Smart Structures and Systems, *Vol. 24, No. 5 (2019) 617-630*
DOI: https://doi.org/10.12989/sss.2019.24.5.617

A completely non-contact recognition system for bridge unit influence line using portable cameras and computer vision

Chuan-Zhi Dong[1a], Selcuk Bas[1,2b] and F. Necati Catbas*[1]

[1]Department of Civil, Environmental, and Construction Engineering, University of Central Florida, 12800 Pegasus Drive, Suite 211, Orlando, Florida 32816, USA
[2]Department of Civil Engineering, Bartin University, Bartin, Turkey

(Received June 18, 2019, Revised August 12, 2019, Accepted August 25, 2019)

Abstract. Currently most of the vision-based structural identification research focus either on structural input (vehicle location) estimation or on structural output (structural displacement and strain responses) estimation. The structural condition assessment at global level just with the vision-based structural output cannot give a normalized response irrespective of the type and/or load configurations of the vehicles. Combining the vision-based structural input and the structural output from non-contact sensors overcomes the disadvantage given above, while reducing cost, time, labor force including cable wiring work. In conventional traffic monitoring, sometimes traffic closure is essential for bridge structures, which may cause other severe problems such as traffic jams and accidents. In this study, a completely non-contact structural identification system is proposed, and the system mainly targets the identification of bridge unit influence line (UIL) under operational traffic. Both the structural input (vehicle location information) and output (displacement responses) are obtained by only using cameras and computer vision techniques. Multiple cameras are synchronized by audio signal pattern recognition. The proposed system is verified with a laboratory experiment on a scaled bridge model under a small moving truck load and a field application on a footbridge on campus under a moving golf cart load. The UILs are successfully identified in both bridge cases. The pedestrian loads are also estimated with the extracted UIL and the predicted weights of pedestrians are observed to be in acceptable ranges.

Keywords: structural health monitoring; displacement; unit influence line; computer vision; structural identification

1. Introduction

Bridge structures are important components of the transportation systems, and it is important to keep them in safe working condition to ensure the normal operation of the transportation network. With daily traffic and other external effects, bridges are undergoing with structural changes, deterioration and damages over time. Currently, human visual inspection is still a common approach to detect defects and most of the decisions are made by inspectors' experiences (Catbas *et al*. 2017). For safe operation, timely maintenance and convenient management in aspect of structural problems, effective sensing technologies and analytical approaches are necessary to detect the structural changes and damages and give reliable condition assessment and performance evaluation timely and sufficiently (Zaurin *et al*. 2015). To achieve this goal, in last two decades structural health monitoring (SHM) has been widely explored and implemented on bridges all over the world. SHM systems can collect massive valuable information including structural input (loads and other external effects) and structural output (responses such as displacement, strain and acceleration) and make diagnosis and prognosis to support the structural safety and decision making (Chen *et al*. 2019, Ni *et al*. 2010, 2011, Ye *et al*. 2013b, 2015b, 2016e).

With the benefits of interdisciplinary integrations, various advanced sensing technologies such as elasto-magneto-electric (EME) sensor for in-service steel cable forces measurement (Duan *et al*. 2015), wireless sensors for dynamic monitoring (Celik *et al*. 2018b), fiber Bragg grating (FBG) sensor for strain monitoring (Ye *et al*. 2016d, 2017), LiDAR scanning for structural condition assessment (Chen *et al*. 2012), skin-type sensor for strain measurement (Kong *et al*. 2018), infrared thermography for automated concrete deck inspection (Catbas *et al*. 2017) and vision-based bridge monitoring at global level (Catbas *et al*. 2018), etc. have been employed in current research and practice. Among these technologies, vision-based approaches are gathering increasing attention in the field of SHM (Dong and Catbas 2019, Ye *et al*. 2016a) due to the advantages such as non-contact, long distance, low cost, time saving and ease of use. Generally, the studies and practices of vision-based monitoring are divided into two aspects: 1) inspection and condition assessment at local level such as crack, spalling (Karaaslan *et al*. 2018) and delamination detection (Hiasa *et al*. 2017) and 2) structural monitoring at global level such as vibration and deflection monitoring (Dong *et al*. 2015, 2018b, 2019b; Xu and Brownjohn 2018; Ye *et al*. 2013a, 2015a, 2016b, f;), cable force monitoring (Feng *et a*l. 2017; Ye *et al*. 2016c), modal analysis (Chen *et*

*Corresponding author, Professor
E-mail: catbas@ucf.edu
[a] Ph.D. Candidate
[b] Assistant Professor and Visiting Scholar

http://www.techno-press.com/journals/sss&subpage=7
ISSN: 1738-1584 (Print), 1738-1991 (Online)

al. 2018a, Hoskere *et al*. 2019, Yang *et al*. 2017), load estimation (Celik *et al*. 2018a), load rating (Catbas *et al*. 2012) and load capacity estimation (Lee *et al*. 2006) etc. With vision-based inspection at local level, the condition assessment is carried out when damages already appear and are visible and large enough. It is very hard to estimate the tiny deteriorations of structures and give further prediction. Vision-based monitoring at global level mostly collects the structural responses and make evaluation of structural performance and safety based on the time histories such as displacement, acceleration and strain. However, the research about the identification of operating traffic loads as structural input doesn't receive sufficient attention. Even though weigh-in-motion (WIM) systems are installed as parts of the SHM system on bridges, only the weight of vehicles can be estimated and the position information is hard to obtain. Both vehicle loads and position information on the bridge are quite important to structural identification at global level. If only response data are used for structural identification without knowing input force, the structural change and damage has to be large enough to induce significant change to the output responses. Feature extraction from output responses for damage detection is very difficult.

To comprehensively and sufficiently evaluate the structural performance, assess the condition, predict safety and remaining life, the monitoring of structural input and output are necessary. Zaurin and Catbas (2010a, b) combined cameras and conventional sensors such as strain gages to extract the strain unit influence line (UIL) and recognized damages using statistical outlier detection from UIL vector sets and also conduct load rating (Catbas *et al*. 2012). Their work was validated with laboratory experiments on large-scale bridge model and field application on real life bridges. Khuc and Catbas (2018) integrated camera and displacement sensors to obtain displacement unit influence surface and proposed a statistical approach to detect bridge damages. Damage cases were simulated by changing the boundary condition and connection of bridge components of a bridge model in laboratory. Both the two examples extracted the static structural properties as damage features and used cameras for input monitoring and conventional sensors for output. There are other studies focusing on input-output data and evaluating structural dynamic properties. Tian *et al*. (2018) conducted impact test on a small-scale beam in laboratory using camera to capture the human input and accelerometers to collect the output responses. The impact test was also validated on a footbridge and modal parameters such as frequency, mode shape and scaling factor were extracted. In another research, Tian *et al*. (2019) conducted impact test on a small scale beam with moveable camera to collect the beam outputs and impact hammer to give excitation and record the inputs. The major difference between two studies done by Tian *et al*. is just to switch the data collection approaches for input-output data sources. The studies above including static and dynamic structural properties estimation were carried out by combining cameras and conventional sensors. The drawbacks of using conventional sensors are traffic closure, setup time and labor force to deal with the cable wiring work. It is not convenient to conduct such experiments, especially for field application. The synchronization between cameras and sensors are also a big challenge.

In this paper, the study of structural identification using input-output data will further advanced from combining cameras and conventional sensors to a completely non-contact recognition system just using cameras. The input and output data are both obtained from portable cameras and computer vision techniques are employed to process the images and track the structural behaviors. UIL is an effective and sensitive index for monitoring bridge behavior under identified loading conditions and explicit structural feature for efficient structural evaluation and assessment. It is also very intuitive for engineers. The proposed recognition system will take UIL as the target parameter for structural identification and the proposed UIL extraction method can be extended to a fully non-contact damage detection approach.

2. Methodology

2.1 UIL

UIL as shown in Fig. 1 indicates the variation of a response such as moment, force, displacement, strain and acceleration at a given position on a structure due to the imposition of a unit load at any point on the structure (Zaurin and Catbas 2010a). To generate a UIL, a unit load is imposed on the structure and moved on it. The response induced by the load at the selected position on the structure is calculated by structural analysis methods or measured by experimental approaches. The response values are then plotted against with the position of load on the structure to generate the UIL. Mathematically UIL of a selected position is the function of the position of the moving unit load on the structure. The detailed concept and calculations of UIL are discussed in elementary structural analysis courses as basics and here only the procedure of UIL extraction using experimental data is introduced.

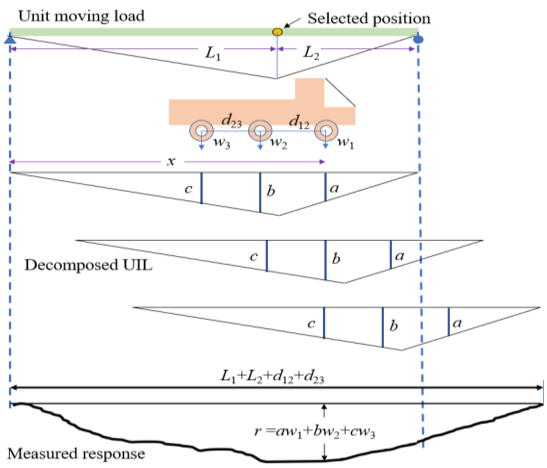


Fig. 1 UIL decomposition

When UIL is extracted from experimental data by processing an inverse problem, UIL provides a signature with a normalized structural response for the selected critical locations instrumented by any type of sensors (selected positions in Fig. 1). To extract UIL with experimental approaches, the weight and location of each axle of a vehicle crossing the bridge has to be known in advance and responses of the selected position are also measured with sensors. Then the UIL of the structure can be extracted using the following equation (Zaurin and Catbas 2010a)

$$\{r\}=[w]\{u\} \tag{1}$$

where $\{r\}$ is the vector containing the responses of the selected position induced by the moving load, $[w]$ is the matrix containing the axle weights with respect to the corresponding distances, and $\{u\}$ is the UIL vector. Fig. 1 gives an example of the extraction of moment UIL. In this example, a specific position determined by L_1 and L_2 is selected and a moving vehicle is imposed on the bridge. The axle weights of the vehicle are w_1, w_2 and w_3. The distance between axles are d_{12} and d_{23}. In this case, one element of Eq. (1) can be expressed as

$$r=aw_1+bw_2+cw_3 \tag{2}$$

When knowing any location of the vehicle on the bridge, Eq. (1) can be written as

$$\begin{Bmatrix} r_1 \\ r_2 \\ . \\ . \\ r_m \end{Bmatrix}_{m\times 1} = \begin{bmatrix} w_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & . \\ . & . & . & . & . & . & . & . & . & . \\ w_2 & 0 & 0 & 0 & w_1 & 0 & 0 & 0 & 0 & . \\ . & . & . & . & . & . & . & . & . & . \\ w_3 & 0 & 0 & w_2 & 0 & 0 & 0 & w_1 & 0 & . \\ . & . & . & . & . & . & . & . & . & . \\ 0 & w_3 & 0 & 0 & w_2 & 0 & 0 & 0 & w_1 & . \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ . & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & w_3 \end{bmatrix}_{m\times n} \times \begin{Bmatrix} u_1 \\ u_2 \\ . \\ u_n \end{Bmatrix}_{n\times 1} \tag{3}$$

where n is the moving steps of the vehicle and also the number of discretized coefficients for unit influence along the actual length of the bridge, and m is the number of the samples of the measured responses. The UIL is calculated as an inverse problem using the equation below

$$\{u\}=[w]^{-1}\{r\} \tag{4}$$

In the study, displacement UIL is targeted and the displacement is regarded as the response that can be measured by vision-based methods. The weight and distance between axles are predesignated and the location of the vehicle is estimated by vehicle tracking from images. In the followings, the vision-based structural input estimation (vehicle location) and vision-based structural output estimation method (displacement responses) are introduced respectively.

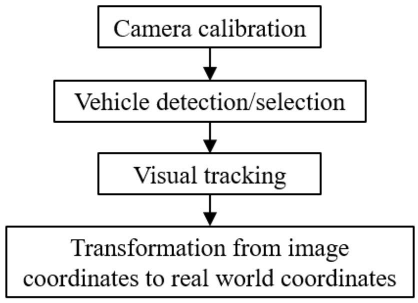


Fig. 2 General procedure of vehicle tracking

2.2 Vision-based structural input estimation: vehicle location

2.2.1 General procedure of vehicle tracking

To identify the vehicle location on bridge surface, in general there are four steps as shown in Fig. 2. At first, the camera is calibrated to rectify the distortions such as projective distortion caused by camera pose and radial distortion caused by lenses. Then the object detection algorithms are implemented to detect the category of the vehicles and give the initial bounding boxes of detected vehicles and they will be regarded as the tracking targets. The tracking targets can also be selected manually. In the third step, the visual tracking algorithms are implemented to tracking the detected or selected vehicles and the vehicle location in each frame of the image sequence or video can be estimated. At last, the vehicle location in the image coordinates is transformed to the real-world coordinates to estimate the vehicle location on bridges.

2.2.2 Camera calibration

During digital recording, three-dimensional (3D) objects in the real world are projected onto the image plane (two-dimensional, 2D) of the camera. The camera calibration is to estimate the projection process including camera intrinsic and extrinsic parameters and distortion parameters. Fig. 3 illustrates a pinhole camera model.

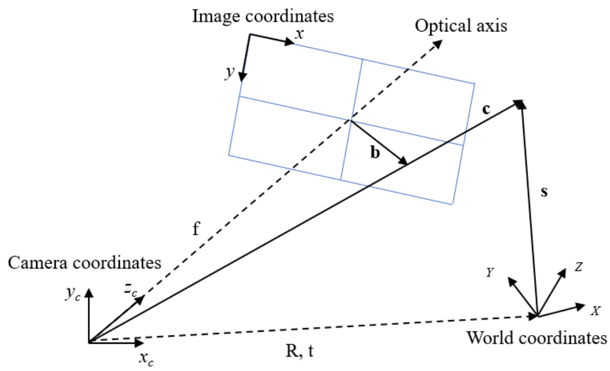


Fig. 3 Pinhole camera model

The projection from world coordinates to the image coordinates through camera coordinates can be expressed by the formula below

$$s\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (5)$$

and it can be simplified as

$$s\mathbf{x} = \mathbf{K}\left[\mathbf{R} \mid \mathbf{t}\right]\mathbf{X} \quad (6)$$

where s is the scale factor, $\mathbf{x} = (x, y, 1)^T$ are image coordinates, $\mathbf{X} = (X, Y, Z, 1)^T$, are world coordinates, and $\mathbf{K}$ is the camera intrinsic parameters which represent the projective transformation from the three-dimensional (3D) real world to the two-dimensional (2D) image. In the intrinsic parameters, f_x and f_y are the focal lengths of the lens in horizontal and vertical directions, c_x and c_y are offsets of the optical axis in horizontal and vertical directions, and γ is the skew factor of the lens. $\mathbf{R}$ and $\mathbf{t}$ are camera extrinsic parameters which represents the rigid rotation and translation from the 3D real world coordinates to the 3D camera coordinates, and r_{ij} (i, j = 1, 2, 3) and t_i (i=1, 2, 3) are the elements of $\mathbf{R}$ and $\mathbf{T}$ respectively. From Eq. (5) and (6), it is indicated that the camera intrinsic parameters are relevant to the camera and lens and the camera extrinsic parameters are relevant to the relative position between the camera-lens and real objects. Once the camera is calibrated with specific lens, as soon as the focal lens doesn't change, the intrinsic parameters don't change. While the extrinsic parameters should be calibrated in different application scenarios. The black and white chessboard is used to do calibration and Zhang's practical approach (Zhang 2002) is always implemented. Various commercial software such as MATLAB, NI VISION and Halcon and open source software such as OpenCV provide libraries to complete the calibration quickly.

2.2.3 Vehicle detection/selection

There are various algorithms for automated vehicle detection. (Zaurin and Catbas 2010a) applied background subtraction to detect the vehicles and made classifications. (Khuc and Catbas 2018) implemented AdaBoost technique and Cascade classifier using histograms of oriented gradients (HOG) features to train and detect vehicle types. With the application of deep learning in computer vision, deep learning-based visual tracking has made great progress. The classical studies in this area are R-CNN (Regions with Convolutional Neural Networks) (Girshick *et al*. 2012) and its successors such as Fast R-CNN (Girshick 2015), Faster R-CNN (Ren *et al*. 2017), Mask R-CNN (He *et al*. 2017), YOLO (You only look once) (Redmon et al. 2015) and SSD (Single shot multibox detector) (Liu et al. 2016). As stated in Section 2.2.1, the vehicle targets can also be selected manually. It all depends on the experimental requirement and application scenarios. If during the time of the experiment there is only one vehicle crossing the bridge, manual selection is good enough to deal with this work. While, if multiple vehicles crossing, pretraining and using deep learning-based vehicle detection algorithms are the more convenient options. In this study, the demonstration is designated for the experiments of UIL extraction and predefined vehicles are selected for the experiments, so that the tracking targets are manually selected from images. Also, using one vehicle to extract UIL faces fewer influencing problems that would happen in multiple vehicle cases. In the real bridge application, automated vehicle detection should be applied to adopt the cases of multiple vehicle crossing bridges.

2.2.4 Visual tracking

Once the vehicle is detected or selected in the first frame of the video or image sequence, visual tracking is necessary to track the location of the vehicle in the successive images. Up to now in the field of computer vision, there are many algorithms for visual tracking and more studies are developed every year (Kristan *et al*. 2018). However not all the algorithms are suitable for the vehicle localization on bridge for load distribution information extraction. As illustrated in Fig. 4(a) and 4(b), due to the camera angle and view depth, the scale of the vehicle and the view changes from the beginning to the end even the camera is stationary. In addition, since this is a truss bridge, the vehicle is occluded during crossing the bridge. The visual tracking algorithm has to satisfy the requirements of scale invariant and view robustness and can predict target location when occlusion happens.

(a) Status of vehicle at the beginning of a truss bridge

(b) Status of vehicle at the end of a truss bridge

Fig. 4 Vehicle tracking example

In this research, the Discriminative Correlation Filter tracker with Channel and Spatial Reliability (CSR-DCF, also called CSRT tracker) (Lukezic *et al*. 2017) is employed to do vehicle visual tracking. CSRT is one of the algorithms using discriminative correlation filter (DCF) which shows great performance. In CSRT, channel and spatial reliability maps are implemented, and a learning progress is applied to update the filter during tracking. This enlarges the search region and improves tracking accuracy of non-rectangular objects. The channel reliability map considers multiple features such as Histogram of Oriented Gradient (HOG), color names and grayscale template to learn and update better filter and spatial reliability map reflect weighting effects in target localization. With the integration of channel and spatial reliability, CSRT achieves state-of-art performance in various popular datasets for visual tracking (Lukezic *et al*. 2017). CSRT satisfies the requirements aforementioned and it is implemented for vehicle tracking in this study.

2.2.5 Coordinate transformation

After the vehicle location (coordinates) in the image is estimated, it needs to be transformed to the real-world coordinates. In this study, the bridge deck is assumed as a plane so that the question is to transform the vehicle from image plane to the deck plane. As shown in Fig. 5, the real-world objects (the bridge and the vehicle) are projected to the image plane. As a result, the shape that is determined by the four points (A, B, C, D) on the real-world plane is distorted due to the projection. According to the work of (Hartley and Zisserman 2003), the projection from the real world plane to the image plane is expressed by the linear transform

$$\mathbf{X} = s\mathbf{H}\mathbf{x} \tag{7}$$

where $\mathbf{X}$ is degraded to $(X, Y, 1)^T$. In this formulation, $\mathbf{H}$ is the 3×3 homography matrix which transforms the real-world plane to the image plane.

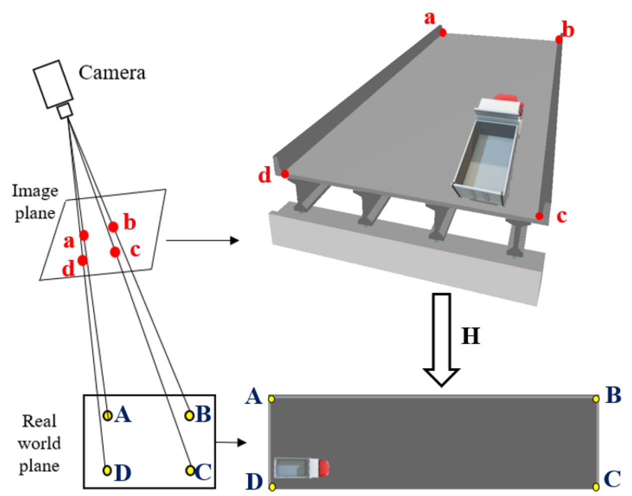


Fig. 5 In-plane transformation using Homography matrix

The scale of the matrix does not affect the equation, so only the eight degrees of freedom corresponding to the ratio of the matrix elements are significant. Eq. (7) can be simplified as

$$\mathbf{X} = \mathbf{H}\mathbf{x} \tag{8}$$

The homography matrix $\mathbf{H}$ has 9 unknowns but only 8 of them are independent. Eq. (8) can be formed by

$$\begin{Bmatrix} X \\ Y \\ 1 \end{Bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{Bmatrix} x \\ y \\ 1 \end{Bmatrix} \tag{9}$$

This matrix is computed directly from image-to-world point correspondences. From Eq. (9) each image-to-world point correspondence provides two linear equations in the $\mathbf{H}$ matrix elements. For n point correspondences, a system of $2n$ equations with 8 unknowns is obtained. This means that at least four point correspondences are needed to solve the problem. If more than four point correspondences are provided, Eq. (9) becomes over-determined and a homogeneous estimation method is implemented to estimate the optimal $\mathbf{H}$ (Dong et al. 2019a). Writing the homography matrix, $\mathbf{H}$ in vector form as, $\mathbf{h} = \{h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, h_9\}^{\mathrm{T}}$, Eq. (9) for n points becomes

$$\mathbf{Ah} = \begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1X_1 & -y_1X_1 & -X_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1Y_1 & -y_1Y_1 & -Y_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2X_2 & -y_2X_2 & -X_2 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2Y_2 & -y_2Y_2 & -Y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & 1 & 0 & 0 & 0 & -x_nX_n & -y_nX_n & -X_n \\ 0 & 0 & 0 & x_n & y_n & 1 & -x_nY_n & -y_nY_n & -Y_n \end{bmatrix} \times \begin{Bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ h_9 \end{Bmatrix} = \mathbf{0} \tag{10}$$

It is a standard result of linear algebra that the vector $\mathbf{h}$ that minimizes the algebraic residuals $|\mathbf{Ah}|$, subject to $|\mathbf{h}|=1$, is given by the eigenvector of least eigenvalue of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$. This eigenvector is obtained directly from the singular value decomposition (SVD) of $\mathbf{A}$. Writing $\mathbf{h}$ back in matrix form, the homography matrix, $\mathbf{H}$ is obtained. The scale s can be calculated by substituting the point correspondences, $\mathbf{X}$, $\mathbf{x}$, and the homography matrix $\mathbf{H}$, into Eq. (7).

At the end, the vehicle location can be transformed from the image coordinates to the real-world coordinates, i.e., the location on the bridge deck.

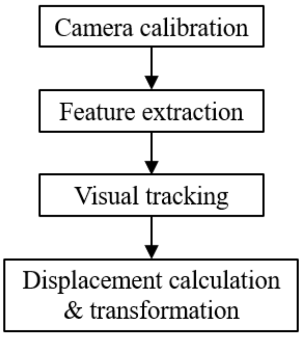


Fig. 6 General procedure of vision-based displacement measurement

2.3 Vision-based structural output estimation: displacement responses

2.3.1 General procedure of vision-based displacement measurement

The structural output estimation carried out in this study is vision-based displacement measurement. Usually there are four steps to estimate displacement from videos of structures using vision-based method (Dong *et al*. 2018a, Dong and Catbas 2019) as shown in Fig. 6. Firstly, camera calibration is done to calculate the geometric relationship between the image coordinates and the real-world coordinates. Secondly, measurement targets are selected from images as images subregions and features in the subregions are extracted for object tracking. Thirdly, visual tracking algorithms are employed combining with the selected features to do object tracking and the locations of measurement targets are updated in consecutive image sequence. At the end, the displacement in image coordinates is calculated by comparing the location change of the measurement targets in each image and the final displacement in real-world coordinates is obtained by combining the camera calibration information and displacement in image coordinates.

The general procedure introduced here is very similar with the vehicle tracking. While in displacement measurement, some different requirements should be addressed. In the following, all of them will be introduced separately.

2.3.2 A practical way of camera calibration for vision-based displacement measurement

For vision-based displacement measurement, the camera calibration is the same with vehicle tracking as introduced in section 2.2.2. Or if the motion of selected location on structures moving in a plane and a lens without distortion is used, the homography matrix can also be applied to simplify the calibration procedure. In this study, the major displacement direction of the selection bridges is vertical so that here a more practical and simplified version of calibration is applied: scale ratio. Scale ratio, *SR* is expressed as the formula when the optical axis is perpendicular to the displacement direction (Dong *et al*. 2018b)

$$SR = \frac{D}{d} \tag{11}$$

where D is the actual dimension of the selected object such as millimetre and d is the dimension of it in image with the unit of pixel. If the optical axis is not perpendicular to the displacement direction, Eq. (11) has to be modified. Detailed discussion can be found in the previous study (Dong *et al.* 2018a). With the scale ratio, the displacement in pixel level can be easily converted to the real-world dimension.

2.3.3 Normalized cross-correlation coefficient using edge map (NCCEM) for template matching

Unlike vehicle tracking, the monitoring target of vision-based displacement measurement is simpler and limited in a specific region. The view and scale do not change too much. Although the tracking for displacement measurement is much easier than that cases in vehicle tracking, it needs much higher accuracy. In general, tracking result should be in sub-pixel level. In previous studies, the authors proposed keypoint matching-based methods (Khuc and Catbas 2016, 2017) and optical flow with keypoints methods (Dong et al. 2018a) to achieve the subpixel level results. All of them showed good measurement results comparing with the conventional displacement sensor. While the processing speed is too slow when using keypoint-based methods. In this study, the normalized cross-correlation coefficient using edge map (NCCEM) for template matching is implemented to obtain the displacement from images. NCCEM is an improved version of digital image correlation (DIC) based template matching methods. The most popular DIC based template method is the normalized cross-correlation coefficient (NCC) method (Chen *et al.* 2018b, Zhong *et al.* 2019a, 2018, 2019b, Zhong and Quan 2018). The NCC coefficient is expressed as

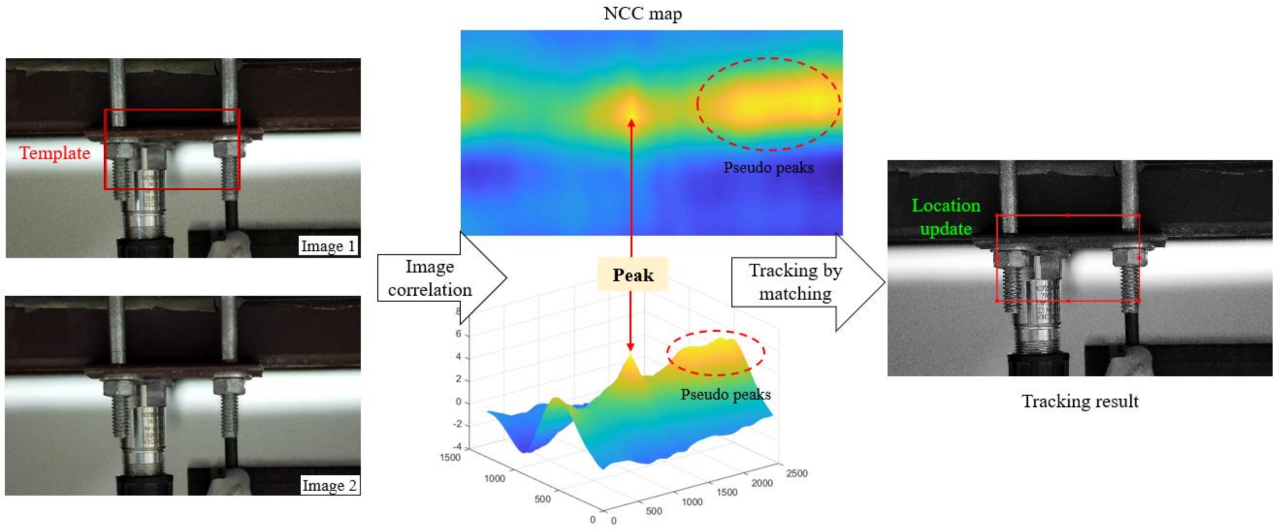
$$R(x,y) = \frac{\sum_{x',y'}\left(T'(x',y')\cdot I'(x+x',y+y')\right)}{\sqrt{\sum_{x',y'}T'(x',y')^2\sum_{x',y'}I'(x+x',y+y')^2}} \tag{12}$$

where

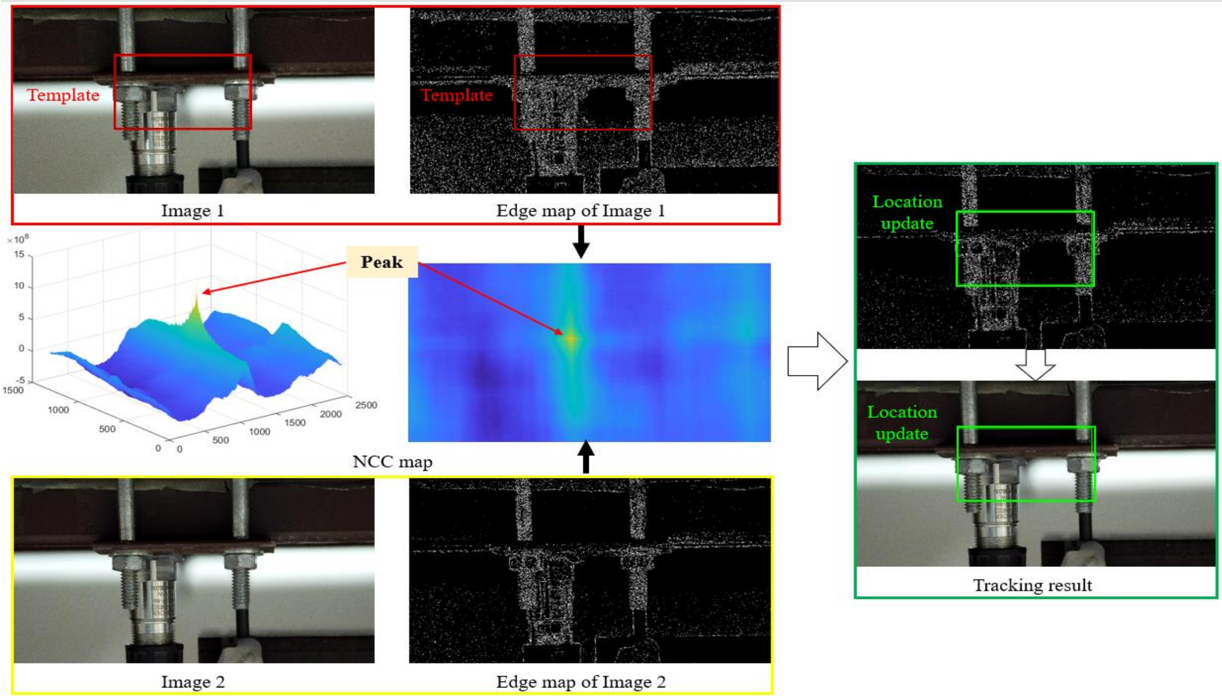
$$T'(x',y') = T(x',y') - \frac{\sum_{x'',y''}T(x'',y'')}{wh} \tag{13}$$

$$I'(x+x',y+y') = I(x+x',y+y') - \frac{\sum_{x'',y''}I(x+x'',y+y'')}{wh} \tag{14}$$

In these formulas above, T is the grayscale image intensity of selected template, I is the grayscale image intensity of the image searching region, (x, y), (x', y') and (x'', y'') represent the location coordinates in image searching regions and template. The NCC coefficient is normalized with image mean value and standard deviation so that it assures the matching result is not affected by the light changing.



(a) Normalized cross-correlation coefficient of grayscale image



(b) Normalized cross-correlation coefficient of Canny edge map

Fig. 7 Digital correlation-based template matching

As shown in Fig. 7(a), when the NCC coefficient achieves the maximum, the matching target is located, i.e., at the peak of the map of NCC coefficient. However, the regular NCC coefficient method is not accurate sometimes. For example, in Fig. 7a, there are some pseudo peaks in the map of NCC coefficient, which make distractions of accurate matching. In this study, the grayscale image is replaced with edge map before template matching using NCC coefficient. The edge map is extracted from grayscale image using Canny operator (Canny 1986). Fig. 7(b) shows the NCC coefficient of Canny edge map, which can be seen that the peak in it is very clear and there is no pseudo peak. With the peak in NCC map of Canny edge map, the matching target is first located in the Canny edge map of the image searching region. Then it is updated in the image searching region. The comparison of NCC maps in Figs. 7(a) and 7(b) indicates that using edge map gives more accurate, reliable and robust results. The NCC methods including regular one and the one using edge map, only give measurement results in pixel level. To achieve the sub-pixel level, a refined searching progress is necessary. In this study, the local pixel upsampling and interpolation operations are applied to do searching refinement. Depends on the required accuracy, a specific iteration number needs to be preset.

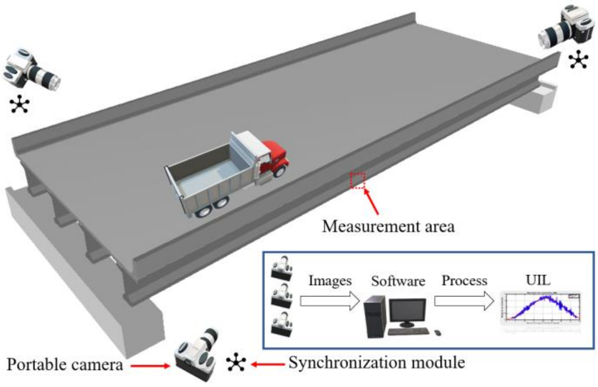


Fig. 8 System configuration

2.3.4 Displacement calculation and transformation

After getting the matching of the target template in consecutive images within sub pixel level, the centre of the template is regarded as the tracking location, (x, y). Assuming the initial location is (x_0, y_0) and with the scale ratio, SR, the displacement of the selected target is $SR\times(x - x_0)$ in horizontal direction and $SR\times(y - y_0)$ in vertical direction.

3. System configuration

The proposed system consists of a set of portable cameras, synchronization modules, a computer and a suite of processing software. Fig. 8 shows the system configuration. The portable cameras are divided into two groups, one is for bridge displacement measurement, and the other is for the vehicle tracking. The synchronization modules are applied to synchronize the image sequences captured from different cameras. All the image sequences are transferred to the computer and processed by the predesignated software. At the end, the UIL is obtained as output.

4. Laboratory demonstration

4.1 Experimental setup

The proposed system is verified on the two-span bridge model (UCF two-span bridge) constructed in the University of Central Florida's Civil Infrastructure Technologies for Resilience and Safety (CITRS) Experimental Design and Monitoring (EDM) laboratory. As shown in Fig. 9, The bridge is a scaled down model of a mid-sized real-life structure and toy trucks with variable weights are used to model moving loads. The bridge consists of two 300-cm main continuous spans. The bridge deck includes a 3.18-mm steel sheet at 120 cm wide, which makes the deck 600 cm long by 120 cm wide.

To view the whole bridge deck and track the vehicle during the whole loading process, a fisheye camera is mounted on the tripod which is 2 m the middle of the bridge. The fisheye camera used here is a Raynic 4K Sports Action Camera with 170-degree wide angle lens. The camera can be connected with a smart phone through the Ez iCam App for remote controlling. This fisheye camera can capture full 1080p (1920 × 1080 pixels) high-density (HD) video clips at a speed of 30 frames per second (30 FPS).
The reason why the fisheye camera is used is that fisheye camera provides a wide angle and can broaden the field of view to assure whole bridge is in the image.

Another portable camera is mounted on the tripod which is close to the midspan of the left span of the bridge to measure the bridge displacement. The distance from the camera to the measurement region, P1, is around 0.8 m. The camera used here is a Z-CAM E1 action camera with a 75-300 mm zoom lens. The camera can also be connected with a smart phone through the Z-CAM official application. The video format set here is 4K (3840 × 2160 pixels) resolution at a speed of 30 FPS. A potentiometer is mounted under the deck to measure the displacement of P1 and is assumed as the ground truth. The model No. of the potentiometer is BEI Duncan 9615. The sampling rate of the data acquisition system for the potentiometer is 200 Hz, which is then downsampled to 30 Hz during post processing. During the experiments, the toy truck moves from one side of the bridge to the other while the potentiometer and the camera record the motion of P1 (midspan of the left span) synchronously.

As shown in Fig. 10, since images captured by the fisheye camera have a severe radical distortion and the straight bridge in the image becomes a curved bridge.

The fisheye camera has to be calibrated. The calibration procedure follows the steps presented in Section 2.2.2 and a white black chess board as shown in Fig. 11 is employed to complete the calibration.

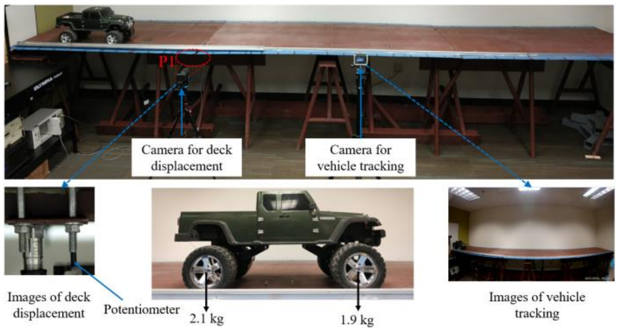


Fig. 9 Experimental setup in laboratory

Fig. 10 Image from fisheye camera

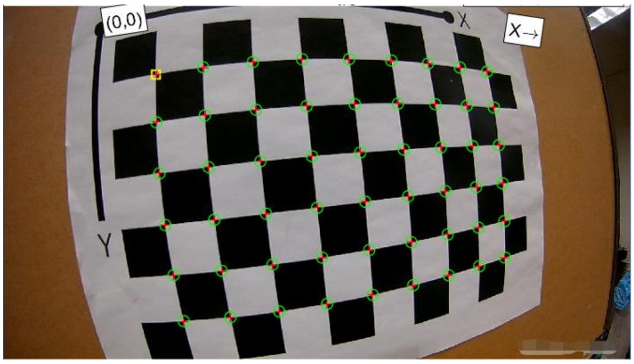


Fig. 11 Camera calibration using a black white chessboard

Fig. 12 Rectified image after camera calibration

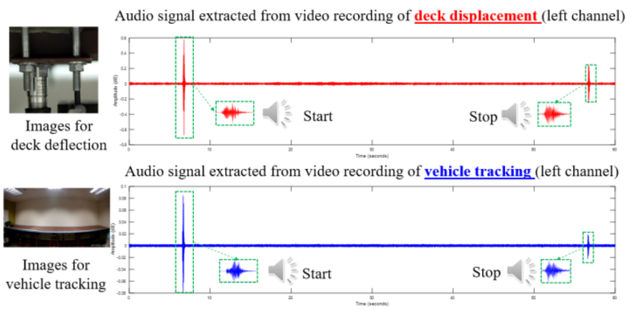


Fig. 13 Video time synchronization using normalized cross-correlation based pattern matching of audio signals

The intrinsic parameters of the fisheye camera are (1) mapping coefficients are [1.03×10^3, -2.49×10^{-4}, -4.89×10^{-7}, 3.55×10^{-10}]; (2) distortion center is [951.95, 577.98]; and (3) stretch matrix is [1, 0; 0, 1]. The extrinsic parameters, i.e., **H** is [-0.100, -1.982, 1107.9; 0.0136, -1.240, 684.14; 1.86×10^{-5}, -0.002, 1]. Fig. 12 shows the rectified image after camera calibration.

In this experiment, the video recorded by the two different portable cameras are synchronized by using cross-correlation based pattern matching of audio signals. During the video recording, the portable cameras also record the audio signals and within the same camera, the images and audio signals are synchronized by the internal clock. As shown in Fig. 13, the audio signals of the two cameras start at different time as the two cameras started recording with different smart phones. During the recording, a voice, “Start”, is called at the beginning of the experiment and a voice, “Stop”, is called at the end of the experiment. The two audio signals are first realigned and synchronized with one dimensional normalized cross-correlation based pattern matching. The pattern is the signal “Start”. Then the signal synchronization is validated by another signal pattern “Stop”. Finally, the two videos are synchronized with the synchronized audio signals.

4.2 Result analysis

Fig. 14 shows the tracking results of the toy truck in the rectified images obtained from the fisheye camera after calibration. During the loading process, although the view and scale of the truck changes, the CSRT tracker can still successfully estimate the location of the toy truck in each image. And eventually the locations in the rectified images are converted to the location on the bridge deck using homography matrix.

Fig. 15 shows the displacement comparison between the proposed vision-based method and the potentiometer. The calibration method for the camera used for displacement measurement is scale ratio and in this experiment, it is 0.0316 mm/pixel. From Fig. 15, it is easy to see that the result obtained from the proposed method is quite consistent with those obtained from the potentiometer. The normalized cross-correlation (NCC) (Dong *et al*. 2019a) is calculated to evaluate the similarities between them. The NCC between the two methods is 99.91%, which shows a very high fit of goodness between the test method (the proposed vision-based method) and the ground truth (potentiometer).

Combining the displacement obtained from the vision-based method and the vehicle location information estimated using vehicle tracking, the UIL is built. In Fig. 16, the blue curve (UIL-raw) is the extracted UIL without any post-processing and filtering. As this bridge displacement is the response under the moving load, it also includes the high vibration modes in the response signal. By applying the Fourier filter, the high vibration modes are removed and the final UIL is shown the red curve (UIL-Fourier).

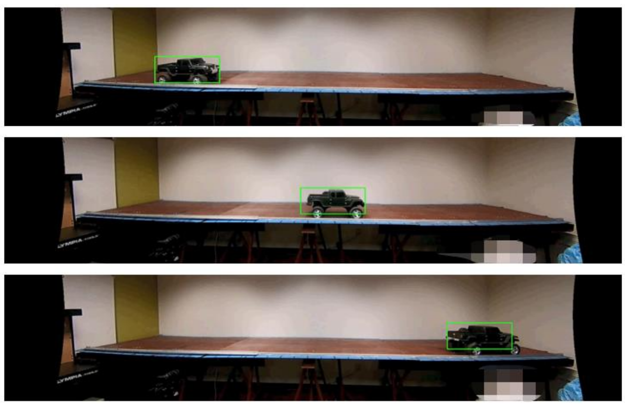


Fig. 14 Vehicle tracking in the rectified images from fish camera

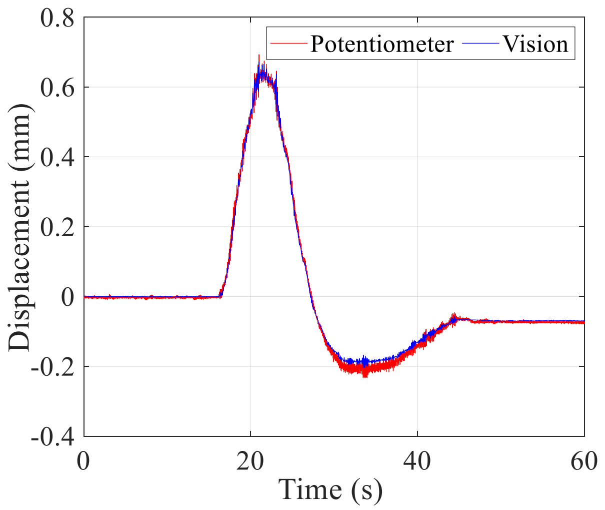


Fig. 15 Displacement comparison between the proposed vision-based method and the potentiometer

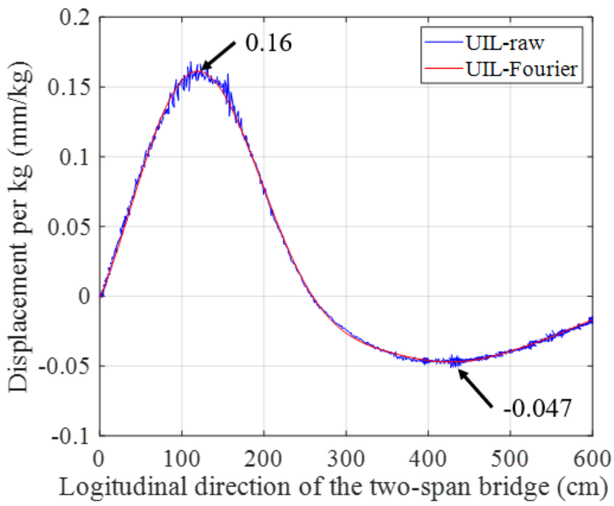


Fig. 16 Extracted UIL using the proposed system

The maximum value of the UIL is 0.16 mm/kg and minimum value is -0.047 mm/kg. Here the downward direction of deck motion is the positive direction and it means the displacement has a positive sign. The negative portion of UIL is obtained when the truck is located on the other span next to the one has the measurement point.

5. Field application

5.1 Experimental setup

A field application is demonstrated on a footbridge under small scale vehicle (golf cart) load as shown in Fig. 17. The bridge comprises of 19.5 m long vertical truss frames that are connected via splice connection in the middle and spans an entire length of 39 m over a pond. The width of the bridge is 4.17 m. The vertical truss members on the left and the right side have HSS 10×10×3/8 top and

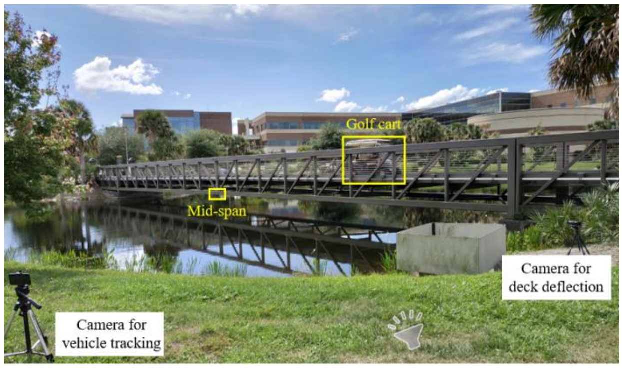


Fig. 17 Experimental setup of a footbridge

bottom chords and are stabilized with HSS 6×4×3/8 type vertical and HSS 4×4×1/4 type diagonal steel members. The lateral stability is provided by another truss frame that is 3.65m wide which is constructed with HSS 3×3×1/4 type diagonal cross braces, W12×22 type lateral members. Two separate spans are spliced in the middle and the entire frame holds a thin layered aluminum-concrete composite deck. The bridge is located at a university campus and is generally under a light human traffic and small-scale vehicles.

The experimental setup is shown in Fig. 17. To track the vehicle during the whole loading process in this experiment, the iPhone XS MAX is employed. The homography transform matrix of the iPhone camera, **H**, is [3.27, 0.15, 289.09; 0.59, 1.23, -398.85; 5.94 × 10^{-4}, 4.75 × 10^{-5}, 1]. The camera used for displacement measurement of the midspan is also Z-CAM E1 camera with a 75-300 mm lens, the same with the one in the laboratory experiment. The scale ratio of this camera is 0.302 mm/pixel. The video format of both cameras are 4K resolution at a speed of 30 FPS. The videos from the two different cameras are also synchronized with the normalized cross-correlation based pattern matching of audio signals introduced in Fig. 13.

A golf cart with three people including the driver drove through the bridge back and forth from one end to the other. The weight of the golf cart is 496.69kg and the weights of the three people are 94.34 kg, 78.47 kg, 75 kg respectively. Before starting the golf cart, there was a group of people coming and crossing the bridge. The cameras also recorded this event.

5.2 Result analysis

Fig. 18 shows the tracking results of the golf cart in the rectified images obtained from the iPhone camera after planar transformation. The original images captured by the iPhone camera are shown in Fig. 4. The scale and view angle of the golf cart changes during the loading process because the iPhone camera is not perpendicular to the longitudinal direction of the footbridge. The images in Fig. 4 are first transformed to the fashion in Fig. 18 using homography matrix. Then the vehicle tracking is performed in the transformed images.

Fig. 18 Vehicle tracking in the planar transformed images: (top) tracking when the vehicle starts from the left end of the footbridge; (middle) tracking when the vehicle is at the midspan; (bottom) tracking when the vehicle arrives the right end

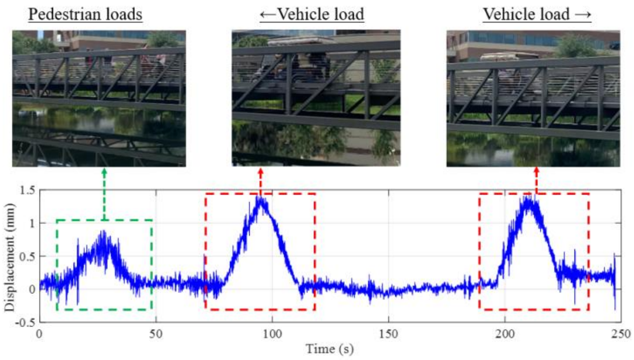


Fig. 19 Displacement of the midspan under different loads

During the loading process, even though the view and scale of the truck changes and the truss part of the footbridge occludes the golf cart, the CSRT tracker can still successfully estimate the location of the golf cart in each image.

The displacement of the midspan during loading process is shown in Fig. 19. The proposed vision-based method successfully recognized the pedestrian loads when a group of people crossed the bridge before the starting of the golf cart. The maximum static displacement response (removing the high vibration modes) of the midspan under the pedestrian load is about 0.6 mm. From the displacement time histories, it also shows the displacement responses when the golf cart crossed the bridge back and forth and both of them are very similar with almost the same maximum static response (removing the high vibration modes), around 1.35 mm. It is reasonable because the weight of the golf cart is constant during the experiment and enough time is spent to let the vibration of the bridge attenuate after the golf cart drove from one end to the other.

In this experiment, the video clips when the golf cart from left end to the right are used to extract the UIL. As shown in Fig. 20, the original UIL is represented with the blue curve and removing the high vibration modes with Fourier filter, the UIL is obtained as represented as the red

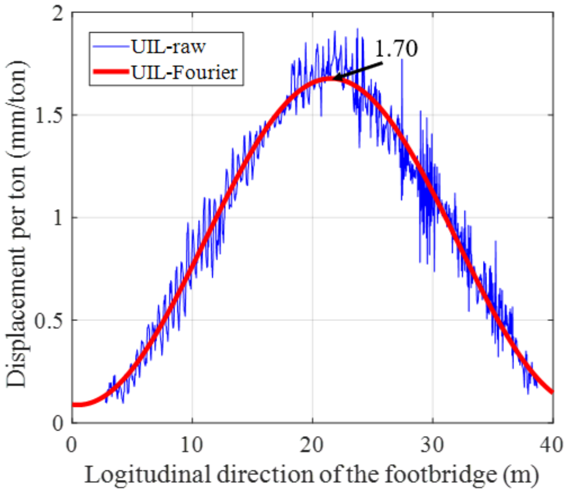


Fig. 20 Extracted UIL of the midspan of the footbridge using the proposed system

curve. The maximum value of the UIL is 1.70 mm/ton. With the UIL and the maximum static displacement of the bridge under pedestrian loads, the maximum static pedestrian loading is calculated, and it is 353 kg. Recalling the pedestrian loading event from the iPhone video, there are six middle-aged female people crossing the bridge. With the predicted total load, 353 kg, the average weight of each pedestrian is around 58.8 kg, which is acceptable.

6. Conclusions

To overcome the inconveniences and disadvantages of the conventional structural health monitoring practices such as high cost, excessive setup time, labor forces with cable wiring work, it would be important and useful to build a structural identification framework with a normalized structural response indicator irrespective of the type and/or the loads for better decision making with a completely non-contact recognition system. In this study, bridge unit influence line (UIL) using only portable cameras and computer vision is proposed. The feasibility of the proposed method is verified through a comparative study of a series of laboratory experiments and a field application. The main approaches, findings, and conclusions are as follows:

- A five-step general procedure for vision-based structural input (vehicle location) recognition is presented. CSRT tracker is implemented to track the vehicle successfully even the scale and view changes and occlusion happens during the visual tracking process.
- To broaden the field of view of camera and to track the vehicle during the whole process, a fisheye camera with wide angle is employed and the full camera calibration is carried out to rectify the radial distortion for accurate vehicle localization.
- A normalized cross-correlation coefficient using edge map (Canny) for template matching is proposed to achieve reliable displacement measurement. The proposed

method avoids the pseudo peaks in NCC map when doing template matching using the traditional NCC based template matching using grayscale images. The displacement results obtained from the proposed method have high consistency with that obtained from conventional displacement sensor with an NCC coefficient of 99.91%.

- The two video recordings from two different portable cameras are successfully synchronized by using the normalized cross-correlation based pattern matching of audio signals.
- The displacement UIL is successfully identified by combining the vehicle location estimated using visual tracking and homography transformation and the displacement record obtained with vision-based method. It makes the whole identification process in a completely non-contact fashion and UIL is extracted in daily traffic flow.
- The extracted displacement UIL is employed for pedestrian load estimation and the predicted weights of pedestrians are observed to be in acceptable ranges. It makes the proposed system work as non-contact weigh-in-motion (WIM) system as presented in this example.

The proposed UIL recognition system also shows great probability to detect damage by using statistical analysis of UILs, bridge load capacity evaluation by regarding UIL as a normalized structural performance indicator and load rating by extracting UILs with the daily traffic flow. The future work will focus on the investigation of them and extend to more possible aspects of structural condition assessment at global level.

Acknowledgments

The financial support for this research was provided by NSF Division of Civil, Mechanical and Manufacturing Innovation [grant number 1463493]. The statements made herein are solely the responsibility of the authors. The authors would like to acknowledge members of the Civil Infrastructure Technologies for Resilience and Safety (CITRS) research group at University of Central Florida for their endless support in creation of this work. The second author would like to kindly acknowledge the Scientific and Technological Research Council of Turkey (TUBITAK) through grant number 2219.

References

Canny, J.F. (1986), "A computational approach to edge detection", *IEEE T. Pattern Anal. Mach. Intell.*, **8**(6), 679-698.

Catbas, N., Zaurin, R., Gul, M. and Gokce, H.B. (2012), "Sensor networks, computer imaging, and unit influence lines for structural health monitoring: case study for bridge load rating",*J. Bridg. Eng.*, **17**(4), 662-670. DOI: 10.1061/(ASCE)BE.1943-5592.0000288.

Chen, S.-E., Liu, W., Bian, H. and Smith, B. (2012), "3D liDAR scans for bridge damage evaluations", *Sixth Congr. Forensic Eng.*, San Francisco, California, United States, 487–495.

Catbas, F.N., Hiasa, S., Dong, C., Pan, Y., Celik, O. and Karaaslan, E. (2017), "Comprehensive Structural Health Monitoring at Local and Global Level with Vision-based Technologies", *26th ASNT Res. Symp.*

Catbas, F.N., Dong, C.Z., Celik, O. and Khuc, T. (2018), "A vision for vision-based technologies for bridge health monitoring", *Maintenance, Safety, Risk, Manag. Life-Cycle Perform. Bridg. Proc. Ninth Int. Conf. Bridg. Maintenance, Saf. Manag. (IABMAS 2018),* Melbourne, Australia, 54.

Celik, O., Dong, C.Z. and Catbas, F.N. (2018a), "A computer vision approach for the load time history estimation of lively individuals and crowds", *Comput. Struct.*, **200**, 32-52. DOI:10.1016/j.compstruc.2018.02.001.

Celik, O., Terrell, T., Gul, M. and Catbas, F.N. (2018b), "Sensor clustering technique for practical structural monitoring and maintenance", *Struct. Monit. Maint.*, **5**(2), 273-295. DOI: 10.12989/smm.2018.5.2.273.

Chen, J.G., Adams, T.M., Sun, H., Bell, E.S. and Büyüköztürk, O. (2018a), "Camera-based vibration measurement of the world war I memorial bidge in portsmouth, new hampshire", *J. Struct. Eng.*, **144**(11), 04018207. DOI:10.1061/(ASCE)ST.1943-541X.0002203.

Chen, Y., Joffre, D. and Avitabile, P. (2018b), "Underwater dynamic response at limited points expanded to full-field strain response", *J. Vib. Acoust.*, **140**(5), 051016. DOI:10.1115/1.4039800.

Chen, Y., Logan, P., Avitabile, P. and Dodson, J. (2019), "Non-model based expansion from limited points to an augmented set of points using chebyshev polynomials", *Exp. Tech.*, 1-23. DOI:10.1007/s40799-018-00300-0.

Dong, C.Z., Ye, X.W. and Liu, T. (2015), "Non-contact structural vibration monitoring under varying environmental conditions", *Vib. Pro.*, **5**, 217–222.

Duan, Y.F., Zhang, R., Dong, C.Z., Luo, Y. Z., Or, S.W., Zhao, Y. and Fan, K.-Q. (2015), "Development of elasto-magneto-Electric (EME) sensor for in-Service cable force monitoring", *Int. J. Struct. Stab. Dyn.*, **16**(04), 1640016. DOI:10.1142/S0219455416400162.

Dong, C.Z., Celik, O. and Catbas, F.N. (2018a), "Marker free monitoring of the grandstand structures and modal identificationusing computer vision methods", *Struct. Heal. Monit.*, 1-19. DOI:10.1177/1475921718806895.

Dong, C.Z., Ye, X.W. and Jin, T. (2018b), "Identification of structural dynamic characteristics based on machine vision technology", *Meas. J. Int. Meas. Confed.*, **126**, 405-416. DOI: 10.1016/j.measurement.2017.09.043.

Dong, C.Z. and Catbas, F.N. (2019), "A non-target structural displacement measurement method using advanced feature matching strategy", *Adv. Struct. Eng.* DOI: 10.1177/1369433219856171.

Dong, C.Z., Celik, O., Catbas, F.N., O'Brien, E.J. and Taylor, S. (2019a), "Structural displacement monitoring using deep learning-based full field optical flow methods", *Struct. Infrastruct. Eng.*, 1-21. DOI: 10.1080/15732479.2019.1650078.

Dong, C.Z., Celik, O., Catbas, F.N., OBrien, E.J. and Taylor, S.E. (2019b), "A robust non-contact displacement measurement method for long-term bridge monitoring using spatio-temporal context learning and taylor approximation", *Sensors*, **19**(14), 3197.

Feng, D., Scarangello, T., Feng, M.Q. and Ye, Q. (2017), "Cable tension force estimate using novel noncontact vision-based sensor", *Meas. J. Int. Meas. Confed.*, **99**, 44-52. DOI:10.1016/j.measurement.2016.12.020.

Girshick, R., Donahue, J., Darrell, T., Berkeley, U.C. and Malik, J. (2012),"(rcnn) Rich feature hierarchies for accurate object detection and semantic segmentation", *2014 IEEE Conf. Comput. Vis. Pattern Recognit.*, 2-9.

Girshick, R. (2015), "Fast r-cnn", *Int. Conf. Comput. Vis.*, 1440-1448. DOI:10.1109/ICCV.2015.169.

Hartley, R. and Zisserman, A. (2003), “Multiple view geometry in computer vision”, Cambridge University Press. DOI:10.1109/ICCV.2009.5459414.

He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2017), “Mask r-cnn”, *IEEE Int. Conf. Comput. Vis.*, 2961-2969. DOI:10.1109/TPAMI.2018.2844175.

Hiasa, S., Necati Catbas, F., Matsumoto, M. and Mitani, K. (2017), “Considerations and issues in the utilization of infrared thermography for concrete bridge inspection at normal driving speeds”, *J. Bridg. Eng.*, **22**(11), 04017101. DOI: 10.1061/(ASCE)BE.1943-5592.0001124.

Hoskere, V., Park, J., Yoon, H. and Spencer Jr., B.F. (2019), “Vision-based modal survey of civil infrastructure using unmanned aerial vehicles”, *J. Struct. Eng.*, **145**(7), 04019062. DOI: 10.1061/(ASCE)ST.1943-541X.0002321.

Khuc, T. and Catbas, F.N. (2016), “Computer vision-based displacement and vibration monitoring without using physical target on structures”,*Struct. Infrastruct. Eng.*, Taylor & Francis, **13**(4), 1-12. DOI:10.1080/15732479.2016.1164729.

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. DOI:10.1002/stc.1852.

Karaaslan, E., Bagci, U. and Catbas, F.N. (2018). “Artificial intelligence assisted infrastructure assessment using mixed reality systems”, *arXiv Prepr.*, arXiv:1812.05659.

Khuc, T. and Catbas, F.N. (2018). “Structural identification using computer vision–based bridge health monitoring”, *J. Struct. Eng.*, **144**(2), 04017202.

Kong, X., Li, J., Collins, W., Bennett, C., Laflamme, S. and Jo, H. (2018), “Sensing distortion-induced fatigue cracks in steel bridges with capacitive skin sensor arrays”, *Smart Mater. Struct.*, **27**(11), 115008. DOI: 10.1088/1361-665X/aadbfb.

Kristan, M., Leonardis, A., Matas, J., Felsberg, M., Pfugfelder, R., Zajc, L.C., Vojir, T., Bhat, G., Lukezic, A., Eldesokey, A., Fernandez, G., *et al*. (2018), “The sixth visual object tracking VOT2018 challenge results”, *VOT2018 Work. ECCV2018*, Munich, Germany.

Lee, J.J., Cho, S. and Shinozuka, M. (2006), “Evaluation of bridge load carrying capacity based on dynamic displacement measurement using real-time image processing techniques”, *Steel Struct.*, **6**, 377-385.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y. and Berg, A.C. (2016). “SSD: single shot multibox detector”, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, **9905 LNCS**, 21-37. DOI:10.1007/978-3-319-46448-0_2.

Lukezic, A., Vojir, T., Zajc, L.C., Matas, J. and Kristan, M. (2017), “Discriminative correlation filter tracker with channel and spatial reliability”, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 6309-6318. DOI:10.1007/s11263-017-1061-3.

Ni, Y.Q., Ye, X.W. and Ko, J.M. (2010), “Monitoring-based fatigue reliability assessment of steel bridges: analytical model and application”, *J. Struct. Eng. - ASCE*, **136**(12), 1563-1573. DOI:10.1061/(ASCE)ST.1943-541X.0000250.

Ni, Y.Q., Ye, X.W. and Ko, J.M. (2011), “Modeling of stress spectrum using long-term monitoring data and finite mixture distributions”, *J. Eng. Mech. - ASCE*, **138**(2), 175-184. DOI: 10.1061/(ASCE)EM.1943-7889.0000313.

Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2015), “You only look once: unified, real-time object detection”, *Proc. 2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR 2016)*, 779-788. DOI:10.1109/CVPR.2016.91.

Ren, S., He, K., Girshick, R. and Sun, J. (2017), “Faster r-cnn: towards real-time object detection with region proposal networks”, *IEEE T. Pattern Anal. Mach. Intell.*, **39**(6), 1137-1149. DOI:10.1109/TPAMI.2016.2577031.

Tian, Y., Zhang, J. and Yu, S. (2018), “Rapid impact testing and system identification of footbridges using particle image velocimetry”, *Comput. Civ. Infrastruct. Eng.*, 1-16. DOI: 10.1111/mice.12390.

Tian, Y., Zhang, J. and Yu, S. (2019), “Vision-based structural scaling factor and flexibility identification through mobile impact testing”, *Mech. Syst. Signal Pr.*, **122**, 387-402. DOI: 10.1016/j.ymssp.2018.12.029.

Xu, Y. and Brownjohn, J.M.W. (2018), “Review of machine-vision based methodologies for displacement measurement in civil structures”, *J. Civ. Struct. Heal. Monit.*, Springer Berlin Heidelberg, **8**(1), 91-110. DOI:10.1007/s13349-017-0261-4.

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

Ye, X.W., Ni, Y.Q., Wong, K.Y. and Ko, J.M. (2013b), “Statistical analysis of stress spectra for fatigue life assessment of steel bridges with structural health monitoring data”, *Eng. Struct.*, **45**, 166-176. DOI: 10.1016/j.engstruct.2012.06.016.

Ye, X.W., Yi, T.-H., Dong, C.Z., Liu, T. and Bai, H. (2015a), “Multi-point displacement monitoring of bridges using a vision-based approach”, *Wind Struct.*, **20**(2), 315-326. http://dx.doi.org/10.12989/was.2015.20.2.315.

Ye, X.W., Yi, T.H., Wen, C. and Su, Y.H. (2015b), “Reliability-based assessment of steel bridge deck using a mesh-insensitive structural stress method”, *Smart Struct. Syst.*, **16**(2), 367-382. http://dx.doi.org/10.12989/sss.2015.16.2.367.

Ye, X.W., Dong, C.Z. and Liu, T. (2016a), “A review of machine vision-based structural health monitoring: methodologies and applications”, *Sensors*, **2016**. DOI: 10.1155/2016/7103039.

Ye, X.W., Dong, C.Z. and Liu, T. (2016b), “Image-based structural dynamic displacement measurement using different multi-object tracking algorithms”, *Smart Struct. Syst.*, **17**(6), 935-956. http://dx.doi.org/10.12989/sss.2016.17.6.935.

Ye, X.W., Dong, C.Z. and Liu, T. (2016c), “Force monitoring of steel cables using vision-based sensing technology: methodology and experimental verification”, *Smart Struct. Syst.*, **18**(3), 585-599. http://dx.doi.org/10.12989/sss.2016.18.3.585.

Ye, X.W., Liu, T., Dong, C.Z. and Chen, B. (2016d), “FBG-based structural performance monitoring and safety evaluation of a composite arch bridge”, *Struct. Heal. Monit. Compos. Struct. Using Fiber Opt. Methods*, CRC Press, 269-297.

Ye, X.W., Su, Y.H., Xi, P.S., Chen, B. and Han, J.P. (2016e). “Statistical analysis and probabilistic modeling of WIM monitoring data of an instrumented arch bridge”, *Smart Struct. Syst.*, **17**(6), 1087-1105. http://dx.doi.org/10.12989/sss.2016.17.6.1087.

Ye, X.W., Yi, T.H., Dong, C.Z. and Liu, T. (2016f), “Vision-based structural displacement measurement: system performance evaluation and influence factor analysis”, *Meas. J. Int. Meas. Confed.*, **88**, 372-384. DOI: 10.1016/j.measurement.2016.01.024.

Yang, Y., Dorn, C., Mancini, T., Talken, Z., Nagarajaiah, S., Kenyon, G., Farrar, C. and Mascarenas, D. (2017), “Blind identification of full-field vibration modes of output-only structures from uniformly-sampled, possibly temporally-aliased (sub-Nyquist), video measurements”, *J. Sound Vib.*, **390**, 232-256. DOI:10.1016/j.jsv.2016.11.034.

Ye, X.W., Yi, T.H., Su, Y.H., Liu, T. and Chen, B. (2017), “Strain-based structural condition assessment of an instrumented arch bridge using FBG monitoring data”, *Smart Struct. Syst.*, **20**(2), 139-150. http://dx.doi.org/10.12989/sss.2017.20.2.139.

Zhang, Z. (2002). “A flexible new technique for camera calibration (technical report)”, *IEEE T. Pattern Anal. Mach. Intel.*, **22**(11), 1330-1334.

Zaurin, R., and Catbas, F. (2010a), “Structural health monitoring using video stream, influence lines, and statistical analysis”, *Struct. Health. Monit.*, **10**(3), 309-332. DOI: 10.1177/1475921710373290.

Zaurin, R. and Catbas, F. (2010b), “Integration of computer imaging and sensor data for structural health monitoring of bridges”, *Smart Mater. Struct.*, **19**, 1-15. DOI:10.1088/0964-1726/19/1/015019.

Zaurin, R., Khuc, T. and Catbas, F.N. (2015), “Hybrid sensor-camera monitoring for damage detection: case study of a real bridge”, *J. Bridg. Eng.*, **21**(6), 1-27. DOI: 10.1061/(ASCE)BE.1943-5592.0000811.

Zhong, F.Q., Indurkar, P.P. and Quan, C.G. (2018). “Three-dimensional digital image correlation with improved efficiency and accuracy”, *Meas. J. Int. Meas. Confed.*, **128**, 23-33. DOI:10.1016/j.measurement.2018.06.022.

Zhong, F., Kumar, R. and Quan, C. (2019a), “A cost-effective single-shot structured light system for 3D shape measurement”, *IEEE Sens. J.*, IEEE, **19**(17), 7335-7346. DOI: 10.1109/JSEN.2019.2915986.

Zhong, F. and Quan, C. (2018), “Efficient digital image correlation using gradient orientation”, *Opt. Laser Technol.*, **106**, 417-426. DOI:10.1016/j.optlastec.2018.04.024.

Zhong, F.Q., Shao, X.X. and Quan, C. (2019b), “A comparative study of 3D reconstruction methods in stereo digital image correlation”, *Opt. Lasers Eng.*, **122**, 142-150.