

Article

A Vision-Based Procedure with Subpixel Resolution for Motion Estimation

Samira Azizi ^{1,2,*}, Kaveh Karami ² and Stefano Mariani ¹

¹ Department of Civil and Environmental Engineering, Politecnico di Milano, 20133 Milano, Italy; stefano.mariani@polimi.it

² Department of Civil Engineering, University of Kurdistan, Sanandaj 6617715175, Iran; ka.karami@uok.ac.ir

* Correspondence: samira.azizi@polimi.it

Abstract: Vision-based motion estimation for structural systems has attracted significant interest in recent years. As the design of robust algorithms to accurately estimate motion still represents a challenge, a multi-step framework is proposed to deal with both large and small motion amplitudes. The solution combines a stochastic search method for coarse-level measurements with a deterministic method for fine-level measurements. A population-based block matching approach, featuring adaptive search limit selection for robust estimation and a subsampled block strategy, is implemented to reduce the computational burden of integer pixel motion estimation. A Reduced-Error Gradient-based method is next adopted to achieve subpixel resolution accuracy. This hybrid Smart Block Matching with Reduced-Error Gradient (SBM-REG) approach therefore provides a powerful solution for motion estimation. By employing Complexity Pursuit, a blind source separation method for output-only modal analysis, structural mode shapes and vibration frequencies are finally extracted from video data. The method's efficiency and accuracy are assessed here against synthetic shifted patterns, a cantilever beam, and six-story laboratory tests.

Keywords: vision-based identification; modal analysis; block matching; population-based optimization



Academic Editor: Francesc Pozo

Received: 28 March 2025

Revised: 6 May 2025

Accepted: 13 May 2025

Published: 14 May 2025

Citation: Azizi, S.; Karami, K.; Mariani, S. A Vision-Based Procedure with Subpixel Resolution for Motion Estimation. *Sensors* **2025**, *25*, 3101. <https://doi.org/10.3390/s25103101>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Civil structures and infrastructures are continuously exposed to different hazards. Structural Health Monitoring (SHM) has therefore emerged as a necessary tool to predict possible structural failures and help prevent them [1]. As the local dynamic behavior of structures is sensitive to variations in their health, any deviation away from the normal or healthy condition may be exploited by the monitoring procedure. This is customarily performed by way of experimental modal analysis or operational modal analysis, which require the identification of relevant natural frequencies, mode shapes, and damping ratios of the monitored structure. Methods based on operational modal analysis process measurements collected through a network of sensors [2], which has to be deployed as widely as possible over the structure [3]. Even in the case of pavement monitoring, wireless sensor networks have been exploited to sense vibrations and estimate displacements/deformations induced by vehicles; see [4]. However, large-scale deployments are still a challenge, due to practical and cost-related constraints that can represent a barrier to the spread of related SHM strategies. Noncontact sensors, such as radar interferometers and laser vibrometers [5], are instead characterized by easier installation. On the other hand, they need a restricted measurement distance that prevents their use in case of large-scale structures [6].

In recent years, vision-based measurements have emerged as an effective method for full-field identification [7,8], damage detection [9,10], model updating [11], and motion estimation [12–14]. This novel method takes advantage of image data to obtain valuable insights into the structural behavior. By providing high-resolution response measurements, it also avoids any effect on the structural dynamics, due to, e.g., the weight of traditional sensing systems to be mounted on the structure, and also reduces the expenses related to the sensor network.

Block matching (BM) is one of the most effective motion estimation methods in digital video processing. With this method, images of a region of the structure under investigation (termed the block) are compared to a reference one. The structural motion is then obtained by minimizing the difference or by maximizing the cross-correlation between the current and the reference frames, within a predefined search area. The way this search for the best match is carried out represents a critical step of BM. A full search (FS) of all of the possible locations within the search area is highly effective but also highly expensive, featuring a computational complexity of $O(p^2)$, where p is the size of the search window. For real-time applications, to reduce the mentioned computational cost, algorithms like three-step search [15], diamond search [16], adaptive road pattern search [17], and fast full search [18] have been proposed. They all start with an initial evaluation at a few search points, to next refine the results based on the least block distortion. These methods often rely on fixed patterns and predefined limits of search, which can lead to errors in real-world cases where distortion varies unpredictably.

BM can be seen as an optimization problem, to be solved by approaches like differential evolution [19], genetic algorithms [20], or a population-based method [21], like Particle Swarm Optimization (PSO) [22], that improve the search efficiency and accuracy. Specifically, PSO has been shown to be very effective in solving the issue of distortion, linked to local minima of the objective function to be minimized. Thanks to the improvements provided by these algorithms, BM has been used also for micro/nano systems, aiming at displacement measurement for nano-positioning stages [23,24]. Its high precision in tracking motions has made it the preferred choice within vision-based measurement methods, also offering flexibility in handling complex motions of the structures [25,26]. It is now widely adopted in video coding for motion estimation and compensation [27,28], and to analyze temporal redundancy between frames, enabling video compression in standards such as H.264 and HEVC [29]. BM is also used in image registration to align images from different sources, particularly in fields like medical imaging [30] and remote sensing [31,32].

In relation to structural engineering, in [33], the displacements of a bridge were obtained by placing two LEDs on it, used as targets in the matching process. In [34], displacements were obtained by way of BM by maximizing the correlation when the target is artificial or part of the structure. Other methods were also proposed to extract features from the images that are invariant to brightness intensity and rotation [35–37], and finally lead to displacement time histories related to different collected frames. These methods are considered feature-based, and provide sparse measurements. In this direction, distinct regions of a six-story laboratory structure were studied in [38] and features were extracted [39]; subsequently, the Kanade–Lucas tracking algorithm [40] was exploited to obtain the displacements.

Other BM-based methodologies to estimate motions are digital image correlation [41–43], up-sampled cross-correlation, orientation code matching [44], and edge-enhanced matching [45]. BM techniques typically determine displacements at an integer pixel resolution, as the pixel represents the smallest unit in an image. Displacements at a subpixel resolution can be obtained by ways of curve fitting [46], Newton Raphson with an interpolation estimation [47,48], and gradient-based optical flow (GOF) [24–28]. The

original optical flow technique was based on variational approaches, to obtain motion by solving equations within which assumptions regarding constant brightness patterns [24] or local phase [29,30] were made. In [49], it was demonstrated that the local phase of an image, obtained through quadratic filters [50], represents the motion more robustly than the intensity. By way of this, in [51], mode shapes were obtained from videos of vibrating structures, and they were subsequently employed to identify [52] and locate damage [53]. Motion estimation in a kind of hierarchical or multi-resolution framework, consisting of an integer pixel stage later refined at the subpixel scale, was originally introduced in [46] and expanded in [47], using a modified Taylor approximation. In [54], this two-stage strategy for targetless structures was studied by combining simple BM and phase-based optical flow in the horizontal direction. In [55], subspace identification was applied in different scenarios for healthy and simulated damage structures, with targets on the structural surface. As phase-based optical flow requires velocity-tuned filters proportional to the motion and a texture of the structural component to provide a one-dimensional motion estimation, an additional processing stage looks necessary to extract motion.

The aim of this work is instead to design an enhanced two-dimensional motion estimation framework for structural analysis, regardless of the vibration amplitude. The proposed solution exploits only the intensity for the motion estimation, within a multi-stage frame that integrates PSO-based BM at the pixel level and GOF at the subpixel one. The algorithm, overall, consists of four main steps in a coarse-to-fine framework: (1) Pre-evaluation of zero-motion pixels, to identify static blocks characterized by subpixel motion only and locally avoid the coarse-level estimation with PSO. (2) For the remaining locations, use of an efficient PSO-based BM and two adaptive strategies related to search limits, in place of the customary constant search limits, and pixel-subsampling blocks, in place of the fixed block size. (3) Use of GOF to achieve a fine-level motion estimation starting from that obtained at the previous stage. (4) Reduction of the estimation errors related to GOF and caused by noise in the case of small motions, through an error cancelation step. The proposed framework allows for attacking problems characterized by displacements either featuring a large or small amplitude, and provides robust solutions with lower computational costs than state-of-the-art alternate solutions. In summary, the proposed solution features the following novelties: (i) A hybrid automated coarse-to-fine algorithm adopted for two-dimensional structural motion estimation, without the need for specific targets to be placed on the structural surface. This framework combines the strengths of BM and gradient-based methods, further introducing enhancements to reduce errors and improve the robustness and efficiency of the estimation procedure; (ii) a blind modal analysis of the estimated motion using Complexity Pursuit (CP), for an in-depth assessment of the displacement accuracy aimed at modal identification.

The performance of the offered procedure is claimed to be independent of the motion magnitude, as is going to be validated on video data related to a synthetic shifted pattern, a lab-scale beam, and a six-story structure.

2. Smart BM and Reduced-Error Gradient Method

The algorithm for motion estimation is illustrated in the following. The workflow of the proposed Smart BM with Reduced-Error Gradient (SBM-REG) motion estimation is shown in Figure 1. To this aim, the two-dimensional structural motion (d_u, d_v) is partitioned into the integer pixel part ($\Delta u, \Delta v$) and the subpixel part ($\delta u, \delta v$).

When the structural response is characterized by large motions, adaptive search limits appear necessary for a reliable estimation of the motion itself. First, within the BM strategy, the integer pixel motion is obtained using a stochastic search method, namely, PSO, for an efficient estimation without a pixel-by-pixel full search. An adaptive selection of the

search area and a subsampling strategy for blocks of pixels, which is referred to as a Smart BM solution, lead to an accurate and robust motion estimation with low computational costs. Once the current frame is shifted by the estimated integer pixel motion $(\Delta u, \Delta v)$, just the subpixel motion $(\delta u, \delta v)$ still has to be assessed. The latter is estimated by way of a GOF method, which is effective for guaranteeing accuracy in cases of small displacements; see [56]. Due to the said accuracy of the GOF method, the procedure does not require any interpolation between the intensity of the pixels. Although GOF is characterized by computational efficiency and simplicity, inaccuracies in the gradient computation can lead to biases in the estimation [40]. If accurate subpixel motion tracking is relevant, account must be taken of the fact that the bias is linearly proportional to the motion, so that the error increases with the magnitude of the actual shift [56]. By leveraging this relationship, an error cancelation step has been implemented to refine the GOF estimate, within an REG procedure. By next applying a blind source separation methodology on the extracted motion, modal parameters like modal shapes and frequencies can be estimated. This framework effectively balances the accuracy and computational efficiency, proving to be suitable for problems characterized by either large or small structural motions.

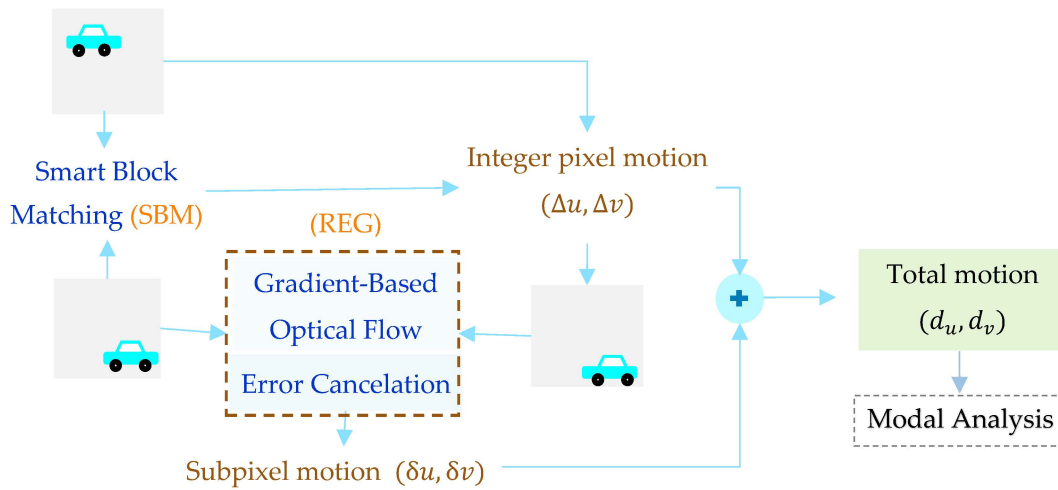


Figure 1. Workflow of the proposed method.

The remainder of this section is structured as follows: Section 2.1 provides details of the PSO-based BM method and of the updating strategy, while Section 2.2 introduces the GOF method along with the error cancelation step.

2.1. Enhanced Population-Based BM for Integer Pixel Motion Estimation

2.1.1. Block Matching

BM is used to quantify the motion between a reference frame $\mathcal{J}_0$ and the current frame $\mathcal{J}_c$ of a video, under the assumption that the intensity of each pixel does not change significantly during motion. A schematic diagram of the process of BM measurement is shown in Figure 2. In it, the video is subdivided into frames and each frame is divided on its own into blocks, with the goal of tracking how each block moves from one frame to the next. A search area consisting of $p \times p$ pixels around the corresponding block of $w \times w$ pixels in the reference frame is selected to find the region that most closely resembles the initial block in the current frame. The similarity between the two blocks is determined through a criterion based on cross-correlation CC, according to

$$CC(\Delta u, \Delta v) = \frac{\sum_{m=1}^w \sum_{n=1}^w [I_0(x_m, y_n) - \bar{I}_0] [I_c(\hat{x}_m, \hat{y}_n) - \bar{I}_c]}{\sqrt{\sum_{m=1}^w \sum_{n=1}^w [I_c(\hat{x}_m, \hat{y}_n) - \bar{I}_c]^2} \sqrt{\sum_{m=1}^w \sum_{n=1}^w [I_0(x_m, y_n) - \bar{I}_0]^2}} \quad (1)$$

where

$$\hat{x}_m = x_m + du, \quad \hat{y}_n = y_n + dv \quad (2)$$

and (x_m, y_n) are the coordinates in a local block system, where the origin $(0, 0)$ is the top-left corner; $I_0(x_m, y_n)$ and $I_c(\hat{x}_m, \hat{y}_n)$ represent the gray (intensity) levels of the blocks in the reference and current frames, respectively; $\bar{I}_0$ and $\bar{I}_c$ are the mean values of the intensity of the block pixels, again in the reference and current frames. The displacement $(\Delta u, \Delta v)$ that maximizes CC represents the estimated motion of the block at the coarse scale. This process is repeated for all of the blocks, allowing motion estimation for the entire structure.

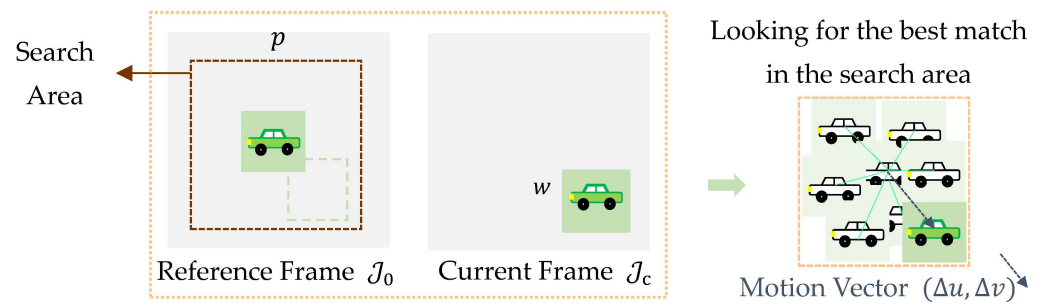


Figure 2. Sketch of the BM-based process.

In traditional BM-based procedures, the search area and the block size are predetermined, and the best match is obtained by searching for all the possible locations within the search area, thereby resulting in being computationally intensive. Additionally, the fixed search limits become unreliable in the case of motions larger than the predefined limits, which therefore become case-dependent features. In SHM, cases are characterized by limited insight into the vibration amplitude at different locations, and a flexible framework for motion estimation appears to be essential. To tackle these challenge, in this work, an advanced motion estimation procedure is proposed. Starting from a pre-evaluation of the similarity at block location $CC(0, 0)$ to identify the zero-pixel motion based on a defined threshold th , static blocks and blocks featuring subpixel motion only are identified. For the remaining blocks, PSO is exploited to maximize CC given by Equation (1) together with two updating strategies to adjust the search area and perform sub-sampling of pixels within the blocks. The pre-evaluation step effectively prevents unnecessary analysis within the regions experiencing subpixel motion, while, in the case of larger motions, the framework dynamically expands the search space to ensure accuracy. Since CC ranges from 0 to 1 and images are typically affected by noise, the aforementioned threshold distinguishing the static blocks from the others is customarily set to 0.9 to point to blocks experiencing subpixel motion only. The chosen threshold represents a balance between computational efficiency and estimation reliability and does not critically affect the overall performance of the framework across different image noise levels or textures, as supported by our validation experiments.

2.1.2. Particle Swarm Optimization

In the proposed solution, for each block, PSO departs from an initial guess of the solution to the maximization of CC, to iteratively refine it [57]. Regarding the notation, a particle is a candidate solution within a D -dimensional space based on the number of parameters to be optimized, while the swarm $\mathbf{Z}$ is formed by n_0 particles and provides the positions of all the particles in the said D -dimensional space. As the solution is sought in

terms of the horizontal and vertical components of the motion $(\Delta u, \Delta v)$, in the considered two-dimensional setting, $D = 2$ one obtains the following:

$$\mathbf{Z} = \left\{ (z_{i1}, z_{i2})_{i=1}, (z_{i1}, z_{i2})_{i=2}, \dots, (z_{i1}, z_{i2})_{i=n_0} \right\} \quad (3)$$

The initial locations $\mathbf{z}_i(0)$ of the particles are randomly deployed with a uniform distribution within an interval upper and lower bounded by Ub and Lb , respectively. At the k -th iteration of the algorithm, $k = 1, \dots, Mi$, to converge towards the optimal solution, the position of the i -th particle is updated according to

$$\mathbf{z}_i(k+1) = \mathbf{z}_i(k) + \mathbf{v}_i(k+1) \quad (4)$$

where $\mathbf{v}_i$ is the velocity of the same particle, given by

$$\mathbf{v}_i(k+1) = \omega \mathbf{v}_i(k) + c_1(\mathbf{pb}_i - \mathbf{z}_i(k))r_1 + c_2(\mathbf{gb} - \mathbf{z}_i(k))r_2 \quad (5)$$

In Equation (5), $\mathbf{pb}_i$ are the coordinates of the i -th particle related to the best solution obtained up to the k -th iteration; $\mathbf{gb}$ is the overall best solution obtained by the swarm up to the same k -th iteration; ω is the inertia weight; c_1 and c_2 are two acceleration constants, which take values in the range [14] and control the motion of the particle within the current iteration, respectively, in the direction of its personal best $\mathbf{pb}_i$ and of the global best $\mathbf{gb}$; r_1 and r_2 are two random variables, showing a uniform distribution in [0 1]. In this work, a linear decreasing function across the iterations has been adopted for ω , moving from $\omega = 0.9$ to $\omega = 0.4$; $c_1 = 2$ and $c_2 = 1$ have been selected instead [58]. The aforementioned best solutions provided by the particle and by the entire swarm are computed by way of the CC metric.

Thanks to pixel subsampling, the computational cost of the optimization process is reduced, as only some pixels in each block are used to compute the cross-correlation between the different frames. Moving from a block of size w , at the l -th algorithmic level of subsampling, $l = 1, 2, \dots, S$, the selected pixels are spaced by $q = 2^{S-l}$ additional pixels that are not allowed for in the cross-correlation computation; see Figure 3. With this subsampling strategy, the (subsampled) cross-correlation SCC is computed according to the following:

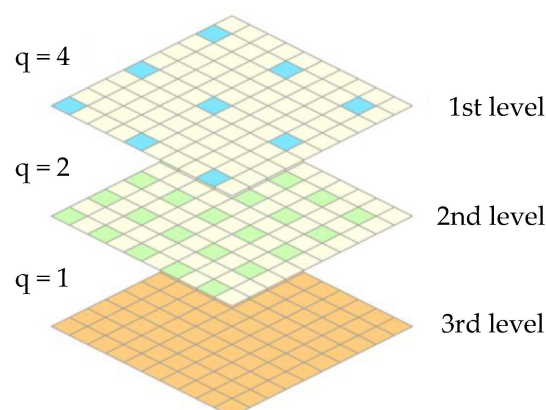


Figure 3. Sketch of three-level pixel subsampling of a block.

$$SCC(\Delta u, \Delta v) = \frac{\sum_{b=0, m=1+bq}^{Ns} \sum_{e=0, n=1+eq}^{Ns} [I_0(x_m, y_n) - \bar{I}_0] [I_c(\hat{x}_m, \hat{y}_n) - \bar{I}_c]}{\sqrt{\sum_{b=0, m=1+bq}^{Ns} \sum_{e=0, n=1+eq}^{Ns} [I_c(\hat{x}_m, \hat{y}_n) - \bar{I}_c]^2} \sqrt{\sum_{b=0, m=1+bq}^{Ns} \sum_{e=0, n=1+eq}^{Ns} [I_0(x_m, y_n) - \bar{I}_0]^2}} \quad (6)$$

As l increases, more particles are accounted for in the evaluation of SCC, until the point that they are all considered at the S -th level. At each iteration, if SCC exceeds the given threshold th , the algorithm is stopped and the **gb** solution is considered as the final response to use as the initial guess for subpixel motion estimation. If the algorithm fails to find a solution attaining the threshold at the S -th level, the search boundaries are updated by increasing the size of the search space (by 20% in the current investigation) for the next generation G . In the results reported in this paper, $S = 3$ and $G = 2$ have been adopted.

The flowchart of the entire SBM-REG procedure is shown in Figure 4.

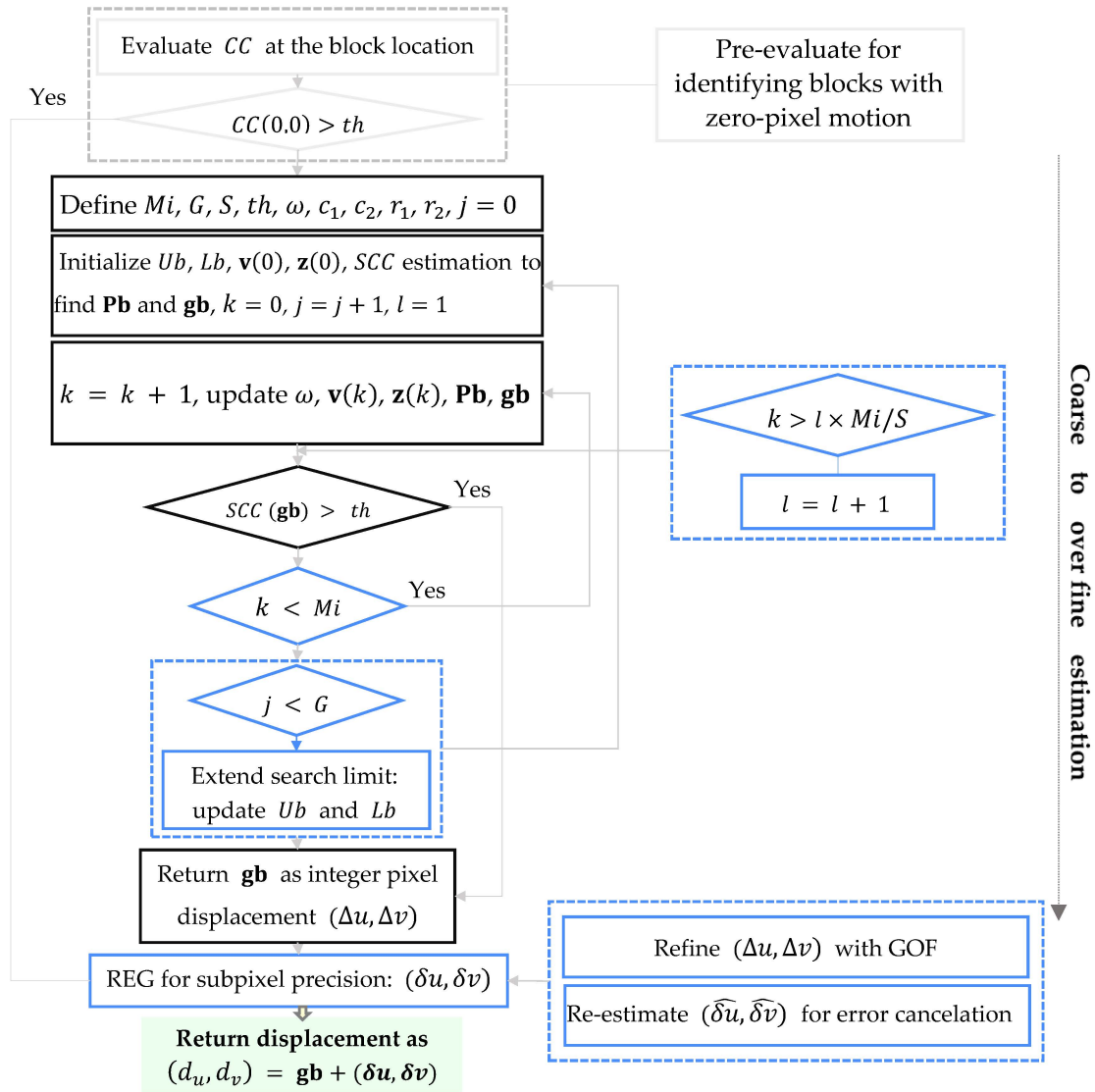


Figure 4. Flowchart of the SBM-REG method.

2.2. Enhanced Gradient-Based Solution with Error Cancellation for Subpixel Motion Estimation

Once the coarse-level estimation $(\Delta u, \Delta v)$ has been obtained, the block in the current frame is shifted accordingly toward the same block in the reference frame, and the estimation of subpixel motion $(\delta u, \delta v)$ can be started on the fine scale. GOF [56] works by assuming a constant local intensity of a point in the blocks of the reference and current images, that is

$$I_0(x + \delta u, y + \delta v) = I_{cr}(x, y) \quad (7)$$

Due to the subpixel displacement amplitude, I_0 is expanded in the Taylor series up to the first order in $(\delta u, \delta v)$, so that

$$I_0(x + \delta u, y + \delta v) = I_0(x, y) + \delta u \cdot I_{0x}(x, y) + \delta v \cdot I_{0y}(x, y) \quad (8)$$

where I_{0x} and I_{0y} represent the spatial gradients of I_0 . Equation (8) can be solved to obtain the estimation $(\hat{\delta u}, \hat{\delta v})$, using the least squares technique in the following form:

$$\begin{bmatrix} \hat{\delta u} \\ \hat{\delta v} \end{bmatrix} = \begin{bmatrix} \sum_{x=1}^W \sum_{y=1}^W (I_{0x})^2 & \sum_{x=1}^W \sum_{y=1}^W (I_{0x} \cdot I_{0y}) \\ \sum_{x=1}^W \sum_{y=1}^W (I_{0x} \cdot I_{0y}) & \sum_{x=1}^W \sum_{y=1}^W (I_{0y})^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{x=1}^W \sum_{y=1}^W (I_0 - I_{cr}) I_{0x} \\ \sum_{x=1}^W \sum_{y=1}^W (I_0 - I_{cr}) I_{0y} \end{bmatrix} \quad (9)$$

in which all of the terms are computed at coordinates (x, y) .

GOF is known to provide errors in the estimates, due to the gradient of image intensity, which increase with the magnitude of the motion or shift in the images [59]. Although the focus here is on subpixel motion estimation, an error cancellation procedure, termed REG, is adopted to obtain the fine-scale estimation. The solution $(\hat{\delta u}, \hat{\delta v})$ provided by Equation (9) is considered a linear transformation of the actual one $(\delta u, \delta v)$, according to

$$\begin{bmatrix} \hat{\delta u} \\ \hat{\delta v} \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{bmatrix} \begin{bmatrix} \delta u \\ \delta v \end{bmatrix} \quad (10)$$

where the parameters $\alpha_i, i = 1, 2, 3, 4$, have to be set.

As the gradient estimation is obtained from the reference image, all of the possible shifts between different frames and the reference frame share the same parameters in Equation (10). By exploiting this assumption, the values of the parameters α_i can be estimated with additional synthetic shifts of the current frame; see Figure 5. Such additional ± 1 -pixel shiftings have a direction opposite to that of the estimated values $(\hat{\delta u}, \hat{\delta v})$. They allow for the estimation of the enhanced $(\delta u, \delta v)$ motion by shifting the image horizontally, vertically, and, finally, both horizontally and vertically, so that a slight redundancy in the equations is achieved in conjunction with Equation (10), to solve for $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ and $(\delta u, \delta v)$.

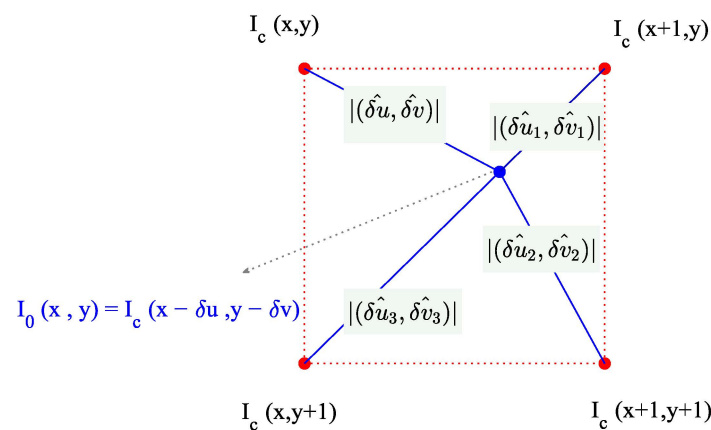


Figure 5. Additional subpixel shifts adopted for error cancelation purposes.

By integrating REG into the proposed motion estimation procedure, in the Results Section, it is shown that the accuracy and reliability of the obtained measures are enhanced across the entire field, and the overall performance of the structural analysis is thereby improved.

3. Blind Modal Analysis with Complexity Pursuit

The video of the vibrating structure is assumed to be decomposed into T frames. As each frame is subdivided into N blocks, and the SBM-REG procedure is used to obtain the motion between the reference and the current frames for each block.

If the motions of all the blocks are collected in the matrix $\mathbf{D} \in \mathbb{R}^{N \times T}$, an output-only modal analysis like blind source separation (BSS) [60,61] can be used to extract the relevant independent modal components. The structural motion $\mathbf{D}$ is accordingly expressed as a linear combination of the modal responses as

$$\mathbf{D}(\mathbf{X}, t) = \boldsymbol{\Phi}(\mathbf{X})\mathbf{q}(t) = \sum_{m=1}^n \boldsymbol{\varphi}_m(\mathbf{X})q_m(t) \quad (11)$$

where $t = 1, 2, \dots, T$ is a time-like variable; $\mathbf{X}$ represents the block location, as defined by the coordinates of the center of the block; n is the number of excited modes; $\boldsymbol{\Phi} \in \mathbb{R}^{N \times n}$ is the matrix collecting the vibration modes; $\mathbf{q} \in \mathbb{R}^{n \times T}$ are the modal responses. By using CP as a BSS technique, mode shapes and modal responses are obtained [62]; by next applying a Fast Fourier Transform (FFT) on the modal responses, the corresponding vibration frequencies are obtained.

CP leverages the temporal predictability of signals. It looks for a de-mixing (row) vector $\mathbf{w}_m$ according to

$$\mathbf{s}_m(t) = \mathbf{w}_m \mathbf{D}(t) \quad (12)$$

and such that the recovered component $\mathbf{s}_m$ possess a temporal structure simpler than the observed mixtures. The solution moves from the temporal predictability of a candidate signal $\mathbf{s}_m(t)$, measured through the following contrast function:

$$F(\mathbf{s}_m) = \log \left(\frac{V(\mathbf{s}_m)}{U(\mathbf{s}_m)} \right) \quad (13)$$

where $V(\mathbf{s}_m)$ captures the overall variability of $\mathbf{s}_m$ while $U(\mathbf{s}_m)$ captures the local smoothness of $\mathbf{s}_m$, respectively defined as

$$\begin{aligned} V(\mathbf{s}_m) &= \sum_{t=1}^T (\mathbf{s}_m(t) - \bar{\mathbf{s}}_m(t))^2 \\ U(\mathbf{s}_m) &= \sum_{t=1}^T (\mathbf{s}_m(t) - \hat{\mathbf{s}}_m(t))^2 \end{aligned} \quad (14)$$

and $\bar{\mathbf{s}}_m$ and $\hat{\mathbf{s}}_m$ are the moving averages:

$$\begin{aligned} \bar{\mathbf{s}}_m(t) &= \lambda_L \bar{\mathbf{s}}_m(t-1) + (1 - \lambda_L) \mathbf{s}_m(t-1) \\ \hat{\mathbf{s}}_m(t) &= \lambda_S \hat{\mathbf{s}}_m(t-1) + (1 - \lambda_S) \mathbf{s}_m(t-1) \end{aligned} \quad (15)$$

where λ_L and λ_S are parameters related to the long-term and short-term half-life.

By way of Equations (12), (14), and (15), Equation (13) can be written:

$$F(\mathbf{w}_m, \mathbf{D}) = \log \left(\frac{V(\mathbf{w}_m, \mathbf{D})}{U(\mathbf{w}_m, \mathbf{D})} \right) = \log \left(\frac{\mathbf{w}_m \bar{\mathbf{R}} \mathbf{w}_m^T}{\mathbf{w}_m \hat{\mathbf{R}} \mathbf{w}_m^T} \right) \quad (16)$$

where $\bar{\mathbf{R}}$ and $\hat{\mathbf{R}}$ are the covariance matrices related to the long-term and short-term variations of the displacement vector, respectively. In relation to the displacement matrix $\mathbf{D}(t)$, CP allows one to solve the problem as an optimization one to determine the de-mixing vector $\mathbf{w}_m$ by maximizing the temporal predictability contrast function F . For additional technical details, readers are referred to [62].

In what follows, the capability of the proposed method is assessed against synthetic patterns and laboratory test setups.

4. Experiments

4.1. Synthetic Shifted Patterns

Sample 7 of the DIC challenge proposed in [63] is considered as a first benchmark for the proposed solution. It consists of 12 speckle images featuring rigid-body subpixel translations ranging between 0 and 1 pixel, with a step of 0.1 pixel, in both the horizontal and vertical directions. The images have a size of 325×487 pixels, with contrast 100 and the noise variance 0.66; see also [64]. The first frame of this sample, featuring 54 blocks of 31×31 pixels, is shown in Figure 6.

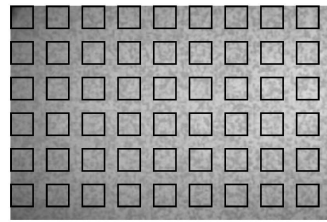


Figure 6. Synthetic shifted patterns: initial frame of sample 7 of the DIC challenge, with reported blocks for subpixel motion estimation.

Figure 7 collects results in terms of the cross-correlation CC, used to recognize the zero-motion at the block locations, for all of the frames. As expected, at lower subpixel displacements, the CC coefficient between the reference and the shifted images exhibits larger values. By setting a threshold to $th = 0.9$, six frames meet the condition for which the coarse estimation step is unnecessary and the REG-based solution can be directly adopted. Nevertheless, to enable a comparative analysis of GOF and REG methods, a threshold $th = 0.8$ has instead been selected.

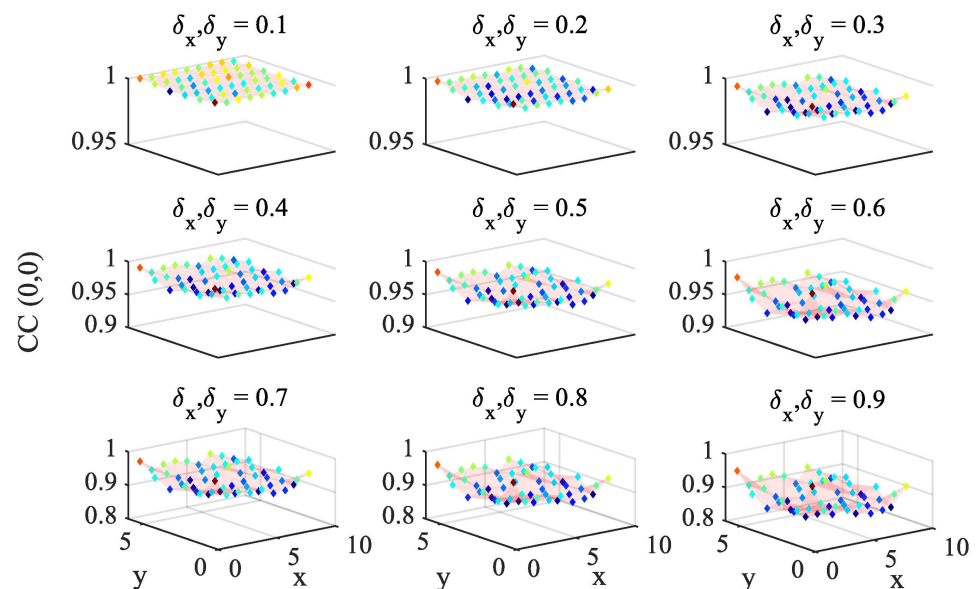


Figure 7. Synthetic shifted patterns: pre-evaluation of zero-motion cross-correlation CC at different block locations for different subpixel displacements.

To assess the accuracy of motion estimation, an error is computed as $\left(\frac{1}{a} \sum_{i=1}^a \Delta_i\right) - \Delta_{applied}$, where $\Delta_{applied}$ is the imposed subpixel displacement while $\Delta_i, i = 1, \dots, a$, are the estimated local values obtained with GOF at the different locations. This error measure is reported in

Figure 8 for the horizontal and vertical components, and for all the steps of the applied shift. The GOF-based error value is shown to be negative for shifts in the range between 0 and 0.6 pixels, indicating that the method underestimates the imposed subpixel shift in both directions, with the error increasing for larger shifts. By applying the bias compensation procedure detailed in Section 2.2, the error in all the solutions is significantly reduced, even by a factor of ten for larger shifts.

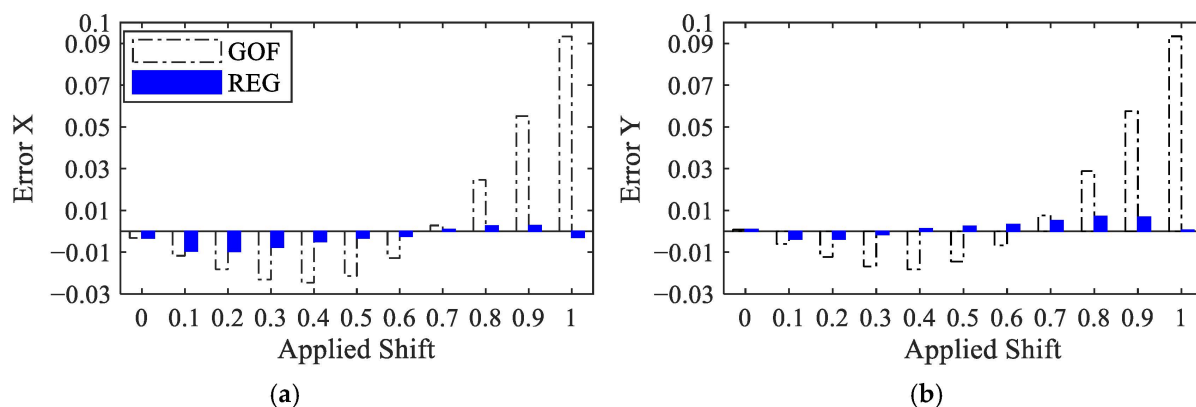


Figure 8. Synthetic shifted patterns: values of the estimation error provided by the GOF and REG methods, relevant to (a) horizontal and (b) vertical shifts (values in pixels).

It is notable that the estimation error is influenced by the spatial frequency of the frames, that is, by the rate of change of the intensity of pixels in the images. Additionally, different shifts in the two in-plane directions can affect the results and the relevant accuracy. Nevertheless, the objective of this study is only the assessment of the performance of the REG method in compensating the errors of the gradient method caused by noise (random bias) and by the gradient interpolator (systematic bias).

Next, to assess the performance of the entire procedure, an 8-bit speckle pattern with a size of 200×200 pixels has been created as an initial frame [65], as shown in Figure 9. The subsequent frames showing horizontal motions have been obtained by applying displacements ranging from 5 to 50 pixels. A block of 31×31 pixels in the center of the frame is used to estimate the motion, with $M_i = 21$, 3 particles; $th = 0.92$; initial search limits set to 30 pixels. Figure 10 shows how the SCC score changes with the iterations: for applied shifts smaller than 30 pixels, the estimation is completed in approximately six iterations at the first algorithmic level; for larger motions, a second generation is instead necessary to extend the search limits, and accurately estimate the motion.

Table 1 shows the estimated values of motion at the coarse, fine, and over-fine stages, along with the corresponding error index (EI) computed as the absolute difference between real and estimated values, divided by the real value. The SBM (coarse) estimation provides the integer part of the applied shift. As the estimations display, in all the cases with a one-pixel error, the selected threshold stops the algorithm properly. In fact, the same estimations are refined in the subsequent stages to attain a maximum error index of around 0.8%.

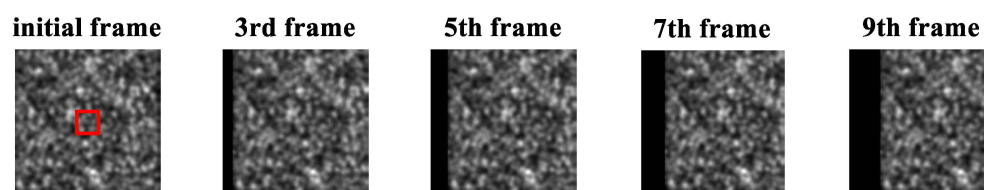


Figure 9. Synthetic shifted patterns: initial speckle pattern, and shifted frames with horizontal displacements for the SBM-REG performance evaluation.

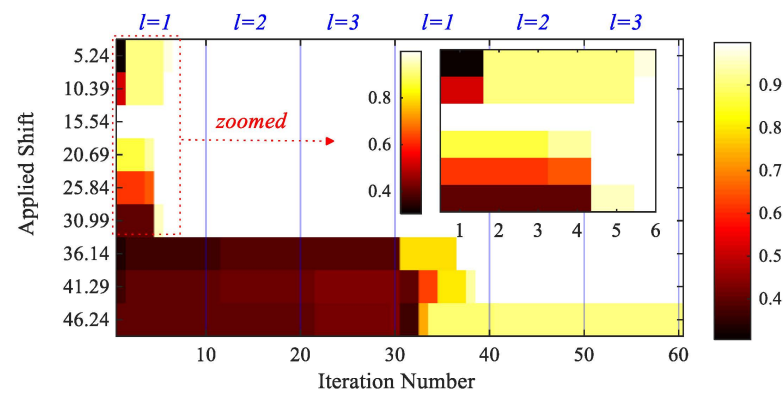


Figure 10. Synthetic shifted patterns: convergence of SCC scores across iterations, for different applied shifts.

Table 1. Synthetic shifted patterns: comparison of shift estimation accuracies at the different procedure stages.

Applied Shift	5.24	10.39	15.54	20.69	25.84	30.99	36.14	41.29	46.44
SBM (coarse)	6	10	16	21	26	30	36	41	45
EI %	14.5	3.75	2.9	1.49	0.6	3.19	0.38	0.7	3.1
SBM-GOF (fine)	5.1584	10.4525	15.4765	20.6465	25.8196	31.1679	36.1450	41.3357	46.9772
EI %	1.67	0.6	0.41	0.21	0.08	0.57	0.013	0.1	1.16
SBM-REG (over fine)	5.2494	10.3964	15.5454	20.6853	25.8456	31	36.1450	41.3003	46.8029
EI %	0.18	0.006	0.034	0.022	0.021	0.032	0.013	0.0249	0.78
FSF/SBM-REG calculation points	5.83	13.35	85	22.02	27.03	25.86	-	-	-

Even if not reported in detail for all the solutions, the ratio between the computation points of the FS and SBM-REG procedures shown at the bottom of Table 1 also highlights the computational efficiency of the offered method.

4.2. Cantilever Beam

The video of a vibrating beam (see [66]) is now considered. This video was captured by a camera with a frame rate of 30 fps, and the beam was initially deflected to induce free vibrations. Figure 11 shows the beam and the experimental setup aiming to simultaneously handle visual (camera-based) and laser measurements. The accessible video was reconstructed to display the beam vibrations related to its first mode, using phase-based motion magnification within the frequency range of 1–3 Hz.

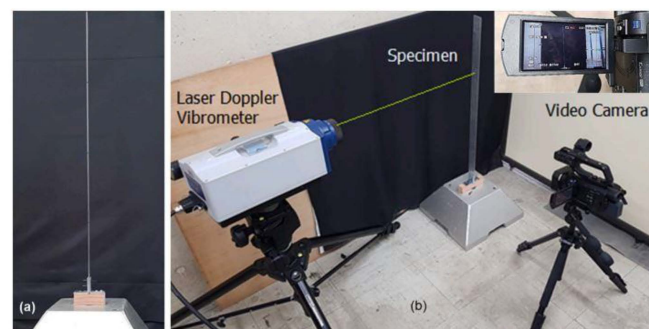


Figure 11. Cantilever beam: (a) initial configuration; (b) experimental setup featuring a video camera and a laser Doppler vibrometer (adapted from [66]).

The proposed algorithm has been adopted to extract the beam motion from the video, along the entire beam length. Figure 12a shows the reference frame of the video, whose size is 50×160 pixels. In the frame, 18 blocks of size 41×41 pixels are identified to extract the displacements. This procedure is at variance with the solution discussed in [66], characterized by 9 picks, as a denser solution is sought here. Since only the horizontal motion has to be extracted, the number of particles is set to 4, with $M_i = 42$ and $th = 0.9$. The initial search limit is set to 40 pixels. Figure 12b displays the extracted displacements for all the blocks and the 293 considered frames.

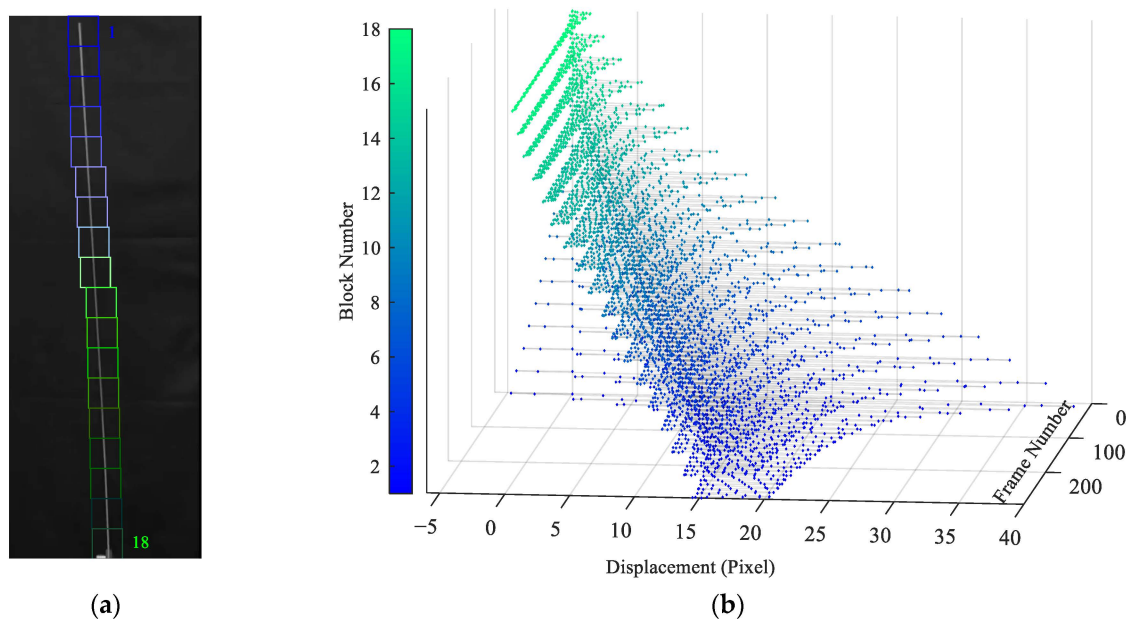


Figure 12. Cantilever beam: (a) reference frame captured by the video camera, and identified blocks along the longitudinal axis of the beam; (b) extracted motion of the 18 blocks in the horizontal direction.

By performing an FFT of the extracted displacements, see Figure 13, the excited vibration frequency of the beam is identified. The obtained frequency amounts to 2.363 Hz, which is in good agreement with the value of 2.43 Hz provided in [44]; the discrepancy between the two values is supposed to be related to the different framerates adopted in the two analyses.

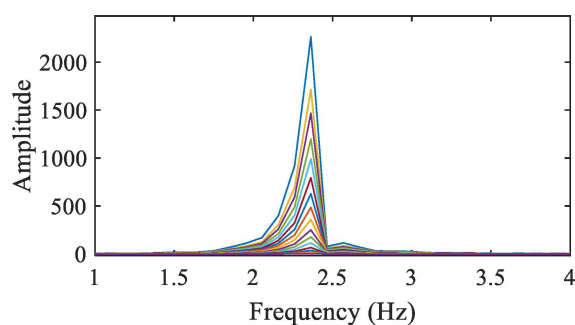


Figure 13. Cantilever beam: FFT of the extracted displacements at all block locations shown in Figure 12a. Different colors correspond to the block numbers 1 to 18 in Figure 12a.

The amplitude of the vibrations at the top of the beam is estimated to be about 40 pixels, while at the bottom of the beam, it is approximately 1 pixel. The SBM-REG procedure can therefore extract motions at different locations and with a noteworthy varying amplitude. To obtain insights into the benefit of the proposed multi-level procedure, Figure 14a shows a comparison between the extracted displacement time histories of the block at the bottom

of the beam: here, integer pixel only and subpixel precisions are, respectively, associated with the SBM and SBM-REG solutions. The relevant FFTs are reported in Figure 14b: it can be seen that the integer pixel solution does not display a frequency peak as clear as the subpixel one, which leads to an estimated value of 2.363 Hz exactly as attained at the tip of the beam, where the magnitude of vibrations is at its maximum.

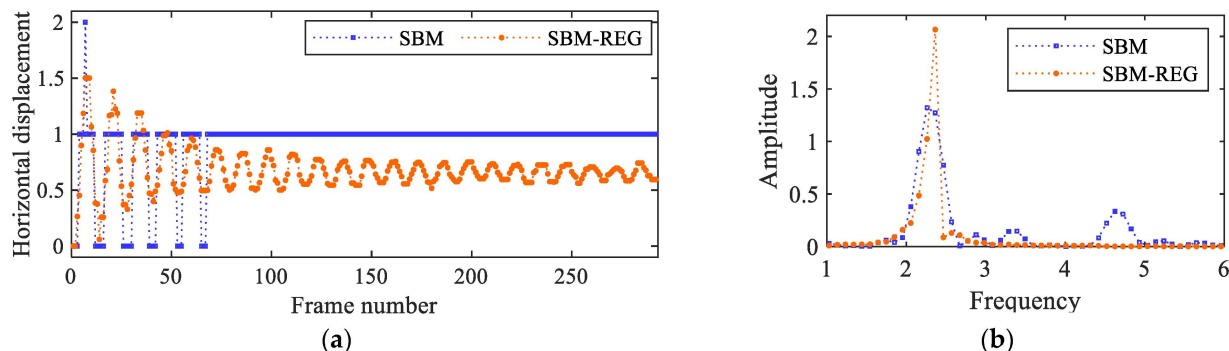


Figure 14. Cantilever beam: (a) extracted horizontal displacement time histories at the bottom of the beam, either allowing for subpixel precision or not; (b) corresponding FFTs of the signals.

The discussion so far has been focused on the estimation of the horizontal displacements but, as the beam looks slightly deflected in its reference state, the vertical motion with subpixel precision can be studied as well. Figure 15a shows the estimated vertical motion obtained by allowing for zero-pixel motion through either GOF or REG, at the top and bottom of the beam. As seen in Figure 15a, the vertical motion estimated by REG at the top of the beam (B_1 /REG) is about 0.6 pixels, while that provided by GOF (B_1 /GOF) is about 0.4 pixels. The FFTs of the two signals, as shown in Figure 15b, also agree even if the spectrum provided by REG displays a clearer peak. The estimated motion at the bottom of the beam has an obvious smaller amplitude, such that a peak of the FFT given by GOF (B_{18} /GOF) is not clear; the other way around, the estimation by REG (B_{18} /REG) again shows a clear frequency peak at the correct value of vibrations.

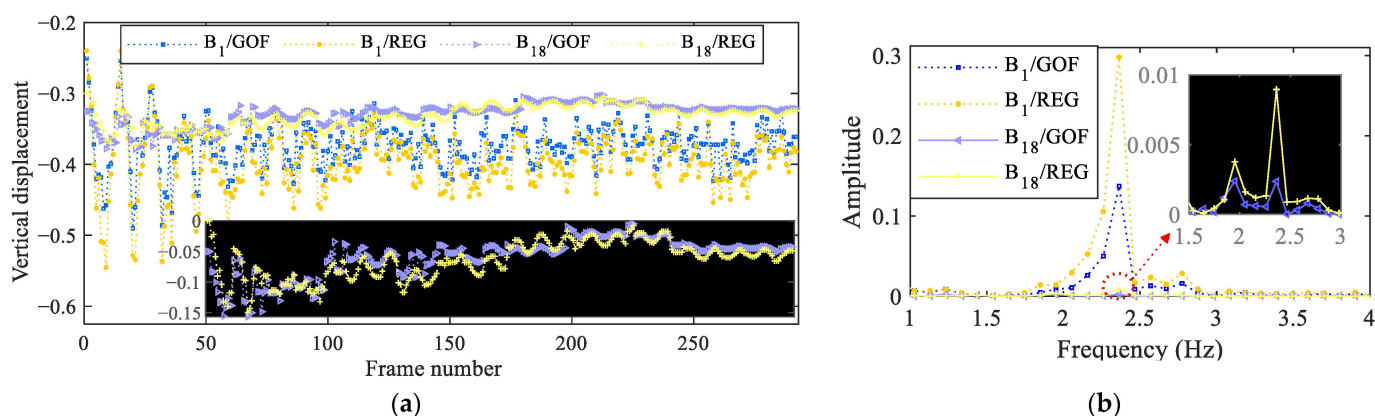


Figure 15. Cantilever beam: (a) extracted vertical displacement time histories at the bottom of the beam, either exploiting GOF or REG; (b) corresponding FFTs of the signals.

To further assess the capability of the proposed SBM-REG method in tracking the two-dimensional motion of structures, the same cantilever beam has been considered and the handled video recording has been fictitiously rotated by $\theta = 20^\circ$. The results can therefore be compared with the former ones in terms of the amplitude of oscillations. In this analysis, parameters have been set as follows: 8 particles, $M_i = 45$, $th = 0.9$. The

initial particle locations have been uniformly distributed around the relevant blocks, with a 30-pixel search limit.

Figure 16 depicts the outcome of the pre-evaluation step, at the top and bottom of the beam. It can be seen that, in almost all of the frames, the bottom block displays a value of CC larger than the threshold due to its subpixel motion. This means that extra processing linked to integer motion estimation with SBM is not necessary, and the subpixel part of the algorithm suffices to provide the entire solution. By way of this pre-evaluation step, the number of computational points is reduced by 99% for block 18 and by about 5% for block 1.

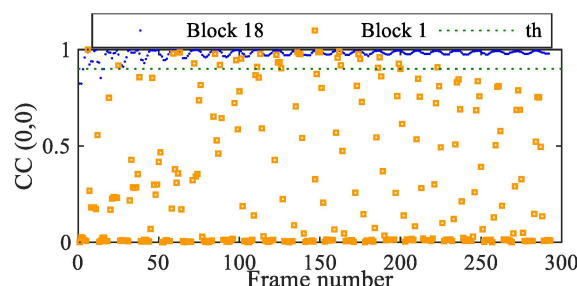


Figure 16. Cantilever beam: pre-evaluation of the cross-correlation at the top and bottom of the beam, for motion estimation with the rotated video.

The estimated horizontal and vertical components of the motion at the top and bottom of the rotated beam are compared with the projected original beam motion in Figure 17. While, at the top, the two solutions are in good agreement, at the bottom, the vertical components do not agree well. The reason for this discrepancy might be attributed to the applied rotation, which leads to a phase shift and therefore affects the results. The relevant FFTs, not reported here for brevity, provide, in all the cases, results in agreement with the former ones, with an estimated frequency of vibrations of 2.36 Hz at all the locations.

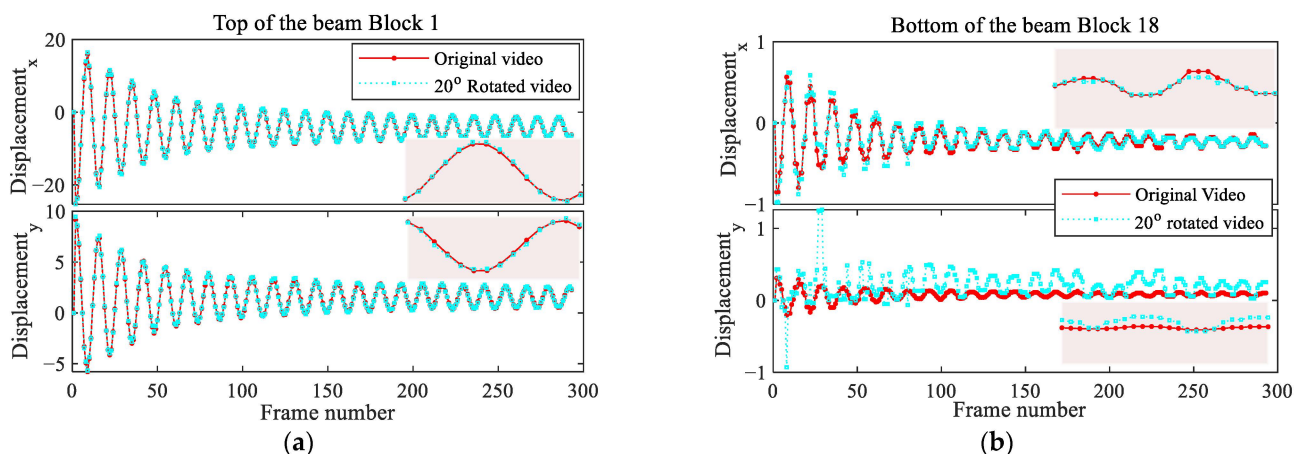


Figure 17. Cantilever beam: estimated motion at the (a) top and (b) bottom of the rotated beam.

Finally, it is worth mentioning that the proposed solution only accounts for the intensity of pixels for BM, whereas existing subpixel estimation methods are phase-based and require the use of a directional filter for the extraction of the phase. Accordingly, the phase-based process leads to one-dimensional results while the proposed SBM-REG solution can provide estimations of two-dimensional motions.

4.3. Six-Story Structure

The six-story building model considered in [38] is now employed to study the effectiveness of blind modal analysis of multi degree-of-freedom structures; CP is adopted to process the extracted displacement field. If compared with the previous beam case, this structure is characterized by a wider frequency spectrum in its dynamics. A band-limited white noise excitation signal was adopted as input in the experiment to excite the different structural modes of vibration. The original video recording lasted 130 s and was taken by a camera recording at 30 fps, but that made available consists only of 50 s of video, comprising 1450 frames.

Figure 18 shows five representative frames of this video, with the first frame on the left displaying the defined blocks, whose size is 41×41 pixels, defined to track the motion of each story. In terms of algorithmic parameters, the number of particles has been set to 6, with $M_i = 60$ and the other PSO parameters set as before.

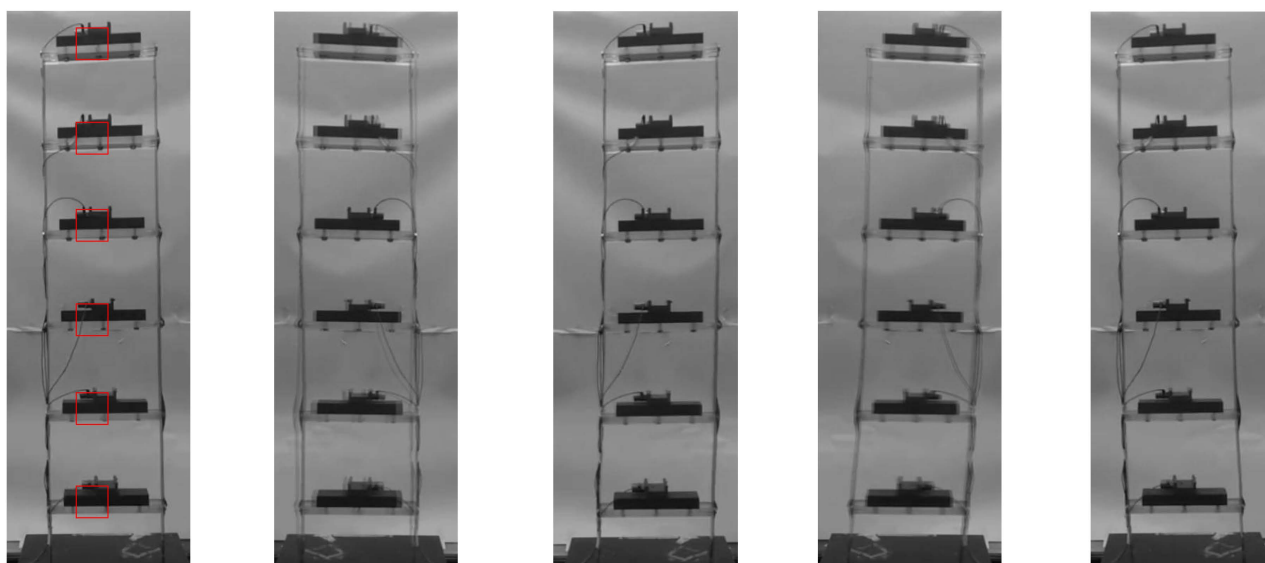


Figure 18. Six-story structure: five frames of the available video. The locations of the different blocks adopted in the analysis are shown in the first frame on the left.

Figure 19 reports the extracted displacements, measured again in pixels, for all of the stories. The maximum amplitude is clearly reported for the (top) 6th story, and amounts to 30 pixels, while the minimum one is related to the 1st story. The displacements obtained at the first level of the procedure have been obtained with a subset of 11×11 pixels, handled by PSO to converge within 14–17 iterations in all the solutions. To go more into the details of the efficacy of the proposed solution, the focus is next on the displacements of the bottom stories, which are characterized by smaller amplitudes and are therefore more in need of subpixel estimations. Figure 20 further illustrates the displacement time histories at the first, second, and third stories, by comparing the outcomes of the full SBM-REG procedure with subpixel resolution (left column), and those of the SBM procedure with integer pixel resolution only (right column). The qualitative improvement of the results is clearly visible, but it becomes even clearer by comparing the estimations of the vibration modes and frequencies.

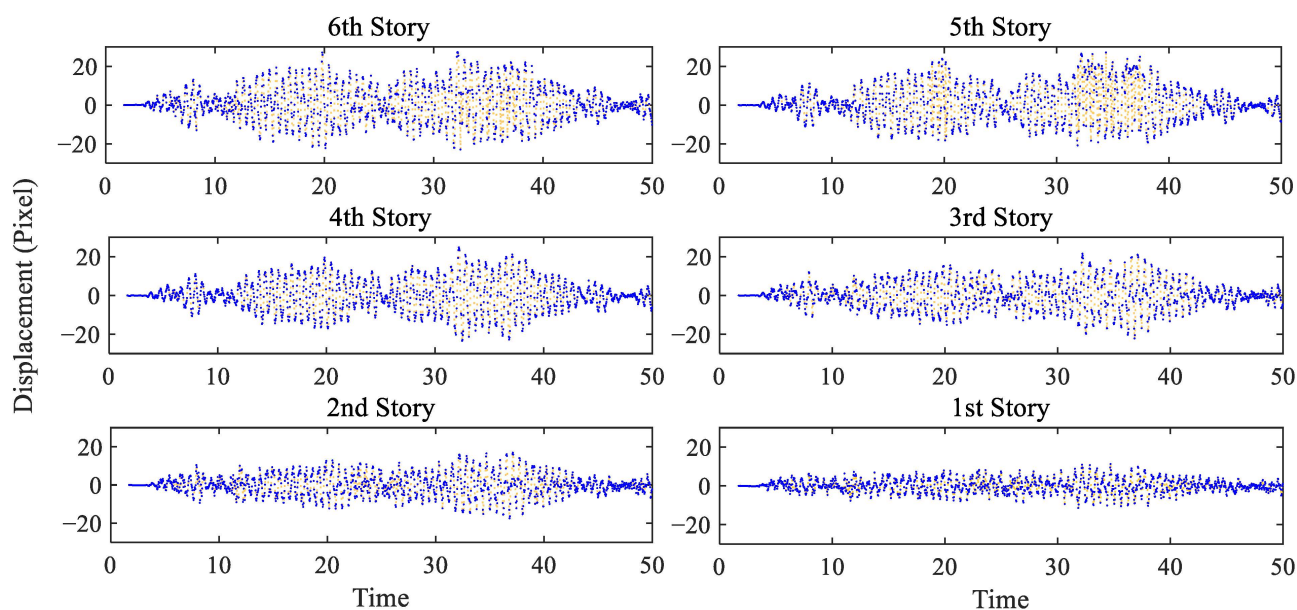


Figure 19. Six-story structure: story–displacement time histories extracted with the SBM-REG procedure.

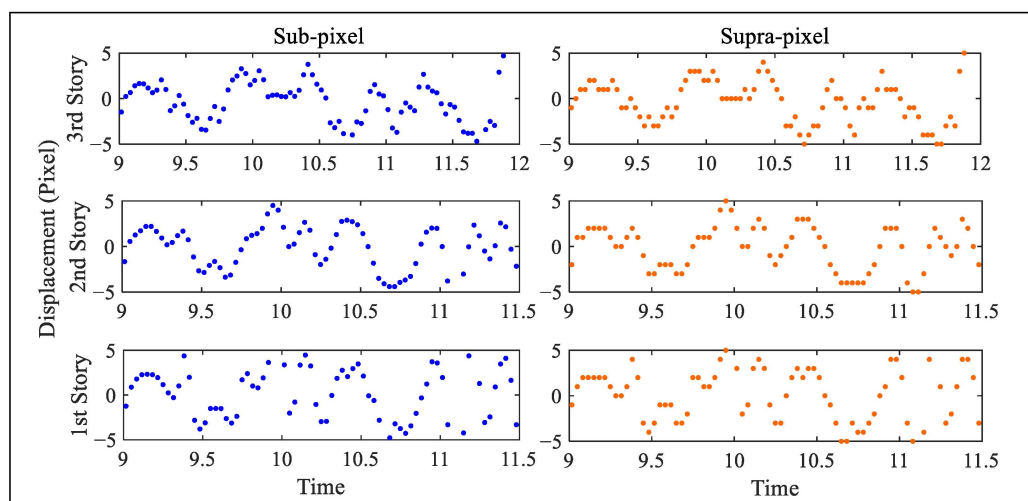


Figure 20. Six-story structure: comparison between the story–displacement time histories extracted with the (left) SBM-REG procedure, and the (right) SBM procedure.

Figure 21 collects the results in terms of the four fundamental modes identified by way of CP. The modes mentioned in [38] are also reported for a direct comparison with the current estimates. As shown in Figure 21a, the first three modes perfectly match the reference ones, while the fourth one shows a reasonably good agreement. The main reason for this difference could be related to the limited number of available frames, even if the solution is still considered accurate enough in overall terms. Figure 21b shows instead the FFTs of the extracted modal responses. The frequency peak values shown in the graphs are compared with the reference values in Table 2, to again show a close match between the values.

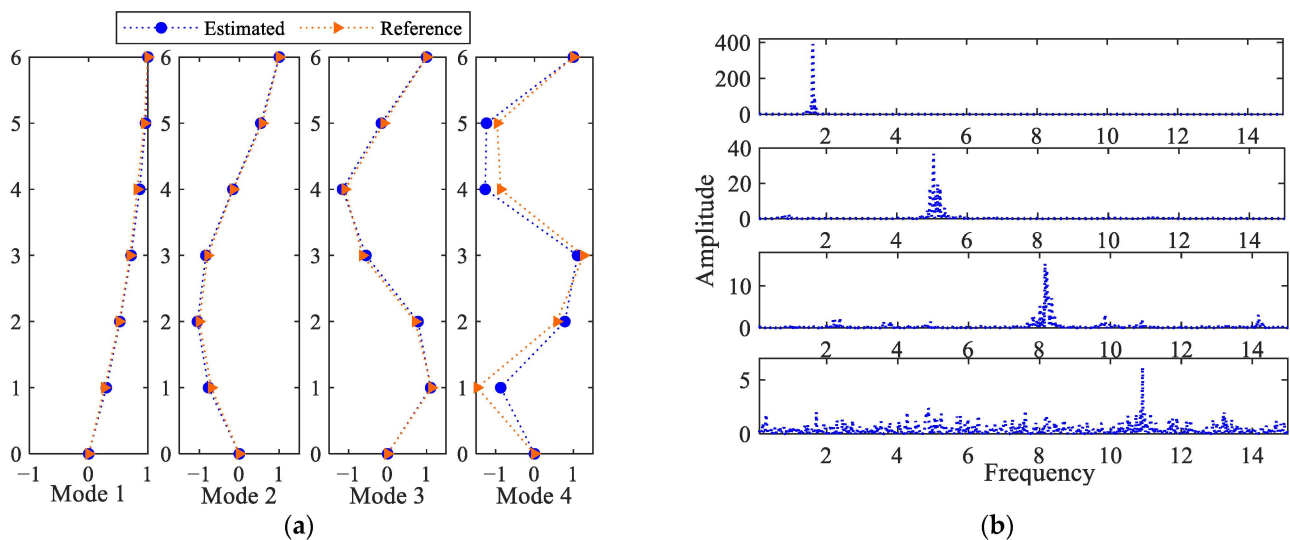


Figure 21. Six-story structure: (a) identified mode shapes, and (b) relevant FFTs.

Table 2. Six-story structure: identified modal frequencies, and comparison with the reference values reported in [38] (values in Hz).

	Mode 1	Mode 2	Mode 3	Mode 4
Reference	1.657	5.038	8.138	10.833
Estimated	1.644	5.045	8.167	10.91
Estimated-Full	1.631	5.045	8.208	10.64

In general, the main goal of vision-based estimation techniques is to obtain full-field information on the vibrations of the structure under investigation. So far, the motion of the six stories has been investigated to compare the results with the data available in the literature. Next, the aim is moved to the estimation of the motion of a column of the building model in a full-field manner. This structural element is divided into 19 blocks, each with a size of 31×31 pixels. The estimated motions of all the blocks, as obtained with the proposed SBM-REG procedure, are shown in Figure 22. The coding for the blocks along the longitudinal axis of the structural element is shown in Figure 23a.

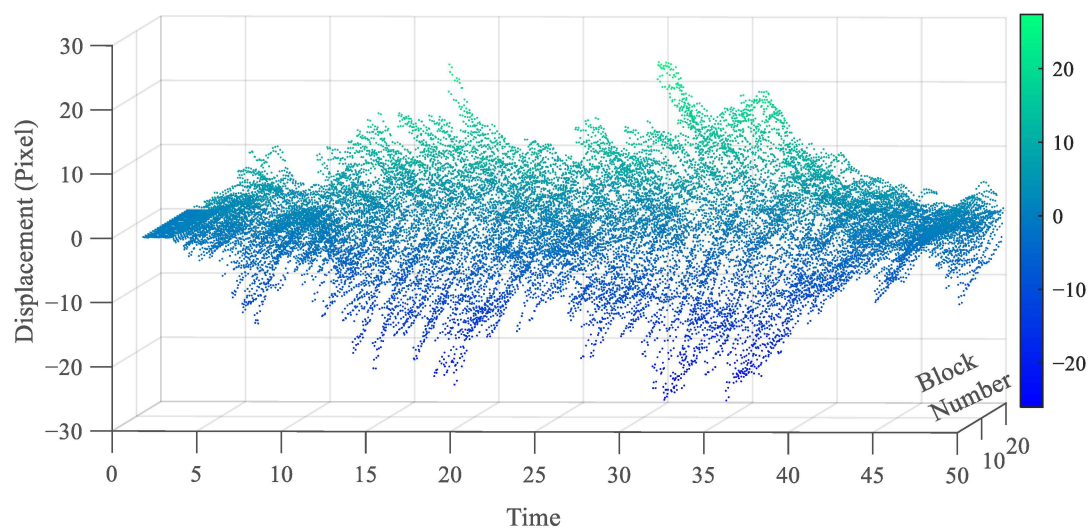


Figure 22. Six-story structure: full-field extracted motion of the 19 blocks in the horizontal direction.

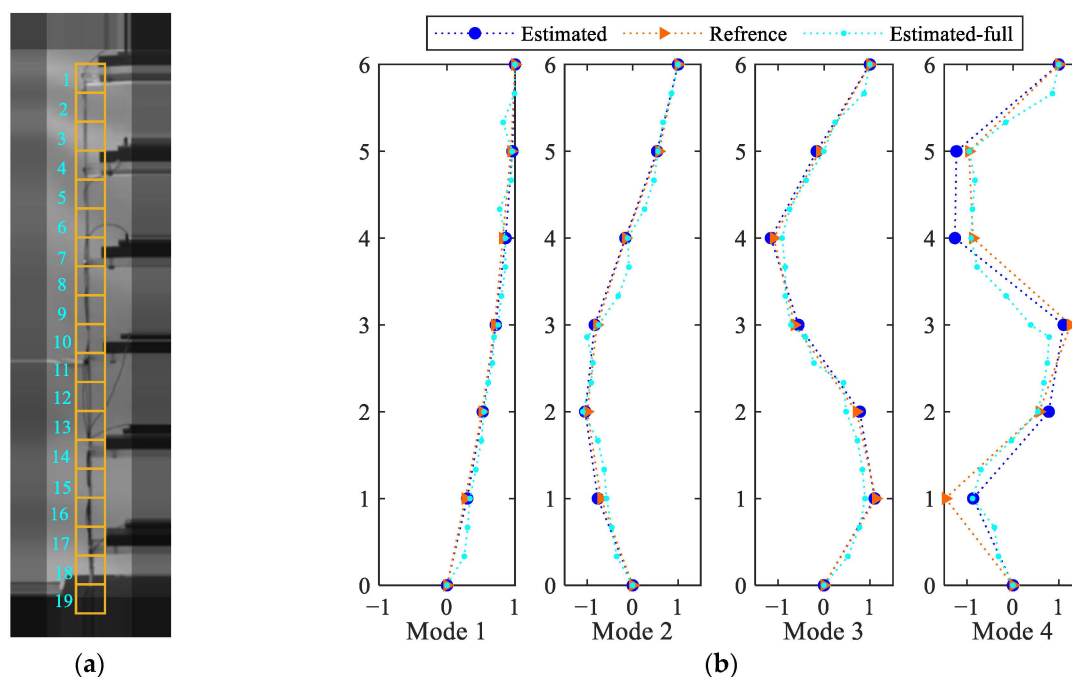


Figure 23. Six-story structure: (a) reference frame captured by the camera, and identified blocks along the left column of the structure; (b) identified mode shapes.

As the depicted results show a remarkable complexity and since modal analysis has to be performed for a number of measured locations that exceeds the number of excited modes, a Principal Component Analysis [51] is adopted first for order reduction purposes. Moving from the original motion matrix, which has size 19×1345 , by projecting the estimations onto the eigenvectors corresponding to eigenvalues that exceed 5% of the maximum one, and by finally exploiting CP, the full-field mode shapes can be estimated and are shown in Figure 23b. The results are compared in the charts with the previous estimations and the reference data, to stress that the extracted modes are in good agreement with those obtained with discrete measurements at story levels. The extracted frequencies, termed in Table 2 as ‘Estimated-Full’, match the reference frequency of the second mode of vibration well and show bounded differences for the other ones, though slightly worse than those obtained with blocks placed exactly at the different stories. Such a result points toward the issue of a proper design of the monitoring system, to assure that the deformation modes can be accurately recognized through the acquired measurements.

5. Conclusions

In this paper, a new video-based processing method for two-dimensional estimations of the motion of civil structures has been proposed. The main goal being system identification of modal parameters, the solution is foreseen to provide a means for future applications related to SHM. The offered Smart BM with Reduced-Error Gradient (SBM-REG) approach combines the strengths of BM and optical flow methods, to achieve accurate motion estimations independently of the vibration amplitude. By employing a coarse-to-fine-scale strategy, the framework integrates enhanced PSO-based BM for coarse-level motion estimation with gradient-based optical flow for fine-level estimations, in a kind a multi-resolution strategy.

Key enhancements to tackle the current challenges are as follows: the exploitation of a stochastic search method, to avoid becoming entrapped into local suboptimal solutions in terms of motion estimation; a strategy for the motion estimation characterized by adaptive search limits, to make the solution more robust; a pixel subsampling strategy with a rele-

vant cost function, to reduce the overall computational burden; and an error cancelation strategy, to remove the systematic error of the gradient-based method. Based on the reported validation outcomes, the proposed SBM-REG framework has been shown to extract displacement time histories with subpixel resolution based on the image intensity only, thereby making it possible to avoid the use of targets or patterns on the monitored structure. The integration of the Smart BM and of the Reduced-Error Gradient-based method provides a capability to deliver accurate motion estimations also for cases characterized by large and small amplitudes of vibration in different regions of the same structure. Hence, the proposed framework lays a foundation for advanced motion estimations and vision-based structural dynamic analysis based on blind modal identification techniques.

It should be noted that the reliability of gradient-based estimation may decrease in low-texture regions or under motion blur, making it necessary to detect unreliable regions as part of a full-field analysis. This limitation will be addressed in future developments by integrating a detection strategy to identify and manage such regions.

Future works will also focus on deformation measurements, which require a six-dimensional optimization frame within PSO, and a Newton–Raphson technique for adaptive refinement. Furthermore, the linear models used in REG will be used to estimate the motion of neighboring blocks and by next leading to damage detection, though a comparison between the estimated motion related to a healthy solution and the actual one observed in the video data, moving towards a digital twin to be adopted in SHM strategies.

Author Contributions: Conceptualization, S.A., S.M. and K.K.; Methodology, S.A.; Software, S.A.; Validation, S.A.; Formal analysis, S.A.; Writing—original draft, S.A.; Writing—review & editing, S.M.; Visualization, S.A.; Supervision, S.M. and K.K.; Project administration, S.M. and K.K.; Funding acquisition, S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Balageas, D.; Fritzen, C.-P.; Güemes, A. *Structural Health Monitoring*; John Wiley & Sons: Hoboken, NJ, USA, 2010; Volume 90.
- Prabatama, N.A.; Nguyen, M.L.; Hornyh, P.; Mariani, S.; Laheurte, J.M. Pavement Monitoring with a Wireless Sensor Network of MEMS Accelerometers. In Proceedings of the 2024 IEEE International Symposium on Measurements & Networking (M&N), Rome, Italy, 2–5 July 2024.
- Capellari, G.; Chatzi, E.; Mariani, S.; Azam, S.E. Optimal design of sensor networks for damage detection. *Procedia Eng.* **2017**, *199*, 1864–1869. [[CrossRef](#)]
- Prabatama, N.A.; Nguyen, M.L.; Hornyh, P.; Mariani, S.; Laheurte, J.-M. Zigbee-Based Wireless Sensor Network of MEMS Accelerometers for Pavement Monitoring. *Sensors* **2024**, *24*, 6487. [[CrossRef](#)] [[PubMed](#)]
- Tzortzinis, G.; Ai, C.; Breña, S.F.; Gerasimidis, S. Using 3D laser scanning for estimating the capacity of corroded steel bridge girders: Experiments, computations and analytical solutions. *Eng. Struct.* **2022**, *265*, 114407. [[CrossRef](#)]
- Lynch, J.P.; Loh, K.J. A summary review of wireless sensors and sensor networks for structural health monitoring. *Shock Vib. Dig.* **2006**, *38*, 91–130. [[CrossRef](#)]
- Feng, D.; Feng, M.Q. Experimental validation of cost-effective vision-based structural health monitoring. *Mech. Syst. Signal Process.* **2017**, *88*, 199–211. [[CrossRef](#)]
- Yang, Y.; Dorn, C. Affinity propagation clustering of full-field, high-spatial-dimensional measurements for robust output-only modal identification: A proof-of-concept study. *J. Sound Vib.* **2020**, *483*, 115473. [[CrossRef](#)]

9. Dworakowski, Z.; Kohut, P.; Gallina, A.; Holak, K.; Uhl, T. Vision-based algorithms for damage detection and localization in structural health monitoring. *Struct. Control Health Monit.* **2016**, *23*, 35–50. [\[CrossRef\]](#)
10. Azizi, S.; Karami, K.; Nagarajaiah, S. Developing a semi-active adjustable stiffness device using integrated damage tracking and adaptive stiffness mechanism. *Eng. Struct.* **2021**, *238*, 112036. [\[CrossRef\]](#)
11. Martini, A.; Tronci, E.M.; Feng, M.Q.; Leung, R.Y. A computer vision-based method for bridge model updating using displacement influence lines. *Eng. Struct.* **2022**, *259*, 114129. [\[CrossRef\]](#)
12. Yang, Y.; Sanchez, L.; Zhang, H.; Roeder, A.; Bowlan, J.; Crochet, J.; Farrar, C.; Mascareñas, D. Estimation of full-field, full-order experimental modal model of cable vibration from digital video measurements with physics-guided unsupervised machine learning and computer vision. *Struct. Control Health Monit.* **2019**, *26*, e2358. [\[CrossRef\]](#)
13. Bhowmick, S.; Nagarajaiah, S.; Lai, Z. Measurement of full-field displacement time history of a vibrating continuous edge from video. *Mech. Syst. Signal Process.* **2020**, *144*, 106847. [\[CrossRef\]](#)
14. Luo, L.; Feng, M.Q.; Wu, Z.Y. Robust vision sensor for multi-point displacement monitoring of bridges in the field. *Eng. Struct.* **2018**, *163*, 255–266. [\[CrossRef\]](#)
15. Li, R.; Zeng, B.; Liou, M.L. A new three-step search algorithm for block motion estimation. *IEEE Trans. Circuits Syst. Video Technol.* **1994**, *4*, 438–442.
16. Zhu, S.; Ma, K.-K. A new diamond search algorithm for fast block-matching motion estimation. *IEEE Trans. Image Process.* **2000**, *9*, 287–290. [\[CrossRef\]](#)
17. Biswas, B.; Mukherjee, R.; Chakrabarti, I. Efficient architecture of adaptive rood pattern search technique for fast motion estimation. *Microprocess. Microsyst.* **2015**, *39*, 200–209. [\[CrossRef\]](#)
18. Al-Najdawi, N.; Al-Najdawi, M.N.; Tedmori, S. Employing a novel cross-diamond search in a modified hierarchical search motion estimation algorithm for video compression. *Inf. Sci.* **2014**, *268*, 425–435. [\[CrossRef\]](#)
19. Cuevas, E.; Zaldívar, D.; Pérez-Cisneros, M.; Oliva, D. Block-matching algorithm based on differential evolution for motion estimation. *Eng. Appl. Artif. Intell.* **2013**, *26*, 488–498. [\[CrossRef\]](#)
20. Jin, H.; Bruck, H.A. Pointwise digital image correlation using genetic algorithms. *Exp. Tech.* **2005**, *29*, 36–39. [\[CrossRef\]](#)
21. Pandian, S.I.A.; Bala, G.J.; Anitha, J. A pattern based PSO approach for block matching in motion estimation. *Eng. Appl. Artif. Intell.* **2013**, *26*, 1811–1817. [\[CrossRef\]](#)
22. Sengar, S.S.; Mukhopadhyay, S. Motion segmentation-based surveillance video compression using adaptive particle swarm optimization. *Neural Comput. Appl.* **2020**, *32*, 11443–11457. [\[CrossRef\]](#)
23. Wu, H.; Zhang, X.; Gan, J.; Li, H.; Ge, P. Displacement measurement system for inverters using computer micro-vision. *Opt. Lasers Eng.* **2016**, *81*, 113–118. [\[CrossRef\]](#)
24. Kim, J.H.; Menq, C.-H. Visual Servo Control Achieving Nanometer Resolution in X-Y-Z. *IEEE Trans. Robot.* **2009**, *25*, 109–116. [\[CrossRef\]](#)
25. Clark, L.; Shirinzadeh, B.; Bhagat, U.; Smith, J. A Vision-based measurement algorithm for micro/nano manipulation. In Proceedings of the 2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Wollongong, Australia, 9–12 July 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 100–105.
26. Anis, Y.H.; Mills, J.K.; Cleghorn, W.L. Visual-servoing of a six-degree-of-freedom robotic manipulator for automated microassembly task execution. *J. Micro/Nanolithogr. MEMS MOEMS* **2008**, *7*, 033017.
27. Babu, D.V.; Subramanian, P.; Karthikeyan, C. Performance analysis of block matching algorithms for highly scalable video compression. In Proceedings of the 2006 International Symposium on Ad Hoc and Ubiquitous Computing, Mangalore, India, 20–23 December 2006; IEEE: Piscataway, NJ, USA, 2006; pp. 179–182.
28. Khawase, S.T.; Kamble, S.D.; Thakur, N.V.; Patharkar, A.S. An Overview of Block Matching Algorithms for Motion Vector Estimation. In Proceedings of the Second International Conference on Research in Intelligent and Computing in Engineering, Gopeshwar, India, 24–26 March 2017; pp. 217–222.
29. Kerfa, D.; Belbachir, M.F. Star diamond: An efficient algorithm for fast block matching motion estimation in H264/AVC video codec. *Multimed. Tools Appl.* **2016**, *75*, 3161–3175. [\[CrossRef\]](#)
30. Commowick, O.; Wiest-Daesslé, N.; Prima, S. Block-matching strategies for rigid registration of multimodal medical images. In Proceedings of the 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI), Barcelona, Spain, 2–5 May 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 700–703.
31. Wang, C.; Yin, Z.; Ma, X.; Yang, Z. SAR image despeckling based on block-matching and noise-referenced deep learning method. *Remote Sens.* **2022**, *14*, 931. [\[CrossRef\]](#)
32. Xu, N.; Ma, D.; Ren, G.; Huang, Y. BM-IQE: An image quality evaluator with block-matching for both real-life scenes and remote sensing scenes. *Sensors* **2020**, *20*, 3472. [\[CrossRef\]](#)
33. Wahbeh, A.M.; Caffrey, J.P.; Masri, S.F. A vision-based approach for the direct measurement of displacements in vibrating systems. *Smart Mater. Struct.* **2003**, *12*, 785. [\[CrossRef\]](#)

34. Feng, D.; Feng, M.Q.; Ozer, E.; Fukuda, Y. A vision-based sensor for noncontact structural displacement measurement. *Sensors* **2015**, *15*, 16557–16575. [[CrossRef](#)] [[PubMed](#)]
35. Mair, E.; Hager, G.D.; Burschka, D.; Suppa, M.; Hirzinger, G. Adaptive and generic corner detection based on the accelerated segment test. In Proceedings of the 11th European Conference on Computer Vision, Heraklion, Crete, Greece, 5–11 September 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 183–196.
36. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [[CrossRef](#)]
37. Vedaldi, A.; Fulkerson, B. VLFeat: An open and portable library of computer vision algorithms. In Proceedings of the 18th ACM International Conference on Multimedia, Firenze, Italy, 25–29 October 2010; pp. 1469–1472.
38. Yoon, H.; Elanwar, H.; Choi, H.; Golparvar-Fard, M.; Spencer, B.F., Jr. Target-free approach for vision-based structural system identification using consumer-grade cameras. *Struct. Control Health Monit.* **2016**, *23*, 1405–1416. [[CrossRef](#)]
39. Harris, C.; Stephens, M. A combined corner and edge detector. In Proceedings of the Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988.
40. Lucas, B.D.; Kanade, T. An iterative image registration technique with an application to stereo vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence, Vancouver, BC, Canada, 24–28 August 1981.
41. Kim, S.W.; Cheung, J.H.; Park, J.B.; Na, S.O. Image-based back analysis for tension estimation of suspension bridge hanger cables. *Struct. Control Health Monit.* **2020**, *27*, e2508. [[CrossRef](#)]
42. Bolognini, M.; Izzo, G.; Marchisotti, D.; Fagiano, L.; Limongelli, M.P.; Zappa, E. Vision-based modal analysis of built environment structures with multiple drones. *Autom. Constr.* **2022**, *143*, 104550. [[CrossRef](#)]
43. Gregorini, A.; Cattaneo, N.; Bortolotto, S.; Massa, S.; Bocciolone, M.F.; Zappa, E. Metrological issues in 3D reconstruction of an archaeological site with aerial photogrammetry. In Proceedings of the 2023 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Kuala Lumpur, Malaysia, 22–25 May 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.
44. Feng, D.; Feng, M.Q. Vision-based multipoint displacement measurement for structural health monitoring. *Struct. Control Health Monit.* **2016**, *23*, 876–890. [[CrossRef](#)]
45. Luo, L.; Feng, M.Q. Edge-enhanced matching for gradient-based computer vision displacement measurement. *Comput. Aided Civ. Infrastruct. Eng.* **2018**, *33*, 1019–1040. [[CrossRef](#)]
46. Pan, B.; Yu, L.; Zhang, Q. Review of single-camera stereo-digital image correlation techniques for full-field 3D shape and deformation measurement. *Sci. China Technol. Sci.* **2018**, *61*, 2–20. [[CrossRef](#)]
47. Xiao, P.; Wu, Z.; Christenson, R.; Lobo-Aguilar, S. Development of video analytics with template matching methods for using camera as sensor and application to highway bridge structural health monitoring. *J. Civ. Struct. Health Monit.* **2020**, *10*, 405–424. [[CrossRef](#)]
48. Feng, D.; Scarangelo, T.; Feng, M.Q.; Ye, Q. Cable tension force estimate using novel noncontact vision-based sensor. *Measurement* **2017**, *99*, 44–52. [[CrossRef](#)]
49. Fleet, D.J.; Jepson, A.D. Computation of component image velocity from local phase information. *Int. J. Comput. Vis.* **1990**, *5*, 77–104. [[CrossRef](#)]
50. Weldon, T.P.; Higgins, W.E.; Dunn, D.F. Efficient Gabor filter design for texture segmentation. *Pattern Recognit.* **1996**, *29*, 2005–2015. [[CrossRef](#)]
51. Yang, Y.; Dorn, C.; Mancini, T.; Talken, Z.; Kenyon, G.; Farrar, C.; Mascareñas, D. Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification. *Mech. Syst. Signal Process.* **2017**, *85*, 567–590. [[CrossRef](#)]
52. Yang, Y.; Dorn, C.; Mancini, T.; Talken, Z.; Theiler, J.; Kenyon, G.; Farrar, C.; Mascareñas, D. Reference-free detection of minute, non-visible, damage using full-field, high-resolution mode shapes output-only identified from digital videos of structures. *Struct. Health Monit.* **2018**, *17*, 514–531. [[CrossRef](#)]
53. Yang, Y.; Jung, H.K.; Dorn, C.; Park, G.; Farrar, C.; Mascareñas, D. Estimation of full-field dynamic strains from digital video measurements of output-only beam structures by video motion processing and modal superposition. *Struct. Control Health Monit.* **2019**, *26*, e2408. [[CrossRef](#)]
54. Luan, L.; Liu, Y.; Sun, H. Extracting high-precision full-field displacement from videos via pixel matching and optical flow. *J. Sound Vib.* **2023**, *565*, 117904. [[CrossRef](#)]
55. Merainani, B.; Xiong, B.; Baltazart, V.; Döhler, M.; Dumoulin, J.; Zhang, Q. Subspace-based modal identification and uncertainty quantification from video image flows. *J. Sound Vib.* **2024**, *569*, 117957. [[CrossRef](#)]
56. Davis, C.Q.; Freeman, D.M. Statistics of subpixel registration algorithms based on spatiotemporal gradients or block matching. *Opt. Eng.* **1998**, *37*, 1290–1298. [[CrossRef](#)]
57. Shi, Y.; Eberhart, R. A modified particle swarm optimizer. In Proceedings of the 1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No. 98TH8360), Anchorage, AK, USA, 4–9 May 1998; IEEE: Piscataway, NJ, USA, 1998; pp. 69–73.

58. Yang, X.; Zou, L.; Deng, W. Fatigue life prediction for welding components based on hybrid intelligent technique. *Mater. Sci. Eng. A* **2015**, *642*, 253–261. [[CrossRef](#)]
59. Brandt, J.W. Analysis of bias in gradient-based optical-flow estimation. In Proceedings of the 1994 28th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 31 October–2 November 1994; IEEE: Piscataway, NJ, USA, 1994; pp. 721–725.
60. Yang, Y.; Xie, R.; Li, M.; Cheng, W. A review on the application of blind source separation in vibration analysis of mechanical systems. *Measurement* **2024**, *227*, 114241. [[CrossRef](#)]
61. Sadhu, A.; Narasimhan, S.; Antoni, J. A review of output-only structural mode identification literature employing blind source separation methods. *Mech. Syst. Signal Process.* **2017**, *94*, 415–431. [[CrossRef](#)]
62. Yang, Y.; Nagarajaiah, S. Blind modal identification of output-only structures in time-domain based on complexity pursuit. *Earthq. Eng. Struct. Dyn.* **2013**, *42*, 1885–1905. [[CrossRef](#)]
63. Reu, P.L.; Toussaint, E.; Jones, E.; Bruck, H.A.; Iadicola, M.; Balcaen, R.; Turner, D.Z.; Siebert, T.; Lava, P.; Simonsen, M. DIC challenge: Developing images and guidelines for evaluating accuracy and resolution of 2D analyses. *Exp. Mech.* **2018**, *58*, 1067–1099. [[CrossRef](#)]
64. Zhao, J.; Pan, B. Smart DIC: User-independent, accurate and precise DIC measurement with self-adaptively selected optimal calculation parameters. *Mech. Syst. Signal Process.* **2025**, *222*, 111792. [[CrossRef](#)]
65. Chen, B.; Pan, B. Camera calibration using synthetic random speckle pattern and digital image correlation. *Opt. Lasers Eng.* **2020**, *126*, 105919. [[CrossRef](#)]
66. Siringoringo, D.M.; Wangchuk, S.; Fujino, Y. Noncontact operational modal analysis of light poles by vision-based motion-magnification method. *Eng. Struct.* **2021**, *244*, 112728. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.