Experimental investigation of structural modal identification using pixels intensity and motion signals from video-based imaging devices: performance, comparison and analysis

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

Experimental investigation of structural modal identification using pixels intensity and motion signals from video-based imaging devices: performance, comparison and analysis

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

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

ABSTRACT

This paper aims to experimentally evaluate a simplified vision-based method for structural health monitoring (SHM). Contrary to conventional solutions that rely on extracting motions through image processing, this paper proposes to conduct the SHM analysis by the direct processing of pixel brightness without extracting the motion signals beforehand. After some pre-processing steps, it is shown that the brightness data reveal essential information about the dynamic characteristics of the monitored vibrating structure. Furthermore, the low-level information of the pixel is compensated by an efficient selection of the so called "active pixels" throughout the image time flow. Finally, a subspace system identification-based method is applied to the brightness data, so that the modal parameters with uncertainty bounds are estimated within a large range of model orders displayed in a stabilization diagram. The experiment database consists of image time flows of a cantilever beam excited by a shake table driven by a band limited random noise. Modal vibrations range from 1 to 173 Hz.

Keywords: Vision modal decomposition, Pixels intensity, Motion extraction, Subspace identification, Uncertainty bounds

1. INTRODUCTION

Many infrastructures of our society, including bridges and buildings, are facing increasing challenges due to climate change. To this end, extensive research works have been carried out over the past decades for structural health monitoring (SHM).[1–3] Operational modal analysis (OMA) is a mature field in SHM. Its goal is to identify inherent structural information such as natural frequencies, damping ratios, and mode shapes from output-only vibration measurements on a structure under unknown ambient excitation. OMA is of high interest for many applications, e.g. damage diagnosis,[4–6] structural model updating,[7] identification of excitation forces.[8]

The process of OMA starts with data collection. Conventional sensing technology was based on setting accelerometers on the structure. Then, methods like frequency domain decomposition,[9,10] auto-regressive moving average (ARMA),[11] stochastic subspace identification[12,13] and the natural excitation technique[14,15] can be used to obtain the modal parameters with high accuracy. However, with traditional sensors, vibrations are measured only at (relatively few) discrete points in the structure, providing limited information for damage diagnosis or model updating.[16,17]

Non-contact sensing technology, such as laser sensor[18] and ground-based synthetic aperture radar,[19] have been successfully applied as an alternative to traditional devices. Whereas numerous applications benefit from these measurement technologies, they are expensive and not adapted to all types of structures.

With both technological and processing improvements, computer vision-based techniques have emerged in the past decade as an alternative, cheaper and easier solution to perform SHM. Commonly, this is done in two

Further author information:
B. Merainani : boualem.merainani@inria.fr
B. Xiong : bian.xiong@inria.fr

steps. First, the vibrating motion is estimated at long-range from video image flows. Various image processing techniques have been developed to extract the local motion signals from subtle changes in the pixel brightness over time. One can classify the approaches to video-based motion extraction into *optical flow*,[20,21] including phase based optical flow,[22] and the *image template matching*, including the widespread digital image correlation (DIC).[23] Second, the most dominant vibrating frequencies of the structure are determined from the power spectral density of the motion time history. More advanced SHM tools rely on using the stochastic subspace-based identification (SSI) method.[24]

By contrast, this paper proposes to conduct SHM by the direct processing of pixel intensity time variations without extracting the motion signals beforehand. The proposed approach is believed to be less computationally expensive. To circumvent the low-level information of the pixel intensity, the proposed approach is able to represent the brightness time variations within each image blocks (ROI : region of interest) by a single time signal obtained after selecting the so called 'active pixels' and averaging their time variations. Then, the stochastic subspace-based identification method[24] that provides more accurate modal parameters along with uncertainty bounds was adopted for refined SHM analysis.

Experiments in a laboratory setting on a vertical cantilever beam are performed to verify the proposed intensity-based processing approach. A comparison has been made with conventional motion-based processing in terms of identified spectral lines and on the stability diagrams of modal identification.

2. EXPERIMENTAL SETUP

A vertical cantilever beam of $2\,m$ long, $30\,mm$ wide, $5\,mm$ thick, and of density $7800 kg/m^3$ was tested to evaluate the performance of the proposed method for vision modal identification. Details of the experimental setup are given bellow and shown in Fig. 2:

- The vertical beam was fixed at its bottom to a shake table (model: NEBS GR-63), which was driven in random mode horizontally in the direction perpendicular to the axis of the camera.
- Video images were recorded at the frame rate $512\,fps$ and at the frame resolution 1920×136 with a PHANTOM camera (model :MIRO 320S), mounted with a NIKKOR lens with a fixed $50\,mm$ focal length. The camera was set on a tripod and installed at a distance of about $5.6\,m$ from the beam, so that the full beam length was in the scene. The scaling factor was $1.02\,mm/pixel$.
- Three additional lighting (Halogen Spotlight, MAZDA CORMORAN type of 1500 W) were used to improve the brightness conditions; in addition, a white curtain was used as a backdrop to enhance the contrast with the structure edges.
- For motion tracking, three paper markers were mounted at some particular positions along the beam, namely, at the top (1/1), middle (1/2) and $0.5\,m$ (1/4) from the cantilever beam bottom. Please see Fig. 4 (top left) for illustration of the marker.
- Three laser displacement sensors were located at the same heights of the markers to serve as the motion benchmark.

3. IMAGE PROCESSING FOR SHM

In this section, the images of the vibrating cantilever are decomposed into image blocks (or ROIs). Two processing are performed on each ROI assuming a solid state motion without deformation. The first processing aims at computing a specific averaged of the brightness time variation. The second one estimates motion signal by the conventional video-based technique. For both cases, The time signals over the set of ROIs will be processed by SSI method for SHM purposes in Section 4.

3.1 Proposed Image Brightness Processing

The proposed image processing includes the selection of active pixels, some phase correction of the brightness time variations, before computing the final averaged time signal.

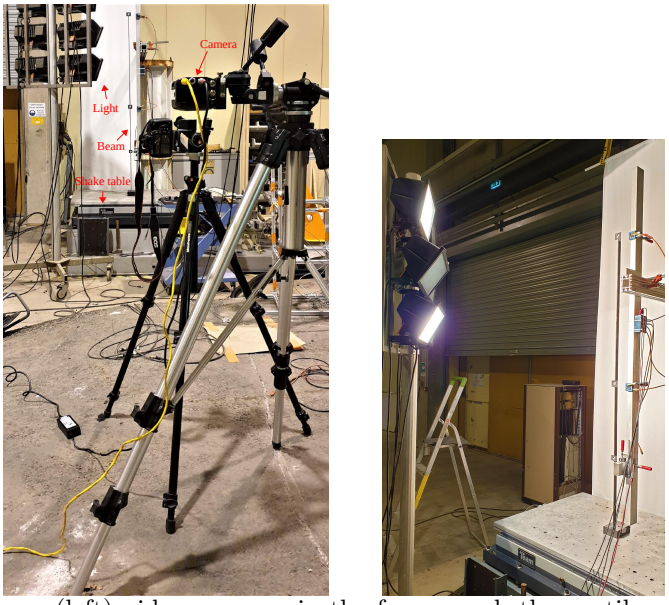

Figure 1: Experimental setup: (left) video cameras in the foreground, the cantilever beam in the background, (right) a close-up view of the lighting conditions, the cantilever beam and the shaking table.

3.1.1 Selection of active pixels

Active pixels are supposed to bring the most information about the vibrating motion. Active pixels usually correspond to the edge between adjacent contrasted areas (The amount of active pixels increase when the image is highly textured, e.g., speckle pattern or random dot pattern). The brightness time variation of such pixels is then closely correlated to the structure motion at that location. According to Fig. 2, this is easily seen when the motion range is within a pixel. When the motion covers a few pixels range, the brightness time variations of each pixel is truncated in amplitude. The correlation with the structure motion is then recovered on the averaged brightness over some pixel neighborhood.

Many algorithms were developed in the literature to locate the edges from the detection of significant local spatial changes in images by either the gradient operators,[25] Gaussian derivatives,[26] or the steerable filters[27] as illustrated in Fig. 4 (top row).

Within the scope of application, the edge computation is usually performed once from the first image frame. These methods are then better tailored to small motion range, namely within one pixel. However, in practical condition, the motion amplitude range is not known beforehand and can be beyond one pixel.

By contrast, a spectral approach is proposed in this section to select the active pixels without any prior knowledge. Indeed, it can be used to determine the pixel-wise motion range. Active pixels are selected within each ROI on the basis of the spectral magnitude of the pixel brightness time history. The proposed selection of pixels is based on the following three-steps processing.

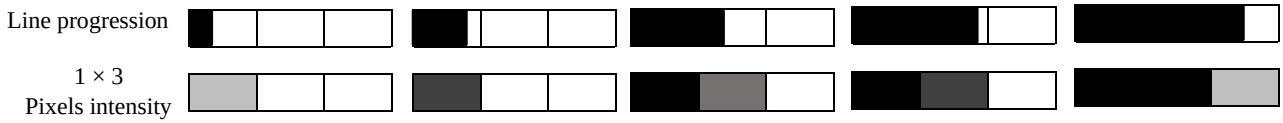

Figure 2: Schematic representation of the pixel brightness variation w.r.t. the motion of some step edge in front of the camera. Top row figures: the right edge of a black object moving from the left to the right along 3 pixels. Bottom row figures: the corresponding pixel brightness variations of the camera.

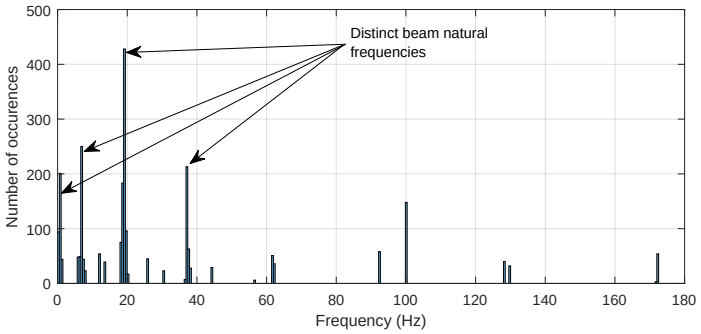

Figure 3: Frequency peaks histogram computed from the brightness time variations of the pixels within the selected ROI.

- First, the Beam Resonance Frequencies (BRF) of the mechanical structure are determined. If the finite element modelling of the structure is available, BRF can be theoretically determined. Otherwise, the main BRFs are determined from the peak magnitudes of the pixel spectrogram. The histogram of the frequency peak location is illustrated in Fig. 3. At this step, the spectrogram is approximated because the pixel brightness may bring truncated information (see Fig. 5). Nevertheless, it is sufficient to determine the set of BRFs.
- Second, the set of BRFs is used to compute the energy of ROI pixels according to Algorithm 1. The maximum energy value serves as the reference level to select active pixels.
- Finally, the pixels having energy above 20% of the latter reference energy level, are selected as active pixels.

Algorithm 1: Active pixel selection within $N \times P$ ROI

Inputs: - (x, y) pixel intensity time history, $\boldsymbol{I}(x, y, \boldsymbol{t})$, where, $\boldsymbol{t}$ is a $L \times 1$ time vector;
- K Beam resonance frequencies BRF;

Output: Pixels energy;
for $x = 1, \ldots, N$ **do**
  for $y = 1, \ldots, P$ **do**
    for $k = 1, \ldots, K$ **do**
      Band-pass filtering around $BRF(k)$: $\boldsymbol{I}_{k,filtered}$;
      Mean intensity computation $\overline{\boldsymbol{I}}_{k,filtered}$;
    end
    Energy computation $E_{x,y} = sum(|\overline{I}_{k,filtered}|^2)$;
  end

The results of active pixels selection procedure are illustrated in Fig. 4 (bottom row) and by the red pixels in Fig. 5. Compared with the conventional edge pixels selection at the top row, the proposed spectral approach achieves more accurate detection. On the one hand, the target edges which are perpendicular to the motion were accurately identified. On the other hand, the width of the detected edge is closely related to the pixel-wise motion amplitude, measured by the laser sensor.

3.1.2 Averaged brightness ROI signal

For convenience, the brightness time variations of the selected active pixels (as shown by red pixels in Fig. 5) are arranged in a matrix form $\mathbf{I}_{P_a}$ where each column $\boldsymbol{I}(x_i, y_i, \boldsymbol{t})$ represents the centered brightness time vector of the i^{th} selected active pixel at the coordinates (x_i, y_i) in the ROI image.

At this stage, the pixel brightness time variations may bring truncated information in case of motion magnitude larger than one pixel, as shown on Fig. 5 (right column). In addition, the brightness time variations have

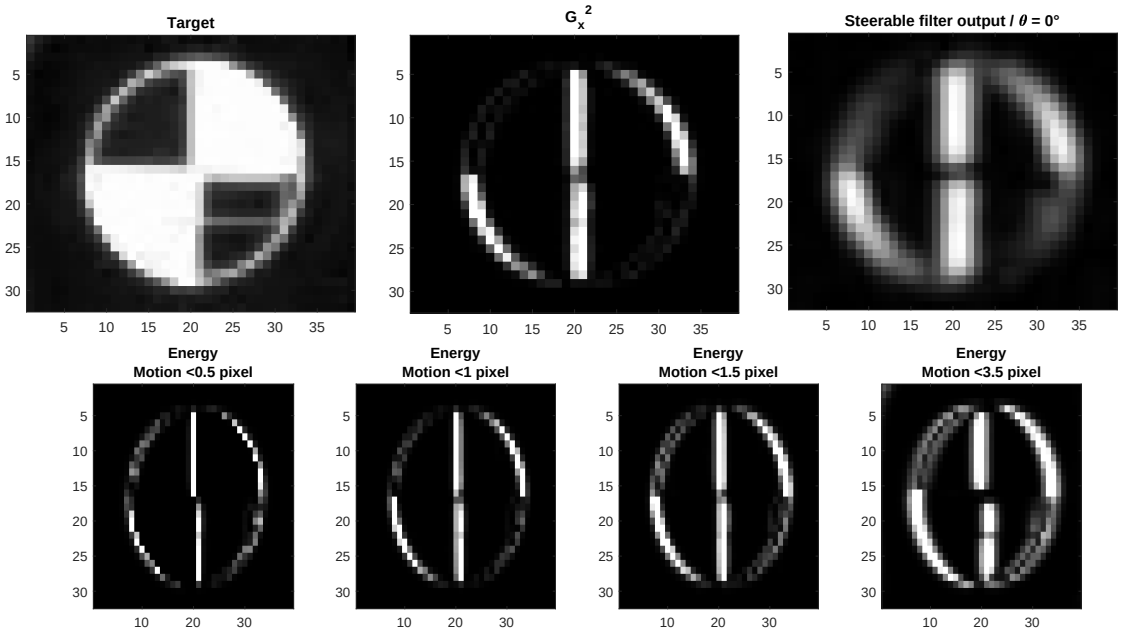

Figure 4: Pixels energy calculation based proposed approach prior active pixels selection in case of 4 different motions amplitude (Bottom), target (Top left), gradient ($G_x{}^2$) (Top middle) and steerable filter (Top right).

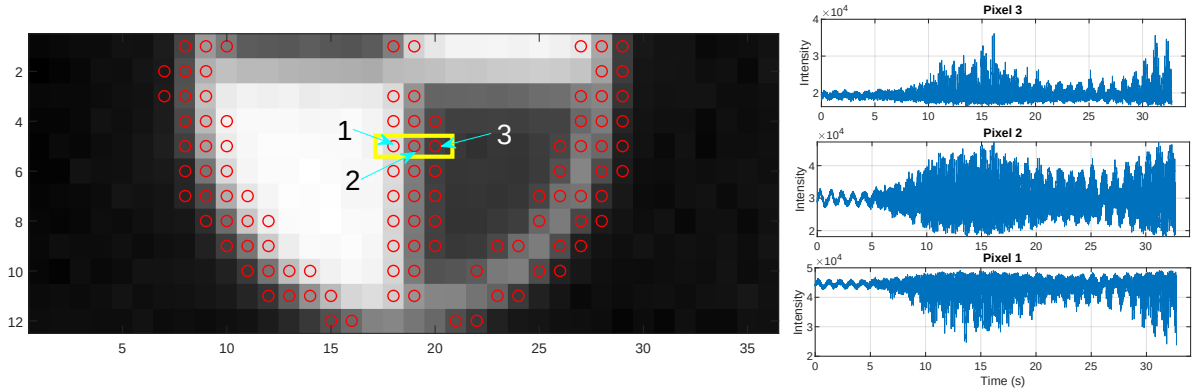

Figure 5: Illustration of the pixel selection: (left) Selected active pixels in red circles are superimposed on the target pattern; (right) Brightness time variations of pixels 1-3 within the boxcar.

the same polarity than the brightness slope at the edges.[21] When the ROI includes many different dark-to-light and light-to-dark transitions, the pixel brightness time variations show various polarity, which tend to cancel when they are summed up. As shown in,[21] phase correction should be performed to achieve a reliable averaged signal with the full vibrating content.
To this aim, the following polarity correction processing is proposed :

$$\forall i \in [2, P_a], \boldsymbol{I}(x_i, y_i, \boldsymbol{t}) = \boldsymbol{I}(x_i, y_i, \boldsymbol{t}) \times sign(\boldsymbol{I}(x_{i-1}, y_{i-1}, \boldsymbol{t})^T \boldsymbol{I}(x_i, y_i, \boldsymbol{t})). \quad (1)$$

where P_a is the amount of selected active pixels. The phase correction ensures that brightness of all selected active pixels are in phase before averaging. For illustration, Fig. 6 represents the processing result on the active pixels in Fig. 5.

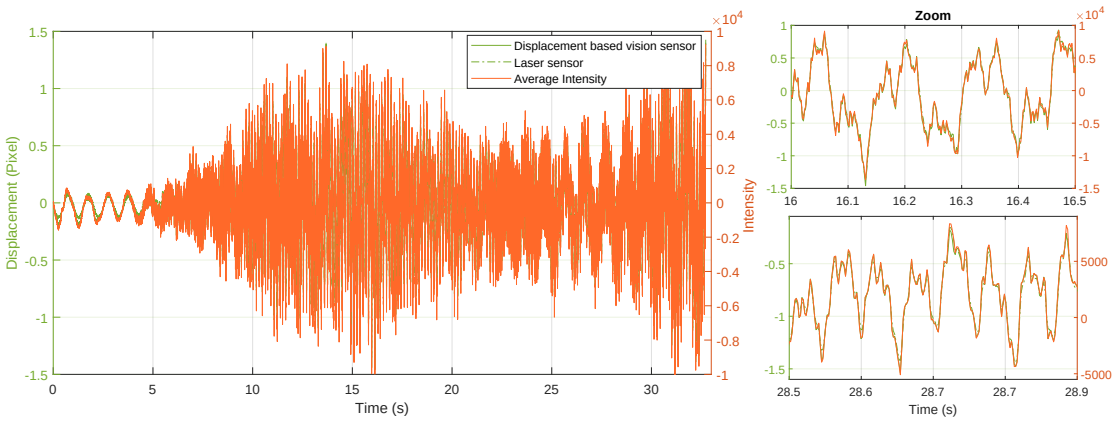

Figure 6: Averaged brightness time variations (right y-axis) superimposed on the displacements measured by both the laser and the vision sensors (left y-axis); full time record on the left, zoom on two short time intervals on the right.

3.2 Conventional video-based motion extraction

Generally, the motion extraction from video images is performed in two steps: pixel shift $(\Delta n, \Delta p)$ estimation using the cross-correlation (CC) and subpixel shift refinement $(\delta x, \delta y)$. Let $I_0 = I(n, p, t0)$ and $I_1 = I(n, p, t)$ be the reference and a template images, respectively. The pixel shifts can be obtained by maximizing the CC as :

$$(\Delta n^*, \Delta p^*) = \arg \max_{\substack{-N \leq \Delta n \leq N \\ -P \leq \Delta p \leq P}} \sum_{u,v} F(u,v) G^*(u,v) e^{i2\pi(\frac{u\Delta n}{N} + \frac{v\Delta p}{P})}, \tag{2}$$

where $F(u, v)$, and $G(u, v)$, are the Fourier transforms of I_0 and I_1 respectively.

Chan et al.[28] proposed a simplified version of classical optical flow estimation[29] to determine the 2D subpixel displacements by solving the system of equations given in 3. It is noteworthy that, in case of large displacements, the reference image must be shifted as $I(n + \Delta n, p + \Delta p, t0)$ to stay in the range of subpixel displacements.

$$\begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = \begin{bmatrix} \sum_{x,y} (\frac{\partial I_0}{\partial x})^2 & \sum_{x,y} \frac{\partial I_0}{\partial x} \frac{\partial I_0}{\partial y} \\ \sum_{x,y} \frac{\partial I_0}{\partial x} \frac{\partial I_0}{\partial y} & \sum_{x,y} (\frac{\partial I_0}{\partial y})^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{x,y} (I_1 - I_0) \frac{\partial I_0}{\partial x} \\ \sum_{x,y} (I_1 - I_0) \frac{\partial I_0}{\partial y} \end{bmatrix}, \tag{3}$$

3.3 Comparison between the two processings

Figure 6, shows and compares the time variations of the average ROI brightness (within the second ROI) and the displacements measured by both the vision and the reference laser sensors. The results shows a good consistency between the three signals.

On Fig. 7, the correlation coefficients (CorCoef) between the displacements and the brightness time variations are computed. It is seen that, the CorCoef varies from 0.996 to 0.9999 and shows the strong relationship between the intensities and the displacements as expected. It shall be noted, however, that the correlation is slightly lower for the upper ROIs, in which the structure motion is larger than one pixel. This is likely due to either insufficient active pixels in the ROI, non-uniform lighting conditions in the upper ROIs or inappropriate threshold.

Figure 8, shows the frequency spectrogram associated to the time signals in Fig. 6. For comparison, the spectrum computed from the average of 3 adjacent active pixels, represented in Fig. 5, was also superimposed on that of the average ROI brightness over all active pixels. The peak prominent indicator[30] was adopted to quantify the local signal to noise ratio of each resonance frequency. Figure 9 compares the results. On the one hand, the peaks, related to the first 4 resonant frequencies, in the laser sensor-derived spectrum, have a larger prominence compared to those derived from the spectra of the vision sensor and the proposed LI-derived vision data approach. On the other hand, our approach shows better peaks prominence, related to the fifth and sixth

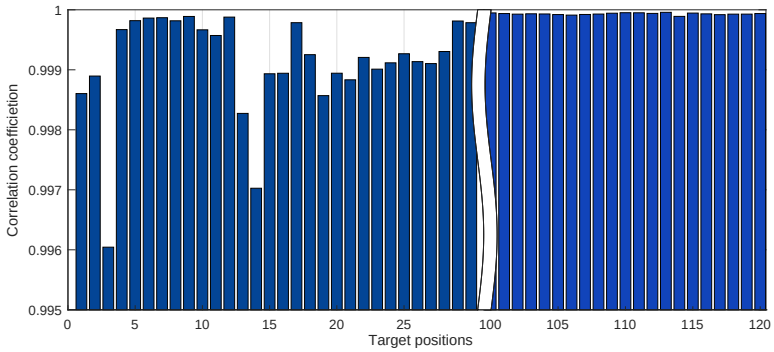

Figure 7: Correlation coefficients between the brightness time variations and vision-based displacements

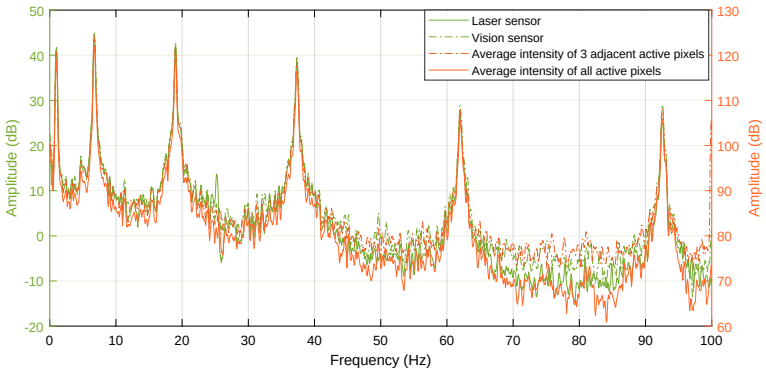

Figure 8: FFT spectra of the average LI time history of all active pixels and 3 adjacent active pixels, showing them on top of each other (right y-axis), FFT spectra of the displacements measured by the laser sensor superimposed on that obtained by the vision sensor (left y-axis).

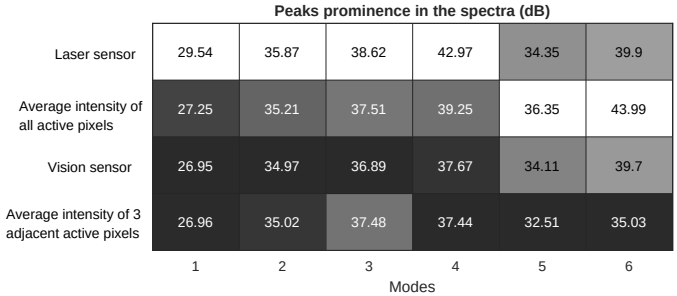

Figure 9: Comparison between the peaks prominence, of the first 6 resonant frequency lines obtained from the spectra of the laser sensor, the vision sensor and the proposed LI-derived vision data

resonant frequencies. Clearly, averaging all active pixel intensities improves the peaks prominence and thus the signal to noise ratio.

It is worth to mention that, higher modes of the beam were detected in the spectra. However, The laser sensor (Model : YP11MGVl80) cut-off frequency was $100Hz$. Beyond this value, the frequency content of the measured displacements may be affected. This is why, the comparison was limited over the $[0-100]Hz$ frequency range.

4. SYSTEM IDENTIFICATION RESULTS

In this section, the computation of the modal parameters of the beam and their uncertainty bounds is performed using the covariance-driven subspace identification algorithm.[24] With this method, the system matrices of a linear time-invariant state-space model are identified from each type of the measurements (LIs, displacement

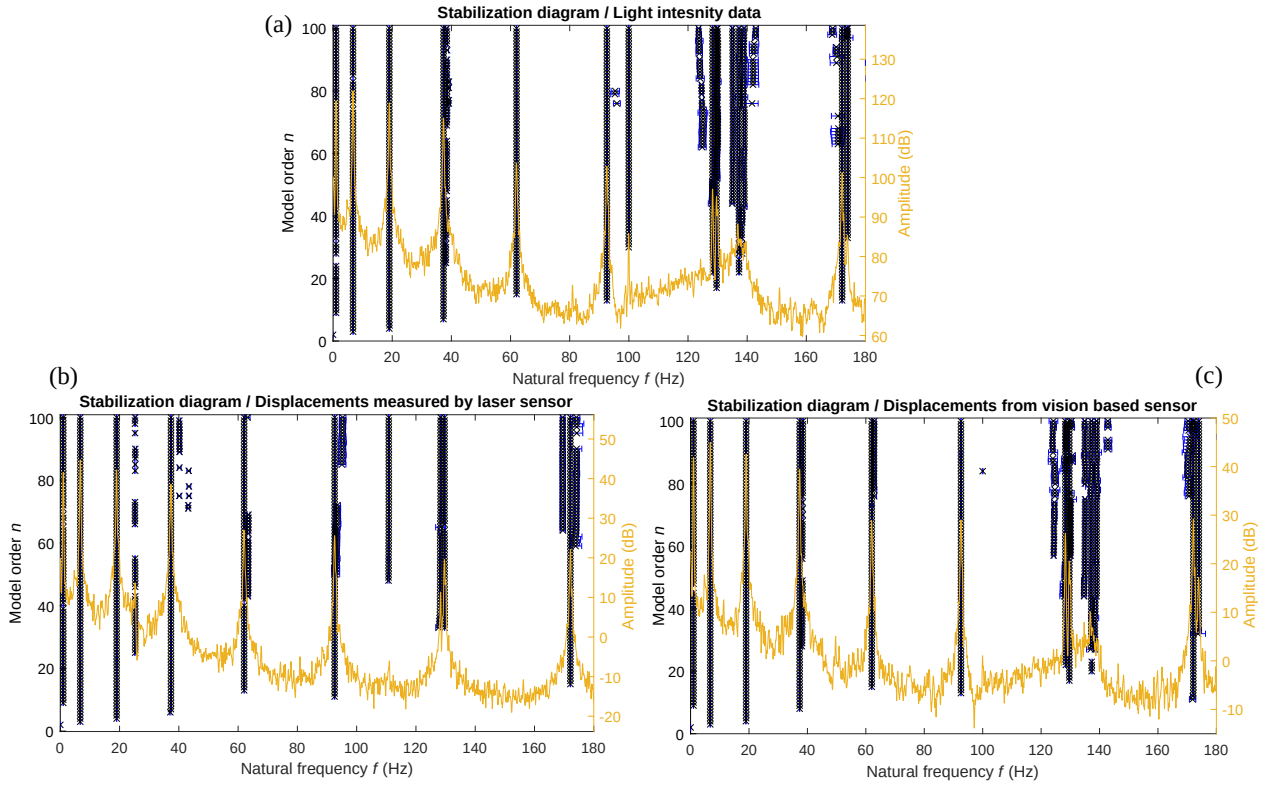

Figure 10: Stabilization diagrams containing the natural frequencies of the first eight modes with their confidence intervals.

extracted from video images and laser sensor signal). From their eigenvalues and eigenvectors, the modal parameters (natural frequencies, damping ratios, mode shapes) are retrieved. Since the model order of the system (corresponding to twice the number of modes that are present in the data) is unknown in practice, the system is identified for many successive model orders. The resulting frequencies are plotted against the model order in a stabilization diagram, which allows to distinguish "true" structural modes (which are stable for successive orders) from spurious noise modes.

In Fig. 10, the stabilization diagrams are shown together with the spectra, where the $\pm\sigma$ confidence interval of each frequency is plotted as a horizontal bar. It can be seen that, for the three analysis, real physical poles clearly manifest as stable poles from low model order (< 40). However, a high order implies the appearance of spurious modes that may be associated to the noise content of the signals. In addition, the stabilization diagram of the LI data, in Fig. 10.a, shows a stable mode at $100Hz$ but it has no physical meaning. In fact, this frequency is caused by the source light flicker.

From the stabilization diagrams, the modal parameters are chosen. In Table 1, the system identification results with their standard deviations are summarized. One can observe that between the identification results from the laser sensor displacements, vision based measurements and the LIs, the maximum error is around 0.6% of in terms of natural frequency estimation.

5. CONCLUSION

This paper presents a framework for structural modal identification using vision measurements. The key idea lies in the fact that the intensities at some pixels (titled active pixels) carry the useful information of the structural vibration. Therefore, an algorithm has been proposed to select such pixels. Modal parameters are then identified from the collected vision data using the output-only subspace identification method. Experiments concerning the video of a vertical cantilever beam excited with a shake table is studied and results show that

Table 1: Overview of the first eight modes

Mode	Intensity				Displacement measured by Laser sensor				Vision-based displacement measurements			
	$f(Hz)$	$\sigma_f(Hz)$	$\zeta(\%)$	$\sigma_\zeta(\%)$	$f(Hz)$	$\sigma_f(Hz)$	$\zeta(\%)$	$\sigma_\zeta(\%)$	$f(Hz)$	$\sigma_f(Hz)$	$\zeta(\%)$	$\sigma_\zeta(\%)$
1	0.9944	0.0149	0.4018	1.6236	1.0006	0.0071	0.0145	0.8407	0.9982	0.0092	0.0476	0.9035
2	6.7410	0.0110	0.2561	0.2118	6.7421	0.0099	0.1881	0.1346	6.7419	0.0124	0.2412	0.1644
3	19.0076	0.0359	0.2196	0.1487	19.0173	0.0127	0.1044	0.0537	18.9960	0.0216	0.2620	0.1547
4	37.3424	0.0195	0.1331	0.0482	37.3443	0.0182	0.1537	0.0465	37.3406	0.0185	0.1319	0.0443
5	62.0247	0.0220	0.1379	0.0363	62.0263	0.0272	0.1263	0.0497	62.0245	0.0221	0.1335	0.0391
6	92.5723	0.0224	0.0823	0.0206	92.5681	0.0168	0.0522	0.0159	92.5742	0.0288	0.0858	0.0456
7	129.6552	0.0312	0.0791	0.0255	129.6541	0.0288	0.0737	0.0244	129.6795	0.0220	0.0749	0.0217
8	172.1089	0.0263	0.0465	0.0146	172.1081	0.0255	0.0570	0.0148	172.1133	0.0322	0.0502	0.0164

- Only pixels that exhibit intensity changes relating to the beam vibration are selected. Therefore, the number of active pixels change with respect to motions amplitude (see Fig. 4).
- Strong relationship exists between the time variation of the intensities and displacements extracted from Vision sensor with a minimum CorCoef of $99,6\%$.
- The beam modal parameters including the natural frequencies and the damping ratios with their uncertainties were identified. A comparison between the SSI results when implemented on the intensity data, laser and vision data were also studied. The results showed around 2% of maximum error in terms of natural frequency estimation. The comparison indicates that the proposed approach has the potential to identify the structure modal parameters with reasonable levels of accuracy.

The proposed approach can estimate the modal properties. However, it is critical to efficiently select the active pixels. If they are insufficient or over-considered, the intensity time variations will be incomplete or corrupted by noise. Furthermore, the proposed approach depends highly on the illumination conditions. A constant illumination during tests is required to obtain reasonable results.

Authors are grateful to Research Fund for Coal and Steel for funding part of this work under grant agreement No 800687 in the framework of DESDEMONA project, and to Ivan Guéguen, Jean-Philippe Gourdon and Fabrice Blaineau for the essential assistance provided for the experimental set up.

REFERENCES

[1] Burgos, D. A. T., Vargas, R. C. G., Pedraza, C., Agis, D., and Pozo, F., "Damage identification in structural health monitoring: A brief review from its implementation to the use of data-driven applications," *Sensors* **20**(3), 733 (2020).

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

[3] Shokravi, H., Shokravi, H., Bakhary, N., Heidarrezaei, M., Rahimian Koloor, S. S., and Petru, M., "Application of the subspace-based methods in health monitoring of civil structures: A systematic review and meta-analysis," *Applied Sciences* **10**(10), 3607 (2020).

[4] Döhler, M., Hille, F., Mevel, L., and Rücker, W., "Structural health monitoring with statistical methods during progressive damage test of s101 bridge," *Engineering Structures* **69**, 183–193 (2014).

[5] Guo, J., Jiao, J., Fujita, K., and Takewaki, I., "Damage identification for frame structures using vision-based measurement," *Engineering Structures* **199**, 109634 (2019).

[6] Allahdadian, S., Döhler, M., Ventura, C., and Mevel, L., "Towards robust statistical damage localization via model-based sensitivity clustering," *Mechanical Systems and Signal Processing* **134**, 106341 (2019).

[7] Doebling, S. W., Farrar, C. R., Prime, M. B., et al., "A summary review of vibration-based damage identification methods," *Shock and vibration digest* **30**(2), 91–105 (1998).

[8] Feng, D. and Feng, M. Q., "Identification of structural stiffness and excitation forces in time domain using noncontact vision-based displacement measurement," *Journal of Sound and Vibration* **406**, 15–28 (2017).

[9] Chen, J. G., Wadhwa, N., Durand, F., Freeman, W. T., and Buyukozturk, O., "Developments with motion magnification for structural modal identification through camera video," in [*Dynamics of Civil Structures, Volume 2*], 49–57, Springer (2015).

[10] Magalhães, F., Caetano, E., and Cunha, Á., "Challenges in the application of stochastic modal identification methods to a cable-stayed bridge," *Journal of Bridge Engineering* **12**(6), 746–754 (2007).

[11] Carden, E. P. and Brownjohn, J. M., "Arma modelled time-series classification for structural health monitoring of civil infrastructure," *Mechanical systems and signal processing* **22**(2), 295–314 (2008).

[12] Peeters, B. and De Roeck, G., "Reference-based stochastic subspace identification for output-only modal analysis," *Mechanical Systems and Signal Processing* **13**(6), 855–878 (1999).

[13] Döhler, M. and Mevel, L., "Fast multi-order computation of system matrices in subspace-based system identification," *Control Engineering Practice* **20**(9), 882–894 (2012).

[14] James III, G. H., Carne, T. G., and Lauffer, J. P., "The natural excitation technique (next) for modal parameter extraction from operating wind turbines," *NASA STI/Recon Technical Report N* **93**, 28603 (1993).

[15] Siringoringo, D. M. and Fujino, Y., "System identification of suspension bridge from ambient vibration response," *Engineering Structures* **30**(2), 462–477 (2008).

[16] Fan, W. and Qiao, P., "Vibration-based damage identification methods: a review and comparative study," *Structural health monitoring* **10**(1), 83–111 (2011).

[17] Ewins, D. J., [*Modal testing: theory, practice and application*], John Wiley & Sons (2009).

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

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

[20] Guo, J., Lu, S., Zhang, D., Zhang, C., et al., "Vision-based measurement for rotational speed by improving lucas–kanade template tracking algorithm," *Applied optics* **55**(25), 7186–7194 (2016).

[21] Javh, J., Slavič, J., and Boltežar, M., "The subpixel resolution of optical-flow-based modal analysis," *Mechanical Systems and Signal Processing* **88**, 89–99 (2017).

[22] Wadhwa, N., Rubinstein, M., Durand, F., and Freeman, W. T., "Phase-based video motion processing," *ACM Transactions on Graphics (TOG)* **32**(4), 1–10 (2013).

[23] Wang, W., Mottershead, J. E., Siebert, T., and Pipino, A., "Frequency response functions of shape features from full-field vibration measurements using digital image correlation," *Mechanical systems and signal processing* **28**, 333–347 (2012).

[24] Döhler, M. and Mevel, L., "Efficient multi-order uncertainty computation for stochastic subspace identification," *Mechanical Systems and Signal Processing* **38**(2), 346–366 (2013).

[25] Hartley, R., "A. zisserman multiple view geometry in computer vision," (2000).

[26] Ji, Y. and Chang, C., "Nontarget stereo vision technique for spatiotemporal response measurement of line-like structures," *Journal of engineering mechanics* **134**(6), 466–474 (2008).

[27] Freeman, W. T., Adelson, E. H., et al., "The design and use of steerable filters," *IEEE Transactions on Pattern analysis and machine intelligence* **13**(9), 891–906 (1991).

[28] Chan, S. H., Võ, D. T., and Nguyen, T. Q., "Subpixel motion estimation without interpolation," in [*2010 IEEE International Conference on Acoustics, Speech and Signal Processing*], 722–725, Citeseer (2010).

[29] Irani, M. and Peleg, S., "Improving resolution by image registration," *CVGIP: Graphical models and image processing* **53**(3), 231–239 (1991).

[30] Kumar, R., Cirrincione, G., Cirrincione, M., Andriollo, M., and Tortella, A., "Accurate fault diagnosis and classification scheme based on non-parametric, statistical-frequency features and neural networks," in [*2018 XIII International Conference on Electrical Machines (ICEM)*], 1747–1753, IEEE (2018).