Subspace-based modal identification and uncertainty quantification from video image flows

Boualem Merainani, Bian Xiong, Vincent Baltazart, Michael Döhler, Jean Dumoulin, Qinghua Zhang

To cite this version:

Boualem Merainani, Bian Xiong, Vincent Baltazart, Michael Döhler, Jean Dumoulin, et al.. Subspace-based modal identification and uncertainty quantification from video image flows. Journal of Sound and Vibration, 2024, 569, pp.117957. 10.1016/j.jsv.2023.117957 . hal-04436527

HAL Id: hal-04436527
https://inria.hal.science/hal-04436527v1

Submitted on 3 Feb 2024

Subspace-based modal identification and uncertainty quantification from video image flows

Boualem Merainani[a,*], Bian Xiong[a], Vincent Baltazart[b], Michael Döhler[c], Jean Dumoulin[a], Qinghua Zhang[c]

[a] *Université Gustave Eiffel, Inria, COSYS-SII, I4S Team, F-44344 Bouguenais, France*
[b] *Université Gustave Eiffel, COSYS-SII, F-44344 Bouguenais, France*
[c] *Université Gustave Eiffel, Inria, COSYS-SII, I4S Team, F-35042 Rennes, France*

Abstract

Vibration measurements for structural health monitoring (SHM) by operational modal analysis (OMA) are classically obtained from sensors that are embedded in or physically attached to the monitored structure, like accelerometers or strain gauges. However, the setup time of these sensors and their restricted number and space coverage limit their monitoring capabilities. Video image-based sensing methods can overcome these shortcomings. With adequate image processing methods, motion signals are extracted from video image flows, which are then processed by system identification methods to estimate modal parameters. In this way, the pixels in selected regions of interest within the images act as a dense network of contactless sensors distributed over the whole structure. In this paper, the efficiency of this video-based approach is demonstrated with laboratory experiments on a cantilever beam, in particular, by evaluating its capability for detecting weak damages mimicked by slight mass modifications. To this end, the steerable filter-based method (ST), that recovers displacements from local phase, is first extended to overcome its motion limitation of one pixel size. Then, the performance of the improved motion extraction method is compared with two other well-established methods in the context of OMA, where natural frequencies, damping ratios and mode shapes with high spatial resolution are estimated together with their uncertainty bounds using covariance-driven subspace identification. The compared methods are evaluated with the help of reference laser displacement measurements as well as a finite element model of the beam, revealing differences in the accuracy of the estimated mode shapes depending on the chosen method for motion extraction. Finally, aiming to investigate early structural damage detection, experiments are carried out under small structural changes and the results are compared to a reference state with the help of estimated uncertainties. Small but statistically significant changes in the modal parameters are detected, showing the potential of the vision based framework for SHM.

Keywords: operational modal analysis; structural health monitoring; computer vision; displacement extraction; uncertainty quantification

---

1. Introduction

Civil engineering infrastructures (e.g. bridges, buildings) are facing increasing challenges due to fast demographic growth, construction densification and climate change. Moreover, many structures exceed their planned lifespan. Therefore, health monitoring of these systems becomes increasingly important. Structural health monitoring aims to assess the structural behavior, to evaluate their performance and safety during the life cycle [1].

Operational modal analysis is a crucial part of vibration-based SHM under operating conditions by extracting modal parameters, which are essential for structural dynamic behavior analysis on the one hand, and for damage detection, localization and quantification on the other hand [2–6]. It is therefore of great interest to provide a simple and affordable technology for reliable OMA. Traditional contact-based vibration measurement systems, e.g. by accelerometers or strain gauges, have practical shortcomings, such as the necessity of installing each of the sensors on

---

[*] Corresponding author
Email address: `boualem.merainani@univ-eiffel.fr` (Boualem Merainani)

the structure as well as the setup of a wired or wireless data transmission system. Furthermore, they can only offer a limited spatial coverage of a structure. In contrast, computer vision-based systems offer a promising alternative with a simpler setup and much higher spatial coverage and resolution. Nevertheless, the methods to process the vision data for OMA have not reached a full maturity yet and have shown some limitations, especially for small amplitude displacements under ambient excitation [7]. This motivates methods for motion extraction with high accuracy on one side, and a systematic analysis of the accuracy in the OMA context on the other side. In particular, the influence of the chosen motion extraction method on the accuracy and precision of OMA is in general unclear, since a systematic evaluation is lacking so far. Besides, the performance of vision-based OMA with respect to traditional measurement techniques is rarely evaluated.

In this paper, a framework for vision-based OMA with high accuracy is proposed by combining improved image processing techniques with state-of-the-art system identification methods. Moreover, its capability to detect small structural changes is demonstrated. In addition to the proposed framework, an important contribution of the paper is a toolset that allows the analysis of the accuracy and precision of the vision-based OMA framework, and subsequently the comparison of performance for different motion extraction methods in the framework. The toolset is based on reference laser measurements, a finite element model of the investigated structure and computed uncertainty bounds that are tied to the used motion extraction and system identification methods.

This paper is organized as follows. Section 2 provides an overview of related works, followed by the objectives. Section 3 briefly presents the theory of image registration and three subpixel refinement approaches for motion extraction, as well as the subspace-based method for modal analysis. In Section 4, the experimental setup is described for vision-based measurements of a cantilever beam in the laboratory. The modal parameters evaluated based on the different approaches for motion extraction are presented in Section 5, where they are analysed with respect to their precision and accuracy, and finally their evolution under structural changes is presented for the most accurate approach in Section 6. Section 7 gives the final conclusions.

2. Related works and objectives

About the context to this work, Section 2.1 introduces existing vibration sensing technologies used for OMA or structural monitoring, including a brief review of vision system technology. Section 2.2 reviews methods used for motion extraction from video images and Section 2.3 details the objectives of the paper.

2.1. Vibration sensing technologies for OMA and SHM

OMA relies on vibration measurements of the monitored structure subject to natural excitations such as wind and road traffic. Conventionally, contact-based sensors are installed on the structure, e.g., linear variable-differential transformers (LVDT), accelerometers or strain gauges [8–10]. Such contact-based measurements are limited to a moderate number of points of the structure. Furthermore, the installation and maintenance of the sensors and the data acquisition system are often difficult and costly, in particular in a SHM context [7]. Furthermore, the additional mass of sensors may induce changes in the dynamic behavior of some structures [11].

Non-contact based vibration measurement techniques include Laser Doppler vibrometry (LDV) [12], terrestrial laser scanner [13, 14], GPS technology [15, 16] and radar interferometry [17]. While offering alternatives to traditional measurement devices, they suffer from some drawbacks that still limit their wide application. Despite the high precision of LDV (higher than 0.05 mm) [18, 19], the price of such devices is very high and only displacements in the direction of the laser beam can be measured [20]. GPS sensor technology may show some limited accuracy for SHM application and/or low acquisition frequency [21, 22] .

Video cameras are an appealing technology for contactless measurements. Compared to traditional point measurement techniques, video cameras can provide full-field vibration measurements, virtually with each pixel acting as a sensor. They can range from industrial cameras with high resolution and sampling rate to consumer grade. The omnipresence of cameras, their flexible measurement distance and commonplace availability make them a promising candidate for widespread usage in SHM applications. A comprehensive review on this topic is given by Feng et al. [7], and an overview on recent monitoring applications is given by Zona [23].

The applied video-based technology depends on the camera system, with or without artificial targets, and on the applied image processing methods to extract the vibration data. For example, Luo et al. [24] adopted a high contrast artificial target vision-based approach for multi-point bridge displacement monitoring, and precision targets illuminated by infrared (IR) lamps or LED have been used in [21, 25]. Other works take advantage from any natural texture in the structure as long as it has a strong contrast compared with the surrounding background. For example, Dorn et al. [26] proposed a target-free approach, which allows full-field vibration measurements and does not require any target installation on the investigated structure. The applied image processing methods for vibration data extraction are diverse and are reviewed in the following section. To validate the vision technology, comparisons to traditional displacement measurement technology are made, e.g., via laser vibrometers [27] or LVDT sensors [21, 28].

2.2. *Motion extraction from video images*

To extract motion signals from video images, most methods follow a two-stage approach. By maximizing the cross-correlation between two images (the reference and shifted images), the integer pixel shift is first estimated. Then, subpixel accuracy can be obtained by various refinement methods classified into three categories: gradient-based, correlation-based interpolation and local phase-based subpixel shifts.

In the first category, the optical flow method [29, 30] has gained great importance for many applications. Chan et al. [31] proposed a simplified version of the optical flow. Based on this work, Javh et al. [32] have extracted subpixel motions of a vibrating steel beam and a cymbal, then performed modal analysis. Lu et al. [33] performed structural modal parameter identification by vision modal analysis. Zhang et al. [11] have combined the cross-correlation with modified Taylor approximation refinement to measure the displacements of a sound barrier on a bridge.

The second category is based on an interpolation in the vicinity of the cross-correlation peak to retrieve the subpixel shift [32, 34]. In [35], the authors have applied the digital image correlation technique to evaluate cable tensile forces on a cable-stayed bridge. In [36], the authors have proposed to up-sample the cross correlation (UCC) between the actual image and the reference one by means of the fast Fourier transform (FFT) to locate its peak and to extract the subpixel shift. Satisfactory results were found in [37] when evaluating the displacements based on UCC and comparing them to high-performance laser measurements. An enhanced normalized cross-correlation method was proposed in [38].

The last category relies on extracting the spatial local phase using the theory of steerable filters firstly introduced by [39]. The phase-based motion magnification algorithm developed by Wadhwa et al. [40] allows to highlight small variation in specific frequency bands. The method decomposes the video into amplitude and phase, at different scales, using the steerable pyramid filter bank. Then, the decomposed phases are band pass filtered, amplified and reconstructed back into video. Many researchers adopted the algorithm, e.g., for use in modal analysis [27, 41] or damage detection and localisation based on operational deflection shapes [42].

2.3. *Objectives*

The main objective of this paper is to further complete and advance the video-based approach to SHM, by combining improved image processing techniques with state-of-the-art system identification methods, demonstrating the capability of such a framework to detect small structural changes. A particular focus is put on the evaluation of the accuracy and precision of the proposed framework in order to allow the comparison of different methods for extracting motion signals from video images in the context of OMA.

First, three methods for motion extraction – one from each of the categories above (see section 2.2) – are selected for a comparison. A constraint for motion extraction from the local phase with the steerable filter is the limitation of less than one pixel displacement in the image coordinates. Beyond this condition, the method breaks down [43]. Since this condition may not always be satisfied for civil structures in operation, an algorithmic extension is proposed in this paper to extract displacements beyond the limitation of one pixel by combining the advantages of the ST with the zero-mean normalized cross-correlation (ZNCC). The approach is named EX-ST.

Second, the modal parameters are obtained by subspace-based system identification [44, 45]. Compared to other methods for modal analysis, subspace methods have favorable properties in terms of accuracy, precision and implementation [45, 46]. Moreover, they allow the computation of uncertainty bounds of the obtained modal parameter estimates [47–51]. These uncertainty quantification techniques are part of our proposed toolset to analyse the accuracy and precision of the vision-based OMA framework. In particular, this enables the evaluation of the precision of the modal parameter estimates obtained from the vision data, the comparison of different motion extraction methods within the framework, as well as the comparison with results from traditional measurement techniques. Up to the

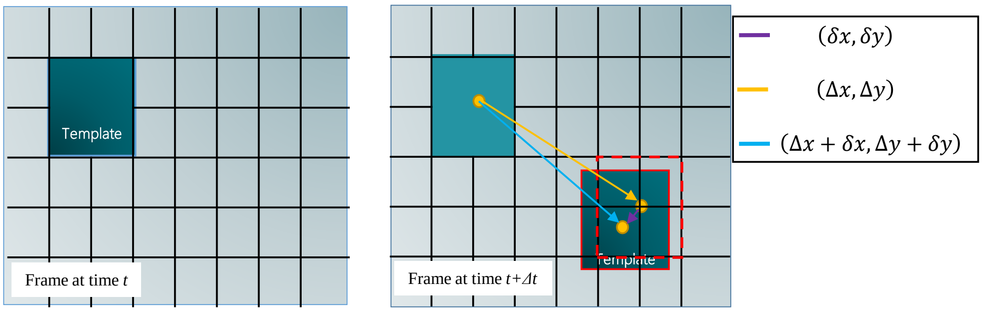


Figure 1: Illustration of overall-shift template image.

author's knowledge, this is the first time that uncertainty bounds of modal parameters are evaluated for vision data, showcasing the precision of the framework. The mode shapes obtained with such an approach are of particular interest due to their high spatial resolution, and thus their special usefulness, e.g. for finite element model updating or damage diagnosis. In this paper, it will be demonstrated that their accuracy depends on the accuracy of the selected method for motion extraction, which will be evaluated with the help of a finite element model.

Finally, the vision-based framework is applied under small structural changes. Such changes are detected based on statistically significant changes in the obtained natural frequencies and mode shapes [52], for which the computed uncertainty bounds are again useful. All the developments of this paper are demonstrated with laboratory experiments on a cantilever beam.

3. Operational modal analysis using vision-based displacement measurement

This section presents the toolset which allowing to perform the vision-based modal analysis at the operational level. First and foremost, we give a presentation of three image processing methods to extract, along the structure, the motion signals from image flows. Motion signals are then processed by subspace-based identification method for modal analysis

3.1. Motion extraction from video frames

To extract motions from video images, the pixel shifts between two successive frames at time t and $t + \Delta t$ in the horizontal and/or vertical directions should be effectively revealed. Such shifts are often estimated through a two-step process, e.g. [11], as illustrated in Fig. 1. First, the cross-correlation matrices between the reference image and every object image in the following video sequence are calculated. The coordinates of the maximum values in the cross-correlation matrices give the pixel level (integer) shifts, as shown in Section 3.1.1. Second, refinements through subpixel shifts estimation $(\delta x, \delta y)$ are performed. In this work, three subpixel accuracy refinement methods are investigated and compared, one from each of the categories described in the introduction: the first one is based on Taylor approximation (Section 3.1.2), the second one is based on quadratic surface fitting (Section 3.1.3) and the last one is a phase-based method using steerable filters (Section 3.1.4). An algorithmic extension of the phase-based method to deal with displacements beyond the limitation of one pixel is proposed in Section 3.1.5.

3.1.1. Image registration using cross correlation

Commonly, the integer pixel displacement $(\Delta x, \Delta y)$ is estimated by locating the maximum value in the cross-correlation matrix calculated between two image templates, of size $X \times Y$, at two different times $I(x, y, t)$ and $I(x, y, t + \Delta t)$ as [53]

$$(\Delta x^*, \Delta y^*) = \arg \max_{\substack{-X \leq \Delta x \leq X \\ -Y \leq \Delta y \leq Y}} \mathrm{CC}(\Delta x, \Delta y), \tag{1}$$

where $\mathrm{CC}(\Delta x, \Delta y)$ is the cross correlation defined as

$$\mathrm{CC}(\Delta x, \Delta y) = \sum_{(x,y)\in\mathbb{T}} I(x - \Delta x, y - \Delta y, t + \Delta t) I(x, y, t). \tag{2}$$

where $\mathbb{T}$ is the set of the pixel indexes of the image templates.

According to [54], the CC is sensitive to brightness and contrast variation. By normalizing Eq. (2) and subtracting the mean of both templates, the problem can be overcome. These operations gives rise to the zero mean normalized cross correlation given by

$$\mathrm{ZNCC}(\Delta x, \Delta y) = \frac{\sum\limits_{(x,y)\in\mathbb{T}} [I(x,y,t)-\bar{I}(t)][I(x-\Delta x,y-\Delta y,t+\Delta t)-\bar{I}(t+\Delta t)]}{\sqrt{\sum\limits_{(x,y)\in\mathbb{T}} [I(x,y,t)-\bar{I}(t)]^2}\sqrt{\sum\limits_{(x,y)\in\mathbb{T}} [I(x-\Delta x,y-\Delta y,t+\Delta t)-\bar{I}(t+\Delta t)]^2}} \tag{3}$$

where

$$\bar{I}(t) = \frac{1}{XY} \sum_{(x,y)\in\mathbb{T}} I(x,y,t) \tag{4}$$

$$\bar{I}(t+\Delta t) = \frac{1}{XY} \sum_{(x,y)\in\mathbb{T}} I(x-\Delta x, y-\Delta y, t+\Delta t) \tag{5}$$

3.1.2. Taylor approximation refinement

Once the pixel shift (Δx, Δy) is estimated, the shift between $I(x, y, t)$ and $I(x + \Delta x, y + \Delta y, t + \Delta t)$ is regarded as less than one pixel. Define $\tilde{x} = x + \Delta x$ and $\tilde{y} = y + \Delta y$, then the relationship between two consecutive templates can be written as [11, 31]

$$\begin{aligned} I(x,y,t+\Delta t) &= I(\tilde{x}+\delta x, \tilde{y}+\delta y, t) \\ &\approx I(\tilde{x},\tilde{y},t) + \frac{\partial I(\tilde{x},\tilde{y},t)}{\partial x}\delta x + \frac{\partial I(\tilde{x},\tilde{y},t)}{\partial y}\delta y, \end{aligned} \tag{6}$$

which is the first order Taylor series approximation. The partial derivatives are estimated by finite difference between two images. In Eq. (6) there are two unknown parameters, δx and δy. The linear least-squares estimator minimizes the square error as

$$(\hat{\delta x}, \hat{\delta y}) = \arg\min_{(\delta x,\delta y)} \Phi(\delta x, \delta y) \tag{7}$$

with

$$\Phi(\delta x,\delta y) \triangleq \sum_{(x,y)\in\mathbb{T}} [I(x,y,t+\Delta t) - I(\tilde{x},\tilde{y},t) - \frac{\partial I(\tilde{x},\tilde{y},t)}{\partial x}\delta x - \frac{\partial I(\tilde{x},\tilde{y},t)}{\partial y}\delta y]^2. \tag{8}$$

The solution to this standard least squares problem can be obtained by solving

$$\frac{\partial\Phi(\delta x,\delta y)}{\partial\delta x} = 0, \quad \frac{\partial\Phi(\delta x,\delta y)}{\partial\delta y} = 0 \tag{9}$$

for δx and δy, which can be rewritten into

$$\begin{bmatrix} \sum_{x,y}(\frac{\partial I(\tilde{x},\tilde{y},t)}{\partial x})^2 & \sum_{x,y}\frac{\partial I(\tilde{x},\tilde{y},t)}{\partial x}\frac{\partial I(\tilde{x},\tilde{y},t)}{\partial y} \\ \sum_{x,y}\frac{\partial I(\tilde{x},\tilde{y},t)}{\partial x}\frac{\partial I(\tilde{x},\tilde{y},t)}{\partial y} & \sum_{x,y}(\frac{\partial I(\tilde{x},\tilde{y},t)}{\partial y})^2 \end{bmatrix} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = \begin{bmatrix} \sum_{x,y}(I(x,y,t+\Delta t) - I(\tilde{x},\tilde{y},t))\frac{\partial I(\tilde{x},\tilde{y},t)}{\partial x} \\ \sum_{x,y}(I(x,y,t+\Delta t) - I(\tilde{x},\tilde{y},t))\frac{\partial I(\tilde{x},\tilde{y},t)}{\partial y} \end{bmatrix}. \tag{10}$$

With the solution for refined subpixel level displacement, the total displacement can be written as

$$d_x = \Delta x + \delta x, \; d_y = \Delta y + \delta y. \tag{11}$$

3.1.3. Quadratic surface fitting

Different from the Taylor approximation refinement method, the quadratic surface equation is fitted within a 3 x 3 neighborhood on the cross-correlation maximum matrix [34] after estimating the pixel shift (Δx, Δy). First, the

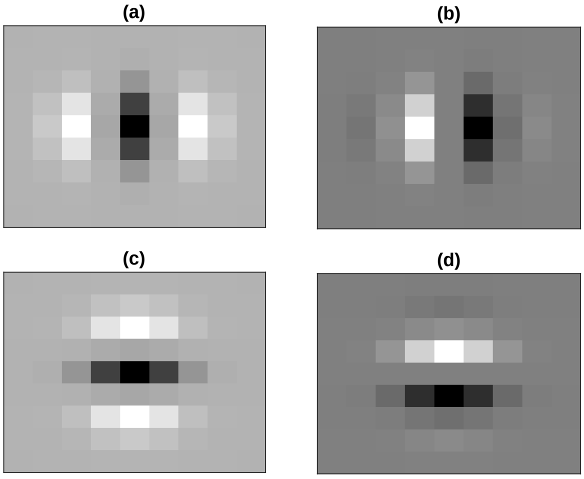


Figure 2: Representation of the Filters used to compute local phase and local amplitude, (a) real horizontal(G_2^0), (b) imaginary horizontal(H_2^0), (c) real vertical ($G_2^{\pi/2}$), (d) imaginary vertical($H_2^{\pi/2}$).

zero-frequency component is moved to the center of the 2D DFT cross-correlation results. The quadratic surface fitted by the nine points can be written as

$$S(x, y) = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2 \tag{12}$$

where x, y are relative coordinates respect to position of the maximum cross-correlation peak. The coefficients $a_i, i = 0 \ldots 5$, can be estimated through a surface model $S(x, y)$ after nine points are substituted into the quadratic surface equation

$$\begin{bmatrix} S(-1,1) & S(0,1) & S(1,1) \\ S(-1,0) & S(0,0) & S(1,0) \\ S(-1,-1) & S(0,-1) & S(1,-1) \end{bmatrix} = \begin{bmatrix} CC(x_i-1, y_i+1) & CC(x_i, y_i+1) & CC(x_i+1, y_i+1) \\ CC(x_i-1, y_i) & CC(x_i, y_i) & CC(x_i+1, y_i) \\ CC(x_i-1, y_i-1) & CC(x_i, y_i-1) & CC(x_i+1, y_i-1) \end{bmatrix} \tag{13}$$

where CC is the cross-correlation matrix between the template image and the object image, and (x_i, y_i) are the coordinates of the maximum value in ZNCC. After obtaining the coefficients of the quadratic surface, the extreme point can be easily calculated by

$$\delta x = \frac{2a_1a_5 - a_2a_4}{a_4{}^2 - 4a_3a_5}, \; \delta y = \frac{2a_2a_3 - a_1a_4}{a_4{}^2 - 4a_3a_5}, \tag{14}$$

and the total displacement follows as in (11).

3.1.4. Phase-based displacement extraction using steerable filters

The steerable filter method have been proposed in [27] to extract the displacement between two images from the phase change. In its original version, the method assumes limited displacement magnitude of the order of one pixel or smaller [43, 55]. As for the Taylor-based method, the steerable filter is based on the assumption of constant brightness of the scene and considers the change of pixel intensity related to image translation [27, 56]. The video frames are convoluted with steerable quadrature filters ($G^\theta + \mathrm{i}H^\theta$) to determine, at each pixel, the local phase and amplitude. In this paper, the real part is selected to be the frequency response of the second derivative of a Gaussian (G_2) and the imaginary part is its Hilbert transform (H_2) [39]. In Fig. 2, the 9-tap steerable filter kernels at orientations 0 and $\pi/2$ are shown.

For an image sequence $I(x, y, t)$, the local amplitude A and the local phase ϕ at the orientation θ are computed as

$$A_\theta(x, y, t)e^{\mathrm{i}\phi_\theta(x,y,t)} = (G_2^\theta + \mathrm{i}H_2^\theta) \otimes I(x, y, t) \tag{15}$$

where $\otimes$ is the convolution operator.

Under some assumptions, the local phases can be used to calculate the velocity in units of pixels, [43, 57, 58], by

$$\begin{aligned} u &= -\left(\tfrac{\partial \phi_0(x,y,t)}{\partial x}\right)^{-1} \tfrac{\partial \phi_0(x,y,t)}{\partial t}, \\ v &= -\left(\tfrac{\partial \phi_{\pi/2}(x,y,t)}{\partial y}\right)^{-1} \tfrac{\partial \phi_{\pi/2}(x,y,t)}{\partial t}, \end{aligned} \tag{16}$$

where u and v are the velocity in the x and y directions, respectively. Based on these expressions, the horizontal and vertical displacements δx and δy are directly obtained by [27, 43]

$$\begin{aligned} \delta x &= -\left(\tfrac{\partial \phi_0(x,y,t)}{\partial x}\right)^{-1} (\phi_0(x,y,t) - \phi_0(x,y,t_0)), \\ \delta y &= -\left(\tfrac{\partial \phi_{\pi/2}(x,y,t)}{\partial y}\right)^{-1} (\phi_{\pi/2}(x,y,t) - \phi_{\pi/2}(x,y,t_0)). \end{aligned} \tag{17}$$

In fact, the obtained displacements are valid in the image regions with strong enough texture. Thus, instead of processing the entire images, the motion computation may be limited to some pixels, labelled as active pixels in the following. These pixels are usually selected on the first frame, from the extremum of the steerable filter output magnitude. These pixels form a mask which is then applied on the whole image frames for motion extraction.

3.1.5. Extension of the local phase-based method to multi-pixel displacement

When the motion magnitude of the structure is beyond one pixel, the phase signal goes beyond a change of π in phase, and the method breaks down. One common adopted solution is to down-sample the video, before being filtered, in order to recover the subpixel condition limit, which however reduces the image contrast and smooths the image texture. As a consequence, the number of active pixels is dramatically reduced. Furthermore, high frequency information may be lost when down-sampling [59]. To address this limitation, we propose to relate the steerable approach to the general two-steps process, presented in Section 3.1.2, in order to determine multi-pixel displacement without down sampling, as follows. After defining the mask, the ZNCC given in Section 3.1.1 is employed to estimate the integer pixel shift $(\Delta x, \Delta y)$ between the reference image $I(x, y, t_0)$ and the actual image $I(x, y, t)$. If a pixel shift exists, the following steps are taken:

① Shift the mask by $(\Delta x, \Delta y)$.

② To compute the phase change on the current frame, Replace temporarily the reference phases $\phi_0(x, y, t_0)$ and $\phi_{\pi/2}(x, y, t_0)$, computed on the first frame and given in Eq. (17) by the phases of the previous frame at time $t - \Delta t$, i.e $I(x, y, t - \Delta t)$, which are $\phi_0(x, y, t - \Delta t)$ and $\phi_{\pi/2}(x, y, t - \Delta t)$.

③ Accumulate the estimated subpixel motion as

$$d_x(t) = \delta x(t) + d_x(t - \Delta t), \;\; d_y(t) = \delta y + d_y(t - \Delta t)$$

Finally, when the pixel shift is zero or comes back to zero, the reference phases are reconsidered as $\phi_0(x, y, t_0)$ and $\phi_{\pi/2}(x, y, t_0)$. The flowchart of the proposed approach is shown in Fig. 3.

3.2. Operational modal analysis with subspace-based system identification

Assume that the vibration behavior of the investigated structure can be modelled by a linear time-invariant system, then its dynamics can be described by the discrete-time state space model [46]

$$\left\{ \begin{array}{rcl} \mathbf{x}_{k+1} & = & \mathbf{A}\mathbf{x}_k + \mathbf{w}_k \\ \mathbf{y}_k & = & \mathbf{C}\mathbf{x}_k + \mathbf{v}_k \end{array} \right. , \tag{18}$$

where k is the integer time step corresponding to the system at time $t = k\Delta t$, $\mathbf{y}_k \in \mathbb{R}^r$ is the output vector containing the computed displacements $d_x(k\Delta t)$ and $d_y(k\Delta t)$ for the r selected templates or active pixels of the video, and $\mathbf{x}_k \in \mathbb{R}^n$ is the state vector. The state noise $\mathbf{w}_k \in \mathbb{R}^n$ is related to the unknown ambient excitation, and vector $\mathbf{v}_k \in \mathbb{R}^r$ describes the measurement noise. The modal parameters are related to the eigenvalues and eigenvectors (λ_i, ϕ_i), $i = 1, \ldots, n$, of $\mathbf{A}$ and to $\mathbf{C}$ by

$$\mu_i = \frac{\log(\lambda_i)}{\Delta t}, \;\; f_i = \frac{|\mu_i|}{2\pi}, \;\; \zeta_i = \frac{-\mathrm{Re}(\mu_i)}{|\mu_i|}, \;\; \varphi_i = \mathbf{C}\phi_i, \tag{19}$$

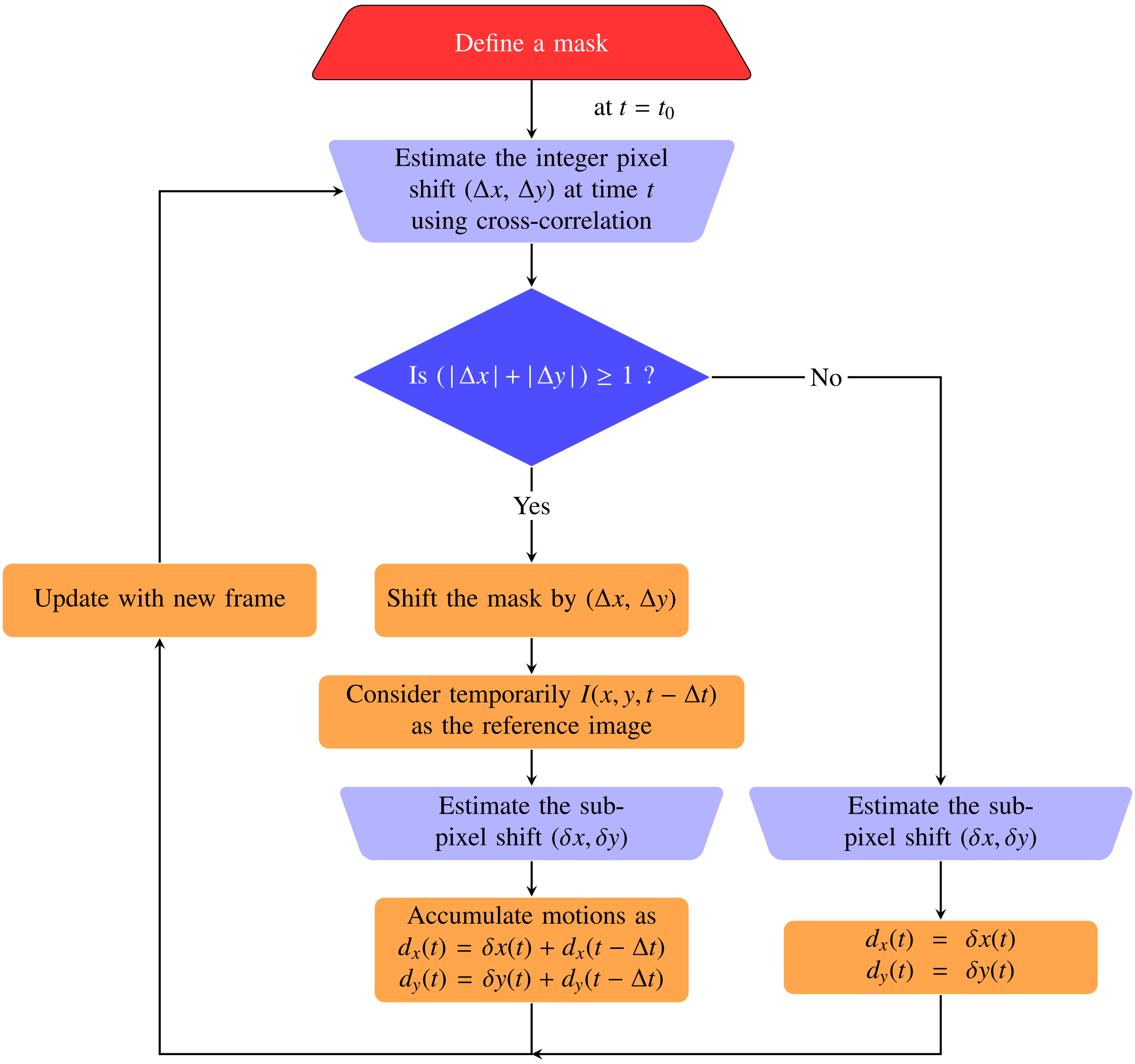


Figure 3: Flowchart of the proposed extension of the phase-based method to multi-pixel displacements.

where μ_i is an eigenvalue of the corresponding continuous-time system, f_i is the natural frequency, ζ_i is the damping ratio and φ_i is the mode shape at the output coordinates.

To estimate the system matrices $\mathbf{A}$ and $\mathbf{C}$ from the output data $\mathbf{y}_k$ of length N, $k = 1, \ldots, N$, and consequently the modal parameters in (19), the reference-based covariance-driven subspace algorithm [44, 45] is used, and the standard deviations of the modal parameter estimates are computed as in [48].

4. Experimental setup for validations

The performance of the different methods for motion extraction and the efficiency of the video-based approach to SHM were verified by a set of experiments conducted on a cantilever beam structure of size 30 mm × 2000 mm × 5 mm mounted on a shake table in the Structures and Integrated Instrumentation Laboratory at Gustave Eiffel University.

4.1. Beam model

The finite element (FE) model of the beam will mainly be used for the comparison of the analytic mode shapes with the estimated ones from data. The steel cantilever beam is modelled using the Timoshenko beam theory. The model is adopted from [60] and illustrated in Fig. 4 (left). The Young's modulus, mass density and the Poisson's ratio of the beam material is $2.1 \cdot 10^{11}$ Pa, $7850\,\mathrm{kg/m^3}$ and 0.21, respectively.

The model is composed of 160 beam elements, which will conveniently correspond to the number of regions of interests (ROIs) into which the video frames will be divided. Every node has three rotational and three translational degrees of freedom.

The beam model is also used to define positions of additional masses that will be used to emulate different small damages in the experiments. With a preliminary analysis, eight positions and an additional mass of 0.044 kg (corresponding to 1.86% of the beam mass) were chosen, such that changes in the modal parameters are moderate. The chosen positions are designated by their distance in mm from the support as as $P2000$, $P1803$, $P1507$, $P1277$, $P949$, $P686$, $P456$ and $P260$, and are illustrated in Fig. 4 (right).

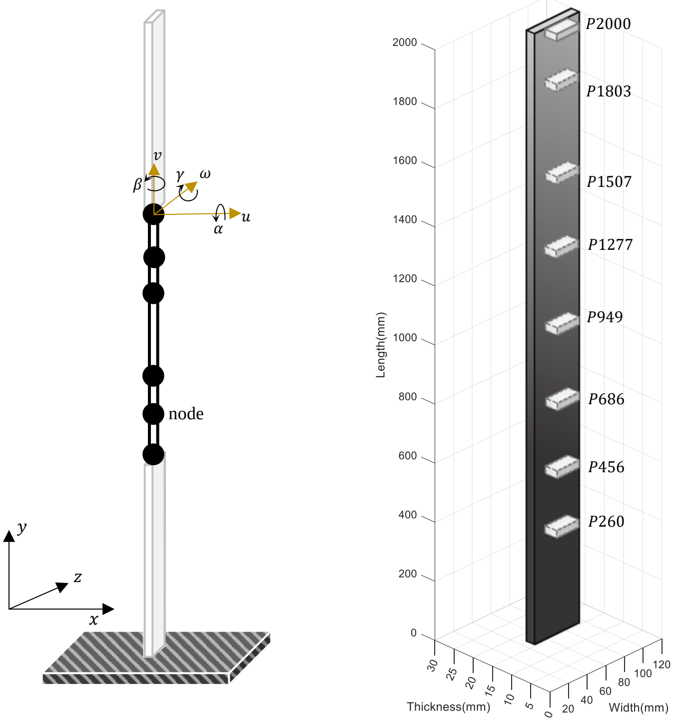


Figure 4: Illustration of cantilever beam FE model (left); choice of positions for adding mass for damage detection experiments (right).

4.2. *Experimental setup*

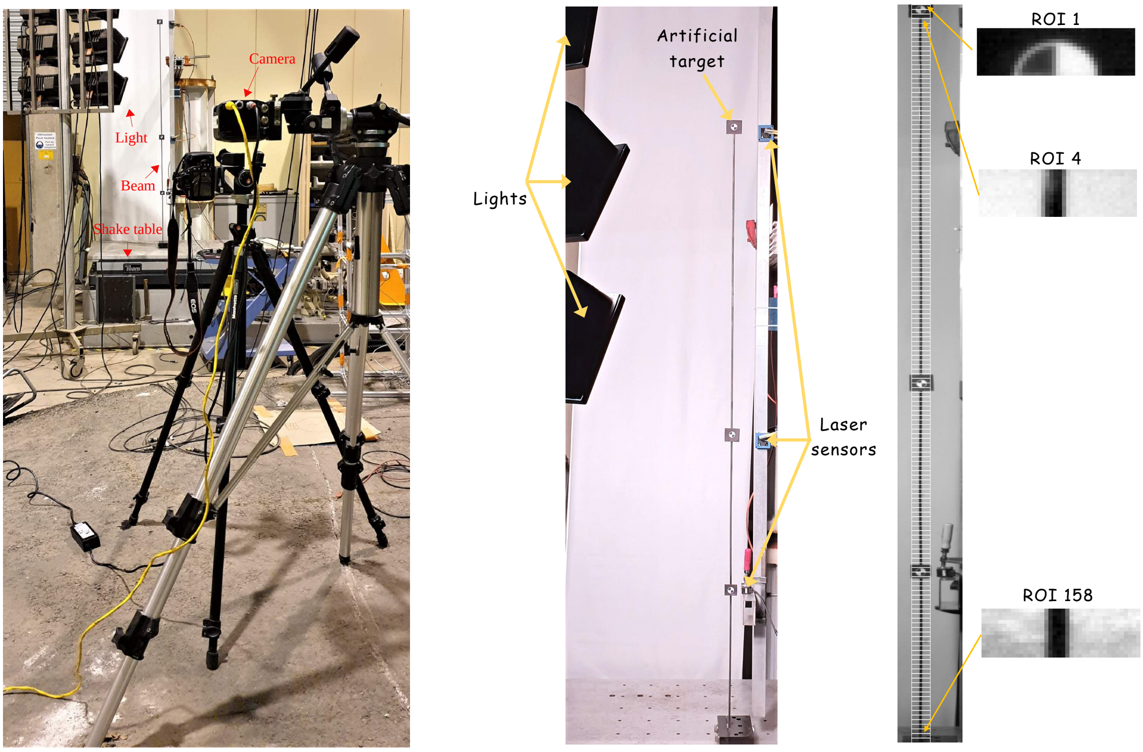
The camera used for recording the video flows was a PHANTOM MIRO 320S, equipped with a CMOS sensor with resolution 1920 × 1200 and mounted with a NIKKOR lens with a fixed 50 mm focal length.

The camera was set on a tripod. The effective distance from the camera to the beam was fixed such that the full beam length will be captured. The camera was used in portrait mode to take advantage of the full 1920 pixels along the beam in vertical direction.

The number of pixels for the horizontal direction was adjusted to 136 pixels to gain space in the camera's dynamic random access memory, while ensuring that the beam remains in the scene when vibrating. The frame rate was chosen as 512 frames per second, ensuring a good identification of the first six modes. The exposure time was selected relatively short, so that the camera can collect enough light. Three additional lights (Halogen Spotlight, MAZDA CORMORAN type of 1500 W) and a white background were used to provide optimal brightness and contrast conditions. Three black and white artificial targets are mounted at three particular positions on the beam (top (1/1), middle (1/2) and 50 cm from the base (1/4)) at the positions of three high-accuracy laser displacement sensors (LDS) for comparison between displacements extracted from vision measurements and displacements measured by LDS.

Table 1: Specifications of the LDS.

Features	Model	
	Keyence IL-030	Wenglor YP11MGVL80
Measuring range	20 to 45 mm	50 to 100 mm
Reference distance	30 mm	50 mm
Wavelength of light source	655 nm	655 nm
Resolution	n/a	25 μm
Repeatability	1 μm	6 μm



(a) (b) (c)

Figure 5: Experimental setup: (a) video cameras in the foreground, the cantilever beam in the background, (b) closer view of the cantilever beam set, (c) raw frame from the video on which the regions of interest are indicated. Three of them are represented on the right hand side.

Two Wenglor (top and middle) sensors and one Keyence (bottom) sensor were used, whose specifications are given in Tab. 1 based on the data sheets, except for the values for the repeatability that are experimentally obtained when the beam was at rest. The experimental installation is shown in Fig. 5.

For data collection with the camera and the LDS, the shake table (Model: NEBS GR-63, by team vibration testing equipment) was driven in random mode horizontally, perpendicular to the camera axis. The excitation noise was band limited between 1 Hz and 200 Hz. Datasets of 33 seconds length were recorded without and with the added mass at the different positions shown in Fig. 4 (right). For processing with the different motion extraction methods, the video frames were separated to 160 ROIs, in total, of size 12×37 pixels. The horizontal displacements from each of the ROIs were estimated by the three introduced methods, namely Taylor approximation (Section 3.1.2), quadratic surface fitting (QSF, Section 3.1.3) and the proposed EX-ST (Section 3.1.5) that all share the ZNCC technique (Section 3.1.1) as a prior step. The computation time per frame of each algorithm is shown in Table 2, indicating that EX-ST takes

Table 2: Computation time to extract displacements with the EX-ST, QSF and Taylor algorithms.

	EX-ST	QSF	Taylor
Time per frame (ms)	1.01	0.53	0.48

about twice the time than Taylor or QSF.

5. Evaluation of motion extraction methods in OMA context

In this section, the different motion extraction algorithms are evaluated in the context of modal analysis using data from the beam in the reference condition (without added mass), and compared to results from the laser data. First, a direct comparison of the extracted displacement data to the laser data is made in terms of the root mean squared error. Second, the estimated natural frequencies and damping ratios and their uncertainty bounds obtained with the covariance-driven subspace algorithm are evaluated. Finally, the quality of the mode shape estimates is evaluated for the different methods based on the FE mode shapes.

5.1. Direct comparison to laser data

Since time synchronization between the laser sensors and the camera was not possible, the laser data was synchronized a posteriori with the image flows by using cross correlation between signals. Furthermore, the extracted displacements from the video were converted from pixels into millimeters by using the pinhole camera model through scaling the distance between two targets in the world coordinate with its pixel distance in the image coordinate. The scaling factor was, thus, estimated to be 1.007 mm/pixel.

In Fig. 6 the motion signals extracted from the vision and laser systems are shown at the upper laser sensor (position 1/1), corresponding to ROI 2 in the motion extraction. The signals are represented on top of each other for comparison. It is observed that the displacements obtained from any of the three motion extraction algorithms seem to match very well with the reference laser data for the whole dataset of 33 s length. To quantify the differences between the vision and the laser data, the root mean squared error

$$\mathrm{RMSE} = \left(\frac{1}{N} \sum_{k=1}^{N} \left(y_k^{(\mathrm{vision})} - y_k^{(\mathrm{LDS})} \right)^2 \right)^{1/2} \tag{20}$$

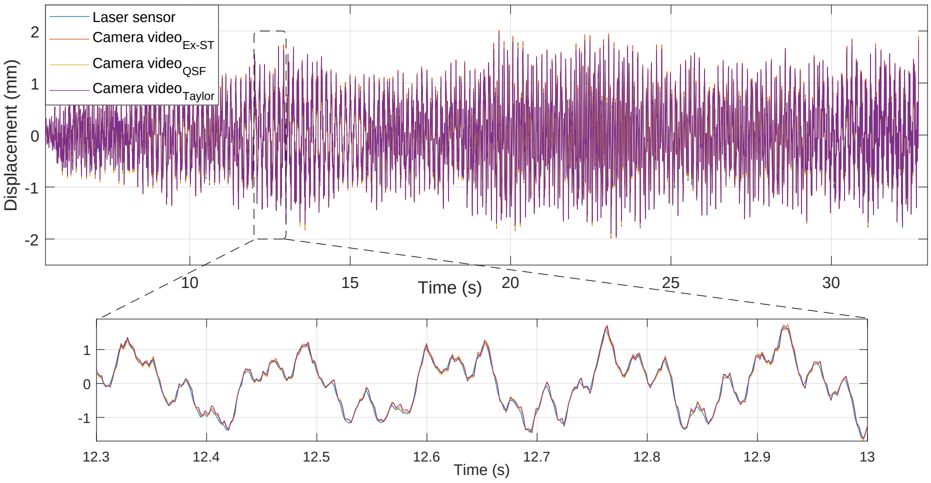
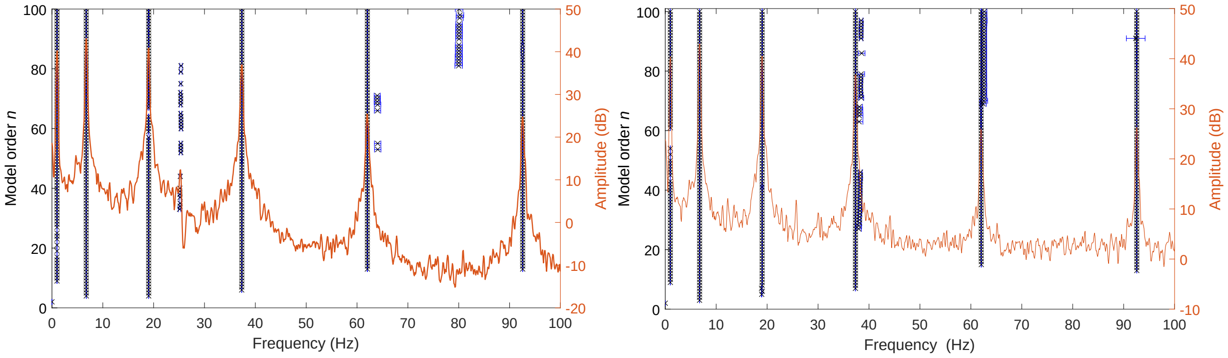


Figure 6: Comparison of the beam displacements at the top (position 1/1) measured by the LDS and extracted from the vision data, with zoom on a time interval.

Table 3: RMSE of the extracted displacements with the EX-ST, QSF and Taylor algorithms with respect to the reference laser data.

RMSE (mm) / Position	Motions extraction algorithms		
	EX-ST	QSF	Taylor
(1/1)	0.0404	0.0430	0.0461
(1/2)	0.0342	0.0348	0.0422
(1/4)	0.0438	0.0451	0.0481



(a)

(b)

Figure 7: Stabilization diagrams with uncertainty bounds (horizontal bars) of the natural frequencies, together with power spectral density. (a): Obtained from laser data, (b): Obtained from vision-data with EX-ST algorithm.

is computed, where $y_k^{(\text{vision})}$ and $y_k^{(\text{LDS})}$ are the extracted displacement data from the vision data and the LDS, respectively. The RMSE values for all three LDS positions (cf. Fig. 5, right) are shown in Table 3, where it is indicated that the proposed EX-ST method provides a lower RMSE compared to ones obtained with the QSF and Taylor methods.

5.2. *Operational modal analysis*

The extracted displacements from the vision data, using the EX-ST, QSF and Taylor algorithms, as well as the LDS data, have been used to obtain the respective sets of modal parameters for each of the algorithms. The reference-based covariance-driven subspace method [44, 45] has been applied for this task, together with the related uncertainty quantification to obtain the confidence intervals of the estimates [48]. With the considerable number of $r = 160$ available virtual sensors from the vision data, three reference sensors were selected for the analysis, namely the ones coinciding with the laser sensor positions. Since the first mode of the structure is expected around 1 Hz and the sampling frequency is at 512 Hz, the number of time lags for the subspace algorithm needed to be set relatively high, and altogether 300 lags have been used for the computation of the covariance-driven Hankel matrix. The identification has been carried out at model orders $1, \dots, 100$, and the resulting stablization diagram of the natural frequencies is shown together with uncertainty bounds and the power spectral density in Fig. 7 for the laser data and the vision data from the proposed EX-ST algorithm. Since physical modes remain stable and have much lower uncertainty bounds than spurious modes due to noise, a threshold of 1.5% was put on the coefficient of variation of the frequencies (standard deviation divided by frequency), leading to a clear and easily interpretable diagram. Obviously, six stable modes are clearly distinguished. Their quality is analysed in the following subsections with respect to the underlying motion extraction methods.

5.2.1. *Precision of the natural frequencies and damping ratio estimates*

The estimated frequencies and damping ratios and their $\pm 2\sigma$ confidence intervals, obtained from the vision data with each of the three motion extraction methods as well as the LDS, are stated in Tab. 4 and visualized in Fig. 8. The confidence intervals are in absolute terms and have the same units as the frequencies and damping ratios, respectively.

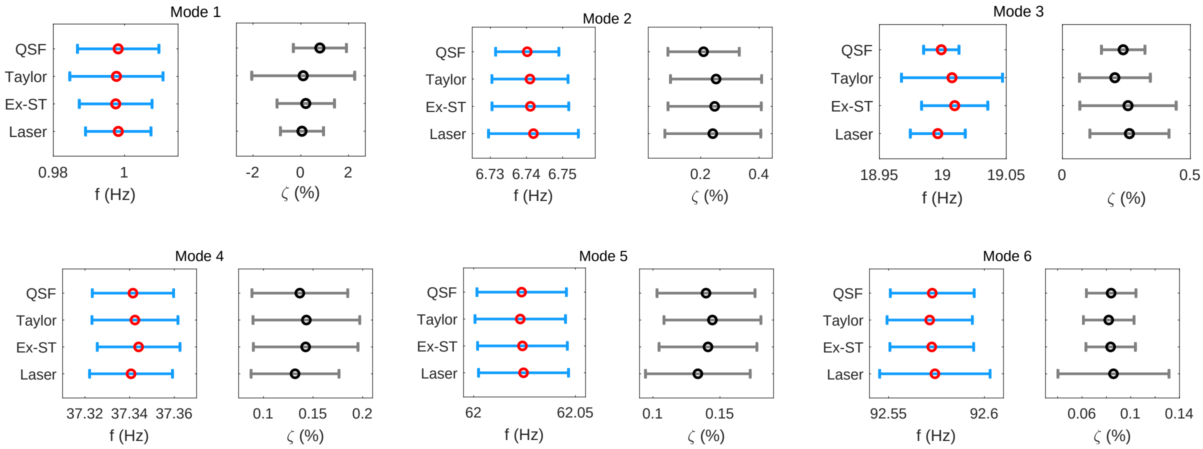


Figure 8: Natural frequencies and damping ratios with their ±2σ confidence intervals obtained from the LDS data and the from the vision data using the EX-ST, QSF and Taylor algorithms.

It can be noted that the first mode has the highest uncertainties, especially its damping ratio. This is due to higher uncertainties in damping estimation in general [61, 62], and due to the difficult estimation of this mode in particular, which is at a considerably lower frequency than all other modes while the dataset is relatively short.

For all six modes the results show close correspondence although the modal parameter estimates vary slightly. The confidence intervals obtained from the different datasets overlap, and each of the frequencies and damping ratios is within all the confidence intervals of the other three estimates, showing that the frequency and damping ratio estimates obtained from the three motion extraction methods as well as the LDS are statistically coherent. Moreover, the size of the confidence intervals is similar for the most part for the different methods, with a bit more variation for some of the damping estimates which is naturally to be expected. This shows that the precision of the modal parameter estimates is also similar for the different methods; hence the choice of the motion extraction method does not seem to have a significant influence on the natural frequency and damping ratio estimates.

5.2.2. *Accuracy of the vision-based mode shapes*

Fig. 9 compares the experimentally obtained mode shapes determined by the LDS and vision data using the three motion extraction methods with those obtained from the FE model. At first glance, the vision-based mode shapes follow a similar pattern as the numerical ones and have close correspondence. Those obtained from the LDS suffer from low spatial resolutions with only three measurement points, while the vision-based mode shapes have a high

Table 4: Natural frequency and damping ratio estimates with their ±2σ confidence intervals.

Mode	Laser sensor		EX-ST		Taylor		QSF	
	f (Hz)	ζ (%)	f (Hz)	ζ (%)	f (Hz)	ζ (%)	f (Hz)	ζ (%)
1	0.998 ±0.018	0.05 ±1.81	0.998 ±0.020	0.21 ±2.40	0.998 ±0.026	0.10 ±4.31	0.998 ±0.023	0.80 ±2.22
2	6.742 ±0.025	0.24 ±0.33	6.741 ±0.021	0.25 ±0.32	6.741 ±0.021	0.25 ±0.31	6.740 ±0.018	0.21 ±0.24
3	19.996 ±0.043	0.26 ±0.31	19.009 ±0.052	0.26 ±0.38	19.007 ±0.080	0.21 ±0.28	18.999 ±0.028	0.24 ±0.17
4	37.341 ±0.037	0.13 ±0.09	37.344 ±0.037	0.14 ±0.11	37.342 ±0.039	0.14 ±0.11	37.342 ±0.036	0.14 ±0.10
5	62.025 ±0.044	0.13 ±0.08	62.024 ±0.044	0.14 ±0.07	62.023 ±0.045	0.14 ±0.07	62.024 ±0.044	0.14 ±0.07
6	92.574 ±0.058	0.09 ±0.09	92.573 ±0.044	0.08 ±0.04	92.572 ±0.045	0.08 ±0.04	92.573 ±0.044	0.08 ±0.04

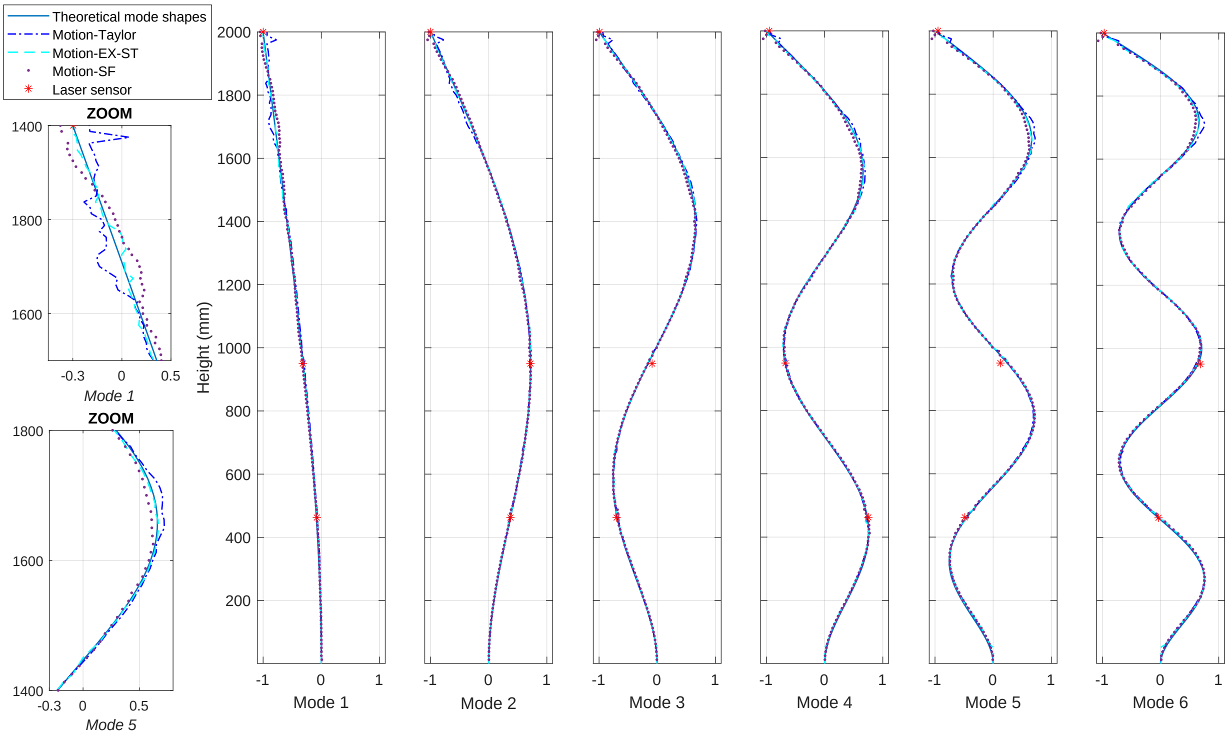


Figure 9: Comparison between the identified mode shapes with those from FE model

resolution of 160 points.

A close examination, however, reveals that errors in motion estimation have a direct impact on the identified mode shapes. For instance, the mode shapes computed from the motion data from the Taylor approximation method present more fluctuations that are visible on the beam top, as can be seen in the zoom in Fig. 9. Those computed based on QSF are smoother but the gap is larger near the anti-node positions. The mode shapes obtained based on the proposed EX-ST method seem to match the best with the FE predicted ones.

These results are confirmed by a comparison of the modal assurance criterion (MAC) between the 6 numerical and the vision-based mode shapes obtained from the three different motion extraction methods in Fig. 10, where the MAC between two complex-valued mode shape vectors φ and ψ is defined by [52]

$$\mathrm{MAC}(\varphi, \psi) = \frac{|\varphi^H \psi|^2}{\varphi^H \varphi \psi^H \psi}. \tag{21}$$

The MAC value is bounded between 0 and 1, and values near 1 indicate a strong similarity between both mode shapes. In Fig. 10 it can be seen that all MAC values are very high (> 0.99); however the mode shapes based on the proposed EX-ST algorithm stand out among the others and offer the highest MAC values among the different motion extraction methods for all modes. Consequently, the proposed EX-ST algorithm was adopted for further analysis in the next section.

6. Application to detect structural changes through added mass

Finally, the vision-based technology is evaluated in a monitoring context, where a mass is attached at different positions to the beam (see Fig. 4, right) to emulate different damages. Based on the chosen EX-ST algorithm, the modal parameters are estimated for each structural state and compared to the reference state.

For the modal parameter comparison in the different structural states, a simple approach is adopted from [52] based on the natural frequencies and mode shapes together with their uncertainty bounds to infer if changes with respect

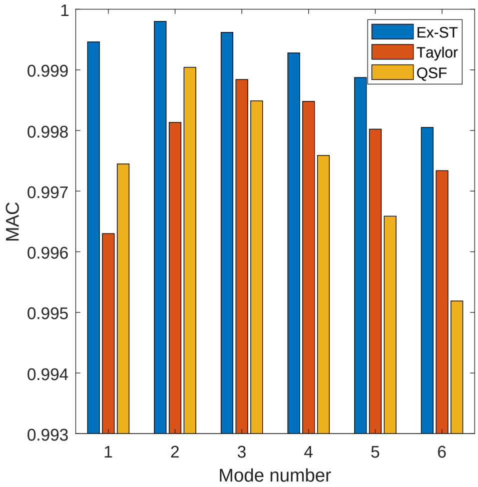


Figure 10: MAC values between vision-based mode shapes based on the three motion extraction methods and numerical mode shapes from FE model.

to the reference are statistically significant. To this end, let the reference modal parameters $(\hat{f}_i^{\text{ref}}, \hat{\varphi}_i{}^{\text{ref}})$, $i = 1, \ldots, 6$, with their uncertainties (frequency standard deviation $\sigma_{f_i}^{\text{ref}}$, mode shape covariance $\Sigma_{\varphi_i}^{\text{ref}}$) be estimated in the mass-free condition, and let the modal parameters $(\hat{f}_i, \hat{\varphi}_i)$ with their uncertainties $(\sigma_{f_i}, \Sigma_{\varphi_i})$ be computed from the vision test data under a structural change. Then, a statistically significant change in frequencies or mode shapes can be inferred for a confidence level of 95% as follows:

- *Frequency change:* The confidence bound for each frequency in the reference state is $[\hat{f}_i{}^{\text{ref}} - 2\sigma_{f_i}^{\text{ref}}, \hat{f}_i{}^{\text{ref}} + 2\sigma_{f_i}^{\text{ref}}]$, since frequency estimates are approximately Gaussian distributed. If $\hat{f}_i$ is not in this interval, then label it as changed.

- *Mode shape change:* The MAC equals one for identical mode shapes. Due to statistical uncertainties it is only close to one when considering estimates of identical mode shapes. To check the hypothesis that $\hat{\varphi}_i^{\text{ref}}$ and $\hat{\varphi}_i$ correspond to the same, unchanged mode shape, $\text{MAC}(\hat{\varphi}_i^{\text{ref}}, \hat{\varphi}_i)$ and its uncertainty is computed. As shown in [52], the MAC distribution is not Gaussian under the hypothesis of identical mode shapes, but approximated by a scaled and shifted χ^2. Then, the quantile t_{MAC} of the distribution can be computed for the chosen 95% confidence level based on the mode shape covariances $\Sigma_{\varphi_i}^{\text{ref}}$ and Σ_{φ_i}, leading to the MAC confidence bound $[t_{\text{MAC}}, 1]$ as detailed in [52]. If the computed MAC is outside this interval, i.e. $\text{MAC}(\hat{\varphi}_i^{\text{ref}}, \hat{\varphi}_i) < t_{\text{MAC}}$, then label $\hat{\varphi}_i$ as changed.

The results of the monitoring using vision data sets of all beam conditions (mass-free reference condition (Ref), and with the added mass at the positions indicated in Fig. 4) are shown for each of the modes in Fig. 11. For each mode, it is evaluated if the frequency estimate lies within the confidence bound of the reference, as outlined above, and a change in the corresponding mode shape as indicated by the confidence bound of the MAC is marked by an empty circle, while no significant mode shape change is marked by a full circle.

First of all, it can be observed that the structural changes induce changes in some of the frequencies and in some of the mode shapes, but not in all of them. For example, there is no significant change in the first mode shape of mode 1, however the frequency changes are significant for some of the mass positions. On the other side, the frequencies of mode 3 show no significant change for the added mass at P1803, P949 and P260 whereas the corresponding mode shapes do change. This also shows the importance of evaluating changes both in the frequencies *and* in the mode shapes, since some damages do not induce change in the frequency but in the mode shape, or vice versa.

From Fig. 11 it can also be observed that for most of the modes the mode shapes change for most of the damage scenarios. It should be noted that only the statistical evaluation of the MAC revealed that these changes are actually

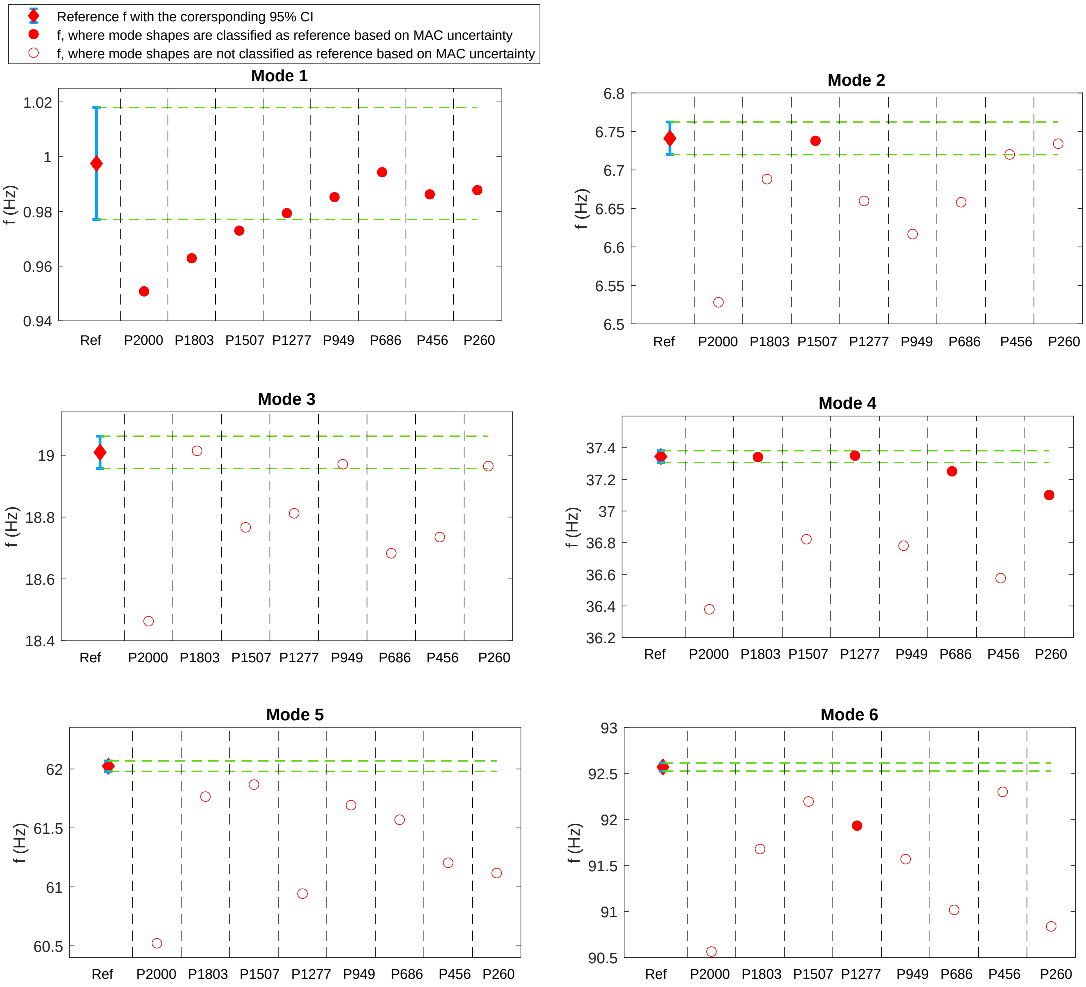


Figure 11: Changes of the natural frequencies related to the first six modes. The green dashed horizontal lines represent the 95% confidence interval of the reference natural frequency

significant, since the MAC values are relatively high, being above 0.97. For the cases where no mode shape change was indicated, a numerical study actually confirmed that the change in the numerical mode shapes is insignificant with mostly MAC > 0.999. To illustrate the actual mode shape changes, Fig. 12 shows the mode shape estimate of mode 4 together with its confidence bound in the reference condition, as well as the mode shape estimates after adding the mass at $P2000$ and at $P1803$. It can be seen that the estimate corresponding to $P1803$ remains in the confidence bound of the reference, and indeed no mode shape change is indicated in Fig. 11. In contrast, the mass attached in $P2000$ has significantly modified the mode shape estimate that is leaving the confidence bound of the reference, and it is correctly classified as changed.

In summary, the structural changes through the added mass could be detected for all of the scenarios in at least some of the modes with the proposed vision technology, either by frequency or mode shape changes, or both. Finally, it should be noted that the change of the frequency and/or MAC of a single mode beyond their reference confidence interval is sufficient to make a decision about the beam health.

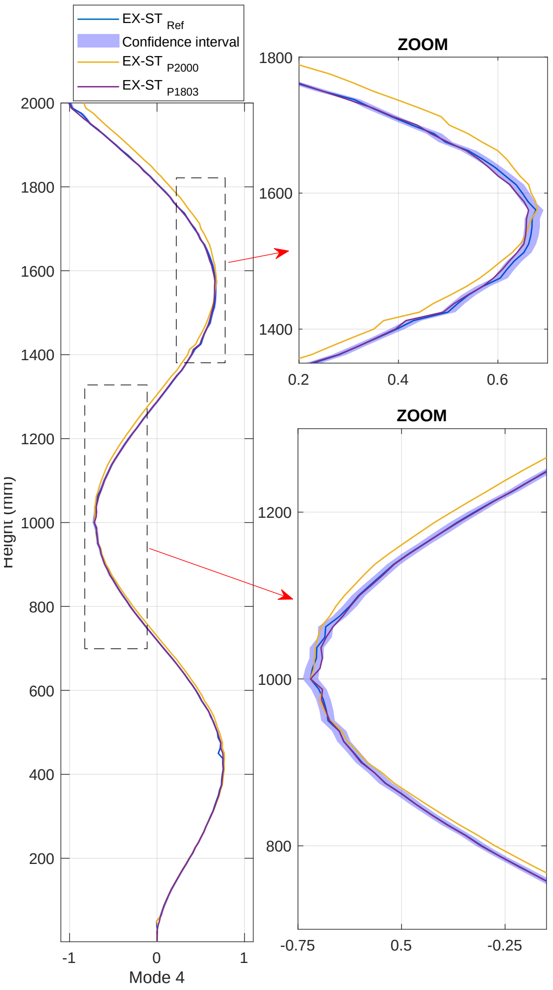


Figure 12: Changes in the fourth mode shape due to added mass on two different positions.

7. Conclusion

In this paper a video-based approach for operational modal analysis was presented and evaluated on experimental data of a cantilever beam. The approach combines an improved motion extraction method with state-of-the-art subspace identification and uncertainty quantification. For motion extraction, the proposed extension of the local phase method that relies on steerable filters (EX-ST) is able to consider both small and large motions in the source videos. Furthermore, two widespread reference motion extraction algorithms namely Taylor and QSF were used for comparison. The observations of this study can be summarized as follows:

- The displacements extracted with the proposed EX-ST method achieve a better accuracy than Taylor approximation and QSF, at the expense of a moderately increased computation time.
- The application of the covariance-driven subspace method with uncertainty quantification allows to evaluate the performance of the vision-based modal parameters estimates. It was concluded that estimates obtained from the displacements extracted from 160 ROIs with motion extraction algorithms showed no statistically significant differences compared to those identified from the reference LDS, and that confidence bounds were similar.
- The identified mode shapes from the vision data have a very high spatial resolution and closely match the numerical ones obtained from the FE model. The mode shape estimates obtained from EX-ST motions were

found smoother and closest to the numerical ones, which was shown by the MAC estimates, while the estimates based on the Taylor approximation and QSF methods were less accurate.

- The proposed framework was able to detect small structural changes by an added mass at different positions of the beam. Statistically significant changes in the frequencies as well as in the high resolution mode shapes were found, considering the estimation uncertainty related to the vision data, showing the potential of the framework for monitoring.

The developments in this paper show that a thorough evaluation of the motion extraction method as well as the consideration of statistical uncertainties is important to ensure good accuracy and precision in video-based operational modal analysis. Such an evaluation has been made in a controlled laboratory environment as a necessary validation step before actual application to real-world civil structures in operation.

After this successful evaluation with laboratory experiments, the presented approach will be tested on a bridge and other outdoor structures in the next step, so that its performance will be evaluated under real world operation conditions. Our future researches will address additional challenges in outdoor environments, including ambient light variations over time and atmospheric distortion, to ensure robust motion extraction for vision-based operational modal analysis.

Acknowledgements

Part of this work was funded by the Research Fund for Coal and Steel under grant agreement No 800687 in the framework of the H2020 DESDEMONA project. The authors are grateful to Ivan Guéguen, Jean Philippe Gourdon and Fabrice Blaineau for the assistance provided for the experimental set up.

References

[1] C. Farrar, K. Worden, An introduction to structural health monitoring, Philos T Roy Soc A 365 (1851) (2007) 303–315.

[2] F. Magalhães, A. Cunha, E. Caetano, Vibration based structural health monitoring of an arch bridge: From automated OMA to damage detection, Mech Syst Signal Pr 28 (2012) 212–228.

[3] J. Brownjohn, Structural health monitoring of civil infrastructure, Philos T Roy Soc A 365 (1851) (2007) 589–622.

[4] O. Avci, O. Abdeljaber, S. Kiranyaz, M. Hussein, M. Gabbouj, D. J. Inman, A review of vibration-based damage detection in civil structures: From traditional methods to machine learning and deep learning applications, Mech Syst Signal Pr 147 (2021) 107077.

[5] M. D. H. Bhuyan, G. Gautier, N. Le Touz, M. Döhler, F. Hille, J. Dumoulin, L. Mevel, Vibration-based damage localization with load vectors under temperature changes, Struct Control Hlth 26 (11) (2019) e2439.

[6] M. Döhler, F. Hille, L. Mevel, W. Rücker, Structural health monitoring with statistical methods during progressive damage test of S101 Bridge, Eng Struct 69 (2014) 183–193.

[7] D. Feng, M. Q. Feng, Computer vision for shm of civil infrastructure: From dynamic response measurement to damage detection–a review, Eng Struct 156 (2018) 105–117.

[8] H. Wang, T. Tao, T. Guo, J. Li, A. Li, Full-scale measurements and system identification on sutong cable-stayed bridge during typhoon fung-wong, Sci World J 2014 (2014).

[9] K.-T. Park, S.-H. Kim, H.-S. Park, K.-W. Lee, The determination of bridge displacement using measured acceleration, Eng Struct 27 (3) (2005) 371–378.

[10] J. Li, H. Hao, Health monitoring of joint conditions in steel truss bridges with relative displacement sensors, Measurement 88 (2016) 360–371.

[11] D. Zhang, J. Guo, X. Lei, C. Zhu, A high-speed vision-based sensor for dynamic vibration analysis using fast motion extraction algorithms, Sensors 16 (4) (2016) 572.

[12] H. H. Nassif, M. Gindy, J. Davis, Comparison of laser doppler vibrometer with contact sensors for monitoring bridge deflection and vibration, NDT & E INT 38 (3) (2005) 213–218.

[13] G. Tzortzinis, C. Ai, S. F. Breña, S. Gerasimidis, Using 3d laser scanning for estimating the capacity of corroded steel bridge girders: Experiments, computations and analytical solutions, Eng Struct 265 (2022) 114407.

[14] A. Bautista-De Castro, L. J. Sánchez-Aparicio, L. F. Ramos, J. Sena-Cruz, D. González-Aguilera, Integrating geomatic approaches, operational modal analysis, advanced numerical and updating methods to evaluate the current safety conditions of the historical bôco bridge, Constr Build Mater 158 (2018) 961–984.

[15] X. Meng, A. Dodson, G. Roberts, Detecting bridge dynamics with gps and triaxial accelerometers, Eng Struct 29 (11) (2007) 3178–3184.

[16] Y. Xu, W. Chan, Wind and structural monitoring of long span cable-supported bridges with gps, in: Proceedings of the 7th Asia-Pacific Conference on Wind Engineering (APCWE'09), 2009.

[17] C. Negulescu, G. Luzi, M. Crosetto, D. Raucoules, A. Roullé, D. Monfort, L. Pujades, B. Colas, T. Dewez, Comparison of seismometer and radar measurements for the modal identification of civil engineering structures, Eng Struct 51 (2013) 10–22.

[18] F. Uehan, Development of the u-doppler non-contact vibration measuring system for diagnosis of railway structures, Quarterly Report of RTRI 49 (3) (2008) 178–183.

[19] S. Ri, Q. Wang, H. Tsuda, H. Shirasaki, K. Kuribayashi, Displacement measurement of concrete bridges by the sampling moiré method based on phase analysis of repeated pattern, Strain (2020) e12351.

[20] U. Gabbert, Vibration analysis of an electric wheel hub motor at stationary operating points, in: Proceedings of the 10th International Conference on Rotor Dynamics–IFToMM: Vol. 4, Vol. 63, Springer, 2018, p. 51.

[21] D. Ribeiro, R. Calçada, J. Ferreira, T. Martins, Non-contact measurement of the dynamic displacement of railway bridges using an advanced video-based system, Eng Struct 75 (2014) 164–180.

[22] T. Kumberg, S. Schneid, L. Reindl, A wireless sensor network using gnss receivers for a short-term assessment of the modal properties of the neckartal bridge, Appl. Sci. 7 (6) (2017) 626.

[23] A. Zona, Vision-based vibration monitoring of structures and infrastructures: An overview of recent applications, Infrastructures 6 (1) (2021) 4.

[24] L. Luo, M. Q. Feng, Z. Y. Wu, Robust vision sensor for multi-point displacement monitoring of bridges in the field, Eng Struct 163 (2018) 255–266.

[25] L. Tian, B. Pan, Remote bridge deflection measurement using an advanced video deflectometer and actively illuminated led targets, Sensors 16 (9) (2016) 1344.

[26] C. Dorn, S. Dasari, Y. Yang, C. Farrar, G. Kenyon, P. Welch, D. Mascareñas, Efficient full-field vibration measurements and operational modal analysis using neuromorphic event-based imaging, J Eng Mech 144 (7) (2018) 04018054.

[27] J. G. Chen, N. Wadhwa, Y.-J. Cha, F. Durand, W. T. Freeman, O. Buyukozturk, Modal identification of simple structures with high-speed video using motion magnification, J Sound Vib 345 (2015) 58–71.

[28] M. Q. Feng, Y. Fukuda, D. Feng, M. Mizuta, Nontarget vision sensor for remote measurement of bridge dynamic response, J Bridge Eng 20 (12) (2015) 04015023.

[29] B. D. Lucas, T. Kanade, et al., An iterative image registration technique with an application to stereo vision, Vol. 81, Vancouver, 1981.

[30] B. K. Horn, B. G. Schunck, Determining optical flow, Artif. Intell. 17 (1-3) (1981) 185–203.

[31] S. H. Chan, D. T. Võ, T. Q. Nguyen, Subpixel motion estimation without interpolation, in: IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2010, pp. 722–725.

[32] J. Javh, J. Slavič, M. Boltežar, The subpixel resolution of optical-flow-based modal analysis, Mech Syst Signal Pr 88 (2017) 89–99.

[33] Z.-R. Lu, G. Lin, L. Wang, Output-only modal parameter identification of structures by vision modal analysis, J Sound Vib 497 (2021) 115949.

[34] B. Pan, B. Xu, D. Chen, J. Feng, Sub-pixel registration using quadratic surface fitting in digital image correlation, AMS 26 (2) (2005) 128–134.

[35] S.-W. Kim, B.-G. Jeon, N.-S. Kim, J.-C. Park, Vision-based monitoring system for evaluating cable tensile forces on a cable-stayed bridge, Struct Health Monit 12 (5-6) (2013) 440–456.

[36] S. T. Thurman, J. R. Fienup, M. Guizar-Sicairos, Efficient subpixel image registration algorithms, Ol/33/2/ol Pdf 33 (2) (2008) 156–0.

[37] D. Feng, M. Q. Feng, E. Ozer, Y. Fukuda, A vision-based sensor for non-contact structural displacement measurement, Sensors 15 (7) (2015) 16557–16575.

[38] E. Psarakis, G. Evangelidis, An enhanced correlation-based method for stereo correspondence with sub-pixel accuracy, in: 10th IEEE International Conference on Computer Vision (ICCV), Beijing, China, 2005, pp. 907–912.

[39] W. T. Freeman, E. H. Adelson, et al., The design and use of steerable filters, IEEE T PATTERN ANAL 13 (9) (1991) 891–906.

[40] N. Wadhwa, M. Rubinstein, F. Durand, W. T. Freeman, Phase-based video motion processing, Acm T Graphic 32 (4) (2013) 1–10.

[41] Q. Zhu, D. Cui, Q. Zhang, Y. Du, A robust structural vibration recognition system based on computer vision, J Sound Vib (2022) 117321.

[42] M. Civera, L. Zanotti Fragonara, C. Surace, An experimental study of the feasibility of phase-based video magnification for damage detection and localisation in operational deflection shapes, Strain 56 (1) (2020) e12336.

[43] J. G. Chen, Video camera-based vibration measurement of infrastructure, Ph.D. thesis, Massachusetts Institute of Technology (2016).

[44] B. Peeters, G. De Roeck, Reference-based stochastic subspace identification for output-only modal analysis, Mech Syst Signal Pr 13 (6) (1999) 855–878.

[45] M. Döhler, L. Mevel, Fast multi-order computation of system matrices in subspace-based system identification, Control Eng Pract 20 (9) (2012) 882–894.

[46] E. Reynders, System identification methods for (operational) modal analysis: review and comparison, Mech Syst Signal Pr 19 (1) (2012) 51–124.

[47] E. Reynders, R. Pintelon, G. De Roeck, Uncertainty bounds on modal parameters obtained from stochastic subspace identification, Mech Syst Signal Pr 22 (4) (2008) 948–969.

[48] M. Döhler, L. Mevel, Efficient multi-order uncertainty computation for stochastic subspace identification, Mech Syst Signal Pr 38 (2) (2013) 346–366.

[49] P. Mellinger, M. Döhler, L. Mevel, Variance estimation of modal parameters from output-only and input/output subspace-based system identification, J Sound Vib 379 (2016) 1–27.

[50] E. P. Reynders, Uncertainty quantification in data-driven stochastic subspace identification, Mech Syst Signal Pr 151 (2021) 107338.

[51] S. Greś, R. Riva, C. Y. Süleyman, P. Andersen, M. Łuczak, Uncertainty quantification of modal parameter estimates obtained from subspace identification: An experimental validation on a laboratory test of a large-scale wind turbine blade, Eng Struct 256 (2022) 114001.

[52] S. Greś, M. Döhler, L. Mevel, Uncertainty quantification of the modal assurance criterion in operational modal analysis, Mech Syst Signal Pr 152 (2021) 107457.

[53] L. Quam, Hierarchical warp stereo. in image understanding workshop, SAIC 1 (1984) 149–155.

[54] M. A. Sutton, J. J. Orteu, H. Schreier, Image correlation for shape, motion and deformation measurements: basic concepts, theory and applications, Springer Science & Business Media, 2009.

[55] S. Collier, T. Dare, Accuracy of phase-based optical flow for vibration extraction, J Sound Vib 535 (2022) 117112.

[56] J. G. Chen, A. Davis, N. Wadhwa, F. Durand, W. T. Freeman, O. Büyüköztürk, Video camera-based vibration measurement for civil infrastructure applications, J Infrastruct Syst 23 (3) (2017) B4016013.

[57] D. J. Fleet, A. D. Jepson, Computation of component image velocity from local phase information, Int J Comput Vision 5 (1) (1990) 77–104.

[58] T. Gautama, M. Van Hulle, A phase-based approach to the estimation of the optical flow field using spatial filtering, IEEE T Neural Networ 13 (5) (2002) 1127–1136.

[59] M. Xiao, S. Zheng, C. Liu, Y. Wang, D. He, G. Ke, J. Bian, Z. Lin, T.-Y. Liu, Invertible image rescaling (2020).

[60] J. S. Przemieniecki, Theory of matrix structural analysis, Courier Corporation, 1985.

[61] W. Gersch, On the achievable accuracy of structural system parameter estimates, Journal of Sound and Vibration 34 (1) (1974) 63–79.

[62] D. Bernal, M. Döhler, S. Mozaffari Kojidi, K. Kwan, Y. Liu, First mode damping ratios for buildings, Earthquake Spectra 31 (1) (2015) 367–381.