

Article

Application Framework and Optimal Features for UAV-Based Earthquake-Induced Structural Displacement Monitoring

Ruipu Ji ¹, Shokrullah Sorosh ¹, Eric Lo ², Tanner J. Norton ², John W. Driscoll ², Falko Kuester ¹, Andre R. Barbosa ³, Barbara G. Simpson ⁴ and Tara C. Hutchinson ^{1,*}

¹ Department of Structural Engineering, University of California San Diego, La Jolla, CA 92093, USA; r5ji@ucsd.edu (R.J.); ssorosh@ucsd.edu (S.S.); fkuester@ucsd.edu (F.K.)

² Qualcomm Institute, University of California San Diego, La Jolla, CA 92093, USA; eklo@ucsd.edu (E.L.); tanorton@ucsd.edu (T.J.N.); jwdrisco@ucsd.edu (J.W.D.)

³ School of Civil and Construction Engineering, Oregon State University, Corvallis, OR 97331, USA; andre.barbosa@oregonstate.edu

⁴ Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305, USA; bsimpson@stanford.edu

* Correspondence: tara@ucsd.edu

Abstract: Unmanned aerial vehicle (UAV) vision-based sensing has become an emerging technology for structural health monitoring (SHM) and post-disaster damage assessment of civil infrastructure. This article proposes a framework for monitoring structural displacement under earthquakes by reprojecting image points obtained courtesy of UAV-captured videos to the 3-D world space based on the world-to-image point correspondences. To identify optimal features in the UAV imagery, geo-reference targets with various patterns were installed on a test building specimen, which was then subjected to earthquake shaking. A feature point tracking-based algorithm for square checkerboard patterns and a Hough Transform-based algorithm for concentric circular patterns are developed to ensure reliable detection and tracking of image features. Photogrammetry techniques are applied to reconstruct the 3-D world points and extract structural displacements. The proposed methodology is validated by monitoring the displacements of a full-scale 6-story mass timber building during a series of shake table tests. Reasonable accuracy is achieved in that the overall root-mean-square errors of the tracking results are at the millimeter level compared to ground truth measurements from analog sensors. Insights on optimal features for monitoring structural dynamic response are discussed based on statistical analysis of the error characteristics for the various reference target patterns used to track the structural displacements.

Keywords: unmanned aerial vehicle (UAV); vision-based sensing; earthquake; target tracking; world point reconstruction; structural displacement monitoring



check for updates

Academic Editor: Arslan Munir

Received: 1 January 2025

Revised: 20 January 2025

Accepted: 21 January 2025

Published: 26 January 2025

Citation: Ji, R.; Sorosh, S.; Lo, E.; Norton, T.J.; Driscoll, J.W.; Kuester, F.; Barbosa, A.R.; Simpson, B.G.; Hutchinson, T.C. Application Framework and Optimal Features for UAV-Based Earthquake-Induced Structural Displacement Monitoring. *Algorithms* **2025**, *18*, 66. <https://doi.org/10.3390/a18020066>

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

1. Introduction

Vision-based sensing has emerged as a cost-effective and non-destructive evaluation technique, capable of supplementing and advancing conventional data collection practices for the monitoring and life-cycle health assessment of civil infrastructure. Image data collected using vision-based sensing systems can be interpreted by image processing and computer vision algorithms. These interpretations can provide insights to support structural risk assessment during construction [1–3], damage detection during operation [4–6], and structural safety evaluation after disasters [7–9]. Similarly, evaluation of the state of civil infrastructure is also important for structural assessments. Dynamic loads due to

earthquakes and other natural hazards pose a significant challenge to the safety of civil infrastructure. To evaluate post-earthquake structural safety and functional recovery, significant efforts have been spent on earthquake-induced structural displacement tracking using images. In conventional practice, stationary cameras, such as surveillance cameras, have been used to monitor structural motion during earthquakes [10]. With stationary cameras, structural motions including displacements, accelerations, and inter-story drift ratios, can be extracted from images or a series of images (videos) recorded by a regular red-green-blue (RGB) camera or other vision-based sensors [11].

While the conventional practice of using stationary cameras has proven to be useful for displacement tracking of infrastructures, the emergence of commercial unmanned aerial vehicles (UAVs) equipped with on-board cameras offers significantly enhanced flexibility and efficiency for image data collection. These aerial platforms are especially advantageous when it is inconvenient or unsafe to obtain the required camera view from ground-based (stationary) cameras (e.g., the roof of the building [12–14]). In addition, the motions of a structure in all six degree-of-freedom (DoFs) can be extracted by combining synchronized video data recorded by multiple UAV platforms obtained from locations surrounding the structure [14].

Although the promise of UAV-based methods is recognized, the accuracy and robustness of UAV vision-based methods to monitor structural displacements are significantly affected by (a) the correction of UAV drift-induced camera movement and (b) the robustness of detection for representative image features for tracked objects. Unlike ground-based cameras, which can be assumed to be stationary, the movement of on-board cameras caused by UAV drift cannot be neglected for UAV vision-based approaches. Therefore, multiple methods have been deployed in practice for camera pose recovery to avoid possible image misalignment. Conventional non-learning-based methods include homography transformation [12,15,16], direct linear transformation (DLT) [17], and embedded inertial measuring unit (IMU) measurement [18]. Learning-based methods are also applied in recent studies. For example, Zhang et al. [19] proposed a two-stage correction method to eliminate UAV drift-induced measurement error using stationary points and variational mode decomposition (VMD).

To enhance the reliable detection and tracking of objects, both reference target-based and target-free methods have been utilized in UAV vision-based monitoring to improve the tracking accuracy of structural displacements. Promising target-free approaches include the work of Khuc et al. [20] who applied scale invariant feature transform (SIFT) and a subsequent Hough Transform for detection and localization of feature points in circular steel plate structures. Wang et al. [12] used Canny edge detection followed by linear regression for straight-line feature detection on the roof of a full-scale six-story cold-formed steel building. Researchers have noted that geo-reference targets on the periphery of the structure could significantly enhance the accuracy and robustness of the UAV vision-based tracking of structural displacements. Thus, different algorithms have been developed to detect and track reference targets with various geometric patterns. Amongst the most common target patterns, black-and-white checkerboards are widely adopted by researchers. In a recent study, Wang et al. [13] developed a frame-by-frame analysis framework using a sub-pixel edge detector followed by edge point clustering and line regression to localize the target image point with sub-pixel accuracy. For applications of other target patterns, image binarization can be applied for patterns with white dots on the black background [21,22]. In addition, color-pass filtering has shown robustness in detecting patterns with specific colors [17].

Although research has emerged to advance the use of UAV vision-based structural displacement monitoring during earthquakes, prior studies have at least one of the following limitations:

- The specimen used for monitoring structural displacements was a reduced-scale structural model rather than a full-scale structure. Compared to a reduced-scale structural model, monitoring the responses of a full-scale structure requires a larger camera-to-scene distance, which may compromise the resolution of the UAV-captured videos.
- The study was conducted in a controlled indoor environment, which is different from actual field applications. Factors such as weather-induced UAV drift and additional image noise caused by non-uniform lighting are therefore not able to be considered in the study.
- Only one or two types of features and reference target patterns were selected for each study. There is no direct comparison of the accuracy or robustness of different features or patterns when conducting structural displacement monitoring.

To address these limitations and to advance the application of UAV vision-based sensing, a framework is proposed for monitoring structural displacements occurring during earthquakes by reprojecting the image points to the 3-D world space based on the world-to-image point correspondences. The tracked structural displacements can provide valuable information in post-disaster safety evaluation and functional recovery analysis of a structure. To provide robust features in the UAV imagery, geo-reference targets with multiple pattern types and colors are installed on a full-scale building specimen, along with the stationary background region. To ensure reliable detection and tracking of multiple image features, a feature point tracking-based algorithm for square checkerboard patterns and a Hough Transform-based algorithm for concentric circular patterns are implemented. Photogrammetry techniques are applied to reconstruct the 3-D world point and extract the structural displacements.

Shake table testing, which has been widely adopted in earthquake engineering to support investigations of structural response by physically simulating ground motion excitations, provides a unique opportunity to explore and evaluate vision-based methods for structural displacement monitoring under earthquakes. In this article, the proposed methodology is validated by monitoring the displacements of a full-scale 6-story mass timber building during a series of shake table test programs conducted at the 6-DoF Large High-Performance Outdoor Shake Table (LHPOST6) [23] at UC San Diego. The building specimen was subjected to a suite of real earthquake motion inputs at its base to emulate a field scenario, with traditional analog sensors deployed on the specimen to measure its dynamic responses. Reasonable accuracy is achieved when implementing the proposed analysis method to capture the building specimen's global behavior, with the overall root-mean-square errors (RMSEs) of the tracking results at the millimeter level compared with ground truth measurements from analog sensors. This study concludes by offering insights on the optimal features for earthquake-induced structural displacement monitoring based on statistical analysis of the error characteristics considering the various reference target patterns used to track the structural displacements.

2. UAV Video Image Collection Program

The proposed UAV vision-based earthquake-induced structural displacement monitoring method and target detection and tracking algorithms applied in the framework, are developed and validated using UAV imagery and other measurements (point cloud model from photogrammetry, analog sensor data) collected during Phase II of the Natural Hazards Engineering Research Infrastructure (NHERI) Converging Design (CD) Project conducted at the LHPOST6 [24–26]. In this section, an overview of the shake table test

program for Phase II of NHERI Converging Design is presented (Section 2.1), followed by the introduction of the adopted UAV vision-based monitoring plan and reference targets layout (Section 2.2).

2.1. Case Study: NHERI Converging Design (Phase II) Shake Table Test Program

To prototype seismic design solutions for mass timber buildings, a series of shake table tests were performed on a full-scale six-story mass-timber building with a seismic lateral-force resisting system using post-tensioned (PT) mass timber rocking walls (Figure 1). The 6-story building was deconstructed and reconstructed from a previous 10-story mass timber building (referred to as the NHERI Tallwood Building [27]). The 6-story mass timber building had a height of 20.73 m and a floor plan of 10.52 m \times 10.46 m. The building used a self-centering seismic lateral-force resisting system composed of cross-laminated timber (CLT) rocking walls in the east-west direction and mass plywood panel (MPP) rocking walls in the north-south direction. To dissipate seismic input energy, the CD Phase II test series incorporated buckling-restrained boundary elements (BRBs) installed at the bottom of the east and west MPP rocking walls. In all phases (10- and 6-story configurations), an operable prefabricated stair tower was designed into the building [28].

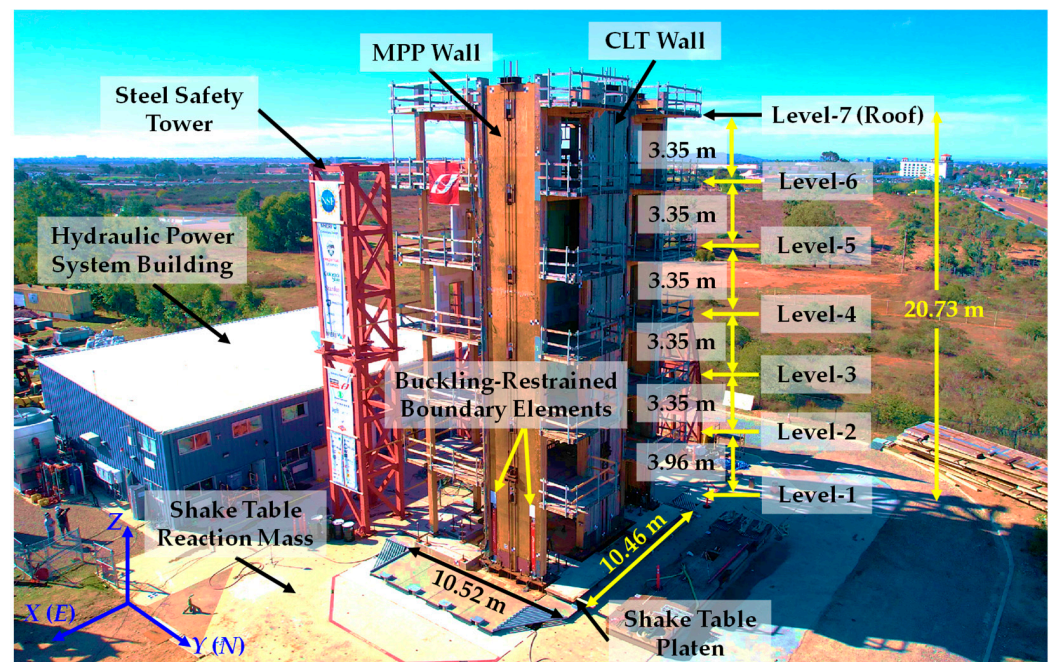


Figure 1. The 6-story mass timber building specimen for the shake table test program of NHERI Converging Design Project (Phase II) at the LHPOST6.

Instrumentation of the test building specimen included 602 channels of analog sensors to record various dynamic responses during shake table testing. All analog measurements were synchronized at a sampling rate of 256 Hz. In the present study, the motion of the roof was recorded using 22 uniaxial micro-electro-mechanical system (MEMS) accelerometers, which served as the ground truth measurements to evaluate the accuracy of the proposed UAV vision-based method. Detailed processing steps for the ground truth measurements and a comparison plan are discussed in Section 5.1.

A total of 18 tests were conducted during the CD phase II test series, corresponding to 18 input ground motions identified with motion identification 1 through 18, i.e. MID 1-18. Among these tests, nine were captured by UAV imagery. Four earthquake events are used to generate the earthquake motion inputs used for these nine tests, namely the 1994 Northridge Earthquake, the 2004 Niigata Earthquake, the 2010 Ferndale Earthquake,

and the 2010 Maule Earthquake. These earthquake motion inputs are scaled to different intensity levels, where each intensity level corresponds to a specific percentage of the Risk-Targeted Maximum Considered Earthquake (MCE_R) that was referenced for the design of the building specimen based on the approach described in Section 21.2.1 in ASCE 7-22 [29]. The MCE_R refers to the earthquake that leads to a 1% probability of structural collapse within a 50-year period [29]. Table 1 shows detailed information of the earthquake motion inputs used in the present UAV-based analysis, including input directions, intensity levels, and achieved peak input accelerations (PIAs), which is a subset of the earthquake motion inputs used in the CD Phase II shake table test program.

Table 1. Earthquake motion inputs used within the present UAV-based analysis. For additional details regarding the test protocol, scaling, and resulting achieved motions, see McBain et al. [25].

Earthquake	MID ¹	Input Direction	Intensity Level	Achieved Peak Input Acceleration (PIA)		
				X [g]	Y [g]	Z [g]
1994 Northridge, USA (Crustal ²) <i>Station: Sun Valley—Roscoe Blvd</i>	7	XYZ	67% MCE_R ³	0.38	0.52	0.45
	8		100% MCE_R	0.59	0.83	0.73
	18		110% MCE_R	0.65	0.92	0.75
2010 Ferndale, USA (Intraslab) <i>Station: 89486</i>	12	XYZ	67% MCE_R	0.38	0.41	0.75
	15		100% MCE_R	0.59	0.63	1.16
2010 Maule, Chile (Interface) <i>Station: CSCH</i>	13	XY	67% MCE_R	0.42	0.34	0.02
	16		100% MCE_R	0.64	0.52	0.03
2004 Niigata, Japan (Crustal) <i>Station: NIGH11</i>	14	XYZ	68.9% MCE_R	0.58	0.40	0.30
	17		100% MCE_R	0.79	0.55	0.45

¹ MID: Motion identification. ² The content in parentheses represents the source type of each earthquake. ³ MCE_R : Risk-Targeted Maximum Considered Earthquake.

The measured peak responses of the building specimen under the nine monitored tests are summarized in Table 2. It is noted that structural responses (accelerations and displacements) discussed in this article are the absolute (total) structural responses, which include the shake table movement. For brevity, the term ‘absolute’ is omitted in subsequent presentation. In addition, the peak responses presented in Table 2 are the peak absolute valued responses in X or Y direction (regardless of +/− in the given direction) calculated based on the measurements from the MEMS accelerometers near the estimated location of the center of mass of the floor plan at roof level. Roof displacement is the total roof movement determined by twice integrating the filtered acceleration time series (details on the displacement calculation are presented in Section 5.1.1). During the nine tests considered, the largest peak roof acceleration (1.32 g) and peak roof displacement (36.28 cm) were both in the Y-direction during MID 18. Although not presented herein, the modal properties of the building are determined using the Frequency Response Function (FRF) [30], as this serves to characterize the transfer function between the input signal on the shake table platen and the output signal on the test specimen (roof level). For this specimen, the fundamental periods for the first three modes of the building are determined from the peaks in the FRF as 1.14 s (X-flexural), 0.95 s (Y-flexural), and 0.71 s (torsional).

Table 2. Summary of structural responses of the building for the tests captured by UAV imagery.




MID	Intensity Level	Peak Roof Acceleration ^{1,2} [g]		Peak Roof Displacement ² [cm]	
		X	Y	X	Y
7	67% MCE _R	0.64	0.89	13.94	22.36
8	100% MCE _R	0.89	1.10	19.72	33.71
12	67% MCE _R	0.74	0.83	20.84	15.99
13	67% MCE _R	0.58	0.85	9.26	17.14
14	68.9% MCE _R	0.77	1.07	19.05	15.99
15	100% MCE _R	0.91	1.11	29.16	24.19
16	100% MCE _R	0.72	1.16	11.61	22.64
17	100% MCE _R	1.01	0.99	28.16	24.46
18	110% MCE _R	0.91	1.32	24.46	36.28

¹ Peak responses are the peak absolute valued response in X or Y direction (regardless of +/- in the given direction). ² Both roof acceleration and displacement responses are absolute (total) values, i.e., including the shake table movement.

2.2. UAV Vision-Based Monitoring Plan and Reference Target Layout

The structural motion of the 6-story mass timber building was captured by monocular cameras on three off-the-shelf commercial UAV platforms in three different views aimed at capturing structural motions in all DoFs. Characteristics of these UAV platforms are summarized in Table 3. The present article focuses on analyzing the plan-view videos, where the DJI Matrice 300 UAV was consistently utilized to capture time-series imagery with a frame rate of 59.94 frames per second (fps). An example camera view of the plan-view videos (the first video frame from MID 15 test video) with a triple zoom-in view for reference targets with various patterns is shown in Figure 2. For all plan-view videos, the UAV was hovering at a position approximately 55 m above the shake table platen (approximately 35 m above the roof of the building specimen) to maintain a similar pixel resolution for all the videos.

Table 3. Characteristics of UAV platforms used for shake table test imagery [31].

Camera View	UAV Platform	Unfolded Size [L × W × H-mm]	Battery Life [min]	Frame Rate [fps]	Resolution
Plan view (XY view)	 DJI Matrice 300 UAV with DJI Zenmuse P1 Camera	810 × 670 × 430	55	59.94	3840 × 2160
East view (YZ view)	 DJI Mavic 2 Pro UAV	322 × 242 × 84	31	29.97	3840 × 2160
North view (XZ view)	 DJI Mavic 3 Enterprise UAV	348 × 283 × 108	45	29.97	3840 × 2160

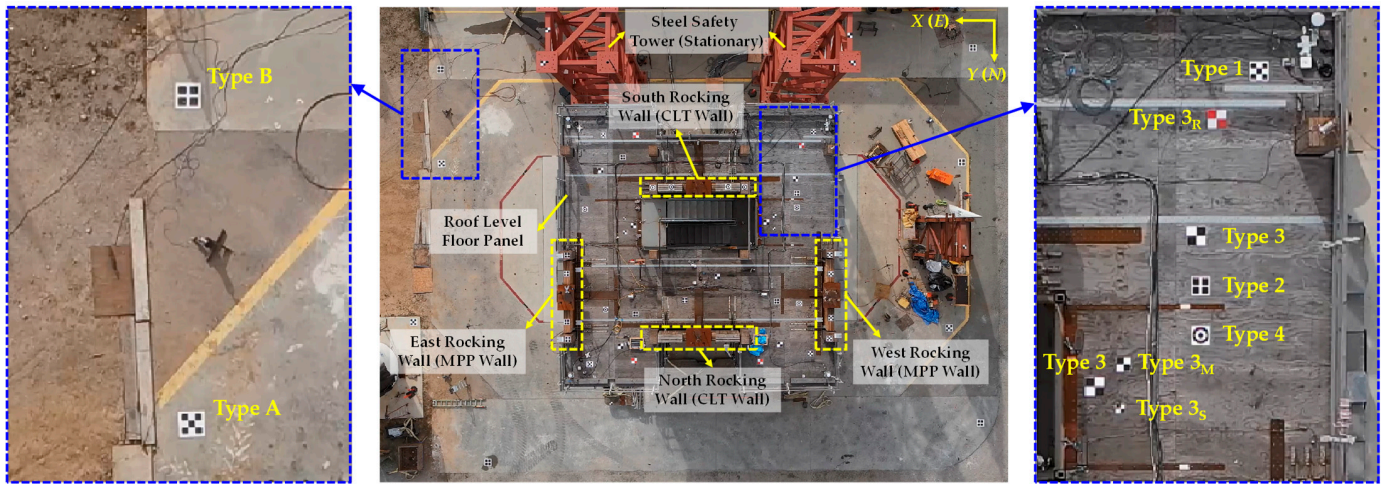


Figure 2. Example camera view of the plan-view videos (the first video frame from MID 15 test video) and zoom-in view ($3\times$) for reference targets with various patterns.

In the UAV vision-based monitoring plan, reference targets were installed on the ground, safety towers, and the building specimen to provide robust image features for detection and tracking. Targets on the ground and safety towers served as stationary features, providing reference points for camera pose recovery and facilitating identification of 3-D world coordinates extracted from the high-resolution point cloud model obtained from photogrammetry (discussed in Section 4). In total, 10 stationary targets (8 on the ground and 2 on the top of safety towers) and 28 moving targets (on the roof panel) were used in the present study. It is noted that the targets installed on the top of the rocking walls (within the yellow dash rectangles in Figure 2) are not considered since there was an elevation difference between the roof panel and the top of the rocking wall; thus, the pixel resolutions and monitored motions of these targets were different from the targets installed on the roof panel. For the installation of targets, either high-strength Velcro tape or high-tack adhesive was applied between each target and its attachment to ensure a reliable installation. Targets were securely attached to the building specimen, ensuring movement of the targets within the plane of the attachment.

Notably, advancing prior work [13], a wide range of reference target patterns were investigated to evaluate the effectiveness of the proposed target detection and tracking algorithm on different patterns and the robustness of different target patterns. Specifically, three patterns were adopted for stationary targets, while seven patterns were adopted for moving targets. A summary of the reference targets is presented in Table 4, with their patterns and dimensions described in Figure 3. The seven patterns adopted for the moving targets on the roof encompass various image features for detection and tracking, considering (a) the number of tiles (two, four, and five), (b) color (black-and-white and red-and-white), (c) shape (square and concentric circles), and (d) dimension. A statistical analysis is conducted to investigate the optimal patterns (features) for structural displacement monitoring under earthquakes by comparing the error characteristics of the displacements extracted from the proposed methodology for each of the moving target patterns. A detailed discussion of this statistical analysis is presented in Section 6.

Table 4. Summary of reference targets used in this study.

Reference Target	Target Pattern Type	Target Dimension [cm × cm]	Number of Targets
Stationary Targets (10 targets)	Type A ¹	45.7 × 45.7	3
	Type B	45.7 × 45.7	5
	Type C	20 × 20	2
Moving Targets (28 targets)	Type 1	20 × 20	4
	Type 2	20 × 20	5
	Type 3	20 × 20	7
	Type 3 _R	20 × 20	4
	Type 3 _M	15 × 15	2
	Type 3 _S	10 × 10	2
	Type 4	20 × 20	4

¹ Type A and B targets were installed on the ground, and Type C targets were installed on the top of the safety towers.

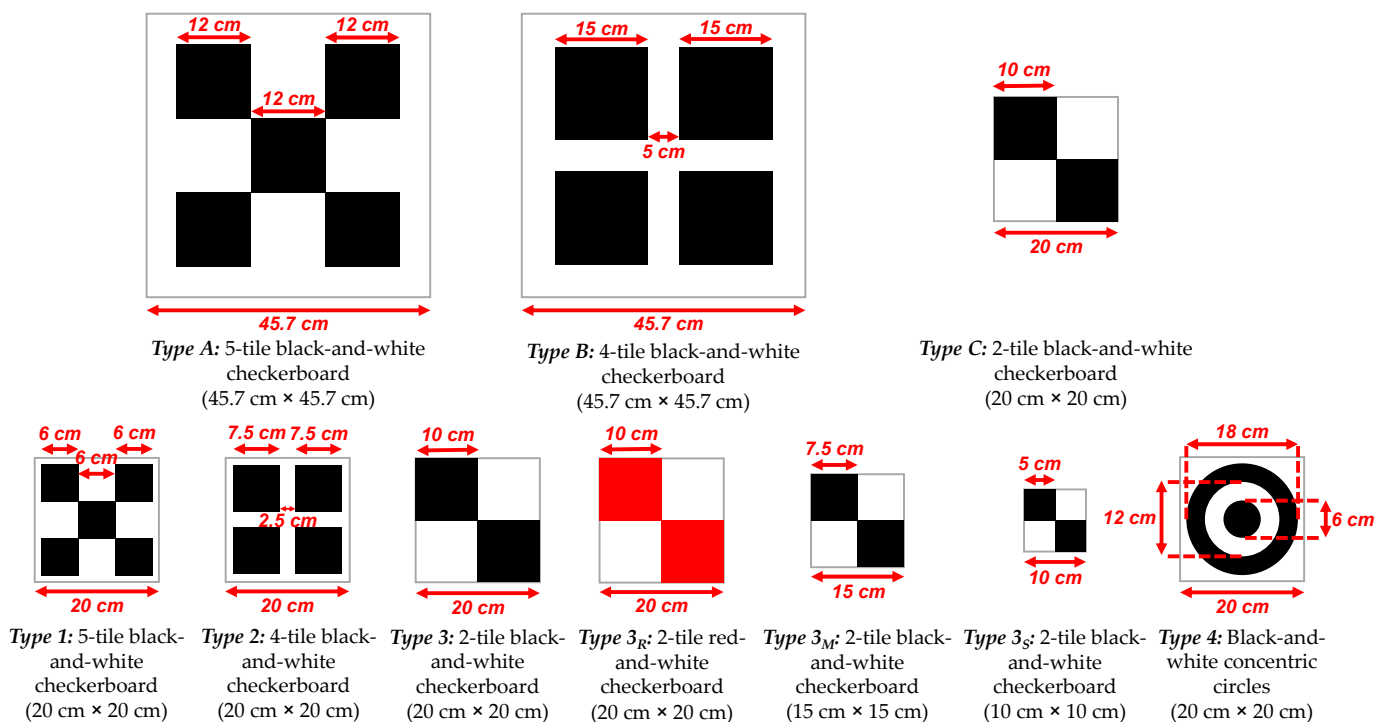


Figure 3. Reference target patterns and dimensions. The three patterns in row one were used as stationary targets and the seven patterns in row two were used as moving targets.

3. Target Detection and Tracking Algorithm

The key approach adopted in the proposed UAV vision-based earthquake-induced structural displacement monitoring framework is to reproject the image points to the 3-D world space based on the world-to-image point correspondence. As such, reliable tracking of certain image feature points is crucial to achieve reasonable accuracy and consistency over all image frames. This section introduces two methods for tracking the movement of the image feature points in the image frame coordinates, namely the feature points tracking-based algorithm for square checkerboard patterns (Section 3.1) and the Hough Transform-based algorithm for concentric circular patterns (Section 3.2).

For both methods presented in this section, there are three initialization steps before applying the algorithms for target detection and tracking:

- Lens distortion correction is applied to each video frame to obtain distortion-free image frames (see Section 4.2).

- Coarse bounding boxes representing the initial regions of interest (ROIs) of each target are manually defined in the reference video frame. The reference video frame is the first frame of each test video, which corresponds to a state at 5 to 10 s before the testing (and hence movement of the specimen) initiates. The pixel dimension of the coarse bounding box in the image frame is 51×51 pixels.
- The colored video frames are converted into grayscale images since black-to-white contrast is easier to identify using a single value threshold compared to a 3-value threshold required with all RGB channels in a colored image. Moreover, grayscale images are of reduced size; thus, the computational cost of the algorithm is reduced.

3.1. Feature Points Tracking-Based Algorithm for Square Checkerboard Patterns

The square checkerboard pattern consists of regularly arranged squares, forming a highly symmetric two-dimensional pattern. The orthogonal geometric structure ensures that each corner point in the pattern has a well-defined position. A square checkerboard pattern usually contains multiple corner points, providing abundant feature points for detection and tracking. A feature points tracking-based algorithm is developed to track the corner points (feature points) and to determine the geometric center in each video frame from the pre-defined initial ROI in the reference video frame for square checkerboard patterns. Following detection of the corner points in the reference video frame, a Kanade-Lucas-Tomasi (KLT) feature tracker [32,33] is initialized with the detected feature points to track the movements of these points. The KLT algorithm aims to minimize computational intensity by directing the search position, as guided by the initial corner points detected for a particular pattern.

All patterns except Type 4 for both stationary and moving targets are square checkerboard patterns. Detection of the corner points based on the pre-defined ROI is successfully implemented based on the four steps described in Figure 4, where a Type 2 target is selected for illustration. Steps 1 to 3 aim to generate a refined ROI to filter out the background noise in the pre-defined ROI. Corner points detection is performed within this refined ROI in Step 4. The details of these analysis steps are presented as follows:

- **Step-1:** The grayscale initial ROI image (Figure 4a) is thresholded into a binary image (Figure 4b) using the Otsu's method [34]. Pixels representing white color in the pattern are defined as foreground pixels. To remove the noise in the background, connected components that have fewer than 10 pixels are removed in the binary image.
- **Step-2:** Edge detection with pixel level accuracy (Figure 4c) is performed on the binary image using the Sobel operator [35]. Although the Canny edge detection method [36] or other sub-pixel edge detection methods [37] may provide better localization of the edge points, the edge detection using the Sobel operator is sufficient for the refined ROI generation in step 3 with a lower computational cost, since this step only aims to generate a refined ROI to filter out the background noise from the initial ROI. In this step, both the edge points representing the edge of the squares (black against white or red against white) and the boundary of the target (white, black, or red against background) are extracted.
- **Step-3:** The boundary point set of all edge points is extracted to form the bounding box of the target, representing the exact target region with pixel level accuracy (blue solid lines in Figure 4d). The approximate center of this bounding box is calculated as the mean of the boundary point set. A scaling operation expanding the bounding box by a factor of 1.1 relative to its approximate center is applied to obtain the refined ROI (red dash lines in Figure 4d), ensuring comprehensive inclusion of all feature points within the refined ROI.

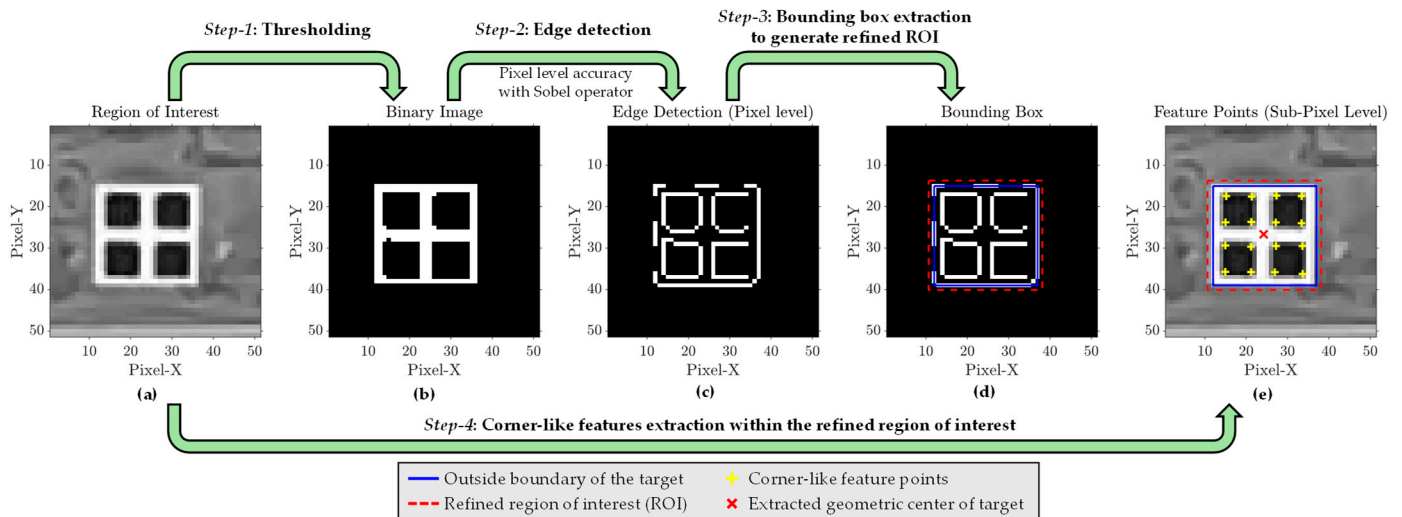


Figure 4. Procedure to detect corner points and extract the target center in the first video frame for square checkerboard patterns (example shown for a Type 2 target).

- Step-4:** The Harris-Stephens feature detection algorithm with quadratic interpolation [38,39] is applied to detect the corner points within the refined ROI in the grayscale image (red dash lines in Figure 4e). The Harris-Stephens feature detection algorithm runs a 3×3 window over the refined ROI in the grayscale image and computes the spatial gradient matrix M at each pixel (x, y) in the refined ROI [38]:

$$M(x, y) = \begin{bmatrix} \sum_x \sum_y I_x(x, y) I_x(x, y) & \sum_x \sum_y I_x(x, y) I_y(x, y) \\ \sum_x \sum_y I_x(x, y) I_y(x, y) & \sum_x \sum_y I_y(x, y) I_y(x, y) \end{bmatrix} \quad (1)$$

where $I_x(x, y)$ and $I_y(x, y)$ are the intensity gradients with respect to x and y (pixel coordinates) at location (x, y) of the grayscale image. Then, a response function $\Gamma(x, y)$ is defined at each pixel based on its spatial gradient matrix $[M]$ [38]:

$$\Gamma(x, y) = \det(M(x, y)) - 0.04(\text{Tr}(M(x, y)))^2 \quad (2)$$

Nonmaximal suppression with a 3×3 window is applied to the calculated response function values to avoid double detection of the corner points. It is noted that the conventional Harris-Stephens feature detection algorithm can only achieve pixel level detection. Therefore, a quadratic interpolation is applied to obtain the sub-pixel coordinates (u, v) of the corner points [39]:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} - \begin{bmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{bmatrix}^{-1} \begin{bmatrix} I_x \\ I_y \end{bmatrix} \quad (3)$$

where (u_0, v_0) are the coordinates of corner points with pixel level accuracy from the conventional Harris-Stephens feature detection; I_x and I_y are the spatial gradients calculated at (u_0, v_0) ; and I_{xx} , I_{yy} , and I_{xy} are the derivatives of the spatial gradients calculated at (u_0, v_0) . The 16 detected corner points with sub-pixel level accuracy for a Type 2 target are plotted with the yellow crosses in Figure 4e. The geometric center of the target in the image frame, which is plotted as the red cross in Figure 4e, is determined by the mean coordinates of all detected corner points.

Given the sub-pixel coordinates of the detected corner points in the reference video frame, a Kanade-Lucas-Tomasi (KLT) feature tracker [32,33] is initialized with these corner points to track their movements in the remaining video frames. With the tracked corner points in each video frame, the geometric center of the target is computed. The algorithm

can then return to a time series of the located geometric center of each target with sub-pixel level accuracy.

Sample corner point detection and target center extraction results for various square checkerboard patterns are shown in Figure 5. Good visual alignments are observed for both the bounding box extraction (blue boxes) and the corner points detection (yellow crosses). It is worth mentioning that the number of extracted corner points is different for each pattern. Type 1 (5-tile) and Type 2 (4-tile) targets have 16 corner points, while the other 2-tile targets have only 9 corner points. During the implementation, some unidentified points are noted for Type 3_R and Type 3_S targets. As a red-and-white pattern, Type 3_R has less contrast between the red square pattern and background after converting the colored image into grayscale. Therefore, detection of the two corner points between the red squares and the background becomes less reliable. For the Type 3_S target, the side length of the black square in the pattern is 5 cm, corresponding to only four to five pixels in the video frame. This small size reduces the values in the spatial gradient matrix of the corner point because of the lower contrast between the pattern region and background, leading to a reduction in the value of the response function $\Gamma(x, y)$ for the corner point. As a result, the proposed algorithm provides less reliable detection for the two corner points between the black squares and the background. Since these unreliable detections are associated with small response function values, only the corner points corresponding to the seven largest response function values after nonmaximal suppression are selected as the detected corner points for Type 3_R and Type 3_S targets. Two corner points between the colored square and the background are excluded in the detection. This operation ensures the robustness of each detected corner point.

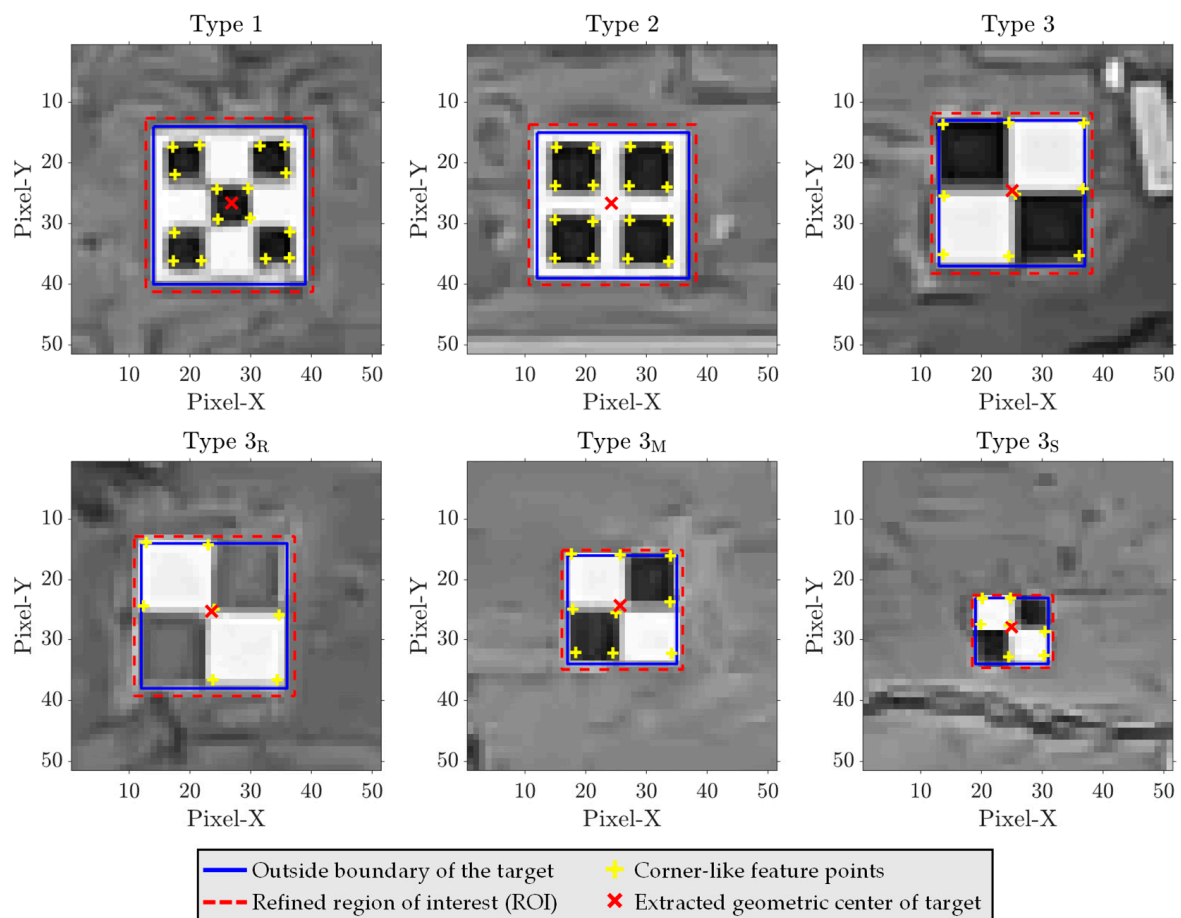


Figure 5. Results of corner points detection and target center extraction for different square checkerboard patterns.

3.2. Hough Transform-Based Algorithm for Concentric Circular Pattern Targets

Different from square checkerboard patterns, the concentric circular pattern does not have any corner-like feature points. Therefore, methods based on the image spatial gradient, including the Harris-Stephens feature detection algorithm and KLT feature tracking algorithm, are no longer suitable for identifying and tracking concentric circular patterns. Rather, the Hough Transform [40] is a common technique for circular shape detection and localization. In 2-D space, a circle with a center point at (a, b) and radius of r can be represented as follows:

$$(x - a)^2 + (y - b)^2 = r^2 \quad (4)$$

If a point (x_0, y_0) is given to be on a circle with unknown center point (a, b) and unknown radius r , the point can be mapped to the parameter space (a, b, r) using the following equations:

$$a = x_0 - r \cdot \cos(\theta) \quad (5)$$

$$b = y_0 - r \cdot \sin(\theta) \quad (6)$$

where θ is an angle ranging from 0 to 2π . For each r value within a given radius range, a series of candidate (a, b) values can be computed as θ varies from 0 to 2π . Therefore, a series of (a, b, r) parameter sets can be obtained. A 3-D accumulator array $A(a, b, r)$ is developed to record the counts of each parameter set for all points (x_0, y_0) on the circle. Finally, the peak value in the accumulator $A(a, b, r)$ corresponds to the most likely circle parameters (a, b, r) .

A Hough Transform-based algorithm is proposed to determine the geometric center of the concentric circular pattern in each video frame based on the pre-defined initial ROI (in the reference video frame). In this study, the Type 4 target is a concentric circular pattern with three circles. Detection of the circular shapes and their center points is implemented based on the following steps:

- **Step-1:** The Canny edge detection [36] is performed on the initial ROI grayscale image to obtain sub-pixel coordinates for the edge points of circles.
- **Step-2:** The Hough Transform is performed in the ROI grayscale image. Because there are three circles in each target, a heuristic clustering algorithm is applied to cluster all circle parameters (a, b, r) into three clusters based on the radius value r . Each cluster center represents the center point and radius of each circle.
- **Step-3:** The geometric distance between the center point of the inner circle and the mean of the center points of the other two circles is determined. If the distance is greater than one pixel, the center point of the inner circle will be suppressed for the calculation of the target geometric center. This step aims to avoid the potential inaccurate detection of circles with small diameters. During the implementation, some unstable detections are observed for the center point of the inner circle in the pattern. The diameter of the inner circle is 6 cm, corresponding to only five to six pixels in the video frame. For circles with a small radius, the number of edge points on the circumference is relatively small. Since the Hough Transform relies on accumulation of the parameter set from edge points, the low number of edge points results in fewer votes in the accumulator. Therefore, the inner circle is more difficult to detect and localize than the other two circles.
- **Step-4:** The geometric center of the target (u, v) is computed as the mean of center points of all the unsuppressed circles. An example of circle detection and target center extraction for Type 4 target is shown in Figure 6.
- **Step-5:** Repeat steps 1 through 4 for all video frames. In each video frame, the ROI is updated as a 51×51 pixel-sized region with the center point (u, v) as the target center

detected in the previous video frame. The algorithm finally returns a time series of the geometric center of the target with sub-pixel level accuracy.

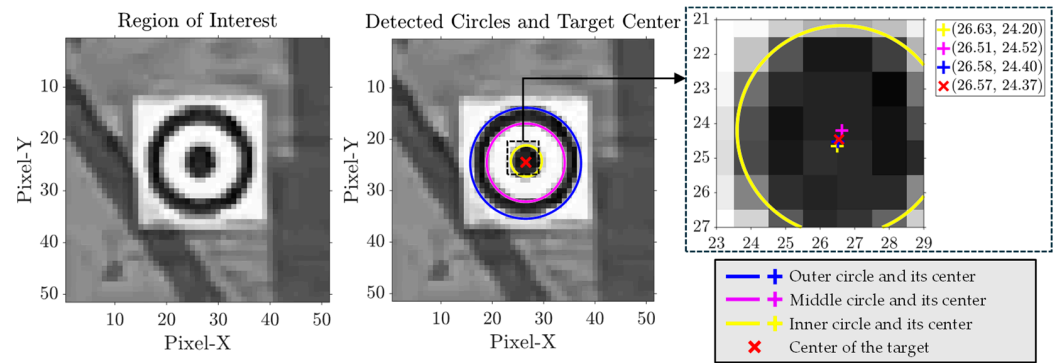


Figure 6. Detected circles with center points by Hough Transform and the final target center for the concentric circular pattern (Type 4 pattern).

Figure 7 shows an example target detection result from the first distortion-corrected plan-view video frame in MID 15. With the feature points tracking-based algorithm for square checkerboard patterns and the Hough Transform-based algorithm for concentric circular patterns, 10 stationary targets and 28 moving targets are successfully tracked for all test videos. The range of pixel resolution is 14–16 mm/pixel for stationary targets on the ground, 9.5–11 mm/pixel for stationary targets on the safety tower, and 8–10.5 mm/pixel for moving targets on the roof panel.

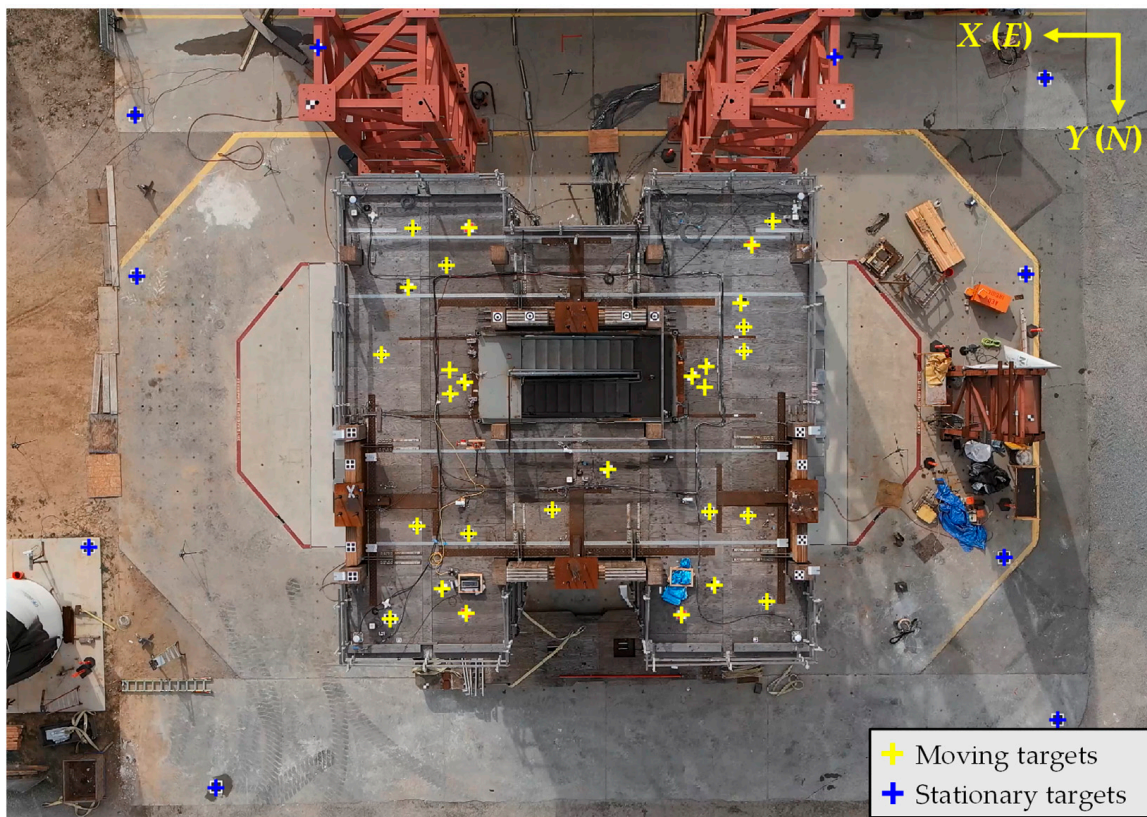


Figure 7. Target detection results in the first video frame (distortion corrected) from plan-view video of MID 15. Note that each cross represents the geometric center of the reference target.

4. Three-Dimensional World Point Reconstruction and Structural Displacement Extraction

A flowchart for the proposed UAV vision-based earthquake-induced structural displacement monitoring framework is shown in Figure 8. Given the sub-pixel coordinates of the reference targets in each video frame from the target detection and tracking algorithms described in Section 3, the 3-D world points are reconstructed based on the following camera projection equation [41]:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[R|t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & \gamma & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{7}$$

where s is an arbitrary scale factor; $[u \ v \ 1]^T$ is the homogeneous representation of the image point coordinates; $[X \ Y \ Z \ 1]^T$ is the homogeneous representation of the world point coordinates; K is the camera intrinsic matrix (f_x and f_y are the focal lengths of the camera, u_0 and v_0 are the principal point coordinates in the image plane, and γ is the image skew parameter); and $[R|t]$ is the camera extrinsic matrix (R is a 3×3 orthogonal matrix representing the camera orientation, and t is a 3×1 vector representing the camera position in world coordinates).

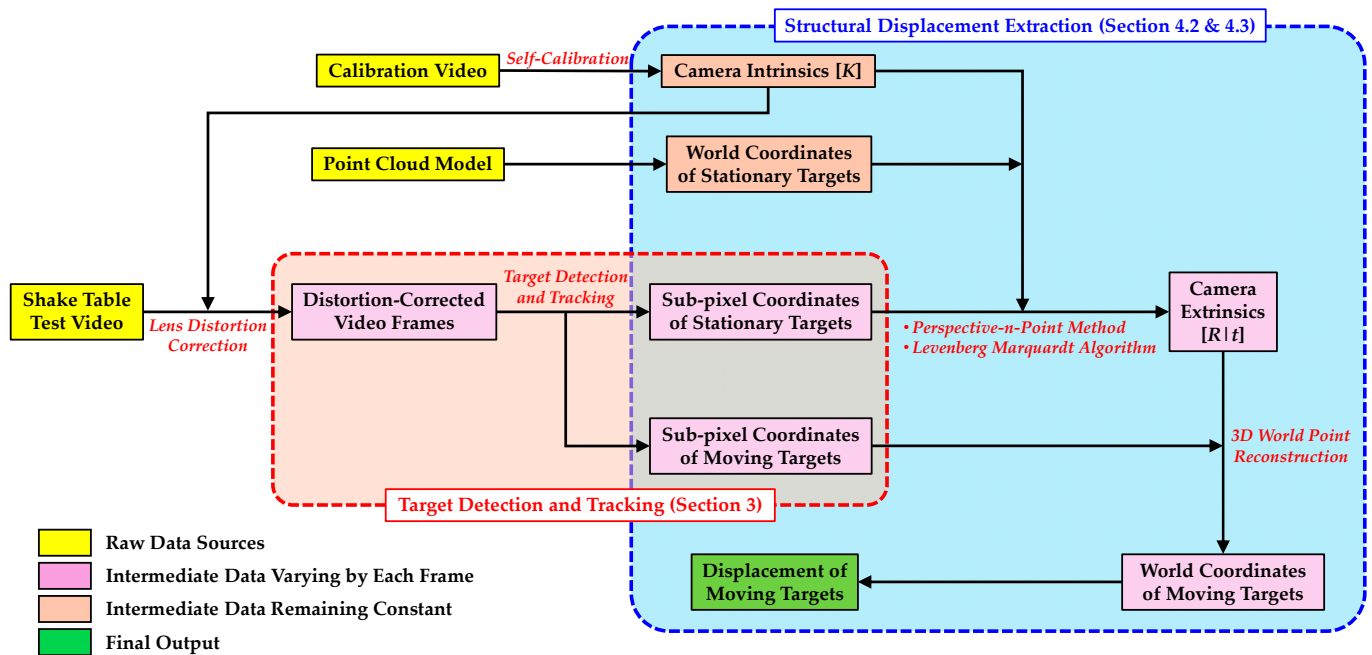


Figure 8. Framework adopted herein for UAV vision-based earthquake-induced structural displacement monitoring.

To solve for the world coordinates of the reference targets, camera calibration is performed first to obtain the camera intrinsics K (Section 4.1). The camera intrinsic parameters are also used to remove the lens distortion effects from the raw video frames. Then, camera pose recovery is applied to estimate the camera extrinsic matrix $[R|t]$ (Section 4.2). The 3-D world points are finally calculated by solving Equation (7) (Section 4.3).

4.1. Camera Calibration and Lens Distortion Correction

Camera lenses, especially wide-angle lenses, introduce radial and tangential distortions in images. Radial distortion causes straight lines to appear as curved in the image,

because light rays passing through the lens refract differently based on their distance from the center of the lens. Tangential distortion occurs when the lens is not perfectly parallel to the image plane, causing the image to be slightly skewed. These distortions lead to inaccurate position and size of objects in the image. Lens distortion correction using accurate camera intrinsics is required to remove this lens distortion effect, making the image more representative of the real scene.

Camera intrinsics can be obtained during the camera calibration process. In the proposed method, which is different from the checkerboard calibration procedure [42], camera intrinsics are estimated from a self-calibration process as part of the bundle adjustment in the structure-from-motion reconstruction using Agisoft Metashape 2.0.2 [43]. Self-calibration of a camera does not depend on any known reference objects. Instead, self-calibration relies solely on multiple images taken by the camera. Self-calibration uses feature points from the images captured from multiple viewpoints, combined with geometric constraints to estimate the camera intrinsics. For more details on this camera calibration procedure refer to Fraser [44] and Westoby et al. [45]. Given accurate camera intrinsics, lens distortions are corrected based on the procedures illustrated in Bouguet [42].

4.2. Camera Pose Recovery

The 6-DoF camera pose is defined by the camera extrinsic matrix $[R|t]$, where R is a 3×3 orthogonal matrix representing the camera orientation and t is a 3×1 vector representing the camera position in world coordinates. In the proposed method, the camera pose in each video frame is recovered based on the Perspective- n -Point (PnP) method followed by the Levenberg-Marquardt algorithm.

The Perspective- n -Point (PnP) method estimates the pose of a calibrated camera from n 3-D world points with known positions and their corresponding 2-D image points. Although the PnP method requires a minimum of only four non-coplanar world-to-image point correspondences (three non-collinear correspondences if the points are coplanar) for obtaining the unique camera pose, more point correspondences can enhance the robustness and accuracy of the estimated camera pose due to the measurement noise in both world points and image points. The proposed method adopts an accurate and scalable solution to the PnP problem (ASP nP) [46] to estimate the camera-pose changes over time. The image and world point coordinates of 10 stationary reference targets (8 targets on the ground and 2 targets on the safety towers, which are identified as the blue crosses in Figure 7) are used as the world-to-image correspondences in each video frame. The image point coordinates are obtained as the target detection and tracking algorithm described in Section 3. The 3-D world point coordinates are extracted from a geo-referenced point cloud model of the building specimen and the background region of the test scene at the LHPOST6 (Figure 9). The point cloud model is generated based on the photogrammetry results from the same UAV and camera combination for the plan-view test video capture (i.e., the DJI Matrice 300 UAV with the DJI Zenmuse P1 camera) using a preprogrammed flight plan. The photogrammetry flight pattern is determined case by case in practice based on the complexity of the scene [47,48]. For the present study, the image acquisition consists of 209 images in a lawnmower pattern using nadir imagery and oblique imagery (40 degrees off the vertical) with a target 75% overlap and sidelap at the roof level. Photogrammetric processing is performed using Agisoft Metashape 2.0.2 [43]. The resulting point cloud consists of around 480 million points with a nominal point spacing of 6 mm.

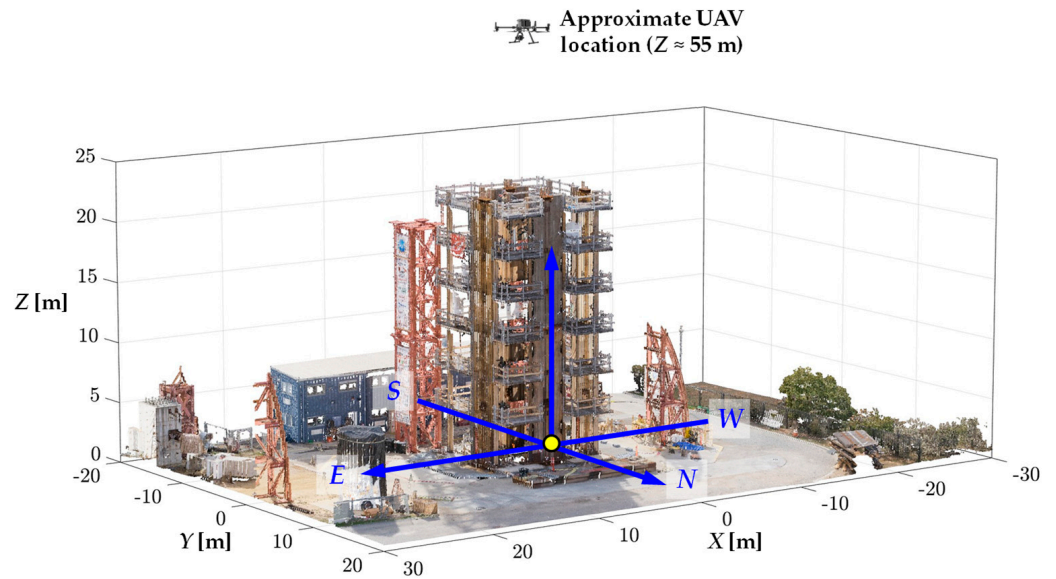


Figure 9. Point cloud model of the test scene at the LHPOST6 from photogrammetry and illustration of the 3-D world coordinates (the yellow dot represents the center of the shake table platen, which is the origin of the world coordinates, and the blue lines represent the XYZ coordinate axes). Note that $Z = 0$ is defined as the top of the shake table platen.

As a non-iterative method for camera pose recovery, $ASPnP$ obtains the optimal camera pose by solving a polynomial system derived from the first-order optimality condition for the image-to-world reprojection error [46]. However, non-iterative PnP solutions can still be improved by some iterative methods such as the Gauss-Newton algorithm or the Levenberg-Marquardt algorithm [49,50]. In the proposed method, the Levenberg-Marquardt (LM) algorithm [51,52] initialized with the $ASPnP$ solution is applied as a nonlinear optimization process for the camera pose. The 6-DoF camera pose can be accurately estimated based on this $ASPnP$ -LM process. Figure 10 shows the camera translational motion trajectory of the onboard camera for plan-view video capture during MID 15. According to the estimated camera motion trajectory, the UAV for plan-view video capture hovered at around 55 m above the shake table platen. The UAV had a drift of less than 0.2 m in both horizontal directions and a drift of less than 0.1 m in vertical direction.

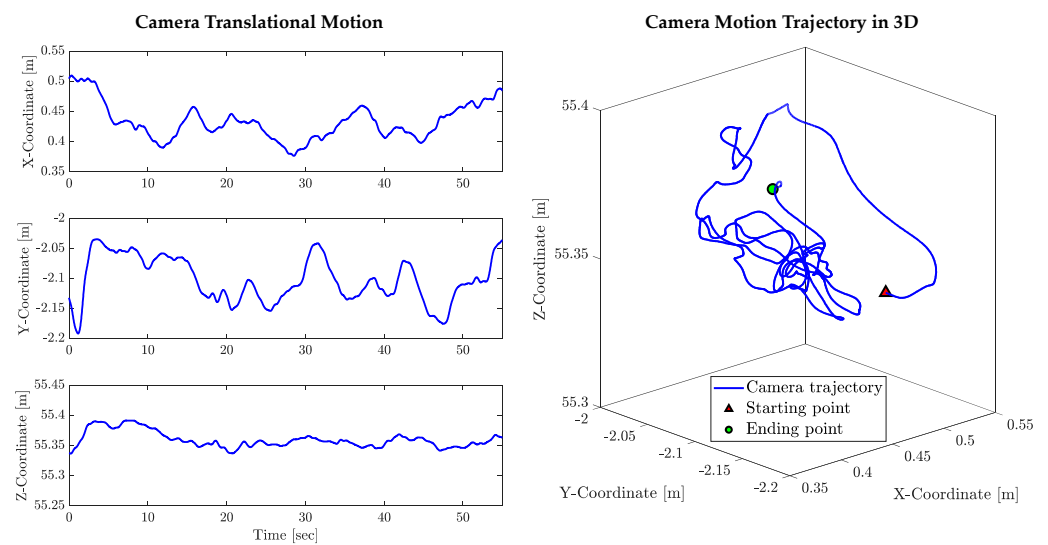


Figure 10. Camera translational motion trajectory for the plan-view video of MID 15.

4.3. Three-Dimensional World Point Reconstruction and Structural Displacement Extraction

Given the reference target image coordinates $[u, v]^T$, camera intrinsic parameters K , and camera extrinsic parameters $[R | t]$, the 3-D world point of the reference target can be reconstructed based on the camera projection equation in Equation (7). The test video is captured by a monocular camera mounted on the UAV. Therefore, scale ambiguity leads to four unknown variables (world coordinates XYZ and the scale factor s) in Equation (7). However, one may assume that the Z -coordinates of all moving targets remained constant as Z_0 during the test (i.e., zero vertical displacements for the targets). This assumption is considered reasonable in the scope of this study because the displacements of the building specimen in the vertical direction are much smaller than the displacements in the lateral direction due to the extremely large self-weight of the structure. In addition, the change in image point coordinates caused by the out-of-plane motion relative to the camera view plane is small because of the large camera-to-scene distance (around 35 m between camera and roof level) [53]. Based on this assumption, Equation (7) can be expanded into the following form:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \left(R \begin{bmatrix} X \\ Y \\ Z_0 \end{bmatrix} + t \right) \quad (8)$$

To solve the world coordinates X and Y , Equation (8) is re-arranged as follows:

$$\begin{bmatrix} X \\ Y \\ Z_0 \end{bmatrix} = sR^{-1}K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} - R^{-1}t \quad (9)$$

The scale factor s can be calculated by solving the equation defined by the 3rd row of Equation (9), where the scale factor is the only unknown variable. Then, world coordinates X and Y can be calculated by back-substituting the scale factor into the first two rows of Equation (9).

Based on the 3-D world point reconstruction procedures above, the world coordinates X and Y of the reference target in each video frame can be solved. The displacement of the reference target at a specific video frame is calculated as the change in the world coordinates of the target relative to the reference video frame (the first frame in the test video). It is noted that the calculated displacement is the absolute (total) displacement of the roof. The motion of the shake table platen is also included in the displacement.

5. Structural Displacement Extraction Validation

In this section, the target detection and tracking algorithm described in Section 3 and the 3-D world point reconstruction and structural displacement extraction framework presented in Section 4 are validated using a series of UAV-captured shake table test videos from the Phase II test series of the NHERI Converging Design Project (refer to Section 2). The proposed methodology is successfully implemented with the Image Processing and Computer Vision Toolbox in MATLAB R2023b [54]. The displacements of all the reference targets on the roof level are extracted, and they are further compared to the ground truth measurement from multiple analog sensors installed on the building specimen to characterize the measurement errors. The effectiveness of the proposed UAV vision-based earthquake-induced structural displacement monitoring method is evaluated in this section.

5.1. Ground Truth Measurements and Comparison Plan

5.1.1. Processing Steps for Ground Truth Measurements

For each reference target, ground truth measurements are roof level displacements, determined by double integrating the accelerations recorded by an accelerometer installed close to the target. The processing steps for ground truth measurements include the following [28]:

- **Baseline correction:** The raw acceleration time series is baseline corrected using the mean value of the pre-event data (the first 100 data points for each test).
- **Zero-padding:** The baseline corrected acceleration data is tapered for the first and last second, then the data is zero-padded for 20 s at both the beginning and the end of the time series.
- **Filtering:** A fourth-order bandpass Butterworth filter with cutoff frequencies of 0.1 and 50 Hz is applied to remove measurement noise above and below those limits.
- **Integration:** The fourth-order Runge-Kutta method is applied to integrate the filtered acceleration to obtain the velocity.

The baseline correction, filtering, and integration steps outlined above are repeated to obtain the calculated velocity and then to obtain the displacement time series. The final displacement result is obtained by baseline correcting, filtering, and truncating (removing the zero-padded portions) of the time series.

5.1.2. Analog Sensor Plan for Results Comparison

Figure 11 shows the distribution of the reference targets and selected analog sensors, which are the source of the ground truth measurements. In this analysis, the roof is divided into five regions. Displacements extracted from targets within each region are primarily considered to be compared with the measurements from the analog sensor within that region, which is the closest sensor to these targets. Therefore, error caused by the location difference between the ground truth measurements and reference targets is minimized. However, large measurement noise or unreasonable discrepancies between the UAV-based analysis results and the ground truth measurements are observed for a few analog measurement channels (specifically, channels 121, 123, and 702) during several tests. For these cases, where reasonable comparison cannot be achieved using measurements from the closest sensor, measurements from another sensor (i.e., channel S433), which is the second closest sensor to the target, are selected for comparison. Table 5 presents the comparison plan of different regions on the roof for each test.

Table 5. Result comparison plan of different regions on the roof panel for each test.

Region Number	Analog Sensors Assumed as Ground Truth for Each Test								
	MID 7	MID 8	MID 12	MID 13	MID 14	MID 15	MID 16	MID 17	MID 18
①	C048: X C050: Y	C048: X C050: Y	C048: X C050: Y	C048: X C050: Y	C048: X C050: Y	C048: X C050: Y	C048: X C050: Y	C048: X C050: Y	C048: X C050: Y
②	121: X 122: Y	S433: X 122: Y	121: X 122: Y	121: X 122: Y	121: X 122: Y	S433: X 122: Y	121: X 122: Y	121: X 122: Y	S433: X 122: Y
③	702: X S436: Y	702: X S436: Y	702: X S436: Y	702: X S436: Y	S433: X S436: Y	S433: X S436: Y	S433: X S436: Y	S433: X S436: Y	S433: X S436: Y
④	C047: X 127: Y	C047: X 127: Y	C047: X 127: Y	C047: X 127: Y	C047: X 127: Y	C047: X 127: Y	C047: X 127: Y	C047: X 127: Y	C047: X 127: Y
⑤	123: X 124: Y	S433: X 124: Y	123: X 124: Y	123: X 124: Y	123: X 124: Y	S433: X 124: Y	123: X 124: Y	123: X 124: Y	123: X 124: Y

Note: Channel numbers with **bold italic font** represent cases where another sensor instead of the closest sensor to the target is selected for displacement result comparison (due to noisy or inadequate quality acceleration measurement).

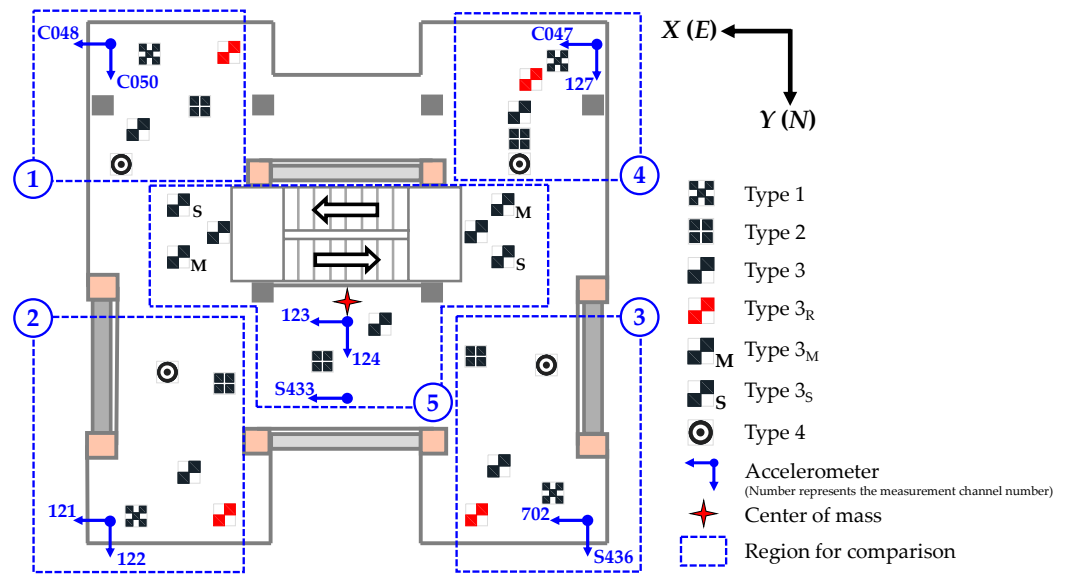


Figure 11. Reference target distribution in five regions of the roof plan and the selected analog sensors utilized as ground truth measurements.

5.2. Structural Displacement Tracking Results and Error Characteristics

In the proposed UAV vision-based earthquake-induced structural displacement monitoring framework, the image points are reprojected to the 3-D world space based on the world-to-image point correspondence at each video frame. Then, the displacement of each reference target is computed as the coordinate changes with respect to its initial world coordinates in the first video frame. The extracted displacements are compared to ground truth measurements based on the comparison plan illustrated in Section 5.1.2. Since the ground truth measurements are the filtered results as described in Section 5.1.1, the same fourth-order bandpass Butterworth filter with cutoff frequencies of 0.1 and 50 Hz is also applied to the extracted displacements to avoid the possible discrepancy caused by the filtering.

For brevity, comparisons of the displacement results of a Type 2 target (4-tile black-and-white checkerboard pattern) from two Ferndale earthquake tests (MID 12 and 15) and a Type 4 target (black-and-white concentric circular pattern) from two Northridge earthquake tests (MID 7 and 8) are presented in Figure 12. These results indicate an overall root-mean-square error (RMSE) between the UAV vision-based analysis results and ground truth measurements of less than 1 cm. In addition, the percentage error at the peak displacement amplitude is also calculated as follows:

$$\text{Error}\% = \frac{\Delta_{\text{UAV}} - \Delta_{\text{GT,max}}}{\Delta_{\text{GT,max}}} \times 100\% \tag{10}$$

where $\Delta_{\text{GT,max}}$ is the maximum displacement from ground truth measurement, and Δ_{UAV} is the displacement value at the timestamp corresponding to $\Delta_{\text{GT,max}}$. The overall percentage error per this computation at the peak displacement amplitudes is less than 10%. The results indicate that the displacements extracted from the proposed method show good alignment with the ground truth measurements for both the overall trend and the peak displacements regardless of locations of measurements, reference target patterns, displacement directions, earthquake motion input intensities, and earthquake events.

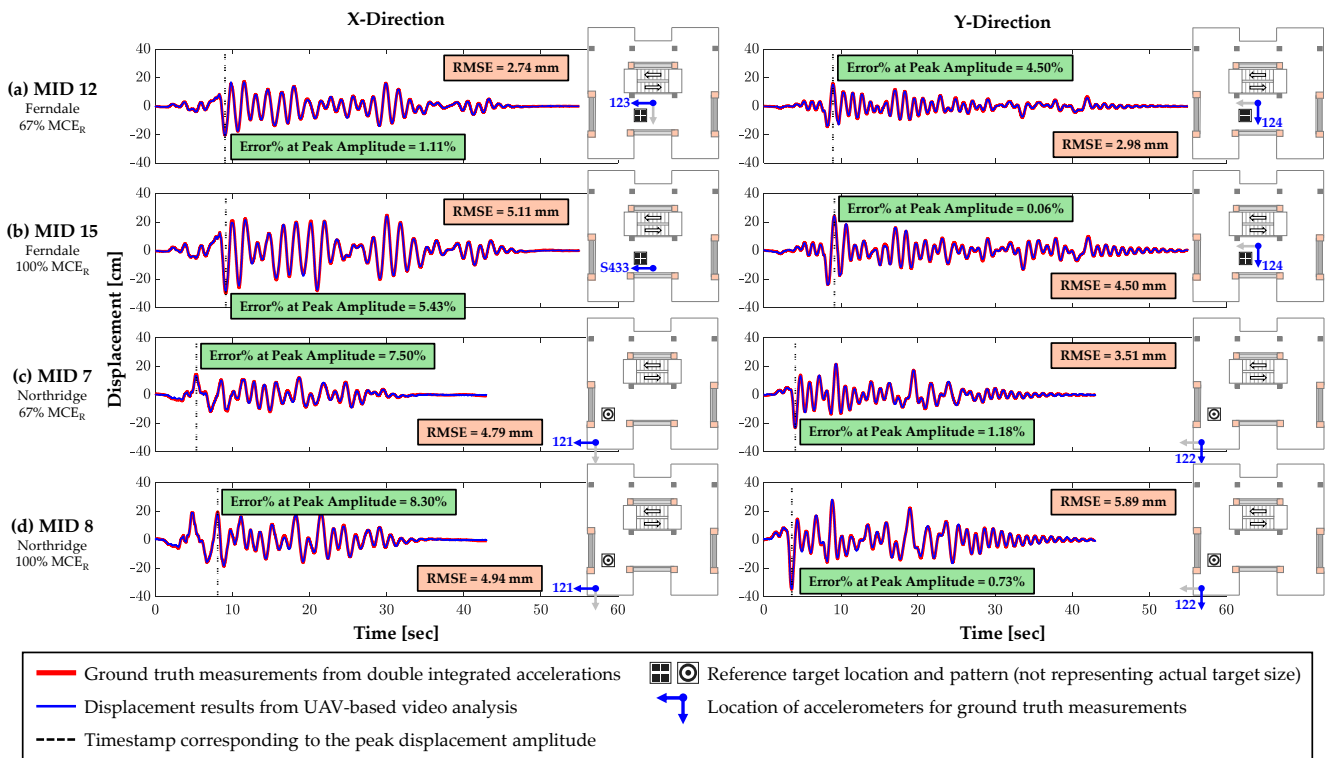


Figure 12. Displacement time series comparison between UAV-based video analysis and ground truth measurements obtained from double-integrated accelerations using analog sensors. (a,b) are results of a Type 2 target (black-and-white checkerboard) close to the center of mass of the building for Ferndale earthquake tests. (c,d) are results of a Type 4 target (black-and-white concentric circles) close to the northeast corner of the building for Northridge earthquake tests.

Figure 13 summarizes the average RMSEs for tracking the roof-level displacements across the various reference target patterns for each test. Each light-colored data point in this plot represents the average RMSE calculated from multiple displacement time series comparisons in the X or Y direction of one test (results from each single target forms a comparison) with a specific pattern. The average RMSEs from all seven patterns are also presented as filled symbols within the plots. For earthquake inputs in the given direction from a specific earthquake event (e.g., X-direction inputs of the Ferndale earthquake), larger average RMSEs are observed in the test with a greater earthquake input motion intensity, i.e., when the achieved peak input acceleration (PIA) is larger. To provide a generalized quantitative illustration of the effect of earthquake input motion intensity on structural displacement tracking errors, the PIAs in the X and Y directions for each test, along with the corresponding average RMSEs over all reference target patterns, are normalized by the achieved PIA and average RMSE values from the 100% MCE_R test in the given direction of that specific earthquake event. Mathematical expressions of this normalization are presented in Equations (11) and (12), where PIA_{MCE_R} and RMSE_{MCE_R} in the denominator are the achieved PIA and average RMSE from the 100% MCE_R test for the same earthquake event and input direction as the corresponding achieved PIA and average RMSE in the numerator. The relationship between normalized average RMSEs and normalized achieved PIAs are presented in Figure 14. Because all average RMSEs and achieved PIAs are normalized by the results from the 100% MCE_R tests, the normalized data for the 100% MCE_R tests are exactly at (1, 1), which is indicated by the black filled symbol within the plot. The normalized average RMSEs for the 67%, 68.9%, and 110% MCE_R tests are all close to the normalized achieved PIAs of these tests. Therefore, the average RMSEs are proportional to the achieved PIAs (i.e., proportional to earthquake

motion input intensity), though this proportionality holds separately for each earthquake event and input direction.

$$\text{Normalized PIA} = \frac{\text{PIA}}{\text{PIA}_{\text{MCER}}} \tag{11}$$

$$\text{Normalized RMSE} = \frac{\text{RMSE}}{\text{RMSE}_{\text{MCER}}} \tag{12}$$

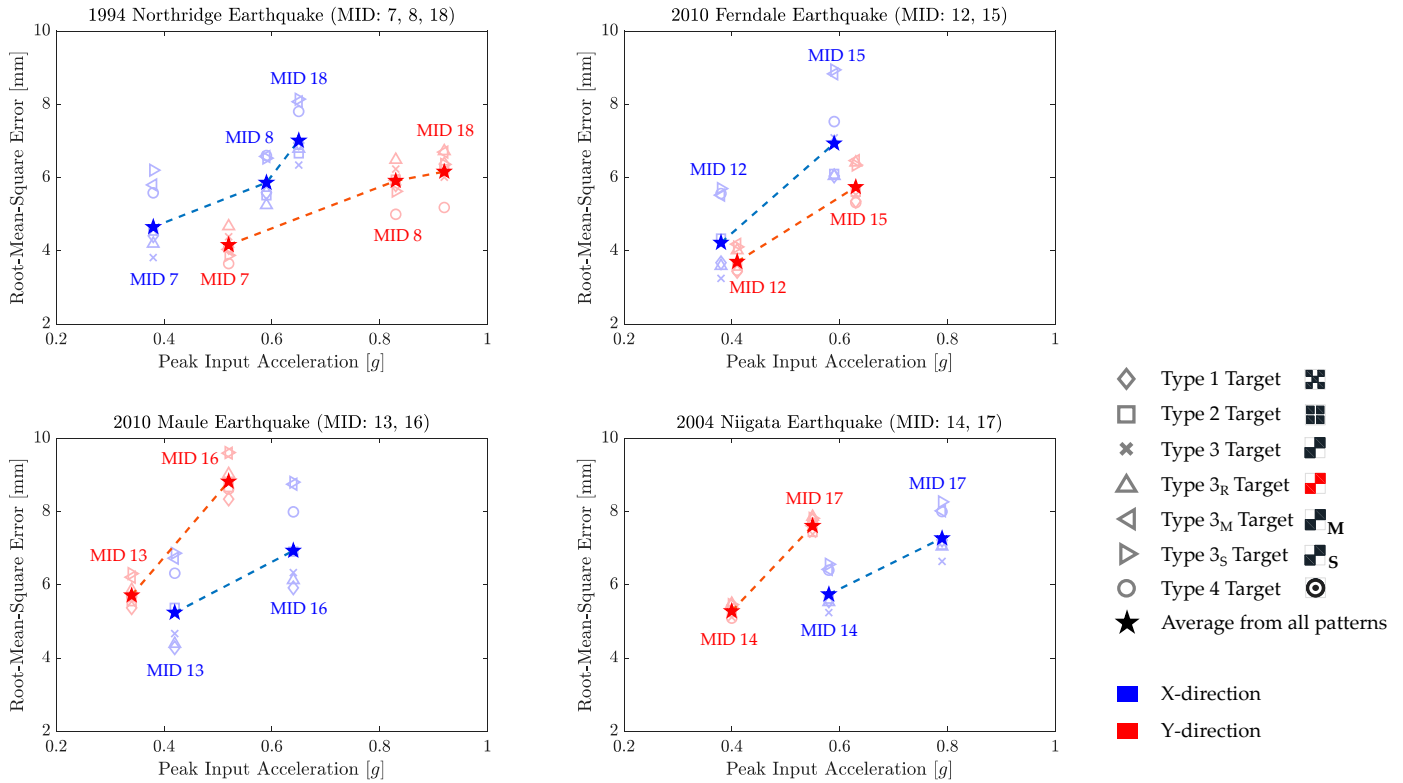


Figure 13. Average root-mean-square errors (RMSEs) of the tracking results for the roof-level displacement for various reference target patterns under each test.

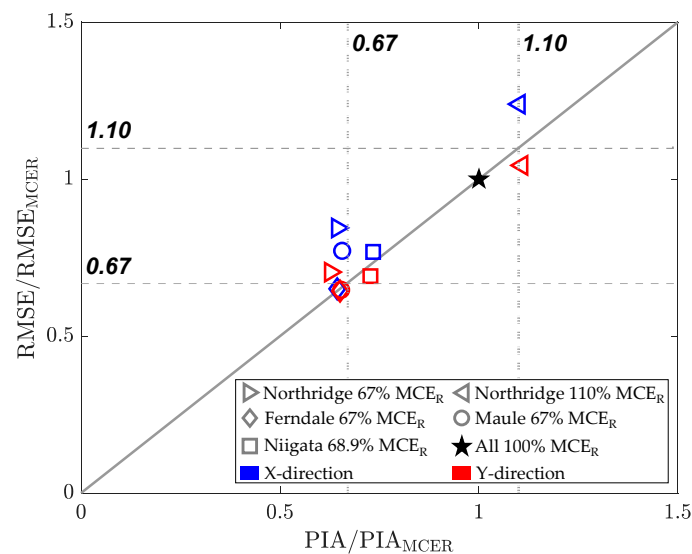


Figure 14. Relationship between normalized average root-mean-square errors (RMSEs) and normalized achieved peak input accelerations (PIAs).

6. Optimal Features for Earthquake-Induced Structural Displacement Monitoring

As the proposed UAV vision-based earthquake-induced structural displacement monitoring method is evaluated using the shake table test results of a full-scale 6-story mass timber building, a statistical analysis of the error characteristics across various patterns is conducted to aid in identifying the optimal features for structural displacement tracking under earthquakes. As a metric, the statistical analysis uses the root-mean-square error (RMSE) of the displacement tracking result from each individual target. Figure 15 presents the statistics of RMSEs for tracking the displacements of the 6-story mass timber building specimen, focusing on the comparison between the different reference target patterns. The RMSEs of displacement tracking from three 67% MCE_R tests (i.e., MID 7, 12, and 13) and four 100% MCE_R tests (i.e., MID 8, 15, 16, and 17) are included in this result. For a particular test, every target provides two measurements (one in the X-direction and the other one in the Y-direction), and each measurement leads to a data point in the RMSE statistics. Therefore, the total number of data points n for each pattern in the RMSE statistics can be calculated as follows:

$$n = 2 \cdot n_{\text{target}} \cdot n_{\text{test}} \quad (13)$$

where n_{target} is the number of targets with a specific pattern, and n_{test} is the number of tests with a specific earthquake motion input intensity level. Several insights are observed from the RMSE statistics:

- **Overall Tracking Accuracy Evaluation:** RMSEs in these analyses are consistently less than 8 mm for the 67% MCE_R tests and less than 10 mm in general for the 100% MCE_R tests (only 6 data points have errors that exceed 10 mm among the 224 data points in total for the MCE_R tests). Thus, the proposed structural displacement monitoring method can achieve an overall tracking accuracy at the millimeter level.
- **Feature Size:** For a certain pattern shape (2-tile black-and-white checkerboard), significantly larger RMSEs are observed for smaller pattern sizes (Type 3_M and Type 3_S) compared to a larger pattern size (Type 3). Smaller reference targets have less pixel dimensions in the video frame. For example, the side length of the black square in the Type 3_S pattern is 5 cm, which corresponds to only four to five pixels in the video frame. The intensity contrast between the target region and the background is significantly reduced, leading to reduced performance on the refined ROI generation with Otsu's thresholding and edge detection. Additionally, as previously mentioned in Section 3.1, the small pixel dimensions of reference targets reduce the values in the spatial gradient matrix of the corner point in Harris-Stephens feature detection. Therefore, it is observed that targets with smaller dimensions have larger RMSEs for displacement tracking by comparing 2-tile black-and-white checkerboard targets with different sizes. Aiming to investigate the minimum required feature (pattern) size for reasonable displacement tracking under different intensities of earthquakes, a statistical analysis on the normalized RMSEs for displacement tracking with respect to normalized tile dimensions of targets is conducted for all the black-and-white checkerboard targets (specifically, Type 1, 2, 3, 3_M , and 3_S) in all nine tests. For each individual target during a specific test, both the tracking RMSE (in X/Y-direction) and the tile dimension D (the side length of each black square) are normalized by the peak displacement $\Delta_{GT,\max}$ (in the X/Y-direction) in each ground truth measurement. Figure 16 presents a compilation of the statistical analysis results. Each data point of the normalized RMSE with respect to the normalized tile dimension ($D/\Delta_{GT,\max}$) is depicted with gray symbols. As previously presented in Table 2, a larger overall peak roof displacement in the Y-direction is observed compared to the X-direction

for the nine tests in this study. In addition, very large peak roof displacements in the Y-direction are recorded during the 100% MCE_R Northridge earthquake tests (MID 8 and 18). Therefore, both RMSEs and tile dimensions after normalization (in the Y-direction) are less than the values in the X-direction. For each direction, data points are divided into multiple bins based on the normalized tile dimension with a fixed bin width of 10%, which are indicated by the red and white background in the plot. The statistics of the data within each bin are presented by a box and whisker plot in Figure 16. As the normalized tile dimensions ($D/\Delta_{GT,max}$) increase, the normalized RMSEs decrease from 5.78% to 3.56% in the X-direction and decrease from 2.68% to 1.46% in the Y-direction. For both X- and Y-direction, a relatively stable RMSE can be achieved when the normalized tile dimensions ($D/\Delta_{GT,max}$) are greater than 50% (i.e., the actual tile dimension of the target is greater than 50% of the peak displacement during the earthquake).

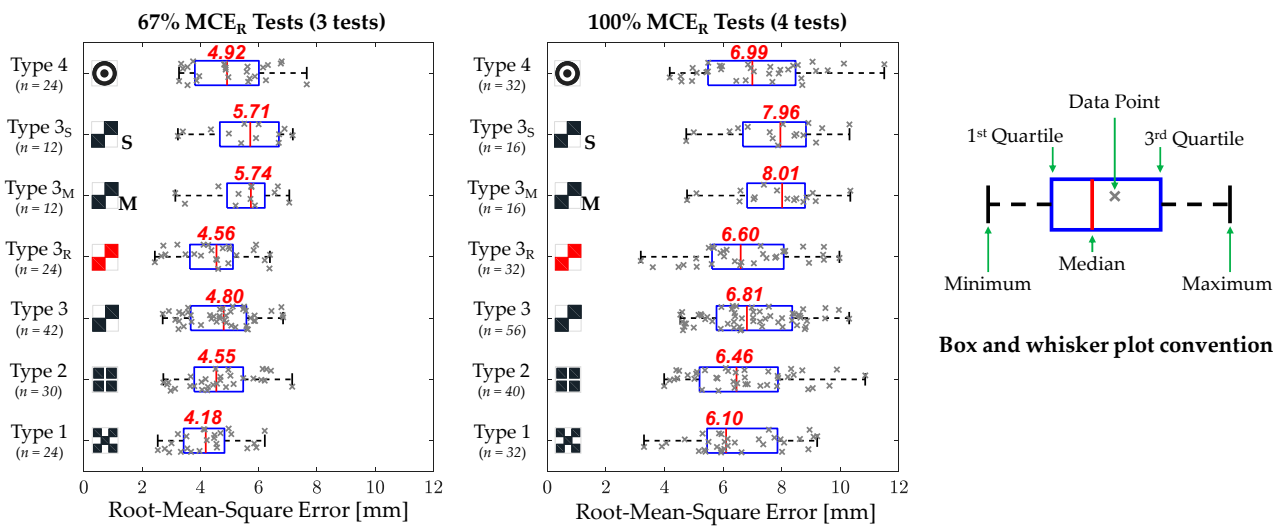


Figure 15. Root-mean-square error (RMSE) statistics for various reference target patterns in displacement tracking of the 6-story mass timber building specimen under 67% and 100% MCE_R tests (n is the number of RMSE data points for each pattern; numbers in red color indicate the median value of each data group).

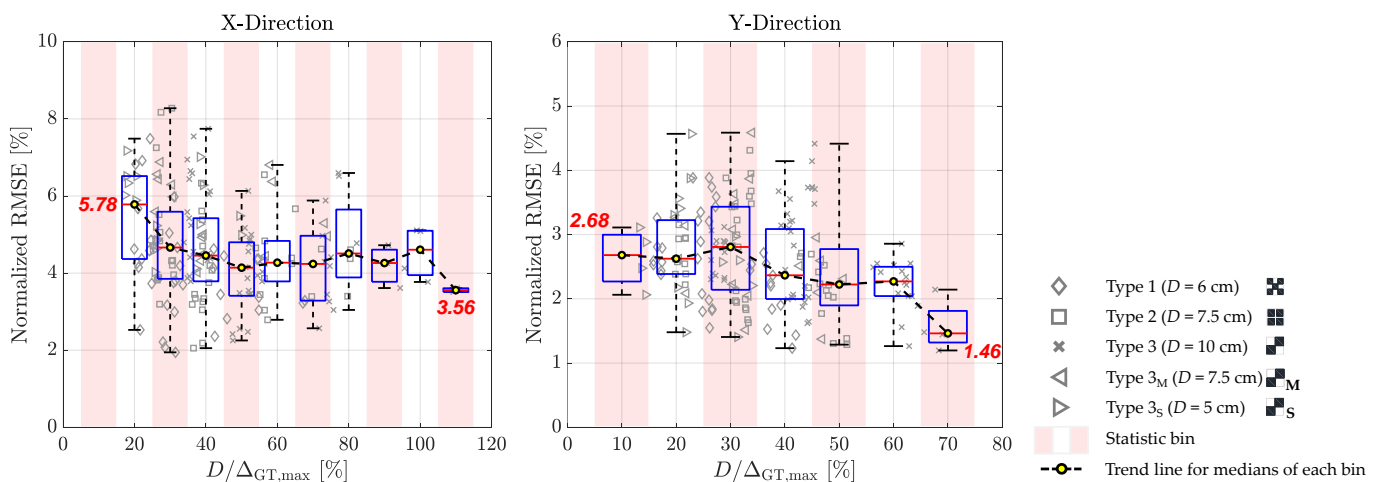


Figure 16. Statistics for normalized root-mean-square errors (RMSEs) with respect to normalized tile dimension ($D/\Delta_{GT,max}$) of black-and-white checkerboard targets. Note that the box plot is the statistics for the data points within each bin. Numbers in red represent the medians for data in the first and last bin.

- **Intensity Contrast:** In the proposed target detection and tracking algorithm, the colored ROI image is converted into grayscale for the convenience of thresholding and the reduced computational cost. Type 3 (black-and-white) and Type 3_R (red-and-white) targets are both 2-tile checkerboard patterns with the same geometric shape and dimension but with different colors. Five pairs of contrasting colors involved in target detection and tracking for these two patterns are listed below in descending order of intensity difference within the color pairs in the grayscale image: (1) black and white, (2) red and white, (3) white and background, (4) black and background, and (5) red and background. As previously discussed in Section 3.1, the two corner points between red squares and background are suppressed for the Type 3_R pattern in feature point detection and tracking. Therefore, displacement tracking for the Type 3_R pattern relies largely on the red-to-white and white-to-background contrast which have relatively large intensity differences within each color pair. However, displacement tracking for the Type 3 targets relies on black-to-white, white-to-background, and black-to-background contrasts, where the intensity difference between the black region and the background is relatively low and it may reduce the performance of feature point localization. This insight is reflected by the slightly larger results in the RMSE statistics for the Type 3 targets compared to the Type 3_R targets.
- **Pattern shape:** Square checkerboard patterns and concentric circular patterns are included for the structural displacement monitoring of the 6-story mass timber building specimen. The square checkerboard patterns provide orthogonal features with explicit feature points, while the concentric circular pattern only provides circular shapes without any explicit point for tracking. Slightly larger RMSEs are observed for the concentric circular pattern (Type 4) compared to square checkerboard patterns with the same geometric dimension and color (Type 1, 2, and 3). Furthermore, additional oscillations are observed in tracking results from the concentric circular pattern. Figure 17 presents the comparison of the displacement results between the multiple reference targets at the northeast corner of the roof level (region 2) from the 67% MCE_R Ferndale earthquake test (MID 12). From the overlay of the displacement time series, there are good visual alignments between the ground truth measurements and UAV vision-based tracking results from all the reference targets in the region. However, additional oscillations can be clearly observed in the tracking results of the concentric circular pattern from the zoom-in view. These additional oscillations occur within low-amplitude regions of the displacement time series (e.g., $t = 7\text{--}8.5$ s in the X-direction and $t = 4\text{--}5.5$ s in the Y-direction for MID 12). These oscillations are not observed in the tracking results from the square checkerboard patterns. This difference is caused by the different algorithms applied in the detection and tracking of the two different types of features (i.e., orthogonal features and circular features). Orthogonal features (corner points) in the square checkerboard patterns are detected first in the reference frame. Then, sub-pixel coordinates of the feature points in the previous video frame are used as the input of the Kanade-Lucas-Tomasi (KLT) feature tracker in the following video frame for displacement tracking. However, the KLT algorithm cannot be applied to circular features since there are no explicit feature points. A Hough Transform is applied to each video frame with a coarsely updated ROI for circular features. The potential error in the ROI updating can be propagated in the Hough Transform in the subsequent video frame. Therefore, compared to concentric circular patterns, orthogonal-shaped patterns with explicit feature points that can be tracked by the Kanade-Lucas-Tomasi (KLT) algorithm are observed to reduce error propagation during the video sequence and reduce the displacement measurement noise.

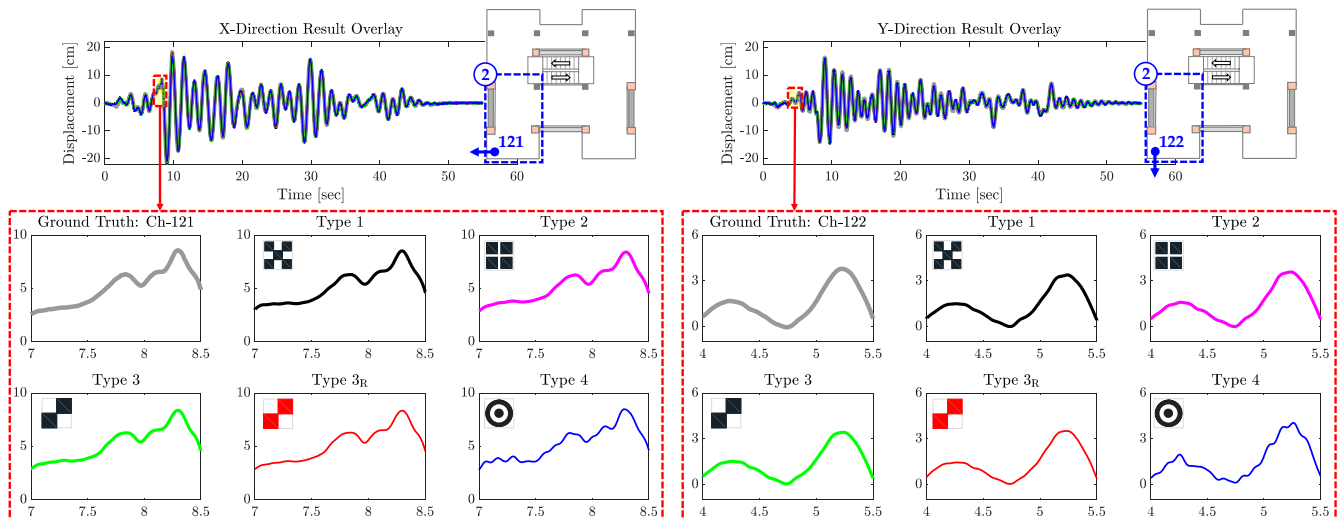


Figure 17. Displacement results comparing multiple reference targets at the northeast corner of the roof level (region 2) from the 67% MCE_R Ferndale earthquake test (MID 12). Robustness of the orthogonal-shaped patterns is readily revealed by comparing Type 1–Type 3 to Type 4 patterns.

7. Conclusions

Vision-based sensing, facilitated by unmanned aerial vehicle (UAV) platforms, has emerged as an effective technology that can enhance life-cycle health monitoring and assessment of civil infrastructures. Such remote, non-destructive measurement approaches offer significant promise in capturing the displacements of structures during natural or manmade hazards. This work aims to improve the robustness of UAV vision-based sensing, focusing on applications of capturing the structural displacements of buildings during earthquakes.

In particular, a capture and analysis framework is proposed based on two target detection and tracking algorithms, considering a variety of feature patterns. Patterns in the scene of interest are either square checkerboards or concentric circles, with black-white or red-white color schemes. Several photogrammetry techniques are applied to extract structural displacements, including camera calibration, lens distortion correction, camera pose recovery, and 3-D world point reconstruction. The proposed methodology is evaluated using ground-truth sensor measurements from shake table tests of a full-scale 6-story mass timber building. Geo-referenced targets with various patterns are placed on the building specimen, along with the stationary background region, to provide robust features in the UAV imagery. Statistical analysis of the error characteristics across the various reference target patterns is conducted to investigate the optimal features for use in displacement tracking during earthquakes. The main takeaways from this study include the following:

- The proposed UAV vision-based method demonstrates reasonable accuracy in tracking structural displacements during a wide range of earthquake motion inputs, with overall root-mean-square errors (RMSEs) at the millimeter level compared to the ground truth measurements from analog sensors.
- Given a specific earthquake event and input direction, the average displacement tracking RMSEs are proportional to the achieved peak input accelerations (PIAs).
- Based on the statistical analysis of the error characteristics across the various reference target patterns, it is observed that the pattern sizes, pattern shapes, and intensity contrast in the region of interest can affect the accuracy of structural displacement monitoring. Orthogonal-shaped patterns (e.g., straight line intersections or squares) with explicit feature points that can be tracked by the Kanade-Lucas-Tomasi (KLT) algorithm are observed to reduce the error propagation during the video sequence and reduce the displacement measurement noise.

- To characterize the effect of pattern size on the robustness of structural displacement monitoring, a relatively stable RMSE can be observed when the normalized tile dimension ($D/\Delta_{CT,max}$) is greater than 50% for black-and-white checkerboard patterns. Therefore, the actual tile dimension of the black-and-white checkerboard pattern is suggested to be greater than 50% of the peak displacement expected during an earthquake to ensure reasonable accuracy in tracking structural displacements.

In summary, this article proposes a methodology to utilize UAV vision-based sensing for monitoring the displacements of full-scale structures during earthquakes. The tracked structural displacements can provide valuable information in post-disaster safety evaluation and functional recovery analysis of the structure. It is worth mentioning that the analysis framework proposed requires post-processing of video data; therefore, the present study does not achieve real-time monitoring during the earthquake event. Although the UAV vision-based capture and analysis framework described herein are based on artificial reference targets rather than natural features of a case-study building specimen, the framework can also be advanced to incorporate target-free structural displacement monitoring approaches. Insights gained from the investigation of optimal features in this article can guide future analysis utilizing natural features of buildings or other structures of interest subjected to dynamic loading.

Author Contributions: Conceptualization, R.J., E.L., F.K. and T.C.H.; methodology, R.J., S.S., E.L. and T.C.H.; investigation, R.J., S.S., E.L., T.J.N. and J.W.D.; data curation, R.J. and S.S.; software, R.J. and T.J.N.; formal analysis, R.J.; validation, R.J.; visualization, R.J.; writing—original draft preparation, R.J.; writing—review and editing, R.J., F.K., A.R.B., B.G.S. and T.C.H.; supervision, F.K. and T.C.H.; resources, F.K., A.R.B., B.G.S. and T.C.H.; project administration, F.K. and T.C.H.; funding acquisition, F.K. and T.C.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the California Governor’s Office of Emergency Services (CalOES) and Seismic Safety Commission (Agreement: A211006635). In addition, the first author is supported by a first-year graduate fellowship from the Structural Engineering Department at UC San Diego. The 6-story NHERI Converging Design Project is supported by the National Science Foundation (Award No. 2120683, 2120684, 2120692), with an important supplement provided by the National Institute of Standards and Technology under the NSF-NIST inter agency agreement. Support from the TallWood Design Institute and USDA Agricultural Research Service (Award No. #58-0204-9-165) are also acknowledged. Findings, opinions, and conclusions are those of the authors and do not necessarily reflect those of the sponsoring organizations.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: In addition to the sponsor organizations mentioned in the Funding section, the authors acknowledge the support from ALERTCalifornia for the UAV platforms used in this study. Collaboration with the NHERI Converging Design team (listed at: <https://tallwoodinstitute.org/converging-design-home-5663/>, accessed on 19 September 2024) and technical support for the shake table test program provided by the NHERI@UCSD staff (listed at: <https://nheri.ucsd.edu/about/personnel>, accessed on 19 September 2024) are greatly appreciated.

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

References

1. Wu, H.; Zhong, B.; Li, H.; Love, P.; Pan, X.; Zhao, N. Combining Computer Vision with Semantic Reasoning for On-Site Safety Management in Construction. *J. Build. Eng.* **2021**, *42*, 103036. [[CrossRef](#)]
2. Hou, X.; Li, C.; Fang, Q. Computer Vision-Based Safety Risk Computing and Visualization on Construction Sites. *Autom. Constr.* **2023**, *156*, 105129. [[CrossRef](#)]
3. Shi, M.; Chen, C.; Xiao, B.; Seo, J. Vision-Based Detection Method for Construction Site Monitoring by Integrating Data Augmentation and Semisupervised Learning. *J. Constr. Eng. Manag.* **2024**, *150*, 04024027. [[CrossRef](#)]

4. Zhang, X.; Wogen, B.E.; Liu, X.; Iturburu, L.; Salmeron, M.; Dyke, S.J.; Poston, R.; Ramirez, J.A. Machine-Aided Bridge Deck Crack Condition State Assessment Using Artificial Intelligence. *Sensors* **2023**, *23*, 4192. [CrossRef] [PubMed]
5. Tang, W.; Jahanshahi, M.R. Active Perception Based on Deep Reinforcement Learning for Autonomous Robotic Damage Inspection. *Mach. Vis. Appl.* **2024**, *35*, 110. [CrossRef]
6. Yao, Z.; Jiang, S.; Wang, S.; Wang, J.; Liu, H.; Narazaki, Y.; Cui, J.; Spencer, B.F., Jr. Intelligent Crack Identification Method for High-Rise Buildings Aided by Synthetic Environments. *Struct. Des. Tall Spec. Build.* **2024**, *33*, e2117. [CrossRef]
7. Liu, Z.; Xue, J.; Wang, N.; Bai, W.; Mo, Y. Intelligent Damage Assessment for Post-Earthquake Buildings Using Computer Vision and Augmented Reality. *Sustainability* **2023**, *15*, 5591. [CrossRef]
8. Kustu, T.; Taskin, A. Deep Learning and Stereo Vision Based Detection of Post-Earthquake Fire Geolocation for Smart Cities within the Scope of Disaster Management: İstanbul Case. *Int. J. Disaster Risk Reduct.* **2023**, *96*, 103906. [CrossRef]
9. Cheng, M.-Y.; Sholeh, M.N.; Kwek, A. Computer Vision-Based Post-Earthquake Inspections for Building Safety Assessment. *J. Build. Eng.* **2024**, *94*, 109909. [CrossRef]
10. Cheng, C.; Kawaguchi, K. A Preliminary Study on the Response of Steel Structures Using Surveillance Camera Image with Vision-Based Method during the Great East Japan Earthquake. *Measurement* **2015**, *62*, 142–148. [CrossRef]
11. Hutchinson, T.C.; Kuester, F. Monitoring Global Earthquake-Induced Demands Using Vision-Based Sensors. *IEEE Trans. Instrum. Meas.* **2004**, *53*, 31–36. [CrossRef]
12. Wang, X.; Wittich, C.E.; Hutchinson, T.C.; Bock, Y.; Goldberg, D.; Lo, E.; Kuester, F. Methodology and Validation of UAV-Based Video Analysis Approach for Tracking Earthquake-Induced Building Displacements. *J. Comput. Civ. Eng.* **2020**, *34*, 04020045. [CrossRef]
13. Wang, X.; Lo, E.; De Vivo, L.; Hutchinson, T.C.; Kuester, F. Monitoring the Earthquake Response of Full-Scale Structures Using UAV Vision-Based Techniques. *Struct. Control Health Monit.* **2022**, *29*, e2862. [CrossRef]
14. Cao, P.; Ji, R.; Ma, Z.; Sorosh, S.; Lo, E.; Norton, T.; Driscoll, J.; Wang, X.; Hutchinson, T.; Pei, S. UAV-Based Video Analysis and Semantic Segmentation for SHM of Earthquake-Excited Structures. In Proceedings of the 18th World Conference of Earthquake Engineering, Milan, Italy, 30 June–5 July 2024.
15. Weng, Y.; Shan, J.; Lu, Z.; Lu, X.; Spencer, B.F., Jr. Homography-Based Structural Displacement Measurement for Large Structures Using Unmanned Aerial Vehicles. *Comput.-Aided Civ. Infrastruct. Eng.* **2021**, *36*, 1114–1128. [CrossRef]
16. Shan, J.; Huang, P.; Loong, C.N.; Liu, M. Rapid Full-Field Deformation Measurements of Tall Buildings Using UAV Videos and Deep Learning. *Eng. Struct.* **2024**, *305*, 117741. [CrossRef]
17. Perry, B.J.; Guo, Y. A Portable Three-Component Displacement Measurement Technique Using an Unmanned Aerial Vehicle (UAV) and Computer Vision: A Proof of Concept. *Measurement* **2021**, *176*, 109222. [CrossRef]
18. Ribeiro, D.; Santos, R.; Cabral, R.; Saramago, G.; Montenegro, P.; Carvalho, H.; Correia, J.; Calçada, R. Non-Contact Structural Displacement Measurement Using Unmanned Aerial Vehicles and Video-Based Systems. *Mech. Syst. Signal Process.* **2021**, *160*, 107869. [CrossRef]
19. Zhang, C.; Lu, Z.; Li, X.; Zhang, Y.; Guo, X. A Two-Stage Correction Method for UAV Movement-Induced Errors in Non-Target Computer Vision-Based Displacement Measurement. *Mech. Syst. Signal Process.* **2025**, *224*, 112131. [CrossRef]
20. Khuc, T.; Nguyen, T.A.; Dao, H.; Catbas, F.N. Swaying Displacement Measurement for Structural Monitoring Using Computer Vision and an Unmanned Aerial Vehicle. *Measurement* **2020**, *159*, 107769. [CrossRef]
21. Fukuda, Y.; Feng, M.Q.; Shinozuka, M. Cost-Effective Vision-Based System for Monitoring Dynamic Response of Civil Engineering Structures. *Struct. Control Health Monit.* **2010**, *17*, 918–936. [CrossRef]
22. Han, Y.; Wu, G.; Feng, D. Vision-Based Displacement Measurement Using an Unmanned Aerial Vehicle. *Struct. Control Health Monit.* **2022**, *29*, e3025. [CrossRef]
23. Van Den Einde, L.; Conte, J.P.; Restrepo, J.I.; Bustamante, R.; Halvorson, M.; Hutchinson, T.C.; Lai, C.-T.; Lotfizadeh, K.; Luco, J.E.; Morrison, M.L.; et al. NHERI@UC San Diego 6-DOF Large High-Performance Outdoor Shake Table Facility. *Front. Built Environ.* **2021**, *6*, 580333. [CrossRef]
24. Barbosa, A. NHERI Converging Design Project: Overview of 6-Story Shake Table Test Program. In Proceedings of the 2024 EERI Annual Meeting, Seattle, WA, USA, 9–12 April 2024.
25. McBain, M.; Pieroni, L.; Araujo, R.; Simpson, B.G.; Barbosa, A. Full-Scale Shake Table Testing of a Six-Story Mass Timber Building with Post-Tensioned Rocking Walls and Buckling-Restrained Boundary Elements. *J. Struct. Eng.* **2025**, to be submitted.
26. Barbosa, A.; Simpson, B.; van de Lindt, J.; Sinha, A.; Field, T.; McBain, M.; Uarac, P.; Kontra, S.; Mishra, P.; Gioiella, L.; et al. Shake Table Testing Program for Mass Timber and Hybrid Resilient Structures Datasets for the NHERI Converging Design Project. In Shake Table Testing Program of 6-Story Mass Timber and Hybrid Resilient Structures (NHERI Converging Design Project) 2025. DesignSafe-CI. Available online: <https://www.designsafe-ci.org/data/browser/public/designsafe.storage.published/PRJ-5736/#detail-86b00b74-105f-4c13-a63f-594b32c52444> (accessed on 9 January 2025).
27. Pei, S.; Ryan, K.L.; Berman, J.W.; van de Lindt, J.W.; Pryor, S.; Huang, D.; Wichman, S.; Busch, A.; Roser, W.; Wynn, S.L.; et al. Shake-Table Testing of a Full-Scale 10-Story Resilient Mass Timber Building. *J. Struct. Eng.* **2024**, *150*, 04024183. [CrossRef]

28. Sorosh, S.; Hutchinson, T.C.; Ryan, K.L.; Smith, K.W.; Kovac, A.; Zabet, S.; Pei, S. Experimental Characterization of a Full-Scale Stair System Detailed to Achieve Seismic Resiliency. *Earthq. Eng. Struct. Dyn.* 2025, *submitted*.
29. American Society of Civil Engineers. *Minimum Design Loads and Associated Criteria for Buildings and Other Structures*; American Society of Civil Engineers: Reston, VA, USA, 2021; ISBN 978-0-7844-1578-8.
30. Edwins, D.J. *Modal Testing: Theory, Practice and Application*; John Wiley & Sons: Hoboken, NJ, USA, 2009.
31. DJI. Available online: <https://www.dji.com> (accessed on 19 September 2024).
32. Lucas, B.D.; Kanade, T. An Iterative Image Registration Technique with an Application to Stereo Vision. In Proceedings of the IJCAI'81: 7th International Joint Conference on Artificial Intelligence, Vancouver, BC, Canada, 24–28 August 1981; Volume 2, pp. 674–679.
33. Shi, J.; Tomasi, C. Good Features to Track. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 21–23 June 1994; pp. 593–600. [[CrossRef](#)]
34. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *Automatica* **1975**, *11*, 23–27. [[CrossRef](#)]
35. Kanopoulos, N.; Vasanthavada, N.; Baker, R.L. Design of an Image Edge Detection Filter Using the Sobel Operator. *IEEE J. Solid-State Circuits* **1988**, *23*, 358–367. [[CrossRef](#)]
36. Canny, J. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *PAMI-8*, 679–698. [[CrossRef](#)]
37. Trujillo-Pino, A.; Krissian, K.; Alemán-Flores, M.; Santana-Cedr s, D. Accurate Subpixel Edge Location Based on Partial Area Effect. *Image Vis. Comput.* **2013**, *31*, 72–90. [[CrossRef](#)]
38. Harris, C.; Stephens, M. A Combined Corner and Edge Detector. In Proceedings of the Alvey Vision Conference 1988, Manchester, UK, 31 August – 2 September 1988; Alvey Vision Club: Manchester, UK, 1988; pp. 147–151.
39. Zhang, Z.; Lu, H.; Li, X.; Li, W.; Yuan, W. Application of Improved Harris Algorithm in Sub-Pixel Feature Point Extraction. *Int. J. Comput. Electr. Eng.* **2014**, *6*, 101–104. [[CrossRef](#)]
40. Chen, X.; Lu, L.; Gao, Y. A New Concentric Circle Detection Method Based on Hough Transform. In Proceedings of the 2012 7th International Conference on Computer Science & Education (ICCSE), Melbourne, VIC, Australia, 14–17 July 2012; pp. 753–758.
41. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*, 2nd ed.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2003; ISBN 978-0-521-54051-3.
42. Bouguet, J.Y. Camera Calibration Toolbox for MATLAB. 2015. Available online: <https://data.caltech.edu/records/jx9cx-fdh55> (accessed on 19 September 2024).
43. Agisoft Metashape User Manual—Professional Edition. Version 2.1. 2024. Available online: <https://www.agisoft.com/downloads/user-manuals/> (accessed on 19 September 2024).
44. Fraser, C.S. Digital Camera Self-Calibration. *ISPRS J. Photogramm. Remote Sens.* **1997**, *52*, 149–159. [[CrossRef](#)]
45. Westoby, M.J.; Brasington, J.; Glasser, N.F.; Hambrey, M.J.; Reynolds, J.M. ‘Structure-from-Motion’ Photogrammetry: A Low-Cost, Effective Tool for Geoscience Applications. *Geomorphology* **2012**, *179*, 300–314. [[CrossRef](#)]
46. Zheng, Y.; Sugimoto, S.; Okutomi, M. ASPnP: An Accurate and Scalable Solution to the Perspective- n -Point Problem. *IEICE Trans. Inf. Syst.* **2013**, *E96.D*, 1525–1535. [[CrossRef](#)]
47. Colomina, I.; Molina, P. Unmanned Aerial Systems for Photogrammetry and Remote Sensing: A Review. *ISPRS J. Photogramm. Remote Sens.* **2014**, *92*, 79–97. [[CrossRef](#)]
48. De Fino, M.; Galantucci, R.A.; Fatiguso, F. Condition Assessment of Heritage Buildings via Photogrammetry: A Scoping Review from the Perspective of Decision Makers. *Heritage* **2023**, *6*, 7031–7066. [[CrossRef](#)]
49. Hesch, J.A.; Roumeliotis, S.I. A Direct Least-Squares (DLS) Method for PnP. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 383–390.
50. Lepetit, V.; Moreno-Noguer, F.; Fua, P. EPnP: An Accurate $O(n)$ Solution to the PnP Problem. *Int. J. Comput. Vis.* **2009**, *81*, 155–166. [[CrossRef](#)]
51. Schonberger, J.L.; Frahm, J.-M. Structure-From-Motion Revisited. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4104–4113.
52. Zach, C. Robust Bundle Adjustment Revisited. In Proceedings of the 13th European Conference on Computer Vision—ECCV 2014, Zurich, Switzerland, 6–12 September 2014; pp. 772–787.
53. Sutton, M.A.; Yan, J.H.; Tiwari, V.; Schreier, H.W.; Orteu, J.J. The Effect of Out-of-Plane Motion on 2D and 3D Digital Image Correlation Measurements. *Opt. Lasers Eng.* **2008**, *46*, 746–757. [[CrossRef](#)]
54. MathWorks. *Computer Vision Toolbox (R2023b)*; MathWorks Inc.: Natick, MA, USA, 2023. Available online: <https://www.mathworks.com/products/computer-vision.html> (accessed on 19 September 2024).

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.