11th International Operational Modal Analysis Conference
20 - 23 May 2025 | Rennes, France

Multi-target computer-vision dynamic monitoring of displacements in a full-scale building

Laura Gioiella[1], Fabio Micozzi[1], Morgan McBain[2], Michele Morici[1], Alessandro Zona[1], Andrea Dall'Asta[1], Barbara G. Simpson[2] and Andre R. Barbosa[3]

[1] University of Camerino, Italy, {laura.gioiella, fabio.micozzi, michele.morici, alessandro.zona, andrea.dallasta}@unicam.it
[2] Stanford University, USA, {mmcbain, bsimpson}@stanford.edu
[3] Oregon State University, USA, andre.barbosa@oregonstate.edu

ABSTRACT

Computer-vision methods for structural monitoring have attracted much attention in recent years thanks to the possibility of measuring an unprecedented number of displacement time histories from a single sensor in the absence of stationary points close to the structure, using hardware with accessible costs that is simple to set up and operate by structural engineers with basic video photography knowledge and programming skills. Past experimental studies showed the potentialities of computer-vision algorithms to provide accurate results that can be equivalent to those obtained from accelerometers and displacement transducers, even when very small displacements are experienced, as in the case of stiff structures and/or low load levels typical of operational conditions. Such achievement is made possible thanks to the development in the last decade of sophisticated and computationally efficient computer-vision algorithms able to achieve resolutions that are a small fraction of the dimension of a single pixel in the captured video footages, while being able to track large displacements. This opens unexplored territories in structural identification and condition assessment of civil structures, freeing structural engineers from the inevitable limitations imposed by a reduced number of pointwise contact sensors, as is commonly the case in many temporary or permanent structural monitoring setups. Accordingly, the objective of this article is to provide a critical discussion of the potential benefits of computer-vision structural monitoring applied to a full-scale building where the use of traditional contact sensors faces major difficulties and limitations. Insight is given to the extraction of the time-history of displacements, with compensation of vibrations and external disturbances in the video camera.

Keywords: Computer-Vision, Displacement Measurements, Modal Analysis, Structural Monitoring.

1. INTRODUCTION

In dynamic testing of civil constructions, as for example is the case of shake table testing in earthquake engineering [1][2], displacement measurements are crucial for assessing the structural performance. Commonly displacements are obtained using contact sensors [3] such as Linear Variable Displacement Transducers (LVDTs) and/or String-Potentiometers (SPs), or double integration of accelerometer data [4]. LVDTs and SPs need stationary points close to the specimen, usually to be realized to this purpose, thus, affecting budget. The use of accelerometers requires double integration, implying several delicate operations that may induce errors in the displacement estimates [5][6] even though baseline corrections and filtering procedures are applied. Moreover, the use of contact sensors requires their use in large amounts, which means also time-consuming operations related to cables running. GPS-based monitoring has also been attempted but lacks sufficient accuracy [7]. Computer-vision [8] emerged as an effective alternative for displacement monitoring. While mostly used for bridges [8][9], its application in buildings grew [10]-[20], especially in the last five years. Traditional vision-based approaches rely on external cameras [10]-[14] or Unmanned Aerial Vehicles (UAVs) [16], requiring multiple cameras, large data storage, synchronization issues, compensation of the UAV own movement, making real-time processing difficult. Internal cameras face challenges with motion-induced disturbances [17]-[20].

This study introduces a novel, simple vision-based methodology combining internal and external cameras installed, respectively, at the roof edge and on the ground (fixed-base) to measure horizontal displacements in multi-story buildings during shake table testing. The roof camera captures floor displacements by recording only the movements of dedicated targets located throughout the building height, while the ground camera provides redundancy for the top-level motion and improves accuracy. The second acquisition of roof displacements provided by the ground camera is useful to compensate for possible sources of noise (vibrations) induced by the playback of the input in the video camera installed on the roof. The proposed methodology was validated through shake table testing of a six-story mass timber structure at University California at San Diego (UCSD) 6-DOFs Large High Performance Outdoor Shake Table (LHPOST6) facility, showing strong potential as a primary displacement monitoring tool. However, the proposed methodology is not confined to shake table testing, since it can cover different monitoring needs in the dynamic response of multi-story buildings, as is for example the case of push-and-release tests [21].

2. PROPOSED VISION-BASED METHODOLOGY

The proposed vision-based methodology can be applied either in planar configuration with only two cameras, or in three-dimensional one with triplets of cameras. In both cases the methodology exploits the Upsampling Cross-Correlation (UCC) algorithm [22] for the extraction of displacements, that can achieve sub-pixel resolution (indicatively up to 1/100 of a pixel) when used with high-contrast targets, as those using a chessboard square shape with different patterns and dimensions [23][24].

2.1. Planar configuration

It is possible to have an insight into the proposed methodology when applied into its planar configuration working with only two cameras, considering a four-story building subjected to a monodirectional input (Figure 1). The internal camera is installed on the roof corner looking downward. The camera acquires the motion of targets T0-T3 (red lines in Figure 1) designed and located so that they do not overlap during motion. The same camera can catch also the shake table motion as difference between the displacements T0 and T1, providing the chance to compare it with the controller response. The second camera, the external one, is located on the reaction mass at ground floor (L0), considered as a fixed base. Such a camera acquires only roof displacement in the proximity of the internal camera through the movement of the target T4 installed nearby. The roof (L4) displacement is acquired twice, from the external camera by pointing at T4 and from the internal one, by following T0. In the hypothetical condition of no noise, the top floor displacements should coincide. If differences are detected they descend from the vibrations and rotations induced on the roof camera. The redundancy of cameras allows correcting displacements through the subtraction of the motion of T0 from T4.

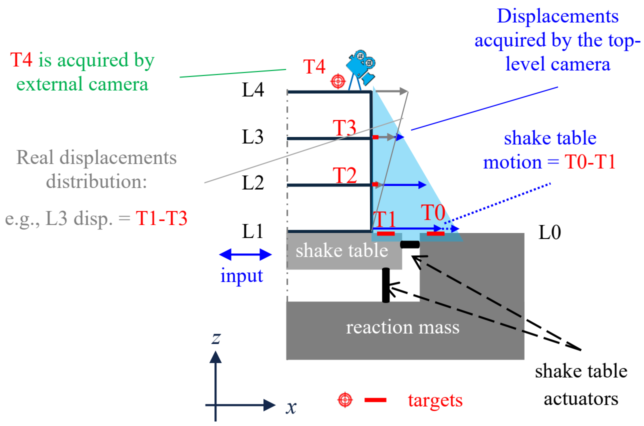

Figure 1. Basics of the proposed methodology during a shake table monodirectional input.

2.2. Three-dimensional configuration

The proposed setup can be extended to capture movement in three dimensions. For this aim two sensors are installed on opposite corners of the building roof to track multi-level horizontal displacements along the global x and y directions, as well as the motion of the shake table. Placing the cameras at the opposing corners enhances data accuracy, particularly in the event of floor rotation, assuming the floors behave as rigid diaphragms. This arrangement allows for duplicate displacement measurements at the greatest possible horizontal distance. The two roof cameras, directed downward, acquire the movement of multiple targets, which are strategically sized and positioned along the building height to prevent overlapping during ground motions playback. In addition to the internal cameras, the proposed vision-based system is completed by four additional external cameras, which serve a dual function: they provide redundancy by also recording roof-level horizontal displacements and are useful to mitigate data noise. The acquisition of roof displacements by external cameras is achieved thanks to two targets, one per each global axis x and y, positioned near each internal camera.

3. APPLICATION

3.1. Testbed building and camera setup

During phase II shake table testing of the NHERI Converging Design Project [25][26] the first application of the proposed vision-based methodology took place. The specimen is a six-story mass timber building (Figure 2a) whose structural layout consists of beams and columns with pinned joints that sustain gravitational loads, while the resistance to horizontal actions is ensured by four perimetral post-tensioned rocking walls, detailed in turns with three different lateral force resisting systems in the lateral (*y* North-South) direction, each of which was tested during one of the three phases of the project. During phase II the lateral resisting system consisted of Buckling-Restrained Braces (BRBs).

The objective of the vision-based methodology was the measurement of the displacements in the lateral direction (*y*) of the building with only two video cameras, one internal and the other external. The internal roof camera was installed in the North-East corner of the roof (L7) pointing downward. The external ground camera was in-stalled at nearly 26 m from the East side of the building at the ground level (L0). The field of view (FOV) of each camera was sized considering different factors: *i*) optical hardware specifications; *ii*) inter-story heights of the building that coincide with the distances from the sensor; *iii*) upsampling factor (assumed equal to 100) for the UCC algorithm; *iv*) magnitude of the expected displacements; *v*) size and number of targets based on the requirements of the previous factors; *vi*) capacity of the adopted laptop to manage the stream of data from the camera sensor.

Two video cameras (Teledyne Blackfly S BFS-U3-23S3M-C) connected each to a laptop through USB3.0 were used for the application. The roof camera was equipped with a lens having 16 mm focal length. The FOV was sized 800 × 700 pixels (width × height) to acquire the displacements of targets located at levels L0 (reaction mass, size 200 mm × 200 mm), L1 (shake table, size 200 mm × 200 mm), L2 (height of +4.32 m above ground, size 100 mm × 100 mm), L3 (+7.67 m, size 100 mm × 100 mm) and L5 (+14.38 m, size 50 mm × 50 mm). The displacements of the levels L0 and L1 are, respectively, the displacement of the roof with and without the shake table motion. Since the ground camera acquires only the lateral *y* roof displacement for providing redundancy and control of the eventual noise, a FOV of 900 × 300 pixels with a single target (T7 100 mm × 100 mm) located on the thickness of roof floor (+21.08 m) was designed, while the lens used has 50 mm focal length. Figure 2b) and Figure 2c) show, respectively, the FOV and targets used for the roof and for the ground camera. The acquisitions reported in this study were recorded at 100 frames per second as a balanced compromise between hardware video recording limits, dimension of the resulting video and expected performances.

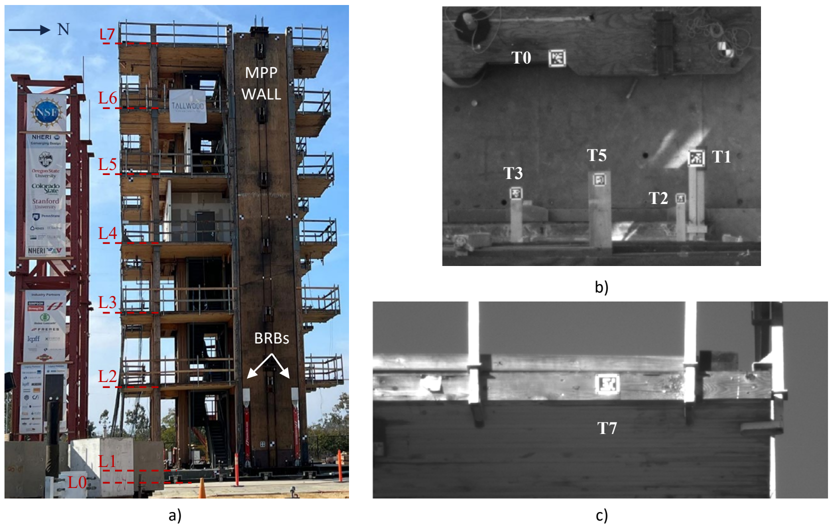

Figure 2. Testbed structure (a), view from the roof camera (b), and from the ground camera (c).

3.2. Selected results

The results refer to a 5-minute ambient vibration (AV) test performed at very beginning of the test sequence and the first white noise performed in the lateral *y* direction (WNY). Figure 3 shows the comparison of the frequency contents between the signals of displacements acquired by the roof (T0 and T1) and the ground camera (T7) with reference to the AV test (Figure3a) and to the WNY (Figure3b) underlining that the analysed signals are characterized by different energy contents.

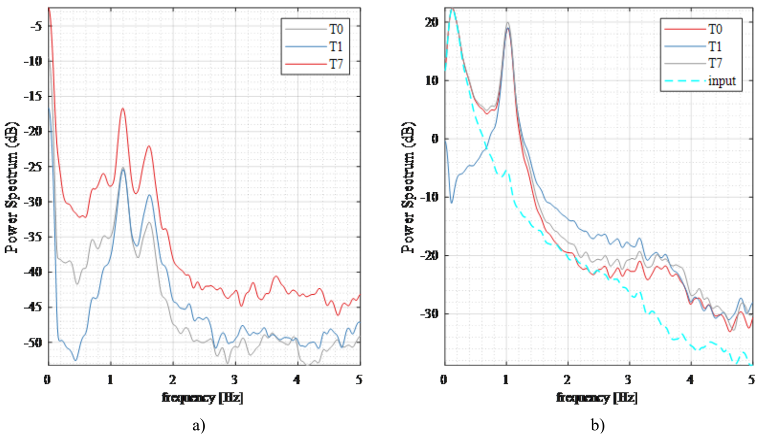

Figure 3. Fundamental frequencies identified in the range 0 Hz – 5 Hz during a) AV test and b) WNY test.

It is worth noting that the cameras can identify the frequencies even though the absolute displacements achieved at the roof level with respect to the ground are significantly different between the two tests, nearly 3 mm for the AV test and nearly 80 mm for the WNY. The figures analyse the signals in the range 0 Hz-5 Hz. For the AV test there is a first frequency at nearly 1.19 Hz and a second one at nearly

1.61 Hz, while during the WNY the first frequency becomes nearly 1.02 Hz. The shift of the first frequency is related to a lower contribution in terms of stiffness of non-structural components during the WNY and above all to the rocking behaviour of the walls that do not occur during an AV test. Form the power spectrum of the WNY excitation it is highlighted only one of the two frequencies identified during the AV test. This aspect can be explained by the fact that the WNY excitation involves only the lateral Y direction consequently it is possible to identify just the frequency of the vibration mode belonging to that direction. Differently, the AV test can highlight the fundamental frequencies belonging to vibration modes of both the horizontal directions X and Y. An interesting aspect is related to the difference in the frequency content between T0 and T7 from one side and T1 in the other (Figure 3b). Since both T0 and T7 acquire the shake table motion, they can also show the frequency content of the input, reported in cyan dotted line.

The Wavelet synchrosqueezed transform [27] provides a deeper insight into the trend of the identified frequencies in the time-frequency plane. Figure 4 shows the frequencies identified in the analyzed tests with reference to the signal acquired by the ground camera through target T7. It is worth noting that for both tests the first frequency, 1.19 Hz (AV) and 1.02 Hz (WNY), highlighted through a red dashed line, remains constant over time.

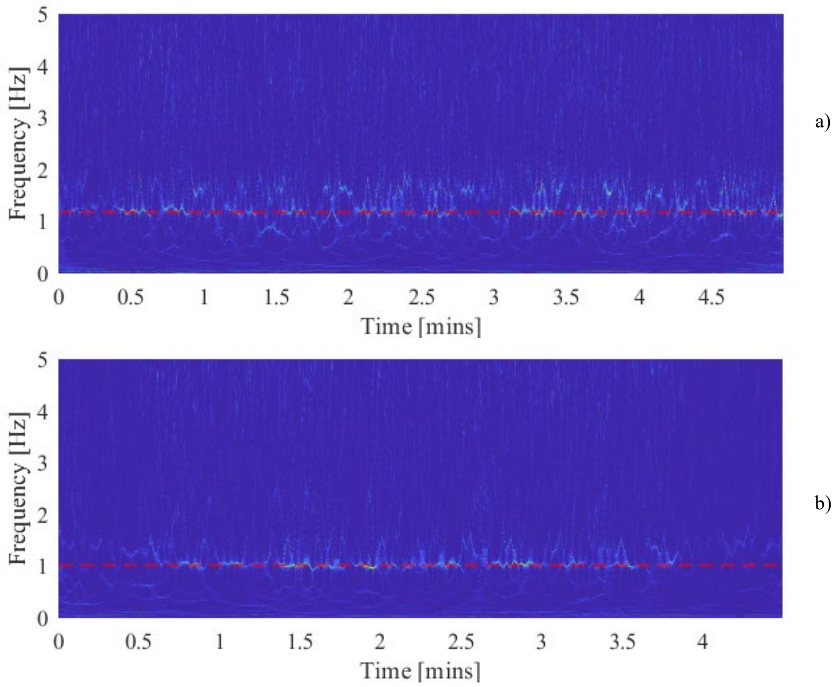

Figure 4. Wavelet synchrosqueezed transform of the T7 signal during a) AV test and b) WNY test.

The roof lateral displacement derived from the roof camera (T0, red solid line) and the ground camera (T7, grey dash-dot line) are compared in Figure 5. It can be noted that both signals have similar noise, and the peaks are almost perfectly overlapped among them, as pointed out in the close-up.

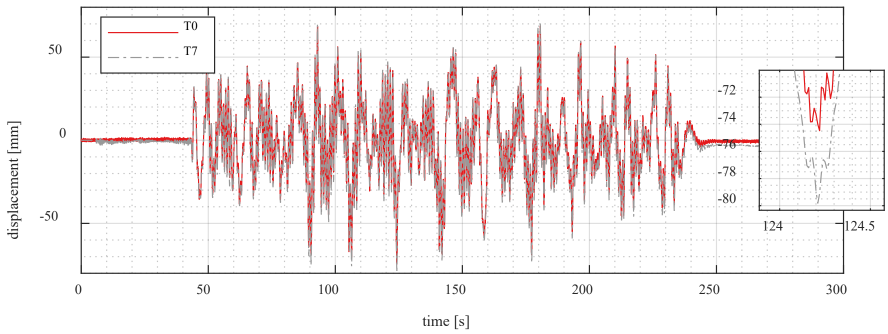

Figure 5. Comparison of the roof lateral displacement from the roof camera (T0) and the ground camera (T7).

The platen lateral displacements measured by the shake table controller (acquired at 512 Hz) and by the roof camera are compared in Figure 6 (blue solid line and cyan dot-solid line, respectively). It is worth noting that, except for the resampling of the recording of the controller at the same frequency of the cameras (100 Hz) and alignment of the signals, no other treatment was done. Given that the displacements extracted from the roof camera are obtained as difference of displacements between two targets (T0 and T1) having nearly the same distance from the camera, the disturbance highlighted in the signal acquired by the roof camera, as shown for the target T0 in Figure 5, are automatically compensated. The peak values (positive and negative) of the platen displacement are reported in Table 1 to quantify the precision of the vision-based acquisition with respect to the controller, as described through the Root Mean Square Error (RMSE). Moreover, it is useful to also introduce the Normalized RMSE (RMSEN) that provides the same information in terms of percentage. The RMSE and RMSEN are, respectively, 0.59 mm and 3.20%. By considering only the percentage errors calculated in correspondence of the highest positive and negative displacements Max Error and min Error, respectively, the values notably decrease to 1.19% and -1.39%.

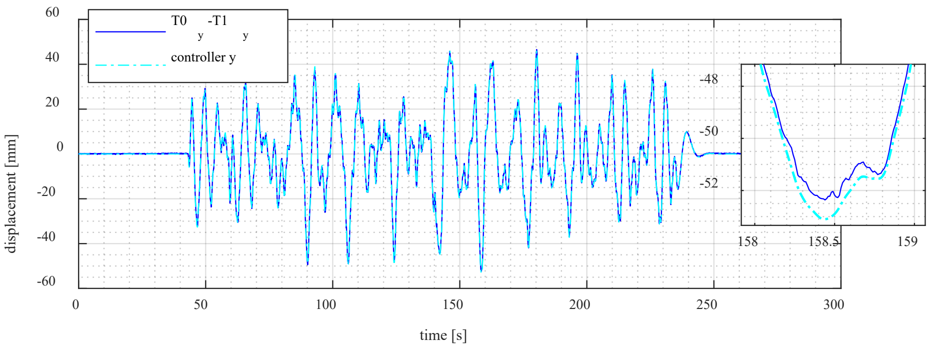

Figure 6. Comparisons of platen displacements measured by the shake table controller and by the roof camera.

Table 1. Statistics of the shake table motion acquired by roof camera and controller.

Source	Max [mm]	min [mm]	RMSE [mm]	RMSEN [%]	Max Error [%]	min Error [%]
roof camera	46.58	-52.36	0.59	3.20	1.19	-1.39
controller	46.03	-53.10				

4. CONCLUSIONS

The application of the proposed vision-based methodology resulted in the following observations:

- The proposed approach exploits the possibilities of computer-vision to monitor displacements using a limited number of video cameras, obtaining very accurate results as compared to the displacements reading of the shake table controller.
- The combined use of internal and external cameras permits compensation for spurious displacements (noise) due to the vibration induced on the internal camera by the playback of the input motion.
- The video cameras can properly estimate the first frequency of the building either during ambient vibration and white noise excitation, which have very different order of magnitude of the absolute displacements at the roof level (nearly 3 mm and nearly 80 mm, respectively).
- The proposed methodology is convenient since it exploits a small number of cameras when compared to the usual number of contact sensors installed during testing operations, with extremely reduced time-consuming operations.
- The proposed methodology exploits cost-effective hardware together with very efficient software for video processing, potentially ready for real-time extraction of displacement time-histories of multiple points within the field of view of the adopted video cameras.

ACKNOWLEDGEMENTS

The Authors wish to thank Joel P. Conte at the University of California San Diego (UCSD) and the staff of the UCSD LHPOST for their support during testing. The Authors of the University of Camerino acknowledge funding by the European Union - NextGenerationEU, Mission 4, Component 2, under the Italian Ministry of University and Research (MUR) National Innovation Ecosystem grant ECS00000041 - VITALITY - CUPJ13C22000430001, the Italian Government, Fondo Complementare Programma unitario di intervento per le aree del terremoto del 2009 e 2016 Misura B - Sub-misura B.4 Centri di ricerca per l'innovazione Centro internazionale per la ricerca sulle Scienze e Tecniche dalla RICostruzione fisica, economica e sociale STRIC e STRIC+", and Fondazione ICSC, "HPC – National Centre for HPC, Big Data and Quantum Computing", Codice CN00000013, Spoke 5 "Environment & Natural Disasters" PNRR, Missione 4, Componente 2, Investimento 1.4 – European Union – NextGenerationEU; CUP H93C22000450007, project FABRICH.

REFERENCES

[1] Van Den Einde L, Conte JP, Restrepo JI, Bustamante R, Halvorson M, Hutchinson TC, Lai CT, Lotfizadeh K, Luco JE, Morrison ML, Mosqueda G, Nemeth M, Ozcelik O, Restrepo S, Rodriguez A, Shing PB, Thoen B, Tsampras G. NHERI@UC San Diego 6-DOF Large High-Performance Outdoor Shake Table Facility. *Frontiers in Built Environment*, 6, 580333, 2021.

[2] Nakashima M, Nagae T, Enokida R, et al. Experiences, accomplishments, lessons, and challenges of E-defense—Tests using world's largest shaking table. *Japan Architectural Review*, 1(1), 4-17, 2018.

[3] Pei S, Ryan KL, Berman JW, Van De Lindt JW, Pryor S, Huang D, Wichman S, Busch A, Roser W, Wynn SL, Ji Y, Hutchinson T, Sorosh S, Zimmerman RB, Dolan J. Shake-Table Testing of a Full-Scale 10-Story Resilient Mass Timber Building. *Journal of Structural Engineering*, 150, 04024183, 2024.

[4] Bock Y, Melgar D, and Crowell BW. Real-time strong-motion broadband displacements from collocated GPS and accelerometers. *Bulletin of the Seismological Society of America*, 101(6), 2904-2925, 2011.

[5] Skolnik DA, Wallace JW. Critical Assessment of Interstory Drift Measurements. *Journal of Structural Engineering*, 136, 1574-1584, 2010.

[6] Dai Z, Li X, Chen S, et al. Baseline correction based on L1-Norm optimization and its verification by a computer vision method. *Soil Dynamics and Earthquake Engineering*, 131, 106047, 2020.

[7] Häberling S, Rothacher M, Zhang Y et al. Assessment of high-rate GPS using a single-axis shake table. *Journal of Geodesy*, 89, 697-709, 2015.

[8] Feng D, Feng MQ. *Computer Vision for Structural Dynamics and Health Monitoring*. Wiley, New York, 2021.

[9] Zona A. Vision-Based Vibration Monitoring of Structures and Infrastructures: An Overview of Recent Applications. *Infrastructures*, 6, 4, 2021.

[10] Ngeljaratan L, Moustafa MA. Structural health monitoring and seismic response assessment of bridge structures using target-tracking digital image correlation. *Engineering Structures*, 213, 110551, 2020.

[11] Wang J, Zaho J, Yuwen L, Shan J. Vision-based displacement and joint rotation tracking of frame structure using feature mix with single consumer-grade camera. *Structural Control and Health Monitoring*, 28(12), e2832, 2021.

[12] Choi I., Kim J, Sohn J. Automated framework for monitoring building structures through derivation of lateral stiffness using marker-free vision-based displacement sensor. *Measurement*, 194, 111062, 2022.

[13] Ji X, Gao X, Zhuang Y, Qu Z. Enhanced measurements of structural inter-story drift responses in shaking table tests. *Engineering Structures*, 278, 115508, 2023.

[14] Pan X, Yang TY, Xiao Y, Yao H, Adeli H (2023) Vision-based real-time structural vibration measurement through deep-learning-based detection and tracking methods. *Engineering Structures*, 281, 115676, 2023.

[15] Sun P., Vasef M, Chen L. Multi-vision-based displacement monitoring using global-local deep deblurring and Rauch-Tung-Striebel smoother. *Measurement*, 242, 116292, 2025.

[16] Wang X, Lo E, De Vivo L, Hutchinson TC, Kuester F. Monitoring the earthquake response of full-scale structures using UAV vision-based techniques. *Structural Control and Health Monitoring*, 29, e2862, 2022

[17] Park JW, Lee JJ, Jung HJ, Myung H. Vision-based displacement measurement method for high-rise building structures using partitioning approach. *NDT & E International*, 43, 642–647, 2010.

[18] Petrone F, Perez R, Coates J, McCallen D. A Biaxial Discrete Diode Position Sensor for Rapid Postevent Structural Damage Assessment. *Journal of Structural Engineering*. 149(3), 11776, 2023.

[19] Harvey Jr PS, Elisha G. Vision-based vibration monitoring using existing cameras installed within a building. *Structural Control and Health Monitoring*, 25, e2235, 2018.

[20] Zhou J, Huo L, Huang C, Yang Z, Li H. Feasibility study of earthquake-induced damage assessment for structures by utilizing images from surveillance cameras. *Structural Control and Health Monitoring*, 4993972, 2024.

[21] Dall'Asta A, Leoni G, Gioiella L, Micozzi F, Ragni L, Morici M, Scozzese F, Zona A. Push-and-release tests of a steel building with hybrid base isolation. *Engineering Structures*, 272, 114971, 2022.

[22] Guizar-Sicairos M; Thurman ST, Fienup JR. Efficient subpixel image registration algorithms. *Optics Letters*, 33, 156-158, 2008.

[23] Micozzi F, Morici M, Zona A, Dall'Asta A. Vision-Based Structural Monitoring: Application to a Medium-Span Post-Tensioned Concrete Bridge under Vehicular Traffic. *Infrastructures*, 8, 152, 2023.

[24] Micozzi F, Morici M, Zona A, Dall'Asta A. Vision-based dynamic monitoring of a post-tensioned concrete bridge under vehicular traffic. *Procedia Structural Integrity*, 62C:848-855, 2024.

[25] https://tallwoodinstitute.org/converging-design-home-5663/

[26] Barbosa A, Simpson B, van de Lindt J, Sinha A, Field T, McBain M, Uarac P, Kontra S, Mishra P, Gioiella L, Pieroni L, Pryor S, Saxey B. Shake table testing program for mass timber and hybrid resilient structures datasets for the NHERI Converging Design project, in *Shake table testing program of 6-story mass timber and hybrid resilient structures* (NHERI Converging Design project). DesignSafe-CI, 2025.

[27] Thakur G, Brevdo E, Fučkar NS, Wu H-T. The Synchrosqueezing algorithm for time-varying spectral analysis: Robustness properties and new paleoclimate applications. *Signal Processing* 93, 1079-1094, 2013.