infrastructures

MDPI

Article

Target-Free Stereo-Vision-Based Modal Identification of a Cantilever Beam Using Physics-Aware Deep-Feature Tracking

Sherbaz Khan [1,*], Afsar Ali [1], Awaiz Noor [2] and Li Hui [1]

[1] Department of Civil and Environmental Engineering, University of Louisiana at Lafayette, Lafayette, LA 70504, USA; afsar.ali1@louisiana.edu (A.A.); li.hui@louisiana.edu (L.H.)
[2] Department of Computing and Informatics, University of Louisiana at Lafayette, Lafayette, LA 70504, USA; awaiz.noor1@louisiana.edu
* Correspondence: sherbaz.khan1@louisiana.edu

Abstract

Reliable target-free vision-based modal identification remains challenging because accurate image correspondence does not necessarily guarantee physically reliable structural measurements. This study presents a Physics-Aware Adaptive Deep-Feature Tracking (PA-ADFT) framework that integrates learned feature extraction, stereo-temporal correspondences, physics-aware measurement validation, and triangulation to reconstruct three-dimensional camera-coordinate displacement trajectories for structural dynamic analysis. The framework was experimentally validated under controlled laboratory conditions through five repeated free-vibration tests on a single healthy aluminum cantilever beam at 10 mm tip-release amplitude, using synchronized stereo cameras and strain gauges. PA-ADFT maintained 7–9 accepted structural measurement points with 90–95% median tracking coverage across five experiments and accurately identified the first bending natural frequency with a median value of 7.02 Hz, compared to 7.03 Hz from strain measurements and 7.06 Hz from the finite-element model. A staged comparison showed that this frequency is already recovered by SuperPoint–LightGlue–Lucas–Kanade tracking; the subsequent geometric, temporal, and kinematic gates preserve track coverage and three-dimensional point identity. Calibrated stereo triangulation recovered the camera-coordinate trajectory in three dimensions, preserving the dominant first-mode bending behavior in the vertical displacement component. Vertical displacement was validated quantitatively against reference measurements; the depth component was reconstructed but not independently validated. Together with physics-aware measurement checks, this supports target-free stereo vision as a viable method for identifying first-mode bending in a single-specimen, single-amplitude laboratory setup.

Keywords: target-free stereo vision; structural health monitoring; modal identification; deep feature tracking; physics-aware measurement validation

Check for updates

Academic Editor: Joan R. Casas

Received: 30 July 2026
Revised: 30 August 2026
Accepted: 16 September 2026
Published: 20 September 2026

Copyright: © 2026 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.

1. Introduction

1.1. Motivation for Reliable Target-Free Structural Measurement

Structural health monitoring (SHM) increasingly relies on vibration-based techniques because modal characteristics provide sensitive indicators of changes in structural condition. The reliability of modal identification, however, depends fundamentally on the quality of the measured structural response [1–3]. Although conventional sensing technologies, including accelerometers, strain gauges, displacement transducers, and laser vibrometers,

provide highly accurate measurements, they are often limited by sparse spatial sampling, installation complexity, high deployment costs, or restricted accessibility, particularly for large-scale civil infrastructure [4,5]. These limitations have driven growing interest in vision-based sensing, which offers dense, non-contact measurements using relatively inexpensive systems and has emerged as a promising alternative for full-field structural monitoring.

Recent advances in digital imaging and computer vision have significantly expanded the capability of optical sensing, enabling target-free measurement of structural motion from naturally occurring surface features [6–8]. Nevertheless, reliable structural measurement remains considerably more challenging than reliable image tracking. Visually consistent feature trajectories do not necessarily correspond to the motion of persistent points, and image correspondence alone cannot guarantee physically meaningful displacement measurements. Feature instability, illumination variation, viewpoint changes, tracking drift, and geometric inconsistency can all introduce errors that propagate directly into the estimated modal properties [9–12]. Consequently, the value of vision-based SHM should be assessed not by image-matching performance alone, but by the physical reliability of the recovered structural response. While modern tracking algorithms have substantially improved feature detection and matching, these gains have not consistently translated into robust structural measurements under practical conditions.

1.2. Current State of Target-Free Vision-Based Modal Identification

Vision-based modal identification has evolved from target-based displacement measurement and digital image correlation (DIC) toward target-free techniques capable of recovering structural motion directly from naturally occurring surface features. While target-based systems and speckle-based DIC can achieve high measurement accuracy under controlled laboratory conditions, their reliance on surface preparation, artificial markers, stable illumination, and favorable imaging geometry limits their practicality for operational civil infrastructure [13,14]. Target-free approaches overcome many of these limitations but introduce a more fundamental measurement challenge: preserving the physical identity of measurement points when the sensing process depends on naturally occurring image features rather than predefined structural targets [15,16].

Accordingly, recent research has focused on improving the robustness of feature correspondence throughout the vibration measurement process [17–19]. Phase-based motion estimation, motion magnification, optical flow, and virtual point tracking have demonstrated that subtle structural vibrations can be recovered from ordinary video, while learned feature detectors and matching networks, including SuperPoint, SuperGlue, LoFTR, and LightGlue, have substantially improved feature matching under variations in viewpoint, illumination, and local appearance [20–24]. Nevertheless, these approaches primarily optimize visual correspondence rather than structural fidelity. Feature trajectories remain susceptible to drift, identity switching, background interference, and accumulated tracking errors, such that visually consistent correspondences do not necessarily represent physically meaningful structural motion [25,26].

A second challenge concerns the transition from image correspondence to engineering measurement. Stereo vision enables metric three-dimensional reconstruction and reduces many limitations associated with monocular imaging, but its accuracy depends on camera calibration, synchronization, epipolar consistency, triangulation quality, and the persistence of left–right correspondences [27,28]. Existing studies have generally improved feature detection, tracking, stereo matching, or displacement reconstruction as separate stages, with structural validation performed on displacement responses. Consequently, limited attention has been devoted to ensuring that geometric consistency, temporal continuity,

and structural plausibility are jointly satisfied throughout the measurement process before image trajectories are interpreted as modal responses.

Deep-learning feature matching has recently been used for structural vibration displacement measurement [17], and spatio-temporal refinement has been applied to stabilize markerless photogrammetric tracking [15]. These studies address components also used in PA-ADFT, but they do not provide staged decomposition showing which processing steps govern modal-frequency accuracy, spatial coverage, or preservation of tracked-point identity. Training-free vision foundation model approaches have also been proposed for structural displacement measurement [28]. However, the absence of metric calibration limits their ability to enforce physically interpretable acceptance criteria, such as kinematic displacement thresholds derived from the expected structural response.

PA-ADFT differs from this body of work by separating image-correspondence confidence from structural-measurement confidence and by using staged ablation of deep feature tracking, geometric and temporal gating, and kinematic identity gating to identify the contribution of each stage to measurement reliability. The remaining challenge, therefore, is the generation of displacement measurements whose geometric integrity, temporal consistency, and physical validity are established throughout the sensing process. Addressing this challenge requires a unified measurement framework that bridges robust computer vision and reliable structural sensing.

1.3. Research Gap and Scientific Hypothesis

These observations indicate that the central limitation of current target-free vision-based modal identification lies not in feature correspondence itself, but in the absence of an integrated measurement process that continuously enforces structural validity. Although substantial progress has been achieved in feature detection, matching, temporal tracking, stereo reconstruction, and displacement estimation, these components are generally developed and optimized independently. Consequently, measurement reliability depends on the combined performance of separate processing stages rather than on a unified framework.

This limitation is particularly important because target-free image features are not predefined structural measurement points. Instead, they are visual proxies whose suitability for structural sensing must be continuously verified throughout the measurement process. While modern learned feature matching algorithms substantially improve correspondence robustness, they do not guarantee that tracked features remain attached to the same material neighborhood or that the reconstructed displacement histories remain physically representative of structural behavior.

Accordingly, this study is based on the hypothesis that computer vision and structural measurement principles must be integrated within a unified physics-aware framework. Specifically, continuously enforcing feature persistence, calibrated stereo geometry, temporal consistency, and physics-based validation throughout the measurement process produces more reliable structural displacement measurements than conventional image correspondence alone. This hypothesis forms the foundation of the proposed Physics-Aware Adaptive Deep-Feature Tracking (PA-ADFT) framework.

1.4. Objectives and Contributions

The objective of this study is to develop and experimentally validate a physics-aware target-free stereo vision framework for reliable vibration-based modal identification. The proposed Physics-Aware Adaptive Deep-Feature Tracking (PA-ADFT) framework addresses the broader challenge of structural measurement reliability by integrating feature correspondence, temporal tracking, stereo geometry, and physics-based validation within a unified sensing framework.

The principal contributions of this work are summarized as follows:

1. A physics-aware target-free stereo framework is proposed that integrates learned feature matching, adaptive tracking, calibrated stereo reconstruction, and structural quality control for vibration-based structural measurement.
2. A staged internal ablation on the five validation trials shows that deep feature tracking recovers the first-mode frequency, geometric and temporal filtering restores track coverage, and a kinematic identity gate preserves three-dimensional correspondence integrity.
3. Experimental validation is performed using synchronized stereo cameras, strain gauges, and a finite-element model to verify the capability of the proposed framework for first-mode identification of an aluminum cantilever beam.

2. Experimental Program

2.1. Experimental Configuration

The experimental program was designed to generate synchronized vision and strain measurements for evaluating the proposed target-free stereo sensing framework under controlled laboratory conditions. As shown in Figure 1, the system comprised an aluminum cantilever beam instrumented with three uniaxial foil strain gauges (120 Ω resistance), a calibrated stereo camera system, a hardware synchronization circuit, and a common data acquisition platform. Figure 2 summarizes the experimental workflow, including system initialization, hardware synchronization, data acquisition, and transfer of synchronized measurements for subsequent processing. The same specimen, instrumentation, and acquisition settings were maintained throughout all experiments to ensure repeatable measurement conditions.

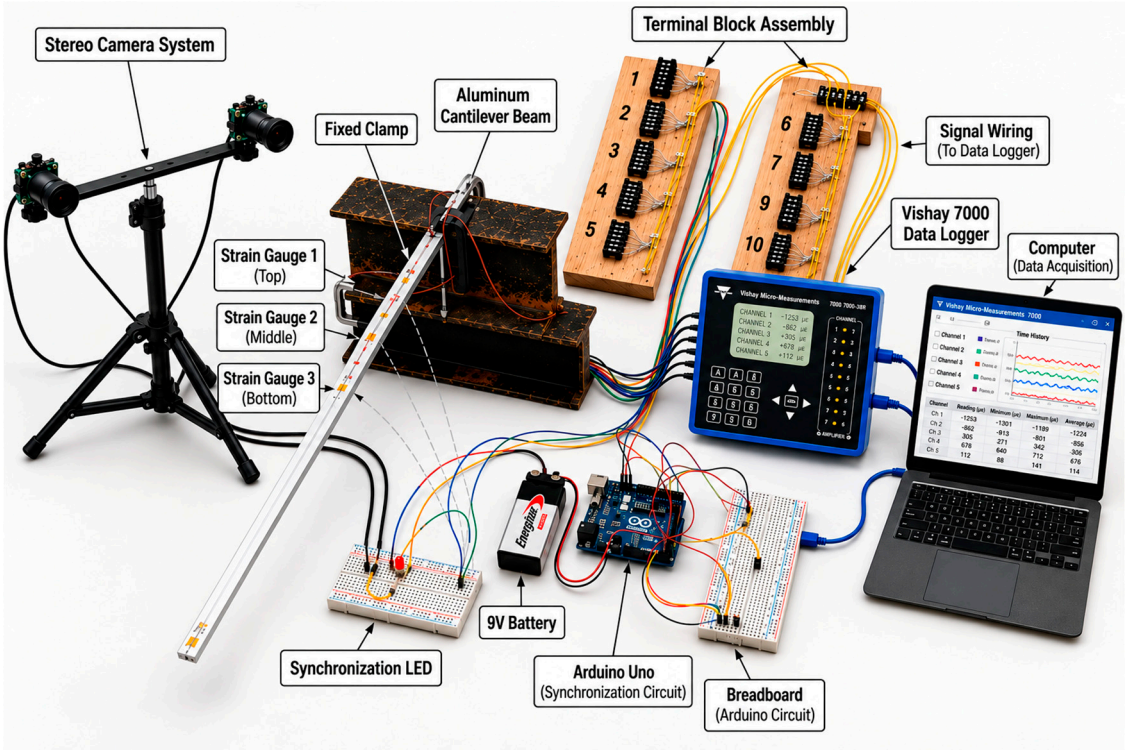


Figure 1. Experimental setup used to validate the proposed PA-ADFT framework through synchronized stereo vision and strain-gauge measurements on an instrumented aluminum cantilever beam.

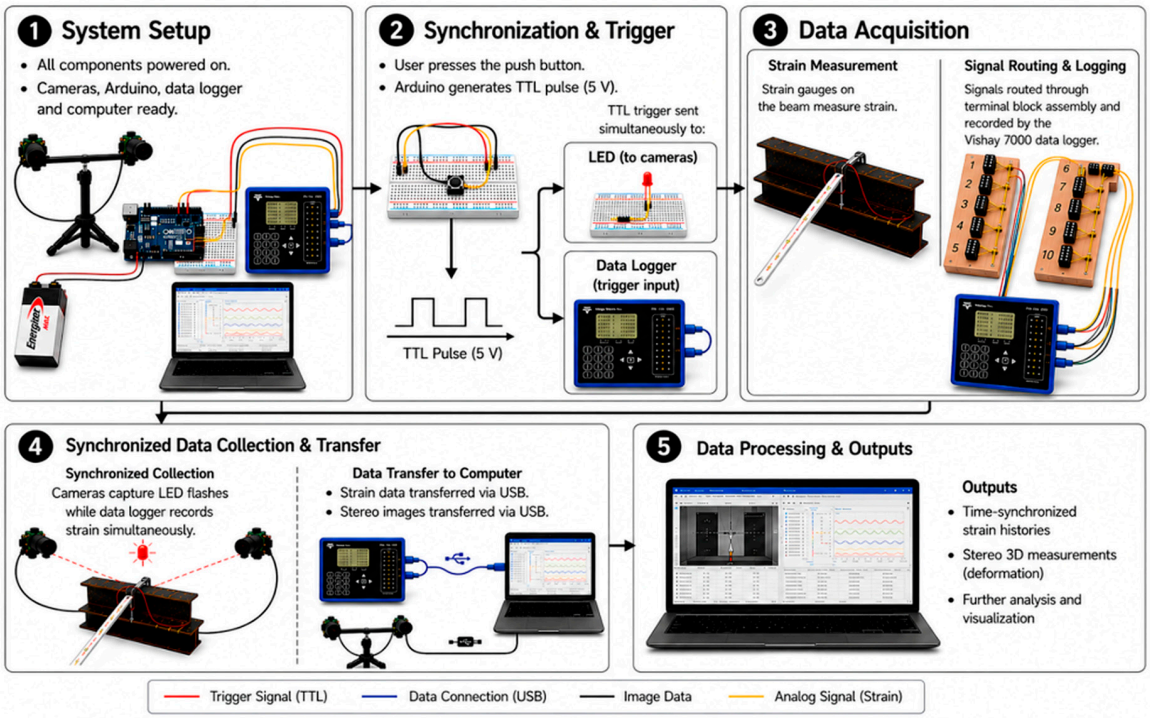


Figure 2. Experimental workflow for synchronized stereo vision and strain-gauge measurements, from system initialization and hardware triggering to data acquisition and processing for target-free modal identification.

The test specimen was a uniform aluminum cantilever beam with a length of 609.6 mm, a rectangular cross-section of 25.35 mm × 3.25 mm, Young's modulus of 69 GPa, Poisson's ratio of 0.33, and density of 2700 kg/m^3. One end of the beam was rigidly clamped to establish a fixed-free boundary condition, while free vibration was initiated by releasing the beam from an initial tip displacement of 10 mm. Structural motion was recorded using two synchronized high-definition Arducam IMX477 HQ cameras (Arducam Technology Co., Ltd., Nanjing, China), each equipped with a 12.3-megapixel Sony IMX477 rolling-shutter image sensor (Sony Group Corporation, Tokyo, Japan), operating at 1920 × 1080 resolution and 60 fps, mounted on a rigid stereo bar with overlapping fields of view covering the entire measurement region. Contact-reference measurements were obtained from three bonded strain gauges positioned 27 mm, 210 mm, and 360 mm from the fixed support, with all signals sampled at 1000 Hz using a DAQ System 7000 (Micro-Measurements, Vishay Precision Group, Raleigh, NC, USA), as illustrated in Figure 1.

Temporal synchronization between the image sequences and strain measurements was achieved through a custom hardware-triggering circuit based on an Arduino-controlled transistor–transistor logic (TTL) pulse that produced an LED event visible to both cameras while simultaneously triggering the data acquisition system, as shown in Figure 2. This configuration ensured a common temporal reference for both sensing modalities throughout each experiment. Following synchronization, free-vibration tests were performed under identical operating conditions, and synchronized stereo images together with reference strain measurements were transferred to the computer to form the experimental dataset used throughout this study.

2.2. Stereo Camera Calibration

Prior to structural testing, the stereo camera system was calibrated to establish a common metric coordinate system for three-dimensional measurement reconstruction. As illustrated in Figure 3a, multiple checkerboard images were acquired throughout the

measurement volume using different orientations and positions to ensure robust estimation of the intrinsic and extrinsic camera parameters over the entire field of view. This calibration procedure defines the geometric relationship between the two cameras, compensates for lens distortion, and provides the foundation for accurate stereo triangulation. The resulting calibration model was subsequently used for all experiments without modification.

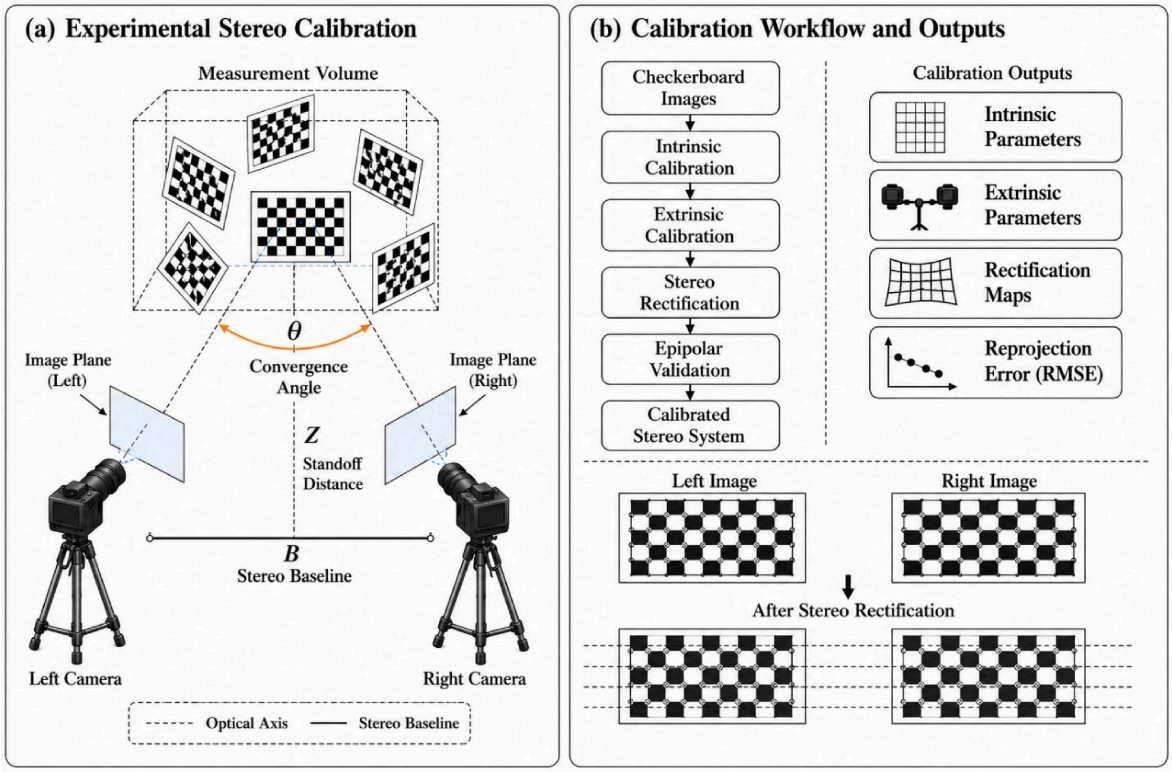


Figure 3. Stereo camera calibration showing (**a**) the experimental calibration geometry and (**b**) the calibration workflow and outputs for the calibrated stereo vision system.

The image formation process for each camera is described by the pinhole projection model, as expressed in Equation (1),

$$s\,\mathrm{x} = \mathrm{K}\,[\mathrm{R} \;\; | \;\; \mathrm{t}]\,\mathrm{X}, \tag{1}$$

where $\mathrm{X} = [X, Y, Z, 1]^T$ denotes a three-dimensional point in the world coordinate system, $\mathrm{x} = [u, v, 1]^T$ is its corresponding image projection, K is the intrinsic camera matrix, R and t represent the extrinsic rotation and translation, respectively, and s is the projective scale factor. Following intrinsic calibration and distortion correction, stereo calibration estimates the relative orientation and baseline between the cameras, enabling stereo rectification such that corresponding image points satisfy the epipolar geometry. The geometric consistency of a stereo correspondence is described by the epipolar constraint given in Equation (2),

$$\mathrm{x}_R^T \mathrm{F} \mathrm{x}_L = 0, \tag{2}$$

where F is the fundamental matrix relating to the left and right image planes. Stereo rectification transforms the image pair so that epipolar lines become horizontally aligned, substantially simplifying correspondence verification and improving reconstruction robustness, as illustrated in Figure 3b.

The calibration procedure produces the intrinsic parameters, extrinsic transformation, rectification maps, and reprojection error that collectively define the calibrated stereo

measurement system. These parameters remain fixed throughout the experimental program and provide the geometric basis for all subsequent three-dimensional point reconstruction. Rather than serving as part of the proposed PA-ADFT framework, the calibrated stereo model establishes an independent measurement infrastructure upon which the target-free tracking methodology is built.

The calibrated system parameters define the metric precision available to the measurement framework. Stereo calibration converged to a reprojection root-mean-square error of 0.312 px, with a recovered baseline of 229.8 mm and a mean focal length of 5395 px for the left camera and 5355 px for the right. At the nominal working distance of 1828.8 mm, one pixel of image-space displacement corresponds to 0.339 mm in the in-plane directions, whereas one pixel of disparity corresponds to 2.697 mm in the depth direction. The ratio of these two scales is fixed by the stereo geometry, Z/B = 7.96, so the depth coordinate is approximately eight times less precise than the in-plane coordinates. This asymmetry is a property of the imaging configuration, and it is used in Section 4.4 to interpret the relative magnitudes of the reconstructed displacement components. Temporal alignment between the image and strain records is established by the hardware trigger described in Section 2.1; the residual alignment error is bounded by one frame interval, 16.7 ms at 60 fps.

2.3. Validation Framework

The experimental program was designed not only to acquire synchronized visual and structural measurements but also to establish independent references for objectively evaluating the proposed PA-ADFT framework. Rather than relying on a single benchmark, the validation strategy integrates complementary analytical, numerical, experimental, and geometric references, as illustrated in Figure 4. The analytical reference is derived from the classical Euler–Bernoulli beam theory, which describes the free vibration of a slender cantilever beam under the assumptions of linear elasticity, small deflections, and negligible shear deformation. Under these assumptions, the transverse vibration of the beam is governed by

$$EI\frac{\partial^4 w(x,t)}{\partial x^4} + \rho A\frac{\partial^2 w(x,t)}{\partial t^2} = 0, \tag{3}$$

where E is Young's modulus of the beam material, I is the second moment of area of the beam cross-section, w(x,t) is the transverse displacement at spatial position x and time t, ρ is the material density, and A is the cross-sectional area.

The corresponding first natural frequency is expressed as

$$f_1 = \frac{\beta_1^2}{2\pi L^2}\sqrt{\frac{EI}{\rho A}}, \tag{4}$$

where f_1 is the first natural frequency, $\beta_1 L = 1.8751$ is the dimensionless eigenvalue for the fundamental bending mode of a cantilever beam, and L is the beam length. This analytical solution provides a physics-based baseline for interpreting the experimentally identified modal characteristics.

An independent numerical reference is established through a three-dimensional finite-element model constructed using the measured beam geometry, material properties, and boundary conditions described in Section 2.1. The model predicts the natural frequencies and mode shapes of the cantilever beam, complementing the analytical solution by accounting for the actual specimen geometry while remaining independent of the vision-based measurements. Experimental validation is provided by three synchronized strain gauges distributed along the beam span and recorded through the common data acquisition system. As direct contact-based sensors, the strain gauges serve as an independent benchmark for

the beam's dynamic response and are used exclusively for validation, without contributing to feature tracking, displacement reconstruction, or modal identification.

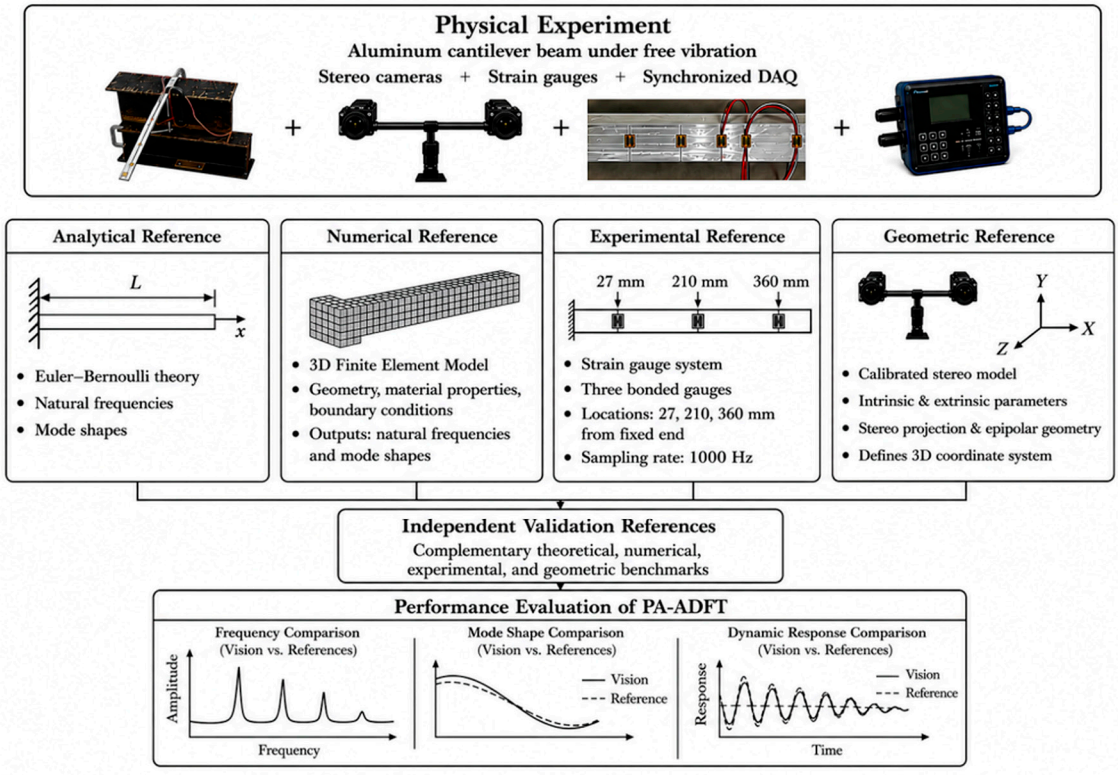


Figure 4. Independent validation framework integrating analytical, numerical, experimental, and geometric references for evaluating the performance of the proposed PA-ADFT framework.

Metric reconstruction of visual measurements is established through calibrated stereo imaging, in which the intrinsic and extrinsic camera parameters define a common three-dimensional coordinate system constrained by stereo projection and epipolar geometry. This geometric reference provides the metric basis for recovering structural displacements, while the analytical model, finite-element model, and strain-gauge measurements remain independent validation benchmarks. Consequently, the proposed PA-ADFT framework is evaluated against multiple theoretical, numerical, experimental, and geometric references rather than calibrated to them, ensuring an objective assessment of the recovered structural response under controlled laboratory conditions.

3. Proposed Physics-Aware Adaptive Deep-Feature Tracking Framework

Following the establishment of the experimental configuration, stereo camera calibration, and independent validation references presented in Section 2, this section introduces the proposed PA-ADFT framework for target-free stereo structural measurement. As illustrated in Figure 5, the proposed framework converts synchronized stereo image sequences into three-dimensional camera-coordinate displacement histories whose geometric and temporal consistency is verified during measurement through four sequential computational stages: (i) shared-weight deep feature extraction, (ii) stereo-temporal association, (iii) physics-aware measurement validation, and (iv) stereo reconstruction with displacement estimation. Unlike conventional vision-based approaches that primarily rely on descriptor similarity or optical-flow continuity, PA-ADFT explicitly distinguishes visual correspondence from measurement reliability by integrating multiple geometric and physics-based consistency constraints into the tracking process. Consequently, visually plausible yet structurally inconsistent observa-

tions are progressively eliminated, allowing only high-confidence structural measurements to propagate through the computational pipeline and providing a reliable foundation for subsequent modal identification and experimental validation.

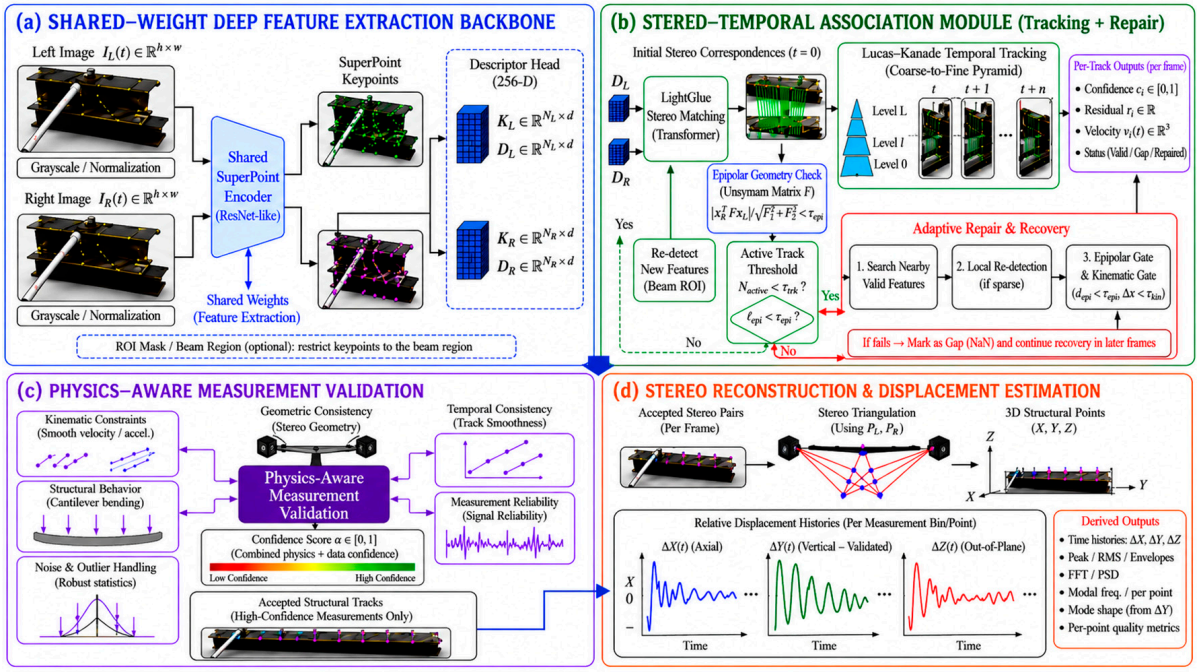


Figure 5. Overview of the proposed PA-ADFT framework: (**a**) deep feature extraction, (**b**) stereo-temporal association, (**c**) physics-aware measurement validation, and (**d**) stereo reconstruction for displacement estimation and modal identification.

The first stage, as illustrated in Figure 5a, extracts robust deep visual features exclusively from the structural region. To suppress background interference and ensure that only beam features participate in subsequent matching, a beam region of interest (ROI) is first identified from each grayscale stereo image using adaptive intensity thresholding followed by morphological refinement. The beam mask is defined as

$$B = \{(u, v) \quad | \quad I(u, v) > T\}, \tag{5}$$

where I(u, v) denotes the grayscale image intensity. The adaptive threshold T used in Equation (5) is computed as

$$T = \max(T_{\min}, P_{99}(I)), \tag{6}$$

where $T_{\min}$ is the minimum allowable threshold and $P_{99}(I)$ denotes the 99th percentile of the image intensity distribution.

The segmented beam regions from the synchronized stereo pair are subsequently processed using a shared-weight SuperPoint network, which simultaneously detects repeatable keypoints and generates locally invariant feature descriptors. Employing identical encoder weights for both camera views ensures descriptor consistency across the stereo pair while improving robustness to illumination variations and moderate viewpoint changes. Restricting feature extraction to the beam region further suppresses background-induced mismatches and ensures that subsequent tracking is initialized using stable structural observations.

The second stage, shown in Figure 5b, establishes persistent stereo-temporal correspondences by combining descriptor-based stereo matching with optical-flow tracking. Initial correspondences between the left and right images are generated using LightGlue,

after which each matched feature is propagated independently through successive image frames using pyramidal Lucas–Kanade optical flow. Assuming brightness constancy within a local neighborhood, the tracking residual is expressed as

$$r(\mathrm{d}) = I_{t+1}(\mathrm{p} + \mathrm{d}) - I_t(\mathrm{p}), \tag{7}$$

where p denotes the feature location and d is the image displacement. The displacement vector is then estimated by minimizing the residual defined in Equation (7), yielding

$$\Delta \mathrm{d} = \left(\sum_{\Omega} \nabla, I, \nabla, I^T \right)^{-1} \sum_{\Omega} \nabla I \left[I_t(\mathrm{p}) - I_{t+1}(\mathrm{p} + \mathrm{d}) \right], \tag{8}$$

where Ω denotes the local tracking window. A forward–backward consistency check subsequently removes unreliable trajectories by verifying reciprocal tracking agreement between consecutive frames. To maintain adequate spatial coverage throughout the measurement sequence, an adaptive feature replenishment strategy continuously introduces newly detected beam features whenever the active track population decreases below a predefined threshold. Rather than replacing failed trajectories indiscriminately, the framework preserves persistent structural measurement bins, enabling temporary feature loss to be recovered without compromising the physical identity of the tracked structural locations.

The principal contribution of PA-ADFT is realized in the third stage as shown in Figure 5c, where every candidate trajectory is evaluated using a physics-aware validation strategy before being accepted as a structural measurement. Instead of relying solely on descriptor similarity, each trajectory is assessed against the geometric, temporal, and kinematic criteria implemented in the tracker, together with a coverage and amplitude check applied when measurement points are accepted. Stereo geometric consistency is first enforced through the epipolar relationship introduced in Equation (2), restated here in the notation of the validation stage,

$$\mathrm{x}_R^T \mathrm{F} \mathrm{x}_L = 0, \tag{9}$$

where F denotes the calibrated fundamental matrix relating to the two stereo cameras. Based on the epipolar constraint in Equation (9), the geometric consistency of each correspondence is quantified by the epipolar distance.

$$d_{\mathrm{epi}} = \frac{| \mathrm{x}_R^T \mathrm{F} \mathrm{x}_L |}{\sqrt{(F\mathrm{x}_L)_1^2 + (F\mathrm{x}_L)_2^2}}, \tag{10}$$

It measures the deviation in each matched feature from its corresponding epipolar line and provides a quantitative indicator of stereo matching accuracy. Temporal consistency is enforced as a forward–backward Lucas–Kanade check, in which a track is retained only if the round-trip re-tracking error between consecutive frames is at most 2.0 px, with triangulation restricted to temporally simultaneous stereo frame pairs. Geometric consistency requires that tracked and repaired points lie within 6 px of the per-frame beam mask, that stereo seeds satisfy a point-to-epipolar-line distance of at most 8 px at 1920×1080, and that each subsequent sample remain within 4 px of the epipolar distance established by its own seed. Physical plausibility is enforced through the kinematic identity gate, which accepts a repaired sample only if its triangulated position lies within 20 mm of the last valid position of the same measurement bin, an allowance that scales with the length of the preceding gap by 5 mm per additional frame and is capped at 60 mm.

Measurement quality is assessed at the point-acceptance stage rather than per sample: a point is accepted only if its track coverage is at least 80% of frames, its record extends at least 3 s after synchronization, and its vertical amplitude remains within plausibility

bounds of 80 mm peak-to-peak and 120 mm maximum. Plausibility with respect to the expected structural response is enforced through the kinematic gate rather than as a separate criterion. In the implementation evaluated here, each criterion is applied as an independent acceptance test against a fixed threshold, and a candidate trajectory is retained only if it satisfies every test. The criteria comprise both upper-bound and lower-bound conditions, so that acceptance is expressed as the conjunction,

$$a_i = \prod_{j \in U} I\left(C_{j,i} \leq t_j\right) \cdot \prod_{k \in L} I\left(C_{k,i} \geq t_k\right) \tag{11}$$

where $C_{j,i}$ is the value of the j-th acceptance criterion evaluated for candidate trajectory i, t_j is the corresponding fixed threshold, U and L denote the disjoint sets of upper-bound and lower-bound criteria, respectively, I(·) is the indicator function returning one when its argument is true and zero otherwise, and a_i identifies a trajectory accepted as a structural measurement. The set U comprises the beam-region, forward–backward, epipolar seed, epipolar deviation, and amplitude criteria together with the kinematic identity gate, while the set L comprises the track-coverage and record-duration criteria; U and L together partition the full set of implemented acceptance criteria, so that acceptance requires every upper-bound and every lower-bound test to be satisfied simultaneously. Because the criteria are combined as a product of indicator functions rather than as a weighted sum, the implemented procedure involves no normalized scores, no weighting coefficients, and no single tunable confidence threshold.

Finally, the accepted structural tracks are transformed into metric structural measurements through calibrated stereo reconstruction, as shown in Figure 5d. The stereo projection model described by Equation (1) is then applied to establish the correspondence between the three-dimensional structural points and their image observations. Here, K_i, R_i, and t_i denote the intrinsic matrix, rotation matrix, and translation vector of the i^{th} camera, respectively. The three-dimensional coordinates of each structural point are subsequently recovered through stereo triangulation,

$$\mathrm{X} = \mathrm{triangulate}(\mathrm{x}_L, \mathrm{x}_R), \tag{12}$$

and the relative structural displacement history is obtained by subtracting the initial reference configuration,

$$\mathrm{u}(t) = \mathrm{X}(t) - \mathrm{X}(t_0). \tag{13}$$

The resulting synchronized three-dimensional displacement histories constitute the primary output of the proposed PA-ADFT framework. To quantify the contribution of each processing stage, the five experiments summarized in Table 1 were reprocessed using three fixed configurations. The first retained SuperPoint feature extraction, LightGlue stereo matching, and pyramidal Lucas–Kanade tracking with the physics-aware gates disabled. The second restored epipolar seed gating (8 px at 1920 × 1080), per-bin epipolar-deviation gating (4 px), and pair–repair. The third corresponds to the PA-ADFT pipeline, which additionally applies the kinematic identity gate defined above. For the tested 10 mm release, first-mode response near 7.03 Hz, and 60 fps acquisition, the expected maximum inter-frame displacement is approximately 7.4 mm; the 20 mm limit therefore provides a conservative allowance for physical motion while rejecting repaired correspondences that jump to incompatible structural locations. These configurations differ by enabling or disabling sequential acceptance criteria in the tracker; they do not correspond to a re-weighting of any confidence score, consistent with the conjunctive acceptance rule of Equation (11). A continuous threshold-sensitivity sweep was not performed; instead, the robustness of the implemented criteria is assessed through the staged ablation reported in Section 4.1 and Figure 6.

Table 1. Tracking performance and first-mode identification for the PA-ADFT pipeline across five repeated free-vibration experiments. Coverage is computed among accepted measurement points.

Trial	Accepted Points	Median Coverage (%)	Minimum Coverage (%)	Stereo Vertical Frequency (Hz)	Strain Frequency (Hz)
1	7	94	87	7.02	7.03
2	7	95	89	7.01	7.03
3	9	91	87	7.01	7.03
4	9	93	88	7.02	7.03
5	8	90	84	7.04	7.03

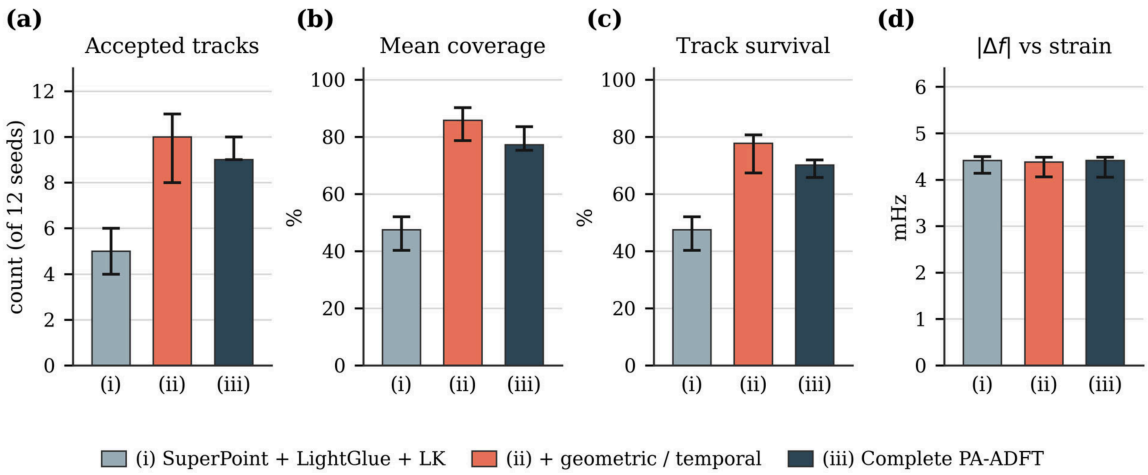


Figure 6. Staged internal ablation of the PA-ADFT pipeline. Configurations are (i) SuperPoint, LightGlue and pyramidal Lucas–Kanade tracking with the physics-aware gates disabled; (ii) the same tracker with epipolar gating and pair–repair restored; and (iii) the PA-ADFT pipeline including the kinematic identity gate.

4. Results and Performance Validation

This section presents the experimental validation of the PA-ADFT framework using repeated free-vibration experiments on the laboratory cantilever beam described in Section 2. The results are organized progressively to establish confidence in the complete measurement pipeline before evaluating structural response and modal-identification accuracy. The validation first examines the repeatability and robustness of the target-free tracking process, followed by assessment of the reconstructed structural response, quantitative verification of the identified modal characteristics against synchronized strain measurements and finite-element analysis, and finally an evaluation of the three-dimensional camera-coordinate displacement information provided by the stereo measurement system.

4.1. Measurement Reliability and Tracking Performance

The reliability of a target-free vision-based measurement framework depends on its ability to consistently establish and maintain persistent structural observations throughout repeated experiments. Table 1 summarizes the tracking performance obtained from five synchronized stereo free-vibration tests recorded at 1920 × 1080 resolution and 60 fps. Across all trials, PA-ADFT retained between 7 and 9 accepted structural measurement points, with median point coverage ranging from 90% to 95% and minimum accepted-point coverage remaining at or above 84%. The workflow initially evaluates twelve target-free seeded locations, and the final accepted points are those that satisfy the complete geometric, temporal, kinematic, and measurement-quality checks. The present point set is therefore

interpreted as adequate for first-mode validation, while higher-mode shape identification or more complex deformation fields would require denser accepted measurement coverage and a different acquisition and excitation strategy.

The repeatability reported in Table 1 was achieved using the fixed acquisition, synchronization, tracking, and signal-processing settings summarized in Table 2. The complete measurement pipeline, including SuperPoint feature extraction, LightGlue stereo matching, pyramidal Lucas–Kanade temporal tracking, forward–backward consistency verification, adaptive feature replenishment, and identical spectral-processing parameters, was maintained throughout all experiments. Reporting these settings ensures that the observed repeatability is attributed to the PA-ADFT framework rather than to the isolated performance of individual computer vision algorithms.

Table 2. Summary of the experimental configuration, processing parameters, and analysis settings used for target-free stereo modal identification.

Analysis Item	Setting Used in This Study	Purpose
Specimen state	Healthy aluminum cantilever beam	Establishes the baseline modal-identification case
Release condition	10 mm vertical tip displacement	Provides repeatable free-decay excitation
Video subset	Five synchronized stereo trials at 1920 × 1080 and 60 fps	Defines the vision dataset for the reported results
Strain reference	Three 120 Ω linear strain gauges sampled at 1000 Hz	Supplies the contact-reference bending response
Gauge stations	27 mm, 210 mm, and 360 mm from the fixed end	Provides spatially separated strain evidence along the beam
Synchronization	Shared electrical pulse with visible stereo LED event	Aligns strain and video time axes
Feature detector	SuperPoint keypoints and descriptors	Seeds target-free beam-region features
Stereo matcher	LightGlue matching with epipolar consistency filtering	Forms calibrated left–right correspondences
Temporal tracker	Pyramidal Lucas–Kanade tracking with forward–backward rejection	Propagates measurement points through the free decay
Adaptive repair	Local feature replenishment is constrained by image proximity, epipolar deviation, and three-dimensional jump limits	Preserves measurement-bin identity during brief feature loss

Table 1 reports tracking completeness and modal identification for the PA-ADFT pipeline. To isolate the contribution of each processing stage, the same five experiments were reprocessed under the three configurations defined in Section 3, with successive stages of PA-ADFT enabled in turn; the results are summarized in Figure 6. This staged comparison is an internal ablation of the proposed processing chain rather than a benchmark against external methods, and comparative performance against monocular displacement estimation, conventional optical-flow tracking, and other published vision-based frameworks remains to be established. Coverage in Table 1 is evaluated among accepted points, whereas coverage in Figure 6 is evaluated over the same twelve initial seeds, including trajectories that fail subsequent gates. SuperPoint–LightGlue–Lucas–Kanade tracking already recovers the first-mode frequency to within 4.4 mHz of the strain-gauge reference, but retains a median of only five of the twelve initial seeds, with a mean coverage of 48% and 58% of trajectories rejected (Figure 6). Restoration of epipolar gating and pair–repair increases the median number of accepted tracks to ten and the mean coverage to 86% and

reduces the rejected fraction to 17%. Inclusion of the kinematic identity gate then returns a median of eight accepted tracks at 77% mean coverage, with 25% of trajectories rejected, and discards a median of 144 repaired jumps per experiment. This reduction in coverage follows from the rejection of kinematically implausible repaired correspondences rather than from a loss of tracker performance.

The same effect is visible in the reconstructed depth channel. The depth-to-vertical ratio used here is the peak-to-peak depth displacement divided by the peak-to-peak vertical displacement of the same measurement point, each taken as the difference between the 99th and 1st percentiles of the raw triangulated component over the full record. It is evaluated at a point retained under all three configurations and is used only to compare configurations, not to characterize the accepted measurement set. On this basis, enabling the kinematic identity gate reduces the ratio from 13.1 to 6.3 in Experiment 3 at representative point 4, and from 11.6 to 8.9 in Experiment 5 at representative point 2, confirming that the gate suppresses repaired correspondences in the depth channel. These values are not comparable with the complete measurement-point result over the 0–10 s window after synchronization, where the corresponding ratio is 0.41. The median frequency difference from the strain-gauge reference is 4.41 mHz, 4.38 mHz, and 4.41 mHz for configurations (i), (ii), and (iii), respectively. Across the five experiments this difference, evaluated from the median frequency of the accepted measurement points, varies between 3.7 mHz and 5.3 mHz, but within any single experiment the three configurations differ by at most 0.08 mHz, so the residual difference is a property of the individual experiment rather than of the measurement gates. The median epipolar residual remains between 1.70 px and 1.95 px in all three configurations.

Figure 7 provides tracking provenance supporting the reported measurements. Candidate features are first selected exclusively from the beam region, subsequently verified through calibrated stereo correspondences, and finally retained as accepted structural measurement bins following the physics-aware validation procedure. This audit trail demonstrates that the reconstructed displacement histories originate from persistent structural image features rather than arbitrary background texture or transient feature matches. Collectively, the consistently high measurement coverage, stable accepted-point population, and successful operation across repeated trials establish the reliability of PA-ADFT for target-free structural measurement and provide the foundation for the response reconstruction and modal validation presented in the following sections.

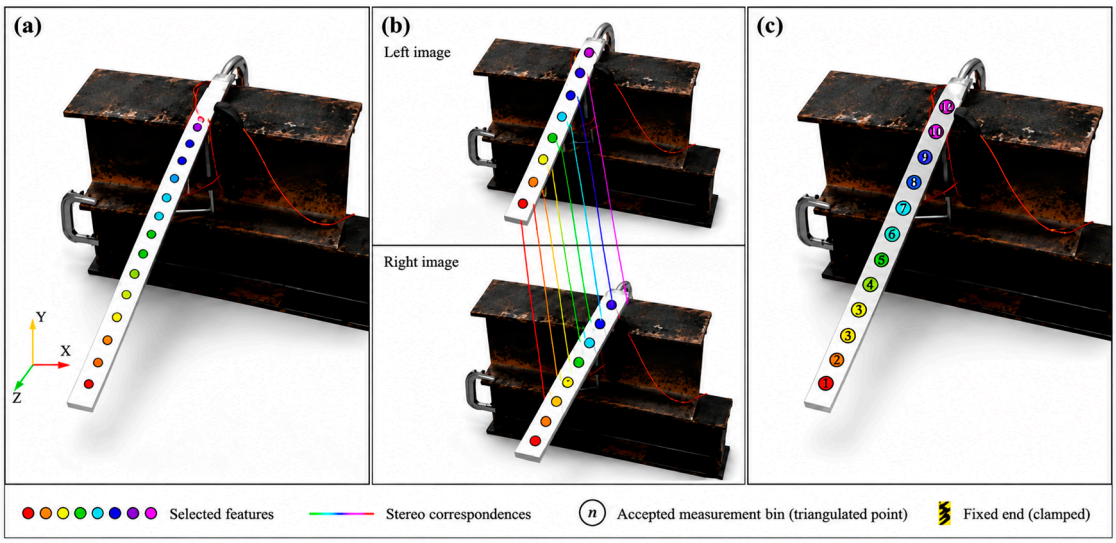


Figure 7. Target-free measurement generation showing (**a**) beam-region feature selection, (**b**) stereo feature matching, and (**c**) accepted structural measurement bins.

4.2. Structural Response Reconstruction

The reconstructed stereo displacement histories exhibit the expected response of a lightly damped cantilever beam subjected to an initial tip release. As shown in Figure 8, all representative measurement locations display a smooth free-vibration response characterized by oscillatory motion with a progressively decreasing amplitude over time. The nearly exponential decay observed in each trace is consistent with the dissipation of vibration energy under free-decay conditions, indicating that the reconstructed displacement histories preserve the temporal characteristics of the structural response without introducing visible tracking discontinuities or non-physical oscillations. The theoretical free-decay envelope overlaid in Figure 8 corresponds to a damping ratio of ζ = 0.185% and an identified first-mode frequency of 7.027 Hz. The corresponding strain-gauge measurements in Figure 9 exhibit the same decaying behavior, providing an independent contact-based reference for the observed structural motion.

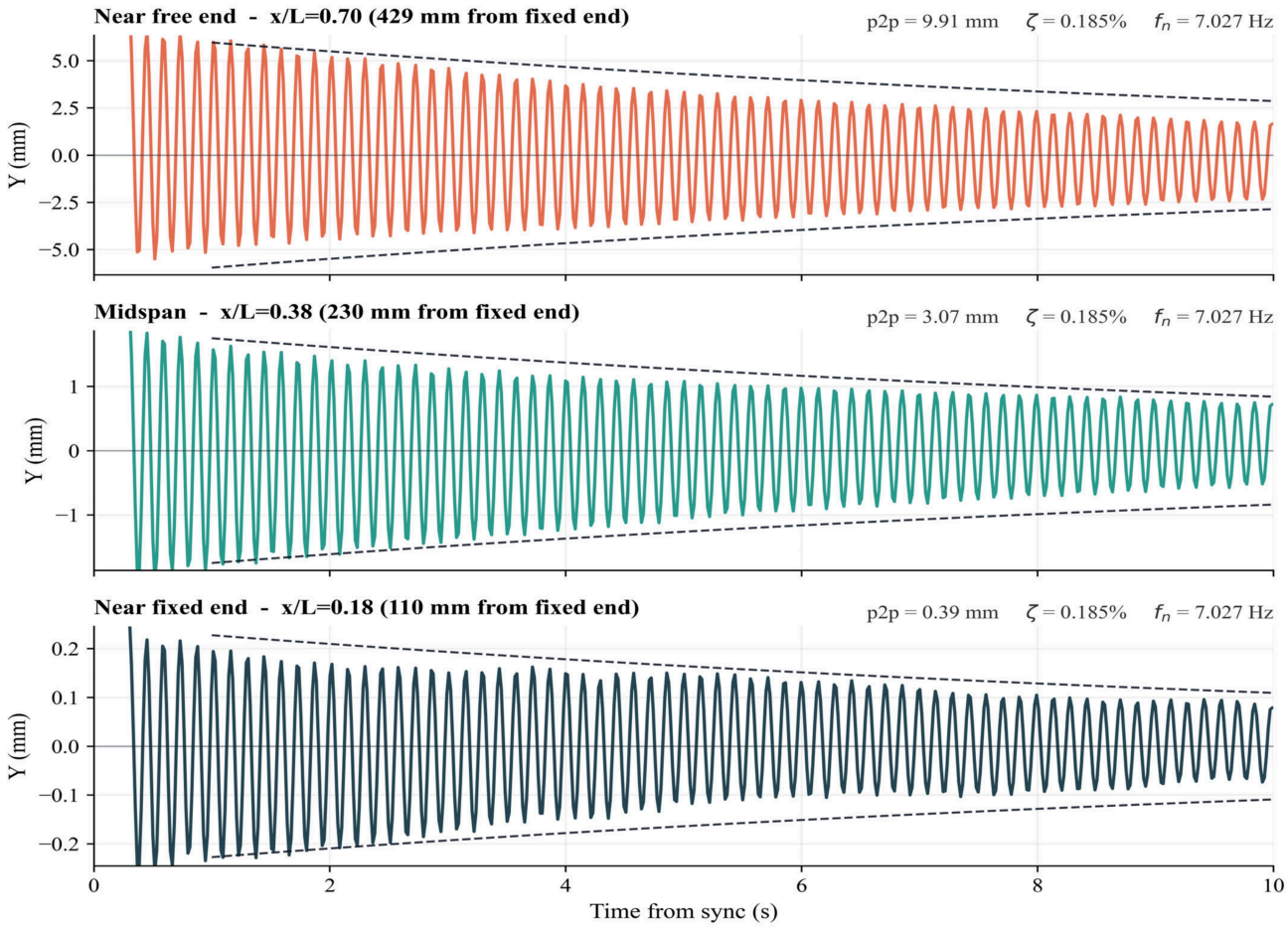


Figure 8. Representative vertical displacement histories reconstructed by PA-ADFT at the near free end, midspan, and near the fixed end during first-mode free vibration.

The spatial variation in the reconstructed displacement also follows the expected deformation pattern of the first bending mode. The largest displacement occurs near the free end of the beam, decreases at midspan, and becomes smallest near the fixed support, with representative peak-to-peak amplitudes of approximately 9.9 mm, 3.1 mm, and 0.4 mm, respectively, as shown in Figure 8. The displacement trend approaching the clamp was assessed rather than the zero-displacement boundary condition being verified directly, because the fixed end itself was not monitored. The closest seeded locations lie within approximately 5% of the beam length of the calibrated fixed end and are tracked with 87–100%

mean coverage across all five experiments. The seeded location closest to the clamp shows a peak-to-peak vertical displacement of 0.05–0.24 mm across the five experiments, 0.5–2.4% of the free-end amplitude, with the next seeded location outward showing a correspondingly larger displacement of 0.35–1.06 mm, consistent with the expected monotonic growth of the first-mode shape away from the fixed support. This monotonic amplitude ordering is consistent with classical cantilever bending behavior, where structural displacement increases continuously from the clamped boundary toward the free end. In contrast, the strain-gauge responses exhibit the opposite amplitude distribution because bending strain is proportional to curvature rather than displacement, producing the largest strain near the fixed support and progressively smaller responses toward the free end, as illustrated in Figure 9. Together, the complementary displacement and strain measurements confirm that the reconstructed stereo motion is physically consistent with the expected first-mode deformation of the healthy cantilever beam.

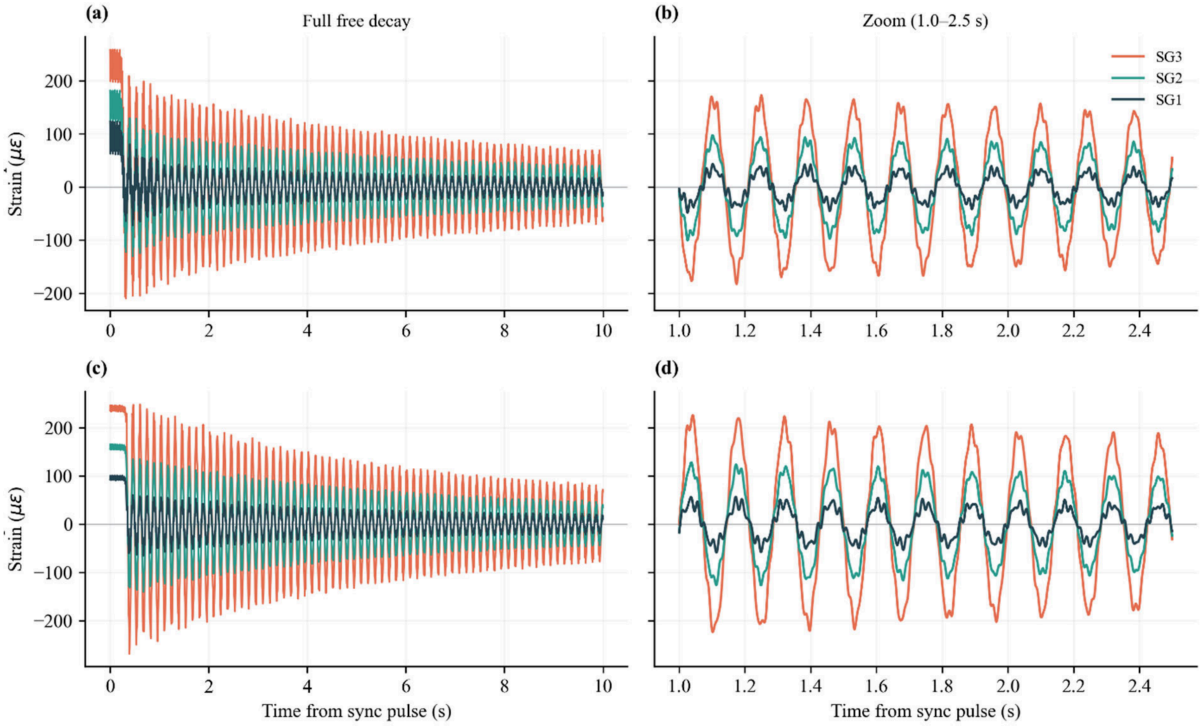


Figure 9. Representative strain-gauge responses measured at three beam locations during free vibration:(**a**) full free-decay response; (**b**) zoomed response from 1.0–2.5 s; (**c**) full free-decay response from the synchronized measurement; and (**d**) corresponding zoomed response from 1.0–2.5 s.

4.3. Modal Identification Accuracy and Experimental Validation

The primary objective of PA-ADFT is to accurately recover the first bending modal frequency from target-free stereo measurements. The repeated trial statistics summarized in Table 1 demonstrate that this objective was achieved consistently across all five trials. The median stereo-derived frequency was 7.02 Hz, compared with a median strain-gauge frequency of 7.03 Hz. Evaluated from the unrounded values, the median difference is 4.41 mHz, or approximately 0.06% of the measured frequency. These results indicate that the proposed tracking and validation strategy produces highly repeatable modal estimates rather than trial-specific measurements.

A representative experimental comparison is presented in Figure 10, where the reconstructed stereo displacement history is directly compared with the synchronized strain-derived displacement estimate at the beam midspan. The time histories exhibit the same free-decay behavior, with a common decay envelope and consistent cycle timing throughout

the oscillation. This consistency extends to the frequency domain, where both measurements produce dominant spectral peaks near 7.0 Hz. Although amplitude differences remain because the strain response is converted to displacement through Euler–Bernoulli beam theory whereas the stereo system measures displacement directly, the correlation coefficient of r = 0.738 indicates a clear positive relationship between the two records and supports the conclusion that both techniques capture the same dominant mode. The residual difference is attributable to the Euler–Bernoulli conversion of the strain response and to the residual frame-level synchronization offset bounded in Section 2.2, rather than to a discrepancy in the identified modal behavior. Quantitative agreement between the two modalities is therefore established in the frequency domain, where the dominant peaks coincide, rather than through point-by-point amplitude agreement in the time domain.

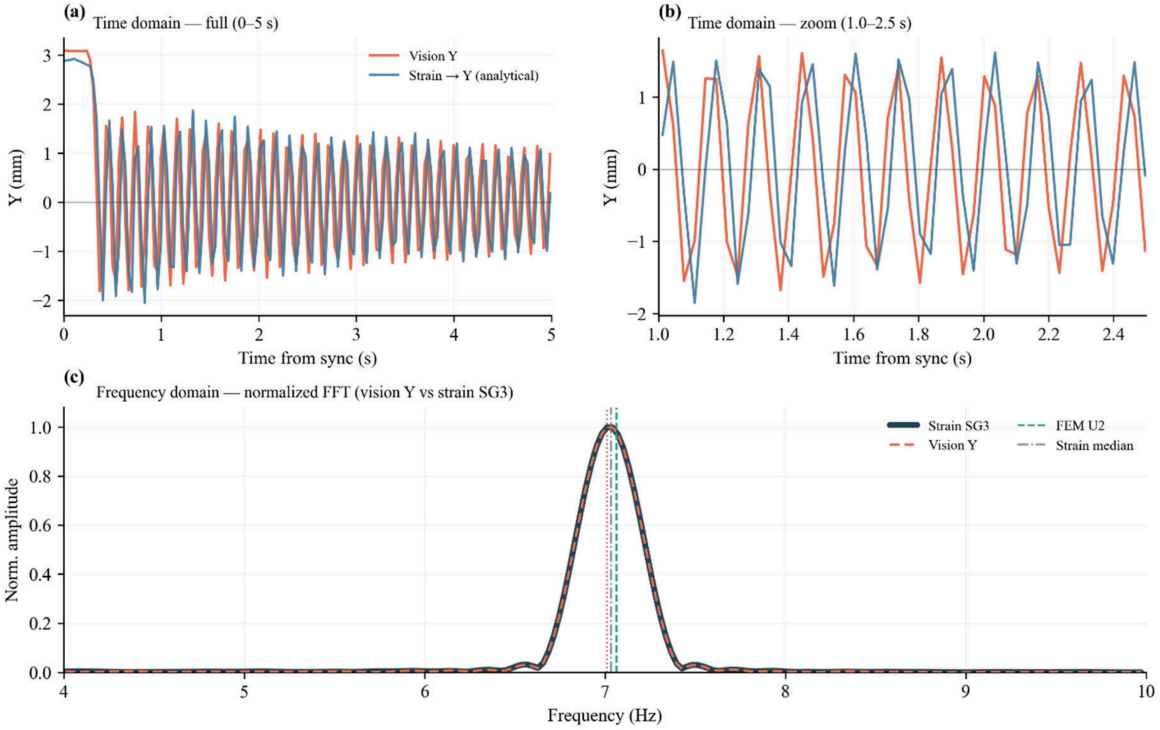


Figure 10. Experimental validation of the proposed PA-ADFT framework through comparison of reconstructed vision-based and strain-gauge responses: (**a**) full time-domain response (0–5 s); (**b**) zoomed time-domain response (1.0–2.5 s); and (**c**) normalized frequency-domain comparison.

Independent numerical and analytical references further support experimental validation. The median stereo frequency (7.02 Hz) and synchronized strain frequency (7.03 Hz) agreed with the finite-element prediction (7.06 Hz), while also agreeing with the first bending frequency expected from Euler–Bernoulli beam theory. The convergence of experimental, numerical, and theoretical references demonstrates that PA-ADFT reliably identifies the correct first bending mode, providing consistent modal estimates from target-free stereo measurements without requiring artificial tracking markers or sensor-based frequency calibration.

4.4. Three-Dimensional Motion Characterization

While the preceding subsections validated the vertical camera-coordinate response against synchronized strain measurements, the stereo reconstruction simultaneously provides the three-dimensional camera-coordinate displacement history of each accepted measurement point. Figure 11 presents the X, Y, and Z displacement components reconstructed at a representative measurement point during free vibration. The Y-component

exhibits the dominant oscillatory response, with a peak-to-peak displacement of approximately 5.82 mm, and displays the same exponentially decaying behavior observed in the preceding validation results. In comparison, the X-component remains comparatively small (approximately 0.6 mm peak-to-peak) and consists primarily of low-amplitude motion about the mean position, whereas the Z-component exhibits variations of approximately 2.41 mm peak-to-peak, corresponding to approximately 41% of the vertical amplitude. As quantified below, this variation lies within one pixel of depth-reconstruction uncertainty. These amplitudes are evaluated over the 0–10 s window after synchronization.

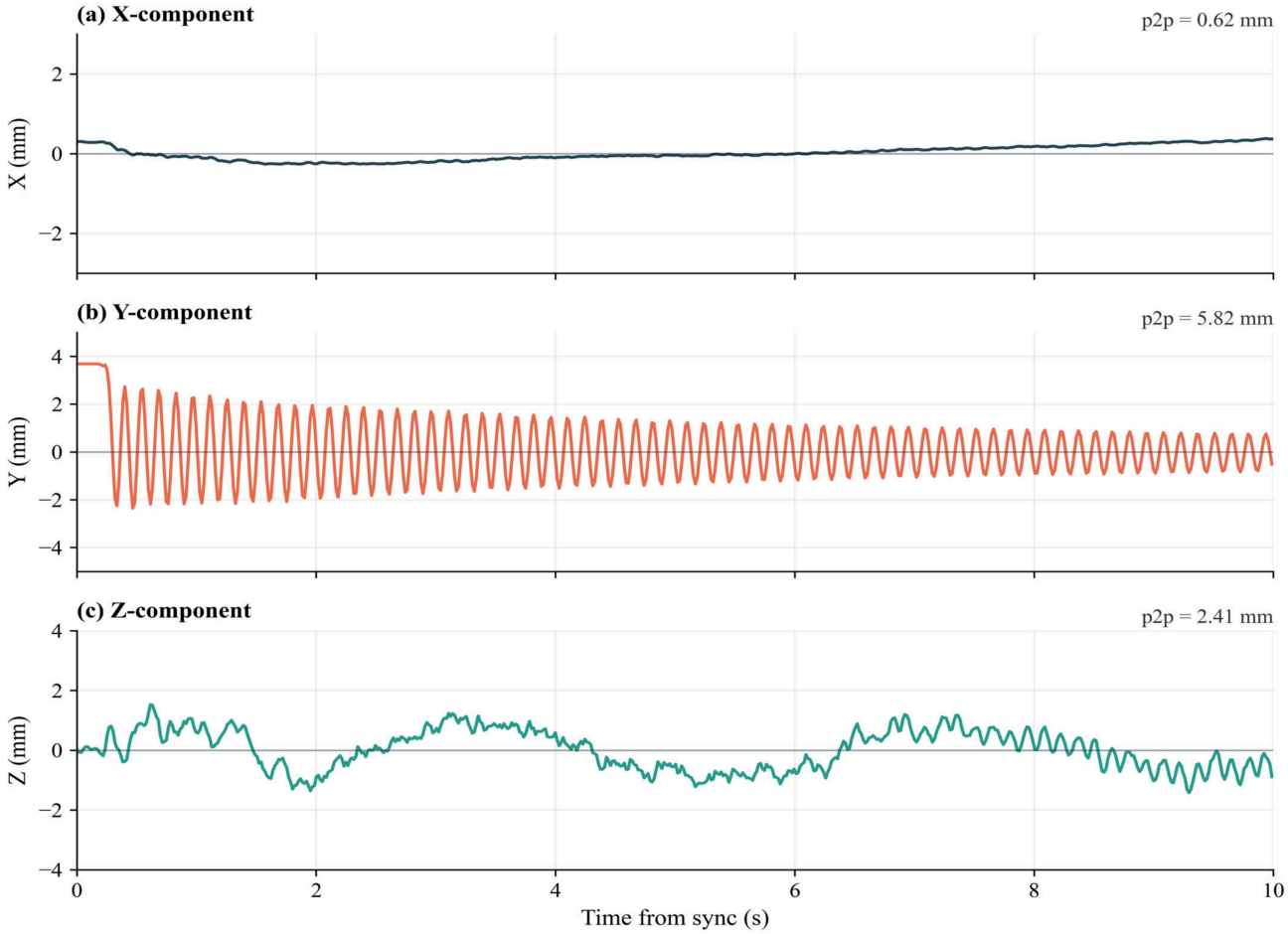


Figure 11. Reconstructed three-dimensional structural response at a representative beam location, showing the simultaneously reconstructed displacment time histories (**a**) X-component; (**b**) Y-component; (**c**) Z-component. Only the Y component is quantitatively validated against the independent references.

The corresponding component-wise frequency spectra are shown in Figure 12. The vertical (Y) spectrum is characterized by a single dominant peak at 7.02 Hz, consistent with the first bending frequency established in Section 4.3. The X-component also contains a peak near the same frequency (7.04 Hz), although with considerably lower displacement amplitude, while the Z-component exhibits a dominant peak at 6.98 Hz together with broader low-amplitude spectral content. The X and Z spectral peaks are compared with the finite-element model (FEM) first-mode frequency in Figure 12, while the Y component remains the primary component validated against both the strain-gauge and finite-element references. The deviations annotated in Figure 12 are evaluated for the single representative measurement point plotted. These single-point values are marginally larger than the per-experiment differences reported in Section 4.1, which are obtained from the median frequency across all accepted measurement points.

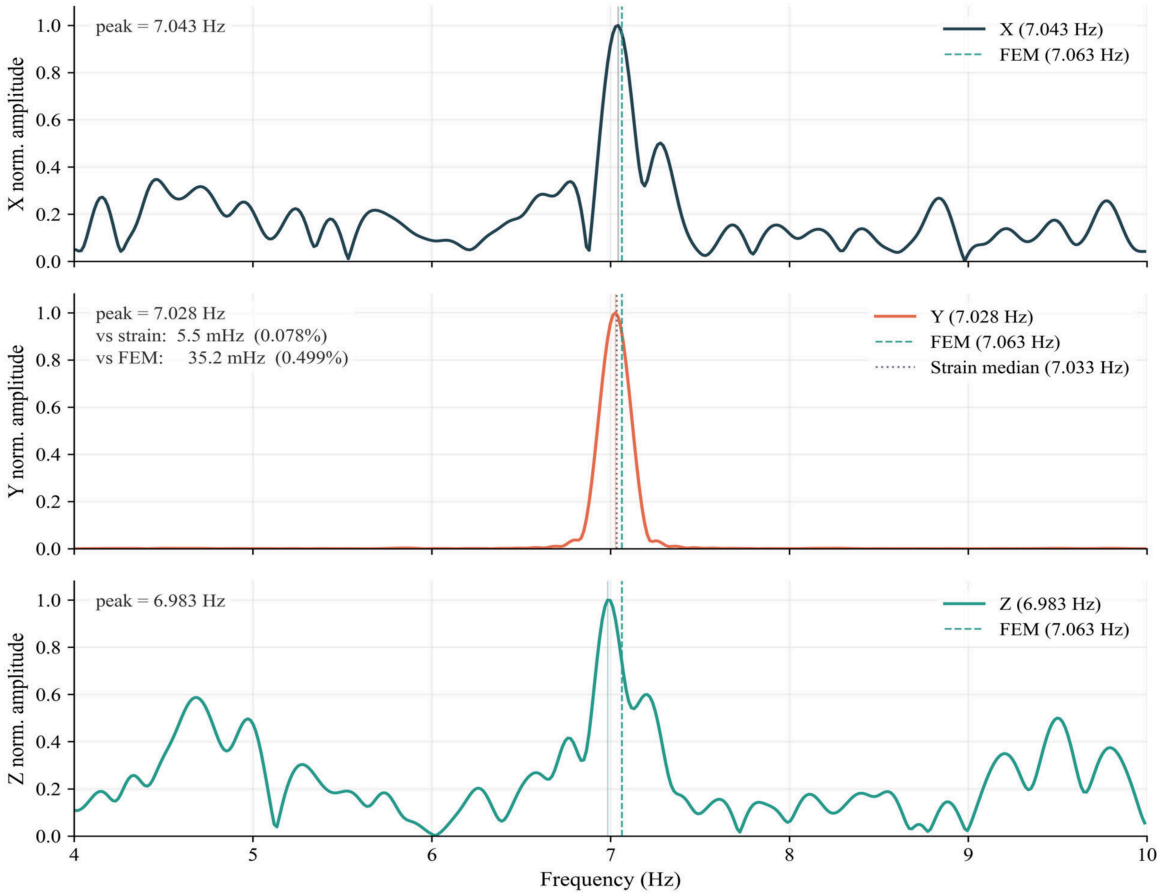


Figure 12. Frequency spectra of the reconstructed three-dimensional displacement components at a representative beam location. The dominant peaks of the X, Y, and Z components are shown relative to the FEM first-mode reference, and the vertical component is additionally compared with the strain-gauge reference.

As established in Section 2.2, the depth direction of a stereo pair carries an uncertainty larger than the in-plane directions by the ratio of working distance to baseline, which for the present configuration is $Z/B = 7.96$. One pixel of image-space error therefore maps to 0.339 mm in the in-plane directions but to 2.697 mm in depth. This geometric amplification explains the broader low-amplitude content and larger apparent variation in the depth component. The observed Z-axis variation of approximately 2.41 mm peak-to-peak corresponds to approximately 0.89 pixel of disparity and therefore lies entirely within the depth uncertainty associated with a single pixel of image-space error. The reported depth variation consequently cannot be interpreted as resolved out-of-plane structural motion, and no physical out-of-plane response is claimed from it.

The cross-axis spectra are useful for confirming that the dominant first-mode frequency is present in the reconstructed camera-coordinate components, while the vertical component provides the primary amplitude and modal-response validation against the independent strain-gauge and finite-element references. Accordingly, the present configuration reconstructs the camera-coordinate trajectory of each accepted point in three dimensions. The vertical component is validated quantitatively against the strain-gauge and finite-element references. The lateral and depth components are therefore reported as reconstructed quantities rather than validated measurements. In particular, the depth component cannot be separated from depth-reconstruction uncertainty and cross-axis leakage at the present baseline-to-working-distance ratio.

5. Discussion

5.1. Measurement Reliability and Physical Consistency

The principal objective of PA-ADFT was not to improve visual feature matching, but to improve the reliability of target-free structural measurements. Image correspondence alone cannot ensure that tracked points remain physically attached to the same material region throughout dynamic excitation. PA-ADFT addresses this limitation by integrating geometric consistency, temporal persistence, and kinematic plausibility within a unified measurement framework before displacement reconstruction. Consequently, structural measurement confidence is established through cumulative physical evidence rather than descriptor similarity alone.

The experimental results demonstrate that this measurement strategy produces stable and repeatable structural observations across independent vibration tests. The consistently high measurement coverage and sustained tracking continuity indicate that the accepted trajectories remained physically representative throughout the free-vibration response despite natural variability in target-free feature detection. More importantly, the close agreement between the recovered modal frequency and the independent strain-gauge and finite-element references confirms that the proposed validation strategy preserves the structural information required for accurate modal identification rather than merely improving image tracking. These findings support the central hypothesis of this study that measurement quality must be verified continuously throughout the sensing process rather than only after displacement reconstruction. As the staged comparison in Section 4.1 shows, the first-mode frequency is already recovered by deep feature correspondence and optical-flow tracking, whereas geometric and temporal filtering is what maintains a spatially distributed set of continuous trajectories, and the kinematic identity gate what prevents repaired correspondences from contaminating the triangulated displacement history. The hypothesis is therefore supported with respect to measurement reliability rather than first-mode frequency identification alone.

5.2. Advantages and Limitations of PA-ADFT

The proposed framework demonstrates that combining modern learned feature correspondence with physics-aware measurement validation enables target-free stereo vision to function as a structural sensing system rather than solely as a computer vision pipeline. Unlike methods that optimize individual stages independently, PA-ADFT integrates feature extraction, stereo correspondence, adaptive tracking, and structural quality control within a single workflow, producing three-dimensional camera-coordinate displacement histories whose vertical component is validated against independent references and is suitable for first-mode analysis. Independent agreement with synchronized strain measurements and a calibrated finite-element model further strengthens confidence that the recovered response reflects structural behavior instead of image artifacts.

The scope of the present validation is limited, and the following boundaries define the conditions under which the reported performance applies rather than the general capabilities of the framework. All five reported experiments were performed on a single healthy aluminum cantilever beam, under a single 10 mm tip-release amplitude, and no parametric sensitivity to excitation amplitude has been established. The free-decay excitation used here produces essentially single-mode response, so the reported accuracy is specific to first-mode identification and does not extend to higher-order modes without a different excitation strategy.

The present 60 fps image acquisition is appropriate for the first bending mode near 7 Hz, but higher modes would require higher frame-rate acquisition and excitation that deliberately stimulates those modes; modes approaching or exceeding the camera Nyquist

limit cannot be identified reliably from these image records. Higher-mode shapes and more complex deformation fields would also require denser accepted measurement coverage than the present first-mode study. As with all stereo vision systems, the measurement accuracy further depends on camera calibration, synchronization quality, and image fidelity. The depth coordinate is constrained by the stereo geometry: for the working distance and baseline used here, depth precision is approximately eight times poorer than in-plane precision, so a shorter working distance or wider baseline would be required to improve depth-resolved measurements.

Taken together, these boundaries limit the quantitative claims of this study to the vertical displacement response, the first-mode frequency, and the preservation of stereo correspondence identity. The depth component is reported as reconstructed rather than validated, and no quantitative conclusion is drawn from its amplitude. The staged comparison in Figure 6 is an internal ablation of the proposed chain and not a benchmark against external vision-based methods; comparison with monocular, optical-flow, and other published target-free frameworks remains to be established. Extension to damaged structures, complex structural geometries, and field conditions, including illumination variability, camera motion, and partial occlusion, requires additional validation and is discussed in Section 5.3.

5.3. Future Research Directions

Future research should extend PA-ADFT beyond the single-amplitude, single-specimen validation reported here toward broader SHM applications. Important directions include validation on damaged configurations, higher-order modal responses, variable excitation amplitudes, and larger civil infrastructure operating under environmental variability. These extensions require additional experiments and analysis beyond the controlled first-mode laboratory validation reported in this manuscript.

Extension to different structures and boundary conditions requires distinguishing which components of the framework transfer directly and which require adaptation. The calibrated stereo geometry, epipolar consistency checks, temporal tracking logic, and sequential acceptance concept are not specific to a cantilever boundary condition. Three components require adaptation: the beam-region segmentation, which presently relies on intensity thresholding against a controlled background and would need a learned or manually defined ROI for complex field imagery; the kinematic identity limit, which is derived from the expected motion amplitude and frame rate and must be recomputed for a different structure or excitation; and the measurement-quality acceptance thresholds, which presently reflect the first bending-mode validation and would need to be widened or removed for multi-mode identification. Field deployment also requires correction for camera motion, for example, through common-mode compensation using stationary background features or near-boundary reference points. Surface occlusion would reduce spatial coverage and may require additional synchronized stereo views. For larger structures, maintaining comparable depth precision over a larger measurement volume requires either a longer baseline or multiple synchronized stereo pairs.

6. Conclusions

Reliable target-free vision-based modal identification requires not only robust image correspondence but also measurements whose geometric and kinematic consistency is verified during acquisition. This study presented a PA-ADFT framework that integrates learned feature extraction, stereo-temporal association, physics-aware measurement validation, and calibrated stereo reconstruction within a unified sensing pipeline to generate three-dimensional camera-coordinate displacement histories for structural dynamic analy-

sis. The framework was experimentally validated through repeated free-vibration tests on an aluminum cantilever beam using synchronized strain-gauge measurements, an independently calibrated finite-element model, and Euler–Bernoulli beam theory as complementary validation references.

Experimental validation provides several important findings. First, PA-ADFT consistently generated reliable target-free measurements, maintaining 7–9 accepted structural measurement points with 90–95% median tracking coverage across repeated experiments and demonstrating robust measurement continuity despite variations in the detected feature population. Second, the reconstructed displacement histories reproduced the expected first-mode free-vibration response, including exponential amplitude decay and the characteristic cantilever deformation pattern, with amplitudes decreasing monotonically from the free end toward the fixed support. Third, the framework accurately identified the first bending modal frequency, yielding a median frequency of 7.02 Hz, compared with 7.03 Hz from synchronized strain measurements and 7.06 Hz from the finite-element model. Fourth, staged reprocessing of the same five experiments showed that SuperPoint–LightGlue–Lucas–Kanade tracking already recovers this frequency, that geometric and temporal filtering maintains track coverage, and that the kinematic identity gate prevents repaired correspondences from contaminating the reconstructed three-dimensional trajectories. Finally, stereo reconstruction recovered the three-dimensional camera-coordinate trajectory of each accepted point. Within this trajectory, the vertical component was quantitatively validated in both amplitude and frequency against the independent references, the identity of the stereo correspondences was preserved, and the lateral and depth components are reported as reconstructed quantities whose amplitudes were not independently validated.

This study demonstrates that reliable target-free modal identification depends on continuous enforcement of geometric, temporal, and kinematic consistency throughout the measurement process rather than on image correspondence alone. By integrating modern deep feature matching with physics-aware measurement validation, PA-ADFT provides a framework in which the vertical displacement response, the first-mode frequency, and the identity of stereo correspondences are established with physical evidence, while the lateral and depth components are reconstructed without independent quantitative validation. The present investigation was limited to the first bending mode of a healthy laboratory cantilever beam under controlled laboratory conditions; application to higher-order modes, damaged structural systems, and field-scale monitoring requires the additional validation steps identified in the discussion.

Author Contributions: Conceptualization, S.K. and L.H.; methodology, S.K. and A.A.; software, S.K. and A.A.; validation, S.K., A.A. and A.N.; formal analysis, S.K. and A.A.; investigation, S.K., A.A. and A.N.; resources, L.H.; data curation, S.K. and A.A.; writing—original draft preparation, S.K. and A.A.; writing—review and editing, S.K., A.A. and L.H.; visualization, S.K. and A.A.; supervision, L.H.; project administration, L.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The numerical values needed to interpret the reported comparisons are provided in the tables and figures. Raw stereo video records are not included with the manuscript because of data-volume constraints, but they can be made available from the corresponding author upon reasonable request and subject to institutional data-management requirements.

Acknowledgments: The authors acknowledge the University of Louisiana at Lafayette for institutional support.

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

References

1. Farrar, C.R.; Worden, K. An Introduction to Structural Health Monitoring. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2007**, *365*, 303–315. [CrossRef]
2. Doebling, S.W.; Farrar, C.R.; Prime, M.B.; Shevitz, D.W. *Damage Identification and Health Monitoring of Structural and Mechanical Systems from Changes in Their Vibration Characteristics: A Literature Review*; A-13070-MS; Los Alamos National Laboratories: Los Alamos, NM, USA, 1996.
3. Hwang, J.; Kareem, A.; Kim, W. Estimation of Modal Loads Using Structural Response. *J. Sound Vib.* **2009**, *326*, 522–539. [CrossRef]
4. Mardanshahi, A.; Sreekumar, A.; Yang, X.; Barman, S.K.; Chronopoulos, D. Sensing Techniques for Structural Health Monitoring: A State-of-the-Art Review on Performance Criteria and New-Generation Technologies. *Sensors* **2025**, *25*, 1424. [CrossRef]
5. Sony, S.; Laventure, S.; Sadhu, A. A Literature Review of Next-generation Smart Sensing Technology in Structural Health Monitoring. *Struct. Control Health Monit.* **2019**, *26*, e2321. [CrossRef]
6. Sabato, A.; Dabetwar, S.; Kulkarni, N.N.; Fortino, G. Noncontact Sensing Techniques for AI-Aided Structural Health Monitoring: A Systematic Review. *IEEE Sens. J.* **2023**, *23*, 4672–4684. [CrossRef]
7. Neyestani, A.; Picariello, F.; Tudosa, I.; Monaco, M.; De Vito, L.; D'Arco, M. Vision-Based Displacement Measurement for Structural Health Monitoring: A Metrology-Oriented Review of Uncertainty Quantification. *Buildings* **2026**, *16*, 2659. [CrossRef]
8. Shao, Y.; Li, L.; Li, J.; An, S.; Hao, H. Computer Vision Based Target-Free 3D Vibration Displacement Measurement of Structures. *Eng. Struct.* **2021**, *246*, 113040. [CrossRef]
9. Weng, J.; Huang, T.S.; Ahuja, N. *Motion and Structure from Image Sequences*; Springer Science & Business Media: New York, NY, USA, 2013; Volume 29.
10. Liu, C. Beyond Pixels: Exploring New Representations and Applications for Motion Analysis. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2009.
11. Zhang, Y.; Liang, X. Domain Adaptation in Structural Health Monitoring of Civil Infrastructure: A Systematic Review. *arXiv* **2025**, arXiv:2512.18780.
12. Liew, J.; Rashidi, M.; Khoa, L.; Matin, N.A.; Ehsan, S. Reviewing a Decade of Structural Health Monitoring in Footbridges: Advances, Challenges, and Future Directions. *Remote Sens.* **2025**, *17*, 2807. [CrossRef]
13. Shao, Y.; Li, L.; Li, J.; Li, Q.; An, S.; Hao, H. DIMMC: A 3D Vision Approach for Structural Displacement Measurement Using a Moving Camera. *Eng. Struct.* **2025**, *338*, 120566. [CrossRef]
14. Cui, D.; Wang, W.; Peng, J.; Zhang, Y.; Zhao, Y.; Chen, B.; Liu, Y.; Wang, J. Enhancing Cable Vibration Measurement at Long Distances through Super-Resolution Reconstruction and Target Foreground Segmentation. *Int. J. Struct. Stab. Dyn.* **2025**, *25*, 2550039. [CrossRef]
15. Tee-Ann, T.; Ko-Hsin, M.; Yuen, T.Y. A Markerless Photogrammetric Framework with Spatio-Temporal Refinement for Structural Deformation and Strain Monitoring. *Buildings* **2025**, *15*, 3584. [CrossRef]
16. Lam, P.T.; Nguyen, X.T.; Vy, V.; Lee, J.; Park, Y.; Jeong, J.-H.; Yoon, H. A Large Language Model-Aided Vision-Based Framework for Automated Structural Displacement Measurement. *Comput.-Aided Civ. Infrastruct. Eng.* **2026**, *48*, 100072. [CrossRef]
17. Li, J.-H.; Zhu, L.; Zhu, Y.-Y.; Su, R.K.-L.; Zhao, G.-Y.; Wu, H.-L.; Wang, X. A Feature Point Matching Approach Based on Deep Learning for Structural Vibration Displacement Measurement. *Measurement* **2026**, *259*, 119665. [CrossRef]
18. Guang, R.; Li, X.; Lei, Y.; Yang, B.; Li, N. Dynamic Vision-Based Machine Vibration Sensing and Fault Diagnosis with Signal Alignment and Feature Clustering. *Eng. Appl. Artif. Intell.* **2025**, *162*, 112445. [CrossRef]
19. Hong, Y.; Zhao, Z.; Tian, Q.; Li, S. Vibration Measurement in Machine Vision Based on ROI Gray-Scale Projection Feature Matching Algorithm. *Adv. Struct. Eng.* **2025**, *28*, 2488–2509. [CrossRef]
20. Guo, L.; Guo, W.; Chen, D.; Duan, B.; Shi, Z. Structural Vibration Measurement Based on Improved Phase-Based Motion Magnification and Deep Learning. *Mech. Syst. Signal Process.* **2025**, *224*, 111945. [CrossRef]
21. Shang, Z.; Shen, Z. Multi-Point Vibration Measurement and Mode Magnification of Civil Structures Using Video-Based Motion Processing. *Autom. Constr.* **2018**, *93*, 231–240. [CrossRef]
22. Cai, E.; Zhang, Y.; Quek, S.T. Visualizing and Quantifying Small and Nonstationary Structural Motions in Video Measurement. *Comput. Civ. Infrastruct. Eng.* **2023**, *38*, 135–159. [CrossRef]
23. Liu, W.; Zhou, W.; Liu, J.; Hu, P.; Cheng, J.; Han, J.; Lin, W. Modality-Aware Feature Matching: A Comprehensive Review of Single-and Cross-Modality Techniques. *arXiv* **2025**, arXiv:2507.22791.
24. Lindenberger, P.; Sarlin, P.-E.; Pollefeys, M. *Lightglue: Local Feature Matching at Light Speed*; IEEE: New York, NY, USA, 2023; pp. 17581–17592.
25. Cornelis, K.; Verbiest, F.; Van Gool, L. Drift Detection and Removal for Sequential Structure from Motion Algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 1249–1259. [CrossRef]
26. Li, D.; Shi, H.; Cai, B.; Bai, X.; Chen, L.; Han, G.; Mi, C. A Review of Technical Advances and Applications of Intelligent Inspection Robots in Structural Health Monitoring. *SmartBot* **2025**, *1*, e70000. [CrossRef]

27. Cao, H.; Qiao, D.; Han, M.; Yu, W.; Wang, B.; Shen, Y. Geometry-Aware Super-Resolution Fusion Calibration for Binocular Structured Light 3D Reconstruction. *Commun. Phys.* **2026**, *9*, 297. [CrossRef]
28. Xian, Q.; Cheng, H.; van der Zwaag, B.J.; Kromanis, R.; Incel, O.D. VFM-SDM: A Vision Foundation Model-Based Framework for Training-Free, Marker-Free, and Calibration-Free Structural Displacement Measurement. *arXiv* **2026**, arXiv:2605.09677.

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.