IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 24 March 2026, accepted 6 May 2026, date of publication 11 May 2026, date of current version 14 May 2026.

Digital Object Identifier 10.1109/ACCESS.2026.3692387

RESEARCH ARTICLE

A ROS-Based Multi-Camera Framework for Real-Time 3-D Structural Displacement Tracking

AMMAR AJMAL[1], TAHIR MEHMOOD[2], JONGBIN WON[3], AND CHANGSU SHIM[4]
[1]Department of Smart Cities, Chung-Ang University, Seoul 06974, Republic of Korea
[2]Department of Computer Science, Hochschule Bonn-Rhein-Sieg, 53757 Sankt Augustin, Germany
[3]Integrated Road Management Research Center, Korea Institute of Civil Engineering and Building Technology, Goyang 10223, Republic of Korea
[4]Department of Civil and Environmental Engineering, Chung-Ang University, Seoul 06974, Republic of Korea

Corresponding author: Changsu Shim (csshim@cau.ac.kr)

This work was supported by the National Research and Development Project for Smart Construction Technology, funded by Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure and Transport, and managed by Korea Expressway Corporation under Grant RS-2020- KA156007.

ABSTRACT Structural Health Monitoring (SHM) requires precise displacement measurements to ensure infrastructure safety and longevity. Vision-based methods offer a scalable, non-contact solution but are often constrained by synchronization and real-time processing challenges. This paper presents a Robot Operating System (ROS)-based framework for real-time 3D displacement tracking that leverages software-synchronized multi-camera systems and AprilTag fiducial pose estimation. The framework makes three tightly coupled contributions: 1) an end-to-end software synchronization pipeline that integrates host clock discipline, SDK capture-time timestamp propagation, and middleware tuple formation for real-time cross-host camera alignment without specialized hardware or post-hoc alignment; 2) a scalable architecture for distributed, multi-point tracking; and 3) a modular real-time pipeline enabling parallel acquisition and extensibility. In the primary LDS-referenced displacement validation experiments, the system achieved sub-millimeter accuracy for the validated in-plane displacement components (RMSE 0.14–0.40 mm) and maintained residual timing offsets below 5 ms across the tested configurations. Additional repeated-run LDS-referenced sinusoidal tests further confirmed repeatable best-view performance across excitation frequencies, with best-view mean RMSEs ranging from 0.096 to 0.261 mm. Subsequent modal testing verified relative frequency errors under 4% across all modes and MAC values exceeding 0.85 for the first four modes. These results demonstrate that consumer-grade cameras, when combined with software-only synchronization, can provide laboratory-grade precision in compact, field-oriented setups for real-time displacement tracking.

INDEX TERMS 3D displacement tracking, modal analysis, multi-camera ROS, robot operating system, structural health monitoring, time synchronization, vision-based displacement.

I. INTRODUCTION

Structural Health Monitoring (SHM) plays a critical role in ensuring the long-term safety, serviceability, and resilience of civil infrastructure, where failures in bridges, buildings, and other systems can have severe societal and economic impacts [1]. SHM enables continuous assessment of structural conditions over a system's service life, facilitating the early detection of anomalies and supporting condition-based maintenance strategies [2]. Among various response metrics, structural displacement has emerged as a particularly important indicator, as it directly reflects global deformation and stiffness degradation [3]. Displacement data are also central to serviceability evaluation and structural control, and are codified in design standards across major jurisdictions (e.g., the United States, South Korea, China, and the European Union) [3]. Consequently, accurate displacement tracking is foundational to both real-time decision-making and long-term performance validation. However, achieving such accuracy in real-world scenarios remains a significant challenge.

Despite the critical role of displacement monitoring in SHM, accurately capturing displacement in practice is hindered by

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Moinul Hossain.

© 2026 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

the limitations of the conventional sensing methods. Contact-based sensors such as linear variable differential transformers (LVDTs) provide high precision but require a stable reference frame, rendering them impractical for large or moving structures, including long-span bridges [4]. Accelerometer-based systems are widely used, but displacement estimation via double integration of acceleration signals suffers from cumulative drift and poor sensitivity to low-frequency or quasi-static movements [3]. Global Navigation Satellite System (GNSS) methods, while effective for long-term displacement tracking, are limited by low sampling rates and limited precision, restricting their use in dynamic monitoring or for capturing small-amplitude or high-frequency structural responses [6], [7]. Similarly, strain gauges and inclinometers can infer displacements only with dense sensor networks and complex data fusion, significantly increasing deployment complexity and cost [3], [8]. These limitations collectively underscore the need for flexible, scalable, and non-invasive sensing solutions capable of providing accurate displacement measurements in both static and dynamic environments across diverse infrastructure types.

To overcome the constraints of contact and inertial sensors, vision-based techniques have emerged as promising non-contact, scalable alternatives for displacement monitoring in SHM [4], [9]. High-resolution digital cameras can remotely track structural targets without requiring physical access, offering a cost-effective and adaptable solution for large-scale deployments. A single camera can monitor multiple points within its field of view, and prior studies have demonstrated accurate displacement measurements under controlled laboratory conditions [10], [11]. Recent advances in computer vision including improved 3D reconstruction, robust pattern tracking, and deep learning-based methods have further enhanced the accuracy and versatility of vision-based displacement estimation, making such approaches increasingly viable for SHM applications, though challenges remain for field deployment [9]. Recent reviews in construction and civil vision monitoring further emphasize the scalability and robustness challenges encountered in field environments, underscoring an ongoing transition toward automated, AI-driven monitoring frameworks [28].

Despite these advances, vision-based SHM systems continue to face critical challenges that limit their deployment in long-term or dynamic field environments. Displacement accuracy is highly sensitive to external factors such as lighting variation, occlusions, weather conditions, and unintended camera motion [9], [10], [12]. Maintaining line-of-sight and a stable reference frame is particularly difficult for flexible or mobile structures, often resulting in additional measurement errors [10], [12], [13]. While high-frame-rate imaging is essential for capturing dynamic displacements, real-time processing of high-speed videos, especially on edge devices, imposes significant computational demands [14]. Among these challenges, the lack of precise temporal synchronization across distributed camera systems remains a critical unresolved issue. Even small phase misalignments can undermine the reconstruction of coherent displacement fields during transient events. However, synchronization is often inadequately addressed in current SHM frameworks [5], [10], [15], underscoring the urgent need for real-time, robust, and scalable multi-camera solutions.

Although many vision-based SHM systems demonstrate strong performance under controlled laboratory conditions, their deployment at field scale remains limited by unresolved architectural and integration challenges [9]. Practical workarounds such as manual triggering or post-hoc alignment highlight the inherent difficulty of achieving precise, frame-level synchronization in distributed camera streams [15]. Conventional stereo-vision systems further impose strict spatial constraints by requiring overlapping fields of view and precise calibration, conditions that are difficult to maintain at scale and largely applicable only to stationary monitoring [5], [10], [13].

Alternative approaches also face limitations: GNSS-disciplined camera clocks provide high-precision timing but introduce complex and costly hardware, whereas low-cost GNSS solutions are insufficient for dynamic monitoring owing to low sampling rates and meter-level positioning accuracy [6], [16]. Advanced software-only synchronization methods, such as time-calibrated video with sub-frame interpolation and coded-light timestamping, have been proposed, yet they are rarely integrated into real-time SHM frameworks [17], [18]. Moreover, the existing vision pipelines often lack the modularity and extensibility needed for seamless integration into broader monitoring platforms. While the Robot Operating System (ROS) offers a robust foundation for managing real-time, multi-sensor systems, its potential for synchronized, vision-based SHM remains underexplored, particularly in achieving millisecond-level temporal alignment with low-cost, off-the-shelf hardware. These limitations collectively underscore the urgent need for a ROS-integrated, scalable framework capable of real-time, multi-point displacement monitoring in field-oriented scenarios.

For distributed vision-based SHM, temporal coherence depends on complementary synchronization functions at different system layers. Host-level clock discipline improves timestamp consistency across acquisition units, while middleware-level association groups temporally corresponding observations into synchronized tuples for downstream processing. Prior multi-camera vision studies such as Panigati et al. [15] typically address synchronization through *a priori* triggering or *a posteriori* alignment. In contrast, the present work integrates clock discipline, capture-time timestamp propagation, and middleware-level tuple formation into an integrated real-time pipeline tailored to synchronized displacement tracking and modal analysis.

To address these challenges, this paper introduces three tightly coupled contributions for real-time multi-camera displacement monitoring in SHM.

1) **End-to-end software synchronization for distributed vision-based SHM.** The framework integrates host clock discipline, capture-time timestamp propagation, and middleware-level tuple formation into an integrated real-time synchronization pipeline, enabling cross-host comparable timestamps and synchronized observation tuples without specialized hardware.
2) **Scalable multi-camera architecture.** A distributed design that supports multi-point displacement tracking across large structures, while also enabling redundant multi-view observation where needed, thereby overcoming the spatial constraints of conventional stereo-vision and single-camera systems.
3) **Modular real-time processing pipeline.** A framework designed for parallel acquisition and extensibility in practical monitoring scenarios, improving robustness under varying operating conditions and enabling immediate structural response analysis.

These contributions are realized within a modular ROS architecture that integrates synchronized video streams from multiple cameras into a unified displacement monitoring framework. By combining host clock discipline, capture-time timestamp propagation, and middleware-level tuple formation, the proposed method achieves millisecond-level temporal coherence entirely in software, avoiding hardware triggering and post-hoc alignment commonly adopted in prior multi-camera vision systems [15]. The distributed architecture supports parallel acquisition and processing, while a custom graphical interface facilitates network control, experiment execution, and data logging. The framework is rigorously validated through laser triangulation displacement sensor (LDS)-referenced displacement benchmarking and multi-floor modal analysis of a five-story shear building model, demonstrating laboratory-grade displacement tracking performance within a scalable, field-oriented monitoring framework.

The remainder of this paper is organized as follows: Section II details the methodology and system architecture, including the calibration, synchronization, and experimental setup. Section III reports the experimental findings, followed by performance analysis and discussion in Section IV. Section V concludes the paper with key takeaways and directions for future research.

II. METHODOLOGY

The proposed methodology integrates the hardware, software, and validation components into a comprehensive framework for vision-based structural monitoring. Our design explicitly addresses three critical challenges: (i) spatial accuracy through distributed multi-camera coordination, (ii) temporal coherence via software-based synchronization, and (iii) real-time performance through parallel and modular processing. This section presents these components in logical order, beginning with the system architecture, followed by the sensing and estimation, synchronization, hardware setup, and evaluation protocols.

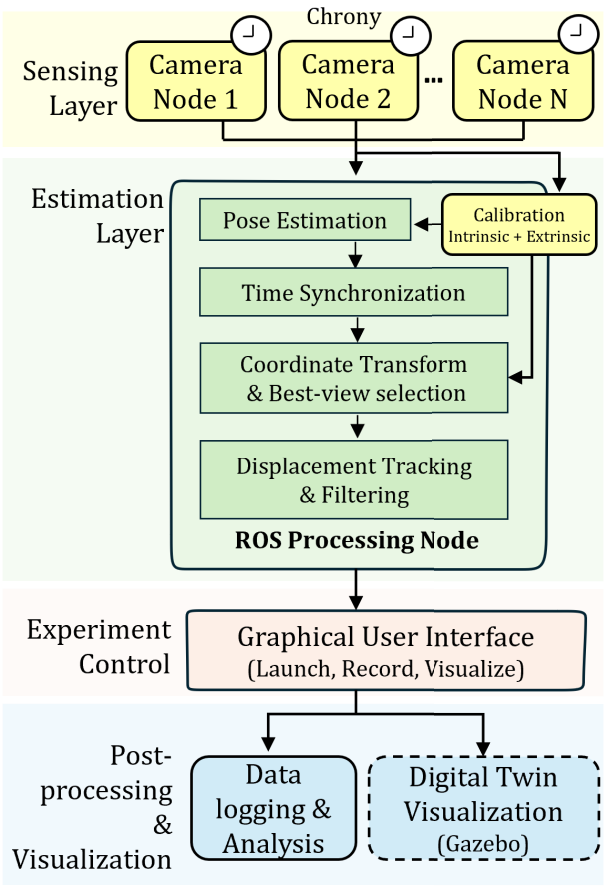


FIGURE 1. System architecture of the ROS-based multi-camera framework.

A. SYSTEM ARCHITECTURE

We implemented a ROS-native multi-camera pipeline for real-time 3D displacement monitoring. The architecture comprises four functional layers: (1) sensing, (2) estimation, (3) experimental control, and (4) post-processing. ROS topics and node composition provide modularity by separating acquisition, estimation, and logging, thereby supporting scalable, low-latency operation. A Gazebo-based visualization component offers real-time 3D rendering of structural motion through ROS's native simulation integration. Temporal alignment is handled entirely in software via host clock discipline, capture-time timestamp propagation, and middleware-level synchronization, eliminating the need for hardware triggers. Fig. 1 summarizes the workflow.

At the ROS graph level, the implementation consists of one acquisition node and one AprilTag-based relative motion estimation node per camera host, together with a shared control/logging layer for synchronization, recording, and experiment management. Each acquisition node publishes raw image frames and camera calibration data, while the corresponding relative motion estimation node subscribes

TABLE 1. Principal ROS nodes, topics, and message types used in the implementation.

Node name	Topic name(s)	Message type	Function

`/sony_cam1`
`/sony_cam2`
`/sony_cam3`
	`/sony_cam`*i*`/image_raw`
	`/sony_cam`*i*`/camera_info`
		`sensor_msgs/Image`
		`sensor_msgs/CameraInfo`
			Per-camera acquisition nodes publishing image frames and calibrated camera information.

`/sony_cam1_apriltag_node`
`/sony_cam2_apriltag_node`
`/sony_cam3_apriltag_node`
	Sub.: `/sony_cam`*i*`/image_raw`,
	`/sony_cam`*i*`/camera_info`
	Pub.: `/sony_cam`*i*`/fiducial_transforms`
		`sensor_msgs/Image`
		`sensor_msgs/CameraInfo`
		`fiducial_msgs/`
		`FiducialTransformArray`
			AprilTag-based relative motion estimation nodes performing detection, pose recovery, reference initialization, and outlier handling.

`/multi_camera_sync_logger`	Sub.: `/sony_cam1/fiducial_transforms`
	`/sony_cam2/fiducial_transforms`
	`/sony_cam3/fiducial_transforms`
		`fiducial_msgs/`
		`FiducialTransformArray`
			Master-host synchronization and logging node using ROS `ApproximateTime` (queue size 10, slop $\Delta t = 10$ ms) to form synchronized tuples and record aligned outputs.

TABLE 2. Calibration settings and intrinsic accuracy.

Parameter	Value
Target board	ChArUco 9×6 (8 mm squares), AprilTag marker 6 mm
Views per camera	20–25 from 0.5–2.5 m
Mean reprojection error	< 0.25 px (median), max accepted 0.5 px
Model	Pinhole + 5-parameter radial–tangential

to these streams, performs AprilTag detection and relative transform recovery, and publishes stabilized relative transform estimates used for downstream displacement analysis. The control/logging layer then subscribes to the per-camera relative transform topics, forms synchronized multi-camera tuples at the middleware level, and records the aligned outputs for downstream displacement analysis, visualization, and validation. Table 1 summarizes the principal ROS nodes, topics, and message types used in the acquisition, relative motion estimation, and control/logging layers of the implementation.

The framework was deployed on three Intel NUC PCs connected over Ethernet under ROS Noetic, with one camera-processing pipeline running per host and the central ROS node hosted on the first NUC.

B. SENSING AND CALIBRATION

Each camera produces timestamped image streams that are calibrated before the run-time operation. Calibration comprises two stages: (i) intrinsic parameter estimation and (ii) initialization of camera-to-world extrinsics. Intrinsic parameters (f_x, f_y, c_x, c_y) and lens distortion coefficients $(k_1, k_2, p_1, p_2, k_3)$ are estimated from multi-view captures of a ChArUco target using standard OpenCV procedures [20], [21], [22]. Frames with large reprojection residuals or incomplete AprilTag marker visibility are discarded, and accepted images are undistorted before downstream processing.

A summary of the calibration configuration, acceptance thresholds, and resulting intrinsic accuracy is provided in Table 2.

The calibration quality is quantified by the root-mean-square (RMS) reprojection error, defined as:

$$\mathrm{RMS} = \sqrt{\frac{1}{N}\sum_{k=1}^{N} \left\| \hat{\mathbf{x}}_k - \mathbf{x}_k \right\|^2}, \tag{1}$$

where N denotes the number of detected corners. Frames with reprojection error above 0.5 px were discarded, consistent with thresholds commonly applied in vision-based calibration workflows to prevent geometric distortion artifacts [22]. The retained frames produced a median reprojection error of less than approximately 0.25 px across all cameras, corresponding to sub-pixel calibration accuracy widely regarded as sufficient for reliable pose estimation in displacement measurement applications [22].

Intrinsic and extrinsic parameters were fixed for the duration of each run and subsequently used to produce metric, registered displacements for downstream analysis.

C. 3D DISPLACEMENT ESTIMATION

The displacement estimation pipeline comprises four stages: (i) AprilTag marker detection and pose recovery, (ii) extrinsic initialization, (iii) view quality assessment with best-view selection, and (iv) temporal stabilization through outlier rejection and filtering.

1) DETECTION AND POSE RECOVERY

AprilTag corners are detected in rectified images and mapped to known tag coordinates:

$$\mathbf{X}_k = \left[\pm s/2,\ \pm s/2,\ 0 \right]^\top, \quad k = 1 \ldots 4,$$

where s denotes the tag size. With calibrated intrinsics $\mathbf{K}$ and distortion coefficients $\mathbf{d}$, the tag-to-camera pose ${}^{c_i}T_t$ is estimated using a Perspective-n-Point (PnP) formulation:

$$\min_{R,t} \sum_{k=1}^{4} \left\| \pi(\mathbf{K}, \mathbf{d},\ R\mathbf{X}_k + t) - \mathbf{x}_k \right\|^2, \tag{2}$$

where $\pi(\cdot)$ denotes the camera projection function and $\mathbf{x}_k$ are the observed image points. This step establishes metric pose estimates, forming the basis for displacement tracking.

2) EXTRINSIC INITIALIZATION

A short initialization window of the first M synchronized detections is averaged to define camera-to-world transformations ${}^{w}T_{c_i}$, with the world frame anchored at the tag pose at $t = 0$. Averaging across multiple frames reduces the effect of noise while ensuring a consistent reference frame for subsequent displacement estimation.

3) VIEW QUALITY ASSESSMENT AND BEST-VIEW SELECTION

In multi-camera configurations, per-frame observations are evaluated to mitigate degradation arising from foreshortening or partial occlusion. Each candidate view is assigned a quality score

$$q_i(k) = \frac{A_i(k)}{\rho_i(k)},$$

where $A_i(k)$ is the projected tag area and $\rho_i(k)$ is the reprojection RMS. Views failing basic quality gates (e.g., excessively small area or unstable reprojection residuals) are discarded, and the highest-scoring observation $i^\star(k)$ is selected as the displacement estimate for that frame. This lightweight selection strategy enhances robustness while avoiding the computational cost of full multi-view fusion. The rationale is to prioritize the most geometrically reliable view at each instant, thereby suppressing degraded observations without imposing an additional processing overhead.

4) TEMPORAL STABILIZATION

Frame-to-frame pose estimates are further refined by rejecting abrupt outliers exceeding a displacement threshold τ, chosen in accordance with expected structural displacement bounds. Minimal smoothing is then applied using a Savitzky–Golay filter to suppress noise while preserving dynamic response content.

While the individual stages of the pipeline build on established computer vision methods, their integration within a synchronized multi-camera framework tailored to real-time SHM forms a key contribution of this work. The inclusion of runtime best-view selection further improves robustness without the computational burden of full multi-view fusion. Together, these adaptations support real-time displacement tracking with sub-millimeter measurement performance and provide the basis for the temporal synchronization strategy described in the following subsection.

D. TEMPORAL SYNCHRONIZATION

Temporal coherence across distributed acquisition units was achieved through two complementary synchronization layers: (i) host clock discipline and (ii) middleware-level tuple formation. In the implemented ROS graph, middleware synchronization was applied to the per-camera `fiducial_msgs/FiducialTransformArray` relative transform streams rather than to raw image topics, thereby reducing bandwidth while preserving timing information required for aligned multi-camera displacement analysis.

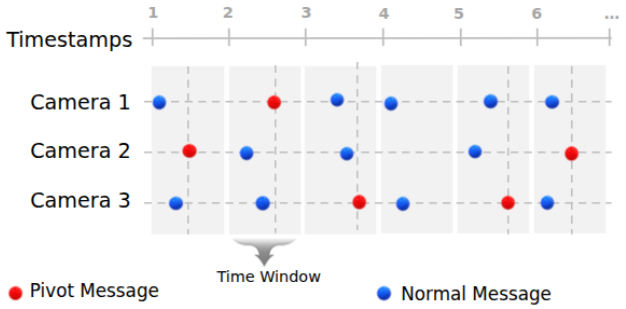


FIGURE 2. ROS `ApproximateTime` synchronization across three cameras, where pivot messages (red) are matched with normal messages (blue) within a tolerance window.

1) HOST CLOCK DISCIPLINE

Each acquisition unit (Intel NUC) was synchronized using Chrony [23], which maintained millisecond-level alignment with a common network reference. This step improved the consistency of capture-time timestamps provided by the camera software development kit (SDK) across distributed devices, establishing a common temporal basis for subsequent processing.

2) MIDDLEWARE TUPLE FORMATION

At the ROS middleware layer, timestamped relative transform messages were grouped into synchronized tuples using the `ApproximateTime` policy. By propagating SDK capture times into ROS message headers, tuple formation was performed using capture-referenced timestamps rather than publication-time variability. A tuple was accepted when

$$\max_{i,j} |t_i - t_j| \leq \Delta t, \tag{3}$$

where t_i and t_j are SDK capture-time timestamps and Δt is the allowable offset. This rule ensures that grouped observations correspond to the same physical event, a requirement that is particularly critical for phase-coherent displacement and modal analysis in SHM applications.

For the experiments reported in this paper, synchronized tuples were formed using ROS `ApproximateTime` with a queue size of 10 and a synchronization slop of $\Delta t = 10$ ms.

Fig. 2 illustrates this process: each pivot message (red) is matched with messages from other cameras (blue) that fall within the tolerance window. If all streams satisfy the condition, the tuple is published as a synchronized set, maintaining coherence despite residual offsets or network jitter.

To minimize capture-to-publish jitter, SDK-level capture timestamps were propagated into ROS headers and used in place of ROS publication times [15], [19]. In this work, $\Delta t = 10$ ms ($\approx$ 60% of a 60 frames per second (FPS) frame interval), which represents a practical compromise: large enough to absorb residual offsets, yet below one frame period to avoid grouping adjacent frames [17], [18]. At 60 FPS,

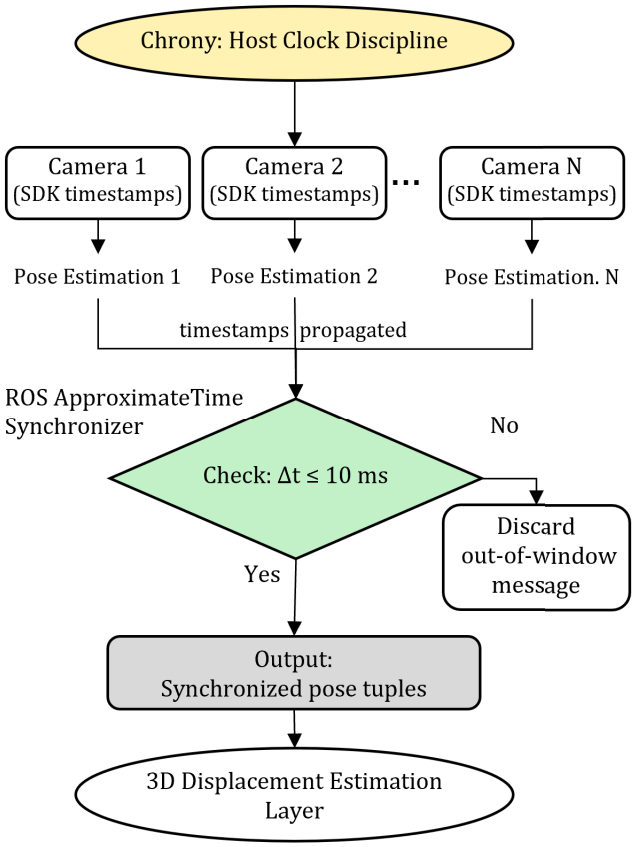


FIGURE 3. Workflow for temporal synchronization across distributed cameras.

the system supports dynamic response analysis up to 30 Hz under the tested conditions, covering the fundamental and low-order modes of typical laboratory-scale and many full-scale structures [24], [25].

3) PHASE ERROR BOUND

For harmonic motion at frequency f, a misalignment of Δt corresponds to a worst-case phase error:

$$\Delta\phi_{\max} \leq 2\pi f \, \Delta t. \tag{4}$$

For the 0–30 Hz excitations considered in this paper, the resulting phase error remained within the synchronization tolerance used for the reported modal identification tests.

Fig. 3 summarizes the overall workflow, where host clock discipline and SDK timestamps enable ROS `ApproximateTime` tuple formation, producing synchronized relative transform data for downstream displacement estimation.

E. HARDWARE AND EXPERIMENTAL SETUP

The experimental platform combined a distributed camera network with an LDS as a high-frequency reference for validation. AprilTag markers were mounted on test specimens at stand-off distances of 2–5 m to maintain visibility from various perspectives. Hardware and software specifications are presented in Table 3.

TABLE 3. Hardware and software specifications used in the experimental setup.

Component	Specification
Cameras	Sony RX10 IV (nominal 60 FPS, manual exposure)
Lenses	Built-in zoom; manual focus set per test
Video capture	AVerMedia Live Streamer Cap 4K (USB 3.0 HDMI input)
Processing units	Intel NUC 13 Pro (Core i7, 16 GB RAM)
Synchronization	Chrony Network Time Protocol (NTP) (1 server, multiple clients); ROS `ApproximateTime` (slop $\Delta t = 10$ ms)
Markers	AprilTag markers (36h11, 20–35 mm side length)
Reference sensor	optoNCDT 1420 LDS (1000 Hz)
Stand-off distance	2–5 m (experiment dependent)
Software	Ubuntu 20.04 LTS, ROS Noetic, OpenCV, AprilTag library

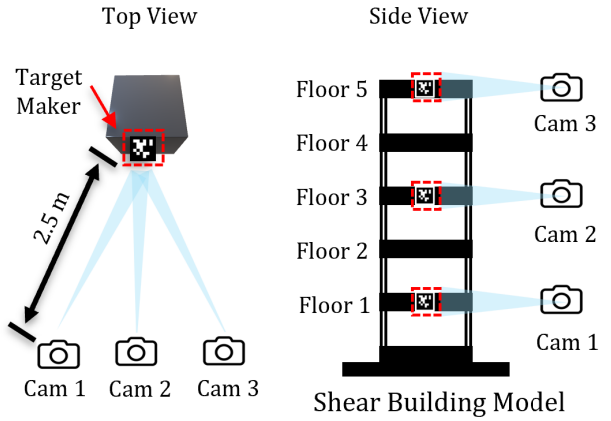


FIGURE 4. Camera configurations: (left) triangulated multi-view setup tracking a single AprilTag marker; (right) distributed setup with cameras tracking AprilTag markers on separate floors.

Two complementary camera configurations were employed (Fig. 4): (i) a *multi-view triangulated setup*, where three cameras simultaneously tracked a single AprilTag marker to provide redundant coverage, and (ii) a *distributed setup*, where each camera tracked a different AprilTag marker attached to separate floors of the shear building model.

Beyond the individual camera configurations, the complete experimental environment is shown in Fig. 5, where the shear building specimen is mounted on a shaking table and instrumented with AprilTag markers, three cameras, and the LDS reference sensor for simultaneous measurement.

Three specimen setups were investigated to evaluate system performance: (a) a cantilever beam with a single AprilTag marker at the free tip to capture free-vibration response along the X-axis under impulse excitation, (b) an acrylic box affixed to the shaking table with a front-facing AprilTag marker for measuring Y-axis displacements under sinusoidal excitation, and (c) a five-story shear building model mounted on the shaking table with AprilTag markers on the 1st, 3rd, and 5th floors for distributed modal response measurement. The representative configurations are shown in Fig. 6. Together, these setups enabled evaluation of displacement accuracy, synchronization fidelity, and modal response, forming the basis for the evaluation protocol in the next subsection.

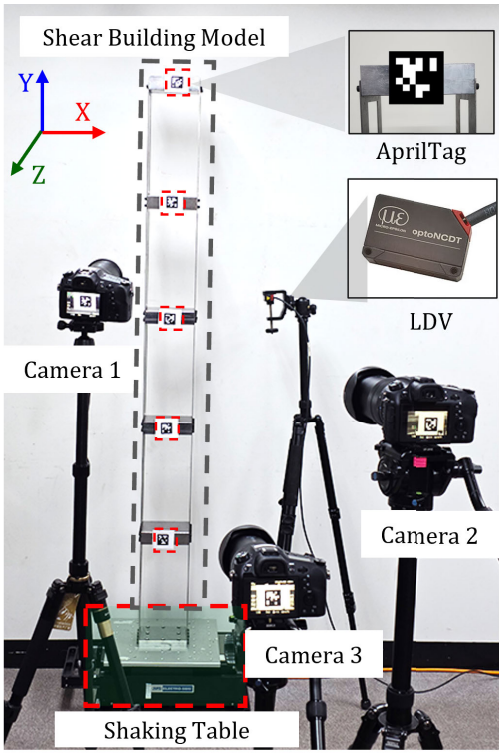
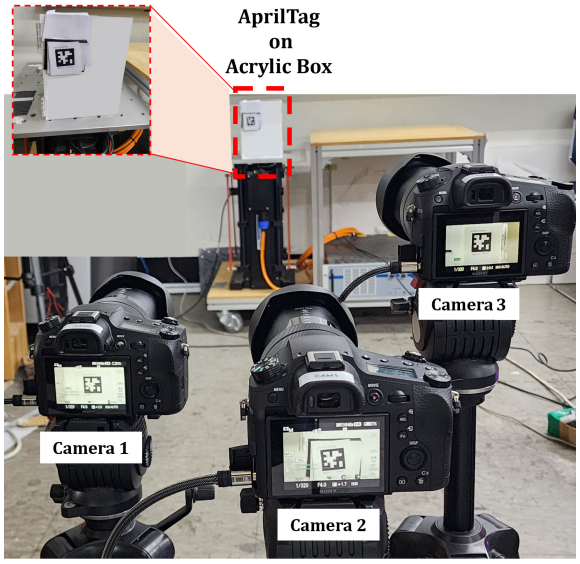
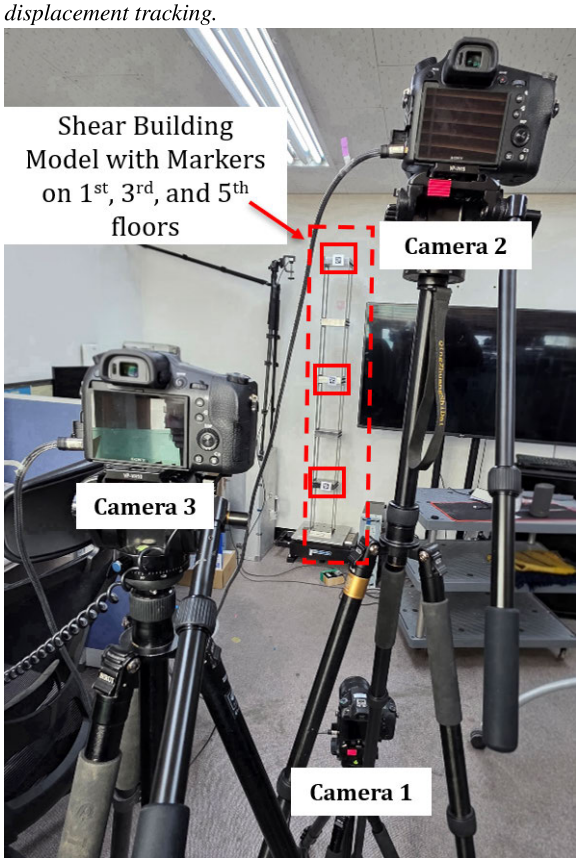


FIGURE 5. Complete experimental environment with shear building model mounted on a shaking table and instrumented with AprilTag markers, three cameras, and an LDS reference sensor.



(a) *Acrylic box with front-facing AprilTag marker for horizontal displacement tracking.*



(b) *Shear building with AprilTag markers on 1st, 3rd, and 5th floors observed by distributed cameras.*

FIGURE 6. Test configurations for system validation.

F. EVALUATION PROTOCOL

The system performance was evaluated across three key aspects: displacement accuracy, temporal synchronization, and modal response validation. Each dimension targets a distinct component of the proposed framework and is assessed using established metrics with appropriate reference baselines.

To assess repeatability and uncertainty under controlled dynamic conditions, an additional repeated-run LDS-referenced sinusoidal validation was performed. Sinusoidal excitation tests at 1, 3, 5, 7, 9, and 11 Hz were each repeated ten times under the same camera configuration, calibration, and acquisition settings, yielding 60 runs in total. For the repeated-run statistical analysis only, LDS-referenced displacement metrics were computed separately for each camera view in each run, and the view yielding the lowest run-level RMSE was retained as a post hoc best-view summary for aggregation across repetitions. Across repeated runs, RMSE and correlation were summarized using the mean and standard deviation, and a 95% confidence interval was additionally reported for RMSE to quantify uncertainty in repeatability performance.

1) DISPLACEMENT ACCURACY

Root-mean-square error (RMSE) was computed against LDS reference measurements:

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{k=1}^{T}\left(y_{\mathrm{vision}}(k) - y_{\mathrm{ref}}(k)\right)^2}, \tag{5}$$

where $y_{\mathrm{vision}}(k)$ and $y_{\mathrm{ref}}(k)$ denote the vision-based and LDS displacements at time k, respectively. RMSE was expressed in millimeters and evaluated per axis. Although the framework estimates full relative 3D motion from monocular AprilTag pose recovery, the quantitative displacement accuracy

TABLE 4. Evaluation metrics and corresponding validation methods.

Analysis Type	Test Configuration	Primary Metrics	Validation
Displacement	Cantilever Beam	RMSE (X-axis), NRMSE, correlation, PSD	LDS reference
	Acrylic Box	RMSE (Y-axis), NRMSE, correlation, PSD	LDS reference
	Shear Building	RMSE (X-axis), NRMSE, correlation, PSD	LDS reference
Temporal Synchronization	Cantilever Beam Acrylic Box Shear Building	Cross-correlation lag, Coherence, Phase Error, Jitter, Residual Disp. Std. Dev.	LDS reference
Modal Analysis	Shear Building (full structure)	Natural frequencies, Absolute Error (Hz), MAC	FE model

assessment in this study was restricted to in-plane components for which a reliable LDS-based reference was available under the selected experimental configurations. The frequency content was analyzed using Welch power spectral density (PSD), and peak frequency alignment between vision and LDS was examined. Correlation and normalized RMSE (NRMSE) were also reported to assess agreement with the LDS reference.

2) TEMPORAL SYNCHRONIZATION

Synchronization performance was assessed in terms of inter-camera alignment, phase accuracy, and dynamic coherence relative to the LDS reference. Cross-correlation lag was measured between synchronized camera pairs to quantify residual timing offsets. Phase error was analyzed across excitation frequencies to evaluate dynamic accuracy, while jitter was characterized to capture variability in alignment. These metrics provide a practical assessment of whether millisecond-level coherence was maintained across cameras, which is critical for reliable displacement and modal analysis in SHM applications.

3) MODAL RESPONSE VALIDATION

Modal parameters of the shear building were identified experimentally and compared with finite element (FE) predictions. Validation focused on natural frequency and mode shape comparison, with correlation quantified by the Modal Assurance Criterion (MAC). Mode shape visualization and frequency response analysis supported this comparison under multiple excitation conditions. The MAC, widely adopted in structural dynamics, was used here to benchmark experimental modes against theoretical predictions. For comparison with analytical predictions, story stiffness and mass parameters were updated using a constrained optimization to align simulated and experimentally identified natural frequencies.

TABLE 5. Displacement accuracy metrics for the cantilever beam test (X-axis).

Camera	RMSE (mm)	NRMSE (%)	Correlation
Cam1	0.40	3.8	0.991
Cam2	0.35	3.2	0.993
Cam3	0.28	2.5	0.995

Mode-shape correlation was quantified using the MAC, evaluated only at the measured floors (1, 3, 5). Experimental and theoretical modes were normalized and sign-aligned for consistency, and results were visualized using the same measured DOFs reported in Section III.

Table 4 summarizes the evaluation metrics and their corresponding test configurations. All metrics were computed using time-synchronized data streams, with the LDS serving as both accuracy reference and timing baseline. Results are presented in Section III, with acceptance thresholds defined by structural monitoring requirements ($<$0.5 mm RMSE, $<$5 ms synchronization error).

These analyses together formed a structured evaluation approach that connected spatial accuracy, timing consistency, and modal behavior. Overall, the tests provided a clear picture of how accurately and reliably the system measures time-synchronized structural displacements at sub-millimeter scale under the tested configurations. The detailed results are presented in Section III.

III. RESULTS

The system performance was evaluated across three validation axes: displacement accuracy, temporal synchronization, and modal response identification. These dimensions are aligned with the two experimental configurations illustrated in Fig. 4. The *multi-view triangulated setup* (left), in which three cameras tracked a single AprilTag marker, underpins the displacement accuracy tests (cantilever, acrylic box, and shear building single-marker), whereas the *distributed setup* (right), in which cameras tracked markers on different floors, enables modal analysis of multi-story response.

A. DISPLACEMENT ACCURACY VALIDATION

The displacement tracking accuracy was evaluated in three experimental configurations (cantilever beam, acrylic box, and shear building) using the *multi-view triangulated setup* (Fig. 4, left). All three cameras observed the same AprilTag marker, enabling direct comparison of vision-based estimates against the high-precision LDS reference. Reported results therefore focus on in-plane displacement components (X and Y axes), consistent with the validation scope defined in Section II-F. While the framework also estimates out-of-plane motion through monocular pose recovery, that component was not quantitatively benchmarked against an external reference in the present study. The following subsections report time- and frequency-domain comparisons together with quantitative error metrics.

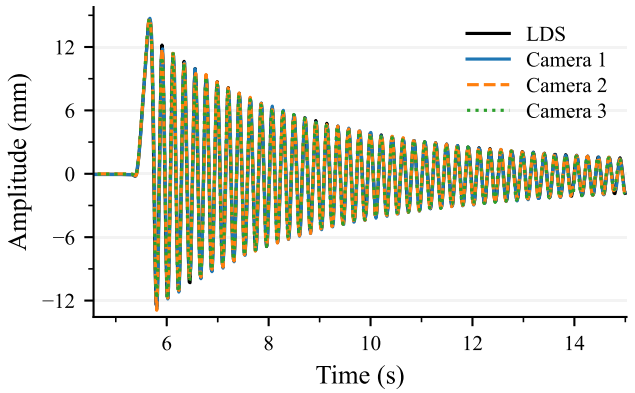


FIGURE 7. Time-domain displacement of the cantilever beam (X-axis) measured by all cameras compared with the LDS reference.

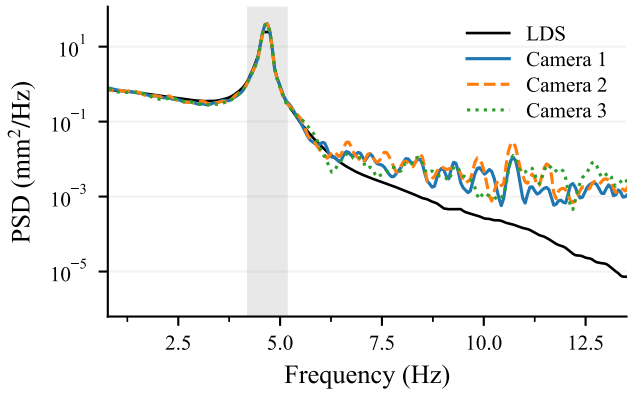


FIGURE 8. Frequency-domain PSD of cantilever beam displacement (X-axis) from cameras and LDS reference, highlighting the dominant resonance at 4.69 Hz.

1) CANTILEVER BEAM (X-AXIS DISPLACEMENT)

The first test examined free vibration of a vertical cantilever beam following an impulse excitation. Figure 7 shows the time-domain displacement history, while Fig. 8 presents the corresponding PSD spectra, confirming accurate capture of the dominant resonance at 4.69 Hz.

Table 5 summarizes the displacement accuracy metrics for this configuration.

2) ACRYLIC BOX (Y-AXIS SINUSOIDAL EXCITATION)

The second validation used an acrylic box mounted on a shaking table and excited with 1–5 Hz sinusoidal inputs along the Y-axis. Figure 9 compares the time-domain displacements. Fig. 10 shows the PSD analysis, which indicates close agreement of camera measurements with the LDS reference.

Table 6 summarizes the quantitative metrics.

3) SHEAR BUILDING (BROADBAND SINE SWEEP, X-AXIS DISPLACEMENT)

To evaluate the broadband performance, a five-story shear building was subjected to a 0–30 Hz sine sweep with a single AprilTag marker placed on the top floor. Figure 11 shows the synchronized displacements captured by the cameras compared with those of the LDS across the full sweep.

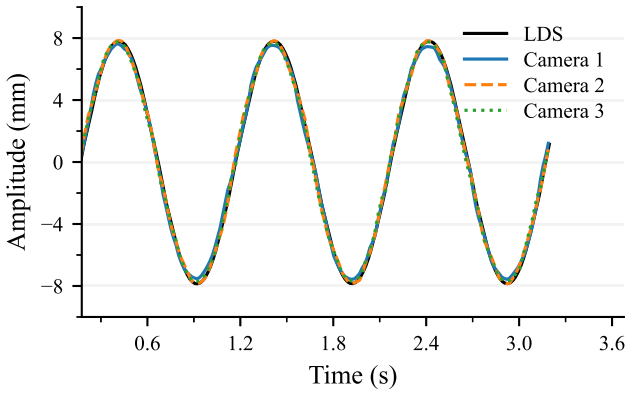


FIGURE 9. Time-domain displacement of the acrylic box (Y-axis sinusoidal excitation) from cameras compared with LDS reference.

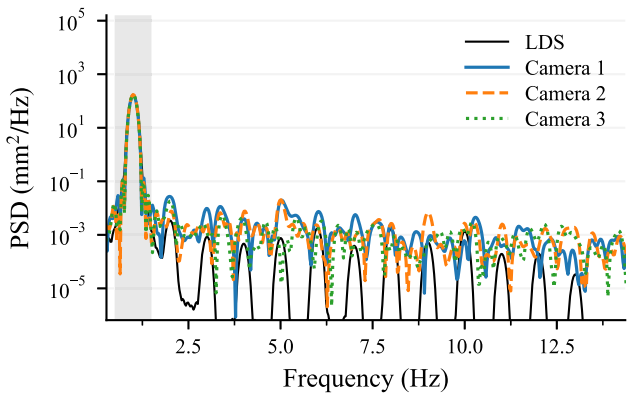


FIGURE 10. Frequency-domain PSD of acrylic box displacement (Y-axis sinusoidal excitation) from cameras and LDS reference, highlighting the dominant resonance at 0.996 Hz.

TABLE 6. Displacement accuracy metrics for the acrylic box test (Y-axis sinusoidal excitation).

Camera	RMSE (mm)	NRMSE (%)	Correlation
Cam1	0.32	3.0	0.998
Cam2	0.26	2.5	0.999
Cam3	0.29	2.7	0.998

TABLE 7. Displacement accuracy metrics for the shear building sine-sweep test (X-axis).

Camera	RMSE (mm)	NRMSE (%)	Correlation
Cam1	0.15	1.3	0.999
Cam2	0.18	1.5	0.999
Cam3	0.14	1.1	0.999

Table 7 summarizes the accuracy metrics used in the test.

4) CONSOLIDATED BEST-VIEW SUMMARY

Table 8 shows the best-performing camera views from each validation test. In all cases, displacement accuracy was within the sub-millimeter RMSE range, with normalized errors below 3% and correlation coefficients approaching unity.

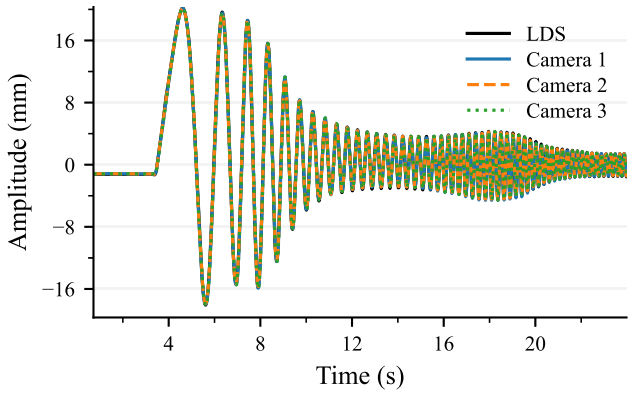


FIGURE 11. Time-domain displacement of the five-story shear building under 0–30 Hz sine sweep (X-axis, top-floor AprilTag marker) from cameras compared with the LDS reference.

TABLE 8. Summary of best-view displacement accuracy across all validation tests.

Test Configuration	RMSE (mm)	NRMSE (%)	Correlation
Cantilever beam (X)	0.28	2.5	0.995
Acrylic box (Y, 1–5 Hz)	0.26	2.5	0.999
Shear building (sweep)	0.14	1.1	0.999

To further quantify statistical robustness, repeated-run LDS-referenced sinusoidal tests were analyzed at excitation frequencies of 1, 3, 5, 7, 9, and 11 Hz, with ten runs per frequency for a total of 60 runs. Table 9 reports the best-view repeated-run statistics. Across all frequencies, the best-view RMSE remained within the sub-millimeter range, with mean values between 0.096 and 0.261 mm, while correlation remained high, ranging from 0.976 to 0.999. The overall best-view RMSE across all repeated runs was 0.181 ± 0.086 mm with a 95% confidence interval of [0.159, 0.203] mm, and the overall correlation was 0.989 ± 0.010. These repeated-run results strengthen the single-run validation results by showing that the proposed framework maintains stable LDS-referenced displacement accuracy across repeated trials under controlled sinusoidal excitation.

B. TEMPORAL SYNCHRONIZATION

Temporal synchronization performance was assessed across the same cantilever, acrylic box, and shear building experiments. The camera displacement traces were compared with the LDS reference to evaluate the residual timing offsets, coherence, phase fidelity, and related synchronization metrics. Cross-correlation analysis indicated offsets in the range of 1–3 ms. Coherence at the dominant frequency was at least 0.98 in all representative scenarios, while phase deviations remained small overall, with a maximum observed deviation of 6.5°. The signal-to-noise ratio (SNR) varied from 16–20 dB in the cantilever test to more than 30 dB in the sine-sweep test, while the residual standard deviations remained within 0.1–0.5 mm. Pairwise inter-camera lag, inferred from LDS-referenced lags, was consistently within 1.5 ms across all the experiments. Representative metrics for temporal synchronization are listed in Table 10.

TABLE 9. Best-view repeatability statistics from repeated LDS-referenced sinusoidal tests, reported as mean ± standard deviation with 95% confidence intervals for RMSE.

Freq. (Hz)	Runs	RMSE (mm)	95% CI (mm)	Correlation
1	10	0.188 ± 0.058	[0.146, 0.229]	0.999 ± 0.000
3	10	0.245 ± 0.087	[0.183, 0.308]	0.993 ± 0.007
5	10	0.261 ± 0.113	[0.180, 0.342]	0.992 ± 0.007
7	10	0.158 ± 0.042	[0.128, 0.188]	0.993 ± 0.005
9	10	0.138 ± 0.027	[0.118, 0.157]	0.982 ± 0.008
11	10	0.096 ± 0.012	[0.087, 0.105]	0.976 ± 0.007
Overall	**60**	0.181 ± 0.086	[0.159, 0.203]	0.989 ± 0.010

TABLE 10. Temporal synchronization metrics for representative experiments relative to the LDS reference.

Test	Camera	Lag (ms)	Coherence	Phase (°)	SNR (dB)	Residual Std. Dev. (mm)
Cantilever	Cam1	1.2	0.99	–5.1	17.1	0.42
	Cam2	0.8	0.99	1.5	19.8	0.31
	Cam3	2.3	0.98	6.5	16.5	0.46
Sine (1 Hz)	Cam1	1.5	0.99	2.2	23.1	0.39
	Cam2	0.9	0.99	1.4	26.5	0.26
	Cam3	1.8	0.98	2.4	25.0	0.31
Sweep (0–30 Hz)	Cam1	2.0	0.99	–0.7	25.5	0.15
	Cam2	1.1	0.99	–0.2	31.1	0.08
	Cam3	2.5	0.98	–1.2	23.6	0.19

C. MODAL ANALYSIS RESULTS

The modal properties of the five-story shear building were identified using a distributed multi-camera setup, with AprilTag markers placed on the 1st, 3rd, and 5th floors. A 0–30 Hz sine sweep excitation was applied, and the responses were processed using Frequency Domain Decomposition (FDD).

The singular value spectrum (Fig. 12) exhibited five resonance peaks corresponding to the fundamental vibration modes. The identified natural frequencies are listed in Table 11, along with the theoretical predictions, absolute errors, and MAC values. Absolute errors remained below 0.15 Hz, and relative errors were under 4% across all modes (Fig. 13). Although the frequency estimates remained accurate overall, higher-order modal behavior should be interpreted together with the sparse spatial sampling used in this study, since only floors 1, 3, and 5 were instrumented for mode-shape recovery.

The experimental mode shapes at the instrumented floors were compared with the theoretical predictions in Fig. 14. Owing to the sparse spatial sampling, higher-mode curvature and nodal behavior between instrumented locations were only

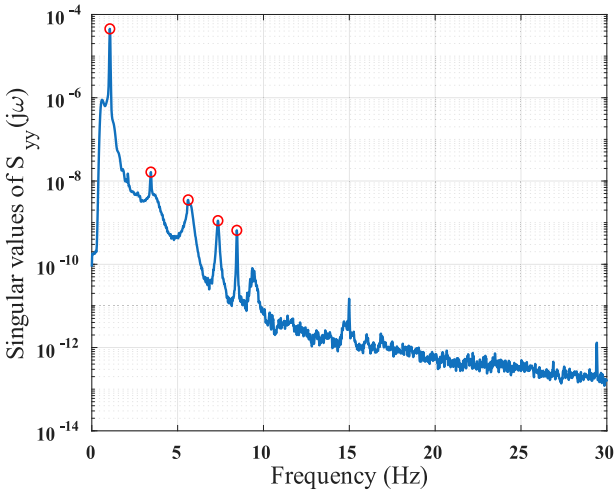


FIGURE 12. Singular value spectrum from FDD analysis of shear building response, showing identified modal frequencies (red circles).

TABLE 11. Experimentally identified and theoretical natural frequencies, absolute errors, and MAC values.

Mode	Theoretical (Hz)	Experimental (Hz)	Abs. Error (Hz)	MAC
1	1.09	1.05	0.04	0.981
2	3.45	3.44	0.01	0.887
3	5.67	5.59	0.08	0.879
4	7.30	7.35	0.05	0.851
5	8.30	8.45	0.15	0.649

partially observed. The mode-shape correlation, quantified by the Modal Assurance Criterion (MAC), is reported in Table 11 and illustrated in Fig. 15. The diagonal MAC values were 0.981, 0.887, 0.879, 0.851, and 0.649 for modes 1–5, respectively.

In addition to the quantitative validation, a visual twin was generated to illustrate the structural response in the simulation environment.

D. VISUAL TWIN DEMONSTRATION

A visual twin of a five-story shear building was generated in the Gazebo environment to illustrate the structural response. The CAD-based model was driven by the measured displacements from the synchronized camera system, producing real-time animation of the building's motion. Figure 16 presents the physical specimen alongside its Gazebo-based counterpart during sine-sweep excitation. The visual twin is intended for demonstration purposes only and does not contribute to the quantitative validation of displacement accuracy or modal identification.

IV. DISCUSSION

The experimental results demonstrate that the proposed ROS-based multi-camera system consistently achieves sub-millimeter displacement accuracy and synchronization

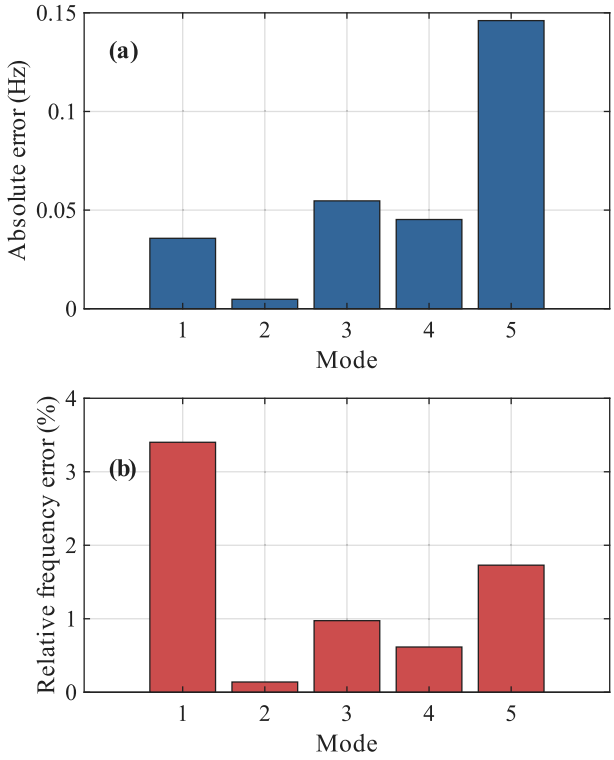


FIGURE 13. Frequency error between experimental and theoretical natural frequencies: (a) absolute error in hertz, and (b) relative frequency error in percentage.

performance, which are sufficient for reliable low-order modal identification and accurate frequency estimation across the tested modes. This section interprets these findings in the context of prior vision-based SHM work, highlights the implications of real-time operation and synchronization, and discusses the advantages and current limitations.

A. PERFORMANCE BENCHMARKING

Across the cantilever, acrylic box, and shear building tests, the proposed system achieved RMSEs of 0.14–0.40 mm across all camera views (Tables 5–7) and residual synchronization offsets well under 5 ms relative to the LDS reference (Table 10). These results are on par with the upper range of the reported accuracies for vision-based SHM systems, despite relying only on consumer-grade cameras and real-time processing.

For displacement tracking, Peng et al. [14] reported 0.38–0.66 mm RMSE on a six-story tower using an edge-computing SURF-based tracker, while Du et al. [26] achieved an error of less than 0.3% in bridge-deck displacement using smartphone video. Our framework achieved sub-millimeter accuracy for the validated in-plane displacement components, which is broadly comparable to prior vision-based studies at similar motion scales. For modal analysis, Panigati et al. [15] identified bridge modal frequencies with errors below 3.5%, whereas our framework produced relative frequency errors under 4% across all modes and MAC values above

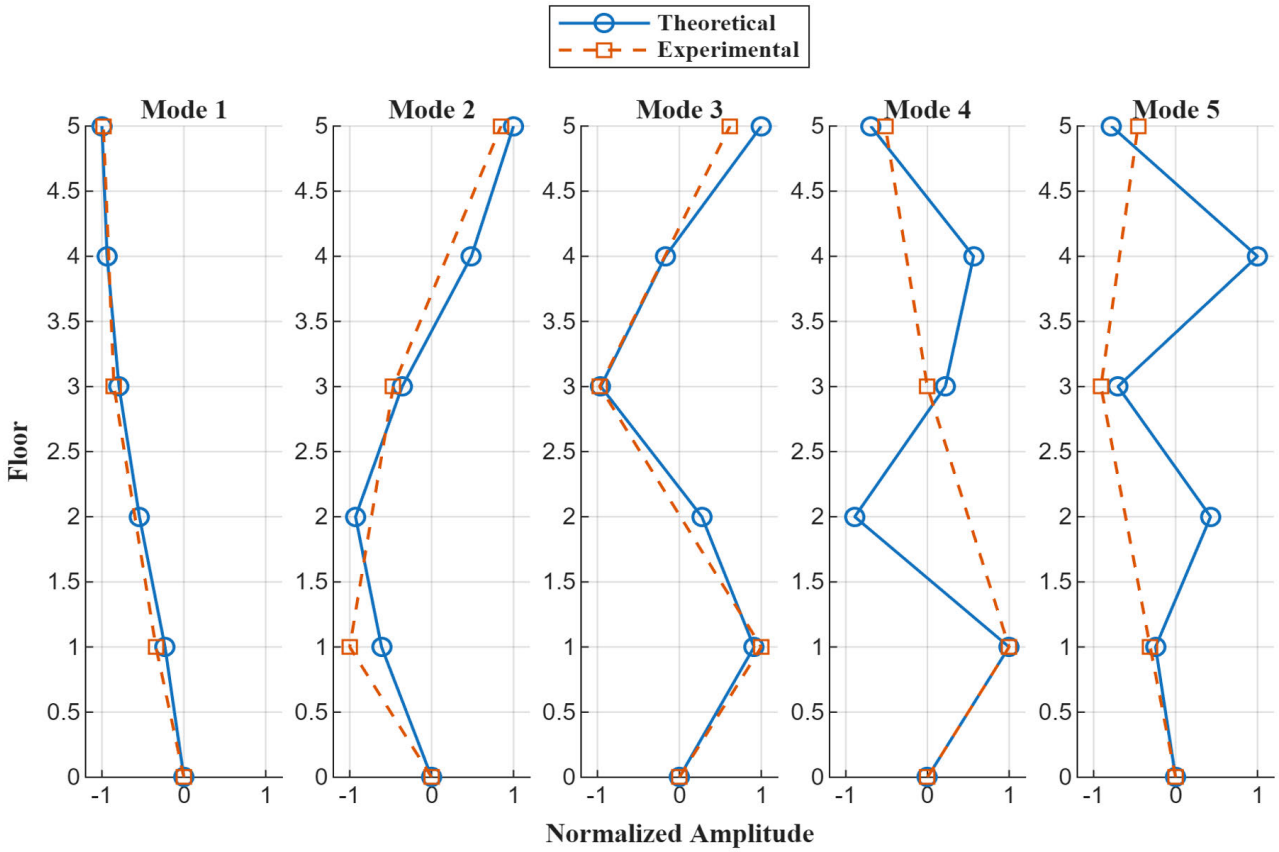


FIGURE 14. Comparison of experimental and theoretical mode shapes (normalized) for five modes at floors 1, 3, and 5.

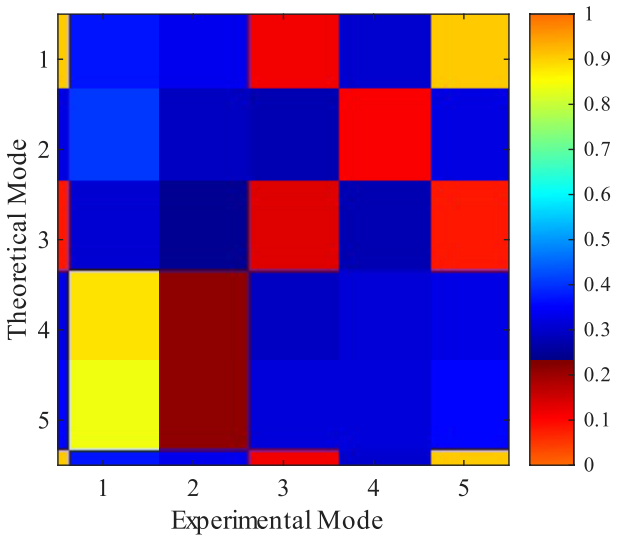


FIGURE 15. Modal Assurance Criterion (MAC) matrix comparing experimental and theoretical mode shapes.



FIGURE 16. Shear building specimen (left) and its Gazebo-based visual twin (right) animated with measured displacements during a sine-sweep excitation.

0.85 for the first four modes (Table 11). These results indicate comparable modal-identification accuracy, while also demonstrating strong mode-shape agreement for the first four modes in a real-time multi-camera tracking framework.

In addition to the single-run validation results, the repeated-run LDS-referenced sinusoidal tests demonstrated stable best-view performance across multiple excitation frequencies, with sub-millimeter mean RMSE, small between-run variability, and consistently high correlation. This additional analysis

strengthens the evidence that the reported accuracy is not limited to a single representative trial but is repeatable across repeated measurements.

B. REAL-TIME OPERATION AND SYNCHRONIZATION

The experiments confirmed that the framework can operate in real-time, with residual synchronization offsets below 5 ms across all tests (Table 10). This level of temporal coherence enables continuous displacement streaming and reliable modal identification across the tested frequency range. In comparison, many vision-based SHM frameworks synchronize frames only during post-processing or require hardware triggers and coded-light timestamps [17], [18]. By combining Chrony-based host clock discipline with ROS `ApproximateTime` tuple formation and capture-time timestamp propagation, the present approach achieves millisecond-level temporal fidelity without the added cost and complexity of the Precision Time Protocol (PTP) or GNSS-disciplined clocks.

High-precision synchronization architectures such as White Rabbit [29] and time-synchronized wireless sensor networks developed for SHM applications [30] have demonstrated precision ranging from sub-nanosecond to sub-millisecond levels in distributed measurement systems. However, these frameworks rely on specialized hardware or embedded synchronization protocols, which increase system complexity and limit scalability for camera-based deployments. In contrast, the proposed software-only synchronization provides millisecond-level coherence that is sufficient for dynamic modal identification while maintaining scalability and simplicity.

The complete ROS pipeline sustained a continuous acquisition rate of 60 FPS from all cameras and an effective processing throughput of 55–58 FPS during the experiments. The end-to-end latency from the image capture to displacement output remained consistently below 100 ms, which was measured through ROS timestamp logging. The average CPU utilization across the three Intel NUC units remained below 70%, and less than 55% memory usage confirmed stable system operation in real-time under all tested conditions.

C. ADVANTAGES AND LIMITATIONS

The use of AprilTag markers within a scalable ROS framework provides significant benefits in the experiments, including straightforward calibration, modular deployment, and the ability to track multiple points from heterogeneous viewpoints. Compared with accelerometer-based methods, the vision-based approach avoids integration drift and provides direct displacement measurements without requiring fixed reference frames.

The framework is inherently capable of estimating full 3D relative motion through monocular AprilTag pose recovery; however, the present study reports quantitative validation only for the in-plane (X–Y) displacement components. This scope was chosen for two complementary reasons. First, the out-of-plane component is more sensitive to monocular depth uncertainty because it is inferred more indirectly through perspective scaling and is therefore more strongly affected by corner-localization noise, stand-off distance, and calibration error propagation. Second, the LDS reference used in the experiments provided high-precision ground truth only along the primary measurement axis of each test configuration, which was aligned with the dominant in-plane structural response. Out-of-plane estimates were available during the experiments, but they were not included in the quantitative accuracy claims because an equally reliable external reference for direct validation was not available. A dedicated out-of-plane validation study therefore remains an important direction for future work.

Although the present experiments were conducted under controlled laboratory conditions, field deployment would introduce additional sources of variability that were not exhaustively assessed in this study. In particular, changing illumination, glare, shadows, and reduced image contrast can affect AprilTag detection quality and corner localization, while partial occlusions may interrupt target visibility and lead to intermittent measurements. The current best-view selection strategy improves robustness by favoring the most reliable observation at each instant and was effective for the tested camera configurations; however, in more complex environments with denser camera networks or more frequent visibility changes, additional mechanisms such as occlusion-aware logic, temporal consistency constraints, or probabilistic multi-view fusion may further improve stability. Long-term field deployment may also require more explicit treatment of camera mounting stability, communication reliability, and calibration maintenance.

Beyond this, similar sources of limitation were observed in the modal identification results, where Modes 1–3 were reconstructed with visually coherent shapes and high MAC values, whereas Modes 4 and 5 showed larger discrepancies despite a MAC of approximately 0.85 for Mode 4. This behavior can be attributed mainly to limited spatial sampling, as only three degrees of freedom (floors 1, 3, and 5) were used for identification and MAC computation. This configuration may overestimate the correlation for higher modes, where the intermediate curvature and nodal behavior are not captured. In addition, the 60 FPS acquisition limits the reliable frequency band to about 30 Hz, and residual inter-camera lags of 0.8–2.5 ms introduce phase errors at higher frequencies. These factors collectively reduce higher-mode consistency and indicate that denser spatial sampling or higher-frame-rate acquisition would improve the reconstruction accuracy. The chosen scope prioritizes real-time operation and practical deployment over exhaustive validation.

Best-view selection was employed in place of full multi-view fusion to ensure scalability and minimize real-time computational load. These scope choices limited certain analyses but were sufficient to demonstrate sub-millimeter displacement tracking performance and reliable modal identification, providing a robust foundation for further development.

Alternative strategies in the literature illustrate possible extensions. Wu et al. [27] achieved ±0.3 mm displacement accuracy by fusing camera data with accelerometers, but this required additional sensor installation, wiring, and fusion algorithms. Similarly, Zhou et al. [17] and Han et al. [18] demonstrated sub-millisecond synchronization by employing coded light panels and time-calibrated sequences. Although these approaches improve fidelity, they increase hardware and setup complexity and are typically applied in post-processing. In contrast, the proposed framework achieves competitive accuracy and synchronization performance using only consumer-grade cameras and software-based time alignment, highlighting its efficiency for real-time SHM deployments.

These results demonstrate the practical effectiveness of software-driven temporal synchronization and vision-based tracking for real-time SHM applications, while also highlighting opportunities for future extension in out-of-plane validation and field-scale deployment.

V. CONCLUSION

This paper presents a ROS-based framework for real-time 3D structural displacement monitoring using a synchronized multi-camera vision system. The architecture integrates host clock discipline, capture-time timestamp propagation, and ROS middleware-level tuple formation to achieve millisecond-level temporal coherence without specialized hardware. The framework is modular and scalable, supporting distributed multi-point tracking for large-scale structural health monitoring applications.

Validation through a high-precision laser displacement sensor and modal testing of a five-story shear building verified sub-millimeter accuracy for the in-plane displacement components and successful identification of modal parameters.

The findings show that consumer-grade cameras, when combined with software-only synchronization, can achieve high-accuracy displacement tracking and millisecond-level temporal coherence while enabling continuous real-time structural monitoring. Additional repeated-run LDS-referenced sinusoidal analysis further confirmed the repeatability of the proposed system by demonstrating stable best-view RMSE and correlation statistics across multiple excitation frequencies.

The contributions of this work are threefold: (i) an end-to-end software synchronization pipeline that integrates distributed acquisition, capture-time timestamp propagation, and middleware-level association for real-time analysis; (ii) a scalable multi-camera architecture for distributed multi-point displacement tracking; and (iii) a modular real-time processing framework validated through displacement benchmarking and modal identification.

Future work will focus on extending the framework to support probabilistic multi-view fusion and experimental validation of out-of-plane displacements. This will facilitate more comprehensive 3D motion quantification and support future field-scale deployments. Further developments may include hybrid synchronization methods and adaptive fusion between cameras and other sensor modalities. The framework is planned for open-source release to encourage community-driven enhancements and broader adoption in dense structural monitoring, adaptive sensing, and integration with complementary SHM technologies.

ACKNOWLEDGMENT

The authors gratefully acknowledge Chung-Ang University for providing the necessary resources and institutional support for this research. Special thanks are extended to the Department of Smart Cities and the Department of Civil and Environmental Engineering for their collaboration and guidance throughout this work.

They would also like to acknowledge the late Prof. Joong-Woong Park, whose guidance and mentorship were invaluable to this research.

AI Disclosure: Language editing and figure caption polishing were assisted by an AI tool (ChatGPT, OpenAI); they reviewed and take responsibility for all content.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest regarding the publication of this article.

REFERENCES

[1] C. R. Farrar, N. Dervilis, and K. Worden, "The past, present and future of structural health monitoring: An overview of three ages," *Strain*, vol. 61, no. 1, p. 12495, 2025.

[2] Y. Deng, Y. Zhao, H. Ju, T. Yi, and A. Li, "Abnormal data detection for structural health monitoring: State-of-the-art review," *Develop. Built Environ.*, vol. 17, Jan. 2024, Art. no. 100337.

[3] Z. Ma, J. Choi, and H. Sohn, "Structural displacement sensing techniques for civil infrastructure: A review," *J. Infrastruct. Intell. Resilience*, vol. 2, no. 3, Sep. 2023, Art. no. 100041.

[4] C. Dong, S. Baş, and F. N. Çatbaş, "Applications of computer vision-based structural monitoring on long-span bridges in Turkey," *Sensors*, vol. 23, no. 19, p. 8161, 2023.

[5] H. Shi, P. Chen, X. Liu, Z. Hong, Z. Ye, Y. Gao, Z. Liu, and X. Tong, "Distributed high-speed videogrammetry for real-time 3D displacement monitoring of large structure on shaking table," *Remote Sens.*, vol. 16, no. 23, p. 4345, Nov. 2024.

[6] S. C. Stiros, "GNSS (GPS) monitoring of dynamic deflections of bridges: Structural constraints and metrological limitations," *Infrastructures*, vol. 6, no. 2, p. 23, Feb. 2021.

[7] R. R. Rabi, M. Vailati, and G. Monti, "Effectiveness of vibration-based techniques for damage localization and lifetime prediction in structural health monitoring of bridges: A comprehensive review," *Buildings*, vol. 14, no. 4, p. 1183, Apr. 2024.

[8] K. Zeng, S. Zeng, H. Huang, T. Qiu, S. Shen, H. Wang, S. Feng, and C. Zhang, "Sensing mechanism and real-time bridge displacement monitoring for a laboratory truss bridge using hybrid data fusion," *Remote Sens.*, vol. 15, no. 13, p. 3444, Jul. 2023.

[9] Y. Zhuang, W. Chen, T. Jin, B. Chen, H. Zhang, and W. Zhang, "A review of computer vision-based structural deformation monitoring in field environments," *Sensors*, vol. 22, no. 10, p. 3789, May 2022.

[10] Y. Xu and J. M. W. Brownjohn, "Review of machine-vision based methodologies for displacement measurement in civil structures," *J. Civil Structural Health Monitor.*, vol. 8, no. 1, pp. 91–110, Jan. 2018.

[11] B. F. Spencer, V. Hoskere, and Y. Narazaki, "Advances in computer vision-based civil infrastructure inspection and monitoring," *Engineering*, vol. 5, no. 2, pp. 199–222, Apr. 2019.

[12] B. Lu, B. Bai, and X. Zhao, "Vision-based structural displacement measurement under ambient-light changes via deep learning and digital image processing," *Measurement*, vol. 208, Jun. 2023, Art. no. 112480.

[13] J. Guo, Y. Xiang, K. Fujita, and I. Takewaki, "Vision-based building seismic displacement measurement by stratification of projective rectification using lines," *Sensors*, vol. 20, no. 20, p. 5775, Oct. 2020.
[14] Z. Peng, J. Li, H. Hao, and Y. Zhong, "Smart structural health monitoring using computer vision and edge computing," *Eng. Struct.*, vol. 319, Aug. 2024, Art. no. 118809.
[15] T. Panigati, A. Abbozzo, M. A. Pace, E. Temur, F. Cigan, and R. Kromanis, "Dynamic identification of bridges using multiple synchronized cameras and computer vision," *Infrastructures*, vol. 10, no. 2, p. 37, Feb. 2025.
[16] (2022). *GPS Accuracy*. Accessed: Aug. 12, 2025. [Online]. Available: https://www.gps.gov/systems/gps/performance/accuracy
[17] X. Zhou, Y. Dai, H. Qin, S. Qiu, X. Liu, Y. Dai, J. Li, and T. Yang, "Subframe-level synchronization in multi-camera system using time-calibrated video," *Sensors*, vol. 24, no. 21, p. 6975, Oct. 2024.
[18] Y. Han, S. Lo Feudo, G. Cumunel, and F. Renaud, "Sub-frame timestamping of a camera network using a coded light signal," *Measurement*, vol. 236, Aug. 2024, Art. no. 115046.
[19] Open Robotics. *Message_filters — Time Synchronization (ApproximateTime/ExactTime) [ROS]*. Accessed: Aug. 12, 2025. [Online]. Available: https://wiki.ros.org/message_filters/ApproximateTime
[20] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognit.*, vol. 47, no. 6, pp. 2280–2292, Jun. 2014.
[21] F. J. Romero-Ramirez, R. Muñoz-Salinas, and R. Medina-Carnicer, "Speeded up detection of squared fiducial markers," *Image Vis. Comput.*, vol. 76, pp. 38–47, Aug. 2018.
[22] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.
[23] Chrony Project. *Chrony: A Versatile Implementation of the Network Time Protocol (NTP)*. Accessed: Sep. 8, 2025. [Online]. Available: https://chrony-project.org/
[24] C. R. Farrar and K. Worden, *Structural Health Monitoring: A Machine Learning Perspective*. Hoboken, NJ, USA: Wiley, 2012.
[25] S. W. Doebling, C. R. Farrar, M. B. Prime, and D. W. Shevitz, "Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: A literature review," Los Alamos National Laboratory, Los Alamos, NM, USA, Tech. Rep. LA-13070-MS, 1996. [Online]. Available: https://www.osti.gov/biblio/249299
[26] W. Du, H. Wang, D. Lei, Z. Hang, C. Cui, and J. He, "Cable vibration and girder displacement measurement via morphology recognition and feature matching methods in bridge structure," *Sci. Rep.*, vol. 15, no. 1, p. 11131, Apr. 2025.
[27] T. Wu, L. Tang, S. Shao, X. Zhang, Y. Liu, Z. Zhou, and X. Qi, "Accurate structural displacement monitoring by data fusion of a consumer-grade camera and accelerometers," *Eng. Struct.*, vol. 262, Jul. 2022, Art. no. 114303.
[28] J. Li, Q. Miao, Z. Zou, H. Gao, L. Zhang, Z. Li, and N. Wang, "A review of computer vision-based monitoring approaches for construction workers' work-related behaviors," *IEEE Access*, vol. 12, pp. 7134–7155, 2024.
[29] Y. Liu, K. I. Ahmed, N. L. Thotakura, J. Yang, Y. Wu, H. Wang, B. Sun, W. Wang, Y. Zhang, and Y. Liu, "Enhancing time synchronization in smart grid with White Rabbit: Theory, architecture, and challenges," *IEEE Access*, vol. 13, pp. 146707–146721, 2025.
[30] I. Val, A. Arriola, C. Cruces, R. Torrego, E. Gomez, and X. Arizkorreta, "Time-synchronized wireless sensor network for structural health monitoring applications in railway environments," in *Proc. IEEE World Conf. Factory Commun. Syst. (WFCS)*, May 2015, pp. 1–9.



AMMAR AJMAL received the B.E. and M.S. degrees in computer engineering from the National University of Sciences and Technology (NUST), Pakistan. He is currently pursuing the Ph.D. degree in smart cities with Chung-Ang University, Seoul, South Korea. He is an Assistant Professor with Bahria University, Pakistan. His research interests include structural health monitoring, computer vision and vision-based sensing, autonomous systems, and real-time sensing and AI-assisted data processing.



TAHIR MEHMOOD received the B.E. degree in electronics engineering from Iqra University, Pakistan, and is currently pursuing the M.Sc. degree in autonomous systems from Hochschule Bonn-Rhein-Sieg (H-BRS), Germany. He is currently a Senior Software Developer with Innok Robotics, Germany. His professional experience includes leading mobile-robot development with MYBOT-SHOP and extensive work in embedded systems and ESP32-based microcontrollers. His research and development interests include robotics, computer vision, embedded control, and autonomous navigation for intelligent robotic platforms.



JONGBIN WON received the Ph.D. degree in civil and environmental engineering from Chung-Ang University, Seoul, South Korea. He is currently a Postdoctoral Researcher with Korea Institute of Civil Engineering and Building Technology (KICT), Goyang, South Korea. He has authored and co-authored several journal articles on vision-based sensing, system identification, and data-driven approaches for infrastructure assessment. His work has been cited widely in the structural monitoring community. His research interests include computer vision-based structural health monitoring, vibration analysis, and digital twin applications for smart infrastructure.



CHANGSU SHIM was the President of Korean Institute of Building Information Modeling (KIBIM). He is currently a Professor with the School of Civil and Environmental Engineering, Urban Design and Studies, Chung-Ang University, Seoul, South Korea. He led a six-year national research project (2020–2025) on the development of smart construction technology, funded by Korean government in cooperation with the University of Cambridge. His research work focuses on advancing smart construction, structural health monitoring, and digital infrastructure technologies for next-generation civil engineering systems. His research interests include BIM-based bridge design, construction and maintenance, composite structures, digital twin technology, prefabricated bridge construction, and 3-D concrete-printed structures.