Measuring Vibrations with Event Cameras

Sofia Baldini, Riccardo Bernardini, Andrea Fusiello, Paolo Gardonio, Roberto Rinaldo

DPIA, Università degli Studi di Udine, Udine, Italy - name.surname@uniud.it

Keywords: Event cameras, Structural vibrations, Vibrometer.

Abstract

This paper presents a study of the use of synchronous event cameras to measure mechanical vibrations. The proposed methodology is based on two-dimensional point tracking and triangulation and achieves a precision comparable to that obtained with an expensive laser Doppler vibrometer.

1. Introduction

A considerable body of research has been conducted on the subject of vibration measurement with optical cameras (Baqersad et al., 2017). This research has encompassed a variety of photogrammetry approaches, with a primary focus on 2D-Point-Tracking (2DPT) (Wahbeh et al., 2003, Ryall and Fraser, 2002). In addition, studies have been conducted on 3D-Point-Tracking (3DPT) (Park et al., 2015), 3D-Digital-Image-Correlation (3D-DIC) (Helfrick et al., 2011, Pappa et al., 2003), and targetless approaches (Bartilson et al., 2015). The initial studies used single camera arrangements, with the camera oriented at a grazing angle relative to the surface of the flexural vibration field. Subsequently, vibration measurements with stereo-camera setups employing triangulation techniques (stereo-DIC) were also utilized (Beberniss and Ehrhardt, 2017, Helfrick et al., 2011, Reu et al., 2017). In more recent times, investigations have been conducted into the use of stereo acquisitions based on setups with moving single cameras or multiple mirrors focusing multi-images on single cameras. These have been explored in conjunction with the use of various techniques, as evidenced by (Gorjup et al., 2019, Barone et al., 2019).

In general, structural vibration measurements cover a rather wide frequency range, which may extend up to a few kHz. To ensure the accuracy of the proposed approach, it is necessary to employ cameras with a high frame rate.

Using *event cameras* (Conradt et al., 2009, iniVation, 2020) provides a low-cost, full-frame solution for vibration measurement, unlike expensive laser devices (Castellini and Tomasini, 2004, Ismail et al., 2019) that track single point, or even more expensive high-speed conventional cameras.

Event cameras, also known as neuromorphic or dynamic vision sensors, operate differently from traditional frame-based cameras. They asynchronously detect changes in the scene with high temporal resolution, instead of capturing images at fixed time intervals. An event is generated when there is a significant change in brightness at a pixel. It typically convey information about the pixel location, the sign of the brightness change (increase or decrease), and the precise timing of the event occurrence, with latency in the order of microseconds.

Event cameras are robust in varying lighting conditions and find applications where high temporal resolution and low latency are crucial, such as robotics, high-speed motion analysis, and autonomous vehicles.

Although event cameras theoretically encode the visual signal within their event streams, it is impossible to apply existing computer vision techniques to these data directly. Consequently, a variety of algorithms have been developed to effectively utilise event data, either by processing the event stream directly (Conradt et al., 2009, Cook et al., 2011, Benosman et al., 2014, Kim et al., 2016, Gallego et al., 2018) or by generating intermediate intensity representations from event data (Lagorce et al., 2017, Sironi et al., 2018, Zhu et al., 2018, Zhou et al., 2018). The former category of approaches presents inherent challenges and is not as extensively researched as its counterpart for intensity images. However, reconstructing an intensity image from event data presents practical challenges because of its inherently ill-posed nature.

In (Rebecq et al., 2019), the authors introduce a novel recurrent network architecture specifically designed to reconstruct videos from event streams. They utilise a perceptual loss function to encourage reconstructions to conform to the statistical properties observed in natural images. The authors demonstrate the efficacy of their approach by showcasing the network's ability to synthesise high-framerate videos, surpassing 5,000 frames per second, capturing high-speed phenomena with remarkable accuracy. Furthermore, they demonstrate that conventional computer vision algorithms can be seamlessly applied to the reconstruction (Muglikar et al., 2021), consistently outperforming algorithms specifically tailored for event data.

In this paper, we will show how event cameras applied to the measurement of mechanical vibrations can achieve results comparable to those obtained with an expensive laser Doppler vibrometer. The study builds on (Del Sal et al., 2021), which is focused on measuring the flexural vibrations of a cantilever beam by triangulation of optical measurements taken with conventional cameras.

2. Method

2.1 From events to intensity images

Intensity images are generated from the event stream using the E2VID (Rebecq et al., 2019) pre-trained neural network. E2VID is a recurrent, fully convolutional model inspired by UNet (Ronneberger et al., 2015). It has a head layer, several encoder and decoder layers, and a final prediction layer, which performs a depth convolution followed by a sigmoid to generate an image prediction. Skip links connect the encoder and decoder layers. The architecture includes downsampling and ConvLSTM (Shi et al., 2015) in the encoders and bilinear up-

sampling in the decoders, with ReLU activation and batch normalisation throughout, except in the final layer.

E2VID can operate in two distinct modes: constant or variable frame rate. In the constant frame rate mode, a video is generated with specific time stamps for each frame (this feature is available in the E2CALIB (Muglikar et al., 2021) distribution), ensuring synchronization among the cameras. In contrast, in the variable frame rate mode, frames are produced based on the occurrence of a sufficient number of events. This mode ensures that intensity images maintain their quality regardless of the amount of motion present. Consequently, this mode is particularly useful for independently calibrating interior parameters. In contrast, the first mode is suitable for multicamera triangulation scenarios. The selection of the frame rate is crucial in relation to the events generated. A frame rate that is too low would result in a significant number of events being sent to E2VID, which could be compensated for by adjusting the "upsampling" parameter. Conversely, a frame rate that is too high would result in a deficiency of events being fed to E2VID, which would result in a loss of detail and a dim image. This sets a practical limit to the highest achievable frame rate, which also depends on the magnitude of the image motion.

2.2 Camera calibration

Cameras are characterized by two sets of parameters, usually called interior and exterior. The former includes the focal length and other parameters that does not change when the camera moves. The latter describes the position and angular attitude of the camera, collectively termed *exterior orientation.* The interior parameters, including radial distortion, of each camera have to be calibrated separately in advance. The exterior orientation of the cameras are recovered during the experiment, as part of the Bundle Adjustment (BA), which optimizes a cost function with respect to the exterior orientation of the cameras and the coordinates of 3D points.

2.3 Tracking

The surface of the vibrating object – let us assume it is a beam, in the following – has been marked with white circles (or "dots" as they will henceforth be referred to) arranged on a grid with dimensions $N \times M$, with the centers of the dots representing the grid points. As the beam is not a perfectly planar surface, and the positioning of the dots has been carried out manually, the nominal coordinates inferred from assuming a planar surface and a regular grid does not correspond to the true ones. It is therefore understood that the coordinates of the grid points are only approximately known from their nominal value, even when the beam is at rest.

The first objective of the proposed methodology is to detect and track the 2D pixel coordinates $(u_j^i(t),\ v_j^i(t))$ of the centers of the NM dots in the sequence of images recorded by the K cameras, where $j = 1 \dots NM$ indexes the dot and $i = 1 \dots K$ runs over the cameras. To this end, the following steps were implemented for the recordings of each camera.

In general, the perspective image of a circle is an ellipse. Therefore, to simplify the dot detection task, a rectification is performed, which removes the perspective effects on the images acquired by the cameras. To determine the homography H_i that rectifies each of the K images, the user must manually identify four points of known coordinates in the first frame of the video sequence. In this study, the center positions of the dots at the four corners of the grid were used.

In the rectified image, the centers of the dots were identified by template matching, which involved taking the maxima of the normalized cross correlation of the image with a small patch that represents a template of the bright dot on a dark background. The output of this stage were the 2D pixel coordinates of the dot centers in the image planes of the cameras (please refer to Fig. 1), denoted by $u_j^i(t)$ and $v_j^i(t)$.

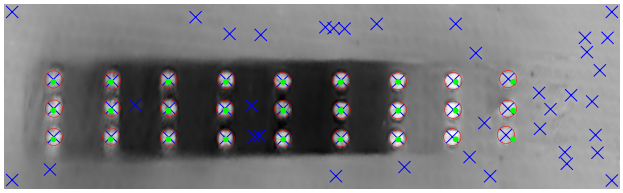


Figure 1. The rectified image in which circles are detected. The green dots depicts the nominal positions (set A), while blue crosses represent spurious circles that have been detected by template matching but not matched (set B). The red circles are the remaining ones after the assignment problem (set $\phi(A)$).

We should be aware that some outliers might be caused by clutter in the images which can manifest as bright spots. To address this issue, an assignment problem was solved, which is formally defined as follows. Given two sets A and B of equal size and a function $\mathcal{C}(a, b)$ that measures the cost of matching a with b, the goal is to find the bijection ϕ that minimises:

$$\sum_{a \in A} \mathcal{C}(a, \phi(a)). \tag{1}$$

The cost is the Euclidean distance. Set A contains the nominal positions of the dots, whereas set B contains the coordinates of the dots detected in the images. This is the unbalanced version of the assignment problem, as the set B can be larger than set A. The Kuhn–Munkres algorithm (Kuhn, 1955), was used to solve this problem. Only those points in the set $\phi(A)$ were retained, while the remainder were discarded.

In order to enhance precision, a subpixel refinement of the center of the dots was conducted by fitting a parabola to the point of maximum correlation and to its two neighboring points along the horizontal and vertical axes, and taking its vertex as the new position, thereby obtaining a fractional part.

Finally, since the coordinates of the detected points are defined in a transformed image space, the inverse of H_i was applied to bring them back to the original image space.

In order to update the homography H_i, image space positions are used in conjunction with the nominal grid-point coordinates. This process, starting with rectification, is then repeated for the subsequent frame.

In summary, dots tracking follows the following steps:

1. Compute the perspective rectification homography for the beam (user input needed only in first frame)
2. Apply the homography to the current frame
3. Extract tentative grid-points coordinates by template matching
4. Match coordinates of tentative grid-points with nominal ones by solving an assignment problem (Kuhn, 1955)
5. Sub-pixel refinement of coordinates
6. Map back to original image-space coordinates
7. Update the homography
8. Go to step 2.

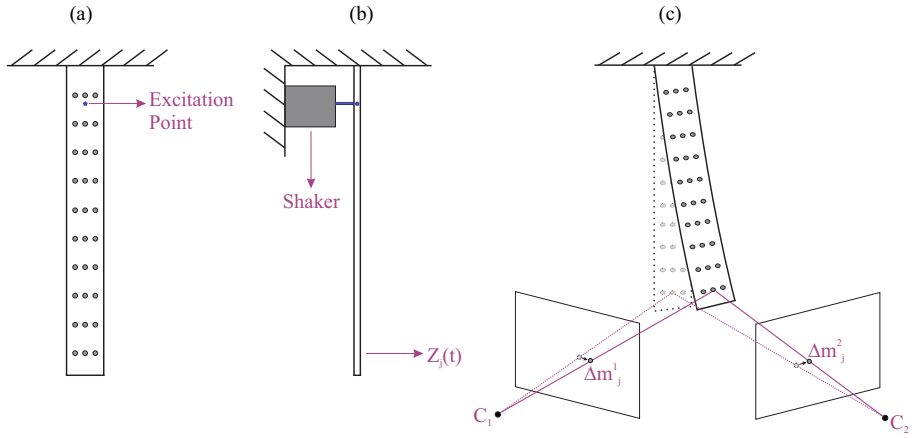


Figure 2. Experimental setup based on triangulation (c). The position od the shaker is shown in (a) and (b)

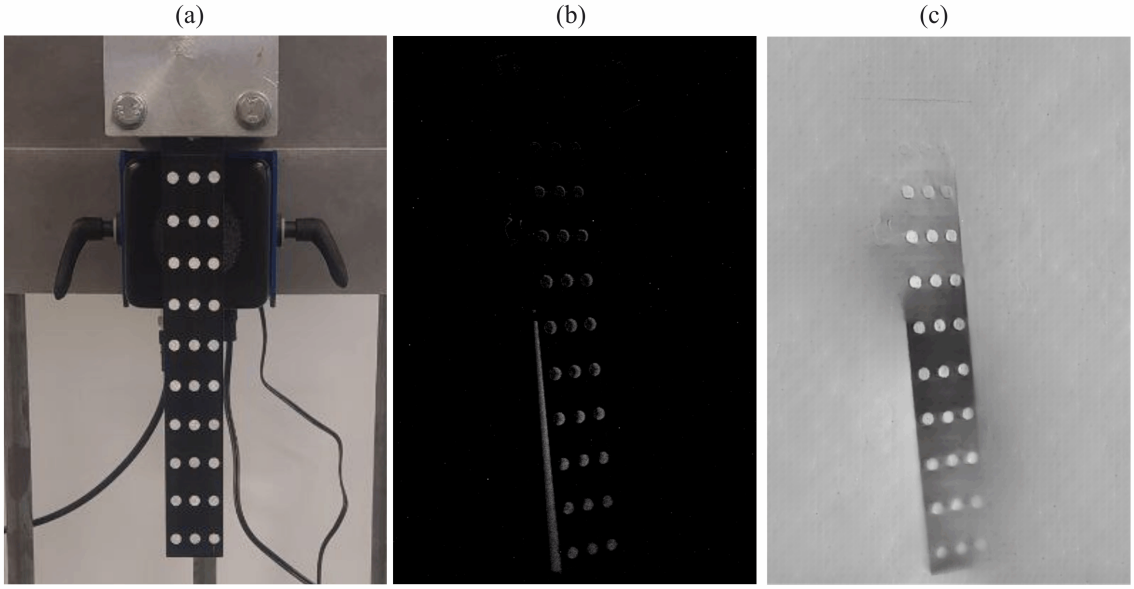


Figure 3. (a) An intensity image of the cantilever beam (taken with a traditional camera), (b) a frame generated from the DVXplorer by accumulation of events and (c) the same frame reconstructed by E2VID. Images are rotated by 90 degrees

It is understood that if the homography does not change (e.g. in the case of a vibrating plate where the frame is fixed), step 2 can be skipped.

The output of this stage are thus the 2D coordinates in pixels of the centers of the MN dots in the image plane of each camera i at each time instant t, which have been collected in the following vector $\mathbf{m}^i_j(t) = [u_j(t),\ v_j(t)]$.

2.4 Triangulation

Having identified the 2D pixel coordinates of the center positions of the dots in the sequence of images for each camera, the next step was to recover by triangulation the 3D coordinates $\mathbf{M}_j(t) = [X_j(t),\ Y_j(t),\ Z_j(t)]$ of the center of the dots. For a given time instant t, the relationship between the 3D coordinates of the dots centers $\mathbf{M}_j(t)$ and the 2D coordinates of the dots centers in the i-th image plane $\mathbf{m}^i_j(t)$ is given by the perspective projection equation:

$$\mathbf{m}^i_j(t) = f(\mathbf{M}_j(t), \mathbf{g}_i), \tag{2}$$

where $\mathbf{g}_i$ is a vector containing the exterior parameters of i-th camera. Bundle Adjustment (BA) simultaneously recovers the exterior orientation of the K cameras and the coordinates of the centers of the dots (at a given time) by solving a non-linear least squares problem, which consists in the minimisation of the cost function

$$\sum_{i,j} ||\mathbf{m}^i_j(t) - f(\mathbf{M}_j(t), \mathbf{g}_i)||^2_2 \tag{3}$$

with respect to $\mathbf{M}_j(t)$ and $\mathbf{g}_i$. The BA can be also used to compute only one set of unknowns when the other is given, and when this is the case, it will be referred to as a *partial* BA. The minimisation of Eq. (3) can be done via iterative techniques, such as the Levenberg-Marquardt algorithm, which should be initialised close to the optimum.

It is assumed that a frame is identified as the *reference frame*. Ideally, this should be the one in which the beam is in the rest position ($Z = 0$). To initialize the external orientation of the cameras, K separate resections with the Direct Linear Transform (DLT) algorithm are carried out using the image coordinates $\mathbf{m}^i_j(t_0)$ of the dots detected in the reference frame and

the nominal rest coordinates of the dots $\hat{\mathbf{M}}_j(t_0)$. The DLT algorithm linearly solves Eq. (2) for $\mathbf{g}_i$ $(i = 1 \dots 6)$. The exterior orientations found by DLT were then refined by a partial BA with fixed points $\hat{\mathbf{M}}_j(t_0)$. This means that all the NM grid points are treated as known "ground control points". Since these ground control points have all $Z = 0$ this sets the zero for the vertical oscillation.

The actual computation of the 3D center coordinates $\mathbf{M}_j(t)$ of the dots at each time instant t was carried out in the final BA that comprises all frames of all cameras, where exterior orientations were also considered among unknowns. The initialization consists in the nominal coordinates of the dots centers and the exterior orientation computed in the previous step.

The unknowns are the six exterior parameters of each of the K cameras (the interior parameters are fixed) and the 3D coordinates of each grid point (NM) in each frame. In this phase the grid points have unknown coordinates, and the nominal values are used only for initialization.

In summary, these are the steps of the triangulation stage:

1. For each camera separately, compute exterior orientation with DLT using nominal grid points coordinates and measured grid points coordinates in the reference frame;
2. partial BA (exterior orientation only) with all cameras and $t = 0$ (the reference frame), fixed nominal grid points coordinates in object space;
3. full BA with all cameras and all frames ($t > 0$). Coordinates of 3D points and orientation parameters are estimated simultaneously.

Appendix A reports the code for the triangulation, that uses functions implemented in (Fusiello, 2022, Fusiello, 2024).

3. Experiments

In this section, we report on experiments conducted with two synchronous Inivation DVXplorer event cameras on a steel cantilever beam. The two cameras are arranged on a circular sector centered on the beam, with a radius of approximately 50 cm, and are positioned at slightly different heights (see Fig 2). As depicted in Fig. 3a, the beam is equipped with a grid of 10 × 3 silver circular markers, of which only a 9 × 3 subset will be used, since the row closest to the clamped end was not detectable due to its negligible motion. In fact, the reconstructed image (Fig. 3c) gets dimmer as it approaches the clamped end.

The beam is excited in bending by a shaker connected to the beam via a stinger at 45 mm from the clamped end, activating the first resonance frequency of the structure, which is around 35 Hz. The cantilever beam was recorded under bending excitation to obtain the deflection shape of the first flexural mode. Following the event recordings, the E2VID neural network was used for frame reconstruction.

Figure 3 shows a picture of the cantilever beam with the circular markers and an example of acquisition by the DVXplorer event camera.

3.1 Assessing camera synchronization

The cameras are synchronized in hardware. The clock synchronization feature ensures precise timing among multiple cameras connected in a daisy-chain. The first camera is designated as the master, with subsequent cameras automatically synchronized. Resetting the master camera's timestamp synchronizes all connected devices within a microsecond-precision.

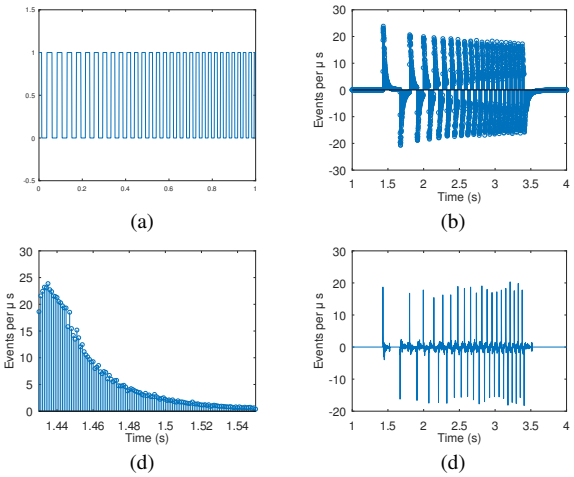


Figure 4. (a) Example of "square wave chirp" used in the assessment experiments. (b) Signal obtained from an event sequence with bin width T =0.1 ms. (c) Zoom on the first impulse of (b). (d) Signal obtained by filtering (b) with a filter designed to remove the exponential tails

To assess this property, an experiment was conducted in which both cameras frame the same portion of a diffusive surface illuminated with a modulated light source. The modulation signal is a chirp similar to the one shown in Fig. 4a, which allowed us to align the event histograms of the cameras via correlation. While the timestamps are in microseconds, we were unable to test this precision with the histogram technique, as the bins should be large enough to collect a significant number of events. A suitable compromise was set at a bin width of 100 µs, which would allow temporal alignment verification with a resolution of 1/10 ms. An issue that reduces the time resolution even further can be seen in Fig. 4b which represents the histogram of the events, normalized in order to have the number of events per microseconds, that it is obtained by using the chirp, Fig. 4c shows a zoom on the first impulse. One would expect a very short peak of positive events (a pixel with increased luminosity) at the time the light is turned on and then nothing; instead, we observed an "exponential tail" whose origin is probably due to the cameras timing precision loss in the presence of simultaneous events and detection threshold (iniVation, 2020, Purohit and Manohar, 2022). This tail artificially enlarges the signal support and makes its correlation less sharp. In order to improve the resolution, we filtered the signal with the discrete time filter

$$y(n) = x(n) - \alpha x(n-1) \tag{4}$$

where is determined by using an AR model ($\alpha = 0.992$ with bin width $T = 0.1$ ms in the case of Fig. 4). This kind of filter removes exponential tails that decay as α^n, $n > 0$. Fig. 4d shows the result of applying filter (4) to the signal of Fig. 4b.

The maximum correlation occurs at 0.29 ms (with parabolic interpolation), which corresponds to the time shift that best aligns the events of the two cameras. We were therefore able to verify that the synchronization of the two cameras is correct up to tenths of milliseconds.

3.2 Calibration

Cameras have been calibrated as described in Sec. 2.2. In particular, a checkerboard pattern has been moved in front of each camera, and twelve intensity images have been extracted from the video reconstructed by E2VID, paying attention to the di-

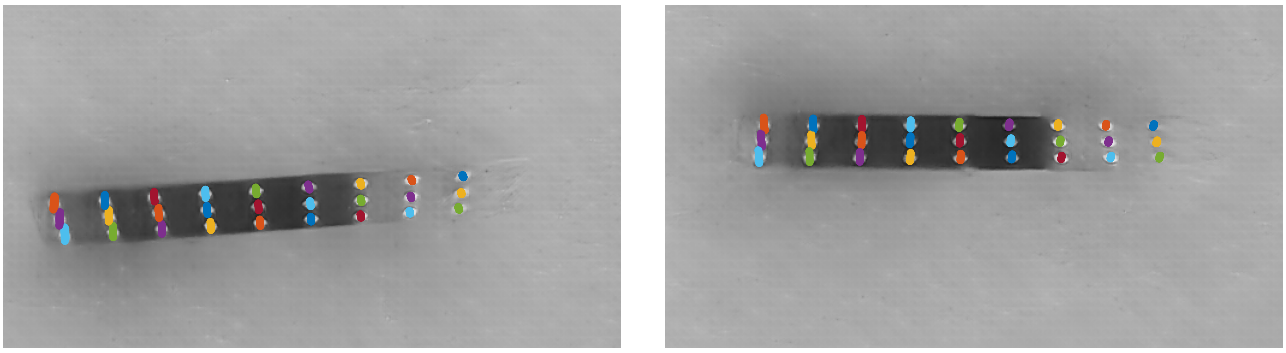


Figure 5. Traces of the markers overlaid on the actual image (cropped for better readability). The clamped end is on the right side in both images

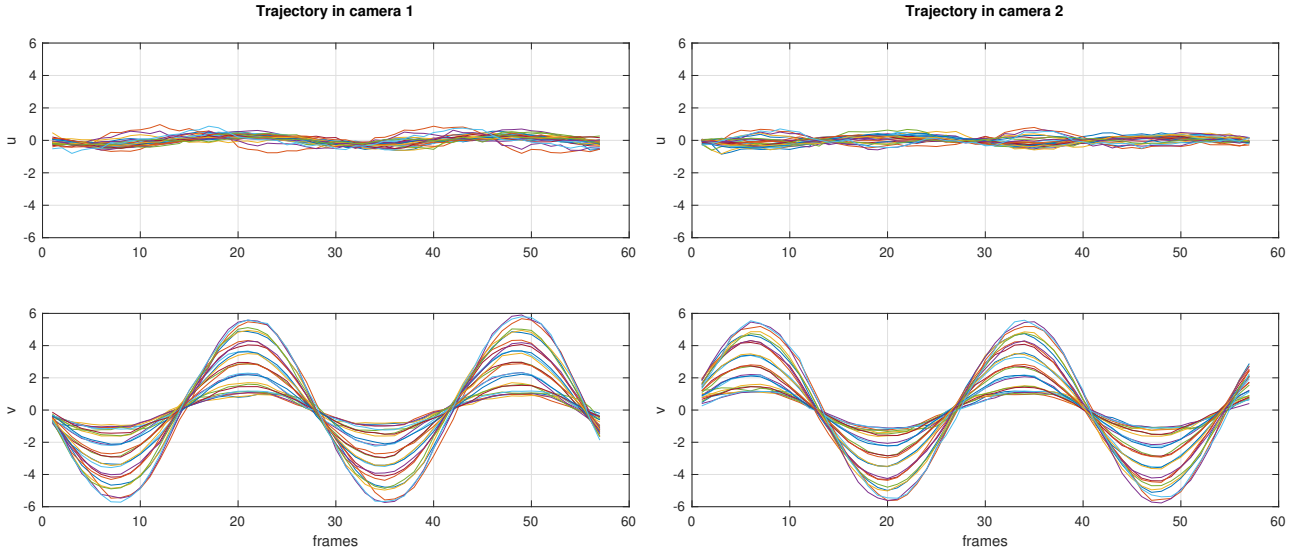


Figure 6. Horizontal and vertical displacements of markers in the images (average has been subtracted to each signal for better readability)

versity of the checkerboard orientations. These images have been fed to the Sturm-Maybank-Zhang algorithm implemented in (Fusiello, 2022).

	α_u	α_v	u_0	v_0	γ	κ_0
camera 1	954	959	385	229	0	-0.3629
camera 2	1059	1055	367	230	0	-0.4130

Table 1. Interior parameters obtained by calibration.

baseline [mm]	α [deg]	D_1 [mm]	D_2 [mm]
743	96	435	541

Table 2. Some exterior figures: D is the distance from the camera to the $Z = 0$ plane, α is the vergence angle

The image size is 640 × 480. Interior parameters of the Brown model are reported in Tab. 1: the principal distance measured in pixels (horizontal α_u and vertical α_v dimensions), coordinates of the principal point (u_0, v_0), first coefficient of the radial distortion polynomial (κ_0). Table 2 shows some data computed from the exterior parameters. From these figures the ground sampling distance (GSD) of the two cameras with respect to the $Z = 0$ plane (that contains the cantilever beam) can be computed, obtaining 0.46 mm and 0.51 mm respectively.

3.3 Measuring vibrations

This study considers the flexural response of a cantilever beam model structure subject to a small transverse force excitation. The beam is thus characterised by small deflections such that its response can be assumed to be linear.

More specifically, the flexural response shows synchronous vibrations such that the transverse oscillations of all points of the beam are in phase (or in phase opposition) but have different amplitudes. In general, the flexural response is given by the superposition of second order modal responses, which are characterised by distinct mode shapes and specific resonant responses. In this study the response to a tonal excitation at the first resonance frequency is considered such that the flexural vibration is controlled by the first mode shape.

The tracking produced 2D trajectories for each of the 27 dots. Fig. 5 shows the traces of markers overlayed on one image. It can be appreciated that the motion is very small, of the order of a few pixels, mainly in the vertical direction. This is confirmed by Fig. 6, where the u and v coordinates of the tracked points are plotted. The maximum displacement in the v direction is about 5 pixels, while in the u direction it is less than one pixel, and although it drowns in noise, a periodic trend is still observable.

The first external orientation obtained by DLT gives a RMS re-

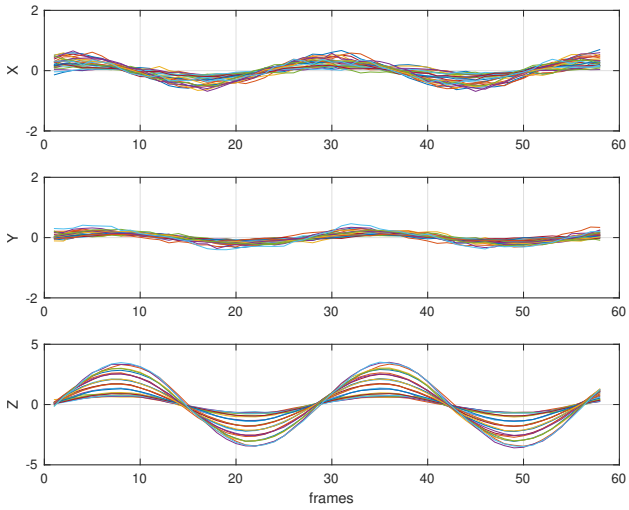
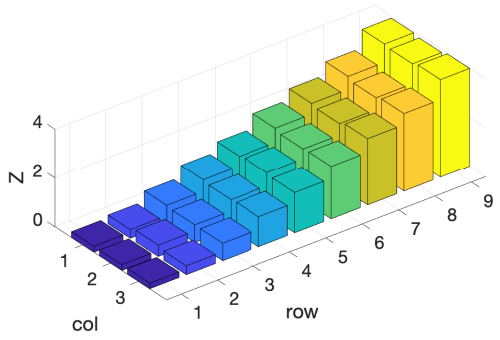
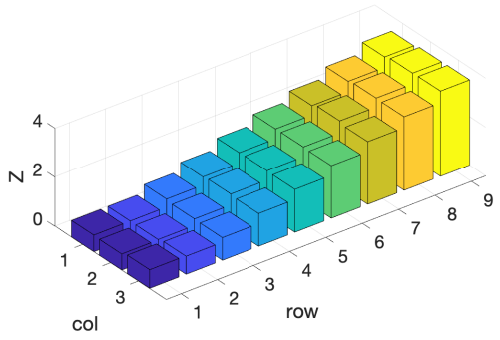


Figure 7. X, Y, and Z displacements (in mm) for all 27 grid points (average has been subtracted to each signal for better readability)

projection error of 1.03 pixels. After refining the exterior parameters with a partial BA the error decreases to 0.66 pixels. The final BA that also computes the position of the 3D points yields an RMS reprojection error of 0.133 pixels.



(a) LDV



(b) our method

Figure 8. Transverse displacement (im mm) measured with with the LDV (a) and our method (b).

The bundle adjustment produces 3D points trajectories for the tracked points. The X, Y, and Z displacements of the 3D points are plotted in Fig. 7. As expected, they follow sinusoids with the same frequency and phase but different amplitudes. The Z components provide the transverse displacements, that is the flexural vibration of the beam. The X and Y components give the in-plane displacements, which, as one would expect, are much smaller, but yet with a periodic time-history.

3.4 Comparison with a laser vibrometer

Furthermore, the experimental rig is equipped with a Polytec PSV-500-A laser doppler vibrometer (LDV), which is employed to obtain reference measurements of the cantilever beam's transverse vibration field (the declared precision of the instrument is on the order of micrometers). The operational principle involves directing a laser beam onto the surface under investigation. As the surface vibrates, the frequency of the reflected laser beam undergoes a Doppler shift. This shift is analysed to determine both the amplitude and frequency of the vibrations along the axis of the laser beam. Amplitudes are shown in Fig. 8a.

In our method, the transverse displacements of the 27 grid points are quantified as the mean of the extreme Z values observed at each period (in the experiments under consideration we used two periods) (Fig. 8b). The residual RMS error with respect to the LDV measures was 0.27 mm, which drops to 0.23 mm if the first row of points, the ones closer to the clamped end, is not considered. In that area, in fact, the transverse displacement is very low, below 0.5 mm, so fewer events are generated, making the reconstructed intensity image dimmer and causing a less accurate tracking.

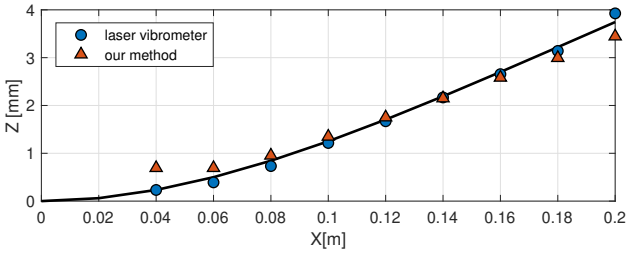


Figure 9. Comparison of the average transverse displacement measured with our method and with a laser vibrometer (LDV). The black line represent the theoretical deflection shape fitted to the LDV measures

Figure 9 shows the average over the rows of the transverse displacements measured with our method and the LDV. It also depicts the theoretical deflection shape – fitted onto the LDV measures – that depends on the boundary conditions and the physical properties of the beam, such as its material properties, geometry, and support conditions.

4. Conclusions

This study demonstrated that event cameras can be used to measure mechanical vibrations with comparable precision to those obtained with an expensive laser Doppler vibrometer. With a GSD of approximately 0.5 mm, achieving a RMS error of 0.27 mm with respect to the LDV, can be considered a remarkable result.

Although the frequency employed in our experiments is still relatively low (36 Hz), the only remaining obstacle to achieving a frequency of up to 500 Hz is the reconstruction of the intensity image, which becomes increasingly dim as the number of accumulated events that generate a frame decreases. In addition, the

vibration amplitude, in response to a given excitation, decreases as the mode frequency increases, making motion tracking more challenging. Our future research will be dedicated to addressing this issue.

Acknowledgments

This work was supported by the University of Udine Research Programme - ESPERT project.

Appendix A: Matlab code for the triangulation function

```
function M_out = reconstruct_mgrid(m_grid,...
                 frames,M_grid,KK,kappa)
% Reconstruct the 3D grid points using BA
% m_grid: 2D coordinates of the grid points
%         in different frames
%         m_grid{camera}{frame}(dim,k);
% frames: indices of the frames to be considered
% M_grid: nominal 3D coordinates of grid points
% KK    : cell array of intrinsis parameters
% kappa : cell array of distorion coefficients
% M_out : 3D coordinates of the grid points
%         in all frames
%
% Author: Andrea Fusiello (2024)
% Depends on functions from the CV Toolkit
% (github.com/fusiello/Computer_Vision_Toolkit)

n_cams = length(m_grid);
n_points = size(M_grid,2);
rest_frame = 1;

%% Camera orientation by resection
fprintf('Orienting %d cameras\n', n_cams);
% initialize by resection with rest frame
m = cell(n_cams);
P = cell(n_cams);
for i=1:n_cams
    % points in the rest frame
    m{i} = m_grid{i}{frames(rest_frame)};
    [R,t] = exterior_lin(m{i},M_grid,KK{i});
    P{i} =  KK{i} * [R,t];
end

% refine with BA
[P,M] = bundleadj(P,M_grid,m,...
    'Verbose','FixedIntrinsic',...
    'FixedPoints',n_points);

%% Bundle adjustment with all the frames
fprintf('Processing all frames at once\n');
m_all = cell(n_cams);
for i=1:n_cams
    m_all{i} = cat(2,m_grid{i}{frames});
end

M_all = repmat(M,1,length(frames));
[P,M_all] = bundleadj(P,M_all,m_all,...
    'Verbose','FixedIntrinsic',...
    'DistortionCoefficients', kappa);

M_out = reshape(M_all,3,[],length(frames));
end
```

References

Baqersad, J., Poozesh, P., Niezrecki, C., Avitabile, P., 2017. Photogrammetry and optical methods in structural dynamics – a review. *Mech. Syst. Signal Pr.*, 86, 17-34.

Barone, S., Neri, P., Paoli, A., Razionale, A. V., 2019. Low-frame-rate single camera system for 3D full-field high-frequency vibration measurements. *Mechanical Systems and Signal Processing*, 123, 143-152.

Bartilson, D. T., Wieghaus, K. T., Hurlebaus, S., 2015. Target-less computer vision for traffic signal structure vibration studies. *Mechanical Systems and Signal Processing*, 60-61, 571-582.

Beberniss, T. J., Ehrhardt, D. A., 2017. High-speed 3D digital image correlation vibration measurement: Recent advancements and noted limitations. *Mechanical Systems and Signal Processing*, 86, 35 – 48.

Benosman, R., Clercq, C., Lagorce, X., Ieng, S.-H., Bartolozzi, C., 2014. Event-based Visual Flow. *IEEE Trans. Neural Netw. Learn. Syst.*, 25(2), 407–417.

Castellini, P., Tomasini, E. P., 2004. Image-based tracking laser Doppler vibrometer. *Review of Scientific Instruments*, 75(1), 222-232.

Conradt, J., Cook, M., Berner, R., Lichtsteiner, P., Douglas, R. J., Delbruck, T., 2009. A pencil balancing robot using a pair of AER dynamic vision sensors. *IEEE Int. Symp. Circuits Syst. (ISCAS)*, 781–784.

Cook, M., Gugelmann, L., Jug, F., Krautz, C., Steger, A., 2011. Interacting maps for fast visual interpretation. *Int. Joint Conf. Neural Netw. (IJCNN)*, 770–776.

Del Sal, R., Dal Bo, L., Turco, E., Fusiello, A., Zanarini, A., Rinaldo, R., Gardonio, P., 2021. Structural vibration measurement with multiple synchronous cameras. *Mechanical Systems and Signal Processing*, 157, 107742.

Fusiello, A., 2022. Computer Vision Toolkit for Matlab. `https://github.com/fusiello/Computer\_Vision\_Toolkit`. Accessed: 2022.

Fusiello, A., 2024. *Computer Vision: Three-dimensional Reconstruction Techniques*. Springer Cham.

Gallego, G., Lund, J. E. A., Mueggler, E., Rebecq, H., Delbruck, T., Scaramuzza, D., 2018. Event-based, 6-DOF Camera Tracking from Photometric Depth Maps. *IEEE Trans. Pattern Anal. Mach. Intell.*, 40(10), 2402–2412.

Gorjup, D., Slavič, J., Boltezar, M., 2019. Frequency domain triangulation for full-field 3D operating-deflection-shape identification. *Mech. Syst. Signal Pr.*, 133 106287, 143-152.

Helfrick, M. N., Niezrecki, C., Avitabile, P., Schmidt, T., 2011. 3D digital image correlation methods for full-field vibration measurement. *Mechanical Systems and Signal Processing*, 25(3), 917-927.

iniVation, A., 2020. Understanding the performance of neuromorphic event-based vision sensors. *Tech. Rep.*

Ismail, H., Signal, M., Amies, A., Haggers, M., Chase, J., 2019. Laser doppler vibrometer validation of an optical flow motion tracking algorithm. *Biomed. Signal Pr. and Control*, 49, 322-327.

Kim, H., Leutenegger, S., Davison, A. J., 2016. Real-time 3d reconstruction and 6-DoF tracking with an event camera. *Eur. Conf. Comput. Vis. (ECCV)*, 349–364.

Kuhn, H. W., 1955. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2), 83-97.

Lagorce, X., Orchard, G., Gallupi, F., Shi, B. E., Benosman, R., 2017. HOTS: A Hierarchy of Event-based Time-surfaces for Pattern Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(7), 1346–1359.

Muglikar, M., Gehrig, M., Gehrig, D., Scaramuzza, D., 2021. How to calibrate your event camera. *IEEE Conf. Comput. Vis. Pattern Recog. Workshops (CVPRW)*.

Pappa, R. S., Black, J. T., Blandino, J. R., Jones, T. W., Danehy, P. M., Dorrington, A. A., 2003. Dot-Projection Photogrammetry and Videogrammetry of Gossamer Space Structures. *Journal of Spacecraft and Rockets*, 40(6), 858-867.

Park, S., Park, H., Kim, J., Adeli, H., 2015. 3D displacement measurement model for health monitoring of structures using a motion capture system. *Measurement*, 59, 352-362.

Purohit, P., Manohar, R., 2022. Field-programmable encoding for address-event representation. *Front.in Neurosci.*, 16.

Rebecq, H., Ranftl, R., Koltun, V., Scaramuzza, D., 2019. High Speed and High Dynamic Range Video with an Event Camera. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43, 1964-1980.

Reu, P. L., Rohe, D. P., Jacobs, L. D., 2017. Comparison of DIC and LDV for practical vibration and modal measurements. *Mechanical Systems and Signal Processing*, 86, 2 – 16.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, Springer, 234–241.

Ryall, T. G., Fraser, C. S., 2002. Determination of Structural Modes of Vibration Using Digital Photogrammetry. *Journal of Aircraft*, 39(1), 114-119.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., WOO, W.-c., 2015. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, 28.

Sironi, A., Brambilla, M., Bourdis, N., Lagorce, X., Benosman, R., 2018. HATS: Histograms of averaged time surfaces for robust event-based object classification. *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 1731–1740.

Wahbeh, A. M., Caffrey, J. P., Masri, S. F., 2003. A vision-based approach for the direct measurement of displacements in vibrating systems. *Smart Materials and Structures*, 12(5), 785–794.

Zhou, Y., Gallego, G., Rebecq, H., Kneip, L., Li, H., Scaramuzza, D., 2018. Semi-dense 3d reconstruction with a stereo event camera. *Eur. Conf. Comput. Vis. (ECCV)*, 242–258.

Zhu, A. Z., Yuan, L., Chaney, K., Daniilidis, K., 2018. EV-FlowNet: Self-supervised optical flow estimation for event-based cameras. *Robotics: Science and Systems (RSS)*.