**Smart Structures and Systems**, *Vol. 31*, *No. 4* (*2023*) *409-419*
https://doi.org/10.12989/sss.2023.31.4.409


# Deformation estimation of truss bridges using two-stage optimization from cameras

Jau-Yu Chou and Chia-Ming Chang*

*Department of Civil Engineering*, *National Taiwan University*, *Taipei*, *Taiwan*

(*Received September 13*, *2022*, *Revised December 16*, *2022*, *Accepted February 8*, *2023*)

**Abstract.** Structural integrity can be accessed from dynamic deformations of structures. Moreover, dynamic deformations can be acquired from non-contact sensors such as video cameras. Kanade-Lucas-Tomasi (KLT) algorithm is one of the commonly used methods for motion tracking. However, averaging throughout the extracted features would induce bias in the measurement. In addition, pixel-wise measurements can be converted to physical units through camera intrinsic. Still, the depth information is unreachable without prior knowledge of the space information. The assigned homogeneous coordinates would then mismatch manually selected feature points, resulting in measurement errors during coordinate transformation. In this study, a two-stage optimization method for video-based measurements is proposed. The manually selected feature points are first optimized by minimizing the errors compared with the homogeneous coordinate. Then, the optimized points are utilized for the KLT algorithm to extract displacements through inverse projection. Two additional criteria are employed to eliminate outliers from KLT, resulting in more reliable displacement responses. The second-stage optimization subsequently fine-tunes the geometry of the selected coordinates. The optimization process also considers the number of interpolation points at different depths of an image to reduce the effect of out-of-plane motions. As a result, the proposed method is numerically investigated by using a truss bridge as a physics-based graphic model (PBGM) to extract high-accuracy displacements from recorded videos under various capturing angles and structural conditions.

**Keywords:** computer vision; deformation estimation; improved Kanade-Lucas-Tomasi algorithm; motion tracking; physics-based graphics model

## 1. Introduction

Damage assessment plays an important role in structural health monitoring (SHM). Displacement measurements can directly diagnose structures and inform structural soundness (Xu and Mita 2021). In conventional structural health monitoring systems, accelerometers and displacement transducers are commonly attached to the structural surface to capture vibrational measurements. However, the installation process is labor-intensive and costly due to the large dimension and high complexity of as-built infrastructures (Feng *et al*. 2015). Moreover, selecting a reference point for a displacement transducer is difficult in structural monitoring applications (Li *et al*. 2014). With computer vision techniques, displacement responses are obtained using contact-free optical measurements without reference point requirements.

To obtain measurements that can provide accurate damage information, the errors between the estimated and exact measurements should be negligible, and the dynamic content in the vibrational responses of structures should be consistent. For instance, vibration-based damage detection widely relies on modal properties obtained from operational modal analysis (OMA) (Liu 1995, Farrar *et al*. 2001, Carde and Fanning 2004). Thus, the estimated measurements must be accurate enough to identify modal properties, such as natural frequencies and mode shapes. In addition, correlations between exact and estimated measurements should also be investigated to prove the reliability of measurements. However, exact displacements and mode shapes are impossible to be measured because of uncertainties (i.e., depth estimation from video cameras) in measurements. Therefore, developing advanced measurement techniques to acquire structural responses should consider the precision and accuracy for use in damage detection of structures.

Optical sensors can yield displacements to be obtained from recorded videos (Patsias and Staszewskly 2002, Yoon *et al*. 2016, Cha *et al*. 2017). Recently, many researchers have focused on capturing vibrational measurements using different computer vision techniques. For instance, image feature points are extracted from video and tracked by the Kanade-Lucas-Tomasi (KLT) algorithm (Narazaki *et al*. 2020, Zhu *et al*. 2020). In addition, Zhu *et al*. (2020) and Dong *et al*. (2020) applied the optical flow method to recorded video to estimate structural vibrational response. Luu *et al*. (2011) combined digital image correlation (DIC) with B-Spline interpolation for accurate displacement measurement. Du *et al*. (2020) employed DIC to capture vibrations of stay cables and estimated the cable forces. Chou and Chang (2021) compared the performance of optical flow, DIC, and edge detection with phase-based

*Corresponding author, Associate Professor,
E-mail: changcm@ntu.edu.tw



ISSN: 1738-1584 (Print), 1738-1991 (Online)

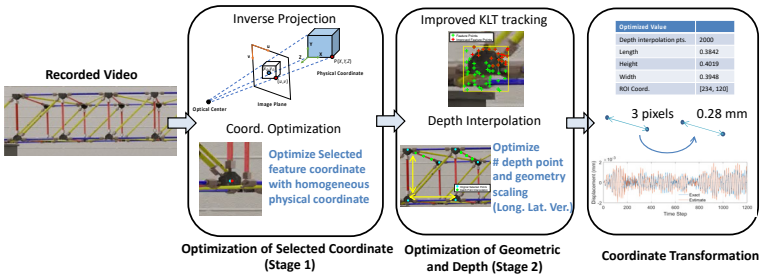



Fig. 1 Flowchart of the proposed approach

motion magnification using a scaled structure. These optical measurements can be directly exploited in vibration-based damage detection algorithms to assess structural integrity. Still, the out-of-plane motions cannot be separated flawlessly due to the lack of depth information in marker-free, contact-free videos.

Recently, digital twins have been proposed to construct a virtual model that can represent the dynamic behavior of actual structures with high accuracy (Ye *et al*. 2019, Ritto and Rochinha 2021). Moreover, Hoskere *et al*. proposed the physics-based graphics models (PBGM), representing the physical properties of a structure and its structural appearance (Hoskere *et al*. 2019). By applying camera intrinsic and extrinsic parameters, videos of structure responses from different capture angles and virtual environments can be obtained with the exact measurements provided. These videos are suitable for the validation of camera-based measurement algorithms since exact measurements are not accessible in field tests. In addition to camera and environment settings, the structural health conditions can also be adjusted. By tuning the structural properties to field measurements, the digital twins can be highly consistent with the as-built structure, and synthetic damage states can be further generated by varying the obtained structural properties. Conclusively, digital twins provide a perfect virtual platform for testing developed algorithms without setting up a series of experiments.

The objective of this study is to extract displacement in physical units using the PBGM provided by the IC-SHM committee. This study proposes a two-stage optimization method for computer-vision-based vibration measurements. First, manually selected feature points are optimized to match homogeneous physical coordinates by using the inverse projection matrix. Then, the depth and geometric properties are optimized using the KLT algorithm (Tomasi and Kanade 1991) with proposed filtering techniques. The measurements are evaluated at various camera angles and under a wide range of structural conditions. As seen in the results, the optimization process can improve measurements by around 50% with different camera positions. The optimized measurements successfully identify the modal properties with a certain level of accuracy. Moreover, the optimized measurements can generate digital twins with similar dynamic characteristics to the PBGM.

## 2. Two-stage optimization displacement estimation

A two-stage displacement estimation method for truss bridge structures using computer-vision-based vibration measurement is proposed in this study. First, the selected regions of interest (ROI) are optimized by minimizing the errors from the image plane to the homogeneous physical coordinates. An improved Kanade-Lucas-Tomasi (KLT, Tomasi and Kanade 1991) method is proposed with two additional outlier elimination criteria that extract displacement responses from the optimized ROIs. Finally, the geometry and interpolated depth points are optimized to reduce out-of-plane motion errors, and the displacements are converted into physical units. Fig. 1 shows the flowchart of the proposed approach. Each component will be introduced in the following subsections.

### *2.1 Optimization of motion tracking and physical unit conversion*

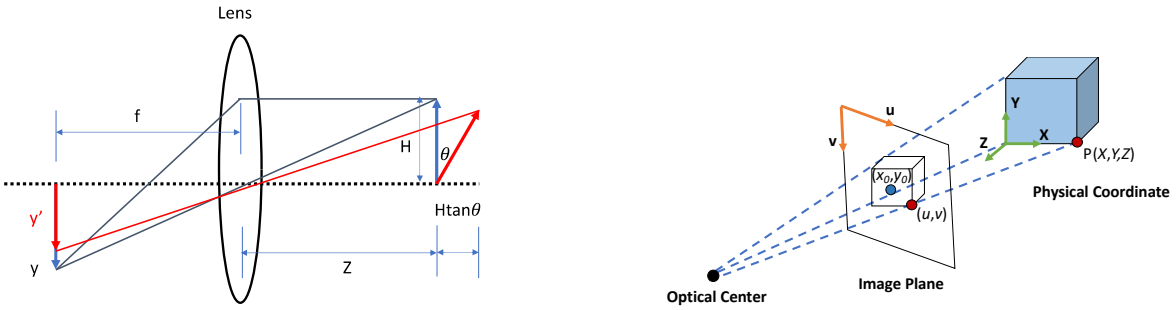
Structural integrity can be accessed by measuring structural displacements. Thus, conversion between the physical and pixel units is critical. In this study, the inverse camera projection matrix (Strat 1987) is applied to the pixel measurements obtained from the improved KLT tracking method and transforms the image coordinates into physical coordinates.

To eliminate the distortion from out-of-plane motions, the camera system in Fig. 2(a) is considered. As mentioned in Sutton *et al*. (2008), the relationship between the original and rotated objects can be represented as

$$\frac{y'}{y} \approx 1 - \frac{H\tan\theta}{Z} \tag{1}$$

where $y$ and $y'$ denote the height of the object without and with a rotation angle $\theta$ in the image plane, respectively; $H$ denotes the object height in the physical coordinate, and $Z$ is the object-to-camera distance. This factor, $y'/y$, can be calculated by differentiating the object height measured from the optical center in the image plane within each frame. The unrotated responses are recovered by dividing the measured pixel response by the factor.

Assuming a simple pinhole camera model, as shown in Fig. 2(b), the point P is located at the homogeneous physical coordinate ($X$, $Y$, $Z$) and is mapped to the image plane with pixel coordinates (u, v). The linear transformation between the image coordinates and the




(a) Out-of-plane rotation effect to camera system (b) Camera model

Fig. 2 Illustration of a pinhole camera

physical coordinates is given by

$$w\begin{bmatrix}\mathrm{u}\\ \mathrm{v}\\ 1\end{bmatrix}=\begin{bmatrix}f_{\mathrm{x}} & 0 & x_0\\ 0 & f_{\mathrm{y}} & y_0\\ 0 & 0 & 1\end{bmatrix}[\mathbf{R}_{3\times3} \quad \mathbf{T}_{3\times1}]\begin{bmatrix}\mathrm{X}\\ \mathrm{Y}\\ \mathrm{Z}\\ 1\end{bmatrix} \tag{2}$$

where $f_{\mathrm{x}}$ and $f_{\mathrm{y}}$ are the focal lengths in the x and y directions, respectively; **R** and **T** are the rotational and translational matrices of the camera pose based on the origin of the homogeneous physical coordinate; $w$ is the scaling ratio that converts physical units to pixel units that can be calculated through the projection matrix and the known homogeneous physical coordinate (See Eq. (2)); $x_0$ and $y_0$ are the principal point coordinates of the camera.

With an unknown camera intrinsic matrix (i.e., focal length and principal points), Eq. (2) can be rewritten as

$$w\begin{bmatrix}u\\ v\\ 1\end{bmatrix}=\mathbf{Proj}_{3\times4}\begin{bmatrix}X\\ Y\\ Z\\ 1\end{bmatrix} \tag{3}$$

where **Proj** is the projection matrix. By selecting 5 or more points in the image plane with corresponding homogeneous physical coordinates, the projection matrix is determined through the direct linear transformation (DLT, Abdel-Aziz and Karara 1971). Moreover, the projection matrix **Proj** is expanded into a square matrix, **S**, by adding a row of [0 0 0 1] at the bottom of **Proj**. The expansion of **Proj** does not affect the transformation but allows the inverse projection matrix to be calculated. The inversed projection matrix can be determined, and the transformation from the image coordinates to the homogeneous physical coordinates is given by

$$\mathbf{S}_{4\times4}{}^{-1}\begin{bmatrix}wu\\ wv\\ w\\ 1\end{bmatrix}=\begin{bmatrix}X_{\mathrm{est}}\\ Y_{\mathrm{est}}\\ Z_{\mathrm{est}}\\ 1\end{bmatrix} \tag{4}$$

where $X_{\mathrm{est}}$, $Y_{\mathrm{est}}$, $Z_{\mathrm{est}}$ are the estimated physical coordinates.

Accuracy of the image to physical conversion can be affected by the manually selected image points. That is, errors are induced by the selected points with minor differences in the corresponding homogeneous physical coordinates. Thus, an optimization using the genetic algorithm (GA, Whitley 1994) is employed. For example, the selected pixel coordinates are considered as the variable, and the objective function is given by

$$(\hat{u}\,,\hat{v}\,)=\operatorname{argmin} f(u\,,v\,)=\sqrt{\frac{1}{N}\sum_{i=1}^{i=N}\left\|\mathbf{P}_i-\mathbf{P}_{\mathrm{est},i}\right\|^2} \tag{5}$$

where $\hat{u}$ and $\hat{v}$ are the optimized locations in $u$ and $v$ coordinates of the image plane; $N$ is the total number of selected points; **P** is the selected homogeneous physical coordinate $(X\,,Y\,,Z\,)$; $\mathbf{P}_{\mathrm{est}}$ is the estimated physical coordinates using the inverse projection matrix in Eq. (4). Finally, the optimized locations are determined and selected as references for motion tracking.

### *2.2 Optimization of motion tracking*

The ROIs are determined using a rectangular box with a fixed length and width, with the optimized location point as the centroid. For each ROI, the Kanade-Lucas-Tomasi (KLT, MATLAB 2020) tracking method is applied along with feature point extraction using the minimum eigenvalue algorithm (Shi and Tomasi 1994). These feature points are ideally assumed to have similar motions and can be averaged to determine a more accurate movement within two frames. However, extracted feature points include motionless points in the background and feature points at a different depth, resulting in slight deteriorations in the estimation. Thus, two criteria are applied in this study to eliminate unideal feature points. First, the root mean square errors (RMSE) between the mean response and the response of the feature points are calculated by

$$R_i=\sqrt{\frac{1}{t}\sum_{j=1}^{t}\left(d_{ji}-\frac{1}{N}\sum_{i=1}^{N}d_{ji}\right)^2} \tag{6}$$

where $d_{ji}$ is the motion response of the $i$-th feature point in pixels at the $j$-th time step; $t$ is the total time step; $R_i$ is the RMSE of the $i$-th feature point. Because the responses within the ROI are assumed to be similar, only the

responses with $R$ satisfying Eq. (7) are selected as candidates.

$$\mathrm{E}[\mathbf{R}] - \sigma_{\mathbf{R}} < R_i < \mathrm{E}[\mathbf{R}] + \sigma_{\mathbf{R}}$$
$$\sigma_{\mathbf{R}} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(R_i - \mathrm{E}[R])^2} \tag{7}$$

where $\mathrm{E}[\mathbf{R}]$ is the mean value of the RMSEs over all feature points, and $\sigma_{\mathbf{R}}$ is the standard deviation of $\mathbf{R}$. In addition to error magnitude similarity, singular value decomposition (SVD) is applied to extract shape similarity, and the dominated motion is obtained by the first left singular vector (Wall *et al*. 2003), such as

$$\mathbf{d} = \mathbf{U}_{\mathrm{m}}\mathbf{S}_{\mathrm{m}}\mathbf{V}_{\mathrm{m}}^{\mathrm{T}} + \mathbf{U}_0\mathbf{S}_0\mathbf{V}_0^{\mathrm{T}} \tag{8}$$

where $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{S}$ are the left singular vector, right singular vector, and singular value matrix, respectively, and $\mathbf{d}$ represents a response matrix with a dimension of $t$ by $N$. The subscription "m" and "0" indicates the main and the less influential components of the response. Then, a similarity index, $\Gamma_i$, for the $i$-th feature point response is calculated by

$$\Gamma_i = \frac{1}{\|\mathbf{d}_i\|}\mathbf{d}_i^T\mathbf{U}_{\mathrm{m}} \tag{9}$$

where $\mathbf{d}_i$ is $i$-th feature point response, and $\mathbf{U}_{\mathrm{m}}$ only considers the first left singular vector. In this equation, if $\Gamma_i$ is close to 1, a high similarity is obtained. By Eq. (7) and Eq. (9), the outliers of the feature points are eliminated.

In addition, because the dimension of the homogeneous physical coordinate is approximated, the exact scale of the structure may differ from the selected points. Also, points from only two depth levels are used to construct the projection matrix, and then large errors are induced when estimating out-of-plane motion. In this study, depth point interpolation is utilized to increase the accuracy of depth (out-of-plane) measurements. As a result, the spatial dimensions (i.e., length, width, and depth of structural components) along with the number of depth interpolations are optimized by the objective function given by

$$\underset{l_{\mathrm{long}},l_{\mathrm{let}},l_{\mathrm{ver}},N_{\mathrm{int}}}{\operatorname{argmin}} f\left(l_{\mathrm{long}}, l_{\mathrm{lat}}, l_{\mathrm{ver}}, N_{\mathrm{int}}\right) = \sqrt{\frac{1}{p}\sum_{i=0}^{p}\left(\boldsymbol{y}_{\mathrm{exact}}^{i} - \boldsymbol{y}_{\mathrm{est}}^{i}\right)^2} \tag{10}$$

where $f(*)$ is the objective function; $\boldsymbol{y}_{\mathrm{exact}}$ is the exact measurements of $p$ points provided by the IC-SHM committee, and $\boldsymbol{y}_{\mathrm{est}}$ represents the measurements converted from pixel units using the inverse projection matrix, such as

$$\boldsymbol{y}_{\mathrm{est}} = \boldsymbol{S}^{-1}(l_{\mathrm{long}}, l_{\mathrm{lat}}, l_{\mathrm{ver}}, N_{\mathrm{int}})w\mathbf{d}_{\mathrm{est}}$$

where $l_{\mathrm{long}}$, $l_{\mathrm{lat}}$, and $l_{\mathrm{ver}}$ are the dimensionless scaling factors of the longitudinal, latitudinal, and vertical directions, respectively. $N_{\mathrm{int}}$ denotes the number of depth interpolations. The lower and upper bounds of the dimension factors are set to 0.5 to 1.5, and the limit of depth interpolations is 30,000. $\mathbf{d}_{\mathrm{est}}$ is the estimated measurement in pixel unit. The depth interpolation can be expressed as

$$\mathbf{P}_{\mathrm{depth},k} = \mathbf{P}_{\mathrm{F},k} + (\mathbf{P}_{\mathrm{B},k} - \mathbf{P}_{\mathrm{F},k})k/N_{\mathrm{int}} \quad k \in [1,\, N_{\mathrm{int}} - 1] \tag{11}$$

where $\mathbf{P}_{\mathrm{depth}}$ is the interpolated depth point in physical coordinated; $\mathbf{P}_{\mathrm{F}}$ and $\mathbf{P}_{\mathrm{B}}$ are the node points close to and far away from the camera along the same longitudinal axis of the truss bridge. The pixel coordinates of the depth points are calculated using Eq. (11) but with the physical points replaced by the pixel points. Note that if the dimensions of the structure are accurate, the optimization process in Eq. (10) can be neglected.

## 3. Displacement estimation using video motion

In this section, the setup of the physics-based graphics model (PGBM, Hoskere *et al*. 2019) is first introduced. The results of optimized measurements are compared with different approaches to investigate performance improvement by multiple criteria, including standard deviation, mode shape consistency, power spectral density, and correlation coefficients. Future opportunities, such as the construction of digital twins and the damage assessment of structures, are also introduced.

### *3.1 PBGM setup*

This study employs a simply-supported truss bridge as a physics-based graphics model (PBGM) to simulate three-dimensional displacement measurements captured by a camera. The datasets are provided by the IC-SHM committee. PBGM is a synthetic graphics model which can represent realistic deformations and geometry based on the simulated numerical model (Narazaki *et al*. 2020). As shown in Fig. 3, this model consists of 14 bays with hollow cylinder truss elements of outer and inner diameters of 0.01554 m and 0.01087 m. Young's modulus is approximately 18.5 GPa, and the shear modulus and material density are approximately 79.3 GPa and 8000 kg/m$^3$, respectively. The dimension of a single bay for the longitudinal, vertical, and transversal directions are 0.3937 m, 0.4000 m, and 0.3937 m. In the simulation, only the selected region is recorded with a camera, as shown in the bottom left of Fig. 3. The indices of each selected truss element are also represented in this figure. The structure is excited under ambient vibrations with a captured frame rate of 120 fps.

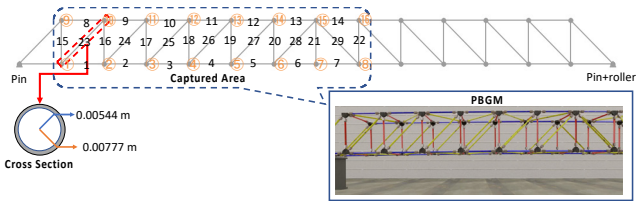
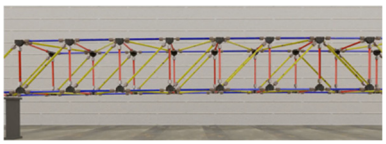


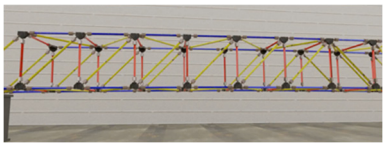

Fig. 3 Illustration of the PBGM

Table 1 Ground truth of all cases

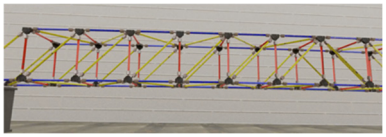
| Case | Damage location (index) | Stiffness reduction [%] |
|---|---|---|
| Undamaged | 0 | 0 |
| Damage 1 | 3 | 0.527 |
| Damage 2 | 18 | 0.527 |
| Damage 3 | 6 | 0.425 |
| Damage 4 | 16 | 0.725 |
| Damage 5 | 11 | 0.600 |
| Damage 6 | 20 | 0.88 |
| Damage 7 | 5 | 0.67 |
| Damage 8 | 9 | 0.48 |
| Damage 9 | 2 | 0.565 |
| Damage 10 | 12 | 0.62 |



(a) Undamaged



(b) Damage 4



(c) Damage 10

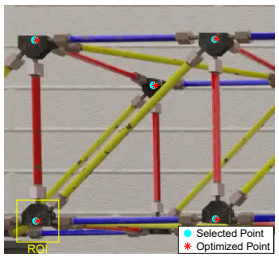
Fig. 4 Video with different capture angles

The objective of this study is to accurately extract the displacement response of the 16 nodes (sensors) under various structural conditions in terms of physical units (i.e., millimeters) and to investigate the dynamic behavior of the structure from recorded videos. In the provided dataset, 11 videos are generated using the PBGM, with each having a different capture angle, as shown in Fig. 4. In addition, each video has a specific damage state, which contains a single damage location and a stiffness reduction percentage. The damage is mimicked by reducing the Young's modulus per truss element. All damage cases are listed in Table 1.

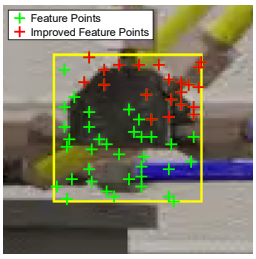
### *3.2 Measurement optimization*

In this study, the measured points and ROI are first manually selected, and the inverse projection matrix is sequentially obtained, as shown in the cyan dots in Fig. 5(a). Using the genetic algorithm (GA) with the objective function in Eq. (5), the optimal measured points are determined by minimizing the transformation errors from pixel to physical coordinates, as shown in the red stars in Fig. 5(a). For the GA optimization, the population size is 200 with a cross-over rate of 0.8. The manually selected points are allowed to move within 10 pixels. The mutation function generates adaptive random directions for the last generation and satisfies the constraints (MATLAB 2020). Within a specific ROI, the feature points (FP) can be extracted using the minimum eigenvalue algorithm (Shi and Tomasi 1994), as shown in Fig. 5(b). As introduced in the previous section, the outliers of the feature points can deteriorate the measurement, and therefore further processing steps such as Eqs. (6)-(9) is applied to filter out the outliers, as shown in the red dots in Fig. 5(b). Note that some background feature points are accepted by the proposed filters due to magnitude and motion similarity with the majority of responses. Moreover, the discontinuity in different background texture, specifically the blurred and smoothen surrounding areas around the truss structure, allows these movement to be tracked by the proposed algorithm. The final displacement response is calculated using the mean response of the filtered FP and the inverse projection matrix.

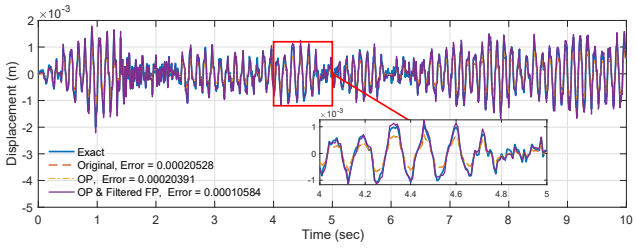
Fig. 5(c) shows the tracking results of a single point using three setups, including without any optimization and FP selection, using the optimized point (OP) selection with all feature points, and using the OP selection with filtered FP. As shown in this figure, the results indicate the improved performance by the point location optimization.




(a) Selected points, optimized points, and selected ROI

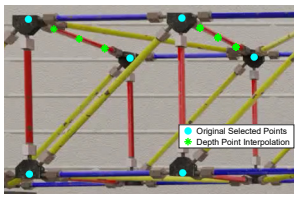



(b) Feature points

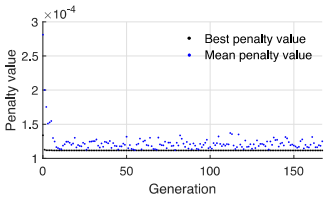



(c) Comparison among different settings

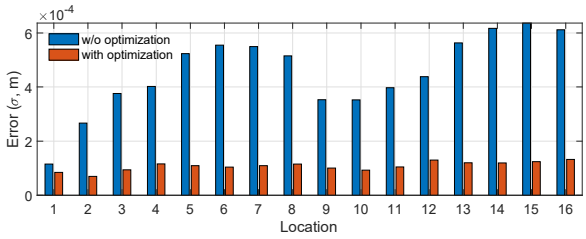
Fig. 5 Demonstration of optimized displacement measurements




(a) Illustration of additional points




(b) Optimization using the genetic algorithm




(c) Comparison of errors at nodes

Fig. 6 Discussions of depth point interpolation

Moreover, using filtered FP can result in a more accurate measurement. Note that instead of tracking the FP frame by frame, the first frame is considered as the reference to reduce error accumulation during the tracking process (Narazaki *et al*. 2020). As a result, the displacement measurement using the optimal location and filtered FPs performs better by investigating error standard deviations among the three approaches.

Table 2 Initial guess and optimized results using the intact case

| Parameters | Initial value | Optimized value |
|---|---|---|
| Longitudinal | 0.3937 m | 0.3925 m |
| Lateral | 0.3937 m | 0.3593 m |
| Vertical | 0.4 m | 0.4226 m |
| Depth interpolation | 2 points (original selected) | 21177 points |
| Error (Standard Deviation) | $4.7503 \times 10^{-4}$ m | $1.0923 \times 10^{-4}$ m |

However, errors exist in the optimized points because the exact physical coordinates may differ from the selected points, especially the transversal (out-of-plane) direction. Also, only two kinds of depths (i.e., at homogenous physical coordinates 0 and 0.3937) are considered to determine the inverse projection matrix, which can enlarge the errors in the out-of-plane displacement measurements due to the lack of depth information. Thus, the optimization by the objective function in Eq. (10) is employed to minimize the errors between the estimated and exact deformations in the undamaged case.

Fig. 6(a) illustrates the depth point interpolation of the two selected bars in the optimization. The interpolated points are added to the projection matrix with additional information in the estimation. As shown in Fig. 6(b), the units of the longitudinal, vertical, and latitude directions of a single bay and the points of interpolations by Eq. (10) become optimal at 167 generations with the error of $1.0923 \times 10^{-4}$ mm. Using the exact response of the intact structure as a reference, the optimized parameters by Eq. (10) are used in all the other cases. Moreover, Fig. 6(c) represents the error standard deviations of the intact structure (i.e., the undamaged case). The blue bars indicate the measurement errors without optimization, while the orange bars indicate the measurement errors using the proposed approach. The results show that all nodes have better accuracy after optimization, and the errors are larger when estimating responses of the upper nodes (i.e., nodes 9-16) due to out-of-plane motion. As a result, the initial and optimized parameters are listed in Table 2, where errors have decreased by around 70% after optimization. These optimized parameters are used in the other 5 cases.

### *3.3 Displacement comparison*

Table 3 lists the comparison of error calculations by standard deviations using different feature point extraction approaches and projection methods. For feature point extraction, the Harris corner (HA, Derpanis 2004) and the minimum eigenvalue (ME, Shi and Tomasi 1994) are compared. One approach is to convert pixel measurements into physical units through coordinate transformation (TR, Narazaki *et al*. 2020), where the non-parallel plane is first transformed into the parallel plane. Then, a pixel-meter factor is determined by the truss width and the correspon-

Table 3 Error magnitude standard deviation using different feature point extractions and coordinate transformations

| Case | Tutorial (HA/TR) | OPTME /TR | OPTME /INV | HA /OPTTR | OPTME /OPTTR | Proposed optimization |
|---|---|---|---|---|---|---|
| Undamaged | 4.04 | 4.53 | 4.75 | 2.64 | 2.04 | 1.09 |
| Damage 1 | 4.45 | 4.57 | 4.78 | 2.50 | 2.01 | 1.16 |
| Damage 2 | 4.59 | 4.45 | 4.58 | 2.61 | 2.09 | 1.27 |
| Damage 3 | 4.49 | 4.90 | 4.84 | 2.60 | 2.13 | 1.48 |
| Damage 4 | 3.92 | 4.93 | 4.90 | 2.51 | 2.17 | 1.53 |
| Damage 5 | 4.57 | 5.07 | 5.13 | 2.55 | 2.23 | 1.76 |

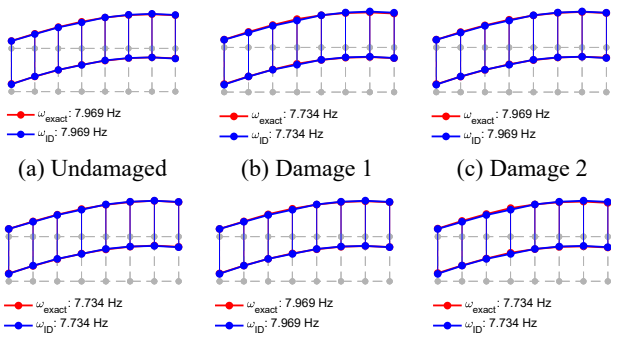
(Unit: $10^{-4}$ m; HA: Harris Corner; TR: coordinate transformation; ME: minimum eigenvalue; OPT: optimized; INV: inverse projection)

ding number of pixels multiplied to convert units. The other approach is carried out by the inverse projection matrix (INV) in Eq. (4), where the truss dimension is embedded in the projection matrix, and the physical coordinates of each pixel measurement are directly converted into physical units (i.e., meters).

Both TR and INV can be optimized using different parameters. The pixel-meter factor is optimized for the coordinate transformation by minimizing errors between the ground truth and the measurement. Note that the optimized parameters for all cases employ the dimensions of the undamaged case. For inverse projection, Eq. (10) is utilized. As seen in Table 3, the coordinate transformation has smaller errors than the inverse projection. After executing optimization, the proposed optimized inverse projection results in the smallest errors among all approaches because the depth information is considered. The overall mean error (i.e., mean value of error standard deviations) of 11 cases is $1.4256 \times 10^{-4}$ m.

### 3.4 Modal property evaluation

Modal properties can be identified by the displacements obtained from the optical measurements of the proposed method. Fig. 7 represents the comparison of the first vertical mode shape using the frequency-domain decomposition (Brincker *et al*. 2000) with the ground truth (red curve) and optical measurement (blue curve). The results indicate accurate mode shape identification with the modal assurance criterion (MAC, Pastor *et al*. 2012) close




(a) Undamaged (b) Damage 1 (c) Damage 2

(d) Damage 3 (e) Damage 4 (f) Damage 5

Fig. 7 Comparison between exact and identified modal properties in the 1st mode

to 1. Moreover, Table 4 lists the natural frequencies of the first three vertical modes using the ground truth and optical measurement, along with the mode shapes compared using MAC. As listed, the optical measurement can extract near identical modal properties with the ground truth and have natural frequency errors of less than 1% and an overall MAC larger than 0.96, indicating that the measurement errors are small enough in modal identification applications.

### 3.5 Power spectral density and correlation coefficients

One important objective in vibrational measurements is to extract the dynamic characteristics of structures and investigate structural integrity. Also, in addition to error standard deviations, calculating the correlation coefficients can provide insight into the consistency between the estimated and ground truth responses. The correlation coefficient $\rho$ is given by

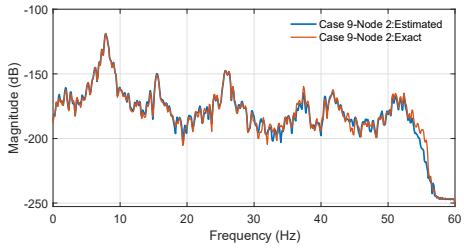
$$\rho = \frac{1}{n-1}\sum_{k=1}^{n}\left(\frac{\boldsymbol{y}_{\text{exact}}^{k} - \mu_{\text{exact}}}{\sigma_{\text{exact}}}\right)\left(\frac{\boldsymbol{y}_{\text{est}}^{k} - \mu_{\text{est}}}{\sigma_{\text{est}}}\right) \tag{12}$$

where $n$ is the number of measured points; $\sigma$ is the standard deviation; $\mu$ is the mean value. When the correlation coefficient is close to one, the estimation has a high similarity with the ground truth. On the contrary, the estimation fails to meet the ground truth when the correlation coefficient is close to zero. For an example of the case Damage 9, Fig. 8(a) represents the power spectral density of the second node of the structure. As shown in the figure, the spectral densities achieve a good agreement, implying that the image-based displacement estimation is capable of providing similar frequency contents as the ground truth.

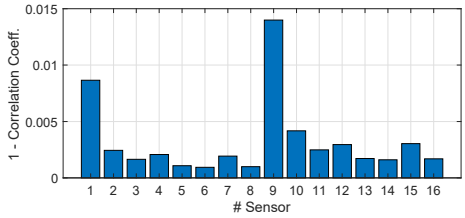
Meanwhile, Fig. 8(b) represents the correlation coefficients of all nodes. Note that the coefficients are subtracted by 1 for better observation. The results show that the estimation has all correlation coefficients over 0.985 with the ground truth. Although nodes 1 and 2 have a little bit larger variances as compared to the ground truth, the results are still comparable. Moreover, the correlation coefficients of all cases are represented in Fig. 9 using a heat map. As shown in this figure, all correlation coefficients are still larger than 0.985. Thus, the accuracy of the proposed image-based displacement estimation is stable

Table 4 Comparison between exact and identified modal properties using first 1200 data points

| Case | 1st vertical mode | | | 2nd vertical mode | | | 3rd vertical mode | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\omega_{\text{exact}}$ (Hz) | $\omega_{\text{ID}}$ (Hz) | MAC | $\omega_{\text{exact}}$ (Hz) | $\omega_{\text{ID}}$ (Hz) | MAC | $\omega_{\text{exact}}$ (Hz) | $\omega_{\text{ID}}$ (Hz) | MAC |
| Undamaged | 7.97 | 7.97 | 1.00 | 26.72 | 26.72 | 1.00 | 52.27 | 51.80 | 0.99 |
| Damage 1 | 7.73 | 7.73 | 1.00 | 26.25 | 26.25 | 0.97 | 52.27 | 51.80 | 0.99 |
| Damage 2 | 7.97 | 7.97 | 1.00 | 26.72 | 26.72 | 1.00 | 51.09 | 51.09 | 0.99 |
| Damage 3 | 7.73 | 7.73 | 1.00 | 26.72 | 26.72 | 1.00 | 51.09 | 51.09 | 1.00 |
| Damage 4 | 7.97 | 7.97 | 1.00 | 26.72 | 26.72 | 1.00 | 52.27 | 51.80 | 0.99 |
| Damage 5 | 7.73 | 7.73 | 1.00 | 26.25 | 26.19 | 0.96 | 51.09 | 51.09 | 0.99 |




(a) Power spectral density




(b) Correlation coefficients

Fig. 8 Displacement estimation performance in the case of Damage 9

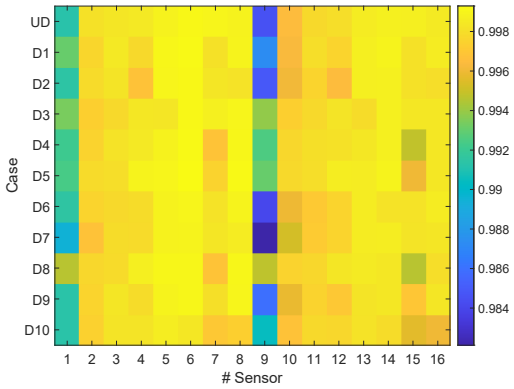



Fig. 9 Correlation coefficients of all undamaged and damaged cases

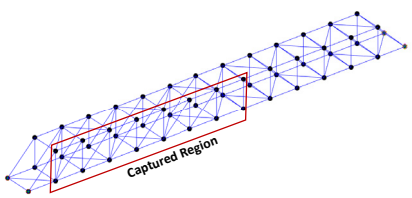
for various structural conditions and different camera angles. The slightly decreased correlation coefficients in nodes 1 and 9 are mainly due to smaller responses near the fixed end. Moreover, larger out-of-plane motions are observed in the videos.

## 4. Potential in finite element-based digital twin updating

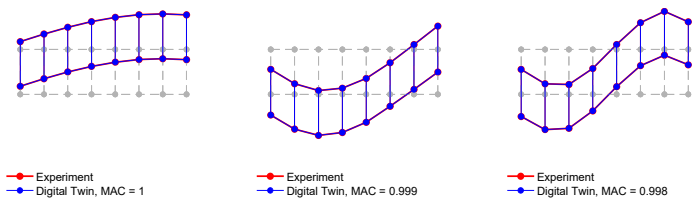
In addition to measurement evaluation, the estimated displacement can further be utilized in model updating. In this study, a digital twin is established using finite element (FE) modeling with known approximated element properties (e.g., Young's modulus, truss width, material density, etc.) provided by the IC-SHM committee. This FE model consists of 14 bays with 160 elements and 56 nodes in total. Each element is considered as a beam element (Paz and Kim 2019), and the boundary conditions of the bridge are set using the information in Fig. 3. The constructed digital twin is shown in Fig. 10(a), where the red box indicates the captured region from the camera. With the displacement measurements obtained from the videos and the geometry of the truss bridge, the numerical reference model or a digital twin can be updated. Next, the natural frequencies and mode shapes of the undamaged structure obtained in Table 4 are selected to tune the model properties of the digital twin. A restricted exhaustive search (Nievergely 2000) is applied, and the factors of Young's modulus, shear modulus, and material density are bounded between 0.95 to 1.2 with a search increment of $10^{-3}$. The objective function is given by

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmin}} f(\boldsymbol{\alpha}) = \frac{\sqrt{\frac{1}{k}\sum_{i=1}^{k}\left(\omega_{\mathrm{n},i} - \omega_{\mathrm{exp},i}\right)^2}}{\mathcal{L}_1} + \frac{\sqrt{\frac{1}{k n_{\mathrm{node}}}\sum_{i=1}^{k}\sum_{j=1}^{n_{\mathrm{node}}}\left(\varphi_{i,j} - \varphi_{i,j}^{\mathrm{exp}}\right)^2}}{\mathcal{L}_2} \tag{13}$$

where $\boldsymbol{\alpha}$ is the element properties, such as Young's modulus, shear modulus, and material density; $\omega_{\mathrm{n}}$ and $\varphi$ are the natural frequency and the mode shape using the information in $\boldsymbol{\alpha}$, while $\omega_{\mathrm{exp}}$ and $\varphi^{\mathrm{exp}}$ are the modal properties identified from the experimental measurement; $k$ is the number of identified modes, and this study only exploits 3 modes; $n_{\mathrm{node}}$ is the total number of nodes used in the identification (i.e., those nodes in the captured region of Fig. 10(a)); $\mathcal{L}_1$ and $\mathcal{L}_2$ are the weighting factors, and an equal weighting is employed for both factors in this study. After tuning the digital twin by Eq. (13), the optimized parameters are listed in Table 5. Moreover, Fig. 10(b) represents the comparison of the first three vertical mode shapes between the identification results and the tuned digital twin. As seen, each mode shape in the tuned digital twin has a MAC value higher than 0.99. This digital twin enables the representation of the dynamic characteristics of the experimental structure.




(a) Illustration of the captured region used in the model tuning




(b) Comparison of mode shapes (left to right: the 1st mode to 3rd mode)

Fig. 10 Finite element-based digital twin

Table 5 Parameters for the digital twin before and after tuning

| Parameters | Initial value | Optimized value |
|---|---|---|
| Young's modulus (E) | 18.5 GPa | 18.87 GPa |
| Shear modulus (G) | 79.3 GPa | 75.34 GPa |
| Density (R) | 8000 kg/m$^3$ | 8120 kg/m$^3$ |
| 1st mode frequency (7.97 Hz) | 7.951 Hz | 7.9707 Hz |
| 2nd mode frequency (26.7 Hz) | 26.910 Hz | 26.9764 Hz |
| 2nd mode frequency (51.8 Hz) | 53.670 Hz | 53.8019 Hz |
| 1st mode MAC | 1.00 | 1.00 |

To investigate the consistency between the updated FE-based digital twin and the experimental specimen (i.e., the PBGM), the mode shapes and natural frequencies of damage cases in Table 1 are considered. The modal properties of the digital twin are determined by adjusting the element stiffness given in Table 1. As cases 6 to 10 as examples, the results are shown in Fig. 11, where the red and green dash lines indicate the undamaged case and modal properties from ground truth displacement; the blue and cyan lines indicate the modal properties from the camera-based estimated measurement and the digital twins. The MAC values are calculated based on the ground truth

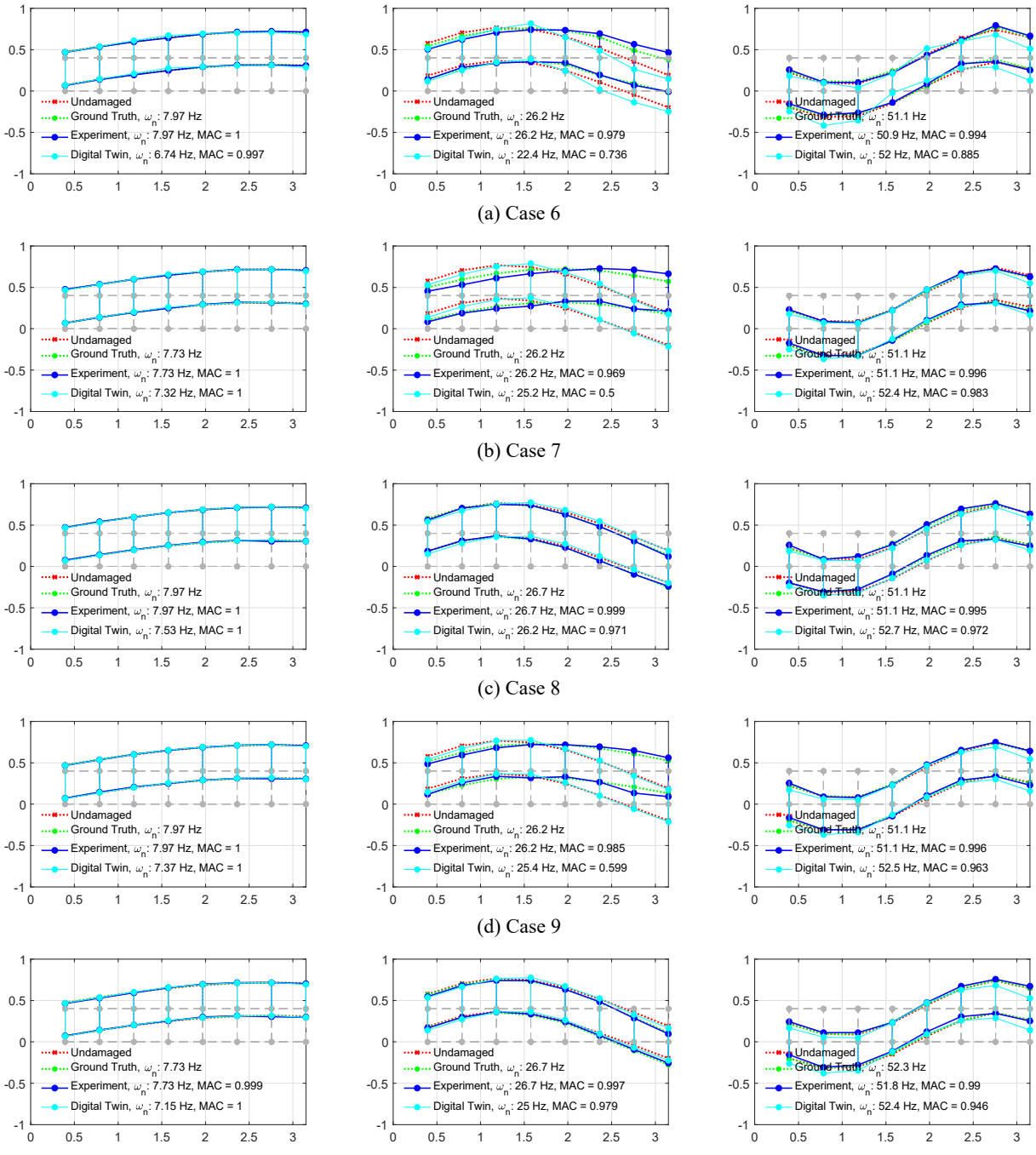



Fig. 11 Comparison of modal properties between the PBGM and the FE-based digital twin

results. This figure shows that the first and third mode shapes are consistent among all inputs (i.e., ground truth, experiment, and digital twin) with MAC close to 1. For the second mode, the mode shapes of the ground truth and the experiment meets a good agreement, indicating accurate displacement estimation based on the proposed approach. As for damage assessment, the mode shapes show negligible variation in the first and third modes and are more sensitive to the second mode. However, minor variations are observed using the digital twin, especially for the second mode. The reasons that may introduce the incorrect estimation are listed in the following.

- Out-of-plane motion: as shown in Fig. 6(c), out-of-plane motions can significantly affect the quality of displacement estimation, which varies in different locations within the image. For instance, upper nodes tend to have larger variations because the out-of-plane motion is compensated as a height difference, resulting in less accurate displacement estimations. The out-of-plane motion also yields a more significant keystone effect (Sukthankar *et al*. 1999) and then deteriorates the measurement. Although the error is highly reduced after carrying out the optimization of the geometry and depth interpolation points, some errors still exist.
- Modal properties: although the proposed method can yield entirely accurate displacement responses; however, the identification errors are still contained in mode shapes and can significantly affect the tuning process of the digital twin, as shown in Fig. 11.
- Digital twin model: in this study, only three vertical modes are used to tune the model properties. Nevertheless, three vertical modes may not be sufficient to represent the dynamic behavior of the truss bridge, resulting in mismatching modal properties, such as in the second mode in Fig. 11.

## 5. Conclusions

In this study, a two-stage optimization for image-based displacement estimation was proposed. First, the video-based vibration measurements using the KLT method were optimized by minimizing the location errors to construct an efficient inverse projection matrix. This inverse projection matrix can directly convert pixel coordinates into physical coordinates. An improvement of the KLT method was proposed using two filtering techniques to eliminate outlier feature points. Then, the physical coordinates of the structure were optimized by using the exact dimensions of the intact structure. Moreover, depth interpolation was employed in the inverse projection matrix to improve measurement accuracy. As a result, the proposed approach was shown to have reduced 70% of errors after applying the proposed optimization process.

To investigate the proposed approach, videos capturing a PBGM of a truss structure were provided by the IC-SHM committee. The videos were captured from different angles, and the structure was mimicked to have different damage conditions. The results showed that the proposed approach was capable of extracting displacement with high accuracy, and the error standard deviation over 11 cases was $1.4256 \times 10^{-4}$ m. Moreover, the power spectral density and correlation coefficients indicated a high consistency with the ground truth. In addition, modal properties were compared to further investigate the dynamic characteristics identified from the estimated displacements. The identified mode shapes were comparable with those obtained from ground truth measurements provided by the IC-SHM committee. A digital twin was constructed by tuning the natural frequencies and mode shapes extracted from the estimated displacements. The results showed that the digital twin was capable of providing similar dynamic behavior according to the resulting modal properties. As a result, the proposed improved KLT method, along with coordinate refinement, can enhance the performance of camera-based vibrational measurements. Moreover, the usage of depth interpolation points induces more information in coordinate transformation and optimization, resulting in more accurate displacement measurements. Although depth interpolation increases measurement accuracy, still the motions are deteriorated by the out-of-plane motions. Future studies should focus on estimating out-of-plane motions, eliminating keystone effects, and obtaining three-dimensional measurements to construct a more representative digital twin.

## Acknowledgments

The structural health monitoring data of this research are obtained from the organizers of the 2nd International Competition for Structural Health Monitoring (IC-SHM), 2021 (http://sstl.cee.illinois.edu/ic-shm2021/).

## References

Abdel-Aziz, Y.I. and Karara, H.M. (1971), "Direct linear transformation from comparator coordinates into object-space coordinates in close-range photogrammetry", *Proceedings of the ASP/UI Symposium on Close-Range Photogrammetry*, American Society of Photogrammetry, Falls Church, VA, USA, pp. 1-18.

Brincker, R., Zhang, L.M. and Andersen, P. (2000), "Modal identification from ambient responses using Frequency Domain Decomposition", *Proceedings of the 18th International Modal Analysis Conference*, San Antonio, TX, USA.

Carden, E.P. and Fanning, P. (2004), "Vibration based condition monitoring: a review", *Struct. Health Monit.*, **3**(4), 355-377. https://doi.org/10.1177/1475921704047500

Cha, Y.J., Chen, J.G. and Büyüköztürk, O. (2017), "Output-only computer vision based damage detection using phase-based optical flow and unscented Kalman filters", *Eng. Struct.*, **132**, 300-313. https://doi.org/10.1016/j.engstruct.2016.11.038

Chou, J.Y. and Chang, C.M. (2021), "Image motion extraction of structures using computer vision techniques: A comparative study", *Sensors*, **21**, 6248. https://doi.org/10.3390/s21186248

Derpanis, K.G. (2004), "The harris corner detector", York University, Toronto, ON, Canada.

Dong, C.Z., Celik, O., Necati Catbas, F., O'Brien, E.J. and Taylor,

S. (2020), "Structural displacement monitoring using deep learning-based full field optical flow methods", *Struct. Infrastruct. Eng.*, **16**, 51-71. https://doi.org/10.1080/15732479.2019.1650078

Du, W., Lei, D., Bai, P., Zhu, F. and Huang, Z. (2020), "Dynamic measurement of stay-cable force using digital image techniques", *Measurements*, **151**, 107211. https://doi.org/10.1016/j.measurement.2019.107211

Farrar, C.R., Doebling, S.W. and Nix, D.A. (2001), "Vibration-based structural damage identification", *Philosoph. Transact. Royal Soc. A*, **359**, 131-149. https://doi.org/10.1098/rsta.2000.0717

Feng, M., Fukuda, Y., Mizuta, M. and Ozer, E. (2015), "*Citizen Sensors* for SHM: Use of accelerometer data from smartphones", *Sensors*, **15**, 2980-2998. https://doi.org/10.3390/s150202980

Hoskere, V., Narazaki, Y. and Spencer Jr, B.F. (2019), "Learning to detect important physics-based graphics models", *Proceedings of the 9th International Conference on Structural Health Monitoring of Intelligent Infrastructure* (*ISHMII*), St. Louis, MO, USA.

Li, J., Hao, H., Fan, K. and Brownjon, J. (2014), "Development and application of a relative displacement sensor for structural health monitoring of composite bridges", *Struct. Control Health Monit.*, **22**, 726-742. https://doi.org/10.1002/stc.1714

Liu, P.L. (1995), "Identification and damage detection of trusses using modal data", *J. Struct. Eng.*, **121**(4), 599-608. https://doi.org/10.1061/(ASCE)0733-9445(1995)121:4(599)

Luu, L., Wang, Z., Vo, M., Hoang, T. and Ma, J. (2011), "Accuracy enhancement of digital image correlation with B-spline interpolation", *Opt. Lett.*, **36**, 3070-3072. https://doi.org/10.1364/OL.36.003070

MATLAB and Computer Vision Toolbox Release (2020), The MathWorks, Inc., Natick, Massachusetts, United States.

Narazaki, Y., Gomez, F., Hoskere, V., Smith, M.D. and Spencer Jr, B.F. (2020), "Efficient development of vision-based dense three-dimensional displacement measurement algorithms using physics-based graphics models", *Struct. Health Monit.*, 147592172093952. https://doi.org/10.1177/1475921720939522

Nievergely, J. (2000), "Exhaustive search, combinatorial optimization and enumeration: exploring the potential of raw computing power", In: *SOFSEM2000*: *Theory and Practice of Informatics* (Proceedings of the 27th Conference on Current Trends in Theory and Practice of Informatics, Milovy, Czech Republic, November-December), pp. 18-35.

Pastor, M., Binda, M. and Harčarik, T. (2012), "Modal assurance criterion", *Procedia Eng.*, **48**, 543-548. https://doi.org/10.1016/j.proeng.2012.09.551

Patsias, S. and Staszewskly, W.J. (2002), "Damage detection using optical measurements and wavelets", *Struct. Health Monit.*, **1**(1), 5-22. https://doi.org/10.1177/147592170200100102

Paz, M. and Kim, Y.H (2019), *Structural Dynamics*, Springer. ISBN 978-3-319-94743-3

Ritto, T.G. and Rochinha, F.A. (2021), "Digital twin, physics-based model, and machine learning applied to damage detection in structures", *Mech. Syst. Signal Process.*, **155**, 107614. https://doi.org/10.1016/j.ymssp.2021.107614

Shi, J. and Tomasi, C. (1994), "Good features to track", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, June, pp. 593-600. https://doi.org/10.1109/CVPR.1994.323794

Strat, T.M. (1987), "Recovering the camera parameters from a transformation matrix", In: *Readings in Computer Vision*, Morgan Kaufmann, pp. 93-100. https://doi.org/10.1016/B978-0-08-051581-6.50017-9

Sukthankar, R., Stockton, R. and Mullin, M. (1999), "Automatic keystone correction", Provisional U.S. Patent Filing, 614.

Sutton, M.A., Yan, J.H., Tiwari, V., Schreier, H.W. and Orteu, J.J. (2008), "The effect of out-of-plane motion on 2D and 3D digital image correlation measurements", *Opt. Lasers Eng.*, **46**, 746-757. https://doi.org/10.1016/j.optlaseng.2008.05.005r

Tomasi, C. and Kanade, T. (1991), "Detection and Tracking of Point Features", Technical Report: CMU-CS-91132, Carnegie Mellon University, Pittsburgh, PA, USA.

Wall, M.E., Rechtsteiner, A. and Rocha, L.M. (2003), "Singular value decomposition and principal component analysis", In: *A Practical Approach to Microarray Data Analysis*, Kluwer: Norwell, MA, USA, pp. 91-109. https://doi.org/10.1007/0-306-47815-3_5

Whitley, D. (1994), "A genetic algorithm tutorial", *Stat. and Comput.*, **4**, 65-85. https://doi.org/10.1007/BF00175354

Xu, K. and Mita, A. (2021), "Absolute displacement-based formulation for peak inter-story drift identification of shear structures using only one accelerometer", *Sensors*, **21**, 3629. https://doi.org/10.3390/s21113629

Ye, C., Butler, L., Calka, B., Iangurazov, M., Lu, Q., Gregory, A., Girolami, M. and Middleton, C. (2019), "A digital twin of bridges for structural heath monitoring", *Proceedings of the 12th International Workshop on Structural Health Monitoring*.

Yoon, H., Elanwar, H., Choi, H., Golparvar-Fard, M. and Spencer Jr, B.F. (2016), "Target-free approach for vision-based structural system identification using consumer-grade cameras", *Struct. Control Health Monit.*, **23**, 1405-1416. https://doi.org/10.1002/stc.1850

Zhu, J., Lu, Z. and Zhang, C. (2020), "A marker-free method for structural dynamic displacement measurement based on optical flow", *Struct. and Infrastruct. Eng.*, **18**(1), 84-96. https://doi.org/10.1080/15732479.2020.1835999