Peer-reviewed version available at Sensors 2019, 19, 3197; doi:10.3390/s19143197

Article

A Robust Vision-based Method for Displacement Measurement under Adverse Environmental Factors using Spatio-Temporal Context Learning and Taylor Approximation

Chuan-Zhi Dong [1], Ozan Celik [2], F. Necati Catbas [3,*], Eugene OBrien [4] and Su Taylor [5]

[1] Ph.D. Candidate, Department of Civil, Environmental, and Construction Engineering, University of Central Florida, 12800 Pegasus Drive, Suite 211, Orlando, Florida 32816, USA; ceczdong@knights.ucf.edu
[2] Ph.D., Department of Civil, Environmental, and Construction Engineering, University of Central Florida, 12800 Pegasus Drive, Suite 211, Orlando, Florida 32816, USA; ozan.celik@knights.ucf.edu
[3,*] Professor, Department of Civil, Environmental, and Construction Engineering, University of Central Florida, 12800 Pegasus Drive, Suite 211, Orlando, Florida 32816, USA; catbas@ucf.edu
[4] Professor, The School of Civil Engineering, University College Dublin D04V1W8, Ireland; eugene.obrien@ucd.ie
[5] Professor, The School of Natural and Built Environment, Queens University Belfast, BT95AG, UK; S.E.Taylor@qub.ac.uk
* Correspondence: catbas@ucf.edu

Received: date; Accepted: date; Published: date

Abstract: Currently the majority of studies on vision-based measurement has been conducted under ideal environments so that an adequate measurement performance and accuracy is ensured. However, vision-based systems may face some adverse influencing factors such as illumination change and fog interference, which can affect the measurement accuracy. This paper develops a robust vision-based displacement measurement method which can handle the two common and important adverse factors given above and achieve sensitivity at the subpixel level. The proposed method leverages the advantage of high-resolution imaging incorporating spatial and temporal context aspects. To validate the feasibility, stability and robustness of the proposed method, a series of experiments was conducted on a two-span three-lane bridge in the laboratory. The illumination change and fog interference are simulated experimentally in the laboratory. The results of the proposed method are compared to conventional displacement sensor data and current vision-based method results. It is demonstrated that the proposed method gives better measurement results than the current ones under illumination change and fog interference.

Keywords: structural health monitoring, displacement measurement, non-contact, computer vision, environmental factors, spatio-temporal context, Taylor approximation

1. Introduction

1.1. Background

Computer vision-based displacement measurement using cameras has attracted increasing attention in the community of structural health monitoring (SHM) because of its being a non-contact, long distance, multi-point, high precision, time saving and cost effective sensing technique [1,2,11–16,3–10]. Structural displacement is a critical indicator for evaluating performance, identifying and determining the effects of damage/change under external loads. For instance, during the regular operation of a structure, the displacement can be monitored to ensure that it stays within a specified tolerance and safety range [17]. Once the displacement time histories from the monitored

structures are extracted using vision-based methods, traditional structural health monitoring and behavior analysis [18] can easily be done. Vision-based displacement measurement methods are also applied for bridge load testing to evaluate the bridge load carrying capacity [19] and have even been used for contactless bridge weigh-in-motion [20]. Combining the multi-point displacement response with structural input data extracted from vehicle tracking, structural identification can be carried out using traditional structural indicators such as the unit influence line (UIL) and unit influence surface (UIS) [21,22]. Without the need for the deployment of conventional sensor networks, operational modal analysis can be performed using vision-based displacement measurement methods, which may provide multi-point synchronization and therefore a much denser spatial resolution than is practical with conventional sensors [4,23–27]. Full field motion estimation and instantaneous mode shapes can even be obtained with high spatial and temporal resolution [28–31]. Modal properties and other indices derived from the vision-based displacement time histories can be turned into sensitive indicators for structural damage detection and model updating [32–34]. There are also numerous studies relating to the estimation of stay cable forces that use vision-based displacement measurement [35,36]. In addition to the structural response monitoring, the external loading information can be predicted. Celik et al. [14] estimated the load time histories of individuals and crowds with the displacement time histories obtained using computer vision-based. These successful research applications make the computer vision-based displacement methods a very promising complementary tool to conventional structural health monitoring practices, particularly for bridges.

1.2. Motivations and objectives

The majority of applications and experiments in the literature are conducted in an ideal measurement environment so that an adequate measurement performance and accuracy is ensured. In addition, when these experiments are performed for the purpose of new method verification or comparison, the measurement time span is generally short and the adverse factors which can influence the measurement accuracy and stability are mostly avoided. For a general proof of concept, it makes sense to conduct such studies. However, when vision-based systems are intended for long-term deployment, either as standalone or to complement a conventional SHM system, some unfavorable contingencies may affect the measurement quality. Even in short term, the accuracy and stability of a vision-based system can be affected adversely. In a review of current literature, Feng and Feng [8] summarize the possible measurement error sources in vision-based methods, including: 1) camera motion; 2) coordinate conversion; 3) hardware limitations; and 4) environmental sources. Brownjohn et al. [12] investigate the challenges in field application of a commercial vision-based system resulting from camera instability, the nature of the target (artificial or structural feature), and illumination. Ye et al. [37] review the state-of-art on systematic errors, assessment and reduction, including: 1) target size and texture, 2) camera alignment; 3) motion blur; and 4) the ratio between target size and full view. Xu and Brownjohn [11] review subpixel techniques used in vision-based displacement measurement methods. Ma et al. [38] study the measurement error in the digital image correlation method caused by self-heating of digital cameras. Ye et al. [39] conduct a series of shaking table experiments in the laboratory to examine the environmental influence factors which affect the accuracy and stability of vision-based systems. The target used in the experiments are QR (quick response) code and the texture of QR codes shows rich sparkle patterns. It is suggested that the measurement results are adversely affected by illumination and vapor. Subsequently, Dong and Ye [40,41] investigate the possibility of improving the accuracy and the stability of vision-based system with the adverse factor of vapor. They use light emitting diodes (LED) and infrared emitting diodes as the measurement target and the experimental results show that these emitting diodes can mitigate the adverse effects of vapor. However, installing these kinds of target on the structure can be difficult, perhaps requiring wiring and a mains power supply, which may not be feasible for a bridge.

These problems may decrease the accuracy of the measurement results and affect the evaluation of structural performance and health condition by monitoring using vision-based systems in a

long-time span. In the literature, there are lots of studies on the analysis of sources of error, but only a few [37,39] seek to improve system performance under adverse influencing factors. Therefore, it is essential to develop a robust vision-based displacement measurement method for long-term structural monitoring, which can handle some of these adverse factors.

While one study cannot address all issues related to computer vision-based monitoring, this paper focuses on the mitigation of environmental factors such as illumination change and fog interference, and improvement of the measurement sensitivity at subpixel level. A robust vision-based displacement measurement method is developed, leveraging the advantages of high-resolution imaging and computer vision techniques to mitigate the interferences induced by illumination change and fog and be adapted for long-term bridge monitoring. The proposed method utilizes the spatio-temporal context (STC) learning algorithm to track the measurement objects in image sequences and obtain the locations. The STC algorithm [42] builds the spatio-temporal relationships between the measurement target and its local context based on a Bayesian framework, which models the statistical correlation between the low-level features (i.e., image intensity and position) from the measurement target and its surrounding regions. The tracking problem is solved by computing a confidence map and obtaining the best target location by maximizing an object location likelihood function. Combining this with the Taylor approximation [43], the accuracy of the proposed method achieves subpixel level without sacrificing processing speed. The objectives of this study are: 1) developing a new vision-based displacement method using spatio-temporal context learning; 2) achieving a subpixel level estimation based on a Taylor approximation for the new vision-based method; and 3) verifying the feasibility, stability and robustness of the proposed method via comparison with the current vision-based methods and conventional displacement sensor (Linear Variable Differential Transformer, LVDT) by conducting a series of experiments under two adverse environmental factors (illumination change and fog) on a two-span three-lane model bridge in the laboratory.

2. Methodology

2.1. General procedure of the vision-based displacement measurement methods

The key aspect of vision-based displacement measurement methods is to convert the measurement of the target motion in the image into actual motion with physical units such as millimeters. In the literature, researchers propose different procedures of vision-based displacement measurement. In this study, the authors summarize the state-of-the-art and present a general procedure as illustrated in Figure 1. There are four steps in this procedure, namely: 1) camera calibration; 2) initial image capture and target selection; 3) visual tracking and 4) scale transformation and displacement calculation.

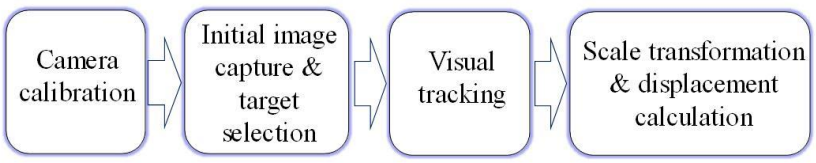

Figure 1. General procedure of the vision-based displacement measurement methods.

In the first step, with the camera calibration, the relationship between the image entity and physical world object is built. A camera is set up with the measurement targets included in the field of view. This is also the step where camera calibration is conducted. If the camera is wide angle, it can cause image distortion [44]. Yoon et al. [13] and Brownjohn et al. [12] implement a flexible technique which was proposed by using the camera to observe a planar calibration object in a few different views to easily calibrate a camera and resolve the distortion problems. The authors also recommend implementing this method to resolve the distortion problems. In this study the authors select a distortion-free camera so that the distortion problems are not an issue. With Zhang's method [44] it is

also possible to convert the image entities into physical world objects since, during the camera calibration, the planar homography matrix between the image and the physical world is obtained. In this paper, instead of planar homography matrix, for convenience, the authors use the scale factor to convert the target motion in pixels to physical units of millimeter. When the optical axis is perpendicular to the object motion plane, the scale factor, s, is

$$s = \frac{D}{d}, \tag{1}$$

where D is the physical length of the object on the motion plane and d is the length in pixels of its corresponding image part. When there is an angle between the target motion plane and optical axis, θ, the scale factor has to be modified by:

$$s = \frac{D}{d\cos\theta}, \tag{2}$$

More details about the scale factor calculation can be found from the literature [9,14,17,35,37,39–41,45,46].

In the second step, the region to be tracked which includes the measurement targets in the field of view of the initial image is selected. According to the visual tracking methods used in the third step, image preprocessing is utilized to extract useful features from the tracking region. Researchers who use digital image correlation (DIC) either in the frequency domain [6] or the spatial domain [27] select the target regions as the patterns and the low-level features of images play the role of visual tracking features. Other researchers extract key points and descriptors such as Shi-Tomasi corners, SURF (Speeded Up Robust Features), SIFT (Scale-Invariant Feature Transform), FREAK (Fast REtinA Keypoint), etc. as the tracking features and the corresponding visual tracking method includes Lucas-Kanade optical flow estimation and key point matching based on nearest neighborhood approximation [9,10,13]. The third step is to track the selected targets in subsequent images captured by the camera continuously and locate the targets' new positions. In this paper, the authors implement the spatio-temporal context (STC) learning method to do the visual tracking. The horizontal and vertical displacements in pixels – x_t - x_0 and y_t - y_0 respectively – are found by subtracting the coordinates of the initial target position (x_0, y_0) from the current target position, (x_t, y_t). When pattern matching methods such as DIC, are used in this step, the displacements in pixels are always integer values. One way to increase the sensitivity and to improve the measurement accuracy is by applying subpixel techniques. For instance, Feng et al. [6] implement upsampled cross correlation in local region to get the displacement at a subpixel level. In this paper, the authors utilize the Taylor approximation method to achieve the subpixel level without upsampling the images and without sacrificing the image processing speed. Finally, with the scale factor, s, and the displacement in pixels, the actual displacement at time t of the physical unit is obtained: (x_t - x_0)s, horizontally, and (y_t - y_0)s, vertically. The visual tracking method and subpixel estimation used in this paper are introduced in detail in the next sections.

2.2. Visual tracking using spatio-temporal context (STC) learning

The spatio-temporal relationship between the local scenes containing the target in consecutive frames can be used to model the statistical correlation between the low-level features, such as image intensity and position, extracted from the target and its local context [42]. As illustrated in the footbridge example of Figure 2., the yellow (smaller) box is the target to be tracked and the red (larger) box is the local context.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 June 2019 doi:10.20944/preprints201906.0023.v1

Peer-reviewed version available at Sensors 2019, 19, 3197; doi:10.3390/s19143197

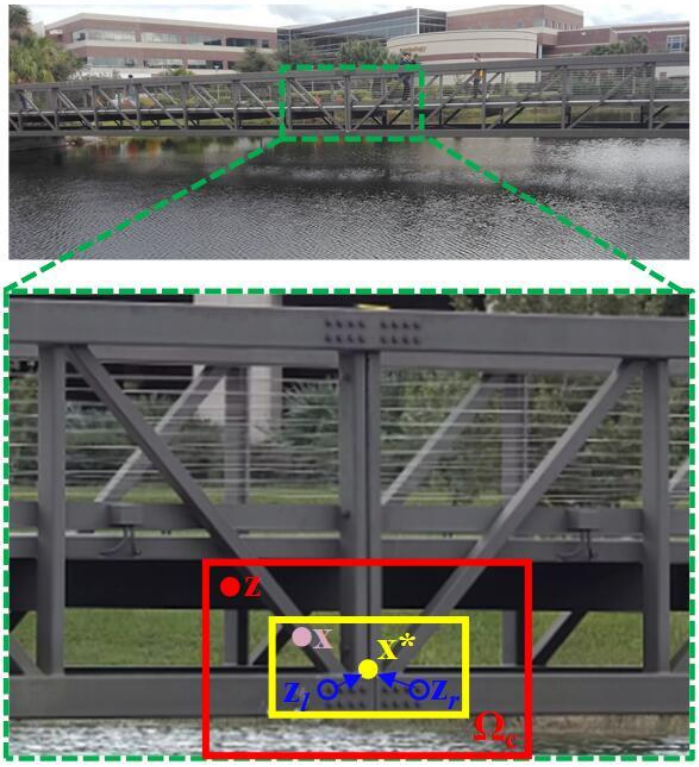

Figure 2. Graphical model of spatial relationship between the target and its surrounding context.

The visual tracking task can be obtained by maximizing an object location likelihood function c(**x**) as [42]:

$$\begin{aligned} c(\mathbf{x}) &= P(\mathbf{x}\,|\,o) = \sum_{c(\mathbf{z})\in X^c} P(\mathbf{x}, \mathbf{c}(\mathbf{z})\,|\,o) \\ &= \sum_{c(\mathbf{z})\in X^c} P(\mathbf{x}\,|\,\mathbf{c}(\mathbf{z}), o) P(\mathbf{c}(\mathbf{z})\,|\,o)' \end{aligned} \tag{3}$$

where **x** is the target location which can be represented with the coordinates defined above, (x, y) and o denotes the target present in the scene. The context feature set, X^c, is defined as:

$$X^c = \{\mathbf{c}(\mathbf{z}) = (I(\mathbf{z}), \mathbf{z}) \,|\, \mathbf{z} \in \Omega_c(\mathbf{x}^*)\}, \tag{4}$$

where I(**z**) denotes the image intensity at location **z** and Ω_c(**x***) is the neighborhood of location **x***. The conditional probability P(**x**|**c**(**z**), o) in Eq. (3) models the spatial relationship between the object location and its context information. It can help to resolve ambiguities when the image measurements allow different interpretations which are introduced in the following parts. P(**c**(**z**)|o) is a context prior probability which models the appearance in the local context.

The conditional probability P(**x**|**c**(**z**), o) in Eq. (3) is defined as:

$$P(\mathbf{x}\,|\,\mathbf{c}(\mathbf{z}), o) = h^{sc}(\mathbf{x} - \mathbf{z}), \tag{5}$$

where h^{sc}(**x**-**z**) is a function only of the direction and the relative distance between the target location **x** and its local context location **z**, which means this function contains the spatial relationship between the target and its spatial context. Eq. (5) defines the spatial context model. It is worth noting that Eq. (5) is not a radially symmetric function which means that h^{sc}(**x**-**z**) is not equal to h^{sc}(|**x**-**z**|). It considers different spatial relationships between the target and its local context, which facilitates the solving of the ambiguities when similar objects appear in close proximity. As shown in Figure 2, when a visual tracking method tries to track a bolt based only on the appearance denoted by $\mathbf{z}_l$, it might be distracted to the right one denoted by $\mathbf{z}_r$ because both bolts and the local surroundings have a similar appearance. This would cause ambiguities and consequently decrease the tracking accuracy, especially when the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 June 2019 doi:10.20944/preprints201906.0023.v1

Peer-reviewed version available at Sensors 2019, 19, 3197; doi:10.3390/s19143197

target moves fast and the search region is not small. With the non-radially symmetric characteristics of $h^{sc}(\mathbf{x}\text{-}\mathbf{z})$, the ambiguities can be resolved.

In Eq. (3), $P(\mathbf{c}(\mathbf{z})|o)$ can be calculated according to the target location that has been initialized manually in the first frame. It is modeled by:

$$P\left(\mathbf{c}\left(\mathbf{z}\right)|o\right)=I\left(\mathbf{z}\right)w_{\sigma}\left(\mathbf{z}-\mathbf{x}^{*}\right), \tag{6}$$

where $w_\sigma(\cdot)$ is a weighted function defined by:

$$w_{\sigma}\left(\mathbf{z}\right)=ae^{-\frac{|\mathbf{z}|^{2}}{\sigma^{2}}}, \tag{7}$$

In Eq. (7) a is the normalization constant which restricts $P(\mathbf{c}(\mathbf{z})|o)$ to be in the range from 0 to 1 and σ is a scale parameter. Eq. (6) ensures that, the closer the context location $\mathbf{z}$ is to the current tracking target location $\mathbf{x}^*$, the more important it is to predict the location and a greater weight is set.

The confidence map of an object location is modeled as:

$$c\left(\mathbf{x}\right)=P\left(\mathbf{x}|o\right)=be^{-\left|\frac{\mathbf{x}-\mathbf{x}^{*}}{\alpha}\right|^{\beta}}, \tag{8}$$

where b is a normalization constant, α is a scale parameter and β is a shape parameter. According to the literature [42] robust results can be obtained when $\beta = 1$. Based on the context prior model in Eq. (6) and the confidence map function in Eq. (8), the objective is to learn the spatial context model, i.e. Eq. (3). Combining Eqs. (3), (5), (6) and (8), gives:

$$\begin{aligned} c\left(\mathbf{x}\right)&=P\left(\mathbf{x}|o\right)=be^{-\left|\frac{\mathbf{x}-\mathbf{x}^{*}}{\alpha}\right|^{\beta}}\\ &=\sum\nolimits_{\mathbf{z}\in\Omega_{c}\left(\mathbf{x}^{*}\right)}h^{sc}\left(\mathbf{x}-\mathbf{z}\right)I\left(\mathbf{z}\right)w_{\sigma}\left(\mathbf{z}-\mathbf{x}^{*}\right),\\ &=h^{sc}\left(\mathbf{x}\right)\otimes I\left(\mathbf{x}\right)w_{\sigma}\left(\mathbf{x}-\mathbf{x}^{*}\right) \end{aligned} \tag{9}$$

where $\otimes$ denotes the convolution operator. The Fast Fourier Transform (FFT), Eq. (9) transforms the function to the frequency domain:

$$\mathcal{F}\left(be^{-\left|\frac{\mathbf{x}-\mathbf{x}^{*}}{\alpha}\right|^{\beta}}\right)=\mathcal{F}\left(h^{sc}\left(\mathbf{x}\right)\right)\odot\mathcal{F}\left(I\left(\mathbf{x}\right)w_{\sigma}\left(\mathbf{x}-\mathbf{x}^{*}\right)\right), \tag{10}$$

where $\mathcal{F}$ denotes the FFT function and $\odot$ is the element-wise product. Then the spatial context model is:

$$h^{sc}\left(\mathbf{x}\right)=\mathcal{F}^{-1}\left(\frac{\mathcal{F}\left(be^{-\left|\frac{\mathbf{x}-\mathbf{x}^{*}}{\alpha}\right|^{\beta}}\right)}{\mathcal{F}\left(I\left(\mathbf{x}\right)w_{\sigma}\left(\mathbf{x}-\mathbf{x}^{*}\right)\right)}\right), \tag{11}$$

where $\mathcal{F}^{-1}$ denotes the inverse FFT function. Then in the whole image sequence, the spatio-temporal context model of the $(t+1)^{\text{th}}$ frame, H_{t+1}^{stc}, can be updated using the spatio-temporal context model, H_t^{stc}, and the spatial context model, h_t^{sc}, of the t^{th} frame. It is formulated as:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 June 2019 doi:10.20944/preprints201906.0023.v1

Peer-reviewed version available at Sensors 2019, 19, 3197; doi:10.3390/s19143197

$$H_{t+1}^{stc} = (1-\rho) H_t^{stc} + \rho h_t^{sc}, \tag{12}$$

where ρ is a learning parameter and t denotes the t^{th} frame. It should be noted that in the first frame, i.e., when t is equal to 1, the spatio-temporal context model H_t^{stc} is equal to the spatial context model, h_t^{sc}.

Finally, the target location $\mathbf{x}_{t+1}^*$ in the $(t+1)^{th}$ frame is determined by maximizing the new confidence map:

$$\mathbf{x}_{t+1}^* = \arg\max_{\mathbf{x} \in \Omega_c(\mathbf{x}_t^*)} c_{t+1}(\mathbf{x}), \tag{13}$$

Deduced from Eq. (10), the new confidence map $c_{t+1}(\mathbf{x})$ is represented as:

$$c_{t+1}(\mathbf{x}) = \mathcal{F}^{-1}\left(\mathcal{F}\left(H_{t+1}^{stc}(\mathbf{x})\right) \odot \mathcal{F}\left(I_{t+1}(\mathbf{x}) w_{\sigma_t}(\mathbf{x} - \mathbf{x}^*)\right)\right), \tag{14}$$

As the scale of the target may tend to change over time, the scale parameter σ in the weight function w_σ in Eq. (7) is updated by:

$$\begin{cases} s_t' = \sqrt{\dfrac{c_t(x_t^*)}{c_{t-1}(x_{t-1}^*)}} \\ \overline{s}_t = \dfrac{1}{n}\sum_{i=1}^{n} s_{t-i}' \\ s_{t+1} = (1-\lambda) s_t + \lambda \overline{s}_t \\ \sigma_{t+1} = s_t \sigma_t \end{cases}, \tag{15}$$

Where s_t' is the estimated scale between two consecutive frames. The estimated target scale s_{t+1} is obtained through filtering in which $\overline{s}_t$ is the average of the estimated scale from n consecutive frames to avoid oversensitive adaptation and to reduce noise, and $\lambda > 0$ is a fixed parameter. More details about the derivation can be found in the literature [42]. In general, the scale updating should be considered, but in this study, only in-plane motion is considered for two-dimensional displacement measurement, so that scale updating is neglected. If this method is used to do three-dimensional displacement measurement which means there is out-plane motion, scale updating has to be considered.

To obtain robust tracking results, the reference gives rules of thumb regarding the selection of the parameters used in STC tracking: $\alpha = 2.25$, $\beta = 1$, $\rho = 0.075$, $s_1 = 1$, $\lambda = 0.25$, and $n = 5$. Additionally, for Eq. (12), with the deductions,

$$H_{t+1}^{stc} = (1-\rho) H_t^{stc} + \rho h_t^{sc};$$

$$\int H_{t+1}^{stc} e^{-jwt} dt = \int (1-\rho) H_t^{stc} e^{-jwt} dt + \int \rho h_t^{sc} e^{-jwt} dt;'$$

$$\begin{aligned} \text{LHS} &= \int H_{t+1}^{stc} e^{-jwt} dt \overset{t \to t-1}{=} \int H_t^{stc} e^{-jw(t-1)} d(t-1) \\ &= e^{jw} \int H_t^{stc} e^{-jwt} dt = e^{jw} H_w^{stc}; \\ \text{RHS} &= \int (1-\rho) H_t^{stc} e^{-jwt} dt + \int \rho h_t^{sc} e^{-jwt} dt \\ &= (1-\rho) \int H_t^{stc} e^{-jwt} dt + \rho \int h_t^{sc} e^{-jwt} dt \\ &= (1-\rho) H_w^{stc} + \rho h_w^{sc}; \end{aligned}$$

Peer-reviewed version available at Sensors 2019, 19, 3197; doi:10.3390/s19143197

$$\text{LHS} = \text{RHS};$$

$$H_w^{stc} = \frac{\rho}{e^{\mathrm{j}w} - (1-\rho)} h_w^{sc}$$

238 a temporal filtering procedure can be easily obtained in the frequency domain, which is,

$$H_w^{stc} = \frac{\rho}{e^{\mathrm{j}w} - (1-\rho)} h_w^{sc}, \tag{16}$$

239 where

$$H_w^{stc} = \int H_t^{stc} e^{-\mathrm{j}wt} dt, \tag{17}$$

240 is the temporal Fourier transform of H_t^{stc} and similar to h_w^{sc}. The temporal filter can be represented
241 by,

$$F_w = \frac{\rho}{e^{\mathrm{j}w} - (1-\rho)}, \tag{18}$$

242 which is a low-pass filter [47]. With this low-pass filter, the spatio-temporal context model is able to
243 filter out image noise caused by appearance variations and this leads to more stable measurement
244 results. The properties of the spatio-temporal model contribute to the resolution of the adverse effects
245 induced by environmental factors such as illumination change and fog.

246 *2.3. Subpixel level estimation using Taylor approximation*

247 With Eq. (13), the targets can easily be tracked in the image sequence, but the change of locations
248 can only be obtained with integer pixel values. To achieve subpixel level motion, the Taylor
249 approximation method is applied to solve the optical flow estimation. Assuming there are two
250 consecutive images, $f(x, y)$ and $g(x, y)$, with a shift $(\Delta x, \Delta y)$, the following estimation applies:

$$\begin{aligned} g(x,y) &= f(x+\Delta x, y+\Delta y) \\ &\approx f(x,y) + \Delta x \frac{\partial}{\partial x} f(x,y) + \Delta y \frac{\partial}{\partial y} f(x,y)' \end{aligned} \tag{19}$$

251 which is the first order Taylor series approximation. The shift in the image can be calculated by
252 minimizing the sum of squared errors (SSE):

$$\arg\min \Phi(\Delta x, \Delta y), \tag{20}$$

253 where

$$\Phi(x,y) = \sum_{x,y} \left[g(x,y) - f(x,y) - \Delta x \frac{\partial}{\partial x} f(x,y) - \Delta y \frac{\partial}{\partial y} f(x,y) \right]^2, \tag{21}$$

254 Using the ordinary least squares (OLS) method to solve Eq. (20), the optimal Δx and Δy can be
255 determined by setting the partial derivatives of Eq. (21) to zero, i.e.,

$$\begin{cases} \dfrac{\partial \Phi}{\partial \Delta x} = 0 \\ \dfrac{\partial \Phi}{\partial \Delta y} = 0 \end{cases}, \tag{22}$$

256 Combining Eqs. (21) and (22) gives the system of linear equations:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 June 2019 doi:10.20944/preprints201906.0023.v1

Peer-reviewed version available at Sensors 2019, 19, 3197; doi:10.3390/s19143197

$$\begin{bmatrix} \sum_{x,y}\left(\frac{\partial f}{\partial x}\right)^2 & \sum_{x,y}\frac{\partial f}{\partial x}\frac{\partial f}{\partial y} \\ \sum_{x,y}\frac{\partial f}{\partial x}\frac{\partial f}{\partial y} & \sum_{x,y}\left(\frac{\partial f}{\partial y}\right)^2 \end{bmatrix} \begin{Bmatrix} \Delta x \\ \Delta y \end{Bmatrix} = \begin{Bmatrix} \sum_{x,y}(g-f)\frac{\partial f}{\partial x} \\ \sum_{x,y}(g-f)\frac{\partial f}{\partial y} \end{Bmatrix}, \quad (23)$$

The optimal shift, (Δx, Δy), is obtained by solving Eq. (23). It should be noted that to make the Taylor approximation valid, the assumptions, $|\Delta x| \ll 1$ and $|\Delta y| \ll 1$ has to be satisfied. When small motion is estimated, i.e. motion less than one pixel, the assumption holds. The procedure simplified from optical flow estimation is called Taylor approximation here and it will be utilized to solve the subpixel level motion estimation [43].

Figures 3. and 4. illustrate the proposed motion estimation at the subpixel level. At first, the spatio-temporal context (STC) tracking method is employed to determine the integer pixel displacements, ($\overline{\Delta x}$, $\overline{\Delta y}$).

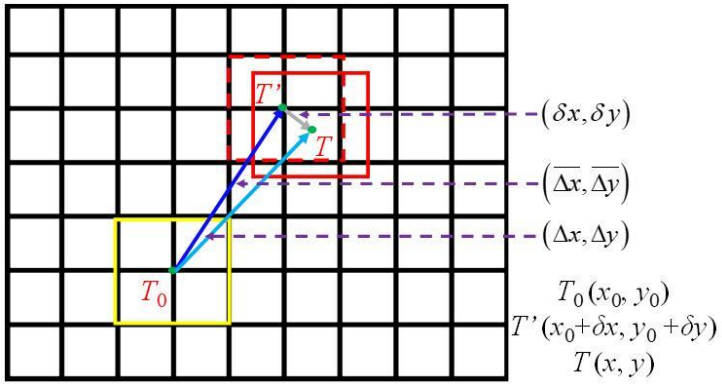

Figure 3. Sketches of motion estimation using STC tracking and Taylor approximation.

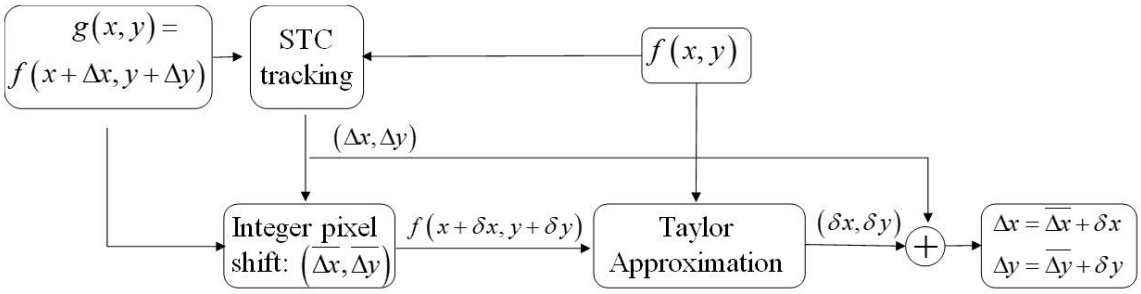

Figure 4. Flowchart of STC based subpixel tracking using Taylor approximation.

In Figure 3., the yellow solid line box represents the original target location in the initial frame and the red dashed line box represents the target recognized in the current frame using STC tracking which has an accuracy at subpixel level. Here the centers of the targets are used to represent their locations, i.e., T_0 and T'. Assuming the real target in the current frame is the red solid line box at location T, the true displacements are (Δx, Δy). Then the displacements ($\overline{\Delta x}$, $\overline{\Delta y}$) are the integer estimations of the true displacements, (Δx, Δy). The differences between ($\overline{\Delta x}$, $\overline{\Delta y}$) and (Δx, Δy) are (δx, δy), from T' to T, where $|\delta x| < 1$ and $|\delta y| < 1$. And the assumption of using the Taylor approximation is satisfied with the condition of $|\delta x| < 1$ and $|\delta y| < 1$. Secondly, the Taylor approximation is employed to estimate the displacement between the target tracked by STC (red dashed line box) and the real target (in red solid line box), i.e., (δx, δy). Finally, the total displacements are

Peer-reviewed version available at Sensors 2019, 19, 3197; doi:10.3390/s19143197

$$\begin{cases} \Delta x = \overline{\Delta x} + \delta x \\ \Delta y = \overline{\Delta y} + \delta y \end{cases}, \tag{24}$$

According to the literature [43], the Taylor approximation gives an error bound of less than 0.0125 pixel, without any image upsampling and the error is much smaller than that of the normal template matching methods using image upsampling. The feasibility and performance of the proposed method for structural displacement will be verified through laboratory experiments in the next sections.

3. Laboratory verification

3.1. Experimental setup

Figure 5 shows the two-span bridge model constructed in the University of Central Florida's Civil Infrastructure Technologies for Resilience and Safety (CITRS) Experimental Design and Monitoring (EDM) laboratory. The bridge is a scaled down model of a mid-sized real-life structure and toy trucks with variable weights are used to model moving loads.

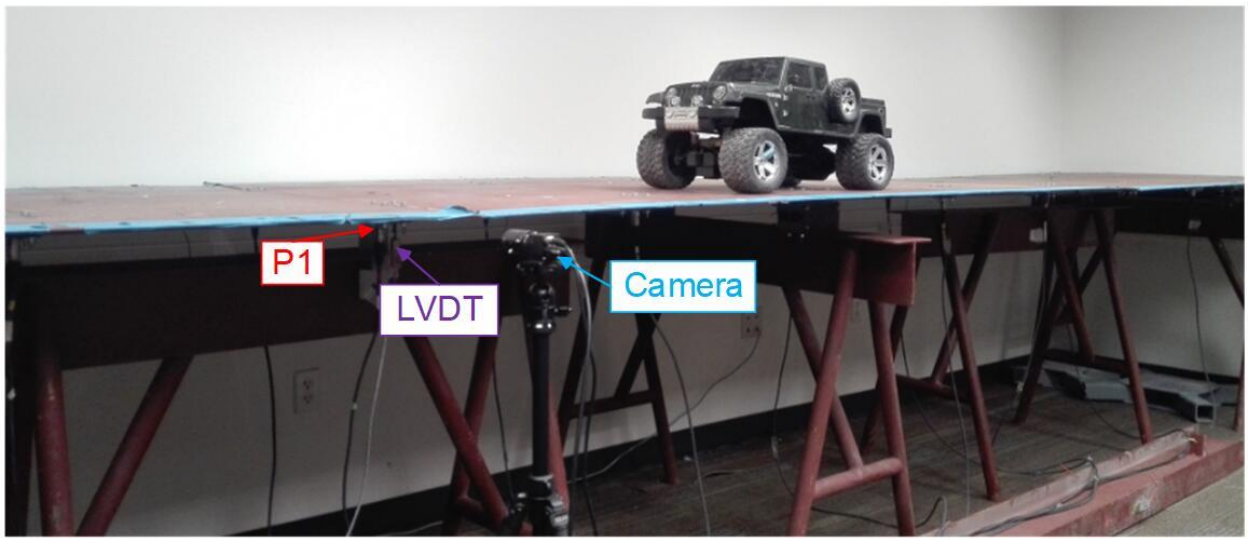

Figure 5. Experimental setup.

An industrial camera is set up in front of the bridge to record images at a measuring point (P1) during the moving load trials. An LVDT is mounted under the deck to measure the displacement of P1 and is assumed as the ground truth. During the experiments, the truck moves from one side of the bridge to the other while the LVDT and the camera record the motion of P1 (midspan of the left span) synchronously. The resolution of the camera is 1280 × 960 with the maximum frame rate of 60 FPS (frame per second). The focal length of the lens is within a zoom range of 6 ~ 60 mm. The sampling rate of the data acquisition system for LVDT is 300 Hz, which is then downsampled to 60 Hz during post processing. Three experiment cases are specified to achieve the objectives of this paper:

- Case 1: the truck moves on the bridge in ideal conditions and no adverse factors are imposed in the measuring environment;
- Case 2: the truck moves on the bridge while the illumination of the laboratory is changed several times by switching a manual controller (shown in Figure 6.). A light meter (Dr. Meter LX1010B Digital Illuminance) is used here to measure the illumination change. Normally, 9 light panels in the lab are on and the illumination is 34 lux. By turning off the 3 light panels which are close to the measurement target, the illumination drops to 18 lux. As shown in Figure

6, the left image is lighter which is taken when the illumination is 34 lux, while the right image is darker which is taken when the illumination is 18 lux;

- Case 3: a humidifier (Honeywell HUL520B Mistmate Cool Mist Humidifier) is placed between the camera and measuring targets (shown in Figure 7). The humidifier produces a mist at the maximum status to simulate natural fog. Normally, the temperature is 24 °C and the relative humidity is 49%. While in the center of the mist, the temperature is 20.3 °C and the relative humidity is 49%.

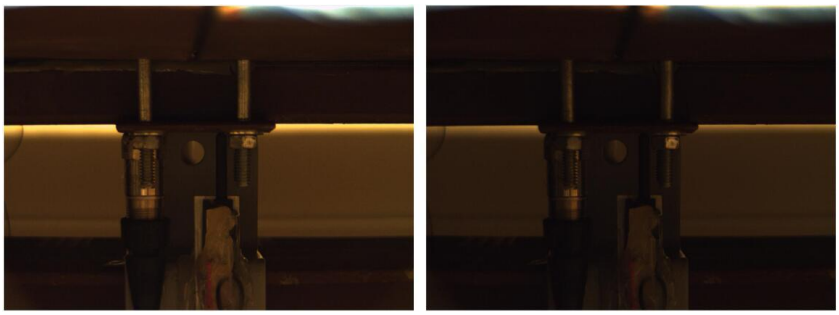

Figure 6. Illumination change.

Figure 7. Fog simulation.

3.2. Results analysis and comparative study of Case 1

The objectives for Case 1 are:

1. To evaluate the performance of the subpixel estimation method implemented in this paper;
2. To verify the feasibility and performance of the proposed method (i.e, STC tracking plus Taylor Approximation, STC-Taylor App) by comparing the conventional LVDT data with the current vision-based displacement methods, e.g., Lucas-Kanade optical flow with SURF features (LK-SURF), key point matching with Fast Library for Approximate Nearest Neighbors and SURF features (FLANN-SURF), digital image correlation (DIC). Figure 8. illustrates the vertical displacement time histories of P1 in pixel units induced by the vehicle loading when the toy truck travels along the two-span bridge.

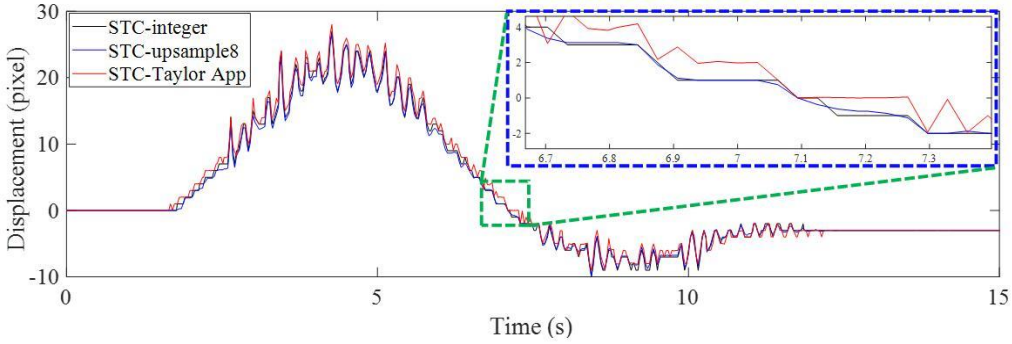

Figure 8. Vertical displacement time histories of P1 in pixel unit using non-subpixel technique, image upsampling technique and Taylor approximation technique.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 June 2019 doi:10.20944/preprints201906.0023.v1

Peer-reviewed version available at Sensors 2019, 19, 3197; doi:10.3390/s19143197

The vision-based methods used here include STC tracking with non-subpixel technique (STC-integer), STC tracking with image upsampling technique (STC-upsample8) and STC tracking with Taylor approximation (STC-Taylor App). In Figure 8, by zooming in the green dashed line box area of the displacement time histories, it is clear that the results of the STC-integer approach are rounded to integer values, i.e., (4, 3, 3, 3, 2, 1, 1, 1, 1, 1, 1, 0, 0, (-1), (-1) pixels, …). The image upsampling technique means that each image recorded during the experiment is upsampled to 8 times in the horizontal and vertical directions using bicubic interpolation. Then the minimum resolution is 1/8 = 0.125 pixel. The result of using image upsampling is a much smoother curve and shows more subpixel level displacement records. However, it still cannot provide more details about the small motion [1], especially at the very beginning and at the end. When there are no apparent loads on the structure, there is still very small structural motion induced by the random environmental loads such as wind, or machine operations nearby, ambient ground vibration etc. As illustrated in Figure 9, during the first several seconds before the toy truck moves, the displacements measured by both STC-integer and STC-upsample8 are exactly zero, which might not be true. Even though the structure is not loaded, it can still vibrate under random environmental loading. STC-Taylor App indicates the small motions of the structure caused by random environmental loadings. It is indicated that the proposed method which combines STC tracking and the Taylor approximation, has a higher sensitivity, resolution and accuracy.

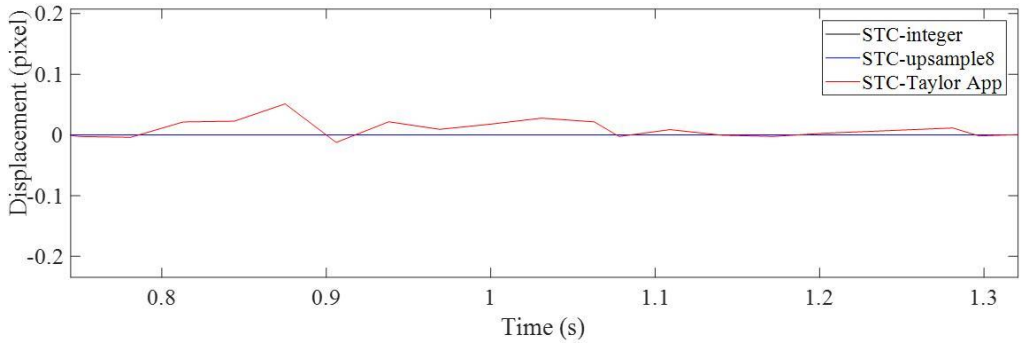

Figure 9. Zoom in the beginning part of the vertical displacement time histories of P1.

Figure 10 verifies the previous findings. In this figure, the horizontal time displacement histories show the bridge motion in the longitudinal direction induced by the moving truck impact. The motion is very small, around 1 pixel. The result from the proposed method (STC-Taylor App) gives very detailed information about the vehicle impact while the results of STC-integer and STC-upsample8 are almost zero except for one or two points, which means the bridge is stationary in the longitudinal direction. Figure 11. is a zooming into the area of the green dashed line box in Figure 10.

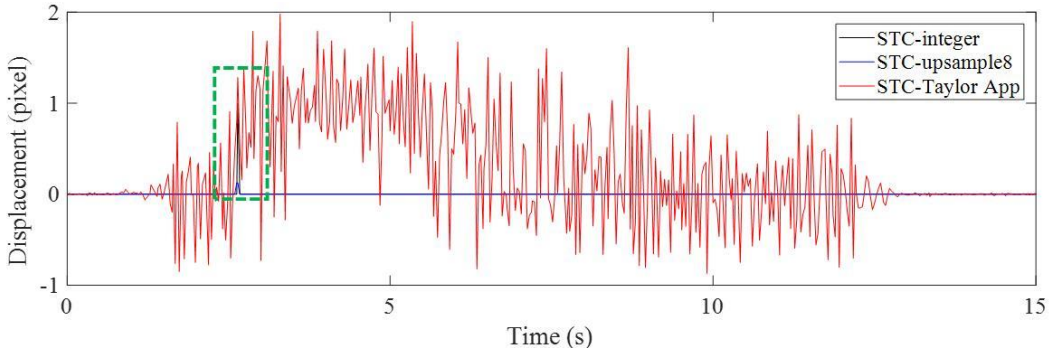

Figure 10. Horizontal displacement time histories of P1 in pixel unit using non-subpixel technique, image upsampling technique and Taylor approximation technique.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 June 2019 doi:10.20944/preprints201906.0023.v1

Peer-reviewed version available at Sensors 2019, 19, 3197; doi:10.3390/s19143197

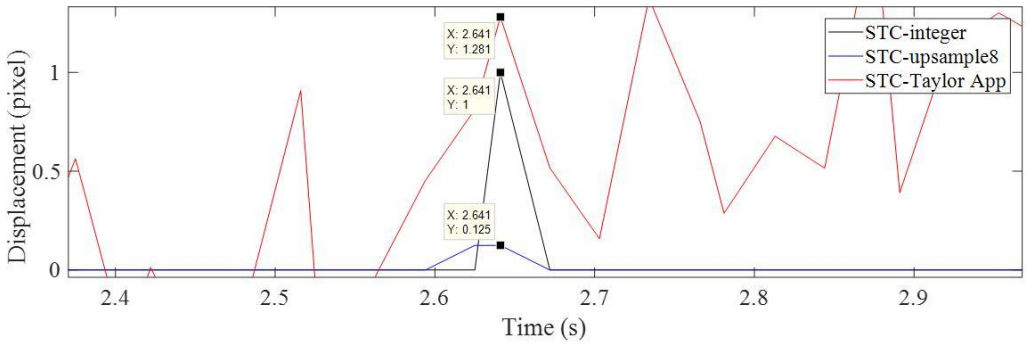

Figure 11. Zoom in the horizontal displacement time histories of P1 in pixel unit.

In the zoomed figure, the displacements of the non-zero points measured by STC-integer, STC-upsample8 and STC-Taylor App are 1, 0.125 and 1.281 pixels. For STC-integer and STC-upsample8, 1 and 0.125 are their minimum measurement resolutions and statistically these points are outliers which should be removed from the displacement time histories. In addition, the image processing speed of the proposed method is much faster than using image upsampling. Table 1. shows the elapsed processing time of one image using three different STC-based methods. The program environment is MATLAB running on a computer with the CPU of i7, 8 processors and 16G RAM. The original image has a resolution of 1280 × 960. It takes 0.0481 seconds to process one image to obtain the displacement at integer pixel level (STC-integer). However, when doing subpixel level estimation using image upsampling, it takes 2.4895 seconds, which is about (2.4895-0.0481)/0.0481 = 50.76 times that of the STC-integer. It takes only 0.0495 seconds to do this and gives even better subpixel results when using STC-Taylor App. The proposed method is about 50 times faster than using image upsampling techniques.

Table 1. This is a table. Tables should be placed in the main text near to the first time they are cited.

Methods	STC-integer	STC-upsample8	STC-Taylor App
Time (s)	0.0481	2.4895	0.0495

Overall, it is suggested that the proposed method using STC tracking and Taylor approximation can provide displacement measurements at subpixel level with high sensitivity, resolution, accuracy, and faster speed.

The next step is to convert the displacement in pixel units to physical unit, e.g., millimeter, and verify the feasibility and performance of displacement measurement by comparing the vision-based methods with the conventional displacement sensor. As illustrated in Figure 12, four vision-based displacement measurement methods (i.e., LK-SURF, FLANN-SURF, DIC and STC-Taylor App) and one conventional displacement sensor (i.e., LVDT) are used to obtain the displacement time histories of P1 when the toy truck passes over the bridge. At the very beginning, the toy truck stands at the left end of the first span, then moves to the right and approaches the measurement point P1. In the meantime, the displacement of P1 (the downward direction is positive) gradually increases to a maximum when the truck is located at P1. Then the toy truck begins to drive off P1 and keep heading to the right, while the displacement of P1 gradually decreases. When the toy truck moves to the right span, the displacement begins to be negative (i.e., upward displacement) due to loading on the other span of the two span bridge. As it approaches the right end of the right span, the absolute value of the displacement at P1 first increases and achieves a maximum and then decreases. When the toy truck arrives at the right end of the bridge, the displacement of P1 becomes stable but does not go back to zero. This is because the rear axle still rests on the bridge.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 June 2019 doi:10.20944/preprints201906.0023.v1

Peer-reviewed version available at Sensors 2019, 19, 3197; doi:10.3390/s19143197

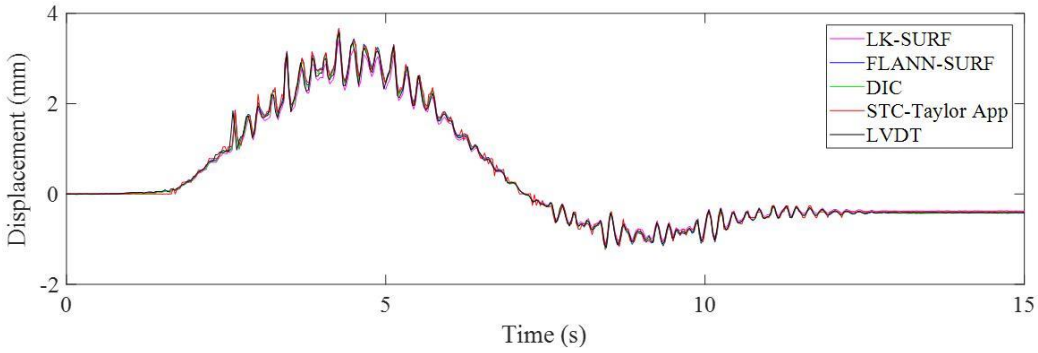

Figure 12. Case 1: displacement time histories of P1 obtained from different methods.

By comparing the displacement time histories, it is easy to see that the result obtained from the proposed method (i.e., STC+Taylor App) is quite consistent with those obtained from the LVDT and other three vision-based methods. Figure 13 illustrates the correlation matrix of these time displacement histories for Case 1. The figures on the diagonal of the correlation matrix are the histograms of the displacement time histories whereas the others are data plots and linear fits between the displacement time histories from the two methods. The correlation matrix is symmetric, and the last row and the last column give the correlation coefficients between the displacement data obtained from the vision-based methods and the conventional displacement sensor, i.e. LVDT. The correlation coefficients of the LK-SURF, FLANN-SURF and DIC with the ground truth, i.e. LVDT are all 0.99, while the correlation coefficient between the proposed method, i.e., STC-Taylor App, and LVDT is 0.98, which is also quite good. The performance of the vision-based displacement measurement methods can also be obtained from the similarity of the histograms of each method comparing with the one of LVDT.

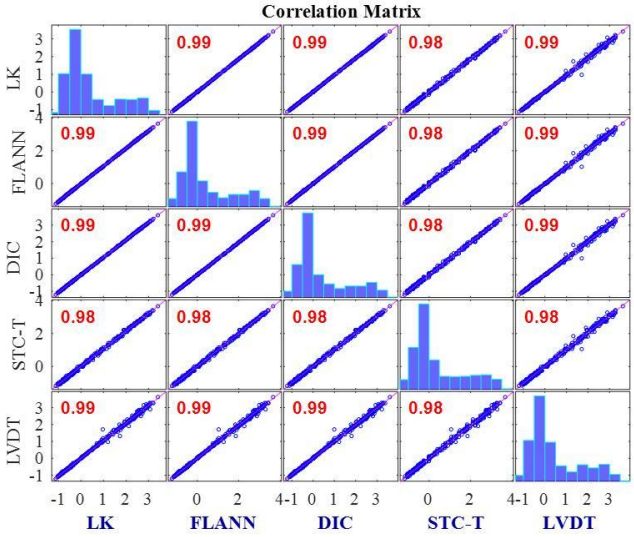

Figure 13. Correlation matrix of time displacement histories of Case 1

Here from the diagonal element of the correlation matrix, it is indicated that the histograms of these time displacement histories are highly consistent with each other. In this case, under ideal experimental conditions and no adverse factors added to the experiment, the robustness and advantages of the proposed method (STC-Taylor App) don't reveal itself. However, the displacement measurement result shows that the proposed method is immensely powerful competitor comparing with the vision-based methods. In the next two cases, the robustness and advantages of the proposed method will be verified.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 June 2019 doi:10.20944/preprints201906.0023.v1

Peer-reviewed version available at Sensors 2019, 19, 3197; doi:10.3390/s19143197

3.3. Results analysis and comparative study of Case 2

This case is designated to verify the robustness of the proposed vision-based displacement method under the adverse environmental condition: illumination change. For vision-based methods, illumination is a serious problem when conducting field applications since the image quality is easy to be affected by the illumination change. Consequently, the visual tracking performance and the displacement measurement accuracy are affected by the poor quality in the formation of images. In this experiment, the environmental illumination is determined by the fluorescent light in the lab. By turning the light switches in the laboratory on and off, the image quality is changed as shown in Figure 6. The time histories of P1 obtained from different vision-based methods and LVDT under environmental illumination change are illustrated in Figure 14.

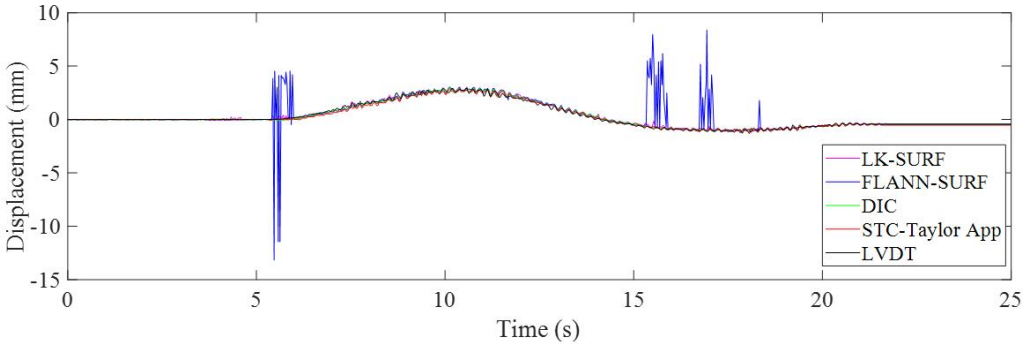

Figure 14. Case 2: displacement time histories of P1 obtained from different methods

The spikes in Figure 14 show that the vision-based method, FLANN-SURF is apparently influenced by the illumination change, which means FLANN-SURF cannot handle this kind of situation comparing to other vision-based methods. As shown in Figure 15, the correlation coefficient of the time histories between that obtained from FLANN-SURF and the ground truth, LVDT, drops to 0.84, while LK-SURF's and DIC's drops from 0.99 to 0.98 and from 0.99 to 0.97 when compared with the correlation matrix obtained in Case 1 shown in Figure 13. However, the correlation coefficient of the time histories between that obtained from the proposed method, STC-Taylor App, and the ground truth, LVDT, is still 0.98 comparing that of Case 1. From Figure 14 and Figure 15, it is indicated that the illumination change does have a significant negative implication effect on the FLANN-SURF and might also influence LK-SURF and DIC slightly. On the other hand, the proposed method, STC-Taylor App shows great robustness and is almost not influenced by the illumination change. STC-Taylor App could be a good option for long-term vision-based displacement measurement since the illumination change is a common problem in field applications.

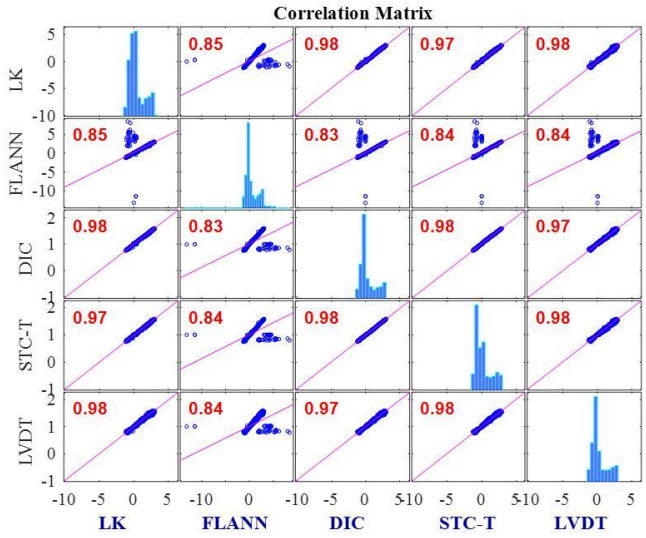

Figure 15. Correlation matrix of time displacement histories of Case 2

3.4. Results analysis and comparative study of Case 3

This case is designated to verify the robustness of the proposed vision-based displacement method under the adverse environmental condition: fog. In this experiment, the fog is simulated by the mist produced by the humidifier as shown in Figure 7. The fog not only does affect the image quality but also contaminates the image features which is the basic foundation of target recognition for visual tracking. In addition, the fog is not still but has a random motion. DIC might perform undesirably because it highly relies on the image intensity to do pattern matching and the intensity would always change under this situation. Due to the random motion of the fog, a false optical flow would be added to the real target motion which causes the optical flow method (e.g. LK method) to fail. Even though feature points, e.g., Shi-Tomasi corners, SURF, SIFT, FREAK, etc., are very robust and distinctive, their use with feature-based methods (e.g., LK-SURF and FLANN-SURF) still can have bad performance due to bad matches. The mist could block features and induce more bad matches as shown in Figure 16. It causes displacement measurement to have errors, especially when there are not enough feature points to describe the tracking targets.

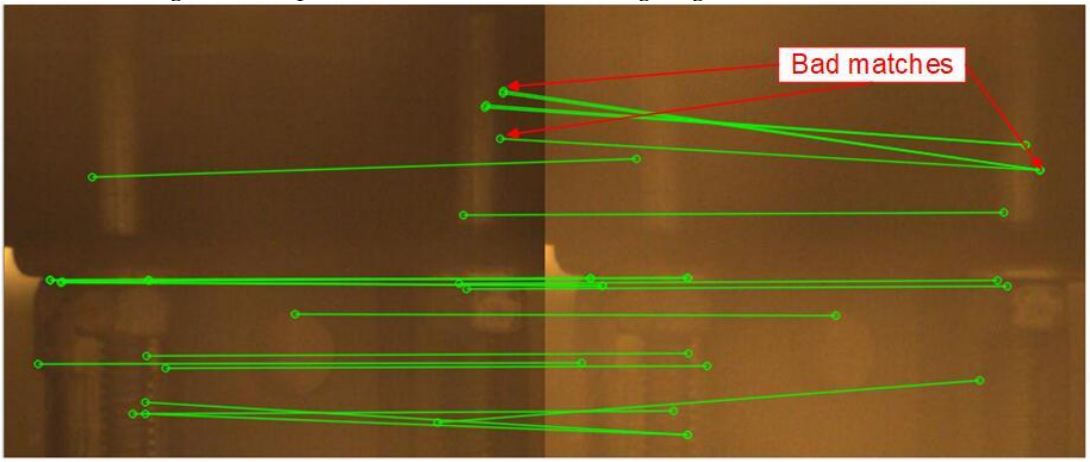

Figure 16. Poor matches when using feature-based methods

Figure 17 illustrates the time histories of P1 obtained from different vision-based methods and LVDT under fog interference. When the fog is imposed to the measurement environment, the displacement results obtained from LK-SURF and DIC provide very poor performance resulting in lots of spikes appearing in the displacement time histories. Only the results from the proposed method (STC-Taylor App) and FLANN-SURF show satisfactory performance. Zooming in the purple dashed line box area of Figure 17, more details are shown in Figure 18. In this figure, except the spikes, some data is also lost from the displacement time history when using. It is because the visual tracker based on LK-SURF loses the targets due to the fog interference. In general view, even though FLANN-SURF gives a good result, it still has outliers. Figure 18 shows an example of the outlier when using FLANN-SURF. The outlier causes more than one-millimeter error comparing with the ground truth and the result from STC-Taylor App. Statistically, it can be removed.

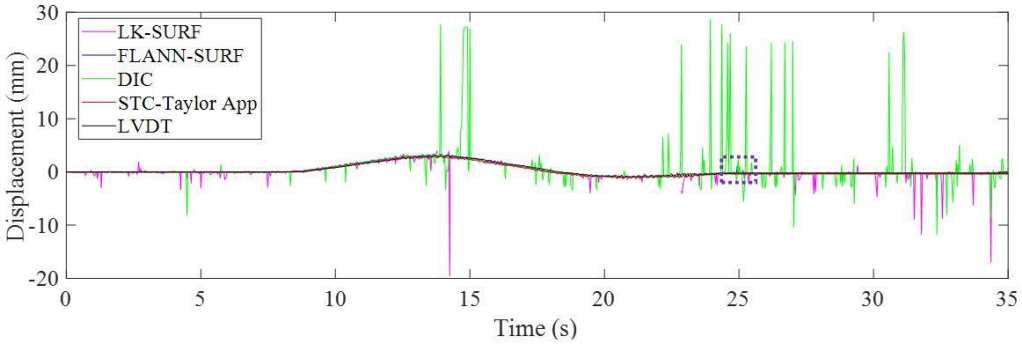

Figure 17. Case 3: displacement time histories of P1 from different methods

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 June 2019 doi:10.20944/preprints201906.0023.v1

Peer-reviewed version available at Sensors 2019, 19, 3197; doi:10.3390/s19143197

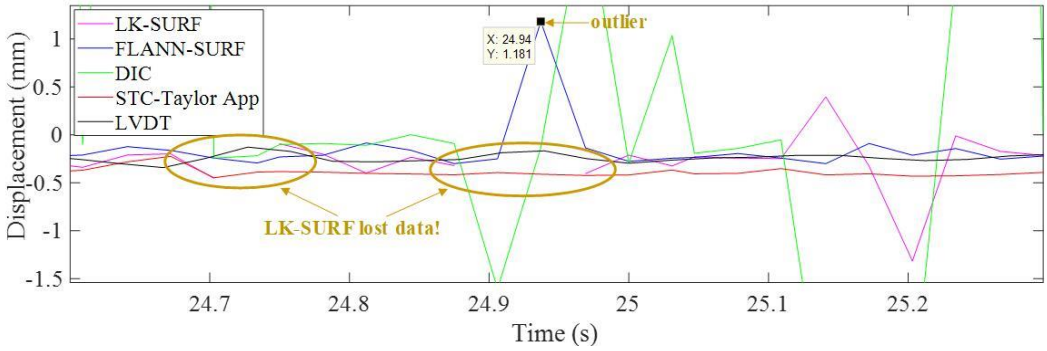

Figure 18. Zoom in the horizontal displacement time histories of P1

Figure 19 shows the correlation matrix between the vision-based methods and the conventional displacement method, i.e., LVDT. The correlation coefficient between LK-SURF and LVDT drops from 0.99 (in Case 1) to 0.84, which means the measurement error of LK-SURF increases. It is even worse for DIC, whose correlation coefficient drops from 0.99 (in Case 1) to 0.78. The linear fit plots between LK-SURF and LVDT and that between DIC and LVDT are hard to be interpreted as correlation. The correlation coefficient between the proposed method, STC-Taylor App and LVDT also drops from 0.98 (in Case 1) to 0.92. Considering the initial status, it is a little bit better than that of FLANN-SURF, since the correlation coefficient between the proposed method, STC-Taylor App and LVDT also drops from 0.99 (in Case 1) to 0.92.

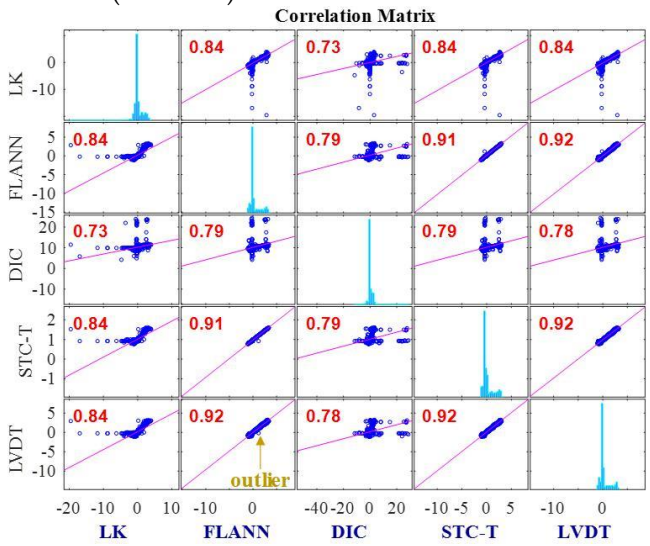

Figure 19. Correlation matrix of time displacement histories of Case 3

The outlier in displacement time history obtained by using FLANN-SURF also shows correlation in matrix plot. The fog indeed has undesirable effects on all of these vision-based methods at different levels. These bad effects might not be easy to find or be quantified in the time histories, but they apparently reveal themselves in correlation matrix. Comparing with the other three vision-based methods, the proposed method gives the best performance. The proposed vision-based displacement measurement method, i.e., STC-Taylor App, shows great robustness under fog interference. STC-Taylor App could be a good option for long-term vision-based bridge displacement measurement since the fog is a common weather problem in field application, especially when the bridge crosses a river and during the foggy season.

Considering the result analysis of Case 2 and Case 3, the proposed method shows the best performance under the two adverse environmental factors.

4. Conclusions

In this study, a robust non-contact displacement measurement method using spatio-temporal context learning and Taylor approximation is proposed. This study aims to resolve the adverse effects induced by the environmental factors such as illumination change and fog interference when using vision-based methods to conduct displacement measurements without adding manual markers or artificial light source for long-term bridge monitoring. The first method that is proposed, namely spatio-temporal context learning, leverages the advantage of image in high resolution of spatial and temporal aspects, which can be used in long-term bridge monitoring. Then, as an extension, the Taylor approximation technique is implemented into the proposed method and to improve the accuracy of the displacement at subpixel level without sacrificing the processing speed. The performance of the proposed subpixel estimation method is compared with the general image upsampling techniques and results shows that the proposed subpixel estimation method is faster than the general image upsampling technique about 50 times. Also, the precision of the proposed method is much better than the general image upsampling technique. To validate the feasibility, stability and robustness of the proposed method, a series of experiments on a two-span three-lane bridge in laboratory under the adverse environmental factors such as illumination change and fog interference are conducted. The illumination change is achieved by turning on and off the light switches in the room and the fog interference is simulated with a humidifier which can produce mist. The results from the proposed method show that:

1. In Case 1, there is no adverse environmental factors and the measurement condition is desirable for vision-based systems. The correlation coefficients of the LK-SURF, FLANN-SURF and DIC with the ground truth, i.e. LVDT are all 0.99, while the correlation coefficient between the proposed method, i.e., STC-Taylor App, and LVDT is 0.98, which is also quite good. It is indicated that that at least in the desirable measurement environment, the proposed method is a strong competitor of the current methods.
2. In Case 2, with the illumination change, the correlation coefficient of the time histories between that obtained from FLANN-SURF and the ground truth, LVDT, drop to 0.84, while LK-SURF's and DIC's just drop from 0.99 to 0.98 and from 0.99 to 0.97, respectively, comparing with the correlation matrix obtained from Case 1. However, the correlation coefficient of the time histories between that obtained from the proposed method, STC-Taylor App, and the ground truth, LVDT, is still 0.98 comparing to that of Case 1.
3. In Case 3, with the fog interference, the correlation coefficient between LK-SURF and LVDT drops from 0.99 to 0.84, while DIC's drops from 0.99 to 0.78 which is the worst. FLANN-SURF's drops from 0.99 to 0.92 and the proposed method, STC-Taylor App, drops from 0.98 to 0.92;

Combining the result analysis of the experimental results, the proposed method shows the best performance under the two adverse environmental factors, and it gives an accuracy at subpixel level without sacrificing the processing speed. By considering the spatial and temporal context learning process, the proposed method in this paper successfully mitigates the effects induced by illumination change and fog interference. Although, the benefits of the proposed method to address other real-world challenges is not explored in this paper, the proposed method may be applied to solve other adverse influencing factors such as motion blur, rain, object occlusion, out of plane movement, orientation of the camera relative to the bridge and camera motion, etc., by taking the advantage of the high spatio-temporal resolution. The computer vision-based approach along with the proposed method can be a good alternative and complementary approach to the conventional structural health monitoring practices. In the future, more studies will be carried out on real bridges to validate the feasibility of the proposed method and also to investigate other relevant challenges for long-term bridge monitoring using computer vision. Besides, in this study only one camera was used, and the proposed method was verified by tracking motion of bridge deck in two-dimensional (2D) plane, which is the limitation. In future study, the feasibility of 3D motion tracking using the proposed method will be investigated and will be tested on other applications such as long span bridge monitoring and cable vibration monitoring.

Author Contributions: The conceptualization and methodology were proposed by Dong; the validations were carried out by Dong and Celik; the formal analysis was done by Dong; the resources was provided by Catbas, writing—original draft preparation was done by Dong; writing—review and editing were done by Celik, Catbas, OBrien and Taylor; funding acquisition was did by Catbas, Obrien and Taylor.

Funding: This research was funded by NSF Division of Civil, Mechanical and Manufacturing Innovation [grant number 1463493].

Acknowledgments: The authors would like to acknowledge the 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

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dong, C.Z.; Celik, O.; Catbas, F.N. Marker free monitoring of the grandstand structures and modal identification using computer vision methods. *Struct. Heal. Monit.* **2018**, 1–19.
2. Feng, M.Q.; Fukuda, Y.; Feng, D.; Mizuta, M. Nontarget Vision Sensor for Remote Measurement of Bridge Dynamic Response. *J. Bridg. Eng.* **2015**, *20*, 4015023.
3. Fukuda, Y.; Feng, M.Q.; Shinozuka, M. Cost-effective vision-based system for monitoring dynamic response of civil engineering structures. *Struct. Control Heal. Monit.* **2010**, *17*, 918–936.
4. Feng, D.; Feng, M.Q. Vision-based multipoint displacement measurement for structural health monitoring. *Struct. Control Heal. Monit.* **2016**, 876–890.
5. Fukuda, Y.; Feng, M.Q.; Narita, Y.; Tanaka, T. Vision-Based Displacement Sensor for Monitoring Dynamic Response Using Robust Object Search Algorithm. *IEEE Sens. J.* **2013**, *13*, 4725–4732.
6. Feng, D.; Feng, M.Q.; Ozer, E.; Fukuda, Y. A vision-based sensor for noncontact structural displacement measurement. *Sensors (Switzerland)* **2015**, *15*, 16557–16575.
7. Feng, D.; Feng, M.Q. Experimental validation of cost-effective vision-based structural health monitoring. *Mech. Syst. Signal Process.* **2017**, *88*, 199–211.
8. Feng, D.; Feng, M.Q. Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection – A review. *Eng. Struct.* **2018**, *156*, 105–117.
9. Khuc, T.; Catbas, F.N. Computer vision-based displacement and vibration monitoring without using physical target on structures. *Struct. Infrastruct. Eng.* **2016**, *13*, 1–12.
10. Khuc, T.; Catbas, F.N. Completely contactless structural health monitoring of real-life structures using cameras and computer vision. *Struct. Control Heal. Monit.* **2017**, *24*, e1852.
11. Xu, Y.; Brownjohn, J.M.W. Review of machine-vision based methodologies for displacement measurement in civil structures. *J. Civ. Struct. Heal. Monit.* **2018**, *8*, 91–110.
12. Brownjohn, J.M.W.; Xu, Y.; Hester, D. Vision-Based Bridge Deformation Monitoring. *Front. Built Environ.* **2017**, *3*, 1–16.
13. Yoon, H.; Elanwar, H.; Choi, H.; Golparvar-Fard, M.; Spencer, B.F. Target-free approach for vision-based structural system identification using consumer-grade cameras. *Struct. Control Heal. Monit.* **2016**, *23*, 1405–1416.
14. Celik, O.; Dong, C.Z.; Catbas, F.N. A computer vision approach for the load time history estimation of lively individuals and crowds. *Comput. Struct.* **2018**, *200*, 32–52.
15. Lydon, D.; Taylor, S.; Martinez del Rincon, J.; Hester, D.; Lydon, M.; Robinson, D. Field investigation of contactless displacement measurement using computer vision systems for civil engineering applications. In Proceedings of the Irish Machine Vision and Image Processing Conference - NUIG; Galway, United Kingdom, 2016.
16. Chen, Y.; Joffre, D.; Avitabile, P. Underwater Dynamic Response at Limited Points Expanded to Full-Field Strain Response. *J. Vib. Acoust.* **2018**, *140*, 051016.
17. Ye, X.W.; Yi, T.-H.; Dong, C.Z.; Liu, T.; Bai, H. Multi-point displacement monitoring of bridges using a vision-based approach. *Wind Struct. An Int. J.* **2015**, *20*.
18. Chen, Y.; Logan, P.; Avitabile, P.; Dodson, J. Non-Model Based Expansion from Limited Points to an Augmented Set of Points using Chebyshev Polynomials. *Exp. Tech.* **2019**, 1–23.

19. Lee, J.J.; Cho, S.; Shinozuka, M. Evaluation of Bridge Load Carrying Capacity Based on Dynamic Displacement Measurement Using Real-time Image Processing Techniques. *Steel Struct.* **2006**, *6*, 377–385.
20. Ojio, T.; Carey, C.H.; OBrien, E.J.; Doherty, C.; Taylor, S.E. Contactless Bridge Weigh-in-Motion. *J. Bridg. Eng.* **2016**, *21*, 04016032.
21. Khuc, T.; Catbas, F.N. Structural Identification Using Computer Vision–Based Bridge Health Monitoring. *J. Struct. Eng.* **2018**, *144*, 04017202.
22. Catbas, F.N.; Dong, C.Z.; Celik, O.; Khuc, T. A Vision for Vision-based Technologies for Bridge Health Monitoring. In Proceedings of the Maintenance, Safety, Risk, Management and Life-Cycle Performance of Bridges: Proceedings of the Ninth International Conference on Bridge Maintenance, Safety and Management (IABMAS 2018),; Melbourne, Australia, 2018; p. 54.
23. Ji, Y.F.; Chang, C.C. Nontarget image-based technique for small cable vibration measurement. *J. Bridg. Eng.* **2008**, *13*, 34–42.
24. Ji, Y.F.; Chang, C.C. Nontarget Stereo Vision Technique for Spatiotemporal Response Measurement of Line-Like Structures. *J. Eng. Mech.* **2008**, *134*, 466–474.
25. Chen, C.C.; Wu, W.H.; Tseng, H.Z.; Chen, C.H.; Lai, G. Application of digital photogrammetry techniques in identifying the mode shape ratios of stay cables with multiple camcorders. *Meas. J. Int. Meas. Confed.* **2015**, *75*, 134–146.
26. Poozesh, P.; Baqersad, J.; Niezrecki, C.; Avitabile, P.; Harvey, E.; Yarala, R. Large-area photogrammetry based testing of wind turbine blades. *Mech. Syst. Signal Process.* **2017**, *86*, 98–115.
27. Dong, C.Z.; Ye, X.W.; Jin, T. Identification of structural dynamic characteristics based on machine vision technology. *Meas. J. Int. Meas. Confed.* **2018**.
28. Chen, J.G.; Wadhwa, N.; Cha, Y.J.; Durand, F.; Freeman, W.T.; Buyukozturk, O. Modal identification of simple structures with high-speed video using motion magnification. *J. Sound Vib.* **2015**, *345*, 58–71.
29. Yang, Y.; Dorn, C.; Mancini, T.; Talken, Z.; Theiler, J.; Kenyon, G.; Farrar, C.; Mascarenas, D. Reference-free detection of minute , high-resolution mode shapes output- only identified from digital videos of structures. **2017**.
30. Yang, Y.; Dorn, C.; Mancini, T.; Talken, Z.; Nagarajaiah, S.; Kenyon, G.; Farrar, C.; Mascarenas, D. Blind identification of full-field vibration modes of output-only structures from uniformly-sampled, possibly temporally-aliased (sub-Nyquist), video measurements. *J. Sound Vib.* **2017**, *390*, 232–256.
31. Yang, Y.; Dorn, C.; Mancini, T.; Talken, Z.; Kenyon, G.; Farrar, C.; Mascarenas, D. Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification. *Mech. Syst. Signal Process.* **2017**, *85*, 567–590.
32. Cha, Y.-J.; Chen, J.G.; Büyüköztürk, O. Output-only computer vision based damage detection using phase-based optical flow and unscented Kalman filters. *Eng. Struct.* **2017**, *132*, 300–313.
33. Yang, Y.; Nagarajaiah, S. Dynamic Imaging: Real-Time Detection of Local Structural Damage with Blind Separation of Low-Rank Background and Sparse Innovation. *J. Struct. Eng.* **2016**, *142*, 04015144.
34. Feng, D.; Feng, M.Q. Model Updating of Railway Bridge Using In Situ Dynamic Displacement Measurement under Trainloads. *J. Bridg. Eng.* **2015**, *20*, 1–12.
35. Ye, X.W.; Dong, C.Z.; Liu, T. Force monitoring of steel cables using vision-based sensing technology : methodology and experimental verification. *Smart Struct. Syst.* **2016**, *18*, 585–599.
36. Feng, D.; Scarangello, T.; Feng, M.Q.; Ye, Q. Cable tension force estimate using novel noncontact vision-based sensor. *Meas. J. Int. Meas. Confed.* **2017**, *99*, 44–52.
37. Ye, X.W.; Dong, C.Z.; Liu, T. A Review of Machine Vision-Based Structural Health Monitoring: Methodologies and Applications. *J. Sensors* **2016**, *2016*.
38. Ma, S.; Pang, J.; Ma, Q. The systematic error in digital image correlation induced by self-heating of a digital camera. *Meas. Sci. Technol.* **2012**, *23*, 025403.
39. Ye, X.W.; Yi, T.H.; Dong, C.Z.; Liu, T. Vision-based structural displacement measurement: System performance evaluation and influence factor analysis. *Meas. J. Int. Meas. Confed.* **2016**, *88*, 372–384.
40. Dong, C.Z.; Ye, X.W.; Liu, T. Non-contact structural vibration monitoring under varying environmental conditions. *Vibroengineering Procedia* **2015**, *5*, 217–222.
41. Ye, X.W.; Dong, C.Z.; Liu, T. Environmental effect on vision-based structural dynamic displacement monitoring. In Proceedings of the Proceedings of the Second International Conference on Performance–based and Life-cycle Structural Engineering (PLSE 2015); Dilum Fernando, Jin-Guang Teng, J.L.T., Ed.; Brisbane, Australia, 2015; pp. 261–265.

42. Zhang, K.; Zhang, L.; Yang, M.-H.; Zhang, D. Fast Tracking via Spatio-Temporal Context Learning. *arXiv Prepr. arXiv1311.1939* **2013**, 1–16.
43. Chan, S.H.; Võ, D.T.; Nguyen, T.Q. Subpixel motion estimation without interpolation. *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.* **2010**, 722–725.
44. Zhang, Z. A Flexible New Technique for Camera Calibration (Technical Report). *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *22*, 1330–1334.
45. Ye, X.W.; Dong, C.Z.; Liu, T. Image-based structural dynamic displacement measurement using different multi-object tracking algorithms. *Smart Struct. Syst.* **2016**, *17*, 935–956.
46. Ye, X.W.; Dong, C.Z. Fatigue cracking identification of reinforced concrete bridges using vision-based sensing technology. In Proceedings of the 24th Australasian Conference on the Mechanics of Structures and Materials; Taylor & Francis Group: Perth, 2017; pp. 1141–1146.
47. Oppenheim, A. V.; Willsky, A.S.; Nawab, S.H. *Signals and Systems*; 2nd ed.; Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1996; ISBN 0-13-814757-4.