

Title: A Stereo Synchronization Method for Consumer-Level Video Cameras to Measure Multi-Target 3D Displacement Using Digital Image Correlation (DIC) in Shake Table Experiments.

Mearge Kahsay Seyfu, Yuan-Sen Yang*

Department of Civil Engineering, National Taipei University of Technology, No. 1., Sect. 3, Zhongxiao E. Road, Taipei, 106, Taiwan, t112429402@ntut.org.tw, *ysyang@ntut.edu.tw

Abstract

Stereovision-based digital image correlation (DIC) measurement using consumer-level cameras has become a reliable method for measuring three-dimensional displacement in civil engineering applications. This method is cost-effective, non-contact, reasonably accurate, and requires less effort and time to set up compared to conventional measurement methods. Three-dimensional displacement measurements require multiple cameras to simultaneously capture image frames of the same scene, ensuring the frames correspond to the exact moment. However, most consumer-level cameras do not support hardware synchronization. Researchers have proposed various software-based synchronization approaches to address this problem. However, most of the proposed methods can only achieve frame-level precision, which is not accurate enough for shake table experiments. Even tiny timing differences between stereo videos can significantly affect the accuracy of triangulated results. Numerous existing approaches are not suitable for shake table experiments. In addition, the absence of user-friendly software for 3D stereovision-based displacement measurement also poses a challenge for engineers and researchers. This paper proposes a novel method for synchronizing consumer-level video cameras to enable accurate 3D displacement measurements of multiple targets using digital image correlation for shake table experiments. The approach determines the time lag between the two videos by minimizing triangulation errors and optimizing them using polynomial interpolation to achieve subframe accuracy. A software with a graphical user interface is developed to analyze the three-dimensional displacement results. The accuracy of the proposed method was validated using a motion capture system and dial gauges with different cases of shake table experiments. The proposed synchronization method optimizes the frame-level triangulation root mean squared error of target points by a maximum of 9.67% (0.02 pixels). The maximum error in the displacement of the proposed approach was less than a millimeter for a three-degree-of-freedom motion in all experiments. Generally, the proposed synchronization method and the developed application software with the easy graphical user interface are viable and accurate with a submillimeter measurement error for shake table experiments.

Keywords: DIC, Synchronization, 3D displacement, Accuracy, Shake table, Software

1. Introduction

In recent years, the advancements in computational capacity and camera technology have made vision-based measurement methods increasingly popular across various fields as a viable alternative to contact-based measurement methods [1] [2]. This emerging technology has also attracted researchers and engineers in the structural and earthquake engineering field. Such technology can capture the deformation, vibration, strain, and displacement of structures and non-structural components subjected to dynamic loading by leveraging the power of computer vision and image processing techniques. These technologies are essential tools for dynamic structural response analysis, design verification and structural health monitoring [3]. While conventional measurement methods are recognized for their accuracy [4], they have some notable drawbacks. The mounting of such devices is often complicated, requiring significant effort and time to set up properly [5]. Additionally, the dynamic range for these devices is commonly limited and this may restrict their effectiveness when employed in certain testing scenarios [6]. These methods are also susceptible to damage, particularly during dynamic tests such as shake table experiments where devices can be subjected to extreme structural conditions. Moreover, the cost associated with these devices is relatively higher than image-based measurements [7] [5]. It's important to note that traditional measurement methods only capture unidirectional motion, which can restrict their applicability in more complex motion scenarios. It is not practical to use conventional measuring devices for freely moving structural or non-structural components, such as suspended ceilings that may encounter large displacement, overturning, bending, collapsing, and twisting during seismic excitation [8]. Vision-based 3D measurements can track and measure the three-dimensional positions of moving targets without making direct contact with the experimental specimen. As a result, the measurement devices are safe from damage. Also, setting up the cameras for the experiment is straightforward, and a data acquisition system is not required in vision-based measurements.

Digital image correlation (DIC) is an emerging method for measuring displacement using vision-based technology. Few studies have utilized DIC in shake table experiments to analyze the response of structures and structural components. For instance, Ngeljaratan and Moustafa [9] used a four-story steel frame structure to assess the response of the structure subjected to vibration. They reported that DIC overestimates the peak values of the response by 5-7% compared to LVDT. Wani et al. [10] used a 5-story damped steel frame structure subjected to shake table ground acceleration by employing a template match algorithm. The maximum difference in displacement measurement was 6.142 mm. A timber-framed structure filled with stone and earth was used by Sieffert et al. [11] to study the response of a single-story building under seismic load. They adopted a

DIC software named “Tracker” for their analysis. The difference between DIC and LVDT displacement measurements was less than 5%. The accuracy of dynamic DIC measurement is dependent on many factors. Such as the selection of tracking algorithm [12], marker size, marker type [13], camera resolution [14], video sampling rate, lighting condition, camera lens quality, calibration [15], speed of target [16], specimen properties, environmental factors [17] [18], camera stability [19], camera distance, and camera viewing angle [20] [21].

Image-based 3D measurement techniques generally have four steps. The first step is calibration, the second is tracking, the third is synchronization, and the final is stereo triangulation. Calibration is correcting the camera for lens distortion and other parameters related to the position of cameras. Tracking is determining the position of targets in an image frame concerning its specific time history. Synchronization coordinates the recording of a scene to occur concurrently on all cameras, and stereo triangulation is computing the 3D coordinates of points. However, not all image-based measurement methods need multiple cameras for 3D measurement [4] [6]. Multi-camera synchronization is one of the key aspects that require special attention to improve the accuracy of the 3D position of targets. When structures experience rapid loading or vibrational effects, even tiny timing differences in video capture can cause significant inaccuracies in the triangulation results of targets [22] [23]. Multiple methods for synchronizing cameras were devised, each having merits and demerits. Lavezzi et al. [24] Employs a manual synchronization technique by adjusting the timestamp of one camera to align with the other. However, the displacement results were noisy. Hardware-based multi-camera synchronization methods were utilized by Malowany et al. [20] and Wei et al. [25]. This synchronization method may yield the highest accuracy. However, it requires a data acquisition system, which makes it costly and impractical for onsite civil engineering structures measurement. Igniting a light source caught by all cameras can serve as a synchronization marker. But it may not yield frame-level precision [26]. The cross-correlation method was used by Yang [15] to synchronize the time difference between two cameras. But this method can only synchronize with a precision of integer frame resolution. A wavelet transform approach was used by Nikfar and Konstantinidis [27] to synchronize the video signal of consumer-level cameras. A software-based approach that operates in two stages was proposed by Ansari et al. [28] In their approach, the clocks of all devices were synchronized to the clock of a leader device using a variant of the Network Time Protocol (NTP). In the second stage, the cameras were instructed to capture a continuous stream of images, and the phase of all client streams was shifted to align with the leader's stream. They managed to achieve a synchronization accuracy of less than 250 microseconds. However, this method requires hardware camera timestamps and image streaming capabilities, which may not be supported by all devices.

A vision-based measurement method with precise synchronization is an effective approach for measuring torsion and rigid body rotation. However, improper synchronization can lead to phase mismatches between rotational data and other system parameters, resulting in incorrect interpretations of the structure's behavior. This misalignment can either amplify or attenuate perceived torsional effects, leading to erroneous conclusions about the system's stability and its response to external forces. It is crucial to address synchronization issues to ensure the precision and validity of the measurements in torsional studies. Without it, discrepancies in frame capture timing can lead to errors in interpreting motion or deformation, emphasizing the critical role of precise timing in motion analysis. Despite the various approaches that have been proposed for camera synchronization, there remains a significant need for an alternative method that effectively accommodates structural experiments. This study aims to develop a synchronization method with sub-frame accuracy and graphic user interface (GUI)-based software for 3D displacement measurement of multiple targets in shaking table experiments utilizing consumer-level cameras. The accuracy of the proposed synchronization method and the DIC measurement was assessed with a series of experiments.

2. Methodology

This section describes the methodological workflow and experimental program used in this study. The research encompassed two distinct experimental phases. The first phase involved a physical experiment, consisting of video recordings of experiments conducted on moving targets. The second phase comprised a non-physical experiment, centered on image analysis using the developed application software.

2.1 Experimental program

This study employed three distinct experimental setups to evaluate the accuracy of the proposed synchronization method and the developed application software in determining the 3D coordinates of multiple moving targets under various conditions. The types of experimental setups used are listed as follows.

- a) A small-scale frame structure (Fig. 1 (a)) fixed at its base with dimensions 26 cm wide, 35 cm long, and 36 cm high was used to test the accuracy of the proposed stereo synchronization method. A grid of twenty tracking markers was fixed and arranged in a four-by-five configuration to the ceiling of the specimen. Figure 1(a) shows an image of the structure's ceiling with tracking markers on its surface. An excitation force was applied by pushing the top edge of the structure's ceiling in the horizontal direction, then it vibrated freely until it stopped, and video was recorded using two SjCam consumer-level action

cameras situated beneath the ceiling of the structure with a recording video quality of 4K and 60 frames per second (fps) sampling rate. Three camera orientations were employed to assess the effect of camera orientation on the accuracy of stereo triangulation: parallel (Fig. 1(b)), opposite (Fig. 1(c)), and angular (Fig. 1(d)).

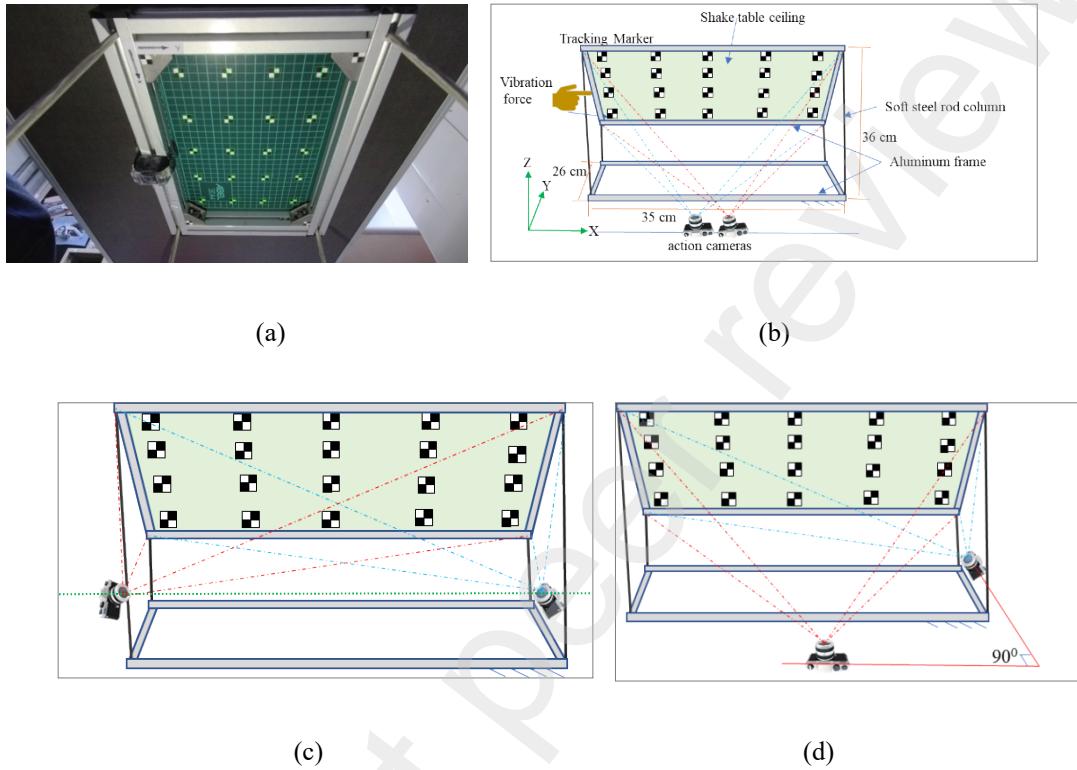


Fig. 1. Pictorial illustration of the experimental setup of the small small-scale frame structure used for assessing the accuracy of the synchronization method in different camera directions: (a) image of the shake table ceiling, (b) parallel camera setup, (c) 90-degree camera setup, (d) opposite camera setup

b) A shake table with a known fixed displacement of $-/+ 3.00$ mm, The OptiTrack motion capture system [29] and, a dial gauge was used to validate the accuracy of displacement measurement of the image-based measurement system. OptiTrack is a high precision industrial level motion capture system. In the past, studies investigated the accuracy of the OptiTrack motion capture system using different approaches, and they proved its measurement error is less than a millimeter [30] [31]. The motion capture system used for validation has six Optiflex 13 cameras positioned at a distance of 2 m from the targets. Two SjCam action cameras were used to record experimental videos with 4K video quality and 60 fps frame rate. Fig. 2. Shows the schematic representation of the experimental setup used. Nine and three points were used for calibration and tracking respectively.

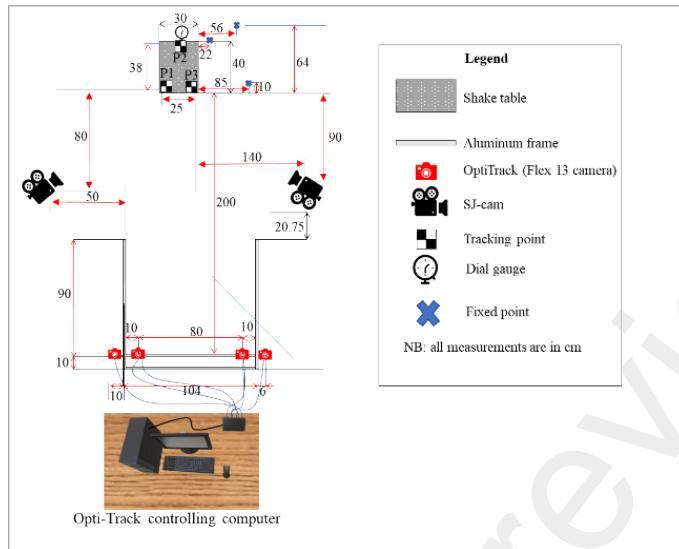
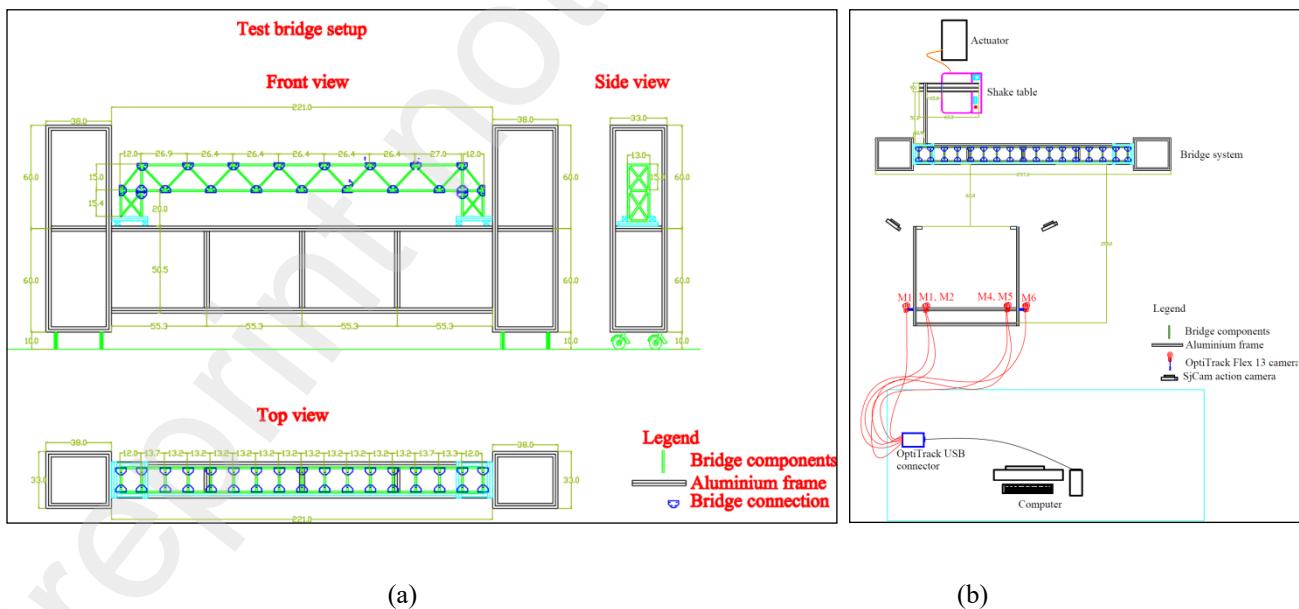
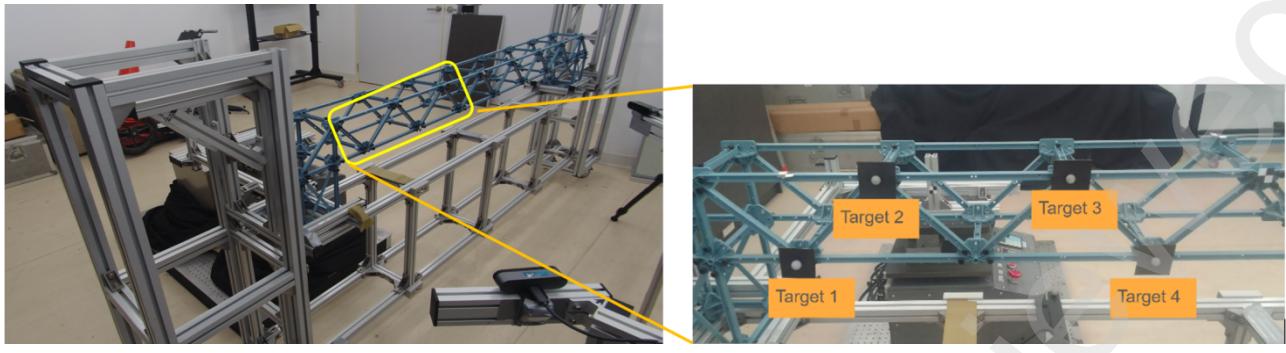


Fig. 2. Schematic of the experimental setup of the shake table.

c) For this experiment, a truss bridge model was built using a plastic frame and connected to a shake table, as illustrated in Figures 3(a, b and c). To verify the accuracy of the displacement measurements obtained by the proposed method, we used an industrial level motion capture system. This experiment differs from the others because the tracking points were located on a secondary object, and the specimen is not rigid. This secondary object had its unique vibration pattern, distinct from the primary vibration source (the shake table).





(c)

Fig. 3. schematic presentation of the bridge system used for the experiment: (a) front, side, and top view of the bridge system (b) whole bridge vibration experimental setup, (c) photograph of the bridge system and targets used.

2.2 Image based measurement mathematical models and procedures

This study employs a stereo-vision system to capture the 3D coordinates of target points in images [32], generating time-series data for each target across the measurement specimen. The time history of these targets allows for estimating displacement and vibration modes. Deformation modes, including rigid body displacement, torsion, shear, bending, and crack patterns, can be derived from the 3D coordinates of multiple points over time [15]. In a stereo vision system, transferring a point from three-dimensional space to two-dimensional images and camera coordinates necessitates geometric transformations [19]. To perform these transformations, three coordinate systems must be established as world coordinates $P_w = (X_w, Y_w, Z_w)$, camera coordinates $P_c = (x_c, y_c, z_c)$ and image coordinate $P_i = (u, v)$ as shown in Fig. 4.

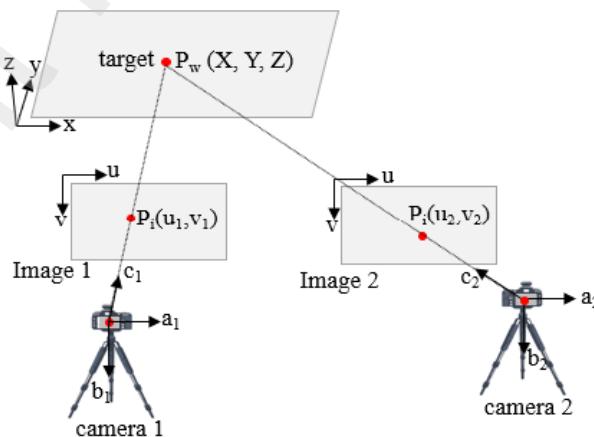


Fig. 4. Illustration of stereo camera setup

Coordinate transformation from the image coordinate to the world coordinate can be expressed in terms of a 4×4 matrix as indicated in equation 1. This matrix is employed to account for the rotation, translation, affine, and perspective transformation [33]. The 3×3 matrix located in the upper left corner of the 4×4 matrix is the rotational matrix, and the vector $T = (T_x, T_y, T_z)^T$, represents the translation of the world coordinate (P_w) to the origin of the camera coordinate system (x_c, y_c, z_c). Mathematically, the relation between camera coordinate and world coordinate can be defined using equation 1 [33] [34].

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R_{xx} & R_{yx} & R_{zx} & T_x \\ R_{xy} & R_{yy} & R_{zy} & T_y \\ R_{xz} & R_{yz} & R_{zz} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (1)$$

The geometric coordinate transformation from the camera coordinate $P_c = (x_c, y_c, z_c)$ to the image coordinates $P_i = (u, v)$ can be represented using equation 2. In this kind of transformation, the computation is easy when the dimension of coordinates of the projective space is equivalent to the projected space. So, the coordinates on the camera screen $P_i = (u, v)$ will be represented by $P_i = (u, v, w)$. The proportion of values for all points originating in a three-dimensional space, when projected onto a two-dimensional screen, remains consistent and equivalent [35, p. 3]. If we divide the image coordinates (u, v, w) by a unit length $c = 1$ we get the normalized coordinate $(u_n, v_n, 1)^T$ [15]. So, image coordinates after the effect of distortion are accounted for as given by equation 2.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_d \\ v_d \\ 1 \end{bmatrix} \quad (2)$$

Where f_x , f_y , c_x , and c_y are the intrinsic camera parameters that define the characteristics of the camera and the 3×3 matrix containing these values is known as the camera matrix, and $(u_d, v_d, 1)^T$, is a matrix which accounts the effect of lens distortion on normalized coordinates. The relationship between distorted coordinates and normalized coordinates can be defined using equation 3 [15] [35, p. 3].

$$\begin{bmatrix} u_d \\ v_d \end{bmatrix} = \begin{bmatrix} k + 2p_1y_n + 3p_2x_n & p_2x_n \\ p_1x_n & k + 2p_2x_n + 3p_1y_n \end{bmatrix} \begin{bmatrix} u_n \\ v_n \end{bmatrix} \quad (3)$$

Where:

$$k = \frac{1 + k_1(x_n^2 + y_n^2) + k_2(x_n^2 + y_n^2)^2 + k_3(x_n^2 + y_n^2)^3}{1 + k_4(x_n^2 + y_n^2) + k_4(x_n^2 + y_n^2)^2 + k_6(x_n^2 + y_n^2)^3} \quad (4)$$

Where $k_1, k_2, k_3, k_4, k_5, k_6, p_1$, and p_6 are distortion coefficients.

The image-based measurement was conducted by the following four steps: (1) camera calibration, (2) tracking, (3) synchronization, and (4) stereo triangulation.

2.2.1. Camera calibration

Images captured by a camera often exhibit distortion due to the perspective features of camera lenses. The process of correcting distortion and other parameters to get a rectified image is called camera calibration. There are two types of camera parameters to consider: intrinsic parameters and extrinsic parameters. Intrinsic parameters define how the camera captures images, and extrinsic parameters defines where the camera is located in the 3D environment. Most commonly, chessboards are used to calibrate cameras [15] [36] [37] [38] [39] because corners are easily identifiable by calibration algorithms. However, chessboards are not reliable for experiments with large areas since it is not practical to use large rigid calibration chessboards. In this study different types of markers with known world coordinates of a single image frame selected from a video were used as input for our calibration software developed using a computer vision library OpenCV [40]. Templates belonging to respective targets were picked using the calibration algorithm to get the image coordinates. Each camera was calibrated separately. The intrinsic and extrinsic camera parameters were calculated using our program based on the relations defined in equations (1-4).

2.2.2 Tracking

The process of detecting and following the movement of a specific target within a sequence of images in computer vision-based measurement is called target tracking. To identify the target with tracking algorithms, it's crucial to have tracking points within a small region of interest (ROI), commonly known as a template. This template should exhibit visual characteristics that differ from the surrounding elements in the parent image. Templates can be created by applying paint, attaching markers, or utilizing natural features of the measurement area such as holes, bolts, or joints. In this study, we employed several types of templates for the tracking task, including black-and-white high-contrast markers, holes, backgrounds with distinguishable patterns, and stickers as a template for the tracking task. Specifically, black and white high-contrast markers, holes, backgrounds with distinguishable patterns, and stickers (see Fig. 5). We used an enhanced correlation coefficient (ECC) tracking method to track the targets. ECC is a computer vision approach that tracks objects or features across a sequence of images. It seeks to search for the best correspondence by identifying the highest correlation in the reference image for a given ROI. This method is especially useful in situations where the object's appearance changes over time, such as due to lighting, occlusion, or deformation.

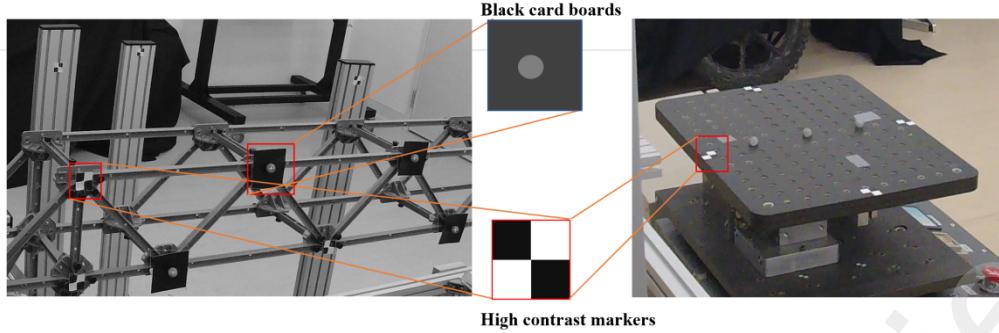


Fig. 5. Types of templates used in this study

2.2.3. Camera synchronization

Camera synchronization is a prerequisite to finding a precise value of 3D coordinates of targets. A little difference in time can cause unexpectedly a large error in the triangulated 3D coordinates can cause unexpectedly a large error in the triangulated 3D coordinates. The experimental cameras used do not support trigger synchronization. As a result, synchronization of the video streams acquired by the cameras is mandatory to get accurate vision-based measurements. Manual synchronization was usually performed with a precision of up to a frame level, though it typically achieved an accuracy of only 2 to 3 frames. This is primarily due to the challenges of manually synchronizing videos by eye. For many consumer video cameras, a frame usually ranges from 1/30 to 1/60 seconds. The synchronization method employed in this study is determining the time lag between the two videos by minimizing triangulation errors. The proposed algorithm tries to find the time lag where the mean root square error (RMSE) of the triangulation projection error of all points is minimum and then optimizes the result to get a sub-frame level time lag using polynomial interpolation. Shimizu [41] proposed a similar method for determining the optimal time lag; however, our method differs in how the time lag is applied. Instead of interpolating between two unsynchronized 3D points, we interpolate between image points, which enables us to get an improved triangulation result. The assumption is that the triangulation projection error (e) is a function of time lag (t_{lag}) expressed by a quadratic equation. If (e_1, e_2, e_3) are RMSE of triangulation at a time lag of (t_1, t_2, t_3) respectively, we can find the optimum time lag (t_{opt}) by derivation of equation (5).

$$e = at^2 + bt + c \quad (5a)$$

$$t_{opt} = \frac{de}{dt}[at^2 + bt + c] = -b/2a \quad (5b)$$

We can find the coefficients (a) and (b) by solving equation (6).

$$\begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} = \begin{bmatrix} t_1^2 & t_1 & 1 \\ t_2^2 & t_2 & 1 \\ t_3^2 & t_3 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \quad (6)$$

The total video length recorded by camera 1 is denoted as ($t_{1\text{total}}$) and for camera 2 as ($t_{2\text{total}}$). For cameras initiated manually, there will always be a time lag (t_{lag}) between the recordings of the two cameras. As shown in Fig. 6, camera 2 begins recording later than camera 1. Hence, the video time of camera 2 (t_{v2}) is the sum of time lag (t_{lag}) and the video time of camera 1(t_{v1}) which can be expressed using equation (7).

$$t_{v2} = t_{v1} + t_{\text{lag}} \quad (7)$$

Where $t_{v1} = a + t_{\text{interest}}$, $t_{v2} = a + t_{\text{lag}} + t_{\text{interest}}$ and $t_{\text{lag}} = t_{20} - t_{10}$

It is not always appropriate to use the very first frame as a starting point for synchronization analysis the range of frames needed for analysis (t_{interest}) can be defined as per the users' criteria. (F_{10} and F_{20}) are (a) frames away from the very first frame (t_{10} and t_{20}) of camera1 and camera2 respectively. (F_{1n} and F_{2n}) are the first and last frames required for synchronization analysis in camera1 and camera2 respectively. In other words, it is the time range in frames (t_{interest}) utilized for analysis, as shown in Fig.6.

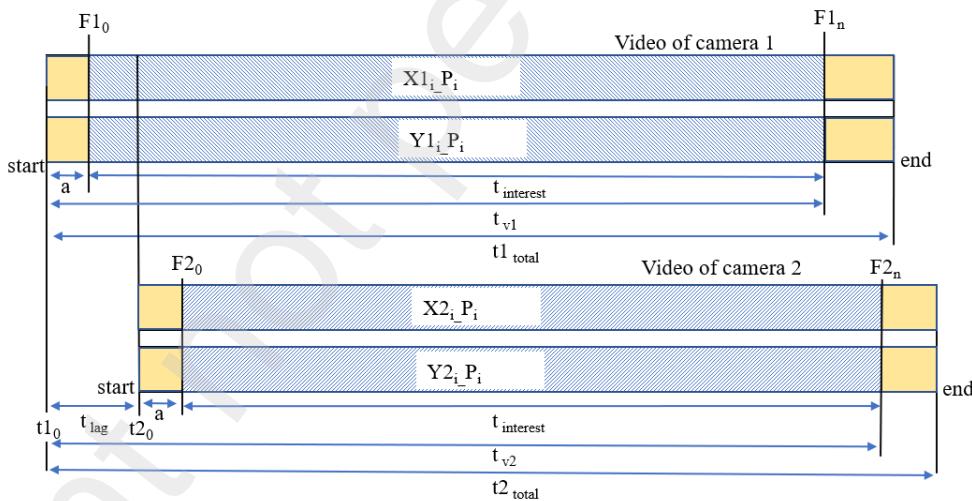


Fig. 6. Visual illustration of time lag between two videos.

2.2.4. Stereo triangulation

Stereo triangulation is a process used in computer vision to determine the 3D position of a point in space by using its projections in two or more images captured from different viewpoints. By analyzing the disparities between the images, systems can accurately reconstruct the spatial coordinates of objects. Ideally, the exact 3D position of a target is located at the intersection of the line of sight and the position of the two cameras. However, when lines do not intersect, an error occurs. To reduce the

error the midpoint between the two lines is used as the target position [41]. The mid-point approach is employed in this study as a triangulation method.

2.3. Software development

The software is developed in Python and designed to analyze two unsynchronized videos of dynamic experiments. The software captures the three-dimensional coordinates of multiple moving targets over a designated period. Tkinter is used to create a GUI of the program (Fig. 8), OpenCV is used as the main computer vision library, and NumPy and SciPy are used to execute mathematical operations. Users can access the folders containing their video files, calibration files, and templates. Users can define the range of tracking, synchronization, and triangulation points based on their needs, and it allows users to select any frame required for tracking from the video file by just inputting the number of frames. It allows users to save the analysis results in a CSV file, and they can visualize the time series displacements. The methodological flow chart of the 3D displacement measurement algorithm is illustrated in Fig. 7. This software will be provided free of charge to encourage its use in the field of structural engineering.

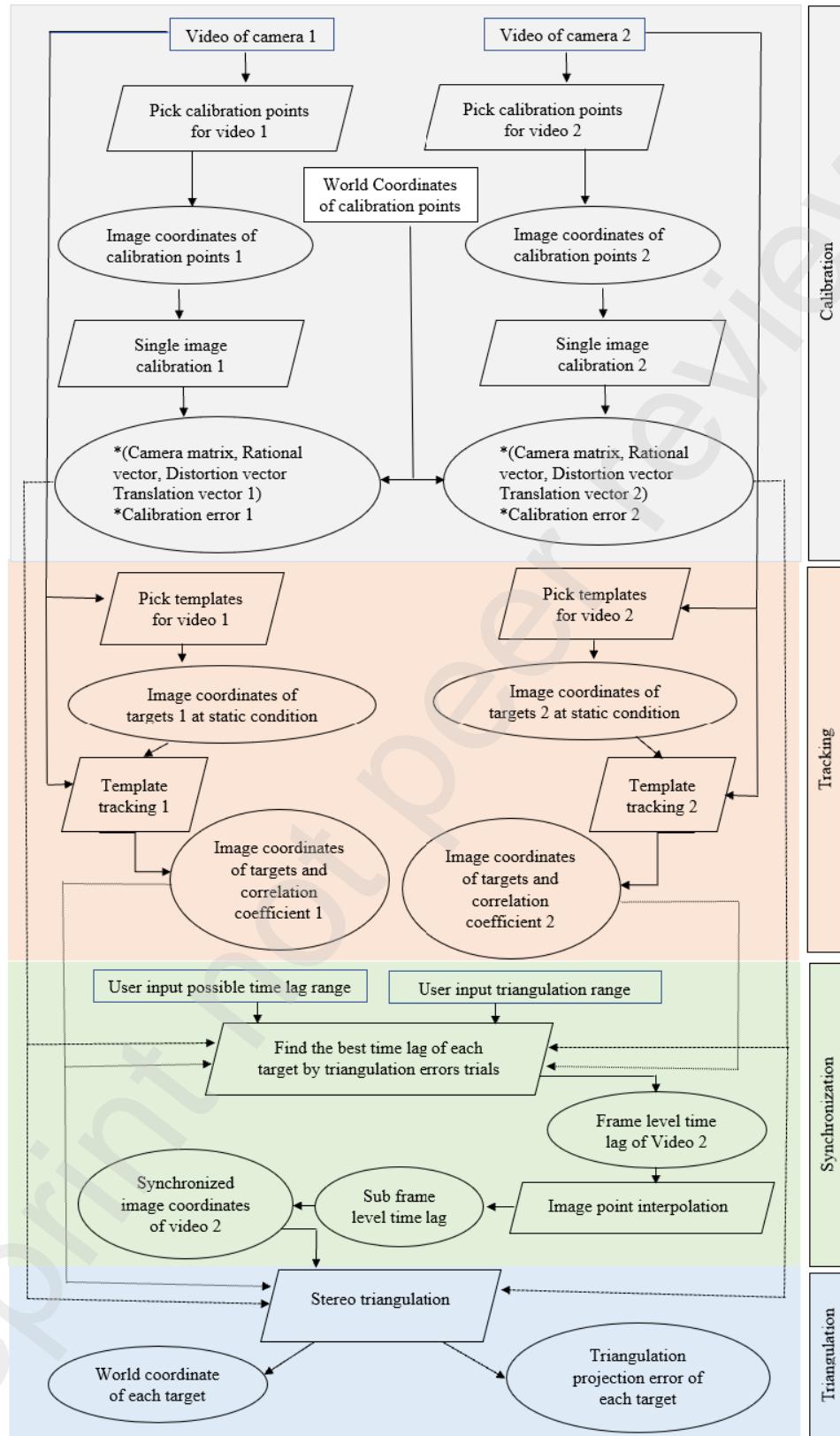


Fig. 7. Flowchart of methodologies of the developed application software.

2.4. Potential area of application

The problem with contact-based motion sensors such as LVDT and accelerometers is that they provide motion data in only two directions, whereas vision sensors that employ stereo cameras have the benefit of providing spatial coordinates in three-dimensions. So, this software could be employed as additional measuring method in the field of earthquake engineering. six degrees of freedom can be captured which include three degree of freedom translation, rocking, rotation and torsion with this software. Also, through studying the distribution of three-dimensional positions of targets on the structure, the modes of deflection and strain of a particular sets of the structural components due to seismic loading can be computed. So, this software may also be applied in construction engineering testing facilities as a technique for non-destructive evaluation and can be applied to conduct experiments for vibration, displacement, deflection and even bending measurements. In addition to this, it can be used to determine the deflection of the structure due to load or due to other environmental conditions, check the alignment of the structure, track down the cracks growth for the purpose of evaluating the condition and the deterioration of the structure.

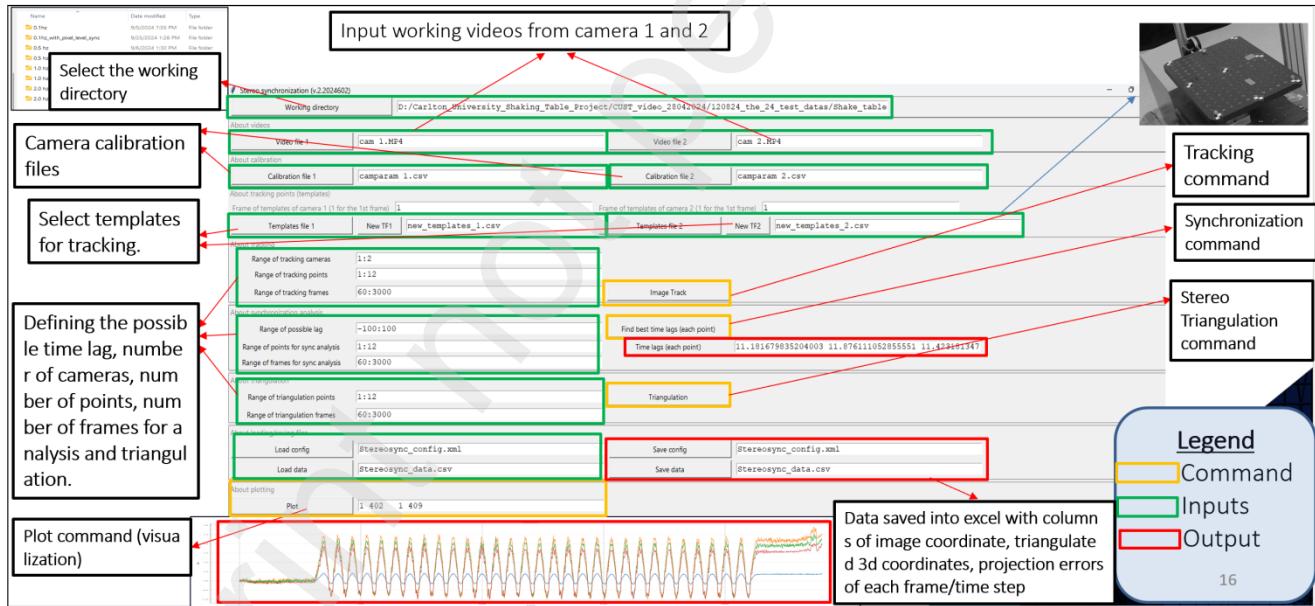


Fig. 8. Graphical user interface of the developed vision-based 3D displacement measurement software.

3. Experimental results

This section presents the experimental and theoretical results for the proposed synchronization approach and vision-based displacement measurement. The 3D position of points was measured for different scenarios and experimental setups. As

mentioned in the experimental setup section, three types of experimental setups were used to evaluate the accuracy of the proposed method. The first experiment is the small-scale frame structure used to assess the accuracy of the synchronization method by interpreting the triangulation projection error results. Ideally, two cameras are synchronized if the RMSE of triangulation is zero. However, zero error is ideal, not practical. The small-scale structure was pushed laterally at one of its corners to initiate the vibration and allow it until it comes to rest, as shown in Fig. 9. The structure exhibits an average vibration frequency of 2.5 Hz over time, the vibrations gradually diminish, and the structure eventually comes to a complete rest after an average of 75 seconds of motion.

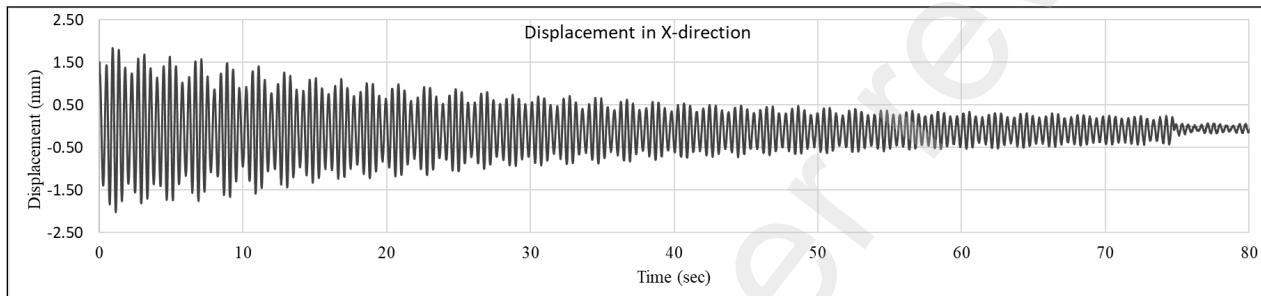


Fig. 9. Vibration trend of the small-scale structure.

Synchronization is critical in ensuring accurate measurements, particularly in dynamic systems where precise time alignment of data from multiple sources is required. The triangulation RMSE of all points (20 targets) for the opposite, 90-degree, and parallel camera setup was 0.209 pixels (improved by 9.67%), 0.663 pixels (enhanced by 2.48%), and 0.363 pixels (improved by 0.54%) when compared with frame level synchronization respectively. Ideally, after synchronization, the time lag for all points should be identical. Figure 10 (b, c, d) illustrates the time lag for each point post-synchronization. This error exists due to the rolling shutter effect. If a time lag of the same value for all points is considered, the rolling shutter error will be high. Hence, to minimize this kind of error, the proposed algorithm computes a time lag for each point. The cameras used are the rolling shutter type, which captures an image line by line like scanners that cause an image to distort if one of the cameras or the scene is in motion. Most rolling shutter cameras scan images from top to bottom with a fixed line delay[42] [43].

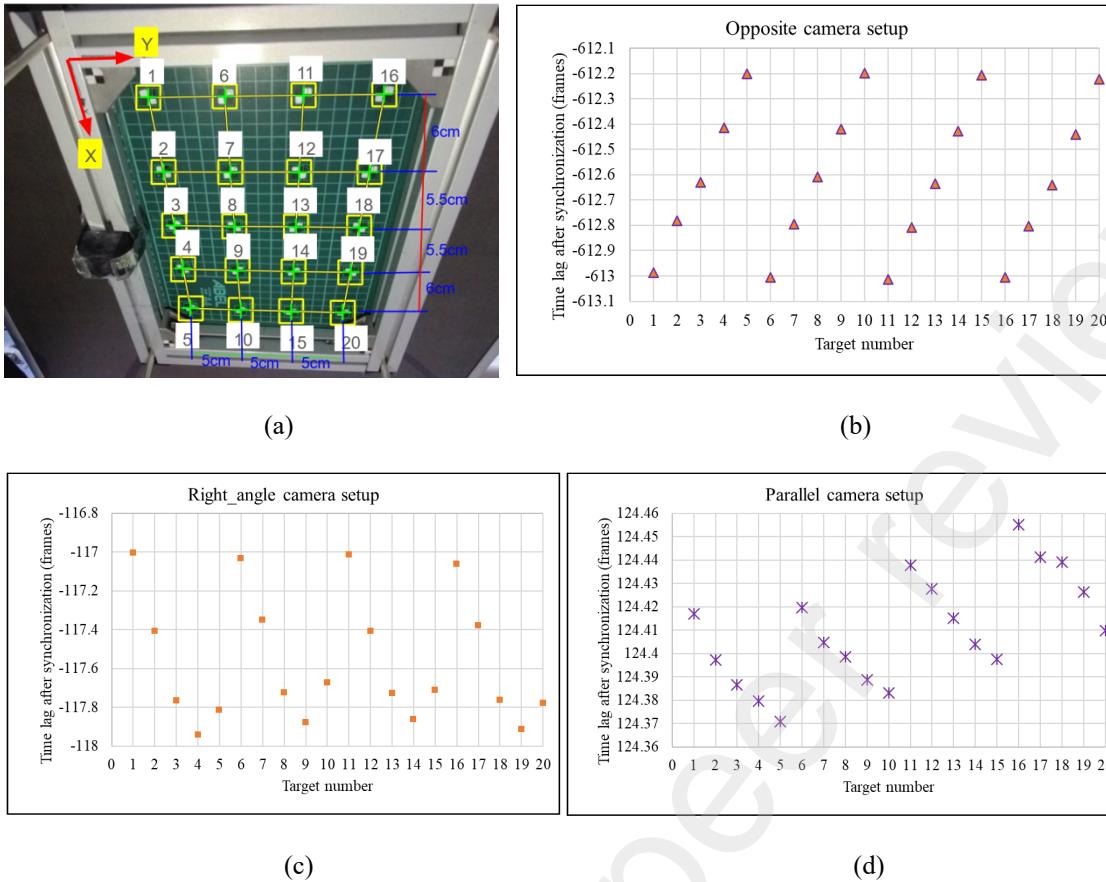


Fig. 10. Time lag (frames) of target points after synchronization: (a) Target alignment and spacing, (b) opposite camera setup, (c) right-angle camera setup, (d) parallel camera setup

The time shift between targets is affected by the orientation of the cameras and is observed in Fig. 10 (a, b, c) by analyzing the time lags. The time differences between the lower targets (5, 10, 15, 20) (Fig. 10 (a)) and the upper targets (1, 6, 11, 16) (Fig. 10 (a)) were approximately 0.795 frames (0.013 seconds) for the opposite camera setup, 0.715 frames (or 0.0119 seconds) for the 90-degree camera setup, and 0.042 frames (0.0007 seconds) for the parallel camera setup, the time differences are notably larger for the opposite and 90-degree camera configurations. For the opposite camera setup, any target captured on the first camera will be captured later on the second camera (Fig. 11 (b)); this is a situation like taking the picture in two cycles while the target is still vibrating; this causes a relatively larger time difference between targets. The time difference is minimal for the parallel camera setup, as both cameras scan the target almost simultaneously (Fig. 11 (a)).

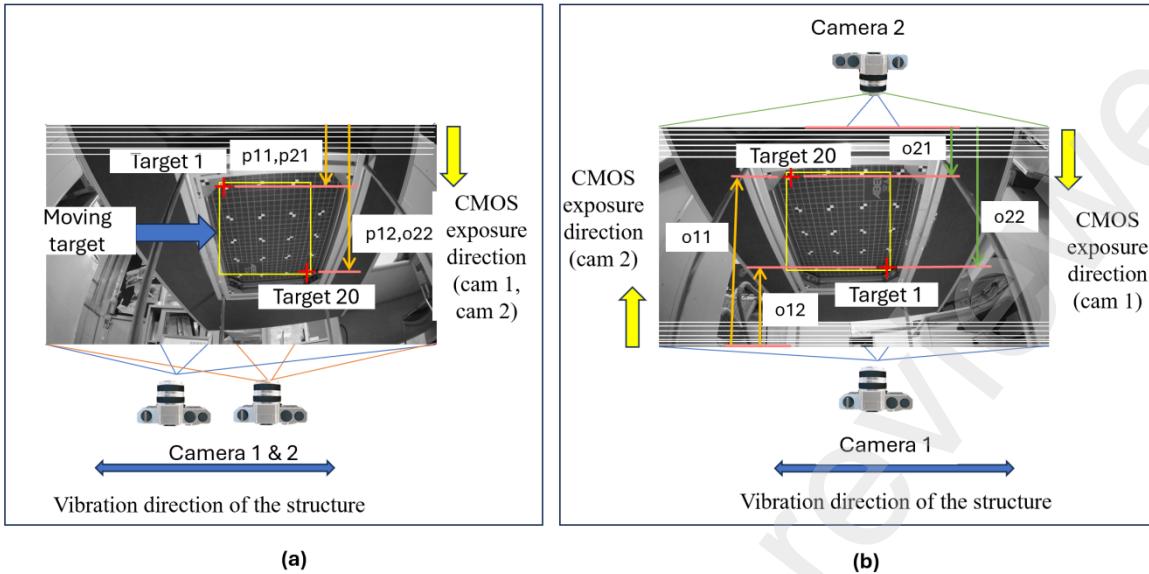


Fig. 11. Effect of camera orientation in time lag difference between targets: (a) parallel camera setup, (b) opposite camera setup.

Additionally, the distance of targets from the top of an image plays a significant role in influencing the time lag difference between targets. If the vertical component of the distance between two targets in the first image is shorter or longer than in the second image, it can result in a noticeable phase shift between the two sets of target points.

Torsional effects on structures can be measured using vision-based methods by comparing the displacement results of different directions. Accurate measurement of rotation angles and positions depends on the proper synchronization of multiple cameras. The results of this study indicate that frame-level synchronization causes displacement measurements to shift temporally, either to the right or left (Fig. 12 (a)). This shift can significantly impact the accuracy of torsion and rigid body rotation measurements, distorting the overall system analysis. Furthermore, improper synchronization may introduce systematic errors that compromise the reliability of predictions related to dynamic behavior under varying conditions. The displacement results of the proposed method were compared to both frame-level synchronization and the displacement results of the industrial motion capture system. displacement results. Results show that the proposed method improved the frame level synchronization up to 0.17 mm (80%), and the displacements for the proposed method are well matched with the displacement results of the industrial level measurement, as shown in Fig. 12 (b).

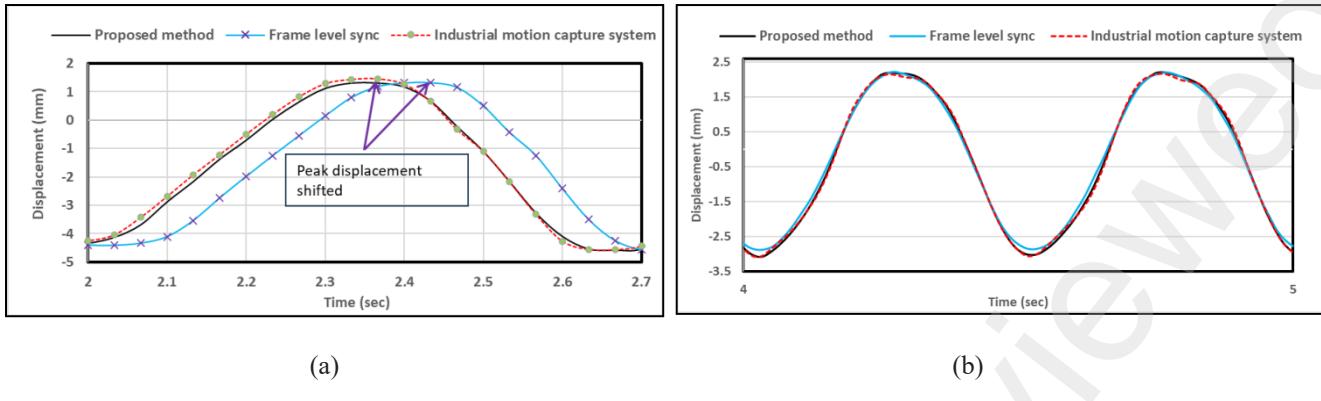


Fig. 12. Accuracy comparison between the proposed method and other methods: (a) displacement result shifted to the right due to frame level synchronization, (b) displacement measurement improvement using the proposed method.

The displacement results from the shake table, measured using the proposed method, were compared to those from the Opti-Track system to evaluate synchronization accuracy. The Opti-Track results served as validation. A dial gauge was employed alongside the Opti-Track system. The accuracy assessment was conducted for different cases of shake table frequency. 0.1 Hz, 0.5 Hz, 1.0 Hz, and 2.0 Hz shake table frequencies were used to assess the effect of object speed on the measuring ability of the developed software. The frequency and amplitude of the shaking table were constant throughout the video recording period. The maximum relative displacement measurement deviations observed in the comparison between the proposed method and Opti-Track were 0.038 mm in the X direction and 0.097 mm in the Z direction. The Y-axis is the direction where the shake table moves back and forth. In this direction, an additional validation tool, a dial gauge, was used. The maximum error recorded was 0.10 mm and 0.20 mm in comparison with Opti-Track and dial gauge, respectively. The summary of relative displacement measurements is presented in Table 1.

Table1: displacement measurement results comparison of the shake table experiment

Shake Table frequency	Measurement method	Average		Average		Average	
		relative	Difference	relative	Difference	relative	Difference
		displacement		displacement		displacement	
		(Ux)		(Uy)		(Uz)	
Industrial motion capture		-/+ 0.075 mm		-/+ 2.90mm	0.05 mm	-/+0.06 mm	
0.1 Hz	system		0.05 mm			0.025 mm	

	Proposed	-/+ 0.125 mm	-/+ 2.95mm	-/+ 0.085 mm	
	method				
	Dial-Gauge		-/+ 2.80 mm	0.15 mm	
	Industrial	-/+ 0.095 mm	-/+ 2.90 mm	0.10 mm	-/+ 0.065 mm
	motion capture				
0.5 Hz	system	0.038 mm			
	Proposed	-/+ 0.133 mm	-/+ 3.00 mm	-/+ 0.125 mm	0.06 mm
	method				
	Dial-Gauge		-/+ 2.90 mm	0.10 mm	
	Industrial	-/+ 0.10 mm	-/+ 2.95 mm	0.05 mm	-/+ 0.048 mm
	motion capture				
1.0 Hz	system	0.05 mm			0.097 mm
	Proposed	-/+ 0.15 mm	-/+ 3.05 mm	-/+ 0.145 mm	
	method				
	Dial-Gauge		-/+ 2.90 mm	0.15 mm	
	Industrial	-/+ 0.125 mm	-/+ 3.20 mm	0.15 mm	-/+ 0.190 mm
2.0 Hz	motion capture				
	system	0.03 mm			0.005 mm
	Proposed	-/+ 0.155 mm	-/+ 3.20 mm	-/+ 0.195 mm	
	method				
	Dial-Gauge		-/+ 3.00 mm	0.20 mm	

According to the manufacturers specification of the shake table used for this experiment the average relative displacement in the Y- direction is -/+ 3.0mm this means the measurement error of Proposed method is 0.20mm. Fig. 13 (a, b, and c) shows sample displacement time series graph for 1hz table frequency of point 1 for three degree of freedom motion.

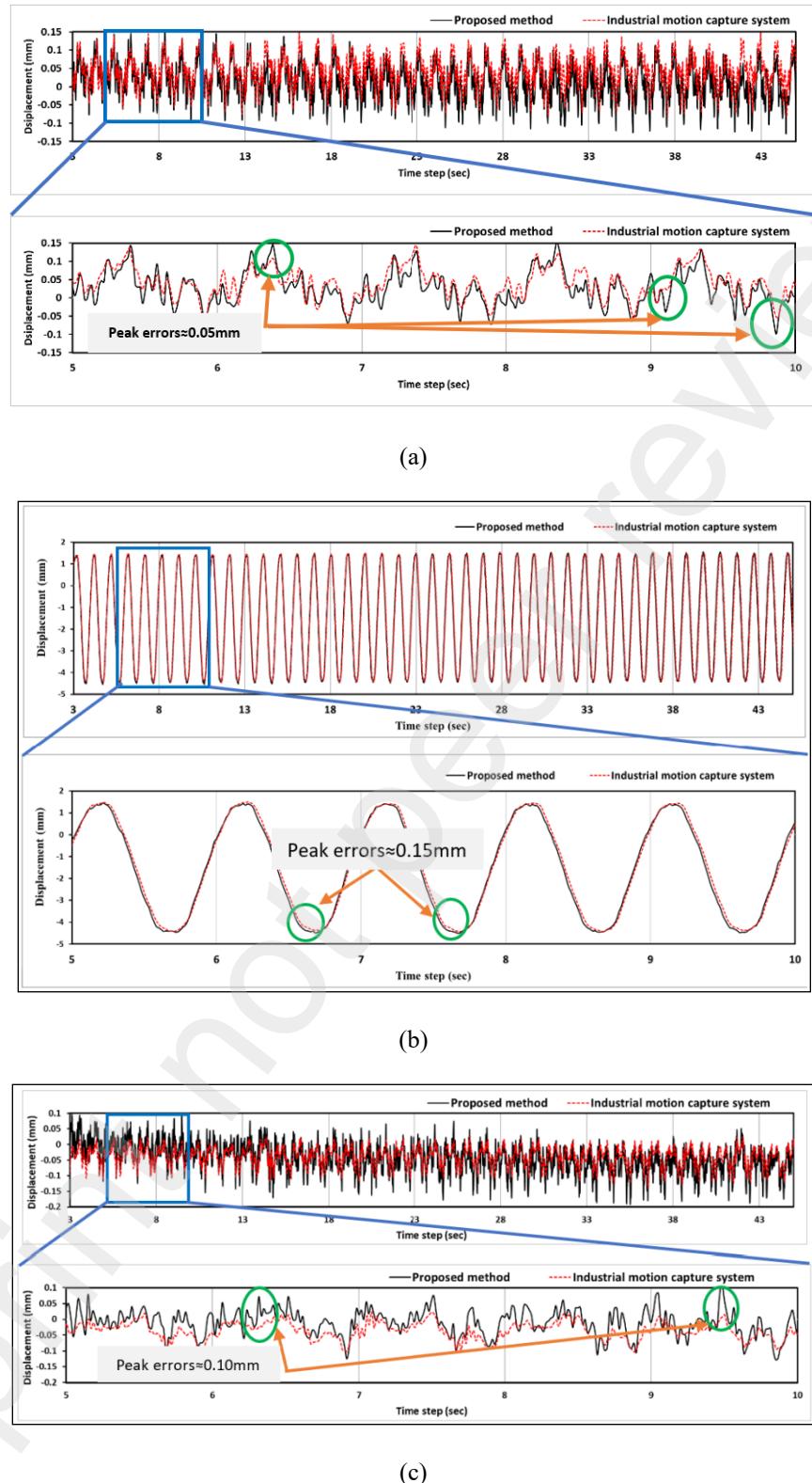
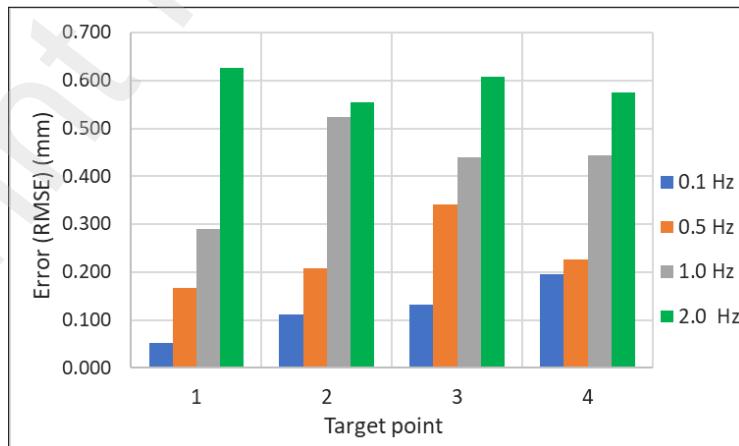


Fig. 13. Time series displacement graph comparison: (a) x-direction, (b) y-direction, (c) z-direction.

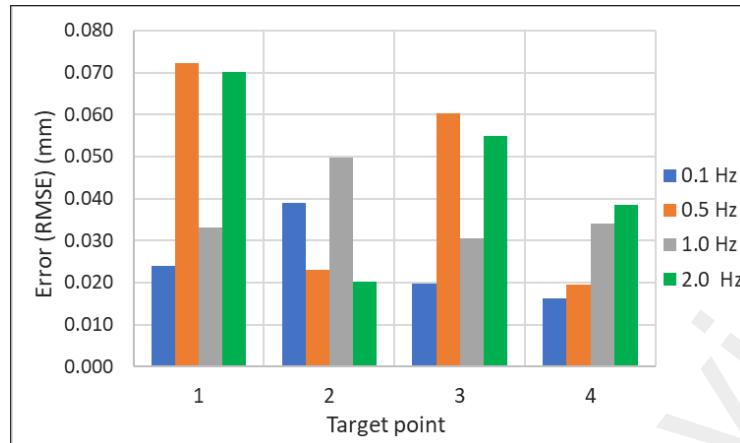
An additional experiment was conducted to validate the accuracy of the DIC measurement software on a bridge system connected to a shaking table (experiment type (c)), as shown in Fig. 3. The 3D displacement errors were computed from a synchronized time series displacement graph of the Industrial motion capture system and Stereo-sync using the RMSE method by considering the time series displacement result of the Industrial motion capture system as the observed value and Stereo-sync as the predicted value. Fig. 14 shows the RMSE results of four targets for 0.1 Hz, 0.5 Hz, 1.0 Hz, and 2.0 Hz shake table frequency. The maximum estimated errors were 0.082 mm in the X direction, 0.626 mm in the Y direction, and 0.072 mm in the Z direction. The accuracy of the measurement was affected by the frequency of the shake table, especially for the Y-axis, which is the main direction of motion. The measurement error increases as the speed of the shake table increases. For the X and Z axes, the effect is minimal because, in these directions, the movement of the specimen is almost stationary with measurement results majorly affected by noise.



(a)



(b)



(c)

Fig. 14: accuracy assessment of the proposed method for different shake table frequencies: (a) X-direction, (b) Y-direction, (c) Z-direction.

Additionally, the experimental results of the bridge vibration measurement also indicate that the proposed approach has the capability to capture high-frequency vibrations with a very small displacement, as shown in Fig. 15. The bridge undergoes a high-frequency internal vibration that is 80 times larger than the shake table frequency at some instants. The estimated frequency of the bridge vibration is approximately 8.0Hz for 0.1Hz shake table frequency, and the maximum amplitude is approximately 0.1 mm at that specific time range, as shown in Fig. 15. The average and maximum difference in displacement measurement between the Proposed method and Industrial motion capture system for this particular phenomenon is 0.0471mm and 0.108mm respectively.

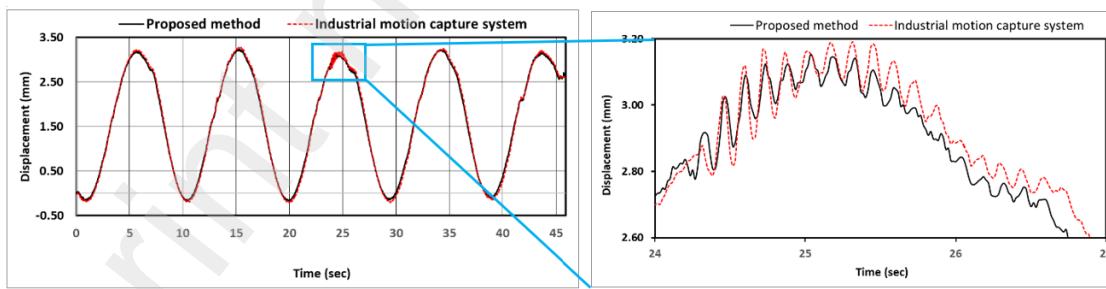


Fig. 15. Accuracy of the proposed method at higher vibration frequencies.

The proposed method provides a reliable way to observe and analyze the structural response of a system under varying excitation conditions. Figure 16 illustrates the time history of the structural vibration of the bridge system, captured using the proposed method alongside the Industrial motion capture system motion capture system for comparison. The results reveal that

the structural frequency of the bridge system increases proportionally as the excitation frequency is raised. Specifically, Figure 15(a) demonstrates that the structural vibration exhibits a lower frequency when the shake table operates at 0.5 Hz. Conversely, Figure 15(b) shows a higher structural vibration frequency corresponding to an increased shake table frequency of 1 Hz. These observations highlight the capability of the proposed method to accurately track dynamic changes in structural behavior across different excitation scenarios.

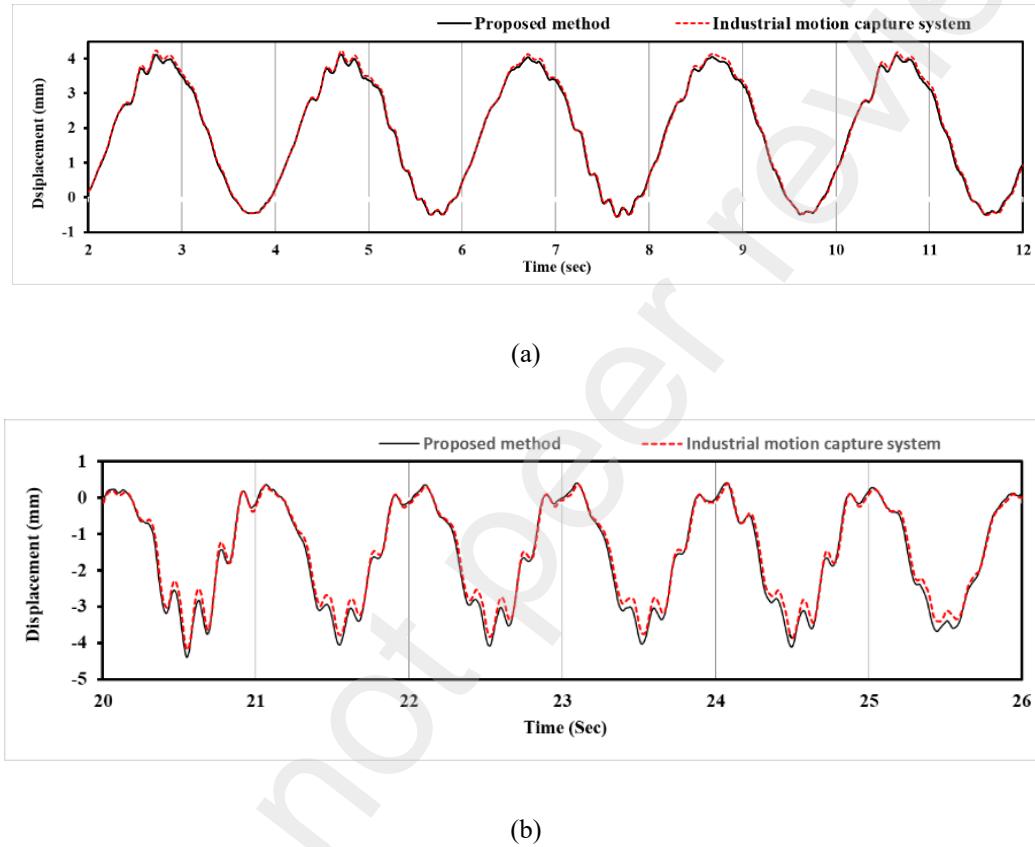


Fig. 16. Bridge vibration response comparison between proposed method and Industrial motion capture system: (a) 0.5 Hz, (b) 1.0 Hz.

4. Summary and conclusions

In this paper, we have developed a stereo synchronization method and application software for image-based dynamic three-dimensional displacement measurement for unsynchronized consumer-level cameras. The prototyped software includes a single-image camera calibration method and synchronization method with subframe-level accuracy, and it can triangulate the 3D position of multiple target points with reasonable accuracy. The synchronization method determines the time lag between the two videos by minimizing triangulation errors. The software developed by this study has the potential to support

conventional measurement methods by providing fairly accurate data at a reasonable cost. Application software specifically prototyped for image-based structural measurements is limited. This software can contribute its share to the structural engineering field by providing a cheap, suitable, and precise platform for dynamic response analysis of structural elements. The software will be available to all users at no cost.

Various experiments were conducted to verify the proposed synchronization approach and the developed 3D displacement analysis software. In the free vibrating shaking table test, used to verify the synchronization method, the maximum triangulation error of all points (20 points) was 0.663 pixels. The maximum triangulation RMSE minimization of all points achieved was 0.02 pixels (lowered by 9.67%). The results demonstrate that the proposed synchronization method can synchronize two video outputs of the same dynamic scene with subframe accuracy (less than one frame). It is also possible to observe the time shift between targets due to the rolling shutter effect using the proposed method.

Experiments using a shaking table and a bridge specimen were conducted with four frequency scales (0.10, 0.50, 1.0, and 2.0 Hz) to verify the accuracy of the developed image-based 3D displacement measurement tool. The maximum error in relative displacement in the main direction of motion (Y-axis) when comparing the proposed method and Industrial motion capture system was 0.10 mm, and it was 0.20 mm when compared with the results of the dial gauge. The maximum deviation in relative displacement in the X (horizontal) and Z (vertical) axes was 0.038 mm and 0.097 mm, respectively. In the bridge system experiment, the maximum errors when comparing the proposed method and Industrial motion capture system were 0.082 mm in width, 0.626 mm in depth, and 0.072 mm in height.

In summary, this study developed application software for image-based 3D displacement analysis based on single image calibration and a new synchronization approach for shake table experiments. The cameras are assumed to be vibration-free. However, in actual shake table tests, cameras may vibrate due to wind flowing into the testing area, ground vibration caused by the shaking table itself, and due to human interventions. Moreover, additional studies are needed to reduce errors further. Additionally, errors may occur due to the rolling shutter effect and camera frame rate inconsistency; developing a method to compensate for these errors and further checking the accuracy assessment based on actual large-scale experimental tests with complex motions different from the experiment used in this study are needed.

References

- [1] D. Feng and M. Q. Feng, "Identification of structural stiffness and excitation forces in time domain using noncontact vision-based displacement measurement," *Journal of Sound and Vibration*, vol. 406, pp. 15–28, 2017.
- [2] J. Baqersad, P. Poozesh, C. Nizrecki, and P. Avitabile, "Photogrammetry and optical methods in structural dynamics—A review," *Mechanical Systems and Signal Processing*, vol. 86, pp. 17–34, 2017.
- [3] J. Lv, P. He, X. Hou, J. Xiao, L. Wen, and M. Lv, "A target-free vision-based method for out-of-plane vibration measurement using projection speckle and camera self-calibration technology," *Engineering Structures*, vol. 303, p. 117416, 2024.
- [4] Y. Shao, L. Li, J. Li, Q. Li, S. An, and H. Hao, "3D displacement measurement using a single-camera and mesh deformation neural network," *Engineering Structures*, vol. 318, p. 118767, 2024.
- [5] J.-F. Lin, Y.-L. Xu, and S. Zhan, "Experimental investigation on multi-objective multi-type sensor optimal placement for structural damage detection," *Structural Health Monitoring*, vol. 18, no. 3, pp. 882–901, May 2019, doi: 10.1177/1475921718785182.
- [6] Z. Ma, J. Choi, and H. Sohn, "Three-dimensional structural displacement estimation by fusing monocular camera and accelerometer using adaptive multi-rate Kalman filter," *Engineering Structures*, vol. 292, p. 116535, 2023.
- [7] C.-X. Qu, J.-Z. Jiang, T.-H. Yi, and H.-N. Li, "Computer vision-based 3D coordinate acquisition of surface feature points of building structures," *Engineering Structures*, vol. 300, p. 117212, 2024.
- [8] D. Feng and M. Q. Feng, "Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection—A review," *Engineering Structures*, vol. 156, pp. 105–117, 2018.
- [9] L. Ngeljaratan and M. A. Moustafa, "Digital image correlation for dynamic shake table test measurements," in *7th International Conference on Advances in Experimental Structural Engineering (7AESE)*, September, 2017, pp. 6–8. Accessed: Dec. 16, 2024. [Online]. Available: https://www.researchgate.net/profile/Mohamed-Moustafa-22/publication/326016564_Digital_Image_Correlation_for_Dynamic_Shake_Table_Test_Measurements/links/5f01492545851550508d84ee/Digital-Image-Correlation-for-Dynamic-Shake-Table-Test-Measurements.pdf
- [10] Z. R. Wani, M. Tantry, and E. N. Farsangi, "In-Plane measurements using a novel streamed digital image correlation for shake table test of steel structures controlled with MR dampers," *Engineering Structures*, vol. 256, p. 113998, 2022.
- [11] Y. Sieffert, F. Vieux-Champagne, S. Grange, P. Garnier, J. C. Duccini, and L. Daudeville, "Full-field measurement with a digital image correlation analysis of a shake table test on a timber-framed structure filled with stones and earth," *Engineering Structures*, vol. 123, pp. 451–472, 2016.

- [12] P. Bing, X. Hui-Min, X. Bo-Qin, and D. Fu-Long, “Performance of sub-pixel registration algorithms in digital image correlation,” *Measurement Science and Technology*, vol. 17, no. 6, p. 1615, 2006.
- [13] M. A. Sutton, J. H. Yan, V. Tiwari, H. W. Schreier, and J.-J. Orteu, “The effect of out-of-plane motion on 2D and 3D digital image correlation measurements,” *Optics and Lasers in Engineering*, vol. 46, no. 10, pp. 746–757, 2008.
- [14] R. Kromanis and C. Forbes, “A low-cost robotic camera system for accurate collection of structural response,” *Inventions*, vol. 4, no. 3, p. 47, 2019.
- [15] Y.-S. Yang, “Measurement of dynamic responses from large structural tests by analyzing non-synchronized videos,” *Sensors*, vol. 19, no. 16, p. 3520, 2019.
- [16] E. Zappa and N. Hasheminejad, “Digital image correlation technique in dynamic applications on deformable targets,” *Exp Tech*, vol. 41, no. 4, pp. 377–387, Aug. 2017, doi: 10.1007/s40799-017-0184-3.
- [17] S. Das and P. Saha, “A review of some advanced sensors used for health diagnosis of civil engineering structures,” *Measurement*, vol. 129, pp. 68–90, 2018.
- [18] J. S. Lyons, J. Liu, and M. A. Sutton, “High-temperature deformation measurements using digital-image correlation,” *Experimental Mechanics*, vol. 36, no. 1, pp. 64–70, Mar. 1996, doi: 10.1007/BF02328699.
- [19] M. A. Michael A., J.-J. Orteu, and H. W. Schreier, “Practical considerations for accurate measurements with DIC,” in *Image Correlation for Shape, Motion and Deformation Measurements*, Boston, MA: Springer US, 2009, pp. 1–24. doi: 10.1007/978-0-387-78747-3_10.
- [20] K. Malowany, M. Malesa, T. Kowaluk, and M. Kujawinska, “Multi-camera digital image correlation method with distributed fields of view,” *Optics and Lasers in Engineering*, vol. 98, pp. 198–204, 2017.
- [21] Z. Zhang, B. Pan, M. Grédiac, and W. Song, “Accuracy-enhanced constitutive parameter identification using virtual fields method and special stereo-digital image correlation,” *Optics and Lasers in Engineering*, vol. 103, pp. 55–64, 2018.
- [22] F. Bottalico, C. Niezrecki, and A. Sabato, “Next Generation 3D-DIC technique with sensor-based extrinsic parameter calibration and natural pattern tracking,” *Structural Health Monitoring 2023*, 2023, Accessed: Dec. 03, 2024. [Online]. Available: <https://dpi-proceedings.com.destechpub.a2hosted.com/index.php/shm2023/article/view/36867>
- [23] X. Li, X. Zhang, Z. Wang, D. Zheng, and P. Li, “Experimental study with DIC technique on the bond failure of structural concrete under repeated reversed loading with different amplitude,” *Advances in Structural Engineering*, vol. 26, no. 16, pp. 3089–3111, Dec. 2023, doi: 10.1177/13694332231208261.

- [24] G. Lavezzi, M. Ciarcià, K. Won, and M. Tazav, “A DIC-UAV based displacement measurement technique for bridge field testing,” *Engineering Structures*, vol. 308, p. 117951, 2024.
- [25] K. Wei, F. Yuan, X. Shao, Z. Chen, G. Wu, and X. He, “High-speed multi-camera 3D DIC measurement of the deformation of cassette structure with large shaking table,” *Mechanical Systems and Signal Processing*, vol. 177, p. 109273, 2022.
- [26] T. Pribanic, A. Hedi, and V. Gracanin, “Harmonic oscillations modelling for the purpose of camera synchronization,” in *International Conference on Computer Graphics Theory and Applications*, SCITEPRESS, 2012, pp. 202–205. Accessed: Dec. 03, 2024. [Online]. Available: <https://www.scitepress.org/PublishedPapers/2012/39277/>
- [27] F. Nikfar and D. Konstantinidis, “Evaluation of vision-based measurements for shake-table testing of nonstructural components,” *J. Comput. Civ. Eng.*, vol. 31, no. 2, p. 04016050, Mar. 2017, doi: 10.1061/(ASCE)CP.1943-5487.0000615.
- [28] S. Ansari, N. Wadhwa, R. Garg, and J. Chen, “Wireless software synchronization of multiple distributed cameras,” in *2019 IEEE International Conference on Computational Photography (ICCP)*, IEEE, 2019, pp. 1–9. Accessed: Dec. 03, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8747340/>
- [29] “OptiTrack Documentation | EXTERNAL OptiTrack Documentation.” Accessed: Dec. 25, 2024. [Online]. Available: <https://docs.optitrack.com>
- [30] J. S. Furtado, H. H. T. Liu, G. Lai, H. Lacheray, and J. Desouza-Coelho, “Comparative analysis of OptiTrack motion capture systems,” in *Advances in Motion Sensing and Control for Robotic Applications*, F. Janabi-Sharifi and W. Melek, Eds., in Lecture Notes in Mechanical Engineering. , Cham: Springer International Publishing, 2019, pp. 15–31. doi: 10.1007/978-3-030-17369-2_2.
- [31] C. Hansen, D. Gibas, J.-L. Honeine, N. Rezzoug, P. Gorce, and B. Isableu, “An inexpensive solution for motion analysis,” *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, vol. 228, no. 3, pp. 165–170, Sep. 2014, doi: 10.1177/1754337114526868.
- [32] M. N. Helfrick, C. Niezrecki, P. Avitabile, and T. Schmidt, “3D digital image correlation methods for full-field vibration measurement,” *Mechanical systems and signal processing*, vol. 25, no. 3, pp. 917–927, 2011.
- [33] H.-T. Chen, “Geometry-based camera calibration using five-point correspondences from a single image,” *IEEE Transactions on Circuits and systems for Video Technology*, vol. 27, no. 12, pp. 2555–2566, 2016.

- [34] Z. Zhang, “A flexible new technique for camera calibration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [35] A. Kaehler and G. Bradski, *Learning OpenCV 3: computer vision in C++ with the OpenCV library*. O’Reilly Media, Inc., 2016. Accessed: Dec. 16, 2024. [Online]. Available: https://books.google.com/books?hl=en&lr=&id=LPm3DQAAQBAJ&oi=fnd&pg=PP1&dq=computer+vision+in+C%2B%2B+with+the+OpenCV+library&ots=2xKmUa9hC8&sig=w-MxAfIvnBMFLyE_0gfXHv4tmxg
- [36] F. Chen, X. Chen, X. Xie, X. Feng, and L. Yang, “Full-field 3D measurement using multi-camera digital image correlation system,” *Optics and Lasers in Engineering*, vol. 51, no. 9, pp. 1044–1052, 2013.
- [37] Y.-S. Yang, C.-H. Chang, and C. Wu, “Damage indexing method for shear critical tubular reinforced concrete structures based on crack image analysis,” *Sensors*, vol. 19, no. 19, p. 4304, 2019.
- [38] L. Yang, M. Li, X. Song, Z. Xiong, C. Hou, and B. Qu, “Vehicle speed measurement based on binocular stereovision system,” *IEEE Access*, vol. 7, pp. 106628–106641, 2019.
- [39] Z. Zhang, “Flexible camera calibration by viewing a plane from unknown orientations,” in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, IEEE, 1999, pp. 666–673. Accessed: Dec. 16, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/791289/>
- [40] A. Mordvintsev and K. Abid, “Opencv-python tutorials documentation. Obtenido de.” 2022.
- [41] M. Shimizu, “Simple triangulation for asynchronous stereo cameras,” in *2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, IEEE, 2019, pp. 269–274. Accessed: Dec. 16, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8977741/>
- [42] L. Oth, P. Furgale, L. Kneip, and R. Siegwart, “Rolling shutter camera calibration,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1360–1367. Accessed: Dec. 16, 2024. [Online]. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2013/html/Oth_Rolling_Shutter_Camera_2013_CVPR_paper.html
- [43] B. Fan, Y. Dai, and M. He, “Rolling shutter camera: Modeling, optimization and learning,” *Machine Intelligence Research*, vol. 20, no. 6, pp. 783–798, 2023.