Citation
Kuddus, M.A. and Li, J. and Hao, H. and Li, C. and Bi, K. 2019. Target-free vision-based technique for vibration measurements of structures subjected to out-of-plane movements. Engineering Structures. 190: pp. 210-222. http://doi.org/10.1016/j.engstruct.2019.04.019

Target-Free Vision-based Technique for Vibration Measurements of Structures subjected to Out-of-Plane Movements

Mir Abdul Kuddus[1], Jun Li[1, 2], Hong Hao[1, 2,*], Chao Li[1], Kaiming Bi[1]

[1]Centre for Infrastructure Monitoring and Protection, School of Civil and Mechanical Engineering, Curtin University, WA 6102, Australia

[2]School of Civil Engineering, Guangzhou University, Guangzhou 510006, China

Email: mirabdul.kuddus@postgrad.curtin.edu.au; junli@curtin.edu.au; hong.hao@curtin.edu.au; chao.li11@postgrad.curtin.edu.au; kaiming.bi@curtin.edu.au

Abstract: Vibration measurements have been widely used for structural health monitoring (SHM). Usually, wired sensors are required to attach on the testing structure, which may be arduous, costly and sometimes impossible to install those sensors on the remote and inaccessible part of the structure to be monitored. To overcome the limitations of contact sensors based vibration measurement methods, computer vision and digital image processing based methods have been proposed recently to measure the dynamic displacement of structures. Real-life structure subjected to bi-directional dynamic forces is susceptible to significant out-of-plane movement. Measuring the vibrations of structures under the out-of-plane movements using target-free vision-based methods have not been well studied. This paper proposes a target-free vision-based approach to obtain the vibration displacement and acceleration of structures subjected to out-of-plane movements from minor level excitations. The proposed approach consists of the selection of a region of interest (ROI), key-feature detection and feature extraction, tracking and matching of the features along the entire video, while there is no artificial target attached on the structure. The accuracy of the proposed approach is verified by conducting a number of experimental tests on a reinforced concrete structural column subjected to bi-directional ground motions with peak ground accelerations (PGA) ranging from 0.01g to 1.0g. The results obtained by the proposed approach are compared with those measured by using the conventional accelerometer and laser displacement sensor (LDS). It is found that the proposed approach accurately measures the displacement and acceleration time histories of the tested structure. Modal identification is conducted using the measured vibration responses, and natural frequencies can be identified accurately. The results demonstrate that the proposed approach is reliable and accurate to measure the dynamic responses and perform the system modal identification for structural health monitoring.

Keywords: Structural health monitoring; Targetless; Vision-based; Vibration response measurement; Image processing; Modal identification

*Corresponding author: Email: hong.hao@curtin.edu.au; Tel: +61 8 9266 4762; Fax: +61 8 9266 2381; Centre for Infrastructure Monitoring and Protection, School of Civil and Mechanical Engineering, Curtin University, WA 6102, Australia.

1 Introduction

Timely structural condition inspection, monitoring and safety evaluation is very important for cost-effective repair and maintenance planning of civil engineering structures. In industry practices, structural condition monitoring of civil engineering infrastructure, such as buildings, bridges and roads, mainly depends on the visual inspection carried out by the professional inspectors to obtain structural health information and rate structural conditions. Physical inspection and non-destructive testing provide quantitative data of structural conditions, but may not be feasible for the inaccessible part of large-scale structures or at least are very labour intensive, such as long-span bridges and high rise buildings. Though the existing physical inspection and condition assessment are mostly carried out by well trained and qualified inspectors, several remarks have been recognized, such as: (i) Manual inspection depends on the personal judgment of inspector and may be lack of consistency [1, 2]. The accuracy and trustworthiness of the inspection results rely mainly on the expert knowledge and professional experience of the inspectors. Consequently, the structural health condition evaluation outcomes sometimes encompass errors and extensive deviations lead to misguide for appropriate decision making and cost-effective maintenance design; (ii) The visual inspection is time consuming and labour intensive [3, 4], specifically for long span bridges and high rise buildings, which significantly increases the expenses; (iii) Manual inspection may not be feasible for quantitative condition monitoring and maintenance design [5]. In addition, a number of life safety hazards are posed to the inspectors [6]. In a word, lack of proper training, vision problems, location and accessibility may cause error prone results from visual inspections [7].

According to a study by Kim [8], the aim of Structural Health Monitoring (SHM) can be stated as a cost-effective evaluation of structural performance and level of service, detection and location of damage, structural lifetime prognosis as well as maintenance management. Structural system identification is one of the key components of SHM that can be used to update the analytical model and detect the damage. Conventional sensors like accelerometer, strain gauge and linear variable differential transducer (LVDT) [9-14] are widely used in SHM systems to measure vibration acceleration, strain and displacement for structural performance and serviceability assessment. However, such conventional sensory system could be highly expensive to operate and maintain with a good service condition, mainly due to labour-oriented, time-consuming and expensive to install a large number of cables involved and the possibility of sensor failures or malfunctions. In addition, the installation of the contact type sensors is sometimes difficult or impossible for inaccessible parts of the structure. For a small structure, the additional mass of the contact sensors may change structural

properties and responses, resulting in error-prone measurements. Therefore, the potential use of non-contact based vibration measurement techniques has received a significant interest in the field of SHM and damage detection. Non-contact measurement methods, such as Global Positioning System (GPS), Laser Doppler Vibrometer, Radar Interferometry, Speckle Photography and Hologram Interferometry, provide significant advantages with ease of operation and flexibility to extract structural displacements of multiple points from a single record [15, 16]. However, most of the non-contact type measurement equipment require high setup cost and complex data interpretation to reflect structural conditions. This may constraint the wide applications of these techniques for SHM and damage assessment of practical engineering structures [17].

In the recent years, as an innovative approach, the vision-based techniques enable to extract object vibrations from video images. A vision-based measurement system consists of a camera, zoom lens and computer installed with post-processing software. Zhang et al. [17] used fast motion magnification algorithm and artificial target based image processing approach for vibration measurement of structure in the laboratory condition. The same procedure was applied in the field for vibration analysis of a railway bridge. Further challenges that need to be overcome in future are complex geometries of structure, closely spaced modes and outdoor measurement perspectives such as wind, camera movement, lighting conditions, the variation of temperature and index of refraction. Lee et al. [18] presented a computer vision-based approach for structural dynamic displacement measurement with artificial pattern or targets fixed on the surface of the structure. The robust light-induced image degradation method was introduced in the field testing condition to evaluate the performance of the technique under strong sunlight. By using the precise target, dynamic displacement of a railway bridge was measured by Ribeiro et al. [19]. However, the method was limited to the application with the use of artificial illuminated target and camera distance 25 m from the target location. The vertical displacement of the bridge was measured by the camera installed at 5m, 10m, 15m and 25m from the bridge, and the obtained displacement results were compared with those measured by LVDT. An experimental investigation was carried out by Yang et al. [20], where speckle pattern was used for image-based strain fields measurement of a reinforced concrete wall. Luo et al. [21] adopted high contrast artificial target vision-based approach for multi-point bridge displacement monitoring. A realistic method was developed by tracking both the structure and stationary reference point to eliminate the error due to camera vibration. Ye et al. [22] developed a vision-based system for displacement measurement of long-span bridges based on digital image processing technology. The proposed algorithm was verified with experimental tests and exhibited an excellent capability in measuring the dynamic displacements under the uni-axial seismic motions

and in-situ bridge tests. Guo and Zhu [23] proposed a computer vision-based Lucas-Kanade template tracking approach to measure the large-scale structural dynamic displacement. The proposed method used pre-designed black and white circular target fixed on the surface of the structure, and the video was captured by a stationary camera 3m away from the target structure. Dong et al. [24] proposed a machine vision technology for structural system identification. Multi-point structural dynamic displacements were measured by using LED lamp and black spots target. Ostrowska et al. [25] developed a vision based digital image correlation technique to measure the deflection of a beam in the laboratory. Image registration algorithm was used, and the measured deflection values obtained in this study were compared with those from the laser tracker system with a mean percentage error of 1.11%. The applied pickle marks and pattern are the important factors for accurate measurement of deflection by using digital image correlation. Ye et al. [26] measured structural dynamic displacement using different multi-object tracking algorithms. Multi-point structural dynamic displacements can be obtained effectively with three different image processing and object tracking algorithms. Shao et al. [27] demonstrated a vision-based motion tracking algorithm to measure the human-induced vibration of civil engineering infrastructure. The developed vision-based system used artificial targets to extract good results in the field of human-induced vibrations when compared to classical sensors.

To remove the limitation of using artificial targets in the above mentioned techniques, studies are being conducted to develop target-free vision-based approaches for vibration displacement measurements. Ji et al. [28] proposed a target-free image based optical flow estimation method for small cable vibration measurement. The proposed method was verified both in the laboratory and in the field on a pedestrian bridge. No camera calibration was required, however, the method was restricted for the long-span bridge cable vibration measurement. Dorn et al. [29] proposed phase-based video motion magnification for blind identification of full-field vibration modes and vision based operational modal analysis. Multi-scale image processing technique was applied on the video frames of vibrating structure to blindly extract natural frequencies, damping ratios and full-field mode shapes from the recorded video. Davis et al. [30] developed a target free vision-based motion magnification algorithm for modal identification of a simple beam structure and visualizing the operational deflection shapes. Elanwar et al. [31] used a consumer grade camera as a vision-based tool for structural system identification. Consumer grade GoPro Hero3 camera and LG G3 smartphone were used in this study, and the captured images were processed by the feature extraction algorithm developed by Harris and Stephens [32]. Daniel et al. [33] developed the target-less computer vision method for the vibration measurement of traffic signal structure. Ambient wind excitation was used, and measured dynamic displacement and identified structural system parameters

obtained from the vision-based technique were compared with those obtained with the traditional sensory system. Feng and Feng [34] performed laboratory and field investigations to study the effectiveness of using vision based techniques for damage detection. Feng et al. [35] developed a non-target vision sensor to measure the bridge dynamic response. The proposed approach was validated in the laboratory shaking table test and in the field test on a railroad bridge. Moreover, the method is capable of measuring multi-point displacements from a single recorded video. Chen et al. [36] developed a computer vision system to measure structural displacement and identify structural properties such as stiffness and damping coefficient of a cantilever beam in the laboratory condition by using phase-based optical flow and unscented Kalman filter. In addition, Chen et al. [37] proposed high speed video based motion magnification technique to obtain modal identification parameter of a simple cantilever steel column structure. Furthermore, video camera based motion magnification technique was demonstrated to measure the vibration of an antenna tower from a distance over 175 m [30]. The camera motion was eliminated by capturing the reference object in the same image and video down sampling enhanced Signal-to-Noise Ratio (SNR). Hu et al. [38] used target free high speed video with consensus based matching and tracking algorithm to measure the vibration of the bus rapid transit viaduct. The performance of the method was verified in the laboratory with a normalized mean squared error (NRMSE) of 2.092 %. Choi et al. [39] proposed a target-less vision based sensor to measure the dynamic response of building structure by using convex hull optimization algorithm. The structure was excited by white noise vibration and structural response obtained from the developed vision-based method was compared with that measured from the conventional sensory system. Khuc and Catbas [40] proposed a camera based structural health monitoring system to measure structural displacement and vibration without using any physical target. Khuc and Catbas [41] also developed a vision based method for dynamic measurement of a railway bridge with an acceptable error compared to using the conventional LVDT.

Despite that the existing vision-based techniques have received successes for the applications of measuring structural vibration displacements of the experimental and in-field structures, several limitations may still exist. For example, many existing methods require to attach physical targets on the surface of the structure, which could make the system setup time consuming and even practically impossible for the inaccessible locations of the real large scale structures, i.e. the long-span bridges and high rise buildings as mentioned above. Physical targets and template matching algorithm have been used to extract vibration displacements, which is also computationally expensive. The existing vision-based methods depend on the high frame rate and high-resolution camera or sometimes demand additional equipment and lenses, which may restrict the system as a low-cost real-life

implementation. Besides these, issues such as lens distortion, temporal aliasing, the field of view (FOV) and sampling frequency variation have not been fully discussed and investigated. Furthermore, the current target-free vision-based vibration measurement methods are usually used to extract high level vibrations, which produce large displacements. However in the real-world practical applications, low level vibration measurement is more challenging for structural health monitoring and damage detection. Most of the existing vision-based studies extract vibration information when the structure is subjected to significant and uni-directional movements. No research has yet been carried out to investigate the applicability of vision based vibration displacement technique when the structure is subjected to minor ambient vibrations and bi-directional out-of-plane movements, which is very much realistic for the real word structure vibrations.

This paper proposes a target-free vision-based approach for dynamic vibration measurement when the structure is subjected to bi-directional excitations and out-of-plane movements. The proposed approach is capable of identifying the time domain minor displacement histories under low level excitations. The proposed approach uses a consumer-grade video camera to derive the vibration displacement, removing the hassle of installing an artificial target on the structure. The camera calibration is carried out to remove the radial distortion from the raw video images, which is enormously crucial for accurate displacement measurement. After the elimination of distortion, the dynamic displacement of the structure is obtained by analysing the calibrated video frame by frame. The region of interest (ROI) of the first frame of the video is selected by image segmentation algorithm. Key features are detected and extracted by using the binary robust invariant scalable keypoints (BRISK) feature detection and extraction algorithm which is invariant to scale, rotation and orientation. This is one of the advantages of this study to detect and extract key-features while the images are captured with a certain camera angle. The fast BRISK algorithm also reduced the computational time of large volume image processing. The extracted key features points are used to calculate the displacement of each frame of the entire video by using Lucas-Kanade tracking (KLT) algorithm which ensure the maximum number of matching of key-points between consecutive image. The false match outliers are discarded by using the maximum likelihood estimation sample consensus (MLESAC) algorithm which produced maximum error-free dynamic displacement. The accuracy of the proposed approach is validated experimentally on a reinforced concrete column model subjected to bi-directional ground motion excitations. The results obtained from the proposed approach are compared with those measured from the high accuracy laser displacement sensors (LDSs) to validate the effectiveness and accuracy of the proposed approach.

2 Target-free vision-based technique for vibration measurement

The proposed target-free vision-based technique for vibration measurements of structures consists of the following significant components as shown in Figure 1, namely, (i) Acquisition of raw video clip; (ii) Removal of geometric distortion from images; (iii) Image segmentation and selection of ROI in the first frame; (iv) Feature detection, extraction, tracking, matching and outlier discarding to measure dynamic displacement; and (v) System identification based on the measured displacements and derived acceleration.

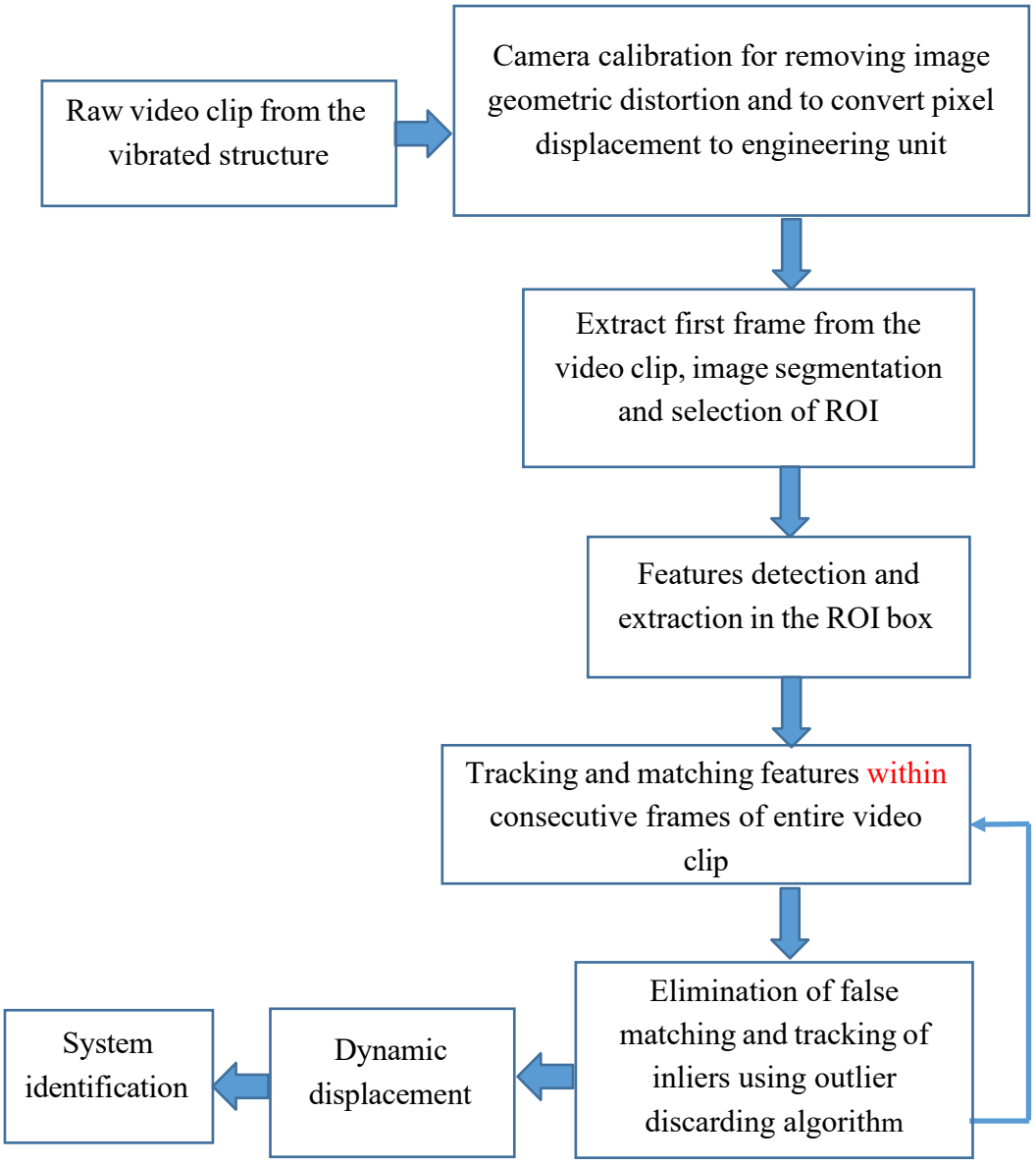


Figure 1: Overview of the proposed target-free vision-based approach for vibration measurements.

2.1 Camera Calibration

As the beginning step, camera calibration is performed to: (i) Remove the radial distortion from the captured images; (ii) Obtain camera intrinsic parameters; and (iii) Calculate the conversion factor from the pixel to engineering unit. Although consumer grade camera lenses are upgraded with the advanced modern camera technology intensely in the recent years, the FOV of low-cost and lightweight cameras may be introduced with considerable radial distortion. This will affect the accuracy of dynamic displacement measurements. From this point of view, camera calibration is essential to remove the geometric distortion from the images for the physical displacement measurement. The camera calibration method developed by Zhang [42] is adopted in this study by capturing images of known geometric points with the checkerboard pattern, from different points of view to estimate the camera intrinsic parameter matrix including camera focal length, lens axis offset, principal points, skewness and lens distortion characteristics. This permits the uncalibrated video images to be converted into calibrated images to measure dynamic displacements precisely. The conducted camera calibration uses a pinhole camera model and precisely estimate the camera intrinsic parameters. The camera pinhole model is expressed as follows: a two dimensional (2D) projected point $mm_{oo} = [uu,\ vv]^{TT}$ on an image plane obtained from a three dimensional (3D) point $MM_{oo} = [XX,\ YY,\ ZZ]^{TT}$ in the world coordinate system. The augmented vectors by adding a unity as the last element are denoted by $mm = [uu,\ vv,\ 1]^{TT}$ $aaaaaa$ $MM = [XX,\ YY,\ ZZ,\ 1]^{TT}$. The relationship between 3D points and its image projection is given as

$$SS \times mm = AA \times [RR\ tt] \times MM \tag{1}$$

$$\text{Or, } SS \times \begin{bmatrix} uu \\ vv \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha\alpha & \gamma\gamma & uu_0 \\ 0 & \beta\beta & vv_0 \\ 0 & 0 & 1 \end{bmatrix} [\,RR_{3x3}\ \ tt_{3x1}\,] \begin{bmatrix} XX \\ YY \\ ZZ \\ 1 \end{bmatrix}; \tag{2}$$

$$AA = \begin{bmatrix} \alpha\alpha & \gamma\gamma & uu_0 \\ 0 & \beta\beta & vv_0 \\ 0 & 0 & 1 \end{bmatrix}$$

where SS and AA are the arbitrary scale factor and camera intrinsic matrix, respectively; RR is the rotation matrix with three degrees of freedom; tt is the translation vector; $(uu_0,\ vv_0)$ is the coordinate of the principal point (optical centre); $\alpha\alpha$ and $\beta\beta$ are the scale factors in the uu and vv axes of the image, respectively; and $\gamma\gamma$ is the parameter describing the skewness of these two image axes. Camera

intrinsic matrix is solved by attaching the checkerboard pattern and analysing the captured image from different points of view.

To deal with the radial distortion, $(uu,\ vv)$ is considered as the ideal pixel image coordinates and $(uu_1,\ vv_1)$ the corresponding real observed image coordinates. The ideal points are also considered as the projection of the model points according to camera pinhole model. Likewise, $(xx,\ yy)$ and $(xx_1,\ yy_1)$ are defined as the normalized ideal and real image coordinates, respectively. Then we have

$$xx_1 = xx + xx[kk_1(xx^2 + yy^2) + kk_2(xx^2 + yy^2)^2] \quad (3)$$

$$yy_1 = yy + yy[kk_1(xx^2 + yy^2) + kk_2(xx^2 + yy^2)^2] \quad (4)$$

where kk_1 and kk_2 are known as the radial distortion coefficients and the principal point $(uu_0,\ vv_0)$ is the centre of the radial distortion. Based on $uu_1 = uu_0 + \alpha\alpha xx_1$ and $vv_1 = vv_0 + \beta\beta yy_1$ with the assumption $\gamma\gamma = 0$, we have

$$uu_1 = uu + (uu - uu_0)[kk_1(xx^2 + yy^2) + kk_2(xx^2 + yy^2)^2] \quad (5)$$

$$vv_1 = vv + (vv - vv_0)[kk_1(xx^2 + yy^2) + kk_2(xx^2 + yy^2)^2] \quad (6)$$

From Equations (5) and (6), two equations are obtained for each point in each image, that is

$$\begin{bmatrix} (uu - uu_0)(xx^2 + yy^2) & (uu - uu_0)(xx^2 + yy^2)^2 \\ (vv - vv_0)(xx^2 + yy^2) & (vv - vv_0)(xx^2 + yy^2)^2 \end{bmatrix} \begin{bmatrix} kk_1 \\ kk_2 \end{bmatrix} = \begin{bmatrix} uu_1 - uu \\ vv_1 - vv \end{bmatrix} \quad (7)$$

For mm points of aa images, $2 \times mm \times aa$ equations will be obtained to form a matrix, i.e. $DDkk = aa$, where $kk = [\ kk_1, kk_2]^{TT}$. Then the least square solution can be found by

$$kk = (DD^{TT}DD)^{-1}\ DD^{TT}aa \quad (8)$$

By using Equation (8), the radial distortion coefficients kk_1 and kk_2 are obtained to remove the geometric distortion from the video images. It is noted that the distortion function is mandatory to be computed once for each camera, and is used to remove distortion from the image frame obtained from video images taken by this camera.

In this study, the conversion ratio from the image pixel to an engineering unit, i.e. *mm* is calculated by using the known distance of two pre-selected points on the target between the image pixel coordinate and the world coordinate systems as below

$$RR_{ss} = \frac{DD}{NN} \text{ unit: } \frac{mmmm}{ppppxxpppp} \tag{9}$$

where RR_{ss} is the conversion factor from the image coordinate to the world coordinate; NN is the pixel distance of the object in the image coordinate and DD is the world coordinate distance of the object in mm.

It is simple to have the image coordinate pixel distance on an image, but to obtain the world coordinate on an image, especially when the structure to be investigated is out of reach which is very common for real structures, might be difficult. To resolve this issue, an alternative approach is suggested by forming a relationship between the conversion ratio RR_{ss} and camera distance ZZ from the aiming point to the camera position. The pinhole camera model is illustrated in Figure 2.

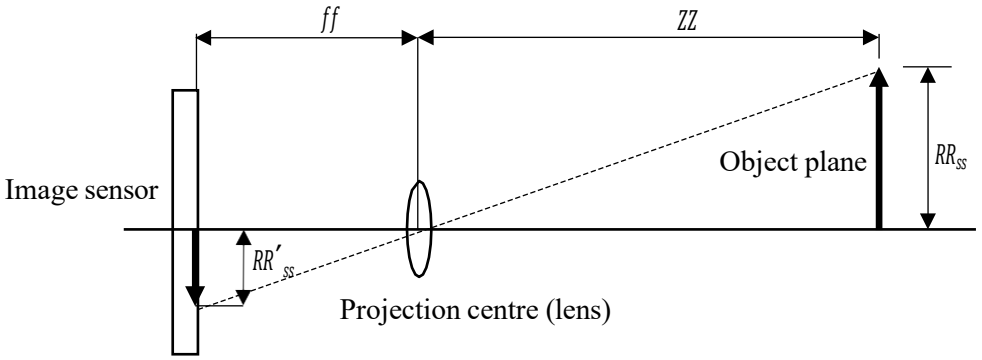


Figure 2: Camera calibration to determine the conversion factor.

From the similar triangle which is made by the image sensor and object plane with the lens projection centre, we have

$$\frac{RR'_{ss}}{RR_{ss}} = \frac{ff}{ZZ} \tag{10}$$

It can be simplified as

$$ZZ = \frac{ff}{RR'_{ss}} RR_{ss} \tag{11}$$

$$ZZ = KKRR_{ss} \tag{12}$$

$$RR_{ss} = \frac{ZZ}{KK} \tag{13}$$

where $KK = \frac{ff}{RR'_{ss}}$ is the camera constant which needs to be determined; ff is focal length and RR'_{ss} is the image sensor spatial resolution; ZZ is the distance from the object plane to the projection centre of the camera lens. For a known dimension object in the image coordinate and world coordinate systems, RR_{ss} is to be calculated by using Equation (9). With calculated RR_{ss} and known distance ZZ, the camera constant KK can be computed by using Equation (13). Once KK is found, for any known distance from the camera to the object, the conversion ratio RR_{ss} for any unknown object dimension can be calculated by using Equation (13).

For professional cameras with higher specifications, the values of camera constants may be provided by its manufacturing company. The camera calibration is necessary to obtain the camera constant values for most of the consumer level cameras. This is why the values of distance ZZ and conversion factor RR_{ss} are obtained by using the above mentioned camera calibration algorithm in this study. In this research, SONY PXW-FS5 4K XDCAM video camera is calibrated by using a checkerboard of 7x10 square shape of alternate black and white colours with a square unit dimension of 21 x 21 mm. A series of videos are taken by targeting the checkerboard from different known distances, denoted as $ZZ_{mmmmmmss}$. The value of $NN_{mmmmmmss}$ is obtained from the checkerboard image in the pixel by using an image processing technique known as Harris corner detection algorithm [32]. It should be noted that the maximum zoom factor is used during all the camera implementation while the focus is adjusted manually to eliminate blurriness from images. The conversion ratio $RR_{mmmmmmss}$ is calculated as

$$RR_{ss_{mmmmmmss}} = \frac{DD_{mmmmmmss}}{NN_{mmmmmmss}} \text{unit: } \frac{mmmm}{ppppxxpppp} \quad (14)$$

where $NN_{mmmmmmss}$ is the pixel distance of two corner points on the checkerboard image; $DD_{mmmmmmss}$ is the world dimension of the unit square on the checkerboard (21mm). A series of images are captured and measured values of $RR_{ss_{mmmmmmss}}$ corresponding to various camera distances $ZZ_{mmmmmmss}$ are given in Table 1. To obtain the relationship between the conversion ratio and distance, the results obtained from calibration tests are curve fitted. The fitted curves in Figure 3 show that the conversion factor and distance are proportional, which is supported by Equation (13). The relationship function derived from the curve fitting is provided as

$$RR_{ss} = \frac{ZZ^{1.1442}}{33333} \text{unit: } \frac{mmmm}{ppppxxpppp} \quad (15)$$

When the conversion ratio is calculated, the FOV of the camera in the horizontal and vertical directions are also computed by using the following relationship

$$FFFFFF = FF * RR_{ss} \quad (16)$$

where FF is the number of sensor resolution in pixel. In this study, $FF = [1920xx1080]$ pixel is adopted for the calculation of FOV in both horizontal ($FFFFFF_h$) and vertical ($FFFFFF_{vv}$) directions, respectively. The results are shown in Figure 3 and Table 1.

Table 1: Measured conversion factor corresponding to camera distance and field of view.

Test No.	$ZZ_{mmmmmmss}$ (*mmmm*)	$DD_{mmmmmmss}$ (*wwwwwppaa*, *mmm*),	$NN_{mmmmmmss}$ (*ppmmaaiipp*, *ppppxxpppp*)	$RR_{ss_{mmmmmmss}}$ $\frac{mmmm}{ppppxxpppp}$	$FFFFFF_h$ (*mmmm*)	$FFFFFF_{vv}$ (*mmmm*)
1	1810	21.0	113.190	0.186	356.215	200.371
2	2860	21.0	70.350	0.299	573.134	322.388
3	3750	21.0	53.550	0.392	752.941	423.529
4	4850	21.0	39.060	0.538	1032.258	580.645
5	5500	21.0	35.490	0.592	1136.095	639.053
6	6900	21.0	27.090	0.775	1488.372	837.209
7	7970	21.0	25.200	0.833	1600.000	900.000
8	8970	21.0	22.680	0.926	1777.778	1000.000
9	10500	21.0	18.795	1.117	2145.251	1206.704
10	12170	21.0	15.813	1.328	2549.801	1434.263
11	13880	21.0	11.739	1.789	3434.705	1932.021
12	15890	21.0	10.059	2.088	4008.351	2254.697
13	17640	21.0	8.946	2.347	4507.042	2535.211
14	19690	21.0	7.665	2.740	5260.274	2958.904
15	21830	21.0	7.119	2.950	5663.717	3185.841
16	23920	21.0	5.502	3.817	7328.244	4122.137

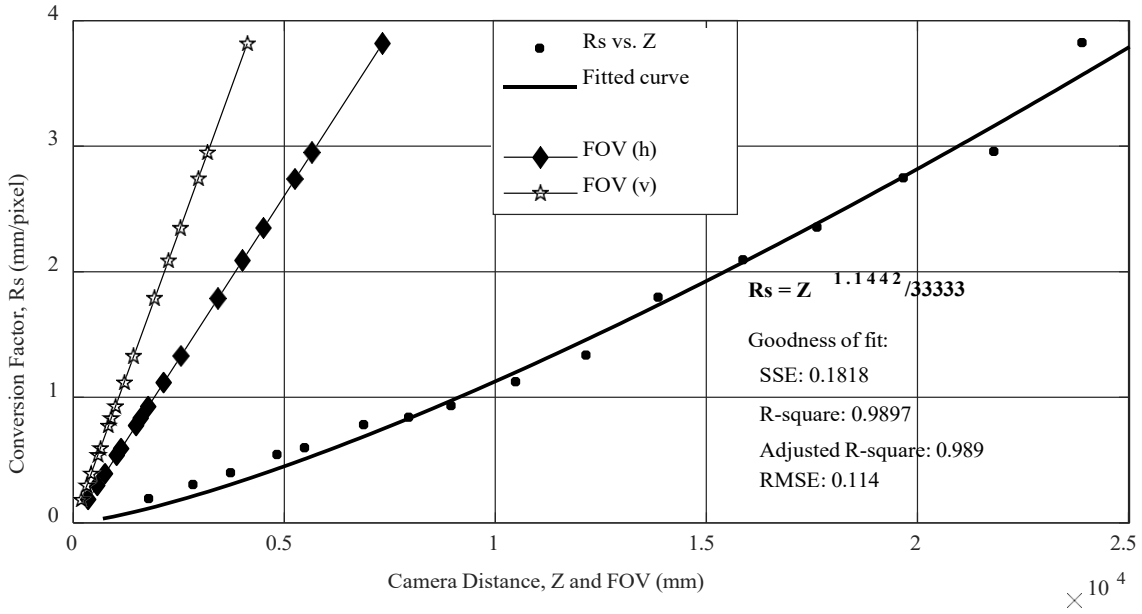


Figure 3: Relationship between camera distance and FOV and the conversion factor for SONY PXW-FS5 4K XDCAM.

2.2 Image processing for displacement measurement

By using the camera intrinsic parameters such as focal length, skewness and principle points, the captured video is calibrated to eliminate the geometric distortion from images and the vibration displacement is calculated by analysing the captured video clip frame-by-frame. In the beginning, the first frame is extracted from the video clip and is segmented using an image segmentation algorithm [43] to select ROI by drawing a user-defined box. The ROI covers the object of interest to be targeted and tracked along the entire video clip. In the field of image processing, ROI should consist of some special image features or key points such as image corner points and the image edges, which have dominant characteristics compared with its neighbour. Key features in the images could be identified by using different feature detection algorithms, such as Harris corner features [32], scale invariant feature transformation (SIFT) [44], BRISK [45], speed-up-robust features (SURF) [46], fast retina key points (FREAK) [47] and features from accelerated segment test (FAST) [48] etc. Their applicability is varying by their capability of detection and invariant to scale, translation, rotation, light and illumination along with other environmental perspectives for accurate dynamic displacement monitoring. Conventional vision-based methods for vibration displacement measurement employ physical targets such as known dimension circles or rectangles, which are attached on the surface of the testing structure. The motions of these targets are tracked, and the

displacement is measured by means of different image processing algorithms. In this study, the requirement of an artificial target is eliminated by using natural features of the testing structure as targets. The BRISK algorithm [45] is adapted to detect and extract key features from the images, for calculating the dynamic displacement of the vibrating test structure. The powerful and fast BRISK algorithm reduced the computational time for large volume image processing while it is invariant to scale and rotation. This is one of the advantages of this study to detect and extract key-features while the images are captured with a certain camera angle. The BRISK detector is used to estimate the true scale of each key point in a continuous scale space and uses a pattern for sampling the neighbourhood of the key point. For positioning and scaling of a key point kk in the image, considering $NN \times (aa - 1)/2$ sampling-point pairs $(uu_{ii},\ uu_{jj})$, the intensity values of these points are $II(uu_{ii},\ \sigma\sigma_{ii})$ and $II(uu_{jj},\ \sigma\sigma_{jj})$ and then the local gradient $ii(uu_{ii}, uu_{jj})$ is estimated as

$$ii(uu_{ii}, uu_{jj}) = (uu_{jj} - uu_{ii}). \frac{II(uu_{jj}, \sigma\sigma_{jj}) - II(uu_{ii}, \sigma\sigma_{ii})}{\|uu_{jj} - uu_{ii}\|^2} \quad (17)$$

where $(\sigma\sigma_{ii}, \sigma\sigma_{jj})$ is the standard deviation of intensity which is proportional to the distance between the points. If AA is a set of all sampling pairs

$$AA = \{(uu_{ii}, uu_{jj}) \in \mathbb{R}^2 | \text{pp} < NN \wedge jj < pp \wedge \text{pp}, jj \in \mathbb{N}\} \quad (18)$$

SS is a short distance sub-set pairs and another subset of LL is the long-distance pair $\mathcal{L}$, and they can be expressed as

$$SS = \{uu_{ii}, uu_{jj}) \in AA \| uu_{jj} - uu_{ii} \| < \delta\delta_{mmmmxx}\} \subseteq AA \quad (19)$$

$$\mathcal{L} = \{uu_{ii}, uu_{jj}) \in AA \| uu_{jj} - uu_{ii} \| < \delta\delta_{mmiimm}\} \subseteq AA \quad (20)$$

where $\delta\delta_{mmmmxx} = 9.75tt$ and $\delta\delta_{mmiimm} = 13.67tt$; tt is the scale of kk. Then the overall characteristics of the gradient of the key point kk is

$$ii = \begin{matrix} ii_{xx} \\ ii_{yy} \end{matrix} = \frac{1}{LL} \sum_{(uu_{ii}, uu_{jj}) \in \mathcal{L}} ii(uu_{ii},\ uu_{jj}) \quad (21)$$

The pattern rotation $\alpha\alpha = aawwaattaaaa2(ii_{yy}, ii_{xx})$ is estimated around the key point kk by considering the short distance intensity comparison of point pairs $(uu_{ii}^{\alpha\alpha}, uu_{jj}^{\alpha\alpha}) \in SS$, and each bit bb corresponds to

$$bb = \begin{cases} 1, II\left(uu_{jj}^{\alpha\alpha}, \sigma\sigma_{jj}\right) > II\left(uu_{ii}^{\alpha\alpha}, \sigma\sigma_{ii}\right) \\ 0, \ wwtt\boldsymbol{h}ppwwwwppeepp \end{cases} \quad (22)$$

$$\forall \left(uu_{ii}^{\alpha\alpha}, uu_{jj}^{\alpha\alpha}\right) \in SS$$

When key points detection and extraction are performed from the first frame in the selected ROI, the robust KLT algorithm [49] is used to track the key features in the images of entire video clip which ensure the maximum number of matching of key-points between consecutive image. Considering that the intensity of a point $XX = (xx, yy)$ in the current image frame at the time instant tt is (xx, yy, tt) , the displacement of point XX is $aa = (\mu\mu, \rho\rho)$ and the intensity of point XX at the time instant $tt + \tau\tau$ is

$$II(xx, yy, tt + \tau\tau) = II(xx - \mu\mu, yy - \rho\rho, tt) \quad (23)$$

The intensity of current image frame $JJ_{xx} = II(xx, yy, tt + \tau\tau)$ and the previous image frame $II\ (XX - aa) = II(xx - \mu\mu, yy - \rho\rho, tt)$ for a small motion is represented by

$$JJ_{xx} = \ II\ (XX - aa) + aa(XX) = II\ (XX) - \ ii.aa \quad (24)$$

where ii is the gradient vector; d is the displacement vector between two frames and aa is the noise assumed to zero.

The residual of the intensity change for a small window ww can be found as

$$\varepsilon\varepsilon = \int_{ww}[II(XX) - ii.aa - JJ(XX)]^{2} wwaaXX = \int_{ww}(h - ii.aa)^{2} wwaaXX \quad (25)$$

where $h = II(XX) - JJ(XX)$. The residual value when differentiating Equation (26) with respect to aa, is equal to zero

$$\int_{ww}(h - ii.aa)iiwwaaAA = 0 \quad (26)$$

With $(ii.aa)ii = (iiii^{TT})aa$ and aa assumed to be constant within the window ww, we have

$$aa = \ ppGG^{-1} \quad (27)$$

where $pp = \int_{ww}(II - JJ)iiwwaaAA$ and $GG = \int_{ww} iiii^{TT}wwaaAA$, the displacement vector aa is calculated for each matched key features XX in each frame of the video.

During the matching and tracking, the outlier of the matching features is discarded by using outlier removing algorithm. Several methods have been developed for geometric transformation and estimation of multiple view relationship between corresponding features points, such as random sample consensus (RANSAC) developed by Fischler and Bolles [50], least median square (LMS) by Rousseeuw [51] and Maximum likelihood estimation of sample consensus (MLESAC) by Torr and Zisserman [52]. The motion of points between two views is estimated by the fundamental matrix, planar projective transformation and quadratic transformation. If the sets of image feature points $x\bar{x}_{ii}$ in the first image frame and the transformed feature points xx'_{ii} in the second image frame with the motion, their positions related for the translation between views can be expressed by

$$xx'_{ii}FFx\bar{x}_{ii} = 0 \tag{28}$$

where $x\bar{x}_{ii} = (xx,\ yy,\ 1)^{TT}$ is homogeneous image coordinate and FF is the fundamental matrix. Fundamental matrix is a $3xx3$ singular matrix, which describes the epipolar geometry by which two perspective images of a single scene are related. The fundamental matrix can be estimated by parameterization and optimization procedure of image matching algorithm. If all points lie on a plane and the camera rotation instead of the translation occurs, a planar projective transformation is given as

$$xx'_{ii} = HHx\bar{x}_{ii} \tag{29}$$

where HH is the transformation matrix. Quadratic transformation is a combination of camera motion and scene with all scene points. The camera optic centre lying on the quadric surface and points are related by using

$$xx'_{ii} = FF_1x\bar{x}_{ii}\ \mathrm{x}\ FF_2x\bar{x}_{ii} \tag{30}$$

where F_1 and F_2 are also the fundamental matrices.

In this study, MLESAC is adopted which produced maximum error-free dynamic displacement. The projective transformation matrix is taken into account for the out-of-plane motion of the structure. Projective transformation requires a minimum of four pairs of key features to track and match between the consecutive frames. Gaussian noise is considered on the image of two images coordinate $(uu,\ vv)$ and $(xx,\ yy)$ with zero mean and uniform standard deviation $\sigma\sigma$, and the probability density function is

$$PP_{rr} = (DD|MM) = \prod_{ii=1,\ldots,mm} \frac{1}{\sqrt{2\pi\pi\sigma\sigma}} pp^{\frac{\sum_{jj=1,2}(uu^{jj}_{ii}-xx^{jj}_{ii})^2+(vv^{jj}_{ii}-yy^{jj}_{ii})^2}{2\sigma\sigma^2}} \quad (31)$$

where M , n and D are the view relation projective matrix, the number of correspondences and the match sets, respectively. The residual error of the projective transformation is the sum of the distance between the original and transformed features. For correspondence $xx^{1,2}$, with the maximum likelihood estimate $\hat{x}^{1,2}$ of the true position $xx^{1,2}$, the maximum likelihood error (MLE) for any point is calculated by

$$pp_{ii}^2 = \sum_{jj=1,2} (\hat{x}^{jj}_{ii} - xx^{jj}_{ii})^2 + (\hat{x}^{jj}_{ii} - yy^{jj}_{ii})^2 \quad (32)$$

The error is minimized by using expectation maximization algorithm, when points are fitted more closely to the given transformation. The error minimization is given as

$$LL = \sum_{ii} ppwwii\, \gamma\gamma \frac{1}{\sqrt{2\pi\pi\sigma\sigma}}^{mm} ppxxpp\, \exp\left(-\sum_{jj=1,2} (\hat{x}^{jj}_{ii} - xx^{jj}_{ii})^2 + (\hat{x}^{jj}_{ii} - yy^{jj}_{ii})^2 \right)(2\sigma\sigma^2) + \left(1 \frac{1}{\vartheta\vartheta}\gamma\gamma\right) \quad (33)$$

where $\gamma\gamma$ is the mixing parameter, and $\vartheta\vartheta$ is a constant. The same procedure is repeated until all the outliers are removed, and the inlier is tracked by using the KLT algorithm from the first frame image to the end of the video frame. Finally, the displacement of the object is found as a pixel unit which is converted to engineering unit by using the conversion factor obtained from camera calibration results, as described in Section 2.1. Technical difficulties involved with this study are camera vibration during image acquisition, processing of large volume of images, camera configuration and functional setting for high-quality image acquisition, and synchronization of displacement measurements between laser sensor and proposed vision-based technique. Details of these difficulties and their corresponding technical solution are described in the results and discussion section.

2.3 Obtaining the acceleration and modal Identification

Dynamic vibration displacement obtained from the proposed approach is used to identify the dynamic characteristic of the structure. The analysis consists of two parts: at the beginning, time domain acceleration responses are derived from the time domain displacement data obtained from

the proposed approach. The time domain displacement history is low-pass filtered, and numerical central difference algorithm is implemented for the second derivative to calculate the time domain acceleration. In the second step, natural frequencies of the structure are identified with fast Fourier Transformation (FFT). The results are compared with those obtained from the wired sensory systems with installed accelerometers and LDSs.

3 Experimental Setup for Validations

A series of experimental tests are conducted to verify the accuracy and effectiveness of using the proposed approach in measuring the vibration displacement and acceleration of civil engineering structures subjected to out-of-plane movements. Figure 4 shows the testing reinforced concrete column model subjected to the different magnitudes of bi-directional ground motions in the laboratory. The reinforced concrete column has a diameter of 100 mm and a height of 600 mm, which is supported on a footing with a dimension of 400 mm x 400 mm x 150 mm. The dimension of the column top cap is 500mm x 500mm x 150 mm, which is used to support a top slab of 1000 x 1000 x 150 mm to simulate the mass on the column. Moreover, the column was constructed with concrete of compressive strength 37 MPa and reinforced with 6 mm diameter longitudinal rebar with 3 mm diameter stirrup provided yield strength of 555 Mpa and 346 Mpa, respectively. Furthermore, a prestress force was applied to the column model by using 9.3 mm diameter tendon of yield strength 1674 Mpa and modulus of elasticity 195 GPa. The foundation of the structural column model is fixed on four separate shaking tables, which can be used to provide bi-directional ground motion excitations. In this study, the band limited white noises are used with the peak ground acceleration (PGA) magnitudes from 0.01g to 1.0g, which represent the structural vibration conditions from very minor level excitations to strong excitations. To validate the accuracy of using the proposed approach for displacement and acceleration measurements, LDSs and accelerometers are installed to take the vibration measurements for comparison analyses.

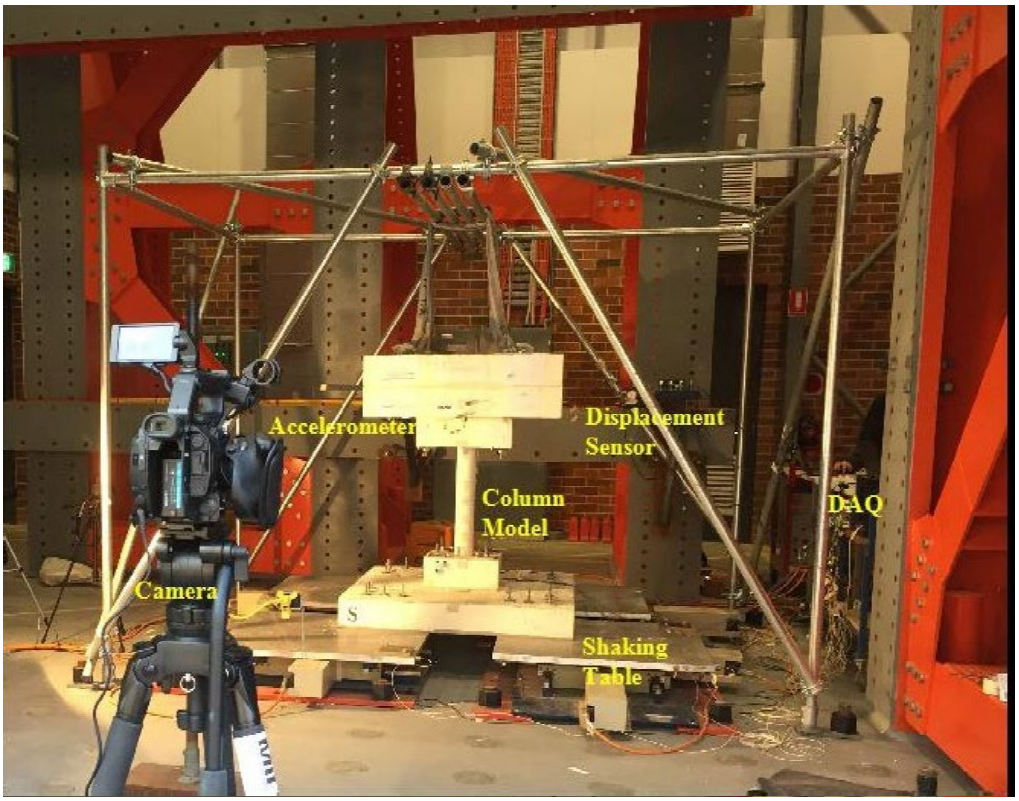


Figure 4: Experiment setup for the proposed vision-based vibration measurement.

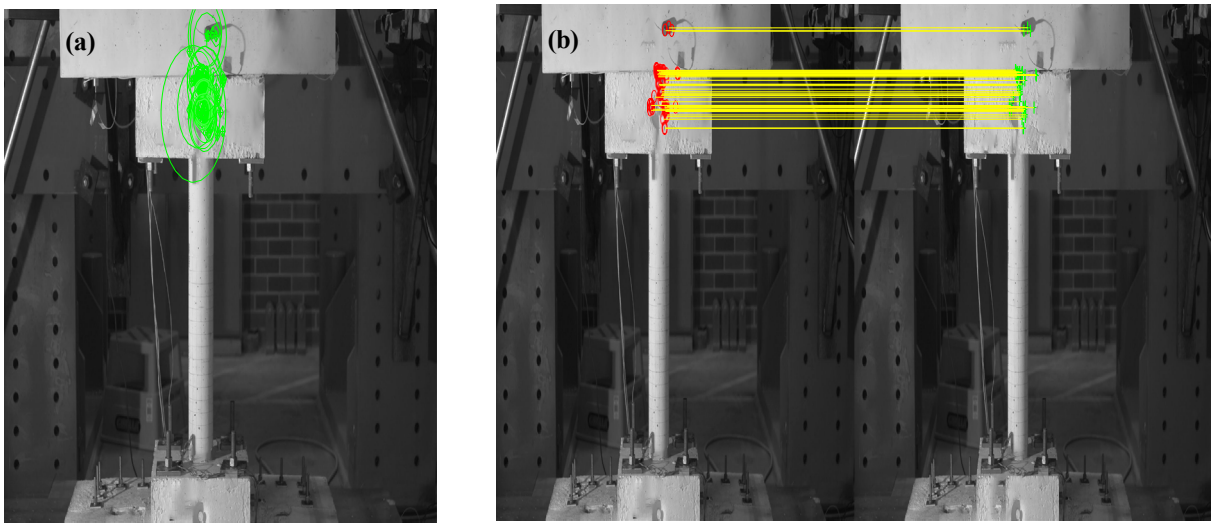


Figure 5: Features detection and tracking: (a) Features detection and extraction; (b) Features tracking and matching between consecutive frames.

Sony PXW-FS5 4K XDCAM camera is used to record the video of the structural vibrations. The camera can record videos with a frame rate of 50 frame per second (fps) and a spatial resolution of 1920x1080 pixel in slow and quick motion modes. However, the camera is also capable of recording

videos with a maximum frame rate of 960 fps at a reduced spatial resolution of 1280x720 pixel in the super slow motion mode. In this study, videos of the structural vibration during the shaking table tests are captured with a frame rate of 50 fps and 100 fps with a resolution of 1920x1080 pixel, since the use of high resolution and frame rate is important for accurate feature detection and tracking in vibration measurement as well as the subsequent modal identification. The BRISK algorithm [45] is adapted to detect and extract key features from the images, and the robust KLT algorithm [49] is used to track the key features in the images of entire video clip between the consecutive frames for calculating the dynamic displacement of the vibrating test structure as shown in Figure 5. The camera calibration described in Section 2.1 is conducted to remove the radial distortion from image geometry and to obtain the unit conversion factor. The camera is focused on the structure with a vertical angle of 9 degree and horizontal angle 5 degree to the perpendicular line of sight of the testing structure. The distance from the testing structure to the camera was 4 m in different tests.

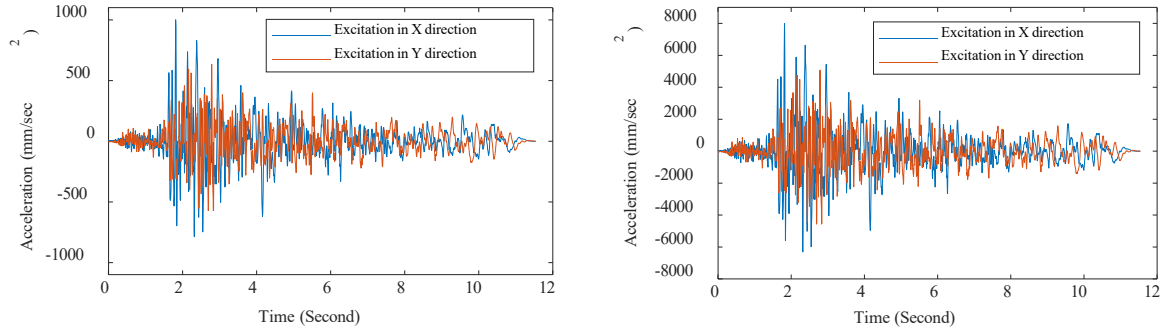
Three CMOS multi-function LDSs (Keyence IL-300) are installed on a fixed steel frame to measure the translational displacement of the column model at the top of the column. Seven accelerometers (PCB 393B04) are installed on the testing structural model at different heights and sides. All the accelerometers and LDSs are connected to a HBM data acquisition (DAQ) system, which is connected to a laptop to collect and store the vibration responses. The sampling frequency is set as 200 Hz with the active anti-aliasing filter.

4 Experimental Results

4.1 Displacement measurement

The accuracy of the proposed approach is experimentally investigated in this study to measure the structural dynamic displacement. The videos are captured and analyzed to obtain the structural vibration when the structure is subjected to different bi-directional ground motions. To further demonstrate the potential applicability of the measured displacements from the proposed method, structural modal identification is also performed. The obtained time domain displacement responses from the proposed approach are compared with those by using the LDSs. Although eleven tests are carried out, four of them are analysed in this study with ground motions of different PGAs, i.e., 0.01g, 0.1g, 0.3g and 0.8g. These four tests represent the structural vibration conditions from very minor ambient excitations to strong excitations. Figures 6(a) and (b) show the measured input ground motions for cases with a PGA of 0.1g and 0.8g, respectively. It can be observed that the inputs along

X- and *Y*- directions are significantly different. Dynamic displacement results in these four tests are shown in Figure 7 to Figure 10.



(a) (b)

Figure 6: Input ground motions in the both directions of the testing: (a) 0.1g; (b) 0.8g.

The time domain displacement resposnes measured from the LDS and calculated from the proposed approach are shown in Figure 7(a) and (b), respectively, for the testing case with a PGA of 0.01g. It is observed that the displacement responses are less than 1mm, which means minor vibrations of the structure. Figure 7(a) shows the comparsion between the dynamic displacements obtained from the two methods. The comparson indicates that the displacment response time history obtained from the proposed approach has a very good match with that measured from LDS up to 48 seconds. Afterwards, there is some visible difference between the two displacement results due to the rotation of the structure, evidenced by the different ground motions in both directions as shown in Figure 6. The reason for this difference is that the video is captured from the front side of the structure while LDS is located on the right side of the model to measure the vibration displacement. At the end of each excitation test, the structure experiences certain rotations, which cannot be properly measured by the LDS because it only measures the displacement perpendicular to the transducer. This measurement inaccuracy causes a difference at the end of vibration test. Even though the maximum displacement amplitude obtained from the both methods is less than 1 mm, the proposed image based method still gives accurate measurement with minimum differences from those recorded by LDS. This demonstrates that the proposed approach is powerful to capture the minor displacement responses accurately.

Accurate displacement measurement results are also observed for the test with excitations of 0.1g, as shown in Figure 8. The dynamic displacement results from the proposed approach show a good agreement with those from LDS. A minor difference is observed from Figure 8(a), at the end of vibration test after 20 seconds because of structural rotation as mentioned above. It is found that the rotational response always occurs, but when response is large its influence is insignificant because it is relatively small compared to the lateral displacement. It is also noted that for the above mentioned two cases, the videos are captured with a frame rate of 50 fps and a spatial resolution of 1920 x 1080 pixel. The videos of the structure under the ambient excitations with PGA magnitudes of 0.3g and 0.8g are captured with a higher frame rate of 100 fps and the same resolution of 1920 x 1080 pixel. The displacement results obtained from the proposed method for the testing case of 0.3g well match with those measured by using LDS, as shown in Figure 9. In addition, for the large excitation test of 0.8g, the proposed approach can measure dynamic displacement precisely until the rotation of the structure is observed after 13 seconds, as shown in Figure 10. For all the test cases, there is about 32% difference of excitation input in the both directions. When the structure experiences a certain degree of rotation due to the bi-direction excitation force and out-of-plane movement, this will cause the inaccurate displacement measurement from LDS. Furthermore, there is a bit phase difference observed between two displacement signals at the beginning of bi-directional ground motion as shown in Figure 9(a) and Figure 10(a). The reason for this phase difference is a degree of rotation of the structural model as the model subjected to 32% variation of bi-directional ground motion in both directions and out-of-plane movement. It is also noted that the displacement measurement obtained from the proposed vision-based method and laser sensor were synchronized by using cross-correlation function between two signals. Timeshift is determined corresponding to the maximum cross-correlation function and the obtained time shift is used to align two displacement signals during analysis.

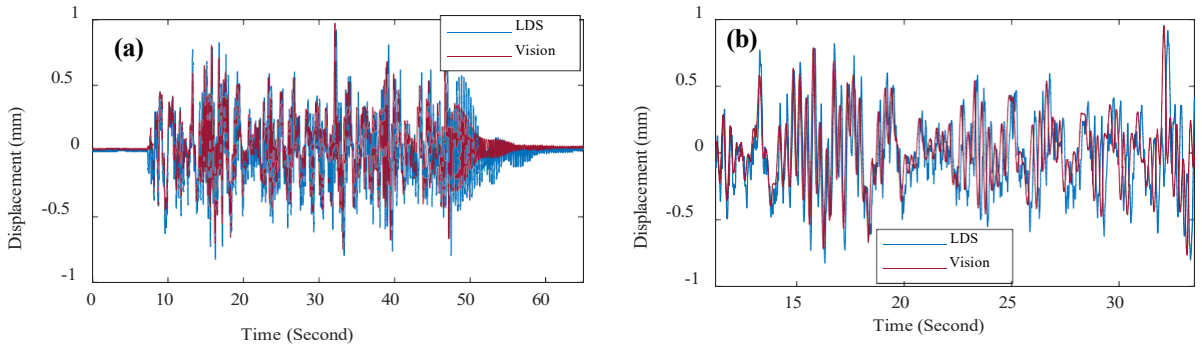


Figure 7: For bi-directional excitations with a PGA of 0.01g: (a) comparison of LDS and vision-based displacement; (b) a zoom window for comparison of displacement.

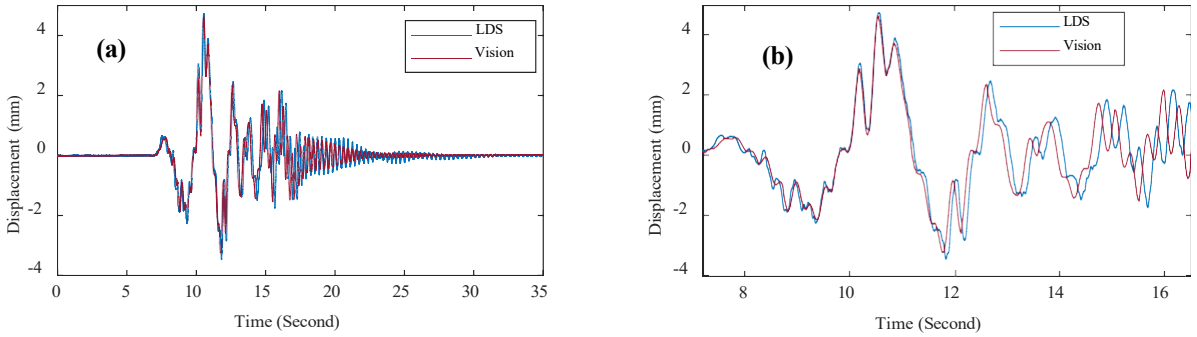


Figure 8: For bi-directional excitations with a PGA of 0.1g: (a) comparison of LDS and vision-based displacement; (b) a zoom window for comparison of displacement.

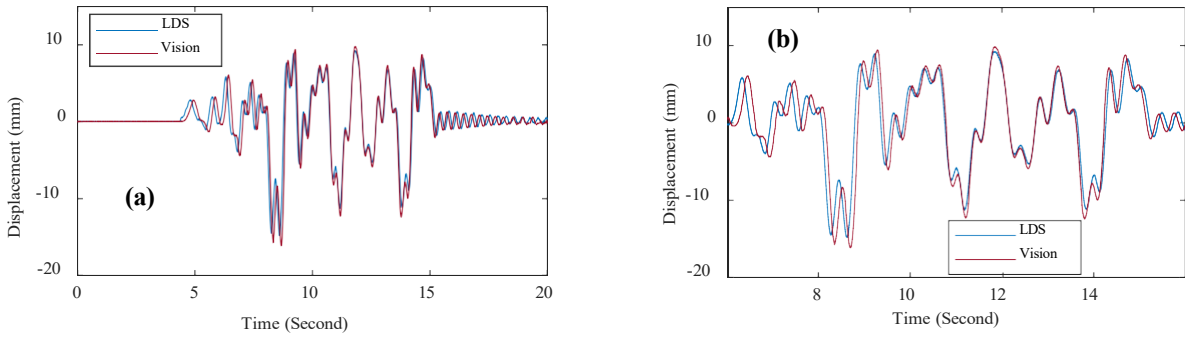


Figure 9: For bi-directional excitations with a PGA of 0.3g: (a) comparison of LDS and vision-based displacement; (b) a zoom window for comparison of displacement.

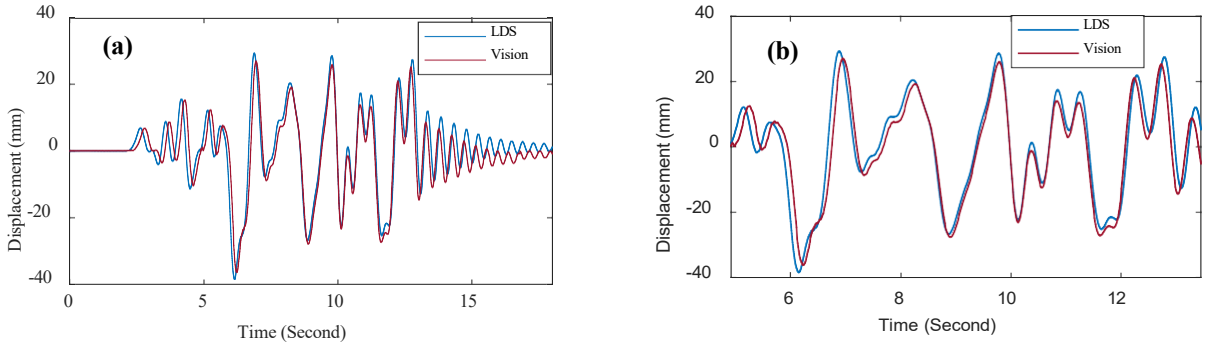


Figure 10: For bi-directional excitations with a PGA of 0.8g: (a) comparison of LDS and vision-based displacement; (b) a zoom window for comparison of displacement.

More experimental tests with bi-directional excitations from 0.01g to 1.0g are conducted, and the videos are analyzed to extract the displacement, which are compared with those from the conventional LDS measurements. The error analysis of the obtained results is presented in Table 2**.** The errors in the maximum displacement, the correlation between two measurements are calculated. It is observed from Table 2 that the proposed approach is able to accurately extract the maximum dynamic displacement with a minimum and a maximum error of 2.2% and 5.7%, respectively when compared to the conventional displacement sensor. It is also noted that when compared with LDS measurements, the proposed approach is capable of measuring the dynamic displacements of the structure subjected to different excitations with a Root-Mean-Square-Error (RMSE) of 1.14 mm. The maximum and minimum errors are observed when the videos are captured in 50 fps for excitations of 0.2g and 0.1g, respectively.

The correlation between measured displacement from sensor and obtained displacement from the proposed vision-based method is established by determining the coefficient of correlation ($\rho\rho$) and coefficient of determination (RR^2) using Equations (34) and (35), respectively.

$$\rho\rho = \frac{|\sum_{ii}(xx_{LL}(\text{pp}) - \mu\mu_{LL}) \times (xx_{vv}(\text{pp}) - \mu\mu_{vv})|}{\sqrt{\sum_{ii}(xx_{LL}(\text{pp}) - \mu\mu_{LL})^2}\sqrt{(xx_{vv}(\text{pp}) - \mu\mu_{vv})^2}} \tag{34}$$

$$RR^2 = 1 - \frac{\sum_{ii}(xx_{LL}(\text{pp}) - xx_{vv}(\text{pp}))^2}{\sum_{ii}(xx_{LL}(\text{pp}) - \mu\mu_{LL})^2} \tag{35}$$

where xx_{LL} and xx_{vv} are the dynamic displacements obtained from LDS and the proposed vision-based method, respectively; $\mu\mu_{LL}$ and $\mu\mu_{vv}$ are the mean values of two displacement data sets. The peak values of the whole time history displacement of LDS and vision based method data sets are used to measure the correlation coefficient. From Table 2**,** it is seen that the correlation coefficient varies from 0.9215 to 0.9787 and the coefficient of determination values ranges from 0.9831 to 0.9994, which shows the strong relationship and similarity between the conventional LDS and the proposed vision-based displacement responses. It is concluded that the displacement obtained from the proposed approach has a very good agreement and correlation with that measured by using the conventional displacement sensor.

The accurate displacement measurement results demonstrate that the proposed target-free vision-based appraoch can be used as an accurate and efficient technique to measure the dynamic vibration displacements for SHM and other purposes. It shall be noted that one significant advantage of the proposed method is that no target is required in the experimental setup on the testing structure. This may significantly reduce the complexity of using the vision-based techniques for the vibration measurements of real word engineering structures, particularly for those structures which might be very difficult or inaccessible to install the artificial targets.

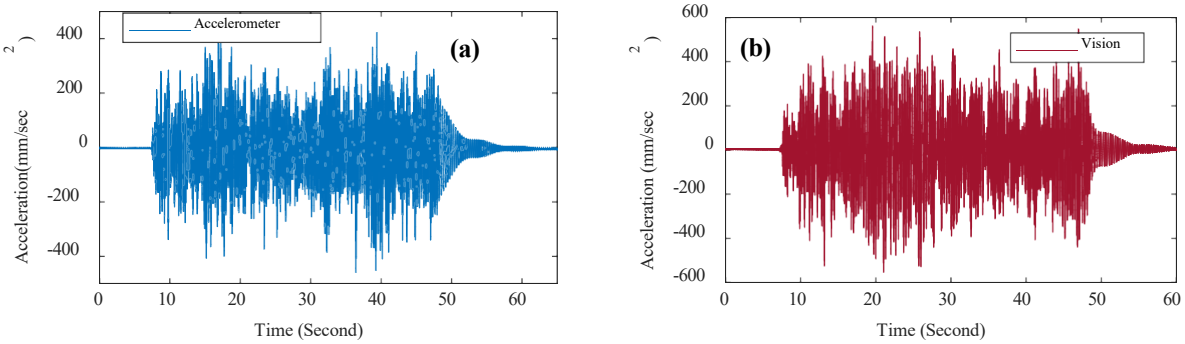
Table 2: Comparison of displacement results and error analysis of the proposed method.

Test No.	Excitation Force (g)	Camera Frame Rate (fps)	Vision Based Method				
			LDS	Maximum displacement (mm)	Error (%)	Corr.($\rho\rho$)	RR^2
1	0.01	50	0.9151	0.9535	-4.1	0.9628	0.9984
2	0.1	50	4.7195	4.6110	2.2	0.9531	0.9925
3	0.2	50	14.3357	13.5090	5.7	0.9384	0.9912
4	0.3	100	16.8042	16.1425	3.9	0.9725	0.9989
5	0.4	100	18.6223	19.2752	-3.5	0.9787	0.9994
6	0.5	100	23.7109	22.5254	4.9	0.9618	0.9943
7	0.6	50	29.3151	30.8187	-5.1	0.9135	0.9831
8	0.7	50	32.3695	33.2156	-2.6	0.9364	0.9876
9	0.8	100	35.2334	36.3925	-3.2	0.9687	0.9969

10	0.9	100	38.4591	40.1044	-4.2	0.9358	0.9928
11	1.0	100	53.2391	55.3295	-3.9	0.9215	0.9953

4.2 Obtaining acceleration and modal identification

Vibration acceleration is obtained by using numerical central difference method to differentiate the displacement obtained from the proposed approach for the testing structural model under bi-directional excitations and out-of-plane movements. These acceleration responses are also compared with that measured from the installed accelerometers to investigate whether accurate acceleration measurement can be achieved. Modal identification is conducted to identify the vibration characteristics of the structure. It is expected that extracting the displacement under lower level vibrations is more challenging than that under the higher level vibrations. To investigate the effectiveness of using the vibration measurement from the proposed approach for modal identification, two tests with PGA magnitudes of 0.01g and 0.3g are considered for determining acceleration and natural frequency. It is noted in Table 2 that the videos of the experimental tests are taken with 50 fps with a spatial resolution of 1920x1080 pixel for the case with 0.01g excitation and 100 fps with the same spatial resolution for the case with 0.3g excitation. The acceleration responses measured from the accelerometer and obtained from the proposed vision-based approach are compared, and the results from modal identification are shown in Figure 11 and Figure 12 for the testing cases of 0.01g and 0.3g, respectively.



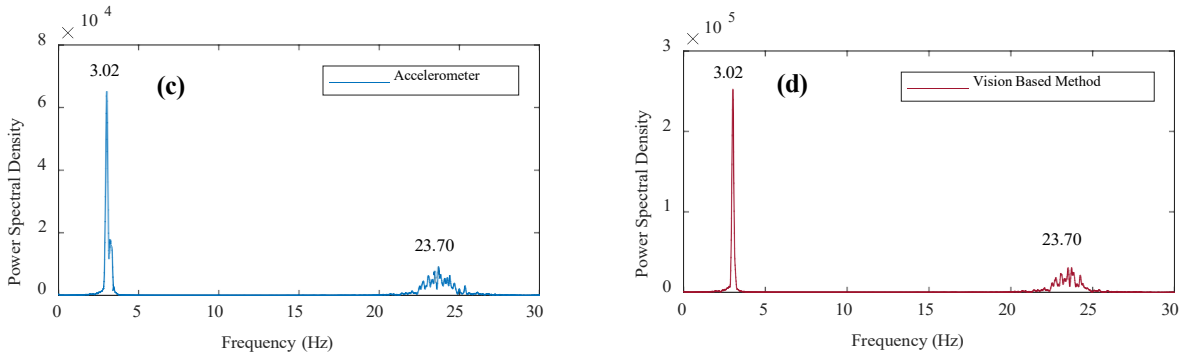


Figure 11: Results for the test with 0.01g excitation: (a) acceleration from the installed accelerometer; (b) derived acceleration from the proposed method; (c) natural frequency obtained from the measured acceleration; (d) natural frequency obtained from the derived acceleration with the proposed method.

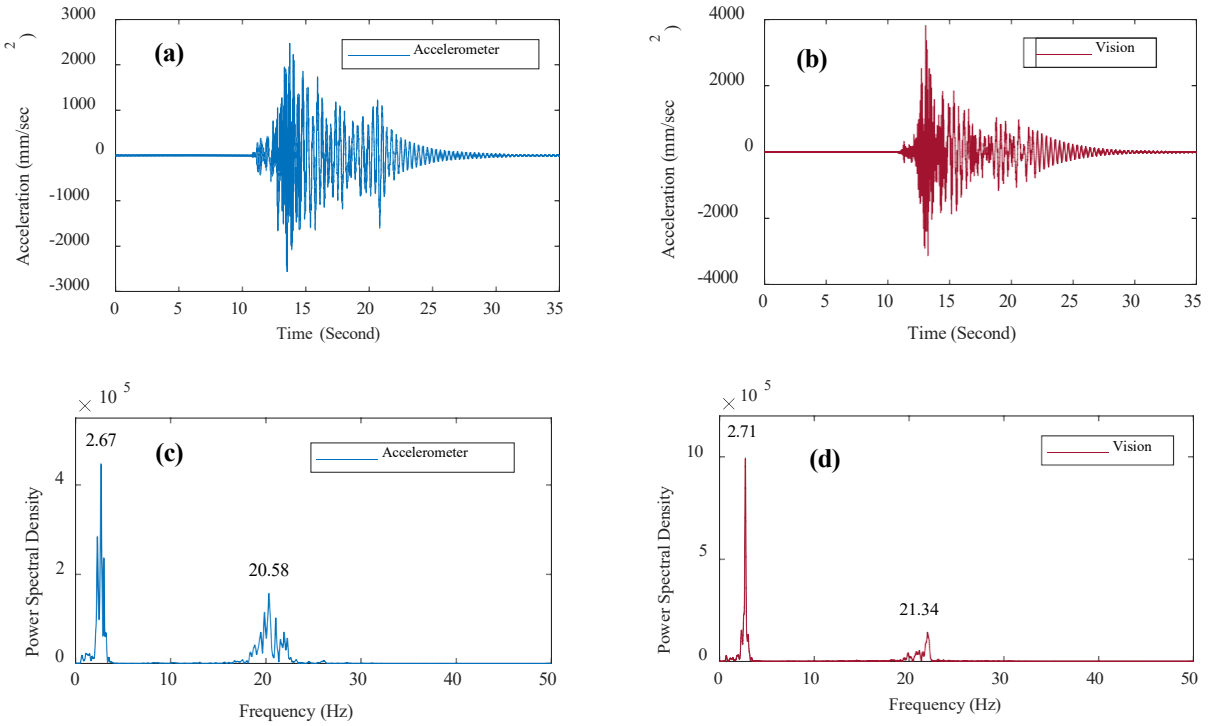


Figure 12: Results for the test with 0.3g excitation: (a) acceleration from the installed accelerometer; (b) derived acceleration from the proposed method; (c) natural frequency obtained from the measured acceleration; (d) natural frequency obtained from the derived acceleration with the proposed method.

Figure 11(a) and (b) show the acceleration responses collected from the installed accelerometer and calculated from the proposed method, respectively. The natural frequencies of the structural

model can be identified with these two acceleration responses, as shown in Figure 11(c) and (d). Two significant modal frequencies can be identified at 3.02Hz and 23.70Hz, and the results from acceleration responses measured from the conventional accelerometer and calculated from the proposed approach match very well. For the test with a PGA of 0.3g, Figure 12(b) shows the acceleration calculated with the displacement from the proposed method, which has a good agreement with that measured by the installed accelerometer as shown in Figure 12(a). The first two natural frequencies in this test obtained from the measured acceleration are 2.67Hz and 20.58Hz as shown in Figure 12(c), which are very close to the natural frequencies obtained from the proposed approach at 2.71Hz and 21.34Hz respectively as shown in Figure 12(d). These results demonstrate that the proposed vision-based approach can efficiently identify the structural vibration frequencies with the errors of 1.49% and 3.69% for the first and the second modes, respectively, in this case. To explain more clearly about the source of error, the rotations as mentioned above cause the variation in the displacement and acceleration responses measured by the vision-based method and conventional sensors. A closer investigation reveals that with the increase of excitation force intensity, the vision-based method produces the natural frequency with a higher percentage of error than low level excitation. This may be because of the significant rotation introduced by the large ground motions and errors in the numerical calculations from displacement to obtain the acceleration. Nevertheless, the results in this section demonstrate that the proposed method can identify structural natural frequencies accurately compared with the traditional measurements from the accelerometer. The proposed approach has a promising performance in measuring real engineering structural vibrations.

5 Discussions

This paper presents a decent target-free vision-based technique to measure structural vibrations in terms of dynamic displacement, acceleration and natural frequency when the structure is exposed to real-world excitation forces and out-of-plane movements. In spite of favourable results, some challenges need to be considered carefully during the implementation of each involved steps of the proposed method to achieve the precise results. These concerns will be described and discussed as follows.

i. During video acquisition, the captured video image may be polluted by unwanted camera vibrations from human induced vibrations and environmental conditions such as strong winds. This camera vibration can produce error prone displacement results as well as structural modal

identification. To remove the camera vibration associated with the camera operation, additional pre-processing may be required which could be time-consuming. In this study, the remote control is used to operate the camera from a certain distance during the test to overcome certain unwanted camera vibration that might be induced by the manual operations.

ii. Accurate vibration measurement through vision-based techniques mostly depends on the quality of video images, which can be obtained by using adequate camera configuration such as frame rate, resolution, lens focal length, zoom factor and shutter specification. The professional high-speed camera is very useful for structural vibration measurement, yet, it is far expensive. The low-cost consumer-grade commercial camera has a lower frame rate varying from 25 fps to 60 fps. However, in most cases the camera frame rate provided by the manufacturer may not be exactly accurate. For example, a camera shows the specification of a frame rate of 25 fps. However, the test from measured data shows that the actual frame rate is at 24.48 fps, which exposes an error of 2%. In this investigation, camera configuration such as frame rate, resolution, lens focal length, zoom factor, and shutter specifications are confirmed by using video metadata before the proposed vision-based method is implemented.

iii. The video captured by a low frame rate camera will produce temporal aliasing effect, and measured displacement is contaminated with noise and even not accurate in the high frequency range. In the vision-based methods, the camera frame rate should be at least two times of the structural vibration frequency to accurately measure the dynamic displacement as well as to identify the frequency effectively. In addition, the adequate resolution should be maintained along with the higher frame rate as the poor resolution image is not suitable for feature detection, extraction, tracking and matching. It is also noticed that most of the commercial video cameras designed in such a way that the resolution drops significantly while increasing the frame rate. Furthermore, to obtain a clear information from the large area of the structure with different FOV, the camera should have a good lens with a high focal length, zoom factor and shutter specification.

iv. The proposed method requires manual pre-processing of the video clip to extract frame and to select ROI for the detection of key features, which could be time-consuming. Moreover, to store and handle a large volume of video images is also another challenging task in the vision-based vibration measurement, since it requires high configuration hardware that increases the computational cost. In this study, a powerful BRISK algorithm is used to detect

and extract key-features which dramatically reduced computational time. Moreover, sometimes the recorded video images need to be converted to another video format for further processing and to store within a relatively small capacity without losing any information. To get more precise results from the vision-based method, light and illumination, wind, weather condition, and other environmental perspectives need to be considered very carefully as they may have noticeable influences on the results.

v. In most of the previous studies, the position of the camera should be perpendicular to the axis of vibration direction of the structure. The concept is not always applicable in the real-life structures as there may be some obstacles or place is inaccessible due to the geographic point of view. In this situation, further investigation needs to be carried out to verify the capability of feature detection and tracking algorithm for a larger deviation from the perpendicular line of sight in both horizontal and vertical directions. A powerful BRISK feature detection and extraction algorithm used in the present study, which is robust and invariant to rotation, scale, and orientation. However, the camera is positioned 5^0 horizontal and 9^0 vertical angle with the perpendicular line of sight to the vibration axis of the structure from 4 m distance. The results derived from the proposed vision-based method shows very good agreement when compared with the traditional sensor system. For this small variation of the angle from the perpendicular axis of the structure with small camera distance, there is no significant variation of results observed when compared with the traditional sensor technique. However, for small image rotation angle, the BRISK algorithm can successfully detect and extract key-features from images. The present study is greatly focused on the vision-based vibration measurement of the structure which is subjected to different (very low to high) level of bi-directional ground motion and out-of-plane movements while the effect of camera frame rates is also investigated. To investigate the effect of a different angle of image acquisition, a future intensive study has planned to verify the proposed vision-based technique when the image will be captured from 0^0 to $\pm 80^0$ horizontal and vertical angle with the perpendicular line of sight to the vibration axis of the structure along with the variable distance between the camera and target structure.

6 Conclusion

This paper proposes a target-free vision-based approach to measure the dynamic displacement of the structure subjected to out-of-plane movements due to the bi-directional excitations. A targetless vision-based technique is developed by camera calibration, detection and extraction of BRISK features, tracking of features between consecutive frame by KLT algorithm, and detection of false matching along with outlier discarding by using MLESAC algorithm. Currently most vision-based methods for displacement measurement require the target fixed on the structure. Many of the current targetless vision-based researches are conducted on the structures in the laboratory with uni-directional excitation force to measure the vibration response, which is not always realistic. The real-world structures subjected to dynamic loads experience multi-directional out-of-plane movement. The proposed approach is able to accurately extract structural vibration without using any artificial target while the structure is subjected to bi-directional excitations which is more realistic in real situations. The proposed method is validated by conducting a series of experimental tests, and the results are compared with those obtained from the traditional sensory system in terms of the time domain dynamic displacement, maximum displacement, acceleration and natural frequencies. The experimental results show that the proposed method can accurately measure the dynamic displacement and effectively identify the natural frequencies of structures. Moreover, the limitations of the proposed method such as camera frame rate, camera configuration, camera vibration, pre-processing of data and data storage are discussed. The future study may include the further development based on the proposed approach for 3D dynamic displacement measurement and using the measurements for structural condition assessment and damage detection.

Acknowledgement

The work described in this paper was supported by Australian Research Council Laureate Fellowships FL180100196. The first author would like to acknowledge the International Postgraduate Research Scholarship and Strategic International Research Scholarship, provided by Curtin University, Australia.

References

1. Phares, B.M., et al., *Routine highway bridge inspection condition documentation accuracy and reliability.* Journal of Bridge Engineering, 2004. **9**(4): p. 403-413.
2. Moore, M., et al., *Reliability of visual inspection for highway bridges, volume I*. 2001.
3. Sanford, K.L., P. Herabat, and S. McNeil. *Bridge management and inspection data: Leveraging the data and identifying the gaps*. in *Proc., 8th Int. Bridge Management Conf*. 1999.
4. Chase, S. and M. Edwards, *Developing a Tele-Robotic Platform for Bridge Inspection.* 2011.
5. Metni, N. and T. Hamel, *A UAV for bridge inspection: Visual servoing control law with orientation limits.* Automation in construction, 2007. **17**(1): p. 3-10.
6. Lim, R.S., et al. *Developing a crack inspection robot for bridge maintenance*. in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. 2011. IEEE.
7. Sterritt, G., *Review of Bridge Inspection Competence and Training.* Project Report, viewed, 2012. **23**.
8. Kim, J., *System Identification of Civil Engineering Structures through Wireless Structural Monitoring and Subspace System Identification Methods.* 2011.
9. Park, K.-T., et al., *The determination of bridge displacement using measured acceleration.* Engineering Structures, 2005. **27**(3): p. 371-378.
10. Guo, T. and Y.-W. Chen, *Field stress/displacement monitoring and fatigue reliability assessment of retrofitted steel bridge details.* Engineering Failure Analysis, 2011. **18**(1): p. 354-363.
11. Wang, H., et al., *Full-scale measurements and system identification on sutong cable-stayed bridge during typhoon Fung-Wong.* The Scientific World Journal, 2014. **2014**.
12. Feng, D., et al., *A vision-based sensor for noncontact structural displacement measurement.* Sensors, 2015. **15**(7): p. 16557-16575.
13. Li, J., et al., *Development and application of a relative displacement sensor for structural health monitoring of composite bridges.* Structural Control and Health Monitoring, 2015. **22**(4): p. 726-742.
14. Li, J. and H. Hao, *Health monitoring of joint conditions in steel truss bridges with relative displacement sensors.* Measurement, 2016. **88**: p. 360-371.
15. Fukuda, Y., M.Q. Feng, and M. Shinozuka, *Cost-effective vision-based system for monitoring dynamic response of civil engineering structures.* Structural Control and Health Monitoring, 2010. **17**(8): p. 918-936.
16. Park, J.-W., et al., *Vision-based displacement measurement method for high-rise building structures using partitioning approach.* Ndt & E International, 2010. **43**(7): p. 642-647.
17. Zhang, D., et al., *A High-Speed Vision-Based Sensor for Dynamic Vibration Analysis Using Fast Motion Extraction Algorithms.* Sensors, 2016. **16**(4): p. 572.
18. Lee, J., et al., *Computer Vision-Based Structural Displacement Measurement Robust to Light-Induced Image Degradation for In-Service Bridges.* Sensors, 2017. **17**(10): p. 2317.
19. Ribeiro, D., et al., *Non-contact measurement of the dynamic displacement of railway bridges using an advanced video-based system.* Engineering Structures, 2014. **75**: p. 164-180.
20. Yang, Y.S., C.W. Huang, and C.l. Wu, *A simple image-based strain measurement method for measuring the strain fields in an RC-wall experiment.* Earthquake Engineering & Structural Dynamics, 2012. **41**(1): p. 1-17.
21. Luo, L., M.Q. Feng, and Z.Y. Wu, *Robust vision sensor for multi-point displacement monitoring of bridges in the field.* Engineering Structures, 2018. **163**: p. 255-266.
22. Ye, X.W., et al., *A vision-based system for dynamic displacement measurement of long-span bridges : algorithm and verification.* Smart Structures and Systems, 2013. **12**(3-4): p. 363-379.

23. Guo, J., Zhu, C., *Dynamic displacement measurement of large-scale structures based on the Lucas–Kanade template tracking algorithm.* Mechanical Systems and Signal Processing, 2016. **66**: p. 425-436.
24. Dong, C., X. Ye, and T. Jin, *Identification of structural dynamic characteristics based on machine vision technology.* Measurement, 2018. 126: p. 405-416.
25. Sładek, J., et al., *Development of a vision based deflection measurement system and its accuracy assessment.* Measurement, 2013. **46**(3): p. 1237-1249.
26. Ye, X.W., C.Z. Dong, and T. Liu, *Image-based structural dynamic displacement measurement using different multi-object tracking algorithms.* Smart Structures and Systems, 2016. **17**(6): p. 935-956.
27. Zheng, F., et al., *Measuring human-induced vibrations of civil engineering structures via vision-based motion tracking.* Measurement, 2016. **83**: p. 44-56.
28. Ji, Y.F. and C.C. Chang, *Nontarget Image-Based Technique for Small Cable Vibration Measurement.* Journal of Bridge Engineering, 2008. **13**(1): p. 34-42.
29. Yang, Y., et al., *Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification.* Mechanical Systems and Signal Processing, 2017. **85**: p. 567-590.
30. Chen, J.G., et al., *Video Camera–Based Vibration Measurement for Civil Infrastructure Applications.* Journal of Infrastructure Systems, 2016. **23**(3): p. B4016013.
31. Yoon, H., et al., *Target-free approach for vision-based structural system identification using consumer-grade cameras.* Structural Control and Health Monitoring, 2016. **23**(12): p. 1405-1416.
32. Harris, C. and M. Stephens. *A combined corner and edge detector*. in *Alvey vision conference*. 1988. Citeseer.
33. Bartilson, D.T., K.T. Wieghaus, and S. Hurlebaus, *Target-less computer vision for traffic signal structure vibration studies.* Mechanical Systems and Signal Processing, 2015. **60**: p. 571-582.
34. Feng, D. and M.Q. Feng, *Experimental validation of cost-effective vision-based structural health monitoring.* Mechanical Systems and Signal Processing, 2017. **88**: p. 199-211.
35. Feng Maria, Q., et al., *Nontarget Vision Sensor for Remote Measurement of Bridge Dynamic Response.* Journal of Bridge Engineering, 2015. **20**(12): p. 04015023.
36. Cha, Y.-J., J. Chen, and O. Büyüköztürk, *Output-only computer vision based damage detection using phase-based optical flow and unscented Kalman filters.* Engineering Structures, 2017. **132**: p. 300-313.
37. Chen, J.G., et al., *Modal identification of simple structures with high-speed video using motion magnification.* Journal of Sound and Vibration, 2015. **345**: p. 58-71.
38. Hu, Q., et al., *A High-Speed Target-Free Vision-Based Sensor for Bus Rapid Transit Viaduct Vibration Measurements Using CMT and ORB Algorithms.* Sensors, 2017. **17**(6): p. 1305.
39. Choi, I., J. Kim, and D. Kim, *A Target-Less Vision-Based Displacement Sensor Based on Image Convex Hull Optimization for Measuring the Dynamic Response of Building Structures.* Sensors, 2016. **16**(12): p. 2085.
40. Khuc, T. and F.N. Catbas, *Completely contactless structural health monitoring of real-life structures using cameras and computer vision.* Structural Control and Health Monitoring, 2017. **24**(1).
41. Khuc, T. and F.N. Catbas, *Computer vision-based displacement and vibration monitoring without using physical target on structures.* Structure and Infrastructure Engineering, 2017. **13**(4): p. 505-516.
42. Zhang, Z., *A flexible new technique for camera calibration.* IEEE Transactions on pattern analysis and machine intelligence, 2000. **22**.

43. Liu, D., et al. *Robust interactive image segmentation with automatic boundary refinement*. in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. 2010. IEEE.
44. Lowe, D.G., *Distinctive image features from scale-invariant keypoints.* International journal of computer vision, 2004. **60**(2): p. 91-110.
45. Leutenegger, S., M. Chli, and R.Y. Siegwart. *BRISK: Binary robust invariant scalable keypoints*. in *Computer Vision (ICCV), 2011 IEEE International Conference on*. 2011. IEEE.
46. Bay, H., et al., *Speeded-up robust features (SURF).* Computer vision and image understanding, 2008. **110**(3): p. 346-359.
47. Vandergheynst, P., R. Ortiz, and A. Alahi. *Freak: Fast retina keypoint*. in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. 2012. Ieee.
48. Trajković, M. and M. Hedley, *Fast corner detection.* Image and vision computing, 1998. **16**(2): p. 75-87.
49. Tomasi, C. and T. Kanade, *Detection and tracking of point features.* 1991.
50. Fischler, M.A. and R.C. Bolles, *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography.* Communications of the ACM, 1981. **24**(6): p. 381-395.
51. Rousseeuw, P.J., *Least median of squares regression.* Journal of the American statistical association, 1984. **79**(388): p. 871-880.
52. Torr, P.H. and A. Zisserman, *MLESAC: A new robust estimator with application to estimating image geometry.* Computer vision and image understanding, 2000. **78**(1): p. 138-156.