Engineering Structures 292 (2023) 116535

Contents lists available at ScienceDirect

Engineering Structures

journal homepage: www.elsevier.com/locate/engstruct

Three-dimensional structural displacement estimation by fusing monocular camera and accelerometer using adaptive multi-rate Kalman filter

Zhanxiong Ma, Jaemook Choi, Hoon Sohn [*]

Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology, Daejeon, South Korea

ARTICLE INFO

Keywords:
Three-dimensional structural displacement estimation
In-plane displacement
Out-of-plane displacement
Accelerometer
Computer vision
Data fusion

ABSTRACT

Structural displacements play an important role in the health monitoring of civil structures; however, the accurate measurement of structural displacements remains a difficult task. Previous efforts have combined a monocular camera and an accelerometer to estimate structural displacement, but only in-plane displacements could be estimated in this way. In this study, the fusion of a monocular camera and an accelerometer was further extended for out-of-plane or three-dimensional displacement estimation. A computer vision algorithm and an adaptive multi-rate Kalman filter were integrated to efficiently estimate high-sampled displacements from low-sampled vision images and high-sampled acceleration measurements. All parameters associated with the computer vision algorithm were automatically calibrated without using any user-defined thresholds. Experimental validation was performed on two building structures and a 10-m-long bridge structure, and the proposed method accurately estimated the displacement for all three structures with a root mean square error of less than 1 mm.

1. Introduction

Monitoring the displacement of civil structures is important because it plays a vital role in structural health monitoring. Displacement helps classify the global behavior of a structure and evaluate its safety. For instance, many countries, including the United States [1] and the Republic of Korea [2], have adopted displacement as a safety indicator in their structural design specifications. In addition, displacement has been widely employed to evaluate the vibration serviceability of pedestrian bridges [3], identify the modal parameters of light poles [4] and building structures [5], and update finite element models of bridges [6]. Linear variable differential transformers (LVDT) [7,8], accelerometers [9,10], and the real-time kinematic global navigation satellite system (RTK-GNSS) [11,12] are traditional sensors used for structural displacement monitoring. However, the field installation of LVDT is cumbersome, unexpected scaffold vibration may lead to inaccurate displacement measurement, the accelerometer cannot estimate important low-frequency displacement, and RTK-GNSS has a limited sampling rate (less than 20 Hz) and limited accuracy (approximately 7–10 mm). Note that all of these sensors must be installed at the displacement estimation point of a target structure; therefore, methods that use these sensors are classified as contact-type methods.

In recent decades, non-contact structural displacement estimation methods using laser Doppler vibrometers (LDV) [13] and radar systems [14,15] have attracted attention. High-accuracy displacement can be measured by LDV and radar systems at a high sampling rate; however, both LDV and radar systems are expensive. Vision cameras are also widely used for non-contact structural displacement monitoring, owing to their low cost. In these applications, a vision camera is mounted at a fixed point to track a target structure at its displacement estimation point. The structural displacement is first extracted from vision measurements using various algorithms, such as optical flow algorithms [16,17], feature-matching algorithms [18,19], and deep-learning algorithms [20,21], and then converted from pixel units to length units using a scale factor pre-estimated from the target dimensions [22] or the target-to-camera distance [23]. Note that most of these studies focused only on in-plane displacement. As shown in Fig. 1, the out-of-plane displacement of the target structure relative to the camera (u_y) can theoretically be estimated from the temporal changes in the target height and width in the image plane (Δw and Δh, respectively) as follows:

$$u_y = \frac{D}{w}\Delta w = \frac{D}{h}\Delta h \tag{1}$$

where D denotes the target-to-camera distance. w and h denote the original target's width and height, respectively, in the image plane.

[*] Corresponding author.
E-mail addresses: mazhanxiong@kaist.ac.kr (Z. Ma), cjmook@kaist.ac.kr (J. Choi), hoonsohn@kaist.ac.kr (H. Sohn).

https://doi.org/10.1016/j.engstruct.2023.116535
Received 29 January 2023; Received in revised form 31 May 2023; Accepted 28 June 2023

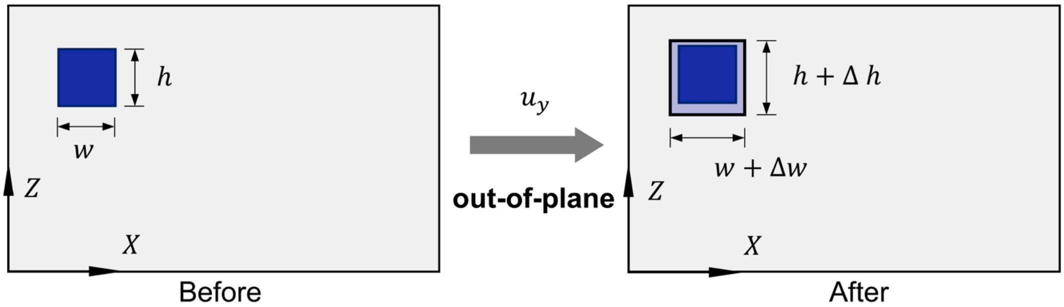


Fig. 1. Target size change caused by target out-of-plane motion.

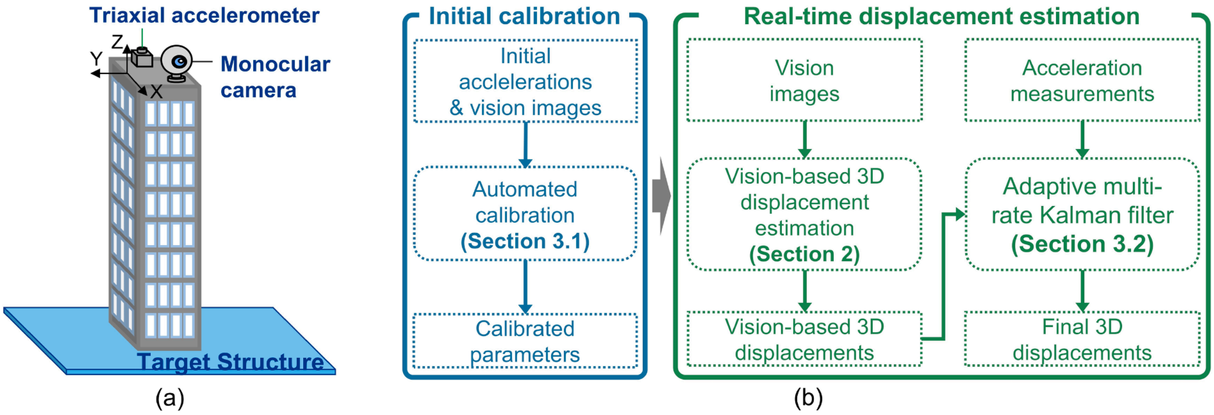

Triaxial accelerometer
Z
Y
X
Monocular camera
Target Structure
(a)
Initial calibration
Initial acclelerations & vision images
Automated calibration (Section 3.1)
Calibrated parameters
Real-time displacement estimation
Vision images
Acceleration measurements
Vision-based 3D displacement estimation (Section 2)
Adaptive multi-rate Kalman filter (Section 3.2)
Vision-based 3D displacements
Final 3D displacements
(b)

Fig. 2. Overview of the proposed structural displacement estimation method: (a) sensor setup and (b) overall flowchart.

However, D, which is at least a few meters, is much larger than u_y, which is on the millimeter or centimeter scale, leading to inaccurate out-of-plane displacement estimation. A sampling moiré method has been proposed for out-of-plane displacement estimation, but the method estimates single-directional displacement, and requires an additional artificial target [24,25]. There have been a few attempts at vision-based three-dimensional (3D) displacement estimation, but they require either a projector [26,27] or a binocular camera [28,29], and may have limited performance for long-distance displacement estimation.

Recently, heterogeneous sensor fusion-based methods have been investigated for structural displacement estimation [30–33], and the fusion of accelerometers and vision cameras has been extensively studied. For example, Park *et al.* [34] and Xu *et al.* [35] estimated high-sampled structural displacement by fusing vision-based displacement and acceleration using a finite impulse response (FIR) filter and Kalman filter, respectively. The authors previously [36,37] combined a monocular camera and an accelerometer installed at a target structure to estimate the structural displacement at the installation point. The scale factor was automatically estimated using initial accelerations and vision images, and high-sampled structural displacement was estimated in real-time using adaptive multi-rate Kalman filter-based fusion of high-sampled acceleration measurement and low-sampled vision-based displacement estimated by an improved feature-matching algorithm [36] or a hybrid computer vision algorithm [37]. However, these methods estimate only the in-plane displacements. Some structures, such as buildings, often vibrate in two horizontal directions. Although some structures, such as bridges, may only vibrate in one direction, it may be difficult to find a nearby fixed target with which the structures have in-plane displacements.

Using a monocular camera and triaxial accelerometer installed at the displacement estimation point of a target structure (as shown in Fig. 2(a)), this study proposes a structural displacement estimation method that is suitable for in-plane, out-of-plane, or even 3D displacement estimation. Fig. 2(b) shows a flowchart of the proposed method. First, the unknown parameters necessary for displacement estimation using a monocular camera (explained in Section 2) are calibrated using initial accelerations and vision images (Section 3.1). Then, after the first step, the vision images are used to estimate the vision-based displacements. Finally, an adaptive multi-rate Kalman filter combines the vision-based displacements with acceleration measurements to improve displacement estimation accuracy (Section 3.2). The performance of the proposed method was validated through a series of laboratory tests, as described in Section 4. Finally, concluding remarks are provided in Section 5. This study offers the following contributions: (1) 3D structural displacement estimation using a monocular camera and an accelerometer, (2) separation of in-plane and out-of-plane displacements using two targets within the field of view (FOV) of the monocular camera, (3) automated calibration of unknown parameters involved in the vision-based displacement estimation, and (4) accurate out-of-plane displacement estimation even at a long target-to-camera distance of 30 m.

2. Structural displacement estimation using a monocular camera

This section explains the working principle of structural displacement estimation using a monocular camera. The estimated vision-based displacement can be fused with acceleration measurements to obtain the final displacement, as explained in Section 3. Note that it is assumed that the camera is rigidly installed on the target structure without any ego-motion and the three axes of the camera are parallel to the three vibration directions of the structure. In addition, the selected targets tracking by the camera are stationary.

2.1. In-plane or out-of-plane structural displacement estimation

When a vision camera is mounted at the displacement estimation point of a target structure, as shown in Fig. 2(a), the structural displacement can be estimated by tracking a nearby fixed target in the

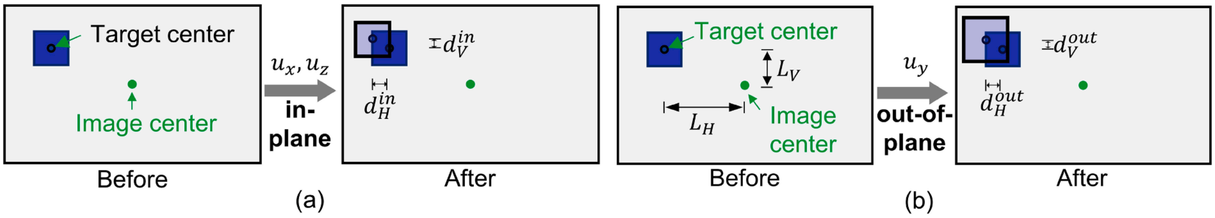


Fig. 3. Target movements caused by structural (a) in-plane (u_x and u_z) and (b) out-of-plane (u_y) displacements.

surroundings of the target structure using a feature-matching algorithm. The structural in-plane displacements relative to the target (u_x and u_z) can cause a target movement in the image plane as shown in Fig. 3(a), and the following relation can be found between the target movement in a pixel unit (d_H^{in} and d_V^{in}) and the structural displacement in a length unit,

$$u_x = \alpha_H^{in} d_H^{in};\ u_z = \alpha_V^{in} d_V^{in} \tag{2}$$

where α_H^{in} and α_V^{in} denote the scale factors for converting in-plane horizontal and vertical displacements from pixel units to length units, respectively, and their values can be estimated from the target dimensions [22] or the target-to-camera distance [23].

If the structure has an out-of-plane displacement (u_y) relative to the target, two-dimensional target movements (d_H^{out} and d_V^{out} in the horizontal and vertical directions, respectively) can be observed, as shown in Fig. 3(b), and the following relationship can be found between the target movements and the structural displacement:

$$u_y = \alpha_H^{out} d_H^{out} = \frac{D}{L_H} d_H^{out};\ u_y = \alpha_V^{out} d_V^{out} = \frac{D}{L_V} d_V^{out}; \tag{3}$$

where α_H^{out} and α_V^{out} denote the scale factors for converting the out-of-plane displacement from pixel units to length units when using horizontal and vertical target movements, respectively. D denotes the target-to-camera distance. L_H and L_V denote the horizontal and vertical distances between the image center and the target center in the image plane, respectively.

2.2. Three-dimensional structural displacement estimation

If the target structure has 3D displacements, the target movements induced by the in-plane and out-of-plane structural displacements are mixed:

$$d_H = d_H^{out} + d_H^{in} = \frac{1}{\alpha_H^{out}} u_y + \frac{1}{\alpha_H^{in}} u_x;\ d_V = d_V^{out} + d_V^{in} = \frac{1}{\alpha_V^{out}} u_y + \frac{1}{\alpha_V^{in}} u_z \tag{4}$$

This causes difficulties in 3D displacement estimation. However, multiple targets are commonly available in the FOV of the camera when a vision camera is installed at the displacement estimation point of the target structure. Assuming that two targets are selected, Equations (4) can be rewritten as

$$\begin{bmatrix} d_{H,1} \\ d_{H,2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\alpha_{H,1}^{out}} & \frac{1}{\alpha_{H,1}^{in}} \\ \frac{1}{\alpha_{H,2}^{out}} & \frac{1}{\alpha_{H,2}^{in}} \end{bmatrix} \begin{bmatrix} u_y \\ u_x \end{bmatrix};\ \begin{bmatrix} d_{V,1} \\ d_{V,2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\alpha_{V,1}^{out}} & \frac{1}{\alpha_{V,1}^{in}} \\ \frac{1}{\alpha_{V,2}^{out}} & \frac{1}{\alpha_{V,2}^{in}} \end{bmatrix} \begin{bmatrix} u_y \\ u_z \end{bmatrix} \tag{5}$$

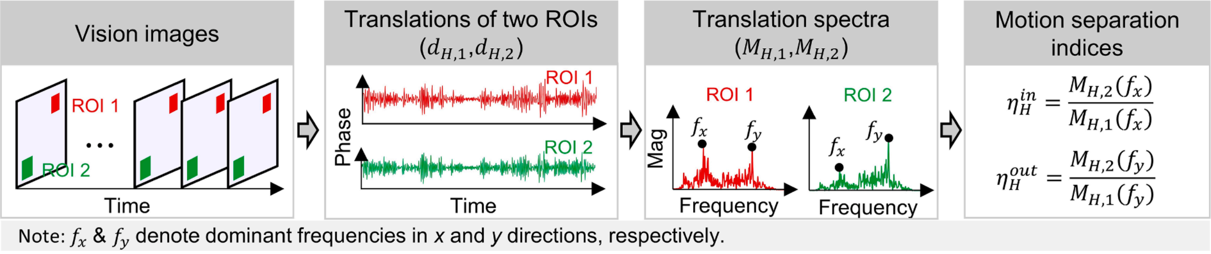
Then, 3D displacements can be estimated as

$$\begin{bmatrix} u_y \\ u_x \end{bmatrix} = \begin{bmatrix} \frac{1}{\alpha_{H,1}^{out}} & \frac{1}{\alpha_{H,1}^{in}} \\ \frac{1}{\alpha_{H,2}^{out}} & \frac{1}{\alpha_{H,2}^{in}} \end{bmatrix}^{-1} \begin{bmatrix} d_{H,1} \\ d_{H,2} \end{bmatrix};\ \begin{bmatrix} u_y \\ u_z \end{bmatrix} = \begin{bmatrix} \frac{1}{\alpha_{V,1}^{out}} & \frac{1}{\alpha_{V,1}^{in}} \\ \frac{1}{\alpha_{V,2}^{out}} & \frac{1}{\alpha_{V,2}^{in}} \end{bmatrix}^{-1} \begin{bmatrix} d_{V,1} \\ d_{V,2} \end{bmatrix} \tag{6}$$

where the subscripts 1 and 2 denote the first and second targets, respectively. Here, the distance variation caused by the out-of-plane displacement is ignored, considering that the out-of-plane displacement at the millimeter or centimeter scale is much smaller than the target-to-camera distance of at least several meters. Note that these two targets should be far away from the image center in the image plane and they should be selected from opposite sides of images for better out-of-plane displacement estimation. In addition, these two targets should be close to the camera for best displacement estimation and have sufficient features.

3. Three-dimensional structural displacement estimation using a monocular camera and a triaxial accelerometer

This section proposes a structural displacement estimation method using a monocular camera and triaxial accelerometer installed at the displacement estimation point of a target structure, as shown in Fig. 2 (a). Low-sampled 3D displacements are first estimated from vision images using two selected ROIs and then fused with high-sampled acceleration measurements with an adaptive multi-rate Kalman filter, thereby obtaining the final high-sampled displacements. Because several parameters are associated with vision-based 3D displacement estimation (Equation (6)), an algorithm was proposed to automatically calibrate them using initial accelerations and vision images. Therefore, the proposed method consists of two stages: (1) initial calibration (Section 3.1) and (2) real-time structural displacement estimation (Section 3.2). Note that the proposed method is explained here for displacement estimation in the x and y directions using target horizontal movement, but it can also be easily extended to 3D displacement estimation using target vertical movement.



Note: f_x & f_y denote dominant frequencies in x and y directions, respectively.

Fig. 4. Flowchart of motion separation index estimation.

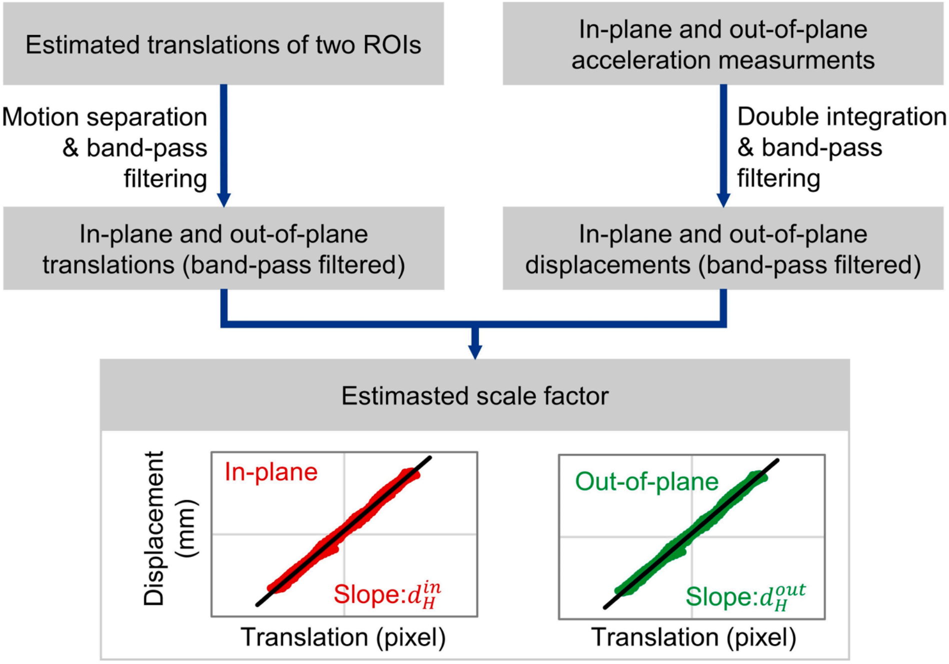

Estimated translations of two ROIs
In-plane and out-of-plane acceleration measurments
Motion separation & band-pass filtering
Double integration & band-pass filtering
In-plane and out-of-plane translations (band-pass filtered)
In-plane and out-of-plane displacements (band-pass filtered)
Estimasted scale factor
Displacement (mm)
In-plane
Slope:d_H^{in}
Out-of-plane
Slope:d_H^{out}
Translation (pixel)
Translation (pixel)

Fig. 5. Flowchart of scale factor estimation.

3.1. *Initial calibration*

3.1.1. *Motion separation index and scale factor estimation*

To estimate the structural displacement from vision images using Equation (6), the parameters associated with the equation, such as the target-to-camera distance and target dimensions, should be known, but may not be readily available in some applications. Two motion separation indices are defined as

$$\eta_H^{out} = \frac{\alpha_{H,2}^{out}}{\alpha_{H,1}^{out}}; \eta_H^{in} = \frac{\alpha_{H,2}^{in}}{\alpha_{H,1}^{in}} \tag{7}$$

and Equation (6) can be rewritten by introducing Equation (7):

$$\begin{bmatrix} u_y \\ u_x \end{bmatrix} = \begin{bmatrix} \alpha_{H,1}^{out} & 0 \\ 0 & \alpha_{H,1}^{in} \end{bmatrix} \begin{bmatrix} d_1^{out} \\ d_1^{in} \end{bmatrix} = \begin{bmatrix} \alpha_{H,1}^{out} & 0 \\ 0 & \alpha_{H,1}^{in} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \eta_H^{out} & \eta_H^{in} \end{bmatrix}^{-1} \begin{bmatrix} d_{H,1} \\ d_{H,2} \end{bmatrix} \tag{8}$$

Then, the motion separation indices (η_H^{out} and η_H^{in}) and scale factors ($\alpha_{H,1}^{out}$ and $\alpha_{H,1}^{in}$) can be separately calibrated using initial accelerations and vision images.

Fig. 4 shows a flowchart of the motion separation index estimation using the initial vision images. Two ROIs were selected from the FOV of the vision camera to cover two targets and translations ($d_{H,1}$ and $d_{H,2}$) are estimated using the feature-matching algorithm from the initial Q vision images. The estimated translations are then transformed into the frequency domain using Fourier transform ($M_{H,1}(f)$ and $M_{H,2}(f)$), and the following equation is obtained,

$$\begin{bmatrix} M_{H,1}(f) \\ M_{H,2}(f) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \eta_H^{out} & \eta_H^{in} \end{bmatrix} \begin{bmatrix} M_1^{out}(f) \\ M_1^{in}(f) \end{bmatrix} \tag{9}$$

where $M_1^{out}(f)$ and $M_1^{in}(f)$ denote the Fourier transform of d_1^{out} and d_1^{in}, respectively. Considering that most structures have different first natural frequencies in different directions, the target structure has different dominant frequencies in two different directions (i.e., f_x and f_y, respectively),

$$M_1^{in}(f_y) = M_1^{out}(f_x) = 0 \tag{10}$$

Then, the following equation can be obtained from Equations (9) and (10),

$$M_{H,1}(f_x) = M_1^{in}(f_x); M_{H,1}(f_y) = M_1^{out}(f_y); \tag{11}$$

$$M_{H,2}(f_x) = \eta_H^{in} M_1^{in}(f_x); M_{H,2}(f_y) = \eta_H^{out} M_1^{out}(f_y);$$

and two motion separation indices can be estimated as

$$\eta_H^{out} = \frac{M_{H,2}(f_y)}{M_{H,1}(f_y)}; \eta_H^{in} = \frac{M_{H,2}(f_x)}{M_{H,1}(f_x)} \tag{12}$$

If two ROIs are selected from opposite sides of the FOV, η_H^{out} should become

$$\eta_H^{out} = -\frac{M_{H,1}(f_y)}{M_{H,2}(f_y)} \tag{13}$$

Note that for the structures with the same fundamental frequencies in different directions, the proposed technique requires that they have different excitation frequencies during the initial short period. Then, motion separation indices can be estimated in the same way.

Fig. 5 shows a flowchart of the scale factor estimation using the initial accelerations and vision images. The translations estimated from the two ROIs are separated using the estimated motion separation indices:

$$\begin{bmatrix} d_1^{out} \\ d_1^{in} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \eta_H^{out} & \eta_H^{in} \end{bmatrix}^{-1} \begin{bmatrix} d_{H,1} \\ d_{H,2} \end{bmatrix} \tag{14}$$

and the separated in-plane and out-of-plane translations (d_1^{out} and d_1^{in}) are band-pass filtered. In contrast, in-plane and out-of-plane displacements in the same frequency band are estimated from the corresponding acceleration measurements through double integration and band-pass filtering. Finally, the two scale factors are estimated from the least-squares regression of the translation and displacements in the in-plane and out-of-plane directions, respectively. Note that the lower cut-off frequency of the band-pass filter should be sufficiently high to remove large low-frequency drifts in the acceleration-based displacement, whereas the upper cut-off frequency should be 1/10 of the sampling frequency of vision measurements [36].

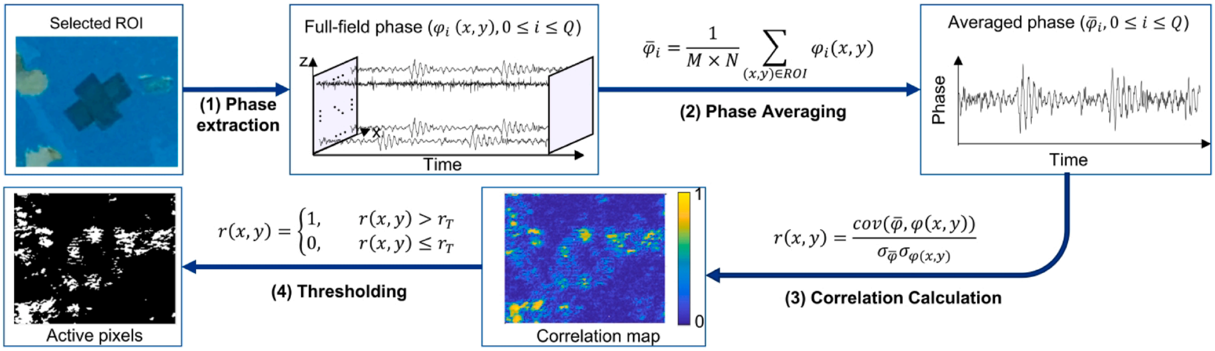

Selected ROI
(1) Phase extraction
Full-field phase ($\varphi_i\ (x,y), 0 \le i \le Q$)
Z
X
Time
$\bar{\varphi}_i = \frac{1}{M \times N} \sum_{(x,y)\in ROI} \varphi_i(x,y)$
(2) Phase Averaging
Averaged phase ($\bar{\varphi}_i, 0 \le i \le Q$)
Phase
Time
$r(x,y) = \frac{cov(\bar{\varphi}, \varphi(x,y))}{\sigma_{\bar{\varphi}}\sigma_{\varphi(x,y)}}$
(3) Correlation Calculation
Correlation map
1
0
$r(x,y) = \begin{cases} 1, & r(x,y) > r_T \\ 0, & r(x,y) \le r_T \end{cases}$
(4) Thresholding
Active pixels

Fig. 6. Flowchart of active pixel selection: (1) extraction of full-field phases within the ROI, (2) averaging of the full-field phases to obtain an averaged phase, (3) calculation of correlation coefficients between the averaged phase and phases of all pixels to obtain a correlation map, and (4) thresholding of the calculated correlation coefficients.

3.1.2. *Active pixel selection for phase-based algorithm*

If the displacement of the target structure is small or the target-to-camera distance is long, only subpixel target movements can be observed. Thus, the phase-based algorithm performs better than the feature-matching algorithm [37] and should be used for vision-based displacement estimation. For any given ROI, the algorithm first extracts full-field phases by calculating the spatial convolution between the ROI and a complex Gabor filter $(G_2^v + jH_2^v)$ [38]. Assuming pixels within the ROI share the same motion pattern induced by structural vibration, an average phase is calculated as the spatial averaging of the extracted full-field phases, and the phase variation compared with the initial average phase extracted from the 1st ROI is finally converted to displacement.

Note that, unlike other computer vision algorithms that estimate target movement in a pixel unit, the phase-based algorithm estimates the target movement as phase variation. Therefore, a different scale factor is required in the phase-based algorithm to convert phase to displacement. However, it can be estimated similarly, as explained in Section 3.1.1. In addition to scale factor estimation, active pixels should be selected within the selected ROIs for the phase-based algorithm. The authors previously proposed an acceleration-aided algorithm for active pixel selection [37]; however, the algorithm cannot be applied here because of the mixture of in-plane and out-of-plane motions. Therefore, a new automatic active pixel selection algorithm is proposed that does not use acceleration measurements or any ad-hoc threshold. Fig. 6 shows a flowchart of the active pixel selection process. The full-field phase within the selected ROI ($\varphi_i(x,y),\ 0 \le i \le Q$) is first extracted from the initial vision images. The averaged phase is then calculated as

$$\overline{\varphi}_i = \frac{1}{M \times N} \sum_{(x,y)\in ROI} \varphi_i(x,y), 0 \le i \le Q \tag{15}$$

where M and N denote the dimensions of the ROI. After that, a correlation coefficient is calculated between the averaged phase and the phase of each pixel, and then a correlation map is generated:

$$r(x,y) = \frac{cov(\overline{\varphi}, \varphi(x,y))}{\sigma_{\overline{\varphi}}\sigma_{\varphi(x,y)}} \tag{16}$$

where cov and σ denote the covariance and standard derivation, respectively. The correlation coefficient map is eventually binarized, and the active pixels are selected as pixels with correlation coefficients larger than the threshold (r_T),

$$r(x,y) = \begin{cases} 1, r(x,y)\rangle r_T \\ 0, r(x,y) \le r_T \end{cases} \tag{17}$$

To avoid using any user-defined thresholds, the OSTU algorithm [39] is were introduced. The OSTU algorithm is a commonly used image thresholding algorithm. Using any given intensity threshold, an image can be divided into two classes. Without using any a priori knowledge, the OSTU algorithm automatically computes such a threshold by maximizing inter-class variance. In this study, the OSTU algorithm was applied to the correlation map, which was constructed by calculating the correlation coefficient between the phase of each pixel and the average phase of all pixels. Since the phases of active pixels represent structural vibration while the phases of inactive pixels are basically noise, active pixels will have high correlation coefficients while inactive pixels will have low correlation coefficients. The threshold of the correlation coefficient (r_T) can be calculated automatically using the OSTU algorithm.

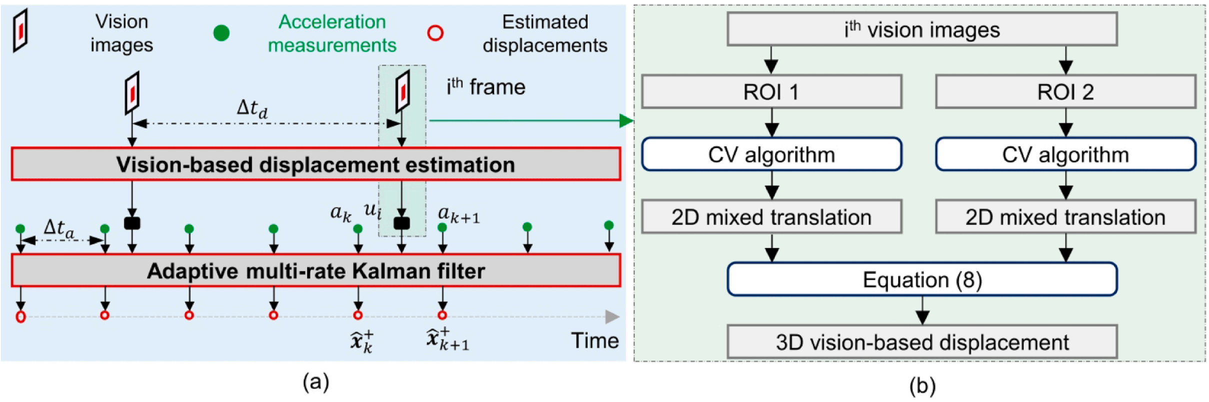


Fig. 7. Real-time 3D structural displacement estimation using an adaptive multi-rate Kalman filter: (a) overview and (b) flowchart of 3D vision-based displacement estimation.

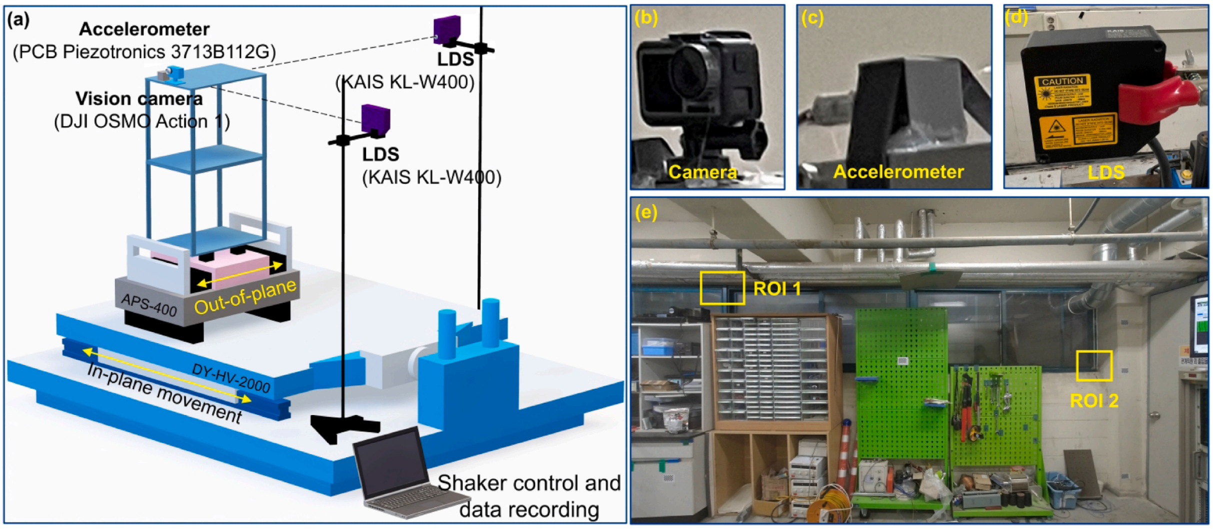


Fig. 8. Overall configuration of an indoor two-story building structure test: (a) sensor setup, (b) DJI OSMO Action camera, and (c) PCB Piezotronics 3713B112G triaxial accelerometer for displacement estimation, (d) laser-based displacement sensor (LDS) to measure reference displacement, and (e) the FOV of the monocular camera and the two ROIs cropped for displacement estimation.

3.2. *Real-time three-dimensional structural displacement estimation*

After the initial calibrations, the structural displacement can be continuously estimated through the adaptive multi-rate Kalman filter-based fusion of asynchronous vision and acceleration measurements, as shown in Fig. 7(a). The 2D mixed translations are independently estimated from each of the two selected ROIs, and the 3D vision-based displacements are estimated using Equation (8), as shown in Fig. 7(b). The low-sampled 3D vision-based displacements are then fused with the high-sampled 3D acceleration measurements to obtain the final displacement, which has the same high sampling rate as the acceleration measurements. Here, the vision-based displacement in each direction should be independently fused with the corresponding acceleration measurement using an adaptive multi-rate Kalman filter. The fusion details are briefly explained below for displacement estimation in one direction, and can be easily extended to the other two directions.

Assuming that the state estimate and acceleration measurement at the (k-1)th time step ($\hat{x}_{k-1}$ and a_{k-1}, respectively) are available, if the vision image is not obtained in the period $[(k-1)\Delta t_a, k\Delta t_a]$, a state ($\hat{x}_k$) at the kth time step is estimated as

$$\hat{\boldsymbol{x}}_k = \boldsymbol{A}(\Delta t_a)\hat{\boldsymbol{x}}_{k-1} + \boldsymbol{B}(\Delta t_a)a_{k-1} \tag{18}$$

$$\boldsymbol{A}(\Delta t_a) = \begin{bmatrix} 1 & \Delta t_a \\ 0 & 1 \end{bmatrix}; \boldsymbol{B}(\Delta t_a) = \begin{bmatrix} \Delta t_a^2/2 \\ \Delta t \end{bmatrix} \tag{19}$$

Here, $\hat{x}_k$ includes two entities, corresponding to the displacement and velocity estimated at$t =$ k Δt_a. Δt_a denotes the acceleration sampling interval. If a vision image is obtained in the period $[(k-1)\Delta t_a, k\Delta t_a]$, $\hat{x}_k$ is estimated as

$$\hat{\boldsymbol{x}}_k = \boldsymbol{A}(\Delta t_{k,i})\left\{(\boldsymbol{I} - \boldsymbol{KH})\left(\boldsymbol{A}(\Delta t_{i,k-1})\hat{\boldsymbol{x}}_{k-1} + \boldsymbol{B}(\Delta t_{i,k-1})a_{k-1}\right) + \boldsymbol{K}u_i\right\} + \boldsymbol{B}(\Delta t_{k,i})a_{k-1} \tag{20}$$

$$\Delta t_{i,k-1} = i\Delta t_d - (k-1)\Delta t_a; \Delta t_{k,i} = k\Delta t_a - i\Delta t_d; \boldsymbol{H} = [1 \quad 0]^T \tag{21}$$

where $\boldsymbol{K}$ denotes the Kalman gain and $\boldsymbol{I}$ denotes a 2 × 2 identity matrix. Δt_d denotes the sampling interval of the vision images. The derivation of Equations (18) to (21) and more details of the adaptive multi-rate Kalman filter can be found in the study by Ma et al. [36]. Note that the final estimated displacement has the same sampling rate as the acceleration measurement, which is higher than that of vision-based displacement.

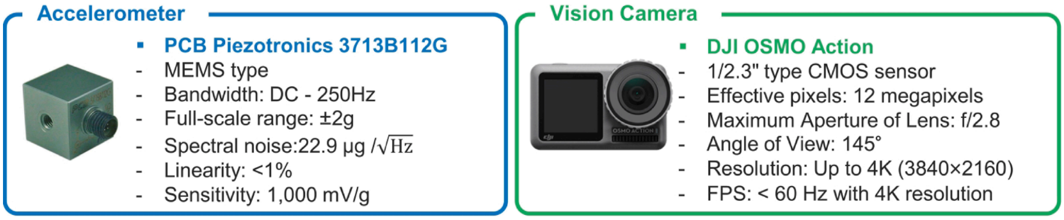
4. Experimental validation

4.1. *Indoor two-story building structure test*

4.1.1. *Test setup*

The absence of a 3D shaker makes it difficult to perform tests considering 3D vibrations. However, in-plane (even bi-directional) displacement estimation has been extensively studied using monocular cameras, and the main difficulty in 3D displacement estimation is the simultaneous estimation of in-plane and out-of-plane displacements. Therefore, the proposed technique was first validated on a two-story building structure by considering simultaneous in-plane and out-of-plane excitations.

Fig. 8 shows the overall configuration of the test. As shown in Fig. 8 (a), the two-story building structure was installed on an APS-400 shaking table placed on a DY-HY-2000 shaking table. The moving directions of the two shaking tables were perpendicular, and so the two-story building structure had two-dimensional (2D) horizontal



Accelerometer

- PCB Piezotronics 3713B112G
- MEMS type
- Bandwidth: DC - 250Hz
- Full-scale range: ±2g
- Spectral noise:22.9 μg /$\sqrt{\text{Hz}}$
- Linearity: <1%
- Sensitivity: 1,000 mV/g

Vision Camera

- DJI OSMO Action
- 1/2.3" type CMOS sensor
- Effective pixels: 12 megapixels
- Maximum Aperture of Lens: f/2.8
- Angle of View: 145°
- Resolution: Up to 4K (3840×2160)
- FPS: < 60 Hz with 4K resolution

Fig. 9. Detailed specifications of the accelerometer and monocular camera used in the indoor two-story building structure test.

Table 1
Description of excitations in the indoor two-story building structure test.

		Out-of-plane excitations				
		None	Recorded signal	0.3 Hz	1 Hz	2 Hz
In-plane excitation	None	–	Test 1	Test 2	Test 3	Test 4
	0.2 Hz	Test 5	–	–	–	–
	0.8 Hz	Test 6	Test 8	Test 9	Test 10	Test 11
	1.5 Hz	Test 7	Test 12	Test 13	Test 14	Test 15

vibration. A triaxial accelerometer (Fig. 8(b)) and monocular camera (Fig. 8(c)) were installed on the top of the two-story building structure for displacement estimation at the same point. Detailed specifications of the monocular camera and accelerometer are listed in Fig. 9. The displacements of the building structure were also measured using two laser-based displacement sensors (LDS) (Fig. 8(d)) with micrometer-level accuracy [40]. Acceleration and LDS data were recorded at a sampling rate of 100 Hz, whereas vision images were recorded at 29.97 frames per second (FPS) with a resolution of 2880 × 3440. Fig. 8(e) shows the FOV of the vision camera. Two ROIs were selected to cover parts of different window borders at a distance of approximately 2.7 m. Note that the horizontal movements of the window borders in the image plane were used for 2D displacement estimation in this test. To fully validate the displacement estimation performance of the proposed method, 15 tests were performed considering different combinations of in-plane and out-of-plane excitations, as listed in Table 1. Note that only the horizontal movement of two targets in the image plane was used in this test. If the structure also has vertical vibration, the vertical displacement can be easily estimated from the vertical movement of the targets in the image plane by repeating the same procedure.

4.1.2. *Motion separation index and scale factor estimation results*

Considering the relatively short target-to-camera distance and relatively large structural displacement, a feature-matching algorithm, i.e., the KAZE algorithm [41], was adopted to estimate displacements from vision measurements. The motion separation indices and scale factors were estimated using vision and acceleration measurements in Test 10, where 0.8 Hz and 1 Hz sinusoidal signals were inputted to shaking tables to generate in-plane and out-of-plane excitations, respectively. Fig. 10 shows the motion separation index results. The translations estimated from the two ROIs were mixed using in-plane and out-of-plane motion, as shown in Fig. 10(a). The corresponding frequency spectra of the estimated translations are shown in Fig. 10(b), and the two motion separation indices (i.e.,η_H^{in} and η_H^{out}) were estimated as −0.878 and 0.934, respectively. Note that two ROIs were selected from opposite sides of the FOV and that η_H^{out} is a negative number.

Fig. 11 shows the scale factor estimation results. Using the estimated motion separation indices, the in-plane and out-of-plane translations were separated from the translation estimated from the two ROIs, as shown in Fig. 11(a). At the same time, the in-plane and out-of-plane displacements were estimated from the acceleration measurements, as shown in Fig. 11(b). Note that both displacements and translations were filtered by a band-pass filter with a lower cutoff frequency of 0.5 Hz and an upper cutoff frequency of 3 Hz. Finally, the scale factors were estimated as 0.8353 pixels/mm and 1.1523 pixels/mm for the in-plane and

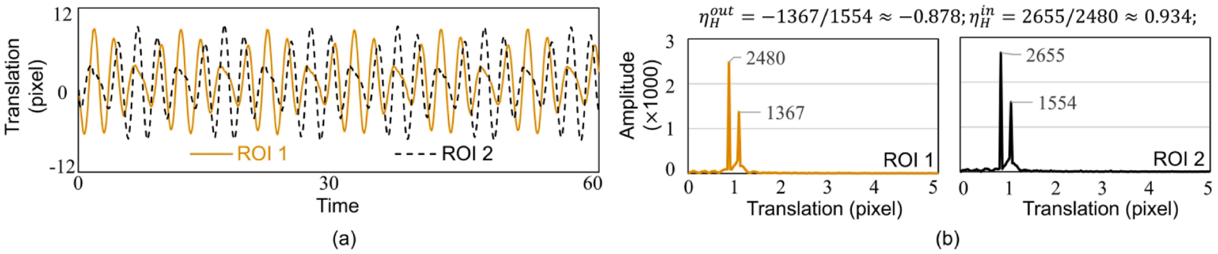


Fig. 10. Motion separation index estimation results in the indoor two-story building structure test: (a) translations extracted from ROIs 1 and 2 and (b) their frequency spectra and the estimated motion separation indices.

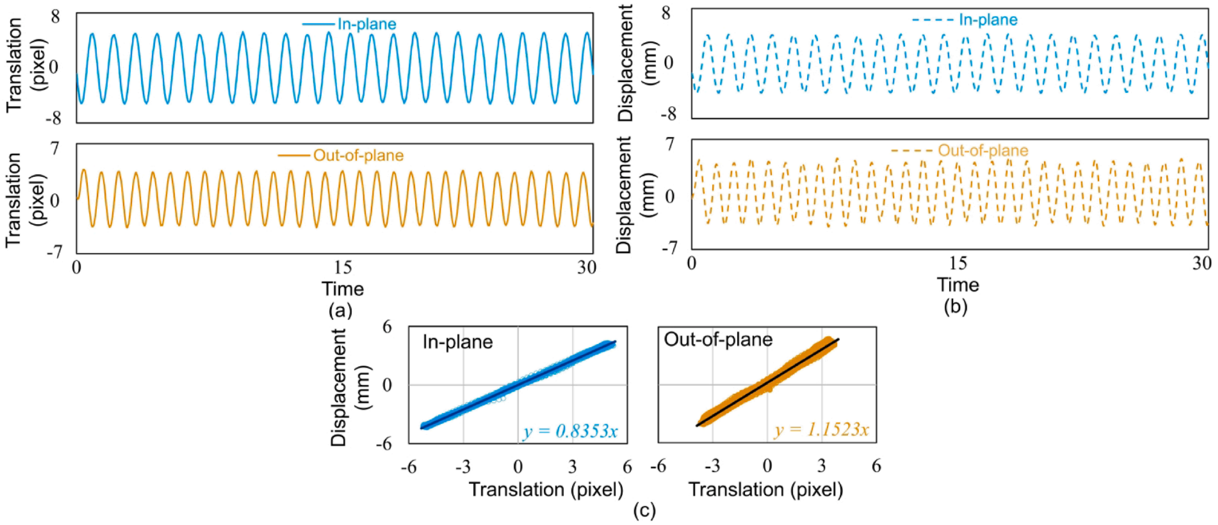


Fig. 11. Scale factor estimation results in the indoor two-story building structure test: (a) separated in-plane and out-of-plane translations (band-pass filtered), (b) in-plane and out-of-plane displacements estimated from acceleration measurements (band-pass filtered), and (c) estimated scale factors between translation and displacement.

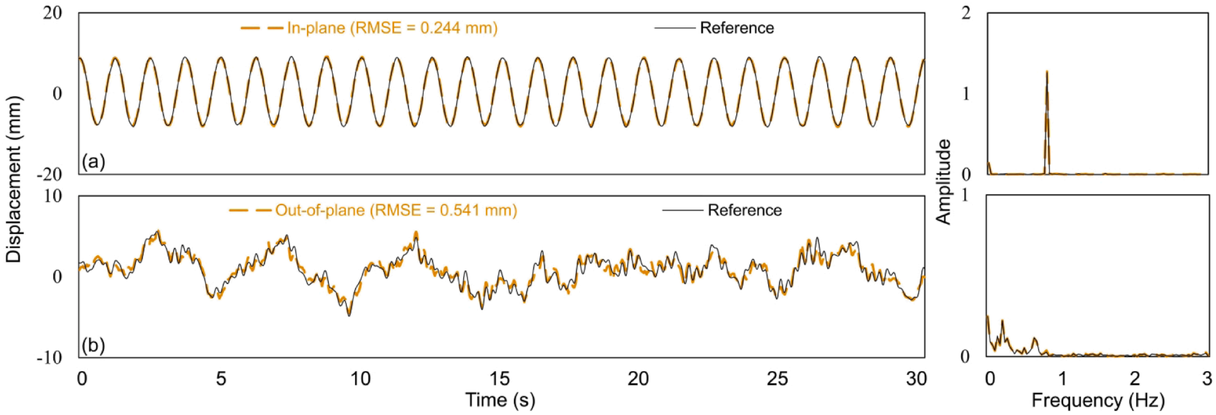


Fig. 12. Displacements estimated in Test 8 of the indoor two-story building structure test: (a) in-plane and (b) out-of-plane.

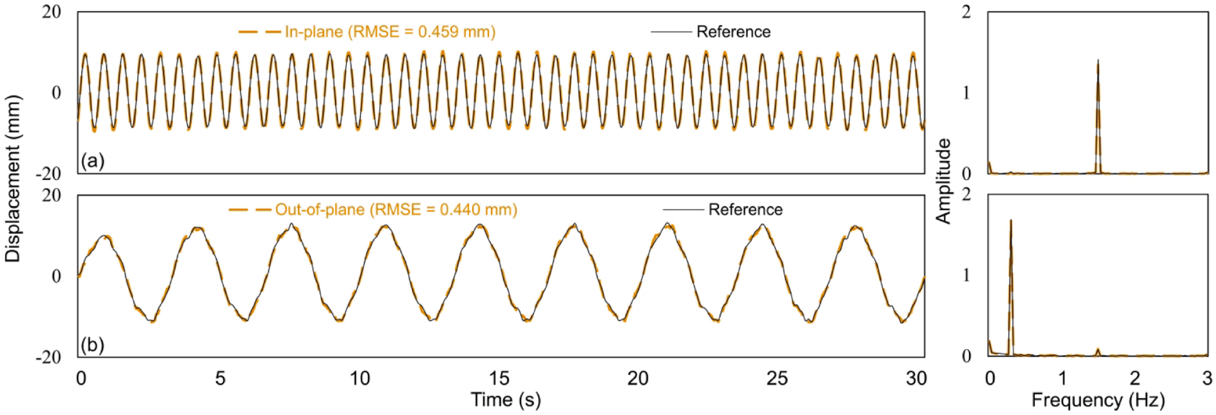


Fig. 13. Displacements estimated in Test 13 of the indoor two-story building structure test: (a) in-plane and (b) out-of-plane.

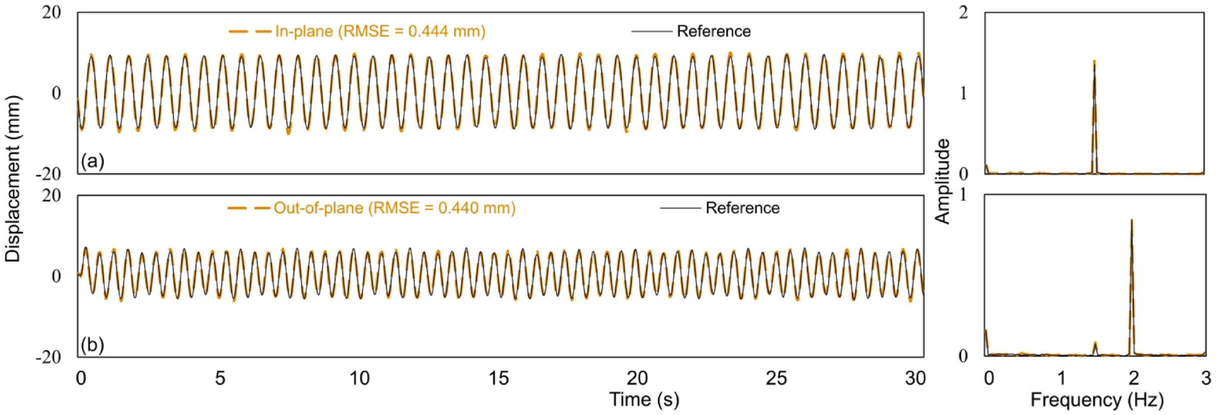


Fig. 14. Displacements estimated in Test 15 of the indoor two-story building structure test: (a) in-plane and (b) out-of-plane.

out-of-plane displacements, respectively, as shown in Fig. 11(c). Note that here vision and acceleration measurements were manually aligned using a correlation-based algorithm [34].

4.1.3. Displacement estimation results

For all 15 tests, displacements were first estimated from the vision images using the estimated motion separation indices and scale factors. The vison-based displacements were then fused with the acceleration measurements to estimate the final displacements. Representative displacement estimation results, that is, displacements estimated in Tests 8, 13, and 15, are shown in Figs. 12-14, and corresponding displacement estimation errors are shown in Fig. 15. Both the in-plane and out-of-plane displacements estimated by the proposed method coincide with those measured by LDS in both frequency and time domains, and the root mean square errors (RMSEs) of the estimated displacements were less than 0.6 mm.

The RMSEs of the displacements estimated in all 15 tests are summarized in Table 2. In all tests, both the in-plane and out-of-plane displacements were estimated accurately. The RMSEs were in the ranges of [0.231 mm, 0.459 mm] and [0.234 mm, 0.552 mm] for the in-plane and out-of-plane displacements, respectively.

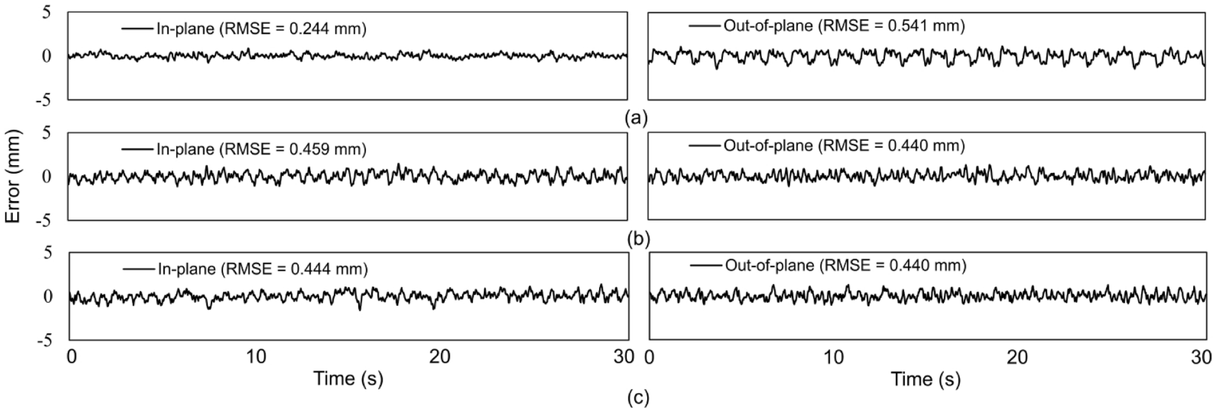


Fig. 15. Displacement estimation errors: (a) Test 8, (b) Test 13, and (c) Test 15.

Table 2
RMSEs (mm) of displacements estimated under different excitations in the indoor two-story building structure test.

		Out-of-plane excitation									
		None		Recorded signal		0.3 Hz		1 Hz		2 Hz	
		In	Out	In	Out	In	Out	In	Out	In	Out
In-plane excitation	None	–	–	–	0.234	–	0.279	–	0.303	–	0.259
	0.2 Hz	0.231	–	–	–	–	–	–	–	–	–
	0.8 Hz	0.269	–	0.244	0.541	0.265	0.552	0.233	0.503	0.278	0.524
	1.5 Hz	0.397	–	0.403	0.398	0.459	0.440	0.431	0.424	0.444	0.440

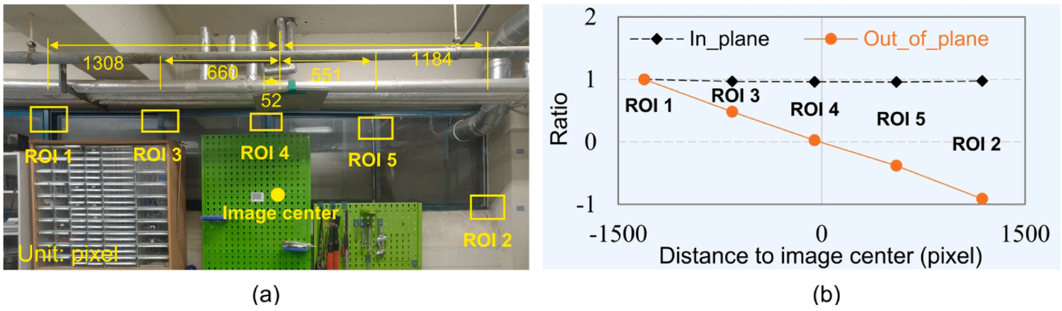


Fig. 16. Effect of ROI location: (a) five ROIs at almost the same distance to the camera and (b) scale factor ratio.

4.1.4. *Effect of ROI location on scale factor estimation*

Considering that vision-based displacement estimation is estimated as the translation multiplied by a scale factor, a small scale factor is beneficial for displacement estimation. In this section, three additional ROIs were selected, as shown in Fig. 16(a), and the effect of the ROI location in the image plane on the scale factor estimation was investigated. The target-to-camera distances for all targets included in the five ROIs were approximately 2.5 m, thus focusing mainly on the effect of different ROI locations in the image plane. Note that this distance is the distance between the surface plane of the camera and the surface plane of the target. In addition, the measurements recorded in tests 5 and 2 were used for the investigation, and the displacements could then be estimated independently from each of these five ROIs because unidirectional excitation was considered in these two tests. The ratios between the scale factor of ROI 1 and the scale factors of all five ROIs were calculated. As shown in Fig. 16(b), the scale factor ratio was almost constant for in-plane displacement. However, the scale factor ratio was highly dependent on the location of the ROI in the image plane, and ROIs far from the image center were preferred.

4.2. *Indoor 10-m-long bridge structure test*

4.2.1. *Test setup*

The proposed method was then validated on a 10-m-long beam-type structure, as shown in Fig. 17(a). Fig. 17(b) shows the overall setup of the 10-m-long beam-type structure test. The same vision camera (Fig. 17(c)) and accelerometer (Fig. 17(d)) used in the previous test were installed at the center of the span for displacement estimation at the same point. A Polytech RSV-150 LDV (Fig. 17(e)) was also used to measure the displacement, and the results were used to evaluate the accuracy of displacements estimated by the proposed method. The LDV and acceleration data were recorded at a sampling rate of 100 Hz, whereas vision measurements were recorded at an FPS of 29.97 Hz with a resolution of 2880 × 3440. Fig. 17(e) shows the FOV of the monocular camera and selected ROI. During the test, the structure had millimeter-level vertical vibration caused by a researcher jumping or walking on the structure, and the vertical vibration became out-of-plane vibration relative to the target included in the ROI. Note that the horizontal movement of the target in the image plane was used for the displacement estimation in this test (Equation (3)). In total, four different excitations were considered in the 10-m-long bridge structure test, and detailed descriptions of the excitations are summarized in Table 3.

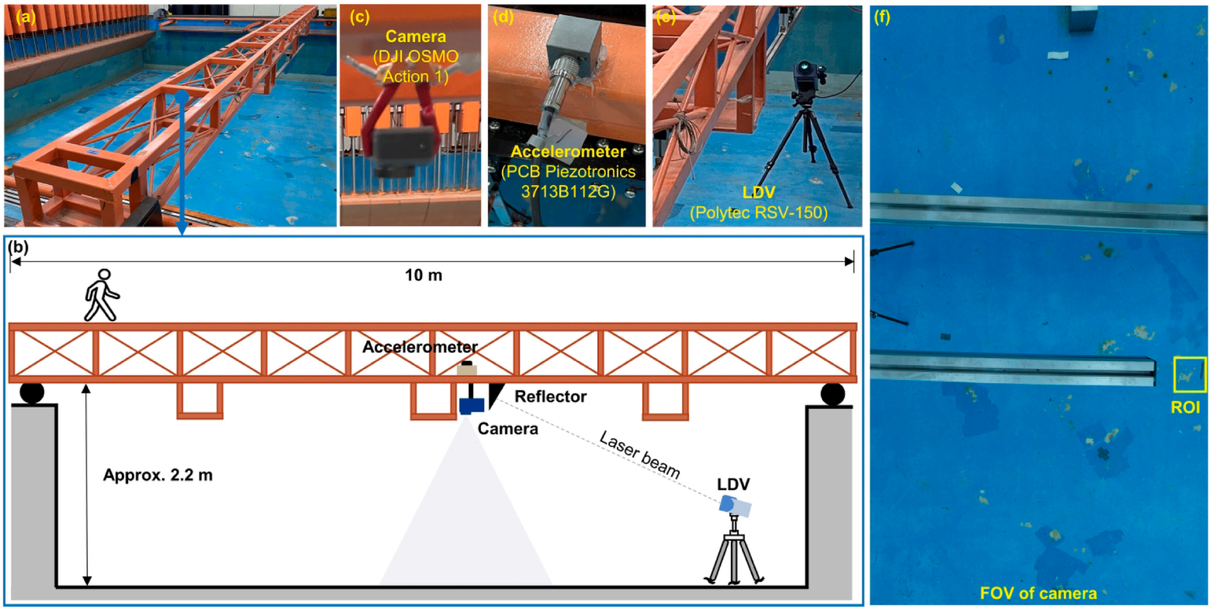


Fig. 17. Overall configuration of the indoor 10-m-long bridge structure test: (a) overview of the test structure, (b) sensor setup, (c) DJI OSMO Action 1 camera, and (d) PCB Piezotronics 3713B112G triaxial accelerometer used for displacement estimation, (e) Polytec RSV-150 LDV used to measure reference displacement, and (f) the FOV of the monocular camera and the cropped ROI for displacement estimation.

Table 3
Description of excitations considered in the indoor 10-m-long bridge structure test.

# of excitations	Descriptions
1, 2	A person slowly passed through the structure and jumped at random locations
3	A person quickly moved to the center of the structure, then stayed static for approximately 50 s, and finally quickly left the structure
4	A person slowly passed through the structure while jumping at random locations. At around 90 s, the person jumped to an adjacent structure and jumped back after 5 s.

4.2.2. *Estimation results*

Considering the relatively small bridge displacement, the phase-based algorithm was adopted for this test. The initial calibration was performed using accelerations and vision images recorded under excitation 2 in this test, and the results are shown in Fig. 18. Fig. 18(a)–(c) show the selected ROI, generated correlation map, and locations of the selected active pixel, respectively, whereas Fig. 18(d) shows the estimated scale factor, which is 3.5767 rad/mm. Although several outliers can be observed in Fig. 18(d), the scale factor was still stably estimated, with a high R^2 value of 0.9259.

Fig. 19 shows the displacement estimation results obtained using a vision camera only or using both a vision camera and an accelerometer. For all four excitations, displacements were estimated from vision measurements with an RMSE of less than 0.4 mm. Fusing vision-based displacement and acceleration measurements further improved the displacement estimation performance; however, the improvement was not significant.

Additionally, the displacement estimation errors under excitation 3 using vision images only or using both acceleration measurements and vision images are compared in Fig. 20. The adaptive multi-rate Kalman filter-based fusion significantly suppressed high-frequency errors; however, the errors were dominated by a low-frequency component. Therefore, the suppression of high-frequency errors did not significantly reduce the RMSE of the estimated displacement. There are two potential reasons for the low-frequency error. First, the recorded images had severe distortion problems. Though automatic calibration was performed through the camera's built-in features, the distortion was not completely eliminated, especially in the marginal portions of the images, which may lead to this error. Using an industrial camera with less distortion may help to reduce this error. Second, the proposed technique only considered structural translation and ignored structural rotation, but a tiny structural rotation may cause large errors. To generate relatively large displacements, structural rotation may be unavoidable in this test, which may lead to this error.

4.3. *Outdoor four-story building structure test*

4.3.1. *Test setup*

Finally, the proposed method was applied to a four-story building structure in an outdoor setting, as shown in Fig. 21. The four-story

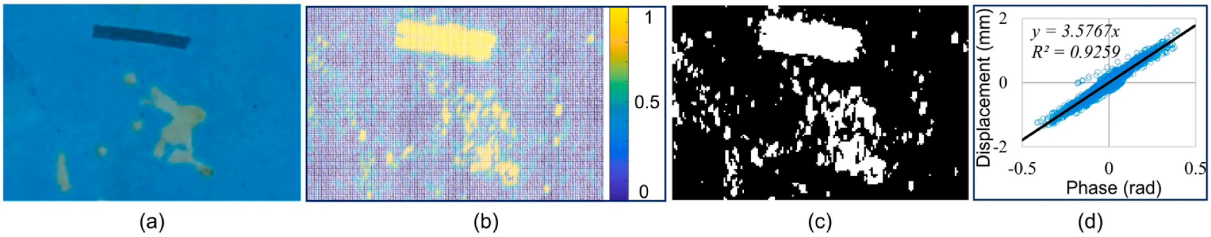


Fig. 18. Active pixel selection and scale factor estimation results using vision and acceleration measurements of excitation 2 in the 10-m-long bridge structure test: (a) selected ROI, (b) generated correlation map, (c) locations of the selected active pixel, and (d) estimated scale factor.

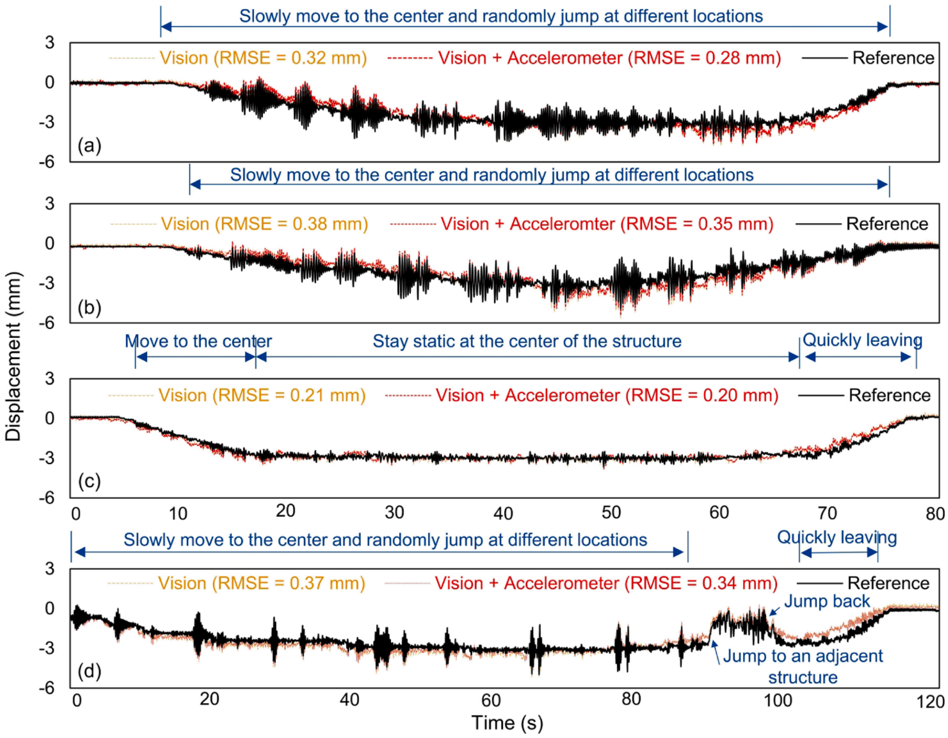


Fig. 19. Displacement estimation results in the 10-m-long bridge structure test: (a) excitation 1, (b) excitation 2, (c) excitation 3, and (d) excitation 4.

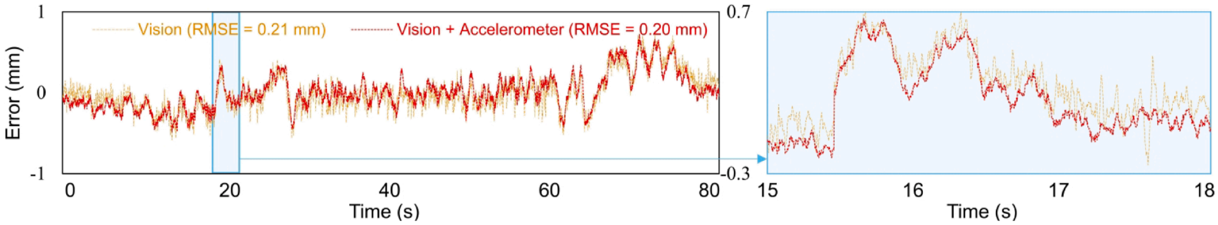


Fig. 20. Displacement estimation errors under excitation 3 in the 10-m-long bridge structure test.

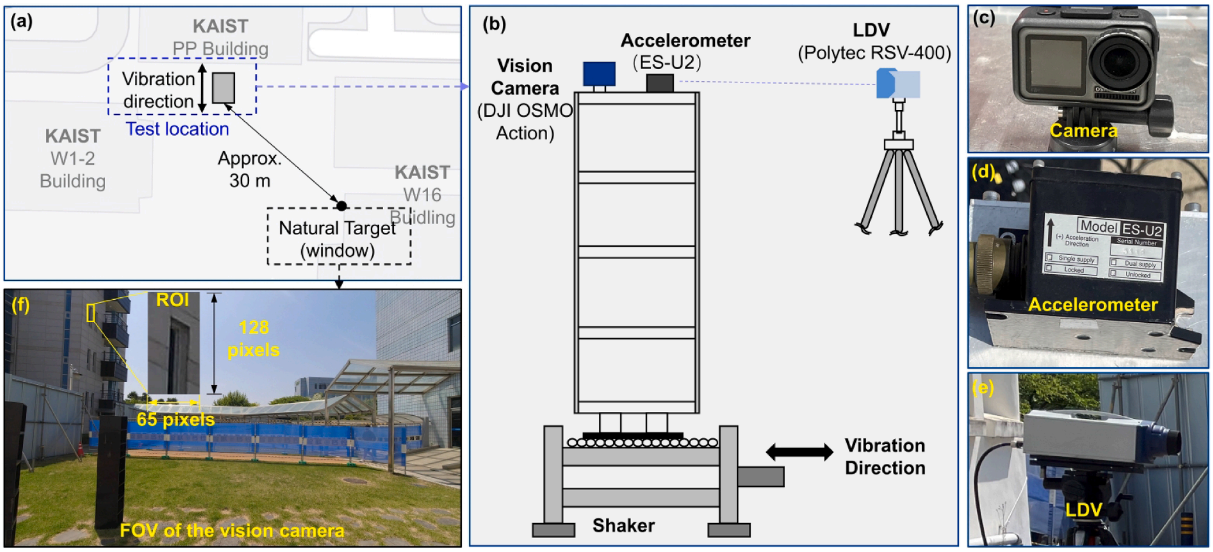


Fig. 21. Overall configuration of an outdoor four-story building structure test: (a) 30 m distance between the sensor location and the selected natural target (i.e., a window from a nearby building), (b) sensor setup, (c) DJI OSMO Action camera and (d) ES-U2 axial accelerometer used for displacement estimation, (e) Polytec RSV-150 LDV used to measure reference displacement, and (f) the FOV of the monocular camera and the cropped ROI for displacement estimation.

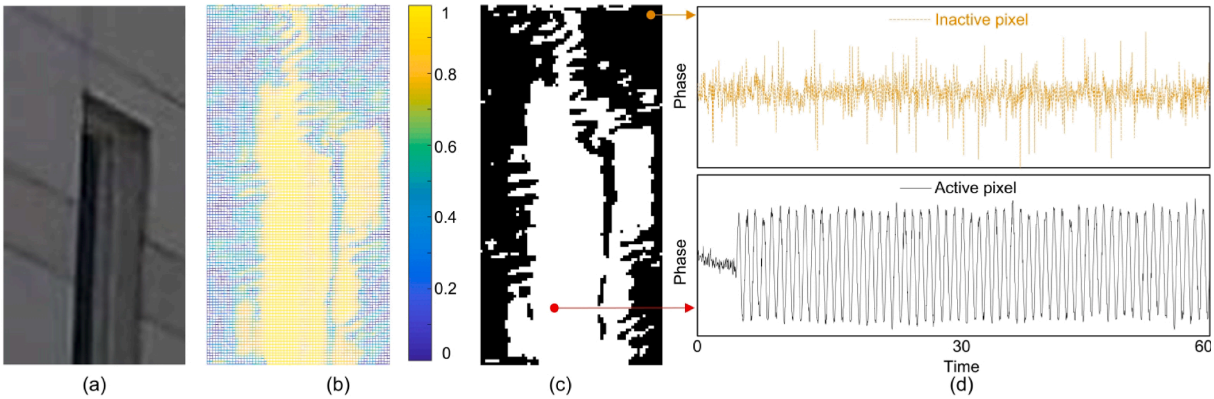


Fig. 22. Active pixel selection results in the outdoor four-story building structure test: (a) selected ROI, (b) generated correlation map, (c) locations of selected active pixels, and (d) phases of an active pixel and an inactive pixel.

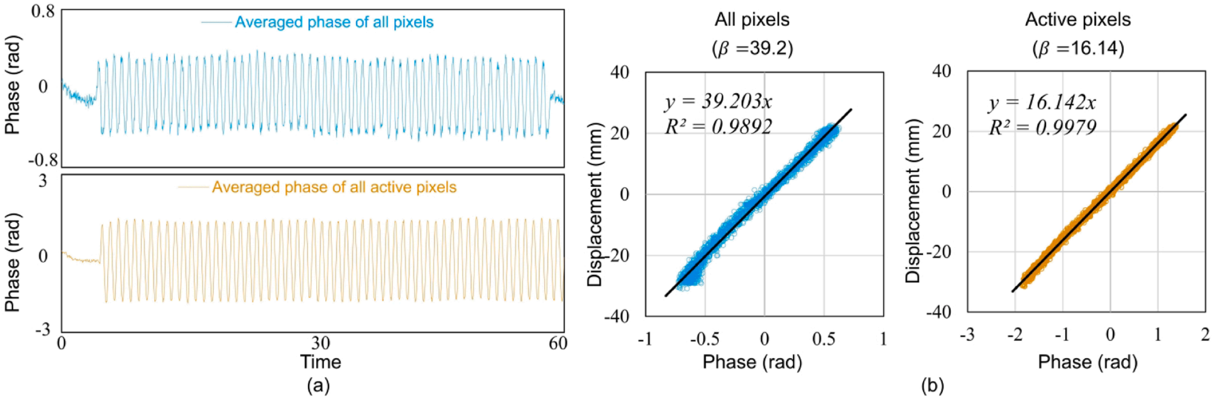


Fig. 23. Scale factor estimation results in the outdoor four-story building structure test: (a) averaged phase of all pixels and all active pixels and (b) scale factors estimated using all pixels and all active pixels.

building structure was installed on an APS-400 shaking table, which generated horizontal movement for the building structure (Fig. 21(b)). A DJI OSMO Action camera (Fig. 21(c)) and EpiSensor ES-U2 force-balance-type uniaxial accelerometer (Fig. 21(d)) were rigidly mounted on the top of the building structure for displacement estimation at the same point. A Polytech RSV-150 LDV was used to measure the displacements of the building structure (Fig. 21(e)). Other experimental setups were the same as in the indoor 10-meter-long bridge structure test. Fig. 21(e) shows the FOV of the vision camera. An ROI was selected to cover a window at a distance of approximately 30 m, and the horizontal movements of the window borders in the image plane were used for displacement estimation. Five different excitation signals were inputted to the shaking table in this test: (1) actual recorded bridge vibration signal, (2) 0.1 Hz sinusoidal signal, (3) 0.3 Hz sinusoidal signal, (4) 0.5 Hz sinusoidal signal, and (5) 1 Hz sinusoidal signal.

4.3.2. Initial calibration results

Considering the long target-to-camera distance, the phase-based algorithm was adopted for this test. Fig. 22 shows the active pixel selection results using accelerations and vision images recorded under a 1 Hz sinusoidal signal excitation in the outdoor four-story building structure test. Fig. 22(a)–(c) show the selected ROI, generated correlation map, and locations of the selected active pixels, respectively, while Fig. 22(d) shows representative phases extracted from the active and inactive pixels. A 1 Hz sinusoidal wave was observed from the phase of the active pixel, but was not observed from the phase of the inactive pixel.

Fig. 23 shows the scale factor estimation results using all pixels and active pixels selected by the proposed algorithm. The average phase of the active pixels has a larger amplitude and lower noise level than the average phase of all pixels, as shown in Fig. 23(a). Therefore, a smaller scale factor was estimated using active pixels with a better goodness of fit (i.e., a larger R^2 value) than that obtained using all pixels. The estimated scale factors using all pixels and active pixels were 39.2 mm/rad and 16.14 mm/rad, respectively.

4.3.3. Displacement estimation results

To more intuitively show the advantage of using active pixels selected by the proposed algorithm than using all pixels, the displacements estimated using all pixels and active pixels are compared in both frequency and time domains in Fig. 24 under all five excitations, and the corresponding errors are compared in Fig. 25. When all pixels were used, displacements were estimated with a maximum RMSE of 2.59 mm and minimum RMSE of 1.32 mm. More accurate displacements were estimated using active pixels, and the RMSEs of the estimated displacements were less than 1 mm in all cases.

The displacement estimation performance can be further improved by fusing vision-based displacement with acceleration using an adaptive multi-rate Kalman filter. Considering that the displacement estimation performance using only vision images was already sufficiently good, the RMSE reduction achieved by the adaptive multi-rate Kalman filter-based fusion was not significant. However, the RMSEs were still reduced by up to 16% (Table 4). Note that the Kalman filter was used to improve displacement estimation accuracy and increase the sampling rate of the estimated displacement. If only low-frequency displacement estimation is needed and short-distance targets are available for the camera, the use of the Kalman filter may not be necessary.

Additionally, the target size change under 1 Hz sinusoidal excitation was estimated using a feature-matching algorithm, and the results are

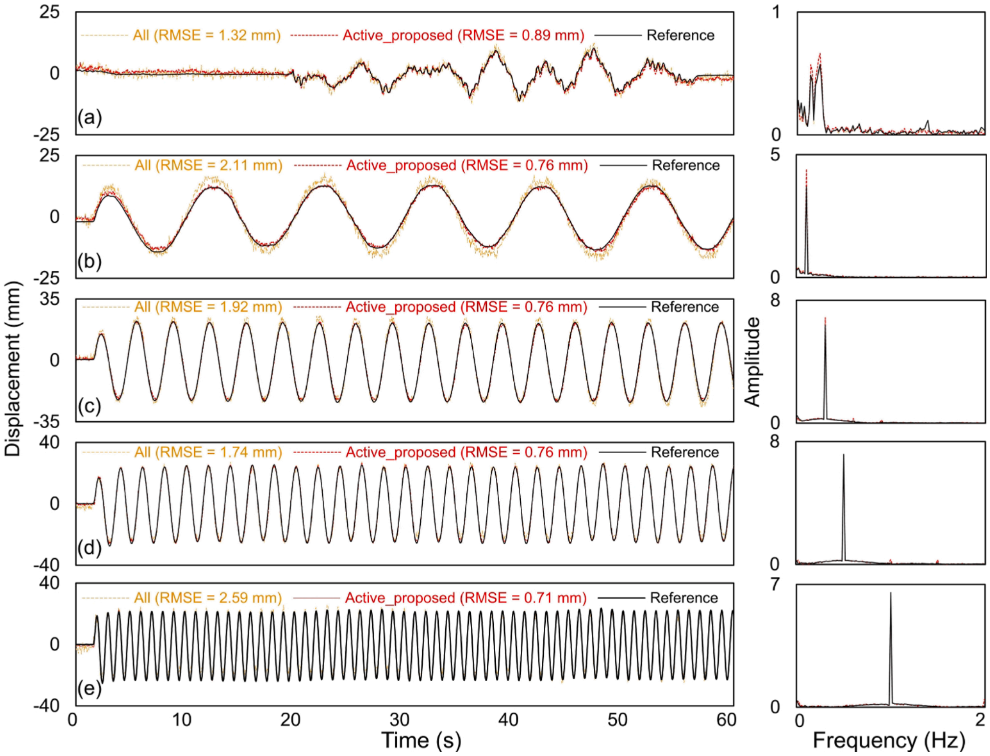


Fig. 24. Out-of-plane displacements estimated using all pixels and active pixels in the outdoor four-story building structure test: (a) actual recorded bridge vibration signal, (b) 0.1 Hz, (c) 0.3 Hz, (d) 0.5 Hz, and (e) 1 Hz.

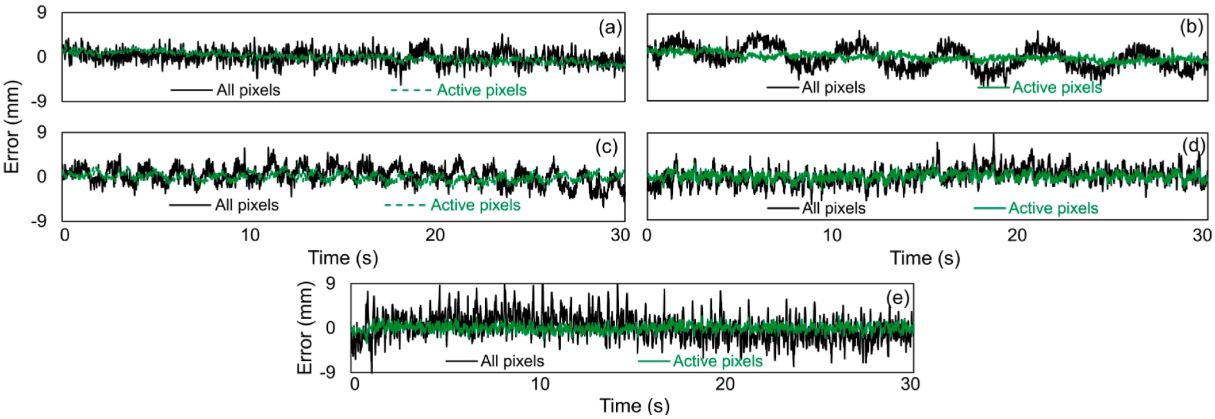


Fig. 25. Displacement estimation errors using all pixels and active pixels in the outdoor four-story building structure test: (a) actual recorded bridge vibration signal, (b) 0.1 Hz, (c) 0.3 Hz, (d) 0.5 Hz, and (e) 1 Hz.

Table 4
Comparison of displacement estimation performance (i.e., RMSE) using vision measurement only and the fusion of vision and acceleration measurement.

Excitations	Vision only (mm)	Vision + Acceleration (mm)	Difference (%)
Recorded bridge vibration signal	0.89	0.85	4.49
0.1 Hz sinusoidal	0.76	0.75	1.32
0.3 Hz sinusoidal	0.76	0.64	15.79
0.5 Hz sinusoidal	0.76	0.66	13.46
1 Hz sinusoidal	0.71	0.60	15.50
Average	0.78	0.72	7.69

shown in Fig. 26. Owing to the long target-to-camera distance, the estimated target size change did not have a 1 Hz sinusoidal wave, indicating difficulty in estimating the out-of-plane structural displacement from the target size change.

5. Conclusions

This paper describes a structural displacement estimation method that fuses measurements from a monocular camera and an accelerometer mounted on a target structure. A computer vision algorithm and adaptive multi-rate Kalman filter are integrated to efficiently estimate high-sampling displacements from low-sampling vision measurements and high-sampling acceleration measurements. All parameters associated with the computer vision algorithm are automatically calibrated, and the proposed method is suitable for in-plane, out-of-plane, and 3D displacement estimation. A two-story building structure test was first

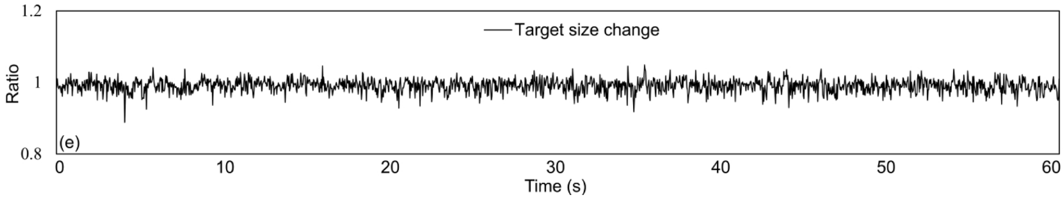


Fig. 26. Estimated target size change in the outdoor four-story building structure test under 1 Hz sinusoidal signal excitation.

conducted with bidirectional excitations, and the in-plane and out-of-plane displacements of the structure were simultaneously estimated using the proposed method with RMSEs below 0.6 mm. The proposed method was then applied to a 10-m-long bridge structure. Although the structure had tiny out-of-plane displacements of a few millimeters, they were still well estimated with RMSEs of less than 0.4 mm. Finally, a four-story building structure test was conducted. Even when a target at a distance of approximately 30 m was used, the proposed methods accurately estimated the displacements, with RMSEs below 0.9 mm. However, there are still the following issues that need to be addressed in future works:

(1) when estimating the structural displacement using a vision camera installed on a target structure, even a tiny structural rotation may cause large errors in the estimated structural displacements. Future studies should investigate the simultaneous estimation of 3D displacements and 3D rotations.
(2) the proposed technique aims at real-time structural displacement estimation. However, because the accelerometer and camera used in this study did not support real-time data streaming, acceleration and vision measurements were separately recorded and then post-processed on the desktop PC. We are now developing a displacement sensor module by integrating a photodetector for vision imaging, an accelerometer, and a microcontroller into a single unit. After the hardware development, the real-time estimation ability of the proposed technique needs to be further validated.

CRediT authorship contribution statement

Zhanxiong Ma: Conceptualization, Methodology, Software, Validation, Writing – original draft. **Jaemook Choi:** Validation, Writing – review & editing. **Hoon Sohn:** Supervision, Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This work was supported by National Research Foundation of Korea (NRF) grants funded by the Korean government (MSIT) [grant numbers 2017R1A5A1014883 and 2022R1C1C2008186).

References

[1] AASHTO. AASHTO LRFD Bridge Design Specifications, Washington, D.C.: the American Association of State Highway and Transportation Officials; 2017.
[2] Mltm. Korea highway bridge design code (in Korean), Seoul: Ministry of Land. Infrastructure and Transport 2010.
[3] Dong C-Z, Bas S, Catbas FN. Investigation of vibration serviceability of a footbridge using computer vision-based methods. Eng Struct 2020;224:111224. https://doi.org/10.1016/j.engstruct.2020.111224.
[4] Siringoringo DM, Wangchuk S, Fujino Y. Noncontact operational modal analysis of light poles by vision-based motion-magnification method. Eng Struct 2021;244:112728. https://doi.org/10.1016/j.engstruct.2021.112728.
[5] Zare Hosseinzadeh A, Tehrani MH, Harvey PS. Modal identification of building structures using vision-based measurements from multiple interior surveillance cameras. Eng Struct 2021;228:111517.
[6] Martini A, Tronci EM, Feng MQ, Leung RY. A computer vision-based method for bridge model updating using displacement influence lines. Eng Struct 2022;259:114129. https://doi.org/10.1016/j.engstruct.2022.114129.
[7] Nassif HH, Gindy M, Davis J. Comparison of laser Doppler vibrometer with contact sensors for monitoring bridge deflection and vibration. NDT and E Int 2005;38:213–8. https://doi.org/10.1016/j.ndteint.2004.06.012.
[8] Santhosh KV, Roy BK. Online implementation of an adaptive calibration technique for displacement measurement using LVDT. Appl Soft Comput 2017;53:19–26. https://doi.org/10.1016/j.asoc.2016.12.032.
[9] Gomez F, Park J-W, Spencer BF. Reference-free structural dynamic displacement estimation method. Struct Control Health Monit 2018;25(8):e2209.
[10] Gindy M, Vaccaro R, Nassif H, Velde J. A state-space approach for deriving bridge displacement from acceleration. Comput Aided Civ Inf Eng 2008;23:281–90. https://doi.org/10.1111/j.1467-8667.2007.00536.x.
[11] Tamura Y, Matsui M, Pagnini L-C, Ishibashi R, Yoshida A. Measurement of wind-induced response of buildings using RTK-GPS. J Wind Eng Ind Aerodyn 2002;90:1783–93. https://doi.org/10.1016/S0167-6105(02)00287-8.
[12] Nakamura S. GPS measurement of wind-induced suspension bridge girder displacements. J Struct Eng 2000;126:1413–9. https://doi.org/10.1061/(ASCE)0733-9445(2000)126:12(1413).
[13] Reu PL, Rohe DP, Jacobs LD. Comparison of DIC and LDV for practical vibration and modal measurements. Mech Syst Sig Process 2017;86:2–16. https://doi.org/10.1016/j.ymssp.2016.02.006.
[14] Gentile C, Bernardini G. An interferometric radar for non-contact measurement of deflections on civil engineering structures: laboratory and full-scale tests. Struct Infrastruct Eng 2010;6:521–34. https://doi.org/10.1080/15732470903068557.
[15] Rodrigues DVQ, Zuo D, Li C. Wind-induced displacement analysis for a traffic light structure based on a low-cost Doppler radar array. IEEE Trans Instrum Meas 2021;70:1–9. https://doi.org/10.1109/TIM.2021.3098380.
[16] Dong C-Z, Celik O, Catbas FN. Marker-free monitoring of the grandstand structures and modal identification using computer vision methods. Struct Health Monitor-an Int J 2019;18:1491–509. https://doi.org/10.1177/1475921718806895.
[17] Yoon H, Shin J, Spencer Jr BF. Structural displacement measurement using an unmanned aerial system. Comput Aided Civ Inf Eng 2018;33:183–92. https://doi.org/10.1111/mice.12338.
[18] Khuc T, Catbas FN. Computer vision-based displacement and vibration monitoring without using physical target on structures. Struct Infrastruct Eng 2017;13:505–16. https://doi.org/10.1080/15732479.2016.1164729.
[19] Yu S, Zhang J. Fast bridge deflection monitoring through an improved feature tracing algorithm. Comput Aided Civ Inf Eng 2020;35:292–302. https://doi.org/10.1111/mice.12499.
[20] Xu Y, Zhang J, Brownjohn J. An accurate and distraction-free vision-based structural displacement measurement method integrating Siamese network based tracker and correlation-based template matching. Measurement 2021;179:109506. https://doi.org/10.1016/j.measurement.2021.109506.
[21] Jeong J-H, Jo H. Real-time generic target tracking for structural displacement monitoring under environmental uncertainties via deep learning. Struct Control Health Monit 2022;29:e2902.
[22] Lee JJ, Shinozuka M. A vision-based system for remote sensing of bridge displacement. NDT and E Int 2006;39:425–31. https://doi.org/10.1016/j.ndteint.2005.12.003.
[23] Feng D, Feng MQ. Experimental validation of cost-effective vision-based structural health monitoring. Mech Syst Sig Process 2017;88:199–211. https://doi.org/10.1016/j.ymssp.2016.11.021.
[24] Ri S, Fujigaki M, Morimoto Y. Sampling Moiré Method for Accurate Small Deformation Distribution Measurement. Exp Mech 2010;50:501–8. https://doi.org/10.1007/s11340-009-9239-4.
[25] Ri S, Wang Q, Tsuda H, Shirasaki H, Kuribayashi K. Displacement measurement of concrete bridges by the sampling Moiré method based on phase analysis of repeated pattern. Strain 2020;56:e12351.
[26] Felipe-Sesé L, Siegmann P, Díaz FA, Patterson EA. Simultaneous in-and-out-of-plane displacement measurements using fringe projection and digital image correlation. Opt Lasers Eng 2014;52:66–74. https://doi.org/10.1016/j.optlaseng.2013.07.025.

[27] Siegmann P, Álvarez-Fernández V, Díaz-Garrido F, Patterson EA. A simultaneous in- and out-of-plane displacement measurement method. Opt Lett, OL 2011;36:10–2. https://doi.org/10.1364/OL.36.000010.

[28] Shao Y, Li L, Li J, An S, Hao H. Target-free 3D tiny structural vibration measurement based on deep learning and motion magnification. J Sound Vib 2022;538:117244. https://doi.org/10.1016/j.jsv.2022.117244.

[29] Shao Y, Li L, Li J, An S, Hao H. Computer vision based target-free 3D vibration displacement measurement of structures. Eng Struct 2021;246:113040. https://doi.org/10.1016/j.engstruct.2021.113040.

[30] Ma Z, Chung J, Liu P, Sohn H. Bridge displacement estimation by fusing accelerometer and strain gauge measurements. Struct Control Health Monit 2021;28:e2733.

[31] Ma Z, Choi J, Jang J, Kwon O, Sohn H, Yang Y. Simultaneous estimation of submerged floating tunnel displacement and mooring cable tension through FIR filter based strain and acceleration fusion. Struct Control Health Monit 2023;2023:1–21.

[32] Ma Z, Choi J, Yang L, Sohn H. Structural displacement estimation using accelerometer and FMCW millimeter wave radar. Mech Syst Sig Process 2023;182:109582. https://doi.org/10.1016/j.ymssp.2022.109582.

[33] Ma Z, Choi J, Sohn H. Continuous bridge displacement estimation using millimeter-wave radar, strain gauge and accelerometer. Mech Syst Sig Process 2023;197:110408.

[34] Park J-W, Moon D-S, Yoon H, Gomez F, Spencer Jr BF, Kim JR. Visual–inertial displacement sensing using data fusion of vision-based displacement with acceleration. Struct Control Health Monit 2018;25:e2122.

[35] Xu Y, Brownjohn JM, Huseynov F. Accurate deformation monitoring on bridge structures using a cost-effective sensing system combined with a camera and accelerometers: Case study. J Bridg Eng 2019;24:05018014. https://doi.org/10.1061/(ASCE)BE.1943-5592.0001330.

[36] Ma Z, Choi J, Sohn H. Real-time structural displacement estimation by fusing asynchronous acceleration and computer vision measurements. Comput Aided Civ Inf Eng 2022;37:688–703. https://doi.org/10.1111/mice.12767.

[37] Ma Z, Choi J, Liu P, Sohn H. Structural displacement estimation by fusing vision camera and accelerometer using hybrid computer vision algorithm and adaptive multi-rate Kalman filter. Autom Constr 2022;140:104338. https://doi.org/10.1016/j.autcon.2022.104338.

[38] Wadhwa N, Rubinstein M, Durand F, Freeman WT. Phase-based video motion processing. ACM Trans Graph 2013;32(1–80):10. https://doi.org/10.1145/2461912.2461966.

[39] Otsu N. A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybernet 1979;9(1):62–6.

[40] KAIS Co., Ltd. KL3 series Laser Displacement Sensor Datasheet 2019.

[41] Alcantarilla PF, Bartoli A, Davison AJ. KAZE Features. In: Fitzgibbon A, Lazebnik S, Perona P, Sato Y, Schmid C, editors. Computer Vision – ECCV 2012, Berlin, Heidelberg: Springer; 2012, p. 214–27. Doi: 10.1007/978-3-642-33783-3_16.