Video Analysis of Nonlinear Systems with Extended Kalman Filtering for Modal Identification

Stefania Lo Feudo, Jean-Luc Dion, Franck Renaud, Gaëtan Kerschen, Jean-Philippe Noël

Video Analysis of Nonlinear Systems with Extended Kalman Filtering for Modal Identification

Stefania Lo Feudo[1*], Jean-Luc Dion[1], Franck Renaud[1], Gaëtan Kerschen[2] and Jean-Philippe Noël[3]

[1*]QUARTZ Laboratory (EA7393), Vibrations, Acoustics, Mechanical Structures and Shapes, ISAE-Supméca, Saint-Ouen-sur-Seine, France.
[2]Space Structures and Systems Laboratory (S3L), Aerospace and Mechanical Engineering Department, University of Liège, Liège, Belgium.
[3]Mechanical Engineering Department, KU Leuven, Leuven, Belgium.

*Corresponding author(s). E-mail(s): stefania.lofeudo@isae-supmeca.fr;
Contributing authors: jean-luc.dion@isae-supmeca.fr; franck.renaud@isae-supmeca.fr; g.kerschen@uliege.be; jp.noel@kuleuven.be;

Abstract

This study proposes to carry out the experimental modal analysis of nonlinear systems under the assumption of almost invariant modal shapes by coupling video analysis from a high speed/resolution camera and extended Kalman filtering. A clamped-clamped beam with a local nonlinearity is considered, and its vibrations are measured by detecting and tracking a large set of (virtual) sensors bonded to the beam outer surface. Specific image processing and video tracking techniques are employed and detailed herein. Then, the instantaneous natural frequencies and modal amplitudes are identified by means of a data assimilation method based on extended Kalman and modal filters. Finally, the proposed method of identification is assessed using a numerical example possessing 3 degrees of freedom and a strong nonlinearity. The performance and limits of the identification process are discussed.

Keywords: Video analysis, Extended Kalman filter, Modal filter, Nonlinear vibrations

1 Introduction: current and future needs in an industrial context

Most industrial systems feature nonlinear dynamical behavior due to assemblies, contacts and geometries that depend directly on the amplitudes of their vibrations. These nonlinearities can be locally "strong", e.g., contact dynamics, but eventually lead to weakly nonlinear effects at the system scale. In that sense, natural frequencies and damping ratios exhibit a strong dependence on the amplitude of vibration, but the mode shapes resemble those of the underlying linear system. In specific industrial applications such as health monitoring, the real-time identification of modal contributions is required in order to track specific nonlinearities or particular dysfunctions. Such applications usually require numerous sensors that may be advantageously replaced by a high speed camera.

To process a video sequence and measure a displacement in the image, several algorithms exist [1]. Digital image correlation (DIC) is widespread [2–5]. By comparing two sub-images taken at two successive instants, it is possible to find the optimal transformation for passing from one to the other and thus to deduce a local movement in the overall image. Optical flow is an interesting alternative [6–8]. This technique is based on an Eulerian approach applied to the light intensity of the pixels and calculates the local speed in the image thanks to the intensity gradient. The method of phase-based motion magnification [9–12] allows to amplify the motions inside a video sequence by analyzing the phase of its Fourier transform.

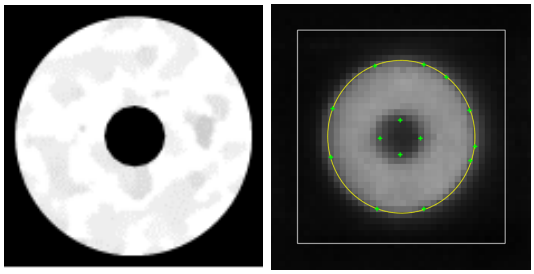
Classical surface treatments for features detection in computer vision consist in white painting, speckle patterns or glued targets. The choice of the surface treatment depends on a number of factors such as the resolution of the image, the physical quantity sought, and the type of computer vision algorithm applied to the video sequence. For example, the speckle pattern is suitable for high-resolution images on which to apply the DIC for the full-field estimation of displacement and strain fields, or for damage detection. On the other hand, the use of discrete targets is suitable for large-scale structures in order to avoid painting large surfaces. Moreover, the use of targets is advantageous in those cases where the structural motion can be reconstructed through measurement of the displacement field at some discrete points placed along the structure. In these cases, employing targets entails significant benefits. Not only in terms of ease of installation and uncertainty, but also in terms of computational costs since it reduces the number of pixels considered during the image processing phase.

In this framework, Ref. [13] measured the horizontal displacements and rotations of each floor of a tall building by placing several cameras and targets across the building width. Ref. [14] performed several experiments showing the effectiveness of camera-based metrology for the motion analysis of flexible beams and cables. In this study, image processing was performed with the aid of the commercial software EVaRT4.6, such that no specific information are provided about the algorithm employed for features detection and tracking. The mode shapes of a cantilever beam were extracted from video records by Romaszko et al. [15].

Other studies focused on the estimation of the accuracy and uncertainty of target-based measurements under uncontrolled environmental conditions, see for example Refs. [16] and [17]. As an alternative to classical glued stickers for targeting, one can use projection of laser beams as in [18–20] or emitting diodes, [21]. Ref. [22] tracked elliptical targets with sub-pixel accuracy on a video recorded by a commercial pocket digital camera. Ref. [23] shown that region-based local detectors are more performant than corner-based ones. Ref. [24] proved that the maximally stable external regions (MSER) detector, first proposed in [25], provides good repeatability and matching scores against viewpoint changes. Ref. [26] verified that MSER is better suited to indoors, simple and human-made dominated scenes. Recently, Renaud et al. [27] detected oscillations of elliptical targets bonded to a wing scale model by the MSER algorithm. The 3D reconstruction of the wing dynamics was obtained by projecting numerical 3D deflection shapes into the image frame of a single camera.

Generally, modal analysis using camera measurements is based on the principle of stereo vision which requires a minimum of two cameras, see [2, 3, 28]. Some authors tried to obtain stereo vision with a single camera using a prism or a set of mirrors [4, 5, 29]. Mode shapes were reconstructed by recording two images of a scene at two different locations and instants with a single camera in [30]. More interestingly, by moving a single camera relative to the object during shooting, it was possible to reconstruct the eigenmodes [8, 31].

This paper proposes to couple video analysis with extended Kalman filtering for real-time experimental modal analysis of a system with a local nonlinearity. Specifically, a video sequence from a nonlinear beam mimicking the European project COST F3 benchmark [32] was recorded using a camera under impulsive excitation. The present work is an extension of a previous work of some of the authors [33]. The present study details the image processing and the video analysis techniques employed and developed, not addressed in [33], and includes new analyses and discussions regarding the accuracy and the feasibility of the real-time implementation of the video analysis



(a) (b)

Fig. 1: (a) Circular target glued on the COST F3 beam. The nominal value of the internal and external diameters are $\bar{D}_{int}$ =1,5 mm and $\bar{D}_{ext}$ =10,5 mm, respectively [35]. (b) Target detected and extracted with the MSER algorithm, feature points and surrounding box [25, 36].

method. Moreover, this study focuses on two well-known test benches previously published [32, 34], in order to ensure repeatability and further results comparison. The two chosen and studied benchmarks are particularly nonlinear, whereas the benchmarks in [33] are linear or weekly nonlinear.

The paper is organized as follows. Section 2 presents the image processing and the video tracking method employed for the video-based vibration measurements. Section 3 describes the data assimilation method based on extended Kalman filtering that is used for the identification of the modal contributions of nonlinear systems. In Section 4, the experimental test case, the clamped-clamped nonlinear beam, is described. Measurements analysis and results are discussed. Because the beam only features a weakly nonlinear behavior, the limits of the proposed method of identification are highlighted through a numerical example possessing a strong nonlinearity [34] in Section 5. Finally, the conclusions are drawn in Section 6.

2 Image processing and video tracking

In this study, the motion of a clamped-clamped beam with a weak nonlinearity resulting from free vibrations was recorded by a camera at middle speed (1000 fps). The beam motion is extracted by tracking during the video sequence discrete targets bonded to the visible outer surface, see Fig. 1 and Fig. 4. Feature extraction is performed with the MSER algorithm using the Image Processing & Computer Vision toolboxes of MATLAB. The result for one target is shown in Fig. 1(b). The choice of this local detector is justified by the fact that is very performing in detect simple features on an indoor scene, and because of its stability against viewpoints changes, blurring and light reduction, [24, 26]. This feature is of particular interest in video-based vibrations measurements since targets may experience variable lighting conditions and focusing due to the structure oscillations. Moreover, the MSER algorithm extracts the detected features in the form of ellipses. This aspect is well suited for the present case study since circular targets are distorted in the image plane by the camera lens and the variation of the viewpoint angle. In the next step of the image processing, some feature points are detected within the ellipses and tracked, [36]. This is possible because of the specific texture of the bonded target (see Fig. 1(a)). The estimation of a geometric transform from matching point pairs at the frame i and $i+1$ allows updating the surrounding box position and size, [37]. Finally, the coordinates of a target are estimated by the box's center.

The algorithm presented in this study aims to detect the targets in the video sequence by looking for the MSERs and feature points not within the whole picture, but only inside virtual rectangular boxes placed around the circular targets. These boxes are drawn on the first frame, and their position is updated at each frame according to the in-plane motion of the target. The protocol is easily adaptable to different set-ups, both in research and industrial settings. This is because of the fact that the specific treatments necessary to extract the features from the filmed scene are performed only on the first image. Indeed, the noise and reflection suppression are performed only on one frame. This procedure requires low computational costs since the features are sought only in several small regions of interest (ROI) which move accordingly with the beam oscillations. In other words, a multi-target tracking is performed in the first image to initialize the virtual targets position, whereas multiple single-target tracking are performed during the video sequence, since each feature is searched in a specific ROI. It is worth noting that this method does not require neither

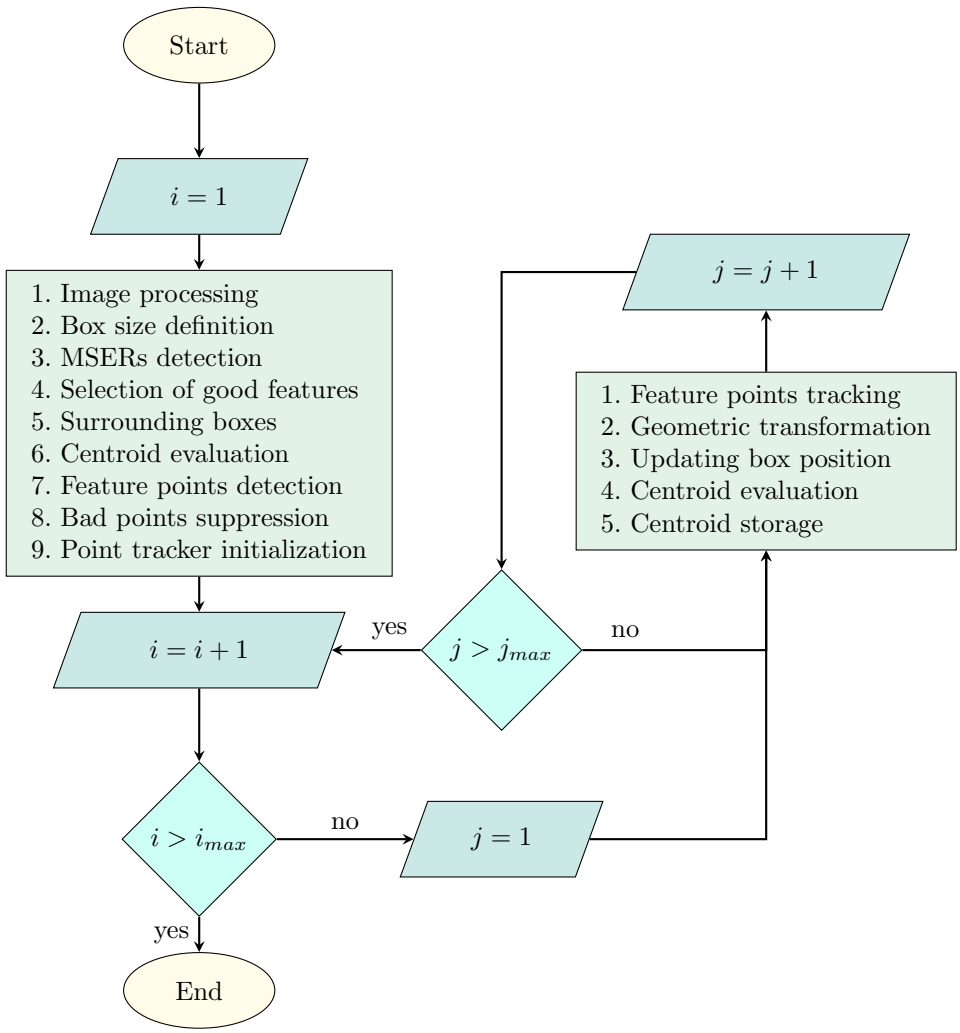


Fig. 2: Flowchart for feature detection and tracking in a video sequence. The bonded targets are extracted with a MSER feature detector from a binary version of the first frame. Each feature is described by an ellipse and a surrounding bounding box. Then, some feature points are detected inside each region. For every next frame, feature points are tracked and the position of the surrounding boxes is updated by applying a geometric transformation estimated by matching point pairs at the frame i and $i + 1$. The output consists in the coordinates of the centroid of each box for all frames. i_{max} is the last frame of the video sequence and j_{max} the maximum number of detected targets, respectively.

a learning process nor a physical model describing the structural behavior. The algorithm for the target detection and tracking is detailed in the Section 2.1.

2.1 Target detection and tracking

On the first frame, the following operations are performed:

1. **Image processing**. Contrast enhancement and adaptive histogram equalization are performed on the original gray-scale image. Then, manual segmentation is applied in order to highlights the bonded targets. Residual noise is filtered by removing small objects from the binary image.
2. **Box size definition**. The typical size of a box is determined by manually drawing on the

image a rectangle surrounding one of the visible targets.

3. **MSERs detection**. MSERs are detected by using the Matlab function `detectMSERFeatures`.
4. **Selection of good features**. Among all localized regions, only those corresponding to a target are retained. This process can be automated by applying geometric conditions defined according to the type of target used. In the present case, for each target the detection of two concentric ellipses is expected. Therefore, detected regions are considered as valid if they meet the following criteria:
   - region's area is smaller than that of the box;
   - the region's center is close to that of another region;
   - both of them have a similar ratio between their major and minor axis;
   - one region is smaller than the other.
5. **Surrounding boxes**. For each region, unrotated rectangular boxes circumscribed to the outer ellipse are calculated.
6. **Centroid evaluation**. The boxes' center represents an estimate of the position of the target in the image coordinates.
7. **Feature points detection**. They are detected inside each box by using the Matlab function `detectMinEigenFeatures`.
8. **Bad points suppression**. Points outside of the ellipses are excluded.
9. **Point tracker initialization**. Initialization of the feature-tracking algorithm (`vision.PointTracker`), [36]. In this case study, the *MaxBidirectionalError* parameter is set equal to 2.

For the j–th box, the following operations are carried out at the i–th frame:

1. **Feature points tracking**. The feature points tracking is performed on a cropped version of the image including only the structure under study.
2. **Geometric transformation**. Estimation of a similarity transformation which maps the greatest number of point pairs the frames i and $i-1$ (`estimateGeometricTransform`), [37]. In this case, the *MaxDistance* parameter is set equal to 4.
3. **Updating box position**. The position of the box is updated by applying the transformation to the original bounding box (`transformPointsForward`).
4. **Centroid evaluation**. The coordinates of the target is estimated as center of the box.
5. **Store centroid**. This represents the position of the target at the current frame.

These operations are summarized in the Figure 2.

Finally, the trajectories of the beam are interpolated according to the discretization of a finite element (FE) model. These data are processed with the aid of the extended Kalman filter (EKF) for the identification of the modal frequencies and amplitudes. The formulation of the data assimilation method is detailed in the next section.

3 Real-time modal filter: a solution based on extended Kalman filtering

The displacement $u(t)$ of a specific target can be decomposed on a real modal basis:

$$u(t) = \Re\left(\sum_{k=1}^{M} \alpha_k(t)\phi_k\right) \tag{1}$$

where $\alpha_k(t)$ is the instantaneous amplitude and ϕ_k the mode shape. In the case of oscillatory motion it can be rewritten as:

$$u(t) = \Re\left(\sum_{k=1}^{M} A_k e^{2i\pi f_k t}\phi_k\right) \tag{2}$$

where f_k is the modal frequency and A_k the modal amplitude. The displacement of the target between two time instants is defined as:

$$u(t+dt) = \Re\left(\sum_{k=1}^{M} \alpha_k(t+dt)\phi_k\right) \tag{3}$$

The assumption that f_k and A_k are almost equal between two consecutive instants leads to:

$$u(t+dt) = \Re\left(\sum_{k=1}^{M} A_k e^{2i\pi f_k (t+dt)}\phi_k\right) \tag{4}$$

By replacing Eq. 2 into Eq. 4, the equation becomes:

$$u(t+dt) = \Re\left(\sum_{k=1}^{M} \alpha_k(t) e^{2i\pi f_k dt} \phi_k\right) \tag{5}$$

Thus, by expressing the instantaneous amplitude in a complex form, $\alpha_k(t) = a_{r,k}(t) + i a_{i,k}(t)$, the real and the imaginary part of $\alpha_k(t+dt)$ can be obtained:

$$\begin{aligned} a_{r,k}(t+dt) &= a_{r,k}(t)\cos(2\pi f_k dt) \\ &\quad - a_{i,k}(t)\sin(2\pi f_k dt), \\ a_{i,k}(t+dt) &= a_{r,k}(t)\sin(2\pi f_k dt) \\ &\quad + a_{i,k}(t)\cos(2\pi f_k dt) \end{aligned} \tag{6}$$

To identify the modal contributions at time $t_n = n.dt$, we consider that each natural mode $k \in [1 \dots M]$ has 3 components in the state vector: its real and observable part $x_{1+3(k-1),n} = a_{r,k}(t_n)$, its imaginary part $x_{2+3(k-1),n} = a_{i,k}(t_n)$, which can be seen as the Hilbert transform of $x_{1+3(k-1),n}$, and its phase variation over one time step $x_{3+3(k-1),n} = f_k dt$, which leads to its instantaneous natural frequency (normalized). The state vector is a vector containing all the variables of the state system and is built as:

$$\begin{aligned} X_n^T = (&x_{1,n} \quad x_{2,n} \quad x_{3,n} \quad \dots x_{1+3(M-1),n} \\ &x_{2+3(M-1),n} \quad x_{3+3(M-1),n}) \end{aligned} \tag{7}$$

In the context of nonlinear systems, state and observation equations are respectively defined as:

$$\begin{cases} X_{n+1} = F_n(X_n) + W_n \\ Z_n = H\ X_n + V_n \end{cases} \tag{8}$$

with W_n the state noise and V_n the observation noise. In other words, the state equation describes the dynamics of the system and the observation equation links the state vector (composed with state variables) to the observation vector (sensors). The modal filter consists in projecting measurements (observation vector, sensors) on natural modes (states). The specific set of equations used for the extended Kalman filter is detailed in our previous work [38]. The filter was there designed to track harmonic components with amplitude and frequency modulations.

Denoting $\hat{x}$, the estimation of u, for M natural modes the nonlinear state matrix is defined as:

$$\begin{aligned} F_n(X_n) &= \begin{bmatrix} \ddots & 0 & 0 \\ 0 & F_{n,k} & 0 \\ 0 & 0 & \ddots \end{bmatrix} X_n \\ \text{with} \quad F_{n,k} &= \begin{bmatrix} c_{n,k} & -s_{n,k} & 0 \\ s_{n,k} & c_{n,k} & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ \text{and} \quad &\begin{cases} c_{n,k} = \cos(\hat{x}_{3+3(k-1),n}) \\ s_{n,k} = \sin(\hat{x}_{3+3(k-1),n}) \end{cases} \end{aligned} \tag{9}$$

As proposed in A. Goeller's PhD thesis [35], the observation matrix H is composed of the modal bases $\Phi = [\phi_1, \phi_2, ..., \phi_M]$:

$$H = [\phi_1, 0, 0, \phi_2, 0, 0, ..., \phi_M, 0, 0] \tag{10}$$

In other terms, only real signals are observed by the sensors, and each sensor is able to see the local contribution of each state variable. The observation matrix H is invariant in time. This choice is directly linked to the assumption of weak nonlinearity, i.e., the damping ratios and natural frequencies depend on the displacement amplitude but the mode shapes are assumed to be invariant vectors. The limits of this assumption will be discussed in Section 5.

The transition (state) matrix F_n depends on the instantaneous frequencies and its evolution can be estimated at first order by its Jocabian matrix:

$$\tilde{F}_{n|n} = \left(\nabla_X \left(F(X)\, X\right)^T\right)^T \Big|_{X=\hat{X}_{n|n}} \tag{11}$$

This Jacobian matrix has the following structure:

$$\tilde{F}_n = \begin{bmatrix} \ddots & 0 & 0 \\ 0 & \tilde{F}_{n,k} & 0 \\ 0 & 0 & \ddots \end{bmatrix} \tag{12}$$

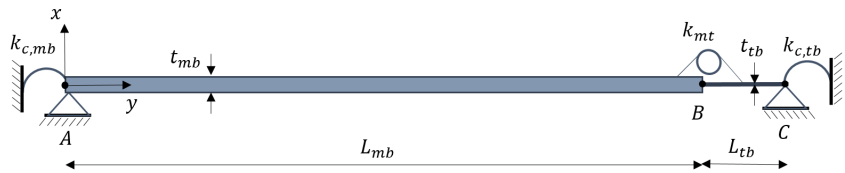


Fig. 3: Schematic of the nonlinear COST F3 beam.

with the elementary bloc defined as follows:

$$\tilde{F}_{n,k} = \begin{bmatrix} c_{n,k} & -s_{n,k} & -a_{r,n,k}s_{n,k} - a_{i,n,k}c_{n,k} \\ s_{n,k} & c_{n,k} & a_{r,n,k}c_{n,k} - a_{i,n,k}s_{n,k} \\ 0 & 0 & 1 \end{bmatrix}$$
$$\text{with} \quad \begin{cases} a_{r,n,k} = \hat{x}_{1+3(k-1),n} \\ a_{i,n,k} = \hat{x}_{2+3(k-1),n} \\ c_{n,k} = \cos(\hat{x}_{3+3(k-1),n}) \\ s_{n,k} = \sin(\hat{x}_{3+3(k-1),n}) \end{cases} \tag{13}$$

The state F_n and observation H matrices are built in order to ensure the robustness of the numerical process of the EKF. In fact, there are several methods to describe the same system; we selected a state system based on real/imaginary parts and instantaneous frequencies. Another way could be based on magnitudes and phases. Additional state variables, e.g. the instantaneous damping ratio, could also be introduced. This choice is a result of our numerical investigations and represents the most robust and accurate solution which we have developed for the state prediction with EKF.

The state matrix F_n tracks the time responses based on frequency decomposition, and the observation matrix H tracks the kinematics and space description with a modal decomposition. This decomposition allows to extract small levels of vibrations with low signal-to-noise ratio. This method, termed Kinematic-SAMI [33], was successfully applied to a linear cantilever beam. In Section 4, it is employed on a beam with a local nonlinear stiffness under free vibrations. Each mode, extracted from the linear FE model is introduced in the observation matrix, Eq. (10).

The advantage of coupling this identification technique with video analysis is that the state vector introduced in Eq. 7 can be filled with experimental data resulting from the virtual sensors motion. In contrast to traditional sensors such as accelerometers, the targets do not affect the dynamics of the considered structure and can be employed in large numbers. The details of the video measurements and the results obtained for the identification of a clamped-clamped beam with a weak nonlinearity are presented in Section 4. The method is also applied to a strongly nonlinear numerical example in Section 5.

4 Experimental test case on a weakly nonlinear system

The COST F3 beam schematized in Figure 3 is used in this paper as a benchmark for the proposed data assimilation method based on the extended Kalman filter. Due to the presence of a thin beam at the end of the cantilever beam, this benchmark is known to exhibit geometrically nonlinear effects, particularly around the first mode.

The geometrical and mechanical properties of the beam are presented in Table 1. The main AB and thin BC beams have different lengths L_{mb}, L_{tb}, thicknesses t_{mb}, t_{tb}, equal width h_b and material properties E_b, ρ_b. Non-ideal clamping is taken into account by two torsional springs. The rotational stiffness at points A and C is equal to $k_{c,mb} = 114700$ Nm/rad and $k_{c,tb} = 40$ Nm/rad, respectively. An additional torsional spring with $k_{mt} = 42,2$ Nm/rad is located between the main and thin beams at point B.

A FE model of the underlying linear beam was formulated according to the Euler-Bernoulli assumption. The main and thin beams were discretized into n_{mb} and n_{tb} 2-node Hermite finite elements, respectively. Two degrees of freedom at each node were considered, namely the transverse displacement w_x in the x direction and the rotation θ_z around the z axis (see Figure 3). The eigenvectors estimated by this model fill the observation matrix, see Eq. (10), needed by the

COST F3 beam	Length $[m]$	Width $[mm]$	Thickness $[mm]$	Young's Modulus $[GPa]$	Density $[kg/m^3]$
Main	0.7	14	14	205	7800
Thin	0.04	14	0.05	205	7800

Table 1: Geometrical and mechanical properties of the nonlinear COST F3 beam [39, 40].

EKF to identify the evolution of the beam modal frequencies and amplitudes.

4.1 Video measurements

Video measurements were performed with a digital camera Vieworks VC-12MX-M/C 180. The hardware and the software HIRIS developed by the R&D Vision company were used for image acquisition. Frames were transmitted in real-time to a PC via a CoaXPress link (CXP-6). Further details about the equipment and settings are given in Table 2 by following the recommendation provided from good-practice guidelines [41].

The beam was clamped on both sides to two rigid blocks fixed on an optical table, as depicted in Figure 4(a). As shown in Figure 5, the camera looked at the beam from above, and the tripod was placed on the table. Figure 4(b) presents a picture taken by the camera, with a 4096 x 500 pixels matrix. In order to reduce the optical distortion, the ROI containing the whole beam was centered with respect to the image sensor. Noteworthy is that the thickness in pixel of the main beam is very small (~ 90) compared to the maximum resolution of the camera in the same direction (3072), and the outer surfaces of the beam and the image sensor are almost parallel.

The tracked features consist of a series of circular targets bonded to the beam surface, as shown in Figure 1(a). $\bar{D}_{int}$ =1,5 mm and $\bar{D}_{ext}$ =10,5 mm are the nominal value of the internal and external diameters. The distance between the camera and the filmed mechanical setup leads to a resolution $\mathcal{R}$ close to 5.52 px/mm. This value is evaluated as the arithmetic mean between the resolution calculated from the targets diameter, $\mathcal{R}_{\mathcal{D}}$, and from the main beam thickness, $\mathcal{R}_{\mathcal{L}}$. $\mathcal{R}_{\mathcal{D}}$ is evaluated as the mean of the ratio between the mean of inner diameter in pixels in two perpendicular directions and the nominal value, taken over three targets placed near the main beam edge. $\mathcal{R}_{\mathcal{L}}$ is the mean of the ratio between the beam length and thickness in pixels and the nominal value (Table 1):

$$\begin{aligned} \mathcal{R}_{\mathcal{D}} &= \frac{1}{3}\sum_{i}\left[\frac{1}{2}\left(\frac{D_{i,e_x}+D_{i,e_y}}{\bar{D}_{int}}\right)\right], \\ \mathcal{R}_{\mathcal{L}} &= \frac{1}{2}\sum_{j}\left(\frac{L_j}{L_{mb}}\right)+\frac{1}{2}\sum_{k}\left(\frac{t_k}{t_{mb}}\right) \end{aligned} \tag{14}$$

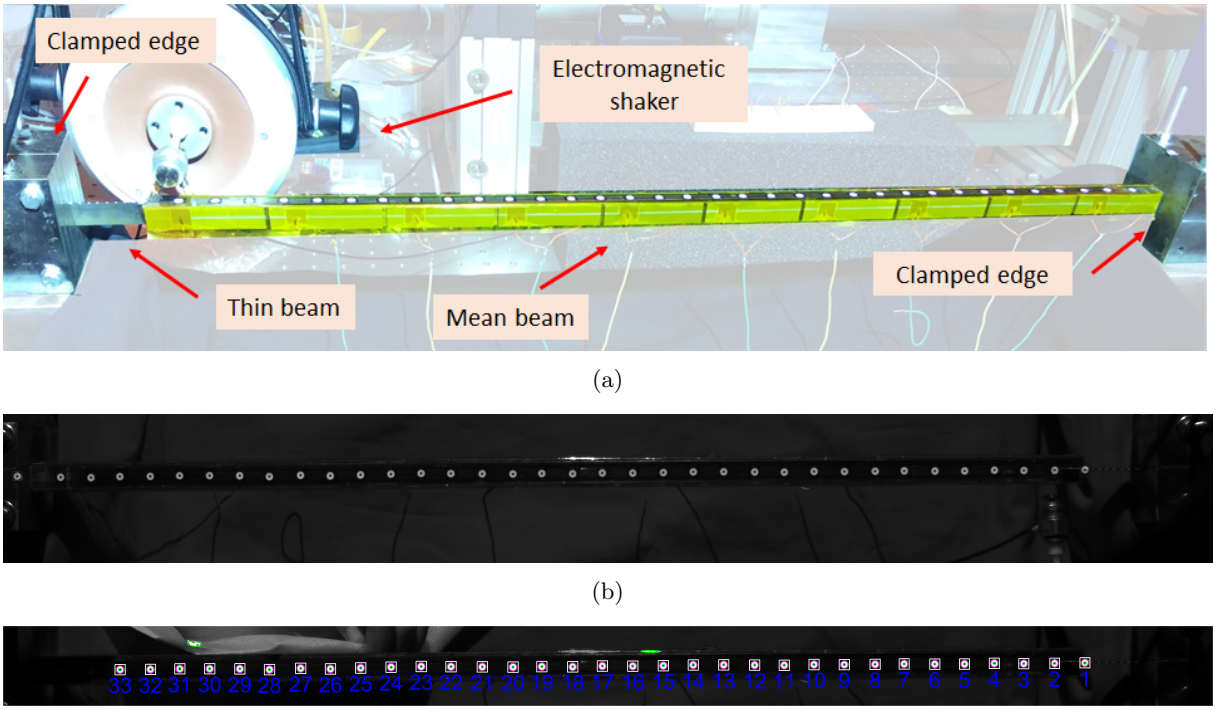
where $i = 1, 2, 3$, and (e_x, e_y) are the axes of the image coordinate systems, the origin of the latter being at the bottom-left corner of the image. $j = u, b$ stands for upper and bottom edges, and $k = l, r$ stands for left and right edges. It is worth noting that several errors may result from the video analysis due for example to poor image resolution and high lens distortion. Also, the feature detection can be less accurate on blurred targets and due to improper lighting conditions. The Kalman filtering technique allows to overcome this issue since all inaccuracies and uncertainties resulting from video analysis are taken into account in Eq. 8 in the observation noise term, V_n. Please note that the piezoelectric patches glued to the beam (in yellow in Figure 4(a)) are not of interest in this experimental campaign. Similarly, the electromagnetic shaker was disconnected from the beam during the shock test.

The beam vibrations were recorded after a shock given in the x direction on the main beam at about 15 cm from point A. The movie frame rate is 1000 fps. A 8 s duration test needs 15,2 Gb of memory to store all images in BMP format.

4.1.1 Video analysis

As depicted in the Figure 4(c), of the 36 bonded targets, 33 were detected and tracked on the video sequence. This is due to poor light conditions on the left side of the image. The figure highlights also the presence of some regions (in green) detected by the MSER algorithm in the first image that do not corresponds to a target. Thanks to the algorithm developed and detailed in Figure 2, these wrong regions are not tracked in the following images.

e



(a)

(b)

(c)

Fig. 4: (a) Experimental setup of the COST F3 beam [32], (b) picture (4096 x 500 pixels) of the beam taken by the Vieworks VC-12MX-M/C 180 digital camera and (c) detected targets, labels and surrounding boxes.

Parameter	**Value**
Camera	Vieworks VC 12MX-M 180
Lens	Samyang 35 mm F1.4 AS UMC
Lighting	DESISTI Rembrandt Piccolo 1200W MK2
Image resolution	4096 x 500 pixels
Field-of-View	742.0 x 90.6 mm
Image scale	5.52 px/mm
Patterning technique	Circular targets
Exposure time	700 μs
Image acquisition rate	1000 fps
Image acquisition platform	HIRIS R&D Vision

Table 2: Image acquisition equipment, settings and software information.

In this study, the target tracking was not performed on real-time, and so neither the identification procedure based on the EKF. Indeed, for the video analysis the computation time is equal to about 0.3 s per image by using a PC with Intel(R) Xeon(R) CPU E5-1607 v4 @ 3.10GHz 3.10 GHz and 32.0 Go of RAM. Therefore, for this study a real-time implementation of the identification procedure would not be possible, being the movie recorded at 1000 fps. The duration of the image processing is related to the processor, to the code performance, and mostly on the feature points tracking algorithm employed. In addition, the image resolution and the number

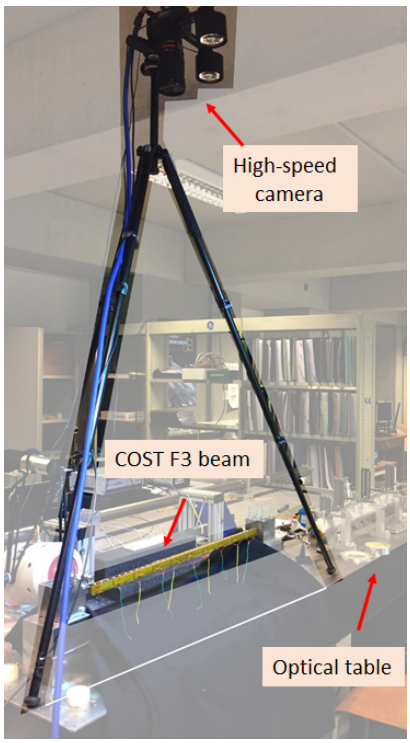


Fig. 5: Digital camera position.

of bonded targets have a large impact. Further studies can be conducted to speed up the video tracking algorithm, being beyond the scope of this paper.

4.2 Vibration measurement analysis

This section describes the identification process (K-SAMI [33], Section 3) performed with a modal filter led by EKF on the experimental data obtained by the video analysis. The first 3 modes were studied. Despite the heavy optical table, the camera linked to the table through a tripod moved slightly when the shock occurred. In order to remove the camera motion in the analysis, a rigid body mode was added to the modal basis.

Figure 6 depicts the time evolution of the natural frequencies of the beam. The (linear) mode shapes associated with each natural frequency are inset. As studied in previous works [32, 39, 40], the first bending mode is the most nonlinear one, with an evolution of the natural frequency from 26 Hz for small vibration amplitudes to 32 Hz for large vibration amplitudes. The second bending mode

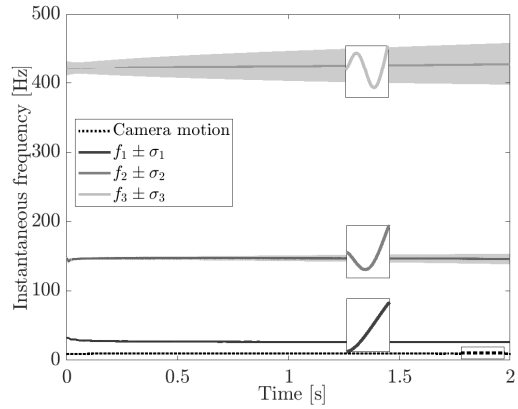


Fig. 6: Evolution of the natural frequencies over time for the first 3 bending modes of the beam (and the rigid body motion of the camera).

behaves in the opposite way with a natural frequency close to 147 Hz for small amplitudes and 142 Hz for large amplitudes. The softening effect of the second mode is attributed to the boundary conditions. The third mode behaves linearly. The rigid body motion is associated with a natural frequency lower than 10 Hz.

The corresponding modal amplitudes are represented in Figure 7. The amplitude of the rigid body mode is close to the amplitude of the first bending mode (∼1 mm) and much greater than the second (∼0,1 mm) and third (0,1 mm) bending modes. The proposed method is particularly suitable for separating the contributions of each mode, and the rigid body mode does not affect the accuracy in the tracking of the flexible modes. One can also observe that the modal amplitudes are modulated. These modulations are not "physical" but only due to their dynamics that interacts in a frequency range which is too close from the natural frequency of the mode. The assumption of narrow bandwidth signal theory used in the Hilbert Transform is not exactly valid in this context. However, the phenomenon and its effects are well understood, and we decided to keep a state system based on the Hilbert Transform assumption instead of modifying the analytical state system or using an EMD-like method [42].

Uncertainties measurements are quantified for each natural mode in Figure 8, which highlights that the kinematics of each natural mode is known with an accuracy (standard deviation) close to

0.01 pixel (∼2 micrometers). Standard deviation are quantified for each variables with the Kalman process assuming that uncertainties follow a Gaussian probability law. These uncertainties concern the whole numerical process but do not include intrinsic optic bias.

The frequency shift of the first bending mode is due to a stiffening effect and is depicted in Figure 9. The continuous wavelet transform (CWT) in this Figure was calculated using the Cauchy wavelet [43]. This frequency tracking was also framed by $\pm 3\sigma_1$ (σ_1 is the standard deviation of the estimation of the first bending mode frequency). One can observe the correspondence between the proposed method and the classical technique based on the wavelet transform at low amplitudes. On the other hand, the stiffening effect is tracked by the EKF from the very beginning at more than 30 Hz, whereas the greatest frequency identified by the CWT is only close to 27 Hz.

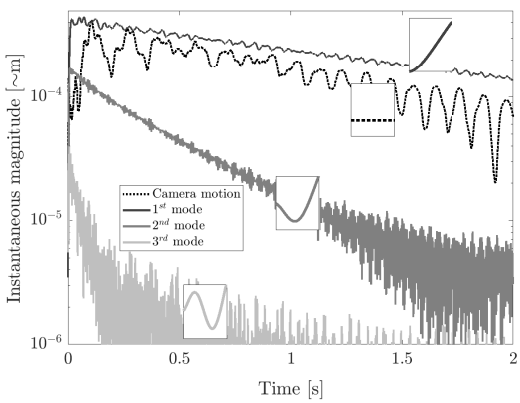


Fig. 7: Evolution of the modal amplitudes along time for the first 3 bending modes of the beam (and the rigid body motion of the camera).

In order to identify the modal stiffness of the first bending mode, it is plotted against the modal displacement in Figure 10. The directly-measured stiffness is the black curve, built from the tracking of the instantaneous frequency and amplitude (the modal stiffness is assumed as the squared frequency). Following previous works on nonlinear normal modes (NNMs), this stiffness can be fitted with a polynomial function and is represented by the gray curve in Figure 10.

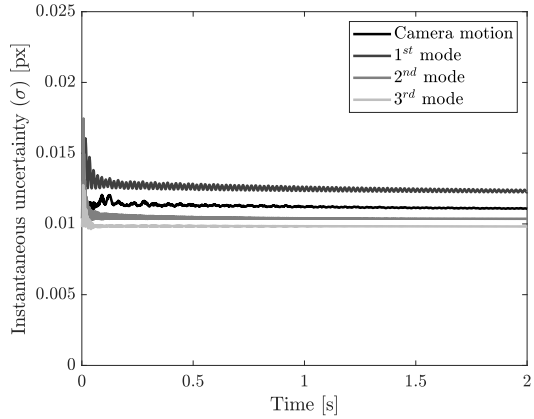


Fig. 8: Uncertainties measurements quantified for each natural mode.

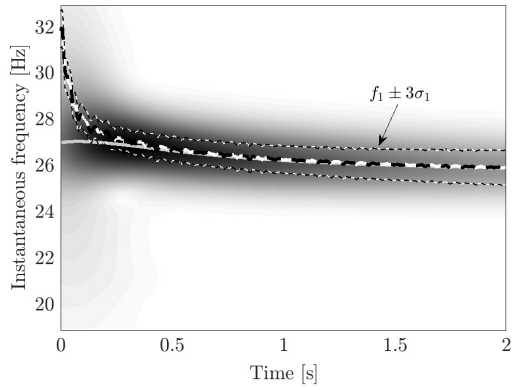


Fig. 9: Instantaneous frequency of the first bending mode: comparison between the proposed method (black and white dotted line) and the CWT (gray shadow and light grey line).

The proposed technique is based on extracting the spatial and temporal coordinates of the beam by locating specific features (the bonded targets) in a video sequence. The coupling of this information with the EKF, through a priori knowledge of the modal shapes, allows the identification of the modal stiffness of the system, in both the linear and nonlinear cases. The EKF does not require knowledge of the complete response history to perform the identification, but only knowledge of the system state and the uncertainty at the previous instant due to the measurement and prediction errors. Indeed, due to its recursive

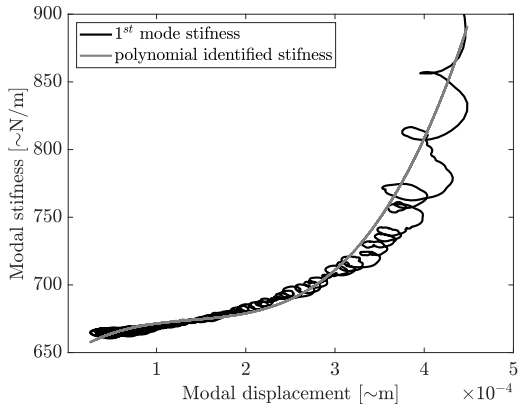


Fig. 10: Identification of the nonlinear stiffness of the first bending mode.

nature, the Kalman filter allows for a real–time implementation, [44].

5 Numerical test case on a strongly nonlinear model

Even if the proposed method is geared towards weakly nonlinear systems in conjunction with video analysis, the main goal of this section is to assess its potential in the presence of strongly nonlinear effects. The aim is to understand the ability of the EKF to track nonlinear resonant frequencies and modal amplitudes with a decomposition based on a linear modal basis.

The chosen nonlinear system [34] possesses 3 degrees of freedom, and Rayleigh damping is introduced in the model:

$$M\ddot{x} + C\dot{x} + Kx + F_{nl} = F \tag{15}$$

where

$$M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ K = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix},$$
$$F_{nl} = \begin{pmatrix} 0 \\ 0.5x_2^3 \\ 0 \end{pmatrix},\ F = \begin{pmatrix} 0 \\ 0 \\ f(t) \end{pmatrix} \tag{16}$$

where C is

$$C = 2.10^{-3}M + 2.10^{-2}K \tag{17}$$

The shock applied to the system is defined as a flat-top window in order to have a controlled and accurate frequency bandwidth. The flat-top window is calculated from:

$$\begin{cases} f(t_n) = a_0 - a_1 \cos\left(\frac{2\pi n}{N-1}\right) + \\ a_2 \cos\left(\frac{4\pi n}{N-1}\right) - a_3 \cos\left(\frac{6\pi n}{N-1}\right) + \\ a_4 \cos\left(\frac{8\pi n}{N-1}\right), \quad 0 \le t \le Ndt \\ f(t_n) = 0, t > Ndt \end{cases} \tag{18}$$

where $t_n = n.dt$, dt is the time step and N is the total number of points during the shock, see [45, 46]. The parameters a_i are listed in Tab. 3. The frequency bandwidth was chosen in order to encompass the 3 linear natural frequencies and their evolution for higher energy levels (see [34], Fig. 11).

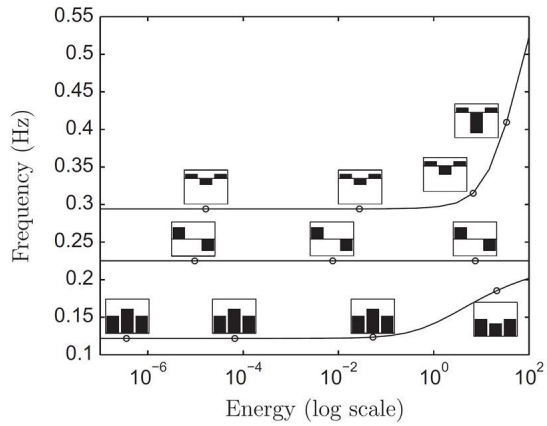


Fig. 11: Frequency–energy plot of the 3 DOF system (Eq. 15) defined by [34]. NNMs represented by bar graphs are inset.

The displacement vector x is obtained here though the numerical integration of the nonlinear system introduced in Eq. 15. These displacements would be obtained experimentally by video analysis by employing at least three virtual sensors placed on each mass. The identification process performed with K-SAMI is depicted in Figure 12. 3 linear modes were used as modal filters for the real-time identification performed by EKF in order to track the 3 NNMs. Kinematic initial conditions of the K-SAMI were randomized and initial

i	0	1	2	3	4
a_i	0.21557895	-0.41663158	0.277263158	-0.083578947	0.006947368

Table 3: Fourier factors of the flat-top window used as a shock function.

frequencies were chosen close to real values (error of 5%).

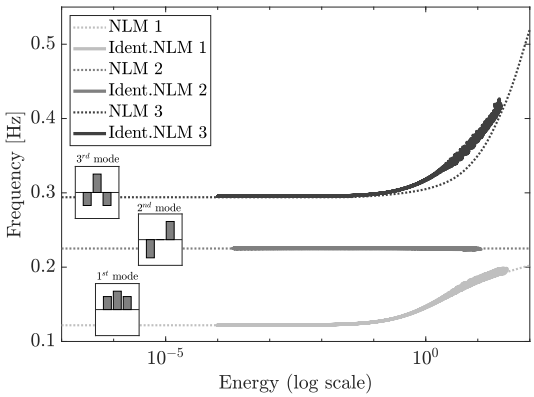


Fig. 12: Frequency–Energy Plot of identified natural frequencies with K-SAMI. The nonlinear first natural frequency is well identified. The linear second natural frequency is accurately tracked. The third nonlinear mode is tracked with a bias.

Despite a large evolution (over 50%), the first natural frequency is well-tracked whereas the mode shape undergoes important changes. However, for high energy levels, the identification process shows larger variance and oscillations in the identified frequency curve. The second mode is quickly and accurately identified as a linear mode without any perturbation along the natural frequency curve. The third mode is strongly nonlinear with great changes in the mode shape and natural frequency. The third natural frequency evolves faster than the first natural frequency, and the identification process leads to a biased natural frequency for high energy levels. This third mode highlights the limits of the K-SAMI process, which assumes that mode shapes are invariant and natural frequencies exhibit "small and smooth" changes.

6 Conclusion

The tracking of nonlinear modes based on the coupling of video analysis with a data assimilation method was successfully performed on an experimental beam featuring a geometrical nonlinearity. The kinematic measurements are obtained with a high resolution/middle speed camera and advanced techniques are used in order to facilitate the tracking process. Specifically, the modal contribution identification is performed with a modal filter combined with an extended Kalman filter and a state space with Hilbert transform of signals. The uncertainty resulting from video analysis is estimated through the observation noise, providing an accuracy about equal to 0.01 pixel for each mode.

The chosen contactless measurement technique and the modal tracking process were shown to be effective for weakly nonlinear modes. Such modes are classically observed on most industrial systems. In order to identify the limits of the method, the tracking was also attempted for a strongly nonlinear numerical example. As further development, an experimental campaign on a strongly nonlinear system is planned to test our findings.

In order to deepen the target-based video analysis reliability for vibrations measurements, the next developments will focus on the estimation of the uncertainty resulting from video recording conditions and image post-processing. Indeed, the EKF method enables to quantify this uncertainty in the observation noise term without identify the error sources. At first, more numerical example will be implemented with the aim of quantifying the amount of error coming from the feature detection and tracking. Then, further experimental tests will be conducted by coupling virtual and physical sensors to address the uncertainty related to the camera and lens properties, the camera position, its rigid motion and the lighting conditions.

Acknowledgment

We thank Tilàn Dossogne (University of Liège) and Reza Babajanivalashedi (ISAE-Supméca) for their time and help during the experimental campaign.

Data availability statement

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

[1] J. Baqersad, P. Poozesh, C. Niezrecki, P. Avitabile, Photogrammetry and optical methods in structural dynamics – a review. Mechanical Systems and Signal Processing **86**, 17–34 (2017). https://doi.org/https://doi.org/10.1016/j.ymssp.2016.02.011

[2] W. Wang, J.E. Mottershead, T. Siebert, A. Pipino, Frequency response functions of shape features from full-field vibration measurements using digital image correlation. Mechanical Systems and Signal Processing **28**, 333–347 (2012). https://doi.org/10.1016/j.ymssp.2011.11.023

[3] T.J. Beberniss, D.A. Ehrhardt, High-speed 3D digital image correlation vibration measurement: Recent advancements and noted limitations. Mechanical Systems and Signal Processing **86**(B), 35–48 (2016). https://doi.org/10.1016/j.ymssp.2016.04.014

[4] B. Pan, L. Yu, Q. Zhang, Review of single-camera stereo-digital image correlation techniques for full-field 3D shape and deformation measurement. Sci. China Technol. Sci. **61**, 2–20 (2018)

[5] L. Yu, B. Pan, Single-camera high-speed stereo-digital image correlation for full-field vibration measurement. Mechanical Systems and Signal Processing **94**, 374–383 (2017). https://doi.org/10.1016/j.ymssp.2017.03.008

[6] J. Javh, J. Slavič, M. Boltežar, The subpixel resolution of optical-flow-based modal analysis. Mechanical Systems and Signal Processing **88**, 89–99 (2017). https://doi.org/10.1016/j.ymssp.2016.11.009

[7] J. Javh, J. Slavič, M. Boltežar, Experimental modal analysis on full-field dslr camera footage using spectral optical flow imaging. Journal of Sound and Vibration **434**, 213–220 (2018). https://doi.org/10.1016/j.jsv.2018.07.046

[8] D. Gorjup, J. Slavič, A. Babnik, M. Boltežar, Still-camera multiview spectral optical flow imaging for 3D operating-deflection-shape identification. Mechanical Systems and Signal Processing **152**, 107,456 (2021). https://doi.org/10.1016/j.ymssp.2020.107456

[9] 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. Journal of Sound and Vibration **345**, 58–71 (2015). https://doi.org/10.1016/j.jsv.2015.01.024

[10] N. Wadhwa, H.Y. Wu, A. Davis, M. Rubinstein, E. Shih, G.J. Mysore, J.G. Chen, O. Buyukozturk, J.V. Guttag, W.T. Freeman, F. Durand, Eulerian video magnification and analysis. Commun. ACM **60**(1), 87–95 (2016). https://doi.org/10.1145/3015573

[11] Y. Yang, C. Dorn, T. Mancini, Z. Talken, G. Kenyon, C. Farrar, D. Mascareñas, Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification. Mechanical

System and Signal Processing **85**, 567–590 (2017). https://doi.org/10.1016/j.ymssp.2016.08.041

[12] Z. Shang, Z. Shen, Multi-point vibration measurement and mode magnification of civil structures using video-based motion processing. Automation in Construction **93**, 231–240 (2018). https://doi.org/10.1016/j.autcon.2018.05.025

[13] J.W. Park, J.J. Lee, H.J. Jung, H. Myung, Vision-based displacement measurement method for high-rise building structures using partitioning approach. NDT & E International **43**(7), 642–647 (2010). https://doi.org/10.1016/j.ndteint.2010.06.009

[14] P.F. Pai, S. Ramanathan, J. Hu, D.K. Chernova, X. Qian, G. Wu, Camera-based non-contact metrology for static/dynamic testing of flexible multibody systems. Measurement Science and Technology **21**(8), 085,302 (2010). https://doi.org/10.1088/0957-0233/21/8/085302

[15] M. Romaszko, B. Sapiński, A. Sioma, Forced vibrations analysis of a cantilever beam using the vision method. Journal of Theoretical and Applied Mechanics p. 243 (2015). https://doi.org/https://doi.org/10.15632/jtam-pl.53.1.243

[16] G. Busca, A. Cigada, P. Mazzoleni, E. Zappa, Vibration monitoring of multiple bridge points by means of a unique vision-based measuring system **54**, 255–271 (2014). https://doi.org/https://doi.org/10.1007/s11340-013-9784-8

[17] B. Ferrer, D. Mas, J. García-Santos, et al., Parametric study of the errors obtained from the measurement of the oscillating movement of a bridge using image processing. J Nondestruct Eval **35**(53) (2016). https://doi.org/https://doi.org/10.1007/s10921-016-0372-6

[18] H. Jeon, Y. Bang, H. Myung, A paired visual servoing system for 6-dof displacement measurement of structures. Smart Materials and Structures **20**, 045,019 (2011). https://doi.org/10.1088/0964-1726/20/4/045019

[19] D. Ribeiro, R. Calçada, J. Ferreira, T. Martins, Non-contact measurement of the dynamic displacement of railway bridges using an advanced video-based system. Engineering Structures **75**, 164–180 (2014). https://doi.org/10.1016/j.engstruct.2014.04.051

[20] K. Henke, R. Pawlowski, P. Schregle, et al., Use of digital image processing in the monitoring of deformations in building structures. J Civil Struct Health Monit **5**, 141–152 (2015). https://doi.org/https://doi.org/10.1007/s13349-014-0091-6

[21] H.V. de Figueiredo, D.F. Castillo-Zúñiga, N.C. Costa, O. Saotome, R.G.A. da Silva, Aeroelastic vibration measurement based on laser and computer vision technique. Exp Tech **45**, 95–107 (2021). https://doi.org/https://doi.org/10.1007/s40799-020-00399-0

[22] D. Mas, J. Espinosa, A.B. Roig, B. Ferrer, J. Perez, C. Illueca, Measurement of wide frequency range structural microvibrations with a pocket digital camera and sub-pixel techniques. Appl. Opt. **51**(14), 2664 (2012). https://doi.org/10.1364/ao.51.002664

[23] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(10), 1615–1630 (2005). https://doi.org/10.1109/TPAMI.2005.188

[24] F. Fraundorfer, H. Bischof, in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops* (2005), pp. 33–33. https://doi.org/10.1109/CVPR.2005.393

[25] J. Matas, O. Chum, M. Urban, T. Pajdla, Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing **22**(10), 761–767 (2004). https://doi.org/10.1016/j.imavis.2004.02.006. British Machine Vision Computing 2002

[26] B. Ferrarini, S. Ehsan, A. Leonardis, N.U. Rehman, K.D. McDonald-Maier, Performance characterization of image feature detectors in relation to the scene content utilizing a large image database. IEEE Access **6**,

8564–8573 (2018). https://doi.org/10.1109/ACCESS.2018.2795460

[27] F. Renaud, S. Lo Feudo, J.L. Dion, A. Goeller, 3D vibrations reconstruction with only one camera. Mechanical Systems and Signal Processing **162**, 108,032 (2022). https://doi.org/https://doi.org/10.1016/j.ymssp.2021.108032

[28] P.L. Reu, D.P. Rohe, L.D. Jacobs, Comparison of DIC and LDV for practical vibration and modal measurements. Mechanical Systems and Signal Processing **86**, 2–16 (2017). https://doi.org/10.1016/j.ymssp.2016.02.006

[29] T. Durand-Texte, M. Melon, E. Simonetto, S. Durand, M.H. Moulet, Single-camera single-axis vision method applied to measure vibrations. Journal of Sound and Vibration **465**, 115,012 (2020). https://doi.org/10.1016/j.jsv.2019.115012

[30] P. Kohut, P. Kurowski, Application of modal analysis supported by 3D vision-based measurements. Journal of Theoretical and Applied Mechanics **47**(4), 855–870 (2009)

[31] D. Gorjup, J. Slavič, M. Boltežar, Frequency domain triangulation for full-field 3D operating-deflection-shape identification. Mechanical Systems and Signal Processing **133**, 106,287 (2019). https://doi.org/https://doi.org/10.1016/j.ymssp.2019.106287

[32] J.C. Golinval, P. Argoul, COST ACTION F3 "structural dynamics" 1997-2001 presentation and some preliminary results. Structural Control for Civil and Infrastructure Engineering pp. 263–280 (2001)

[33] A. Goeller, J.L. Dion, R. Le Breton, T. Soriano, Kinematic SAMI : a new real-time multi-sensor data assimilation strategy for nonlinear modal identification. Mechanics & Industry **21**(4), 413 (2020). https://doi.org/10.1051/meca/2020035

[34] G. Kerschen, M. Peeters, J.C. Golinval, A.F. Vakakis, Nonlinear normal modes, part I: A useful framework for the structural dynamicist. Mechanical Systems and Signal Processing **23**(1), 170–194 (2009). https://doi.org/10.1016/j.ymssp.2008.04.002. Special Issue: Non-linear Structural Dynamics

[35] A. Goeller, Modèles réduits dynamiques pour la synthèse de séquences vidéo - application à la compression de données, la réalité augmentée et la prédiction. Ph.D. thesis, Univ. Paris Saclay SMEMAG (2018)

[36] J. Shi, C. Tomasi, Good features to track. IEEE Conference on Computer Vision and Pattern Recognition pp. 593–600 (1994)

[37] P. Torr, A. Zisserman, Mlesac: A new robust estimator with application to estimating image geometry. Computer Vision and Image Understanding **78**(1), 138–156 (2000). https://doi.org/https://doi.org/10.1006/cviu.1999.0832

[38] J.L. Dion, C. Stephan, G. Chevallier, H. Festjens, Tracking and removing modulated sinusoidal components: A solution based on the kurtosis and the extended kalman filter. Mechanical Systems and Signal Processing **38**(2), 428–439 (2013). https://doi.org/10.1016/j.ymssp.2013.04.001

[39] M. Peeters, G. Kerschen, J.C. Golinval, Modal testing of nonlinear vibrating structures based on nonlinear normal modes: Experimental demonstration. Mechanical Systems and Signal Processing **25**(4), 1227–1247 (2011). https://doi.org/10.1016/j.ymssp.2010.11.006

[40] M. Peeters, G. Kerschen, J.C. Golinval, Dynamic testing of nonlinear vibrating structures using nonlinear normal modes. Journal of Sound and Vibration **330**(3), 486–509 (2011). https://doi.org/https://doi.org/10.1016/j.jsv.2010.08.028

[41] E. Jones, M.E. Iadicola (eds.), *A Good Practices Guide for Digital Image Correlation* (International Digital Image Correlation Society, 2018). https://doi.org/10.32720/idics/gpg.ed1

e

[42] M. Feldman, *Hilbert Transform Applications in Mechanical Vibration* (Wiley, 2011)

[43] T.P. Le, P. Argoul, Continuous wavelet transform for modal identification using free decay response. Journal of Sound and Vibration **277**(1), 73–100 (2004). https://doi.org/10.1016/j.jsv.2003.08.049

[44] C.K. Chui, G. Chen, *Kalman Filtering with Real-Time Applications*. Springer Series in Information Sciences (Springer Berlin, Heidelberg, 2013)

[45] S. Gade, H. Herlufsen, Use of weighting functions in DFT/FFT analysis (part I). Tech. Rep. 3, Brüel & Kjær Technical Review (1987)

[46] G. Heinzel, A.O. Rüdiger, R. Schilling, Spectrum and spectral density estimation by the discrete fourier transform (DFT), including a comprehensive list of window functions and some new at-top windows. Tech. rep., Max Planck Institute für Gravitationsphysik / Laser Interferometry & Gravitational Wave Astronomy (2002)