Smart Structures and Systems, *Vol. 29, No. 5 (2022) 715-728*
https://doi.org/10.12989/sss.2022.29.5.715

Vision-based Input-Output System identification for pedestrian suspension bridges

Jeonghyeok Lim [a] and Hyungchul Yoon*

Department of Civil Engineering, Chungbuk National University,
1 Chungdae-ro, Seowon-gu, Cheongju-si, Chungcheongbuk-do, Republic of Korea

(*Received June 15, 2021, Revised January 25, 2022, Accepted February 25, 2022*)

Abstract. Recently, numbers of long span pedestrian suspension bridges have been constructed worldwide. While recent tragedies regarding pedestrian suspension bridges have shown how these bridges can wreak havoc on the society, there are no specific guidelines for construction standards nor safety inspections yet. Therefore, a structural health monitoring system that could help ensure the safety of pedestrian suspension bridges are needed. System identification is one of the popular applications for structural health monitoring method, which estimates the dynamic system. Most of the system identification methods for bridges are currently adapting output-only system identification method, which assumes the dynamic load to be a white noise due to the difficulty of measuring the dynamic load. In the case of pedestrian suspension bridges, the pedestrian load is within specific frequency range, resulting in large errors when using the output-only system identification method. Therefore, this study aims to develop a system identification method for pedestrian suspension bridges considering both input and output of the dynamic system. This study estimates the location and the magnitude of the pedestrian load, as well as the dynamic response of the pedestrian bridges by utilizing artificial intelligence and computer vision techniques. A simulation-based validation test was conducted to verify the performance of the proposed system. The proposed method is expected to improve the accuracy and the efficiency of the current inspection and monitoring systems for pedestrian suspension bridges.

Keywords: computer vision; deep learning; pedestrian suspension bridge; structural health monitoring; system identification

1. Introduction

Recently, numbers of long span pedestrian suspension bridges have been constructed around the world. Mishima Skywalk (total length of 400 m) was constructed in Japan in 2015, Tianmeng bridge (total length of 505 m) was constructed in China in 2016, Charles Kuonen Suspension Bridge (total length of 494 m) was constructed in Switzerland in 2017, Sogeum Bridge (total length of 200 m) was constructed in Korea in 2018, SkyBridge (total length of 207m) was constructed in USA in 2019, and 516 Arouca Bridge (total length of 516m) was constructed in Portugal in 2020. Furthermore, Topjeong Reservoir Bridge (total length of 600m) will be constructed in Korea in 2021.

Recent report by the Board of Audit and Inspection of Korea indicated that maintenance for the pedestrian suspension bridges in Korea are not being conducted properly. Corrosions were detected in Mangho and Jeodu bridge, cable fastening problem was detected in Cheonjangho bridge, and bolt loosening problem was detected in Yeonhwa bridge (The Board of Audit and Inspection of Korea 2018). Furthermore, recent tragedies around the world showed how the safety of the pedestrian suspension bridge can wreak havoc on the society. In 2015, the Waikaremoana Lake Hopu Ruahine bridge in New Zealand overturned, causing four injuries (Cooke 2015). In 2016, Yellow Bridge near Bali, Indonesia, collapsed, killing eight people (Mortimer 2016). In 2017, a large crowd gathered at the pedestrian suspension bridge on the Carrizal River in Tosagua, Ecuador, causing an overturning accident that caused 32 people to fall (Mills 2017). In addition, in 2020, in Bengkulu, Sumatra Island in western Indonesia, a pedestrian suspension bridge collapsed, killing nine people (Widianto *et al.* 2020).

Despite these accidents, there are no specific design standards or safety standards for the pedestrian suspension bridge in Korea. Pedestrian suspension bridges are not categorized in road bridges nor pedestrian bridges. In addition, separate design standards, safety standards, nor legal definitions for the pedestrian suspension bridge are not yet prepared. Therefore, it is necessary to develop a system that can ensure the safety and evaluate the deterioration of the pedestrian suspension bridge.

One of the popular methods being used for diagnosing the safety of structures is vibration-based structural health monitoring (SHM) system. Vibration-based SHM ensures the integrity of a structure based on the vibration response of the structure such as displacement or acceleration. SHM systems include multiple applications such as finite element model updating, damage detection, and system

*Corresponding author, Associate Professor,
E-mail: hyoon@cbnu.ac.kr
[a] Master of Civil Engineering
E-mail: lim0827@g.cbnu.ac.kr

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

identification. Finite element model updating is a process that updates the design stage finite element model into as-is current condition that reflects the deterioration (Oh *et al.* 2017). Damage detection is a method to estimate the location and the of size of the damage (i.e., reduction in the stiffness) in the structure (Eftekhar Azam *et al.* 2019). In addition, system identification is also one of the popular applications of SHM, which predicts the mathematical system information of a structure, including dynamic characteristics of the structure, by measuring the structural response. System identification can be categorized into output-only system identification and input-output system identification; output-only system identification uses the response of the structure to estimate the structural system, whereas the input-output system identification considers not only the response of the structure, but also the dynamic load of the structure. In most cases, the output-only method has been applied for monitoring bridges by assuming the input load to be a white noise (Altunişik *et al.* 2011). This assumption has been applied not because it is more accurate, but because of the difficulty of measuring the input load being applied to the structure. For most cases, the assumption of idealizing the input load to a white noise is reasonable, especially when the input load is within a large range of frequency.

However, assuming the dynamic load to be a white noise may reduce the accuracy, especially when applied to pedestrian suspension bridges. The dominant dynamic load in pedestrian suspension bridge is the pedestrian load which is within a specific frequency area of 1.8 to 2.2 Hz (Kim *et al.* 2015). The accuracy of the output-only system identification, which assumes the dynamic load to be a white noise, will significantly reduce due to the improper assumption for dynamic load. Therefore, input-output based system identification considering both the response of the structure and the dynamic load applied to the structure should be performed to improve the accuracy of system identification.

Various methods for measuring structural response have been introduced by researchers. The traditional method for measuring the structural response was to use wired sensors including displacement transducers, accelerometers, and/or strain gauges (Kang *et al.* 2006, Kim and Feng 2007, Li *et al.* 2015). However, wired sensor-based sensing systems have problems such as consuming huge costs for cable installation and maintenance (Kim *et al.* 2016). To overcome the limitation of the wired sensor system, a wireless sensor-based monitoring system was developed. Kim *et al.* (2007) applied the wireless sensor network-based health monitoring technology to real structures and confirmed the cost reduction effect. Cho *et al.* (2008) developed an automated wireless sensor-based tension estimation system for suspension and cable-stayed bridges. However, there are still problems such as traffic control to install sensors and loss in data transmission process (Song *et al.* 2008). Recently, more convenient methods such as using computer vision techniques are being introduced. Yoon *et al.* (2016) introduced a smartphone-based tracking system for structural system identification. Park *et al.* (2014) proposed a displacement measurement method using CCTV for structural health monitoring. Yoon *et al.* (2018) enabled the measurement by using the camera installed in unmanned aerial vehicles.

On the other hand, only a few studies were conducted for estimating the dynamic load of the structure. Tradition method for estimating the dynamic load is to use an impact hammer. Jo *et al.* (2009) proposed a method to measure the dynamic load using an impact hammer and a transfer function. Recently, new technologies such as Weigh In Motion (WIM) are being utilized to obtain more information related to the in-service load. Kwon and Suh (2009) developed an unmanned, non-stop high-speed load measurement system using WIM. Qin *et al.* (2018) conducted a study to estimate the static weight of a vehicle while moving using a multi-sensor WIM system. However, WIM systems cannot estimate the exact dynamic load of moving objects, and also are difficult to be applied at pedestrian suspension bridges. Therefore, a new method for dynamic load estimation specialized for pedestrian suspension bridge is needed.

Computer vision and deep learning techniques can provide an opportunity to measure the in-service load of the pedestrian suspension bridges. Velardo and Dugelay (2012) proposed the method that estimate human weight using Kinect RGBD sensor, Jiang *et al.* (2020) studied a method of estimating human weight by extracting a skeleton from a 2D image. Recently Lim and Yoon (2019) introduced a method for finding the location of the pedestrian using a commercial camera. However, studies to estimate the size and location of pedestrian loads on the pedestrian suspension bridge have not been conducted yet.

Therefore, this study aims to develop an input-output system identification method for pedestrian suspension bridges using computer vision and deep learning. The proposed method is comprised of three sub-systems: (1) pedestrian load estimation, (2) structural response measurement, (3) input-out system identification. Pedestrian load estimation predicts the location and the weight of pedestrian loads using computer vision and deep learning, and structural response measurement measures the displacement of pedestrian suspension bridges using commercial cameras. Finally, input-output system identification is performed using the two information: pedestrian load and the structural response. To validate the system, a simulation-based experiment using a lumped-mass pedestrian bridge model was conducted. The adequacy of a data acquisition system can only be measured in terms of how the data is going to be used, and thus Modal Assurance Criteria (MAC) was used to assess the efficacy of the proposed method.

2. System development

2.1 System overview

Fig. 1 shows the overview of the proposed system developed in this research. The proposed system consists of three phases: vision-based pedestrian load estimation, vision-based structural displacement measurement, and

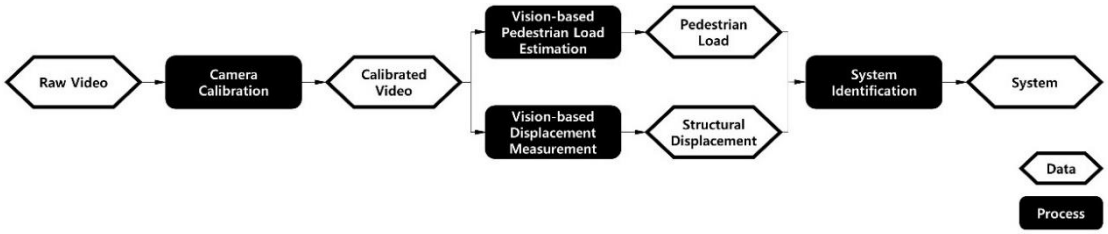


Fig. 1 System overview

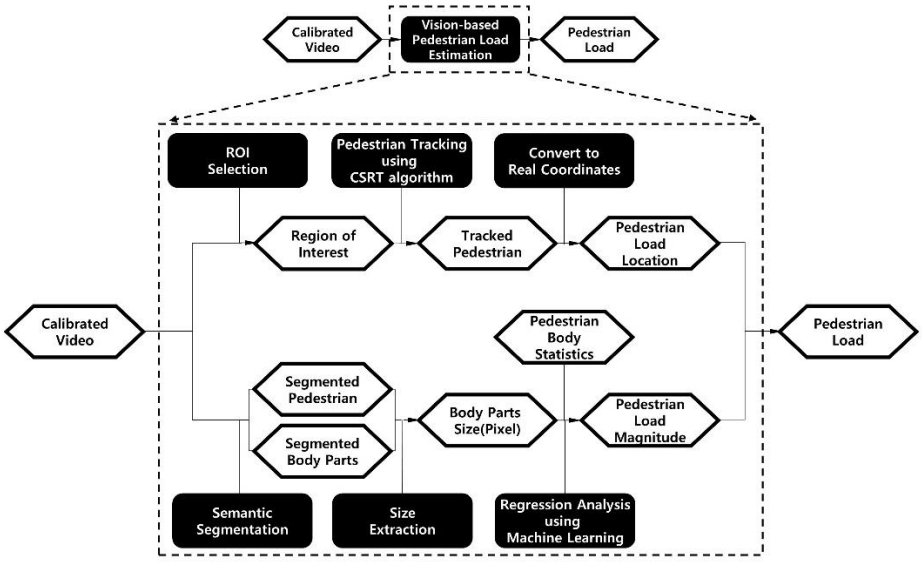


Fig. 2 Flowchart for Phase 1: vision-based pedestrian load estimation

system identification. Phase 1 is vision-based pedestrian load estimation, where the location and magnitude of the pedestrian load is estimated by using computer vision and deep learning techniques, Phase 2 is structural displacement measurement, where the structural response due to the pedestrian load is measured using computer vision techniques. Finally, the two information (i.e. pedestrian load and structural response) are combined in Phase 3 to estimate the dynamic system for the pedestrian suspension bridge.

2.2 Phase 1: Vision-based pedestrian load estimation

Fig. 2 shows the detailed process for Phase 1. There are two main objectives for Phase 1: (1) to estimate the location of the pedestrian load, and (2) to estimate the magnitude of the pedestrian load. Object tracking techniques such as CSRT (Channel Spatial Reliability-Discriminative Correlation Filter Tracker) was used for estimating the location of the pedestrian load. The location of the pedestrian is then transformed from the image coordinate to the world coordinate. Next, a semantic segmentation is applied to classify the body parts of the pedestrian from the pedestrian images. Once the body part is classified from pedestrian images, the weight of each pedestrian is estimated by using machine learning techniques. The detailed information for each process is described in the following sections.

2.2.1 Estimation of pedestrian load location

The proposed method utilizes the video of a pedestrian crossing the pedestrian suspension bridge to estimate the location of the pedestrian. The proposed method for estimating the pedestrian load location is composed of 3 steps: region of interest selection, object tracking, and coordinate transformation.

The first step for tracking the pedestrian is to select the region of interest (ROI) in the first frame of the video. The ROI box is implemented in GUI for users to select over the area of pedestrian. The ROI box should be selected to cover the entire body, including the arms and legs of the pedestrian to be tracked. The reduction of tracking accuracy due to the movement of the body part of the pedestrian can be decreased.

Next step is to track the pedestrian using CSRT algorithm (Lukezic *et al*. 2017). Once the ROI is selected, features such as HOG (Histogram of Oriented Gradients, (Dalal and Triggs 2005) and Color names (Van De Weijer *et al*. 2009) are extracted from the selected ROI, and a discriminative correlation filter is constructed using extracted features. To optimize the filter, the spatial reliability map that identifies pixels that are likely to contain tracking objects is used. The spatial reliability map is estimated using the output of a graph labeling problem solved efficiently in each frame. The Channel reliability is utilized for use in filter response weights for each channel. The Channel reliability is estimated using the properties of the constrained least squares solution to the filter design.

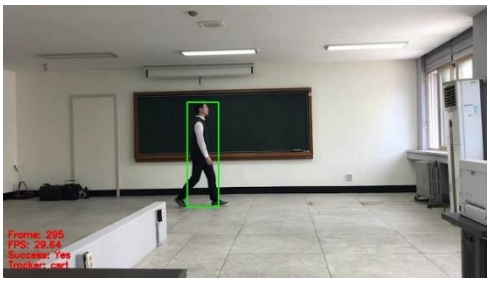


Fig. 3 A screenshot for pedestrian tracking using CSRT algorithm

The location of the tracking object is estimated by summing the features extracted in the previous time step and the correlation response weighted by the estimated channel reliability. In this process, the filter response for each channel is used to calculate the detection reliability value. The location of the pedestrian can be estimated in pixels by using the coordinates of the center of the ROI estimated by the CSRT algorithm. Fig. 3 shows the pedestrian tracking process through the CSRT algorithm.

The pedestrian load location obtained from the previous step is in the image coordinate system with units of pixels. The pedestrian location should be transformed into the world coordinate to be used in system identification. In this study, a coordinate transformation method suggested by Lim and Yoon (2019) was adopted to convert the pedestrian location in image coordinate to the world coordinate. The scale factor of pixel-length was calculated by measuring the length of a known object in pixels, and all the points in the image coordinate system was converted to the world coordinate.

2.2.2 Estimation of pedestrian weight

The proposed method utilizes the video of a pedestrian crossing the pedestrian suspension bridge to estimate the pedestrian weight as well. Because the weight of the pedestrian cannot be directly obtained, the proposed method estimates the weight of the pedestrian inferring to the visual information and by applying deep learning and computer vision techniques. This study assumed that the visual information will be obtained from the side when the pedestrians cross over the pedestrian suspension bridges. The proposed method for estimating the pedestrian weight is composed of 4 steps: pedestrian segmentation, body part segmentation, body measurement, and weight estimation. First, the pixels containing the pedestrian information is separated from the video frame, and next the pedestrian image is partitioned into multiple human body parts. Once the human body part has been segmented, the body measurements can be estimated for each body parts. Finally, the weight of the pedestrian can be estimated by inferring to the body measurement.

The first step is to find the pixels that contains the pedestrian information from the video frame using a semantic segmentation algorithm. Semantic segmentation is a process partitioning an image into multiple segments of pixels and provide semantic information to each segment. In this study, Pedestrian Segmentation Network (PSN) has been developed through transfer learning by applying initialized weights based on ResNet-50 (He *et al.* 2016) and Deeplab-V3+ (Chen *et al.* 2017). The structure of PSN is composed of an encoder and a decoder as shown in Fig. 4.

It consists of a structure in which the input image is dimensionally lowered as it passes through the convolutional layer of the encoder, learns features for the image, and recovers to the original dimension through the decoder. The Deeplab-V3+ used in this study improved its performance by using Atrous Convolution, which has space inside the filter during encoding, as opposed to the general segmentation network structure. The classification layer is modified to segment the pixels for pedestrian from the other pixels. 14,400 images from the 3D People Dataset (Pumarola *et al.* 2019) was used to train the Pedestrian Segmentation Network. 3D People Dataset is a dataset containing over 2 million image data for 72 human motions as shown in Fig. 5. In this study, the images of “walking” and “running”, which are the most representative actions that pedestrians can take on the pedestrian suspension bridge, were selected as the training data.

Next step is to partition the human body into multiple segments that represents the human body parts. In this study, Body-part Segmentation Network (BSN) has been developed through transfer learning by applying initialized weights based on ResNet-50 and Deeplab-V3+ as well. The

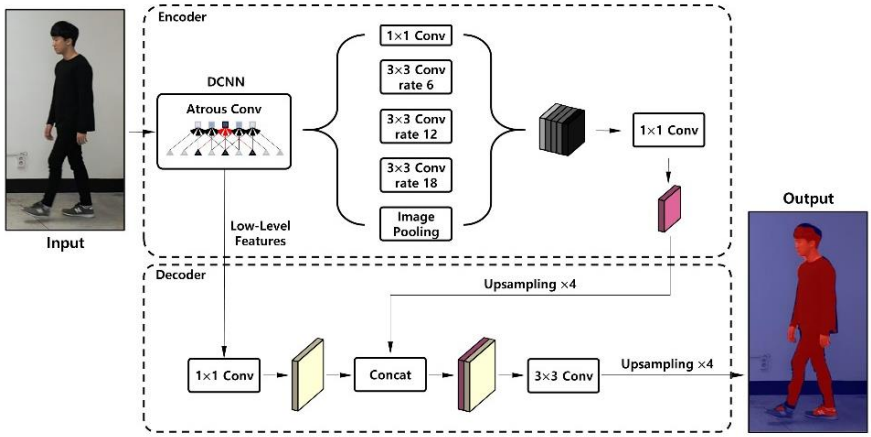
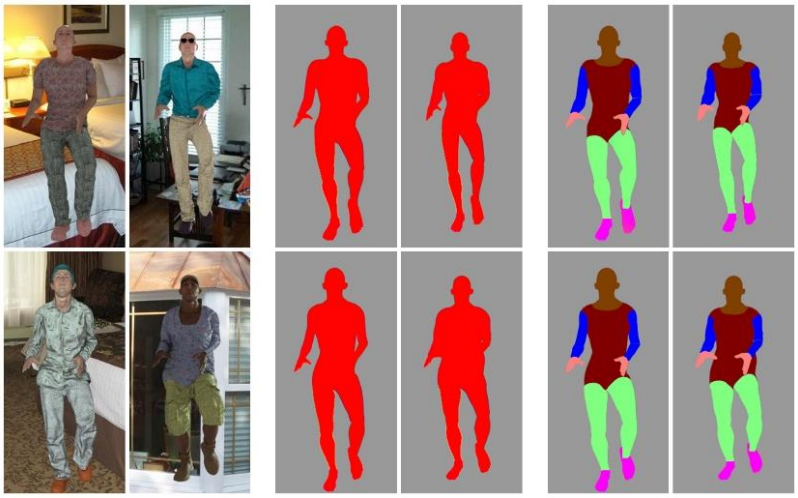


Fig. 4 Architecture for the proposed Pedestrian Segmentation Network (PSN)



(*a*) (*b*) (*c*)

Fig. 5 3D People Dataset (*a*) raw image for 3D people dataset, (*b*) labeled image for PSN, (*c*) labeled image for BSN

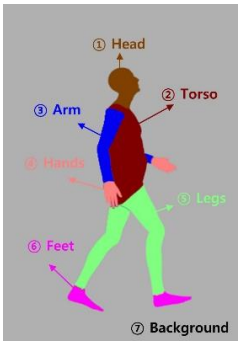
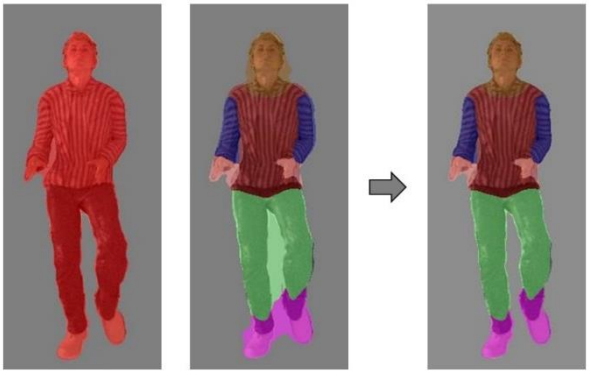


Fig. 6 Seven classes used for BSN: Head, Torso, Arms, Hands, Legs, Feet and Backgrounds



(*a*) (*b*) (*c*)

Fig. 7 An example of pedestrian segmentation result; (*a*) PSN, (*b*) BSN, (*c*) union of PSN and BSN

structure of BSN is like that of PSN, but the classification layer is modified to partition the pixels into 7 human body parts as shown in Fig. 6: head, torso, arm, hand, leg, feet and background.

14,400 images from the 3D People Dataset were also used to train the BSN as well. The proposed system combines the PSN and the BSN by taking the union of each result to improve the segmentation accuracy. The accuracy of the semantic segmentation generally decreases by thenumber of classes, and thus the proposed system takes the unions of the results from two networks as illustrated in Fig. 7, instead of using a single network.

Once the human body is partitioned from the video frame, the proposed system estimates the body measurement for each body part. Through the previous segmentation step, all pixels were classified into 7 classes. Body measurements are made using the coordinates of the pixels in each class in the segmented image. For the estimation of pedestrian load through the correlation between body size and weight, the size of each body part was extracted from the divided pedestrian image such as Fig. 8. To minimize the error in the body area size extraction process, the extraction process was automated. The body part size changes according to the movement of pedestrians, so the body part size was extracted from all frames of the video, and the accuracy was improved by removing the outlier.

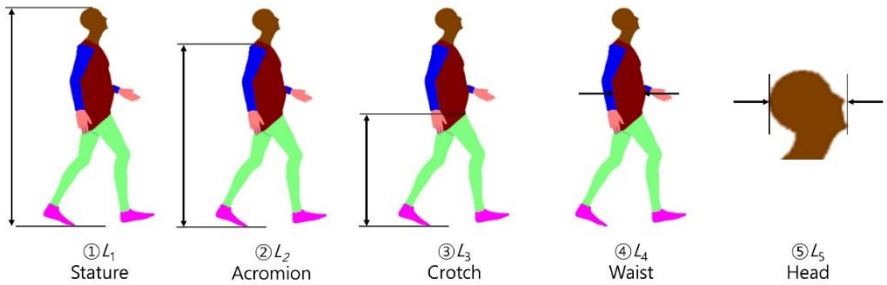


Fig. 8 Features for Gaussian Process Regression in weight estimation

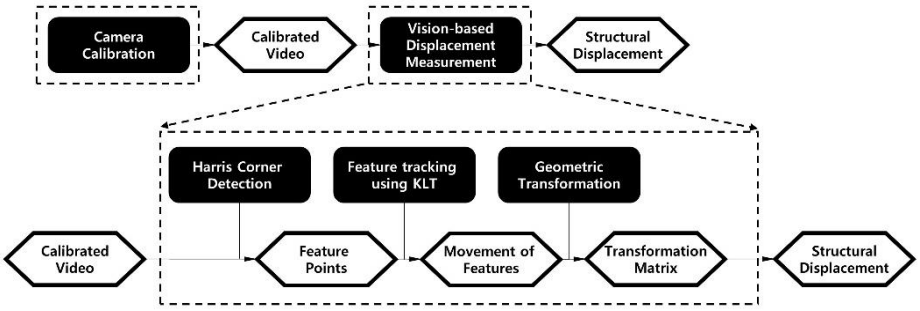


Fig. 9 Flowchart for Phase 2: vision-based structural displacement measurement

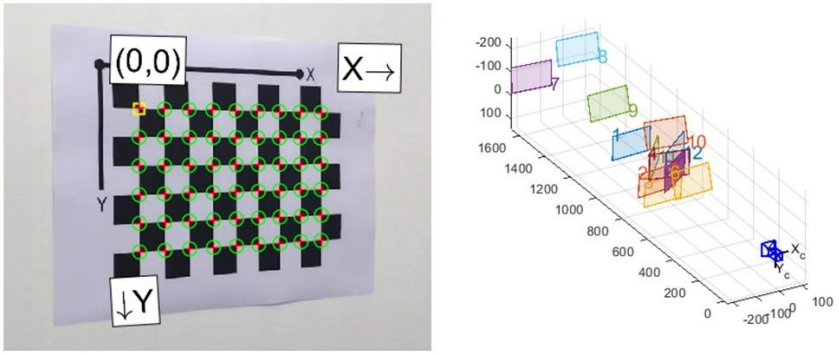


Fig. 10 An example of camera calibration process

Finally, the weight of the pedestrian can be estimated from the body measurements. The human body measurements have strong correlation with the weight, and therefore the proposed method utilized various Machine Learning algorithms to estimate the human body weight. 5 numbers of features were selected to estimate the weight of the pedestrian as shown in Fig. 8, and total of 3,190 training data were collected from the human body size survey project conducted by KOSIS (Korean Statistical Information Service 2018). The data set collected by KOSIS data includes information on gender, age, weight and 120 body parts for 6,140 men and women. Among them, 3190 men data were used as a training data of regression models to obtain information of body parts from pedestrian images. The units for the body dimensions of the training data collected by KOSIS is in 'mm', whereas the units for estimated body dimensions from PSN and BSN is in 'pixel'. Therefore, the proposed method normalized the units, dividing the 4 features by the stature.

The performance of 19 different Machine Learning algorithms including Linear Regression, Decision Tree, Support Vector Machine (SVM) and Gaussian Process Regression (GPR) compared, and "Matern 5/2 GPR" algorithm which showed the highest accuracy was adopted in the proposed method. Using the adopted algorithm and normalized training data, the regression model was trained to output the pedestrian weight when the body measurement was input.

2.3 Phase 2: Vision-based structural displacement measurement

Fig. 9 shows the detailed process for Phase 2. There are two main objectives for Phase 2: to calibrate the camera and measure the structural displacement. The camera calibration process was performed prior to phase 1 and phase 2, and was inserted into phase 2 for convenience of explanation. Camera calibration was performed using the principle of a pinhole camera. The movement of the structure was tracked using the Harris corner detection algorithm and the KLT feature tracking algorithm. Next, the geometric transform was used to measure the structural displacement. The detailed information for each process is described in the following sections. Phase 2 was conducted in reference to Yoon *et al.* (2016)'s study, "Target-free approach for vision-based structural system identification using consumer-grade cameras".

2.3.1 Camera calibration

Distortion can occur in the image due to the imperfections of lens. Such distortion becomes a problem when accurate visual information is needed. When converting points from the image coordinate into the world coordinate, the location of the points may contain large error depending on the degree of distortion. Therefore, camera calibration is essential for vision-based displacement measurements.

The image captured by the camera is obtained by projecting points in a three-dimensional space onto a two-dimensional plane. In the image processing field, all geometrical analysis of images is made based on a pinhole camera model, and a general pinhole camera model can be expressed as Eq. (1) follows.

$$sp = K[R\ t]P, \qquad K = \begin{bmatrix} f_x & skew & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

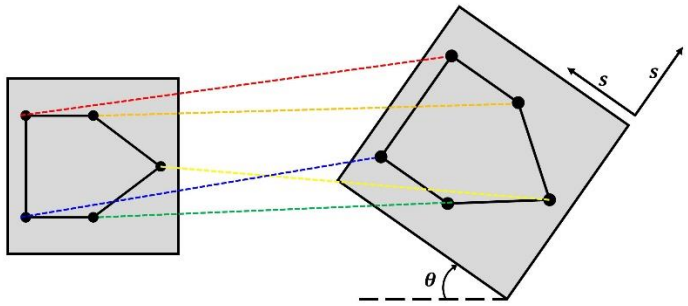


Fig. 11 Similarity transformation

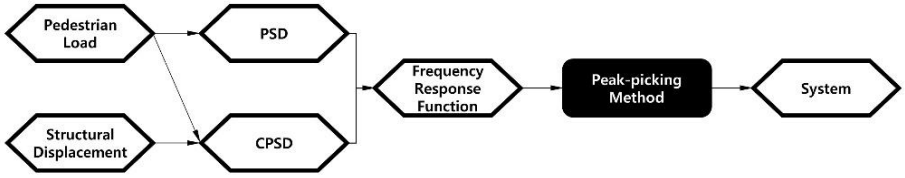


Fig. 12 Flowchart for Phase 3: input-output system identification

K is an Intrinsic Matrix containing information about the internal parameters of the camera, and R and t are rotation matrix and translation vector for converting the world coordinate to the camera coordinate, respectively. Also, p and P are the image coordinate and the world coordinate, respectively. Intrinsic matrix containing camera internal parameter information can be obtained through Camera calibration by installing a Checkerboard to be used as a reference point as shown in Fig. 10, and using images taken from various fields. The distorted image can be corrected using the acquired intrinsic matrix.

2.3.2 Displacement measurement

In this study, feature detection algorithm was used to find the feature point which is the object to be measured in the image. The feature point should be able to be easily found even if the shape, size, or location of the object changes in the image, or the viewpoint or lighting of the camera changes. There are various algorithms to find feature points, but in this study, the Harris corner detection algorithm (Harris and Stephens 1988) was used. Harris corner shows good performance for image translation, rotation and affine transformation, and lighting change. The KLT feature tracker algorithm (Lucas and Kanade 1981. Tomasi and Kanade 1991) is applied to the acquired feature points to calculate the movement of feature.

The displacement of the structure can be obtained by using the feature point detection algorithm and the movement of the feature point obtained using the KLT feature tracker. The movement of the feature point between two consecutive frames can be expressed as Eq. (2).

$$X_i' = HX_i \tag{2}$$

The correlation between the feature point X_i of the current frame and the feature point X_i' obtained by tracking the change in the next frame is expressed as a transformation matrix H. The images taken by the real structure are expressed in the 2D image coordinate system, where the matching relationship between the two images (the previous frame and next frame) can be modeled using a transformation matrix.

In this study, the goal is to measure the displacement of the structure, assuming that the shape of the structure does not change significantly when the structure moves, and thus the "similarity transformation" was used. Since the feature point X_i of the previous frame and the feature point X_i' of the next frame are known, a matrix H representing the relationship between the two points can be obtained. As shown in Fig. 11, the displacement of the structure can be obtained by re-applying the transformation matrix H obtained using the change of features to the ROI box.

2.4 Phase 3: System identification

Once the pedestrian load and the displacement of a pedestrian suspension bridge are obtained, the next step is to analyze the dynamic characteristics of the pedestrian suspension bridge by using system identification methods. As discussed in the introduction, system identification methods predict the mathematical system representing physical structures. In this study, peak-picking method, one of the most popular method for modal analysis, was used to analyze the dynamic system of pedestrian suspension bridges, which was applied to both input-output method and output-only method for validation. With measured input (i.e., pedestrian load) and output (i.e., dynamic response), peak picking method can analyze the dynamic characteristics of the pedestrian suspension bridges.

Fig. 12 illustrates the overview of the procedure for the phase 3. First, power spectral density (PSD) $S_{ff}(\omega)$ and cross-power spectral density (CPSD) $S_{xf}(\omega)$ are obtained based on the pedestrian load and the displacement of the structure. Next, the frequency response function is obtained by dividing the CPSD with the PSD. The frequency response function of the system can be obtained by combining this two information as the Eq. (3). Finally, the

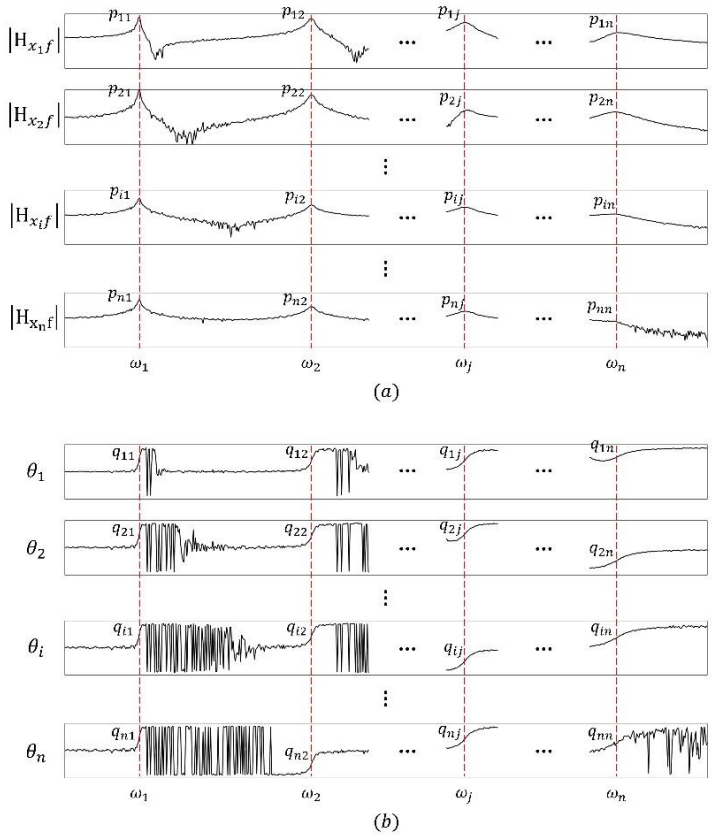


Fig. 13 Peak points located on top of frequency response functions: (*a*) amplitude (*b*) phase

dynamic characteristics of the structure are acquired using the peak-picking method.

$$\mathrm{H}_{xf}(\omega) = \frac{\mathrm{S}_{xf}(\omega)}{\mathrm{S}_{ff}(\omega)} \tag{3}$$

The Peak-picking method predicts the natural frequency and mode shape by extracting the frequency response function (FRF) of system. FRF is the quantitative measurement of the magnitude and phase change of the output spectrum of a dynamic system to input. In single-degree-of-freedom (SDOF) system, FRF can be expressed as Eq. (4).

$$\mathrm{H}_{xf}(\omega) = \frac{p}{-m\omega^2 + ic\omega + k} \tag{4}$$

,where p, m, ω, c, i and k respectively refer to the magnitude of the input, the mass of the structure, the forcing frequency, the damping coefficient of structure, the imaginary number satisfying $i^2 = -1$, and the stiffness of the structure. The amplitude of a FRF $\left|\mathrm{H}_{xf}(\omega)\right|$, and the phase of a FRF $\theta(\omega)$ can be obtained by using the real and imaginary parts of FRF, which can be expressed as Eqs. (5)-(6). $Real$ represents the real parts of FRF, and $Imag$ represents the imaginary part of FRF.

$$\text{Amplitude of FRF}: \left|\mathrm{H}_{xf}(\omega)\right| = \sqrt{Imag^2 + Real^2} \tag{5}$$

$$\text{Phase of FRF}: \theta(\omega) = \tan^{-1}\left(\frac{Imag}{Real}\right) \tag{6}$$

The natural frequency ω_n of system can be obtained by picking the peak point from the FRF plot, where the denominator is minimized in Eq. (4). Also, the phase of the FRF drops by 180 degrees at the location of the natural frequency. However, if the system identification method is used to estimate the dynamic system, the FRF uses the estimated value (Eq. (3)) in the dynamic response.

In multi-degree-of-freedom (MDOF) system, FRF can be expressed for each degree of freedom. The amplitude and phase plots of the frequency response function of a multi-degree of freedom (MDOF) system consisting of n degrees of freedom are illustrated in Fig. 13. Similarly, peak points can be picked from each FRF plots and the natural frequency and phase can be obtained. The mode shape can be obtained using the ratio of phases at the same natural frequency for each FRF. In the MDOF system consisting of n DOF, n frequency response functions ($\mathrm{H}_{x_1f}, \mathrm{H}_{x_2f}, \cdots, \mathrm{H}_{x_if}, \cdots, \mathrm{H}_{x_nf}$) can be obtained by assuming controllability and observability to be equal to the number of DOF. As in Fig. 13, the natural frequency for each mode ω_j $(j = 1, 2, \cdots, n)$ can be estimated by picking the peak points p_{ij} from the amplitudes of FRFs. p_{ij} denotes the jth peak point of in the ith FRF, and the x-coordinate component of the p_{ij} is ω_j, and the y-coordinate component is $\left|\mathrm{H}_{x_if}(\omega_j)\right|$.

The natural frequency for each mode ω_j can be also estimated from the phase of FRFs. q_{ij} denotes the jth point of phase in the ith FRF, and the *x*-coordinate

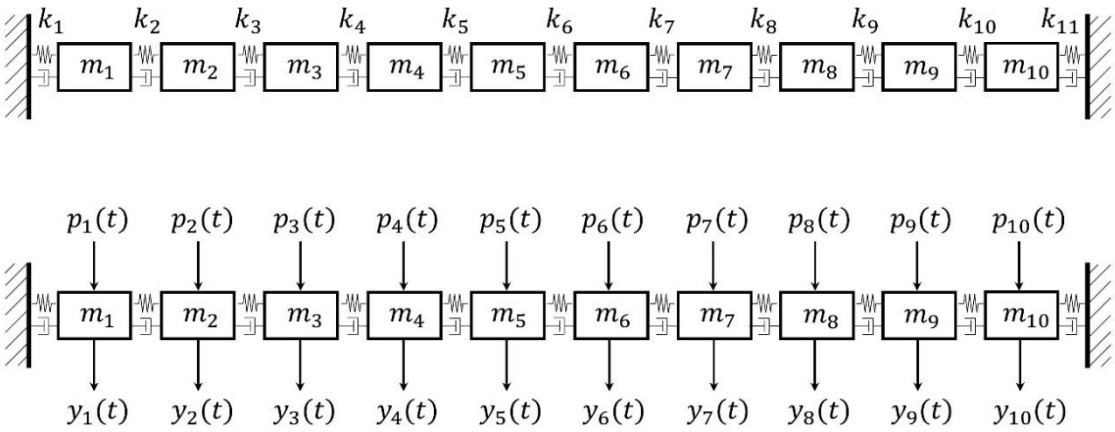


Fig. 14 10 DOF lumped-mass pedestrian suspension bridge model

component of q_{ij} is ω_j, and the *y*-coordinate component is $\theta_i(\omega_j)$. The modal matrix φ_{ij} can be calculated using the relative magnitude of $\left|\mathrm{H}_{x_if}(\omega_j)\right|$ and the relative phase difference $\theta_i(\omega_j)$ as shown in Eqs. (7)-(8).

$$\varphi_{ij} = \begin{bmatrix} \alpha_{11}p_{11} & \alpha_{12}p_{12} & \cdots & \alpha_{1j}p_{1j} & \cdots & \alpha_{1n}p_{1n} \\ \alpha_{21}p_{21} & \alpha_{22}p_{22} & \cdots & \alpha_{2j}p_{2j} & \cdots & \alpha_{2n}p_{2n} \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ \alpha_{i1}p_{i1} & \alpha_{i1}p_{i2} & \cdots & \alpha_{ij}p_{ij} & \cdots & \alpha_{in}p_{in} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1}p_{n1} & \alpha_{n2}p_{n2} & \cdots & \alpha_{nj}p_{nj} & \cdots & \alpha_{nn}p_{nn} \end{bmatrix} \quad (7)$$

where α_{ij} is a value representing the sign of the modal matrix of the ith mode and the jth DOF, and can be simplified by classifying to in-phase and out-of-phase as shown in Eq. (8).

$$\alpha_{ij} = \begin{cases} 1, if \ \left|\theta_i(\omega_j) - \theta_{ref}(\omega_j)\right| = 0 \ or \ 360° \\ -1, if \ \left|\theta_i(\omega_j) - \theta_{ref}(\omega_j)\right| = 180° \end{cases} \quad (8)$$

3. Validation

3.1 Validation setup

To validate the performance of proposed method as a system, a simulation-based validation test was conducted by using a numerical model for a pedestrian suspension bridge. The performance of the proposed method is evaluated by comparing the estimated dynamic characteristics such as natural frequency, mode shape, MAC value with the reference values of the numerical model. The proposed method uses visual information of a pedestrian bridge to estimate the dynamic characteristics. Therefore, a lumped-mass model for a pedestrian suspension bridge was constructed and an animation was generated to verify the performance of the proposed method. The lumped-mass model is a 10-DOF system as show in Fig. 14. Masses (m_1~m_{10}) in each degree of freedom was selected as 2.5 kg each, the spring stiffness for k_1, k_{11} were 3000 N/m, and k_2~k_9 were selected as 2000 N/m. The damping ratio (ζ) of 5% was selected for each mode.

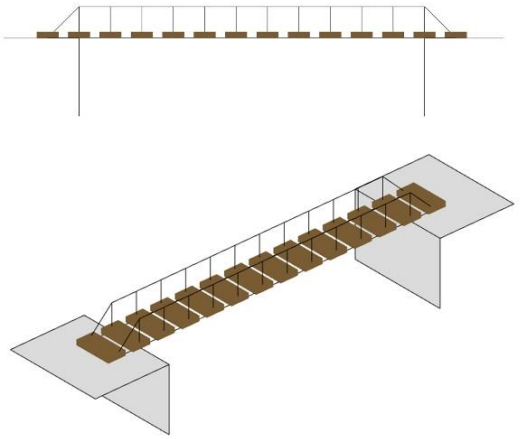


Fig. 15 Animation for the 10 DOF lumped-mass model

The response of the lumped mass model was then visualized with an animation. The animation contains the information of both the dynamic load (i.e., pedestrian load) and the dynamic response of the pedestrian bridge. To visualize the response of the pedestrian bridge, each mass was represented as a plate as shown in Fig. 15. Once a pedestrian cross over the bridge, the response of each degree of freedom was calculated from the lumped-mass model, and animated accordingly.

To visualize the pedestrian load, the images from 3D people dataset was used. Fig. 16 shows the following processes. First, the background was removed from the pedestrian image of the 3D people dataset. Because the pedestrian image without the background has a different scale with the visualized bridge model, the images were converted into a point cloud. Next, the pedestrian image was imported into the bridge model and 3D coordinates were assigned. These steps were repeated for a series of images that contains the movement of the pedestrian. Finally, the location of the dynamic load applied to the bridge model is assigned to each image to synchronize the movement of pedestrian with the location of the load. Through the previous process, the animated video for pedestrian suspension bridge simulation was generated as shown in Fig. 17, with a resolution of 1920×1000p and a frame rate of 100fps.

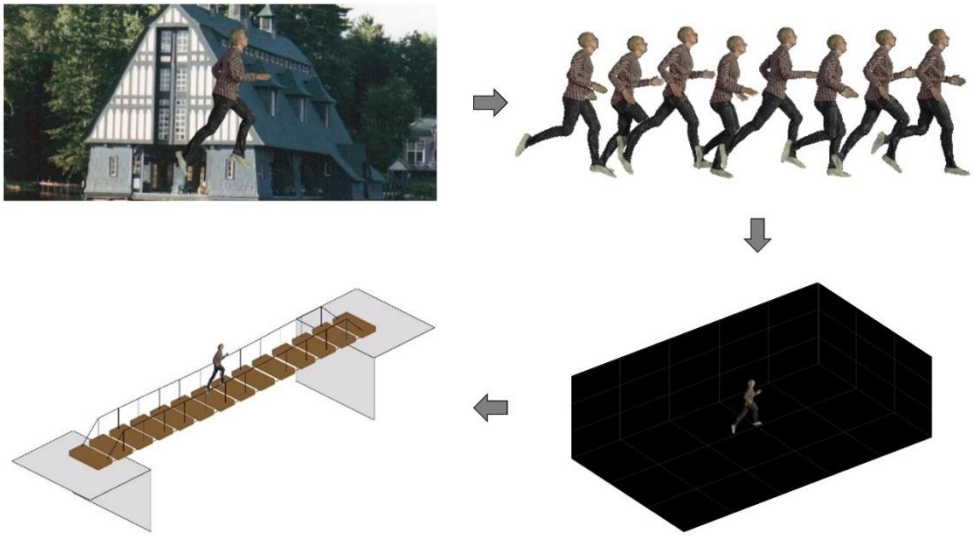


Fig. 16 Procedure for combining pedestrian with the bridge model

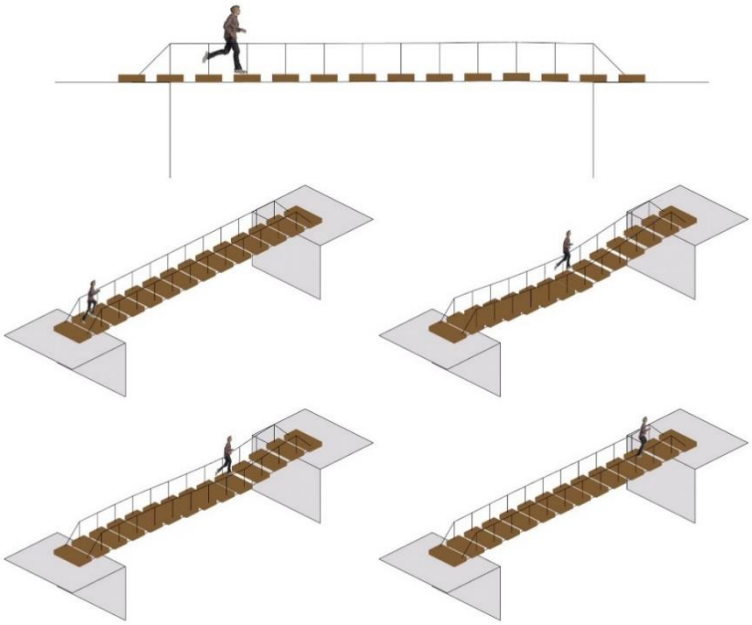


Fig. 17 Animated video for the validation test

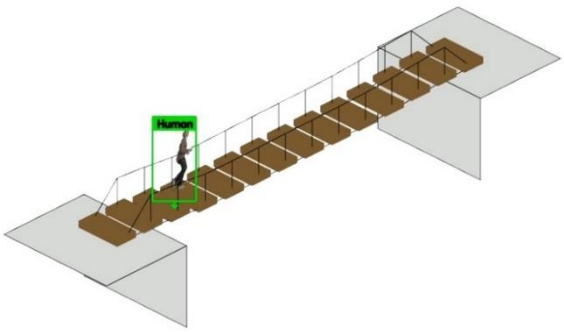


Fig. 18 Example result for pedestrian load estimation

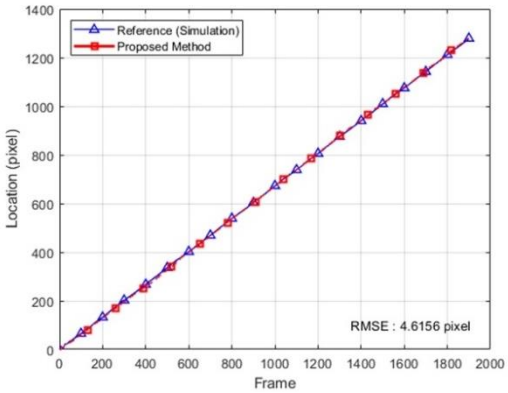
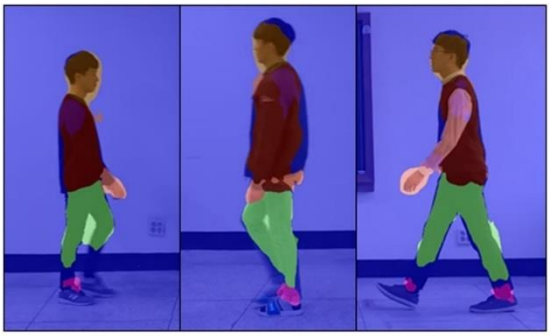


Fig. 19 Estimated pedestrian location using CSRT

3.2 Pedestrian load estimation result

To validate the performance of the pedestrian load estimation, the location and the weight of a pedestrian was estimated from the animated video and compared with reference values. The location of the pedestrian was



(a) (b) (c)

Fig. 20 Pedestrian segmentation result using a pedestrian video

estimated by using the CSRT based tracking algorithm, as shown in Fig. 18. The location of pedestrian recorded in the validation test setup step and the location estimated through the proposed method were compared. As a result, the proposed method was able to estimate the pedestrian with RMSE of 4.62 pixel as shown in Fig. 19.

Next, the weight of the pedestrian was estimated by using the proposed Pedestrian Segmentation Network (PSN), Body-part Segmentation Network (BSN) and the Matern 5/2 GPR model. Because 3D People Data set did not contain information of weight, a pedestrian walking video was recorded as shown in Fig. 20. Table 1 shows the results of estimating the weight using the pedestrian video. The weight estimation errors were 3.06%, 16.92%, and 8.57%, respectively.

Table 1 Weight estimation result

	Feature ① L_2/L_1	Feature ② L_3/L_1	Feature ③ L_4/L_1	Feature ④ L_5/L_1	Weight(kg) Ref	Weight(kg) Proposed method	Error (%)
A	0.7124	0.4198	0.1295	0.1415	63	64.925	3.06%
B	0.7650	0.4018	0.1461	0.1218	72	84.179	16.92%
C	0.7658	0.4064	0.1497	0.1283	80	86.857	8.57%

3.3 Structural displacement result

To validate the performance of the vision-based structural displacement measurement process, the measured response from the animation was compared with the actual response of the numerical model. As shown in Fig. 21, ROI was designated for each plate in the animated video, and the structural displacement was measure by applying the

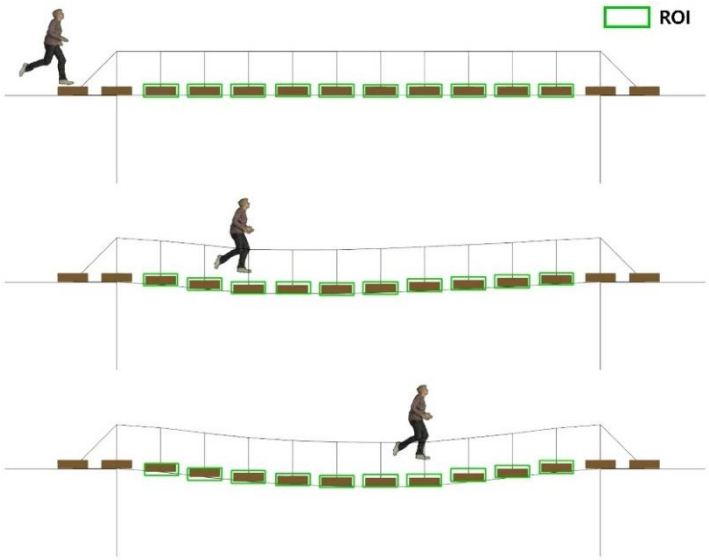
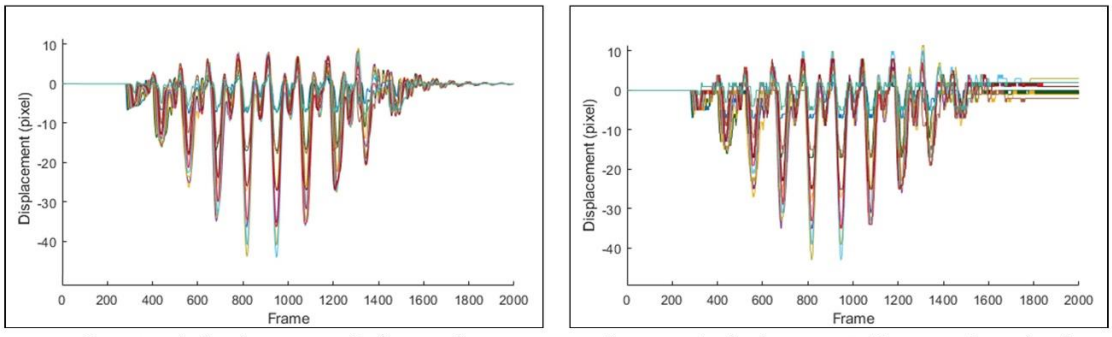


Fig. 21 Structural displacement measurement using KLT algorithm



Structural displacement (Reference) Structural displacement (Proposed method)

Fig. 22 Displacement measurement results

Table 2 Estimated natural frequency

Natural frequency (Hz)	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6	Mode 7	Mode 8	Mode 9	Mode 10
Reference	1.3630	2.6928	3.9571	5.1254	6.1698	7.0665	7.7970	8.3500	8.7255	8.9364
Proposed method	1.3672	2.6367	3.8086	5.0781	6.1523	7.1289	7.9102	8.6914	9.2773	10.4492
Error (%)	0.3100	2.0800	3.7500	0.9200	0.2800	0.8800	1.4500	4.0900	6.3200	16.9300

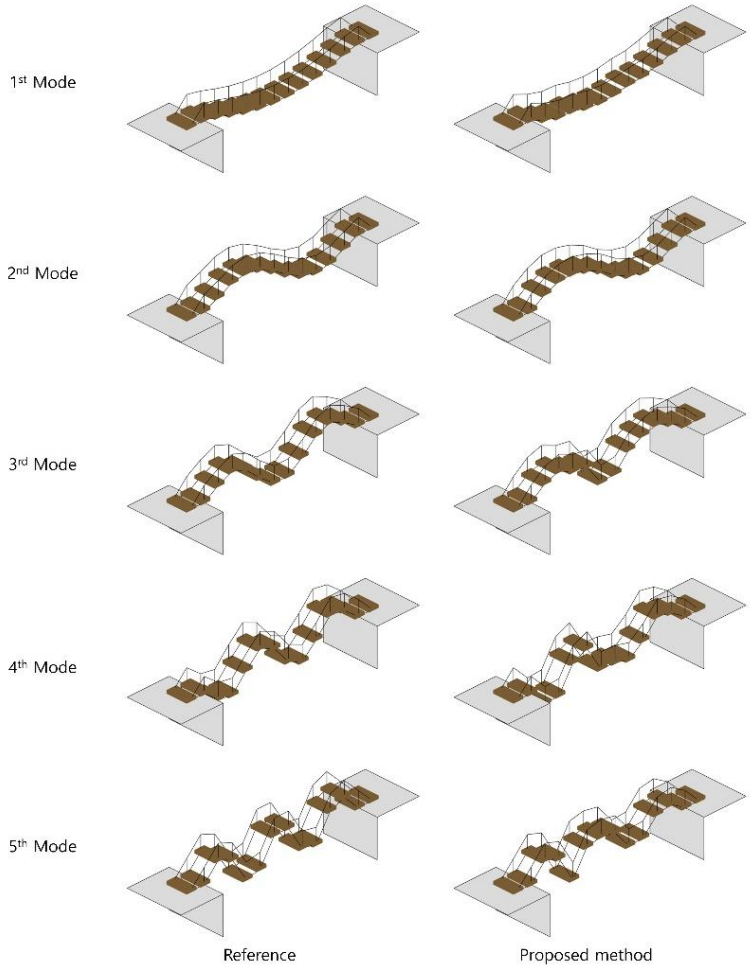


Fig. 23 Estimated mode shapes

proposed method. The proposed method was able to measure the response of the bridge with RMSE of 7.2 pixel as shown in Fig. 22.

3.4 System identification result

The input-output system identification method was applied based on the pedestrian load and the structure displacement information obtained through the previous steps. First, the frequency response function (FRF) was calculated from the estimated input load and measured response. Next, the peak-picking method was applied to estimate the natural frequencies and the mode shape of the pedestrian bridge. The comparison for the estimated natural frequencies with the theoretical natural frequencies are shown in Table 2. The proposed method was able to estimate all 10 natural frequencies with averaged error of 3.701%, while the output-only peak-picking method were unable to estimate any of these modes.

The estimated mode shape compared to the theoretical mode shape is illustrated in Fig. 23. As shown in the figure, the mode shape up to 5th mode showed similar shape with the reference mode shape. To evaluate the similarity between the mode shapes quantitatively, the MAC (Modal Assurance Criterion) value (Pastor *et al*. 2012) was used. MAC value is a statistical indicator for correlation between reference modal vector and compatible analytical modal vector. MAC value is close to 0, representing no consistent correspondence, and If MAC value is close to 1, representing a consistent correspondence. MAC value can be obtained as Eq. (9) below

$$\mathrm{MAC}(i,j) = \frac{\left|\{\varphi_A\}_i^T\{\varphi_X\}_j\right|^2}{\left(\{\varphi_A\}_i^T\{\varphi_A\}_i\right)\left(\{\varphi_X\}_j^T\{\varphi_X\}_j\right)} \tag{9}$$

where $\{\varphi_A\}_i$ is compatible modal vector (mode i), $\{\varphi_X\}_j$ is reference modal vector (mode j), $\{\varphi_A\}_i^T$ is transpose of $\{\varphi_A\}_i$, $\{\varphi_X\}_j^T$ is transpose of $\{\varphi_X\}_j$. Table 3 shows the

Table 3 MAC value for the validation test

Mode	1	2	3	4	5	6	7	8	9	10	
1	0.9833	0.0048	0.0026	0.001	0.0037	0.0043	0.0000	0.0001	0.0000	0.0003	
2	0.0000	0.9894	0.0029	0.0022	0.0027	0.0016	0.0007	0.0000	0.0001	0.0004	
3	0.0438	0.0022	0.8761	0.0005	0.0108	0.0030	0.0465	0.0022	0.0119	0.0030	
4	0.0336	0.0004	0.0329	0.7226	0.0906	0.0762	0.0015	0.0298	0.0009	0.0116	
5	0.0437	0.0388	0.0005	0.0171	0.6940	0.0790	0.0443	0.0042	0.0294	0.0490	Reference
6	0.0673	0.0029	0.0347	0.0295	0.0185	0.6817	0.0656	0.0130	0.0867	0.0001	
7	0.2954	0.2322	0.1911	0.0243	0.0037	0.0699	0.1043	0.0385	0.0209	0.0197	
8	0.0501	0.0359	0.0001	0.0125	0.0053	0.0771	0.1133	0.3257	0.2960	0.0839	
9	0.0025	0.0062	0.0104	0.0509	0.0612	0.0835	0.2076	0.2555	0.2409	0.0812	
10	0.0451	0.0154	0.0129	0.0374	0.0816	0.1369	0.1786	0.3299	0.1014	0.0608	

MAC values between the modal vectors estimated in the proposed method and the reference modal vectors. The result showed 0.9833 for the first mode, 0.9894 for the second mode, and 0.8761 for the third mode. While the MAC values decreased with higher modes, the proposed was able to estimate all 10 modes shape.

4. Conclusions

In this study, we proposed a new method to identify the dynamic characteristics of a pedestrian suspension bridge by using deep learning and computer vision techniques. First, the location of a pedestrian load was tracked in real-time using CSRT algorithm. Next, to estimate the weight of a pedestrian load, the pedestrian image was segmented into multiple body parts using Deeplab v3+ based semantic segmentation network. The weight of the pedestrian was then estimated from the body measurements using Gaussian Process Regression (GPR). To measure the dynamic displacement of a pedestrian bridge, an optical flow based KLT algorithm developed by authors previous work was applied. Finally, input-output system identification method was applied to predict the dynamic characteristics of the pedestrian suspension bridge by combining the information on dynamic load and dynamic response.

To validate the proposed method, a simulation-based experiment test was conducted. An animated video for a pedestrian suspension bridge was generated. The validation result showed that the proposed method was able to estimate the dynamic characteristics of the pedestrian suspension bridge while the output only method could not.

The proposed method is expected to provide an efficient and powerful tool for monitoring pedestrian suspension bridges. Because the proposed method only requires the video of a pedestrian suspension bridge, the method can be utilized easily with a drone system or a CCTV camera system. To improve the accuracy, detail modelling for pedestrian load including impact load and the interaction between the pedestrian with the bridge is planned as a future work.

Acknowledgments

This research was supported by National Research Foundation of Korea (NRF) funded by the Ministry of Education(2019R1I1A3A01044827)

References

Altunişik, A.C., Bayraktar, A. and Sevim, B. (2011), "Output-only system identification of posttensioned segmental concrete highway bridges", *J. Bridge Eng.*, **16**(2), 259-266. https://doi.org/10.1061/(asce)be.1943-5592.0000150.

Chen, L., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A. L. (2017), "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs", *IEEE Trans. Pattern Anal. Mach. Intel.*, **40**(4), 834-848. https://doi.org/10.1109/tpami.2017.2699184.

Cho, S., Yun, C., Lynch, J.P., Zimmerman, A.T., Spencer Jr., B.F. and Nagayama, T. (2008), "Smart wireless sensor technology for structural health monitoring of civil structures", *Steel Struct.*, **8**(4), 267-275.

Cooke, H. (2015), "Lake waikaremoana bridge breaks, sending trampers plummeting", https://www.stuff.co.nz/national/71747665/lake-waikaremoana-bridge-breaks-sending-trampers-plummeting.

Dalal, N. and Triggs, B. (2005), "Histograms of oriented gradients for human detection", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, June.

Eftekhar Azam, S., Rageh, A. and Linzell, D. (2019), "Damage detection in structural systems utilizing artificial neural networks and proper orthogonal decomposition", *Struct. Control Hlth. Monit.*, **26**(2), e2288. https://doi.org/10.1002/stc.2288.

Harris, C.G. and Stephens, M. (1988), "A combined corner and edge detector", *Alvey Vision Conference*, **15**(50) 10-5244. https://doi.org/10.5244/c.2.23.

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*, 770-778.

Jiang, M., Guo, G. and Mu, G. (2020), "Visual BMI estimation from face images using a label distribution based method", *Comput. Vis. Image Understand.*, **197**, 102985. https://doi.org/10.1016/j.cviu.2020.102985.

Jo, E., Chun, H., Lee, S. and Hong, K. (2009), "Estimation of dynamic load and structural vibration on steel plate using transfer function", *J. Asian Arch. Build. Eng.*, **8**(2), 493-500. https://doi.org/10.3130/jaabe.8.493.

Kang, I., Schulz, M.J., Kim, J.H., Shanov, V. and Shi, D. (2006), "A carbon nanotube strain sensor for structural health monitoring", *Smart Mater. Struct.*, **15**(3), 737. https://doi.org/10.1088/0964-1726/15/3/009.

Kim, D. and Feng, M.Q. (2007), "Real-time structural health

monitoring using a novel fiber-optic accelerometer system", *IEEE Sensor. J.*, **7**(4), 536-543. https://doi.org/10.1109/jsen.2007.891988.

Kim, J., Sim, S., Cho, S., Yun, C. and Min, J. (2016), "Recent R&D activities on structural health monitoring in korea", *Struct. Monit. Mainten.*, **3**(1), 91-114. https://doi.org/10.12989/smm.2016.3.1.091.

Kim, N.E., Lee, C.H. and Kim, S.Y. (2015), "Formula for equivalent impulsive force to predict vibrational response of high-frequency staircases", *J. Korean Soc. Steel Constr.*, **27**(2), 181-193. https://doi.org/10.7781/kjoss.2015.27.2.181.

Kim, S., Pakzad, S., Culler, D., Demmel, J., Fenves, G., Glaser, S. and Turon, M. (2007), "Health monitoring of civil infrastructures using wireless sensor networks", *Proceedings of the 6th International Conference on Information Processing in Sensor Networks*, April.

Korean Statistical Information Service (2018), "Statistics for 120 body parts (male)", https://stat.kosis.kr/statHtml_host/statHtml.do?orgId=115&tblId=TX_115190170&dbUser=NSI_IN_115.

Kwon, S. and Suh, Y. (2009), "Development and application of the high speed weigh-in-motion for overweight enforcement", *Int. J. Highw. Eng.*, **11**(4), 69-78.

Li, J., Hao, H., Fan, K. and Brownjohn, J. (2015), "Development and application of a relative displacement sensor for structural health monitoring of composite bridges", *Struct. Control Hlth. Monit.*, **22**(4), 726-742. https://doi.org/10.1002/stc.1714.

Lim, J. and Yoon, H. (2019), "Real-time pedestrian dynamic-load localization using vision-based motion sensing", *J. Korean Soc. Hazard Mitigat.*, **19**(7), 323-330. https://doi.org/10.9798/kosham.2019.19.7.323.

Lucas, B.D. and Kanade, T. (1981), "An iterative image registration technique with an application to stereo vision", *Proceedings DARPA Image Understanding Workshop*, 121-130.

Lukezic, A., Vojir, T., ˇCehovin Zajc, L., Matas, J. and Kristan, M. (2017), "Discriminative correlation filter with channel and spatial reliability", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6309-6318.

Mills, K. (2017), "Suspension bridge with families walking across suddenly collapses sending dozens of people plummeting into the water below", https://www.mirror.co.uk/news/world-news/suspension-bridge-families-walking-across-10614388.

Mortimer, C. (2016), "Bali suspension bridge collapse kills eight people, including three children, and injures 30 others", https://www.independent.co.uk/news/world/asia/bali-suspension-bridge-collapse-eight-people-dead-30-injured-yellow-bridge-klungkung-a7364871.html.

Oh, B.K., Kim, D. and Park, H.S. (2017), "Modal response-based visual system identification and model updating methods for building structures", *Comput.-Aid. Civil Infrastr. Eng.*, **32**(1), 34-56. https://doi.org/10.1111/mice.12229.

Park, J.C., Cho, J.S., Gil, H.B. and Shin, J.I. (2014), "Vision-based technology for structural health monitoring of bridges", *Mag. Korea Inst. Struct Mainten. Inspect.*, **18**(2), 10-16.

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

Pumarola, A., Sanchez-Riera, J., Choi, G., Sanfeliu, A. and Moreno-Noguer, F. (2019), "3dpeople: Modeling the geometry of dressed humans", *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2242-2251.

Qin, T., Lin, M., Cao, M., Fu, K. and Ding, R. (2018), "Effects of sensor location on dynamic load estimation in weigh-in-motion system", *Sensor.*, **18**(9), 3044. https://doi.org/10.3390/s18093044.

Song, J., Jin, H., Chung, Y., Lee, S., Nam, W. and Jang, D. (2008), "Web-based bridge monitoring system with wireless sensor network environment", *J. Korean Soc. Hazard Mitigat.*, **8**(5), 35-44.

The Board of Audit and Inspection of Korea (2018), "On-site inspection of vulnerable leisure facilities", 2018-specific-056.

Tomasi, C. and Kanade, T. (1991), Detection and Tracking of Point Features, CMU-CS-91-132.

Van De Weijer, J., Schmid, C., Verbeek, J. and Larlus, D. (2009), "Learning color names for real-world applications", *IEEE Trans. Image Proc.*, **18**(7), 1512-1523. https://doi.org/10.1109/cvpr.2007.383218.

Velardo, C. and Dugelay, J. (2012), "What can computer vision tell you about your weight?", *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, 1980-1984.

Widianto, S., Davies, E. and Fernandez, C. (2020), "Collapse of footbridge over indonesia river kills at least nine", https://www.reuters.com/article/us-indonesia-bridge-idUSKBN1ZJ0G7.

Yoon, H., Elanwar, H., Choi, H., Golparvar-Fard, M. and Spencer Jr., B.F. (2016), "Target-free approach for vision-based structural system identification using consumer-grade cameras", *Struct. Control Hlth. Monit.*, **23**(12), 1405-1416. https://doi.org/10.1002/stc.1850.

Yoon, H., Shin, J. and Spencer Jr., B.F. (2018), "Structural displacement measurement using an unmanned aerial system", *Comput.-Aid. Civil Infrastr. Eng.*, **33**(3), 183-192. https://doi.org/10.1111/mice.12338.

HJ