DOI: 10.1111/mice.12767

RESEARCH ARTICLE

COMPUTER-AIDED CIVIL AND INFRASTRUCTURE ENGINEERING WILEY

Real-time structural displacement estimation by fusing asynchronous acceleration and computer vision measurements

Zhanxiong Ma | **Jaemook Choi** | **Hoon Sohn**

Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology, Yuseong-gu, Daejeon, Republic of Korea

Correspondence
Hoon Sohn, Department of Civil and Environmental Engineering, Korea Advanced Institute Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea.
Email: hoonsohn@kaist.ac.kr

Funding information
National Research Foundation of Korea, the Korean Government (MSIT), Grant/Award Number: 2017R1A5A1014883

Abstract

Although displacement measurement is essential for many civil infrastructure applications, the precise estimation of structural displacement remains a challenge. In this study, a structural displacement estimation technique was developed by fusing asynchronous acceleration and computer vision measurements using a Kalman filter. First, the scale factor, which converts translation from vision measurements (in pixel units) into displacement (in length units), is automatically computed using a natural target (i.e., without any artificial target or any prior knowledge of the target size). Second, an improved feature matching algorithm was developed to better trace the natural target within the computer vision. Third, an adaptive multirate Kalman filter was formulated such that asynchronous computer vision and acceleration measurements with different sampling rates could be seamlessly combined to improve displacement estimation. The feasibility and effectiveness of the proposed displacement estimation technique were validated by performing shaking table, four-story building model, and steel box girder pedestrian bridge tests. In all tests, the proposed technique was able to accurately estimate displacements with root mean square errors of less than 3 mm.

1 | INTRODUCTION

Displacement plays a vital role in the assessment, monitoring, and control of civil infrastructure, as it provides crucial information relating to structural integrity and health (D. Feng & M. Q. Feng, 2016). For instance, a finite-element model is updated using in situ measurements of dynamic displacement of a bridge under train loads (D. Feng & M. Q. Feng, 2015). Vehicle-induced displacement has been used for bridge damage detection (Feng et al., 2016) and bridge condition assessment (Moreu et al., 2015). Several techniques are available for measuring direct displacement measurements. A linear variable differential transformer (LVDT) is typically placed between a target point on a structure and a fixed reference point to measure the relative displacement (Nassif et al., 2005). Although the LVDT can measure displacement with a high accuracy, its installation requires an additional scaffold, which is not convenient in real application. The use of a real-time kinematic global navigation satellite system (RTK-GNSS) is another option for displacement measurements. A set of GNSS antennae and receiver was placed on a target point, and the other set on a fixed reference point. The antennas at both points simultaneously receive carrier waves from multiple satellites, thereby calculating the displacement at the target point based on phase differences between

the carrier waves received at the two measurement points (Nakamura, 2000). However, RTK-GNSS typically has a low-sampling rate of up to 20 Hz, and its accuracy is limited to approximately 7–10 mm. Moreover, it cannot work in GNSS-denied environments. Terrestrial laser scanning has also been adopted for displacement measurement (Park et al., 2007). It can measure displacement based on the time-of-flight information or the phase-shift of the reflected laser beam from a target surface, but it typically has a high-noise level and a low-sampling rate.

Displacement has also been estimated by double integration of measured acceleration. However, this acceleration-based displacement estimation amplifies the displacement estimation error in a low-frequency range, producing a large low-frequency drift in estimated displacement (Gindy et al., 2008; Park et al., 2005). Several techniques have been proposed to remove such low-frequency drift (Gomez et al., 2018; Lee et al., 2010), but these techniques also remove important structural responses within the same low-frequency range.

With the rapid development of computer vision, vision cameras have been used as noncontact sensors, and applied to bridge component recognition (Narazaki et al., 2020), construction vehicle detection (Arabi et al., 2020), damage detection (Pan & Yang, 2020; Zhang et al., 2020), defect segmentation (Wang & Cheng, 2020), and structural displacement estimation (Park et al., 2015). For displacement estimation, a vision camera tracks the motion of an artificial or natural target on a target structure based on template matching techniques (M. Q. Feng et al., 2015; Luo & Feng, 2018), optical flow techniques (Dong et al., 2018; Yoon et al., 2018), and feature matching algorithms (Khuc & Catbas, 2016; Yu & Zhang, 2019). In vision-based techniques, the translation is first estimated from vision measurements in pixel units and converted into displacement in length units using a scale factor. The scale factor can be estimated by identifying the size of the target in physical units in advance. However, manual measurement of the target size in the field can be challenging. In addition, the high computational cost of these techniques hinders the real-time estimation of displacement with a high-sampling rate.

Data-fusion techniques have been proposed for displacement estimation. These techniques combine multiple measurements to produce more accurate displacement estimation. Examples include the fusion of RTK-GNSS and accelerometers (Kim et al., 2016) as well as the fusion of strain gauges and accelerometers (Park et al., 2013, 2016). However, the former cannot be used in GNSS-denied environments, and the latter requires prior knowledge of the mode shapes of the target structure. Attempts have also been made to fuse vision cameras and accelerometers (Chang & Xiao, 2010; Xu et al., 2019). A low-sampling rate displacement from vision measurement and a high-sampling rate acceleration measurement are fused by a multirate Kalman filter (Smyth & Wu, 2007) to estimate a high-sampling rate displacement. However, these techniques still require prior knowledge of the actual length of a target object to estimate the scale factor, and they are applicable only to synchronous measurements.

The authors previously estimated structural displacement (1) by fusing synchronous acceleration and LiDAR/GNSS displacement measurements using a multirate Kalman filter (Kim et al., 2014, 2016), and (2) by fusing synchronous acceleration and strain measurements using a finite impulse response (FIR) filter (Ma et al., 2021). In this study, a new Kalman filter based technique is proposed for long-term continuous monitoring of structural displacement by fusing asynchronous computer vision and acceleration measurements. First, a scale factor is computed automatically from acceleration and vision measurements. Next, an improved feature matching algorithm is developed and applied to vision measurements to estimate a low-sampling rate displacement. Finally, the low-sampling rate displacement is fused with the high-sampling rate acceleration measurement using an adaptive multirate Kalman filter to estimate a high-sampling rate displacement with improved accuracy. The uniqueness of this study lies in the following: (1) the scale factor is computed automatically, (2) an adaptive multirate Kalman filter is developed to fuse asynchronous acceleration and vision measurements and estimate high-sampling rate displacement, (3) a region-of-interest (ROI) updating technique is proposed to improve the matching of a natural target in sequential vision images, and (4) automatic mismatch rejection algorithms are proposed to further advance feature matching.

The remainder of this paper is organized as follows. The research background is presented in Section 2, and the working principle of the proposed technique is explained in Section 3. In Section 4, the feasibility of the proposed technique is experimentally validated using a shaking table, four-story building model, and steel box girder pedestrian bridge tests. The concluding remarks are presented in Section 5.

2 | RESEARCH BACKGROUND

2.1 | Vision-based displacement estimation using feature matching algorithm

Figure 1 shows an overview of feature matching-based displacement estimation. A vision camera is usually installed at a stationary location near the target structure and is

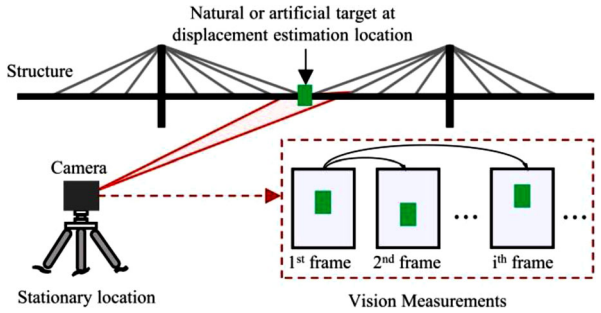

FIGURE 1 Overview of displacement estimation technique using feature matching algorithm

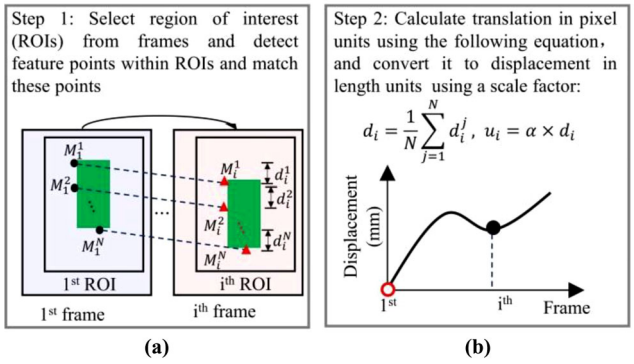

FIGURE 2 Displacement estimation procedure using feature matching algorithm: (a) Step 1: region-of-interest (ROI) selection, and feature detection and matching, and (b) Step 2: displacement estimation in length units

aimed at the displacement estimation location. A natural or artificial target should be available at displacement estimation location, and the target should be within the field of view (FOV) of the camera. Note that natural targets are those already present on the structure such as structural components, while artificial targets are those designed and mounted on the structure specifically for displacement estimation. The detailed procedure for estimating the target displacement at *i*th frame is described in Figure 2. Note that the proposed technique is applicable only to estimation of in-plane motion, and only the vertical movement is considered explicitly in this study. However, the proposed technique can be easily extended to estimate the horizontal movement as well.

As shown in Figure 2a, the size and location of the ROI are first selected from the first frame to include a target of interest (TOI), and the same ROI is maintained within the subsequent frames. Feature points, $M_1^1, M_1^2, \ldots$, are detected within the first ROI and matched with the corresponding features, $M_i^1, M_i^2, \ldots$ within the *i*th ROI using speeded up robust features (SURF) algorithm (Bay et al., 2006). The SURF algorithm is selected here because of its high accuracy and low computational cost. Note that

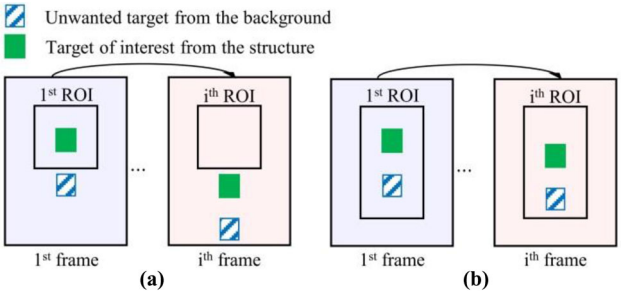

FIGURE 3 Limitation using a fixed ROI: (a) fixed ROI is too small such that the target of interest (TOI) is outside ROI and (b) fixed ROI is too large such that unwanted target is also inside ROI

two matched feature points (e.g., M_1^1 and M_i^1 depicted in Figure 2a) are named as a match.

Then, the relative movement of the *j*th match between the first and *i*th frames, d_i^j, is computed, and the averaged relative translation, d_i, is computed as follows (Figure 2b) (Khuc & Catbas, 2016):

$$d_i = \frac{1}{N}\sum_{j=1}^{N} d_i^j \tag{1}$$

where N denotes the number of matches. Finally, d_i in pixel units is converted to displacement, u_i, in length units using a scale factor α (Figure 2b).

$$u_i = \alpha \times d_i \tag{2}$$

The scale factor is estimated using the following equation (Lee & Shinozuka, 2006):

$$\alpha = \frac{u_{\text{target}}}{d_{\text{target}}} \tag{3}$$

where u_{target} and d_{target} represent the physical size of the artificial or natural target and the corresponding size of the same target in pixel units, respectively.

However, this conventional displacement estimation technique has several limitations. First, the maximum displacement should be estimated in advance to determine the size of the ROI. Because the location of the ROI is fixed for all frames, the TOI can move outside the ROI when the TOI movement is too large compared to the ROI size, as shown in Figure 3a (Yoon et al., 2016). Nevertheless, enlarging the ROI size, as shown in Figure 3b, becomes computationally more demanding and increases the possibility of capturing an unwanted target inside the ROI. Second, conventional feature matching algorithms often produce mismatches that impair the displacement estimation accuracy. When two unrelated features (e.g., M_1^1 and M_i^2 in Figure 2b) are matched for displacement estimation, this

phenomenon is called a mismatch. Although mismatch rejection algorithms, such as KnnMatch (Lowe, 2004) and RANSAC (Fischler & Bolles, 1981), have been proposed, the implementation of these algorithms demands human interactions to establish the ad hoc threshold. In addition, RANSAC is computationally demanding and unsuitable for real-time estimation. Third, the actual length of a target must be known in advance to estimate the scale factor (Lee & Shinozuka, 2006), and such an estimation can be cumbersome for civil infrastructure applications.

2.2 | Fusion of acceleration and vision-based displacement using a multirate Kalman filter

Because displacement estimation based on vision measurement is computationally demanding, vision-based displacement estimation is often limited to a low-sampling rate. Attempts have been made to estimate displacement at a high-sampling rate by fusing low-sampling rate vision-based displacement and high-sampling rate acceleration using a multirate Kalman filter (Chang & Xiao, 2010; Xu et al., 2019). The working principle of the multi-rate Kalman filter is briefly summarized here (Smyth & Wu, 2007).

If $\dot{x}_k$ and x_k are the true velocity and displacement, respectively, at the kth time step, a discrete state space model for the acceleration-displacement relationship can be expressed as

$$x_k = A(\Delta t_a)\, x_{k-1} + B(\Delta t_a)\, a_{k-1} + B(\Delta t_a)\, w_{k-1} \quad (4)$$

$$u_k = H x_k + v_k \quad (5)$$

where x_k is the state variable $\{x_k,\ \dot{x}_k\}^T$, and a_{k-1} and u_k are the measured acceleration and vision-based displacement, respectively. w_{k-1} and v_k are the corresponding noises with variances of Q and R, respectively. Δt_a is the time interval of the acceleration measurement. H is the vector $[\,1\ 0\,]$. A and B are functions of the time interval Δt_a:

$$A(\Delta t_a) = \begin{bmatrix} 1 & \Delta t_a \\ 0 & 1 \end{bmatrix},\ B(\Delta t_a) = \begin{bmatrix} \Delta t_a^2/2 \\ \Delta t_a \end{bmatrix} \quad (6)$$

Based on the state space model, a multirate Kalman filter is formulated for displacement estimation using acceleration and vision-based displacement sampled at different sampling rates, but synchronously. Here, synchronous measurement indicates that the sampling rate of the acceleration measurement is an integer multiple of that of the vision-based displacement. The prior estimate ($\hat{x}_k^-$) is obtained using the posterior estimate and acceleration at the previous time step ($\hat{x}_{k-1}^+$ and a_{k-1}, respectively):

$$\hat{x}_k^- = A(\Delta t_a)\ \hat{x}_{k-1}^+ + B(\Delta t_a)\, a_{k-1} \quad (7)$$

and its error covariance matrix ($\hat{P}_k^-$) is also obtained as

$$\hat{P}_k^- = A(\Delta t_a)\, \hat{P}_{k-1}^+ A^T(\Delta t_a) + QB(\Delta t_a)\, B^T(\Delta t_a) \quad (8)$$

where $\hat{P}_{k-1}^+$ denotes the error covariance matrix of $\hat{x}_{k-1}^+$.

When vision-based displacement, u_k, and acceleration, are available at the kth time step, the Kalman gain (K) is calculated as

$$K = \hat{P}_k^- H^T \left(H \hat{P}_k^- H^T + R\right)^{-1} \quad (9)$$

and a posterior state estimate ($\hat{x}_k^+$) and its error covariance matrix ($\hat{P}_k^+$) are obtained as

$$\hat{x}_k^+ = (I - KH)\ \hat{x}_k^- + K u_k \quad (10)$$

$$\hat{P}_k^+ = (I - KH)\, \hat{P}_k^- \quad (11)$$

If only acceleration is available at the kth time step,

$$\hat{x}_k^+ = \hat{x}_k^- \quad (12)$$

$$\hat{P}_k^+ = \hat{P}_k^- \quad (13)$$

This Kalman filter is applicable only when acceleration and vision measurements are synchronized, and the ratio of their sampling rates is an integer. In addition, the performance of the Kalman filter relies significantly on the estimation of Q and R. Although the Q value for acceleration measurement can be estimated in a laboratory setting, the estimation of the R value for vision-based displacement is difficult because this value is highly influenced by light conditions, distance between the vision camera and the target, and texture of the target. Several algorithms are available for automatic estimation of Q and R (Wang et al., 2017; Xie et al., 2007; Zhi et al., 2016), and a covariance-matching based algorithm (Akhlaghi et al., 2017) is adopted in this study to develop an adaptive multirate Kalman filter. The details are given in Section 3.2.

3 | DEVELOPMENT OF PROPOSED DISPLACEMENT ESTIMATION TECHNIQUE

To address the limitations mentioned in Section 2, a new displacement estimation technique is proposed in

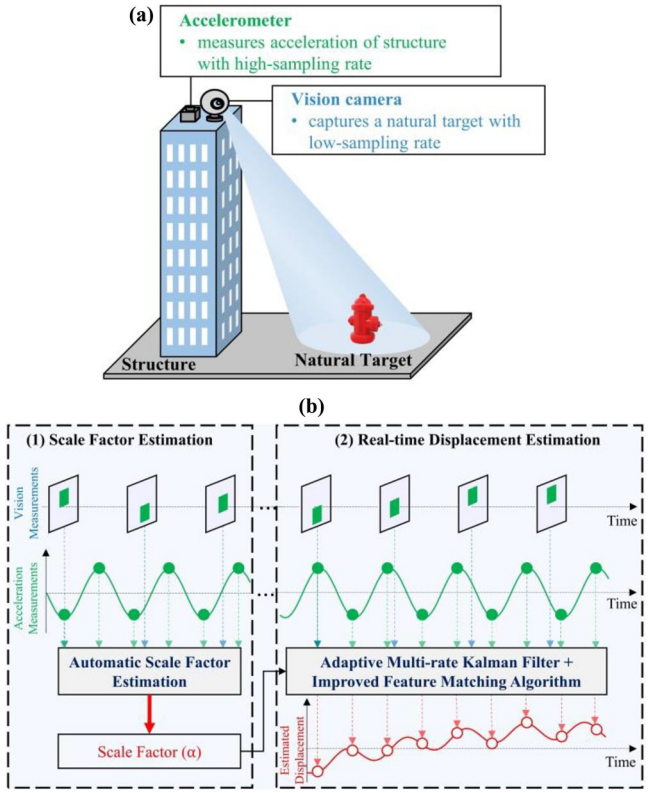

FIGURE 4 Overview of the proposed displacement estimation technique: (a) accelerometer and vision camera installation and (b) overall flowchart

this study, and an overview of the proposed technique is demonstrated in Figure 4. As shown in Figure 4a, a vision camera with a low-sampling rate was installed on a structure to trace a natural target around the structure, and an accelerometer was placed at the same location as the vision camera to measure acceleration with a high-sampling rate. The proposed technique shown in Figure 4b consists of two parts: (1) automatic scale factor estimation and (2) real-time displacement estimation. More details are provided in the following subsections.

3.1 | Automatic scale factor estimation

3.1.1 | Working principle

A flowchart of the proposed automatic scale factor estimation technique is shown in Figure 5. First, translation in the pixel unit is estimated from vision measurement using a feature matching algorithm described in Section 2.1, and displacement in length units is estimated from the acceleration measurement using double integration. Next, a filtered translation, $d^f(t)$, and a filtered displacement, $u^f(t)$, are obtained by band-pass filtering of the previously estimated translation and displacement, respectively. The

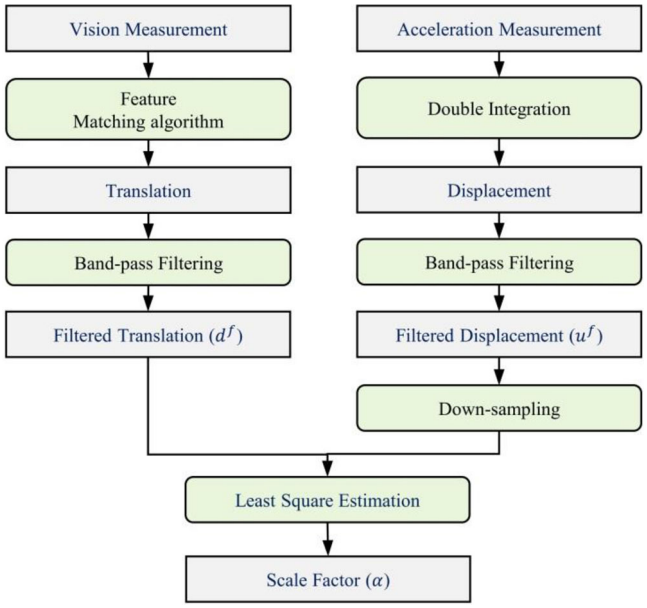

FIGURE 5 Flowchart of the proposed automatic scale factor estimation technique

lower cutoff frequency of the band-pass filter, f_{lc}, was set to be sufficiently high to eliminate the low-frequency drift in the acceleration-based displacement. The upper cutoff frequency, f_{uc}, is set to one-tenth of the vision camera's sampling rate so that the sampling rate can be at least 10 times faster than the highest frequency of the signal (Brandt, 2011). Then, the following relationship exists between the filtered translation and displacement:

$$u^f(t) = \alpha \times d^f(t) \tag{14}$$

where α is a scale factor, which is estimated using a least-squares algorithm. Note that u^f is downsampled to the sampling rate of the vision camera before the least-square estimation. Considering the relatively small target size and the relatively long distance between the camera and the target, an averaged scale factor is estimated here by assuming that the distances of the camera to all features within the small target are approximately equal.

3.1.2 | Development of mismatch rejection algorithm for improved displacement estimation

For translation estimation, it is important to properly match the relevant features, as shown in Figure 2b. However, conventional matching algorithms often produce mismatches (i.e., incorrectly matched features) that deteriorate the translation estimation accuracy and, consequently, the scale factor estimation accuracy. In

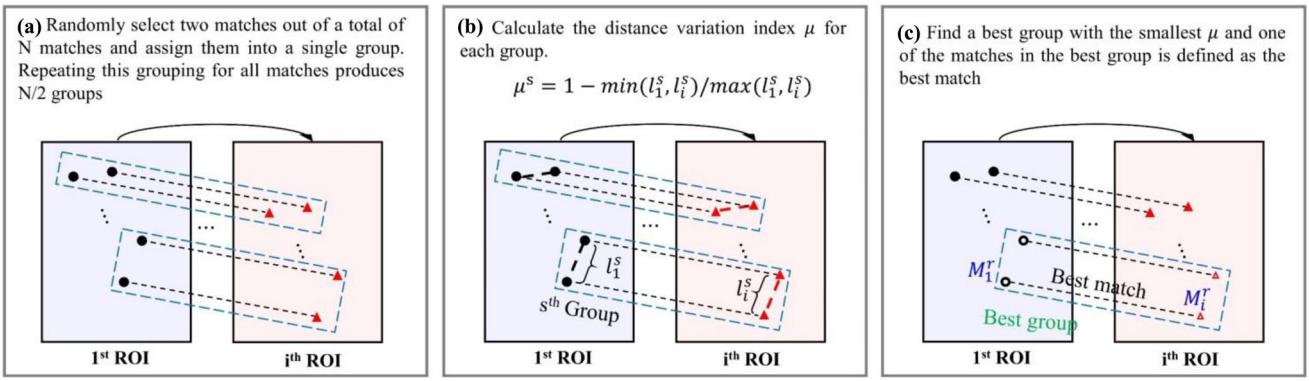

FIGURE 6 Selection of the best match: (a) group matches, (b) calculate length variation index, and (c) find a best match

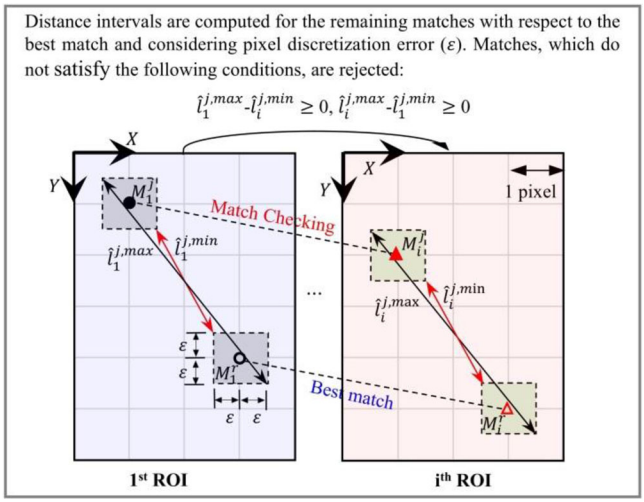

FIGURE 7 Cross-check of the remaining matches with respect to the best match for rejecting mismatches

this study, an automated mismatch rejection algorithm, comprising two steps, is proposed: (1) selecting the best match from the matching results (Figure 6) and (2) cross-checking the remaining matches with respect to the best match for rejecting mismatches (Figure 7). The algorithm is called mismatch rejection algorithm I hereafter, as the other algorithm will be introduced later in Section 3.3. Note that the basic assumption here is that the time variation of the distance between two points on the image plane is much smaller than the displacement at the target measurement point.

Step 1—*Selection of the best match*: Suppose that N matches are obtained between the first and the ith ROIs using a conventional feature matching algorithm. Then, two matches are randomly selected from these matches and assigned to a single group, and this grouping is repeated for all matches, producing $N/2$ groups (Figure 6a). Note that if N is an odd number, the last match is discarded in the grouping, producing$(N-1)/2$ groups. Next, the following distance variation index was calculated for each group (Figure 6b):

$$\mu^s = 1 - \min\left(l_1^s, l_i^s\right) / \max\left(l_1^s, l_i^s\right) \tag{15}$$

where μ^s denotes the distance variation index of the sth group, and l_1^s and l_i^s denote the distance between two feature points in the sth group at the first and ith ROIs, respectively. This index should be close to zero for a group with two good matches. Finally, the group with the smallest μ^s value is selected as the best group, and one out of the two matches in the best group is defined as the best match (denoted by $\{M_1^r, M_i^r\}$ in Figure 6c).

Step 2—*Mismatch rejection*: Once the best match, $\{M_1^r, M_i^r\}$, is selected, the true distance between feature M_1^j and the best match feature M_1^r within the first ROI should fall between $\hat{l}_1^{j,max}$ and $\hat{l}_1^{j,min}$ expressed as

$$\begin{cases} \hat{l}_1^{j,min} = \sqrt{\left(\left|\Delta x_1^j\right| - 2\varepsilon\right)^2 + \left(\left|\Delta y_1^j\right| - 2\varepsilon\right)^2} \\ \hat{l}_1^{j,max} = \sqrt{\left(\left|\Delta x_1^j\right| + 2\varepsilon\right)^2 + \left(\left|\Delta y_1^j\right| + 2\varepsilon\right)^2} \end{cases} \tag{16}$$

where $\hat{l}_1^{j,max}$ and $\hat{l}_1^{j,min}$ denote the maximum and minimum distance limits between features M_1^j and M_1^r, respectively. ε denotes the maximum pixel discretization error in both the x and y directions, and its value is set to 0.5 pixel. Similarly, the maximum and minimum distance limits ($\hat{l}_i^{j,max}$ and $\hat{l}_i^{j,min}$) between feature M_i^j and the best match feature M_i^r are computed for the ith ROI. Finally, any match $\{M_1^j, M_i^j\}$ ($j = 1, \ldots, N,$ and $j \neq r$) that does

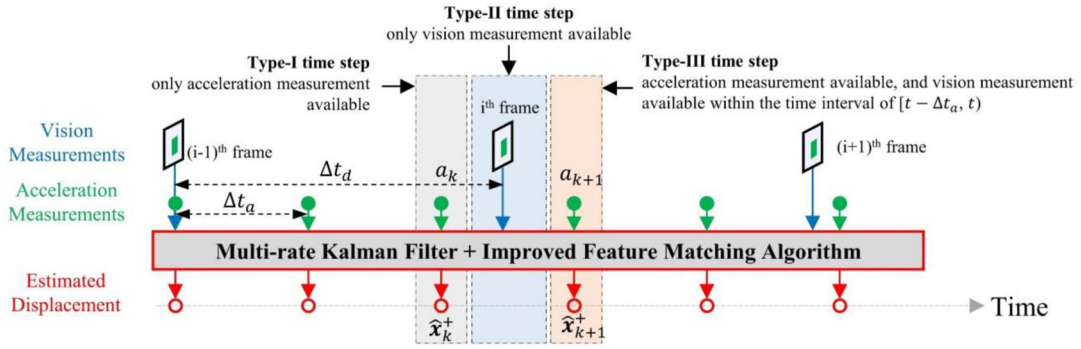

FIGURE 8 Overview of fusion of asynchronous acceleration and vision measurements using adaptive multirate Kalman filter

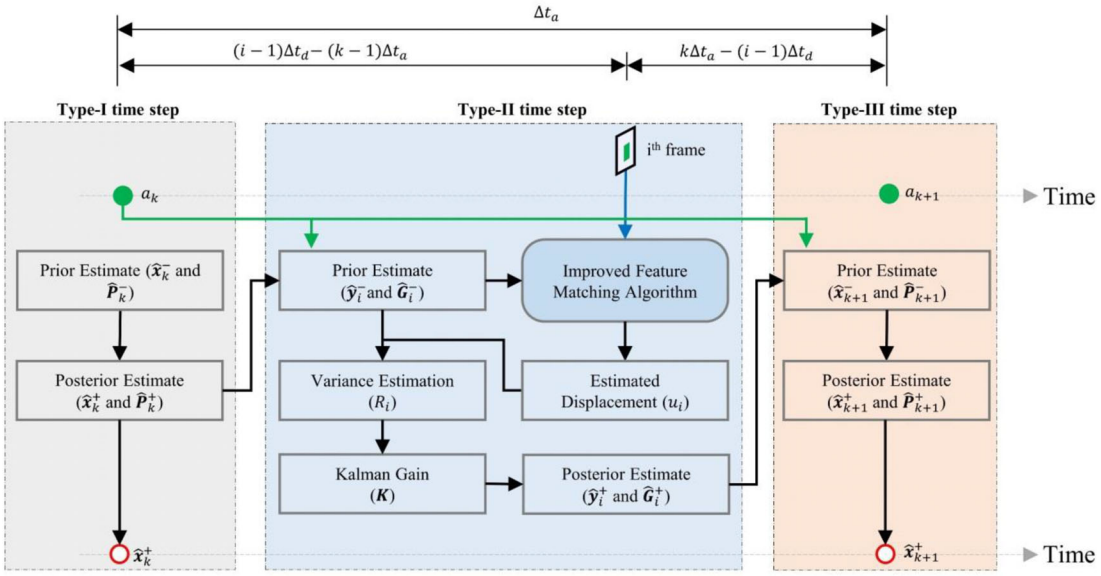

FIGURE 9 Example of fusion of asynchronous acceleration and vision measurements using adaptive multirate Kalman filter

not satisfy Equations (16) and (17) is classified as a mismatch and rejected.

$$\hat{l}_1^{j,max} - \hat{l}_i^{j,min} \geq 0, \ \hat{l}_i^{j,max} - \hat{l}_1^{j,min} \geq 0 \tag{17}$$

Note that the mismatch rejection algorithm occasionally fails to reject some mismatches whose distance satisfies Equation (17), and produces outliers. However, the effect of these few outliers on the scale factor estimation is negligible because the scale factor is estimated by a least-square estimation algorithm with more than 1000 measurement points and there can be only few outliers.

3.2 | Adaptive multirate Kalman filter for asynchronous acceleration and vision measurements fusion

As mentioned in Section 2.2, the conventional Kalman filter assumes that the measurements used for data fusion are synchronous. In this section, a new Kalman filter is developed such that (1) asynchronous measurements can be fused seamlessly, and (2) the noise variance of vision-based displacement, R, can be adaptively estimated.

An overview of the asynchronous fusion of acceleration and vision measurements using the proposed adaptive multirate Kalman filter is illustrated in Figure 8. All time steps are categorized into three types according to the availability of acceleration and vision measurements, and only the state estimates at acceleration steps are retained so that the final estimated displacement has the same sampling rate as the acceleration measurement.

The proposed Kalman filter is formulated for three different types of time steps, and an example is demonstrated in Figure 9 to describe the overall procedure. In Type-I time step, the state variable and its error covariance matrix are estimated using Equations (7) and (8) and Equations (12) and (13), respectively, similar to the conventional Kalman filter.

In Type-II time step, the state variable and its error covariance matrix are again estimated using Equations (7) and (8) by simply changing the time interval from Δt_a to $\{(i-1)\Delta t_d - (k-1)\Delta t_a\}$. Here, k denotes the kth time step of acceleration measurements; i denotes the ith time step of vision measurements; and Δt_a and Δt_d denote the time intervals of acceleration and vision measurement, respectively. Note that $\hat{y}_i^-$ and $\hat{G}_i^-$ denote the prior estimate and its error covariance matrix, respectively, in the Type-II time step. Using the prior estimate, an improved feature matching algorithm, which is later described in Section 3.3, is applied to the vision measurement to estimate displacement u_i. Then, the noise variance of u_i is adaptively estimated using a covariance-matching technique (Mohamed & Schwarz, 1999):

$$R_i = E\left(\eta_i \eta_i^T\right) - H\hat{G}_i^- H^T \tag{18}$$

where R_i denotes the noise variance of u_i, $E(\cdot)$ denotes the expectation operation, and innovation η_i is defined as

$$\eta_i = u_i - H\hat{y}_i^- \tag{19}$$

Although the expectation of $\eta_i \eta_i^T$ can be approximated by averaging η_i^2 within a moving time window, this expectation operation is computationally expensive. To ease the computational demand, a forgetting factor is introduced here for the adaptive estimation of R_i (Akhlaghi et al., 2017):

$$R_i = \beta R_{i-1} + (1-\beta)\left(\eta_i^2 - H\hat{G}_i^- H^T\right) \tag{20}$$

where β denotes the forgetting factor, and $0 < \beta < 1$. Finally, the Kalman gain (K) was calculated using Equation (9), and the posterior estimate and its error covariance matrix were obtained using Equations (10) and (11). Note that $\hat{y}_i^+$ and $\hat{G}_i^+$ denote the posterior estimate and its error covariance matrix, respectively, in the Type-II time step.

In the Type-III time step, the prior and posterior estimates are equal and are estimated as

$$\hat{x}_{k+1}^+ = \hat{x}_{k+1}^- = A\left(k\Delta t_a - (i-1)\Delta t_d\right)\hat{y}_i^+ + B\left(k\Delta t_a - (i-1)\Delta t_d\right)a_k \tag{21}$$

According to the relationship between $\hat{y}_i^+$, $\hat{x}_k^+$, u_i, and a_k, Equation (21) is rewritten as

$$\hat{x}_{k+1}^+ = W_1\,\hat{x}_k^+ + W_2 a_k + W_3 u_i \tag{22}$$

$$W_1 = A\left(k\Delta t_a - (i-1)\Delta t_d\right)(I - KH)\,A\left((i-1)\Delta t_d - (k-1)\Delta t_a\right) \tag{23}$$

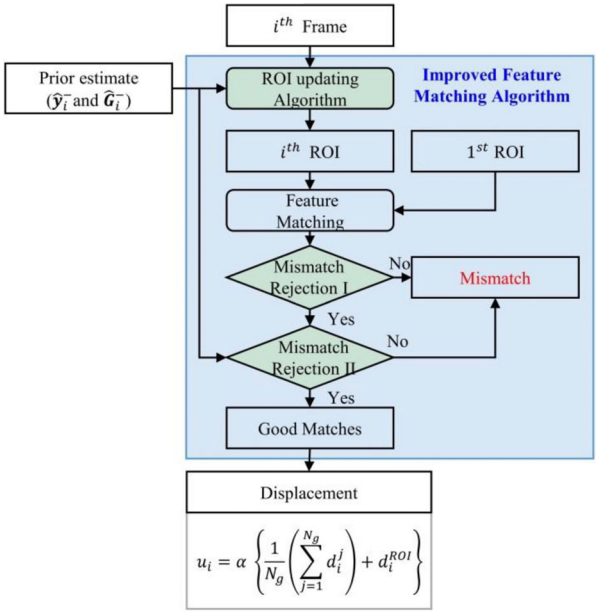

FIGURE 10 Overview of improved feature matching algorithm

$$W_2 = A\left(k\Delta t_a - (i-1)\Delta t_d\right)(I - KH)\,B\left((i-1)\Delta t_d - (k-1)\Delta t_a\right) + B\left(k\Delta t_a - (i-1)\Delta t_d\right) \tag{24}$$

$$W_3 = A\left(k\Delta t_a - (i-1)\Delta t_d\right)K \tag{25}$$

Because $\hat{x}_k^+$, a_k, and u_i are independent of each other, the error covariance matrix of $\hat{x}_{k+1}^+$ is calculated as

$$\hat{P}_{k+1}^+ = W_1\,\hat{P}_k^+ W_1^T + qW_2 W_2^T + R_i W_3 W_3^T \tag{26}$$

3.3 | Improved feature matching algorithm

An improved feature matching algorithm is proposed to better estimate the displacement from vision measurements. An overview of the proposed feature matching algorithm is illustrated in Figure 10. Using the prior estimate, $\hat{y}_i^-$, the ROI at the ith frame is first updated using the proposed algorithm, and then the conventional feature matching algorithm is applied to the first and ith ROIs. Mismatches are rejected by two mismatch rejection algorithms using $\hat{y}_i^-$ and $\hat{G}_i^-$, and only good matches were retained for displacement estimation. The major improvements over the existing feature matching algorithm include the development of ROI updating and two mismatch rejection algorithms. The mismatch rejection algorithm I is explained in Section 3.1.2, and the ROI updating algorithm and mismatch rejection algorithm II are proposed here.

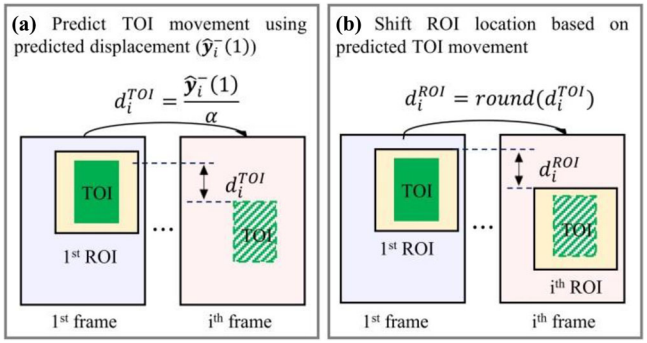

FIGURE 11 Working principle of ROI updating algorithm: (a) predict TOI movement and (b) shift ROI location based on the predicted TOI movement

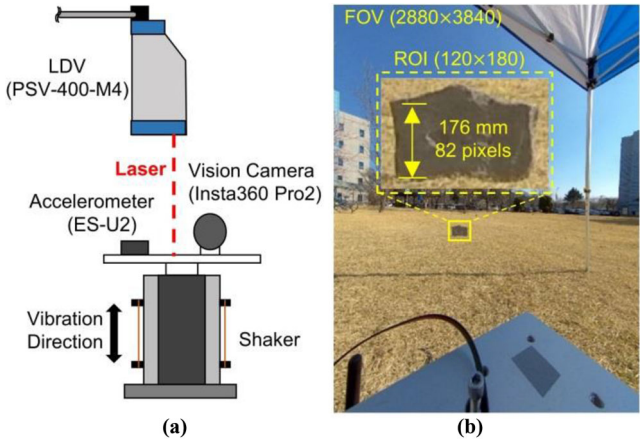

FIGURE 12 (a) Experimental setup of shaking table test and (b) field of view (FOV) of vision camera and initial ROI

Figure 11 demonstrates the working principle of the proposed ROI updating algorithm. First, the TOI movement in pixel units, d_i^{TOI}, is calculated using the predicted displacement in length units, $\hat{y}_i^-(1)$,

$$d_i^{\mathrm{TOI}} = \frac{\hat{y}_i^-(1)}{\alpha} \tag{27}$$

Note that $\hat{y}_i^-(1)$ denotes the first entity of $\hat{y}_i^-$. Then, the ROI within the ith frame is shifted based on d_i^{TOI}. Considering that the ROI movement (d_i^{ROI}) can only be discrete at pixel resolution, a *round* function is included:

$$d_i^{\mathrm{ROI}} = \mathrm{round}\left(\frac{\hat{y}_i^-(1)}{\alpha}\right) \tag{28}$$

After updating the ROI of the ith frame, feature matching between the first and ith ROIs is performed using the conventional feature matching algorithm. Then, the proposed mismatch rejection algorithms I and II were sequentially applied to the matched results. As mentioned in Section 3.1.1, mismatch rejection algorithm I occasionally fails to reject mismatches, and then the combination of the two mismatch rejection algorithms is essential for more robust mismatch rejection. The mismatch rejection algorithm I is described in Section 3.1.2, and the working principle of mismatch rejection algorithm II is summarized here.

After filtering by mismatch rejection algorithm I, translation (d_i^j) is calculated for each of the remaining matches. A 99.7% confidence interval can be estimated for the d_i^j value, assuming that the error in the predicted displacement, $\hat{y}_i^-(1)$, follows a normal distribution with variance of $\hat{G}_i^-(1,1)$.

$$\begin{aligned}\frac{\hat{y}_i^-(1)}{\alpha} - \frac{3\sqrt{\hat{G}_i^-(1,1)}}{\alpha} - 2\varepsilon - d_i^{\mathrm{ROI}} \le d_i^j \\ \le \frac{\hat{y}_i^-(1)}{\alpha} + \frac{3\sqrt{\hat{G}_i^-(1,1)}}{\alpha} + 2\varepsilon - d_i^{\mathrm{ROI}}\end{aligned} \tag{29}$$

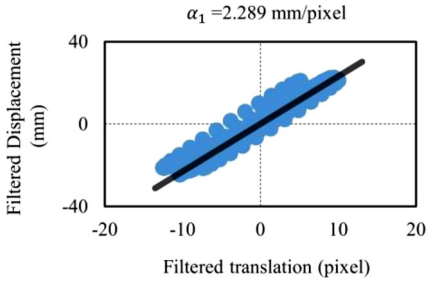

FIGURE 13 Scale factor for shaking table test estimated by the proposed technique

where $\hat{G}_i^-(1,1)$ denotes the first entity of $\hat{G}_i^-$. Note that the pixel discretization error (ε) is considered in the estimation of this confidence interval. Any match that does not fall within this confidence interval is classified as a mismatch and rejected. The final translation is estimated as the summation of the averaged translations of all remaining matches and ROI movement, and it is converted to the displacement (u_i) using the estimated scale factor (α):

$$u_i = \alpha\left\{\frac{1}{N_g}\left(\sum_{j=1}^{N_g} d_i^j\right) + d_i^{ROI}\right\} \tag{30}$$

where N_g is the number of good matches. Because this study focuses on one-directional displacement estimation, the mismatch rejection criterion (Equation 29) is given only for one direction (y direction). The same mismatch rejection criterion can be applied to the x direction for two-directional displacement estimation. In such a case, a two-axial accelerometer should be used, and then displacements can be estimated in both x and y directions.

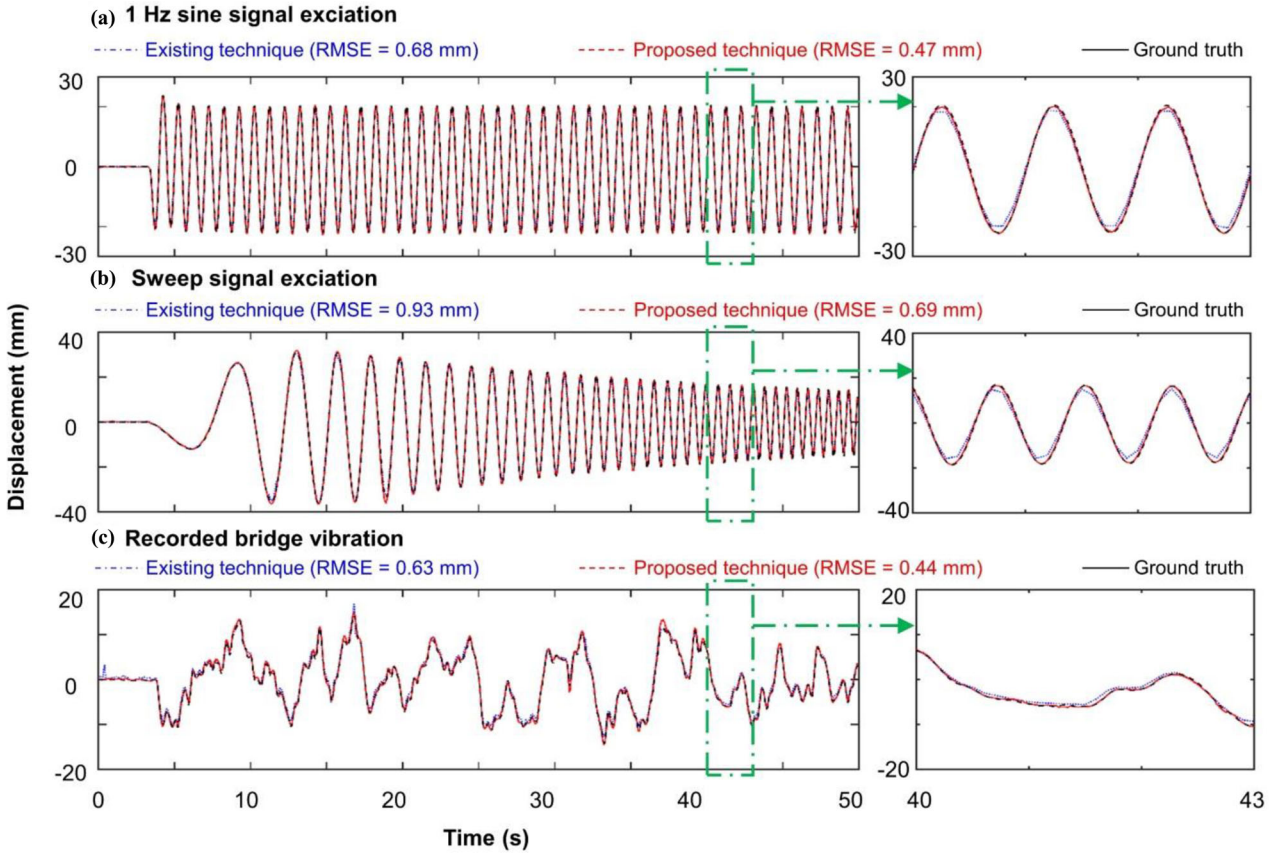

FIGURE 14 Displacements for shaking table test estimated by the proposed and existing techniques (Yu & Zhang, 2019) under: (a) 1 Hz sine, (b) sweep signal, and (c) recorded bridge vibration signal inputs

4 | EXPERIMENTAL VALIDATION

4.1 | Shaking table test

Figure 12a demonstrates the experimental setup of the shaking table test. A uniaxial force balance accelerometer (EpiSensor ES-U2) and a vision camera (Insta360 Pro2) were placed on a steel plate, and the plate was rigidly connected to a shaking table, providing vertical movement for the plate. A PSV-400-M4 laser Doppler velocimeter (LDV), with a resolution of 0.5 pm (Polytech GmbH, 2005), was used to measure the ground truth displacement of the plate. The vision camera was sampled at 29.97 Hz, while acceleration and LDV measurements were sampled at 100 Hz. The FOV of the camera is illustrated in Figure 12b. A concrete block located at approximately 3 m distance was selected as the target, and the initial ROI was selected to include this target, as shown in Figure 12b. Three different excitation signals were used as the input signal to the shaking table in this test: (1) a 1 Hz sine signal, (2) a sweep signal with a frequency bandwidth of DC to 3 Hz, and (3) a recorded bridge vibration signal.

The scale factor was estimated by the proposed technique using vision and acceleration measurements corresponding to a 1 Hz sine input (shown in Figure 14a). As explained in Section 3.1.1, the proposed technique includes a band-pass filter. The lower cutoff frequency of the filter was set to 0.3 Hz after a series of laboratory tests, while the upper cutoff frequency was set to one-tenth of the vision camera's sampling rate. The scaling factor estimated by the proposed technique ($\alpha_1 = 2.289$ mm/pixel) (as shown in Figure 13) was closed to that ($\alpha_2 = 2.146$ mm/pixel) estimated by the existing technique (Equation 3) using the height of the concrete block (as shown in Figure 12b).

Displacements were estimated using the proposed and existing techniques (Yu & Zhang, 2019). The existing technique computes the scale factor based on Equation (3) and combines a prepurification strategy, KnnMatch, and RANSAC to reject mismatches. The thresholds for KnnMatch and RANSAC were set to 0.65 and 5 pixels, respectively. The prepurification strategy requires a prior estimation of the maximum displacement, and this value was set to 45 mm in this test. The vision measurements were downsampled to 10 (≈29.97/3) Hz to reduce the computational time, thus enabling real-time displacement estimation. Note that the downsampled vision measurements were used in both the proposed and existing techniques. Figure 14 compares the displacements estimated using the

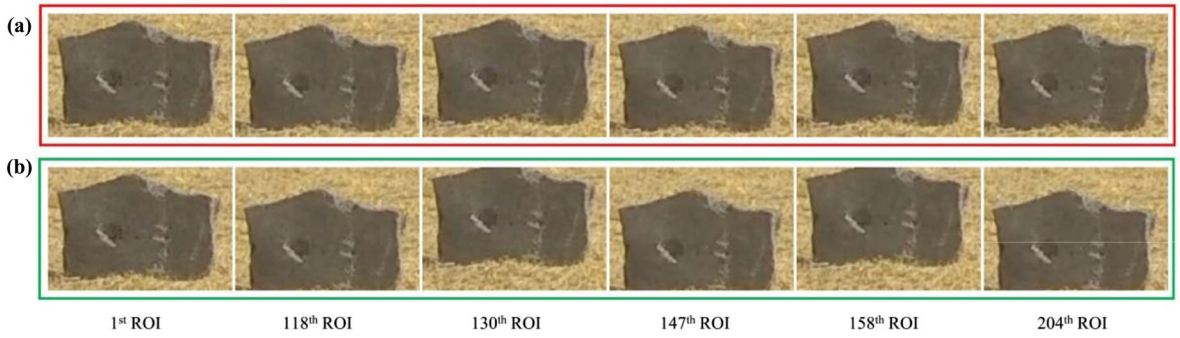

FIGURE 15 Comparison of ROIs at different time steps of the sweep case (a) with and (b) without the proposed ROI updating algorithm

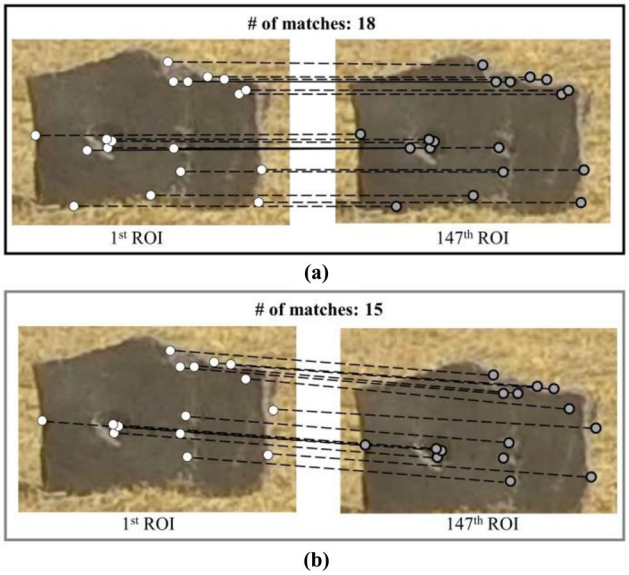

FIGURE 16 Comparison of matching performance between 1st and 147th ROIs (a) with and (b) without the proposed ROI updating algorithm

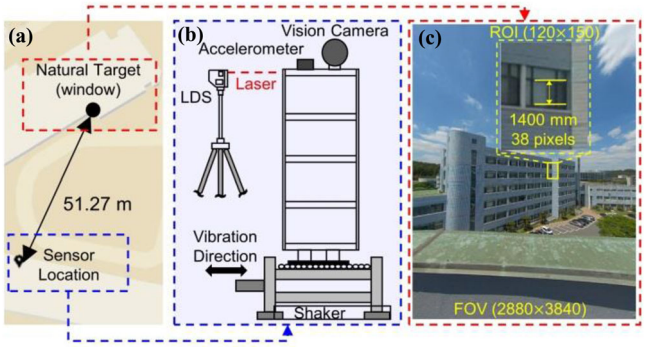

FIGURE 17 Experimental setup of a four-story building model: (a) distance between sensor and natural target, (b) configuration of the four-story building model, and (c) FOV of vision camera and initial ROI for four-story building test

proposed and existing techniques. The root mean square errors (RMSEs) of each technique were calculated, and the RMSE of the proposed technique was less by 30% compared to that of the existing technique.

Figure 15 illustrates that the target (i.e., a concrete block) is always within ROIs using the proposed ROI updating algorithm, while the target is partly out of ROIs at certain time steps without updating the ROI. Note that the target can be fully out of the ROI with a large displacement.

The improvement in the matching performance achieved by the proposed ROI updating algorithm is demonstrated in Figure 16. A total of 18 matches were obtained using the updated ROI algorithm, whereas only 15 matches were obtained using the fixed ROI.

The computation costs of the proposed displacement estimation technique were analyzed using a desktop PC configured with Intel i7-6700 CPU (3.4 GHz) and 8 GB RAM. The acquisition and processing of each vision image took around 0.09 s, allowing real-time vision-based displacement estimation at 10 Hz. Then, by fusing the vision-based displacement at 10 Hz with the acceleration measurement at 100 Hz, the final displacement was estimated at 100 Hz in real time.

4.2 | Four-story building test

The experimental setup of the four-story building model test is demonstrated in Figure 17. The experiment was performed on the roof of a building inside the KAIST campus, and a window of another building, approximately 52 m apart, was used as a natural target. The shaking table moved the building model in a horizontal direction, and a KAIS KL3-W400 laser displacement sensor, which has a resolution of 10 μm, was used to measure the ground truth displacement (KAIS Co., Ltd., 2019). The FOV of the vision camera and the initial ROI are shown in Figure 17c. Five different excitation signals were considered in this test: (1) 0.5 Hz sine signal, (2) 1 Hz sine signal, (3) sweep signal with a frequency bandwidth of DC to 3 Hz, (4) multitone signal with a frequency bandwidth of 0.5–2.5 Hz and a frequency

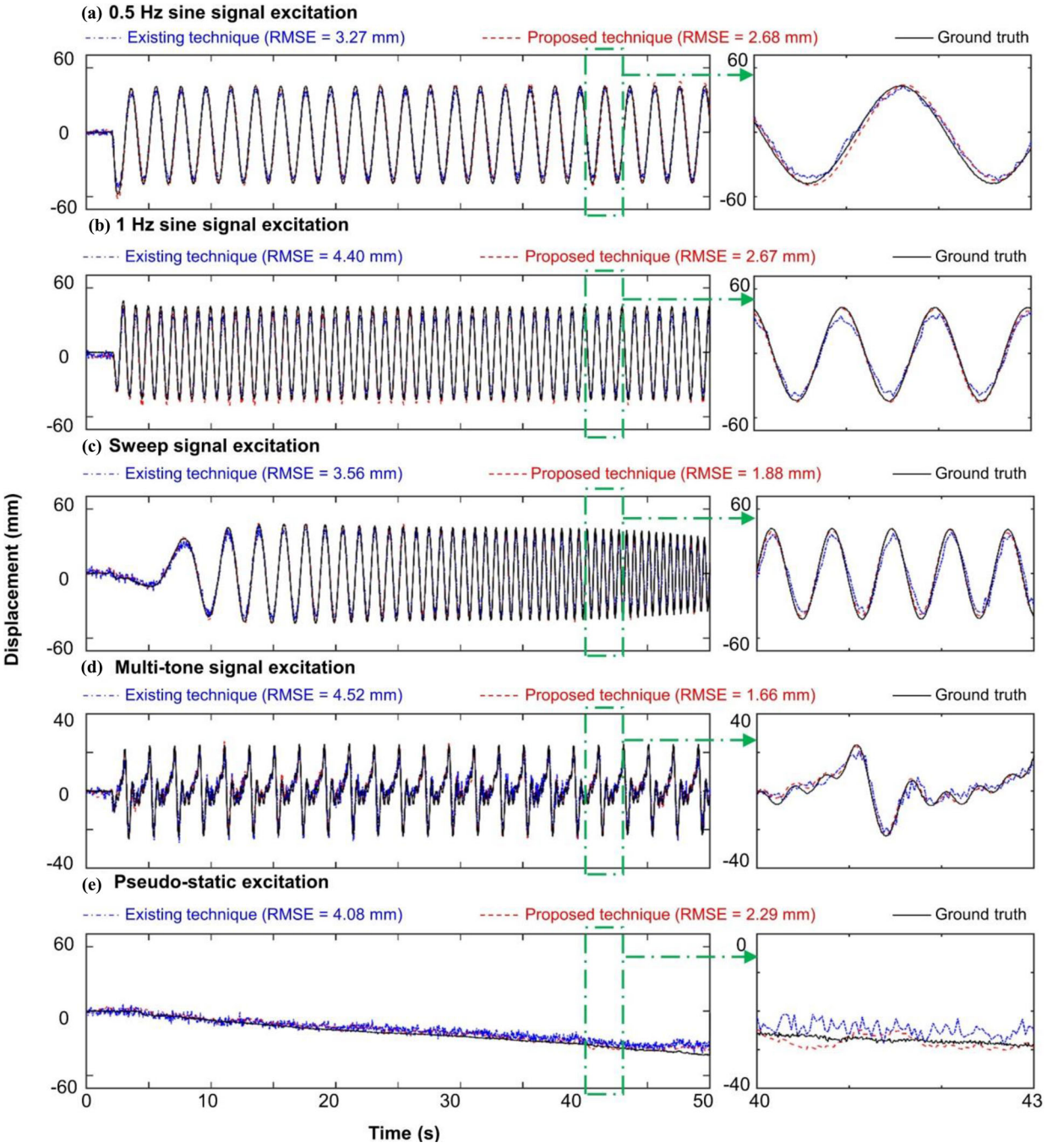

FIGURE 18 Displacements for four-story building test estimated by the proposed and existing techniques under: (a) 0.5 Hz sine signal, (b) 1 Hz sine signal, (c) multitone signal, (d) sweep signal, and (e) pseudo-static excitation

step of 0.5 Hz, and (5) manual pseudo-static excitation. All the other test parameters were identical to those used for the previous shaking table test.

The scale factor estimated by the proposed technique using a 1 Hz sine input was 40.225 mm/pixel (α_1), and the scale factor estimated by Equation (3) and the known window height shown in Figure 17c was 36.842 mm/pixel (α_2). The displacements estimated by the proposed and existing techniques (Yu & Zhang, 2019) are compared in Figure 18, and the estimated maximum displacement, which was required for the pre-purification strategy in the existing technique, was set to 60 mm. Overall, the RMSE of the proposed technique was less by approximately 44% compared to that of the existing technique.

Next, the effectiveness of the proposed scale factor estimation algorithm was examined by comparing the displacement estimated with two different scale factors. Note that only scale factor was changed in this comparison.

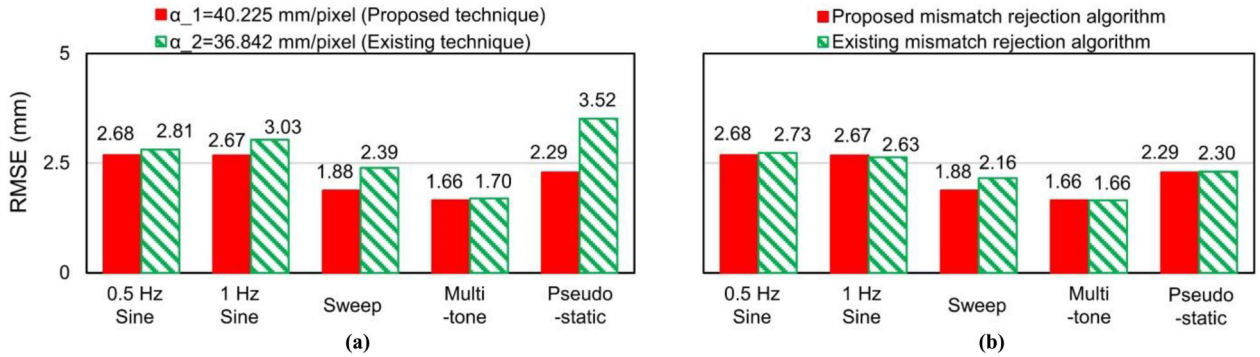

FIGURE 19 Improvement of displacement estimation performance using the proposed (a) scale factor estimation technique and (b) mismatch rejection algorithm for four-story building test

Figure 19a shows that using α_1 produces a smaller RMSE value for the displacement estimation than using α_2, thereby validating the effectiveness of the proposed scale factor estimation algorithm. In addition, the displacements were estimated using α_1 but with two different mismatch rejection algorithms, that is, the proposed and existing mismatch rejection algorithms (Yu & Zhang, 2019). As shown in Figure 19b, the proposed mismatch rejection algorithm achieves a similar matching performance compared to that of the existing mismatch rejection algorithm. Moreover, the proposed algorithm does not require any human interactions for ad-hoc threshold establishment, and its computational time was only approximately 1/20th of that of the existing algorithm.

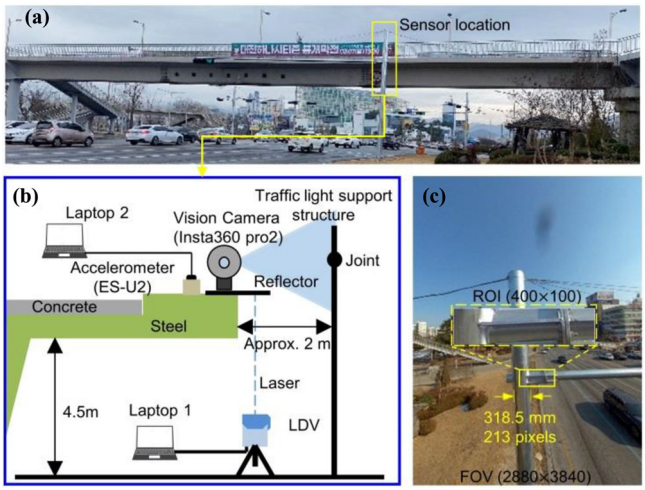

FIGURE 20 Overview of the pedestrian steel box girder bridge test: (a) pedestrian steel box girder bridge, (b) experimental setup, and (c) FOV of vision camera and initial ROI

4.3 | Pedestrian steel box girder bridge test

Figure 20 provides an overview of the pedestrian steel box girder bridge test. The pedestrian bridge, located in Daejeon, Korea, is 45 m long and 8 m wide. As shown in Figure 20b, the vision camera and uniaxial force balance accelerometer identical to the previous tests were installed at the one-fourth span point of the bridge. The ground truth displacement was measured using a Polytech RSV-150 LDV, which has 0.3 nm resolution (Polytech GmbH, 2019). The FOV of the camera is illustrated in Figure 20c. A joint of a traffic light support structure, at approximately 2 m, was selected as a natural target, and an initial ROI was selected to include this target, as shown in Figure 20c. Four different excitations were considered in this test: (1) Case 1: 4 people jumping at one-fourth span point, (2) Case 2: 1 person jumping at one-fourth span point and 16 people slowly crossing the bridge; (3) Case 3: 2 people jumping at one-fourth span point and 16 people slowly crossing the bridge; and (4) Case 4: 16 people slowly crossing the bridge.

The scale factor estimated by the proposed technique using the vision and acceleration measurements of Case 1 was 1.359 mm/pixel (α_1) and that estimated using Equation (3) and the diameter of the support structure shown in Figure 20c was 1.495 mm/pixel (α_2). The displacements estimated by the proposed and existing techniques (Yu & Zhang, 2019) are compared in Figure 21. The maximum displacement, which is required for the prepurification strategy in the existing technique, was set to 4.5 mm in this test. Compared to the existing technique, the proposed technique improved the displacement estimation accuracy by reducing the RMSE by 30%.

The effectiveness of the proposed scale factor estimation and mismatch rejection algorithms is illustrated in Figure 22. Again, α_1 produced RMSEs that were smaller than α_2. In addition, the performance of the proposed mismatch rejection algorithm was comparable to that of the

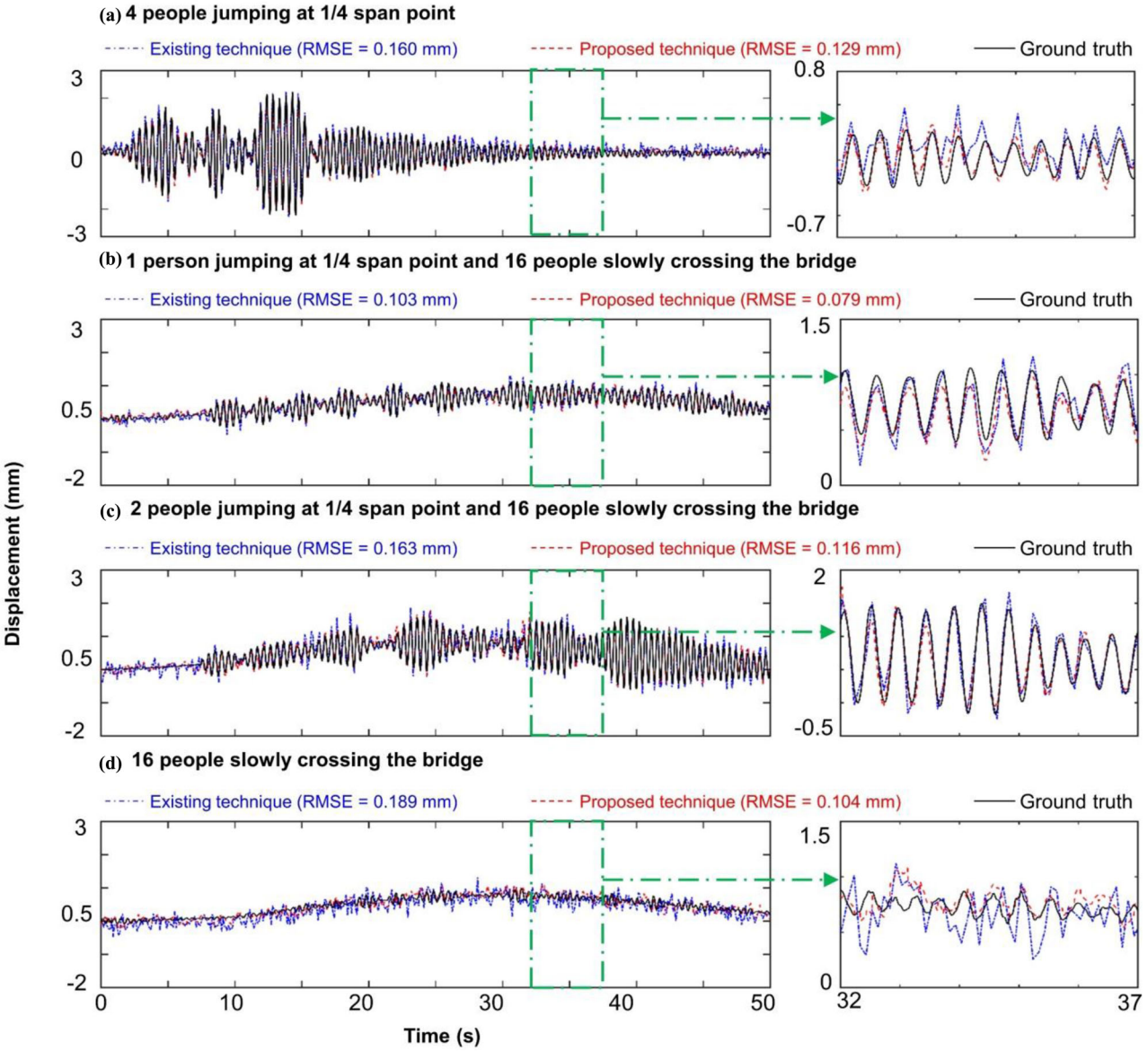

FIGURE 21 Displacements estimated by the proposed and existing techniques: (a) Case 1: 4 people jumping at one-fourth span point; (b) Case 2: 1 person jumping at one-fourth span point and 16 people slowly crossing the bridge; (c) Case 3: 2 people jumping at one-fourth span point and 16 people slowly crossing the bridge; and (d) Case 4: 16 people slowly crossing the bridge

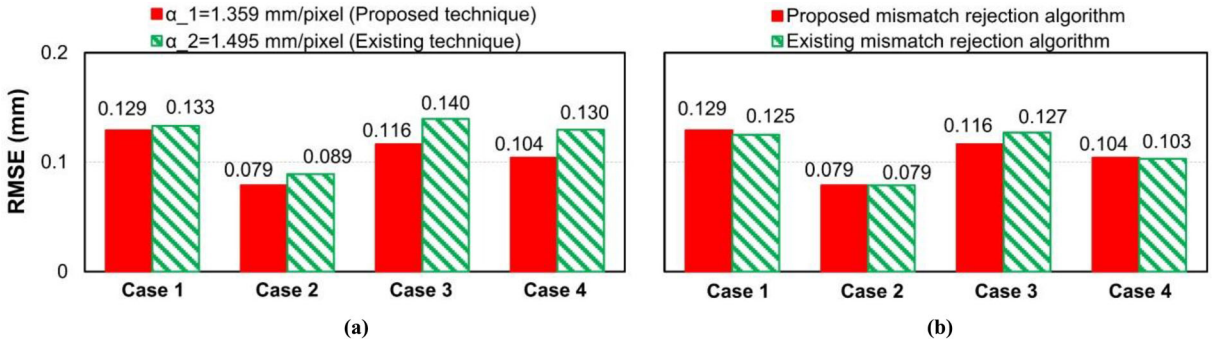

FIGURE 22 Improvement of displacement estimation performance using the proposed (a) scale factor estimation techniques and (b) mismatch rejection algorithm for pedestrian steel box girder bridge test

existing mismatch rejection algorithm, while the computational time was reduced to 5%.

5 | CONCLUSIONS

In this study, a displacement estimation technique was proposed by combining asynchronous acceleration and computer vision measurements. The proposed technique allows real-time structural displacement estimation for long-term continuous monitoring. The primary contribution of this study lies in the automatic scale factor estimation, development of an adaptive multirate Kalman filter to fuse asynchronous acceleration and vision measurements, and improvement of feature matching by proposing an ROI updating algorithm and two automatic mismatch rejection algorithms. The feasibility of the proposed displacement estimation technique for structural displacement measurement was experimentally validated by conducting shaking table, four-story building model, and pedestrian steel box girder bridge tests. The overall RMSEs for these three tests were 0.53, 2.24, and 0.11 mm, respectively. Currently, the proposed technique requires manual selection of an initial ROI from the first frame and estimates only one-directional displacement. A follow-up study is required for the automatic selection of the initial ROI and displacement estimation in all six degrees of freedom.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (no. 2017R1A5A1014883). We also thank the students from Smart Structures and Systems Lab at KAIST for their assistance during the field test.

REFERENCES

Akhlaghi, S., Zhou, N., & Huang, Z. (2017). Adaptive adjustment of noise covariance in Kalman filter for dynamic state estimation. *2017 IEEE power & energy society general meeting* (pp. 1–5). IEEE.

Arabi, S., Haghighat, A., & Sharma, A. (2020). A deep-learning-based computer vision solution for construction vehicle detection. *Computer-Aided Civil and Infrastructure Engineering*, *35*(7), 753–767.

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded up robust features. *European Conference on Computer Vision* (pp. 404–417). Springer.

Brandt, A. (2011). *Noise and vibration analysis: Signal analysis and experimental procedures*. John Wiley & Sons.

Chang, C. C., & Xiao, X. H. (2010). An integrated visual-inertial technique for structural displacement and velocity measurement. *Smart Structures and Systems*, *6*(9), 1025–1039.

Dong, C.-Z., Celik, O., & Catbas, F. N. (2018). Marker-free monitoring of the grandstand structures and modal identification using computer vision methods. *Structural Health Monitoring*, *18*(5–6), 1491–1509.

Feng, D., & Feng, M. Q. (2015). Model updating of railway bridge using in situ dynamic displacement measurement under trainloads. *Journal of Bridge Engineering*, *20*(12). https://doi.org/10.1061/(ASCE)BE.1943-5592.0000765

Feng, D., & Feng, M. Q. (2016). Output-only damage detection using vehicle-induced displacement response and mode shape curvature index. *Structural Control and Health Monitoring*, *23*(8), 1088–1107.

Feng, M. Q., Fukuda, Y., Feng, D., & Mizuta, M. (2015). Nontarget vision sensor for remote measurement of bridge dynamic response. *Journal of Bridge Engineering*, *20*(12). https://doi.org/10.1061/(ASCE)BE.1943-5592.0000747

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, *24*(6), 381–395.

Gindy, M., Vaccaro, R., Nassif, H., & Velde, J. (2008). A state-space approach for deriving bridge displacement from acceleration. *Computer-Aided Civil and Infrastructure Engineering*, *23*(4), 281–290

Gomez, F., Park, J. W., & Spencer, B. F., Jr. (2018). Reference-free structural dynamic displacement estimation method. *Structural Control and Health Monitoring*, *25*(8), e2209.

KAIS Co., Ltd., (2019). *KL3 series laser displacement sensor datasheet*. https://ekais.kr/ekaisbiz/wpcontent/uploads/2019/05/KL3-Series.pdf

Khuc, T., & Catbas, F. N. (2016). Computer vision-based displacement and vibration monitoring without using physical target on structures. *Structure and Infrastructure Engineering*, *13*(4), 505–516.

Kim, J., Kim, K., & Sohn, H. (2014). Autonomous dynamic displacement estimation from data fusion of acceleration and intermittent displacement measurements. *Mechanical Systems and Signal Processing*, *42*(1–2), 194–205.

Kim, K., Choi, J., Koo, G., & Sohn, H. (2016). Dynamic displacement estimation by fusing biased high-sampling rate acceleration and low-sampling rate displacement measurements using two-stage Kalman estimator. *Smart Structures and Systems*, *17*(4), 647–667.

Lee, H. S., Hong, Y. H., & Park, H. W. (2010). Design of an FIR filter for the displacement reconstruction using measured acceleration in low-frequency dominant structures. *International Journal for Numerical Methods in Engineering*, *82*(4), 403–434.

Lee, J. J., & Shinozuka, M. (2006). A vision-based system for remote sensing of bridge displacement. *NDT & E International*, *39*(5), 425–431.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110.

Luo, L., & Feng, M. Q. (2018). Edge-enhanced matching for gradient-based computer vision displacement measurement. *Computer-Aided Civil and Infrastructure Engineering*, *33*(12), 1019–1040.

Ma, Z., Chung, J., Liu, P., & Sohn, H. (2021). Bridge displacement estimation by fusing accelerometer and strain gauge measurements. *Structural Control and Health Monitoring*, *28*(6), e2733.

Mohamed, A., & Schwarz, K. (1999). Adaptive Kalman filtering for INS/GPS. *Journal of Geodesy*, *73*(4), 193–203.

Moreu, F., Jo, H. J. R. S. A. S. H. B., Li, J., Kim, R. E., Cho, S., Kimmle, A., & LaFave, J. M. (2015). Dynamic assessment of timber railroad bridges using displacements. *Journal of Bridge Engineering*, *20*(10). https://doi.org/10.1177/1045389X17721375

Nakamura, S. (2000). GPS measurement of wind-induced suspension bridge girder displacements. *Journal of Structural Engineering*, *126*(12), 1413–1419.

Narazaki, Y., Hoskere, V., Hoang, T. A., Fujino, F., Sakurai, A., & Spencer, B. F. (2020). Vision-based automated bridge component recognition with high-level scene consistency. *Computer-Aided Civil and Infrastructure Engineering*, *35*(5), 465–482.

Nassif, H. H., Gindy, M., & Davis, J. (2005). Comparison of laser Doppler vibrometer with contact sensors for monitoring bridge deflection and vibration. *Ndt and E International*, *38*(3), 213–218.

Pan, X., & Yang, T. Y. (2020). Post-disaster imaged-based damage detection and repair cost estimation of reinforced concrete buildings using dual convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, *35*(5), 495–510.

Park, H. S. Lee, H. M., Adeli, H., & Lee, I. (2007). A new approach for health monitoring of structures: Terrestrial laser scanning. *Computer-Aided Civil and Infrastructure Engineering*, *22*(1), 19–30.

Park, J. W., Lee, K. C., Sim, S. H., Jung, H. J., & Spencer, B. F., Jr. (2016). Traffic safety evaluation for railway bridges using expanded multisensor data fusion. *Computer-Aided Civil and Infrastructure Engineering*, *31*(10), 749–760.

Park, J. W., Sim, S. H., & Jung, H. J. (2013). Displacement estimation using multimetric data fusion. *IEEE/ASME Transactions on Mechatronics*, *18*(6), 1675–1682.

Park, K. T., Kim, S. H., Park, H. S., & Lee, K. W. (2005). The determination of bridge displacement using measured acceleration. *Engineering Structures*, *27*(3), 371–378.

Park, S. W., Park, H. S., Kim, J. H., & Adeli, H. (2015). 3D displacement measurement model for health monitoring of structures using a motion capture system. *Measurement*, *59*, 352–362.

Polytech GmbH. (2005). *RSV-400 scanning vibrometer datasheet*. http://www.vibrometry.co.kr/PSV-400_2005.pdf

Polytech GmbH. (2019). *RSV-150 scanning vibrometer datasheet*. http://hysen.cafe24.com/wpcontent/uploads/2019/10/OM_DS_RSV-150_E_42319.pdf

Smyth, A., & Wu, M. (2007). Multi-rate Kalman filtering for the data fusion of displacement and acceleration response measurements in dynamic system monitoring. *Mechanical Systems and Signal Processing*, *21*(2), 706–723.

Wang, H., Deng, Z., Feng, B., Ma, H., & Xia, Y. (2017). An adaptive Kalman filter estimating process noise covariance. *Neurocomputing*, *223*, 12–17.

Wang, M., & Cheng, J. C. P. (2020). A unified convolutional neural network integrated with conditional random field for pipe defect segmentation. *Computer-Aided Civil and Infrastructure Engineering*, *35*(2), 162–177.

Xie, Y., Zhang, Y., & Ye, Z. (2007). Short-term traffic volume forecasting using Kalman filter with discrete wavelet decomposition. *Computer-Aided Civil and Infrastructure Engineering*, *22*(5), 326–334.

Xu, Y., Brownjohn, J. M. W., & Huseynov, F. (2019). Accurate deformation monitoring on bridge structures using a cost-effective sensing system combined with a camera and accelerometers: Case study. *Journal of Bridge Engineering*, *24*(1), 1–14.

Yoon, H., Elanwar, H., Choi, H., Golparvar-Fard, M., & Spencer, B. F., Jr. (2016). Target-free approach for vision-based structural system identification using consumer-grade cameras. *Structural Control Health Monitoring*, *23*(12), 1405–1416.

Yoon, H., Shin, J., & Spencer, B. F., Jr. (2018). Structural displacement measurement using an unmanned aerial system. *Computer-Aided Civil and Infrastructure Engineering*, *33*(3), 183–192.

Yu, S., & Zhang, J. (2019). Fast bridge deflection monitoring through an improved feature tracing algorithm. *Computer-Aided Civil and Infrastructure Engineering*, *35*(3), 292–302.

Zhang, C., Chang, C. C., & Jamshidi, M. (2020). Concrete bridge surface damage detection using a single-stage detector. *Computer-Aided Civil and Infrastructure Engineering*, *35*(4), 389–340.

Zhi, L., Li, Q. S., & Fang, M. (2016). Identification of wind loads and estimation of structural responses of super-tall buildings by an inverse method. *Computer-Aided Civil and Infrastructure Engineering*, *31*(12), 966–982.

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