Mechanical Systems and Signal Processing 249 (2026) 114102

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Full Length Article

Vibration measurement with neuromorphic vision sensors

Sofia Baldini *, Filippo Stazi, Riccardo Bernardini, Andrea Fusiello, Paolo Gardonio, Roberto Rinaldo

Università degli Studi di Udine - DPIA, Via delle Scienze 206, 33100 Udine, Italy

ARTICLE INFO

Keywords:
Photogrammetry
Event cameras
3D-point-tracking
Multi-view triangulation

ABSTRACT

In this paper, a novel methodology for vibration measurements of flexible structures with neuromorphic vision sensors (i.e. event cameras) is proposed and validated. Unlike conventional frame-based sensors, event cameras asynchronously record changes in scene brightness, enabling high temporal resolution with reduced data redundancy. In this study, a framework is developed for marker tracking and 3D triangulation from event recordings. The experimental setup involves four synchronized event cameras observing a cantilever beam coated with a grid of markers. The beam is excited harmonically at the first two flexural resonance frequencies, and the resulting 3D trajectories of the markers are reconstructed with subpixel accuracy. The resulting flexural deflection shapes are reconstructed through bundle adjustment of the camera recordings. The results have been benchmarked against laser vibrometer measurements, which confirmed the accuracy and applicability of the proposed measurement approach. Overall, the study demonstrates the potential of event cameras for dynamic testing in scenarios where high-speed, low-latency, low-bandwidth acquisitions are required without recording redundant data.

1. Introduction

Vibration measurements play a fundamental role in numerous engineering applications, including structural health monitoring (SHM), mechanical diagnostics, and predictive maintenance [1–4]. Conventional approaches often rely on contact sensors such as accelerometers or strain gauges, which provide accurate measurements but present several limitations: for instance, they require physical bonding to the structure, which could alter the vibration response particularly of lightweight structures. Also, they are sensitive to electromagnetic noise and temperature variations, and they involve complex cabling arrangements [1,5].

Over the past three decades, non-contact optical techniques have emerged as a viable alternative to classic vibration measurement techniques due to their ability to acquire high-resolution images with high frame-rate without influencing the structural behaviour which is especially advantageous when working with lightweight or delicate structures [6–9]. For conventional, frame-based cameras, various computer-vision approaches have been developed to estimate structural displacements from image sequences, including 2D Point Tracking (2DPT) [10,11], 3D Point Tracking (3DPT) [12], and Digital Image Correlation (DIC) methods [13,14] as well as more recent targetless solutions [15,16]. Additionally, in scenarios where computational efficiency is critical, simplified vision techniques, such as optical flow methods, can be used to accelerate post-processing [17]. Early vision-based vibration measurement methods predominantly relied on single-camera setups, which are inherently limited to detecting in-plane (e.g., flexural) vibrations of one-dimensional structures. The reconstruction of full 3D displacements requires at least two independent viewpoints to enable

* Corresponding author.
E-mail address: baldini.sofia@spes.uniud.it (S. Baldini).

https://doi.org/10.1016/j.ymssp.2026.114102
Received 10 October 2025; Received in revised form 31 December 2025; Accepted 27 February 2026
Available online 4 March 2026

triangulation. This limitation can be overcome by employing synchronized stereo or multi-camera systems in conjunction with 3D digital image correlation (3D-DIC) techniques [13,18]. However, several authors have pointed out that such multi-camera configurations may suffer from errors due to imperfect synchronization and geometric misalignment, and have used these concerns to motivate the adoption of single-camera alternatives, based on additional optical components, such as mirror arrays, that project multiple perspectives of the vibrating surface onto a single sensor [19]. While effective in mitigating synchronization errors, these approaches typically involve a trade-off in terms of spatial resolution. More recently, frequency-domain triangulation has been introduced as an alternative means of recovering 3D vibration information from a single conventional camera [20,21]. This technique operates in the frequency domain, exploits the full sensor resolution, and can be implemented with still-frame cameras without requiring inter-frame synchronization [22]. The idea of using multiple-cameras to acquire multi-view images has also been explored [23].

A key limitation of traditional video-based methods is the trade-off between frame rate, resolution, and cost. Capturing high-frequency vibrations requires high-speed cameras operating at thousands of frames per second, often with limited dynamic range and substantial bandwidth and data storage requirements. Furthermore, motion blur and low light sensitivity can affect the accuracy of measurements, particularly for vibration measurements. High-resolution and high-speed cameras are expensive and generate large data volumes, which limit scalability. However, the possibility of measuring high-frequency vibrations with low-speed cameras has been analysed in the literature (see, among others, [24–27]) The proposed techniques allow high-frequency information to be reconstructed following random or periodic sampling at sub-Nyquist frequencies. However, the techniques generally require structured light sources and precise synchronization of sampling instants, which in practice are subject to jitter, as well as specific assumptions about the excitation signal. In particular, subsampling is possible by assuming an exactly periodic signal [24,27], or by assuming an analytical model of the vibration, whose parameters are estimated by minimizing a cost function on the available data [25,26]. In this context, neuromorphic cameras (also known as event cameras or dynamic vision sensors) have recently gained attention as a novel tool for high-speed optical vibration measurement [28–32]. The simplicity of the setup, the lack of specific assumptions required, and the low cost make these a potentially attractive alternative. Unlike conventional frame-based sensors, event cameras asynchronously detect pixel-level brightness changes, generating streams of timestamped "events" with microsecond latency and negligible data redundancy [28,29,33,34]. These features make them particularly suited for monitoring fast structural motions and capturing high-frequency vibrations that traditional sensors may fail to record. While event cameras have been widely adopted in fields such as robotics, autonomous vehicles, and neuroscience [34–36], their application to structural vibration analysis is quite recent and relatively unexplored. For instance, Dorn *et al.* [37] first measured vibrations with one event camera located at grazing angle on a beam structure. Lai *et al.* [38] have developed a physics-informed sparse identification framework that uses event cameras to capture full-field structural vibrations, Shi *et al.* [39] have proposed a laser-assisted event camera method for accurate, non-contact vibration frequency measurement. Also, Zhao *et al.* [40] have presented an event camera–based algorithm to reconstruct vibration signals and extract amplitude–frequency characteristics with experiments on mechanical equipment. In fact, the asynchronous nature and sparse output of event cameras pose challenges for direct application of classical computer vision methods. To bridge this gap, various approaches have been proposed, either to extract motion directly from the event stream or to reconstruct high-speed intensity images that are compatible with conventional tracking pipelines [28,35,41–45]. In particular, neural-network-based reconstruction techniques such as E2VID [46,47] have shown promising results in synthesizing intensity images from event data with sufficient spatial and temporal fidelity offering a practical route to integrate event cameras into established tracking techniques.

Event-based data processing strategies can be broadly divided into two categories. The first processes the asynchronous stream of events directly to estimate motion quantities such as optical flow, velocity, or pose increments. While this event-only paradigm has shown promising results in several vision tasks, its application to quantitative, multi-view structural vibration measurement remains challenging. Direct event-based methods are inherently sensitive to background activity and noise, strongly depend on local image contrast and texture, and may yield spatially inhomogeneous measurements in regions with small motion amplitudes or low brightness gradients. Moreover, many event-only approaches estimate motion incrementally, which complicates the recovery of absolute displacement, phase, and modal shapes over long sequences, and they are not readily compatible with standard photogrammetric pipelines based on triangulation and bundle adjustment.

The second strategy reconstructs intensity images from event data and then applies conventional computer-vision and photogrammetric techniques. In this work, we adopt this approach in order to leverage well-established multi-view geometry tools while retaining the high temporal resolution and low data redundancy of event cameras. Although neural-network-based reconstruction methods such as E2VID may introduce artifacts, including temporal smoothing or reconstruction bias in regions of sparse event activity, these effects are acceptable in the present context because the reconstructed images are used exclusively for marker localization. The subsequent multi-view triangulation and bundle adjustment further mitigate residual reconstruction errors by enforcing geometric consistency across cameras and time.

Building on the early work with two cameras only presented in [31,32], this study investigates the use of multiple event cameras for the non-contact measurement of mechanical vibrations. By combining event-based sensing with stereo photogrammetry, the flexural vibration field of a cantilever beam excited harmonically is reconstructed at the first two resonance frequencies such that the first two flexural deflection shapes are reconstructed. The system tracks a set of discrete high-contrast markers across the reconstructed intensity frames and triangulates their 3D motion. The method is validated through quantitative comparison with a high-precision laser Doppler vibrometer, focusing on the reconstruction of the first and second flexural deflection shape.

This paper presents a novel framework for high-speed, non-contact 3D vibration measurement using stereo event cameras. Also, it provides experimental validation against ground truth measurements from a laser vibrometer, demonstrating high accuracy and applicability for structural dynamic analysis. Hence, rather than on the vision algorithms, this paper contributes at system and

methodological levels by closing the gap between event-vision research and experimental structural dynamics / optical metrology.

This paper is structured into five Sections. To start with, Section 2 describes the event-based measurement system and the experimental setup. Then, Section 3 presents the processing pipeline, including event-to-intensity reconstruction, stereo calibration, and 3D point tracking. Section 4 discusses the experimental results and compares them with laser vibrometer data. Finally, Section 5 concludes the paper with remarks on the proposed approach.

2. Neuromorphic cameras and experimental setup

Vibration measurements in engineering applications typically span in a frequency range between 2 and 5 kHz. To accurately reconstruct the vibrational field using optical sensors, it is therefore essential to utilize cameras capable of capturing fast dynamic changes. For this purpose, event cameras have been investigated, as they offer a cost-effective, full-frame solution for vibration measurement. Unlike traditional laser vibrometers [48], which acquire data sequentially one point at a time, or conventional high-speed cameras that acquire large amount of data from full frame sequences of images, event cameras provide a compelling alternative for high-frequency motion capture. This section introduces the event-based sensing technology employed in the study, focusing on the operating principles and hardware architecture of event cameras, as well as their advantages in high-speed dynamic measurements. Then, a detailed description of the experimental setup is presented, including the configuration of the multi-camera system and the reference instrumentation used to assess measurement accuracy. The test case consists of a cantilever beam undergoing flexural vibrations, which enables the evaluation of the proposed event-based vibration measurement approach in a realistic structural dynamic scenario.

2.1. Neuromorphic cameras

Neuromorphic or event cameras have emerged as a revolutionary advancement in the field of computer vision thanks to their unique architecture that mimics biological vision systems. These tools are also referred to as neuromorphic or dynamic vision sensors and operate on principles that differ fundamentally from those of conventional frame-based imaging systems. Their architecture enables significant advantages in various application domains, including robotics, autonomous vehicles, and eye tracking. Focusing on the hardware architecture, event cameras are built around a unique design that enables each pixel to operate independently, detecting local temporal variations in brightness instead of recording full frames at fixed time intervals. At the core of the sensor is an array of event-sensitive pixels, commonly referred to as Dynamic Vision Sensor (DVS) pixels. Each pixel consists of a photodiode that continuously monitors the incident light energy, coupled with an integrated comparator circuit. As shown in Fig. 1, an event is triggered only when the logarithmic change in brightness exceeds a predefined positive or negative threshold. The resulting event is encoded as a data packet comprising the spatial coordinates of the pixel (x,y), a high-resolution timestamp (t), and the polarity of the change (positive or negative).

This pixel array is interfaced with an Application-Specific Integrated Circuit (ASIC), a dedicated microelectronic component responsible for orchestrating the generation, processing, and transmission of event data. The ASIC performs several critical functions, including high-precision timestamping, on-chip data compression, and communication via standard digital interfaces such as USB, MIPI, or SPI. Due to the asynchronous nature of event generation, the system necessitates an efficient communication protocol capable of sustaining a high-throughput, continuous data stream, while maintaining a substantially lower bandwidth requirement compared to conventional frame-based cameras. Thanks to its neuromorphic design, the event camera system enables ultra-fast visual acquisition, high energy efficiency, and robustness in dynamic and complex visual environments. More in detail, key features of event cameras include their low power consumption and high dynamic range, reaching levels up to 110 dB, compared to the mere 60 dB typical of traditional cameras [34]. These capabilities allow event cameras to operate effectively in challenging lighting conditions and fast-moving environments without suffering from motion blur (a common problem in conventional cameras). For this study, four DVXplorer event cameras by iniVation [49] were employed. Accurate synchronization among multiple Dynamic Vision Sensors (DVS) is a critical requirement in applications where precise temporal alignment of data streams is essential. This is particularly true in multi-view setups used for 3D triangulation, where sub-millimetre displacements of vibrating structures must be accurately reconstructed. In this work, hardware-based synchronization was adopted. The four event cameras were connected via a dedicated trigger cable, with

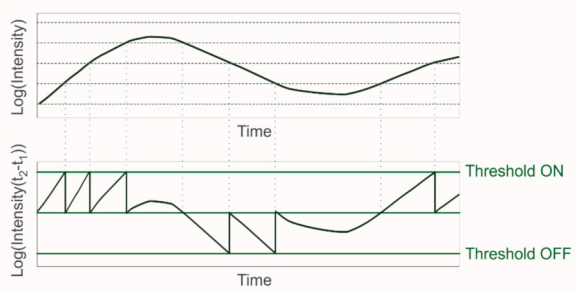


Fig. 1. Principle of operation of an event camera sensor.

one camera designated as the master. This setup ensures high temporal coherence between the clocks of all sensors, enabling accurate temporal alignment of asynchronous event streams across different viewpoints. The accuracy of this synchronization has been validated in the experiment reported below in Section 4.1.

During the vibration experiments, the measured event data rate was approximately 35 Mbytes/s. For comparison, an equivalent frame-based acquisition at 1000 fps would generate approximately 293 Mbytes/s, assuming the same spatial resolution. This comparison highlights the substantially lower data rate of the event-based acquisition under the experimental conditions considered. The storage requirements follow directly from the data rate and the typical experiment duration, which in our case is approximately 10 s. Regarding latency, the iniVation DVXplorer event camera is specified to have sub-millisecond latency. The camera temporal resolution, corresponding to the internal timestamp granularity, was measured in our setup to be approximately 130 μs, which is consistent with the nominal specification of 200 μs. This temporal resolution sets the effective timing precision of the event stream.

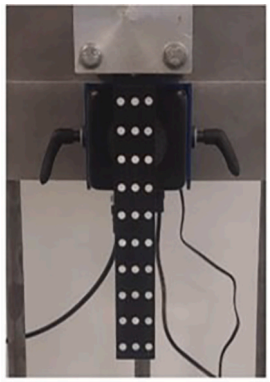
2.2. Experimental setup

This study focuses on measuring with four synchronized event cameras the flexural deflection shapes of a beam excited harmonically by a shaker at the first and second resonance frequencies (corresponding to the primary two resonance frequencies) of a cantilever beam. The beam, fabricated from steel, has its main physical properties summarized in Table 1, alongside the main specifications of the event cameras employed.

As illustrated in Fig. 2a, the beam surface is subdivided into a uniform grid of 10×3 points, each point is denoted with a silver circular marker. There are no strict requirements on the lighting conditions or on the specific marker type. In practice, the markers are selected to ensure a high visual contrast with respect to the beam surface, so that they can be easily and reliably identified by the image processing pipeline. High reflectivity is not a strict requirement; rather, clear contrast between the markers and the background is the key factor for robust detection. The spatial coordinates of these markers are defined with respect to a coordinate system (o, x, y, z) located near the top right corner and offset by 5 mm from the cantilever beam longest edge. As shown in Fig. 2b, the bending excitation is provided by a shaker coupled to the beam through a stinger. Additionally, Fig. 2a highlights a specific marker selected as the reference point for depicting the time-harmonic response discussed in Section 4. The experimental setup includes four synchronized event cameras (see Fig. 2a) with optical axes oriented toward the centre of the cantilever beam. These cameras are positioned approximately along circular arcs centred on the beam and whose radius is approximately 50 cm. The four cameras are mounted at slightly different vertical levels, arranged in two pairs: two cameras are placed at lower elevations with a small vertical offset between them, while the remaining two are positioned at higher levels, also with a slight relative offset. This configuration ensures adequate spatial coverage and triangulation accuracy.

Table 1
Cantilever beam physical properties and event cameras specifications.

Parameter	**Cantilever Beam**
Length	$L = 210mm$
Width	$h = 30mm$
Thickness	$s = 2mm$
Density	$\rho = 7850kg/m^3$
Young's modulus	$E = 21 \times 10^{10} N/m^2$
Poisson ratio	$\nu = 0,3$
Position of the shaker force excitation from the clamp	$y = 45mm$
Grid of measured points	9×3



Parameter	**Camera**
Model	iniVation DVXplorer
Spatial resolution	$640 \times 480\, pixels$
Temporal resolution	$200\mu s$
Typical latency	$< 1ms$
Dynamic range	~ 90 dB (3–100 k lux with 99.9% of pixels respond to 27.5% contrast) ~ 110 dB (0.3–100 k lux with 50% of pixels respond to 80% contrast)
Contrast sensitivity	13% (with 50% of pixels respond) 27.5% (with 99.9% of pixels respond)
Marker diameter	$5mm$
Radial distance	$d \cong 500mm$
Ground sampling distance (average)	0.56

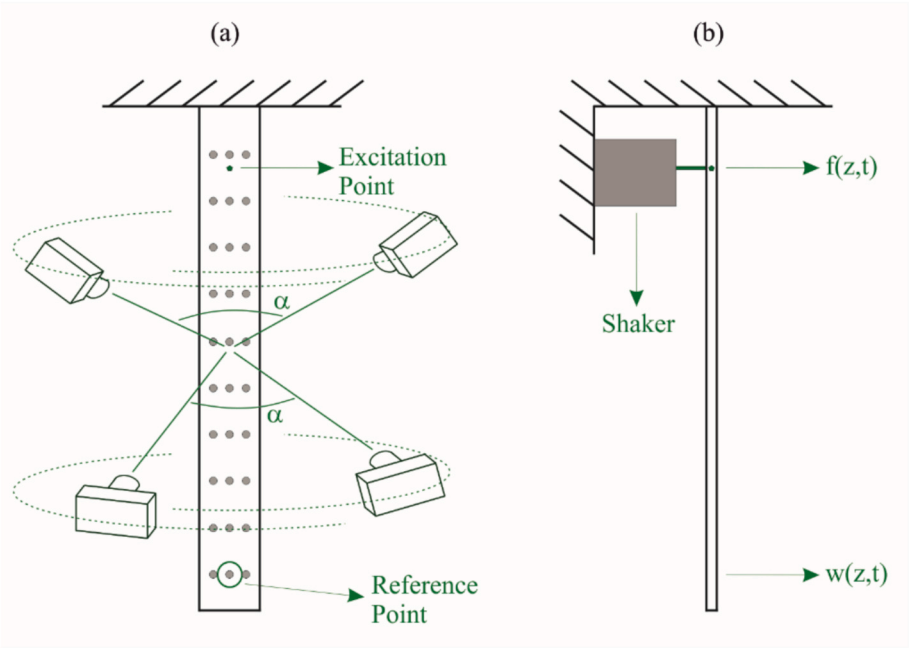


Fig. 2. Sketch of the experimental setup a) front view of the cantilever beam with 4 event cameras and reference point, b) lateral view with excitation position and main direction of the displacements.

In addition to the event cameras, the setup incorporates a scanning laser vibrometer. This equipment is employed to acquire reference measurements of the beam transverse vibration field at the predefined grid of marker positions.

3. Methodology

This section presents the processing pipeline adopted to extract accurate 3D vibration measurements from asynchronous event-based data. First, the set of open-source tools employed to manage and process the raw output from the event cameras are described. Then, the procedures for reconstructing intensity frames from event streams and for performing intrinsic and extrinsic camera calibration is detailed. Finally, the methods used for marker detection, point tracking, and 3D triangulation, which enable the reconstruction of the vibrational field from multiple camera views are outlined.

3.1. Open-source tools for event-based vibration measurements

In this study, the implementation of event-based vibration measurement techniques is supported by a robust collection of open-source tools designed to facilitate efficient data acquisition, processing, and analysis. The processing pipeline integrates well-established platforms such as MATLAB [50,51], Python, and the pre-trained neural network E2VID [46].

The reconstruction of intensity images from raw event streams is carried out using E2VID, a recurrent, fully convolutional neural network tailored for event-based vision. Its architecture is inspired by UNet [52] and includes multiple encoder–decoder layers linked by skip connections. Key components include depth-wise convolutions, ConvLSTM [53] units to capture temporal dependencies, and bilinear upsampling for image refinement. Batch normalization and ReLU activations are employed to ensure stable training and inference.

E2VID operates in two modes: a fixed frame rate mode, where intensity images are reconstructed at regular intervals, and a variable frame rate mode, where frame generation is driven by event density. The fixed-rate mode is particularly suited to multi-camera setups and 3D triangulation, as it guarantees temporal synchronization across sensors. In contrast, the variable frame rate mode, generates frames dynamically based on event density and it is optimized for scenarios with intermittent motion. The selection of frame rate must ensure a balance between event density and image quality: low frame rates may cause excessive event accumulation, while high frames rates lead to insufficient event data, causing loss of detail.

The Computer Vision and Calibration Toolkit for MATLAB [50,51], which is also compatible with Octave, is employed for camera calibration, multi-view geometry, and 3D reconstruction. This self-contained library requires no additional toolboxes and provides all necessary functions for intrinsic/extrinsic calibration, bundle adjustment, and triangulation, making it ideal for vibration-related displacement analysis [54].

The adoption of open-source tools over commercial alternatives is driven by two main advantages. The first one is based on flexibility and customization: unlike commercial solutions, open-source tools allow researchers to modify algorithms, optimize

performance, and tailor processing pipelines to specific experimental needs. The second advantage pertains to community support and continuous development: these tools benefit from active communities that continuously contribute to their development, maintenance, and documentation.

By leveraging open-source resources, the methodologies developed in this work remain fully transparent, reproducible, and readily extendable. The following sub-sections describe how these tools are applied to key stages of the workflow, including data reconstruction, marker tracking, and 3D vibration field estimation.

3.2. Intensity frames reconstruction and calibration procedure

Although, in principle event cameras encode the full spatiotemporal visual signal in their event streams, the data format is fundamentally incompatible with standard vision pipelines. To bridge this gap, two main strategies have been proposed in the literature: direct event-based processing [28,35,41,55] and the reconstruction of intermediate intensity representations [56–58]. Although conceptually elegant, the former presents significant implementation difficulties and remains relatively underexplored. Instead, the latter approach, which is adopted in this study, offers a more practical and widely applicable solution by generating intermediate intensity frames (an example is reported in Fig. 3) from the raw event stream. Nevertheless, although this approach makes it possible to apply conventional computer vision techniques to event data, this is not without challenges. In fact, reconstructing a coherent and structurally meaningful intensity image from these sparse spatiotemporal samples constitutes an ill-posed problem, requiring sophisticated algorithms such as neural networks, to infer a plausible representation of the original scene. Moreover, event cameras produce incomplete and asynchronous information that reflects only local changes in brightness rather than full-frame images.

Transforming the flow of raw event data into usable information requires the aggregation of events over short temporal windows to approximate a visual representation of the scene. The choice of accumulation time is a critical parameter, as it directly influences the temporal resolution and the fidelity of motion representation. A well-calibrated window should satisfy the Nyquist–Shannon sampling criterion to avoid temporal aliasing, and should ensure that a sufficient density of events is available to generate a sequence of intensity frames of adequate quality. These frames must retain enough information to support subsequent processing steps, such as point tracking and 3D triangulation, which are normally applied in frame-based image processing for 3D point tracking.

In the experiments carried out for this research, the E2VID network is used in fixed frame-rate mode with overlapping window. The window size guarantees a sufficient density of events while the window shift determines the frame rate. The E2VID neural network has been employed with zero overlapping window for the detection of the first flexural deflection shape and with an overlapping window of about 50% for the measurement of the second flexural deflection shape. In this way on one hand, it was possible to accumulate the necessary number of events to achieve a frame reconstruction with a good enough definition of the markers to perform the marker tracking procedure and on the other hand, it was possible to see how the overlapping window can improve performances.

Prior to image acquisition, all four event cameras underwent a thorough calibration procedure. As customary in photogrammetry and computer vision [59], each pinhole camera model is defined by two distinct sets of parameters: intrinsic and extrinsic. The intrinsic parameters, such as focal length, principal point, and lens distortion coefficients, characterise the internal geometry and optical properties of the camera. In contrast, the extrinsic parameters define the spatial *pose* (position and orientation) of each camera in the world coordinate system, which is referred to as the exterior orientation. In the present study, during a preliminary setup phase, the intrinsic calibration of each camera, including the correction of radial distortion, was performed individually using the Sturm–Maybank–Zhang method [60,61].

Fig. 3b shows a representative intensity frame reconstructed via the E2VID neural network from a sequence of events recorded on a checkerboard. Such planar calibration targets are widely used in multi-camera calibration tasks due to their geometric regularity. As

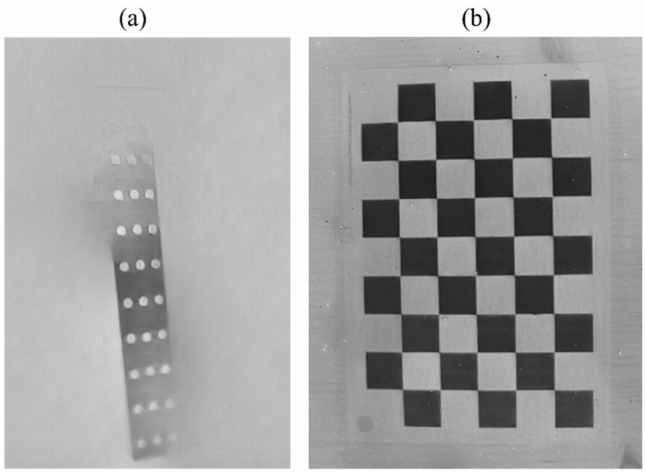


Fig. 3. Intensity frames reconstructed by pre-trained neural network E2VID, a) cantilever beam, b) checkerboard.

illustrated in Fig. 4, to calibrate the four cameras, multiple views of the checkerboard were acquired independently with each device. These views served as input to the calibration algorithm, yielding the intrinsic parameters summarised in Table 2. The computational time of the overall proposed procedure is dominated by the time required to reconstruct the frames from the event sequence. Using a computer equipped with Intel(R) Core(TM) i7-10700KF, CPU clock 3.80 GHz, 32 GB ram, 64 bit Windows 11 Pro operating system, Python 3.10.13 implementation, the time required by E2VID to reconstruct one frame is approximately one second. Real-time processing, not the objective of this contribution, is therefore not feasible.

3.3. Marker tracking and triangulation

Once the intensity frames for each camera were reconstructed, the next essential step involves the detection and tracking of the silver markers glued on the cantilever beam, necessary for the 3D triangulation and flexural vibration reconstruction. As previously introduced, a regular grid of 9×3 circular markers have been stuck on the surface of the beam. Here, the objective is to extract their 2D coordinates at each frame from all four camera views and reconstruct their corresponding 3D trajectories over time. According to Ref. [23], the procedure implemented for each camera involves the following steps:

a) four markers located at the corners of the grid, whose 3D positions are known a priori, are manually identified in the first reconstructed frame to establish a geometric reference.
b) A planar homography transformation is computed to rectify the perspective deformation caused by the camera viewpoint. This transformation maps the original image into a new view in which the elliptical appearance of the markers, caused by perspective, is corrected into circular shapes, facilitating detection. The homography $H \in \mathbb{R}^{3\times3}$ is defined using the known world coordinates X_i and their corresponding image points x_i in homogeneous coordinates, such that ($\sim$ denotes equality up to a scale factor):

$$x_i \sim HX_i \quad \text{for } i = 1, \cdots, 4 \tag{1}$$

c) In each rectified image, candidate marker centres are detected through template matching, using normalized cross-correlation between a predefined patch and the image. Due to noise and illumination artifacts, spurious detections may occur. To robustly associate detected points with theoretical grid positions, the task is formulated as an assignment problem [62]. Given two sets A (nominal marker coordinates) and B (detected coordinates), and a cost function $C(a,b)$ defined as the Euclidean distance, the optimal bijection $\psi : A \rightarrow B$ minimizes:

$$\min_{\psi} \sum_{a\in A} C(a, \psi(a)) = \sum_{a\in A} \|a - \psi(a)\|^2 \tag{2}$$

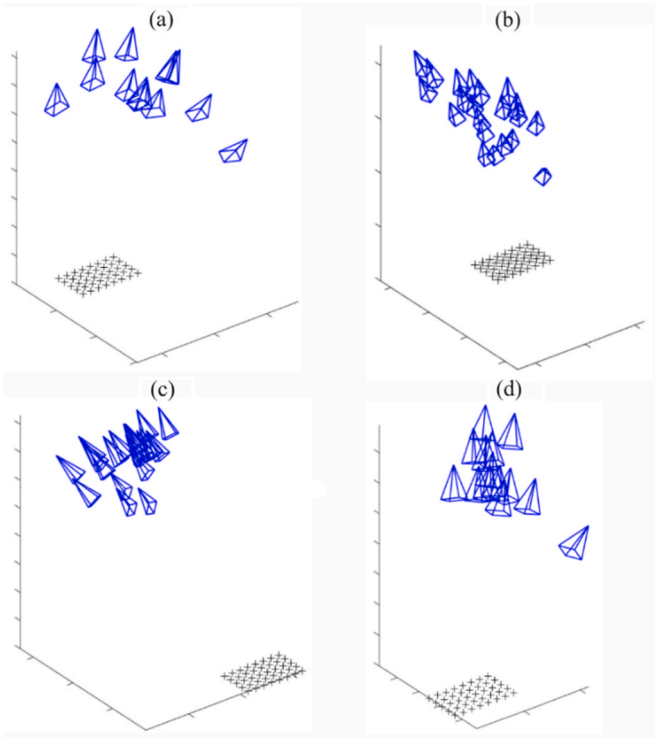


Fig. 4. Multiple view in each event camera to perform calibration, a) camera 1, b) camera 2, c) camera 3, d) camera 4.

Table 2
Internal parameters of each event camera whose calibration mean reprojection error is 0.37 pixels (Ground Sample Distance is given in [mm/pixel]).

Camera 1		**Camera 2**	
Focal Length u direction	$\alpha_{u,1} = -1241$	Focal Length u direction	$\alpha_{u,2} = -1395$
Focal Length v direction	$\alpha_{v,1} = -1244$	Focal Length v direction	$\alpha_{v,2} = -1403$
Distortion	$\kappa_{0,1} = -0,17$	Distortion	$\kappa_{0,2} = -0,31$
Principal Point coordinates	$u_{0,1} = 420,53$ $v_{0,1} = 236,87$	Principal Point coordinates	$u_{0,2} = 280,92$ $v_{0,2} = 263,86$
GSD	$min = 0,642$ $max = 1,13$	GSD	$min = 0,585$ $max = 1,26$
Camera 3		**Camera 4**	
Focal Length u direction	$\alpha_{u,3} = -1287$	Focal Length u direction	$\alpha_{u,4} = -1146$
Focal Length v direction	$\alpha_{v,3} = -1285$	Focal Length v direction	$\alpha_{v,4} = -1141$
Distortion	$\kappa_{0,3} = -0,2$	Distortion	$\kappa_{0,4} = -0,3$
Principal Point coordinates	$u_{0,1} = 307,22$ $v_{0,1} = 171,69$	Principal Point coordinates	$u_{0,1} = 350,28$ $v_{0,1} = 269,81$
GSD	$min = 0,891$ $max = 1,79$	GSD	$min = 0,685$ $max = 1,11$

d) To enhance precision, a subpixel refinement is applied to each detected marker position. Around the peak correlation point, a 1D parabolic fit is performed along both horizontal and vertical directions using the central point and its two neighbours. The vertices of the fitted parabolae provides the refined position.
e) e. The inverse homography H^{-1} is then applied to each marker coordinate, mapping the positions back to the original, unrectified image space. If necessary, the homography is updated in order to adapt to displacements in the following frames.

The proposed method identifies markers through a global, lattice-constrained assignment rather than purely local detection, ensuring robustness to weak, missing, or corrupted markers, which are naturally handled as absent observations in triangulation and bundle adjustment. Per-frame planar homography rectifies the grid to enable stable template matching and topology-aware assignment, improving robustness over standard circle-grid detectors, while not imposing any planarity constraint on the reconstructed 3D motion.

The output of this stage is thus the 2D coordinates in pixels of the centre of each dot j in the image plane of each camera i at each time instant t, which have been collected in the following vector:

$$m_j^i(t) = \left[u_j(t), v_j(t)\right] \tag{3}$$

For time-harmonic small-amplitude vibrations, the marker motion can be approximated by a sinusoidal model. Once the 2D trajectories have been extracted for each marker in each camera view, 3D reconstruction is performed using a classic triangulation procedure based on the following steps:

a. An initial estimate of each camera extrinsic parameters is computed via the Direct Linear Transform (DLT) method [59,63]. This method leverages the known 2D–3D correspondences from the first frame to estimate the camera projection matrix P, assuming a pinhole model:

$$x \sim PX \tag{4}$$

b. A partial bundle adjustment (BA) is then applied to refine the estimated camera exterior orientations using the DLT results and nominal grid coordinates as constraints.
c. A full bundle adjustment procedure [59] is finally executed to jointly optimize both the 3D marker positions $M_j(t) = \left[X_j(t), Y_j(t), Z_j(t)\right]^T$ and the camera extrinsic parameters, thus minimizing the reprojection error across all views and time steps.

It is important to note that, due to the nature of event cameras, which only generate data in response to scene changes, capturing a static reference frame is nontrivial. To overcome this limitation, a reference position for each marker is estimated by computing the average position over time, assuming symmetric harmonic motion about the equilibrium point. This strategy allows for a reliable estimation of the beam absolute deflection shapes without requiring a true rest image. This comprehensive approach, combining event-based vision, intensity frame reconstruction, subpixel tracking, and multi-view triangulation, enables high-fidelity vibration analysis with minimal data, an ideal configuration for dynamic structural monitoring.

3.4. Vibration reconstruction from triangulation

In the case of small-amplitude flexural vibrations of a cantilever beam, the transverse displacement of each marker point depicted on the structure can be readily estimated from the 3D coordinates obtained with the triangulation procedure. Specifically, for the j-th marker point, the transverse displacement component corresponds to the variation in its $Z_j(t)$ coordinate over time.

Assuming a time-harmonic excitation at a known circular frequency ω, the displacement $w\left(x_j, y_j, t\right)$ at each marker position $\left(x_j, y_j, t\right)$ can be written in complex exponential form.

$$w\left(x_j, y_j, t\right) \cong \mathscr{R}e\left\{\widehat{w}\left(x_j, y_j, \omega\right) e^{i\omega t + \varphi\left(\mathrm{x_j},\, \mathrm{y_j},\, \omega\right)}\right\} \tag{5}$$

Here, $\widehat{w}\left(x_j, y_j, \omega\right)$ and $\varphi\left(\mathrm{x_j},\, \mathrm{y_j},\, \omega\right)$ are the complex amplitude and phase of the displacement at each marker point. Since the cameras operate at a fixed frame rate of $1/T$, the amplitude and phase components are extracted from the discrete sequence of positions $Z_j(t_k)$ associated with the j-th marker centre, where $k = 0, \cdots, N-1$. For instance, selecting the number of samples N such that the recorded sequence spans an integer number of sinusoidal cycles, the Discrete Fourier Transform (DFT) can be exploited to accurately retrieve both the magnitude and phase of the oscillation. In practice, this is achieved using the Fast Fourier Transform (FFT) algorithm [64], which yields the spectral coefficients corresponding to the excitation frequency ω, thereby providing a precise estimation of the vibration parameters.

4. Experimental results

This section presents the experimental results obtained with the cantilever beam setup introduced in Section 2 by applying the event-based vibration measurement technique described in Section 3. The reconstructed 3D trajectories of the marker points are used to extract the first two flexural deflection shapes of the beam. To evaluate the accuracy of the proposed method the deflection shapes obtained from the event-cameras measurements, are compared with those acquired by the scanning laser Doppler vibrometer.

4.1. Trajectories of markers in each event camera

In this section, the measurements of the flexural vibration field of the cantilever beam at its first two resonance frequencies, (i.e. the flexural deflection shapes), are analysed in detail. As described in the previous sections, four synchronized event-based cameras (DVXplorer, by iniVation) were employed to record the displacements of the silver markers applied to the surface of the cantilever beam. The main technical specifications of the cameras, are summarized in Table 1. The two dynamic range values reported in the table are not associated with different camera settings or operating modes. Rather, the ~ 90 dB and ~ 110 dB values describe the sensor behaviour under different contrast thresholds and percentages of responding pixels. The camera does not switch between these values; instead, it continuously adapts to the scene dynamics, and its effective dynamic range depends on the actual illumination and contrast conditions. For our application, both values indicate that the sensor provides sufficient dynamic range to robustly detect the markers under varying lighting conditions. A grid of 10 × 3 markers were glued on the cantilever beam. However, the displacements of the first

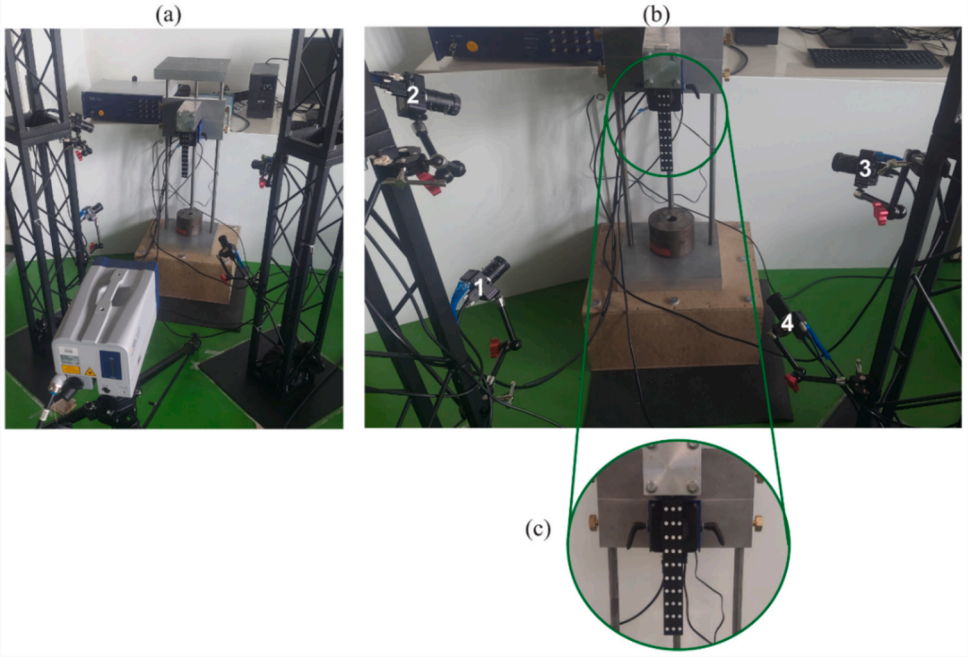


Fig. 5. Experimental setup a) cantilever beam with 4 event cameras and laser vibrometer, b) Event cameras IDs, c) magnification of the cantilever beam with markers.

row (the one closer to the clamped end) were too small to be detected at the first resonance frequency, while the displacements of the first two rows were too small to be detected at the second resonance frequency. Therefore, in the two experiments, the first and first two rows were discarded. The experimental setup is illustrated in Fig. 5a,b, where the four DVXplorer cameras are identified by ID numbers. Fig. 5c provides a magnified view of the cantilever beam, showing the grid of markers used for tracking purposes. Marker identification is formulated as a global, lattice-constrained assignment rather than a purely local detection problem, allowing the method to remain robust when some markers are weak, missing, or corrupted. Missing grid points are naturally handled as absent observations and are simply excluded from triangulation and bundle adjustment, without compromising the reconstruction of the remaining markers.

Before processing the experimental data, a dedicated test was carried out to ensure accurate synchronization among the four cameras. The DVXplorer units are hardware-synchronized through daisy-chained connections, with one camera acting as the master. Thanks to the built-in synchronization feature, the master camera controls the timestamp reset for all connected units, maintaining timing alignment with sub-microsecond accuracy. To validate this synchronization, a verification experiment was performed in which all four cameras observed the same portion of a diffusely reflective surface illuminated by a temporally modulated light source.

The source was driven by a chirp signal (Fig. 6a), designed to produce a known time-varying light pattern. This allowed to align the event histograms generated by the cameras using a cross-correlation procedure. Although camera timestamps are provided with microsecond resolution, the verification method using event histograms required a coarser temporal binning to ensure statistical reliability. A bin width of $100\mu s$ was chosen as a compromise, allowing alignment verification with a temporal resolution of $0.1ms$. However, as shown in Fig. 6b, the resulting event histograms revealed a non-negligible "exponential tail" following the main peak. Ideally, a sharp peak corresponding to the lighting impulse was expected, but instead, a decaying tail was observed. This effect is attributed to limitations in the cameras event-handling precision under high-density stimuli and internal detection thresholds [30,36], which artificially broaden the histogram and reduce the sharpness of correlation peaks. To mitigate this effect and enhance alignment precision, a digital filtering approach was adopted. Specifically, a first-order autoregressive (AR) filter of the form

$$y(n) = x(n) - \alpha x(n-1) \tag{6}$$

was applied to the histogram signal, where the parameter α (for example $\alpha = 0.975$) was identified via model fitting. This filter effectively suppresses exponential decays of the form α^n, thereby improving the temporal resolution of the signal. Fig. 6c illustrates the result of applying the filter to the signal in Fig. 6b. After filtering, the peak of the correlation function, refined using parabolic interpolation, was found at a time offset of approximately $0.13ms$, indicating the temporal shift required to best align the event data from four cameras. This experiment confirmed that the camera synchronization was accurate within tenths of milliseconds, which is very important to perform high-frequency acquisitions.

Once the clock synchronisation of the four event cameras were verified, the recordings of the vibrating cantilever beam begun. Two separate tests were conducted, on the cantilever beam vibrating at the first two flexural resonance frequencies. As shown in Fig. 2b, the flexural vibrations were generated by the electrodynamic shaker connected to the beam near its clamped end via a stinger. The first deflection shape was found around 32 Hz and the second one around 190 Hz. As described in Section 3.2, following the camera acquisitions of both the beam vibration at the first two resonances and the chessboard calibration sequences, intensity images were reconstructed using the E2VID neural network [47]. This step was essential for the marker tracking and subsequent triangulation described in the methodology section. The reconstruction of intensity images was performed using a sampling threshold of 1000 Hz for the first flexural deflection shape and 1600 Hz for the second flexural deflection shape. Although this value largely exceeds the minimum value required by the Nyquist criterion, it was intentionally selected to explore the temporal resolution limits achievable with the proposed approach. Beam-like structures typically exhibit a low modal overlap, which increases proportionally to the square root of the excitation frequency $\sqrt{\omega}$, as discussed in [65]. Consequently, the time-harmonic response at each resonance frequency is predominantly governed by the mode shape associated with the corresponding resonant mode. Furthermore, under the assumption of small-amplitude displacements, the structural response can be accurately approximated as linear, which allows a modal decomposition

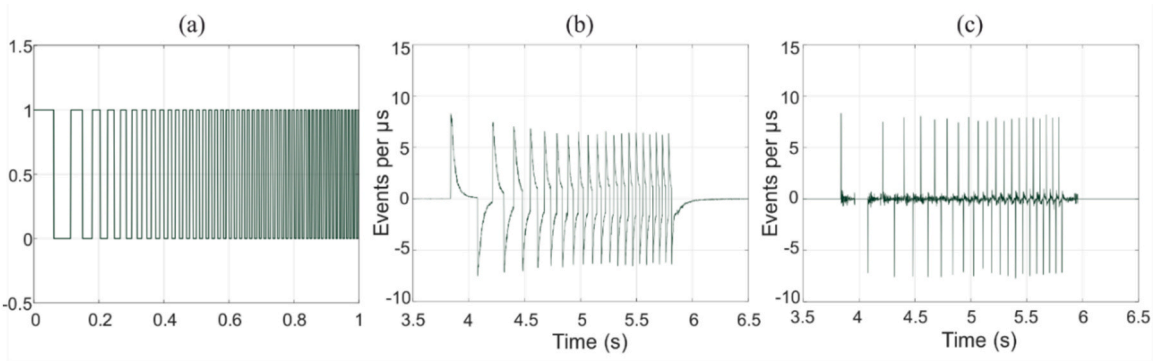


Fig. 6. a) Example of a 'square wave chirp' signal used in the assessment experiments, b) signal obtained from an event sequence with bin width T=0.1 ms, c) signal obtained by filtering the raw signal with a AR filter designed to remove exponential tails.

of the vibration field.

In each rectified frame, candidate marker centres were identified using template matching, implemented via normalized cross-correlation between the reconstructed image and a predefined template patch. However, due to the presence of noise and illumination artifacts, several false positives were observed, as illustrated in Fig. 7. To address this issue and robustly associate the detected points with the corresponding locations in the reference marker grid, the assignment algorithm described in Section 3.3 was employed.

Focusing on the acquisition taken at the first flexural resonance frequency, the tracking process yielded 2D trajectories for each of the 27 markers (grid of 9×3). Fig. 8 shows the tracked paths for each camera (denoted with the ID numbers reported in Fig. 5b), which are superimposed on a single reconstructed frame. It is clear that the displacements are very small, of the order of a few pixels. This observation is corroborated by the plots in Fig. 9, which display the temporal evolution of the u and v coordinates for the tracked points. The signals derived from the analysis of the events recorded by the four cameras exhibit sinusoidal time-histories consistent with the expected motion of the beam. Here, the thicker line highlights the trajectory of the reference marker (the one indicated in Fig. 2), which exhibits the largest oscillation amplitude. A maximum displacement of approximately 7 pixels is observed for the vertical component v, while the peak displacement reaches about 6 pixels in the horizontal component u. The different amplitudes of the displacement u, v components recorded by the four event cameras (a–d) result from their distinct orientations, that is their view-points.

Moving to the recordings taken for vibrations at the second flexural resonance frequency, the tracking process yielded again 2D trajectories for each of the 21 visible markers on a grid of 7×3 (one row of markers coincides with the nodal line of the second deflection shape of the cantilever beam, therefore the displacements are negligible and the row is not detected by the event cameras). Fig. 10 shows the tracked paths by each camera (denoted with ID numbers in Fig. 5b) which are superimposed on a single reconstructed frame. As anticipated in the methodology section, for the second flexural deflection shape an overlapping window of about 50% was employed to reconstruct frames using the E2VID neural network in order to accumulate the number of events necessary to have a sufficiently defined marker in each frame.

The plots in Fig. 11, display the temporal evolution of the u and v coordinates for the tracked points. Here the signals derived from the analysis of the events recorded by the four cameras exhibit sinusoidal trends consistent with the expected motion of the beam also for the second deflection shape here analysed. The thicker line highlights the trajectory of the reference marker (the one indicated in Fig. 2), which should exhibit the largest oscillation amplitude. In the vertical component v, a maximum displacement of approximately 1.5 pixels is observed, while in the horizontal component u, the peak displacement reaches about 1 pixels. The differences observed among the four event cameras (a–d) reflect their distinct orientations and, consequently, their view-points. Contrasting Figs. 9 and 11, it is evident that the displacements, as the frequency of the vibration rises, become smaller and challenging to be detected.

4.2. *Vibration reconstruction of the cantilever beam*

After reconstructing the trajectories of all 27 markers in the image planes of each camera, triangulation was carried out to obtain

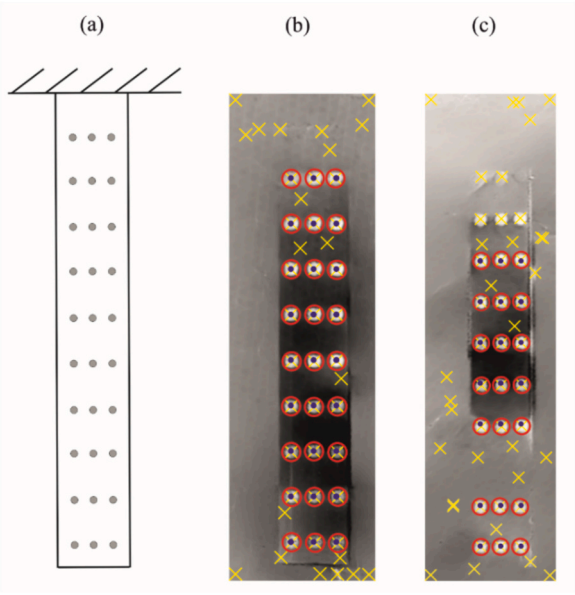


Fig. 7. a) Sketch of the cantilever beam, b) and c) rectified image in which circles are detected. The violet dots depict the nominal positions (set A), while yellow crosses represent spurious circles that have been detected by template matching but not matched (set B). The red circles are the remaining ones after the assignment problem. b) first flexural deflection shape, c) second flexural deflection shape. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

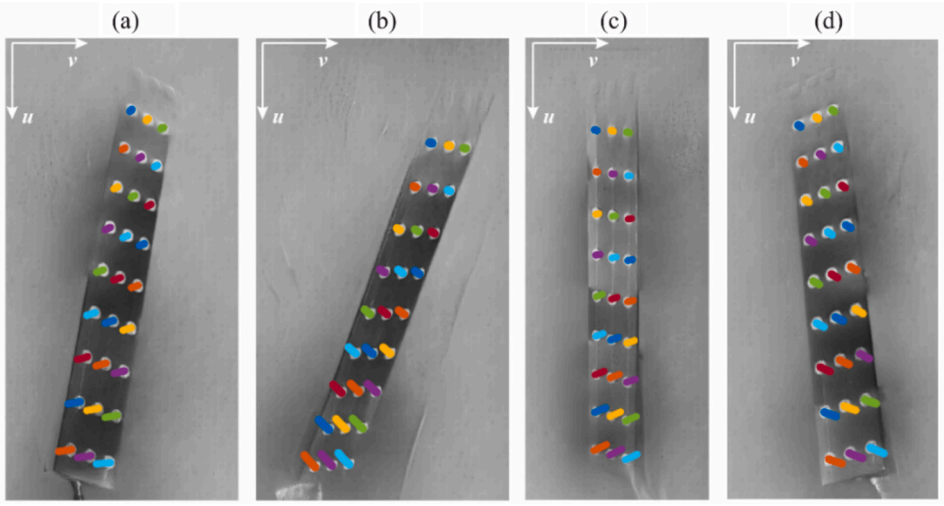


Fig. 8. Traces of markers in each camera with image coordinate system with v vertical and u horizontal components, a) camera 1, b) camera 2, c) camera 3, d) camera 4.

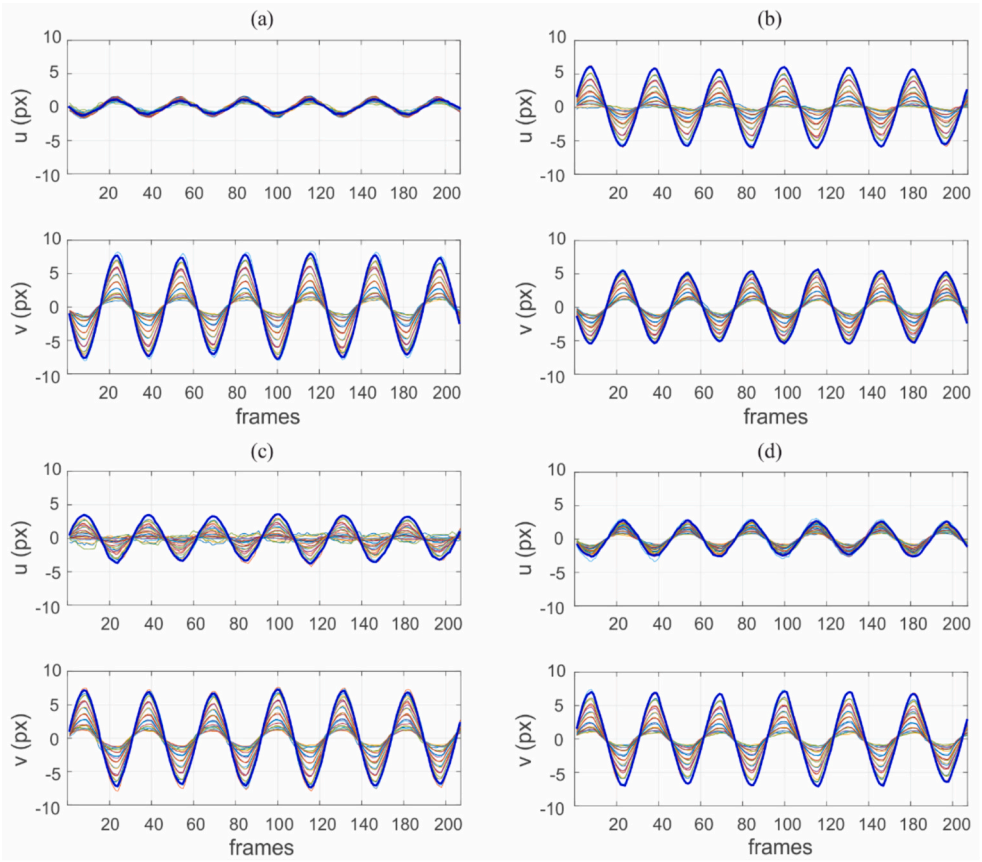


Fig. 9. Horizontal and vertical displacements of 27 markers in the image coordinate system. Thicker line for the reference point at the tip. a) camera 1, b) camera 2, c) camera 3, d) camera 4.

their 3D trajectories of the markers in the physical space coordinates. As illustrated in Fig. 12, the displacements in the X , Y, and Z directions exhibit sinusoidal trends with identical frequency and phase. The frequency is consistent with the applied harmonic excitation. The Z component captures the dominant transverse motion associated with the beam flexural response, while the X and Y

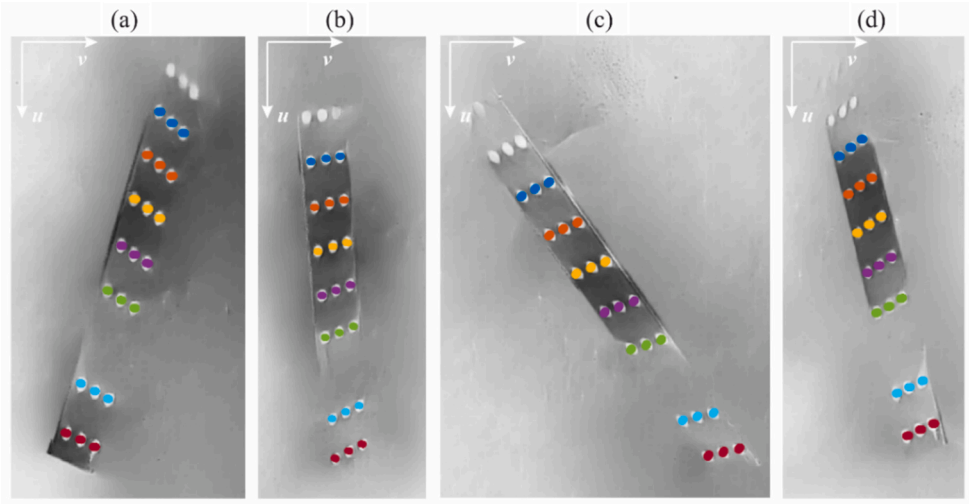


Fig. 10. Traces of markers in each camera with image coordinate system with v vertical and u horizontal components, a) camera 1, b) camera 2, c) camera 3, d) camera 4.

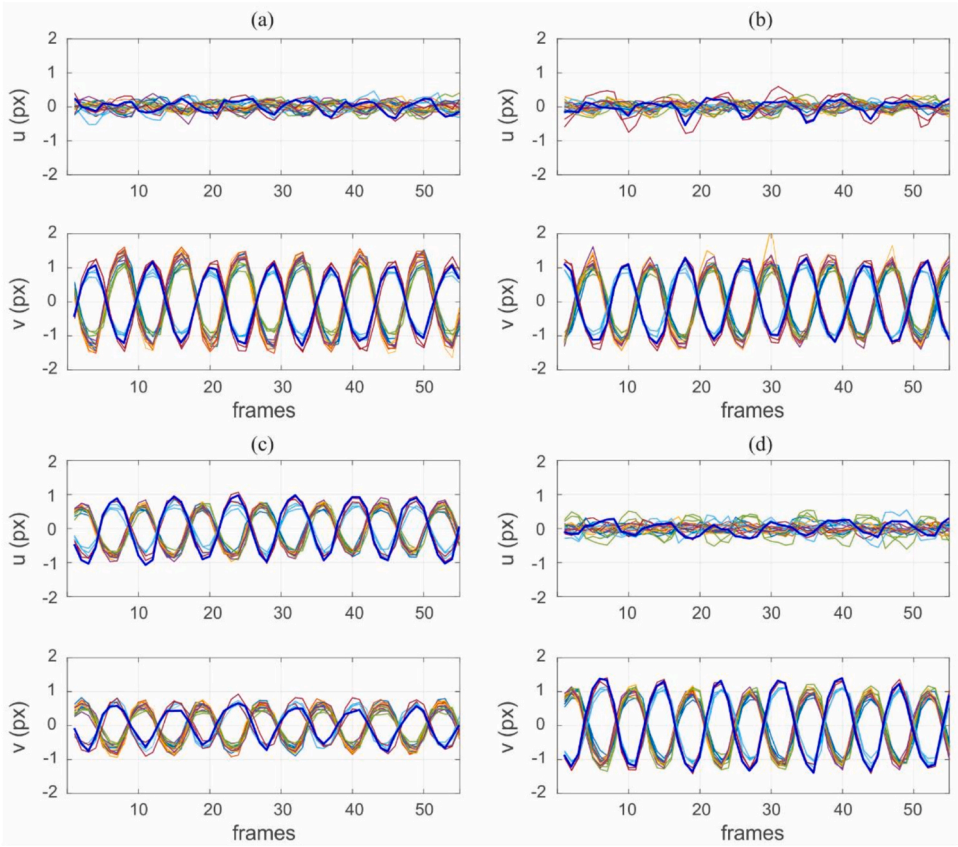


Fig. 11. Horizontal and vertical displacements of 21 markers in the image coordinate system. thicker line for the reference point at the tip. a) camera 1, b) camera 2, c) camera 3, d) camera 4.

components represent in-plane displacements. Although the latter are of much smaller amplitude, they still show a coherent, periodic time-history over time. Fig. 12 also highlights with a thicker curve the trajectory of the reference marker located at the tip of the cantilever beam as depicted in Fig. 2a. For the vibrations at the first flexural resonance frequency, at about 32 Hz, the application of the bundle adjustment algorithm led to an accurate reconstruction of the first deflection shape. In fact, the root-mean-square

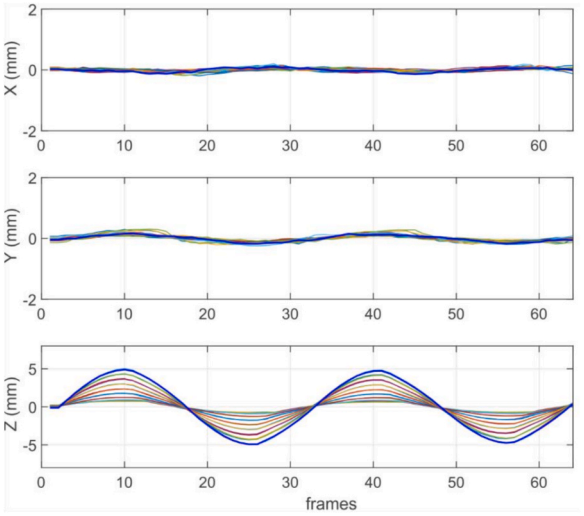


Fig. 12. X, Y, Z displacements for the first flexural deflection shape in millimetres for all 27 markers. Thicker line for the reference point at the tip.

reprojection error was reduced from an initial value of 0.5 pixels to a final value of 0.22 pixels.The bundle adjustment provided accurate 3D trajectories for all tracked markers.

Moving to the second vibrational deflection shape of the cantilever beam, which occurs around 190 Hz, the deflection shape of the beam shows a nodal line coincident with the position of the third row of markers starting from the tip. After reconstructing the trajectories of all 21 visible markers in the image sequences from each camera, the triangulation procedure was carried out to obtain their 3D coordinates. At a higher frequency, the amplitude of the displacements becomes smaller, and more challenging to be detected and reconstructed. As illustrated in Fig. 13, the displacements in the X , Y, and Z directions exhibit sinusoidal time-histories with identical frequency and opposite phase depending on the marker position with respect to the nodal line. In this case too, the frequency of the harmonic response coincides with that of the force excitation exerted by the shaker. The Z component captures the dominant transverse motion associated with the beam flexural response, while the X and Y components represent in-plane displacements. Although the latter are of much smaller amplitude, they still show a coherent, harmonic evolution over time. Fig. 13 also highlights with a thicker curve the trajectory of the reference marker located at the tip of the cantilever beam as depicted in Fig. 2a. In this second case, the application of the bundle adjustment algorithm led to a substantial improvement in the reconstruction accuracy: the root-mean-square reprojection error was reduced from an initial value of 0.66 pixels to a final value of 0.43 pixels. Again, the bundle adjustment provided time-resolved 3D trajectories for all tracked markers.

4.3. Comparison with laser vibrometer measurements

The accuracy of the measurements made with the four event cameras is assessed with respect to additional measurements taken with a Polytec PSV-500-A laser doppler vibrometer (LDV). More specifically, for both type of measurements, the transverse displacements at the grid of points (27 for the first flexural deflection shape and 21 for the second flexural deflection shape) are estimated as the average of the peak-to-peak values in the Z-direction observed over each vibration cycle; in this case, two full periods were analysed. To quantify the difference between the flexural deflection shape reconstructed from the event camera acquisitions and that measured by the laser scanner vibrometer, the following two error were calculated:

$$E_{RMS,w} = \sqrt{\frac{1}{N}\sum_{j}\left[w_c\left(x_j, y_j, t_k\right) - w_v\left(x_j, y_j, t_k\right)\right]^2} \tag{7}$$

$$E_{event-laser} = \frac{1}{3}\sum_{i=1}^{3} w_c(x_i, y_i, t_k) - w_v(x_i, y_i, t_k) \tag{8}$$

Here, N represents the number of measurement points, $w_c\left(x_j, y_j, t_k\right)$ and $w_v\left(x_j, y_j, t_k\right)$ are the transverse displacements derived respectively from the camera and the laser vibrometer measurements at position (x_j, y_j) and time t_k, where the i-index represents points on the same row. Overall, the root mean square deviation $E_{RMS,w}$ between the event-based measurements and those obtained via LDV was calculated to be equal to 0.17 mm for the first deflection shape (the row closest to the clamped end was discharged to have the same number of markers of the second deflection shape) and equal to 0.09 mm for the second deflection shape. Considering the first deflection shape around 32 Hz, Fig. 14a compares the flexural deflection shapes derived from the four event cameras and the laser vibrometer measurements. The black line depicts the first flexural deflection shape of a cantilever beam obtained from the expression

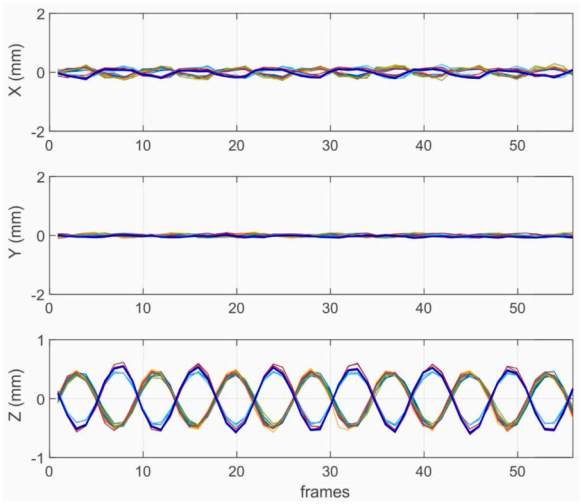


Fig. 13. X, Y, Z displacements for the second flexural deflection shape in millimetres for all 21 visible markers. Thicker line for the reference point at the tip.

reported in Appendix A and fitted in such a way to minimise the mean error with respect to the values measured with the laser (blue circles). Also, Fig. 14b presents histogram bars for the $E_{event-laser}$ error at each marker row. Moreover, the red lines show the positive and negative errors of the measurements with the laser vibrometer, which according to the datasheet is about 3% of the vibration amplitude. This error is also displaced in Fig. 14a at the marker points with vertical red bars. In general figure 14a shows a rather good agreement between the measurement taken with the event cameras and that made with the laser vibrometer. In fact, Fig. 14b shows that the errors between the event cameras and the laser vibrometer measurements are of the order of 0.02 to 0.4 mm for the deflection

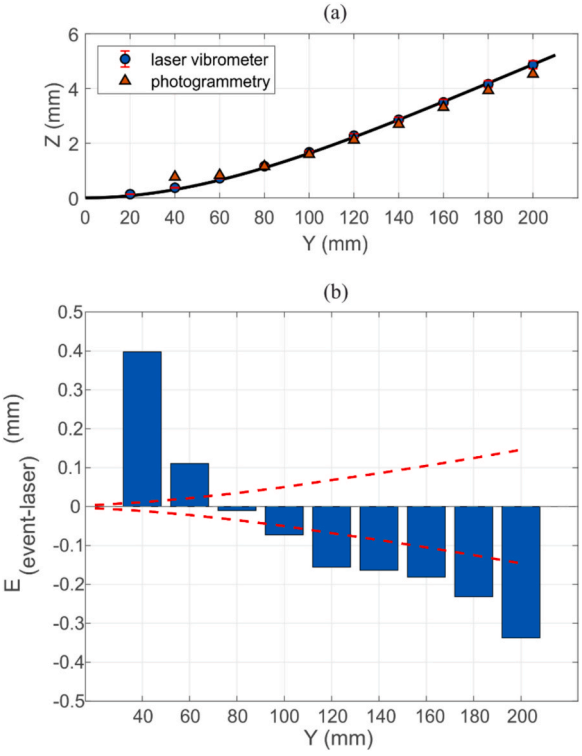


Fig. 14. Comparison of the average transverse displacements measured with event cameras and laser vibrometer for the first flexural deflection shape. a) Orange triangles for event cameras and blue circles for laser vibrometer. b) Difference between event and laser displacements (blue bars) and red line for the vibrometer measurement error. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

shape with tip amplitude of 5 mm. The event camera overestimates the deflection shape close to the beam clamping edge whereas it underestimates it close to the beam free end.

Fig. 15a presents the same type of results as in Fig. 14 for the second flexural deflection shape of the clamped beam. Here, Fig. 15a shows a slightly larger mismatch between the two measurements, although the errors depicted in Fig. 15b are comprised between 0.02 and 0.2 mm. This is because the amplitude of the second deflection shape is comparatively smaller than that of the first one. Indeed, the tip amplitude of the second deflection shape is about 0.6 mm against the 5 mm measured for the first flexural deflection shape. Overall, Figs. 14 and 15 indicate that the measurements performed with the event cameras provided a reasonably good estimate of the first two flexural deflection shapes. The deviations of the measurements from those performed with the laser vibrometer are probably the results of the low spatial resolution of the cameras rather than the proposed methodology to reconstruct the displacements from the flow of events generated by the four cameras. Neuromorphic vision sensors technology is relatively new and it is expected that in the forthcoming year it will progress towards high spatial resolutions such that vibration measurements can be performed with greater accuracy, comparable to that of laser vibrometer.

5. Conclusions

This paper has presented a new approach for the reconstruction of the flexural vibration field of a cantilever beam using photogrammetry applied to event cameras. Experimental validation was performed on a harmonically excited steel cantilever beam. The event-based displacement fields were compared to reference measurements acquired via a laser Doppler vibrometer, resulting in a residual root mean square error ($E_{RMS,w}$) of 0.17 mm for the first flexural deflection shape and equal to 0.09 mm for the second flexural deflection shape. Moreover, the application of bundle adjustment reduced the RMS reprojection error from 0.5 to 0.22 pixels for the first flexural deflection shape and from 0.66 to 0.43 pixels for the second flexural deflection shape, demonstrating the effectiveness of the optimization in enhancing 3D reconstruction accuracy.

Two additional considerations further highlight the value of the proposed framework. First, the experiments were conducted using event cameras with relatively low spatial resolution. It is notable that both the precision and quality of the results could be substantially improved through the adoption of higher-resolution event cameras. Second, although the experimental validation focused on the first and second flexural resonance frequency, the methodology is inherently capable of capturing responses at significantly higher frequencies. Given the high temporal resolution of event cameras, the principal limitation at high frequencies would be their spatial resolution, as the vibration amplitudes tend to decrease with increasing frequency, making displacement detection more challenging.

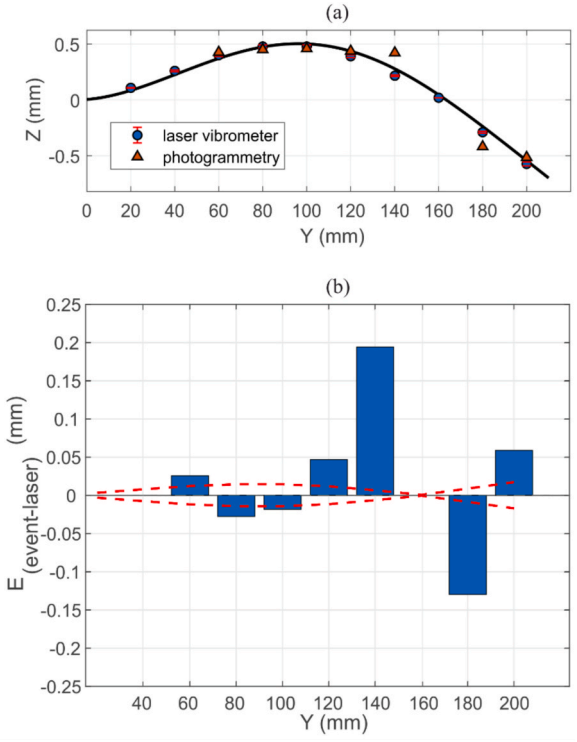


Fig. 15. Comparison of the average transverse displacements measured with event cameras and laser vibrometer for the second flexural deflection shape. a) orange triangles for event cameras and blue circles for laser vibrometer. b) Difference between event and laser displacements (blue bars) and red line for the vibrometer measurement error. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Overall, the findings demonstrate that event cameras, when coupled with an appropriate photogrammetric pipeline, represent a compelling alternative for high-speed vibration measurement. Their ability to provide accurate, high-temporal-resolution data without redundancy positions them as a powerful tool for future research in structural dynamics and optical metrology. Furthermore, the open-source nature of the tools developed ensures reproducibility and encourages continued innovation in event-based measurement systems.

CRediT authorship contribution statement

Sofia Baldini: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization. **Filippo Stazi:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization. **Riccardo Bernardini:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization. **Andrea Fusiello:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization. **Paolo Gardonio:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization. **Roberto Rinaldo:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the ESPERT project of the University of Udine for the funding provided.

Appendix A

The flexural mode shape used to plot the black lines in Fig. 14a and 14b was taken from Ref [66] that gives the following expression

$$\phi_n(x) = (\cosh(k_n x) - \cos(k_n x)) - \sigma_n(\sinh(k_n x) + \sin(k_n x)) \tag{A1}$$

Where

$$\sigma_n(x) = \frac{\sinh(k_n L) - \sin(k_n L)}{\cosh(k_n L) + \cos(k_n L)} \tag{A2}$$

Here k_n is the modal wavenumber, L is the length of the beam and n is the mode.

Data availability

Data will be made available on request.

References

[1] F.N. Catbas, Structural health monitoring: applications and data analysis, in: Structural Health Monitoring of Civil Infrastructure Systems, Elsevier, 2009: pp. 1–39. https://doi.org/10.1533/9781845696825.1.
[2] C. Zhang, A.A. Mousavi, S.F. Masri, G. Gholipour, K. Yan, X. Li, Vibration feature extraction using signal processing techniques for structural health monitoring: a review, Mech. Syst. Signal Process. 177 (2022) 109175, https://doi.org/10.1016/j.ymssp.2022.109175.
[3] M. Basseville, A. Benveniste, B. Gach-Devauchelle, M. Goursat, D. Bonnecase, P. Dorey, M. Prevosto, M. Olagnon, In situ damage monitoring in vibration mechanics: diagnostics and predictive maintenance, Mech. Syst. Signal Process. 7 (1993) 401–423, https://doi.org/10.1006/mssp.1993.1023.
[4] K. Feng, J.C. Ji, Q. Ni, M. Beer, A review of vibration-based gear wear monitoring and prediction techniques, Mech. Syst. Signal Process. 182 (2023) 109605, https://doi.org/10.1016/j.ymssp.2022.109605.
[5] Á.J. Molina-Viedma, E. López-Alba, L. Felipe-Sesé, F.A. Díaz, Operational deflection shape extraction from broadband events of an aircraft component using 3D-DIC in magnified images, Shock Vib. 2019 (2019) 4039862, https://doi.org/10.1155/2019/4039862.
[6] P. Poozesh, J. Baqersad, C. Niezrecki, P. Avitabile, E. Harvey, R. Yarala, Large-area photogrammetry based testing of wind turbine blades, Mech. Syst. Signal Process. 86 (2017) 98–115, https://doi.org/10.1016/j.ymssp.2016.07.021.
[7] S. Tashakori, A. Baghalian, M. Unal, H. Fekrmandi, D.M. Volkan y Senyürek, I.N. Tansel, Contact and non-contact approaches in load monitoring applications using surface response to excitation method, Measurement 89 (2016) 197–203, https://doi.org/10.1016/j.measurement.2016.04.013.
[8] P. Castellini, C. Santolini, Vibration measurements on blades of a naval propeller rotating in water with tracking laser vibrometer, Measurement 24 (1998) 43–54, https://doi.org/10.1016/S0263-2241(98)00044-X.
[9] J. Baqersad, P. Poozesh, C. Niezrecki, P. Avitabile, Photogrammetry and optical methods in structural dynamics – a review, Mech. Syst. Signal Process. 86 (2017) 17–34, https://doi.org/10.1016/j.ymssp.2016.02.011.
[10] A.M. Wahbeh, J.P. Caffrey, S.F. Masri, A vision-based approach for the direct measurement of displacements in vibrating systems, Smart Mater. Struct. 12 (2003) 785–794.

[11] T.G. Ryall, C.S. Fraser, Determination of structural modes of vibration using digital photogrammetry, J. Aircraft 39 (2002) 114–119.
[12] S.W. Park, H.S. Park, J.H. Kim, H. Adeli, 3D displacement measurement model for health monitoring of structures using a motion capture system, Measurement 59 (2015) 352–362.
[13] M.N. Helfrick, C. Niezrecki, P. Avitabile, T. Schmidt, 3D digital image correlation methods for full-field vibration measurement, Mech. Syst. Signal Pr. 25 (2011) 917–927.
[14] R.S. Pappa, J.T. Black, J.R. Blandino, T.W. Jones, P.M. Danehy, A.A. Dorrington, Dot-projection photogrammetry and videogrammetry of gossamer space structures, J. Spacecr. Rockets 40 (2003) 858–867, https://doi.org/10.2514/2.7047.
[15] D.T. Bartilson, K.T. Wieghaus, S. Hurlebaus, Target-less computer vision for traffic signal structure vibration studies, Mech. Syst. Signal Pr. 60–61 (2015) 571–582.
[16] B. Ferrer, P. Acevedo, J. Espinosa, D. Mas, Targetless image-based method for measuring displacements and strains on concrete surfaces with a consumer camera, Constr. Build. Mater. 75 (2015) 213–219, https://doi.org/10.1016/j.conbuildmat.2014.11.019.
[17] J. Javh, J. Slavič, M. Boltežar, The subpixel resolution of optical-flow-based modal analysis, Mech. Syst. Signal Process. 88 (2017) 89–99, https://doi.org/10.1016/j.ymssp.2016.11.009.
[18] B.A. Furman, B.D. Hill, J.R. Rigby, J.M. Wagner, R.B. Berke, Sensor synchronized DIC: a robust approach to linear and nonlinear modal analysis using low frame rate cameras, J. Sound Vib. 584 (2024) 118478, https://doi.org/10.1016/j.jsv.2024.118478.
[19] L. Yu, B. Pan, Single-camera high-speed stereo-digital image correlation for full-field vibration measurement, Mech. Syst. Signal Pr. 94 (2017) 374–383.
[20] D. Gorjup, J. Slavič, M. Boltežar, Frequency domain triangulation for full-field 3D operating-deflection-shape identification, Mech. Syst. Signal Process. 133 (2019) 106287, https://doi.org/10.1016/j.ymssp.2019.106287.
[21] S. Baldini, G. Guernieri, D. Gorjup, P. Gardonio, J. Slavič, R. Rinaldo, 3D sound radiation reconstruction from camera measurements, Mech. Syst. Signal Process. 227 (2025), https://doi.org/10.1016/j.ymssp.2025.112400.
[22] D. Gorjup, J. Slaviu, A. Babnik, M. Boltezar, Still-camera multiview spectral optical flow imaging for 3D operating-deflection-shape identification, Mech. Syst. Signal Pr. 152 107456 (2021).
[23] R. Del Sal, L. Dal Bo, E. Turco, A. Fusiello, A. Zanarini, R. Rinaldo, P. Gardonio, Structural vibration measurement with multiple synchronous cameras, Mech. Syst. Signal Pr. 157 (2021) 107742.
[24] P. Neri, A. Paoli, A.V. Razionale, C. Santus, Low-speed cameras system for 3D-DIC vibration measurements in the kHz range, Mech. Syst. Signal Process. 162 (2022) 108040, https://doi.org/10.1016/j.ymssp.2021.108040.
[25] Y. Kato, S. Watahiki, Vibration mode identification method for structures using image correlation and compressed sensing, Mech. Syst. Signal Process. 199 (2023) 110495, https://doi.org/10.1016/j.ymssp.2023.110495.
[26] Y. Wang, F.S. Egner, T. Willems, M. Kirchner, W. Desmet, Camera-based experimental modal analysis with impact excitation: reaching high frequencies thanks to one accelerometer and random sampling in time, Mech. Syst. Signal Process. 170 (2022) 108879, https://doi.org/10.1016/j.ymssp.2022.108879.
[27] D. Mastrodicasa, E. Di Lorenzo, S. Manzato, B. Peeters, P. Guillaume, 3D-DIC full field experimental modal analysis of a demo airplane by using low-speed cameras and a reconstruction approach, Mech. Syst. Signal Process. 227 (2025) 112387, https://doi.org/10.1016/j.ymssp.2025.112387.
[28] R. Benosman, C. Clercq, X. Lagorce, S.-H. Ieng, C. Bartolozzi, Event-based visual flow, IEEE Trans. Neural Netw. Learn. Syst. 25 (2014) 407–417.
[29] T. Delbruck, M. Lang, Robotic goalie with 3 ms reaction time at 4% CPU load using event-based dynamic vision sensor, Front. Neurosci. (2013).
[30] P. Purohit, R. Manohar, Field-programmable encoding for address-event representation, Front. Neurosci. 16 (2022).
[31] S. Baldini, Bernardini R., Fusiello A., Gardonio P., Rinaldo R., Vibration measurement with event cameras, in: Proceedings of ISMA2024, Katholieke Universiteit Leuven, Leuven, 2024.
[32] S. Baldini, R. Bernardini, A. Fusiello, P. Gardonio, R. Rinaldo, Measuring vibrations with event cameras, in: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Brescia, 2024. https://doi.org/https://doi.org/10.5194/isprs-archives-XLVIII-2-W7-2024-9-2024.
[33] A. Amir, B. Taba, D. Berg, T. Melano, J. McKinstry, C. Di Nolfo, T. Nayak, A. Andreopoulos, G. Garreau, M. Mendoza, J. Kusnitz, M. Debole, S. Esser, T. Delbruck, M. Flickner, D. Modha, A low power, fully event-based gesture recognition system, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, pp. 7388–7397, https://doi.org/10.1109/CVPR.2017.781.
[34] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A.J. Davison, J. Conradt, K. Daniilidis, D. Scaramuzza, Event-based vision: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 44 (2022) 154–180, https://doi.org/10.1109/TPAMI.2020.3008413.
[35] J. Conradt, M. Cook, R. Berner, P. Lichtsteiner, R.J. Douglas, T. Delbruck, A pencil balancing robot using a pair of AER dynamic vision sensors, in: IEEE Int. Symp. Circuits Syst. (ISCAS), 2009, pp. 781–784.
[36] A.G. iniVation, Understanding the performance of neuromorphic event-based vision sensors, Tech. Rep. (2020).
[37] 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 (2018), https://doi.org/10.1061/(ASCE)EM.1943-7889.0001449.
[38] Z. Lai, I. Alzugaray, M. Chli, E. Chatzi, Full-field structural monitoring using event cameras and physics-informed sparse identification, Mech. Syst. Signal Process. 145 (2020) 106905, https://doi.org/10.1016/j.ymssp.2020.106905.
[39] C. Shi, N. Song, B. Wei, Y. Li, Y. Zhang, W. Li, J. Jin, Event-based vibration frequency measurement with laser-assisted illumination based on mixture gaussian distribution, IEEE Trans. Instrum. Meas. 72 (2023) 1–13, https://doi.org/10.1109/TIM.2023.3301911.
[40] M. Zhao, X. Shen, F. Jiang, Research on mechanical vibration measurement method based on event camera, in: 2023 3rd International Conference on Energy Engineering and Power Systems (EEPS), IEEE, 2023, pp. 528–532, https://doi.org/10.1109/EEPS58791.2023.10257083.
[41] M. Cook, L. Gugelmann, F. Jug, C. Krautz, A. Steger, Interacting Maps for Fast Visual Interpretation, in: Int. Joint Conf. Neural Netw. (IJCNN), 2011: pp. 770–776.
[42] H. Kim, S. Leutenegger, A.J. Davison, Real-time 3D Reconstruction and 6-DoF Tracking with an Event Camera, in: Eur. Conf. Comput. Vis. (ECCV), 2016: pp. 349–364.
[43] G. Gallego, J.E.A. Lund, E. Mueggler, H. Rebecq, T. Delbruck, D. Scaramuzza, Event-based, 6-DOF camera tracking from photometric depth maps, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2018) 2402–2412.
[44] K. Chaney, A. Panagopoulou, C. Lee, K. Roy, K. Daniilidis, Self-supervised optical flow with spiking neural networks and event based cameras, in: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2021, pp. 5892–5899, https://doi.org/10.1109/IROS51168.2021.9635975.
[45] A. Zhu, L. Yuan, K. Chaney, K. Daniilidis, EV-FlowNet: Self-supervised optical flow estimation for event-based cameras, in: Robotics: Science and Systems XIV, Robotics: Science and Systems Foundation, 2018. https://doi.org/10.15607/RSS.2018.XIV.062.
[46] H. Rebecq, R. Ranftl, V. Koltun, D. Scaramuzza, High speed and high dynamic range video with an event camera, IEEE Trans. Pattern Anal. Mach. Intell. 43 (2019) 1964–1980. https://api.semanticscholar.org/CorpusID:189998802.
[47] H. Rebecq, R. Ranftl, V. Koltun, D. Scaramuzza, Events-to-video: bringing modern computer vision to event cameras, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2019, pp. 3852–3861, https://doi.org/10.1109/CVPR.2019.00398.
[48] P. Castellini, E.P. Tomasini, Image-based tracking laser Doppler vibrometer, Rev. Sci. Instrum. 75 (2004) 222–232, https://doi.org/10.1063/1.1630859.
[49] IniVation, https://docs.inivation.com/hardware/current-products/dvxplorer.html, (n.d.).
[50] A. Fusiello, Computer Vision Toolkit for Matlab, (n.d.).
[51] Andrea Fusiello, Calibration Toolkit for Matlab, (n.d.).
[52] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: 2015: pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
[53] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W.-C. Woo, Convolutional LSTM network: a machine learning approach for precipitation nowcasting, in: Adv Neural Inf Process Syst, 2015.
[54] A. Fusiello, Computer vision: three-dimensional reconstruction techniques, Springer, Cham, 2024.

[55] E. Mueggler, B. Huber, D. Scaramuzza, Event-based, 6-DOF pose tracking for high-speed maneuvers, in: IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS), 2014: pp. 2761–2768.
[56] X. Lagorce, G. Orchard, F. Gallupi, B.E. Shi, R. Benosman, HOTS: a hierarchy of event-based time-surfaces for pattern recognition, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 1346–1359.
[57] A. Sironi, M. Brambilla, N. Bourdis, X. Lagorce, R. Benosman, HATS: histograms of averaged time surfaces for robust event-based object classification, in: IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2018, pp. 1731–1740.
[58] Y. Zhou, G. Gallego, H. Rebecq, L. Kneip, H. Li, D. Scaramuzza, Semi-dense 3D reconstruction with a stereo event camera, in: Eur. Conf. Comput. Vis. (ECCV), 2018: pp. 242–258.
[59] R. Hartley, A. Zisserman, Multiple view geometry in computer vision, 2nd ed., Cambridge University Press, Cambridge, 2003.
[60] P.F. Sturm, S.J. Maybank, On plane-based camera calibration: a general algorithm, singularities, applications, in: Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149), 1999: pp. 432–437. https://doi.org/10.1109/CVPR.1999.786974.
[61] Z. Zhang, A flexible new technique for camera calibration, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 1330–1334, https://doi.org/10.1109/34.888718.
[62] H.W. Kuhn, The Hungarian method for the assignment problem, Naval Res. Logist. Quart. 2 (1955) 83–97.
[63] I.E. Sutherland, Three-dimensional data input by tablet, Proc. IEEE 62 (1974) 453–461, https://doi.org/10.1109/PROC.1974.9449.
[64] A.V. Oppenheim, R.W. Schafer, Discrete-time signal processing, 3rd ed., Prentice Hall Press, USA, 2009.
[65] P. Gardonio, E. Turco, Tuning of vibration absorbers and Helmholtz resonators based on modal density/overlap parameters of distributed mechanical and acoustic systems, J. Sound Vib. 451 (2019) 32–70, https://doi.org/10.1016/j.jsv.2019.03.015.
[66] P. Gardonio, M.J. Brennman, Mobility and impedance methods in structural dynamics, in: F. Fahy, J. Walker (Eds.), Advanced Applications in Acoustics, Noise and Vibration, 1st ed., CRC Press, London, 2004.