Received: 5 October 2020 | Revised: 18 August 2021 | Accepted: 14 September 2021

DOI: 10.1002/stc.2862

WILEY

RESEARCH ARTICLE

Monitoring the earthquake response of full-scale structures using UAV vision-based techniques

Xiang Wang[1] | **Eric Lo**[2] | **Luca De Vivo**[3] | **Tara C. Hutchinson**[3] | **Falko Kuester**[3]

[1]School of Civil Engineering, Sun Yat-Sen University, Guangzhou, China

[2]Qualcomm Institute, University of California, San Diego, La Jolla, California, USA

[3]Department of Structural Engineering, University of California, San Diego, La Jolla, California, USA

Correspondence
Tara C. Hutchinson, Department of Structural Engineering, University of California, San Diego, La Jolla, CA, USA.
Email: tara@ucsd.edu

Funding information
Division of Civil, Mechanical and Manufacturing Innovation, Grant/Award Numbers: 1663348, 1663569; National Institute of Standards and Technology, Grant/Award Number: 70NANB17H211

Abstract

Vision-based sensing, when utilized in conjunction with camera-equipped unmanned aerial vehicles (UAVs), has recently emerged as an effective sensing technique in a variety of civil engineering applications (e.g., construction monitoring, conditional assessment, and post-disaster reconnaissance). However, the use of these non-intrusive sensing techniques for extracting the dynamic response of structures has been restricted due to the perspective and scale distortions or image misalignments caused by the movement of the UAV and its on-board camera during flight operations. To overcome these limitations, a vision-based analysis methodology is proposed in the present study for extracting the dynamic response of structures using unmanned aerial vehicle (UAV) aerial videos. Importantly, geo-referenced targets were strategically placed on the structures and the background (stationary) region to enhance the robustness and accuracy related to image feature detection. Image processing and photogrammetric techniques are adopted in the analysis procedures first to recover the camera motion using the world-to-image correspondences of the background (stationary) targets and subsequently to extract the dynamic structural response by reprojecting the image feature of the (moving) targets attached to the structures to the world coordinates. The displacement tracking results are validated using the responses of two full-scale test structures measured by analog displacement sensors during a sequence of shake table tests. The high level of precision (less than 3 mm root-mean-square errors) of the vision-based structural displacement results demonstrates the effectiveness of the proposed UAV displacement tracking methodology. Additionally, the limitations and potential solutions associated with the proposed methodology for monitoring the dynamic responses of real structures are discussed.

KEYWORDS

earthquakes, motion tracking, photogrammetry, unmanned aerial vehicles, vision-based sensing

Formerly with Department of Structural Engineering, University of California, San Diego.

1 | INTRODUCTION

Unmanned aerial vehicle (UAV) vision-based sensing has attracted significant research attention among civil engineers in the recent decade due to their cost-effectiveness and convenience for imagery data collection. These camera-equipped aerial platforms are particularly useful when access to camera views is inconvenient or even unsafe for ground-based cameras (e.g., building roof and underside of bridge decks). Recent applications of UAV vision-based techniques involve construction monitoring and management,[1-3] operational condition inspection of civil structures,[4-8] as well as post-disaster assessment of structural damage and geo-hazards.[9-11] Nevertheless, the application of UAV vision for quantitatively tracking the subtle (centimeter-level) variations of dynamic structural responses has been limited, although vision-based structural monitoring using ground-based cameras has been a well-studied topic.[12-19]

Unlike ground-based cameras that remain essentially stationary, the UAV-mounted cameras are subjected to the movement of the UAV (colloquially referred to as drift) as well as the camera relative to the UAV during flight operations. Importantly, the mobility of these sensing platforms introduces image distortion to individual video frames due to the following two sources: (1) UAV drift-induced image misalignment and scale distortions and (2) perspective distortion caused by the non-orthogonal camera orientation (e.g., parallel lines appear to converge). Correcting these image distortion effects requires accurate information regarding the UAV position and camera motion. However, the lack of such information in the majority of the commercial UAV platforms poses major challenges for accurately extracting structural dynamic responses from the aerial imagery. To this end, a handful of recent research efforts have attempted to circumvent the need for resolving the UAV-mounted camera pose and instead employed image processing techniques (e.g., image feature matching, optical flow, and homography transformation) to compensate for the image distortion induced by camera movements.[20-22] Nevertheless, the dynamic responses of the structures extracted using such techniques are contaminated with low-frequency noise caused by the camera motion and need to be further processed via high-pass filtering. In a recent study, Yoon et al[23] resolved the on-board camera motion using the multi-view stereo triangulation of the static image features (on a background reference target). The effectiveness of the displacement tracking performance of their method was validated using lab-scale prototype structure experiments. However, the conditions used for recovering the camera motion (e.g., controlled indoor laboratory environment and close camera-to-scene distance) may not be feasible for monitoring structures in field applications. Therefore, there is a strong need to develop UAV vision-based displacement tracking methodologies for capturing the dynamic responses of large-scale or field structures and to evaluate the benefits and limitations of such methodologies.

In recognition of the restrictions associated with the camera motion-induced perspective and scale distortions or image misalignments as reported in previous studies,[20,21] a UAV vision-based analysis methodology is proposed herein first to recover the camera motion and subsequently to extract the structural displacements under dynamic loading scenarios (e.g., earthquakes). Importantly, geo-referenced checkerboard targets (hereafter referred to as reference targets) were strategically placed on the monitored structures and the background (stationary) region to establish robust feature correspondences between world and image coordinates. The proposed displacement tracking methodology is implemented as a frame-by-frame video analysis procedure in three sequential steps: (1) detecting the image points of the stationary and moving (non-stationary) reference targets; (2) recovering the camera pose using the world-to-image correspondences of the stationary targets; and (3) reconstructing the world points of the moving targets from their detected image points.

The effectiveness of the proposed displacement tracking procedure is assessed using experimental data collected from cold-formed steel (CFS) shear wall specimens (emulating single-story structures) during a shake table test program. These test specimens were subjected to a sequence of earthquake input motions using the Large High Performance Outdoor Shake Table (LHPOST) facility at the University of California, San Diego (UCSD).[24] Facilitated by an outdoor test environment, two unmanned aerial vehicles (UAVs) were deployed to record aerial videos of the test structures during the simulated earthquake tests from two strategically defined camera views. In addition, point cloud data of the test scene was collected using a Light and Ranging Detection (LiDAR) scanner to provide accurate geo-information for characterizing the test structures and the stationary targets in the background region. The structural responses measured by analog displacement sensors during these shake table tests provide a unique dataset for comparison with UAV video analysis results obtained using the proposed displacement tracking methodology. Validation of the UAV video-based results against the analog displacement sensor measurements demonstrates the effectiveness of the proposed analysis methodology for successfully quantifying the dynamic responses of large structures. With the availability of camera pose information estimated using the photogrammetric methods, the video analysis results achieve a sub-pixel level precision (<3 mm), which is substantially smaller than the centimeter-level accuracy as

reported in a prior UAV monitoring study conducted by the authors.[21] This indicates that the camera motion resolved based on the image-to-world correspondence of the reference targets effectively enhances the accuracy of structural dynamic response monitoring results. Additionally, the limitations and potential solutions associated with the application of the proposed technique in monitoring real structures, including the restriction to two-dimensional displacement tracking due to the use of monocular camera systems as well as the ground movement in real earthquake events, are discussed.

2 | SHAKE TABLE TEST PROGRAM

In an effort to advance understanding of the seismic behavior of CFS shear walls, four shake table test series were sequentially conducted at the NHERI@UCSD outdoor shake table test facility in Fall 2018.[24] In each test series, two CFS shear wall specimens (north and south specimens) were installed on the 12.2 m × 7.6 m shake table and simultaneously tested using a sequence of earthquake motions with progressively increased motion intensities. Each test specimen was constructed using a pair of nominally identical CFS shear walls (4.8 m long and 2.7 m tall) placed along the east-west direction (longitudinal direction of the shake table platen) and a concrete mass (5.0 m × 3.0 m × 0.25 m) atop the shear wall pair (see Figure 1a). The total height of each test specimen was ~3.45 m above the shake table platen. All earthquake input motions were applied in the east-west direction using the single-axis shake table, which coincided with the longitudinal axis of the wall specimens. Consequently, the seismic responses of the test specimens in the transverse and vertical directions were significantly smaller than their counterparts in the longitudinal (shaking) direction.

The test specimens were densely instrumented with an array of ~120 analog sensors in each test series, including accelerometers and string potentiometers, for measuring their dynamic responses during the shake table tests. All analog sensors were connected to a multi-node data acquisition system that sampled data at a rate of 256 Hz. Importantly, the longitudinal displacements of the two test specimens and the shake table platen were measured with a total of five string potentiometers, namely, a pair of string potentiometers attached to the top of each specimen and one to the shake table platen. These displacement measurements are considered as ground truth in the displacement tracking result validations. Although the nominal (quantization) error of the string potentiometer measurements was sufficiently small (<0.05 mm), it is important to note that their actual measurement errors may reach several millimeters due to imperfect sensor installation and functionality conditions (e.g., string alignment errors and dynamic string vibration) or accidental torsional and transverse response of the specimens during the earthquake loading.

In the present study, the displacement tracking analysis focuses on Test Series #3 given the following two considerations: (1) completeness of the video dataset in terms of camera viewpoints and motion intensities and (2) robustness of ground truth displacement measurements (baseline fluctuation of analog displacement sensors occurred in several

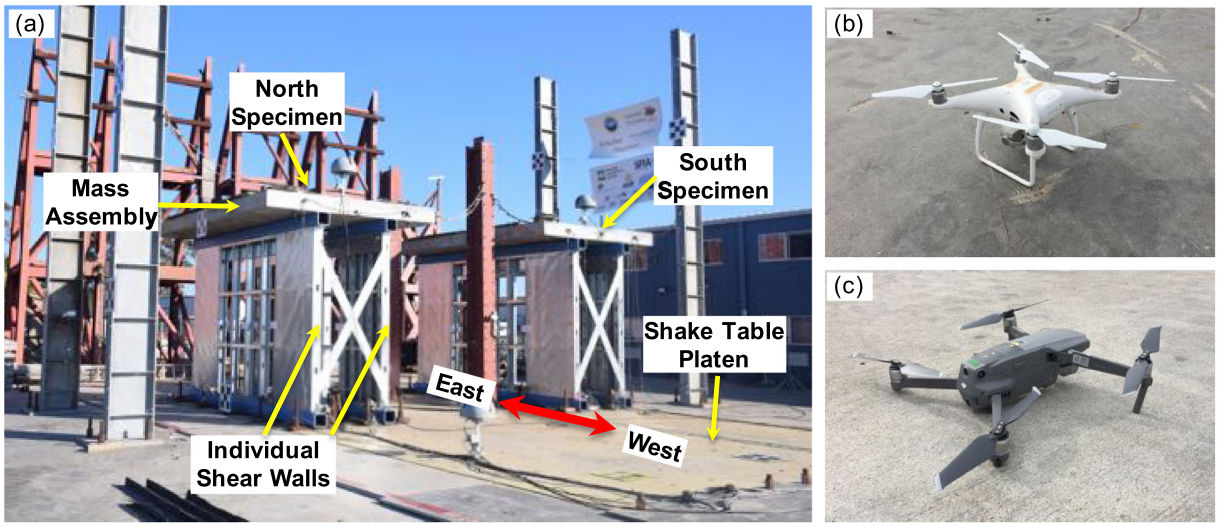

FIGURE 1 Shake table test program and UAV monitoring system: (a) isometric view of the test specimens (arrow denotes the direction of earthquake shaking), (b) Phantom 4 Pro, and (c) Mavic 2 Pro

other test series). Test Series #3 consisted of five earthquake tests with gradually increased motion intensities. The achieved peak responses of the table platen and the test specimens are summarized in Table 1. It is noted that Δ_{abs} represented the absolute displacement achieved at the top of the specimen (the mass center), whereas Δ_{rel} the displacement of the specimen relative to the table platen. Importantly, it is the relative (differential) displacements rather than the absolute displacements that cause damage to structures during earthquakes. Except for Test EQ2 (UAV videos were not recorded during this test), the remaining four tests employed the same earthquake record as the seed motion amplitude-scaled to achieve four distinct earthquake intensities (refer to Table 1 for details regarding the seed motions). As indicated by the table, the amplitude of the achieved peak acceleration of the input motion of Test EQ5 (nominal scale factor of 150%) was approximately six times that of Test EQ1 (nominal scale factor of 25%). Accordingly, the achieved peak absolute displacements of the specimens ranged from several centimeters in the low-intensity tests (EQ1–EQ3) to ~30 cm during the final test (EQ5). Detailed information regarding the shake table test program and test results are available in Singh et al.[24]

3 | MONITORING PLAN AND TARGET DETECTION

Two commercial UAV platforms, namely, DJI Phantom 4 Pro and Mavic 2 Pro models (Figure 1b,c), were consistently deployed to video-record the dynamic responses of the test specimens during the shake table test program. Each UAV platform was equipped with a built-in camera capable of high-resolution video recording (DJI, 2019). Two strategic camera views were consistently employed in the video recordings: (1) a top (overhead) view taken by the Mavic 2 Pro to monitor both the north and south specimens (Figure 2a) and (2) an elevation view adopted by the Phantom 4 Pro to monitor the north specimen (Figure 2b). All the test videos were recorded at a rate of 30 frames per second. Nevertheless, the position information of the UAVs and their on-board cameras during flight operations were not precisely known since these platforms were only equipped with low-resolution position sensors to support flight control and obstacle avoidance. In this regard, the movements of the on-board cameras during video recording are estimated using the Perspective-n-Point (PnP) techniques.[25,26] Detailed discussions of this and other photogrammetric techniques used in the present study are elaborated later in Section 4.

3.1 | Reference targets and camera views

In the UAV vision-based monitoring plan, reference targets with checkerboard patterns were placed on the test specimens and shake table platen as well as at various locations of the stationary (background) region to provide reliable image features for UAV video analysis. Two types of reference targets, namely, the five-tile targets and the four-tile targets (Figure 3), were adopted for the monitoring plan. Although the two types of targets varied in checkerboard pattern and tile size, they offered identical dimensions (45.7 cm × 45.7 cm) and edge length (240 cm). To ensure the robustness of image feature detection, these targets were sized on the basis that the tile edges provide sufficient pixel length

TABLE 1 Achieved peak responses of the table platen and test specimens during Test Series #3 (all responses represent those in the longitudinal direction)

Test motion name (expected performance target)	Shake table platen		North specimen		South specimen	
	PIA (g)	*PID* (cm)	Δ_{abs} (cm)	Δ_{rel} (cm)	Δ_{abs} (cm)	Δ_{rel} (cm)
EQ1:CNP-25 (Elastic)	0.15	3.6	3.7	0.3	3.7	0.7
EQ2:CUR-25 (Elastic)	0.11	1.2	1.2	0.3	1.5	0.7
EQ3:CNP-50 (Quasi-Elastic)	0.29	7.2	7.3	0.5	7.6	1.4
EQ4:CNP-100 (Design)	0.63	16.5	16.8	2.4	17.6	3.2
EQ5:CNP-150 (Above-design)	0.80	25.1	30.6	27.7	27.3	14.0

Note: Nomenclature of test motion is defined as *Sequential EQ #: Seed Record Name–Scale Factor*; CNP = earthquake motion recorded at Canoga Park station during the 1994 Northridge earthquake; CUR = earthquake motion recorded at Curico station during the 2010 Maule earthquake in Chile; *PIA* = peak input acceleration of the earthquake motion; *PID* = peak input displacement of the earthquake motion; Δ_{abs} = peak absolute displacement; Δ_{rel} = peak relative (differential) displacement; UAV test videos were recorded during all earthquake tests except EQ2.

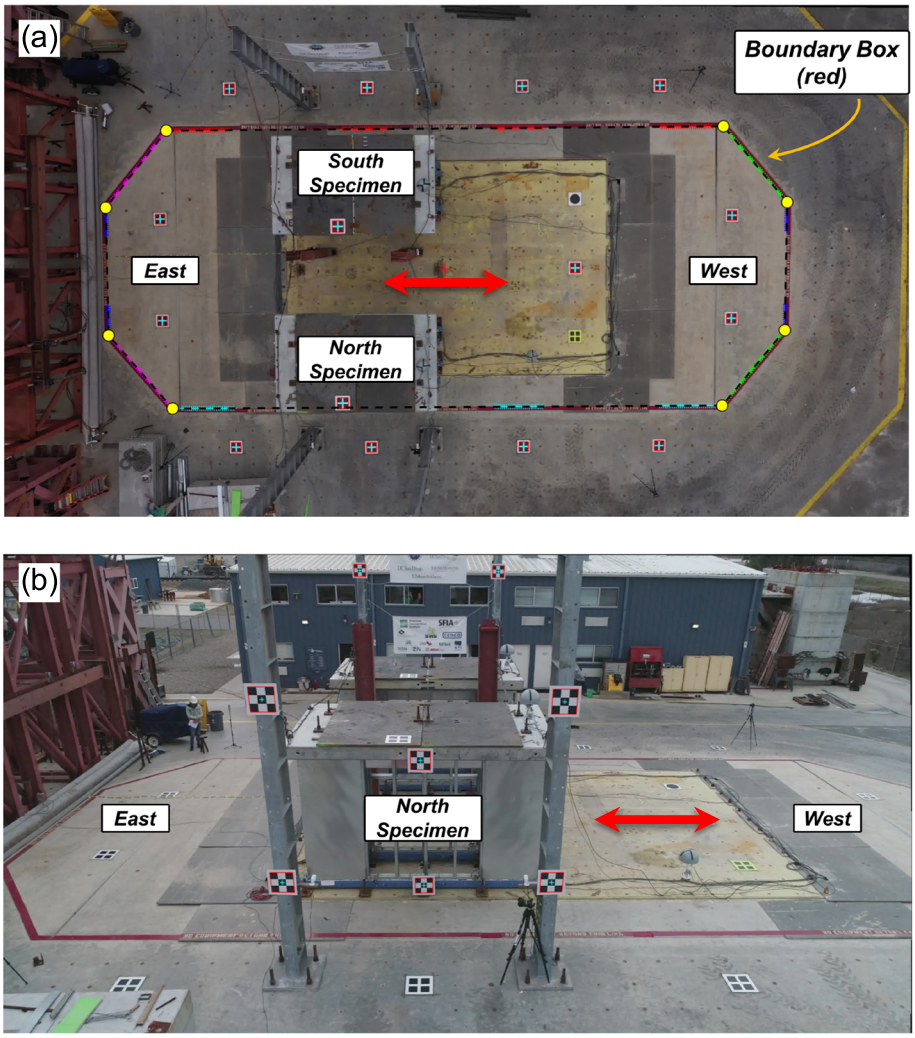

FIGURE 2 Camera views of the aerial videos: (a) top-view video frame (raw image resolution: 3840 × 1920 pixels), and (b) elevation-view video frame (raw image resolution: 4096 × 2160 pixels). Arrows denote the direction of earthquake shaking

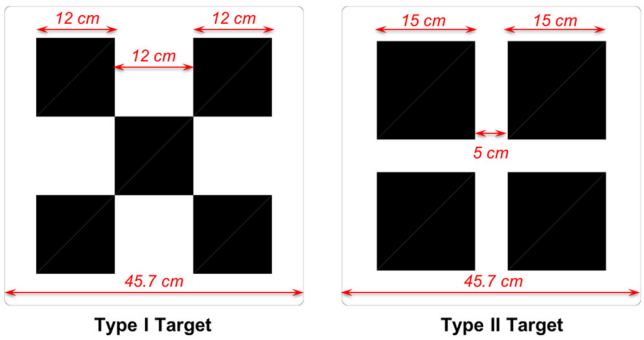

FIGURE 3 Reference target dimensions and checkerboard patterns

(at least 12 pixels) at a given camera-to-scene distance of 30 m (the expected maximum range for the on-board cameras during the flight operations). It is noted that the UAV camera viewpoints and target dimensions adopted in the monitoring plan was determined using a virtual point cloud model of the test scene as well as the intrinsic parameters of the UAV on-board cameras in the test preparation stage.

The top-view videos (Figure 2a) were recorded with the UAV positioned between 25 and 30 m above the center of the shake table platen during the video recording. This camera view involves a total of fifteen reference targets (Type II

targets), including twelve stationary targets around the shake table perimeter and three moving targets at the top of the two specimens and the shake table platen. For the elevation view (Figure 2b), the UAV was located 10 m north of the north specimen (the edge of the specimen was ~4 m north of the center of the table) at a much lower altitude (5–7 m) during the video recording. As a result of smaller camera-to-scene distance, only eight reference targets (Type I targets) were employed in this camera view, including six stationary targets attached to the steel posts and two moving targets at the top and bottom of the north specimen. Furthermore, a terrestrial laser scanner (Faro Focus 3D) was utilized to collect geo-referenced point cloud data of the wall specimens and the background region of the test scene (Figure 4). The point cloud model provided accurate (centimeter-level precision) geo-information for characterizing the locations of the reference targets and other natural features (e.g, shake table boundary lines as shown in Figure 2a). It is noted that the geometric center of the table platen is defined as the origin of the world coordinates throughout this study (see Figure 4).

3.2 | Target detection procedure and validation results

Since the proposed UAV vision-based displacement tracking procedure relies on world-to-image point correspondences of the reference targets for tracking the dynamic displacements of the test specimens, an image feature analysis procedure is implemented in MATLAB[27] for extracting the target locations in the video frames (or referred to as target image points). Detection of target image points is performed successively on pre-defined regions of interest (ROIs) located in the vicinity of individual targets using three sequential steps as follows:

Step 1: sub-pixel edge detection. Instead of the conventional edge detection techniques,[28] a sub-pixel edge detector[29] is adopted for detecting the edge features of the reference targets. The local intensity gradients at the neighborhood of the edge pixels are considered in this sub-pixel edge detector for extracting the edge points and calculating their directional normals (orientations). Compared with conventional edge detectors, the sub-pixel edge detector effectively enhances the smoothness and continuity of the edge detection results particularly in the presence of blurred edges and corners (Figure 5a). Interested readers are referred to Trujillo-Pino et al[29] for detailed information on this sub-pixel edge detection method.

Step 2: edge point clustering. This step aims to classify the detected edge points into different clusters based on their orientations and locations. The edge points are first grouped into two orientation bins (red and green points as shown in Figure 5b) using a heuristic histogram-based clustering strategy. A greedy search is performed over all orientations to identify two orientation bins (with a fixed bin width of 10 degrees) that contain the largest number of edge points. The K-means clustering algorithm[30] is subsequently used to classify the points in each orientation

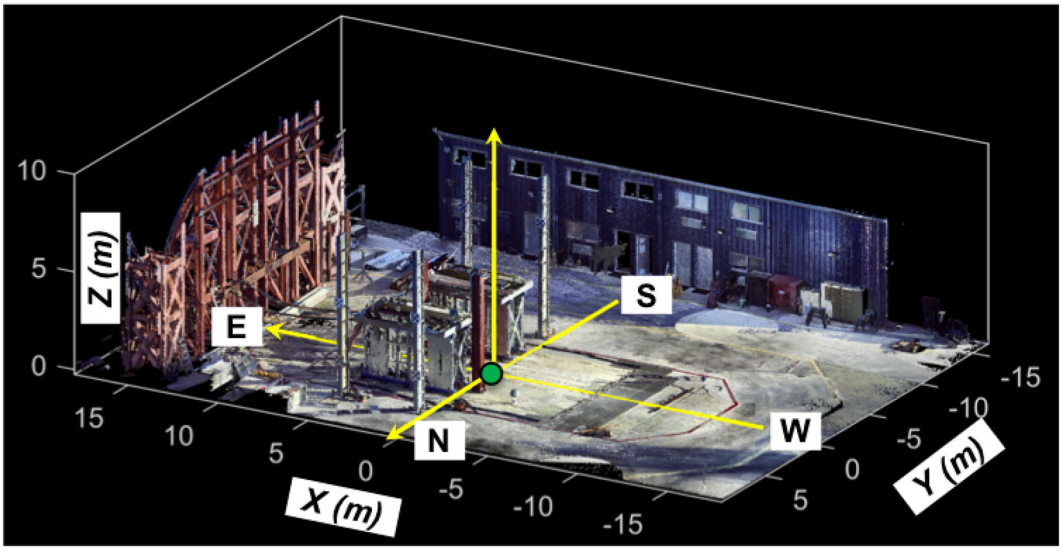

FIGURE 4 Point cloud model of the test scene. Note that the green dot denotes the origin of the world coordinates (center of the shake table platen) and the yellow lines represent the coordinate axes (arrows denote the positive directions)

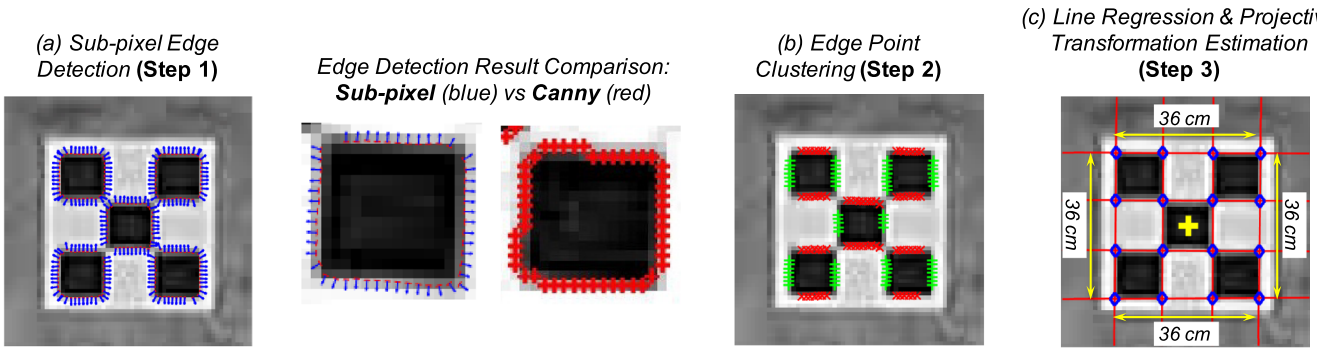

FIGURE 5 Illustration of the target image point detection steps (ROI dimensions: 70 × 70 pixels)

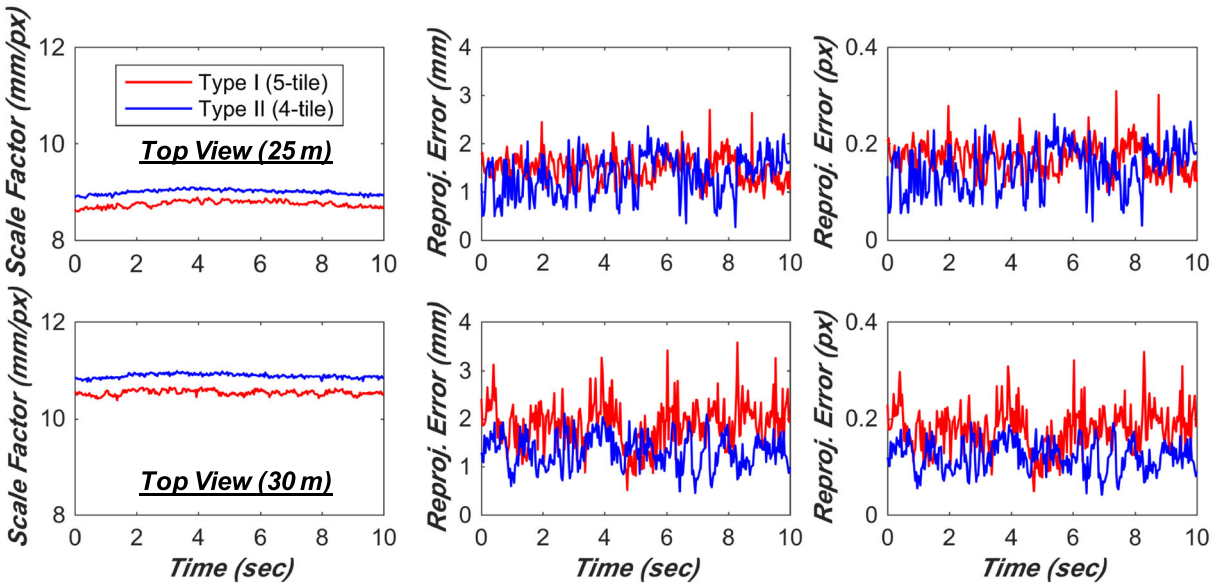

FIGURE 6 Scale factors and reprojection errors of the stationary reference targets

bin into four clusters based on their pixel coordinates (each cluster represents an edge line of the checkerboard target). It is noted that several points around the corners are removed following the orientation clustering as their directional normals are inconsistent with those of the adjoining edge points. This effect is considered reasonable since the pixels around the checkerboard corners typically contain a higher level of image noise.

Step 3: projective transformation and image point extraction. The clustered edge points (red and green points as shown in Figure 5b) allow for linear regression of eight edge lines in the two orthogonal directions (red lines as shown in Figure 5c). The fitted edge lines are then used to determine the pixel coordinates of the line intersections (a total of 16 blue dots as shown in Figure 5c). Since these intersections represent a set of co-planar points on the target, the correspondences between the pixel and metric coordinates of these points are employed for estimating a projective transformation between the two coordinate systems.[31] Consequently, the target image point is determined by projecting the target center point from the metric coordinates to the image coordinates (yellow cross as shown in Figure 5c). Importantly, this transformation defines a scale factor between and pixel coordinates and metric coordinates of an individual target, which represents the pixel resolution of the target. As shown in Figure 5c, the target perimeter (connecting the four bounding corners) had a length of 1440 mm and about 168 pixels in the image, resulting in a pixel resolution of 8.6 mm/pixel for this target. Note that the pixel resolution of a reference target may vary depending on its location relative to the UAV on-board camera as well as low-level (mostly <5%) fluctuations due to the camera motion during video recordings (refer to Figure 6). Additionally, the target reprojection error (defined as the mean error distance between the detected image points of the intersections and their image projections from the metric coordinates) provides a useful metric for evaluating the precision of the target detection results.

To evaluate the effectiveness of the target detection procedure, a UAV platform (Phantom 4 Pro) was deployed to different altitudes and monitored the reference targets placed on the bare table (one for each target as shown in Figure 3). These aerial videos were taken at the test preparation stage while the test scene (including the targets) was in a stationary condition. Figure 6 compares the scale factors and reprojection errors of the reference targets extracted from the videos recorded at two different altitudes, namely, 25 and 30 m above the table platen (representing the lower- and upper-bound position of the top-view camera). The results shown in the figure contain a duration of 10 seconds (or 300 video frames) for both videos. It is noted that the plots in the top and bottom rows correspond to the target detection results obtained from the videos taken at an altitude of 25 and 30 m, respectively. The scale factor results demonstrate that increasing the camera-to-scene distance (from 25 to 30 m in altitude) slightly reduces the pixel resolution (from 8.5–9 mm/pixel for the 25-m results to 10.5–11 mm/pixel for the 30-m results). Additionally, the low-level (<5%) fluctuation of the scale factors is indicative of the presence of camera motion during the video recording. The reprojection errors of the two types of targets are comparable (<2 mm) for the 25-m results, whereas the errors of Type I target (red) are slightly larger (reach as much as 3 mm) with increased camera-to-scene distance (30-m altitude). Nevertheless, these reprojection errors are all well below their corresponding pixel resolutions (between 0.1 and 0.3 pixel for videos taken at both altitudes). These observations demonstrate that the proposed target detection procedure is capable of achieving sub-pixel accuracy for detecting the target image points.

4 | UAV MOTION TRACKING ANALYSIS PROCEDURE

Figure 7 schematically outlines the flowchart of the proposed UAV vision-based displacement tracking framework. Four types of source data, namely, calibration videos, point cloud model of the test scene, UAV image keyframes (optional), and test videos, are required for tracking the displacement responses of the test specimens under dynamic loading scenarios. In addition to the target detection procedure as discussed in the earlier section, several photogrammetric techniques, namely, camera calibration, bundle adjustment, camera pose recovery, and world point reconstruction, are adopted in the displacement tracking procedure. Camera calibration and bundle adjustment are considered as

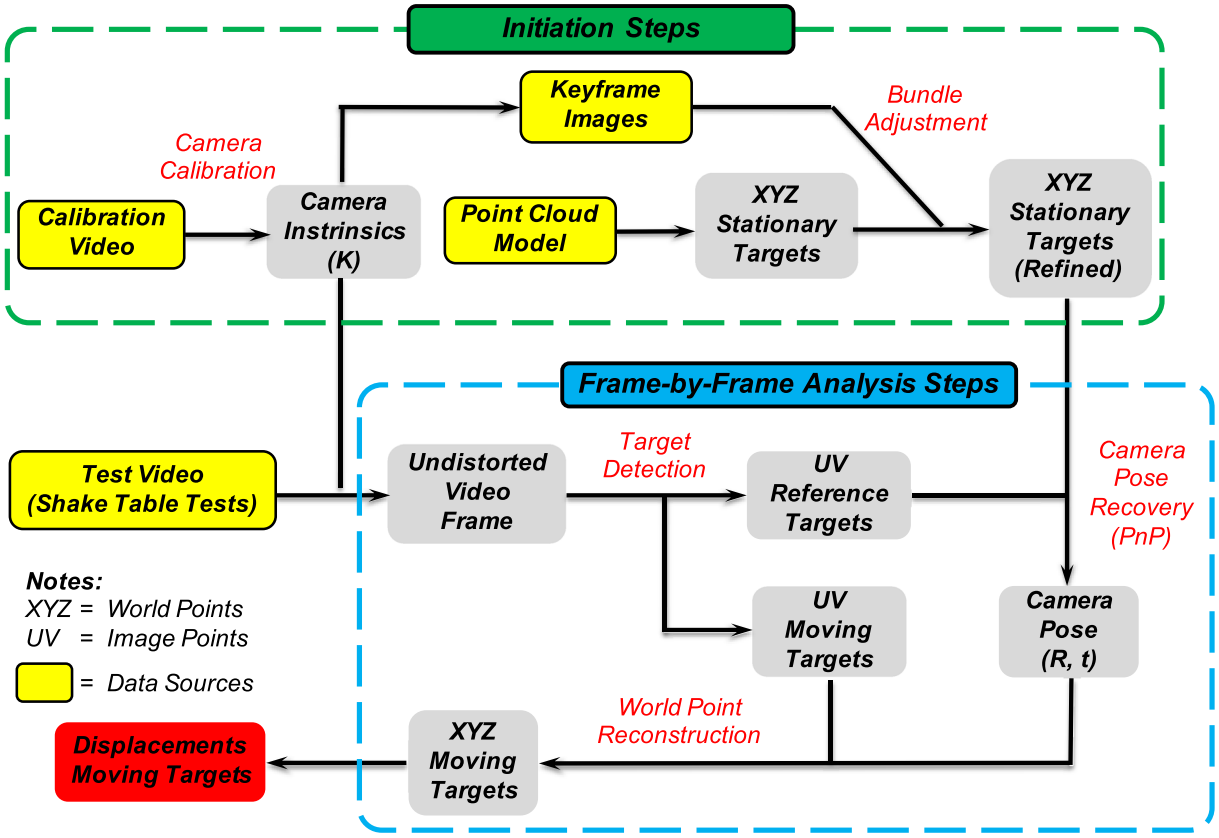

FIGURE 7 Flowchart of the proposed UAV vision-based structural displacement tracking framework

initiation steps in this analysis framework, since they provide the essential information (e.g., camera intrinsics and world points of the reference targets) for initiating the frame-by-frame video analysis. The frame-by-frame video analysis consists of three sequential steps: (1) detecting the image points of the reference targets (as discussed in the previous section), (2) recovering the camera pose using the world-to-image point correspondences of the stationary targets, and (3) reconstructing the world points of the moving (non-stationary) targets from their detected image points. This step-by-step video analysis procedure is implemented as an automatic off-line process in MATLAB.[27]

In the frame-by-frame video analysis, a reference frame is first selected from video footage to represent a stationary condition of the test scene (a few seconds before the onset of an earthquake test), whereas all remaining video frames are referred to as subsequent frames. It is noted that the ROIs associated with individual targets need to be specified only in the reference frame, whereas their locations in each subsequent frame relative to its previous frame are determined successively using a simple pixel-level correlation technique.[32] However, the ROI movements estimated using the image correlation technique may contain errors of several pixels, and therefore, this method is used only for frame-by-frame ROI location adjustment. In what follows, the implementation details of the photogrammetry techniques adopted in motion tracking analysis are discussed.

4.1 | Camera calibration

Camera calibration, or sometimes referred to as camera resectioning, is the process of estimating the lens distortion coefficients and camera projection parameters. The estimated coefficients are used to undistort raw images in order to remove the lens distortion effects (e.g., barrel distortion and tangential distortion). Following the lens distortion correction, the relation between a 3D world point $P_w = [x,y,z,1]^T$ (homogeneous coordinates) and its 2D image projection $P_c = [u,v,1]^T$ (homogeneous coordinates) follows the camera projection equation:[31]

$$sP_c = K[R|t]P_w \tag{1a}$$

or (in expanded form):

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & \gamma & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{1b}$$

where s is an arbitrary scale factor, K denotes the camera intrinsic parameters (in which f_x and f_y are the focal lengths associated with the two orthogonal directions in the image plane, γ is the image skew parameter, and u_0 and v_0 are the principal point coordinates of the image plane), R and t are the camera extrinsic parameters that characterize the rotational and translational transformation from the world coordinates to the camera coordinates. In this study, the lens distortion parameters and camera intrinsic parameters of the UAV on-board cameras are estimated using a well-documented checkerboard calibration procedure.[33] Interested readers are referred to this literature for the implementation details of the camera calibration procedures.

4.2 | Bundle adjustment

Bundle adjustment is an optimization method in computer vision that jointly refines the 3D geometry of a given structure, that is, a set of 3D world points, and the camera pose or camera motion related to multiple viewpoints, where each viewpoint is equivalent to a video frame.[34] It is assumed that the corresponding 2D image points of the world points in all viewpoints and the camera intrinsic parameters are considered as given. In its mathematical formulation, the world points and camera extrinsic parameters are optimized by minimizing the total reprojection error of the image points from all given viewpoints:

$$\min \sum_{i=1}^{n} \sum_{j=1}^{m} ||f(P_{ij}) - x_{ij}||^2 \tag{2}$$

where $f(P_{ij})$ represents the image projection of the ith world point on image j (calculated using Equation 1), x_{ij} represents the ith image point detected from image j, n is the total number of the world points, and m is the total number of frames.

In this study, bundle adjustment is employed to refine the world coordinates of the reference targets manually extracted from the LiDAR point cloud model as shown in Figure 4. The target coordinates associated with each aerial camera view (i.e., top and elevation) are optimized using approximately 20 keyframe images. Nevertheless, the refined target coordinates vary only slightly from their original coordinates (<2 cm error between the initial and refined coordinates) as a result of the accurate initial target locations extracted using the LiDAR data. Although bundle adjustment does not substantially enhance the motion tracking results due to the presence of accurate geo-referenced information of the test scene in the present study, it is considered an optional step in the proposed motion tracking framework, particularly when the estimated (initial) target locations involve larger uncertainties. If precise measurement data of the target locations are not available, it is recommended to increase the number of keyframes (viewpoints) in bundle adjustment in order to reduce the reprojection errors associated with the reference target coordinate refinement.

4.3 | Camera pose recovery

The six degree-of-freedom (DOF) camera pose is characterized using the 3×3 orthogonal rotation matrix R and 3×1 translation vector t to represent the camera orientation and position with respect to the world coordinates. In the present study, the camera pose associated with an individual video frame, or sometimes referred to as viewpoint, is resolved using the Perspective-n-Point (PnP) technique,[25,26] which estimates the pose of a calibrated camera with known intrinsic parameters given a set of n world points in 3D space and the corresponding 2D image points. Although PnP methods require a minimum of only three non-colinear world-to-image point correspondences for recovering the camera pose, increasing the number of point correspondences enhances the robustness and accuracy of the estimated camera pose due to the noisy measurements associated with the image and world points. In the displacement tracking procedure, the world-to-image correspondences are established based on the number of stationary targets placed in the background region, namely, 6 correspondences for the elevation view and 12 correspondences for the top view in consideration of the larger camera-to-scene distance (>25 m).

4.4 | World point reconstruction

Given the camera intrinsic and extrinsic parameters related to a specific video frame, the world point of a moving target is reconstructed using its image point identified from the target detection procedure. Since projecting a 2D image point to a 3D world point using a monocular camera is subjected to scale ambiguity, it is assumed that the z coordinates of all the moving targets, namely, the ones on the shake table platen and the top of the specimens, remained constant during the earthquake tests (zero vertical displacements). To reconstruct the world coordinates of the moving target, the vertical coordinate z in Equation 1 is replaced by a constant value $\tilde{z}$, and therefore, Equation 1 can be rewritten as

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = KR\begin{bmatrix} x \\ y \\ \tilde{z} \end{bmatrix} + K\begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \tag{3a}$$

or:

$$\begin{bmatrix} x \\ y \\ \tilde{z} \end{bmatrix} = sR^{-1}K^{-1}\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} - R^{-1}\begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \tag{3b}$$

According to Equation 3b, reconstructing x and y coordinates (of the world point) is performed in two steps: (a) solve for the unknown scale factor s using the third (last) row of Equation 3b (the scale factor is the only unknown variable), and (b) calculate x and y coordinates by substituting the scale factor s in Equation 3b with the value obtained from the

previous step. In the present study, the constant z coordinate assumption is considered as reasonable due to the following two reasons: (1) the earthquake input motions were applied only along the longitudinal direction of the specimens, and therefore, the specimen responses in the vertical (and transverse) direction were sufficiently small compared with those in the longitudinal (loading) direction, and (2) the large camera-to-scene distance (~10 m for the elevation view and >25 m for the top view) effectively reduces the photogrammetric errors in the presence of accidental out-of-plane displacements, since such errors are proportional to the ratio of the out-of-plane displacement over the camera-to-scene distance.[35]

5 | MOTION TRACKING RESULTS VALIDATION

The dynamic responses of the cold-formed steel wall specimens were recorded by two UAV platforms using the strategically defined camera views during four select earthquake tests of Test Series #3 (refer to Table 1). In this section, the test videos taken from the two pre-defined camera viewpoints are all analyzed using the proposed UAV displacement tracking procedure to extract the dynamic displacements of the test specimens. These video-based displacement results are subsequently validated against the ground truth measurements (measured using the string potentiometers) to assess the effectiveness of the proposed displacement tracking procedure and the error characteristics of the video analysis results. Additionally, the practical limitations and the potential enhancement related to the proposed monitoring technique are summarized and discussed in this section.

5.1 | Motion tracking analysis results

To articulate the video analysis steps as outlined in the previous section, detailed video analysis results of the final earthquake test (EQ5) are discussed herein. Specifically, the result discussions focus on the three sequential steps related to the frame-by-frame video analysis, namely, reference target detection, camera pose recovery, and structural displacement extraction (refer to Figure 7).

5.1.1 | Target detection

As the first step of the frame-by-frame analysis procedure, the image points of all the stationary and moving targets are identified using the target detection algorithm as discussed in Section 3. For brevity, the results of the moving targets associated with the final earthquake test (EQ5) are presented in Figure 8. For the top-view video, the pixel resolution of the target on the table platen is ~8.7 mm/pixel, whereas those attached on the specimens provide slightly higher

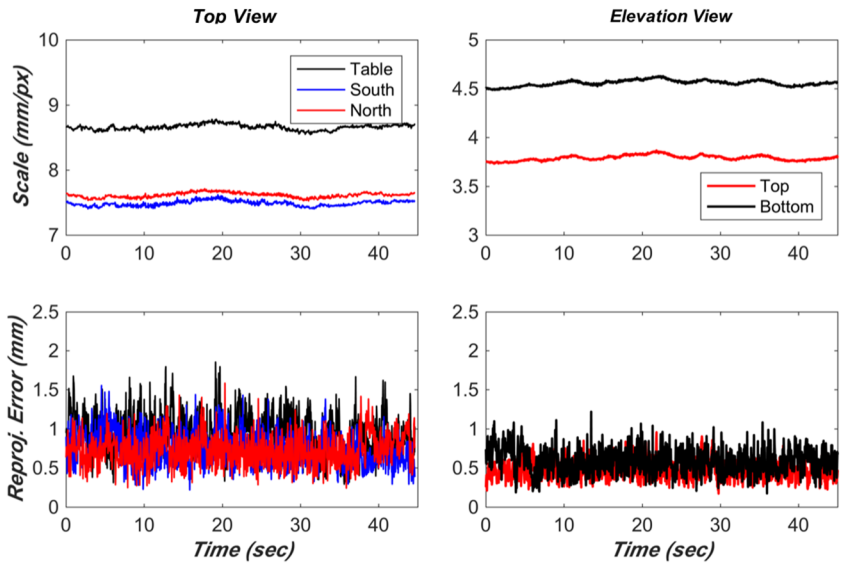

FIGURE 8 Scale factors and reprojection errors of the moving targets during the final earthquake test (EQ5)

resolutions (7.4–7.6 mm/pixel) as a result of slightly smaller distance between the targets and the on-board camera (the top of the wall specimens were ~3.45 m above the table platen). In contrast, the two targets in the elevation view (top and bottom of the north specimen) attain much higher pixel resolutions (~3.8 mm/pixel for the top target and ~4.5 mm/pixel for the bottom target), since the camera was located at a much closer position from the specimen (see detailed discussions in the next subsection). It is noted that the accuracy of the target detection results is corroborated by the fact that these target reprojection errors are substantially smaller than their corresponding pixel resolutions. For

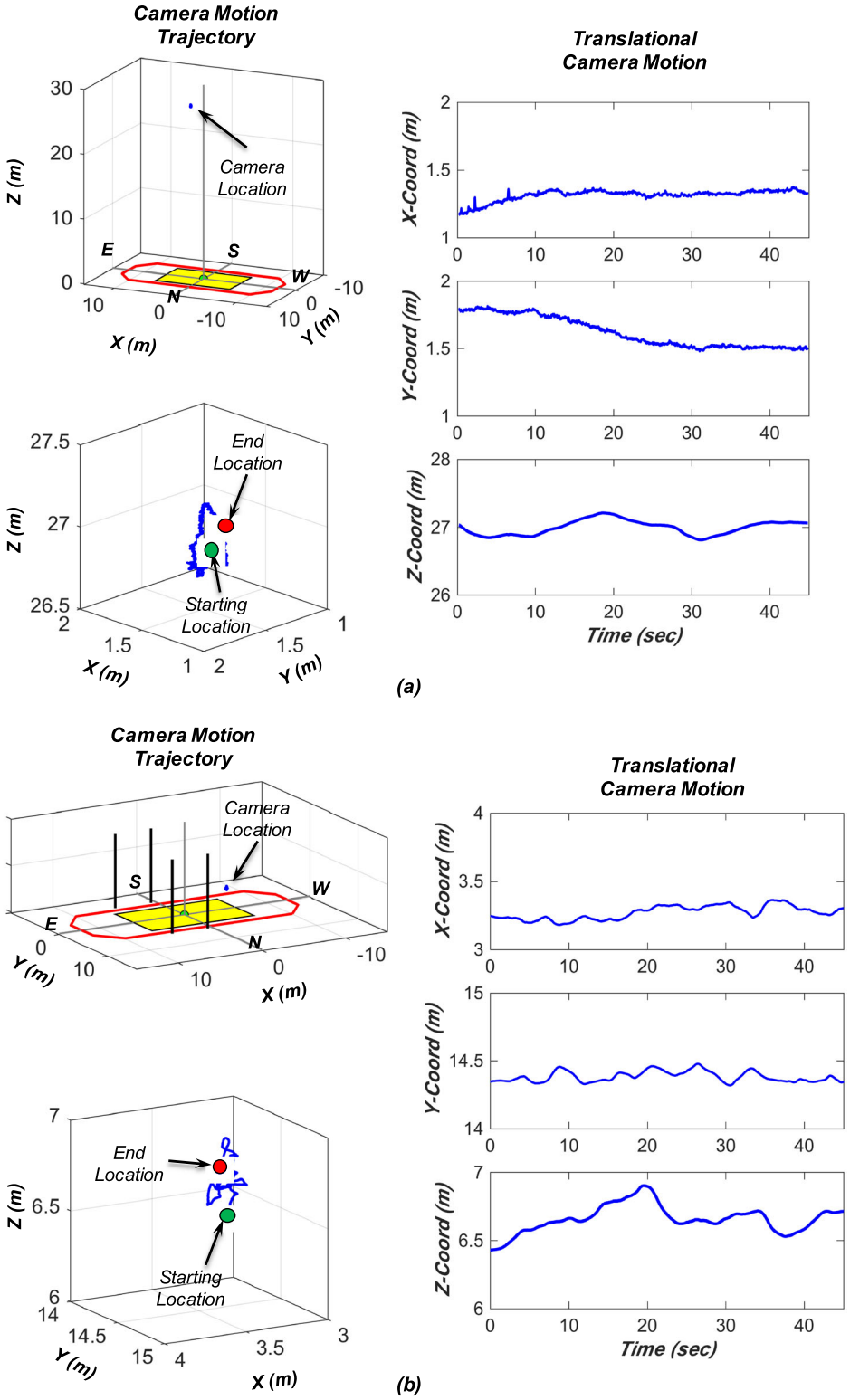

FIGURE 9 Estimated camera motion during the final earthquake test (EQ5): (a) top-view camera, and (b) elevation-view camera

instance, the reprojection errors of the target attached to the table platen (top view) are less than 2 mm for all the video frames, whereas its corresponding resolution is ∼8.7 mm/pixel.

5.1.2 | Camera pose recovery

As discussed in Section 4.3, the camera motion during the video recording is estimated based on the world-to-image point correspondences of the stationary targets. Figure 9 shows the motion trajectories of the two on-board cameras during the final earthquake test (EQ5). It is noted that the origin of the world coordinates is defined as the geometric center of the table platen (refer to Figure 4). According to the estimated motion trajectories, the top-view UAV was positioned at ∼27 m above the center of the shake table platen, whereas the elevation-view UAV hovered at a substantially lower altitude of 6.5–7 m above the ground and ∼10 m away from the north specimen. Additionally, the top-view UAV drifted as much as 0.5 m in all three translational directions during the video recording. The movements of the elevation-view UAV were even smaller (<0.3 m in the horizontal directions) possibly due to the lower operating altitude. Although these UAV motions may be considered sufficiently stable from the flight control perspective, they introduce frame-by-frame perspective and scale distortions to the aerial images, highlighting the need for camera pose recovery in the UAV vision-based displacement tracking.

5.1.3 | Structural displacement extraction

Given the pose information of the on-board camera and the image points of the moving targets, the x (longitudinal) and y (transverse) coordinates of the moving targets can be resolved using Equation 3 (see Section 4.4). In the present study, the z-coordinates of the moving targets are assumed to remain constant during individual earthquake tests, that is, zero vertical displacement of the specimens. Consequently, the absolute displacements of the top of the specimens and the shake table platen, represented by individual moving targets, are determined as the change in the target coordinates of the in the subsequent frames relative to those in the reference frame. It is noted that the reference frame corresponds to a stationary condition a few seconds prior the onset of the earthquake test. Furthermore, the displacements of the specimens relative to the table platen (sometimes referred to as the building drifts) can be derived from the absolute

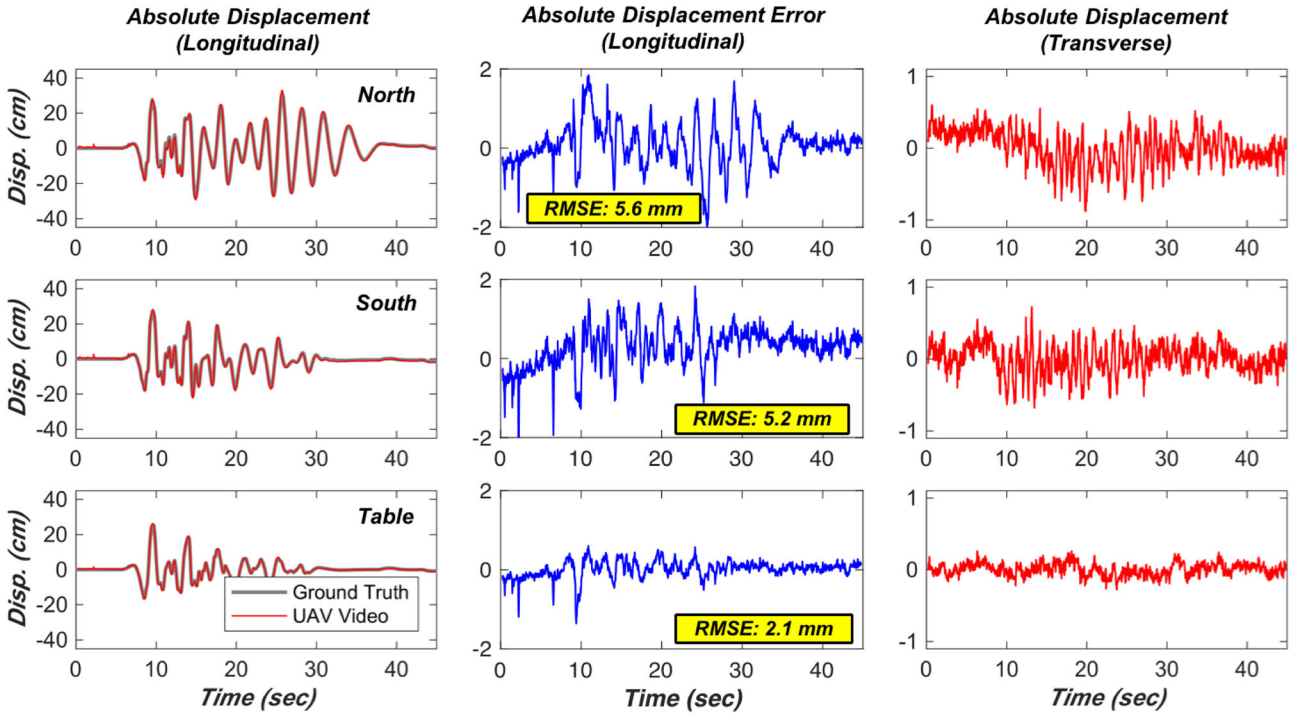

FIGURE 10 Top-view displacement results of the final earthquake test (EQ5): absolute displacements and associated error responses

displacements. As discussed in the earlier section, it is the relative (differential) displacements rather than the absolute displacements that induce damage to structural components during earthquakes or other dynamic loading scenarios.

Figure 10 shows the absolute displacements results of the specimens and the table platen from the top-view video analysis during the final earthquake test (EQ5), and the elevation-view displacement results are shown in Figure 11. These vision-based displacement results are compared with the ground truth data to evaluate their error characteristics. It is noted that the ground truth displacements considered in this study are those measured directly using contact-based analog string potentiometers. This comparison indicates that the vision-based displacements extracted from both the top-view and elevation-view videos are in reasonable agreement with the ground truth measurements, since the largest root-mean-square errors (RMSEs) remained <6 mm for the video analysis results. The errors associated with the top-view results are slightly larger than those of the elevation-view results due to the larger camera-to-scene distance. It is also observed that the errors of the specimen absolute displacements are consistently higher than those of the table. This is likely due to the accidental torsional behavior of the specimens during this high-intensity final earthquake test (see further discussions later in this section). Additionally, the transverse displacements are considered reasonable since they effectively confirm that the displacement responses of the specimens in this non-shaking direction were sufficiently small (<1 cm for the specimens and 1–2 mm for the table platen), however these low-amplitude responses were not measured by the string potentiometers during the earthquake tests. Although not shown herein for brevity, the vision-based relative displacements of the test specimens also agree well with the ground truth measurements. The video analysis errors are assessed comprehensively later in this section.

5.2 | Results validation and assessment of tracking errors

Video-based displacements of the test specimens during the four earthquake tests are compared with the ground truth measurements to assess the effectiveness of the proposed displacement tracking procedure. The two cameras on-board the UAVs maintained consistent viewpoints during all the earthquake tests, and therefore, the targets recorded by different test videos provide comparable pixel resolutions for the video analysis. Figure 12 compares the scale factors (pixel resolutions) of the moving targets attached to the table platen and the top of the specimens during the four earthquake tests, namely, EQ1 and EQ3–EQ5 (refer to Table 1). The peak input acceleration (PIA) of the earthquake motion is adopted as the the x-axis for the plots to demonstrate the gradually increased motion intensities during the earthquake tests. It is noted that while the field-of-view of the top-view videos included both the north and south specimens, the elevation-view videos focused only on the north specimen. Since the scale factors extracted from different video frames varied slightly as a result of UAV camera motion, each data point represents the scale factor of a specific target averaged among all video frames over the entire duration of the earthquake test. The comparisons indicate that the pixel resolutions of the targets remained stable during all the earthquake tests, as the observed discrepancies of the scale factors among different earthquake tests are within 5% for the top-view videos and ~10% for the elevation-view videos.

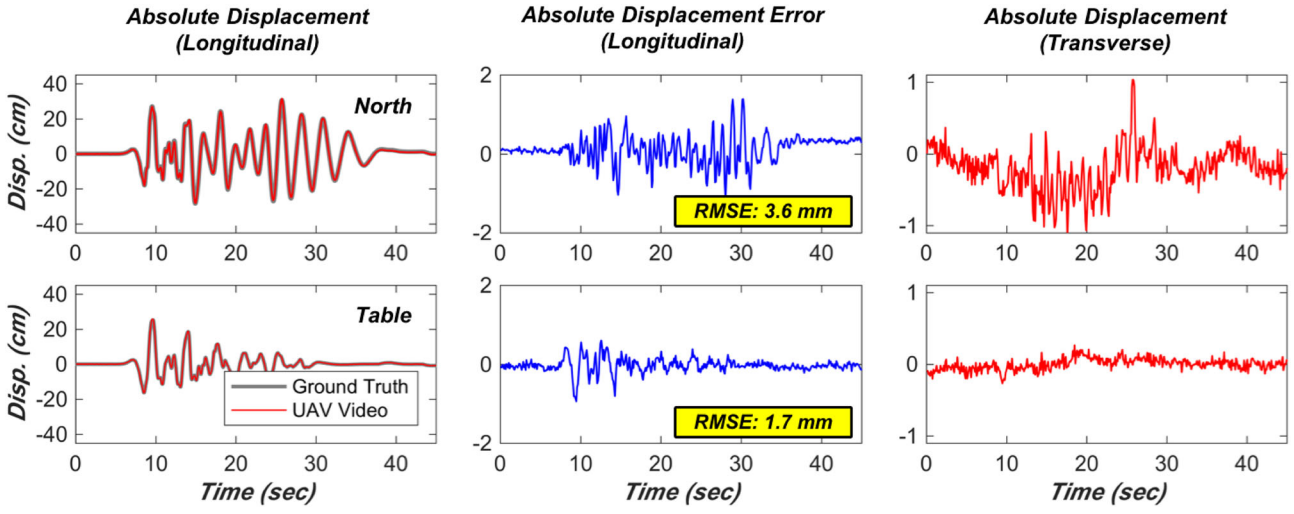

FIGURE 11 Elevation-view displacement results of the final earthquake test (EQ5): absolute displacements and associated error responses

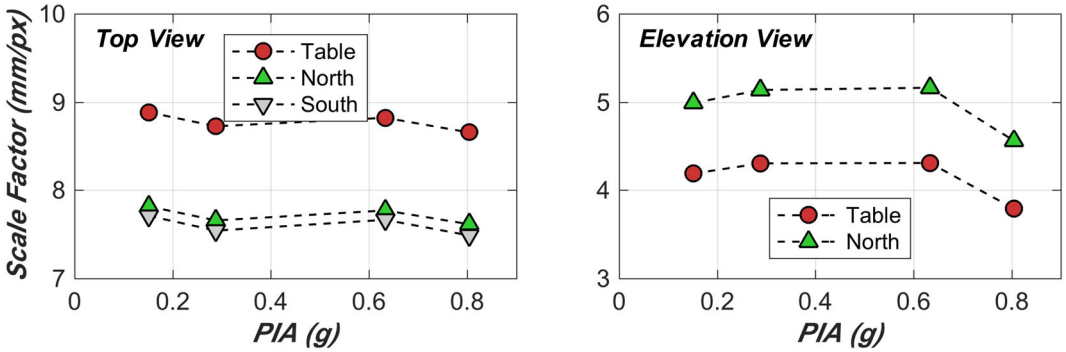

FIGURE 12 Pixel resolutions of the moving targets during the earthquake tests. PIA = peak input acceleration of the earthquake motion

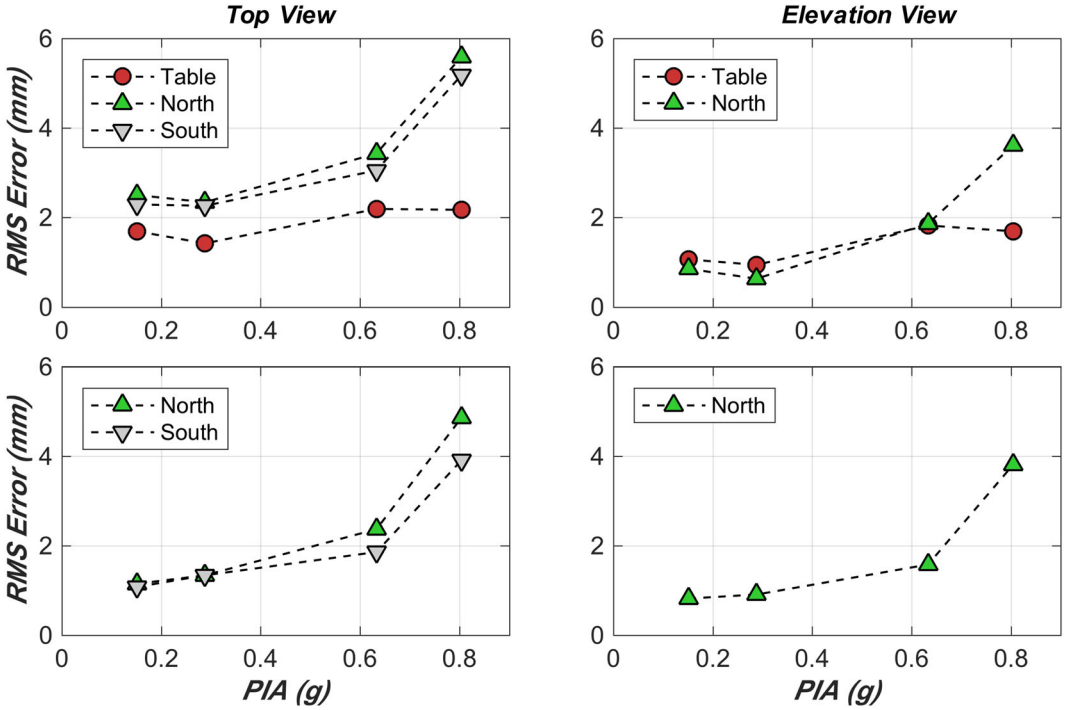

FIGURE 13 Displacement tracking errors of the table platen and the top of the specimens during the earthquake tests. RMS = root-mean-square; PIA = peak input acceleration of the earthquake motion

Additionally, the scale factors of the targets (representing the metric length per pixel at the target locations) extracted from the top view videos are roughly twice as large as those of the elevation view videos, which is attributed to the much larger camera-to-scene distance associated with the top-view camera viewpoints.

Figure 13 summarizes displacement tracking errors associated with the table platen and the top of the specimens during the four earthquake tests. The displacement tracking errors are defined as the root-mean-square (RMS) errors (or discrepancies) between the vision-based analysis results and string potentiometer measurements. It is noted that data points in the first row of the figure represent the errors associated with the absolute displacements of individual targets, whereas those in the second row represent the errors of the displacements at the top of the specimens relative to the table platen. As discussed in Section 2, the displacements measured by the string potentiometers may also contain errors of several millimeters due to the practical limitations associated with the sensor conditions (e.g., string alignment errors and dynamic string vibration) or accidental torsion of the specimens during the tests. The evolution of the displacement tracking errors indicates that the absolute displacement errors of the table platen are consistently small during all four earthquake tests (<2 mm RMSE) as evident in both the top-view and elevation-view results. The specimen displacement tracking errors are moderately (25%–50%) larger than those of the table during the first three

earthquake tests. However, these errors increase abruptly during the last earthquake test (EQ5), whereas the table displacement tracking error remains consistent with the previous tests. Therefore, it is reasonable to maintain that this inconsistent error increase (in the absence of substantial variations of the table displacement tracking errors) is likely attributed to the unmeasured accidental torsional responses at the top of the specimens due to the extremely large displacement demands (and severe structural damage) during the final earthquake test (refer to Table 1). Since the string potentiometers were attached to the middle of the concrete mass, they were not capable of capturing torsion-induced displacements at the edge of the specimens (the target locations). Excluding the inconsistently large errors of the specimen responses during the final earthquake test, the video analysis results achieve half-centimeter accuracy, namely, 3 mm RMSE for the top-view results and 2 mm RMSE for the elevation-view results (due to the smaller camera-to-scene distance).

The displacement tracking errors are further compared with the pixel resolutions of the corresponding targets (see Figure 12) to determine their pixel-level errors, that is, the RMSE in the unit of pixel. These pixel-level errors of the displacement tracking results are summarized in Figure 14. For instance, the (top view) table displacement RMSE is 2.2 mm during the final earthquake and the corresponding target pixel resolution is 8.6 mm, and therefore, the displacement tracking results achieve an accuracy of ∼1/4 pixel (sub-pixel accuracy). As evident in the results, the displacement tracking errors are consistently less than half a pixel for all four earthquake tests (except the specimen displacements at the final test), which further demonstrates the validity of the proposed UAV video analysis procedure in tracking dynamic structural displacements. Although the sub-pixel accuracy (∼3 mm) achieved by the video analysis results is sufficient for capturing the centimeter-level absolute displacements of the specimens, it introduces high noise-to-signal ratios for monitoring very small relative displacements (<5 mm) during low-intensity earthquake tests (EQ1 and EQ2 as shown in Table 1). In such scenarios, the pixel resolution of the video frames needs to be enhanced to reduce the noise-to-signal ratios.

5.3 | Discussion and limitations

Validation of the displacement tracking results using the full-scale shake table experiments demonstrates the effectiveness of the proposed UAV vision-based analysis procedure, notably achieving a sub-pixel level error of <3 mm. With the implementation of the camera motion recovery strategies, the displacement tracking errors in the present study are substantially smaller than the reported errors of 1-2 cm in a previous UAV monitoring study conducted by the authors.[21] This prior study focused on monitoring the roof displacement responses of a full-scale six-story building, with a height of 18.3 m, tested on the same shake table facility. It is noted that the recovering the UAV camera pose was not feasible in the previous study due to insufficient stationary feature points with precisely known positions. As a result, two major improvements are implemented in the present study: (1) strategic placement of a sufficient amount of reference targets with precise position information on the test structures and background region and (2) development of the sub-pixel detection algorithm for accurately extracting image points of the reference targets (refer to Section 3). These strategies substantially enhance the robustness and accuracy of the recovered camera pose and structural displacements achieved from the vision-based analysis procedures.

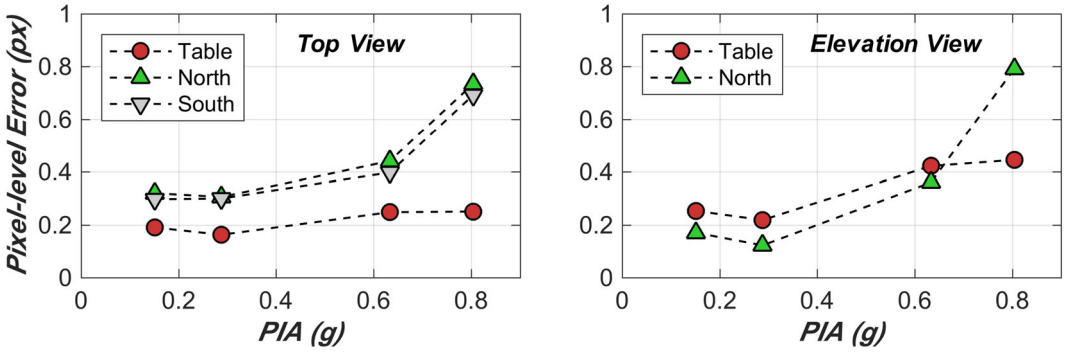

FIGURE 14 Pixel-level precision of the structural displacement tracking results during the earthquake tests. PIA = peak input acceleration of the earthquake motion

Despite the sub-pixel level accuracy for the video analysis results, important limitations exist with regard to the implemented UAV displacement tracking procedure. Besides the practical constraints related to the image sensors (e.g., frame rate and pixel resolution), perhaps two prominent limitations are worth mentioning herein:

1. The use of monocular camera systems allows for resolving structural displacements in at most two dimensions. This is because recovering the world point from the image point requires known coordinate information in at least one dimension (see Section 4.4). Resolving three-dimensional structural displacements simultaneously requires either the use of multi-UAV systems (time synchronization needed for videos taken from different vehicles) or customized UAV platforms equipped with stereo cameras (such cameras are not commonly available in commercial UAV platforms).
2. The accuracy of camera pose (or motion) recovered using the vision-based technique are not validated in the present study due to the absence of auxiliary information associated with the UAV position during the flight. An enhanced confidence level and potential improvement regarding the camera pose estimation may be achieved with the support of auxiliary data collected by on-board positioning sensors (e.g., inertial sensors, range or depth sensors, and real-time kinematic sensors). Customized UAV platforms are needed for acquiring data collected by such sensors.

In the event of real earthquakes, all the reference targets, including those at the ground or on the steel posts, would be subjected to earthquake-induced movements rather than remaining stationary. Although the targets would not remain stationary during real earthquake events, it is reasonable to assume that the ground movement, including the reference targets at the ground level, follows a rigid motion pattern. This assumption is considered reasonable in the absence of ground failure (e.g., surface rupture during an extreme event). In this regard, the proposed methodology remains valid for monitoring the dynamic displacements of structures during real earthquakes, the proposed methodology continues to employ the ground level as the fixed (stationary) reference frame. In such a scenario, the camera position recovered based on the image-to-world correspondence of the reference targets represents the camera position relative to the ground level instead of the camera pose in an absolutely stationary reference frame, and therefore, the dynamic displacements at the top of the specimens reflect those relative to the ground level. Importantly, it is the relative movement of the structure (often referred to as seismic drift) rather than the absolute displacements that result in damage to the structures.

In real structure monitoring applications, practical issues associated with the UAV monitoring plan, including camera viewpoint selection, target placement strategy, and target dimension (or resolution), rely on specific monitoring scenes and the camera intrinsic parameters (e.g., focal length and field of view). As discussed in Section 3.1, a virtual point cloud model of the test scene was developed in the test preparation stage to guide the monitoring plan development in the present study. Such preparation strategies are recommended for enhancing the robustness and accuracy associated with the reference target detection in the UAV displacement tracking methodology.

5.4 | Computational performance

As discussed in Section 4, the frame-by-frame video analysis procedures implemented in MATLAB[27] consist of three sequential steps: (1) reference target detection (from video frames), (2) camera pose recovery, and (3) reconstructing the world points of the moving (non-stationary) targets (also refer to Figure 7). The computational performance of the video frame analysis is predominated by the image processing steps associated with reference target detection, whereas the remaining two steps only involve algebraic equation solving. The CPU processing time for the last two steps is negligible (<3%) in comparison to that allocated to the image detection of the reference targets. The video frames are all analyzed using a laptop computer with a 2.7 Hz Intel Core i5 processor and 8 Gb memory. Depending on the number of reference targets, the average CPU processing time for each video frame is ~1320 ms for the top-view video frames (with 15 reference targets) and ~710 ms for the elevation-view video frames (with 8 reference targets). The present analysis procedures are computationally more efficient than those presented in a prior UAV monitoring study conducted by the authors,[21] since the average processing time reported in the prior study is ~2170 ms per video frame under the same computational environment. This is attributed to the fact that the image processing procedures of the prior study deal with natural feature detection and matching over almost the entire image domain, whereas the present methodology focuses only on several pre-defined regions of interest for reference target detection.

6 | CONCLUSIONS

Unmanned aerial vehicle (UAV) imagery has recently emerged as an effective sensing tool in civil engineering applications such as construction progress monitoring, operational condition inspection, and post-disaster damage assessment of civil structures (e.g., bridges and buildings). However, the use of such techniques for quantitatively tracking the subtle (centimeter-level) variations of structural dynamic responses has occurred in only a handful of prior studies. To this end, a UAV vision-based analysis methodology is developed for tracking the structural dynamic displacements of full-scale single-story cold-formed steel test specimens using test videos during a series of shake table tests at various earthquake intensities. These test videos were consistently collected at two strategically defined camera viewpoints (top and elevation views) during each earthquake test. Importantly, geo-referenced checkerboard targets were strategically placed on the monitored structures and the background (stationary) region to enhance the robustness and accuracy of recovering the camera pose and structural displacements. Furthermore, a sub-pixel detection algorithm is developed for accurately extracting the target image features in the video frames. Validation of the UAV video analysis results against the ground truth (string potentiometer) measurements demonstrates the effectiveness of the proposed methodology for quantifying the dynamic displacement responses of the single-story test structures during low- to high-intensity earthquake excitations. With the availability of camera pose information estimated using the photogrammetric methods, the video analysis results achieve a high level of precision (<3 mm root-mean-square errors) for both the top- and elevation-view test videos, thus demonstrating the potential for the use of UAV to monitor structural dynamic responses in future engineering applications.

ACKNOWLEDGMENTS

This study is supported by NIST award: Improving Disaster Resilience Through Scientific Data Collection with UAV Swarms (Award #: 70NANB17H211). The shake table test program was funded through the National Science Foundation (NSF) Collaborative Research: Seismic Resiliency of Repetitively Framed Mid-Rise Cold-Formed Steel Buildings (CMMI Award #: 1663569 and 1663348). The individuals providing technical support for the shake table test program, including PhD student Amanpreet Singh from UCSD and NHERI@UCSD staff, namely, Robert Beckley, Darren McKay, Jeremy Fitcher, and Alex Sherman, are greatly appreciated. Findings, opinions, and conclusions are those of the authors and do not necessarily reflect those of the sponsoring organizations.

AUTHOR CONTRIBUTIONS

Xiang Wang: conceptualization; methodology; data acquisition; analysis, validation, and visualization; writing—original draft and editing. **Eric Lo**: conceptualization; data acquisition. **Luca De Vivo**: conceptualization; data acquisition. **Tara C. Hutchinson**: conceptualization; methodology; writing—review and editing. **Falko Kuester**: conceptualization; methodology; writing—review.

DATA AVAILABILITY STATEMENT

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

ORCID

Xiang Wang https://orcid.org/0000-0002-9845-1875
Tara C. Hutchinson https://orcid.org/0000-0001-9109-7896

REFERENCES

1. Ham Y, Han KK, Lin JJ, Golparvar-Fard M. Visual monitoring of civil infrastructure systems via camera-equipped unmanned aerial vehicles (UAVs): a review of related works. *Visualization Eng*. 2016;4(1):1-8.
2. Han KK, Golparvar-Fard M. Potential of big visual data and building information modeling for construction performance analytics: an exploratory study. *Autom Construct*. 2017;73:184-198.
3. Asadi K, Suresh AK, Ender A, et al. An integrated UGV-UAV system for construction site data collection. *Autom Construct*. 2020;112:103068.
4. Ellenberg A, Kontsos A, Moon F, Bartoli I. Bridge related damage quantification using unmanned aerial vehicle imagery. *Struct Control Health Monit*. 2016;23(9):1168-1179.
5. Bhowmick S, Nagarajaiah S, Veeraraghavan A. Vision and deep learning-based algorithms to detect and quantify cracks on concrete surfaces from UAV videos. *Sensors*. 2020;20(21):6299.

6. Jiang S, Zhang J. Real-time crack assessment using deep neural networks with wall-climbing unmanned aerial system. *Comput-Aided Civil Infrastruct Eng*. 2020;35(6):549-564.
7. Liu Y-F, Nie X, Fan J-S, Liu X-G. Image-based crack assessment of bridge piers using unmanned aerial vehicles and three-dimensional scene reconstruction. *Comput-Aided Civil Infrastruct Eng*. 2020;35(5):511-529.
8. Tian Y, Zhang C, Jiang S, Zhang J, Duan W. Noncontact cable force estimation with unmanned aerial vehicle and computer vision. *Comput-Aided Civil Infrastruct Eng*. 2021;36(1):73-88.
9. Meyer D, Hess M, Lo E, Wittich CE, Hutchinson TC, Kuester F. UAV-based post disaster assessment of cultural heritage sites following the 2014 South Napa Earthquake. *2015 Digital heritage*, Vol 2. Manhattan, New York: IEEE; 2015:421-424.
10. Dominici D, Alicandro M, Massimi V. UAV photogrammetry in the post-earthquake scenario: case studies in L'Aquila. *Geomatics, Nat Hazards Risk*. 2017;8(1):87-103.
11. Franke KW, Rollins KM, Ledezma C, et al. Reconnaissance of two liquefaction sites using small unmanned aerial vehicles and structure from motion computer vision following the April 1, 2014 Chile earthquake. *J Geotechnical Geoenvironmental Eng*. 2017;143(5):4016125.
12. Lee JJ, Shinozuka M. A vision-based system for remote sensing of bridge displacement. *NDT E Int*. 2006;39(5):425-431.
13. Park SW, Park HS, Kim JH, Adeli H. 3D displacement measurement model for health monitoring of structures using a motion capture system. *Measurement*. 2015;59:352-362.
14. Feng D, Feng MQ. Vision-based multipoint displacement measurement for structural health monitoring. *Struct Control Health Monit*. 2016;23(5):876-890.
15. Yoon H, Elanwar H, Choi H, Golparvar-Fard M, Spencer Jr BF. Target-free approach for vision-based structural system identification using consumer-grade cameras. *Struct Control Health Monit*. 2016;23(12):1405-1416.
16. Xu Y, Brownjohn J, Kong D. A non-contact vision-based system for multipoint displacement monitoring in a cable-stayed footbridge. *Struct Control Health Monit*. 2018;25(5):e2155.
17. Xu Y, Brownjohn JMW. Review of machine-vision based methodologies for displacement measurement in civil structures. *J Civil Struct Health Monit*. 2018;8(1):91-110.
18. Bhowmick S, Nagarajaiah S, Lai Z. Measurement of full-field displacement time history of a vibrating continuous edge from video. *Mech Syst Signal Process*. 2020;144:106847.
19. Bhowmick S, Nagarajaiah S. Identification of full-field dynamic modes using continuous displacement response estimated from vibrating edge video. *J Sound Vibr*. 2020;489:115657.
20. Hoskere V, Park J-W, Yoon H, Spencer BF. Vision-based modal survey of civil infrastructure using unmanned aerial vehicles. *J Struct Eng*. 2019;145(7):4019062.
21. Wang X, Wittich CE, Hutchinson TC, et al. Methodology and validation of UAV-based video analysis approach for tracking earthquake-induced building displacements. *J Comput Civil Eng*. 2020;34(6):928.
22. Weng Y, Shan J, Lu Z, Lu X, Spencer Jr BF. Homography-based structural displacement measurement for large structures using unmanned aerial vehicles. *Comput-Aided Civil Infrastruct Eng*. 2021;36(9):1114-1128.
23. Yoon H, Shin J, Spencer BF. Structural displacement measurement using an unmanned aerial system. *Comput-Aided Civil Infrastruct Eng*. 2018;33(3):183-192.
24. Singh A, Wang X, Hutchinson TC. Lateral response of cold-formed steel framed steel sheathed in-line wall systems detailed for mid-rise buildings. Part I: Shake table test phase. SSRP-2019/05, La Jolla, CA, Department of Structural Engineering, University of California San Diego; 2021.
25. Gao X-S, Hou X-R, Tang J, Cheng H-F. Complete solution classification for the perspective-three-point problem. *IEEE Trans Pattern Anal Machine Intell*. 2003;25(8):930-943.
26. Lepetit V, Moreno-Noguer F, Fua P. Epnp: an accurate o(n) solution to the PnP problem. *Int J Comput Vision*. 2009;81(2):155-166.
27. MathWorks. *Computer vision toolbox (R2019a)*. Natick, MA: the MathWorks Inc.; 2019.
28. Canny J. A computational approach to edge detection. *IEEE Trans Pattern Anal Machine Intell*. 1986;PAMI-8(6):679-698.
29. Trujillo-Pino A, Krissian K, Alemán-Flores M, Santana-Cedrés D. Accurate subpixel edge location based on partial area effect. *Image Vision Comput*. 2013;31(1):72-90.
30. Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu AY. An efficient k-means clustering algorithm: analysis and implementation. *IEEE Trans Pattern Anal Machine Intell*. 2002;24(7):881-892.
31. Hartley R, Zisserman A. *Multiple view geometry in computer vision*. Cambridge, UK: Cambridge University Press; 2003.
32. Yoo J-C, Han TH. Fast normalized cross-correlation. *Circuits, Syst Signal Process*. 2009;28(6):819-843.
33. Bouguet JY. Camera calibration toolbox for Matlab. Retrieved from: http://vision.caltech.edu/bouguetj/calib_doc; 2015.
34. Triggs B, McLauchlan PF, Hartley RI & Fitzgibbon AW Bundle adjustment—a modern synthesis. In: International Workshop on vision algorithms. Berlin, Heidelberg: Springer; 1999:298-372.
35. Sutton MA, Yan JH, Tiwari V, Schreier HW, Orteu J-J. The effect of out-of-plane motion on 2D and 3D digital image correlation measurements. *Optics Lasers Eng*. 2008;46(10):746-757.

How to cite this article: Wang X, Lo E, De Vivo L, Hutchinson TC, Kuester F. Monitoring the earthquake response of full-scale structures using UAV vision-based techniques. *Struct Control Health Monit*. 2021;e2862. doi:10.1002/stc.2862