared on a synthetic video and the influence of the phase-based motion estimation method is studied in Sec. 3.5. A preprocessing step based on a smoothing of the estimated motion is proposed in Sec. 3.6, and the performances reached by modal analysis combined with the preprocessing are assessed against additive gray level noise and mode shape amplitude. Finally, the performances of motion estimators detailed in chapter 2 combined with the video-based OMA methods are compared on synthetic videos of a cantilever beam.

3.2 Frequency response function estimation

In modal analysis, measurements can be affected by noise. To reduce the noise of measurements, stochastic tools such as cross-correlation and spectral densities are used to estimate the FRFs. In the context of OMA, the cross-correlation of two stationary signals x and y is defined by:

$$R_{xy}(\tau) = \mathbb{E}[x(t) \cdot y(t+\tau)]. \tag{3.1}$$

The associated cross-spectral density (CSD) is the Fourier transform of the cross-correlation function that is expressed as:

$$S_{xy}(\omega) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} R_{xy}(\tau) \cdot e^{-j \cdot \omega \cdot \tau} d\tau, \tag{3.2}$$

where ω is the angular frequency. Assuming that data have been acquired over a very long period, the CSD has the following property:

$$S_{xy}(\omega) = \hat{X}^{*}(\omega) \cdot \hat{Y}(\omega), \tag{3.3}$$

with $\hat{X}$ and $\hat{Y}$ the Fourier transform of x and y, respectively. The auto-correlation and power spectral density (PSD) of signal x are denoted as $R_{xx}(\omega)$ and $S_{xx}(\omega)$.

For given measured input force γ and response y, the FRFs can be estimated from the input/output CSDs and PSDs by:

$$\hat{H}_1(\omega) = \frac{S_{\gamma y}(\omega)}{S_{\gamma\gamma}(\omega)}, \tag{3.4}$$

$$\hat{H}_2(\omega) = \frac{S_{yy}(\omega)}{S_{y\gamma}(\omega)}. \tag{3.5}$$

In operational conditions, as the input force vector $\boldsymbol{\gamma}(t)$ is unknown, the FRFs cannot be estimated. Thus, the output cross-correlation matrix $\boldsymbol{R}_{yy}(\tau)$ and the PSD matrix $\boldsymbol{S}_{yy}(\omega)$ are computed. Assuming the input force is a white noise, the two matrices behave respectively like the impulse response function and FRF matrices [19, 33].

In practice, the PSD is estimated using Welch's method [50]. This method splits the response y into overlapping time intervals. The response on each interval is windowed with the Hann window to compute the discrete Fourier transform (DFT). Finally, the PSD is estimated by computing the mean of the DFT squared modulus of each windowed interval response.

3.3 Classical OMA methods

Different OMA methods exist to determine the modal basis from cross-correlation or PSD values of observations. These methods can be subdivided into two categories, according to the number of degrees of freedom (DoFs) they consider. Single DoF methods assume that modes are well separated in frequency and that they can be determined by studying narrow frequency bands independently. On the other hand, multiple DoF methods can estimate modes that are close in frequency space. This section provides details about a simple peak-picking single DoF method and two multiple DoF methods called least-square complex frequency and stochastic subspace identification.

3.3.1 Peak-picking (PP) and Operational deflection shape (ODS)

The peak-picking method is a single DoF method which assumes that, in a given narrow frequency band, the response is only due to one single mode. The influence of other modes in this narrow frequency band is supposed to be negligible.

Around the peak corresponding to the i^{th} mode, the FRF amplitude follows the form:

$$|H(f)| = \frac{1}{k_i \cdot \left(\sqrt{\left(1-\left(\frac{f}{f_i}\right)^2\right)^2 + \left(2\cdot\zeta_i\cdot\frac{f}{f_i}\right)^2}\right)}, \tag{3.6}$$

with $f = \frac{\omega}{2\pi}$ the frequency (Hz), f_i the natural frequency, ζ_i the damping ratio of the mode, and k_i the stiffness associated to the mode.

The FRF amplitude expressed by Eq. (3.6) is maximized at the frequency $f_{\text{max}} = f_i \cdot \sqrt{1 - 2\cdot\zeta_i^2}$. For a low damping ratio $\zeta_i \ll 1$, the natural frequency can be approximated by the peak frequency f_{max}.

According to the half-power method [11], the damping ratio can be estimated by:

$$\zeta_i = \frac{f_{\text{up}}^2 - f_{\text{low}}^2}{4\cdot f_i^2}, \tag{3.7}$$

with f_{low} and f_{up} the half-power frequencies such that $|H(f_{\text{low}})| = |H(f_{\text{up}})| = \frac{|H(f_{\text{max}})|}{\sqrt{2}}$ and $f_{\text{low}} < f_{\text{max}} < f_{\text{up}}$. For this purpose, the FRF is interpolated in the frequency domain. Figure 3.1 displays a close view of an FRF with the frequencies f_{low}, f_{up} and f_{max} associated to a peak.

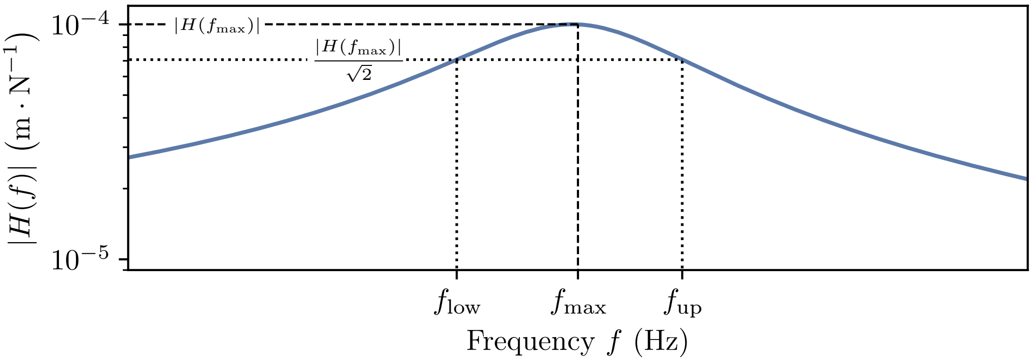

Figure 3.1: Peak-picking and half-power methods

The mode shape can be approximated by the operational deflection shape (ODS) computed at the selected frequency f_{max}. For this purpose, a given line i of the FRF (or PSD) matrix is selected, and every component is set to zero except for $\omega = 2\pi\cdot f_{\text{max}}$. For every DoF $j \in [\![1, N]\!]$, the corresponding ODS is obtained by computing the inverse Fourier transform of FRF $\boldsymbol{H}_{ij}(\omega)$.

The uncertainty of this method is directly linked to the frequency resolution of the FRF and the interpolation method. Furthermore, if several modes are close together, the assumption about single DoF (one mode in a narrow frequency band) is not valid.

3.3.2 Least-square complex frequency (LSCF)

The least-square complex frequency method [13] is a multiple DoF frequency method that aims to fit the FRF (resp. the SD) matrix. As seen in Sec. 1.2 (see Eq. (1.36)), the FRF components follow the form:

$$\boldsymbol{H}_{pq}(\omega) = \sum_{i=1}^{N} \left(\frac{\boldsymbol{\Phi}_{pi}\boldsymbol{L}_{qi}}{\zeta_i \cdot \omega_i + j\left(\omega - \omega_i\sqrt{1-\zeta_i^2}\right)} + \frac{\boldsymbol{\Phi}_{pi}^{*}\boldsymbol{L}_{qi}^{*}}{\zeta_i \cdot \omega_i + j\left(\omega + \omega_i\sqrt{1-\zeta_i^2}\right)} \right), \tag{3.8}$$

with $\boldsymbol{L}_{qi} = \frac{\boldsymbol{\Phi}_{qi}}{\alpha_i}$, and the i^{th} mode being defined by its natural frequency ω_i, its damping ratio ζ_i, and its mode shape $\boldsymbol{\phi}_i$ (the i^{th} column of $\boldsymbol{\Phi}$).

Within the frequency band of interest, the measured FRF $\boldsymbol{H}_{pq}^{\text{m}}$ is assumed to have the following form:

$$\boldsymbol{H}_{pq}^{\text{m}}(\omega) = \sum_{i=1}^{N} \left(\frac{\boldsymbol{\Phi}_{pi}\boldsymbol{L}_{qi}}{\zeta_i \cdot \omega_i + j\left(\omega - \omega_i\sqrt{1-\zeta_i^2}\right)} + \frac{\boldsymbol{\Phi}_{pi}^{*}\boldsymbol{L}_{qi}^{*}}{\zeta_i \cdot \omega_i + j\left(\omega + \omega_i\sqrt{1-\zeta_i^2}\right)} \right) + \boldsymbol{R}_{pq}^{\text{u}} - \frac{\boldsymbol{R}_{pq}^{\text{l}}}{\omega^2}, \tag{3.9}$$

with $\boldsymbol{R}_{pq}^{\text{u}}$ and $\boldsymbol{R}_{pq}^{\text{l}}$ the upper and lower residuals matrices [15].

The method fits the estimated FRFs with a function $\hat{\boldsymbol{H}}$ of the same form as in Eq. (3.9). For this purpose, the difference between the measured and the estimated FRFs at angular frequency ω is defined by:

$$e_{pq}(\omega) = \boldsymbol{H}_{pq}^{\text{m}}(\omega) - \hat{\boldsymbol{H}}_{pq}(\omega). \tag{3.10}$$

The total squared error to minimize is:

$$E = \sum_{p}\sum_{q}\sum_{\omega} e_{pq}(\omega) \cdot e_{pq}^{*}(\omega). \tag{3.11}$$

For each unknown parameter $\{\omega_i, \zeta_i, \boldsymbol{\Phi}_{pi}, \boldsymbol{L}_{qi}, \boldsymbol{R}_{pq}^{\text{u}}, \boldsymbol{R}_{pq}^{\text{l}}\}$, the following set of equations $\frac{\partial E}{\partial \omega_i} = \frac{\partial E}{\partial \zeta_i} = \frac{\partial E}{\partial \boldsymbol{\Phi}_{pi}} = \frac{\partial E}{\partial \boldsymbol{L}_{qi}} = \frac{\partial E}{\partial \boldsymbol{R}_{pq}^{\text{u}}} = \frac{\partial E}{\partial \boldsymbol{R}_{pq}^{\text{l}}} = 0$ minimizes the error. This set of equations is non-linear and is solved iteratively using first-order approximations. As this problem is sensitive to initial values, it is used in practice to refine a model that is coarsely estimated by another method such as Ibrahim's time-domain method [17].

3.3.3 Stochastic subspace identification (SSI)

The covariance-driven stochastic subspace identification is a time-domain method based on a multiple DoF state space formulation. As detailed in Sec. 1.2.4, the multiple DoF problem can be recast into a discrete-time state space form:

$$\boldsymbol{x}[k+1] = \boldsymbol{A} \cdot \boldsymbol{x}[k] + \boldsymbol{w}[k], \tag{3.12}$$

$$\boldsymbol{y}[k] = \boldsymbol{C} \cdot \boldsymbol{x}[k] + \boldsymbol{v}[k], \tag{3.13}$$

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

The objective of stochastic subspace identification (SSI) is to get estimates $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{C}}$ only from the observations $\{\boldsymbol{y}[k]\}_{k=0}^{N_k-1}$ (columns of $\hat{\boldsymbol{\delta}}^h$ and $\hat{\boldsymbol{\delta}}^v$, in our case) to obtain the modes [35]. For this purpose, a set of covariance matrices are defined as follows:

- $\boldsymbol{\Lambda}_i = \mathbb{E}[\boldsymbol{y}[k+i] \cdot \boldsymbol{y}[k]^\intercal] \in \mathbb{R}^{N_o \times N_o}$ the output covariance matrix,
- $\boldsymbol{G} = \mathbb{E}[\boldsymbol{x}[k+1] \cdot \boldsymbol{y}[k]^\intercal] \in \mathbb{R}^{2N \times N_o}$ the next-state covariance matrix.

These matrices are linked by:

$$\boldsymbol{\Lambda}_i = \boldsymbol{C} \cdot \boldsymbol{A}^{i-1} \cdot \boldsymbol{G}. \tag{3.14}$$

Given a parameter $R < N_k/2$, each output covariance matrix can be estimated by:

$$\boldsymbol{\Lambda}_i = \frac{1}{N_k - 2R - 1} \sum_{k=0}^{N_k-2R-2} \boldsymbol{y}[k+i] \cdot \boldsymbol{y}[k]^\intercal. \tag{3.15}$$

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

$$\boldsymbol{T}_{1:R} = \begin{pmatrix} \boldsymbol{\Lambda}_R & \boldsymbol{\Lambda}_{R-1} & \cdots & \boldsymbol{\Lambda}_1 \\ \boldsymbol{\Lambda}_{R+1} & \boldsymbol{\Lambda}_R & \cdots & \boldsymbol{\Lambda}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Lambda}_{2R-1} & \boldsymbol{\Lambda}_{2R-2} & \cdots & \boldsymbol{\Lambda}_R \end{pmatrix} \in \mathbb{R}^{R \cdot N_o \times R \cdot N_o}. \tag{3.16}$$

This block Toeplitz matrix can be decomposed using Eq. (3.14) to give:

$$\boldsymbol{T}_{1:R} = \underbrace{\begin{pmatrix} \boldsymbol{C} \\ \boldsymbol{C}\boldsymbol{A} \\ \vdots \\ \boldsymbol{C}\boldsymbol{A}^{R-1} \end{pmatrix}}_{\boldsymbol{O}_R} \cdot \underbrace{\begin{pmatrix} \boldsymbol{A}^{R-1}\boldsymbol{G} & \boldsymbol{A}^{R-2}\boldsymbol{G} & \dots & \boldsymbol{A}\boldsymbol{G} & \boldsymbol{G} \end{pmatrix}}_{\boldsymbol{C}_R}. \tag{3.17}$$

For a given model order N, the generalized observability matrix $\boldsymbol{O}_R \in \mathbb{R}^{R \cdot N_o \times 2N}$ and the generalized reversed controllability matrix $\boldsymbol{C}_R \in \mathbb{R}^{2N \times R \cdot N_o}$ [46] can be estimated using the singular value decomposition (SVD) of the Toeplitz matrix that is written as:

$$\boldsymbol{T}_{1:R} = \begin{pmatrix} \boldsymbol{U}_{2N} & \overline{\boldsymbol{U}_{2N}} \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{\Sigma}_{2N} & \boldsymbol{0} \\ \boldsymbol{0} & \overline{\boldsymbol{\Sigma}_{2N}} \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{V}_{2N}^{\intercal} \\ \overline{\boldsymbol{V}_{2N}^{\intercal}} \end{pmatrix}, \tag{3.18}$$

$$\approx \boldsymbol{U}_{2N} \cdot \boldsymbol{\Sigma}_{2N} \cdot \boldsymbol{V}_{2N}^{\intercal}. \tag{3.19}$$

Matrices $\boldsymbol{U}_{2N}$ and $\boldsymbol{V}_{2N}$ are both orthonormal, *i.e.*, $\boldsymbol{U}_{2N} \cdot \boldsymbol{U}_{2N}^{\intercal} = \boldsymbol{U}_{2N}^{\intercal} \cdot \boldsymbol{U}_{2N} = \boldsymbol{I}_{2N}$ and $\boldsymbol{V}_{2N} \cdot \boldsymbol{V}_{2N}^{\intercal} = \boldsymbol{V}_{2N}^{\intercal} \cdot \boldsymbol{V}_{2N} = \boldsymbol{I}_{2N}$. The matrix $\boldsymbol{\Sigma}_{2N}$ is diagonal and composed of singular values. From Eqs. (3.17) and (3.19), the observability and reversed controllability matrices can be estimated by:

$$\boldsymbol{O}_R = \boldsymbol{U}_{2N} \cdot \boldsymbol{\Sigma}_{2N}^{1/2}. \tag{3.20}$$

$$\boldsymbol{C}_R = \boldsymbol{\Sigma}_{2N}^{1/2} \cdot \boldsymbol{V}_{2N}^{\intercal}. \tag{3.21}$$

Matrix $\boldsymbol{C}$ is estimated by taking the first N_o rows of $\boldsymbol{O}_R$. To estimate A, the Toeplitz matrix $\boldsymbol{T}_{2:R+1}$ is constructed such that:

$$\boldsymbol{T}_{2:R+1} = \boldsymbol{O}_R \cdot \boldsymbol{A} \cdot \boldsymbol{C}_R, \tag{3.22}$$

$$\hat{\boldsymbol{A}} = \boldsymbol{O}_R^{\dagger} \cdot \boldsymbol{T}_{2:R+1} \cdot \boldsymbol{C}_R^{\dagger} \tag{3.23}$$

$$= \boldsymbol{\Sigma}_{2N}^{-1/2} \cdot \boldsymbol{U}_{2N}^{\intercal} \cdot \boldsymbol{T}_{2:R+1} \cdot \boldsymbol{V}_{2N} \cdot \boldsymbol{\Sigma}_{2N}^{-1/2}, \tag{3.24}$$

with $\boldsymbol{O}_R^{\dagger}$ and $\boldsymbol{C}_R^{\dagger}$ the Moore-Penrose pseudoinverse matrices of $\boldsymbol{O}_R$ and $\boldsymbol{C}_R$.

Then, from the eigenvalue decomposition $\hat{\boldsymbol{A}} = \boldsymbol{\Psi S \Psi^{-1}}$ with $\boldsymbol{S} = \operatorname{diag}(\lambda_i), i \in [\![1, 2N]\!]$, natural frequencies, damping ratios, and mode shapes can be computed as:

$$f_i^N = \frac{1}{2\pi}\left|\frac{\log \lambda_i}{\Delta t}\right|, \tag{3.25}$$

$$\zeta_i^N = \frac{\mathfrak{Re}(\lambda_i)}{|\lambda_i|}, \tag{3.26}$$

$$\boldsymbol{\Phi}^N = \hat{\boldsymbol{C}}\boldsymbol{\Psi}. \tag{3.27}$$

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

In practice, the number N_m of modes in the frequency band of interest is not known. Therefore, the method is run with different model orders N. For a given model order, N *poles* $\left\{p_i^N\right\}_{i=1}^N$ are obtained, each with its natural frequency f_i^N, its damping ratio ζ_i^N, and its mode shape $\boldsymbol{\phi}_i^N$.

3.4 Video-based OMA methods

Let N_{PoI} be the number of pixels of interest where displacement is estimated along N_k frames. The observation matrix $\boldsymbol{y} \in \mathbb{R}^{N_o \times N_k}$ is composed of horizontal and vertical displacements stacked with $N_o = 2N_{\text{PoI}}$. The specificity of video-based modal analysis is the number of observations, which can be very high due to dense motion estimation. Thus, the two following methods first reduce the dimension of the observation matrix before running a modal analysis.

3.4.1 Principal component analysis and complexity pursuit (PCA+CP)

The objective of principal component analysis (PCA) and complexity pursuit (CP) [52] is to decompose the displacement matrix $\boldsymbol{y}$ as:

$$\boldsymbol{y} = \boldsymbol{\Phi} \cdot \boldsymbol{q}, \tag{3.28}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{N_o \times N}$ and $\boldsymbol{q} \in \mathbb{R}^{N \times N_k}$ are the mode shape and modal coordinate matrices, respectively. The method proceeds in two steps. First, PCA is used to reduce the number of displacement matrix rows from N_o to a given model order N. Then, the modal coordinates are separated from mode shapes using a blind source separation by CP algorithm.

When N_{PoI} is high, video-based modal analysis may examine a high-dimensional displacement matrix. To reduce its size, it is factorized by SVD:

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{U} & \overline{\boldsymbol{U}} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \overline{\boldsymbol{\Sigma}} \end{pmatrix} \begin{pmatrix} \boldsymbol{V} \\ \overline{\boldsymbol{V}} \end{pmatrix}. \tag{3.29}$$

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

$$\boldsymbol{\eta} = \boldsymbol{U}^\intercal \cdot \boldsymbol{y}. \tag{3.30}$$

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

$$\boldsymbol{q} = \boldsymbol{W} \cdot \boldsymbol{\eta} \ , \tag{3.31}$$

where $\boldsymbol{W} \in \mathbb{R}^{N \times N}$ is the demixing matrix to retrieve, this blind source separation problem is solved using CP [54]. This method assumes that the modal coordinates $\boldsymbol{q}$ are the signals with the lowest Kolmogorov complexity. To extract them, the method estimates each row $\boldsymbol{w_i}$ of $\boldsymbol{W}$ so that each $\boldsymbol{q_i} := \boldsymbol{w_i} \cdot \boldsymbol{\eta} \in \mathbb{R}^{N_k}$, $i \in [\![1,\ N]\!]$, has the highest temporal predictability, that is defined as:

$$P(\boldsymbol{q}_i) = \log \left(\frac{\sum_{k=0}^{N_k-1} (\bar{q}_i[k] - q_i[k])^2}{\sum_{k=0}^{N_k-1} (\check{q}_i[k] - q_i[k])^2} \right) \ , \tag{3.32}$$

where $\bar{q}_i$ and $\check{q}_i$ are the long and short exponential moving averages of $\boldsymbol{q}_i$.

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

$$f_i^N = \arg\max_f \left|\mathrm{DFT}\{\boldsymbol{q}_i\}[f]\right| \ . \tag{3.33}$$

Each damping ratio ζ_i^N is estimated from q_i using the logarithmic decrement method [18]. Each mode shape $\boldsymbol{\phi}_i^N$ is a column of the mode shape matrix $\boldsymbol{\Phi}^N$ that is computed as:

$$\boldsymbol{\Phi}^N = \boldsymbol{U}\boldsymbol{W}^{-1} \ . \tag{3.34}$$

Yang et al. [52] estimate the displacement at each scale of the complex steerable pyramid (see Sec. 2.3.2) and apply this technique to determine a modal basis at each scale.

3.4.2 Fast SSI (FSSI)

In the state-of-the-art SSI method (see Sec. 3.3.3), the size $R \cdot N_o \times R \cdot N_o$ of the Toeplitz matrix may be huge for video-based modal analysis. To reduce the number of observations, we propose to use a similar projection as in Eq. (3.30). The displacement matrix is first decomposed by SVD, then projected on the first P vectors to obtain a reduced observation matrix $\check{\boldsymbol{y}} \in \mathbb{R}^{P \times N_k}$:

$$\boldsymbol{y} = \left(\boldsymbol{U}_P\, \overline{\boldsymbol{U}_P}\right) \begin{pmatrix} \boldsymbol{\Sigma}_P & \boldsymbol{0} \\ \boldsymbol{0} & \overline{\boldsymbol{\Sigma}_P} \end{pmatrix} \begin{pmatrix} \boldsymbol{V}_P \\ \overline{\boldsymbol{V}_P} \end{pmatrix}, \tag{3.35}$$

$$\check{\boldsymbol{y}} = \boldsymbol{U}_P^{\intercal} \cdot \boldsymbol{y}. \tag{3.36}$$

Given a value of P, SSI is performed on $\check{\boldsymbol{y}}$ with N and R as parameters. The Toeplitz matrix $\check{\boldsymbol{T}}_{1:R} \in \mathbb{R}^{R \cdot P \times R \cdot P}$ 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 N_o pixels can then be computed as:

$$\boldsymbol{\Phi}^N = U_P \cdot \check{\boldsymbol{\Phi}}^N. \tag{3.37}$$

3.5 Operational modal analysis on synthetic videos

To assess the performances reached by video-based modal analysis methods, we consider the displacements estimated from the synthetic videos used in chapter 2 as observations. To do so, a line of $N_{\mathrm{PoI}} = 686$ pixel of interest is defined on the synthetic cantilever beam rotated by 78.75° (see Figs. 3.2 (a) and 3.2 (b)). Then, motion is estimated in each frame k and extracted for each pixel of interest with curvilinear coordinate z (see Figs. 3.2 (c) and 3.2 (d)) to estimate $\boldsymbol{\delta}[z; k]$.

3.5.1 Qualitative comparison of video-based OMA methods

To compare the performances reached by PCA+CP and FSSI methods, the video generated with input force $\gamma(L, 0) = 0.64\,\mathrm{N}$ and without additive gray level noise ($\sigma_{\text{noise}} = 0\,\mathrm{px}$) is analyzed by the SMPC method (see Sec. 2.5.3) to extract displacements. For FSSI, the observations are reduced by keeping the $P = 10$ first principal components, and the block Toeplitz matrix is constructed with $R = 40$ block rows.

3.5.1.1 Stability analysis

As the number N_m of modes in the frequency range of interest is not known, a stability analysis is performed to estimate the modal basis. For this purpose, PCA+CP and FSSI are run with several increasing values of model order. Indeed, irrelevant models produce spurious poles that can be discarded by a stability analysis. This analysis consists in the construction of a stabilization diagram with the poles obtained for each model order N [45]. Physical poles tend to be stable between successive model orders, while spurious ones tend to be unstable (in frequency, damping, and/or mode shape).

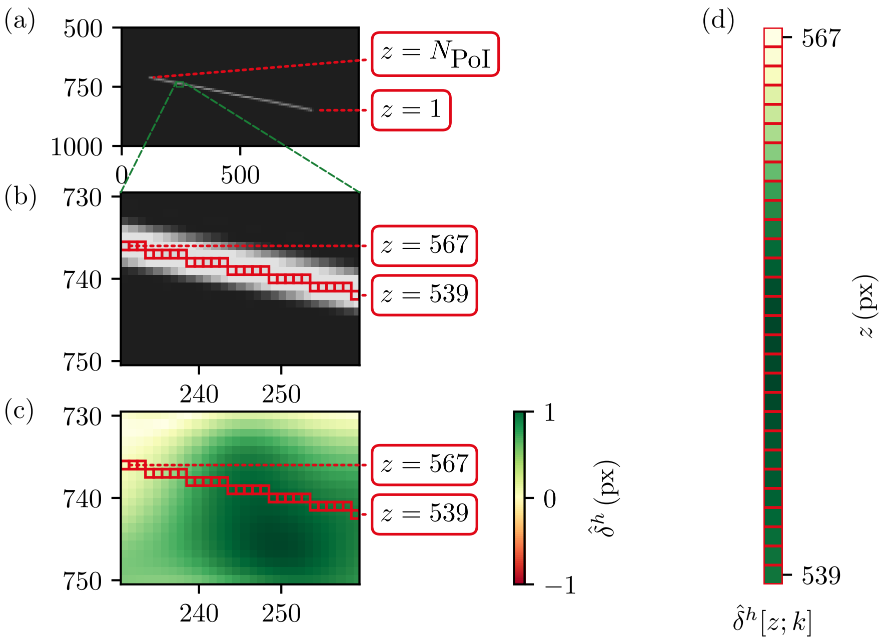

Figure 3.2: Motion extraction process of a cantilever beam. (a) Frame of a synthetic video. (b) Focus on pixels of interest in red with $z \in [\![539, 567]\!]$. (c) Estimated motion $\hat{\delta}^h[z; k]$ in the focus view. (d) Estimated motion $\hat{\delta}^h[z; k]$ at the pixels of interest.

For a pole p_i^N, $i \in [\![1, N]\!]$ of order N with f_i^N, ζ_i^N, and $\boldsymbol{\phi}_i^N$ its natural frequency, damping ratio, and mode shape, the stability criterion is based on the following predicate:

$$S(p_i^N) = \exists p_j^{N-1} \left[\left(\frac{|f_i^N - f_j^{N-1}|}{f_j^{N-1}} < 0.01 \right) \wedge \left(\frac{|\zeta_i^N - \zeta_j^{N-1}|}{\zeta_j^{N-1}} < 0.05 \right) \wedge \left(\mathrm{MAC}(\boldsymbol{\phi}_i^N, \boldsymbol{\phi}_j^{N-1}) > 0.98 \right) \right], \tag{3.38}$$

with the modal assurance criterion (MAC) between any two mode shapes being defined by:

$$\mathrm{MAC}(\boldsymbol{\phi}, \tilde{\boldsymbol{\phi}}) = \frac{(\boldsymbol{\phi}^\intercal \tilde{\boldsymbol{\phi}})^2}{(\boldsymbol{\phi}^\intercal \boldsymbol{\phi})(\tilde{\boldsymbol{\phi}}^\intercal \tilde{\boldsymbol{\phi}})} . \tag{3.39}$$

The stabilization diagram is thus built by representing the stable and unstable poles with different colors, in the frequency and model order plane. The main idea is that spurious modes occur randomly and are not stable at two consecutive model orders. Contrarily, successive stable poles around the same frequency are a good indicator of a physical mode.

An example of stabilization diagram is provided in Fig. 3.3 (a) with model order $N \in [\![2, 50]\!]$ and a frequency band of interest from 0 to 50 Hz for a mass-spring-damper system with 12 masses. In this example, the $N_m = 12$ modes can be identified thanks to vertical series of stable poles, each vertical series corresponding to one mode. The number of identified poles increases with respect to the model order N. However, when N is larger than the number N_m of modes, most of the poles are unstable.

In real experimentation, the stabilization diagram may be difficult to analyze. To help identify modes, the complex mode indicator functions (CMIFs) may also be shown in the diagram [42]. These functions are the squared eigenvalues of the estimated FRF matrix at each frequency, sorted in descending order. The first CMIF is the most important, and each of its peaks indicates the presence of a mode at the associated frequency. Figure 3.3 (b) shows the same stabilization diagram as in Fig. 3.3 (a), together with the first CMIF.

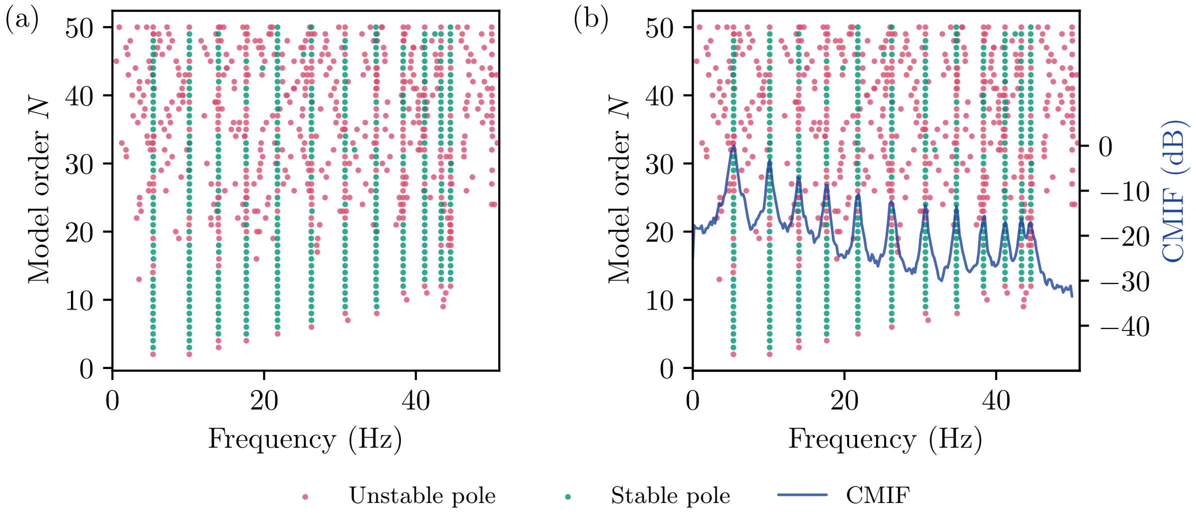

Figure 3.3: Example of stabilization diagram with 12 modes (a), with CMIF (b).

3.5.1.2 Stabilization diagrams of video-based OMA methods

To compare PCA+CP and FSSI methods qualitatively for video-based OMA, the stabilization diagrams of both methods with $N \in [\![4, 50]\!]$ are shown in Fig. 3.4 for our cantilever beam video. The PCA+CP method provides stable poles close to the theoretical natural frequencies of Tab. 3.1. However, the close view between 6 and 7 Hz shows that the poles are doubled and relatively far from the natural frequency $f_1 = 6.19\,\text{Hz}$ of the first mode. Furthermore, a vertical series of stable poles around 18 Hz does not correspond to any theoretical natural frequency. On the contrary, the four modes can be easily identified on the diagram deduced from FSSI, that provides relevant modes.

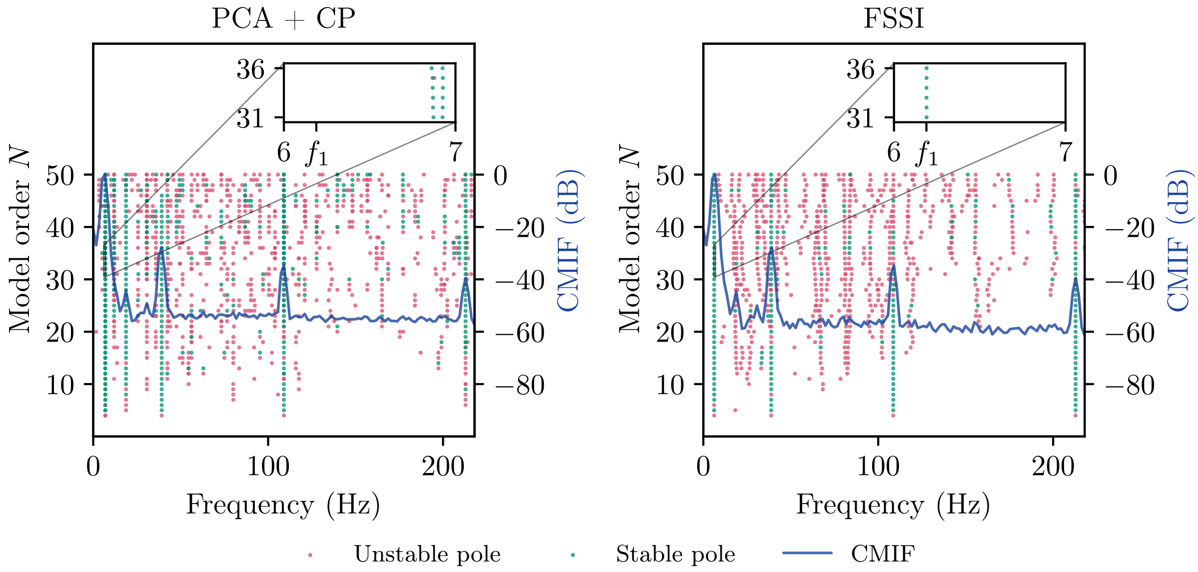

Figure 3.4: Stabilization diagrams provided by PCA+CP and FSSI methods on synthetic video of the cantilever beam.

3.5.2 Quantitative comparison of video-based OMA methods

The stabilization diagram of each OMA method is processed to estimate the modal basis. For each theoretical natural frequency f_i, the stable poles with a relative frequency error lower than 1 % are gathered and their natural frequencies, damping ratios, and mode shapes are averaged to estimate the mode $\{\hat{f}_i, \hat{\zeta}_i, \hat{\phi}_i\}$. To compare the methods quantitatively, the relative errors with respect to theoretical values recalled in Tab. 3.1 are computed as:

$$\begin{cases} \epsilon\{\hat{f}_i\} = 100 \cdot \dfrac{\left|\hat{f}_i - f_i\right|}{f_i} \\ \epsilon\{\hat{\zeta}_i\} = 100 \cdot \dfrac{\left|\hat{\zeta}_i - \zeta_i\right|}{\zeta_i} \\ \epsilon\{\hat{\boldsymbol{\phi}}_i\} = 1 - \mathrm{MAC}(\hat{\boldsymbol{\phi}}_i, \boldsymbol{\phi}_i) \end{cases}, \tag{3.40}$$

and are gathered in Tab. 3.2.

Mode i	1	2	3	4
f_i (Hz)	6.19	38.79	108.60	212.82
ζ_i ($\times 10^2$)	0.11	1.13	0.29	0.13

Table 3.1: Beam theoretical natural frequencies and damping ratios.

Because the frequency estimated by PCA+CP is shifted by more than 1% for the first mode, all the stable poles around 6.89 Hz are selected to estimate its characteristics. The FSSI-based modal basis is overall more accurate than PCA+CP, particularly for natural frequencies and damping ratios, with errors lower than 1 % and 8 %, respectively. The estimated mode shapes are represented in Fig. 3.5. The largest mode shape errors are located near the clamped end of the beam, where motion amplitude is the smallest. The methods provide similar mode shapes, with mode shape errors below 0.1.

Criterion	Frequency relative error (%)				Damping ratio relative error (%)				Mode shape error (1-MAC)			
Error	$\epsilon\{\hat{f}_1\}$	$\epsilon\{\hat{f}_2\}$	$\epsilon\{\hat{f}_3\}$	$\epsilon\{\hat{f}_4\}$	$\epsilon\{\hat{\zeta}_1\}$	$\epsilon\{\hat{\zeta}_2\}$	$\epsilon\{\hat{\zeta}_3\}$	$\epsilon\{\hat{\zeta}_4\}$	$\epsilon\{\hat{\phi}_1\}$	$\epsilon\{\hat{\phi}_2\}$	$\epsilon\{\hat{\phi}_3\}$	$\epsilon\{\hat{\phi}_4\}$
PCA + CP	11.370	0.908	0.354	0.020	> 100	78.157	> 100	> 100	0.002	**0.005**	0.012	0.078
FSSI	**0.018**	**0.018**	**0.000**	**0.007**	**2.677**	**1.919**	**2.663**	**7.389**	**0.001**	0.007	**0.011**	**0.028**

Table 3.2: Modal basis estimation errors on synthetic video for PCA+CP and FSSI methods.

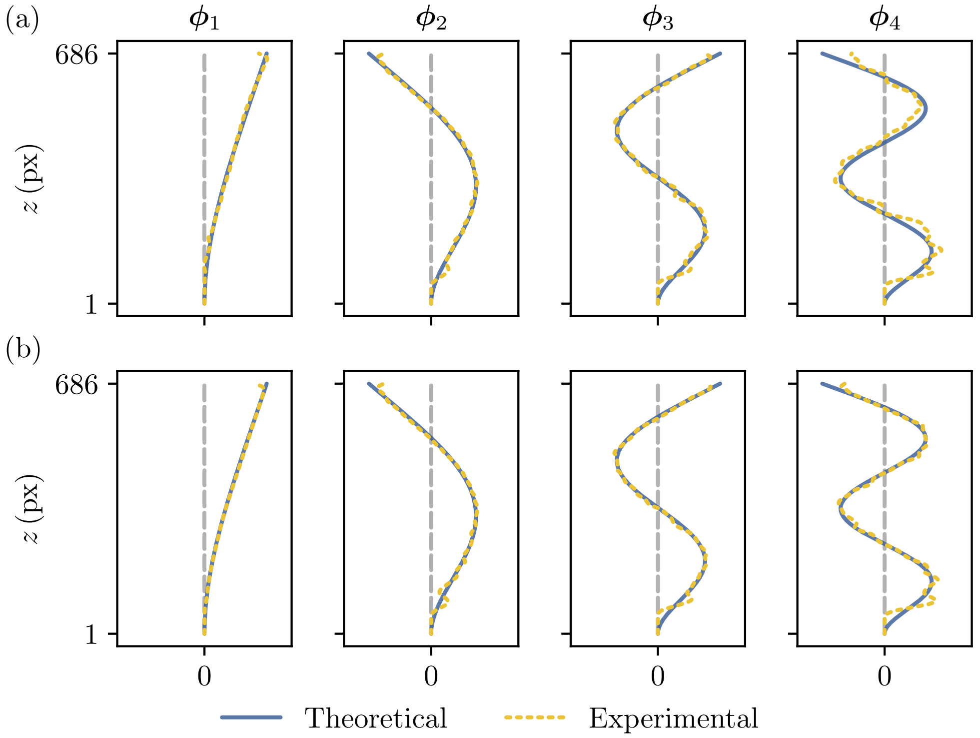

Figure 3.5: Theoretical and estimated mode shapes for PCA+CP (a) and FSSI (b).

In this study, FSSI provides better results than PCA+CP, and is therefore retained in the next experiment. This is confirmed by a more complete study in [30], where we show that FSSI always performs better than PCA+CP by comparing the modes provided by these methods on synthetic videos with different motion amplitudes, different additive gray level noises, and different Gaussian blurs.

3.5.3 Sensitivity against motion estimation quality

In chapter 2, the phase-based motion estimation (PME) methods are assessed by computing the cosine similarity between theoretical and estimated motions (see Eq. (2.56)). To study the influence of motion estimation quality on the performances of the OMA, let now FSSI analyze the motion estimated by different PME methods. For this purpose, the video is generated with input force $\gamma(L, 0) = 0.64\,\mathrm{N}$ and gray level noise standard deviation $\sigma_{\text{noise}} = 2\sqrt{2}$. Gray levels are corrupted by additive noise because this modifies the performances reached by PME methods. For FSSI method, the same parameter values as before are used to construct the stabilization diagram and estimate the modal basis automatically.

The stabilization diagrams of modal analysis based on GLSF, GSCF, and WLS (see Secs 2.4.1, 2.4.2, and 2.5.1) are not shown here because they do not retrieve the four modes. Three modes are missing for GLSF and GLCF, and the last one is missing for WLS. These three methods obtain the lowest cosine similarity (namely 0.072 for GLSF and GLCF, and 0.973 for WLS). Oppositely, MAS, PC, and SMPC (see Secs 2.4.3, 2.5.2, and 2.5.3) methods retrieve the four modes and their cosine similarity reaches 0.989, 0.995, and 0.995, respectively. This shows that high cosine similarity is required to correctly estimate the modal basis.

The modal basis estimation errors obtained using Eq. (3.40) when MAS, PC, and SMPC are used as motion estimation methods are gathered in Tab. 3.3. The results show that the SMPC method provides the best results for 6 among the 12 cases. However, the error on the fourth mode shape is quite high (above 0.2) for the three methods (see Fig. 3.6). For MAS, the fourth mode is completely uncorrelated with the theoretical mode, since the mode shape error reaches 0.689. For PC and SMPC, the fourth mode roughly correlates with the theoretical mode, but suffers from a lack of local consistency. Indeed, the fourth mode provides the lowest contribution to the beam motion and is thus harder to estimate. The motion estimation of quality of MAS method is not sufficient to estimate the mode shape correctly.

Criterion	Frequency relative error (%)				Damping ratio absolute error ($\times 10^2$)				Mode shape error (1-MAC)			
Error	$\epsilon\{\hat{f}_1\}$	$\epsilon\{\hat{f}_2\}$	$\epsilon\{\hat{f}_3\}$	$\epsilon\{\hat{f}_4\}$	$\epsilon\{\hat{\zeta}_1\}$	$\epsilon\{\hat{\zeta}_2\}$	$\epsilon\{\hat{\zeta}_3\}$	$\epsilon\{\hat{\zeta}_4\}$	$\epsilon\{\hat{\phi}_1\}$	$\epsilon\{\hat{\phi}_2\}$	$\epsilon\{\hat{\phi}_3\}$	$\epsilon\{\hat{\phi}_4\}$
MAS	0.053	0.107	0.055	0.117	> 100	**1.296**	**3.360**	79.478	**0.001**	0.077	0.242	0.689
PC	0.046	**0.051**	**0.043**	**0.075**	**15.364**	8.485	10.690	15.783	**0.001**	0.065	0.090	0.275
SMPC	**0.044**	0.057	0.049	0.094	15.972	9.702	7.526	**7.358**	**0.001**	**0.061**	**0.089**	**0.246**

Table 3.3: Modal basis estimation errors on synthetic video obtained by FSSI analysis with three phase-based motion estimation methods.

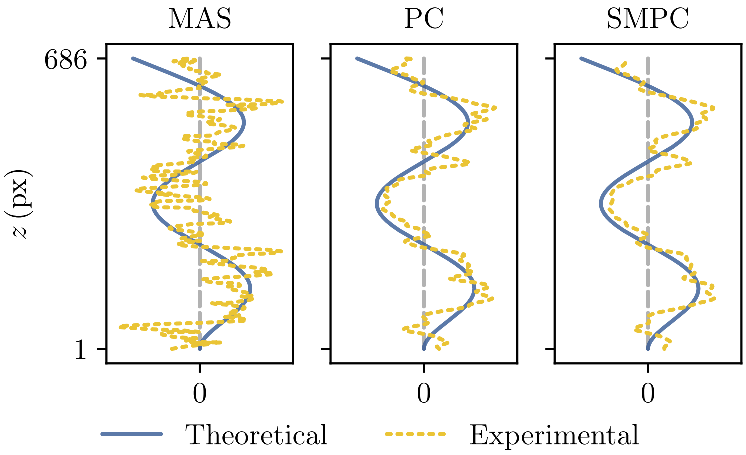

Figure 3.6: Comparison of the shapes of the fourth mode obtained by FSSI analysis with MAS, PC, and SMPC phase-based motion estimation methods.

3.6 Estimated motion smoothing

Video-based operational modal analysis can suffer from poorly estimated motion. As the studied object is continuous, its displacement projected in the frame plane is supposed to be continuous. For this purpose, we propose to preprocess the estimated motion before modal basis determination in order to reduce the number of spurious modes and improve the modal analysis accuracy.

3.6.1 Estimated motion smoothing in the frame plane

For this purpose, the flattened estimated motion at the N_{PoI} pixels of each frame are spatially smoothed by a Gaussian filter G_{smooth} with standard deviation σ_{smooth} (see Fig. 3.7). First, each vector $\hat{\delta}^h$ and $\hat{\delta}^v$ is spatially padded with the mean of the $\lceil \sigma_{\mathrm{smooth}} \rceil$ values close to the beam boundaries ($z = 1$ and $z = 686$). Then the padded vector is convolved with G_{smooth} to give new estimators $\tilde{\delta}^h[z;k] = \hat{\delta}^h[z;k] * G_{\mathrm{smooth}}$ and $\tilde{\delta}^v[z;k] = \hat{\delta}^v[z;k] * G_{\mathrm{smooth}}$.

To assess the improvement brought by this preprocessing, we apply it after SMPC motion estimation and before FSSI modal analysis. We use the video generated with $\gamma(L, 0) = 0.64\,\mathrm{N}$ and gray level noise standard deviation $\sigma_{\mathrm{noise}} = 2\sqrt{2}\,\mathrm{px}$ studied in Sec. 3.5.3. The Gaussian filter standard deviation is set to $\sigma_{\mathrm{smooth}} = 30\,\mathrm{px}$, which is approximately equal to the standard deviation of the Gabor filters at the last scale. The smoothing result is represented in Fig. 3.8 with respect to space and time. This preprocessing improves both the spatial and the temporal consistency of the estimated motion.

The stabilization diagrams obtained without and with our preprocessing are compared in Fig. 3.9 We can see that the stabilization diagram with preprocessing still contains a relevant

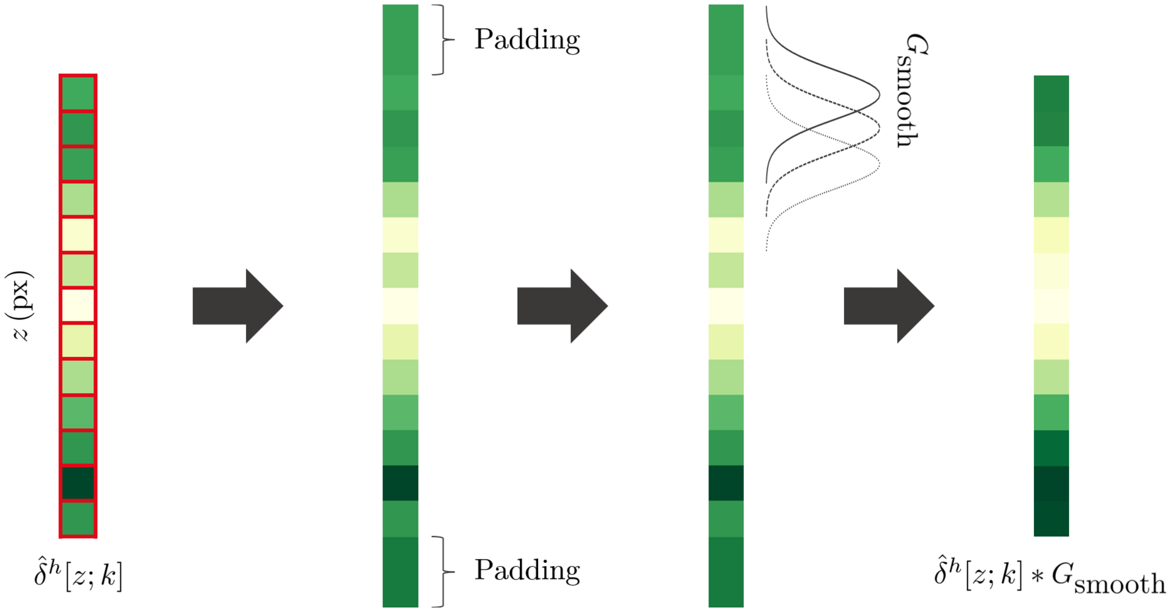

Figure 3.7: Smoothing process (of the estimated horizontal motion).

number of stable poles near each natural frequency. Furthermore, the spurious series of stable poles seem to be largely removed thanks to the preprocessing. The respective errors are listed in Tab. 3.4. The frequency and damping errors obtained with preprocessing are close to those obtained without preprocessing, except for the damping ratio of the first mode. Finally, the preprocessing noticeably improves the mode shapes, the fourth mode error being divided by more than 3 (see Fig. 3.10).

Criterion	Frequency relative error (%)				Damping ratio relative error (%)				Mode shape error (1-MAC)			
Error	$\epsilon\{\hat{f}_1\}$	$\epsilon\{\hat{f}_2\}$	$\epsilon\{\hat{f}_3\}$	$\epsilon\{\hat{f}_4\}$	$\epsilon\{\hat{\zeta}_1\}$	$\epsilon\{\hat{\zeta}_2\}$	$\epsilon\{\hat{\zeta}_3\}$	$\epsilon\{\hat{\zeta}_4\}$	$\epsilon\{\hat{\phi}_1\}$	$\epsilon\{\hat{\phi}_2\}$	$\epsilon\{\hat{\phi}_3\}$	$\epsilon\{\hat{\phi}_4\}$
Without	0.044	0.057	**0.049**	**0.094**	**15.972**	**9.702**	**7.526**	**7.358**	**0.001**	0.061	0.089	0.246
With	**0.021**	**0.045**	0.084	0.119	71.217	11.323	8.043	13.272	**0.001**	**0.029**	**0.056**	**0.075**

Table 3.4: Modal basis estimation errors without and with estimated motion preprocessing.

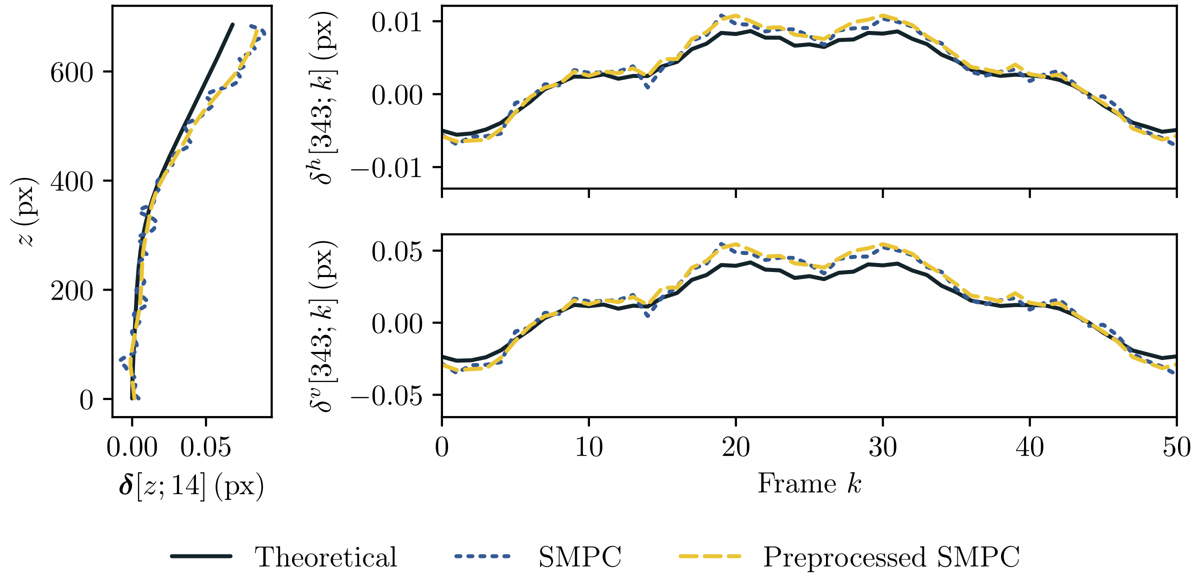

Figure 3.8: Theoretical, SMPC, and preprocessed SMPC displacement in frame $k = 14$ (left), and over the first 50 frames at the middle of the beam $z = 343$ (right).

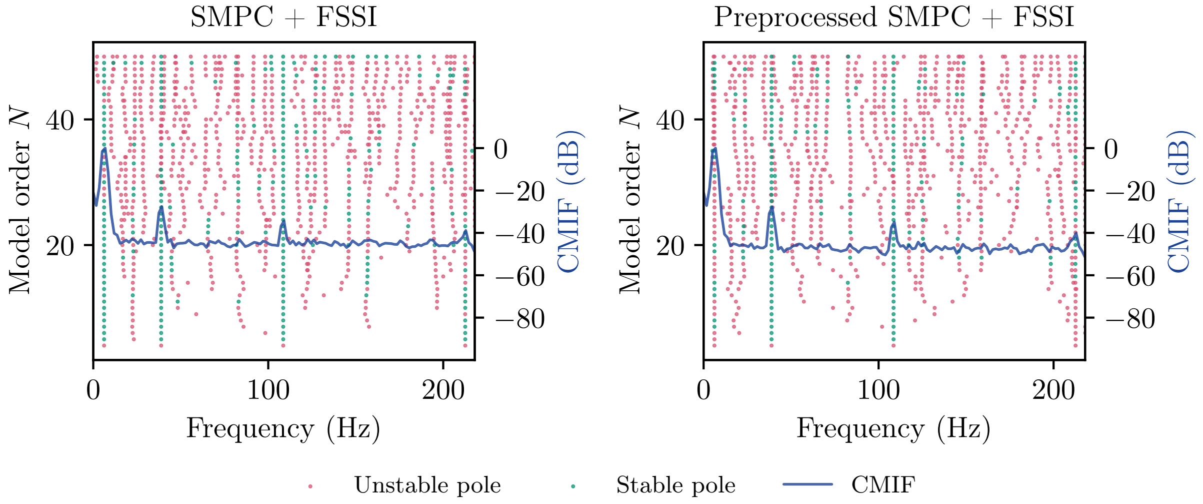

Figure 3.9: Stabilization diagrams obtained without and with preprocessing of the SMPC estimated motion.

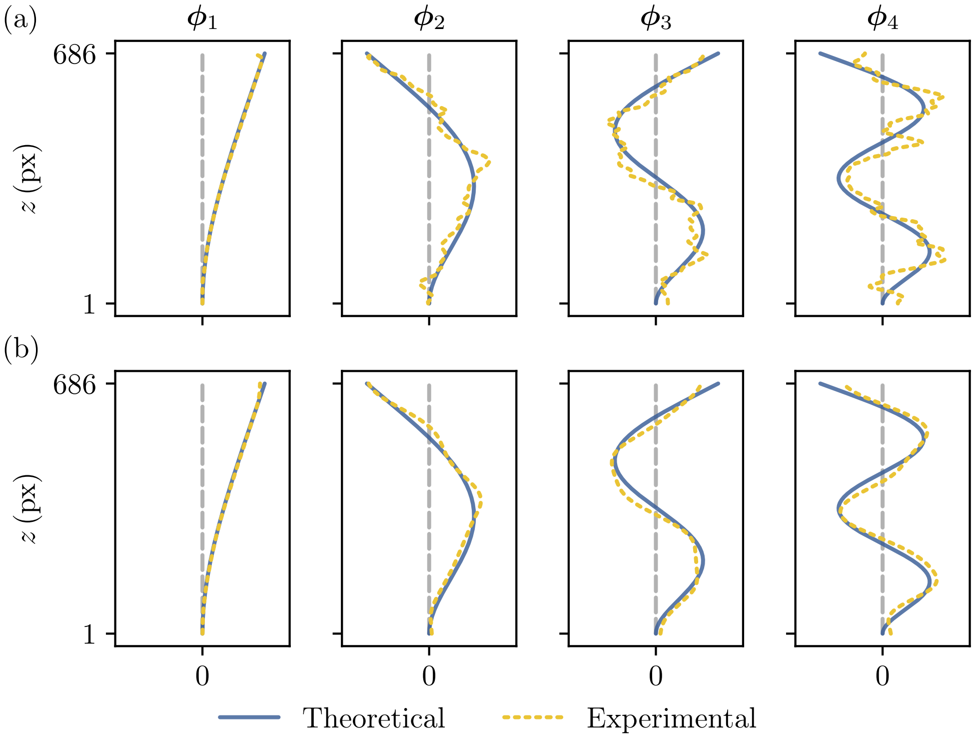

Figure 3.10: Mode shapes obtained without (a) and with (b) preprocessing of the SMPC estimated motion.

3.6.2 Sensitivity against additive noise

To further assess the improvement brought by the preprocessing, we study the fourth mode, which has the lowest amplitude, for different additive gray level noise standard deviation σ_{noise}. For this purpose, we compute the natural frequency, damping ratio, and mode shape error with respect to the noise level. The results gathered in Fig. 3.11 show that the natural frequency errors are below 0.3 % with or without smoothing. The damping ratios errors with smoothing are lower than those without smoothing for three of the six noise levels. For some cases, these errors are significant because they are computed relatively to the low damping value of the fourth mode ($1.3 \cdot 10^{-3}$). Finally, the mode shape error increases with respect to σ_{noise} but preprocessing the estimated motion reduces it whatever the noise level.

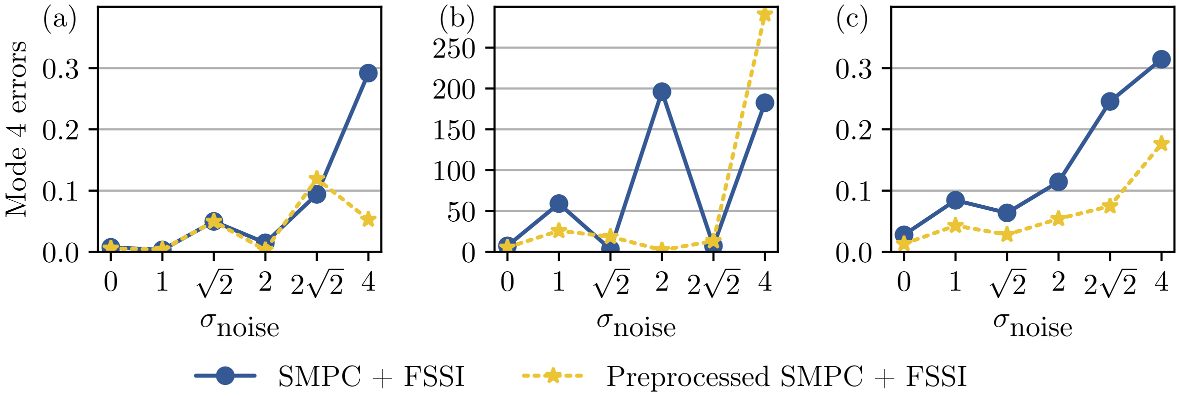

Figure 3.11: Fourth mode errors for natural frequency (a), damping ratio (b), and mode shape (c) with respect to noise standard deviation σ_{noise}.

3.6.3 Sensitivity against mode amplitude

To assess the sensitivity of video-based modal analysis against motion amplitude, four noiseless videos with respective input force $\gamma(L, 0) \in \{0.08, 0.16, 0.32, 0.64\}$ (N) are generated. For each input force, the displacement amplitude of each mode is reported in Tab. 3.5. From each video, the motion is estimated by SMPC, and the modal analysis is performed by FSSI. Each mode

Mode / Force $\gamma(L, 0)$	1	2	3	4
0.64 N	0.22759	0.02244	0.00830	0.00384
0.32 N	0.11380	0.01122	0.00415	0.00192
0.16 N	0.05690	0.00561	0.00208	0.00096
0.08 N	0.02845	0.00280	0.00104	0.00048

Table 3.5: Displacement amplitude (px) of each mode in function of input force.

is automatically estimated thanks to the stabilization diagram, as described in Sec. 3.5.1. The mode shape errors obtained with or without preprocessing are plotted with respect to the mode amplitude in Fig. 3.12. None of the two methods finds the last mode when the input force is $\gamma(L, 0) = 0.08\,\mathrm{N}$. Indeed, the maximum amplitude of $4.8 \cdot 10^{-4}$ px in the video is too small to detect the beam vibration. The figure also shows that the error with preprocessing is lower than 0.05 when mode amplitude is larger than 0.003 px. Furthermore, our preprocessing systematically reduces the error for all the considered mode amplitudes. However, the last mode with a maximum amplitude of $9.6 \cdot 10^{-4}$ px is not retrieved when the preprocessing is applied. The small motion associated to this mode is removed by the Gaussian smoothing.

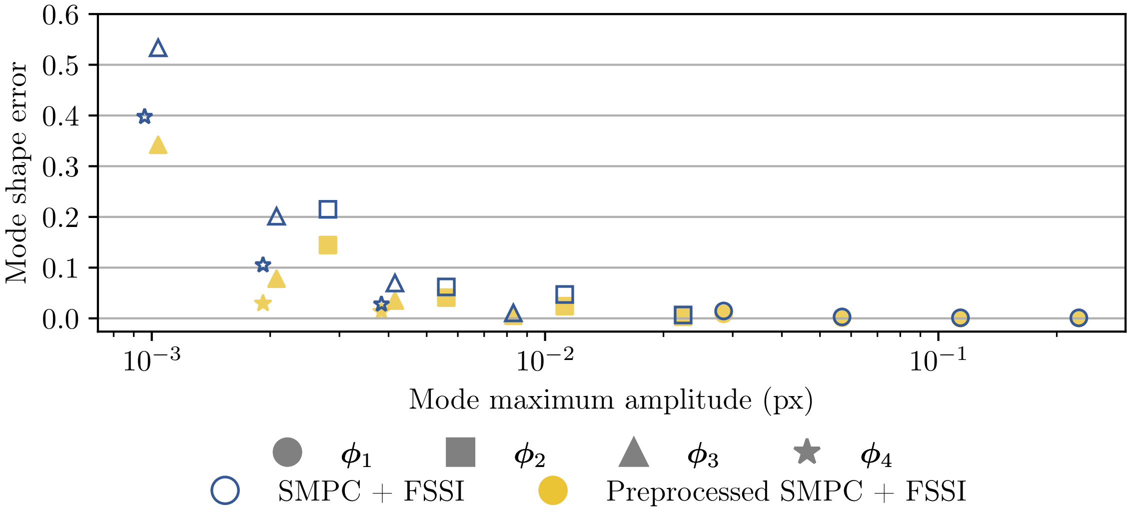

Figure 3.12: Mode shape error with respect to its amplitude.

3.7 Conclusion

In this chapter, we describe different modal analysis methods. State-of-the-art methods based on assumptions about single and multiple degrees of freedom are presented, and two dense video-based methods are detailed. The first one, PCA+CP, uses principal component analysis to reduce dimension of measurements and performs blind source separation using complexity pursuit. The second one, FSSI, is an adaptation of the stochastic subspace identification that also uses a reduction of the observation dimension.

The performances reached by the two methods are compared on a synthetic video of a cantilever beam, generated as detailed in chapter 2. The methods estimate the system poles for different modal orders in order to build a stabilization diagram. Stable poles are automatically extracted from this diagram, which requires the knowledge of the theoretical natural frequencies. When the latter are unavailable, the poles can be extracted by using automatic parameter-less clustering methods [38]. Our FSSI method provides a better stabilization diagram with less spurious modes, and gives better estimation of the modal basis.

We also compare the performances of FSSI operational modal analysis when different phase-based motion estimation methods are used. The tested methods are those described in chapter 2. This confirms that our SMPC method provides the best results.

A preprocessing based on a Gaussian smoothing of the estimated motion is then proposed. This improves modal analysis performances and reduces the mode shape error by up to 70%.

To highlight the limits of the approach, the robustness against additive noise of the video gray levels and against motion amplitude is assessed. The proposed method is robust against strong noise and successfully estimates modes whose amplitude is higher than 0.003 px in the video.

In this chapter, video-based modal analysis is assessed on synthetic videos of a vibrating cantilever beam. Chapter 4 details the experimental study of a real cantilever beam and a bent beam whose vibrations are observed by a high-speed camera in controlled conditions.

4 Application to cantilever beam monitoring

4.1 Introduction

In the previous chapters, phase-based motion estimation and video-based modal analysis are assessed on synthetic videos of a cantilever beam excited by an impact hammer. This chapter is dedicated to the operational modal analysis (OMA) of two cantilever beams observed in controlled conditions and excited by a shaker that provides a white noise input force spectrum. First, the experimental setup of a straight beam is presented in Sec. 4.2.1. To provide ground truth data, an accelerometer-based OMA is conducted, and the estimated mode shapes are compared with the Euler–Bernoulli theoretical ones in Sec. 4.2.2. Among the different subpixel phase-based motion estimation methods tested on synthetic videos, the SMPC method (see Sec. 2.5.3) provides the best results. It is thus retained as our phase-based OMA method in Sec. 4.2.3 and is compared to the simple and state-of-the-art optical flow method (see Sec. 1.5.3.2). This intensity-based method quickly estimates subpixel motion and provides good results for modal identification on noisy videos [20]. The motion estimated by the two video-based methods are compared with accelerometer data in the frequency domain, and their modal basis errors are assessed with regard to the accelerometer-based ones. Finally, as the beams studied so far are straight, they move along one single main direction. A similar study is thus performed on a bent beam in Sec. 4.3 to test the methods on a video with several object motion directions.

4.2 Modal analysis of a cantilever straight beam

To assess the video-based modal analysis, an experimental study with a straight beam observed under controlled conditions is conducted in this section. The modal basis of the cantilever beam is first estimated using accelerometers to get ground truth data. The modal basis is then estimated using our video-based method and is thus compared to that obtained with accelerometers.

4.2.1 Experimental setup

In our experimental setup, a steel beam with rectangular section is clamped at its bottom using a vice such that the free part of the beam measures $0.930 \times 0.030 \times 0.006\,\mathrm{m}$ (see Fig. 4.1). Eight monoaxial accelerometers[1] are fixed to the beam with adhesive wax at the locations 0.09, 0.22, 0.34, 0.47, 0.59, 0.71, 0.83, and 0.91 m to get ground-truth vibration values. The beam is excited punctually at 0.18 m from the clamped end using an electromagnetic vibration generator as shaker[2]. The shaker and the eight accelerometers are connected to a data acquisition system[3] that generates the input white noise of the shaker with 2 V input voltage and converts analog

---

[1] Brüel & Kjaer DeltaTron accelerometer Type 4517

[2] Frederiksen 218500

[3] Dewesoft SIRIUS-HS-ACC8

acceleration values into digital data. We assume that the frequency response of the shaker is constant and that the beam is thus excited by a white noise.

In front of the scene, a camera[4] equipped with a lens of 8 mm focal length is positioned at 1.6 m from the beam. Three mainstream LED bulbs, with a color temperature of 2700 K, are placed close to the beam so that sufficient light energy can be integrated by the camera sensor with low integration time. To ensure that the main beam motion is linearly projected onto the focal image, the camera is oriented such that the image plane is parallel to the face of the beam that measures 0.930×0.006 m. The camera is rotated along the optical axis by an angle of $9.65\,^{\circ}$ to be close to the conditions of the synthetic videos generated in Sec. 2.7. The camera sensor contains 720×540 pixels whose size is $6.9 \cdot 10^{-6} \times 6.9 \cdot 10^{-6}$ m. The size of the region of interest is set to 720×140 px and the frame rate to 500 fps so that the frequency band of interest ranges from 0 to 250 Hz. The pixels of interest (PoI) are selected as a straight line of pixels on the beam (see Fig. 4.2). To facilitate the analysis, the video is processed thanks to a user interface that has been developed during this work, whose main features are summarized in Appendix B.

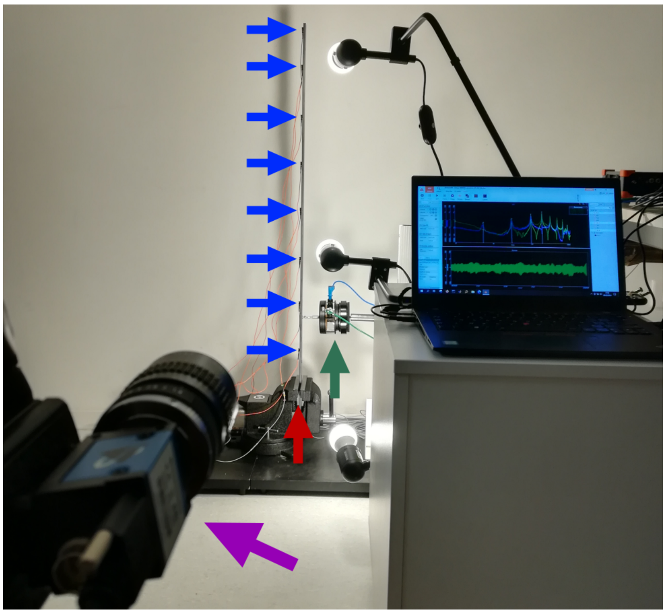

Figure 4.1: Experimental setup for the straight beam with camera (purple arrow), beam (red arrow), accelerometers (blue arrow), and vibration shaker (green arrow).

4.2.2 Accelerometer-based modal analysis

The sampling frequency of the accelerometers is set to 1400 Hz so that the frequency band of interest ranges from 0 to 700 Hz. To estimate the modal basis, the acceleration values recorded

---

[4] TIS DMK 33UX287

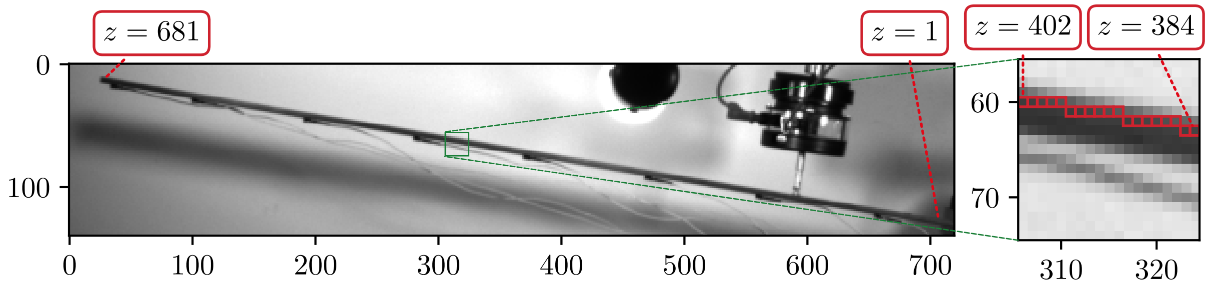

Figure 4.2: Frame from the acquired video of the straight beam (left) and pixels of interest displayed as red (right).

by the eight accelerometers are used as observations. The stochastic subspace identification (SSI) method detailed in Sec. 3.3.3 is performed with the number R of block rows set to 40 and the model order $N \in [\![2, 50]\!]$.

A stability analysis (see Sec. 3.5.1) is conducted to estimate the modal basis manually. For this purpose, a stabilization diagram is built, on which the first complex mode indicator function (CMIF) is shown (see Fig. 4.3).

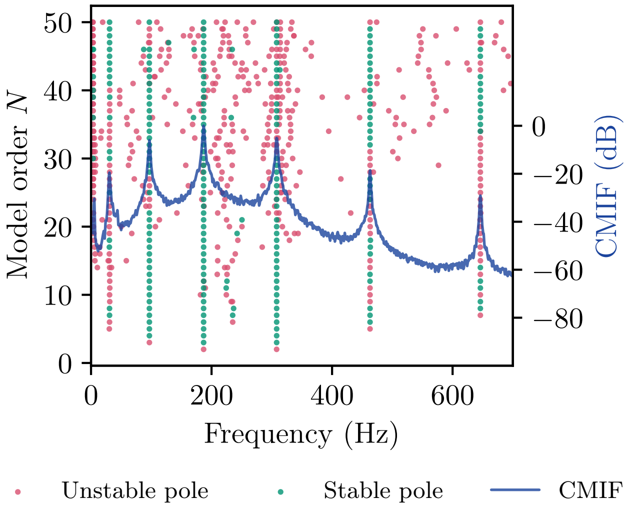

Figure 4.3: Stabilization diagram of the accelerometer-based OMA of the straight beam.

In Fig. 4.3, seven peaks can be identified from the CMIF at frequencies around 5, 31, 97, 188, 308, 462, and 647 Hz. Furthermore, there is a vertical series of stable poles indicating the presence of a mode close to each of these frequencies, except at 5 Hz. A frequency is selected manually by zooming in on the stabilization diagram, and all the stable poles around this frequency (±1 %) are automatically extracted. The mode is then estimated by computing the mean of the extracted poles in terms of natural frequencies, damping ratios and mode shapes.

The first mode (close to 5 Hz) cannot be estimated because no stable pole is sufficiently close to this frequency. This is because the accelerometers do not provide any reliable data below 10 Hz. Hence, the first mode frequency is estimated by computing the peak location of the first CMIF interpolated thanks to cubic spline. The obtained natural frequencies and damping ratios are listed in Tab. 4.1.

Mode i	1	2	3	4	5	6	7
f_i (Hz)	5.361	30.451	96.834	186.839	308.037	463.040	646.117
ζ_i ($\times 10^2$)	–	3.311	1.003	0.578	0.335	0.164	0.087

Table 4.1: Natural frequencies and damping ratios of the straight beam modes obtained from the accelerometer-based OMA.

To ensure that the accelerometer-based OMA provides a reliable modal basis, it is compared to the Euler–Bernoulli model detailed in Sec. 2.6.1. As the theoretical model does not consider any damping and requires the unknown Young modulus of the beam material to estimate natural frequencies, only the mode shapes can be computed. Figure 4.4 shows a comparison between the six mode shapes obtained from the accelerometer-based OMA and the theoretical ones from the Euler–Bernoulli model. As mode shapes are defined up to a factor, they are normalized with their L^2-norm to be compared with similar amplitudes. The mode shapes obtained with data acquired by accelerometers are close to the theoretical ones.

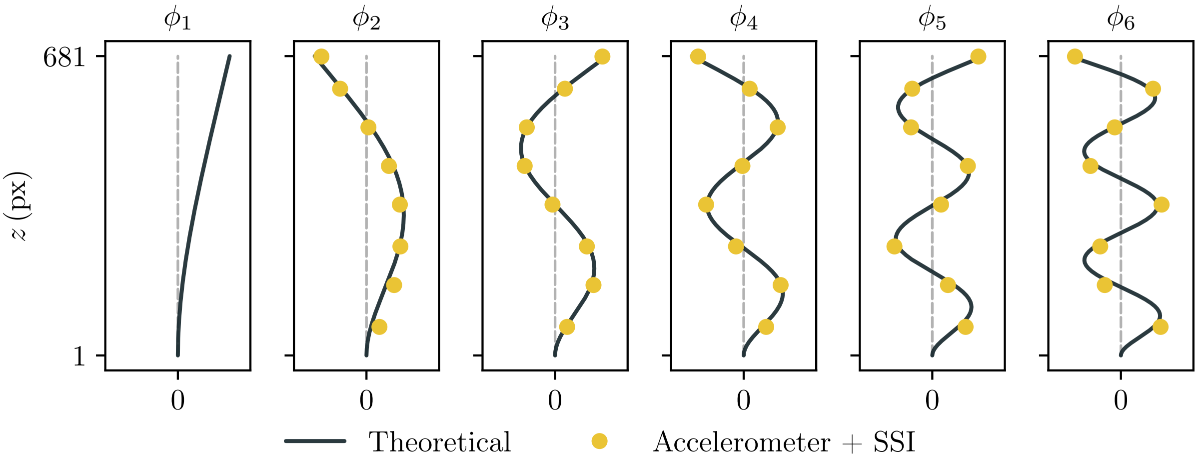

Figure 4.4: Comparison between Euler–Bernoulli theoretical mode shapes and accelerometer-based ones for the straight beam.

To quantify the mode shape errors, both theoretical mode shapes $\boldsymbol{\phi}_i$ and experimental ones $\hat{\boldsymbol{\phi}}_i \in \mathbb{R}^8, i \in [\![1,7]\!]$, are extracted at the accelerometer locations, and the shape errors are computed as:

$$\epsilon\{\hat{\phi}_i\} = 1 - \mathrm{MAC}(\hat{\phi}_i, \phi_i), \tag{4.1}$$

with the modal assurance criterion (MAC) defined by Eq. (3.39). The mode shapes errors, listed in Tab. 4.2, are lower than 0.05. These values are close to the mode shape stability criterion $\epsilon\{\hat{\phi}_i\} < 0.02$ of the stabilization diagram, which highlights the strong correlation between accelerometer-based and theoretical mode shapes. The errors can be attributed to the inaccurate position and orientation of accelerometers. Thus, the accelerometer-based modal basis can be considered as ground truth data to assess the performances of the video-based methods.

Mode i	1	2	3	4	5	6	7
Error $\epsilon\{\hat{\phi}_i\}$	–	0.031	0.004	0.011	0.021	0.032	0.046

Table 4.2: Accelerometer-based mode shape errors for the straight beam.

4.2.3 Video-based modal analysis

A video of the experimental scene is recorded during 8 s at 500 fps. To estimate the beam motion from this video, each of the 4000 frames is decomposed using Gabor filters with $N_r = 5$ scales and $N_\theta = 8$ orientations. The SMPC method detailed in Sec. 2.5.3 is then performed to estimate the motion with a phase confidence threshold $\tau = 0.5$ and a spatial Gaussian kernel $\mathcal{G}$ with standard deviation of 3 px.

4.2.3.1 Gradient-based optical flow (GOF)

To compare our phase-based method with an intensity-based method, we also estimate motion thanks to the gradient-based optical flow (GOF) method [21] (see Sec. 1.5.3.2) by:

$$\hat{\boldsymbol{\delta}}[\boldsymbol{p};k] = \frac{\bar{\tilde{I}}[\boldsymbol{p}] - \tilde{I}[\boldsymbol{p};k]}{|\boldsymbol{\nabla}\bar{\tilde{I}}[\boldsymbol{p}]|} \cdot \begin{pmatrix} \cos\left(\angle_{\boldsymbol{\nabla}\bar{\tilde{I}}[\boldsymbol{p}]}\right) \\ \sin\left(\angle_{\boldsymbol{\nabla}\bar{\tilde{I}}[\boldsymbol{p}]}\right) \end{pmatrix}, \tag{4.2}$$

with $\tilde{I}[\boldsymbol{p};k]$ the filtered gray level intensity computed by the spatial 3×3 average filter, $\bar{\tilde{I}}[\boldsymbol{p}]$ the temporal mean of the filtered gray level intensity, and $\boldsymbol{\nabla}\bar{\tilde{I}}[\boldsymbol{p}]$ its spatial gradient with angle $\angle_{\boldsymbol{\nabla}\bar{\tilde{I}}[\boldsymbol{p}]}$. The method is low time-consuming and produces satisfying subpixel estimations for modal identification in noisy videos [20]. As this method only provides relevant results for

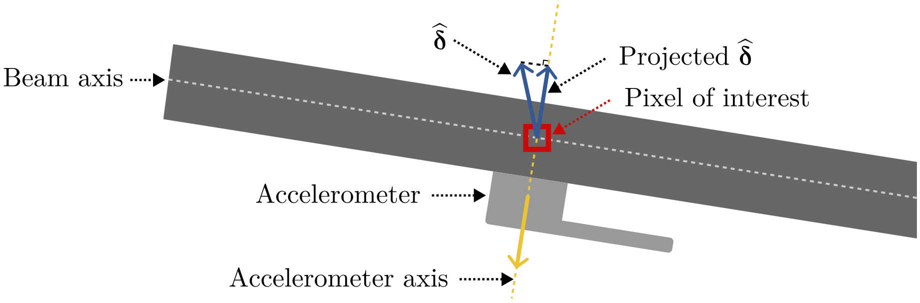

Figure 4.5: Accelerometer axis (yellow) and estimated and projected motion at the pixel of interest (blue) in the image plane.

region with sufficient intensity gradient, its pixels of interest are shifted vertically with respect to SMPC ones to cover the border of the beam.

In each frame k, the motion estimated by both methods is extracted at the curvilinear coordinate z of the pixel of interest to obtain $\hat{\boldsymbol{\delta}}[z;k]$. The smoothing preprocessing detailed in Sec. 3.6 is applied to both horizontal and vertical components of $\hat{\boldsymbol{\delta}}$ with a standard deviation $\sigma_{\text{smooth}} = 30\,\text{px}$ to obtain our SMPC estimated motion $\tilde{\boldsymbol{\delta}}[z;k]$. The subpixel displacement amplitude estimated by SMPC at the free end of the beam is 0.218 px, which is approximately that obtained on the synthetic video with $\gamma(L,0) = 0.64\,\text{N}$ (see Sec. 2.7).

To estimate motion from our straight beam video, the GOF and SMPC respectively require 11.65 s and 4671.72 s on a computer equipped with an Intel(R) Xeon(R) Silver 4216 CPU and 64 GB of RAM. The simple GOF method is thus 400 times faster than SMPC. Note that both method implementations could be speeded up by computing the motion only at the pixels of interest and by taking advantage of processing parallelization.

4.2.3.2 Video-based motion fidelity to accelerometer-based motion

For a first comparison between the video-based estimated displacement and accelerometer data, we focus on the value at the closest pixel of interest from the accelerometer that is located at the free end of the beam, where the motion is the largest. As the accelerometer is monoaxial, the horizontal and vertical components of the estimated displacement at this pixel are projected on the accelerometer axis in the frame plane (see Fig. 4.5). Then, the obtained displacement is temporally derived twice to get acceleration. Assuming that the accelerometer axis is in a plane parallel to the image plane, the accelerometer data are finally converted from $\text{m} \cdot \text{s}^{-2}$ to $\text{px} \cdot \text{s}^{-2}$ to be comparable with the video-based acceleration. The conversion multiplication factor is the pixel resolution, *i.e.*, the ratio between the length of the beam in the frame and the actual beam length.

Because video and accelerometer data are not synchronized and are acquired according to different sampling frequencies, we compare their accelerations in the frequency domain via their power spectral density (PSD) (see Fig. 4.6). The SMPC- and accelerometer-based amplitudes are similar, with peaks at four different frequencies indicated by green arrows. The GOF PSD fits the accelerometer-based PSD at frequencies below 98 Hz but suffers from intensity frequency perturbation at 100 and 200 Hz (indicated by red arrows).

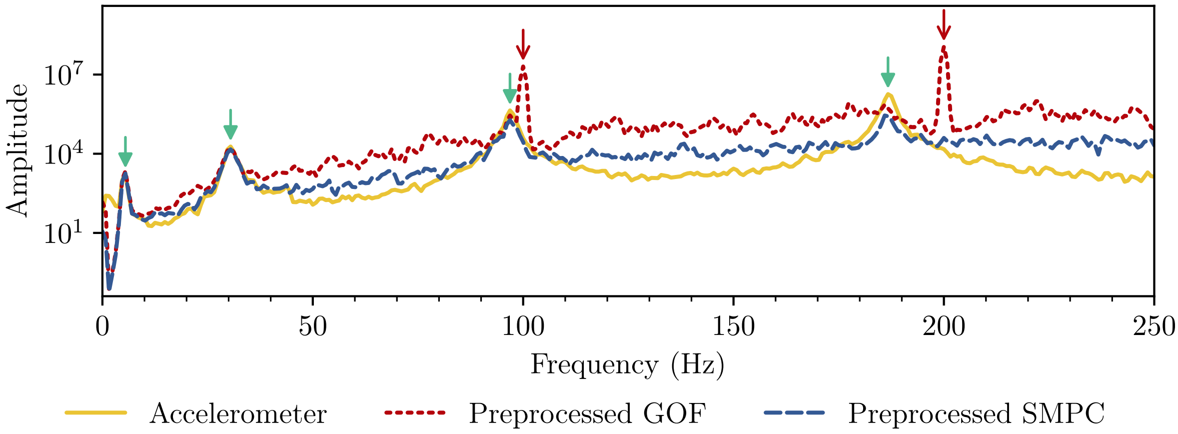

Figure 4.6: Comparison of acceleration power spectral density at the free end of the straight beam according to accelerometer-based data, video-based SMPC, and video-based GOF.

To identify the cause of this perturbation, we extract the PSD of the displacement along the direction normal to the beam, by rotating the displacement components back according to the angle of the camera at each curvilinear coordinate z, for GOF and SMPC methods. As shown at the bottom of Fig. 4.7 for SMPC, four physical mode shapes and their corresponding nodes can be identified by the four vertical lines with high PSD amplitude and low amplitude at the nodes (in dotted circles). However, at 100 and 200 Hz, the PSD amplitude is high (indicated by red arrows) for pixels $z \in [\![200, 300]\!]$. These pixels correspond to the surface elements of the beam that are close to the light bulb visible in the video, as in Fig. 4.2. For GOF PSD (see top of Fig. 4.7), the first three mode shapes are identifiable, but the fourth is barely visible. The spurious peaks at 100 and 200 Hz are present for every pixel, which suggests that GOF is less tolerant than SMPC to flickering intensity perturbation.

To further investigate the cause of the perturbation, the gray level intensity at a pixel between the light bulb and the beam is displayed in time and frequency in Fig. 4.8. The same two peaks at 100 and 200 Hz are present in the frequency domain. This confirms that the perturbation is caused by the flickering frequency of the light bulb. Indeed, such flickering generates gray level variations in the frame pixel values that are interpreted as motion. This phenomenon is mentioned by other works as a limit of video-based methods [5, 9].

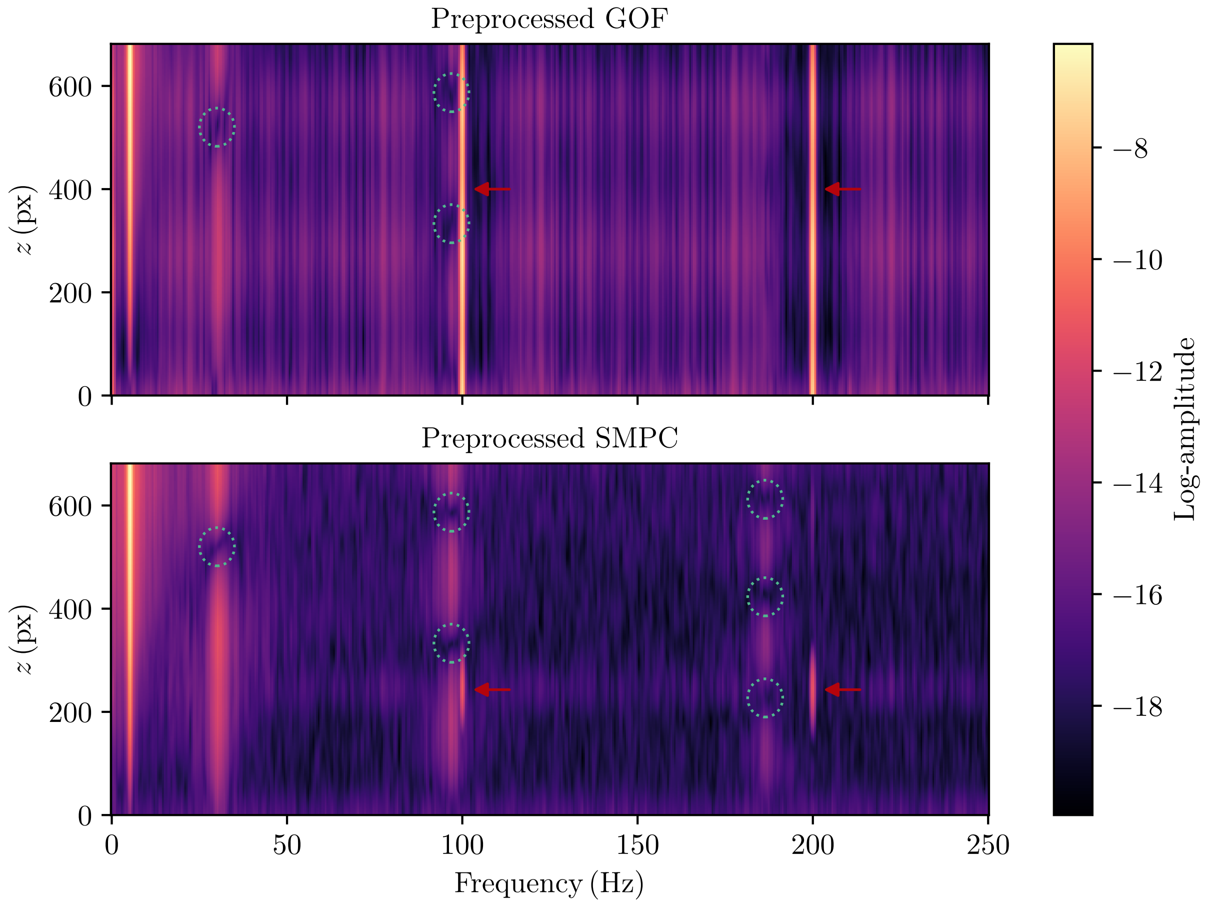

Figure 4.7: Power spectral density amplitude of the video-based motion estimations at each pixel of interest of the straight beam.

4.2.3.3 Modal basis accelerometers vs video

The motion estimated by each method is the input of the FSSI video-based modal analysis procedure detailed in Sec. 3.4 to estimate the modal basis. First, the dimension reduction is performed by keeping the $P = 10$ first principal components and the block Toeplitz matrix is constructed with $R = 40$ block rows. The stabilization diagram is built with the same model order range $N \in [\![2, 50]\!]$ as for the accelerometers. The GOF and SMPC stabilization diagrams are represented in Fig. 4.9. The diagrams on the left correspond to the standard stabilization diagrams with all poles, and those on the right only contain the stable poles retained for the modal basis estimation.

The CMIF deduced from the GOF method has four peaks at the frequencies: 5, 30, 100, and 200 Hz. A small peak of the CMIF and a series of stable pole is also present at 97 Hz, suggesting the presence of a mode. Furthermore, a series of stable poles is present at 186 Hz and is thus

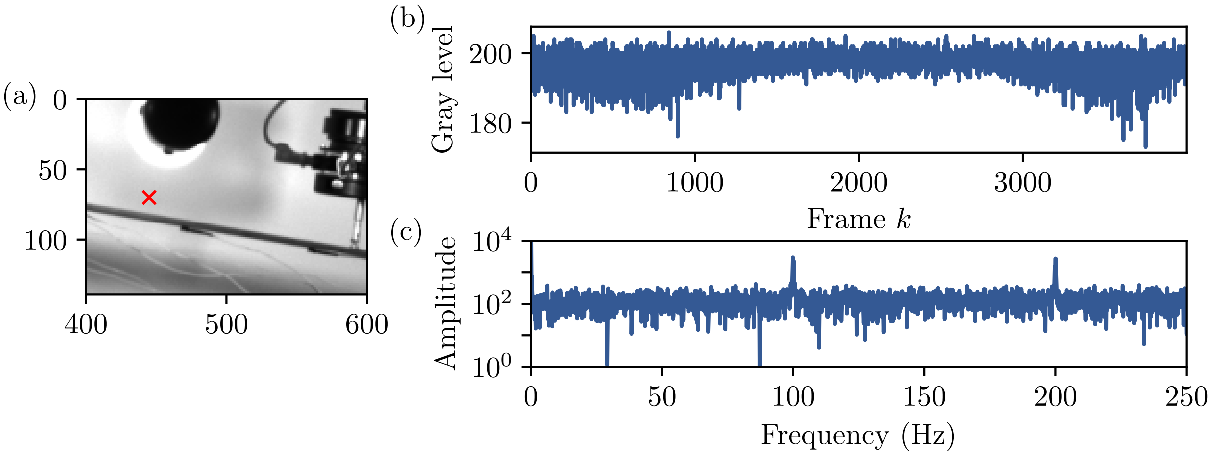

Figure 4.8: Intensity gray level (b) and its Fourier transform amplitude (c) at the pixel represented in red (a).

studied to compare it with SMPC results. On the stabilization diagram constructed with the preprocessed SMPC method, six CMIF peaks are identifiable at the frequencies: 5, 30, 97, 100, 186, and 200 Hz. There is a vertical series of stable poles at each of these frequencies, suggesting the presence of a mode. As 100 and 200 Hz correspond to the light bulb perturbation frequencies, they are discarded. For each frequency of interest, a stable pole is manually selected, and the modal basis is determined by the same procedure as for accelerometer-based data.

To assess the performances reached by the video-based OMA methods, the four physical mode shapes obtained in the frequency range $[0, 250]$ Hz are compared with the accelerometer modal basis. The natural frequencies and damping ratios deduced from GOF and SMPC are compared with accelerometer-based ones in Tab. 4.3. The GOF and SMPC mode shapes are compared with Euler–Bernoulli theoretical mode shapes in Fig. 4.10. We can see that the FSSI modal analysis with preprocessed SMPC provides a good estimation of the four modes. The last mode shape provided by FSSI with preprocessed GOF is strongly biased, which again suggests that GOF motion estimation method is less robust against perturbations.

Criterion	Natural frequency (Hz)				Damping ratio ($\times 10^2$)			
Method	$\hat{f}_1$	$\hat{f}_2$	$\hat{f}_3$	$\hat{f}_4$	$\hat{\zeta}_1$	$\hat{\zeta}_2$	$\hat{\zeta}_3$	$\hat{\zeta}_4$
Accelerometer	5.361	30.451	96.834	186.839	–	3.311	1.003	0.578
GOF	5.275	30.812	96.903	186.296	0.397	3.470	1.085	0.906
SMPC	5.276	30.691	97.010	186.416	0.453	2.673	1.057	0.538

Table 4.3: Accelerometer- and video-based natural frequencies and damping ratios of the straight beam modes.

The video-based OMA performances are objectively assessed by quantitatively comparing the modal basis. For this purpose, we compute the video mode shape error with respect to

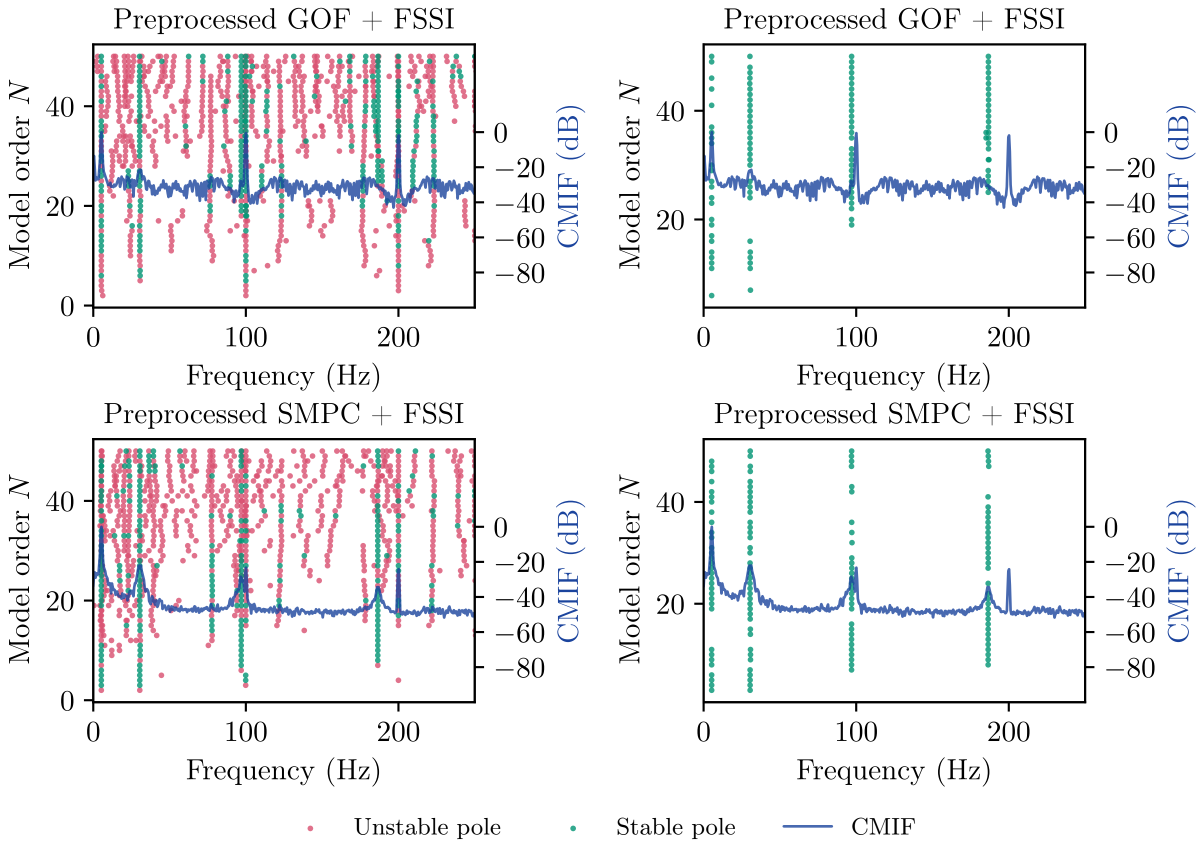

Figure 4.9: Stabilization diagrams of the video-based OMA methods for the straight beam. All poles are shown in the left column and only selected stable poles are shown in the right column.

theoretical ones using Eq. (4.1) and the frequency and damping ratios relative errors with respect to accelerometer-based ones as follows:

$$\begin{cases} \epsilon\{\hat{f}_i\} = 100 \cdot \dfrac{\left|\hat{f}_i - f_i\right|}{f_i} \\ \epsilon\{\hat{\zeta}_i\} = 100 \cdot \dfrac{\left|\hat{\zeta}_i - \zeta_i\right|}{\zeta_i} \end{cases} . \tag{4.3}$$

The errors are reported in Tab. 4.4. Both methods provide good estimators of natural frequencies with relative errors below 2 %. Furthermore, the errors of damping values are below 21 %. The FSSI with preprocessed GOF successfully determines the first three shapes but fails to estimate the fourth mode with an error of about 0.6. Finally, the four mode shapes determined by the

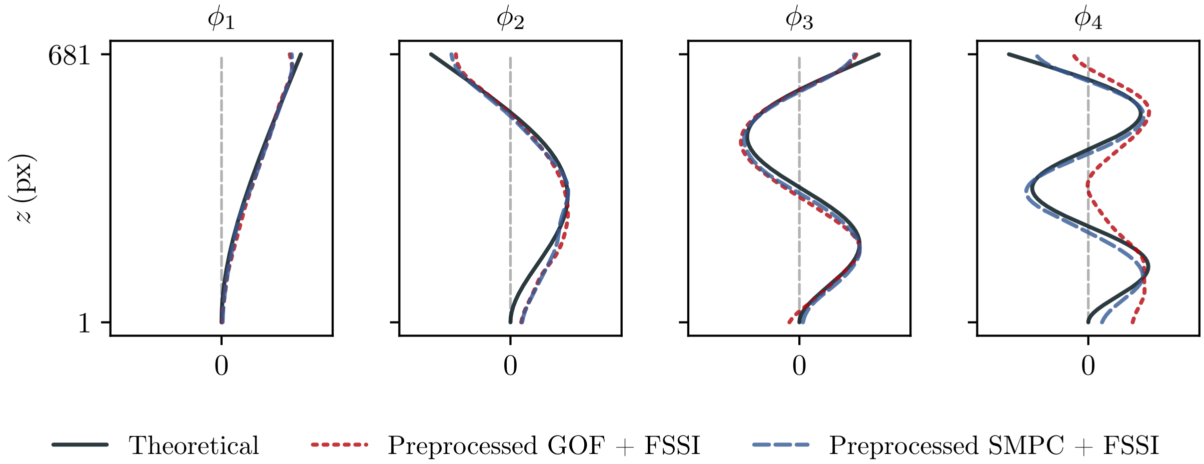

Figure 4.10: Comparison between Euler–Bernoulli theoretical model and video-based mode shapes for the straight beam.

FSSI with preprocessed SMPC are close to the theoretical mode shapes, with errors lower than 0.06. The FSSI with preprocessed SMPC thus provides the best results.

Criterion	Frequency relative error (%)				Damping ratio relative error (%)				Mode shape error (1-MAC)			
Method	$\epsilon\{\hat{f}_1\}$	$\epsilon\{\hat{f}_2\}$	$\epsilon\{\hat{f}_3\}$	$\epsilon\{\hat{f}_4\}$	$\epsilon\{\hat{\zeta}_1\}$	$\epsilon\{\hat{\zeta}_2\}$	$\epsilon\{\hat{\zeta}_3\}$	$\epsilon\{\hat{\zeta}_4\}$	$\epsilon\{\hat{\phi}_1\}$	$\epsilon\{\hat{\phi}_2\}$	$\epsilon\{\hat{\phi}_3\}$	$\epsilon\{\hat{\phi}_4\}$
GOF	**1.612**	**0.076**	**0.071**	0.291	–	**4.209**	12.333	20.515	**0.003**	**0.031**	0.036	0.596
SMPC	1.666	0.317	0.189	**0.190**	–	20.786	**5.383**	**6.944**	**0.003**	0.040	**0.025**	**0.058**

Table 4.4: FSSI modal basis estimation errors with GOF and SMPC motion estimation methods for the straight beam.

The SMPC video-based method provides a coarse estimation of the modal basis. Indeed, most of the errors are above the classical acceptability criteria, namely:

- frequency relative error below 1 %,
- damping ratio relative error below 5 %,
- mode shape error below 0.02.

The frequencies are well estimated, the first one being probably due to the poor reference estimation from coarse accelerometer data. However, all the damping ratio errors range between 5 and 21 %, and the shape errors of three modes are above 0.02. The estimation could potentially be improved by acquiring a longer video. Indeed, the SSI method is based on correlation estimation, which is improved when the duration of the data acquisition is increased.

4.3 Modal analysis of a cantilever bent beam

In the first experimental study, the performances of video-based OMA methods are assessed on an experimental video of a straight cantilever beam. However, the motion of such a structure is one-directional and the direction is known.

Let us now consider a realistic case with several motion directions. For this purpose, accelerometer- and video-based OMAs of a beam with a bend are compared. Indeed, the bend introduces a second main direction of motion. Moreover, the direction strongly varies in the region close to the bend. Furthermore, the camera frame rate is increased comparatively to the previous study to extend the frequency range and capture more modes. Our phase-based motion estimation method is once again compared with the intensity-based motion estimation method.

4.3.1 Experimental setup

The cantilever beam properties are the same as in the previous section, but it is bent such that it forms an angle of 143 ° (see Fig. 4.11). The bend is located at 0.43 m from the bottom clamped end.

Eight accelerometers are fixed on the beam to serve as ground-truth vibration data. They are located at 0.10, 0.20, 0.30, 0.40, 0.46, 0.61, 0.76, and 0.90 m from the clamped end of the beam, and oriented to measure the normal acceleration of the beam. Their sampling frequency is still set to 1400 Hz. As in Sec. 4.2.1, the beam is excited by a shaker fixed at 0.18 m from the clamped end and the data acquisition system is configured to deliver white noise with 2 V input voltage to the shaker.

The camera is located at 1.6 m from the beam and observes its thin side. A region of interest of 720×260 px is selected and 628 PoI are defined along the beam to extract the estimated motion (see Fig. 4.12). The beam bend is located at the PoI $z = 315$. The video lasts 8 s with a frame rate of 1000 fps, so that the frequency band of interest is $[0, 500]$ Hz and $N_k = 8000$ frames are acquired.

The same video-based motion estimation methods as in the previous section are compared, namely the state-of-the-art GOF intensity-based method and our SMPC phase-based method. This GOF intensity-based method provides two-dimensional subpixel motion estimations by assuming that motion occurs along the direction of the local intensity gradient at each pixel. The PoI represented in Fig. 4.12 are slightly shifted vertically for this method to match the beam edge characterized by higher gradient values.

As in the previous study, the phase-based SMPC method is retained because it provides the best results for motion estimation of the synthetic videos of a cantilever beam (see Sec. 2.7). Furthermore, this method analyzes several subband orientations and is thus well-suited for multi-directional motion analysis.

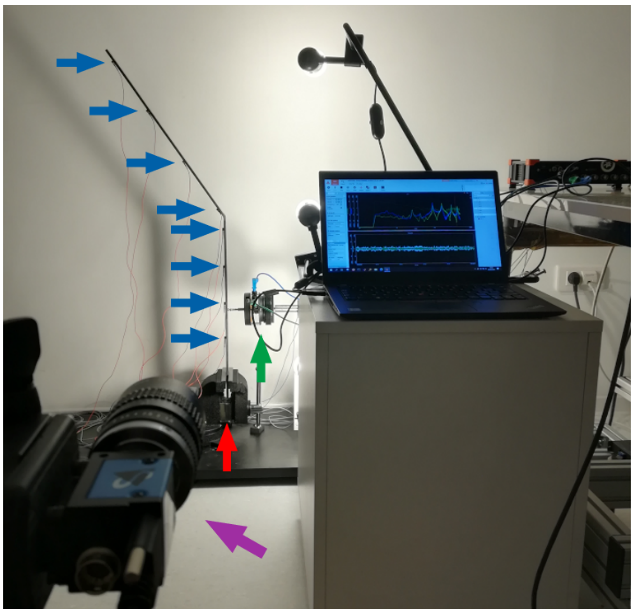

Figure 4.11: Experimental setup for the bent beam, with camera in purple, beam in red, accelerometers in blue, and vibration shaker in green.

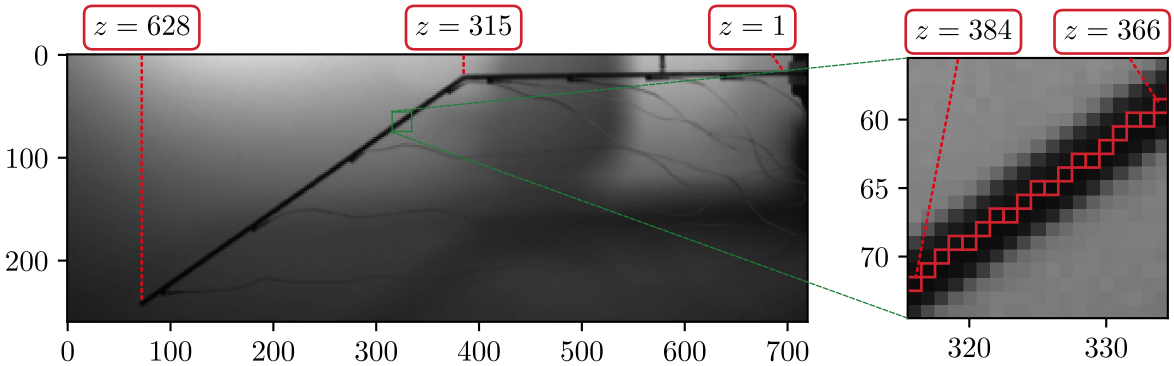

Figure 4.12: Frame from the experimental video of the bent beam (left) and pixels of interest in red (right).

The estimated motion along the curvilinear coordinate z is extracted and spatially smoothed as detailed in Sec. 3.6. The GOF method is once again approximately 400 times faster than SMPC (41.65 s vs. 16402.25 s for the 8000 frames). At the free end of the beam, the maximum displacement amplitude of our SMPC motion estimation method is 0.278 px (corresponding to $0.380 \cdot 10^{-3}$ m), which confirms a subpixel motion.

4.3.2 Video-based motion fidelity to accelerometer-based motion

To compare the video-based estimated motion with accelerometer-based data, the displacement orthogonal to the beam is extracted at the closest pixel to the accelerometer near the free end of the beam, where the motion is the largest. This extracted displacement is temporally derived twice to get acceleration in $\mathrm{px} \cdot \mathrm{s}^{-2}$. For the sake of comparison, the acceleration of the corresponding accelerometer is converted in the image plane from $\mathrm{m} \cdot \mathrm{s}^{-2}$ to $\mathrm{px} \cdot \mathrm{s}^{-2}$.

The PSD amplitude of the accelerometer-based and the two video-based accelerations are represented in Fig. 4.13. The PSD amplitude of the accelerometer data contains six main physical peaks, indicated by green arrows. These peaks seem to correspond to the six first natural frequencies of the beam. These natural frequencies are unknown but close to the straight beam ones.

For the preprocessed GOF, the PSD amplitude is high at the first four physical peak frequencies of the accelerometer, suggesting that the method can identify the first four modes. However, peaks at multiples of 100 Hz caused by the light bulb flickering frequency still occur in the PSD amplitude. The last two physical peaks are not detectable on the preprocessed GOF PSD amplitude. The harmonic perturbation at the flickering frequency may hide the peak of the last two modes.

The PSD amplitude of the preprocessed SMPC acceleration fits the accelerometer one below 200 Hz, with four physical peaks. Beyond this frequency, there is only a small peak at 305 Hz, suggesting that the amplitude of the fifth mode is sufficiently high to be detected. Only the sixth peak is not detectable from the SMPC PSD amplitude. Furthermore, as no 100 Hz harmonic peak is present in SMPC PSD, it confirms that this method is more robust than GOF against such intensity temporal perturbation.

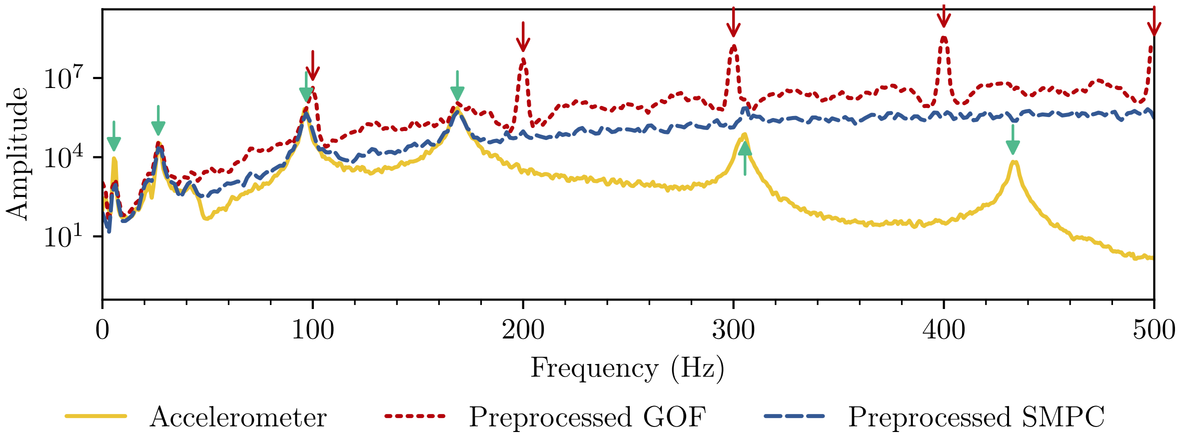

Figure 4.13: Comparison of acceleration power spectral density at the free end of the bent beam according to accelerometer-based data, video-based SMPC, and video-based GOF.

To spatially compare the video-based estimated motions, we compute the PSD of the displacement orthogonal to the beam at each pixel of interest using Welch's method [50]. The log-amplitude of the PSDs of GOF and SMPC motion estimations are represented in Fig. 4.14. The SMPC PSD is sparse in frequency, with low amplitude except around five frequencies. At these five frequencies, the drop in the PSD amplitude at some pixels correspond to the nodes of the modes (in dotted circles). The GOF PSD is much less sparse than SMPC ones. For every pixel, peaks occur at harmonics caused by the light bulb frequencies. The peaks corresponding to the physical modes are less detectable than in SMPC PSDs because of the flickering light perturbation (see red arrows). Furthermore, two low-amplitude horizontal lines at $z \approx 125$ and $z \approx 200$ have no physical cause linked to the beam itself. We explain the error at $z \approx 125$ by the low contrast between the beam and the shaker stem.

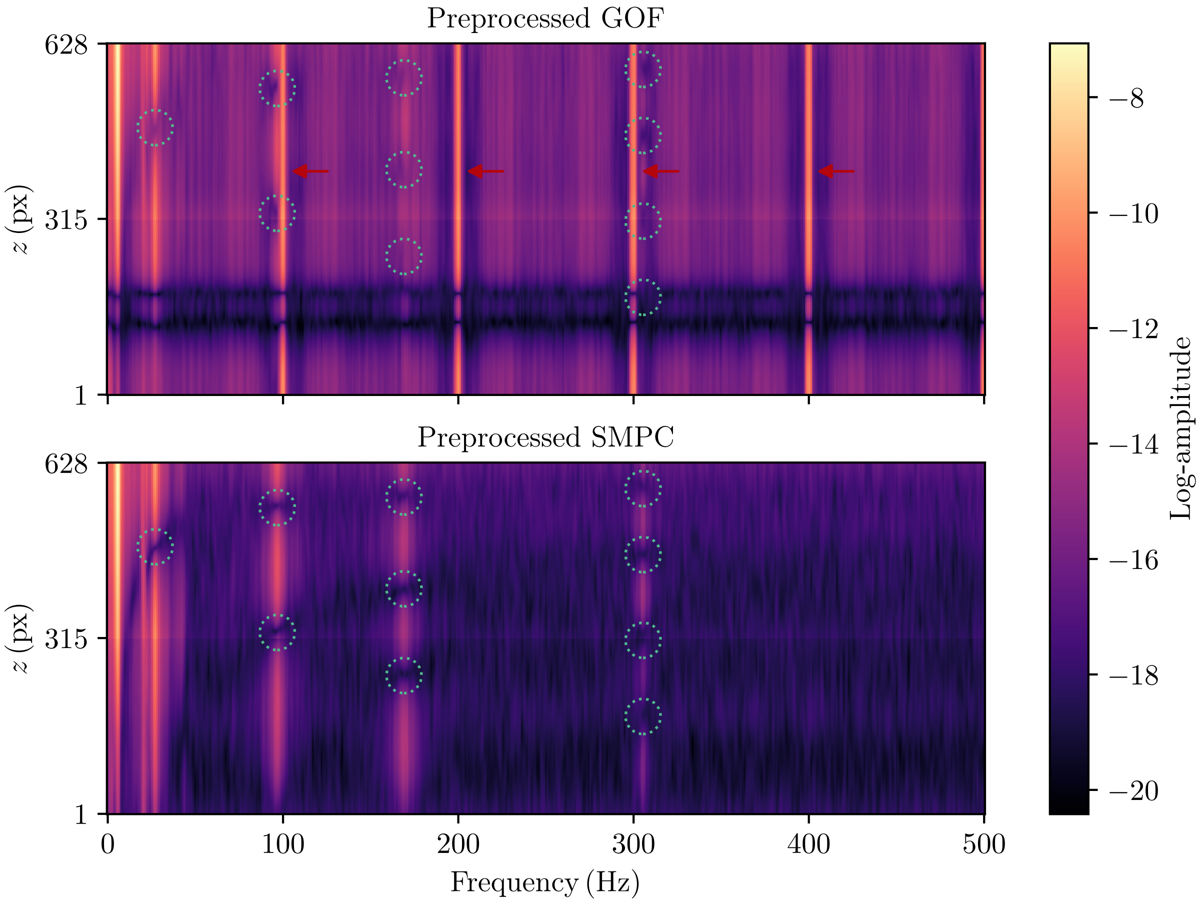

Figure 4.14: Power spectral density amplitude of the video-based estimated motion at each pixel of interest of the bent beam.

4.3.3 Modal basis accelerometers vs video

As no theoretical model is available for the bent beam, we directly compare the accelerometer-based and video-based modal bases. For this purpose, the accelerations provided by the accelerometers are used as observations to determine the ground truth modal basis. The SSI method detailed in Sec. 3.3.3 is used with $R = 40$ block rows for the block Toeplitz matrix. The two video-based modal bases are determined thanks to the FSSI method detailed in Sec. 3.4. The $P = 10$ first principal components are used to reduce the dimension of observations, and the block Toeplitz matrix is still constructed with $R = 40$ block rows. These parameter values are the same as in chapter 3 because they produce satisfying results. However, the influence of parameter adjustment on results is not studied here.

The stabilization diagram of each method is still built with a model order N in the range $[\![2, 50]\!]$. The frequency band of interest of accelerometers is set to $[0, 500]$ Hz to match the video one. The stabilization diagrams are fully shown in the left column of Fig. 4.15.

Six peaks are detectable on the accelerometer-based CMIF. However, due to the poor sensitivity of accelerometers below 10 Hz, no stable pole is sufficiently close to the peak at 5 Hz on the stabilization diagram. Each CMIF peak at frequency above 10 Hz is represented by a vertical series of stable poles, suggesting the presence of a physical mode.

The GOF stabilization diagram is difficult to interpret. Indeed, the CMIF is corrupted by the intensity frequency perturbation, and peaks above 100 Hz correspond to frequencies of the harmonics. Furthermore, many vertical series of stable poles are present above 100 Hz. The three first modes can be selected using the three first peaks. However, as stable poles are present at frequencies associated to the fourth and fifth frequency peaks of accelerometer CMIF, we also estimate the modes at the corresponding frequencies.

The SMPC CMIF shows five peaks whose frequencies match those of the first five accelerometer CMIF peak frequencies. A vertical series of stable poles is also present at each CMIF peak frequency.

To determine the modal basis, the stable poles at each identified frequency of interest are automatically selected (see right column of Fig. 4.15). For a given frequency, the mode is determined by computing the mean natural frequency, damping ratio, and mode shape of selected stable poles.

The natural frequencies and damping ratios of the six modes are listed in Tab. 4.5 and their shapes are represented in Fig. 4.16. The accelerometer + SSI method identifies five modes from the second to the sixth. As the first mode is not identified from the accelerometer data, we estimate its frequency as the location of the first CMIF peak. Both video-based methods

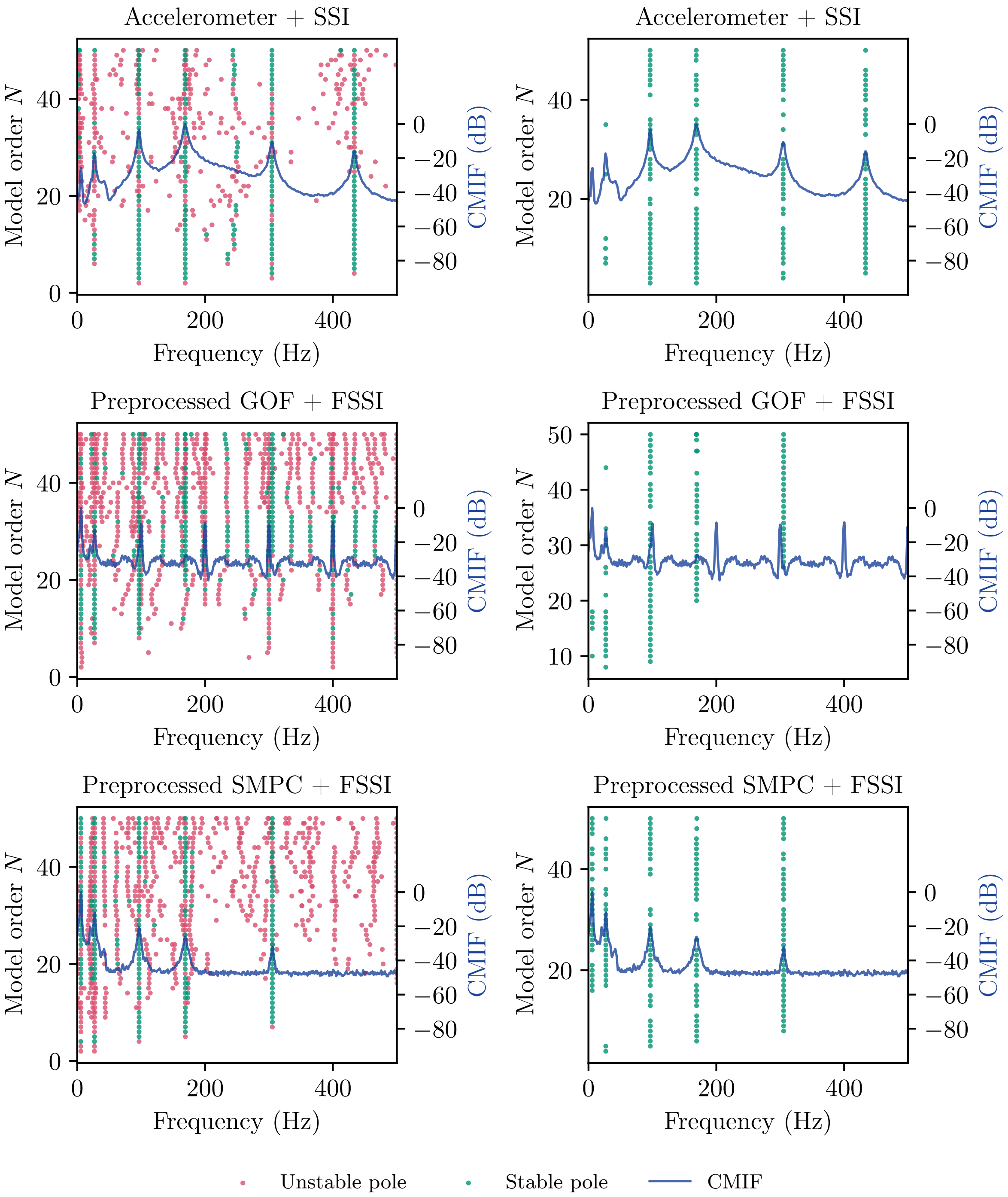

Figure 4.15: Stabilization diagrams of the accelerometer-based and video-based OMA methods for the bent beam. All poles are shown in the left column and selected stable poles are shown in the right column.

identify the five first modes. However, the preprocessed SMPC + FSSI method provides a cleaner stabilization diagram than that of preprocessed GOF + FSSI.

Criterion	Natural frequency (Hz)						Damping ratio ($\times 10^2$)					
Method	$\hat{f}_1$	$\hat{f}_2$	$\hat{f}_3$	$\hat{f}_4$	$\hat{f}_5$	$\hat{f}_6$	$\hat{\zeta}_1$	$\hat{\zeta}_2$	$\hat{\zeta}_3$	$\hat{\zeta}_4$	$\hat{\zeta}_5$	$\hat{\zeta}_6$
Accelerometer	5.683	26.880	96.603	168.885	304.693	433.641	–	3.940	1.142	1.081	0.299	0.050
GOF	5.695	27.016	96.783	169.297	305.327	–	1.864	3.725	0.983	1.201	0.397	–
SMPC	5.662	27.105	96.730	169.042	305.267	–	2.148	3.312	0.927	1.197	0.383	–

Table 4.5: Accelerometer- and video-based natural frequencies and damping ratios of the bent beam modes.

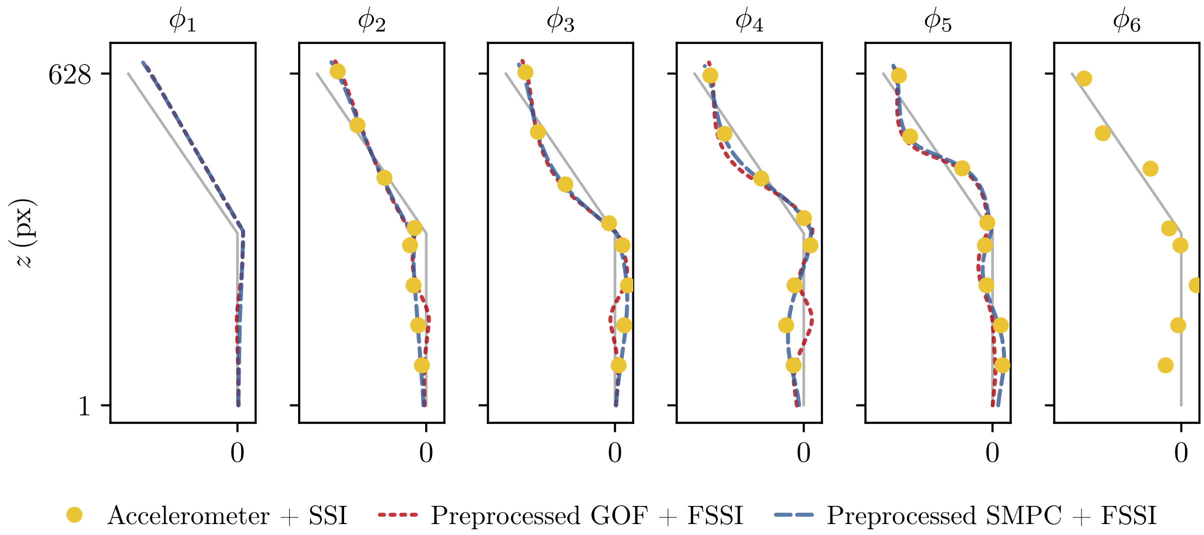

Figure 4.16: Accelerometer- and video-based mode shapes of the bent beam.

Both video-based modal bases are quantitatively compared to the accelerometer-based one by computing the errors (see Tab. 4.6). Both methods provide good estimations of natural frequencies and damping ratios with errors below 1 % and 20 %, respectively. Our preprocessed SMPC + FSSI method provides better mode shapes than the preprocessed GOF + FSSI method, with errors lower than 0.08. However, GOF-based mode shapes fit SMPC-based ones over most pixels of interest. Figure 4.16 shows that the GOF errors mainly lie between $z = 100$ and $z = 250$, probably because the estimated motion is strongly corrupted at these locations. Indeed, the contrast between the beam and background pixel intensities is low in this area, but GOF requires high contrast to produce accurate estimations [21].

Furthermore, neither of the two video-based OMA methods estimates the sixth mode at 433.641 Hz, because no peak is detectable at this frequency in the PSD of Fig. 4.14.

Criterion	Frequency relative error (%)					Damping ratio relative error (%)				Mode shape error (1-MAC)			
Method	$\epsilon\{\hat{f}_1\}$	$\epsilon\{\hat{f}_2\}$	$\epsilon\{\hat{f}_3\}$	$\epsilon\{\hat{f}_4\}$	$\epsilon\{\hat{f}_5\}$	$\epsilon\{\hat{\zeta}_2\}$	$\epsilon\{\hat{\zeta}_3\}$	$\epsilon\{\hat{\zeta}_4\}$	$\epsilon\{\hat{\zeta}_5\}$	$\epsilon\{\hat{\phi}_2\}$	$\epsilon\{\hat{\phi}_3\}$	$\epsilon\{\hat{\phi}_4\}$	$\epsilon\{\hat{\phi}_5\}$
GOF	**0.210**	**0.556**	0.358	0.272	0.225	**8.374**	11.496	2.200	**3.292**	0.213	0.233	0.645	0.226
SMPC	0.365	0.884	**0.303**	**0.121**	**0.205**	18.531	**5.105**	**1.859**	6.745	**0.072**	**0.028**	**0.051**	**0.075**

Table 4.6: FSSI modal basis estimation errors with GOF and SMPC motion estimation methods for the bent beam.

The SMPC method provides a coarse estimation of the modal basis for the beam along several directions. Indeed, natural frequencies are well estimated with relative errors below 1 %. However, damping ratios and mode shapes are above the standard acceptation criteria, with damping relative errors between 5 and 20 % and shape errors larger than 0.02. The errors are similar for both the straight and the bent beam (see Tabs. 4.4 and 4.6), which indicates that the performances of our video-based OMA method do not depend on the motion direction. The high mode shapes errors can be explained by the coarse location of accelerometers, fixed by hand with adhesive wax. Furthermore, the phase-based motion estimation is corrupted by gray level perturbation, which alters the estimation. The last mode is not retrieved since no motion is detected. One of the cause could be that motion amplitude at this frequency is too low. Therefore, the experimental setup should be modified so that the spatial resolution of the scene is increased.

4.4 Conclusion

In this chapter, we conduct the modal analysis of a cantilever straight beam and a cantilever bent beam observed in controlled conditions. Two video-based OMA methods are tested. The first one combines a state-of-the-art intensity-based motion estimation method (GOF) with our modal analysis method, dedicated to video (FSSI). The second one estimates the motion with our phase-based SMPC method. The performances of the video-based OMA methods are compared to the accelerometer-based OMA for both beams. First, the video-based estimated motions are converted into accelerations and compared with accelerometer data in the frequency domain. For each method, a stabilization diagram is then constructed, and the stable poles are identified in order to determine the modal basis. Finally, the video-based estimated modal bases are compared with that determined with accelerometers.

In the first experimental study, the four modes in the frequency band of interest are retrieved by the phase-based video OMA. The experiment is close to the study conducted in Sec. 3.5 on synthetic videos, with similar motion amplitudes, camera frame rate, and motion direction. The four modes are compared with ground-truth accelerometer data with respect to natural frequencies and damping ratios, and with the Euler–Bernoulli theoretical model with respect to mode shapes. The study shows that SMPC provides better motion estimation than GOF. The motion estimated by GOF is more corrupted by the intensity flickering dure to light bulb than SMPC. Both video-based OMA methods determine the four modes in the frequency range of interest, but our SMPC-based method outperforms the GOF-based method for the last mode shape. The SMPC-based natural frequencies are close to the accelerometer-based ones, but the damping ratios and mode shapes are above the classical criteria of acceptance.

The second study focuses on a bent cantilever beam with multiple directions of motion. For this study, the camera frame rate is increased so that there are six modes in the frequency range of interest. The accelerometer-based OMA method correctly estimates five modes but fails to retrieve the lowest frequency mode. On the other hand, both video-based methods determine five modes, but do not retrieve the highest frequency mode. This study confirms that our SMPC method is more robust against flickering intensity perturbation than GOF method, with cleaner stabilization diagrams. Moreover, the mode shapes obtained with GOF are unsatisfying in low-contrast parts of the video. As the GOF method is rather simple, the SMPC method could be compared with a more robust intensity-based motion estimation method. The errors of the subpixel SMPC-based modal basis are close to the criteria of acceptance, which confirms that phase-based methods are promising but improvements are still needed.

Several strategies could be investigated to improve the results. First, longer videos could be studied to get better correlations estimations in the video-based OMA. The influence of the method parameters on performances should be assessed. The performances variability could be studied by running the analysis several times with different videos and estimating the mean and standard deviation of estimated modal basis values. A different camera could be used with a sensor fitted with more pixels to identify smaller motion. Finally, a stabilized lighting could be used to illuminate the scene to avoid light bulb flickering perturbation.

Conclusion

Synthesis

The main objective of this thesis is to develop a video-based operational modal analysis (OMA) method to extract vibration characteristics of mechanical structures. The work can be synthesized into two main contributions.

As vibrations of structures are acquired by a camera with limited spatial resolution, motion must be estimated at subpixel scales. Furthermore, structures may vibrate along several unpredictable directions. Therefore, we propose a new two-dimensional subpixel motion estimation method called weighted spatial mean based on phase confidence (SMPC). This phase-based motion estimation (PME) method analyzes the phase obtained from the decomposition of video frames into space–frequency complex subbands. The subbands with irrelevant phase are discarded thanks to a confidence indicator and the phase of the remaining ones are fused to obtain one single estimated motion. The performances of SMPC are compared with state-of-the-art methods on synthetic videos of a vibrating cantilever beam. SMPC is time-consuming in comparison with other methods, but provides estimations that are more robust against additive intensity noise.

Classical methods of operational modal analysis (OMA) are designed for a low number of sensors. However, video-based methods can potentially estimate motion at each pixel of each frame. To adapt the classical methods to a high number of pixels, a dimension reduction method is proposed to first decrease the size of video-based observations before they are fed into a classical OMA procedure. The method is compared with a state-of-the-art video-based OMA on synthetic videos of a cantilever beam. A preprocessing step based on a spatial smoothing of the estimated motion is also proposed to improve the quality of the determined modal basis. The proposed method outperforms the state-of-the-art method on synthetic videos. Furthermore, the method successfully estimates modes whose amplitude is larger than 0.003 px for noiseless synthetic videos with high contrast between the beam and background.

Finally, we conduct the OMA of two different cantilever beams observed by a high-speed camera in controlled conditions, with both accelerometer-based and video-based methods. A straight cantilever beam is first observed to provide videos whose characteristics are close to the synthetic ones. Then, the OMA of a bent cantilever beam is conducted to study vibration along several unknown directions.

For both experiments, our SMPC method is compared with a state-of-the-art optical flow-based method. We show that our phase-based SMPC method experimentally performs better than the intensity-based optical flow one. Furthermore, SMPC determines all the modes of the straight beam and only misses one mode of the bent beam in the frequency range of interest. However, the modal bases are somewhat coarsely estimated comparatively to accelerometer-based ones, with relative errors that are slightly above the classical criteria of acceptance.

Perspectives

Although this work contains several contributions, further research is required to design an automatic and robust video-based vibration monitoring system. Indeed, the video-based OMA method detailed in this thesis produces promising results, but still needs to be improved to respect the acceptability criteria.

First, the influence of parameters on the performances reached by our SMPC method should be assessed. Indeed, the method relies on several parameters such as the number of scales and orientations of the space–frequency decomposition, the threshold of the phase confidence mask, and the standard deviation of the Gaussian weight kernel. All these parameters have been adjusted using a trial-and-error approach. Furthermore, they depend on the characteristics of the videos and should be defined by considering the video properties, such as the spatial frequencies of its frames. Moreover, an in-depth study about the robustness of performances of our SMPC method should be studied, by analyzing the variability of the results with several videos.

In addition, the comparison of our subpixel phase-based method with an intensity-based method should be extended to more complex methods. Indeed, the gradient-based optical flow method to which we compare ours is rather simple and does not fully emphasize the performances of SMPC.

Our video-based OMA method also requires to adjust several parameters, namely the size of the reduced observations, as well as the number of block rows and the model order of the stochastic subspace identification. In this work, these parameters are also set manually. Their influence on the quality of the determined modal basis should be studied. Futhermore, vibration monitoring of mechanical structures requires accurate modal basis to detect faults. The variability of the results with respect to the video settings and the stochastic subspace identification parameters should also be explored.

Exploiting stabilization diagrams requires that the operator is experienced enough to accurately determine the modal basis. It would be useful to develop an automatic analysis of stabilization diagram to reduce the human errors during this step and to simplify monitoring [55].

This work only presents experimental studies in controlled conditions. Since most of mechanical structures are located outdoor, an experimental study of such structures should be conducted in uncontrolled conditions to assess the influence of light variations and of occlusions in the scene.

To improve the determination of the modal basis, the OMA could be performed using estimated motion in two videos of close views of the structure. However, an overlapping area between the two videos would be necessary to serve as reference. On the other hand, another setup with several synchronized cameras with a focus on different parts of the structure could also be assessed to improve the spatial resolution, hence the estimated motion accuracy.

Our SMPC method estimates the motion projected in the frame plane, and thus the mode shapes according to two directions. However, many mechanical structures are complex and require estimations in three dimensions. Therefore, it would be interesting to combine two cameras to estimate 3D vibrations thanks to stereo-vision.

The developed approaches are based on a theoretical model of vibration. It should be finally interesting to investigate how free-model approaches, based on supervised machine learning, can identify the modal basis of mechanical systems. As vibrations can be decomposed into a finite sum of modes for a given frequency range, videos with subpixel vibrations could be represented by a few number of components. Thus, it could be interesting to train an autoencoder with the video to learn its representation and to perform the modal analysis in the latent space.

Bibliography

[1] Peter Avitabile. “Experimental modal analysis - A simple non-mathematical presentation”. In: *Sound and Vibration* 35 (Jan. 2001), pp. 20–31.

[2] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. “Speeded-Up Robust Features (SURF)”. In: *Computer Vision and Image Understanding* 110.3 (June 2008), pp. 346–359. DOI: 10.1016/j.cviu.2007.09.014.

[3] G. Bradski. “The OpenCV Library”. In: *Dr. Dobb’s Journal of Software Tools* (2000).

[4] P. Castellini, M. Martarelli, and E.P. Tomasini. “Laser Doppler Vibrometry: Development of advanced solutions answering to technology’s needs”. In: *Mechanical Systems and Signal Processing* 20.6 (Aug. 2006), pp. 1265–1285. DOI: 10.1016/j.ymssp.2005.11.015.

[5] Justin G. Chen. “Video Camera-based Vibration Measurements of Infrastructure”. PhD Thesis. MIT, 2016.

[6] Justin G. Chen, Neal Wadhwa, Young-Jin Cha, Frédo Durand, William T. Freeman, and Oral Buyukozturk. “Modal identification of simple structures with high-speed video using motion magnification”. In: *Journal of Sound and Vibration* 345 (June 2015), pp. 58–71. DOI: 10.1016/j.jsv.2015.01.024.

[7] Justin G. Chen, Neal Wadhwa, Frédo Durand, William T. Freeman, and Oral Buyukozturk. “Developments with Motion Magnification for Structural Modal Identification Through Camera Video”. In: *Dynamics of Civil Structures, Volume 2*. Ed. by Juan Caicedo and Shamim Pakzad. Cham: Springer International Publishing, 2015, pp. 49–57. DOI: 10.1007/978-3-319-15248-6_5.

[8] T. C. Chu, W. F. Ranson, and M. A. Sutton. “Applications of digital-image-correlation techniques to experimental mechanics”. In: *Experimental Mechanics* 25.3 (Sept. 1985), pp. 232–244. DOI: 10.1007/BF02325092.

[9] Abe Davis, Katherine L. Bouman, Justin G. Chen, Michael Rubinstein, Oral Büyüköztürk, Frédo Durand, and William T. Freeman. “Visual Vibrometry: Estimating Material Properties from Small Motions in Video”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.4 (2017), pp. 732–745. DOI: 10.1109/TPAMI.2016.2622271.

[10] Charles Dorn, Sudeep Dasari, Yongchao Yang, Charles Farrar, Garrett Kenyon, Paul Welch, and David Mascareñas. "Efficient Full-Field Vibration Measurements and Operational Modal Analysis Using Neuromorphic Event-Based Imaging". In: *Journal of Engineering Mechanics* 144.7 (July 2018), p. 04018054. DOI: `10.1061/(ASCE)EM.1943-7889.0001449`.

[11] D. J. Ewins. *Modal Testing: Theory, Practice and Application*. Wiley, July 2009.

[12] David J. Fleet and Allan D. Jepson. "Computation of component image velocity from local phase information". In: *International Journal of Computer Vision* 5.1 (Aug. 1990), pp. 77–104. DOI: `10.1007/bf00056772`.

[13] Patrick Guillaume, Peter Verboven, S. Vanlanduit, Herman Van der Auweraer, and Bart Peeters. "A poly-reference implementation of the least-squares complex frequency-domain estimator". In: *Proceedings of the 21st IMAC Conference & Exposition, Kissimmee, Florida, USA* 21 (Jan. 2003).

[14] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. "Array programming with NumPy". In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: `10.1038/s41586-020-2649-2`.

[15] Ward Heylen, Stefan Lammens, and Paul Sas. *Modal Analysis Theory and Testing*. 2. ed. Leuven: Katholieke Universiteit Leuven, Department of Mechanical Engineering, 2007.

[16] J. D. Hunter. "Matplotlib: A 2D graphics environment". In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: `10.1109/MCSE.2007.55`.

[17] S. R. Ibrahim and E. C. Mikulcik. "A method for the direct identification of vibration parameters from the free response". In: *Shock and Vibration Bulletin* 47 (1977), pp. 183–196.

[18] D. J. Inman. *Engineering vibration*. Fourth edition. Upper Saddle River, N.J: Pearson Prentice Hall, 2014.

[19] George James, Thomas Carne, and J. Laufer. "The natural excitation technique (NExT) for modal parameter extraction from operating structures". In: *Journal of Analytical and Experimental Modal Analysis* 10 (Jan. 1995).

[20] Jaka Javh, Janko Slavič, and Miha Boltežar. "High frequency modal identification on noisy high-speed camera data". In: *Mechanical Systems and Signal Processing* 98 (Jan. 2018), pp. 344–351. DOI: `10.1016/j.ymssp.2017.05.008`.

[21] Jaka Javh, Janko Slavič, and Miha Boltežar. “The subpixel resolution of optical-flow-based modal analysis”. In: *Mechanical Systems and Signal Processing* 88 (May 2017), pp. 89–99. DOI: `10.1016/j.ymssp.2016.11.009`.

[22] Allan D Jepson and David J Fleet. “Phase singularities in scale-space”. In: *Image and Vision Computing* 9.5 (Oct. 1991), pp. 338–343. DOI: `10.1016/0262-8856(91)90039-R`.

[23] Jer-Nan Juang. *Applied system identification*. Englewood Cliffs, N.J: Prentice Hall, 1994.

[24] Hyoung-Chul Kim, Moo-Hyun Kim, and Do-Eun Choe. “Structural health monitoring of towers and blades for floating offshore wind turbines using operational modal analysis and modal properties with numerical-sensor signals”. In: *Ocean Engineering* 188 (Sept. 2019), p. 106226. DOI: `10.1016/j.oceaneng.2019.106226`.

[25] Sung-Wan Kim and Nam-Sik Kim. “Multi-point Displacement Response Measurement of Civil Infrastructures Using Digital Image Processing”. In: *Procedia Engineering* 14 (2011), pp. 195–203. DOI: `10.1016/j.proeng.2011.07.023`.

[26] David G. Lowe. “Distinctive Image Features from Scale-Invariant Keypoints”. In: *International Journal of Computer Vision* 60.2 (Nov. 2004), pp. 91–110. DOI: `10.1023/B:VISI.0000029664.99615.94`.

[27] Bruce D. Lucas and Takeo Kanade. “An iterative image registration technique with an appli- cation to stereo vision”. In: *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*. IJCAI’81. Vancouver, BC, Canada: Morgan Kaufmann Publishers Inc., Aug. 1981, pp. 674–679.

[28] F. Magalhães, A. Cunha, and E. Caetano. “Vibration based structural health monitoring of an arch bridge: From automated OMA to damage detection”. In: *Mechanical Systems and Signal Processing* 28 (Apr. 2012), pp. 212–228. DOI: `10.1016/j.ymssp.2011.06.011`.

[29] N.M.M. Maia, J.M.M. Silva, E.A.M. Almas, and R.P.C. Sampaio. “Damage detection in structures: from mode shape to frequency response function methods”. In: *Mechanical Systems and Signal Processing* 17.3 (May 2003), pp. 489–498. DOI: `10.1006/mssp.2002.1506`.

[30] Cédric Marinel, Olivier Losson, Benjamin Mathon, Jean Le Besnerais, and Ludovic Macaire. “Comparison of operational modal analysis methods from displacement estimation by video processing”. In: *Proceedings of International Conference on Noise and Vibration Engineering*. Leuven, Belgium, 2022. URL: `https://hal.science/hal-03821159`.

[31] Cédric Marinel, Olivier Losson, Benjamin Mathon, Jean Le Besnerais, and Ludovic Macaire. “Operational modal analysis of a cantilever beam by phase-based sub-pixel motion estimated from experimental videos”. In: *Proceedings of the Sixteenth International Conference on Quality Control by Artificial Vision*. Ed. by Igor Jovančević and Jean-José Orteu. Vol. 12749. Albi, France: SPIE, June 2023, 127490W. DOI: `10.1117/12.2692503`.

[32] Cédric Marinel, Benjamin Mathon, Olivier Losson, and Ludovic Macaire. "Comparison of Phase-based Sub-Pixel Motion Estimation Methods". In: *Proceedings of 2022 IEEE International Conference on Image Processing (ICIP)*. Bordeaux, France: IEEE, Oct. 2022, pp. 561–565. DOI: `10.1109/ICIP46576.2022.9897338`.

[33] Eli Parloo. "Application of frequency-domain system identification techniques in the field of operational modal analysis". PhD thesis. Université Libre de Bruxelles, Belgium, May 2003.

[34] Bart Peeters and Guido De Roeck. In: *Earthquake Engineering & Structural Dynamics* 30.2 (2001), pp. 149–171. DOI: `10.1002/1096-9845(200102)30:2<149::AID-EQE1>3.0.CO;2-Z`.

[35] Bart Peeters and Guido De Roeck. "Reference-based stochastic subspace identification for output-only modal analysis". In: *Mechanical Systems and Signal Processing* 13.6 (Nov. 1999), pp. 855–878. DOI: `10.1006/mssp.1999.1249`.

[36] Peyman Poozesh, Aral Sarrafi, Zhu Mao, Peter Avitabile, and Christopher Niezrecki. "Feasibility of extracting operating shapes using phase-based motion magnification technique and stereo-photogrammetry". In: *Journal of Sound and Vibration* 407 (2017), pp. 350–366. DOI: `10.1016/j.jsv.2017.06.003`.

[37] Javier Portilla and Eero Simoncelli. "A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients". In: *International Journal of Computer Vision* 40.1 (2000), pp. 49–70. DOI: `10.1023/a:1026553619983`.

[38] E. Reynders, J. Houbrechts, and G. De Roeck. "Automated interpretation of stabilization diagrams". In: *Modal Analysis Topics, Volume 3*. Ed. by Tom Proulx. Conference Proceedings of the Society for Experimental Mechanics Series. New York, NY: Springer New York, 2011, pp. 189–201. DOI: `10.1007/978-1-4419-9299-4_16`.

[39] Mark H. Richardson. "Is it a mode shape, or an operating deflection shape?" In: *Sound and Vibration* 30th Anniversary Issue (1997), pp. 1–8.

[40] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. "ORB: An efficient alternative to SIFT or SURF". In: *2011 International Conference on Computer Vision*. Barcelona, Spain: IEEE, Nov. 2011, pp. 2564–2571. DOI: `10.1109/ICCV.2011.6126544`.

[41] Hubert Schreier, Jean-José Orteu, and Michael A. Sutton. *Image Correlation for Shape, Motion and Deformation Measurements: Basic Concepts,Theory and Applications*. Boston, MA: Springer US, 2009. DOI: `10.1007/978-0-387-78747-3`.

[42] C.Y. Shih, Y.G. Tsuei, R.J. Allemang, and D.L. Brown. "Complex mode indication function and its applications to spatial domain parameter estimation". In: *Mechanical Systems and Signal Processing* 2.4 (Oct. 1988), pp. 367–377. DOI: `10.1016/0888-3270(88)90060-x`.

[43] E.P. Simoncelli and W.T. Freeman. “The steerable pyramid: a flexible architecture for multi-scale derivative computation”. In: *Proceedings., International Conference on Image Processing*. Vol. 3. 1995, 444–447 vol.3. DOI: `10.1109/ICIP.1995.537667`.

[44] Andreĭ Nikolaevich Tikhonov and Aleksandr Andreevich Samarskii. *Equations of mathematical physics*. Courier Corporation, 2013.

[45] 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. DOI: `10.1109/imtc.2004.1351525`.

[46] Peter VanOverschee and Bart L. R. De Moor. *Subspace identification for linear systems: Theory – Implementation – Applications*. Boston: Kluwer Academic Publishers, 1996. DOI: `10.1007/978-1-4613-0465-4`.

[47] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17 (2020), pp. 261–272. DOI: `10.1038/s41592-019-0686-2`.

[48] Neal Wadhwa, Justin G. Chen, Jonathan B. Sellon, Donglai Wei, Michael Rubinstein, Roozbeh Ghaffari, Dennis M. Freeman, Oral Büyüköztürk, Pai Wang, Sijie Sun, Sung Hoon Kang, Katia Bertoldi, Frédo Durand, and William T. Freeman. “Motion microscopy for visualizing and quantifying small motions”. In: *Proceedings of the National Academy of Sciences* 114.44 (Oct. 2017), pp. 11639–11644. DOI: `10.1073/pnas.1703715114`.

[49] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T. Freeman. “Phase-based video motion processing”. In: *ACM Transactions on Graphics* 32.4 (July 2013), pp. 1–10. DOI: `10.1145/2461912.2461966`.

[50] P. Welch. “The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms”. In: *IEEE Transactions on Audio and Electroacoustics* 15.2 (June 1967), pp. 70–73. DOI: `10.1109/TAU.1967.1161901`.

[51] Yan Xu and James M. W. Brownjohn. “Review of machine-vision based methodologies for displacement measurement in civil structures”. In: *Journal of Civil Structural Health Monitoring* 8.1 (Jan. 2018), pp. 91–110. DOI: `10.1007/s13349-017-0261-4`.

[52] Yongchao Yang, Charles Dorn, Tyler Mancini, Zachary Talken, Garrett Kenyon, Charles Farrar, and David Mascareñas. “Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification”. In: *Mechanical Systems and Signal Processing* 85 (Feb. 2017), pp. 567–590. DOI: `10.1016/j.ymssp.2016.08.041`.

[53] Yongchao Yang, Charles Dorn, Tyler Mancini, Zachary Talken, Satish Nagarajaiah, Garrett Kenyon, Charles Farrar, and David Mascareñas. “Blind identification of full-field vibration modes of output-only structures from uniformly-sampled, possibly temporally-aliased (sub-Nyquist), video measurements”. In: *Journal of Sound and Vibration* 390 (Mar. 2017), pp. 232–256. DOI: `10.1016/j.jsv.2016.11.034`.

[54] Yongchao Yang and Satish Nagarajaiah. “Blind modal identification of output-only structures in time-domain based on complexity pursuit”. In: *Earthquake Engineering & Structural Dynamics* 42.13 (May 2013), pp. 1885–1905. DOI: `10.1002/eqe.2302`.

[55] Xijun Ye, Peili Huang, Chudong Pan, and Liu Mei. “Innovative stabilization diagram for automated structural modal identification based on ERA and hierarchical cluster analysis”. In: *Journal of Civil Structural Health Monitoring* 11.5 (Aug. 2021), pp. 1355–1373. DOI: `10.1007/s13349-021-00514-8`.

[56] Dashan Zhang, Jie Guo, Xiujun Lei, and Changan Zhu. “A High-Speed Vision-Based Sensor for Dynamic Vibration Analysis Using Fast Motion Extraction Algorithms”. In: *Sensors* 16.4 (Apr. 2016), p. 572. DOI: `10.3390/s16040572`.

[57] Jiwen Zhou, Wendi Zhang, Yun Li, Xiaojian Wang, Li Zhang, and Hongguang Li. “Phase-Based Displacement Sensor With Improved Spatial Frequency Estimation and Data Fusion Strategy”. In: *IEEE Sensors Journal* 22.4 (Feb. 2022), pp. 3306–3315. DOI: `10.1109/JSEN.2022.3141110`.

A Open-source synthetic video generation

A.1 Euler–Bernoulli solution

The spatial and temporal partial derivatives in Eq. (2.50) can be developed to give:

$$\frac{c^2}{\phi(z)}\frac{\partial^4\phi}{\partial z^4}(z) = -\frac{1}{q(t)}\frac{\partial^2 q}{\partial t^2}(t),\ \forall z \in [0, L], t \in [0, \infty), \tag{A.1}$$

$$\frac{c^2}{\phi(z)}\frac{\partial^4\phi}{\partial z^4}(z) = -\frac{1}{q(t)}\frac{\partial^2 q}{\partial t^2}(t) \ = \omega^2. \tag{A.2}$$

Hence,

$$\begin{cases} \dfrac{\partial^4\phi}{\partial z^4}(z) - \beta^2\phi(z) & = 0 \quad \text{with } \beta^2 = \frac{\omega}{c}, \\ \dfrac{\partial^2 q}{\partial t^2}(t) + \omega^2 q(t) & = 0. \end{cases} \tag{A.3}$$

Functions ϕ and q have the following forms:

$$\phi(z) = C_1 \cosh(\beta z) + C_2 \sinh(\beta z) + C_3 \cos(\beta z) + C_4 \sin(\beta z), \tag{A.4}$$

$$q(t) = A\cos(\omega t) + B\sin(\omega t). \tag{A.5}$$

where A, B, C_1, C_2, C_3, and C_4 are constants.

A.1.1 Mode shapes

The constants of ϕ can be computed by adding some boundary conditions. For a cantilever beam, the bottom boundary (*i.e.*, $z = 0$) is clamped and the top boundary (*i.e.*, $z = L$) is free:

$$g(0, t) = 0\ \forall\ t \in [0, \infty), \tag{A.6}$$

$$\frac{\partial g}{\partial z}(0, t) = 0,\ \forall\ t \in [0, \infty), \tag{A.7}$$

$$\frac{\partial^2 g}{\partial z^2}(L, t) = 0,\ \forall\ t \in [0, \infty), \tag{A.8}$$

$$\frac{\partial^3 g}{\partial z^3}(L, t) = 0,\ \forall\ t \in [0, \infty). \tag{A.9}$$

Relationship between constants can be determined using these conditions:

$$\begin{aligned} C_1 + C_3 &\overset{(A.6)}{=} 0, \\ C_3 &= -\,C_1, \end{aligned} \tag{A.10}$$

and,

$$\begin{aligned} C_2 + C_4 &\overset{(A.7)}{=} 0, \\ C_4 &= -\,C_2. \end{aligned} \tag{A.11}$$

Furthermore, combination of those relationship and free end boundary condition gives:

$$\begin{aligned} C_1 \cosh(\beta L) + C_2 \sinh(\beta L) - C_3 \cos(\beta L) - C_4 \sin(\beta L) &\overset{(A.8),(A.10),(A.11)}{=} 0, \\ C_1(\cosh(\beta L) + \cos(\beta L)) + C_2(\sinh(\beta L) + \sin(\beta L)) &= 0, \end{aligned} \tag{A.12}$$

and,

$$\begin{aligned} C_1 \sinh(\beta L) + C_2 \cosh(\beta L) + C_3 \sin(\beta L) - C_4 \cos(\beta L) &\overset{(A.9)-(A.11)}{=} 0, \\ C_1(\sinh(\beta L) - \sin(\beta L)) + C_2(\cosh(\beta L) + \cos(\beta L)) &= 0, \\ C_2 &= C_1 \frac{\sin(\beta L) - \sinh(\beta L)}{\cos(\beta L) + \cosh(\beta L)}. \end{aligned} \tag{A.13}$$

Finally, Eqs (A.12) and (A.13) can be combined to get the following equation:

$$\begin{aligned} & C_1 \cdot \left(\cosh(\beta L) + \cos(\beta L) + \frac{\sin(\beta L) - \sinh(\beta L)}{\cos(\beta L) + \cosh(\beta L)} (\sinh(\beta L) + \sin(\beta L)) \right) = 0, \\ \Leftrightarrow\ & C_1 \cdot \left(\cosh(\beta L)^2 + \cos(\beta L)^2 + 2\cosh(\beta L)\cos(\beta L) + \sin(\beta L)^2 - \sinh(\beta L)^2 \right) = 0, \\ \Leftrightarrow\ & C_1 \cdot (\cos(\beta L)\cosh(\beta L) + 1) = 0. \end{aligned} \tag{A.14}$$

Non-trivial infinity of solutions are found to exist only if $\cos(\beta_i L) \cdot \cosh(\beta_i L) + 1 = 0$. This nonlinear equation can be solved numerically. For each solution β_i, there is a corresponding modal contribution q_i with natural frequency of vibration $\omega_i = \beta_i^2 \sqrt{\frac{EI}{\rho A}}$ and a mode shape ϕ_i of the form:

$$\phi_i(z) = C_i \cdot \left(\cosh(\beta_i z) - \cos(\beta_i z) + \frac{\sinh(\beta_i L) - \sinh(\beta_i L)}{\cos(\beta_i L) + \cosh(\beta_i L)} (-\sinh(\beta_i z) + \sin(\beta_i z)) \right) \tag{A.15}$$

As the modes are given relatively to a constant C_i, it is set to normalize the mode such that:

$$\|\phi_i(z)\|_{L^2} = 1\,. \tag{A.16}$$

These modes are orthogonal to one another:

$$\int\limits_0^L \phi_r(z) \cdot \phi_s(z)\, dx = 0 \quad \forall\, r \neq s \tag{A.17}$$

Indeed, $\forall r \in \mathbb{N}$,

$$\frac{\partial^4 \phi_r(z)}{\partial z^4} = \beta_r^4 \phi(z) \Rightarrow \int\limits_0^L \frac{\partial^4 \phi_r(z)}{\partial z^4} \cdot \phi_s(z)\, dx = \int\limits_0^L \beta_r^4 \cdot \phi(z) \cdot \phi_s(z)\, dx \tag{A.18}$$

Furthermore,

$$\int\limits_0^L \frac{\partial^4 \phi_r(z)}{\partial z^4} \cdot \phi_s(z)\, dx = -\int\limits_0^L \frac{\partial^3 \phi_r(z)}{\partial z^3} \cdot \frac{\partial \phi_s(z)}{\partial z}\, dx + \left[\frac{\partial^3 \phi_r(z)}{\partial z^3} \cdot \phi_s(z)\right]_0^L, \tag{A.19}$$

$$= \int\limits_0^L \frac{\partial^2 \phi_r(z)}{\partial z^2} \cdot \frac{\partial^2 \phi_s(z)}{\partial z^2}\, dx - \left[\frac{\partial^2 \phi_r(z)}{\partial z^2} \cdot \frac{\partial \phi_s(z)}{\partial z}\right]_0^L, \tag{A.20}$$

$$= \int\limits_0^L \frac{\partial^2 \phi_r(z)}{\partial z^2} \cdot \frac{\partial^2 \phi_s(z)}{\partial z^2}\, dx, \tag{A.21}$$

$$= \int\limits_0^L \phi_r(z) \cdot \frac{\partial^4 \phi_s(z)}{\partial z^4}\, dx. \tag{A.22}$$

Hence,

$$\int_0^L \frac{\partial^4 \phi_r(z)}{\partial z^4} \cdot \phi_s(z)\,dx - \int_0^L \phi_r(z) \cdot \frac{\partial^4 \phi_s(z)}{\partial z^4}\,dx = 0, \tag{A.23}$$

$$\beta_r^4 \int_0^L \phi(z) \cdot \phi_s(z)\,dx - \beta_s^4 \int_0^L \cdot\phi(z) \cdot \phi_s(z)\,dx = 0, \tag{A.24}$$

$$\frac{\rho A}{EJ}(\omega_r - \omega_s) \int_0^L \phi(z) \cdot \phi_s(z)\,dx = 0, \tag{A.25}$$

$$\int_0^L \phi(z) \cdot \phi_s(z)\,dx = 0 \quad \forall\, r \neq s. \tag{A.26}$$

A.1.2 Modal contribution

The modal contribution $q_i(t)$ corresponding to the mode shape $\phi_i(z)$ depends on initial condition $g(z,0)$ and force $\gamma(z,t)$ applied to the beam. It can be computed for each mode using the mode shapes orthogonality:

$$EJ\frac{\partial^4 g}{\partial z^4}(z,t) + \rho A\frac{\partial^2 g}{\partial t^2}(z,t) = \gamma(z,t), \tag{A.27}$$

$$\int_0^L \phi_i(z) EJ\frac{\partial^4 g}{\partial z^4}(z,t)\,dx + \int_0^L \phi_i(z)\rho A\frac{\partial^2 g}{\partial t^2}(z,t)\,dx = \int_0^L \phi_i(z)\gamma(z,t)\,dx, \tag{A.28}$$

$$EJ\int_0^L \phi_i(z)\sum_{r=0}^{\infty}\frac{\partial^4 \phi_r(z)}{\partial z^4}q_r(t)\,dx + \rho A\int_0^L \phi_i(z)\sum_{r=0}^{\infty}\phi_r(z)\frac{\partial^2 q_r(t)}{\partial t^2}\,dx = \int_0^L \phi_i(z)\gamma(z,t)\,dx, \tag{A.29}$$

$$EJ\sum_{r=0}^{\infty}q_r(t)\int_0^L \phi_i(z)\frac{\partial^4 \phi_r(z)}{\partial z^4}\,dx + \rho A\sum_{r=0}^{\infty}\frac{\partial^2 q_r(t)}{\partial t^2}\int_0^L \phi_i(z)\phi_r(z)\,dx = \int_0^L \phi_i(z)\gamma(z,t)\,dx, \tag{A.30}$$

$$EJ\sum_{r=0}^{\infty}\beta_r^4 q_r(t)\int_0^L \phi_i(z)\phi_r(z)\,dx + \rho A\sum_{r=0}^{\infty}\frac{\partial^2 q_r(t)}{\partial t^2}\int_0^L \phi_i(z)\phi_r(z)\,dx = \int_0^L \phi_i(z)\gamma(z,t)\,dx, \tag{A.31}$$

$$EJ\beta_i^4 q_i(t) + \rho A\frac{\partial^2 q_i(t)}{\partial t^2} = \int_0^L \phi_i(z)\gamma(z,t)\,dx, \tag{A.32}$$

$$\omega_i^2 q_i(t) + \frac{\partial^2 q_i(t)}{\partial t^2} = \frac{1}{\rho A}\int_0^L \phi_i(z)\gamma(z,t)\,dx. \tag{A.33}$$

B Video-based OMA user interface

During this work, a video processing user interface (UI) has been developed to facilitate the video-based modal analysis. The UI is available at the following link: `https://zenodo.org/record/8376571`. The program is written in Python using high-performance open source projects such as: NumPy [14], SciPy [47], Matplotlib [16], OpenCV [3], and strid[1].

The UI allows the user to load and play a video, select regions or pixels of interest, and denoise the video. The video-based modal analysis can be run by selecting a phase-based method and a video-based OMA procedure. Finally, the stabilization diagram and the mode shapes can be viewed in the UI. The UI is also coupled with the TIS DMK 33UX287 camera API to acquire high-speed videos.

---

[1] `https://doi.org/10.5281/zenodo.6540518`