Smart Structures and Systems, *Vol. 31*, *No. 4* (*2023*) *393-407*
https://doi.org/10.12989/sss.2023.31.4.393

A novel computer vision-based vibration measurement and coarse-to-fine damage assessment method for truss bridges

Wen-Qiang Liu [1,2‡], En-Ze Rui [1,2‡], Lei Yuan [1,2], Si-Yi Chen [1,2], You-Liang Zheng [1,2] and Yi-Qing Ni*[1,2]

[1] *Department of Civil and Environmental Engineering*, *The Hong Kong Polytechnic University*, *Hung Hom*, *Kowloon*, *Hong Kong S.A.R.*
[2] *National Rail Transit Electrification and Automation Engineering Technology Research Center* (*Hong Kong Branch*), *Hung Hom*, *Kowloon*, *Hong Kong S.A.R.*

(*Received August 30*, *2022*, *Revised November 22*, *2022*, *Accepted February 2*, *2023*)

Abstract. To assess structural condition in a non-destructive manner, computer vision-based structural health monitoring (SHM) has become a focus. Compared to traditional contact-type sensors, the advantages of computer vision-based measurement systems include lower installation costs and broader measurement areas. In this study, we propose a novel computer vision-based vibration measurement and coarse-to-fine damage assessment method for truss bridges. First, a deep learning model FairMOT is introduced to track the regions of interest (ROIs) that include joints to enhance the automation performance compared with traditional target tracking algorithms. To calculate the displacement of the tracked ROIs accurately, a normalized cross-correlation method is adopted to fine-tune the offset, while the Harris corner matching is utilized to correct the vibration displacement errors caused by the non-parallel between the truss plane and the image plane. Then, based on the advantages of the stochastic damage locating vector (SDLV) and Bayesian inference-based stochastic model updating (BI-SMU), they are combined to achieve the coarse-to-fine localization of the truss bridge's damaged elements. Finally, the severity quantification of the damaged components is performed by the BI-SMU. The experiment results show that the proposed method can accurately recognize the vibration displacement and evaluate the structural damage.

Keywords: computer vision; damage assessment; deep learning; model updating; structural health monitoring; vibration measurement

1. Introduction

Damage localization and severity quantification of civil infrastructure are critical issues in structural health monitoring (SHM) (Ni *et al*. 2015, Sun *et al*. 2020, Bernagozzi *et al*. 2021, Lei *et al*. 2021). Researchers have proposed sophisticated and dependable methods for structural damage detection (Bakhtiari-Nejad *et al*. 2005, He *et al*. 2022, Mousavi *et al*. 2022, Wang *et al*. 2022a). For example, Bernal (2002) proposed a damage locating vector (DLV) method that extracts a force vector from the null space of the flexibility matrix difference for damage localization. Moreover, DLV-based damage detection methods have indicated their feasibility through various paradigms (Gao *et al*. 2007, Feng and Feng 2017, Frans *et al*. 2017). Due to expensive input excitations that are difficult to implement in practical situations, structural damage detection methods using output-only measurements have been proposed in recent years (Nagarajaiah and Basu 2009, Feng and Feng 2016, Bernagozzi *et al*. 2018). For example, a stochastic DLV (SDLV) method was proposed to make up for the deficiency of the DLV method (Bernal 2006). A surrogate flexibility matrix was established based on the output-only measurements under ambient excitations for damage detection (An *et al*. 2014). However, these methods are difficult to account for measurement noise and quantify uncertainty in identification. Meanwhile, Bayesian inference in the probabilistic context has also been widely adopted for localizing structural damage and assessing structural health (Figueiredo *et al*. 2014, Kim *et al*. 2015, Uzun *et al*. 2019, Eltouny and Liang 2021, Wang *et al*. 2021a). For instance, a sparse Bayesian model updating framework was proposed by Wang *et al*. (2021a) for structural damage detection, and its feasibility was verified through laboratory test. However, the efficiency of damage detection would be seriously affected when parameterized structural models are required for Bayesian methods to predict structural modal characteristics. In view of this, Bayesian probabilistic frameworks free of structural models have recently been proposed for structural damage detection (Wang *et al*. 2018, 2021b, 2022c, Ni *et al*. 2020, Ni and Zhang 2021).

Computer vision (CV) techniques have been increasingly adopted in SHM to capture the vibration motion of civil structures due to their high cost-effectiveness and ease of implementation (Ye *et al*. 2016, Khuc and Catbas 2017, Xu *et al*. 2018, Ngeljaratan and Moustafa 2020, Wang *et al*. 2022b). Furthermore, unlike traditional contact-type sensors, CV-based measurements add no mass to the target structure, which leads to no change in the dynamic properties of the target structure (Ni

*Corresponding author, Chair Professor,
E-mail: ceyqni@polyu.edu.hk
‡ The authors contribute equally to the article.

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

et al. 2019). Therefore, CV-based measurements have the potential to replace the existing contact-type sensors in SHM (Ye *et al*. 2013, Xu *et al*. 2018, Liang 2019, Erdogan and Ada 2020). However, the relevant research is still at an early stage and faces some challenges (Spencer *et al*. 2019, Dong and Catbas 2021).

For example, most of the CV-based methods adopt the traditional optical flow tracking algorithm based on target matching. There are two issues in practical applications of this method: (i) The tracked target needs to be manually selected in the first frame of a video, which largely restricts the automation of implementation, and the manual selection is subjective and operator-dependent, which directly affects the tracking accuracy; (ii) The background information of the target is not considered, and the target tracking task might fail when the target is partially occluded or light changes and blurred motion happen. As a result, practical applications of SHM using CV-based displacement measurements are still limited (Cha *et al*. 2017, Feng and Feng 2017, Xu 2020).

Given the above, this study proposes a CV-based three-step structural damage detection method for a truss bridge, including displacement measurement, coarse-to-fine damage localization, and damage severity quantification, as shown in Fig. 1. The main contribution of this paper is threefold. First, the deep learning-based FairMOT target tracking model (Zhang *et al*. 2021) is introduced to replace the traditional target tracking method used in SHM to achieve the displacement measurement of structural vibration. Second, by leveraging the advantages of the SDLV (Bernal 2006) and the Bayesian inference-based stochastic model updating (BI-SMU) (Wan and Ren 2016), the two methods are combined to achieve effective localization and accurate assessment of structural damage. Lastly, the BI-SMU is extended for damage severity quantification.

The rest of this paper is organized as follows. The proposed method is described in detail in Section II. In Section III, verification experiments are conducted. Section IV summarizes and concludes the research work.

2. Methodology

2.1 Target tracking for displacement measurement

To measure the structural vibration of the truss bridge of concern accurately, the ROIs of the structure should be properly detected and tracked first. Then, the displacement of the tracked ROIs relative to the initial frame will be calculated. In recognizing the aforementioned shortcomings of the traditional target tracking algorithm in application to SHM, a deep learning (DL)-based target tracking method is adopted in this study to achieve better performance.

There are mainly two kinds of DL-based multi-object tracking (MOT) algorithms at present: the first kind is the two-step MOT (Fang *et al*. 2018, Mahmoudi *et al*. 2019), which successively uses a detection model to extract the bounding box position of the target, and then adopts an association model to extract the re-identification (Re-ID) features in each bounding box and links the Re-ID features to one of the existing tracking bounding boxes based on specific metrics defined by those features. The advantage of this kind of methods is that it can build the optimal model for each task; but because the features are not shared between tasks, it cannot meet the performance requirements of real-time inferencing. The second kind is the single-step MOT (Voigtlaender *et al*. 2019, Wang *et al*. 2020), which performs feature extraction at the same time as object detection. The core idea is to complete object detection and Re-ID embedding in a single network so as to share most of the computation to reduce inference time. However, Zhang *et al*. (2021) found that such methods suffer from the incorrect target ID association. Specifically, this kind of methods uses anchor-based object detection. The Re-ID feature of the object is extracted in the anchor area, and the anchor and object areas might misalign, which will lead to serious ambiguity during network training. To tackle this problem, Zhang *et al*. (2021) proposed the FairMOT network. First, the MOT problem is treated as pixel-level key point (target center) estimation and ID classification problem. By adopting the anchor-free object detection method, the target center is estimated on the high-resolution feature map, eliminating the ambiguity problem caused by the anchor-based object detection. Then, deep layer aggregation (DLA) is applied to enhance the backbone feature extraction network ResNet-34 (He *et al*. 2016), so as to integrate the information of multiple feature layers of different scales and improve the ability of the network to deal with the change of object scale. The model achieves state-of-the-art performance on publicly available object tracking datasets. In view of the advantages of this model, it is selected as the target tracking algorithm in lieu of the traditional target tracking algorithm.

2.1.1 ROIs tracking

As shown in Fig. 2, FairMOT consists of three modules: encoder-decoder, detection, and Re-ID.

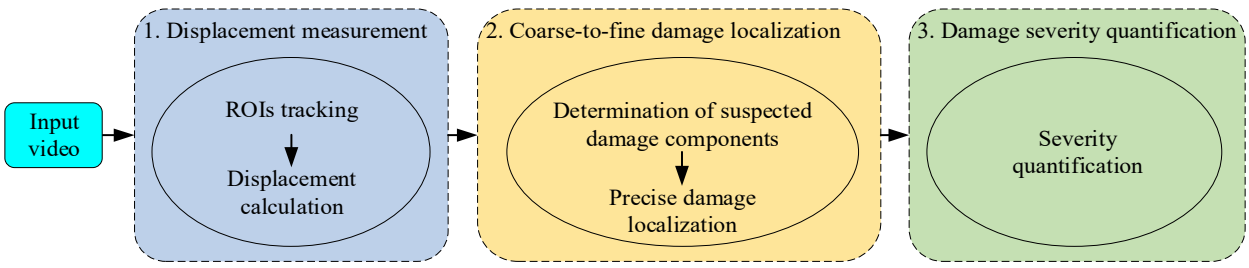


Fig. 1 Flowchart of the proposed method

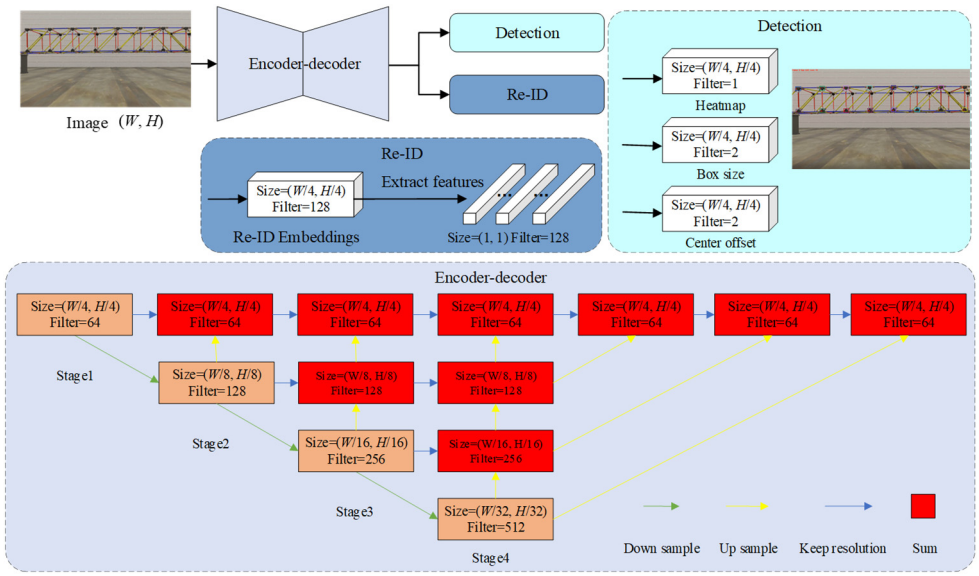


Fig. 2 Schematic diagram of the network structure of the FairMOT tracking model

1) The function of the encoder-decoder module is to extract the feature of each frame of the acquired video. Here, the variant DLA-34 (Zhou *et al*. 2020) of the ResNet (He *et al*. 2016) is adopted as the backbone network of the module, and the network structure is shown in Fig. 2. The DLA-34 model introduces DLA to fuse multi-layer features on the basis of ResNet-34. By adding more skip connections between low-level and high-level features and using the deformable convolution network (DCN) (Dai *et al*. 2017) to replace all Upsampling modules, the receptive field is dynamically adjusted according to the scale and pose of the object. The above treatments can well improve the anchor alignment problem. All relevant dimension information of the input, output, and the intermediate feature map are listed in Fig. 2.
2) The function of the detection module is to detect the target, which comprises three parallel head networks, including a heatmap head, a box size head, and a center offset head.
The heatmap head network is built to estimate the position of the target center. The target center of each ground truth box $b^i = (x_1^i, y_1^i, x_2^i, y_2^i)$ in the current frame is calculated as $c_x^i = \frac{x_1^i + x_2^i}{2}$ and $c_y^i = \frac{y_1^i + y_2^i}{2}$. The heatmap response at the location (x, y) is represented as $M_{xy} = \sum_{i=1}^{N} exp^{-\frac{(x-\tilde{c}_x^i)^2 + (y-\tilde{c}_y^i)^2}{2\sigma_C^2}}$ where N is the number of targets in the frame and σ_C denotes the standard deviation. $(\tilde{c}_x, \tilde{c}_y) = \left(\left\lfloor \frac{c_x^i}{4} \right\rfloor, \left\lfloor \frac{c_y^i}{4} \right\rfloor\right)$ represents the location on the feature map, $\lfloor\ \rfloor$ is the floor function. The loss function is expressed as pixel-wise logistic regression with a focal loss

$$L_{\text{heat}} = \frac{1}{N} \sum_{xy} \begin{cases} (1 - \hat{M}_{xy})^{\alpha} log(\hat{M}_{xy}) & M_{xy} = 1; \\ (1 - M_{xy})^{\beta} (\hat{M}_{xy})^{\alpha} log(1 - \hat{M}_{xy}) & \text{otherwise;} \end{cases} \quad (1)$$

where $\hat{M}$ is the estimated heatmap, α and β are the pre-determined parameters in a focal loss.
The box size head and center offset head are constructed to localize the target's position more accurately. The offset and box size are defined as $o^i = \left(\frac{c_x^i}{4}, \frac{c_y^i}{4}\right) - \left(\left\lfloor \frac{c_x^i}{4} \right\rfloor, \left\lfloor \frac{c_y^i}{4} \right\rfloor\right)$; $s^i = (x_2^i - x_1^i, y_2^i - y_1^i)$, and the box loss is determined as follows

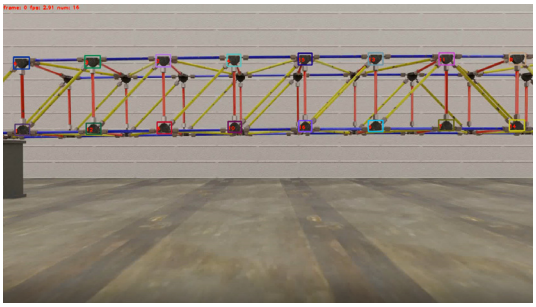
$$L_{\text{box}} = \sum_{i=1}^{N} \left\| o^i - \hat{o}^i \right\|_1 + \lambda_s \left\| s^i - \hat{s}^i \right\|_1 \quad (2)$$

where λ_s is a weighting parameter and set as 0.1, $\hat{o}$ and $\hat{s}$ are the estimated values.
3) The Re-ID module is created to classify the target through learning Re-ID features. The Re-ID feature of the center location $(\tilde{c}_x, \tilde{c}_y)$ of each object on the heatmap is extracted and then put into a fully connected layer and a softmax layer to get a class distribution vector $P = \{p(k), k \in [1, K]\}$. The Re-ID loss is defined as follows

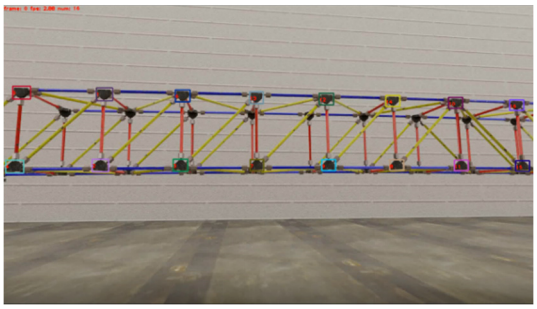
$$L_{\text{identity}} = \sum_{i=1}^{N} \sum_{k=1}^{K} L^i(k)\, log(p(k)) \quad (3)$$

where $L^i(k)$ is the one-hot representation of the ground truth class label, K is the number of all the identities in the training data.
4) The detection and Re-ID losses are then combined to train the network (refer to Zhang *et al*. (2021) on the details). An example of the joints tracked by the FairMOT is shown in Fig. 3.

$$L_{\text{detection}} = L_{\text{heat}} + L_{\text{box}} \quad (4)$$

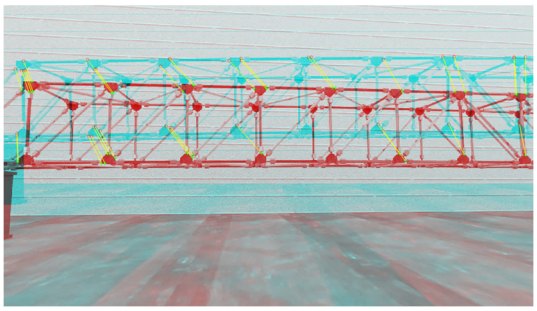


(a) Parallel case

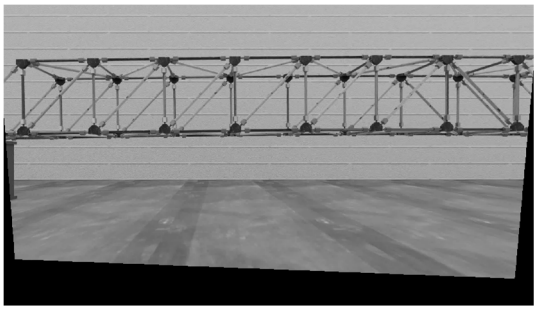


(b) Non-parallel case

Fig. 3 An example of the joints tracked by the FairMOT



(a) Matched points



(b) Recovered image

Fig. 4 An example of recovering the non-parallel image

$$L_{\text{total}} = \frac{1}{2}\left(\frac{1}{e^{w_1}}L_{\text{detection}} + \frac{1}{e^{w_2}}L_{\text{identity}} + w_1 + w_2\right) \quad (5)$$

where w_1 and w_2 are learnable parameters that balance the two tasks.

2.1.2 Displacement calculation

From Fig. 3, it can be seen that the calculated center offsets cannot be as the displacement of targets because the objective of the FairMOT model is to detect and track the whole of the target rather than its center position. Therefore, a high-precision displacement measurement method based on template matching is proposed to detect the displacement. The implementation procedures are as follows.

1) Firstly, all the targets detected in each video's first frame (called as *tFrame*) are used as the initial targets. Starting from the second frame, all the targets tracked in the current frame (called as *cFrame*) are the dynamic targets; and by matching them with their corresponding initial targets, their displacements are calculated. The normalized cross-correlation is used to assess the degree of match or similarity. Before that, to enhance the accuracy, the detected and tracked regions are resized to a larger scale. The formula for match and similarity check is as follows

$$sim = \frac{\sum_{xy}[f(x,y) - \bar{f}_{u,v}][t(x-u,y-v) - \bar{t}]}{\left\{\sum_{x,y}[f(x,y) - \bar{f}_{u,v}]^2 \sum_{x,y}[t(x-u,y-v) - \bar{t}]^2\right\}^{0.5}} \quad (6)$$

where f is the image region of the tracked target, $\bar{t}$ is the mean of the template, $\bar{f}_{u,v}$ is the mean of $f(x,y)$ in the region under the template.

In addition, to ensure that the tracked region can contain the initial targets, the scope of the tracked box $(x_1^i - a, y_1^i - a, x_2^i + b, y_2^i + b)$ is expanded. Here, a and b are set as 20 and 80 pixels, respectively.

2) Secondly, two scale adjustments under the above steps are performed to balance the detection speed and accuracy. The first resize coefficient is set to 16, and the second is expanded four times based on the former.
3) Thirdly, correction is required because it is uncertain whether the camera plane of each video is parallel to the truss plane. To this end, a fast and efficient pairing correction method is proposed based on corner matching. And the calculation process and results are shown below. First, a frame without translation transformation is treated as the standard frame *sFrame*. Then, the feature points in the video's first frame *tFrame* and in the standard frame are detected and matched using the Harris corner feature extraction method. After that, the transformation matrix T is obtained using the spatial relationship of the paired points. Lastly, the detection result of each video is multiplied by the transformation matrix to get accurate displacement information.

$$(s_{\text{Point}_i}, d_{\text{Point}_i}) = Match\begin{pmatrix} Harris(sFrame, ROI_i^s), \\ Harris(tFrame, ROI_i^t) \end{pmatrix} \quad (7)$$

$$T = Transform(s_{\text{Point}_i}, d_{\text{Point}_i}) \tag{8}$$

$$(ROI_i^t{}_{\text{trans}}, ROI_i^c{}_{\text{trans}}) = (ROI_i^t, ROI_i^c) \cdot T \tag{9}$$

4) Finally, according to the scale factor $px2m$ of pixels and physical coordinates, the actual offset value is calculated as follows

$$disp2D = (center(ROI_i^c{}_{\text{trans}}) - center(ROI_i^t{}_{\text{trans}})) \cdot px2m \tag{10}$$

2.2 SDLV for coarse damage localization

A coarse-to-fine damage localization method is proposed. The process of coarse damage localization is as follows: First, the stochastic subspace identification (SSI) is used to extract modal frequency and mode shape information, then calculate the system and observation matrices from the state-space model. Then, the SDLV algorithm is adopted to deal with the system and observation matrices and roughly locate the damage location. The implementation procedures are described below.

2.2.1 SSI

The SSI method is adopted for modal identification using CV-based displacement measurements. It consists of four steps: building the Hankel matrix, calculating the Toeplitz matrix, conducting the singular value decomposition (SVD) on the Toeplitz matrix, and forming the system and observation matrices. The Hankel matrix H is firstly established based on the discrete-time output vector y_k. The Hankel matrix can be expressed as follows

$$H = \frac{1}{\sqrt{j}}\begin{bmatrix} y_0 & y_1 & \cdots & y_{j-1} \\ y_1 & y_2 & \cdots & y_j \\ \vdots & \vdots & \ddots & \vdots \\ y_{i-1} & y_i & \cdots & y_{i+j-2} \\ \hline y_i & y_{i+1} & \cdots & y_{i+j-1} \\ y_{i+1} & y_{i+2} & \cdots & y_{i+j} \\ \vdots & \vdots & \ddots & \vdots \\ y_{2i-1} & y_{2i} & \cdots & y_{2i+j-2} \end{bmatrix} = \begin{bmatrix} Y_p \\ \hline Y_f \end{bmatrix} \tag{11}$$

where Y_p is the Hankel past matrix and Y_f is the Hankel future matrix. The Toeplitz matrix $T_{1/i}$ can be calculated as follows

$$T_{1/i} = Y_f Y_p^T \tag{12}$$

SVD is then performed on the Toeplitz matrix $T_{1/i}$, which can be expressed as

$$T_{1/i} = \begin{pmatrix} S_{T_1} & S_{T_2} \end{pmatrix}\begin{pmatrix} V_{T_1} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} D_{T_1}^T \\ D_{T_2}^T \end{pmatrix} = S_{T_1} V_{T_1} D_{T_1}^T \tag{13}$$

With the SVD results, the following two matrices (O_i and Γ_i) are defined

$$O_i = S_{T_1} V_{T_1}^{1/2} \tag{14}$$

$$\Gamma_i = V_{T_1}^{1/2} D_{T_1}^T \tag{15}$$

After obtaining O_i and Γ_i, the continuous system matrix A and the continuous observation matrix C can be calculated as follows

$$A = O_i^{\dagger} T_{2/i+1} \Gamma_i^{\dagger} \tag{16}$$

$$C = O_i(1{:}\,l, :) \tag{17}$$

where l is the number of observation points, and $(\square)^{\dagger}$ stands for the pseudo-inverse of a matrix. $T_{2/i+1}$ is the shifted block Toeplitz matrix. After conducting eigenvalue decomposition on the matrix A, we can acquire the discrete system matrix A_C and observation matrix C_C.

2.2.2 SDLV

The discrete system and observation matrices obtained from the previous step are then used for coarse damage localization. The SDLV approach is introduced to acquire a group of load vectors causing zero-stress fields over the damaged regions on the truss bridge. The group of the load vectors is generated from the effective null space of the change in the matrix ΔQ^T. Firstly, Q is given by

$$Q = -C_C A_C^{-1} H_P^{\dagger} L \tag{18}$$

where $C_C \in R^{m \times N}$ is the observation matrix and $A_C \in R^{N \times N}$ is the system matrix with m being the number of inputs. $H_P^{\dagger}$ is the pseudoinverse of H_P. $H_P \in R^{2m \times N}$ and $L \in R^{2m \times m}$ are further given by

$$H_P = \begin{pmatrix} C_C A_C \\ C_C \end{pmatrix} \tag{19}$$

and

$$L = \begin{pmatrix} I \\ \hline 0 \end{pmatrix} \tag{20}$$

Let ΔQ^T represent the change in the matrix Q^T due to structural damage, which is given by

$$\Delta Q^T = Q_u^T - Q_d^T \tag{21}$$

where the subscripts u and d denote the undamaged status and the damaged status, respectively. To obtain the damage locating vectors, SVD is conducted on the matrix ΔQ^T, which results in

$$\Delta Q^T = USV^T = \begin{pmatrix} U_1 & U_0 \end{pmatrix}\begin{pmatrix} S_1 & 0 \\ 0 & S_0 \approx 0 \end{pmatrix}\begin{pmatrix} V_1 \\ V_0 \end{pmatrix} \tag{22}$$

where U and V are unitary matrices. With Eq. (22), we can obtain

$$\begin{pmatrix} \Delta Q^T V_1 & \Delta Q^T V_0 \end{pmatrix} \approx \begin{pmatrix} U_1 & U_0 \end{pmatrix}\begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} U_1 S_1 & 0 \end{pmatrix} \tag{23}$$

From the above, it can be found that

$$\Delta Q^T V_0 \approx 0 \tag{24}$$

When the load vector V_0 is applied to the undamaged and damaged structure simultaneously, the structural displacement induced by V_0 remains the same. It also implies that a zero-stress field should appear on the damaged elements when a load vector V_0 is applied to the undamaged structure.

2.3 BI-SMU for fine damage localization and severity quantification

After the suspected damage locations are determined based on the coarse damage localization results, the precise damage localization and severity quantification strategy is adopted to identify the damage location better and quantify the damage severity. First, a Gaussian Process-based metamodel (GPM) is established, which is used to map the relationship between the material properties of the suspected damage elements and the structural frequency reductions caused by the structure damage. Next, the BI-SMU method (Wan and Ren 2016) is incorporated for fine damage localization and severity quantification. The implementation procedures are briefed below.

2.3.1 GPM

Gaussian Process regression is utilized in this study for metamodel establishment. The relationship between the material properties (elastic modulus) of the suspected damage elements and the structural frequency changes caused by structural damage is mapped based on the GPM. Consider a zero-mean Gaussian Process with kernel function $k(x, x'; \sigma)$, which can be described as follows

$$u(x) \sim GP(0, k(x, x'; \sigma)) \tag{25}$$

where σ denotes hyperparameters in the kernel function. In this work, the kernel function takes the squared exponential form. The Limited-Memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) optimizer is used to optimize the hyperparameters in the squared exponential kernel by minimizing the negative log marginal likelihood (NLML).

2.3.2 BI-SMU

The posterior probability density function of the updated parameters θ can be obtained by Bayes' theorem

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int_\theta p(y|\theta)p(\theta)d\theta} = cp(y|\theta)p(\theta) \tag{26}$$

where c is a constant which is not related to θ; $p(\theta)$ denotes the prior probability density function of θ; $p(y|\theta)$ denotes the likelihood function of observations; $p(\theta|y)$ denotes the posterior probability density function of θ. Assuming that k tests are conducted and each test is independent, the likelihood function, which indicates the possibility that the test value is observed, can be expressed as follows

$$p(y|\theta) = \frac{1}{(2\pi)^{2k}\Sigma^{\frac{1}{2}}} exp(-\frac{1}{2}(f_p(\theta) - \hat{y})^T \Sigma^{-1}(f_p(\theta) - \hat{y})) \tag{27}$$

where $f_p(\theta)$ denotes the metamodel prediction; $\hat{y}$ and Σ denote the mean and covariance matrix of the measurement data. If the prior distributions of the updated parameters take the form of uniform distribution, the expression of the posterior probability density function reduces to

$$p(\theta|y) = c' \times exp(-\frac{1}{2}(f_p(\theta) - \hat{y})^T \Sigma^{-1}(f_p(\theta) - \hat{y})) \tag{28}$$

where c' is a constant which is not related to θ. Markov Chain Monte Carlo (MCMC) method is adopted for numerical computation of the posterior probability density functions of the updated parameters. The Metropolis-Hasting (MH) algorithm is utilized to realize the MCMC method.

2.4 General description of the entire framework

As shown in Fig. 5, the entire framework of the proposed method is described as follows:

1) First stage: the ROIs of the structure are tracked and located by the FairMOT, and then vibration displacement is calculated by the template matching method.
2) Second stage: the structure's natural frequencies, system matrix, and observation matrix are extracted by the SSI method. The system and observation matrices of the undamaged and damaged cases are sent to the SDLV to obtain the damage locating force vectors from the null space of the difference of the surrogate matrix. Then, the vectors are applied to the finite element model to calculate the accumulative stress index and determine the suspected damaged locations. After that, the relationship between the

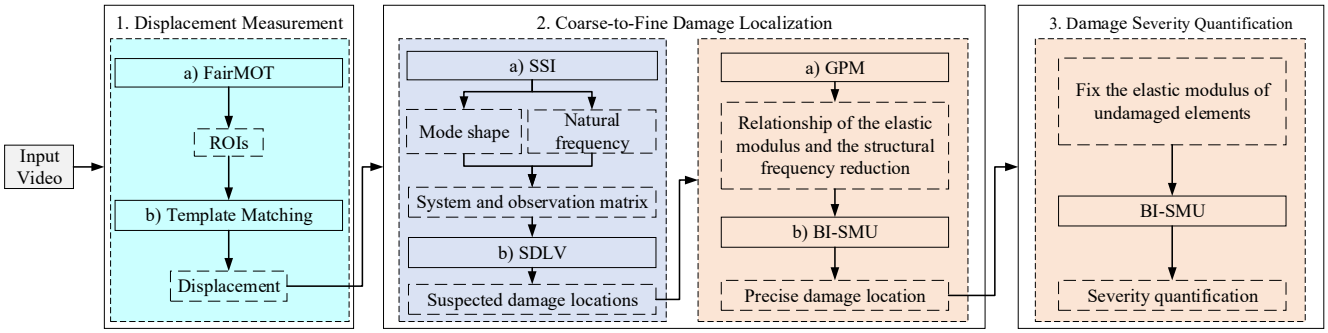


Fig. 5 Schematic diagram of the proposed framework

elastic moduli of the suspected damage elements and the modal frequency reductions is established based on a GPM. Then, the likelihood function is constructed based on metamodel prediction and measured modal frequencies from the previous SSI results. Next, the posterior probabilities of the updated parameters are calculated by the maximum a posteriori (MAP) estimation, and the updated elastic moduli are regarded as the predicted remaining stiffnesses of the structural elements. According to the value order of the detected results, the precise damage location is obtained.

3) Third stage: according to the previously established relationships between the elastic moduli of the damaged elements and the modal frequency reductions as well as the precisely identified damage locations, the damage severities are evaluated using the BI-SMU by quantifying the elastic moduli of damaged elements.

3. Experimental investigation

3.1 Experimental environment and setup

This study uses a physics-based graphic model, namely PBGM, for validation of the proposed method. A finite element model of a truss bridge is utilized to establish the photo-realistic synthetic environment. We focus the damage detection task on the left half of the truss bridge, which includes 29 truss elements and 16 nodes, as shown in Fig. 6. For more information about this PBGM, please refer to the references (Narazaki *et al*. 2021, Gomez *et al*. 2022). All experiments were developed and conducted on a platform with MATLAB, Pytorch v1.9.0. The IC-SHM 2021 dataset is in Table 1, the first five damage cases (Dam1~Dam5) are used as the validation dataset to evaluate the performance of the proposed method, and the other five damage cases (Dam6~Dam10) are used as the testing dataset to verify the effectiveness and generalization of the proposed method.

3.2 Displacement measurement

3.2.1 Performance validation

In order to train the target tracking model and verify the effectiveness of the adopted tracking model, we first randomly select 21,600 images from the videos Dam1, Dam3 and Dam4 for model training, and select 14,400 images from the videos Dam2 and Dam5 for testing. Three target tracking indices defined as Eqs. (29-31) are chosen to evaluate the accuracy of the model. The training and evaluation results are shown in Fig. 7 and Table 2. It is shown that the selected tracking method can accurately track all the targets.

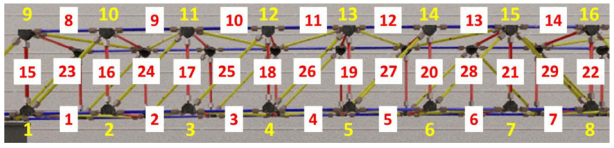


Fig. 6 Element and node numbering of the truss bridge

Table 1 Damage cases

Case	Damaged Element No.	RS*
Dam1	3	47.3%
Dam2	18	47.3%
Dam3	6	42.5%
Dam4	16	27.5%
Dam5	11	40.0%
Dam6	20	12.0%
Dam7	5	33.0%
Dam8	9	52.0%
Dam9	11	43.5%
Dam10	12	38%

*RS stands for remaining stiffness (measured in percent)

Table 2 Evaluation of the target tracking detection results

Dataset	IDF_1	IDP	IDR
Dam2	100%	100%	100%
Dam5	100%	100%	100%
Overall	100%	100%	100%

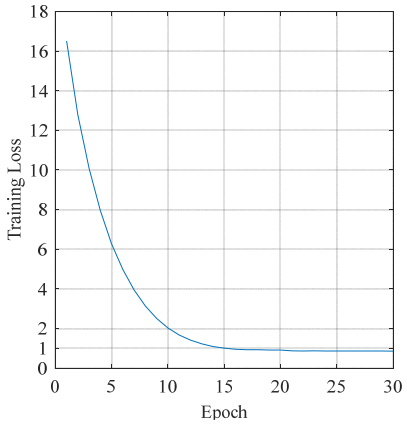


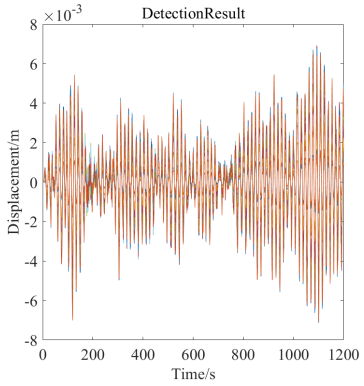
Fig. 7 The training loss curve

$$IDP = \frac{IDTP}{IDTP + IDFP} \tag{29}$$

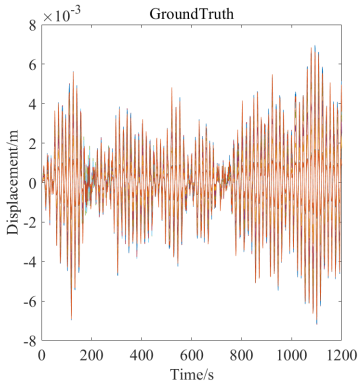
$$IDR = \frac{IDTP}{IDTP + IDFN} \tag{30}$$

$$IDF_1 = \frac{2IDTP}{2IDTP + IDFP + IDFN} \tag{31}$$

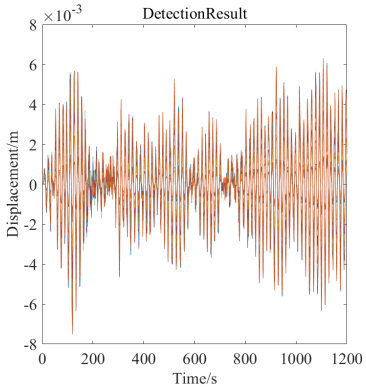
where $IDTP$ is the number of true positive IDs, $IDFP$ is the number of false positive IDs, and $IDFN$ is the number of false negative IDs. IDP represents the identification precision, IDR represents the identification recall, and IDF_1 expresses the F-score of object ID in each bounding box.



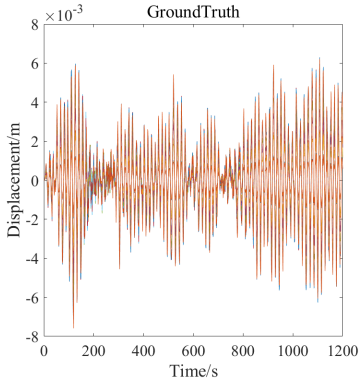
(a) Displacement detection results of case Dam1 (Parallel)



(b) Displacement ground truth results of case Dam1 (Parallel)

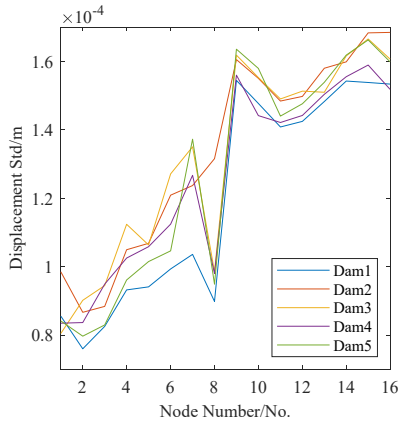


(c) Displacement detection results of case Dam4 (Unparallel)

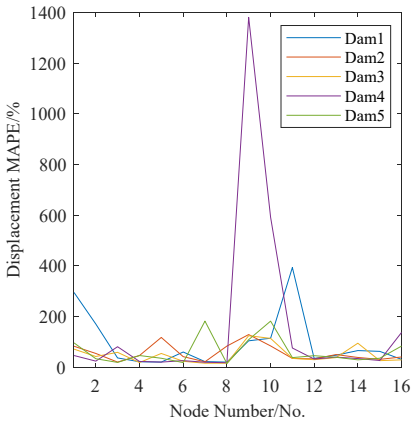


(d) Displacement ground truth results of case Dam4 (Unparallel)

Fig. 8 Detection results and ground truth results of the truss videos



(a) Displacement Std



(b) Displacement MAPE

Fig. 9 Displacement estimation Std and MAPE of each ROI for the validation dataset with the proposed method

Next, to verify the performance of the proposed precise damage localization method, the displacement estimation error standard deviation (Std) and the displacement mean absolute percentage error (MAPE) shown below are selected as the evaluation indices to assess the proposed method.

$$\text{Std} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left|(dis2D_i^d - dis2D_i^g) - \mu\right|^2} \tag{32}$$

$$\text{MAPE} = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{dis2D_i^g - dis2D_i^d}{dis2D_i^g}\right| \tag{33}$$

where $dis2D^d$ denotes the detected displacement offset, $dis2D_i^g$ represents the ground truth displacement offset, $\mu = \frac{1}{N}\sum_{i=1}^{N}(dis2D_i^d - dis2D_i^g)$.

Fig. 8 shows the truss video's detection and ground truth results. In Fig. 9, the performance of the proposed method is evaluated on the validation dataset. From these results,

the following observations are made: (i) For tracking videos, the Std error ranges between 0.07 mm and 0.17 mm, which proves that the method has high accuracy and good stability. The value of MAPE basically fluctuates in a small range, with the largest fluctuation being at Node 9. As illustrated in Fig. 6, Node 9 is at the uppermost and outermost portion of the structure, which is the most susceptible in the structure. Therefore, the experimental observation is consistent with the actual situation, which also proves the rationality of the result; (ii) By comparing the deviations of different positions, it is found that the measurement accuracy of Nodes 1-8 at the bottom level is higher than that of Nodes 9-16 at the top level. The reason is because the upper target is farther away from the center axis of the camera, and the lower resolution results in more significant errors. It also conforms to the actual on-site situation. Finally, the proposed method is tested on the test dataset. The results are shown in Fig. 10, from which consistent conclusions with the validation dataset are obtained, proving the robustness of the method.

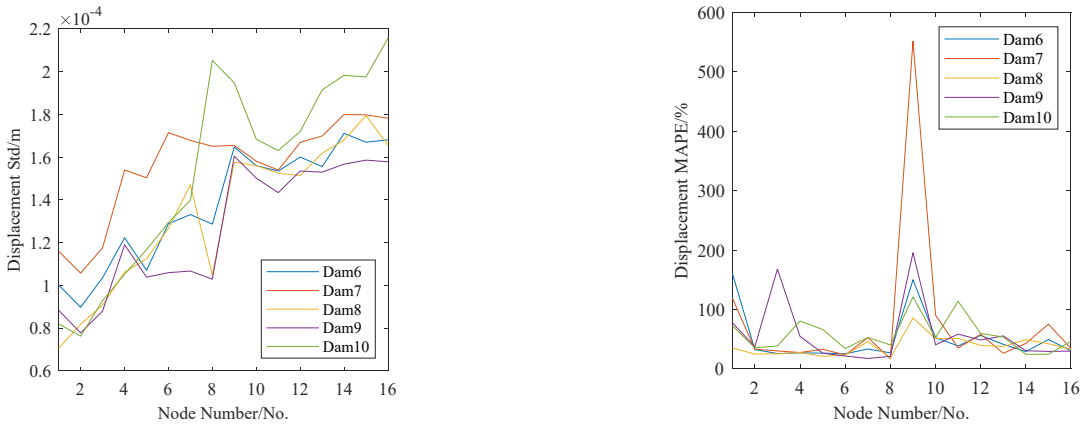
3.2.2 Performance comparison with KLT

In order to compare the performance of the proposed method with traditional target tracking algorithms commonly used in SHM, the classical Kanade-Lucas-Tomasi (KLT) (Shi 1994) target tracking algorithm is also applied, and is evaluated using the same indices given in the preceding subsection. The results obtained by the KLT method are shown in Figs. 11 and 12. By comparing these figures with Figs. 9 and 10, it can be seen that in terms of either Std or MAPE, the accuracy of results obtained by the KLT method is worse than that obtained by the proposed method in both validation dataset and testing dataset scenarios, which proves the superiority of the proposed method over the traditional target tracking algorithm.

3.3 Coarse-to-fine damage localization

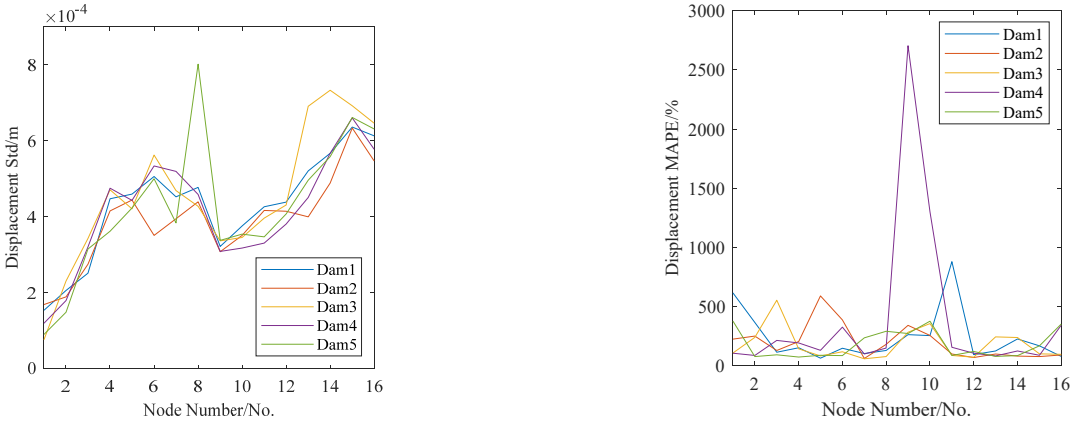
3.3.1 Coarse damage localization

Due to the limited resolution of the video, an error in vertical displacement measurement is inevitable. Node 1 and Node 9 have smaller displacement amplitudes compared with other nodes, which will lead to greater relative errors. Meanwhile, the vertical displacement measurements of the nodes on the upper level, i.e., Nodes 9-16, will be influenced by the motion perpendicular to the visual plane due to the video geometry. Hence, only the vertical displacements of Nodes 2-8 are utilized for further



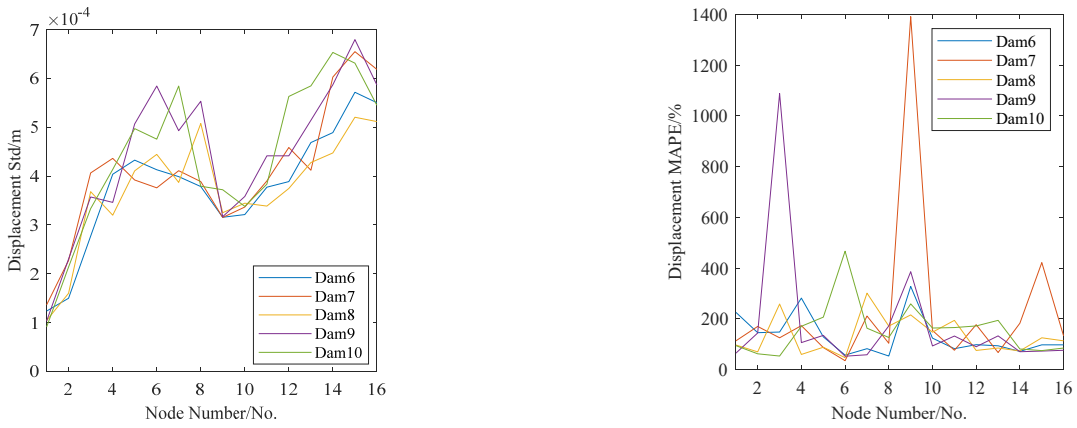
(a) Displacement Std (b) Displacement MAPE

Fig. 10 Displacement estimation Std and MAPE of each ROI for the test dataset with the proposed method



(a) Displacement Std (b) Displacement MAPE

Fig. 11 Displacement estimation Std and MAPE of each ROI for the validation dataset by KLT



(a) Displacement Std (b) Displacement MAPE

Fig. 12 Displacement estimation Std and MAPE of each ROI for the test dataset by KLT

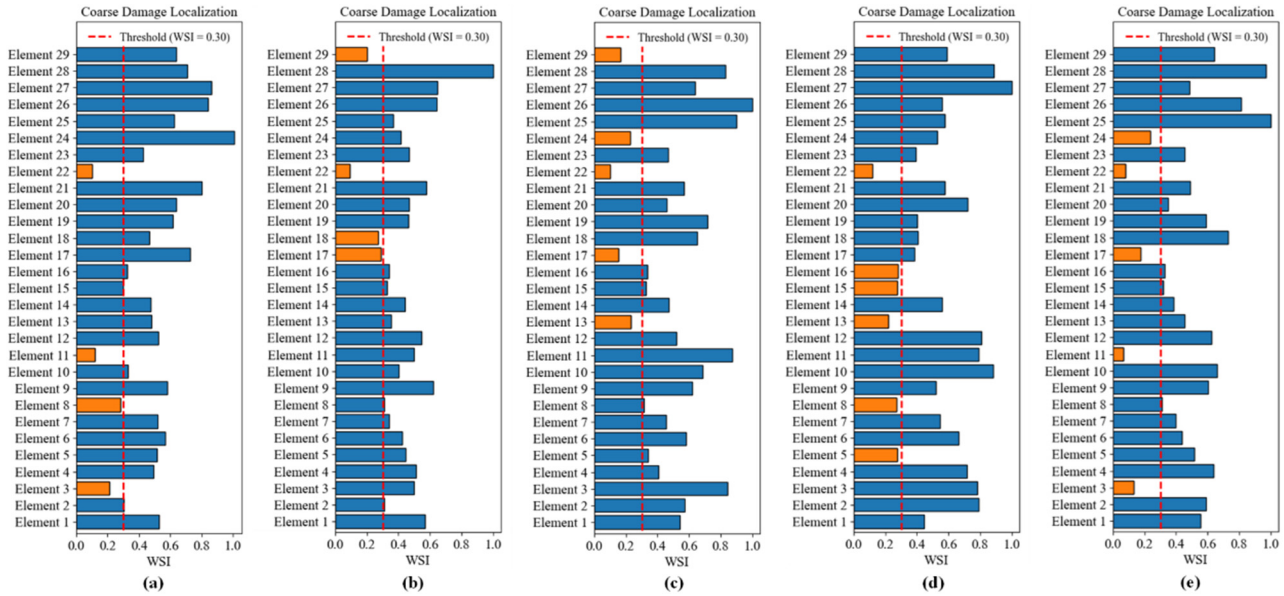
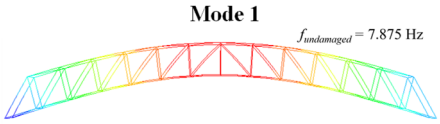


Fig. 13 WSI of each truss element in the truss bridge in the cases of (a) Dam1, (b) Dam2, (c) Dam3, (d) Dam4, and (e) Dam5

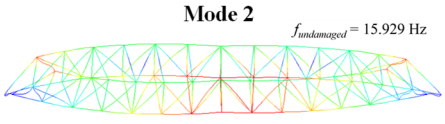
damage detection. First, the SSI method is pursued to the time-series displacement data to extract structural modal information. After an insight into the obtained power spectrum density diagrams, the first two vertical bending vibration modes are selected to formulate the observation matrix and system matrix in the state-space model. Next, the SDLV method is applied with the use of the continuous observation matrix and system matrix derived from the SSI analysis. Damage locating vectors are obtained from the null space of the surrogate flexibility matrix. Then the force vectors are applied to the finite element model of the truss bridge. To screen out suspected damage locations, the weighted stress index (WSI) is introduced (Bernal 2002). A low WSI value indicates a high risk of damage. However, it should be noted that, even if an element has the lowest WSI among all elements, it does not necessarily imply that it is a damaged element. Many factors may induce low WSIs among undamaged elements, such as the demand for static structural equilibrium. The coarse damage localization results of all five damage cases on the validation dataset are shown in Fig. 13.

Based on the results, we may pick out the suspected damaged elements. To include the really damaged truss members, truss elements with WSIs below 0.3 are chosen as the suspected damage elements, which are marked using orange bars in Fig. 13. For instance, in the case of Dam1, elements 3, 8, 11, and 22 are highlighted, indicating their higher damage risks compared with other elements. From the coarse damage localization results, it can be found that the authentic damage components are included in the suspected damage elements in the cases of Dam1, Dam2, Dam4, and Dam5. So coarse damage localization lays a solid foundation for precise damage localization and severity quantification in the next step. However, in the case of Dam 3, we discovered that the truly damaged truss member, i.e., element 6, was excluded from the group of suspected damaged elements. As shown in Fig. 13(c), the WSI of element 6 reached 0.6, much higher than what we expected. When we used ground truth displacement data, the WSI was chopped from 0.6 to around 0.1. One possible



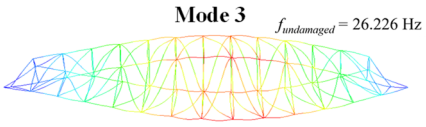
Measured Frequency Reduction

Case	Mean Frequency Reduction (Hz)	Standard Deviation (Hz)
Dam1	7.408×10^{-2}	3.975×10^{-2}
Dam2	3.033×10^{-3}	1.841×10^{-3}
Dam4	9.930×10^{-3}	3.739×10^{-3}
Dam5	7.886×10^{-2}	3.747×10^{-2}



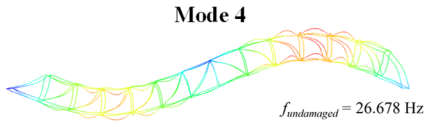
Measured Frequency Reduction

Case	Mean Frequency Reduction (Hz)	Standard Deviation (Hz)
Dam1	1.304×10^{-2}	3.379×10^{-2}
Dam2	-9.169×10^{-3}	2.497×10^{-2}
Dam4	9.495×10^{-3}	1.182×10^{-2}
Dam5	2.347×10^{-1}	1.257×10^{-1}



Measured Frequency Reduction

Case	Mean Frequency Reduction (Hz)	Standard Deviation (Hz)
Dam1	-4.513×10^{-2}	1.917×10^{-1}
Dam2	1.227×10^{-1}	8.446×10^{-2}
Dam4	1.326×10^{-2}	1.511×10^{-2}
Dam5	1.447×10^{-2}	2.074×10^{-2}



Measured Frequency Reduction

Case	Mean Frequency Reduction (Hz)	Standard Deviation (Hz)
Dam1	4.758×10^{-1}	1.062×10^{-1}
Dam2	1.829×10^{-2}	1.855×10^{-2}
Dam4	1.731×10^{-2}	9.342×10^{-3}
Dam5	1.214×10^{-1}	4.023×10^{-2}

Fig. 14 Mode shapes and measured frequency reductions of four selected vibration modes

explanation is that the SDLV method is extremely sensitive to mode shape, while mode shapes obtained from displacement measurements based on a single camera cannot be as accurate and reliable as those from contact-type sensors.

3.3.2 Precise damage localization

On the one hand, the relationship between the elastic moduli of the suspected damage elements and the modal frequency reductions caused by the stiffness degradation of structural elements can be simulated using the finite element model of the truss bridge. A GPM is then established to map this relationship. For instance, in the case of Dam1, the elastic moduli of elements 3, 8, 11, and 22 are regarded as the GPM inputs. The modal frequencies will alter as the elastic moduli vary in the finite element analysis. The modal frequency reductions in four vibration modes are calculated and regarded as the metamodel outputs using the finite element model. Fig. 14 depicts the mode shapes of the four selected vibration modes.

On the other hand, the 'field-measured' modal frequency reductions can be obtained using CV-based displacement measurements. The mean and standard deviation of the modal frequency reductions are obtained based on the first minute of the truss bridge vibration video. The first minute of the video is divided into six ten-second segments. Each segment was treated as an independent experiment, resulting in a total of six independent experiments. Based on the SSI analysis, modal frequency reductions of the four vibration modes are obtained in each independent experiment. By synthesizing all the results the mean and standard deviation of the 'field-measured' modal frequency reductions are yielded as shown in Fig. 14.

Then, the BI-SMU is executed for precise damage localization. The likelihood function is constructed by combining 'field-measured' frequency reductions with the GPM predictions to realize the MAP estimation. The posterior probability distributions of the elastic moduli of the suspected damage elements are simulated using the MH-MCMC algorithm with thirty thousand iterations. Fig. 15 illustrates the precise damage localization results of four damage cases. For instance, in the case of Dam1, as shown in Fig. 15(a), the elastic modulus of truss element 3 reduces to around 50 percent of its initial value after three thousand iterations, while those of truss elements 8, 11, and 22 remain stable at values close to the original one. As a result, truss element 3 is identified to be the damaged element in the case of Dam1. Similarly, truss elements 18, 16, and 11 are picked out as the damaged elements in the cases of Dam2, Dam4, and Dam5, respectively.

3.4 Damage severity quantification

After identifying the damaged elements in the previous phase, the elastic moduli of elements that were not picked out during precise damage localization are fixed as their original values as in undamaged state, and MH-MCMC with thirty thousand iterations is performed again to elicit the posterior probability density distributions of the elastic moduli of the damaged elements. The damage severity quantification results in four damage cases are shown in Fig. 16. In each case, the left panel shows the last ten thousand iterations of the Markov chain while the right panel shows its distribution. For instance, in the case of Dam1, as shown in Fig. 16(a), the elastic modulus of truss element 3 reduces to around 48.9 percent of its initial value. The updated material property can be regarded as the prediction of the remaining stiffness of the damaged element. It should be noted that element 3 is the truly damaged element in the case of Dam1, and its residual stiffness is 47.3 percent of its initial value. The damage severity quantification results on the validation dataset are summarized in Table 3. It is worth noting that in Table 3, the prediction of the residual stiffness is defined as the mean value of the last ten thousand iterations of the Markov Chain. It is seen that the proposed method has a good capacity for damage localization and quantification on the validation dataset cases. However, among the last five testing dataset scenarios, the proposed method proves effective only in the cases of Dam6 and Dam7, but

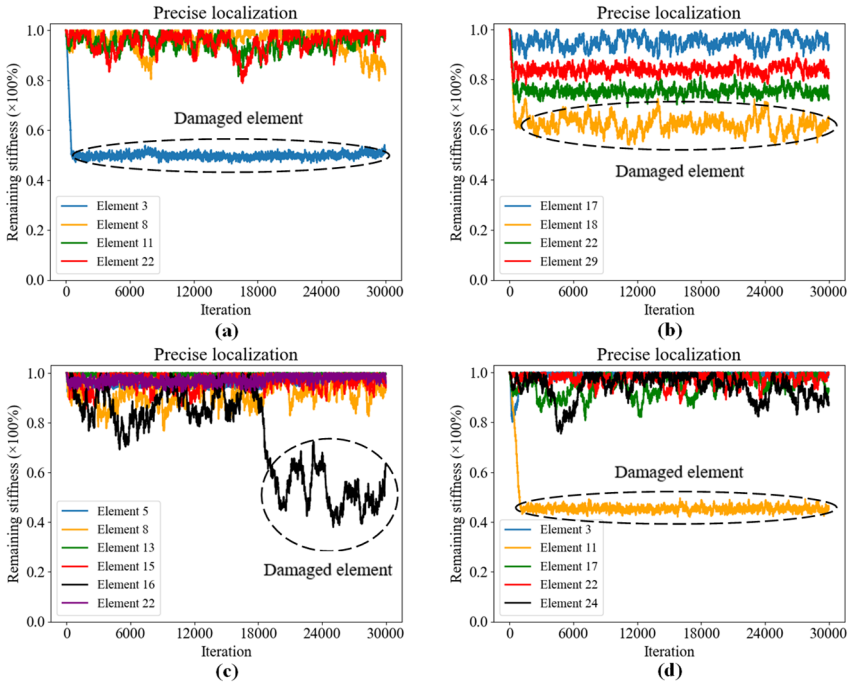


Fig. 15 Precise damage localization results of (a) Dam1, (b) Dam2, (c) Dam4, and (d) Dam5

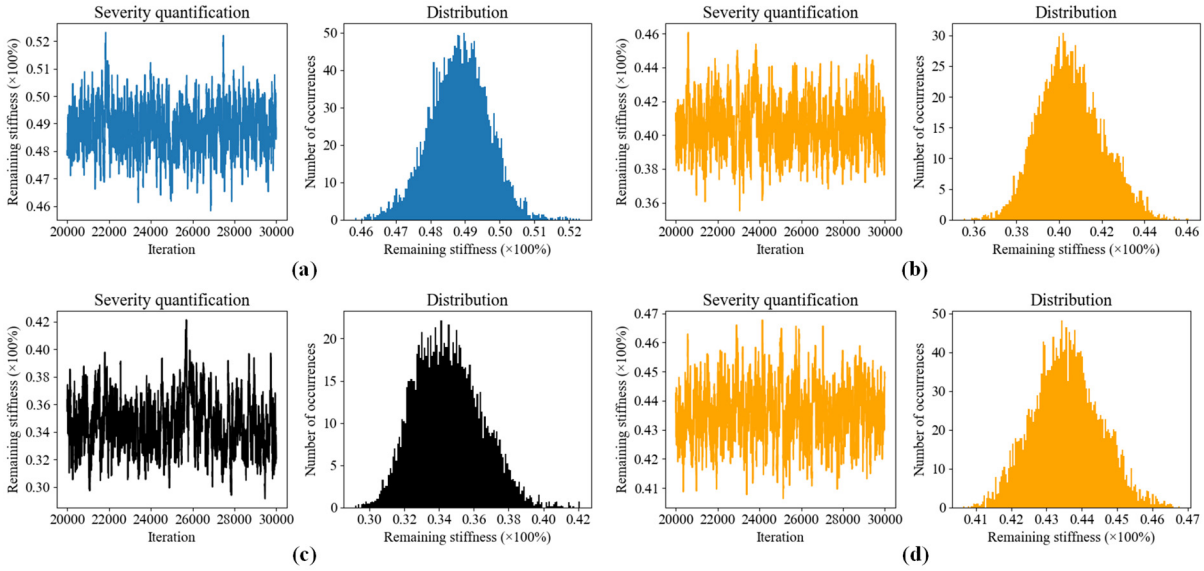


Fig. 16 Damage severity quantification results of (a) Dam1, (b) Dam2, (c) Dam4, and (d) Dam5

ineffective in the remaining three cases. On the whole, the detection rate of the proposed method reaches 60%. In the category of 'parallel cases', the detection rate is 100%. However, when it comes to unparallel cases, the detection rate drops to only 33%. An analysis of the possible reasons for such results is that the experiment uses a monocular camera. When the positions of the camera and the detection plane are relatively fixed, the movement of the target can be accurately detected through simple camera calibration. However, in addition to vibrating up and down in the detection plane, the object studied in this experiment may also swing back and forth. As such, the relationship between the detection plane and the camera plane will change, and the original calibration parameters will become invalid, which will lead to inaccurately calculated offsets that have been observed in the results of Section 3.2.1. Moreover, the deviation caused by this factor is more significant when the detection plane and the camera plane are not parallel. It may greatly affect the damage localization and assessment results. A more feasible solution is to use a binocular camera or a 3D camera for target tracking, which can help overcome the above problem and improve the accuracy of damage assessment.

Table 3 Damage detection results of all ten cases in the IC-SHM 2021 dataset

Case	Category	Prediction		Ground truth	
		Element	RS	Element	RS
Dam1	Parallel	3	48.9%	3	47.3%
Dam2	Parallel	18	40.6%	18	47.3%
Dam3	Unparallel	Fail	Fail	6	42.5%
Dam4	Unparallel	16	34.5%	16	27.5%
Dam5	Unparallel	11	43.7%	11	40.0%
Dam6	Parallel	20	16.5%	20	12.0%
Dam7	Parallel	5	37.8%	5	33.0%
Dam8	Unparallel	Fail	Fail	9	52.0%
Dam9	Unparallel	Fail	Fail	11	43.5%
Dam10	Unparallel	Fail	Fail	12	38.0%

4. Conclusions

The paper proposed a computer vision-based vibration measurement and damage assessment method for truss bridges. The proposed displacement measurement method can make up for the disadvantages of traditional target tracking algorithms and enhance detection efficiency. The proposed damage detection method can accurately construct the relationship between the material or sectional properties of the suspected damage elements and the structural modal frequency reductions of the truss bridge based on a GPM. When the relationship between the detection plane and the camera plane is well defined, the damage location can be precisely identified and the damage severity can be effectively quantified. Although the work presented in this paper is a preliminary study on the IC-SHM 2021 benchmark problem and needs improvement in many aspects, it provides a framework from deep learning-based target tracking to damage localization and to damage severity quantification in the context of CV-based SHM.

Acknowledgments

The work was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region (SAR), China (Grant No. R5020-18) and a grant from The Hong Kong Polytechnic University (Grant No. 1-YW5H). The authors also appreciate the funding support by the Innovation and Technology Commission of Hong Kong SAR Government to the Hong Kong Branch of National Engineering Research Center on Rail Transit Electrification and Automation (Grant No. K-BBY1), and The Hong Kong Polytechnic University's Postdoc Matching Fund Scheme (Grant No. 1-W20D).

References

An, Y., Ou, J., Li, J. and Spencer, B. (2014), "Stochastic DLV method for steel truss structures: simulation and experiment", *Smart Struct. Syst.*, *Int. J.*, **14**(2), 105-128. https://doi.org/10.12989/sss.2014.14.2.105

Bakhtiari-Nejad, F., Rahai, A. and Esfandiari, A. (2005), "A structural damage detection method using static noisy data", *Eng. Struct.*, **27**(12), 1784-1793. https://doi.org/10.1016/j.engstruct.2005.04.019

Bernagozzi, G., Mukhopadhyay, S., Betti, R., Landi, L. and Diotallevi, P.P. (2018), "Output-only damage detection in buildings using proportional modal flexibility-based deflections in unknown mass scenarios", *Eng. Struct.*, **167**, 549-566. https://doi.org/10.1016/j.engstruct.2018.04.036

Bernagozzi, G., Achilli, A., Betti, R., Diotallevi, P.P., Landi, L., Quqa, S. and Tronci, E.M. (2021), "On the use of multivariate autoregressive models for vibration-based damage detection and localization", *Smart Struct. Syst.*, *Int. J.*, **27**(2), 335-350. https://doi.org/10.12989/sss.2021.27.2.335

Bernal, D. (2002), "Load vectors for damage localization", *J. Eng. Mech.*, **128**(1), 7-14. https://doi.org/10.1061/(ASCE)0733-9399(2002)128:1(7)

Bernal, D. (2006), "Flexibility-based damage localization from stochastic realization results", *J. Eng. Mech.*, **132**(6), 651-658. https://doi.org/10.1061/(ASCE)0733-9399(2006)132:6(651)

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

Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H. and Wei, Y. (2017), "Deformable convolutional networks", *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, October.

Dong, C.-Z. and Catbas, F.N. (2021), "A review of computer vision–based structural health monitoring at local and global levels", *Struct. Health Monit.*, **20**(2), 692-743. https://doi.org/10.1177/1475921720935585

Eltouny, K.A. and Liang, X. (2021), "Bayesian-optimized unsupervised learning approach for structural damage detection", *Comput.-Aided Civ. Infrastr. Eng.*, **36**(10), 1249-1269. https://doi.org/10.1111/mice.12680

Erdogan, Y.S. and Ada, M. (2020), "A computer-vision based vibration transducer scheme for structural health monitoring applications", *Smart Mater. Struct.*, **29**(8), 085007. https://doi.org/10.1088/1361-665X/ab9062

Fang, K., Xiang, Y., Li, X. and Savarese, S. (2018), "Recurrent autoregressive networks for online multi-object tracking", *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, Lake Tahoe, USA, March.

Feng, D. and Feng, M.Q. (2016), "Output-only damage detection using vehicle-induced displacement response and mode shape curvature index", *Struct. Control Health Monit.*, **23**(8), 1088-1107. https://doi.org/10.1002/stc.1829

Feng, D. and Feng, M.Q. (2017), "Experimental validation of cost-effective vision-based structural health monitoring", *Mech. Syst. Sig. Process.*, **88**, 199-211. https://doi.org/10.1016/j.ymssp.2016.11.021

Figueiredo, E., Radu, L., Worden, K. and Farrar, C.R. (2014), "A Bayesian approach based on a Markov-chain Monte Carlo method for damage detection under unknown sources of variability", *Eng. Struct.*, **80**, 1-10. https://doi.org/10.1016/j.engstruct.2014.08.042

Frans, R., Arfiadi, Y. and Parung, H. (2017), "Comparative study of mode shapes curvature and damage locating vector methods for damage detection of structures", *Procedia Eng.*, **171**, 1263-1271. https://doi.org/10.1016/j.proeng.2017.01.420

Gao, Y., Spencer Jr, B.F. and Bernal, D. (2007), "Experimental verification of the flexibility-based damage locating vector method", *J. Eng. Mech.*, **133**(10), 1043-1049. https://doi.org/10.1061/(ASCE)0733-9399(2007)133:10(1043)

Gomez, F., Narazaki, Y., Hoskere, V., Spencer Jr, B.F. and Smith,

M.D. (2022), "Bayesian inference of dense structural response using vision-based measurements", *Eng. Struct.*, **256**, 113970. https://doi.org/10.1016/j.engstruct.2022.113970

He, K., Zhang, X., Ren, S. and Sun, J. (2016), "Deep residual learning for image recognition", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, June.

He, Y., Zhang, L., Chen, Z. and Li, C.Y. (2022), "A framework of structural damage detection for civil structures using a combined multi-scale convolutional neural network and echo state network", *Eng. Comput.*, 1-19. https://doi.org/10.1007/s00366-021-01584-4

Khuc, T. and Catbas, F.N. (2017), "Completely contactless structural health monitoring of real-life structures using cameras and computer vision", *Struct. Control Health Monit.*, **24**(1), e1852. https://doi.org/10.1002/stc.1852

Kim, C.-W., Morita, T., Oshima, Y. and Sugiura, K. (2015), "A Bayesian approach for vibration-based long-term bridge monitoring to consider environmental and operational changes", *Smart Struct. Syst.*, *Int. J.*, **15**(2), 395-408. https://doi.org/10.12989/sss.2015.15.2.395

Lei, Y., Liu, L., Mi, J. and Zhang, Y. (2021), "Damage detection of bridge structures under unknown seismic excitations using support vector machine based on transmissibility function and wavelet packet energy", *Smart Struct. Syst.*, *Int. J.*, **27**(2), 257-266. https://doi.org/10.12989/sss.2021.27.2.257

Liang, X. (2019), "Image-based post-disaster inspection of reinforced concrete bridge systems using deep learning with Bayesian optimization", *Comput.-Aided Civ. Infrastr. Eng.*, **34**(5), 415-430. https://doi.org/10.1111/mice.12425

Mahmoudi, N., Ahadi, S.M. and Rahmati, M. (2019), "Multi-target tracking using CNN-based features: CNNMTT", *Multimedia Tools Applicat.*, **78**(6), 7077-7096. https://doi.org/10.1007/s11042-018-6467-6

Mousavi, A.A., Zhang, C., Masri, S.F. and Gholipour, G. (2022), "Structural damage detection method based on the complete ensemble empirical mode decomposition with adaptive noise: a model steel truss bridge case study", *Struct. Health Monit.*, **21**(3), 887-912. https://doi.org/10.1177/14759217211013535

Nagarajaiah, S. and Basu, B. (2009), "Output only modal identification and structural damage detection using time frequency & wavelet techniques", *Earthq. Eng. Eng. Vib.*, **8**(4), 583-605. https://doi.org/10.1007/s11803-009-9120-6

Narazaki, Y., Gomez, F., Hoskere, V., Smith, M.D. and Spencer Jr, B.F. (2021), "Efficient development of vision-based dense three-dimensional displacement measurement algorithms using physics-based graphics models", *Struct. Health Monit.*, **20**(4), 1841-1863. https://doi.org/10.1177/1475921720939522

Ngeljaratan, L. and Moustafa, M.A. (2020), "Structural health monitoring and seismic response assessment of bridge structures using target-tracking digital image correlation", *Eng. Struct.*, **213**, 110551. https://doi.org/10.1016/j.engstruct.2020.110551

Ni, Y.Q. and Zhang, Q.H. (2021), "A Bayesian machine learning approach for online detection of railway wheel defects using track-side monitoring", *Struct. Health Monit.*, **20**(4), 1536-1550. https://doi.org/10.1177/1475921720921772

Ni, Y.Q., Wang, J. and Chan, T.H.T. (2015), "Structural damage alarming and localization of cable-supported bridges using multi-novelty indices: a feasibility study", *Struct. Eng. Mech., Int. J.*, **54**(2), 337-362. https://doi.org/10.12989/sem.2015.54.2.337

Ni, Y.Q., Wang, Y.W., Liao, W.Y. and Chen, W.H. (2019), "A vision-based system for long-distance remote monitoring of dynamic displacement: experimental verification on a supertall structure", *Smart Struct. Syst., Int. J.*, **24**(6), 769-781. https://doi.org/10.12989/sss.2019.24.6.769

Ni, Y.Q., Wang, Y.W. and Zhang, C. (2020), "A Bayesian approach for condition assessment and damage alarm of bridge expansion joints using long-term structural health monitoring data", *Eng. Struct.*, **212**, 110520. https://doi.org/10.1016/j.engstruct.2020.110520

Shi, J. (1994), "Good features to track", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, USA, June.

Spencer Jr, B.F., Hoskere, V. and Narazaki, Y. (2019), "Advances in computer vision-based civil infrastructure inspection and monitoring", *Engineering*, **5**(2), 199-222. https://doi.org/10.1016/j.eng.2018.11.030

Sun, L., Shang, Z., Xia, Y., Bhowmick, S. and Nagarajaiah, S. (2020), "Review of bridge structural health monitoring aided by big data and artificial intelligence: From condition assessment to damage detection", *J. Struct. Eng.*, **146**(5), 04020073. https://doi.org/10.1061/(ASCE)ST.1943-541X.0002535

Uzun, M., Sun, H., Smit, D. and Büyüköztürk, O. (2019), "Structural damage detection using Bayesian inference and seismic interferometry", *Struct. Control Health Monit.*, **26**(11), e2445. https://doi.org/10.1002/stc.2445

Voigtlaender, P., Krause, M., Osep, A., Luiten, J., Sekar, B.B.G., Geiger, A. and Leibe, B. (2019), "Mots: Multi-object tracking and segmentation", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, USA, June.

Wan, H.P. and Ren, W.X. (2016), "Stochastic model updating utilizing Bayesian approach and Gaussian process model", *Mech. Syst. Sig. Process.*, **70**, 245-268. https://doi.org/10.1016/j.ymssp.2015.08.011

Wang, J., Liu, X.Z. and Ni, Y.Q. (2018), "A Bayesian probabilistic approach for acoustic emission-based rail condition assessment", *Comput.-Aided Civ. Infrast. Eng.*, **33**(1), 21-34. https://doi.org/10.1111/mice.12316

Wang, Z., Zheng, L., Liu, Y., Li, Y. and Wang, S. (2020), "Towards real-time multi-object tracking", *Proceedings of the European Conference on Computer Vision*, Glasgow, UK, August.

Wang, X., Hou, R., Xia, Y. and Zhou, X. (2021a), "Structural damage detection based on variational Bayesian inference and delayed rejection adaptive Metropolis algorithm", *Struct. Health Monit.*, **20**(4), 1518-1535. https://doi.org/10.1177/1475921720921256

Wang, Y.W., Ni, Y.Q., Zhang, Q.H. and Zhang, C. (2021b), "Bayesian approaches for evaluating wind-resistant performance of long-span bridges using structural health monitoring data", *Struct. Control Health Monit.*, **28**(4), e2699. https://doi.org/10.1002/stc.2699

Wang, Q.A., Dai, Y., Ma, Z.G., Ni, Y.Q., Tang, J.Q., Xu, X.Q. and Wu, Z.Y. (2022a), "Towards probabilistic data-driven damage detection in SHM using sparse Bayesian learning scheme", *Struct. Control Health Monit.*, **29**(11), e3070. https://doi.org/10.1002/stc.3070

Wang, Y.W., Ni, Y.Q. and Wang, S.M. (2022b), "Structural health monitoring of railway bridges using innovative sensing technologies and machine learning algorithms: a concise review", *Intell. Transport. Infrastr.*, 1, liac009. https://doi.org/10.1093/iti/liac009

Wang, Y.W., Zhang, C., Ni, Y.Q. and Xu, X.Y. (2022c), "Bayesian probabilistic assessment of occupant comfort of high-rise structures based on structural health monitoring data", *Mech. Syst. Sig. Process.*, **163**, 108147. https://doi.org/10.1016/j.ymssp.2021.108147

Xu, Y. (2020), "Photogrammetry-based structural damage detection by tracking a visible laser line", *Struct. Health Monit.*, **19**(1), 322-336. https://doi.org/10.1177%2F1475921719840354

Xu, Y., Brownjohn, J. and Kong, D. (2018), "A non-contact vision-based system for multipoint displacement monitoring in a cable-

stayed footbridge", *Struct. Control Health Monit.*, **25**(5), e2155. https://doi.org/10.1002/stc.2155

Ye, X.W., Ni, Y.Q., Wai, T.T., Wong, K.Y., Zhang, X.M. and Xu, F. (2013), "A vision-based system for dynamic displacement measurement of long-span bridges: algorithm and verification", *Smart Struct. Syst., Int. J.*, **12**(3-4), 363-379. https://doi.org/10.12989/sss.2013.12.3_4.363

Ye, X.W., Dong, C.Z. and Liu, T. (2016), "A review of machine vision-based structural health monitoring: methodologies and applications", *J. Sens.*, 2016. https://doi.org/10.1155/2016/7103039

Zhang, Y., Wang, C., Wang, X., Zeng, W. and Liu, W. (2021), "Fairmot: On the fairness of detection and re-identification in multiple object tracking", *Int. J. Comput. Vision*, **129**(11), 3069-3087. https://doi.org/10.1007/s11263-021-01513-4

Zhou, X., Koltun, V. and Krähenbühl, P. (2020), "Tracking objects as points", *Proceedings of the European Conference on Computer Vision*, Glasgow, UK, August.