Engineering Structures 163 (2018) 255–266

Contents lists available at ScienceDirect

Engineering Structures

journal homepage: www.elsevier.com/locate/engstruct

Robust vision sensor for multi-point displacement monitoring of bridges in the field

Longxi Luo[a,*], Maria Q. Feng[a], Zheng Y. Wu[b]

[a] Department of Civil Engineering and Engineering Mechanics, Columbia University, New York, NY 10027, United States
[b] Bentley Systems, Incorporated, Watertown, CT 06795, United States

ARTICLE INFO

Keywords:
Robust vision sensor
Computer vision
Multi-point displacement monitoring
Low contrast features
Camera vibration cancellation

ABSTRACT

Computer vision sensors have great potential for accurate remote displacement monitoring in the field. This paper presents InnoVision, a video image processing technique developed by the authors to address a number of difficulties associated with the application of the vision sensors to monitoring structural displacement responses in the outdoor condition that are rarely comprehensively studied in literatures. First, limited lighting condition in the field presents a challenge to tracking low contrast features on the structural surface using intensity-based template matching algorithms. For tackling this challenge, a gradient based template matching algorithm is formulated. Second, to cost-effectively monitor structural displacements at multiple points using one camera, widely used interpolation subpixel methods are investigated and incorporated into InnoVision. Third, camera vibration in the field causes displacement measurement errors. A practical solution is proposed by applying the multi-point monitoring to track both the structure and a stationary reference point. The effect of the camera vibration can be canceled by subtracting the reference displacement from the structural displacements. Several laboratory and field tests are conducted to evaluate the InnoVision's performance. One of the field tests is conducted in a challenging low lighting condition at night on a steel girder bridge to validate the robustness of InnoVision in comparison with two other vision sensing methods. Another field test is carried out on the Manhattan Bridge to demonstrate the efficacy of the proposed technique for canceling camera vibration and the capability of InnoVision to simultaneously monitor multiple points under the effect of camera vibration.

1. Introduction

Monitoring of structural health conditions is necessary for early detection of problems and prevention of catastrophic structural failure of aging infrastructure. Structural health monitoring is often based on measurement of structural vibration, acceleration in particular. However, structural displacements are more sensitive than acceleration to measure for low-frequency structures such as high-rise buildings and long-span bridges. Generally, displacement sensors can be categorized into contact type and non-contact type. Linear variable differential transformer (LVDT) [1], and global positioning system (GPS) [2–5] are the widely used contact type sensors. Laser displacement sensor and vision sensor are the main non-contact type sensors.

The LVDT and GPS are limited by the accessibility of the structure and require cumbersome installation. LVDT which measures the differential displacement between the device and the measurement target needs to be installed on a stationary platform that is free of vibration. However, it is hard to find a stationary platform for some of the structures. Other factors such as wind force, measuring distance and installation costs also make it impractical for monitoring large structures especially for long-term monitoring. The challenge of using GPS for displacement monitoring is its high cost and low accuracy that is insufficient for structural dynamic response analysis. The accuracy of GPS-based displacement sensor is around $\pm$ 1.5 cm in horizontal axis and $\pm$ 2 cm in vertical axis [2,3,5].

The non-contact type sensors have advantages over contact type sensors in measuring displacement without accessing the structure. Laser displacement sensor [6] is a high fidelity non-contact sensor and is widely recognized. However, similar to the LVDT, the laser displacement sensor needs a stationary platform for reference. Moreover, strong laser beams required for long distance monitoring may impose danger to human safety/health.

Vision-based displacement sensors provide a simple, cost-effective, and accurate alternative for remote displacement monitoring. Various vision sensors have been developed and applied for displacement monitoring including the widely used digital image correlation (DIC) [7–13], up-sampled cross correlation (UCC) [14], phase-based method [16,17], orientation code matching (OCM) [18,19] and others [20–25].

* Corresponding author.
E-mail address: ll2873@columbia.edu (L. Luo).

https://doi.org/10.1016/j.engstruct.2018.02.014
Received 29 July 2017; Received in revised form 17 December 2017; Accepted 1 February 2018

However, there are a few challenges associated with applying vision sensors in monitoring displacement response in the field that were rarely comprehensively studied in literatures, such as low lighting environment, insufficient camera resolution, and evident camera vibration. A new video image processing technique InnoVision was developed to tackle these difficulties.

First, it is difficult to accurately track the structural displacement of natural targets with low contrast features in low light conditions using intensity-based template matching algorithms (DIC and UCC methods). Inspired by the histogram of oriented gradients (HOG) algorithm [26] and the template matching method – PQ-HOG algorithm [27], a new gradient based template matching algorithm was developed in the InnoVision for monitoring targets with low contrast features.

The second challenge for vision sensors is insufficient resolution when monitoring multi-point structural displacements. A subpixel algorithm needs to be implemented in the vision sensor to increase resolution. There are several kinds of subpixel methods including gradient-based methods [28,29], Newton-Raphson method [28], upsampling cross correlation method, generic methods [30,31], phase correlation method [32], neural network methods [33], and the interpolation methods. Interpolation methods are the most popular subpixel methods in the vision sensors because of their simplicity, accuracy and computational efficiency [9,16,19,34,35]. Three widely used interpolation subpixel methods will be evaluated and incorporated into InnoVision, which are spline interpolation, cubic convolution [36], and paraboloid fitting method [19,35].

Camera vibration is the third challenge when using vision sensor in the field. Only few studies have been published on camera vibration cancellation. A practical technique for camera vibration cancellation using InnoVision was proposed by simultaneously tracking both the structure and a stationary reference point using InnoVision. The camera vibration can be canceled by subtracting the displacements of the reference point from the structural displacements. The InnoVision also has the capability of simultaneous displacement monitoring of multiple points on the structure under the effect of camera vibration. The performance of the InnoVision was evaluated through several laboratory and field tests.

This paper was arranged in the following way: Section 2 covers the configuration and algorithms of InnoVision (Section 2.1), and the practical technique for camera vibration cancellation by applying multi-point monitoring using InnoVision (Section 2.2). In Section 3, three laboratory tests are included. The first test investigates and incorporates three interpolation subpixel methods in InnoVision to evaluate the performance of the subpixel methods (Section 3.2). The second laboratory test evaluates the robustness of InnoVision in low contrast features (Section 3.3), The third test (Section 3.4) demonstrates the efficacy of the practical technique for camera vibration cancellation and further validates the robustness of InnoVision to low contrast features; Section 4 covers the field test conducted on a steel girder bridge in a challenging low lighting condition to demonstrate the robustness of InnoVision in comparison with two other methods; Section 5 covers the other field test conducted on Manhattan Bridge for validating the efficacy of the proposed technique for canceling camera vibration, and the advantage of InnoVision for monitoring multiple points simultaneously under the effect of camera vibration; Section 6 concludes this paper.

2. Multi-point displacement monitoring of low contrast features

2.1. InnoVision system

The InnoVision system contains a video camera and a computing unit. The video camera used in the system is a mono PointGrey USB 3.0 Camera of model FL3-U3-13Y3M-C that has 1280 by 1024 pixels of 4.8 μm in size. The video camera is installed in a remote location to capture the structural vibrations. The captured video is then transmitted to the computing unit installed with a video image processing software for extracting displacements from the video images. The computing unit has the Intel i7 CPU and 9 Gb RAM. The power consumption of the camera is less than 3 Watts. A power supply with the capacity of about 3 Wh is required for the camera for one hour of field test. In the current InnoVision system, the camera was connected to the computing unit by a USD 3.0 connection cable and powered by the battery of the computing unit. Considering the battery capacity of 65 Wh and the power consumption of at most 37 Watts for the computing unit, the battery of the computing unit was able to power both the computing unit and the video camera for about an hour and half. Alternatively, the camera can also be connected by a general-purpose input/output (GPIO) cable to an external power supply.

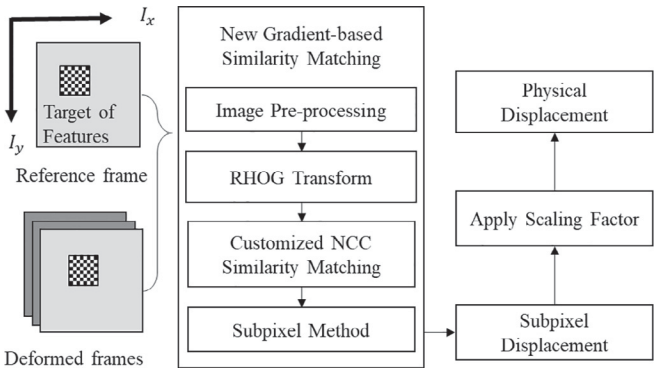

Fig. 1. The overview of the video image processing software.

The overview of the video image processing software was presented in Fig. 1. The InnoVision system uses a video camera to record structural vibrations remotely by tracking a target of natural features on the structure surface. At first, a target of features is selected on the first frame, then the pixel displacement of the target in subsequent frames in the video is obtained by the template matching algorithm, in this case, a robust gradient based similarity matching algorithm developed in InnoVision to cope with low contrast features in low lighting condition. To increase the displacement resolution, a subpixel method is implemented in the InnoVision. Finally, the subpixel displacement of the target is converted into physical displacements by multiplying a scaling factor (SF). By selecting and tracking multiple targets, the vision sensor is able to monitor displacements of multiple points on a structure using only one video camera. To cancel the effect of camera vibration, a new practical technique is developed using InnoVision by tracking the background target and the structure targets simultaneously.

2.1.1. Dense-RHOG code based similarity matching with subpixel resolution

It is inevitable the vision sensor needs to track low contrast features in changing lighting conditions when applied for field measurements. Low contrast features have intensities very similar to that of the background, making it challenging for intensities based template matching algorithms to accurately track the structural vibrations. A new similarity matching algorithm based on the gradient information was developed in InnoVision for tracking low contrast features in challenging low light condition. The proposed similarity matching algorithm is inspired by the sparse HOG feature descriptor [26], and the PQ-HOG template matching algorithm [27]. The new gradient based similarity matching algorithm is based on a new similarity estimation function and the dense rectangular HOG (dense-RHOG) feature descriptor. The dense-RHOG represents the steepest ascent orientation and magnitude estimated from the pixel neighborhoods. The dense-RHOG thus obtained contains information of the texture and shape of the target and is essentially robust in low illumination condition and invariant to changing illumination conditions. The detail of the new similarity matching algorithm with pixel level analysis is shown below.

At first, the densest one-pixel-step HOG code grid is computed. Each pixel is transformed into a four-bin feature descriptor, dense-RHOG

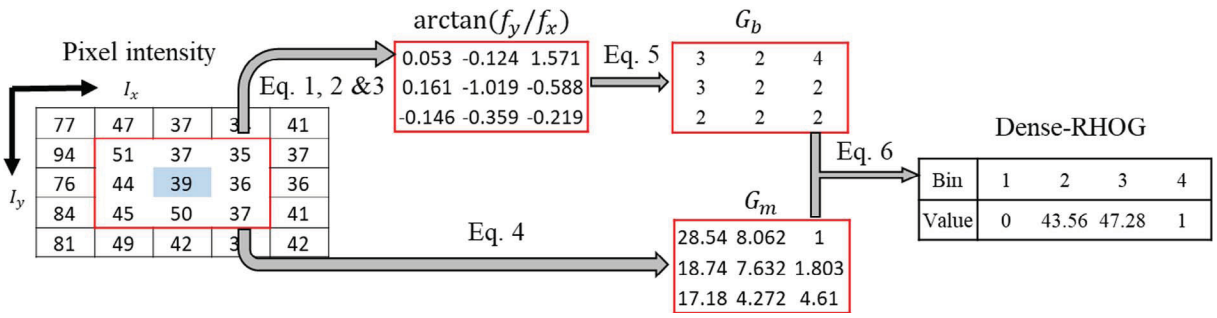

Fig. 2. Example for transforming a pixel into dense-RHOG code.

code, which is estimated from nine pixel neighborhoods. Fig. 2 presents an example for converting a pixel intensity in the blue shade into a four-bin dense-RHOG code estimated from the pixel neighborhoods within the red window.

Assume a discrete digital image is represented by $I(x,y)$, its horizontal and vertical derivatives (f_x,f_y) are computed respectively:

$$f_x(x,y) = \frac{\partial I}{\partial x} = 0.5*(I(x+1,y)-I(x-1,y)), \tag{1}$$

$$f_y(x,y) = \frac{\partial I}{\partial y} = 0.5*(I(x,y+1)-I(x,y-1)). \tag{2}$$

Then the gradient orientation angle θ and gradient magnitude G_m at each pixel are calculated:

$$\theta(x,y) = \arctan\left(\frac{f_y}{f_x}\right), \tag{3}$$

$$G_m(x,y) = \sqrt{f_x^2 + f_y^2}. \tag{4}$$

Since the numerical value of the gradient orientation angle θ is confined to $(-\frac{\pi}{2},\frac{\pi}{2})$, the orientation bin G_b is assigned by quantizing θ into four bins:

$$G_b(x,y) = 4*\frac{\theta(f_x,f_y)}{\pi} + 3; \quad G_b = 1,2,3,4. \tag{5}$$

As presented in Fig. 3, each orientation bin is given a numerical assignment.

The dense-RHOG is calculated from the gradient magnitude G_m of the nine neighborhood pixels and their corresponding orientation bin values G_b:

$$Dense\ RHOG(k) = \sum_{G_b(x,y)=k} G_m(x,y); \quad k = 1,2,3,4. \tag{6}$$

A customized similarity estimation function for dense-RHOG codes is employed to evaluate the similarity between any two images of the same size. The best match between the dense-RHOG code images of the template T and any object image I from the same scene is searched by maximizing the measured similarity γ in the form of mean of normalized cross correlation of each bin of the matching dense-RHOG codes, as shown below:

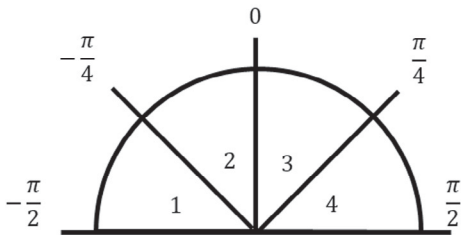

Fig. 3. Orientation bin.

$$\gamma = \frac{1}{4}\sum_{k=1}^{4} s(R_{I_{m,n}}(k),R_T(k)), \tag{7}$$

$$s(R_{I_{m,n}}(k),R_T(k)) = \frac{\sum_{I_{m,n}}[R_{I_{m,n}}(i,j,k)-\overline{R_{I_{m,n}}}][R_T(i,j,k)-\overline{R_T}]}{\left\{\sum_{I_{m,n}}[R_{I_{m,n}}(i,j,k)-\overline{R_{I_{m,n}}}]^2 \sum_T [R_{I_{m,n}}(i,j,k)-\overline{R_T}]^2\right\}^{\frac{1}{2}}}, \tag{8}$$

where

$$\overline{R_{I_{m,n}}} = \frac{1}{MN}\sum_{I_{m,n}}[R_{I_{m,n}}(i,j,k)]; \quad \overline{R_T} = \frac{1}{MN}\sum_{T}[R_T(i,j,k)] \tag{9}$$

where $R_{I_{m,n}}$ and R_T are the dense-RHOG code images of the object image and the template image respectively, and (m,n) shows the position of the object image in the scene. M,N are the sizes of the template image in the axes of both directions.

The template image needs to be compared with the entire scene image to find the best matching point, where the similarity γ reach its maximum. This process can be time consuming especially when the scene image is large. Therefore, the dense-RHOG matching process is carried out within a region-of-interest (ROI) window defined based on the current best matching position of the template image.

In the matching process, the similarity is calculated for each pixel position in the scene image and a similarity map containing the values of the similarity measurements is obtained as the rectangular grid shown in Fig. 4, thus the resolution of the best matching position as well as the displacement obtained is one pixel. To increase the displacement resolution, a subpixel algorithm needs to be implemented and applied to the similarity map in InnoVision. Three interpolation subpixel methods paraboloid interpolation, cubic convolution, and spline interpolation will be implemented and evaluated, and the method with the best performance will be chosen to be employed by InnoVision for subpixel resolution.

The paraboloid interpolation is proposed by Gleason et al. [35] and was applied by the authors [19]. In this method, the value distribution within a small window (3 × 3 pixels) of the similarity map is assumed

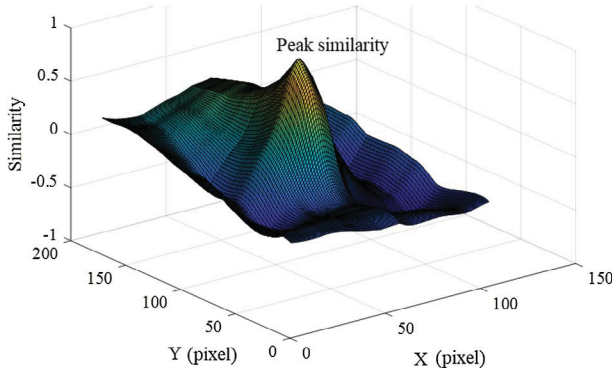

Fig. 4. Example of similarity map.

to be paraboloid. The coefficients of the paraboloid surface which fits all the points in the window is obtained through least squares method. The extreme similarity value and its coordinate are obtained by finding the peak value on the paraboloid surface. The extreme coordinate thus obtained is the subpixel coordinate ($M(x', y')$) used for obtaining the subpixel displacement D by comparing with the coordinate of the template in the reference image $P(x,y)$.

$$D = M(x',y') - P(x,y). \tag{10}$$

In the cubic convolution, the surface of the similarity map was interpolated using bi-cubic convolution by applying a convolution kernel proposed by Keys in 1981 [36] in axes of both directions. In the spline interpolation, the similarity map is fitted by a third order interpolation surface, as that presented in Fig. 4, based on the properties that the surface passes through all the points and the first and second derivatives will be continuous everywhere including the knots to ensure the spline will take a shape that minimizes the bending. The two-dimensional cubic convolution and spline interpolation can be applied using MATLAB built-in function *interp2*.

After an interpolation subpixel method is employed, InnoVision can produce subpixel displacements with sufficient resolution. Then, the obtained subpixel displacement in the image coordinate needs to be converted into displacement in the physical coordinate, as shown below:

$$X_{physical} = X_{image} * SF, \tag{11}$$

where SF is the scaling factor which can be calculated by comparing the physical dimension of a measured object with the pixel dimension in the image plane, as presented below and Fig. 5:

$$SF = \frac{D_{physical}}{D_{image}}. \tag{12}$$

2.2. Multi-point displacement monitoring for camera vibration cancellation

InnoVision can be used for robust multi-point displacement monitoring using only one camera after gaining sufficient resolution by employing one of the interpolation subpixel methods. However, ambient ground vibrations and the wind are inevitable in the field and will cause camera vibrations, therefore causing errors in displacement results. The effects of camera vibration become more significant when the target being monitored is located far away from the camera since the signal to noise ratio will decrease with the increase of target distance.

The effect of the camera vibration can be canceled by a practical technique that applies the multi-point measurement using InnoVision. To cancel the camera vibration, at first, the object targets which are referred as targets are selected on the monitored structure; and the background target (BG target) is selected on the stationary background structure. The displacements of the targets and the BG target are monitored simultaneously. The displacements detected on the BG target can be regarded as the measurement errors due to camera vibration.

The displacement of the BG target is equal to the error displacement of the scene due to camera vibration in the image coordinate:

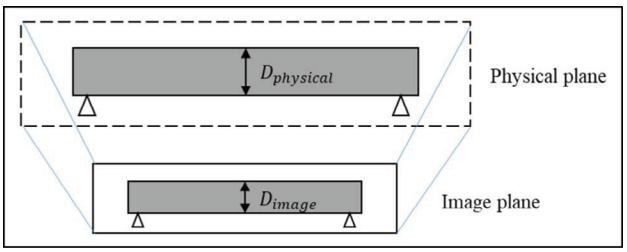

Fig. 5. Physical plane to image plane.

$$X_{image}^{BG} = X_{image}^{camera}. \tag{13}$$

By subtracting the displacements of the BG target (X_{image}^{BG}) from the displacements of the targets (X_{image}^{target}) before applying the SF, the camera vibration can be canceled in the new targets' displacements in image coordinate (X_{image}^{new}), given as below:

$$X_{image}^{new} = X_{image}^{target} - X_{image}^{BG}. \tag{14}$$

Because the displacement subtraction is performed before applying the scaling factor, the SF for the BG target is not required. The displacement in image coordinate is converted into the displacement in physical coordinate using the SF of the target after vibration cancellation.

$$X_{physical}^{new} = X_{image}^{new} * SF. \tag{15}$$

3. Laboratory tests

3.1. Experiment setup

A two-story shear structure was monitored in the experiments, as shown in Fig. 6. Two laser displacement sensors (LDS) were installed at 40 cm next to the structure as reference sensors. The video camera was located 4.175 m away from the structure. A shaking table was placed under the video camera to simulate the camera vibration. The SF of video images was 0.7724 mm/pixel. A hammer was used to induce impact force on the second floor of the structure. The movements of the structure due to the impact force were recorded by the reference laser displacement sensors and the vision sensor. The displacements obtained by the InnoVision were evaluated by comparing with the reference data. Three tests were conducted with different testing scenarios as listed in Table 1: In the first test, to evaluate the necessity and performance of the subpixel methods, the video camera remained stationary without any vibration and three interpolation subpixel methods were investigated and incorporated into InnoVision. The second test was conducted in low light condition also without any camera vibration to validate the robustness of InnoVision to low contrast features in low lighting condition. In the third test, the video camera was placed on a shaking table that vibrated under a white noise signal to demonstrate the efficacy of the practical technique for camera vibration cancellation. The third test was also conducted in low lighting condition to further validate the robustness of InnoVision.

3.2. Lab test #1 - subpixel algorithm for higher resolution

The displacement resolution of the template matching methods is limited to 1 pixel which corresponds to 0.7724 mm in the laboratory test due to the video camera quality and long object-to-camera distance. This resolution is not sufficient for structural dynamic analysis. To increase the displacement resolution, one can either move the video camera closer to the target, purchase an expensive high-resolution video camera, or employ a subpixel registration method. To be able to monitor multiple points, the object-to-camera needs to be very long to ensure the image plane is large enough. If only the video cameras with limited quality are available, the first choice is to employ one of the efficient interpolation subpixel methods. Three of the interpolation subpixel methods including the spline, cubic convolution, and paraboloid methods were tested and evaluated by comparing the results before/after applying the subpixel methods.

In the first laboratory test, the effectiveness of interpolation subpixel methods was evaluated. The results were analyzed then plotted in Fig. 7. The displacement results obtained without subpixel method did not match well with the reference data; on the contrary, the displacements obtained after implementing any of the three interpolation subpixel methods matched very well with the reference data. The root mean squared errors (RMSE) without/with subpixel method were

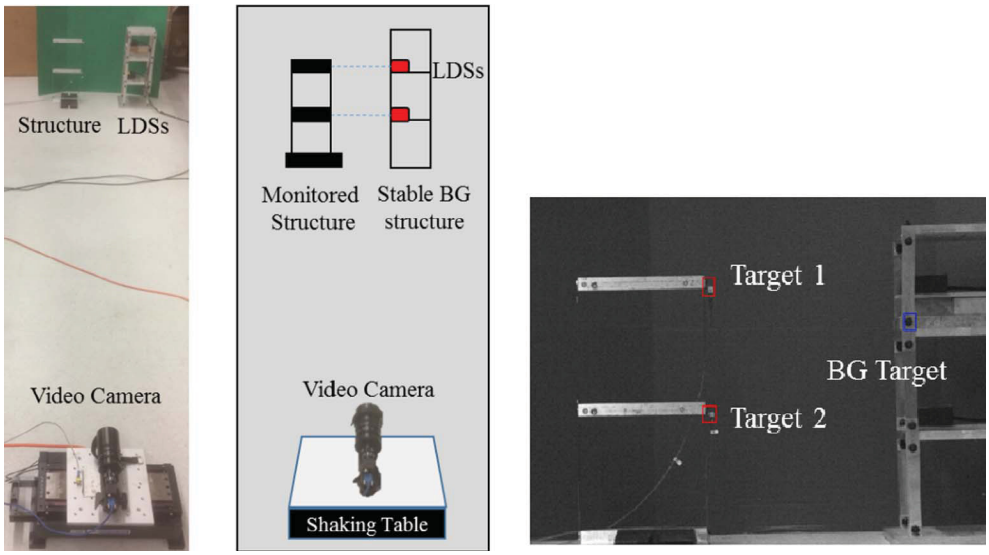

Fig. 6. Experiment setup and target selection.

Table 1
Laboratory testing scenarios.

	Camera vibration condition	Lighting condition	Testing element
First test	Stationary	Good lighting	Subpixel
Second test	Stationary	Low lighting	Low contrast feature
Third test	Shake with white noise	Low lighting	Camera vibration

Table 2
Testing error of subpixel methods.

	No subpixel	Spline	Cubic convolution	Paraboloid
RMSE (mm)	0.245	0.105	0.112	0.144
Reduction in RMSE (%)	–	57.39	54.73	41.67

calculated and listed in Table 2. After implementing the subpixel methods, the RMSE were significantly reduced by up to 57%. The spline interpolation method performed the best in the test, therefore was implemented in the proposed vision sensor, InnoVision. After applying the spline interpolation subpixel method, the displacement resolution of InnoVision was improved significantly from 1 pixel (0.7724 mm) to 1/20 pixel (0.03662 mm) and became sufficient for multi-point displacement monitoring.

When higher resolution is desired, the interpolation subpixel method implemented in InnoVision obviates the need for a more expensive camera. To test the limit of the subpixel methods, the RMSE of the results obtained with different subpixel resolution using the spline interpolation subpixel method were compared. Fig. 8 presents the reduction in RMSE in percentage versus the denominator of the subpixel resolution (D). The subpixel resolution is equal to $1/D$ pixel. As shown in the plot, the RMSE was significantly reduced by the subpixel method up to $D = 20\left(\frac{1}{20}\text{ pixel}\right)$. When the subpixel resolution is less than 1/20

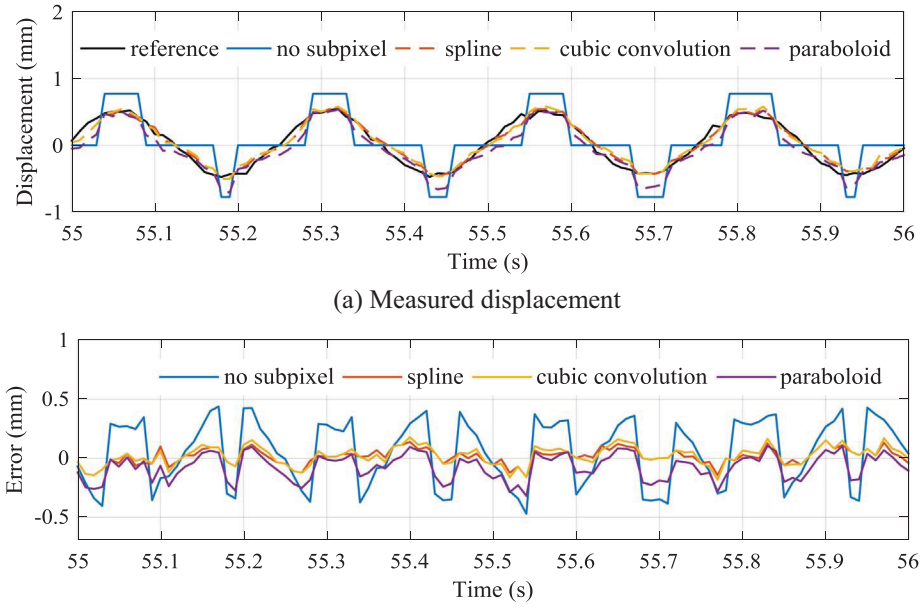

Fig. 7. The measurement results obtained by the vision sensor without /with subpixel methods.

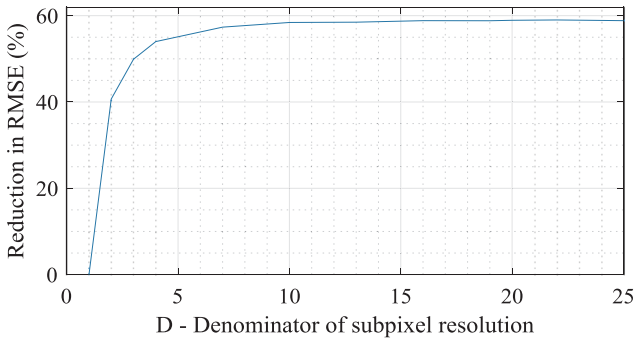

Fig. 8. Reduction of RMSE with improvement of subpixel resolution.

pixel (D > 20), the RMSE does not improve further. Therefore, the subpixel resolution limit of InnoVision after applying the subpixel method is around 1/20 pixel.

To demonstrate the effectiveness of the subpixel method, the field test on the Manhattan Bridge measurement is used as an example, which will be explained in more details in Section 4. In the field test, the scaling factor is 26.76 mm/pixel, therefore displacement resolution before the subpixel method is 26.76 mm. After applying the spline interpolation subpixel method, the displacement resolution is improved significantly to 1.34 mm. The camera pixel dimensions currently used in InnoVisino is 1280 by 1024 pixels. To obtain the same displacement resolution without applying the subpixel method, a camera with pixel dimensions of 25,600 by 20,480 pixels (524 mega pixels) would be required, which would be considerably more expensive than the camera currently used in InnoVision.

3.3. Lab test #2 - robust tracking of low contrast features

Since natural targets on the structure do not have high contrastness sometimes and the lighting conditions always change in the field environment, it is inevitable the structure is monitored by tracking low contrast features in low lighting condition. In the second test, the low lighting condition was simulated by setting the structure in an environment illuminated by dim light. A target with low contrast features was selected on the structure for tracking. The target can hardly be distinguished from the background since their pixel intensities are similar. The same target under good lighting condition had much higher contrast from the background as shown in Fig. 9. The displacements of the target in low lighting condition were obtained by both the proposed InnoVision method and the traditional DIC. The DIC used for comparison employed the normalized-cross-correlation (NCC) algorithm for template matching and the spline interpolation subpixel method for subpixel resolution. The NCC algorithm is implemented by applying the MATLAB function *normxcorr2*.

From Fig. 10, the displacements obtained by InnoVision matched very well with the reference data while the displacements obtained by DIC did not match well with the reference data. The displacement errors obtained by InnoVision was less than 1 mm, while the displacement errors obtained by DIC reached as high as 3 mm. The DIC is expected to fail catastrophically and diverge in this low lighting condition. But since the template matching was carried out within a confined ROI and the displacement was obtained based on the highest correlation point within the ROI, the erroneous results could be discrete and may approach to zero. The RMSE of the displacement measurement obtained by InnoVision (0.431 mm) was much lower than that obtained by DIC method (2.020 mm). This test validated the robustness of the InnoVision in monitoring displacement of low contrast features in low lighting condition.

However, DIC is effective when a high-contrast artificial target is used. The displacement measurements obtained by both InnoVision and DIC through tracking the high-contrast artificial target are presented in Fig. 10(c). As shown in the plots, both the displacements produced by the InnoVision and DIC match very well with the reference data. This validated the accuracy of the InnoVision and DIC techniques used in the paper when tracking a high-contrast target.

3.4. Lab test #3 - camera vibration cancellation through multi-point displacement monitoring

The camera vibration due to ambient ground vibration and the wind is inevitable in the field, and can results in errors in displacement measurement. The third laboratory test was conducted to validate the practical technique for camera vibration cancellation through multi-point displacement monitoring enabled using InnoVision. The camera vibration in the test was simulated by placing the camera on a shaking table which was excited by white noise signal. Two targets on the floors of the monitored structure and one BG target on the background stationary structure were selected as shown in Fig. 6 and their displacements were tracked simultaneously. The effects of camera vibration were canceled by subtracting the displacements of the BG target from the displacements of the targets.

The displacement measurement before and after camera vibration were plotted in Fig. 11. The measurement errors were greatly reduced after camera vibration cancellation. The displacements of the targets after camera vibration cancellation matched very well with the reference data. Table 3 showed the testing errors RMSE were reduced up to 61% after camera vibration cancellation. The laboratory results confirmed the efficacy of the practical camera vibration cancellation technique by applying multi-point vision displacement monitoring. Since the third test is also conducted in low lighting condition, the robustness of InnoVision is further validated.

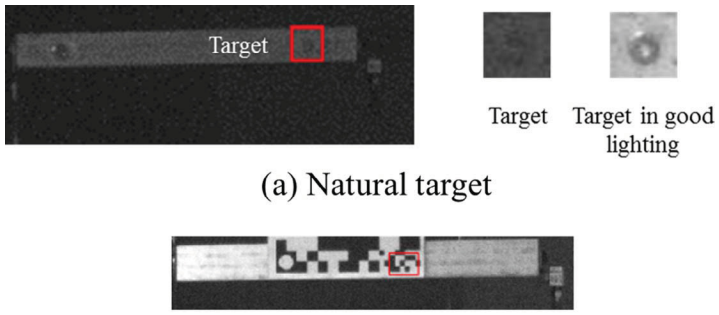

Fig. 9. Target selected for the second laboratory test.

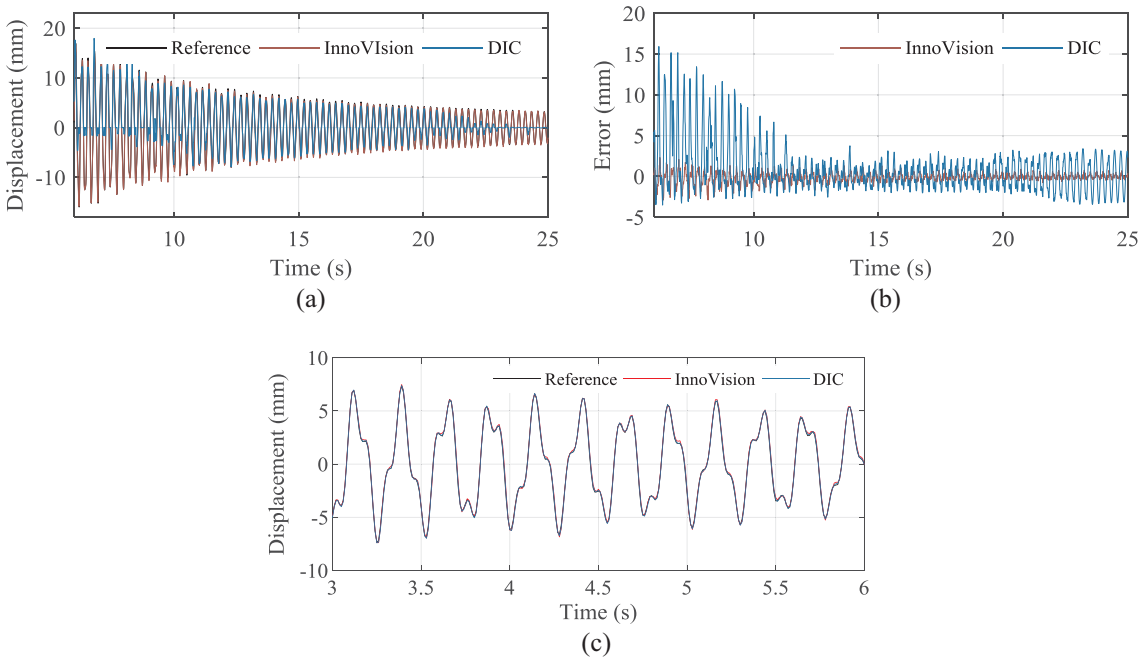

Fig. 10. Comparison of laboratory measurements, (a) Displacements using natural target. (b) Measurement errors using natural target. (c) Displacements using high-contrast artificial target.

4. Field test #1 – Validation of robustness of InnoVision

4.1. Setup for field test #1

The robustness of InnoVision in monitoring of low contrast features was validated in the second laboratory test. A field test was conducted on a 16.9 m long steel girder bridge to further confirm the robustness of InnoVision in comparison with two other methods including DIC method, and UCC method. The field test was conducted at night and illuminated by a dim flashlight.

A video camera was placed on a stationary point which was perpendicular to the bridge at 30 feet (9.14 m) away, as presented in Fig. 12. The vertical displacement responses of the mid-span of the bridge were captured when a train passed through at the speed of 25 mph. A rivet on the mid-span bridge surface was selected as the target, which had low contrastness from the background because the test was conducted at night. An LVDT was placed on a stationary ground under the mid-span of the bridge as the reference sensor. The

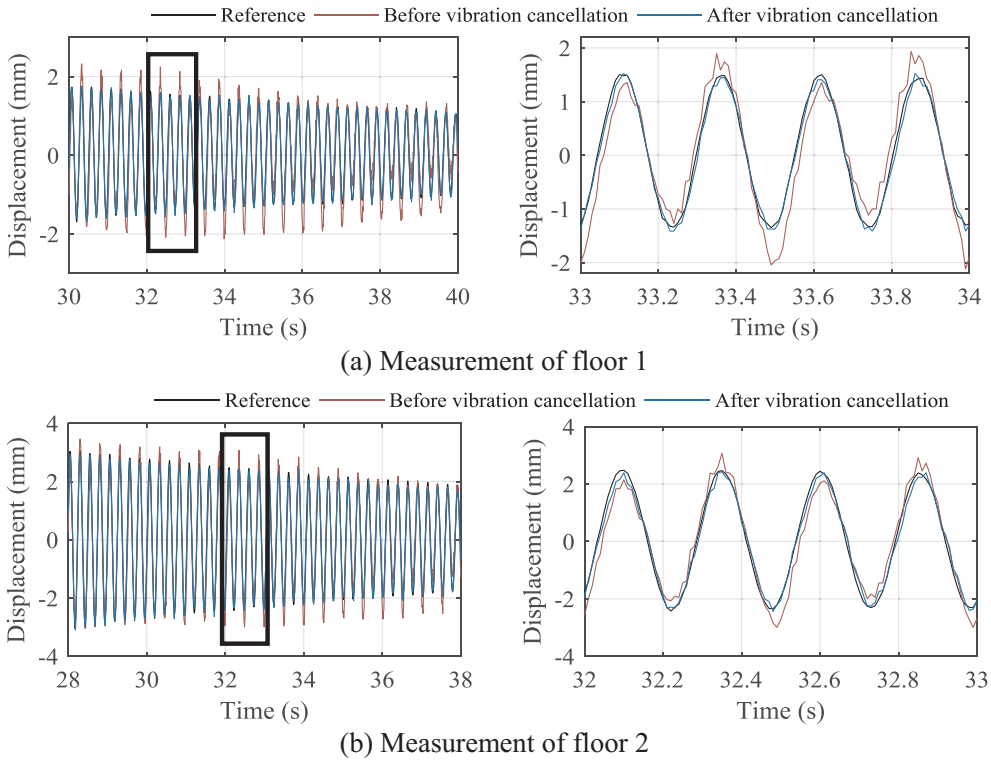

Fig. 11. Displacement measurement before/after camera vibration cancellation.

Table 3
Testing error before/after camera vibration cancellation.

		Before camera vibration cancellation	After camera vibration cancellation	Reduction (%)
Floor 1	RMSE (mm)	0.409	0.158	61.45
Floor 2	RMSE (mm)	0.446	0.206	53.81

reference sensor was utilized to evaluate the performance of InnoVision and the other two methods: DIC and UCC. The sampling frequency of the displacement result was 150 Hz.

4.2. Displacement measurement of low contrast features

The measured displacements obtained by the reference sensor and the vision sensors were depicted in Fig. 13. The measurement of InnoVision matched very well with the reference data while the DIC method could only roughly detect the general trend of the displacement. The displacement obtained by UCC method had errors too large in comparison with the reference data, therefore was not plotted. The RMSE estimations of the displacement measurements were listed in Table 4. The RMSE of the measurement results obtained by the InnoVision (0.28 mm) was much lower than that by DIC method (1.8 mm) and UCC method (29.4 mm). The results demonstrated the InnoVision which is based on the steepest ascent orientation and magnitude is more robust to low contrast features in low lighting conditions than DIC and UCC methods which are directly based on pixel intensities.

4.3. Analyses of robustness of InnoVision in comparison with DIC

InnoVision is not sensitive to changes in lighting conditions since the gradient magnitude G_m and gradient orientation angle θ are decided by the gradients of intensities as shown in Eqs. (3) and (4) in Section 2.1. On the contrary, the DIC method could not track the target accurately because it relies on image intensities for tracking. Changes in the lighting conditions may change the correlation value in the DIC therefore causing errors in the displacement measurement. For example, when the intensities offset by a factor of v, the intensities and the correlation value in the DIC change, but the gradient in either direction will not change. Therefore the orientation angle θ and the orientation magnitude G_m in InnoVision also remain the same as shown in the functions below:

$$f'_x = \frac{\partial I'}{\partial x} = \frac{\partial (I+v)}{\partial x} = \frac{\partial (I)}{\partial x} = f_x \tag{16}$$

$$f'_y = \frac{\partial I'}{\partial y} = \frac{\partial (I+v)}{\partial y} = \frac{\partial (I)}{\partial y} = f_y \tag{17}$$

$$\theta' = \arctan\left(\frac{f_y'}{f_x'}\right) = \theta \tag{18}$$

$$G'_m = \sqrt{(f'_x)^2 + (f'_y)^2} = G_m \tag{19}$$

When the target has high contrast, small changes in image intensities can be negligible since a rectangular target with the pattern of $m \times n$ pixels instead of one pixel is selected for tracking. However, when the target has low contrast, small changes in image intensities may affect the target pattern, therefore resulting in errors in measurement obtained by DIC. Another reason is that in low lighting condition, the image intensities fluctuates since the photo counts for the fixed exposure time may vary. The analyses agree with the results in the field test. As shown in Fig. 13, for the durations of 9.6–10 s in the field test, the lighting condition changed and caused high-variance and high-frequency errors in the displacement obtained by DIC. For other durations, the small-variance errors are probably due to the image intensities fluctuations.

5. Field test #2 – Validation of camera vibration cancellation

5.1. Setup for field test #2

After the robustness of InnoVision is demonstrated in the first field test, the capability of InnoVision to cancel camera vibration through multi-point displacement monitoring was further validated in the second field test. The test was conducted on Manhattan Bridge, a steel suspension bridge with 448 m long span. The dynamic response of the Manhattan Bridge was captured by a video camera that was located on Brooklyn Bridge at 447 m away from the mid-span of Manhattan Bridge as shown in Fig. 14. The camera vibration was introduced by the structural vibration of Brooklyn Bridge which was subjected to constant traffic.

At first, a natural target was selected on the mid-span of the Manhattan Bridge and a BG target was selected on a background building. The displacements of the target and the BG target were obtained by InnoVision. The camera vibration was canceled by subtracting the displacements of the BG target from the displacements of the target before applying the scaling factor. Recall the pixel displacement due to camera vibration is the same on the whole scene image in the image coordinate and the displacement subtraction is performed before applying the scaling factor, the scaling factor for the background building is not required. The scaling factor for the bridge was estimated at 26.76 mm/pixel. The sampling rate of the displacement measurement was 60 Hz. Since there is not a stationary platform for installing high fidelity displacement sensor such as LVDT or laser displacement sensor, the measurement result was validated through dynamic analysis in frequency domain. Acceleration data taken at the mid-span of the bridge were used as reference data to compare with the measurement results in the frequency domain.

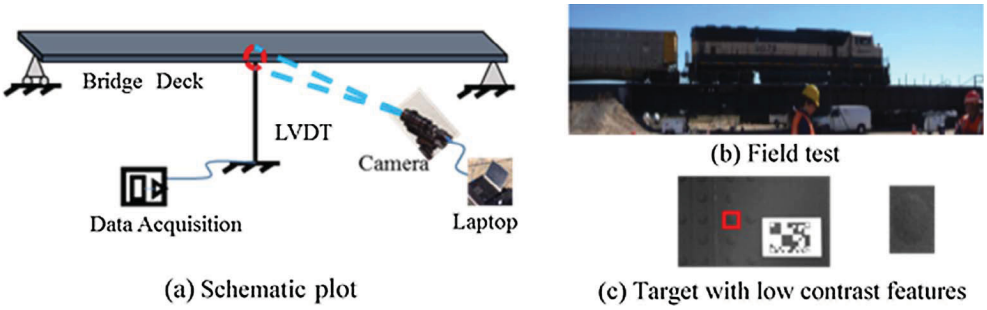

Fig. 12. Setup of a field test conducted on a steel girder bridge.

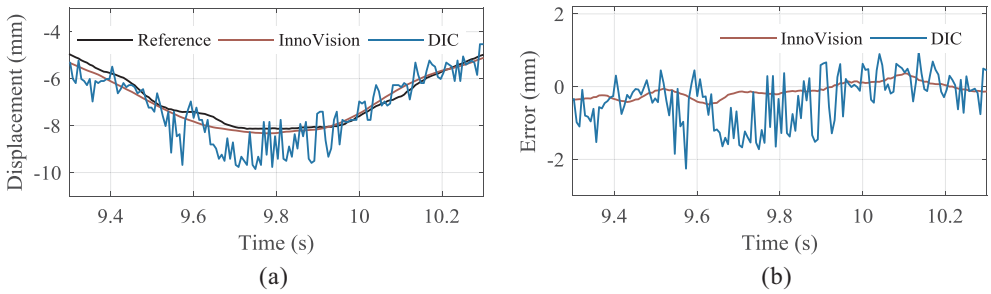

Fig. 13. Comparison of field measurement obtained by different methods; (a) measured displacement; and (b) measurement error.

Table 4
Testing error of different methods.

		InnoVision	DIC	UCC
RMSE (mm)	Natural Target	0.28	1.88	29.4

5.2. Measurement result of camera vibration cancellation

The displacement results before camera vibration cancellation and after vibration cancellation were plotted in Fig. 15. As shown in the figure, errors of the displacements were large due to camera vibration but were significantly reduced after vibration cancellation. The power spectrum density (PSD) of the displacements after camera vibration cancellation was plotted in Fig. 16, from which the first four identified resonant frequencies 0.22 Hz, 0.30 Hz, 0.40 Hz, and 0.50 Hz matched very well with the reference data identified from the accelerations as listed in Table 5. In Table 5, V1, V2 and V3 represent the first three vertical resonant frequencies and T1 represents the first torsional resonant frequency. A high pass filter with cut-off frequency at 0.06 Hz was applied to the displacements. It can also be seen in Table 5 that the resonant frequencies identified from the displacement before camera vibration cancellation did not match with the reference data. This is because the error due to camera vibration is so strong that it suppresses the bridge vibration. Some of the identified resonant frequencies are due to camera vibration and it is hard to decide which frequencies are the natural frequencies of the bridge, therefore the frequency readings can be inaccurate before camera vibration cancellation. The difference of the identified resonant frequencies before/after camera vibration cancellation validates the necessity and the efficacy of the camera vibration cancellation method.

5.3. Simultaneous multi-point displacement monitoring with camera vibration cancellation

The InnoVision was able to accurately monitor bridge displacement after camera vibration cancellation. Moreover, InnoVision has advantage over accelerometers or GPS in bridge monitoring that it can monitor multiple points on the bridge simultaneously without the need of moving the sensor or using multiple sensors.

Five targets were selected along the bridge span and one BG target was selected on the background building as shown in Fig. 17. The displacements of the targets after camera vibration cancellation were extracted and plotted in Fig. 18. The structural deflection responses of multiple points on the bridge under a moving train load were clearly reflected by displacement measurement. Target 1 and target 5 were the first and last to reach the peak displacement respectively, it can be predicted the train passed through the bridge from the left to the right side of the image plane which matched the video footage. The field test validates the InnoVision's advantage to simultaneously monitor structural responses of multiple points on a bridge under the effect of camera vibration.

6. Conclusions

This study contributes to a new video image processing technique InnoVision developed with capabilities of robust tracking of low contrast features, high subpixel resolution, multi-point displacement monitoring, and practical camera vibration cancellation in response to a number of difficulties associated with vision based structural displacement monitoring in the field. Some of the important conclusions of this study are summarized as follows:

1. To enable robust tracking of low contrast features by InnoVision, a

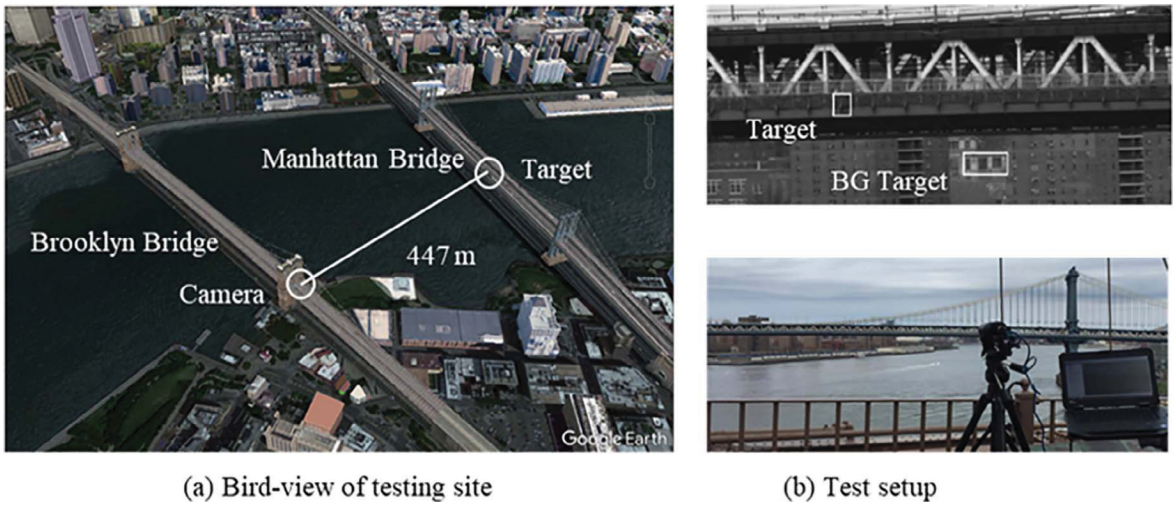

Fig. 14. Field test bird-view and setup.

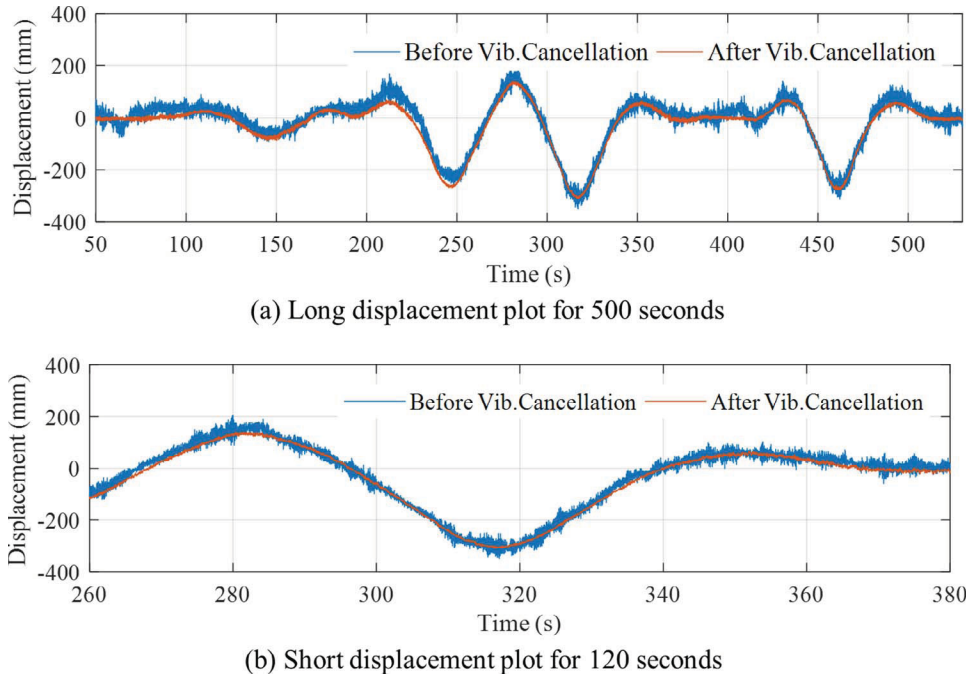

Fig. 15. Displacement measured before/after camera vibration cancellation.

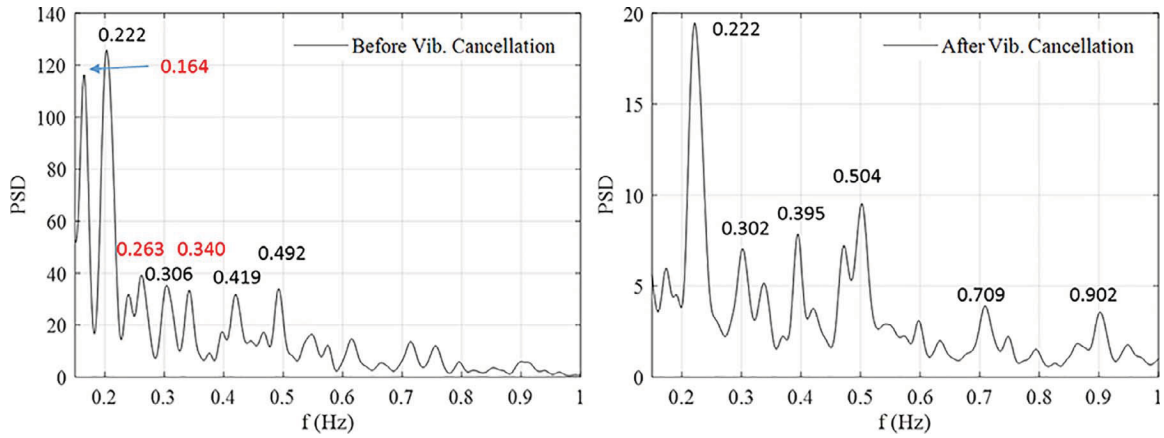

Fig. 16. Frequency plot before/after camera vibration cancellation.

Table 5
Resonant frequencies before/after camera vibration cancellation.

Data type	Data Year	1st resonant V1 (Hz)	2nd resonant V2 (Hz)	3rd resonant T1 (Hz)	4th resonant V3 (Hz)
Reference acceleration data [37]	2009	0.23	0.30	0.37	0.50
Displacement before vibration cancellation	2017	0.16	0.22	0.26	0.31
Displacement after vibration cancellation	2017	0.22	0.30	0.40	0.50
Recent acceleration data	2016–2017	0.23	–	0.40	0.51
GPS data [38]	Before 2009	0.23	–	0.30	0.49

Fig. 17. Targets for monitoring multiple points.

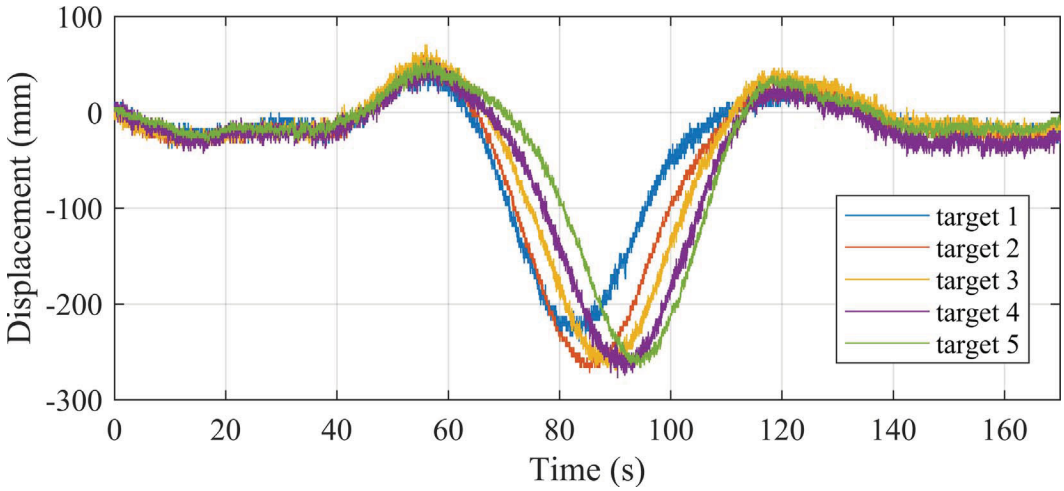

Fig. 18. Multi-point displacement measurement.

new gradient based template matching algorithm was proposed based on the dense-RHOG feature descriptor and a new similarity matching method. The dense-RHOG represents the steepest ascent orientation and magnitude that contains information of the texture and shape of the target and is therefore essentially robust to low contrast features in low illumination condition and invariant to changing illumination conditions. The low contrast features were successfully tracked with high accuracy by InnoVision in both the laboratory tests and the night field tests on a railway bridge.

2. The displacement measurement resolution of the InnoVision vision system was increased significantly by implementing the spline interpolation subpixel method, which showed the highest accuracy among the three interpolation methods investigated.
3. A practical solution was developed using the InnoVision to cancel the camera vibration that is inevitable in the field measurement by simultaneously tracking the displacements of both the structure and a stationary reference point, then subtracting the displacement of the reference point from the structural displacement measurement. The laboratory tests and the field tests on the Manhattan Bridge validated the efficacy of the vibration cancellation technique.
4. InnoVision's advantage of simultaneous monitoring of displacements at multiple points was also demonstrated in the Manhattan Bridge field tests.

Other environmental conditions such as heat haze and high humidity can also affect the measurement of the vision sensor. In the future research, the authors will focus on studying and providing solutions for these challenges. Heat haze induced image distortion due to high temperature in hot weather is a known factor to affect the measurement accuracy [39]. The authors are currently working on providing a framework for heat haze filtering. High intensity of water particles in the air in high humidity (foggy) weather can result in biased and low-contrast images, therefore affecting the performance of computer vision techniques. The effect of humidity on the measurement accuracy of the vision sensor needs to be studied. And whether the techniques in the researches conducted on de-hazing can be applied to InnoVision to help reduce the effect of the humidity needs further investigation.

Owing to its robustness in monitoring low contrast features, high subpixel resolution, efficacy of canceling the effect of camera vibration, and capability of multi-point displacement monitoring, the developed InnoVision has great advantages in multi-point displacement monitoring of bridges in the challenging field conditions.

Acknowledgement

This work is partially supported by NCHRP Highway IDEA Project (No. 20-30/IDEA 189), Bentley Systems, Inc, and NSF IUCRC grant (No. 1738802). The authors would like to thank Dr. L. Li of the Carleton Laboratory of Columbia University for his generous advice and support regarding the laboratory tests and Ryan Leung and Lijun Xie, graduate students at Columbia University for their enthusiastic participation in the field tests. The authors would also like to thank the anonymous reviewers for their constructive comments which helped improving the quality of the manuscript.

References

[1] Novacek G. Accurate linear measurement using LVDTs. Circuit Cellar Ink 1999;106:20–7.
[2] Celebi M. GPS in dynamic monitoring of long-period structures. Soil Dyn Earthquake Eng 2000;20:477–83.
[3] Nakamura S-i. GPS measurement of wind-induced suspension bridge girder displacements. J Struct Eng 2000;126:1413–9.
[4] Figurski M, Gałuszkiewicz M, Wrona M. A bridge deflection monitoring with GPS. Artif Satel 2007;42:229–38.
[5] Mayer L, Yanev B, Olson L, Smyth A. Monitoring of the Manhattan Bridge and interferometric radar systems. Review Literature And Arts Of The Americas; 2010. p. 3378–84.
[6] Nassif HH, Gindy M, Davis J. Comparison of laser Doppler vibrometer with contact sensors for monitoring bridge deflection and vibration. NDT E Int 2005;38:213–8.
[7] Yoneyama S, Kitagawa A, Iwata S, Tani K, Kikuta H. Bridge deflection measurement using digital image correlation. Exp Tech 2007;31:34–40.
[8] Kohut P, Holak K, Uhl T, Ortyl Ł, Owerko T, Kuras P, et al. Monitoring of a civil structure's state based on noncontact measurements. Struct Health Monitor 2013;12:411–29.
[9] Kim S-W, Jeon B-G, Kim N-S, Park J-C. Vision-based monitoring system for evaluating cable tensile forces on a cable-stayed bridge. Struct Health Monitor 2013;12:440–56.
[10] Cigada A, Mazzoleni P, Zappa E. Vibration monitoring of multiple bridge points by means of a unique vision-based measuring system. Exp Mech 2014;54:255–71.
[11] Hoag A, Hoult N, Take W, Le H. Monitoring of rail bridge displacements using digital image correlation. Struct Health Monitor 2015;2015.
[12] Dworakowski Z, Kohut P, Gallina A, Holak K, Uhl T. Vision-based algorithms for damage detection and localization in structural health monitoring. Struct Control Health Monitor 2016;23:35–50.
[13] Mazzoleni P, Zappa E. Vision-based estimation of vertical dynamic loading induced by jumping and bobbing crowds on civil structures. Mech Syst Sig Process 2012;33:1–12.
[14] Guizar-Sicairos M, Thurman ST, Fienup JR. Efficient subpixel image registration algorithms. Opt Lett 2008;33:156–8.
[16] Chen JG, Davis A, Wadhwa N, Durand F, Freeman WT, Büyüköztürk O. Video camera-based vibration measurement for civil infrastructure applications. J Infrastruct Syst 2016;23:B4016013.
[17] Cha Y-J, Chen J, Büyüköztürk O. Output-only computer vision based damage detection using phase-based optical flow and unscented Kalman filters. Eng Struct 2017;132:300–13.
[18] Fukuda Y, Feng MQ, Narita Y, Kaneko Si, Tanaka T. Vision-based displacement sensor for monitoring dynamic response using robust object search algorithm. IEEE

Sens J 2013;13:4725–32.
[19] Luo L, Feng M, Fukuda Y, Zhang C. Micro displacement and strain detection for crack prediction on concrete surface using optical nondestructive evaluation methods. Int J Prognost Health Manag 2015;6:1–12.
[20] Lee JJ, Shinozuka M. A vision-based system for remote sensing of bridge displacement. NDT E Int 2006;39:425–31.
[21] Santos CA, Costa CO, Batista J. A vision-based system for measuring the displacements of large structures: simultaneous adaptive calibration and full motion estimation. Mech Syst Sig Process 2016;72:678–94.
[22] Bartilson DT, Wieghaus KT, Hurlebaus S. Target-less computer vision for traffic signal structure vibration studies. Mech Syst Sig Process 2015;60:571–82.
[23] Oh BK, Hwang JW, Kim Y, Cho T, Park HS. Vision-based system identification technique for building structures using a motion capture system. J Sound Vib 2015;356:72–85.
[24] Ribeiro D, Calçada R, Ferreira J, Martins T. Non-contact measurement of the dynamic displacement of railway bridges using an advanced video-based system. Eng Struct 2014;75:164–80.
[25] Wahbeh AM, Caffrey JP, Masri SF. A vision-based approach for the direct measurement of displacements in vibrating systems. Smart Mater Struct 2003;12:785.
[26] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: IEEE computer society conference on computer vision and pattern recognition, 2005, CVPR 2005. IEEE; 2005. p. 886–93.
[27] Sibiryakov A. Fast and high-performance template matching method. In: 2011 IEEE conference on computer vision and pattern recognition (CVPR). IEEE; 2011. p. 1417–24.
[28] Bing P, Hui-Min X, Bo-Qin X, Fu-Long D. Performance of sub-pixel registration algorithms in digital image correlation. Meas Sci Technol 2006;17:1615.
[29] Pan B, Cheng P, Xu B. In-plane displacements measurement by gradient-based digital image correlation. Third International Conference on Experimental Mechanics and Third Conference of the Asian Committee on Experimental Mechanics. International Society for Optics and Photonics; 2005. p. 544–52.
[30] Pitch A, Mahajan A, Chu T. Measurement of whole-field surface displacements and strain using a genetic algorithm based intelligent image correlation method. J Dyn Syst Meas Contr 2003;126:479.
[31] Jin H, Bruck HA. Theoretical development for pointwise digital image correlation. Opt Eng 2005;44:067003–67014.
[32] Foroosh H, Zerubia JB, Berthod M. Extension of phase correlation to subpixel registration. IEEE Trans Image Process 2002;11:188–200.
[33] Pitter MC, See CW, Somekh MG. Subpixel microscopic deformation analysis using correlation and artificial neural networks. Opt Express 2001;8:322–7.
[34] Debella-Gilo M, Kääb A. Sub-pixel precision image matching for measuring surface displacements on mass movements using normalized cross-correlation. Remote Sens Environ 2011;115:130–42.
[35] Gleason SS, Hunt MA, Jatko WB. Subpixel measurement of image features based on paraboloid surface fit. In: Proceedings of the machine vision systems integration in industry. Boston (MA): SPIE; 1990.
[36] Keys R. Cubic convolution interpolation for digital image processing. IEEE Trans Acoust Speech Signal Process 1981;29:1153–60.
[37] Jang J, Smyth A. Bayesian model updating of a full-scale finite element model with sensitivity-based clustering. Struct Contr Health Monit 2017;24.
[38] Mayer L, Yanev BS, Olson LD, Smyth AW. Monitoring of manhattan bridge for vertical and torsional performance with GPS and interferometric radar systems. In: Transportation research board 89th annual meeting; 2010.
[39] Luo L, Feng MQ. Vision based displacement sensor with heat haze filtering capability. Struct Health Monitor 2017;2017.