

Article

Vision and Vibration Data Fusion-Based Structural Dynamic Displacement Measurement with Test Validation

Cheng Xiu , Yufeng Weng and Weixing Shi *

Department of Disaster Mitigation for Structures, College of Civil Engineering, Tongji University, Shanghai 200092, China; xiucheng@tongji.edu.cn (C.X.)

* Correspondence: swxtgk@tongji.edu.cn

Abstract: The dynamic measurement and identification of structural deformation are essential for structural health monitoring. Traditional contact-type displacement monitoring inevitably requires the arrangement of measurement points on physical structures and the setting of stable reference systems, which limits the application of dynamic displacement measurement of structures in practice. Computer vision-based structural displacement monitoring has the characteristics of non-contact measurement, simple installation, and relatively low cost. However, the existing displacement identification methods are still influenced by lighting conditions, image resolution, and shooting rate, which limits engineering applications. This paper presents a data fusion method for contact acceleration monitoring and non-contact displacement recognition, utilizing the high dynamic sampling rate of traditional contact acceleration sensors. It establishes and validates an accurate estimation method for dynamic deformation states. The structural displacement is obtained by combining an improved KLT algorithm and asynchronous multi-rate Kalman filtering. The results show that the presented method can help improve the displacement sampling rate and collect high-frequency vibration information compared with only the vision measurement technique. The normalized root mean square error is less than 2% for the proposed method.

Keywords: Kanade–Lucas–Tomasi optical-flow method; data fusion; computer vision; Kalman filter



Citation: Xiu, C.; Weng, Y.; Shi, W. Vision and Vibration Data Fusion-Based Structural Dynamic Displacement Measurement with Test Validation. *Sensors* **2023**, *23*, 4547. <https://doi.org/10.3390/s23094547>

Academic Editor: Mohammad Noori

Received: 30 March 2023

Revised: 29 April 2023

Accepted: 5 May 2023

Published: 7 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In structural health monitoring, it is necessary to deploy sensors to monitor the structure's response. These sensors collect important data on various aspects of the structure, such as vibrations and displacements [1,2]. These measurements can provide valuable insights into the structure's integrity and indicate any load anomalies or structural defects. Moreover, displacement monitoring can also be used to update the finite element model of the structure, which is essential for accurately assessing, monitoring, and controlling civil infrastructure [3–7]. For example, peak deformation demands, including peak inter-story drift ratio and peak roof displacement, are essential indicators in earthquake engineering for evaluating structural seismic performance [8–11]. Vehicle-induced displacement is also utilized to detect bridge damage and assess bridge conditions [12]. Additionally, the displacement of a high-rise building is an important indicator of safety [13,14]. Therefore, displacement is critical in ensuring civil infrastructure's health and integrity.

There are many means of directly measuring the displacement response of a structure in the field of structural engineering, which include pull-wire displacement gauges, linear variable differential transformers (LVDT) [15], laser Doppler vibrometers (LDV) [16], Real-Time Kinematic global satellite navigation systems (RTK-GNSS) [17], etc. LVDT usually need to be installed between the target point and a fixed reference point; hence, despite the high accuracy of LVDT measurements, they are not easy to be installed in practical engineering [18,19]. As LVDT is a contact measurement method, any severe structural deformation or breakage during a shaking table test can potentially damage the LVDT. On

the other hand, LDV can remotely measure displacement with high resolution and accuracy; however, it can be expensive and limited to a few measurement points [16]. RTK-GNSS is more accurate than normal GNSS and can provide centimeter-level accuracy; however, it has a lower sampling frequency [20]. Moreover, the GPS method is infeasible for indoor measurement due to the requirement of signal reception [21,22]. Accelerometric integration is also used to measure displacement; however, this method suffers from low-frequency drift and cannot measure residual deformation [23]. Several methods [24–28] have been used to solve the drift problem; however, these methods will remove information about the structure's response.

With the development of high-quality, low-cost optical cameras and lenses in recent years, structural monitoring and inspection-based computer vision have gradually become a hot topic. Numerous displacement estimation methods, such as template matching [29–31], feature matching [18,32–34], digital image correlation (DIC) [35–37], and optical flow methods [38–41], have been proposed. Optical flow methods are widely used among these techniques due to their high accuracy and computational efficiency. Many researchers have utilized the Kanade–Lucas–Tomasi (KLT) tracker, an intensity-based optical flow estimation algorithm, for target-based or target-free structural displacement measurement [42–44]. The concept of optical flow was initially introduced by Gibson [45] and referred to the velocity of a moving object in a time-varying image. Based on this idea, the KLT optical flow method matches and tracks feature points in two adjacent frames to obtain motion information for those points. However, despite its advantages, the KLT method has two primary limitations, loss of feature points during tracking [41,46] and drift-type errors [47,48], respectively. Regarding the first limitation, since the Taylor expansion is used in the derivation of KLT, it is necessary to satisfy the assumption of small deformations, which is described in detail in Section 2. As for the second limitation, the KLT tracker estimates the feature locations by using an image gradient, but errors induced by integration drift can cause inaccuracies in the measurement of residual displacement, leading to deviations from the correct tracking over time. This problem is particularly challenging when tracking long sequences, even though it may not be noticeable in individual image pairs.

To estimate displacement in a high sampling rate, vision-based measurements at a low sampling rate and acceleration measurements at a high sampling rate can be combined as well. The study by Roberts et al. [49] highlighted the importance of displacement fusion in extending the available frequency band, particularly in detecting vibrations of bridges reliably. They found that a minimum sampling rate of 100 Hz is required for bridges. To achieve the sampling rate, several researchers have proposed to fuse low-sampled measurements, such as GPS and strain sensors, with high-sampled measurements (such as acceleration) [50–53]. Efforts were also taken to fuse vision cameras and accelerometers. Park et al. [54] utilized a complementary filter to fuse acceleration and displacement, while Ma et al. [55] employed an adaptive Kalman filter to estimate displacement. These methods mainly used the feature-matching based method to estimate displacement, which takes more time compared with the KLT method.

Utilizing the high dynamic sampling rate of traditional contact acceleration sensors, this paper introduces a data fusion approach for contact acceleration monitoring and non-contact displacement recognition, constructing and validating an accurate estimation method for critical dynamic deformation states in structures. This paper is structured as follows: Section 2 provides a brief overview of the KLT algorithm. Section 3 introduces the algorithm employed in this study. Section 4 presents the study's results, demonstrating the proposed method's high efficiency, accuracy, and robustness in achieving drift-free large structural displacements. The primary limitations of the KLT method were addressed by fusing accelerometer data, which improved the accuracy of feature tracking and reduced errors caused by integration drift. Finally, the concluding remarks are presented in Section 5.

2. A Brief Review of the Kanade–Lucas–Tomasi (KLT) Method

Optical flow refers to the pattern of apparent motion of objects in an image between two frames due to either the motion of the object or the camera. For instance, in Figure 1, three target points in two adjacent images can have their positions in the second image identified by detecting the pixels with consistent intensity values with the corresponding pixels in the first image. It represents the displacement of a 2D vector field (d_x, d_y) when a feature point moves from the first frame $I(x, y, t)$ to the second frame after a time interval of d_t . The optical flow equation assumes that the object's brightness does not change.

$$I_1(x, y, t) = I_2(x + d_x, y + d_y, t + d_t) \quad (1)$$

$I_1(x, y, t)$ represents image pixels from the reference image, and $I_2(x + d_x, y + d_y, t + d_t)$ is the image pixels of the following image. For simplicity, let $\mathbf{d} = [d_x, d_y]^T$, $\mathbf{X} = [x, y]^T$.

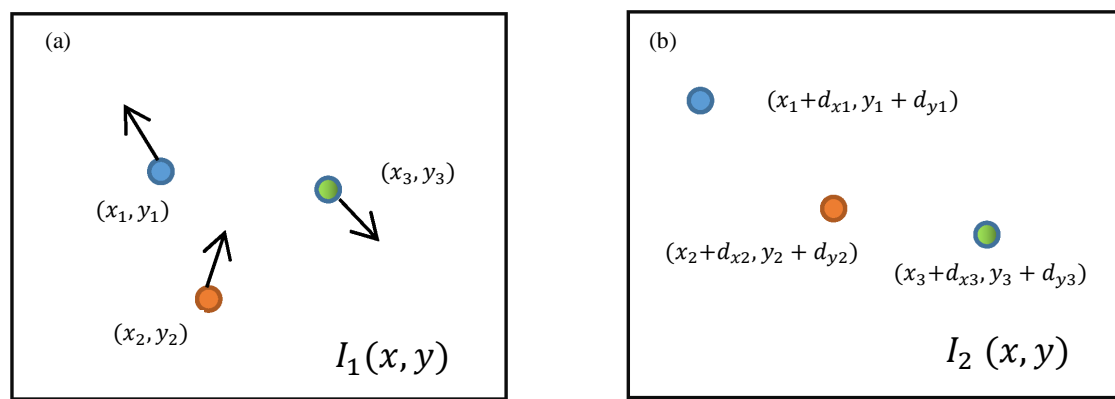


Figure 1. Demonstration of motion of three points in two frames: (a) first frame and (b) second frame.

Under the pixel window, the error function is constructed:

$$\varepsilon = \iint_W [I_2(\mathbf{X} + \mathbf{d}) - I_1(\mathbf{X})]^2 \omega(\mathbf{X}) d\mathbf{X} \quad (2)$$

where a window W , centered on the position of a target point, is established in the first image. $w(\mathbf{X})$ is a weighting function that assigns weight to the surrounding pixel. In the simplest scenario, $w(\mathbf{X}) = 1$. Another commonly used function is the Gaussian function, which addresses the center of the window.

Set the partial derivative of ε with respect to \mathbf{d} as:

$$\iint_W [I_2(\mathbf{X} + \mathbf{d}) - I_1(\mathbf{X})] \left[\frac{\partial I_2(\mathbf{X} + \mathbf{d})}{\partial \mathbf{d}} - \frac{\partial I_1(\mathbf{X})}{\partial \mathbf{d}} \right] \omega(\mathbf{X}) d\mathbf{X} = 0 \quad (3)$$

The following formula can be obtained from Taylor's expansion:

$$I_2(\mathbf{X} + \mathbf{d}) \approx I_2(\mathbf{X}) + d_x \frac{\partial I_2}{\partial x}(\mathbf{X}) + d_y \frac{\partial I_2}{\partial y}(\mathbf{X}) \quad (4)$$

The substitution of Equation (4) into Equation (3) leads to

$$\iint_W [I_2(\mathbf{X}) - I_1(\mathbf{X}) + \mathbf{p}^T \mathbf{d}] \mathbf{p}(\mathbf{X}) \omega(\mathbf{X}) d\mathbf{X} = 0 \quad (5)$$

where:

$$\mathbf{p} = \left[\frac{\partial I_2}{\partial x}, \frac{\partial I_2}{\partial y} \right]^T \quad (6)$$

The following equation can be obtained from Equation (5):

$$Zd = e \quad (7)$$

where $Z = \iint_W \mathbf{p}(\mathbf{X})\mathbf{p}^T(\mathbf{X})\omega(\mathbf{X})d\mathbf{X}$, and $e = \iint_W [I_1(\mathbf{X}) - I_2(\mathbf{X})]\mathbf{p}(\mathbf{X})\omega(\mathbf{X})d\mathbf{X}$.

Equation (7) is solved by an iterative method to obtain the value of d . When the value of e is less than the set threshold, the approximate solution of d can be obtained. In summary, the KLT tracker uses points from the previous and current frames to create motion vectors. Selecting these feature points is an essential part of the KLT method. Normally, a region-of-interest (ROI) is used to focus on a specific part of an image to extract relevant information. Common feature detectors include scale-invariant feature transform (SIFT), speeded-up robust features (SURF), and oriented FAST and rotated BRIEF (ORB) [56]. The Harris point suggested in [44] is an efficient detector in real-time for calculating the optical flow because Harris points are simple, reliable, and efficient corner detection. Traditionally, the KLT algorithm calculates velocity by computing the optical flow between consecutive frames. If the small motion assumption is not satisfied, the traditional way is using an image pyramid, as shown in Figure 2, which is briefly described below.

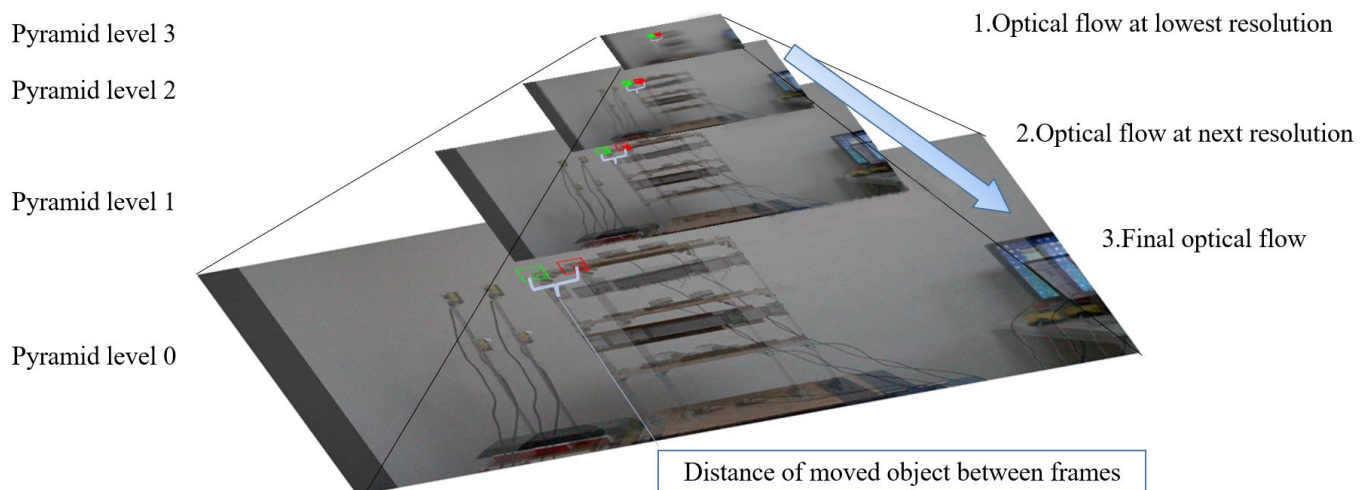


Figure 2. Image pyramid of the KLT method.

The overall pyramidal tracking algorithm proceeds as follows: as an initial layer 0, the original image is used, and the image is reduced by 2^L times in length and width to serve as a layer L . The Gaussian pyramid is generated using the obtained images by superimposing them from bottom to top. The corresponding points are also reduced by 2^L times. The displacement value of the target point on the highest layer is calculated using the method described in the previous section. This value is used in the optical-flow calculation of the next layer as an initial guess to determine the accurate displacement value. Once the displacement value is calculated, it is passed to the following layer as an initial guess and then to the lowest layer (level 0) to obtain the actual displacement value. The work of Kim et al. [57] provides a detailed description of the propagation process. The limitations of the KLT method are discussed and demonstrated by Won et al. [41] The paper demonstrates feature loss and drift occurrence in the KLT method.

3. Methodology

Figure 3 shows an overview of the proposed method. As presented in Figure 3a, one camera is fixed on the ground to trace natural targets on the structure, and an accelerometer is placed on the same floor as the natural targets. Figure 3b illustrates the two stages of the proposed technique for displacement estimation. In the first stage, referred to as the calibration stage, shown in Figure 3, several tasks are accomplished, including the

correction of lens parameters, time synchronization, and scale factor calculation. Following this, the second stage, which is called the displacement estimation stage, is initiated.

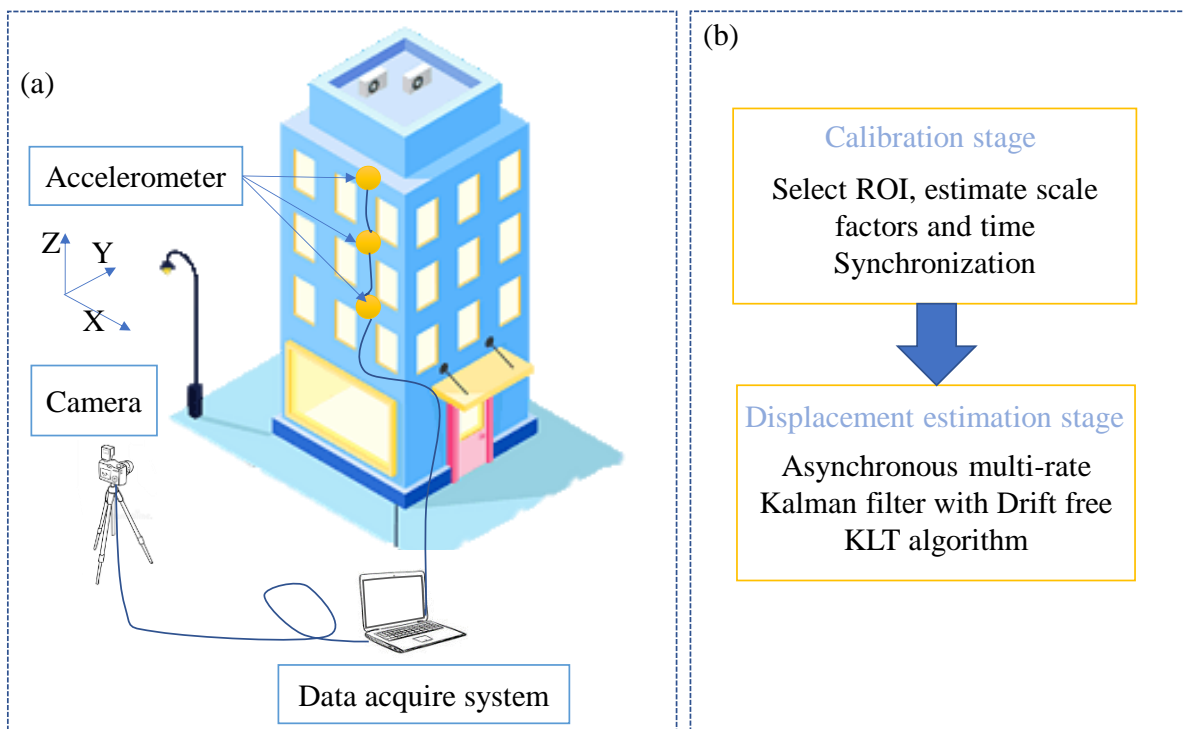


Figure 3. Overall of the proposed displacement estimation technique (a) sensors and camera layout (b) main stages of the proposed technique.

3.1. Calibration Stage

3.1.1. Video Preprocessing and Measurement Conversion

This section uses video preprocessing to correct the distortion caused by the wide-angle lens typically used in consumer-grade cameras. A chessboard pattern is used to calibrate the camera to correct lens distortion [58]. The calibration process involves capturing multiple chessboard images from different angles and orientations, enabling the estimation of the parameters for the lens distortion model. Once the distortion parameters are determined, the images are rectified to remove the distortion and create a rectified image.

3.1.2. Time Synchronization between Vision and Acceleration

This study used two separate acquisition systems to collect data from the camera and the accelerometer. Due to varying sampling rates and data sources, time synchronization is critical before fusing them. As a result, it was necessary to synchronize the data in time. As shown in Figure 4, to avoid the low-frequency drift phenomenon commonly observed in acceleration sensors, the integration results were filtered using a bandpass filter. The lower limit of the passband in bandpass filtering should be sufficiently large to avoid drift, and the upper limit should be at 1/10 of the camera sampling frequency [59]. Additionally, the results of computer vision measurements were resampled to match the sampling frequency of the acceleration measurements. The computer vision measurement results were also filtered using a bandpass filter with the same range as the integration results. This step reduces the impact of frequencies outside the filter range. The cross-correlation analysis was used to finely align the data from the camera and the accelerometer [54]. Here, the time lag is determined at the point where the maximum value of the cross-correlation occurs. This process enabled accurate data matching from both systems and properly synchronized the recorded data.

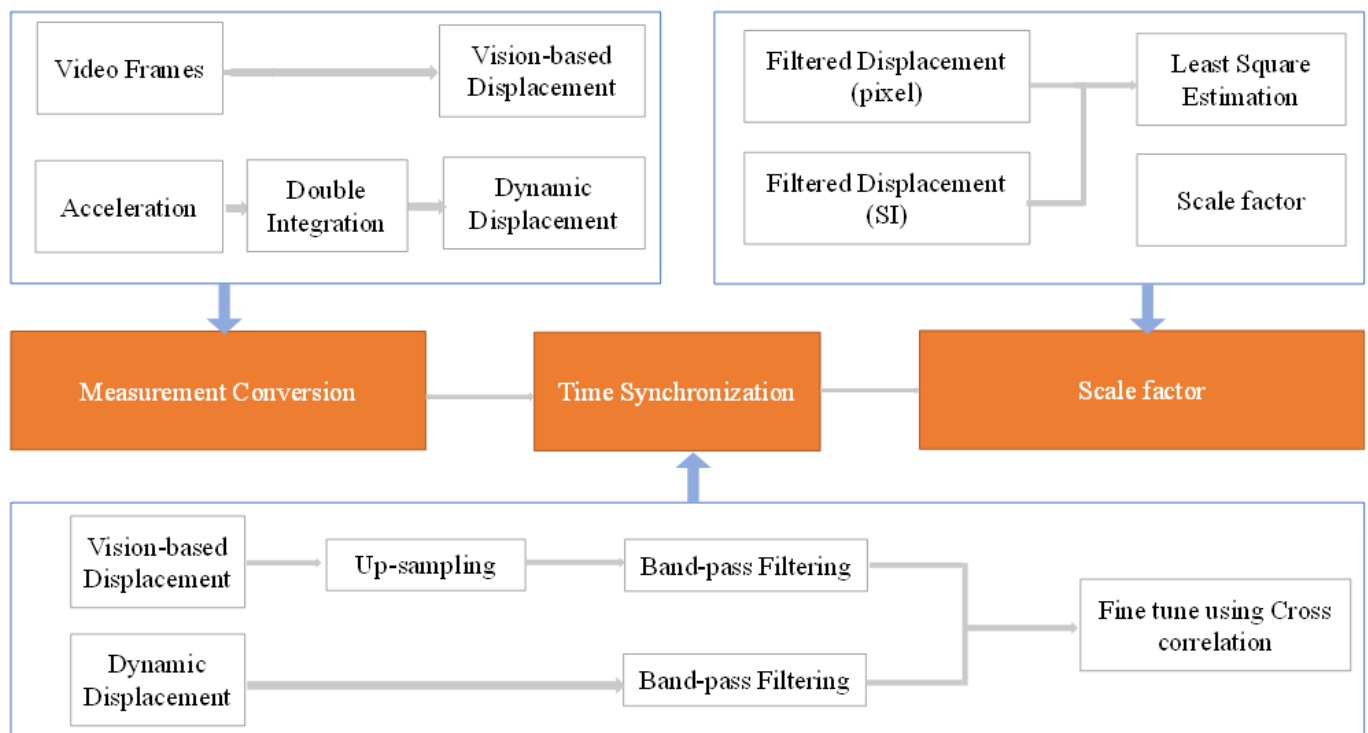


Figure 4. The overall process of time synchronization and calculating the scale factor.

3.1.3. Calculating the Scale Factor

The scale factor λ , determined by the distance between the camera and the target object, translates the image pixel values into real-world metric values, as shown below.

$$\lambda = \frac{D}{d} \left(\text{unit: } \frac{\text{mm}}{\text{pixel}} \right) \quad (8)$$

where D is the actual dimension of the known object, and d is the number of pixels in the image that covers the object.

After time synchronization, the displacements obtained from both methods are truncated to the same length. The scale factor is then estimated using the least squares method. By implementing these steps, potential discrepancies in the displacements can be minimized, and the study results can be reliable.

3.2. Displacement Estimation Stage

3.2.1. Drift-Free KLT Method

Figure 5 describes the detailed procedure for estimating target displacement in the i -th frame. It is important to note that the proposed technique only applies to in-plane motion estimation, and only one direction is considered, though it can be extended in two directions. The method includes the following steps: first, feature points, such as Harris corner points, are selected in the reference frame. Using a priori estimate y , the current frame image is translated. Image translation allows for the adjustment of the images in a way that the displacements fall within the range of small motion, enabling the application of the Taylor expansion of Equation (4).

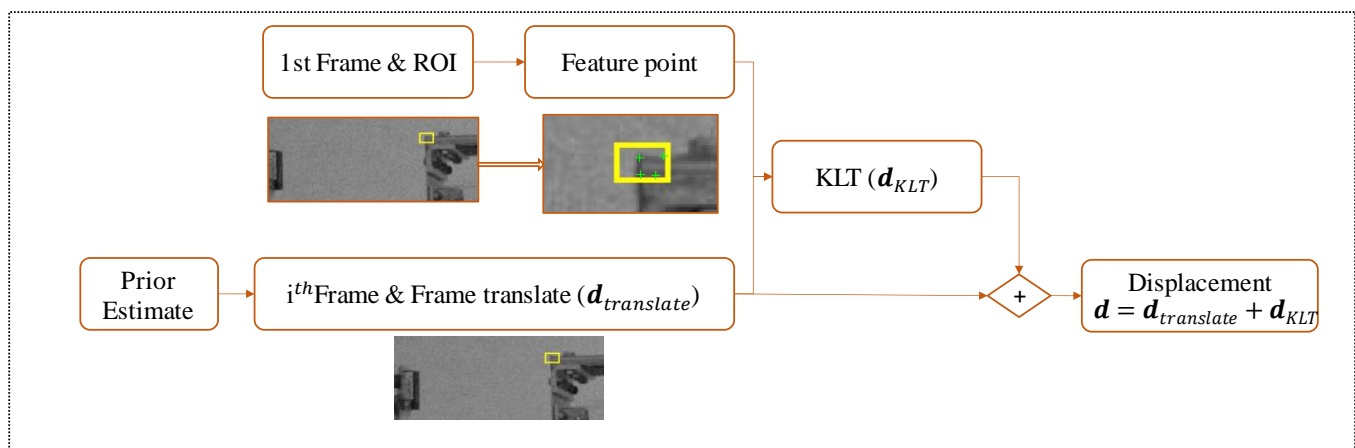


Figure 5. Proposed drift-free KLT method.

Consequently, this approach improves the accuracy and reliability of the displacement estimation, particularly in cases where the initial displacements may not meet the small motion assumption. Furthermore, by incorporating image translation, the proposed method demonstrates its adaptability to various scenarios, enhancing its practical applicability and performance. After translating the image, the KLT algorithm calculates the optical flow between the reference frame and the current frame to obtain the average velocity of the selected feature points, which is used to determine their average displacement.

$$\mathbf{d} = \mathbf{d}_{\text{translate}} + \mathbf{d}_{\text{KLT}} \quad (9)$$

In the Equation (9), \mathbf{d} is the displacement of different frames, and \mathbf{d}_{KLT} is displacement calculated from the drift-free KLT method. $\mathbf{d}_{\text{translate}}$ is the image translate pixel, calculated as follows:

$$\mathbf{d}_{\text{translate}} = \text{round}\left(\frac{D_{\text{predicted}}}{\lambda}\right) \quad (10)$$

where $D_{\text{predicted}}$ is the predicted displacement of the target object. Using a priori estimation in the proposed method improves the accuracy of displacement estimates by minimizing the impact of drift-type errors that can accumulate over time. Furthermore, by selecting feature points with strong texture in the reference frame and employing optical flow to calculate displacement, the method further improves the accuracy of displacement estimates.

3.2.2. Asynchronous Kalman Filter

The Kalman filter is a widely used method for data processing that estimates data by continuously predicting and correcting in the time domain. In general, the sampling frequency of the accelerometer is higher than the frame rate of the video. Smyth and Wu [60] used a multi-rate Kalman filter to fuse acceleration and displacement at different sampling rates to improve the estimation of the displacement signal. Ma et al. [55] proposed an asynchronous Kalman filter to fuse acceleration and displacement with adaptive parameters.

In the case of asynchronous situations, Ma et al. [55] categorized time steps into three types. Figure 6 shows the overview of the proposed methods. Type 1 involves only acceleration updates, while the second type involves visual updates. Type 3 involves acceleration updates following visual updates. Among these three types, only in type 2 are the values and probabilities of displacement fused when computing computer displacement updates.

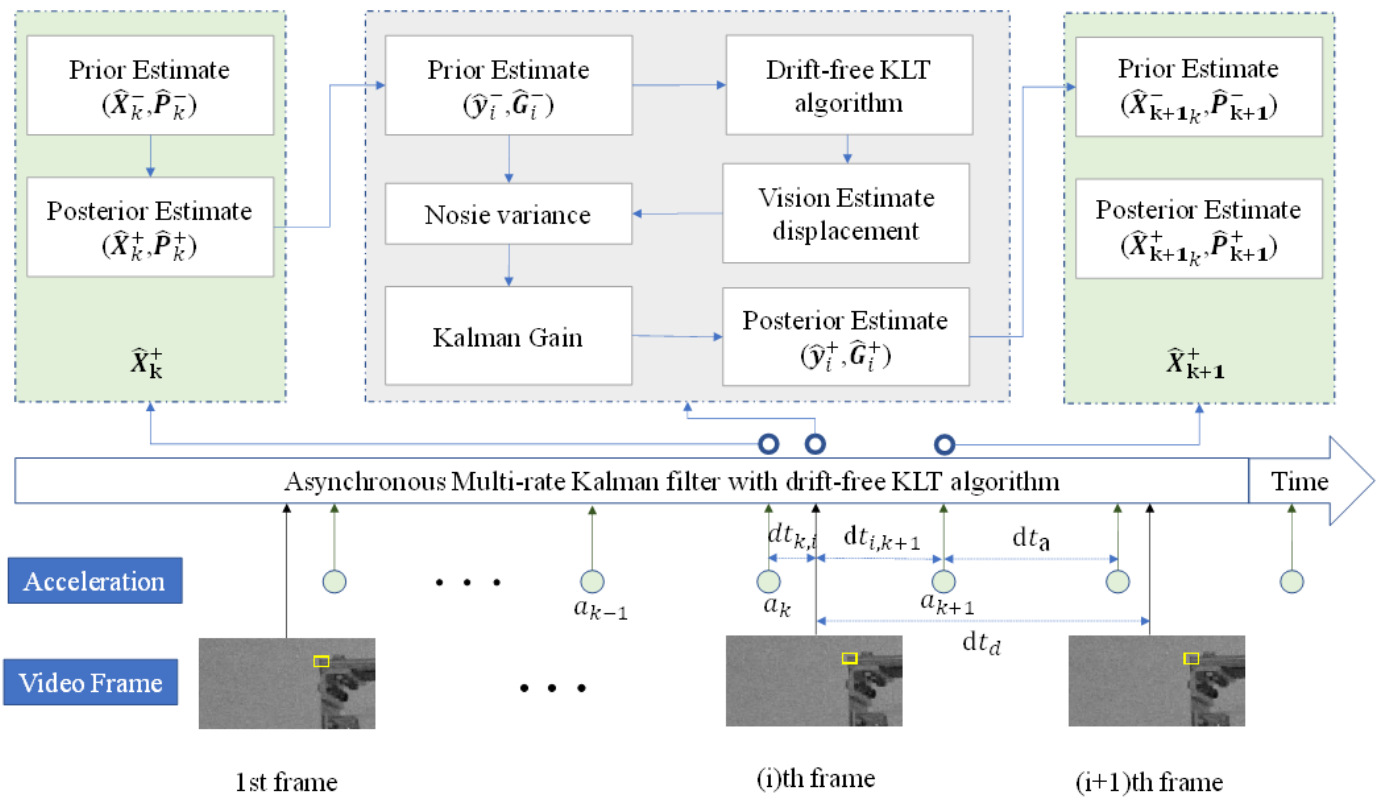


Figure 6. Overview of how to fuse asynchronous vision and acceleration measurement using an asynchronous multi-rate Kalman filter with drift-free KLT algorithm to estimate structural displacement.

Suppose $\mathbf{X}_k = [x_k, \dot{x}_k]^T$ is a state variable, and x_k, \dot{x}_k represents displacement and velocity, respectively, at the k -th time step, then a discrete state space model for the relationship between acceleration and displacement can be described as:

$$\mathbf{X}_k = \mathbf{A}(dt)\mathbf{X}_{k-1} + \mathbf{B}(dt)a_{k-1} + \mathbf{B}(dt)w_{k-1} \quad (11)$$

$$D_k = \mathbf{H}\mathbf{X}_k + v_k \quad (12)$$

where w_k and v_k are the noises of measured acceleration and displacement, respectively. Q and R are the corresponding variances of w_k and v_k , respectively. dt is the time interval of the time step. \mathbf{A} and \mathbf{B} are the state transition matrix and control input matrix, respectively. In this case, they are functions of the time interval:

$$\mathbf{A}(dt) = \begin{bmatrix} 1 & dt \\ 0 & 1 \end{bmatrix}; \mathbf{B}(dt) = \begin{bmatrix} dt^2/2 \\ dt \end{bmatrix}; \mathbf{H} = [\mathbf{1} \quad \mathbf{0}] \quad (13)$$

Assume that during type 1, only acceleration is considered. The $\hat{\mathbf{X}}_k^-$ and its covariance $\hat{\mathbf{P}}_k^-$ were obtained as follows:

$$\hat{\mathbf{X}}_k^- = \mathbf{A}(dt_a)\hat{\mathbf{X}}_{k-1}^+ + \mathbf{B}(dt_a)a_{k-1} \quad (14)$$

$$\hat{\mathbf{P}}_k^- = \mathbf{A}(dt_a)\hat{\mathbf{P}}_{k-1}^- \mathbf{A}^T(dt_a) + \mathbf{Q}(dt_a) \quad (15)$$

$$\mathbf{Q}(dt_a) = q \begin{bmatrix} dt_a^3/3 & dt_a^2/2 \\ dt_a^2/2 & dt_a \end{bmatrix} \quad (16)$$

where dt_a and q denote the time interval and noise variance of the acceleration measurements, respectively. The q value can be easily estimated using laboratory testing.

Since no other measurement is available in this time interval,

$$\hat{\mathbf{X}}_k^+ = \hat{\mathbf{X}}_k^-; \hat{\mathbf{P}}_k^+ = \hat{\mathbf{P}}_k^- \quad (17)$$

In type 2, the prior state $\hat{\mathbf{Y}}_i^-$ and covariance $\hat{\mathbf{G}}_i^-$ can be estimated according to the following state estimation:

$$\hat{\mathbf{Y}}_i^- = \mathbf{A}(dt_{k,i})\hat{\mathbf{x}}_k + \mathbf{B}(dt_{k,i})a_k \quad (18)$$

$$\hat{\mathbf{G}}_i^- = \mathbf{A}(dt_{k,i})\hat{\mathbf{P}}_k\mathbf{A}^T(dt_{k,i}) + \mathbf{Q}(dt_{k,i}) \quad (19)$$

where $dt_{k,i}$ denotes the time interval between the k -th acceleration and the i -th vision measurements. With the $\hat{\mathbf{Y}}_i^-$, the drift-free KLT method was applied to estimate displacement d_i from vision measurements.

The posterior state and its covariance were calculated as follows:

$$\hat{\mathbf{Y}}_i^+ = \hat{\mathbf{Y}}_i^- + \hat{\mathbf{P}}_k\mathbf{H}^T(\mathbf{H}\hat{\mathbf{P}}_k\mathbf{H}^T + \mathbf{R})^{-1}(D_i - \mathbf{H}\hat{\mathbf{Y}}_i^-) \quad (20)$$

$$\hat{\mathbf{G}}_i^+ = (\mathbf{I} - \hat{\mathbf{P}}_k\mathbf{H}^T(\mathbf{H}\hat{\mathbf{P}}_k\mathbf{H}^T + \mathbf{R})^{-1})\hat{\mathbf{G}}_i^- \quad (21)$$

Here, R is calculated as follows:

$$R = \sigma_D^2 / dt_{k,i} \quad (22)$$

where σ_D^2 is the observation noise of displacement measurement.

In type 3, the prior state and covariance are estimated according to the following state estimation:

$$\hat{\mathbf{X}}_{k+1} = \mathbf{A}(dt_{i,k+1})\hat{\mathbf{Y}}_i^+ + \mathbf{B}(dt_{i,k+1})a_k \quad (23)$$

$$\hat{\mathbf{P}}_{k+1} = \mathbf{A}(dt_{i,k+1})\hat{\mathbf{G}}_i^+\mathbf{A}^T(dt_{i,k+1}) + \mathbf{Q}(dt_{i,k+1}) \quad (24)$$

3.2.3. Parameter Estimation

As described in Equation (8), the actual displacement is the product of the scale factor and the pixel displacement. Therefore, according to the law of error transfer, the variance of the displacement measurement can be calculated by the following equation:

$$\sigma_D^2 = \sigma_\lambda^2 \cdot \bar{D}^2 + \sigma_D^2 \cdot \bar{\lambda}^2 \quad (25)$$

where σ_u^2 is the variance of the displacement measurement, \bar{D} and σ_D^2 are the mean and variance of the displacement, respectively, and $\bar{\lambda}$ and σ_λ^2 are the mean and variance of the scale factor, respectively. For structural monitoring, the mean value of displacement \bar{d} can be assumed to be 0, and the variance of displacement σ_d^2 is estimated by calculating the mean of the variance of all frames based on the matching results for each frame.

3.3. Comparison with Conventional Motion Estimation Approaches

This paper compares two commonly used motion estimation methods: (a) a feature-matching-based method [32] and (b) the commonly used KLT tracker, as mentioned in Section 2. The feature-matching-based method consists of the following steps: (1) video preprocessing: the specific step is the same as the method in this article. (2) Feature detection and feature description: this step detects distinctive features or key points in ROI. These features are usually corners, edges, or regions with rich textures. Here, the

Harris corner detection algorithm is used. After detecting the features, a descriptor is computed for each feature. (3) Feature matching: the matching step involves comparing the descriptors from the two images and finding the best match for each feature. (4) Outlier removal: since not all matched features correspond to the same physical point in the scene, some matches might be incorrect or outliers. One effective technique for outlier removal is random sample consensus (RANSAC), which is commonly used in computer vision and image processing. (5) Motion estimation: the relative displacement between the two images can be computed with the set of correctly matched features.

4. Small-Scale Laboratory Validation

4.1. Experimental Setup

The proposed method for drift-free large motion measurement is investigated in a laboratory experiment to determine its performance and its sensitivity to the video's frame rate. Figure 7 illustrates the validation of a three-story steel building model excited by a uniaxial shaking table. The simultaneous measurement of structural responses was conducted using the proposed system and a laser displacement sensor used for ground truthing; the details regarding these devices are in Table 1. The algorithm was implemented in MATLAB, running on a PC with a 2.3 GHz Intel i7 processor and 32 GB of RAM. In this experiment, three different types of excitations were used at the bottom of the structure: (1) 1 Hz sine excitation, (2) 4 Hz sine excitation, and (3) earthquake excitation.

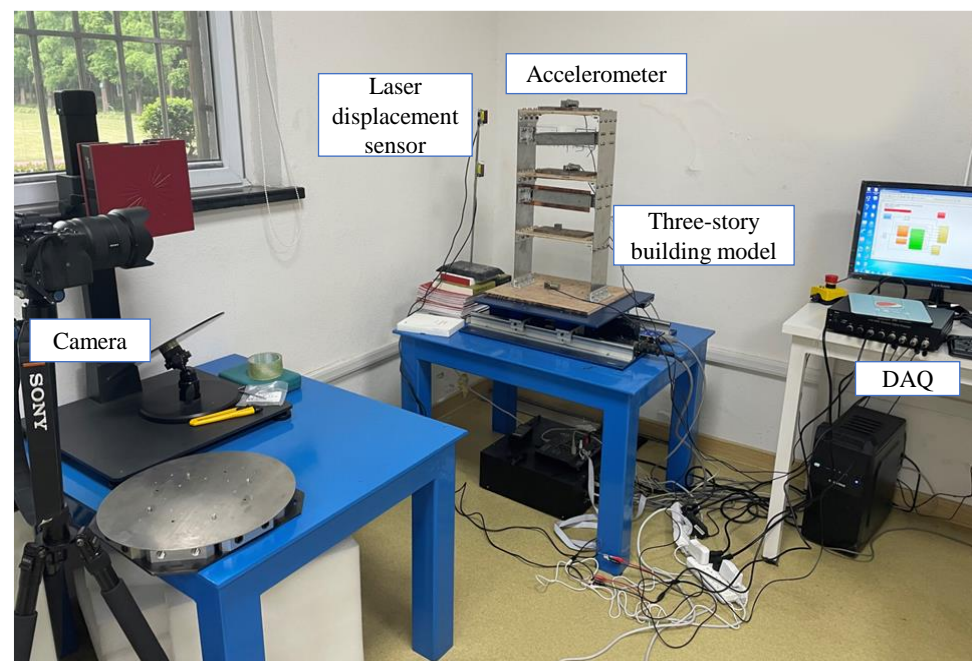


Figure 7. Experimental setup of shaking table test.

Table 1. Details of the cameras and sensors.

Type	Description
Camera	A Sony ILCE-7RM4 camera, featuring a resolution of 1920×1080 p, is utilized to capture the video of the structural vibration at a frame rate of 100 fps.
Laser displacement sensor (LDS)	A Panasonic HG-C 1200 micro laser distance sensor is employed to supply the ground-truth displacement data for the top floor, with a sampling rate of 500 Hz.
Accelerometer	A KT-1100 accelerometer is employed to deliver the acceleration data for the top floor, with a sampling rate of 500 Hz.

4.2. Experimental Result

To quantify the measurement accuracy of the results, the error analysis is conducted using the normalized root-mean-square error (NRMSE):

$$NRMSE = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{x}_i - x_i)^2}}{\max(x_i) - \min(x_i)} \quad (26)$$

where \hat{x} is the estimated displacement; x is the reference displacement; N is the number of displacement measurements.

Figure 8 shows the grayscale initial video frame; the selected target region is framed in a red box containing the salient corner features to be tracked. This figure shows that the Harris detector successfully detects the corner of the structure and other feature points. After selecting the feature points, the feature-matching-based method is employed to estimate the movement of the target object for each frame of the video in the calibration stage.

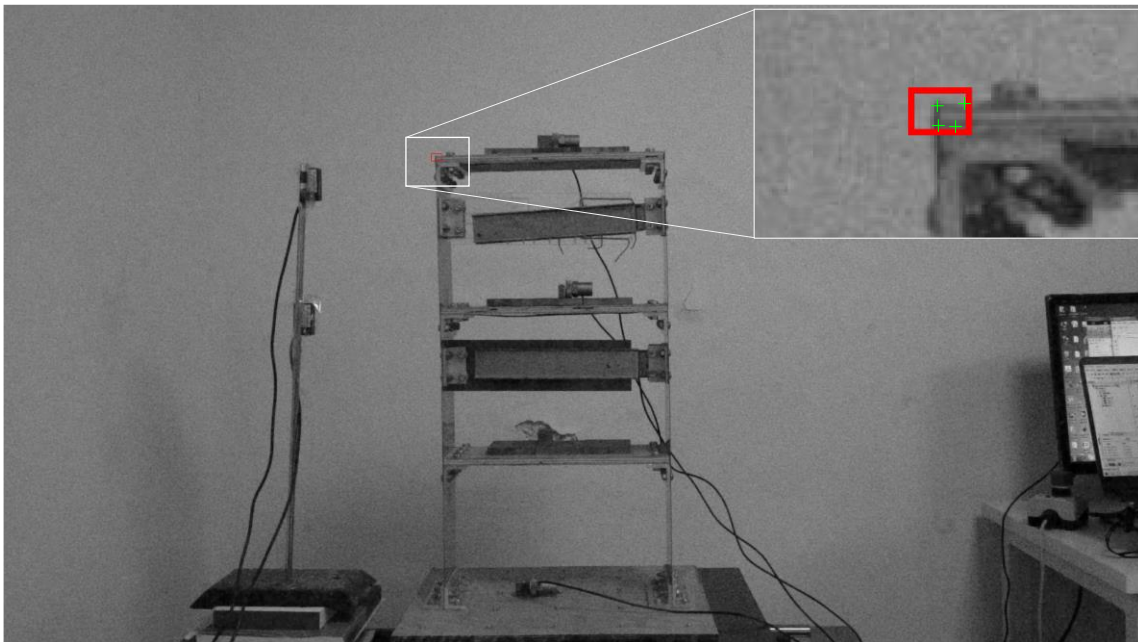


Figure 8. Selected target (red box) at initial video frame and initial track point (green point).

Under case (1), the scale factor calculation results proposed in this paper are shown in Figure 9, while the scale factor obtained through structural size measurement is 0.78 mm/pixel. The two-scale factor shows that the scale method estimated here is effective, so in the absence dimension scenario, the scale factor can be estimated in this way.

In the Kalman filter process, the noise parameter q is selected as $10^4 \text{ mm}^2/\text{s}^2$ in this experiment, which is estimated based on prior experience. For case (1), with a video sampling rate of 100 Hz, the results are shown in Figure 10. As shown in the figure, compared to the feature-based and KLT methods, the KLT method exhibits a significant drift phenomenon, while the drift-free method proposed in this paper does not have this issue. All comparisons are made here by linearly interpolating the data to 500 Hz. The method proposed in this paper can improve the NRMSE value, reducing it by 38% and 83%, respectively.

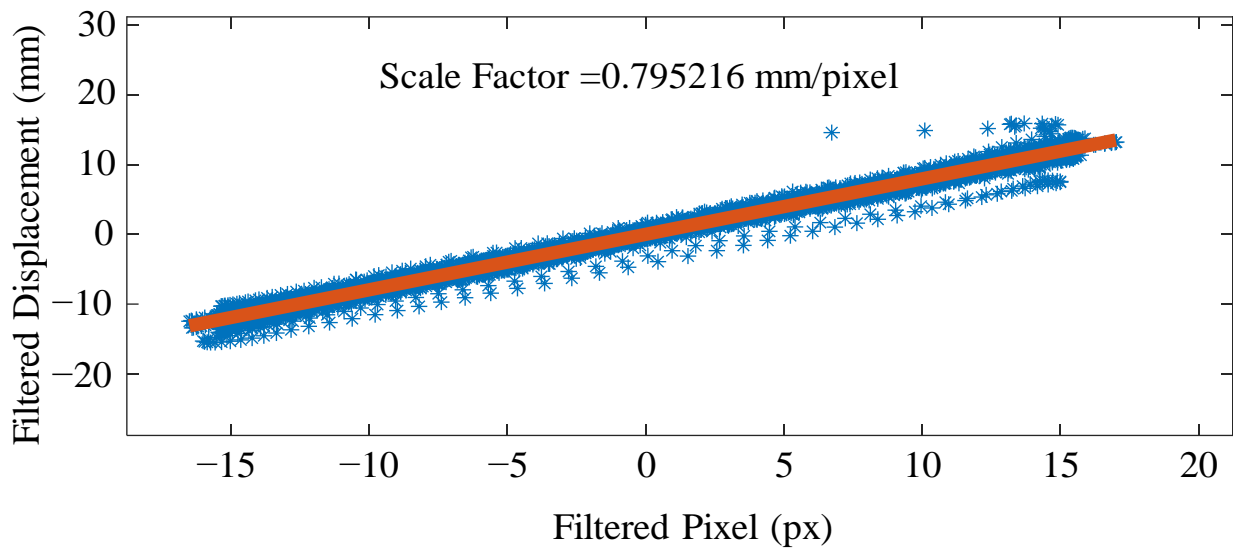


Figure 9. Scale factor of the experiment.

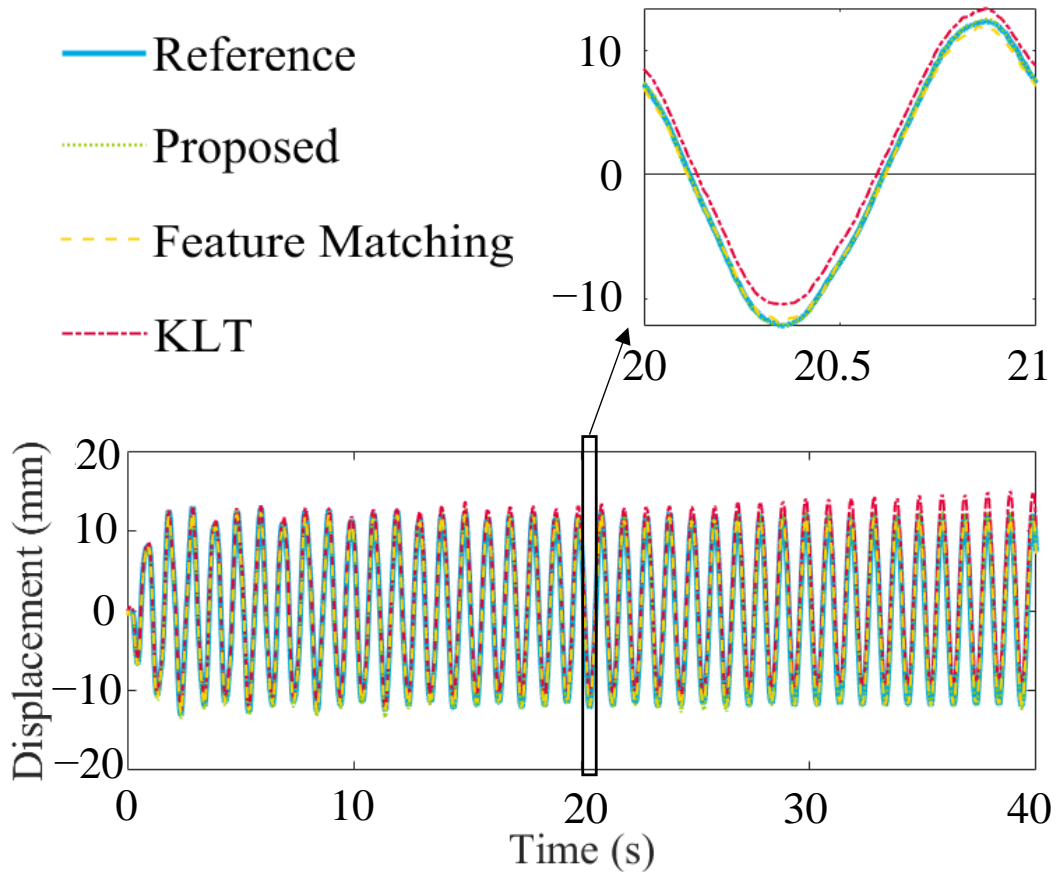


Figure 10. Comparison between ground truth and displacements estimated by vision algorithms.

Figure 11 shows that the target is within ROI using the proposed image translating method. In this figure, the frame below moves 16 pixels, and the target in the ROI roughly remains the same. Thus, the effectiveness of image translation during significant displacement is verified.

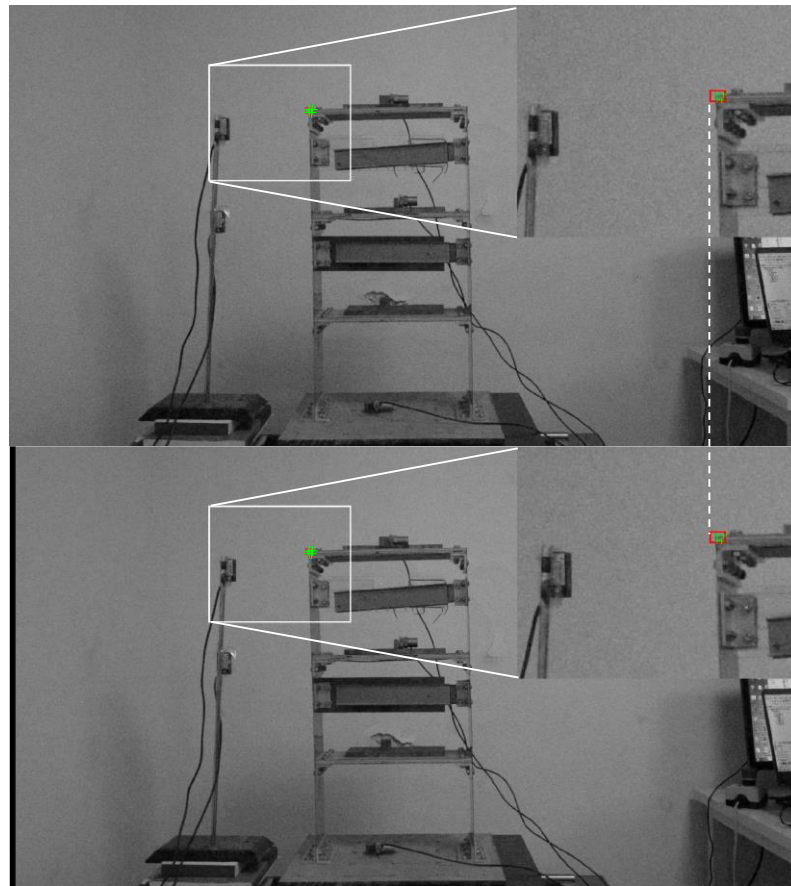


Figure 11. Comparison of ROI at different steps of the case (1). The red box represents the ROI, and the green dots represent the feature points.

The vision sample frequency was modified to investigate the influence of the vision sample frequency further and reduce computation time. In case (1), by resampling the video and reducing the sampling frequency to 50 Hz, 25 Hz, and 10 Hz, it can be found that as the sampling frequency decreases, the NRMSE values increase to 0.91%, 1.52%, and 1.51%, respectively, as shown in Figure 12. In comparison, the feature-matching method changes to 1.57%, 1.66%, and 2.51%, respectively, and the result for the KLT method is higher than the feature-matching method. In case (1), since the excitation frequency is only 1 Hz, the forced vibration frequency can be accurately captured in all cases.

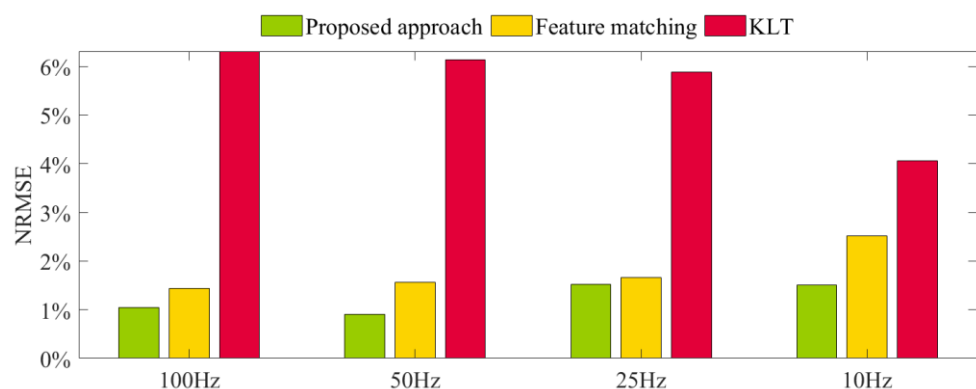


Figure 12. NRMSE of different sample frequency.

In case (2), the input frequency at the base of the structure was set to 4 Hz, with an amplitude of 30 mm, allowing for the evaluation of the effectiveness of the proposed

method under large displacement and high-frequency vibration conditions. As in previous tests, the laser displacement sensor at the top of the structure was used as a reference for calculating the error values. This experimental setup aimed to demonstrate the accuracy and reliability of the proposed method in large amplitude and high-frequency vibrations.

Figure 13 shows the displacement of 100 Hz and 10 Hz sample frequencies. As the figure shows, compared with 100 Hz, the pure 10 Hz vision sample frequency failed to capture several peak values. Under a frequency of 100 Hz, the NRMSE values for the proposed method and the feature matching method were 1.3% and 1.58%, respectively. At a sampling frequency of 10 Hz, the KLT method could not detect displacements and was, thus, omitted from the comparison. The NRMSE values for the proposed and feature matching methods at 10 Hz were 5% and 12%, respectively. These findings indicate that for high-frequency vibrations, the accuracy of purely visual methods is limited due to the constraints imposed by the Nyquist sampling theorem, preventing real-time data acquisition.

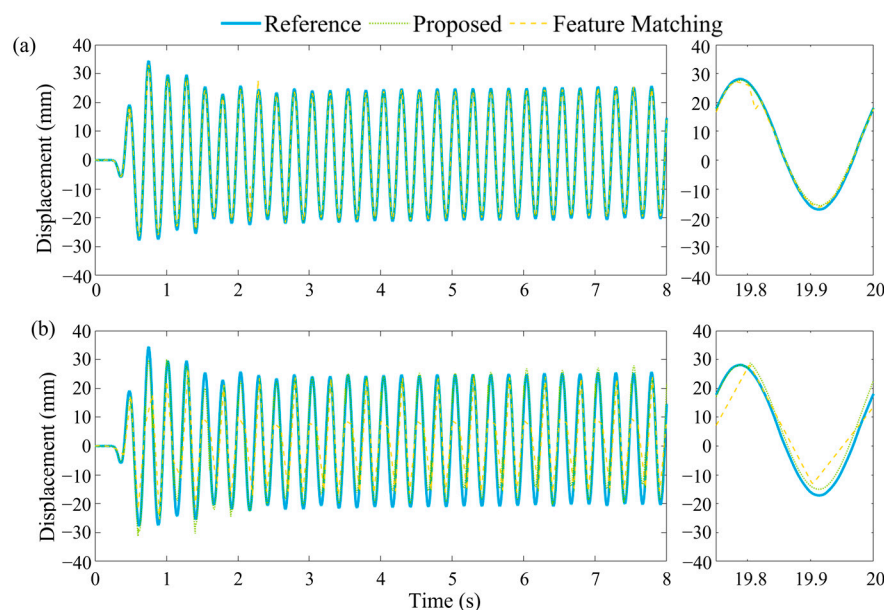


Figure 13. Comparison between ground truth and estimate displacement with different vision sample rates in case (2) (a) 100 Hz and (b) 10 Hz.

The computational time for each frame is presented in Table 2. Table 2 reveals that the proposed method's computation time is shorter than the feature-matching method but longer than the KLT algorithm. In principle, the computation time for the proposed method should be close to that of the KLT algorithm. The discrepancy in computation times may be attributed to the time required for image translation and algorithm initialization. Further investigation into optimizing the proposed method's computation time may help close the gap and make it more comparable to the KLT algorithm, enhancing its practical applicability in real scenarios. Additionally, the proposed method can provide real-time estimations of drift-free displacements due to the reduced computation time. This advantage makes the method more suitable for applications where rapid and accurate displacement measurements are critical. The proposed method can outperform alternative approaches, particularly in scenarios with high-frequency vibrations or large displacements, by offering a balance between accuracy and efficiency.

Case (3) presents the displacement of the frame under the excitation of the El Centro earthquake wave. Due to the frame's flexibility, unlike case (1) and case (2), the top-floor displacement is primarily governed by the frequency of the structure. In Case (3), the time history curves and NRMSE values were calculated for different video sampling rates, as shown in Figure 14. As the sampling rate decreases, the NRMSE increases from 0.83% to

0.93%, 0.91%, and 1.13%. This trend demonstrates the influence of sampling rate on the accuracy of displacement measurements.

Table 2. Compute time per frame of different algorithms.

Algorithm	Time Per Frame(s)
Proposed method	0.035
Feature matching method	0.172
KLT	0.009

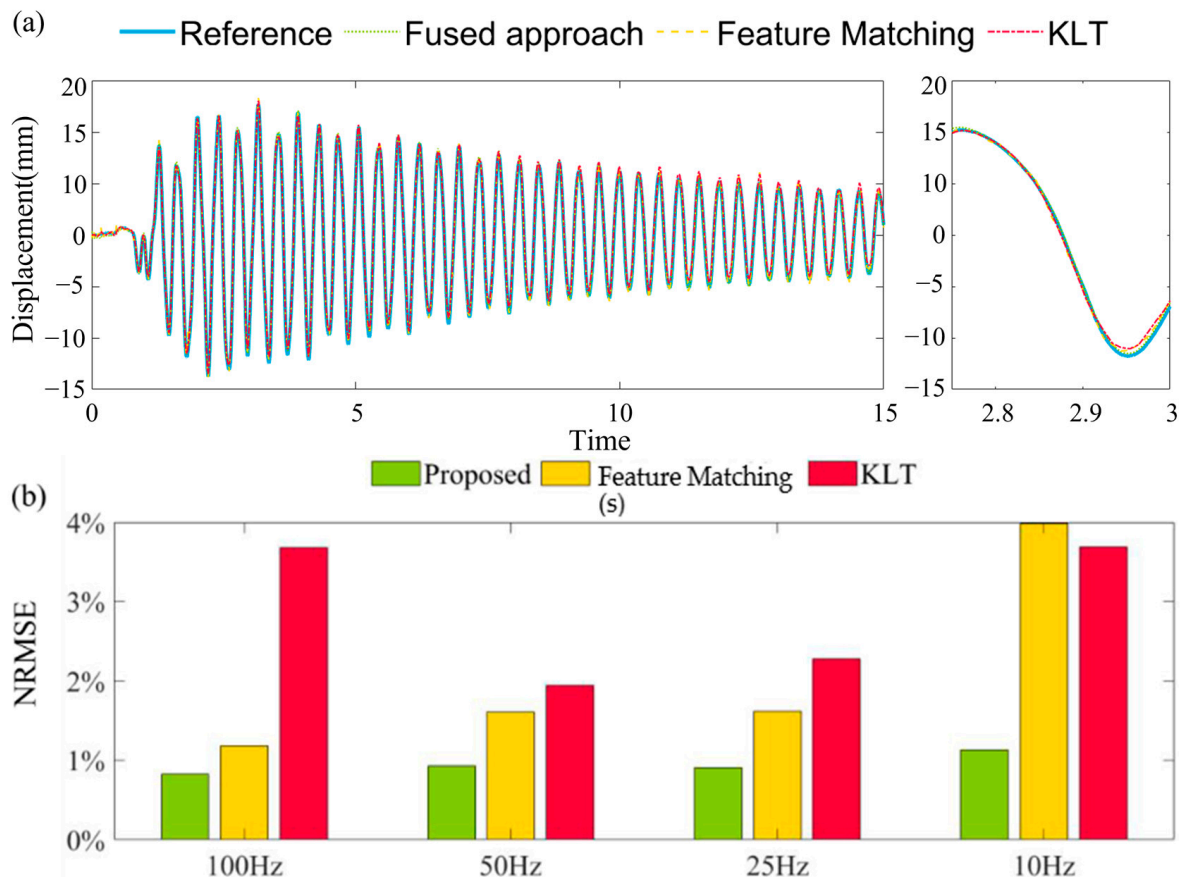


Figure 14. (a) time history curve for earthquake excitation (b) NRMSE of different sample frequency.

Under earthquake conditions, the NRMSE values decrease for the given conditions, indicating that the proposed method exhibits robustness. This improved performance demonstrates the method's ability to maintain accuracy and reliability even in challenging situations. The method's robustness is crucial in practical applications, where dynamic conditions and external disturbances can significantly impact the quality and reliability of displacement estimates.

Power spectral density (PSD) is a function used to describe the energy distribution of a signal in the frequency domain. PSD is frequently used in signal processing and communication systems to describe the spectral characteristics of noise and signals.

Figure 15 shows that the proposed method's PSD is closer to the reference measurement results. Note that here the PSD is calculated without interpretation. The high-frequency information of the structure is more similar to the LDV results, which is beneficial for determining the structure's frequency and mode shapes. For the low-frequency portion, after applying the Kalman filter, the power spectral density curve is closer to the pure visual results. The first frequency, 2.63 Hz, is successfully identified in both scenarios. Under the 10 Hz scenario, the vision method failed to identify the 6.84 Hz, the second

mode. The third frequency, 18.55 Hz, is not apparent from those curves. These results demonstrate the effectiveness of incorporating the Kalman filter in improving the accuracy of the displacement estimates, particularly in capturing the structure's essential dynamic characteristics across various frequency ranges.

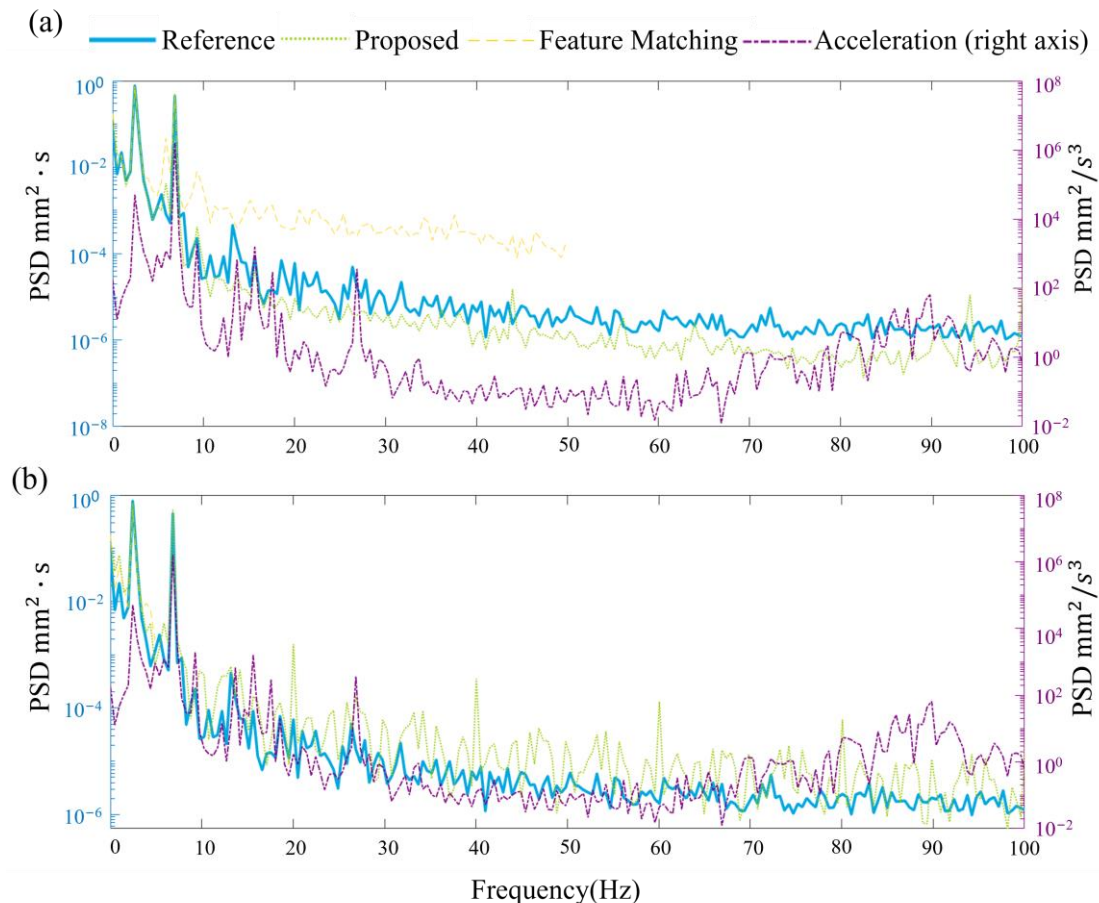


Figure 15. PSD for each measurement under earthquake excitation with different vision sample frequencies (a) 100 Hz (b) 10 Hz.

5. Conclusions

This paper uses an accelerometer and computer vision techniques to fuse contact monitoring and non-contact tracking data of structural dynamics to exploit both advantages fully. In response to the shortcomings that computer vision techniques cannot capture high-frequency vibration information of structures and require additional parameters to estimate the scaling factor, while accelerometers cannot monitor low-frequency displacements and have zero drift, this paper proposes to fuse data from computer vision and accelerometers using Kalman filtering to calculate the scaling factor using the least squares method.

The method's reliability is also verified by using a frame structure shaker table test. The results show that (1) the method can reliably estimate the scale factor. (2) In the time domain, the NRMSE value is effectively reduced, and the overall displacement measurement accuracy is improved. (3) In the frequency domain range, the proposed data fusion method compensates for the low sampling rate of pure computer vision and effectively improves the signal-to-noise ratio of displacement data in the higher-order mode range.

The study also investigated the impact of lowering the sampling frequency on the video vision technique. The findings reveal that the accuracy of the displacements is only slightly affected when the sampling frequency is decreased from 100 to 10 Hz. The fused displacements' power spectral densities remain unchanged, even though the sampling frequency is reduced to a tenth of its original value. This demonstrates that the proposed

fused method is a feasible and efficient alternative for measuring displacement in civil engineering structures.

Author Contributions: Conceptualization, C.X. and Y.W.; methodology, C.X.; software, C.X. and Y.W.; formal analysis, C.X.; investigation, C.X.; resources, W.S.; data curation, C.X.; writing—original draft preparation, C.X.; writing—review and editing, C.X.; visualization, C.X.; supervision, W.S.; project administration, W.S.; funding acquisition, C.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by National Key Research and Development Program of China (Grant no. 2022YFF0608903), Shanghai Pujiang Program (22PJ1413600) and the Fundamental Research Funds for the Central Universities (22120220573).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are available on request due to restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Huang, K.; Yuen, K.-V.; Wang, L. Real-time simultaneous input-state-parameter estimation with modulated colored noise excitation. *Mech. Syst. Signal Process.* **2022**, *165*, 108378. [[CrossRef](#)]
2. Shi, W.X.; Wang, L.K.; Lu, Z.; Wang, H.T. Experimental and numerical study on adaptive-passive variable mass tuned mass damper. *J. Sound Vib.* **2019**, *452*, 97–111. [[CrossRef](#)]
3. Qin, S.; Yuan, Y.; Han, S.; Li, S. A Novel Multiobjective Function for Finite-Element Model Updating of a Long-Span Cable-Stayed Bridge Using In Situ Static and Dynamic Measurements. *J. Bridge Eng.* **2023**, *28*, 04022131. [[CrossRef](#)]
4. Wang, L.K.; Nagarajaiah, S.; Shi, W.X.; Zhou, Y. Semi-active control of walking-induced vibrations in bridges using adaptive tuned mass damper considering human-structure-interaction. *Eng. Struct.* **2021**, *244*, 112743. [[CrossRef](#)]
5. Wang, L.K.; Nagarajaiah, S.; Zhou, Y.; Shi, W.X. Experimental study on adaptive-passive tuned mass damper with variable stiffness for vertical human-induced vibration control. *Eng. Struct.* **2023**, *280*, 115714. [[CrossRef](#)]
6. Shi, W.X.; Wang, L.K.; Lu, Z. Study on self-adjustable tuned mass damper with variable mass. *Struct. Control Health* **2018**, *25*, e2114. [[CrossRef](#)]
7. Shi, W.X.; Wang, L.K.; Lu, Z.; Zhang, Q.W. Application of an Artificial Fish Swarm Algorithm in an Optimum Tuned Mass Damper Design for a Pedestrian Bridge. *Appl. Sci.* **2018**, *8*, 175. [[CrossRef](#)]
8. American Society of Civil Engineering. *Seismic Rehabilitation of Existing Buildings*; American Society of Civil Engineers: Reston, VA, USA, 2007.
9. Wang, L.K.; Nagarajaiah, S.; Shi, W.X.; Zhou, Y. Seismic performance improvement of base-isolated structures using a semi-active tuned mass damper. *Eng. Struct.* **2022**, *271*, 114963. [[CrossRef](#)]
10. Wang, L.K.; Shi, W.X.; Zhou, Y. Adaptive-passive tuned mass damper for structural aseismic protection including soil-structure interaction. *Soil Dyn. Earthq. Eng.* **2022**, *158*, 107298. [[CrossRef](#)]
11. Wang, L.K.; Shi, W.X.; Zhou, Y. Study on self-adjustable variable pendulum tuned mass damper. *Struct. Des. Tall Spec. Build.* **2019**, *28*, e1561. [[CrossRef](#)]
12. Wu, G.-M.; Yi, T.-H.; Yang, D.-H.; Li, H.-N. Damage Identification of Tie-Down Cables in Cable-Stayed Bridges Using Vehicle-Induced Displacement. *J. Perform. Constr. Facil.* **2021**, *35*, 04021011. [[CrossRef](#)]
13. Zhang, Q.; Luo, X.; Ding, J.; Xie, B.; Gao, X. Dynamic response evaluation on TMD and main tower of Shanghai Tower subjected to Typhoon In-Fa. *Struct. Des. Tall Spec. Build.* **2022**, *31*, e1929. [[CrossRef](#)]
14. Wang, L.K.; Shi, W.X.; Li, X.W.; Zhang, Q.W.; Zhou, Y. An adaptive-passive retuning device for a pendulum tuned mass damper considering mass uncertainty and optimum frequency. *Struct. Control. Health* **2019**, *26*, e2377. [[CrossRef](#)]
15. Li, J.; Hao, H. Health monitoring of joint conditions in steel truss bridges with relative displacement sensors. *Measurement* **2016**, *88*, 360–371. [[CrossRef](#)]
16. Nassif, H.H.; Gindy, M.; Davis, J. Comparison of laser Doppler vibrometer with contact sensors for monitoring bridge deflection and vibration. *Ndt E Int.* **2005**, *38*, 213–218. [[CrossRef](#)]
17. Celebi, M. GPS in dynamic monitoring of long-period structures. *Soil Dyn. Earthq. Eng.* **2000**, *20*, 477–483. [[CrossRef](#)]
18. Kuddus, M.A.; Li, J.; Hao, H.; Li, C.; Bi, K. Target-free vision-based technique for vibration measurements of structures subjected to out-of-plane movements. *Eng. Struct.* **2019**, *190*, 210–222. [[CrossRef](#)]
19. Tan, D.; Li, J.; Hao, H.; Nie, Z. Target-free vision-based approach for modal identification of a simply-supported bridge. *Eng. Struct.* **2023**, *279*, 115586. [[CrossRef](#)]
20. Topal, G.O.; Akpinar, B. The efficiency of single base and network RTK for Structural Health Monitoring. *Adv. Geod. Geoinform.* **2022**, *71*, 12351.

21. Park, H.S.; Lee, H.M.; Adeli, H.; Lee, I. A New Approach for Health Monitoring of Structures: Terrestrial Laser Scanning. *Comput. Aided Civ. Infrastruct. Eng.* **2007**, *22*, 19–30. [\[CrossRef\]](#)
22. Xu, Y.; Wei, Y.; Wang, D.; Jiang, K.; Deng, H. Multi-UAV Path Planning in GPS and Communication Denial Environment. *Sensors* **2023**, *23*, 2997. [\[CrossRef\]](#)
23. Zanelli, F.; Castelli-Dezza, F.; Tarsitano, D.; Mauri, M.; Bacci, M.L.; Diana, G. Design and Field Validation of a Low Power Wireless Sensor Node for Structural Health Monitoring. *Sensors* **2021**, *21*, 1050. [\[CrossRef\]](#)
24. Chen, Z.; Fu, J.; Peng, Y.; Chen, T.; Zhang, L.; Yuan, C. Baseline Correction of Acceleration Data Based on a Hybrid EMD-DNN Method. *Sensors* **2021**, *21*, 6283. [\[CrossRef\]](#)
25. Zhang, H.; Mao, J.; Wang, H.; Zhu, X.; Zhang, Y.; Gao, H.; Ni, Y.; Hai, Z. A Novel Acceleration-Based Approach for Monitoring the Long-Term Displacement of Bridge Cables. *Int. J. Struct. Stab. Dyn.* **2023**, *23*, 530. [\[CrossRef\]](#)
26. Pan, C.; Zhang, R.; Luo, H.; Shen, H. Target-based algorithm for baseline correction of inconsistent vibration signals. *J. Vib. Control* **2017**, *24*, 2562–2575. [\[CrossRef\]](#)
27. Wang, Y.; Wang, L.K.; Shi, W.X. Two-dimensional air spring based semi-active TMD for vertical and lateral walking and wind-induced vibration control. *Struct. Eng. Mech.* **2021**, *80*, 377–390.
28. Wang, L.K.; Shi, W.X.; Zhang, Q.W.; Zhou, Y. Study on adaptive-passive multiple tuned mass damper with variable mass for a large-span floor structure. *Eng. Struct.* **2020**, *209*, 110010. [\[CrossRef\]](#)
29. Belongie, S.; Malik, J.; Puzicha, J. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 509–522. [\[CrossRef\]](#)
30. Feng, D.; Feng, M.Q. Vision-based multipoint displacement measurement for structural health monitoring. *Struct. Control Health Monit.* **2016**, *23*, 876–890. [\[CrossRef\]](#)
31. Corona, G.; Maciel-Castillo, O.; Morales-Castaneda, J.; Gonzalez, A.; Cuevas, E. A new method to solve rotated template matching using metaheuristic algorithms and the structural similarity index. *Math. Comput. Simul.* **2023**, *206*, 130–146. [\[CrossRef\]](#)
32. Wang, J.Q.; Zhao, J.; Liu, Y.W.; Shan, J.Z. Vision-based displacement and joint rotation tracking of frame structure using feature mix with single consumer-grade camera. *Struct. Control Health* **2021**, *28*, e2832. [\[CrossRef\]](#)
33. Luo, L.; Feng, M.Q.; Wu, Z.Y. Robust vision sensor for multi-point displacement monitoring of bridges in the field. *Eng. Struct.* **2018**, *163*, 255–266. [\[CrossRef\]](#)
34. Xu, Y.; Brownjohn, J.; Kong, D. A non-contact vision-based system for multipoint displacement monitoring in a cable-stayed footbridge. *Struct. Control Health Monit.* **2018**, *25*, e2155. [\[CrossRef\]](#)
35. Yoneyama, S.; Ueda, H. Bridge Deflection Measurement Using Digital Image Correlation with Camera Movement Correction. *Mater. Trans.* **2012**, *53*, 285–290. [\[CrossRef\]](#)
36. Ghorbani, R.; Matta, F.; Sutton, M.A. Full-Field Deformation Measurement and Crack Mapping on Confined Masonry Walls Using Digital Image Correlation. *Exp. Mech.* **2014**, *55*, 227–243. [\[CrossRef\]](#)
37. Wu, H.F.; Gyekenyesi, A.L.; Shull, P.J.; Yu, T.-Y.; Reagan, D.; Sabato, A.; Niezrecki, C. Unmanned aerial vehicle acquisition of three-dimensional digital image correlation measurements for structural health monitoring of bridges. In *Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, and Civil Infrastructure 2017*; SPIE: Portland, OR, USA, 2017.
38. Diamond, D.H.; Heyns, P.S.; Oberholster, A.J. Accuracy evaluation of sub-pixel structural vibration measurements through optical flow analysis of a video sequence. *Measurement* **2017**, *95*, 166–172. [\[CrossRef\]](#)
39. Sarrafi, A.; Mao, Z.; Niezrecki, C.; Poozesh, P. Vibration-based damage detection in wind turbine blades using Phase-based Motion Estimation and motion magnification. *J. Sound Vib.* **2018**, *421*, 300–318. [\[CrossRef\]](#)
40. Yan, Z.; Jin, Z.; Teng, S.; Chen, G.; Bassir, D. Measurement of Bridge Vibration by UAVs Combined with CNN and KLT Optical-Flow Method. *Appl. Sci.* **2022**, *12*, 5181. [\[CrossRef\]](#)
41. Won, J.; Park, J.W.; Park, K.; Yoon, H.; Moon, D.S. Non-Target Structural Displacement Measurement Using Reference Frame-Based Deepflow. *Sensors* **2019**, *19*, 2992. [\[CrossRef\]](#)
42. Lucas, B.D.; Kanade, T. An Iterative Image Registration Technique with an Application to Stereo Vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence, Vancouver, BC, Canada, 24–28 August 1981; pp. 674–679.
43. Tomasi, C.; Kanade, T. Detection and Tracking of Point Features. *Int. J. Comput. Vis.* **1991**, *9*, 137–154. [\[CrossRef\]](#)
44. Shi, J.; Tomasi, C. Good Features to Track. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 6 August 2002; pp. 593–600.
45. Gibson, J.J. *The Perception of the Visual World*; George Allen & Unwin: London, UK, 1950; pp. 118–131.
46. Liu, Y.M.; Tian, L.F. An Improved Algorithm on Adaptive KLT Vision Tracking. *Adv. Mater. Res.* **2013**, *631*, 1270–1275. [\[CrossRef\]](#)
47. Zhu, J.; Lu, Z.; Zhang, C. A marker-free method for structural dynamic displacement measurement based on optical flow. *Struct. Infrastruct. Eng.* **2020**, *18*, 84–96. [\[CrossRef\]](#)
48. Brox, T.; Malik, J. Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE Trans Pattern Anal. Mach. Intell.* **2011**, *33*, 500–513. [\[CrossRef\]](#)
49. Roberts, G.W.; Meng, X.L.; Dodson, A.H. Integrating a global positioning system and accelerometers to monitor the deflection of bridges. *J. Surv. Eng.* **2004**, *130*, 65–72. [\[CrossRef\]](#)
50. Kim, K.; Choi, J.; Chung, J.; Koo, G.; Bae, I.-H.; Sohn, H. Structural displacement estimation through multi-rate fusion of accelerometer and RTK-GPS displacement and velocity measurements. *Measurement* **2018**, *130*, 223–235. [\[CrossRef\]](#)

51. Chan, W.S.; Xu, Y.L.; Ding, X.L.; Dai, W.J. An integrated GPS-accelerometer data processing technique for structural deformation monitoring. *J. Geod.* **2006**, *80*, 705–719. [[CrossRef](#)]
52. Chang, C.C.; Xiao, X.H. An integrated visual-inertial technique for structural displacement and velocity measurement. *Smart Struct. Syst.* **2010**, *6*, 1025–1039. [[CrossRef](#)]
53. Zhu, H.; Gao, K.; Xia, Y.; Gao, F.; Weng, S.; Sun, Y.; Hu, Q. Multi-rate data fusion for dynamic displacement measurement of beam-like supertall structures using acceleration and strain sensors. *Struct. Health Monit.* **2019**, *19*, 520–536. [[CrossRef](#)]
54. Park, J.W.; Moon, D.S.; Yoon, H.; Gomez, F.; Spencer, B.F.; Kim, J.R. Visual-inertial displacement sensing using data fusion of vision-based displacement with acceleration. *Struct. Control. Health* **2018**, *25*, e2122. [[CrossRef](#)]
55. Ma, Z.X.; Choi, J.; Sohn, H. Real-time structural displacement estimation by fusing asynchronous acceleration and computer vision measurements. *Comput.-Aided Civ. Infrastruct. Eng.* **2022**, *37*, 688–703. [[CrossRef](#)]
56. Saleem, S.; Bais, A.; Sablatnig, R.; Ahmad, A.; Naseer, N. Feature points for multisensor images. *Comput. Electr. Eng.* **2017**, *62*, 511–523. [[CrossRef](#)]
57. Kim, H.; Cho, J.; Jung, Y.; Lee, S.; Jung, Y. Area-Efficient Vision-Based Feature Tracker for Autonomous Hovering of Unmanned Aerial Vehicle. *Electronics* **2020**, *9*, 1591. [[CrossRef](#)]
58. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [[CrossRef](#)]
59. Brandt, A. *Noise and Vibration Analysis*; John Wiley & Sons: Chichester, UK, 2011; pp. 45–46.
60. Smyth, A.; Wu, M. Multi-rate Kalman filtering for the data fusion of displacement and acceleration response measurements in dynamic system monitoring. *Mech. Syst. Signal Process.* **2007**, *21*, 706–723. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.