

Structure and Infrastructure Engineering

Maintenance, Management, Life-Cycle Design and Performance

ISSN: 1573-2479 (Print) 1744-8980 (Online) Journal homepage: <https://www.tandfonline.com/loi/nsie20>

Structural displacement monitoring using deep learning-based full field optical flow methods

Chuan-Zhi Dong, Ozan Celik, F. Necati Catbas, Eugene J. O'Brien & Su Taylor

To cite this article: Chuan-Zhi Dong, Ozan Celik, F. Necati Catbas, Eugene J. O'Brien & Su Taylor (2020) Structural displacement monitoring using deep learning-based full field optical flow methods, Structure and Infrastructure Engineering, 16:1, 51-71, DOI: [10.1080/15732479.2019.1650078](https://doi.org/10.1080/15732479.2019.1650078)

To link to this article: <https://doi.org/10.1080/15732479.2019.1650078>



Published online: 21 Aug 2019.



Submit your article to this journal [↗](#)



Article views: 744



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 15 View citing articles [↗](#)



Structural displacement monitoring using deep learning-based full field optical flow methods

Chuan-Zhi Dong^a, Ozan Celik^a, F. Necati Catbas^a, Eugene J. O'Brien^b and Su Taylor^c

^aDepartment of Civil, Environmental, and Construction Engineering, University of Central Florida, Orlando, Florida, USA; ^bThe School of Civil Engineering, University College, Dublin, Ireland;; ^cThe School of Natural and Built Environment, Queens University Belfast, Belfast, UK

ABSTRACT

Current vision-based displacement measurement methods have limitations such as being in need of manual targets and parameter adjustment, and significant user involvement to reach the desired result. This study proposes a novel structural displacement measurement method using deep learning-based full field optical flow methods. The performance of the proposed method is verified via a laboratory experiment conducted on a grandstand structure with a comparative study, where the same data samples are analysed with a commonly used vision-based method, and a displacement sensor measurement is used as the ground truth. Statistical analysis of the comparative results show that the proposed method gives higher accuracy than the traditional optical flow algorithm and shows consistent results in compliance with displacement sensor measurements. Image collection, tracking, and non-uniform sampling are investigated in the experimental data and suggestions are made to obtain more accurate displacement measurements. A field-validation on a footbridge showed that the measurement error induced by the camera motion is mitigated by a camera motion subtraction procedure. The proposed method has good potential to be applied by structural engineers, who have little or no experience in computer vision and image processing, to do vision-based displacement measurements.

ARTICLE HISTORY

Received 12 December 2018
Revised 17 June 2019
Accepted 4 July 2019

KEYWORDS

Computer vision; deep learning; displacement measurement; grandstand structures; human induced vibration; optical flow; structural health monitoring

1. Introduction

Displacement is a critical indicator for structural performance evaluation and health condition assessment of infrastructure. Static and dynamic characteristics of structures such as bridge load capacity (Lee, Cho, & Shinozuka, 2006; Ojio, Carey, O'Brien, Doherty, & Taylor, 2016), bridge deflections (Moreu et al., 2016) and deformation profiles (Xu, Brownjohn, & Kong, 2018), load distribution (Fuchs, Washer, Chase, & Moore, 2004), load input information, unit influence line (UIL) and unit influence surface (UIS) (Khuc & Catbas, 2018), modal frequency and shape (Chen, Logan, Avitabile, & Dodson, 2019; Chen, Zhang, & Chen, 2014; Chen, Zhang, Zhang, & Zheng, 2015; Dong, Celik, & Catbas, 2018; Dong, Ye, & Jin, 2018; Yang, Dorn, Mancini, Talken, Nagarajiah, et al., 2017; Yoon, Elanwar, Choi, Golparvar-Fard, & Spencer, 2016) can be extracted from displacement data. Currently, displacement measurement is still a difficult task in conventional structural health monitoring (SHM) (Catbas, Dong, Celik, & Khuc, 2018).

Ye, Yi, Dong, Liu, and Bai (2015) summarised the current displacement measurement methods in the field of SHM, including: (1) contact type: linear variable differential transformers (LVDT), double integration of recorded acceleration data, displacement derivation from the strain-deflection relationship; and (2) non-contact type: global positioning systems (GPS) and integration of data from

Laser Doppler Vibrometers (LDV). Non-contact type displacement methods do not need a reference level, and do not need the access to the measured structures and can save on road closures, which are key advantages of this approach.

Ye et al. (2015) also indicated the limitation of conventional non-contact type methods: (1) GPS has low accuracy and sampling rate; and (2) LDV has high cost. A total station is also a non-contact displacement measurement tool. However, the use of total station is unsuitable for bridge monitoring (da Silva, Ibañez, & Poleszuk, 2018) due to difficulties of the installations of control points and continuous automatic monitoring at high frequency. In order to achieve desired continuous measurement, additional equipment has to be added to the original total station (Ehrhart & Lienhart, 2017; Omidalizarandi, Kargoll, Paffenholz, & Neumann, 2018). Extracting displacement measurements from image sequences has become a popular research topic in various applications of civil engineering (Feng & Feng, 2015, 2016; Feng, Fukuda, Feng, & Mizuta, 2015; Pan, Qian, Xie, & Asundi, 2009) since the manufacture of advanced cameras improved and computer vision techniques progressed. The advantages of non-contact, long-distance, high precision, low cost and less time-consuming measurement capabilities has caused vision-based displacement methods to get increasing attention from the community of structural

health monitoring with the potential of becoming an alternative to the conventional displacement measurement methods in SHM as well as to infrastructure inspections (Chen, Adams, Sun, Bell, & Büyüköztürk, 2018; Khuc & Catbas, 2017; Luo & Feng, 2018; O'Byrne et al., 2015; Wu, Casciati, & Casciati, 2014; Xu & Brownjohn, 2018; Ye, Dong, & Liu, 2016a).

In general, current vision-based displacement measurement methods are divided into five categories: (1) image correlation based template matching (Feng & Feng, 2016; Lava, Cooreman, Coppeters, De Strycker, & Debruyne, 2009; Pan, Tian, & Song, 2016; Sutton, Yan, Tiwari, Schreier, & Orteu, 2008; Ye, Dong, & Liu, 2016b), (2) colour based template matching (Ye, Dong, & Liu, 2016c), (3) key point matching (Khuc & Catbas, 2016, 2017; Lydon et al., 2019), (4) Lucas-Kanade optical flow estimation at key points (Celik, Dong, & Catbas, 2018a, 2018b; Dong, Celik, et al., 2018; Lydon et al., 2018; Yoon et al., 2016; Yoon, Shin, & Spencer, 2018) and (5) full field optical flow estimation (Celik et al., 2018a, 2018b; Chen et al., 2015; Khaloo & Lattanzi, 2017).

Image correlation-based template matching is the most popular (Chen, Joffre, & Avitabile, 2018; Zhong, He, & Li, 2017; Zhong, Shao, & Quan, 2018; Zhong & Quan, 2017, 2018). However, it is sensitive to changes in shading, illumination and background condition, especially when used in field applications (Xu & Brownjohn, 2018). To improve the measurement performance, manual light sources or targets are designed to be fixed on the structures and then tracked. Ye and Dong (Dong, Ye, et al., 2018; Dong, Ye, & Liu, 2015; Ye, Dong, et al., 2016c, 2016b; Ye, Yi, Dong, & Liu, 2016) installed light emitting diode (LED) and QR (quick response) codes on structures to improve the texture contrast of the visual tracking area and tried to eliminate the influence of illumination changes.

Tian and Pan (2016) combined the use of LED targets and a coupled bandpass optical filter to mitigate the ambient light interference. Colour based template matching is not robust to colour change and the application is limited to the close-range displacement measurements. For long distances, the colour condition of the measurement area could easily be affected by the light and shading, which makes it hard to get the right measurement results. To improve the measurement performance, artificial targets with specific colours could be utilised. Key point matching is a non-target method which calculates the displacement by averaging location change of the robust key points extracted from images. The method relies on calculating the similarities of the descriptors of key points in consecutive images based on statistical distance. Once similar key point pairs are recognised, the locations are confirmed to be the continuation of the former motion.

Generally, the key points may have robust properties such as being invariant to shading, illumination change, and scale. The most popular key points are Harris corner (Harris & Stephens, 1988), Shi-Tamasi corner (Shi & Tomasi, 1994), Scale-Invariant Feature Transform (Lowe, 2004) (SIFT) feature points and Speeded-Up Robust

Features (SURF) (Bay, Ess, Tuytelaars, & Van Gool, 2008). The performance of key point matching methods is highly dependent on the saliency of the texture of the measurement surface. The number of key point extractions from a measurement area of an image is not easy to decide and it has been an open question as to how many key points should be extracted for displacement measurement to achieve the best performance. Lucas-Kanade optical flow is a sparse flow calculation algorithm and is usually combined with key points to do visual tracking. This displacement measurement methodology involves similar limitations as the key point matching does. Besides, the 'small motion' assumption of optical flow restricts its application for large structural deflections (Dong, Celik, et al., 2018) although a pyramid method is used to refine the displacement evaluation in large displacement cases.

Full field optical flow can calculate the displacement vector of each pixel of images and give the displacement information of the entire structure. Classical full field optical flow estimation algorithms (Sun, Roth, & Black, 2010) originated from the core work of Horn and Schunck (1981). These algorithms are derived from variational methods which are based on the gradient change in images and need filters to smooth the motion in images. They are adversely affected by illumination change and give inaccurate flow estimation at the motion boundaries.

A phase-based optical flow algorithm is another method implemented into some structural displacement measurement problems (Chen et al., 2015; Yang, Dorn, Mancini, Talken, Kenyon, et al., 2017; Yang et al., 2018), but the applications are limited to those cases without background clutter. Parameters in these algorithms have to be adjusted to accommodate the differences in applications and it is too complicated for practical use. Moreover, full field optical flow calculation is a heavy task that needs longer computation time, which makes it unsuitable for continuous structural displacement measurement, especially for real time monitoring. Detailed comparisons can be found in (Dong, Celik, et al., 2018).

In this study, a novel structural displacement measurement method using deep learning based full field optical flow is proposed. A general procedure for vision-based displacement system is presented and the planar homography matrix is applied for camera calibration. By implementing a pre-trained deep neural network for optical flow calculation, i.e. FlowNet2, the full field optical flow is obtained, and the displacement of the measurement region is calculated by using a mean kernel or Gaussian kernel. The proposed method does not need manual target and can be operated with less human participation than the key point matching, Lucas-Kanade optical flow with key points and classical full field optical methods. Image collection strategies, tracking strategies in image sequences, non-uniform image sampling and camera motion problems are also discussed in this paper. Useful strategies are identified to address the problems that could occur in practical application. Laboratory experiments on a grandstand structure and field

experiments on a footbridge are conducted to verify the feasibility of the proposed method.

2. Methodology and system development

2.1. General procedure for vision-based displacement measurement system

Figure 1 illustrates the flowchart for the proposed full field structural displacement measurement method. In the first step, the camera is calibrated to obtain the relationship between the image coordinates and the real-world coordinates, i.e. to find how many physical units (e.g. millimetres) in the real world represent one pixel in the image. In the second step, image data from the structure in question are collected and are transferred to the next step for real time or post processing. In the third step, optical flow algorithms are implemented to do visual tracking and calculate the full field structural motion, thereby obtaining the motion vector at each pixel of the image.

In the fourth step, the false structural motion induced by the camera vibration is mitigated by subtracting the motion of the static parts in the image. In the last step, the full field structural displacement is obtained by converting the displacement in pixels to the displacement in physical units. In the flowchart in Figure 1, the three steps: camera calibration, full field optical flow estimation and mitigation of camera vibration are crucial as they directly affect the measurement accuracy. Each step will be introduced in detail below.

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. In research on vision-based structural displacement measurement, there are three main methods used frequently (Xu et al., 2018): scale factor, full projection matrix and the planar homography matrix. Scale factor is calculated as the ratio of the real world object dimension to the image dimension (Ye, Dong, et al., 2016b), or the ratio of the distance from the camera to measurement target and the focal length (Khuc & Catbas, 2017), when the axis of camera and lens is perpendicular to the motion plane.

If the axis of the camera and lens are not perpendicular to the motion plane, the scale factor has to be modified by the camera angle to the motion plane (Feng, Feng, Ozer, & Fukuda, 2015) and the scale factors in the vertical and horizontal directions are calculated separately. When multiple targets are located in the field of view with different depth,

the scale factors should be considered separately (Dong, Celik, et al., 2018).

If a scale factor is to be applied in vision-based displacement measurement systems, the assumption that the radial distortion is negligible must hold. When consumer grade cameras are utilised, and if the lens has a wide angle, the image has to be rectified to eliminate camera distortion. Objects in an image distorted by radial distortion become more distorted when the object moves further away from the image centre. Under this circumstance, the full projection matrix is usually applied. In general, a two-step calibration process is needed (Xu et al., 2018): (1) camera intrinsic matrix estimation using Zhang's method (Zhang, 2000); (2) camera extrinsic matrix estimation using at least four point correspondences.

Combining the camera intrinsic matrix and extrinsic matrix, the full projection matrix is obtained and applied to transform the motion in the image coordinates to the real-world coordinates. With this method, the camera angle problem and radial distortion problem are solved. It should be noted that the first step is usually completed indoors and the focal length is always fixed to keep the camera intrinsic matrix unchanged. However, when a zoom lens is selected to be used in the field, the focal length of the zoom lens is adjusted to take the best images with respect to the surroundings of the measurement targets, camera location, field of view (FOV), camera resolution and measurement distance, and so forth. Then, the camera intrinsic matrix has to be calibrated in the field, which is not an easy task. The authors recommend using a zoom lens with negligible radial distortion to avoid the tedious calibration work of camera intrinsic matrix for field applications.

Without the first step of the full projection matrix calibration, it is degraded to the third calibration method, the planar homography matrix. The planar homography matrix transforms the 2D plane of the object to the 2D real-world plane which means it can only be applied to estimate motion in two dimensions. For most of the structures in civil engineering, not all the displacements in three directions are needed. For example, the displacement in the vertical direction of bridge structures, is the most dominant as well as the displacement in the transverse direction for some flexible structures such as long span bridges under wind loads.

On the other hand, for high rise buildings, bridge towers and stay cables, two cameras and a full projection matrix are necessary to determine the transformation between 2D image planes and the 3D real world. In this study, only 2D motion is discussed and it is feasible to apply the planar homography matrix to achieve camera calibration. As shown in Figure 2, the image on the top right is the original picture of a footbridge taken by the camera and the axis of the camera lens is not perpendicular to the front side of the

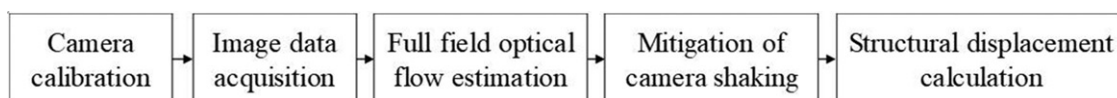


Figure 1. Flowchart for proposed full field structural displacement measurement method.

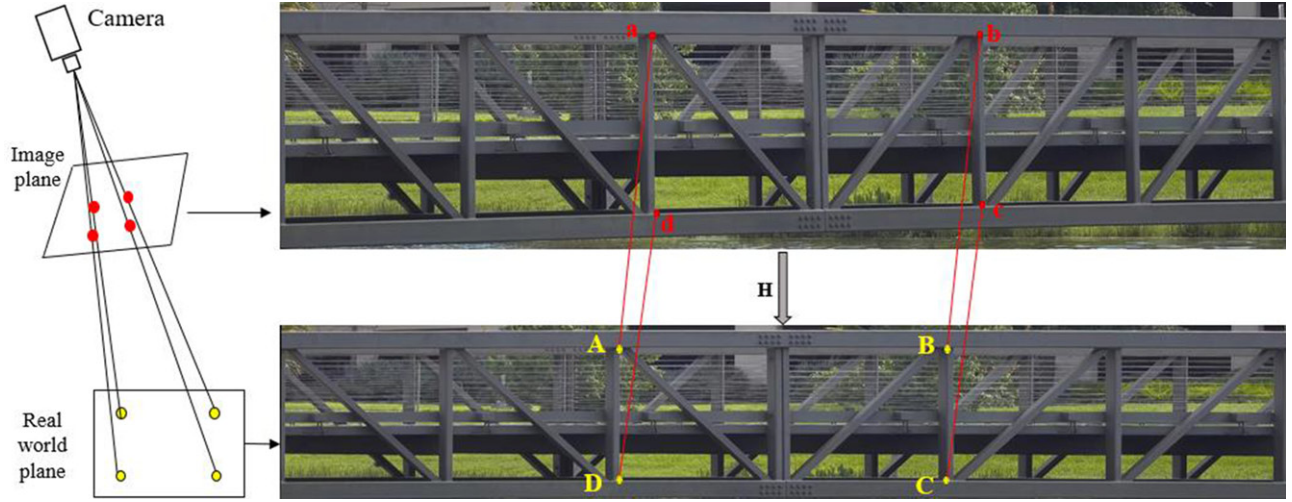


Figure 2. Image projection using the planar homography matrix and four point correspondences.

footbridge, indicating some obvious projection distortions. After using the planar homography matrix and four point correspondences, the original image is re-projected and the image at the lower right shows the rectified effect without projection distortion. The mathematical details of the planar homography matrix method are as follows:

As shown in Figure 2, the real-world object (a footbridge) is 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:

$$\mathbf{X} = \mathbf{H}\mathbf{x} \quad (1)$$

where $\mathbf{X} = \{X, Y, 1\}^T$, $\mathbf{x} = \{x, y, 1\}^T$. In this formulation, (x, y) are the image coordinates, (X, Y) are the original coordinates in the real world and \mathbf{H} is the 3×3 homography matrix which transforms the real-world plane to the image plane.

The equal symbol, '=' of Equation (1) is equality up to scale. If s denotes the scale, Equation (1) is expressed by:

$$\mathbf{X} = s\mathbf{H}\mathbf{x} \quad (2)$$

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 (Hartley & Zisserman, 2003). The homography matrix \mathbf{H} has nine unknowns but only eight of them are independent. Equation (1) can be formed as:

$$\begin{Bmatrix} X \\ Y \\ 1 \end{Bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{Bmatrix} x \\ y \\ 1 \end{Bmatrix} \quad (3)$$

This matrix is computed directly from image-to-world point correspondences. From Equation (1), 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 eight 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, Equation (1) becomes over-determined and a homogeneous estimation method is implemented to estimate the optimal \mathbf{H} . Writing the homography matrix, \mathbf{H} in vector form as, $\mathbf{h} = \{h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, h_9\}^T$, Equation (3) for n points becomes:

$$\mathbf{A}\mathbf{h} = \begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1X_1 & -y_1X_1 & -X_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1Y_1 & -y_1Y_1 & -Y_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2X_2 & -y_2X_2 & -X_2 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2Y_2 & -y_2Y_2 & -Y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & 1 & 0 & 0 & 0 & -x_nX_n & -y_nX_n & -X_n \\ 0 & 0 & 0 & x_n & y_n & 1 & -x_nY_n & -y_nY_n & -Y_n \end{bmatrix} \begin{Bmatrix} h_1 \\ h_2 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ h_9 \end{Bmatrix} \quad (4)$$

It is a standard result of linear algebra that the vector \mathbf{h} that minimises the algebraic residuals $|\mathbf{A}\mathbf{h}|$, subject to $|\mathbf{h}|=1$, is given by the eigenvector of least eigenvalue of $\mathbf{A}^T\mathbf{A}$. 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 Equation (2).

2.3. Image data acquisition

Unlike data from conventional sensors such as displacement sensors, accelerometers, strain gauges or tiltmeters, which provide one dimensional data (i.e. temporal data), image data is two dimensional and contains temporal and spatial information. This increases the demand to sample larger amounts of data and results in a reduction in the sampling rate of image data acquisition systems (i.e. cameras and image grabbers) compared to the conventional sensors. For image data acquisition systems, the sampling rate is referred to as frame rate, expressed in frames per second (FPS). When used within the context of single point or full field displacement time history from a vision-based system, the sampling rate is different from its frame rate.

The frame rate is related to the camera exposure time, time trigger, and so forth. Usually frame rate is a critical factor to be considered when doing image data acquisition in vision-based displacement measurement and it always influences the selection of image acquisition methods and devices. Depending on the monitoring or measurement requirements (whether real time monitoring or not), generally there are two ways to do image data acquisition as shown in Figure 3:

1. If there is no requirement to carry out real time displacement monitoring, the image data acquisition systems grab images continuously (also called video recording) and then processes the image data afterwards (post processing). The image data acquisition system can be a portable camera, or a digital camera which can be connected to a computer through different types of interface such as GigE, USB2/3, Camera Link, FireWire (IEEE 1394/IIDC DCAM Standard), or an analogue camera which needs an image grabber card to be connected with a computer, or even a smart phone. The normal portable cameras and smart phones usually have internal clock and the frame rate can be set as a fixed number such as 30 FPS, 60 FPS or 120 FPS. The images or videos are stored in the on-board storage card. In practical applications, the frame rate is not always fixed. For example, when the frame rate is set to 60 FPS, practically it might be less than that. For

instance, within the scope of this study, a Canon portable camera was tested, capturing video at a frame rate of 60 FPS at a resolution of 1920×1080 pixels. However, analysis showed that the real frame rate was 59.94 FPS on average. The frame rate reduction might be associated with some frames being delayed or dropped as a result of longer exposure time or unsuccessful triggering. As for the off-the-shelf portable cameras, the real and the pre-set frame rates are not distinctly different than each other. This feature makes them a convenient option for the monitoring of structures with low frequency dynamic characteristics. However, it is hard to find information on image time-stamps and dropped images. On the other hand, computer controlled analogue or digital cameras can deliver this information accurately during the exposure, since the image data acquisition procedure is programmed into a software package. When using these kinds of cameras, the exposure time can be auto-adjusted to acquire images with good quality. Nevertheless, image data acquisition is still a non-uniform sampling process. In general, the frame rate is calculated as $1/\Delta t$, where Δt is ideally the time interval for uniform sampling. In reality, the time interval between consecutive images varies every time, as Δt_i ($i = 1, 2, \dots, k$) and the average frame rate is calculated as the ratio of the total number of frames and the total acquisition time. To partially remedy this problem, a triggering function (i.e.

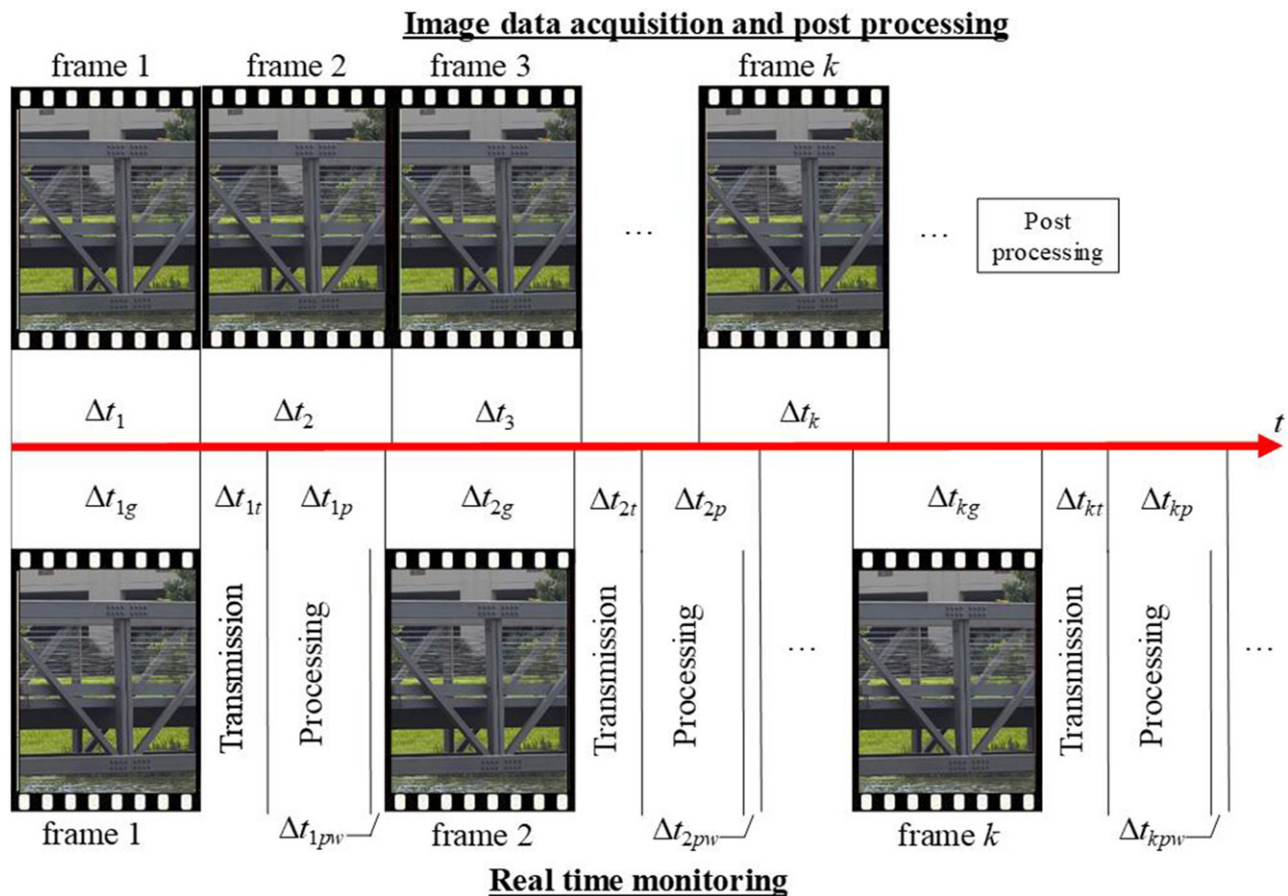


Figure 3. Image data acquisition and processing.

an edge signal) that controls the exposure can be sent to the Input/Output (I/O) interface of the camera satisfying the nominal uniform sampling (Dong, Celik, et al., 2018). It should be noted that the trigger frequency has to be less than the camera's maximum frame rate and a high trigger frequency may lead to an increased probability of frame drop.

2. If there is a requirement to do real time displacement monitoring, then the selection has to be among computer-controlled cameras. The data sampling process is divided into three steps: (1) image grabbing; (2) image transmission to the computer; (3) image processing. The computation time spent in each step (Δt_{ig} , Δt_{ip} , Δt_{pw}) may also vary to produce different numbers of samples because the exposure time, transmission time, and motion tracking time interval (during image processing) at each step may differ. For instance, as described earlier, a triggering function is applied to make the image grabbing time intervals, Δt_{ig} , equal and uniform, while the transmission time, Δt_{ip} , and image processing time, Δt_{pw} , may vary. For these reasons, the data sampling rate is decreased and becomes less than that of (1). Also, it is a non-uniform sampling process. To make this data sampling process uniform, a waiting time, Δt_{ipw} , can be added into the image processing time. This waiting time forces the total sampling time to be fixed and, as a result, the real sampling rate is decreased.

It is essential to know whether or not the application will require real time monitoring as it impacts the selection of motion tracking algorithms. In this study, full field displacement is estimated by computationally demanding optical flow algorithms which means that some of them cannot afford real time monitoring. That is why the selection of the optimal optical flow algorithm is a crucial step. Furthermore, in the cases when multiple cameras are needed, an array of cameras with time synchronisation can be designated to satisfy the measurement requirement (Wilburn, Joshi, Vaish, Levoy, & Horowitz, 2005).

2.4. Full field optical flow estimation using deep learning methods

Optical flow is the distribution of motion velocity vectors in image data. The motion can be an image sequence captured by a single camera or two images captured by two different cameras. The optical flow is usually represented by a 2D vector, i.e. the horizontal and vertical components, along with the two different directions. To estimate the full field structural displacement, optical flow estimation of the image sequences containing the motion of structures, is a good option. In general, there are two different optical flow estimation subsets: (1) local optical flow estimation methods, which calculate the flow vector on the selected pixels, blobs, or key points [e.g. Lucas-Kanade algorithm (Lucas & Kanade, 1981)]; and (2) global optical flow estimation methods, which calculate the flow vector at each pixel of the

image [e.g. Horn-Schunck algorithm (Horn & Schunck, 1981)].

Global optical flow estimation methods are ideal choices for full field structural displacement measurement, while utilising pyramid, window and smoothing techniques. The local optical flow estimation methods can also be used to estimate the flow vector at each pixel. Bouguet (1999) improved the original Lucas-Kanade method by implementing pyramid, feature tracker and interpolation to get the optical flow at each pixel. Sun, Roth, and Black (2014) analysed the current practices in optical flow estimation quantitatively and most of the optical flow methods were developed using the formulation structured by Horn-Schunck.

These methods are called classical methods. Based on the classical optical flow methods, Sun et al. (2014) implemented non-local smoothing techniques to develop a new method named Classic+NL (Classic with non-local). The performance of Classic+NL was validated by comparing with the classical optical flow methods on the popular optical flow datasets and showed better estimation results. Khaloo and Lattanzi (2017) implemented the Lucas-Kanade (LK) method, Horn-Schunck (HS) method, Black and Anandan (BA) method, and Classic+non local smoothing (Classic+NL) methods, which are investigated in Sun's work (Sun et al., 2014), and developed pixel-wise structural motion tracking methods and verified them on two shaking table tests.

Even though these classical optical flow methods are programmed as built-in functions into current computer vision libraries such as MATLAB or OpenCV and successfully implemented for structural displacement monitoring, there are lots of parameters in the functions that need to be adjusted based on experience. Limitations of classical optical flow methods such as the small displacement assumption, brightness consistency, motion boundary problems are still the main sources of errors. For structural engineers without enough experience in the computer vision field, it is difficult to use such methods. Further, some of the classical methods are too slow to satisfy the requirement of real time monitoring.

Instead of the aforementioned classical methods in computer vision, deep learning has been a very popular tool to help address the challenges in the field of computer vision in recent years (Bengio, Goodfellow, & Courville, 2017; Lecun, Bengio, & Hinton, 2015). With pre-trained deep neural networks, the optical flow can be easily estimated (Dosovitskiy et al., 2015; Ilg et al., 2017) by processing the image sequences through the pipeline of the forward propagation of the networks, without adjusting too many parameters as in classical optical flow methods. Since the dataset can be augmented by adding artefact noise, illumination change and other interference factors, the deep learning-based optical flow methods can perform better than the classical methods (Ilg et al., 2017).

With GPU acceleration, the deep learning-based optical flow method can do real-time monitoring. In addition, deep learning based methods such as FlowNet (Dosovitskiy et al.,

2015) and FlowNet2 (Ilg et al., 2017) are good at large displacement estimation, which is one of the drawbacks of the classical methods. In this study, a deep learning-based optical flow method, i.e. FlowNet2, is implemented to achieve full field structural displacement monitoring. FlowNet2 is based on the work of FlowNet that was first proposed by Dosovitskiy et al. (Dosovitskiy et al., 2015). The study represented a paradigm shift in optical flow estimation by allowing the use of a simple Convolutional Neural Network (CNN) architecture to directly learn the concept of optical flow from the dataset.

In FlowNet, Dosovitskiy et al. proposed two CNN architectures: FlowNet-S and FlowNet-C. In FlowNet-S, Dosovitskiy et al. first stacked two input images together and fed them through a generic network with 9 convolutional layers, allowing the network to decide itself how to process the image pair to extract the motion information. The first layer has a CNN kernel size of 7×7 , the second and third layers have kernel sizes of 5×5 , and the fourth to ninth have kernel sizes of 3×3 . The dimensions of each layer are conv1 ($354 \times 512 \times 6$), conv2 ($192 \times 256 \times 64$), conv3 ($96 \times 128 \times 128$), conv3_1 ($48 \times 64 \times 256$), conv4 ($24 \times 32 \times 512$), conv4_1 ($24 \times 32 \times 512$), conv5 ($12 \times 16 \times 512$), conv5_1 ($12 \times 16 \times 512$) and conv6 ($6 \times 8 \times 1024$). Finally, they added a refinement operation of the coarse feature maps to the high-resolution prediction and then provided the optical flow prediction. The detailed CNN architecture can be found in (Dosovitskiy et al., 2015).

In FlowNet-C, instead of directly stacking two images, they first fed the two images to three convolutional layers separately and then combined them together with an explicit correlation layer. After another six convolutional layers, a refinement operation was added and then the optical flow prediction output. The training dataset FlowNet used is their homemade FlyingChair dataset, which simulates the motions and illumination change. The pre-trained FlowNet performs well on the current optical flow dataset, especially for the case of large displacement.

However, FlowNet still cannot compete with variational methods in small displacement and real-world data. Ilg et al. (2017) proposed FlowNet2 based on FlowNet, where multiple FlowNet-S and FlowNet-C networks were stacked and a sub-network specialising in small motions to improve the accuracy and the speed of the original FlowNet was integrated. They trained the new CNN architecture on the FlyingChair and FlyingThings3D datasets. Illumination change, background clutter and other noises were added to the training data set as data augmentation to simulate real scenarios. Combining different training dataset and orders, FlowNet2 finally gave nine different CNN architectures for optical flow prediction.

These CNN architectures are suitable for various application requirements such as small displacement, large displacement, good accuracy and fast speed as well as being capable of dealing with illumination change and background clutter. FlowNet2 performs well on small displacement and real-world data and is fast enough for real-time motion estimation. It should be noted that FlowNet2 will not always be

the strong choice for full field optical flow estimation, and with the development of the computer vision techniques, more advanced optical flow algorithms will come out and be the alternatives for the purpose given in this study.

Figure 4 presents the optical flow estimations of a beam motion in two images using six different methods, namely, Horn-Schunk (HS), Lucas-Kanade with pyramid and sparse to dense interpolation (LKPyrSD), Farneback, BA, Classic + NL and FlowNet2. The reason the specific methods are chosen for comparison is that they are implemented and validated for structural displacement measurement in literature aforementioned. The beam has a downward deflection from Frame 1 to Frame 2 and since there are sensors installed on it, the motion in the images are not just those of the beam but also those of the cables hanging from the sensors.

In the flow field colour coding, the colour indicates the motion direction and the distance away from the centre indicates the motion amplitude. In this case, the beam moves down from Frame 1 to Frame 2 so it should be coloured yellow in the full field optical flow map according to the flow field colour coding. In the optical flow estimation results, HS provides the worst estimate and is not robust in the presence of image noise. The beam motion is interfered with motions in other directions and the background causes an excessive amount of incorrect motion estimation. LKPyrSD gives poor results for motion boundaries and the motions on boundaries are blurred.

The Farneback method cannot give accurate results for the whole beam but only those parts with salient textures. BA and Classic-NL perform better than these three but still give unsatisfactory results on boundaries. FlowNet2 gives the best results for beam motion, especially at boundaries. It even gives more detail about the motion of the cable. On this basis, the authors have selected FlowNet2 for full field optical flow estimation. The results using HS, LKPyrSD, BA and Classic-NL are similar to those by (Khaloo & Lattanzi, 2017) and Classic-NL gives the best prediction among these four classical optical flow methods.

In the comparative studies of the experimental verification section, the authors will compare the results from FlowNet2 with Classic-NL for structural displacement monitoring. Instead of performing image re-projection using the planar homography matrix to mitigate the distortion caused by projection first and then estimating optical flow just as (Khaloo & Lattanzi, 2017) did in their work, in this study the authors directly estimate the optical flow and shift the original points to the new location. Then the authors implement the planar homography matrix to project the location in the image to the real world. The consideration is that if the planar homography matrix is applied to the image first, the re-projection might break the pixel structures and this makes optical flow estimation inaccurate.

2.5. Camera motion subtraction

When using vision-based methods to estimate structural displacement, especially in field application, camera motion is

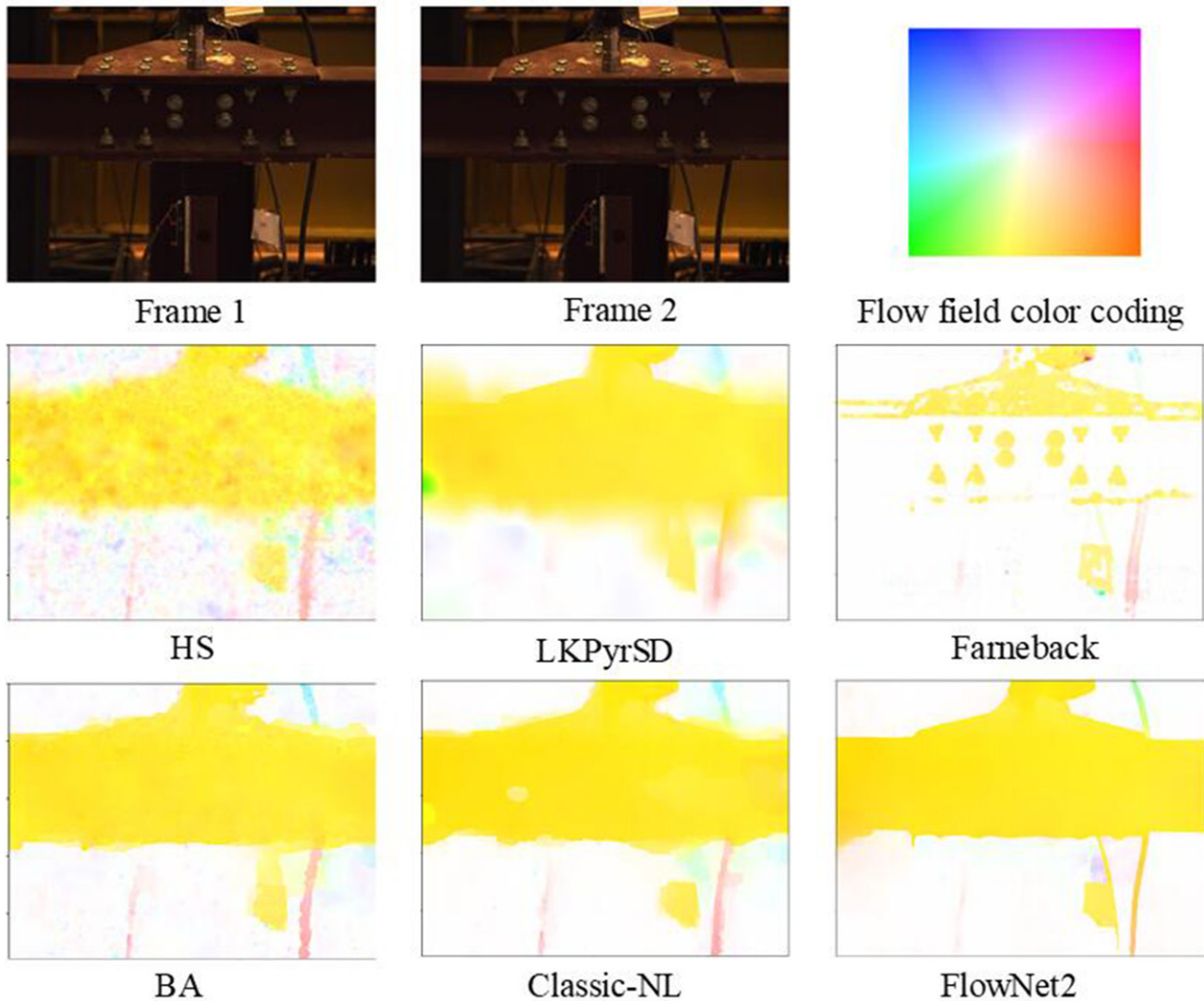


Figure 4. Optical flow estimations using different methods.

always a big issue which can induce displacement measurement errors. Camera motion may be caused by the ground vibration or wind. It will be mixed into the structural motion and has to be removed to rectify displacement measurement. There are two main approaches to mitigate the camera motion: (1) filtering out the displacement components related to the frequencies of the camera motion; (2) directly subtracting the motion of a static object/scene in the video from the total motion.

The first approach is suitable for the case when the frequency of the camera motion is not close to the structural motion with a trade-off that accelerometers are necessary to be installed on the camera to identify the frequencies of camera motion. The second approach is suitable for the case when there are static areas (objects assumed to be static) in the field of view of the camera (Feng & Feng, 2017). In this study, the second approach is implemented to eliminate the displacement errors induced by the camera motion. As shown in Figure 5, the areas A, B, C and D can be assumed to be static. When a camera affected by ambient motion is utilised to measure the structural displacement of the foot-bridge, i.e. the displacement of M, the rectified displacement

is M subtracted by the average displacement of A, B, C and D.

2.6. Structural displacement calculation

Once the rectified optical flow is obtained, the planar homography matrix in Equation (1) is applied to convert the displacement in pixels to their actual physical counterparts. By using full field optical flow and the planar homography matrix, the full field structural displacement is obtained. Theoretically, the displacement of any point of the structure can be acquired by taking the value of the full field structural displacement map. Conversely, in conventional structural health monitoring (SHM), displacement sensors are installed to measure the structural displacements at discrete points. As shown in Figure 6, a displacement sensor is installed to measure the structural displacement of the beam, and the result obtained here is the displacement at a discrete point. The experimental setup and loading condition are outlined in Section 3 and Figure 9.

The structural displacement at a discrete point can be obtained by using the displacement in the area close to the

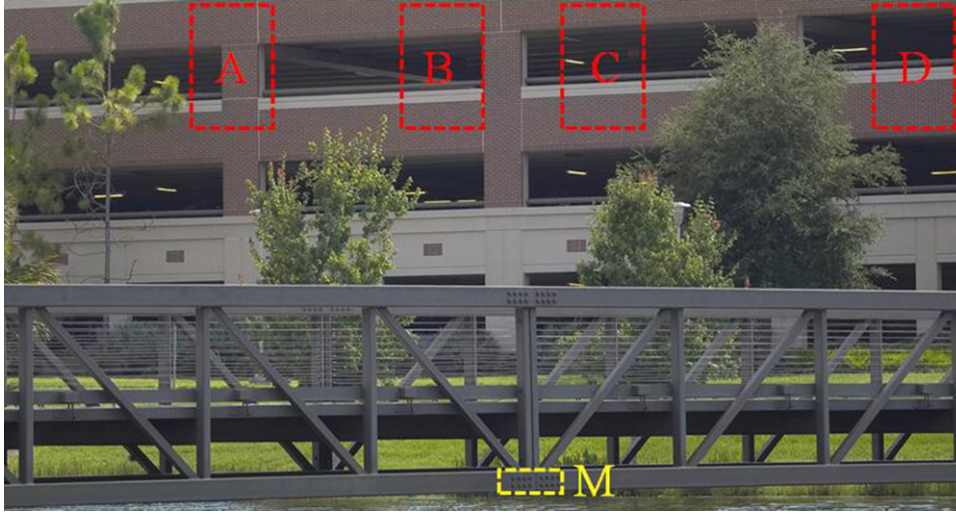


Figure 5. Structural displacement rectification using camera motion subtraction.

single measurement point. For example, the area marked by the red box in the two frames of Figure 6 is used to estimate the structural displacement of the discrete point that the installed displacement sensor measures. Two methods can be applied to estimate the structural displacement: (1) displacement calculation with Gaussian kernel; and (2) displacement calculation with mean kernel. The Gaussian kernel applied in this study is represented by \mathbf{G} , as follows:

$$\mathbf{G}(m, n) = \frac{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(m-\frac{h_c}{2})^2 + (n-\frac{w_c}{2})^2}{2\sigma^2}}}{\sum_{m=1}^h \sum_{n=1}^w \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(m-\frac{h_c}{2})^2 + (n-\frac{w_c}{2})^2}{2\sigma^2}}} \quad (5)$$

$$h_c = \left\lfloor \frac{h+1}{2} \right\rfloor \quad (6)$$

$$w_c = \left\lfloor \frac{w+1}{2} \right\rfloor \quad (7)$$

$$\sigma = \left\lfloor \frac{h+w}{2} \right\rfloor \quad (8)$$

where, m and n are the row column number of the Gaussian kernel, respectively, and h and w are its height and width. The symbol, $\lfloor \cdot \rfloor$ indicates the floor function that takes as input a real number, x , and gives as output the greatest integer less than or equal to x . while the mean kernel is represented by \mathbf{M} , as follows:

$$\mathbf{M}(m, n) = \frac{1}{hw} \quad (9)$$

The Gaussian kernel used in this study actually models the focus of attention that is motivated by the biological visual system which concentrates on certain image regions requiring detailed analysis (Zhang, Zhang, Yang, & Zhang, 2013). The closer to the focus centre, the greater the weight is set. This Gaussian kernel implements the concept of attention guide tracking. Additionally, since most of the classical optical flow methods do not perform well on the motion boundaries, the Gaussian kernel can decrease the weight when calculating the weighted average displacement.

When applying camera motion subtraction, if the displacement of the assumed static areas on the background are also calculated using a Gaussian kernel, the error then can be reduced by giving lesser weight to the parts away from the focus centre of the static areas. This is an indirect way of suppressing the outliers, especially those close to the motion boundaries but far away from the focus centre. When manually selecting the assumed static area, the assumption is more accurate as the kernel is placed closer to the centre. In addition, the mean kernel is a well-known strategy which is applied in key points-based tracking using Lucas-Kanade optical flow (Dong, Celik, et al., 2018), key point matching using Fast Library for Approximate Nearest Neighbors (FLANN) and Kanade-Lucas-Tomasi based template matching (Yoon et al., 2016). Correlation based template matching also uses the mean kernel method to find the best location (Dong, Ye, et al., 2018).

The displacement of a discrete point is estimated by combining the full field displacement of the selected region with either Gaussian kernel, \mathbf{G} , or mean kernel, \mathbf{M} :

$$d_{Gi} = \sum_1^m \sum_1^n (\mathbf{X}_i - \mathbf{X}_0) \circ \mathbf{G} \quad (10)$$

or

$$d_{Mi} = \sum_1^m \sum_1^n (\mathbf{X}_i - \mathbf{X}_0) \circ \mathbf{M} \quad (11)$$

where \circ is the element-wise product operator, d_{Gi} and d_{Mi} are the displacements estimated using Gaussian kernel and mean kernel, \mathbf{X}_0 is the original coordinate of the pixel-wise location in the real world, \mathbf{X}_i is the current (the i th frame) coordinate of the pixel-wise location in the real world and $\mathbf{X}_i - \mathbf{X}_0$ is the displacement vector.

The displacement time history is obtained by calculating the optical flow between the current image (frame k as shown in Figure 7) and the original image (frame 1 in Figure 7, top row). In this strategy, there is neither frame nor tracking location update and the displacement at every single time point is independent of the others. Another

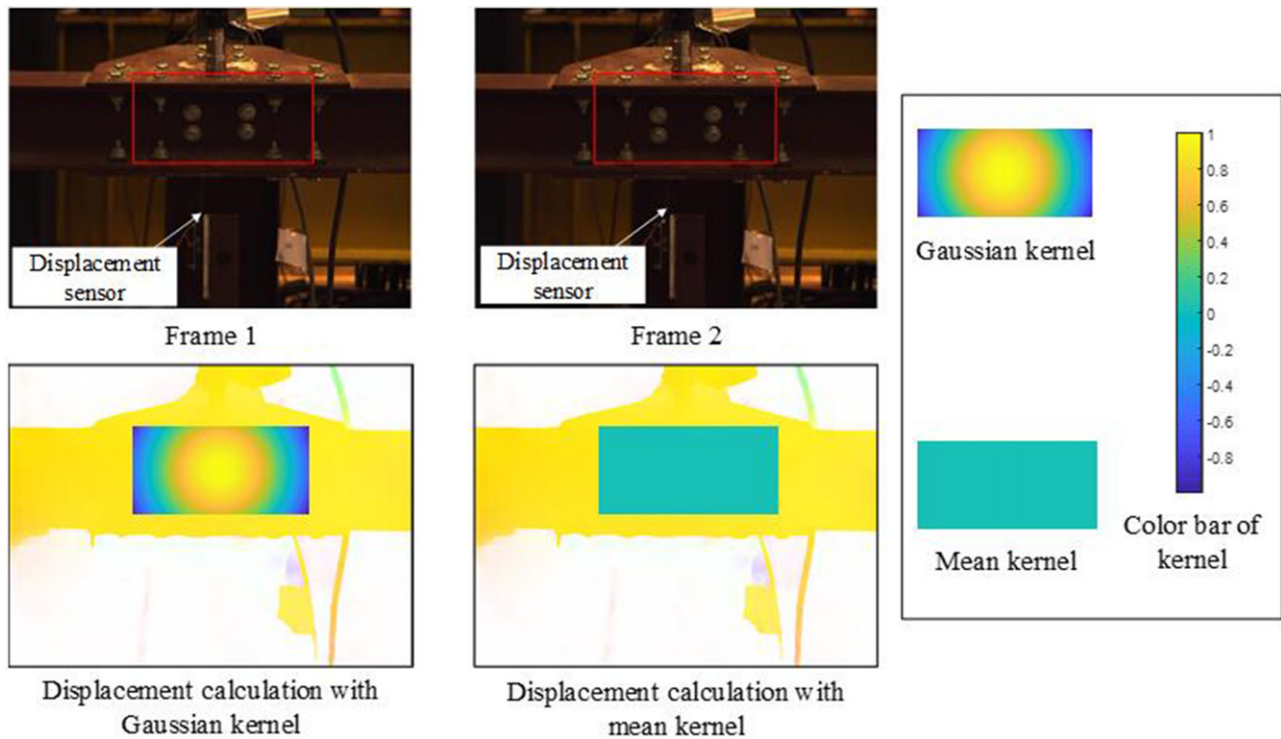


Figure 6. Structural displacement at a discrete point using kernels.



Figure 7. Two different strategies to process image sequences and get displacement time histories.

strategy which is indicated in the bottom row of Figure 7 calculates the optical flow between two consecutive frames with updating the frame each time. In the displacement time history, the displacement at every single time point is dependent on its previous neighbours. The final displacement time history is the cumulative operation of the

incrementals obtained by calculating the displacement of two consecutive frames.

With practical experiences and observations of conducting vision-based displacement measurement, general suggestions are summarised as the pros and cons of the two different strategies, which are listed in Table 1. The main

reason that the first strategy is recommended is that when the second strategy (with frame update) is applied, errors tend to accumulate. Stiros (2008) analysed the accumulated errors in velocities and displacement deduced from accelerographs using numeric integration, which provided a possible way to eliminate the errors in the second strategy. Stiros indicated that the errors depend on the characteristic errors of accelerometers such as the sensitivity/accuracy of the measurements described by standard deviation, duration of the record and instabilities in the sampling rate. Also, the peaks in the accelerographs contribute to the errors during numeric integration.

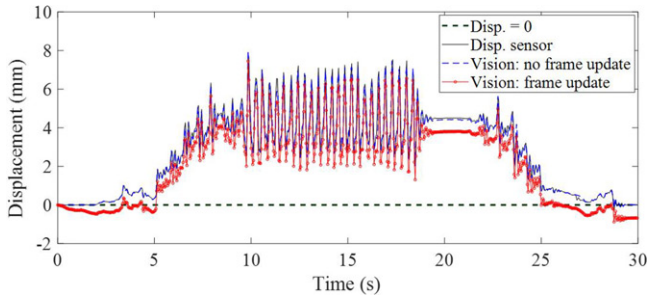


Figure 8. Displacement results obtained from the two different image sequence processing strategies and the displacement sensor.

In this study, when using the second strategy, the process of calculating displacements from consecutive frames is very similar to the process of calculating velocity from accelerations using numeric integration. It may be beneficial for the elimination of errors in the second strategy from the theoretical analysis of Stiros' work. While the formulas summarised by Stiros are limited to the analysis of linear movements and rotations are ignored; at the end, Stiros stated that if baseline corrections are taken into consideration for the formulas of errors, the numeric integration errors may be reduced. In the study, the first strategy to take the first frame as the baseline actually applies the way of baseline correction to some degree and it is more practical than using numerical integration with Stiros' theoretical analysis.

Figure 8 displays the displacement results obtained from the two different image sequence processing strategies and the ground truth (displacement sensor) from the same experiment introduced in Section 3. Due to the accumulation of errors, the displacement result obtained using this strategy with frame update deviates from the ground truth, while the result obtained using the strategy without frame update is consistent with it. While not updating brings inconvenience and possible errors to structural displacement

Table 1. Pros and cons of the two different strategies to process image sequence.

Strategies	Without frame update: optical flow calculation between the current image and the original image	With frame update: optical flow calculation between two consecutive frames
Pros	The displacement at every instant in the time history depends on the tracking between the current image and the original image and is independent of the others. The error in current time instant will not be accumulated.	Target scale changes, deformation, illumination changes and other changes of image quality can be updated in current tracking task and adjusted tracking scenarios give high chances of accurate tracking.
Cons	Without updating the frames, the target scale changes, deformation, illumination changes may affect tracking performance. Classical optical flow may fail to estimate large displacements in non-consecutive frames since there is a small motion assumption in classical optical flow methods. The measurement target may be out of view in the original frame.	The displacement at every instant in the time history is dependent on its previous neighbour. When calculating displacement time history, error in the current instant will be accumulated afterwards. This cumulative effect may cause a drift through the time history and a gradual loss of accuracy.

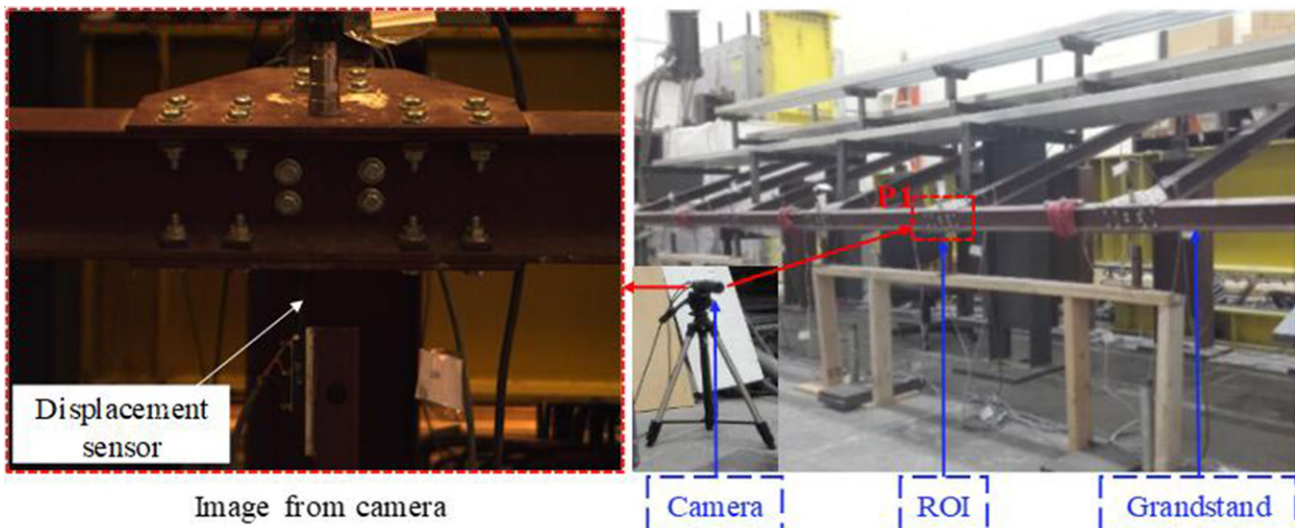


Figure 9. Experimental setup of the grandstand monitoring.

measurement, these can be overcome by controlling the image quality and using pyramid methods for optical flow estimation or visual tracking.

In the practice of conventional object tracking task, the second strategy (with frame update) is more popular and practical when processing the tracking problems with scale/view changes and illumination changes (OpenCV, 2019; Zhang et al., 2013). However, in vision-based displacement measurement, the first strategy (without frame update) is preferred, regardless of approach: digital image correlation based template matching, feature point matching, or optical flow (Dong, Celik, et al., 2018).

3. Laboratory verification

3.1. Experimental setup

In this section, an experiment on a model grandstand in the structures' laboratory of the University of Central Florida is designated to verify the feasibility and performance of the proposed displacement methods. The grandstand, shown in Figure 9, is a scaled model of part of a real American football stadium. Detailed information can be found in previous papers (Celik et al., 2018a, 2018b; Dong, Celik, et al., 2018). One region of interest (ROI), P1, is selected as the measurement point for the proposed method. At this point, a conventional displacement sensor (potentiometer) is installed to measure the displacement for comparison and is used as the ground truth. The cameras are MindVision-MV-GE131gc-t with a maximum frame rate of 60 Hz, a resolution of 1280×960 pixel and a zoom lens with a focal length of $5 \sim 100$ mm. The cameras are connected to the same acquisition system as the displacement sensor.

Unlike previous work (Dong, Celik, et al., 2018), in this study no trigger module is applied to enforce uniform sampling in the image data acquisition. The average frame rate of the camera is around 29 FPS (frame per second). The sampling rate for the displacement sensor is 100 Hz and it is down-sampled for comparison with the camera data. During the experiment, one person stands on the grandstand and jumps as the camera and potentiometer record the structural motion at P1. The acquired image sequence is analysed using the proposed methods. The displacement obtained from image sequences is compared with that of the displacement sensor.

3.2. Comparative study of displacement measurement using different methods

In this study, FlowNet2 is implemented and is verified through comparison with Classic + NL and the displacement sensor. Here, to obtain the displacement, both mean kernel and Gaussian kernel are applied to the full field optical flow results estimated by Classic + NL and FlowNet2. Figure 10 illustrates the comparison of displacement time histories of P1 using displacement sensor (Disp. Sensor) and vision-based methods, i.e. Classic + NL full field optical flow with Mean kernel (C + NL + M), Classic + NL full field optical flow with Gaussian kernel (C + NL + G), FlowNet2 full field optical flow with Mean kernel (FlowNet2 + M), and FlowNet2 full field optical flow with Gaussian kernel (FlowNet2 + G).

The synchronisation of different data sources is done manually. Different segments of the time history plot correspond to different events happening on the grandstand under human load. First, the subject climbs up the grandstand causing an increase in displacement ($0 \sim 5$ s); then walks to P1 causing fluctuations and an increase in the displacement ($5 \sim 9$ s); then begins to jump ($9 \sim 19$ s), which produces a continuous up-and-down pattern; then briefly stops; resumes jumping for two more seconds ($19 \sim 21$ s) and finally climbs down from the grandstand, allowing the displacement to return to zero. The figure indicates that the results obtained from all the vision-based methods are consistent with the benchmark.

Before comparing the vision-based methods and displacement sensor quantitatively, the displacement time histories from the vision-based methods have to be preprocessed. As mentioned, the image sampling is non-uniform in this experiment. Figure 11(a) shows the time spent on the image collection for each frame, Δt and Figure 11(b) gives the histogram and normal distribution fit for the same variable. The mean time interval for a frame is $\mu = 0.0337$ s which gives an average camera frame rate of 29.7 FPS. This frame rate cannot be directly approximated to 30 FPS, because the standard deviation of the time interval, σ is 0.012 s, which is significant. To use the mean, frame rate would cause misalignment problems for displacement time histories. The interval of mean ± 2 standard deviations, $[\mu - 2\sigma, \mu + 2\sigma]$, is, $[0.010\text{s}, 0.057\text{s}]$ at a level of confidence of 95%. It should be noted that the uncertainties of image process measurement are the selection of visual tracking algorithms/optical flow algorithms, region of interest and average methods

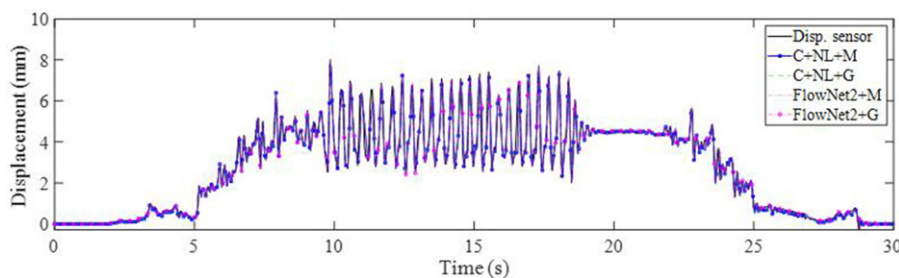


Figure 10. Comparison of displacement time histories from displacement sensor and vision-based methods.

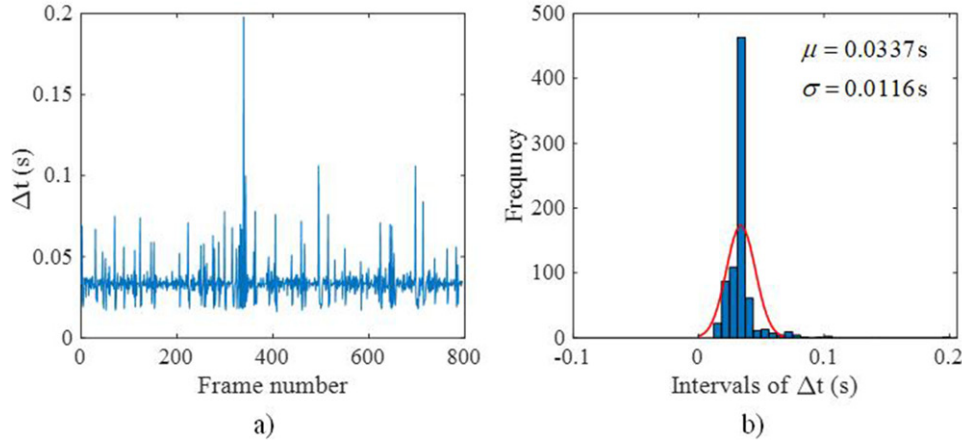


Figure 11. Statistical analysis of time spent on image collection for each frame.

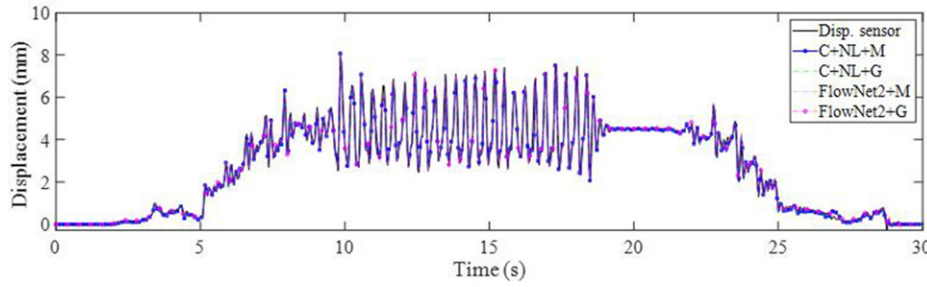


Figure 12. Resampled displacement time histories using displacement sensor and vision-based methods.

(mean kernel or weighted kernels), and so forth. These are also the sources of the uncertainties.

The non-uniformly sampled displacement time histories obtained from the vision-based methods are first resampled at 25 Hz using a cubic spline interpolation, while the uniformly sampled displacement time history obtained from the displacement sensor is directly down-sampled to 25 Hz. Cross correlation (Oppenheim, Willsky, & Nawab, 1996) is applied to synchronise the resampled displacement time histories obtained from the vision based methods and the displacement sensor.

Figure 12 depicts the resampled displacement time histories of P1 using all methods. After resampling, the consistency between the displacement time histories obtained from vision-based methods and displacement sensor are still very good and do not change, compared to Figure 10. Normalised root mean square error (NRMSE) is applied to evaluate the goodness of fit between the signals, and normalised cross-correlation (NCC) is calculated to evaluate the similarities between them:

$$\text{FIT}_{\text{NRMSE}} = 1 - \frac{|d_v(i) - d_s(i)|}{|d_s(i) - \mu_{d_s}|} \quad (12)$$

$$\text{NCC} = \frac{|\sum (d_v(i) - \mu_{d_v}) \times (d_s(i) - \mu_{d_s})|}{\sqrt{\sum (d_v(i) - \mu_{d_v})^2} \sqrt{\sum (d_s(i) - \mu_{d_s})^2}} \quad (13)$$

where $d_v(i)$ and $d_s(i)$ are the displacements from the vision-based methods and displacement sensor, respectively, and μ_{d_v} and μ_{d_s} are their mean values.

Table 2. Normalised root mean square error (NRMSE) and normalised cross-correlation (NCC) of the fit between the vision-based displacements and the benchmark.

Method	FIT _{NRMSE}	NCC
C + NL + M	0.8727	0.9921
C + NL + G	0.8726	0.9921
FlowNet2 + M	0.8758	0.9923
FlowNet2 + G	0.8758	0.9923

The greater the values of FIT_{NRMSE} and NCC, the better the fit. It can be seen in Table 2 that the proposed methods, FlowNet2 + M and FlowNet2 + G, perform a little bit better than the alternatives, C + NL + M implemented in literature (Khaloo & Lattanzi, 2017). For example, the normalised mean square error for the former two is 0.8758 which is slightly better than that of the latter (0.8727). Similarly, the NCC of the former at 0.9923 is slightly better than that of the latter at 0.9921.

The displacement time histories when no loads are present on the structure are extracted as the measurement error distribution and are used to estimate the measurement accuracy and resolution. The index of the measurement accuracy and resolution is defined by ± 2 standard deviations, which corresponds to a level of confidence of 95% (Khuc & Catbas, 2017; Xu et al., 2018). The smaller the standard deviation, the smaller the error and the better the measurement accuracy and resolution. Figure 13 presents the distributions of measurement error from different methods. Table 3 gives the measurement resolution ($\pm 2\sigma$) analysis.

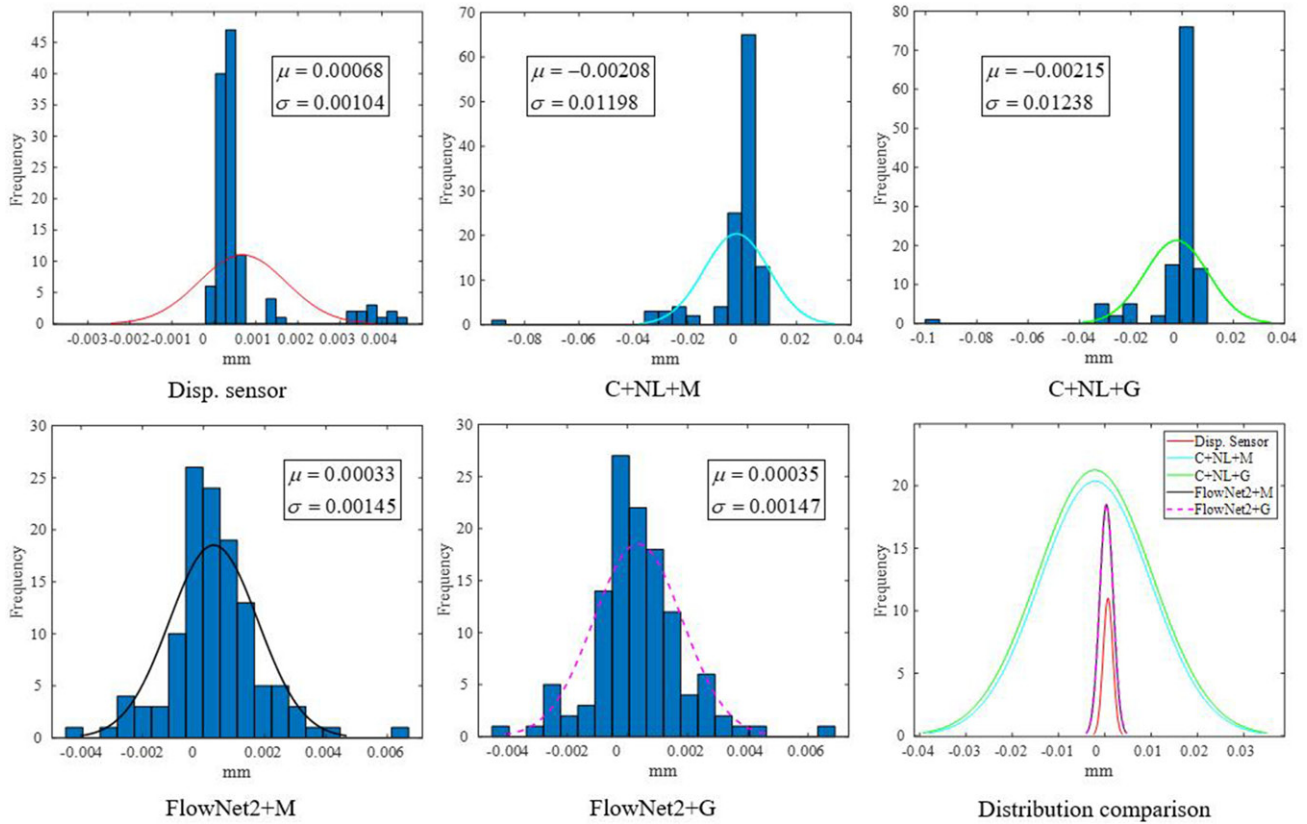


Figure 13. Distributions of measurement error for different methods.

Table 3. Measurement accuracy and resolution ($\pm 2\sigma$) analysis.

Method	Disp. sensor	C + NL + M	C + NL + G	Flow Net2 + M	Flow Net2 + G
$\pm 2\sigma$	± 0.0021	± 0.0240	± 0.0248	± 0.0029	± 0.0029

From Figure 13 and Table 3, it is indicated that the accuracy and resolution of the proposed methods using FlowNet2 is ± 0.0029 mm which is very close to the ground truth (displacement sensor) which is ± 0.0021 mm. The vision-based method using Classic + NL, has a resolution of around 0.0240 mm, which is almost 10 times that of the proposed methods. Further, from Figure 13, it can be seen that the proposed methods using FlowNet2 give much better stability than Classic + NL in displacement measurement, especially from the sixth subplot 'Distribution comparison'.

Since the displacements obtained from the vision-based methods are non-uniformly sampled, they cannot be directly processed by general Fast Fourier Transform (FFT) or Power Spectral Density (PSD). Therefore, in this study, two ways are applied to extract the frequency information from the non-uniformly sampled displacement data: (1) Lomb-Scargle Periodogram for the non-uniformly sampled data (Lomb, 1976); and (2) Power Spectral Density Using FFT for uniformly sampled data from cubic spline data interpolation. Figure 14 shows the comparison of displacement data in the frequency domain. The modal testing of the same structure was conducted in previous work and the first natural frequency is 5.74 Hz (Celik et al., 2018a, 2018b; Dong, Celik, & Catbas, 2017; Dong, Celik, et al., 2018).

From Figure 14, it is suggested that all methods give the operational modes for human jumping (frequency around 2.86 Hz). However, Lomb-Scargle method does not give unique peaks (peak at 2.825 Hz and peak at 2.881 Hz) and the peaks have a clear shift from those obtained from the displacement sensor which is regarded as ground truth. This is a distortion in dynamics which may be induced by non-uniform sampling. The peaks obtained from the cubic spline data interpolation of the original non-uniform data is very close to the ground truth.

By picking the peak of the PSD curve, the frequencies of human jumping load are extracted as shown in Table 4. From this table, it is indicated that using cubic spline data interpolation seems better than directly using the Lomb-Scargle method. Using cubic spline interpolation data gives the exact human jumping mode frequency as the displacement sensor. In this view, using cubic spline interpolation may correct the distortion that may be induced by non-uniform sampling. It might be because the sampling rate is low and in this case cubic spline interpolation works.

However, in Figure 14 there is a discrepancy between spline and displacement sensor results at higher frequencies, which is expected due to the resampling step. This observation reflects the limitation of the vision-based methods: inherently vision-based methods give a lower frequency range than conventional displacement measurement for hardware and software processing reasons, so that over the entire range, especially the high frequency range, vision-based methods are not as sensitive as their conventional counterparts.

From the accuracy and resolution analyses, as shown in Tables 2 and 3, using mean kernel seems to be a little better

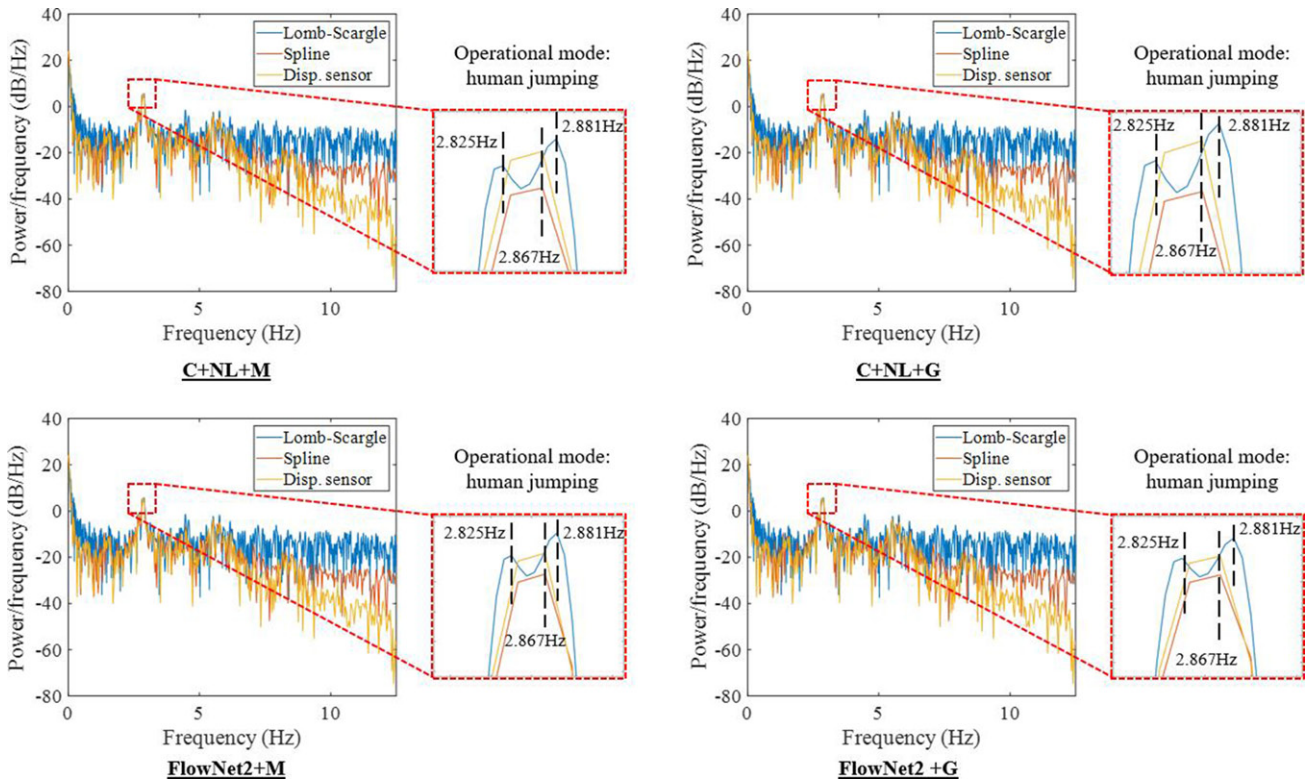


Figure 14. Comparison of displacement data in the frequency domain.

Table 4. Frequencies of human jumping load extracted from displacement data: unit (Hz).

Method	Disp. sensor	C + NL + M	C + NL + G	Flow Net2 + M	Flow Net2 + G
Direct PSD	2.867	–	–	–	–
Spline + PSD	–	2.867	2.867	2.867	2.867
Lomb-Scargle	–	2.881	2.881	2.881	2.881

than the Gaussian kernel, which was not expected from the very beginning. The authors believe this to be because the region using the kernel did not contain too many motion boundaries and the advantage of Gaussian kernel was not manifest. As a result, at this time, it remains an open question whether it is better to use mean kernel or Gaussian kernel.

3.3. Comparison of computation times of vision-based methods using different full field optical flow algorithms

When comparing the image processing time, it takes 1.7 seconds to calculate the full field optical flow of two images with a resolution of 1280×960 pixel using FlowNet2. The computation time is accelerated by a Graphics Processing Unit (GPU) on a Linux system (Ubuntu 18.04) with the AMD Ryzen 5 2600X CPU, 16 Gb RAM, and the NVIDIA GeForce GTX 1080 Graphics Card. It takes about 1600 seconds for the same operation on the same system using Classic + NL. The Classic + NL used in this study is the same as that of Khaloo and Lattanzi and does not implement GPU acceleration. During this experiment, 1159 images were collected, and it took about 32.8 minutes using FlowNet2 to calculate the full

field optical flow of the image sequence whereas it took about 21.5 days for Classic + NL.

The processing tests were conducted on the same computer and computation times were directly extracted from the internal clock when running optical flow codes. This is a limited test, but it is a practical and simpler way to compare the speeds of different algorithms. To date, it is unknown whether Classic + NL can use GPU acceleration. Perhaps in the future, the current Classic + NL can be extended to a GPU version and the processing speed accelerated. At this time, the proposed method of implementing FlowNet2 gives a much higher processing speed and at the same time, provides better accuracy.

As a deep learning-based optical flow estimation algorithm, FlowNet2 can take advantage of GPU acceleration technology, which makes the highly time-consuming optical flow estimation task much faster. As stated in the literature (Ilg et al., 2017), FlowNet2 and its sub-networks can achieve 8 to 140 FPS real time optical flow estimation on the Middlebury data set with an NVIDIA GeForce GTX 1080 Graphics Card. This means that FlowNet2 and its sub-networks can be easily implemented to do real time full field displacement measurement. From this experiment, it can be clearly seen that the vision-based method using Classic + NL algorithm is much slower than using FlowNet2.

4. Field application

4.1. Experimental setup

In this section, a field application is presented that was performed on a footbridge on campus of the University of

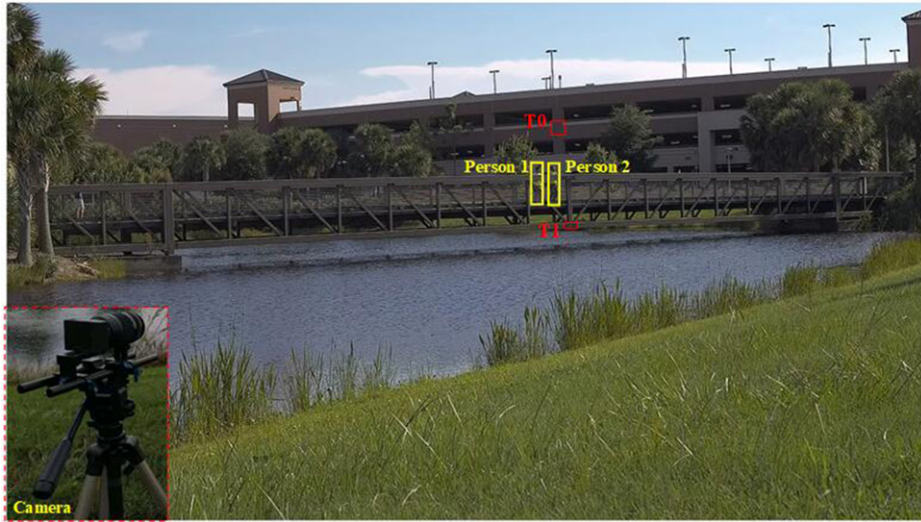


Figure 15. Experimental setup for a footbridge.

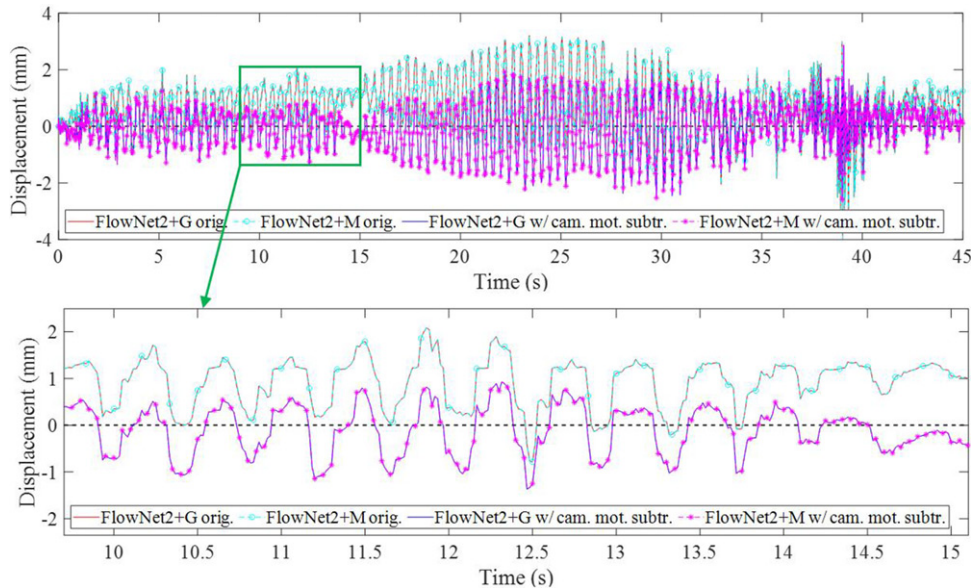


Figure 16. Displacement time histories obtained from the proposed methods. FlowNet2 + G orig and FlowNet2 + M orig represent the original displacement data obtained using FlowNet2 with Gaussian kernel and Mean kernel, respectively. While FlowNet2 + G w/cam. mot. subtr and FlowNet2 + M w/cam. mot. subtr represent those with camera motion subtraction.

Central Florida (UCF) to verify the feasibility of the proposed displacement measurement methods. The structure (Figure 15) is a three-span (7.31 m + 39.01 m + 7.3 m) truss bridge with a width of 3.65 m. A portable camera (Z Camera E1) with a resolution of 1980×1080 pixels, a frame rate of 60 FPS and an Olympus zoom lens with a focal length of 75–300 mm was used to collect images during the experiment. The measurement location is at midspan and T1 (see figure) was selected as the region for the vision-based measurement. The distance from the camera to the measurement region was about 52 m. An accelerometer was installed at midspan to measure the vibration. The sampling rate of the accelerometer was 200 Hz. Point T0 in the background was selected as the static reference and used to eliminate the camera motion caused by ground vibration and wind effects. During the experiment, two persons jumped at

the bridge midspan and both the camera and the accelerometer recorded the vibration of the bridge.

4.2. Analysis and results

Figure 16 illustrates the displacement time histories obtained from the proposed methods. Here both Mean kernel and Gaussian kernel are used to calculate the displacement at midspan. In this experiment, using Mean kernel and Gaussian kernel give almost identical displacement results. The red curve (-) and the cyan dashed curve with circle (-o-) show the original displacement at T1 without doing camera motion subtraction, while the blue curve (-) and the magenta dashed curve with asterisks (-*-) show the displacement at T1 with camera motion subtraction. Further,

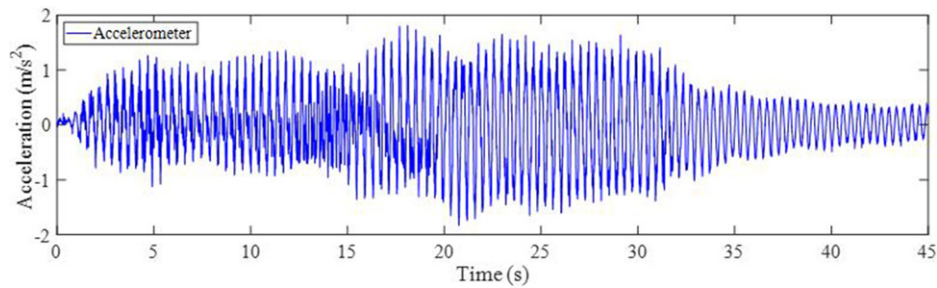


Figure 17. Acceleration time history obtained from accelerometer.

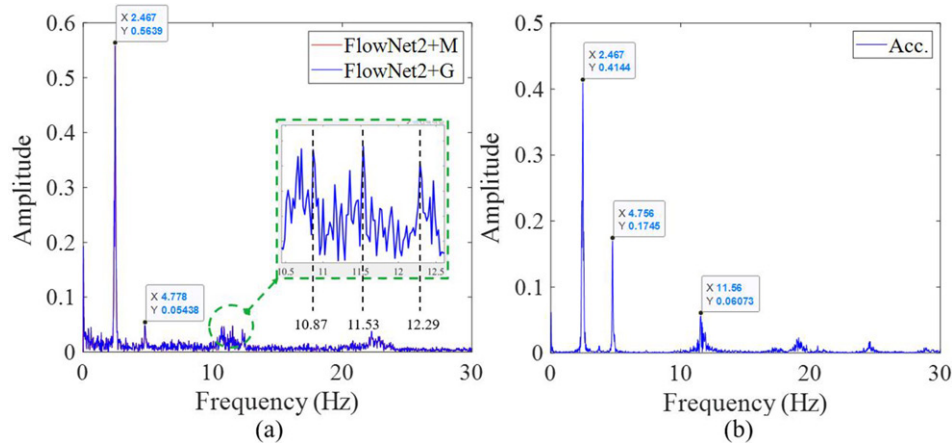


Figure 18. Operational frequencies of bridge under human jumping excitation: (a) vision-based system; (b) accelerometer signal.

the black dashed curve is added to show the zero line. In this figure, FlowNet2 + G org and FlowNet2 + M org represent the original displacement data obtained using FlowNet2 with Gaussian Kernel and mean kernel, respectively. While FlowNet2 + G w/cam. mot. subtr and FlowNet2 + M w/cam. mot. subtr represent the ones with camera motion subtraction. Please note that FlowNet2 + G org and FlowNet2 + M org as well as FlowNet2 + G org and FlowNet2 + M org are almost identical in Figure 16.

After camera motion subtraction, the structural vibration varies up and down by around ± 2 mm. The camera motion subtraction shifts the displacement downwards by about 1.5 mm, especially in the first 32 seconds, and the range is reduced by about 38%. Clearly the camera motions caused by ground vibration and wind have a substantial influence on the displacement measurement and it is necessary to correct for the camera motions. Technological advances in cameras might provide vibration reduction features, which can also be considered for camera shake correction in the future.

Figure 17 illustrates the acceleration time history obtained from the accelerometer installed at T1. A comparison was conducted between the displacement from the proposed methods and the measured acceleration in the frequency domain. The portable camera used in this experiment provides a uniform sampling rate so that there is no need to process the displacement data using non-uniform frequency analysis methods. By directly applying an FFT to the displacement and acceleration data, the frequency

spectra are obtained and are shown in Figure 18. By using a peak-picking method, the operational modal frequencies of the footbridge under human loads are extracted and summarised in Table 5.

In this study, the peak frequencies are simply those observed from displacement and acceleration data. There was no investigation to determine whether they are caused by human jumping or are natural modal frequencies. The first three operational modes are listed and for the modes extracted from the proposed methods, they are very close to those from the accelerometer. However, around the third mode (i.e. 11.53 Hz), there are two more pseudo-modes (10.87 Hz and 12.29 Hz) which make it hard to pick the right operational mode.

Compared with the third natural frequency measured by the accelerometer (i.e. 11.56 Hz), it is assumed that 11.53 Hz in the vision-based signal, is most likely to be the third operational mode. The pseudo-modes may come from the camera motion caused by wind or ground motion. Even through camera motion subtraction is applied in this case, there may still be vibration which cannot be removed completely. From Table 5, it can be seen that the difference between the modal frequencies obtained from the proposed methods and the accelerometer are all less than 0.5%. This gives a lot of confidence that the proposed method is accurate and capable of use in field application.

Table 5 Comparative study of operational modal frequencies from vision and accelerometer signals

Table 5. Comparative study of operational modal frequencies from vision and accelerometer signals.

Operational mode	f (Hz): Vision	f (Hz): Acc.	Difference between Vision and Acc.
1	2.467	2.467	0.00%
2	4.778	4.756	0.46%
3	10.87, 11.53, 12.29	11.56	5.97%, 0.26%, 6.31%

It should be noted that the camera motion effects are not completely eliminated as shown in Figures 17 and 18, even camera motion subtraction is applied. Camera motion subtraction is partially efficient in getting rid of the tripod/camera vibration effects. The pseudo modes can be removed by filtering out the vibration frequencies of the tripod/camera setup. While they are not measured in this study and it is one of the limitations.

4.3. Recommendations for practice in field application

In field application, the measurement environment is different to that in a laboratory, and the following are recommended:

1. Camera motion: The influence of camera motion needs to be minimised. The effect of wind and ground vibration should be reduced by careful selection of the camera location. It is useful to include a stationary object in the field of view to facilitate camera motion subtraction. Putting a triaxial accelerometer on the camera and filtering out the camera motion effects are not inconvenient in field application.
2. Background clutter: Background clutter (e.g. due to leaves moving in the wind) should be avoided because, when calculating full field optical flow, the motions in the background may cause difficulties. It tends to reduce the accuracy of the flow prediction of the measurement.
3. Region selection of target: The target in the image should be sufficiently large and it should be ensured that no other moving object is inside the region that is not part of the target.
4. Kernel selection: For simplicity and convenience, the mean kernel can be selected since there is very little difference in the use of mean kernel and Gaussian kernel in this experiment.
5. Camera calibration: Drawings of the structure should be used or dimensions measured to facilitate image calibration.

5. Conclusions

To achieve non-contact displacement monitoring for civil structures with less user involvement and to overcome the limitations of common vision-based methods, a novel full field structural displacement measurement method using deep learning-based optical flow, 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 conclusions are as follows:

1. A procedure for vision-based displacement measurement is presented and provides a standard reference for future users.
2. A deep learning-based full field optical flow algorithm, FlowNet2, is implemented in the proposed approach. It decreases the requirement for human involvement in the operation and gives more accurate measurement results with less computation time.
3. Issues in vision-based methods for real-time monitoring and post processing are explored and strategies for the use of portable cameras, industrial cameras, triggers and time control are presented. The non-uniform sampling problems are discussed, and camera trigger, spline interpolation, Lomb-Scargle method are recommended to solve the problems.
4. Strategies for displacement calculation in common vision-based methods are discussed, specifically the issue of whether to calculate the motion between consecutive images or between the current image and the initial one. To reduce drift caused by an accumulation of errors in calculating the differences between consecutive images, the authors recommend the latter approach.
5. The camera motion issue is discussed in the context of field application. Camera motion subtraction is proposed to address the errors induced by camera motion.

In the future, further work will be done to process the non-uniform sampled image data and explore the application of kernels in calculating displacements at discrete structural points. Additional study will be focused on the investigation of bridge deflection profile, full field structural modal analysis and distribution factor calculation using the proposed method. Furthermore, how shading and illumination affect the proposed method will be evaluated.

Acknowledgements

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.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The financial support for this research was provided by NSF Division of Civil, Mechanical and Manufacturing Innovation [grant number 1463493].

Nomenclature

X	image coordinates
x	real world coordinates
H	homography matrix
<i>s</i>	scale factor
h	vector form of H
G	Gaussian kernel
M	mean kernel
<i>d</i>	displacement
$\text{FIT}_{\text{NRMSE}}$	goodness of fit using normalised root mean square error
NCC	normalised cross-correlation
σ	standard deviation

References

- Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3), 346–359. doi:10.1016/j.cviu.2007.09.014
- Bengio, Y., Goodfellow, I., & Courville, A. (2017). *Deep learning*. MIT Press. Retrieved from www.deeplearningbook.org (last accessed online on 8/9/2019)
- Bouguet, J. (1999). *Pyramidal implementation of the Lucas Kanade Feature Tracker Description of the algorithm*. Intel Corporation. OpenCV Documents. doi:10.1109/HPDC.2004.1323531 (http://robots.stanford.edu/cs223b04/algo_tracking.pdf; last accessed online on 8/9/2019)
- Catbas, F. N., Dong, C. Z., Celik, O., & Khuc, T. (2018). A vision for vision-based technologies for bridge health monitoring. In Maintenance, Safety, Risk, Management and Life-Cycle Performance of Bridges. Proceedings of the Ninth International Conference on Bridge Maintenance, Safety and Management (IABMAS 2018), (p. 54). Melbourne, Australia. Edited By Nigel Powers, Dan M. Frangopol, Riadh Al-Mahaidi, Colin Caprani; CRC Press.
- Celik, O., Dong, C. Z., & Catbas, F. N. (2018a). Measurement of human loads using computer vision. In *IMAC36: It's not just modal anymore: Engineering extremes: Unifying concepts in shock, vibration, and nonlinear mechanics*. Orlando. Springer Publishers.
- Celik, O., Dong, C. Z., & Catbas, F. N. (2018b). A computer vision approach for the load time history estimation of lively individuals and crowds. *Computers & Structures*, 200, 32–52. doi:10.1016/j.compstruc.2018.02.001
- Chen, J. G., Adams, T. M., Sun, H., Bell, E. S., & Büyüköztürk, O. (2018). Camera-based vibration measurement of the World War I Memorial Bridge in Portsmouth, New Hampshire. *Journal of Structural Engineering*, 144(11), 04018207. doi:10.1061/(ASCE)ST.1943-541X.0002203
- Chen, J. G., Wadhwa, N., Cha, Y. J., Durand, F., Freeman, W. T., & Buyukozturk, O. (2015). Modal identification of simple structures with high-speed video using motion magnification. *Journal of Sound and Vibration*, 345, 58–71. doi:10.1016/j.jsv.2015.01.024
- Chen, Y., Joffre, D., & Avitabile, P. (2018). Underwater dynamic response at limited points expanded to full-field strain response. *Journal of Vibration and Acoustics*, 140(5), 051016. doi:10.1115/1.4039800
- Chen, Y., Logan, P., Avitabile, P., & Dodson, J. (2019). Non-model based expansion from limited points to an augmented set of points using Chebyshev polynomials. *Experimental Techniques*, 1–23, August 2019 (<https://doi.org/10.1007/s40799-019-00340-0>)
- Chen, Y., Zhang, B., & Chen, S. (2014). Model reduction technique tailored to the dynamic analysis of a beam structure under a moving load. *Shock and Vibration*, 2014, 1–13. doi:10.1155/2014/406093
- Chen, Y., Zhang, B., Zhang, N., & Zheng, M. (2015). A condensation method for the dynamic analysis of vertical vehicle-track interaction considering vehicle flexibility. *Journal of Vibration and Acoustics*, 137(4), 041010. doi:10.1115/1.4029947
- da Silva, I., Ibañez, W., & Poleszuk, G. (2018). Experience of using total station and GNSS technologies for tall building construction monitoring. In: Rodrigues H., Elnashai A., Calvi G. (eds). *Facing the Challenges in Structural Engineering*. GeoMEast 2017. Sustainable Civil Infrastructures. Springer, Cham.
- Dong, C. Z., Celik, O., & Catbas, F. N. (2017). Marker free development of computer vision based structural dynamic monitoring and modal analysis. Paper presented at the 2017 Engineering Mechanics Institute Conference (EMI 2017), San Diego.
- Dong, C. Z., Celik, O., & Catbas, F. N. (2018). Marker free monitoring of the grandstand structures and modal identification using computer vision methods. *Structural Health Monitoring*, 1–19. doi:10.1177/1475921718806895
- Dong, C. Z., Ye, X. W., & Jin, T. (2018). Identification of structural dynamic characteristics based on machine vision technology. *Measurement: Journal of the International Measurement Confederation*, 126, 405.
- Dong, C. Z., Ye, X. W., & Liu, T. (2015). Non-contact structural vibration monitoring under varying environmental conditions. *Vibroengineering Procedia*, 5, 217–222.
- Dosovitskiy, A., Fischer, P., Ilg, E., Philip, H., Hazirbas, C., Golkov, V., ... Brox, T. (2015). FlowNet: Learning optical flow with convolutional networks. Paper presented at the IEEE International Conference on Computer Vision (pp. 2758–2766), Santiago, Chile.
- Ehrhart, M., & Lienhart, W. (2017). Accurate measurements with image-assisted total stations and their prerequisites. *Journal of Surveying Engineering*, 143(2), 04016024. doi:10.1061/(ASCE)SU.1943-5428.0000208
- Feng, D., & Feng, M. Q. (2015). Model updating of railway bridge using in situ dynamic displacement measurement under trainloads. *Journal of Bridge Engineering*, 20(12), 04015019. doi:10.1061/(ASCE)BE.1943-5592.0000765
- Feng, D., & Feng, M. Q. (2016). Vision-based multipoint displacement measurement for structural health monitoring. *Structural Control and Health Monitoring*, 23(5), 876–890. doi:10.1002/stc.1819
- Feng, D., & Feng, M. Q. (2017). Experimental validation of cost-effective vision-based structural health monitoring. *Mechanical Systems and Signal Processing*, 88, 199–211. doi:10.1016/j.ymssp.2016.11.021
- Feng, D., Feng, M. Q., Ozer, E., & Fukuda, Y. (2015). A vision-based sensor for noncontact structural displacement measurement. *Sensors (Switzerland)*, 15(7), 16557–16575. doi:10.3390/s150716557
- Feng, M. Q., Fukuda, Y., Feng, D., & Mizuta, M. (2015). Nontarget vision sensor for remote measurement of bridge dynamic response. *Journal of Bridge Engineering*, 20(12), 04015023. doi:10.1061/(ASCE)BE.1943-5592.0000747
- Fuchs, P. A., Washer, G. A., Chase, S. B., & Moore, M. (2004). Laser-based instrumentation for bridge load testing. *Journal of Performance of Constructed Facilities*, 18(4), 213–219. doi:10.1061/(ASCE)0887-3828(2004)18:4(213)
- Harris, C., & Stephens, M. (1988). A combined corner and edge detector. Paper presented at the 4th Alvey Vision Conference (pp. 147–151), Manchester.
- Hartley, R., & Zisserman, A. (2003). *Multiple view geometry in computer vision* (2nd ed). Cambridge University Press. Cambridge, England.
- Horn, B. K. P., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, 17(1-3), 185–203. doi:10.1016/0004-3702(81)90024-2
- Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., & Brox, T. (2017). FlowNet 2.0: Evolution of optical flow estimation with deep networks. Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition, July 21 to July 26, 2017 in Honolulu, Hawaii. (pp. 2462–2470). doi:10.1109/CVPR.2017.179
- Khaloo, A., & Lattanzi, D. (2017). Pixel-wise structural motion tracking from rectified repurposed videos. *Structural Control and Health Monitoring*, 24(11), e2009. doi:10.1002/stc.2009
- Khuc, T., & Catbas, F. N. (2016). Computer vision-based displacement and vibration monitoring without using physical target on structures. *Structure and Infrastructure Engineering*, 13(4), 1–12. doi:10.1080/15732479.2016.1164729

- Khuc, T., & Catbas, F. N. (2017). Completely contactless structural health monitoring of real-life structures using cameras and computer vision. *Structural Control and Health Monitoring*, 24(1), e1852. doi:10.1002/stc.1852
- Khuc, T., & Catbas, F. N. (2018). Structural identification using computer vision-based bridge health monitoring. *Journal of Structural Engineering*, 144(2), 04017202. doi:10.1061/(ASCE)ST.1943-541X.0001925
- Lava, P., Cooreman, S., Coppieters, S., De Strycker, M., & Debruyne, D. (2009). Assessment of measuring errors in DIC using deformation fields generated by plastic FEA. *Optics and Lasers in Engineering*, 47(7-8), 747-753. doi:10.1016/j.optlaseng.2009.03.007
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. doi:10.1038/nature14539
- Lee, J. J., Cho, S., & Shinozuka, M. (2006). Evaluation of bridge load carrying capacity based on dynamic displacement measurement using real-time image processing techniques. *Steel Structures*, 6, 377-385.
- Lomb, N. R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39(2), 447-462. doi:10.1007/BF00648343
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91-110. doi:10.1023/B:VISI.0000029664.99615.94
- Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. Paper presented at the 7th International Joint Conference on Artificial Intelligence (IJCAI), Vancouver, British Columbia.
- Luo, L., & Feng, M. Q. (2018). Edge-enhanced matching for gradient-based computer vision displacement measurement. *Computer-Aided Civil and Infrastructure Engineering*, 33(12), 1019-1040. doi:10.1111/mice.12415
- Lydon, D., Lydon, M., del Rincon, J. M., Taylor, S., Robinson, D., O'Brien, E., & Catbas, N. (2018). Development and field testing of a time-synchronized system for multi-point displacement calculation using low cost wireless vision-based sensors. *IEEE Sensors Journal*, 1748(18), 9744-9754. doi:10.1109/JSEN.2018.2853646
- Lydon, D., Lydon, M., Taylor, S., Martinez, J., Rincon, D., Hester, D., & Brownjohn, J. (2019). Development and field testing of a vision-based displacement system using a low cost wireless action camera. *Mechanical Systems and Signal Processing*, 121, 343-358. doi:10.1016/j.ymsp.2018.11.015
- Moreu, F., Li, J., Jo, H., Kim, R. E., Scola, S., Spencer, B. F., & LaFave, J. M. (2016). Reference-free displacements for condition assessment of timber railroad bridges. *Journal of Bridge Engineering*, 21(2), 04015052. doi:10.1061/(ASCE)BE.1943-5592.0000805
- O'Byrne, M., Ghosh, B., Schoefs, F., O'Donnell, D., Wright, R., & Akrashi, V. (2015). Acquisition and analysis of dynamic responses of a historic pedestrian bridge using video image processing. *Journal of Physics: Conference Series*, 628, 012053. doi:10.1088/1742-6596/628/1/012053
- Ojio, T., Carey, C. H., O'Brien, E. J., Doherty, C., & Taylor, S. E. (2016). Contactless bridge weigh-in-motion. *Journal of Bridge Engineering*, 21(7), 04016032. doi:10.1061/(ASCE)BE.1943-5592.0000776
- Omidalizarandi, M., Kargoll, B., Paffenholz, J. A., & Neumann, I. (2018). Accurate vision-based displacement and vibration analysis of bridge structures by means of an image-assisted total station. *Advances in Mechanical Engineering*, 10(6), 1-19. doi:10.1177/1687814018780052
- OpenCV. (2019). OpenCV 4.1.0 document: Video analysis. Retrieved from https://www.docs.opencv.org/master/d3/db0/tutorial_py_table_of_contents_video.html
- Oppenheim, A. V., Willsky, A. S., & Nawab, S. H. (1996). *Signals and systems* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall, Inc.
- Pan, B., Qian, K., Xie, H., & Asundi, A. (2009). Two-dimensional digital image correlation for in-plane displacement and strain measurement: A review. *Measurement Science and Technology*, 20(6), 062001. doi:10.1088/0957-0233/20/6/062001
- Pan, B., Tian, L., & Song, X. (2016). Real-time, non-contact and targetless measurement of vertical deflection of bridges using off-axis digital image correlation. *NDT & E International*, 79, 73-80. doi:10.1016/j.ndteint.2015.12.006
- Stiros, S., Psimoulis, P., & Kokkinou E. (2008). Errors Introduced by Fluctuations in the Sampling Rate of Automatically Recording Instruments: Experimental and Theoretical Approach. *ASCE Journal of Surveying Engineering* 134(3), [https://doi.org/10.1061/\(ASCE\)0733-9453\(2008\)134:3\(89\)](https://doi.org/10.1061/(ASCE)0733-9453(2008)134:3(89))
- Shi, J., & Tomasi, C. (1994). Good features to track. Proceedings CVPR'94., 1994 IEEE Computer Society Conference (pp. 593-600). Seattle, WA, doi:10.1109/CVPR.1994.323794
- Sun, D., Roth, S., & Black, M. J. (2010). Secrets of optical flow estimation and their principles. Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 2432-2439). San Francisco, CA, doi:10.1109/CVPR.2010.5539939
- Sun, D., Roth, S., & Black, M. J. (2014). A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision*, 106(2), 115-137. doi:10.1007/s11263-013-0644-x
- Sutton, M. A., Yan, J. H., Tiwari, V., Schreier, H. W., & Orteu, J. J. (2008). The effect of out-of-plane motion on 2D and 3D digital image correlation measurements. *Optics and Lasers in Engineering*, 46(10), 746-757. doi:10.1016/j.optlaseng.2008.05.005
- Tian, L., & Pan, B. (2016). Remote bridge deflection measurement using an advanced video deflectometer and actively illuminated LED targets. *Sensors (Switzerland)*, 16(9), 1344-1313. doi:10.3390/s16091344
- Wilburn, B., Joshi, N., Vaish, V., Levoy, M., & Horowitz, M. (2005). High-speed videography using a dense camera array. Paper presented at the ACM SIGGRAPH 2005 (pp. 765-776). Los Angeles, CA.
- Wu, L. J., Casciati, F., & Casciati, S. (2014). Dynamic testing of a laboratory model via vision-based sensing. *Engineering Structures*, 60, 113-125. doi:10.1016/j.engstruct.2013.12.002
- Xu, Y., & Brownjohn, J. M. W. (2018). Review of machine-vision based methodologies for displacement measurement in civil structures. *Journal of Civil Structural Health Monitoring*, 8(1), 91-110. doi:10.1007/s13349-017-0261-4
- Xu, Y., Brownjohn, J., & Kong, D. (2018). A non-contact vision-based system for multipoint displacement monitoring in a cable-stayed footbridge. *Structural Control and Health Monitoring*, 25(5), e2155-23. doi:10.1002/stc.2155
- Yang, Y., Dorn, C., Mancini, T., Talken, Z., Kenyon, G., Farrar, C., & Mascareñas, D. (2017). Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification. *Mechanical Systems and Signal Processing*, 85, 567-590. doi:10.1016/j.ymsp.2016.08.041
- Yang, Y., Dorn, C., Mancini, T., Talken, Z., Nagarajaiah, S., Kenyon, G., Farrar, C., & Mascareñas, D. (2017). Blind identification of full-field vibration modes of output-only structures from uniformly-sampled, possibly temporally-aliased (sub-Nyquist), video measurements. *Journal of Sound and Vibration*, 390, 232-256. doi:10.1016/j.jsv.2016.11.034
- Yang, Y., Dorn, C., Mancini, T., Talken, Z., Theiler, J., Kenyon, G., Farrar, C., & Mascareñas, D. (2018). Reference-free detection of minute, non-visible, damage using full-field, high-resolution mode shapes output-only identified from digital videos of structures. *Structural Health Monitoring*, 17(3), 514-531. doi:10.1177/1475921717704385
- Ye, X. W., Dong, C. Z., & Liu, T. (2016a). A Review of machine vision-based structural health monitoring: Methodologies and applications. *Journal of Sensors*, 2016, 7103039. doi:10.1155/2016/7103039
- Ye, X. W., Dong, C. Z., & Liu, T. (2016b). Force monitoring of steel cables using vision-based sensing technology: methodology and experimental verification. *Smart Structures and Systems*, 18(3), 585-599. doi:10.12989/sss.2016.18.3.585
- Ye, X. W., Dong, C. Z., & Liu, T. (2016c). Image-based structural dynamic displacement measurement using different multi-object

- tracking algorithms. *Smart Structures and Systems*, 17(6), 935–956. doi:10.12989/sss.2016.17.6.935
- Ye, X. W., Yi, T. H., Dong, C. Z., & Liu, T. (2016). Vision-based structural displacement measurement: System performance evaluation and influence factor analysis. *Measurement: Journal of the International Measurement Confederation*, 88, 372–384.
- Ye, X. W., Yi, T.-H., Dong, C. Z., Liu, T., & Bai, H. (2015). Multi-point displacement monitoring of bridges using a vision-based approach. *Wind and Structures*, 20(2), 315. doi:10.12989/was.2015.20.2.315
- Yoon, H., Elanwar, H., Choi, H., Golparvar-Fard, M., & Spencer, B. F. (2016). Target-free approach for vision-based structural system identification using consumer-grade cameras. *Structural Control and Health Monitoring*, 23(12), 1405–1416. doi:10.1002/stc.1850
- Yoon, H., Shin, J., & Spencer, B. F. (2018). Structural displacement measurement using an unmanned aerial system. *Computer-Aided Civil and Infrastructure Engineering*, 33(3), 183–192. doi:10.1111/mice.12338
- Zhang, K., Zhang, L., Yang, M.-H., & Zhang, D. (2013). Fast tracking via spatio-temporal context learning. *ArXiv Preprint ArXiv*. Zurich, Switzerland, Retrieved from https://doi.org/10.1007/978-3-319-10602-1_9
- Zhang, Z. (2000). A flexible new technique for camera calibration (Technical Report). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330–1334. doi:10.1109/34.888718
- Zhong, F. Q., & Quan, C. G. (2017). Digital image correlation in polar coordinate robust to a large rotation. *Optics and Lasers in Engineering*, 98, 153–158. doi:10.1016/j.optlaseng.2017.06.019
- Zhong, F. Q., & Quan, C. G. (2018). A single color camera stereo vision system. *IEEE Sensors Journal*, 18(4), 1474–1482. doi:10.1109/JSEN.2017.2786747
- Zhong, F. Q., He, S. P., & Li, B. (2017). Blob analyzation-based template matching algorithm for LED chip localization. *The International Journal of Advanced Manufacturing Technology*, 93(1-4), 55–63. doi:10.1007/s00170-015-7638-5
- Zhong, F. Q., Shao, X. X., & Quan, C. (2018). 3D digital image correlation using a single 3CCD colour camera and dichroic filter. *Measurement Science and Technology*, 29(4), 045401. doi:10.1088/1361-6501/aaab02