Smart Structures and Systems, Vol. 6, No. 9 (2010) 1025-1039
DOI: http://dx.doi.org/10.12989/sss.2010.6.9.1025

An integrated visual-inertial technique for structural displacement and velocity measurement

C.C. Chang*[a] and X.H. Xiao[b]

Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

(Received August 20, 2009, Accepted May 19, 2010)

Abstract. Measuring displacement response for civil structures is very important for assessing their performance, safety and integrity. Recently, video-based techniques that utilize low-cost high-resolution digital cameras have been developed for such an application. These techniques however have relatively low sampling frequency and the results are usually contaminated with noises. In this study, an integrated visual-inertial measurement method that combines a monocular videogrammetric displacement measurement technique and a collocated accelerometer is proposed for displacement and velocity measurement of civil engineering structures. The monocular videogrammetric technique extracts three-dimensional translation and rotation of a planar target from an image sequence recorded by one camera. The obtained displacement is then fused with acceleration measured from a collocated accelerometer using a multi-rate Kalman filter with smoothing technique. This data fusion not only can improve the accuracy and the frequency bandwidth of displacement measurement but also provide estimate for velocity. The proposed measurement technique is illustrated by a shake table test and a pedestrian bridge test. Results show that the fusion of displacement and acceleration can mitigate their respective limitations and produce more accurate displacement and velocity responses with a broader frequency bandwidth.

Keywords: displacement measurement; videogrammetry; Kalman filter; data fusion.

1. Introduction

Dynamic responses of large-scale civil engineering structures including their displacement/deformation, velocity and acceleration are important information for the health monitoring, safety assessment and vibration control strategy of these structures. The most commonly used sensor for extracting these responses is the accelerometer which can provide accurate acceleration measurement with a broad frequency band. Despite such an advantage, accelerometer suffers from a potential limitation when one intends to obtain velocity and displacement from the measured acceleration. It is well-known that the integration from acceleration to velocity and displacement requires selection of filters and baseline correction that cannot be easily automated (Hudson 1979).

One popular displacement measurement method is the global positioning system (GPS) which was introduced to civil engineering discipline more than a decade ago (Çelebi and Sanli 2002). Psimoulis and

*Corresponding Author, Professor, E-mail: cechang@ust.hk
[a]Professor
[b]Graduate Student

Stiros (2008) explored the possibility of using GPS and robotic total stations (RTS) for measurements of oscillations of relatively rigid structures. Currently, the GPS can provide rather accurate real-time displacement measurement at a sampling frequency up to 20 Hz. Nickitopoulou *et al.* (2006) conducted a series of experiments and found that the standard accuracy of GPS can reach up to 15 mm horizontally and 35 mm vertically at 1.5% outlier level for low vibration frequency. Psimouli *et al*. (2008) reported that GPS is suitable for the identification of dynamic characteristics of relatively rigid civil engineering structures with modal frequencies up to 4 Hz if displacements are above the uncertainty level (≥5 mm). The accuracy of GPS however, depends greatly on many factors such as vibration frequency, satellite coverage, atmospheric effect, multipath effect and data post-processing methods (Roberts *et al*. 2004).

As a results of rapid advancement in electronics and computer vision technology, commercial digital cameras are now equipped with high pixel resolution at a reasonable cost. More and more videogrammetric techniques have been developed using these digital cameras for displacement measurement of civil structures (Olaszek 1999, Patsias and Staszewski 2002, Chang and Ji 2007, Ji and Chang 2008a,b). Recently, Caetano *et al.* (2007) reported an application of a vision-based technique for a target-camera distance of 850 m and a field of view of 300 m. They were able to measure cable vibration with amplitudes in the order of 0.1 m. These video-based techniques are non-contact in nature and can be performed remotely at a distance away from the structure to be measured. These techniques are flexible and do not require complicated setup or calibration. Note that the use of commercial cameras although reduces the cost of the techniques, they however can only record images at a relatively low frequency, typically at 30 Hz. It is also found that the measured results are contaminated with noise which can be attributed to errors associates with corner detection and camera calibration. One possible remedy to improve the accuracy of displacement measurement is to fuse the measured displacement with acceleration. By so doing, the redundancy between each other can be explored to overcome respective disadvantages and lead to displacement and velocity with improved accuracy. This concept has been implemented using the GPS-based displacement measurement recently (Roberts *et al.* 2004, Li *et al.* 2006, Chan *et al.* 2006). It is also noted that multi-sensor systems that combine direct positioning and imaging sensors are rapidly becoming a standard source of information for various aerial and land-based geolocation mapping applications including orthophoto production, feature extraction/vector mapping and surface reconstruction (Grejner-Brzezinska *et al.* 2008).

Some examples, of the state-of-the-art applications presented here, illustrate well the ability of multi-sensor systems as important mapping/GIS tools, enabling the much-desired automation of photogrammetric data collection and interpretation. The aspect of real-time mobile mapping/mobile computing, based on GPS/INS, automatic image processing and telecommunication networks were indicated as the newest trends in MMS technology development.

In this study, an integrated visual-inertial sensing method that combines a monocular videogrammetric displacement measurement technique and a collocated accelerometer is proposed for displacement and velocity measurement of civil engineering structures. The monocular videogrammetric technique extracts three-dimensional (3D) translation and rotation of a planar target from an image sequence recorded by one camera. A multi-rate Kalman filtering technique (Lewis *et al.* 2008, Smyth and Wu 2007) is then adopted to integrate the low-frequency displacement with the high-frequency acceleration. A shake table test and a pedestrian bridge test are conducted to validate the feasibility of the proposed measurement technique.

2. Monocular videogrammetric technique

The proposed monocular videogrammetric technique requires only one camera, one planar target and one planar calibration board. The procedure includes extraction of camera parameters and extraction of translation and rotation response which will be briefly discussed in the following.

2.1 Extraction of camera parameters

A commercial-grade camera needs to be calibrated preferably in-situ to obtain its intrinsic and extrinsic parameters. A pinhole model that describes a general projective camera can be expressed as (Hartley and Zisserman 2003)

$$\lambda \boldsymbol{m} = \boldsymbol{K}[\boldsymbol{R} \quad \boldsymbol{T}]\boldsymbol{M} \tag{1}$$

where $m = (u, v, 1)^T$ is the homogenous vector of a two-dimensional (2D) point in the image coordinate uv system defined in pixels; $\boldsymbol{M} = (X, Y, Z, 1)^T$ is the homogenous vector of a 3D real point in the world coordinate XYZ system; and λ is a scale factor. The 3×3 camera intrinsic matrix $\boldsymbol{K}$ contains five camera intrinsic parameters including two focal lengths, one skew parameter and two coordinates of the principal point. The 3×3 rotation matrix $\boldsymbol{R}$ and the 3×1 translation vector $\boldsymbol{T}$ represent the orientation and the translation between the camera coordinate system and the world coordinate system, respectively. In total, there are eleven parameters (five intrinsic and six extrinsic parameters) which can be determined by providing at least six pairs of correspondence between non-planar 3D points and their 2D images. In this study, the plane pattern used by Chang and Ji (2007) is adopted for camera calibration. This plane calibration pattern is positioned around the location where response is to be measured and is shown at a few arbitrary orientations. To determine coordinates of points from the acquired images, an improved Harris corner detector (Harris and Stephens 1988) is adopted. The algorithm is able to achieve sub-pixel accuracy yet is invariant to rotation, scale, illumination variation and image noise (Schmid *et al*. 2000).

Assume that the planar target is attached to the location where response is to be measured. Let the world coordinate XYZ system be attached to one corner of the planar target. The world coordinates of a point on the planar target can be expressed as $\boldsymbol{M} = (X, Y, 0, 1)^T$. Using this expression in (1) gives

$$\lambda \boldsymbol{m} = \boldsymbol{K}[\boldsymbol{r}_1 \ \boldsymbol{r}_2 \ \boldsymbol{T}]\begin{Bmatrix} X \\ Y \\ 1 \end{Bmatrix} = \boldsymbol{H}\begin{Bmatrix} X \\ Y \\ 1 \end{Bmatrix} \tag{2}$$

where $\boldsymbol{H}$ is a 3×3 planar homographic matrix that transforms points on a world plane to the image plane; $\mathbf{r}_1$ and $\mathbf{r}_2$ are the 1st and the 2nd column of the rotation matrix $\boldsymbol{R}$, respectively. The planar homographic matrix $\boldsymbol{H}$ contains eight unknown parameters. If the coordinates of any four grid points on the planar target are known, $\boldsymbol{H}$ can be estimated using the direct linear transformation method. When the coordinates of more than four grid points are known, $\boldsymbol{H}$ can be estimated by minimizing the reprojection error as follow

$$\boldsymbol{H} = \arg\min_{\boldsymbol{H}} \sum_i \left\| \boldsymbol{H}\boldsymbol{M}_i - \lambda \boldsymbol{m}_i \right\|^2 \tag{3}$$

where the subscript i indicates the number of grid points. Given that the intrinsic matrix $\boldsymbol{K}$ is available from calibration, the rotation vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ and translation vector $\boldsymbol{T}$ can be obtained as

$$[\boldsymbol{r}_1\ \boldsymbol{r}_2\ \boldsymbol{T}] = \boldsymbol{K}^{-1}\boldsymbol{H} \tag{4}$$

As the 3rd column of the rotation matrix is orthogonal to the first two columns, the rotation matrix $\boldsymbol{R}$ can hence be obtained. Note that if the number of grid points on the planar target is sufficient (more than six points when lens distortion is not considered or more than eight points when lens distortion is considered), a separate camera calibration process is not necessary. The five intrinsic parameters of the camera and the six extrinsic parameters of the planar target for each image can be obtained simultaneously. Lepetit and Fua (2005) commented that simultaneous estimation of both intrinsic and extrinsic parameters strongly depends on the geometry and the number of correspondences. To obtain more reliable 3D tracking results, it is always more preferable to use a calibrated camera and estimate only the extrinsic parameters from the image sequence.

2.2 Extraction of rotation and translation responses

For clarity, the first and a subsequent image in an image sequence, denoted as image 1 and image i, are used to illustrate the extraction of rotation and translation response between the two images. Two corresponding world coordinate systems, world coordinate 1 and world coordinate i, are attached to one corner point of the target on each of the two images respectively (see Fig. 1). Assume that the rotation matrixes and the translation vectors of these two images obtained using the above method are represented as $\boldsymbol{R}_1$, $\boldsymbol{T}_1$ and $\boldsymbol{R}_i$, $\boldsymbol{T}_i$, respectively. Denote the camera coordinates of an arbitrary point N as $\boldsymbol{N}_c$ and its coordinates in the world coordinate system 1 and i as $\boldsymbol{N}_1$ and $\boldsymbol{N}_i$, respectively. These coordinates are related via the rotation matrixes and translation vectors as (Hartley and Zisserman 2003)

$$\boldsymbol{N}_c = \boldsymbol{R}_1\boldsymbol{N}_1 + \boldsymbol{T}_1 \tag{5}$$

$$\boldsymbol{N}_c = \boldsymbol{R}_i\boldsymbol{N}_i + \boldsymbol{T}_i \tag{6}$$

Combining these two equations gives

$$\boldsymbol{N}_i = \boldsymbol{R}_{1-i}\boldsymbol{N}_1 + \boldsymbol{T}_{1-i} \tag{7}$$

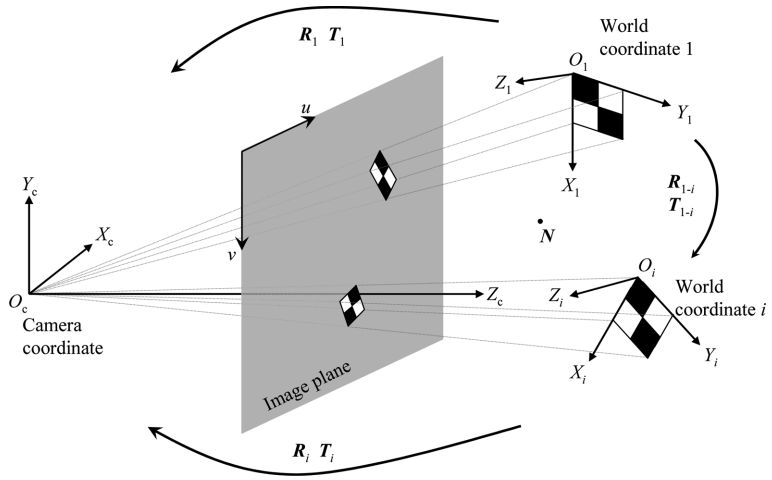


Fig. 1 Transformation between the camera coordinate system and the world coordinate systems

$$\boldsymbol{R}_{1-i} = \boldsymbol{R}_i^{-1}\boldsymbol{R}_1 \quad \boldsymbol{T}_{1-i} = \boldsymbol{R}_i^{-1}(\boldsymbol{T}_1 - \boldsymbol{T}_i) \tag{8}$$

where $\boldsymbol{R}_{1\text{-}i}$ and $\boldsymbol{T}_{1\text{-}i}$ represent the rotation matrix and the translation vector between the world coordinate system 1 and the world coordinate system i, respectively. As shown, the monocular videogrammetric technique can measure 3D translation and rotation of the world coordinate system defined on the planar target. While the determination of displacements from the translation vector is apparent, the relation between the rotation matrix and the rotation angles is not as straightforward. Further discussion on this can be seen in Chang and Xiao (2010).

3. Multi-rate Kalman filter with smoothing

Assume that in addition to the displacement response from the monocular videogrammetric technique, collocated acceleration response at the target location is also available. A multi-rate Kalman filter with smoothing (Smyth and Wu 2007, Lewis *et al*. 2008) can be used to fuse the two types of responses. For simplicity, the following derivation considers only one dimensional displacement and acceleration. Assume that the measured acceleration (denoted as $\ddot{x}^m$) and displacement (denoted as x^m) are sampled at a time interval of T_a and T_d, respectively. Usually, the acceleration can be obtained at a smaller sampling interval, i.e., $T_a<T_d$. For simplicity, it is further assumed that the time ratio T_d/T_a is an integer. Considering the measured acceleration as an input and the displacement obtained from the monocular videogrammetric technique as observed output, the discrete-time state estimates with observer for the collocated acceleration measurement and displacement measurement can be expressed as

$$\boldsymbol{X}_k = \boldsymbol{A}\boldsymbol{X}_{k-1} + \boldsymbol{B}\ddot{x}_{k-1}^m + \boldsymbol{B}w_{k-1} \tag{9}$$

$$x_k^m = \boldsymbol{C}\boldsymbol{X}_{k-1} + v_{k-1} \tag{10}$$

where

$$\boldsymbol{A} = \begin{bmatrix} 1 & T_a \\ 0 & 1 \end{bmatrix}; \quad \boldsymbol{B} = \begin{bmatrix} T_a^2/2 \\ T_a \end{bmatrix}; \quad \boldsymbol{C} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{11}$$

Note that the state vector $\mathbf{X}_k = [x_k \ \dot{x}_k]^T$ and k is the time step when acceleration and displacement are measured. For this discrete system, the measurement noises of acceleration and displacement, denoted as w and v, are assumed to be independent Gaussian white noises with variances of q and r, respectively. The Kalman filter estimates a process using a form of feedback control: the filter estimates the process state at some time and obtains feedback in the form of measurements. As such, the equations for the Kalman filter are separated into two groups: time update and measurement update. The time update equations project forward the state and error covariance to obtain *a priori* estimates for the next time step

$$\hat{\boldsymbol{X}}_k^- = \boldsymbol{A}\hat{\boldsymbol{X}}_{k-1} + \boldsymbol{B}\ddot{x}_{k-1}^m \tag{12}$$

$$\boldsymbol{P}_k^- = \boldsymbol{A}\boldsymbol{P}_{k-1}\boldsymbol{A}^T + \boldsymbol{Q} \tag{13}$$

where $\hat{\boldsymbol{X}}_{\boldsymbol{k}}^{-}$ and $\hat{\boldsymbol{X}}_{\boldsymbol{k-1}}$ are the *a priori* and *a posteriori* state estimates, respectively; and $\boldsymbol{P}_{\boldsymbol{k}}^{-}$ and $\boldsymbol{P}_{\boldsymbol{k-1}}$ are the *a priori* and *a posteriori* error covariance estimates, respectively. The covariance matrix $\boldsymbol{Q}$ is

$$\boldsymbol{Q} = q\begin{bmatrix} T_a^4/4 & T_a^3/3 \\ T_a^3/3 & T_a^2/2 \end{bmatrix} \tag{14}$$

The projected state and error covariance estimates are then updated by measurements as

$$\hat{\boldsymbol{X}}_{\boldsymbol{k}} = \hat{\boldsymbol{X}}_{\boldsymbol{k}}^{-} + \boldsymbol{G}_k(x_k^m - \boldsymbol{C}\hat{\boldsymbol{X}}_k^{-}) \tag{15}$$

$$\boldsymbol{P}_k = (1 - \boldsymbol{G}_k\boldsymbol{C})\boldsymbol{P}_k^{-} \tag{16}$$

where the Kalman gain $\boldsymbol{G}$ is

$$\boldsymbol{G}_k = \boldsymbol{P}_k^{-}\boldsymbol{C}^T(\boldsymbol{C}\boldsymbol{P}_k^{-}\boldsymbol{C}^T + r)^{-1} \tag{17}$$

As the displacement is measured at a much slower rate than the acceleration, only time updating is performed between any two consecutive displacement measurement k-1 and k

$$\hat{\boldsymbol{X}}_k = \hat{\boldsymbol{X}}_k^{-} = \boldsymbol{A}\hat{\boldsymbol{X}}_{k-1} + \boldsymbol{B}\ddot{x}_{k-1}^m \tag{18}$$

$$\boldsymbol{P}_k = \boldsymbol{P}_{\boldsymbol{k}}^{-} = \boldsymbol{A}\boldsymbol{P}_{\boldsymbol{k-1}}\boldsymbol{A}^{\boldsymbol{T}} + \boldsymbol{Q} \tag{19}$$

At time kT_d, both acceleration and displacement are available, hence both time and measurement updates should be performed to obtain the optimal result.

To improve the accuracy of estimation and eliminate the discontinuity caused by the multi-rate Kalman filter, a smoothing technique can be incorporated with the multi-rate Kalman filter. There are three types of smoothing: fixed-point, fixed-interval and fixed-lag smoothing. In this paper, a fixed-interval smoothing which combines the forward Kalman filter and backward filter over the entire time period is adopted. The smoothed estimates $\hat{\boldsymbol{X}}_{k|S}$ for the whole time period can be obtained by

$$\hat{\boldsymbol{X}}_{k|S} = \hat{\boldsymbol{X}}_k + \boldsymbol{F}_k(\hat{\boldsymbol{X}}_{k+1|S} - \hat{\boldsymbol{X}}_{k+1}^{-}) \tag{20}$$

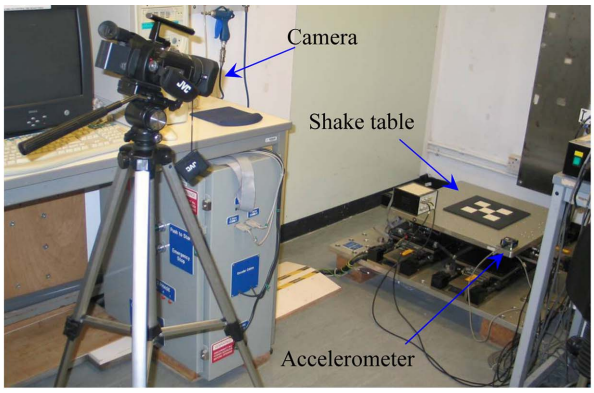
The smoothing gain $\boldsymbol{F}_k$ is given by

$$\boldsymbol{F}_k = \boldsymbol{P}_k\boldsymbol{A}^T(P_{k+1}^{-})^{-1} \tag{21}$$

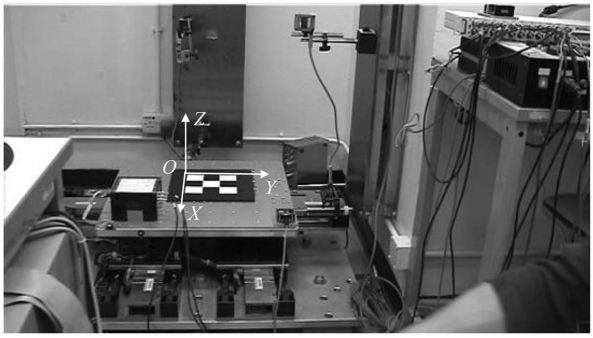
with $k = S$-1, S-2, …, 0 where S is the total number of data point.

4. Experimental demonstration

To validate the accuracy and feasibility of the proposed visual-inertial measurement technique, two tests were conducted. The camera used to capture video had three 1/3 inch 1.18 million-pixel



(a)



(b)

Fig. 2 (a) Setup of shake table test and (b) first image recorded and the defined world coordinate system

progressive charge coupled devices and could record 1,280×720 pixels of images at 30 frames per second. It could provide 10× optical and 200× digital zoom with a focal length ranging from 5.2 mm to 52 mm. A force-balancing accelerometer with a measurement range of ±1 g was used to acquire acceleration. The sampling frequency of the accelerometer was set at 300 Hz for both tests.

4.1 Shake table test

A planar target with nine 60 mm×60 mm black and white squares was attached on a shake table as shown in Fig. 2(a). An accelerometer was also attached to the shake table to obtain the acceleration of the shake table during the test. The camera was placed at about 2.5 meters away and about 1.5 meters high from the shake table. The precise location of the camera was not important as the camera parameters would be determined through on-site calibration. The focal length of camera was set at 5.2 mm. The shake table was programmed to move along the X-axis as shown in Fig. 2(b). Prior to the test, a calibration board used in Chang and Ji (2007) was placed around the planar target. A few photos were taken with this board arbitrarily positioned and were used for calibrating the camera. The shake table was programmed to follow the prescribed displacement time history below for 8 seconds

$$x(t) = 10\sin(2\pi t) + 4\sin(8\pi t) + 1\sin(18\pi t) + 0.2\sin(32\pi t) + 0.3t \text{ (in mm)} \quad (22)$$

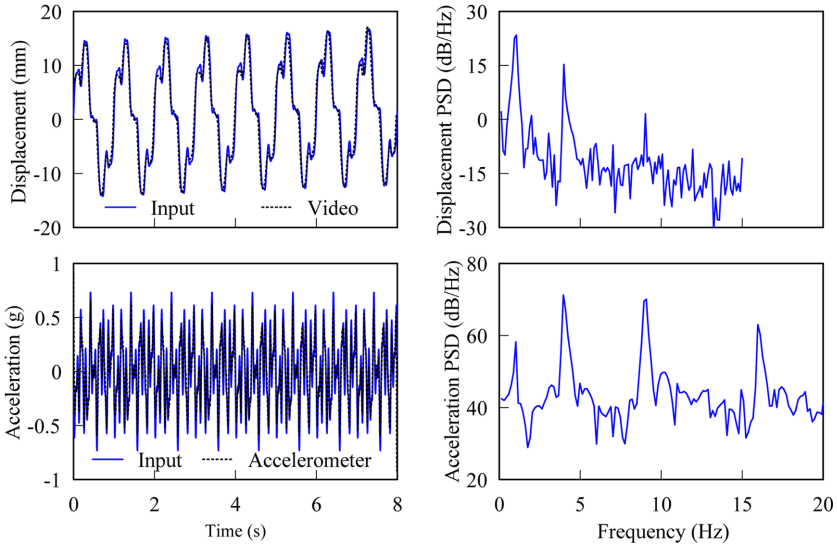


Fig. 3 Measured displacement and acceleration time histories of point O in Fig. 2(b) and their power spectral densities (PSD's) (Video: the monocular videogrammetric technique)

This displacement time history contained four frequencies: 1, 4, 9 and 16 Hz as well as a linear drift. When the shake table stopped after 8 seconds, a displacement drift of 2.4 mm remained. For the Kalman filter fusion, the noise variances of measured acceleration and video displacement were calculated to be 9.97×10^{-5} g^2 and 2 mm^2, respectively. These values were obtained from the image sequence and the accelerometer readings recorded prior to the activation of the shake table.

Fig. 3 shows the displacement time history obtained from the monocular videogrammetric technique (denoted as "video") and the acceleration time history measured by the accelerometer. Also shown are their power spectral densities (PSD's). It is seen that both the measured displacement and acceleration can accurately track the movement of the shake table. As the sampling rate of camera was fixed at 30 Hz, the monocular videogrammetric technique can not capture the last frequency component (16 Hz). On the other hand, the PSD of measured acceleration shows that the four frequencies can be captured distinctively.

Fig. 4 compares the displacement time histories obtained from the monocular videogrammetric technique, the fusion of acceleration and videogrammetric displacement (denoted as "fused"), and the double integration of acceleration (denoted as "integrated"). Also shown are the errors between these displacement time histories and the input displacement time history of the shake table. It is seen that errors of the video displacement and the fused displacement both fluctuate around zero which indicates that the linear drift can be captured by both displacements. For this particular test, the standard deviations of errors for the video displacement and the fused displacement are computed to be 0.75 mm and 0.35 mm, respectively. This indicates that the fusing of video displacement and acceleration could reduce the displacement error by about 50%. It is also seen that the double integration of acceleration produces the largest error when compared to the input displacement. The error seems to drift in time which confirms that the linear trend of the displacement could not be reflected in the acceleration time history recorded by the accelerometer. Hence as expected, the displacement obtained from the double integration of measured acceleration could not track the non-zero displacement drift.

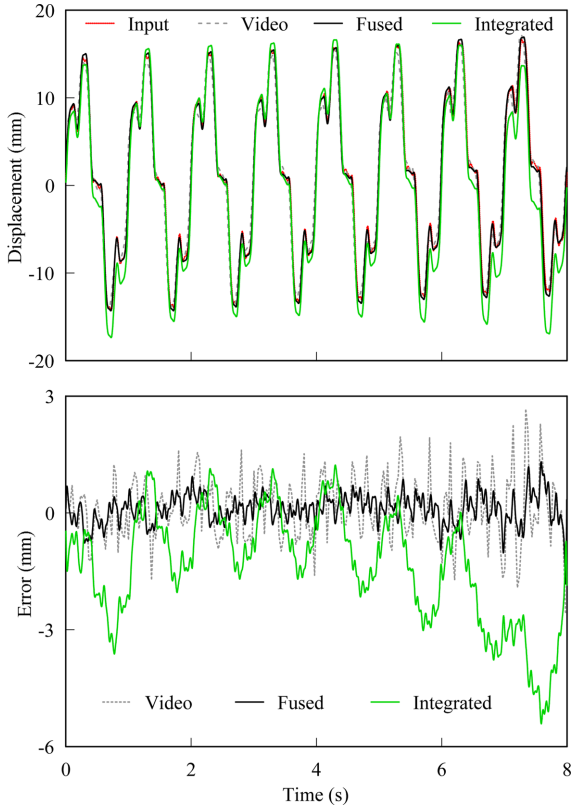


Fig. 4 Comparison of displacement time histories obtained from the videogrammetric technique (denoted as "video"), the fusion of acceleration and videogrammetric displacement (denoted as "fused"), and the double integration of acceleration (denoted as "integrated")

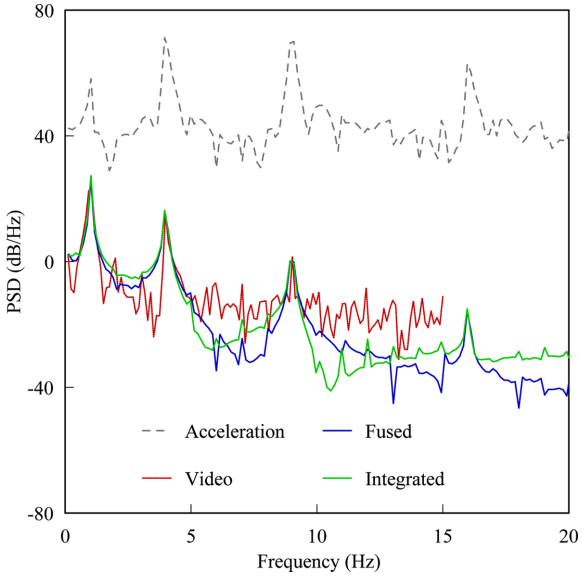


Fig. 5 PSD's of measured acceleration and three displacement time histories shown in Fig. 4

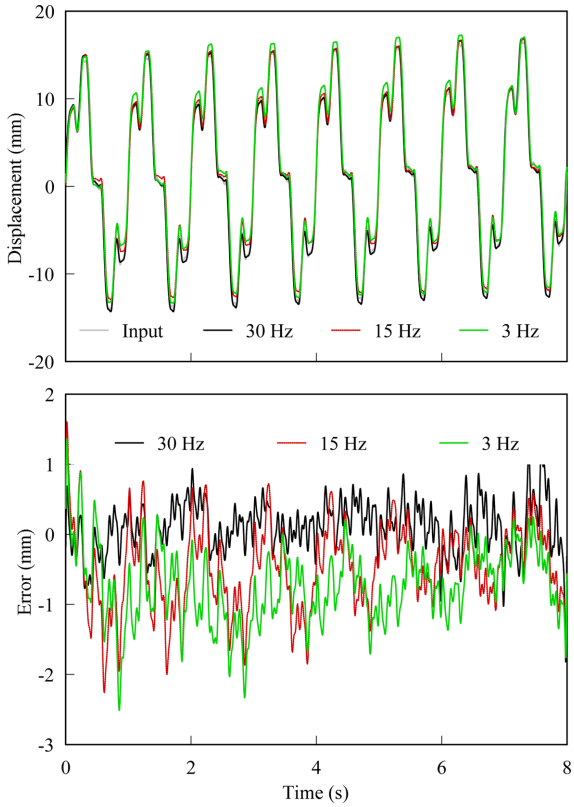


Fig. 6 The effect of reducing sampling frequencies of the video displacement on the fused displacement

The PSD's of the video, fused and integrated displacements are shown in Fig. 5 together with that of the measured acceleration. It is seen that the video displacement although can capture three lower frequencies but seems to be contaminated with noise especially for frequencies higher than 5 Hz. The fused displacement on the other hand not only can capture all four frequencies but also can lower the noise at the high frequency range. These results verified that the fusing of acceleration and displacement measured from the monocular videogrammetric technique can improve the accuracy of displacement measurement while also broaden its frequency bandwidth.

To further study the advantage of Kalma filter data fusion, the video displacement time history was down-sampled at 15 Hz and 3 Hz, respectively. The down-sampled displacement time histories were then fused with the acceleration time history (recorded at 300 Hz). Fig. 6 shows the comparison of fused displacement time histories. It is seen that the error of displacement measurement increases as the sampling frequency of the video displacement reduces. Fig. 7 further shows the PSD's of the three fused displacement time histories. It is seen that although the three fused displacement time histories are different; their PSD's actually match quite well. All three of them show that the highest frequency of 16 Hz which cannot be obtained from the video displacement can now be captured by all three fused displacements.

One side product of the Kalman filter fusion is the velocity which can be obtained simultaneously with the fused displacement. Fig. 8 shows the velocity time histories obtained from fusing the measured acceleration with the video displacement sampled at 30, 15 and 3 Hz, respectively. It is seen

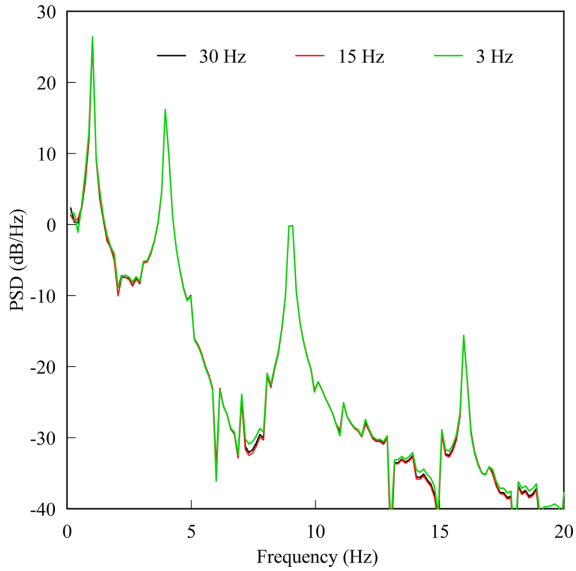


Fig. 7 PSDs of the fused displacement time histories shown in Fig. 6

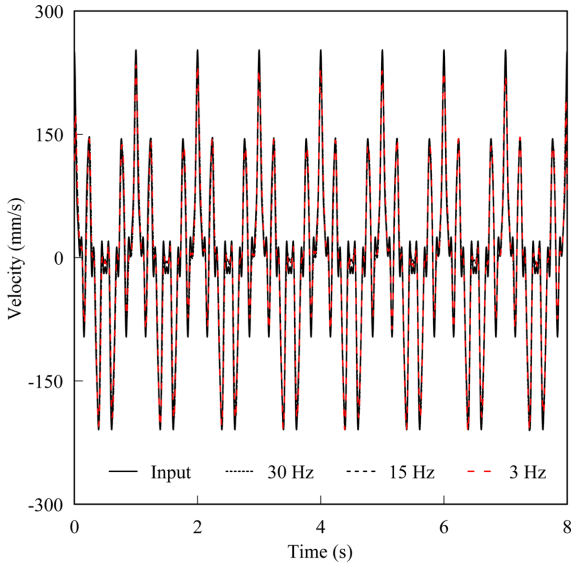


Fig. 8 Velocity time histories obtained from fusing video displacements sampled at 30, 15 and 3 Hz, respectively with acceleration

that all three velocity time histories match quite well. This indicates that reducing the videogrammetric sampling frequency does not seem to affect the accuracy of velocity obtained from the Kalman filter fusion.

4.2 Pedestrian bridge test

The single-tower pedestrian bridge has an overall length of about 20 m and contains one main span and one side span as shown in Fig. 9. The planar target used in the previous test was attached to one side of the bridge deck at about one third location from the left side of the main span (see

Fig. 9 A single-tower pedestrian bridge

(a) (b)

Fig. 10 (a) Test setup and (b) first image recorded by the camera and an enlarged image showing the planar target with a corner-fixed world coordinate system

Fig. 10(a)). The camera was placed at about 6 m horizontally away from the target with an inclined angle of about 30°. The focal length of camera was set at 5.2 mm. Fig. 10(b) shows the first image recorded by the camera as well as an enlarged view of the target with the defined world coordinate system. It was anticipated that the bridge deck would vibrate predominately along the X-axis (vertical) direction. The accelerometer was placed on the main span right on top of the target and was aligned to measure vertical acceleration of the bridge deck. The bridge was excited manually by three people jumping on the middle of the main span for about 10 seconds. Prior to the excitation, the accelerometer and the camera were turned on and recorded a short duration of acceleration and an image sequence, respectively. The information was processed for estimating the noise variances of video displacement and acceleration which were found to be 1 mm^2 and 1.34×10^{-6} g^2, respectively. These values were used in the fusing of acceleration and video displacement using the Kalman filter.

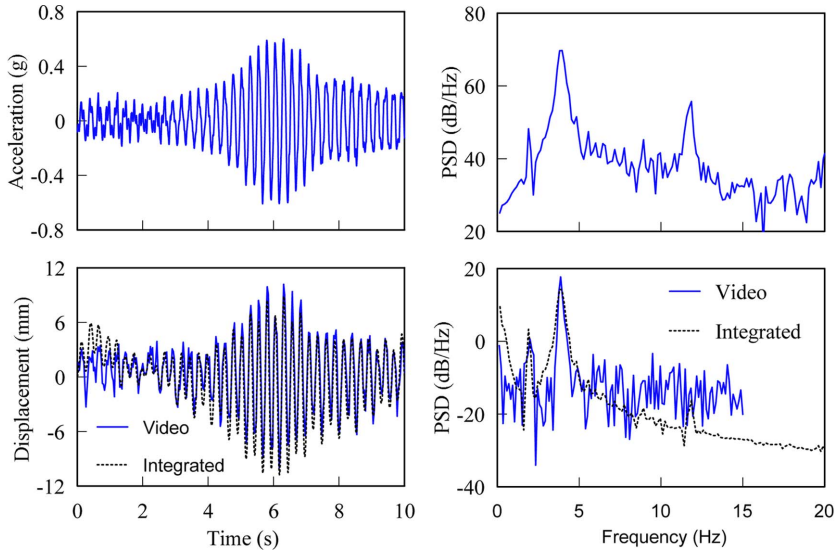


Fig. 11 Measured acceleration time history and displacement time histories of point O in Fig. 10(b) and their PSD's

Fig. 11 shows the displacement time history obtained from the monocular videogrammetric technique (denoted as "video") and the acceleration time history measured by the accelerometer. Also shown are their power spectral densities (PSD's). It is seen from the acceleration PSD that, between the frequency range of 0 to 20 Hz, three natural frequencies of the bridge can be identified: 1.98, 3.88 and 11.72 Hz. The PSD of the video displacement on the other hand can only detect the lowest two natural frequencies. It appears that the video displacement is contaminated with measurement noise beyond 5 Hz frequency; hence the third natural frequency 11.72 Hz cannot be identified. For comparison, the displacement time history obtained from the double integration of measured acceleration and its PSD were also plotted in Fig. 11. It is seen that all three natural frequencies can be identified from the PSD of the integrated displacement. The integrated displacement appears to contain less measurement noise comparing to the video displacement. The PSD of the integrated displacement however shows that it contains low frequency drift; hence the integrated displacement might not be accurate.

Next, fusion of measured acceleration and video displacement was performed using the Kalman filter. The fused displacement and its PSD are shown in Fig. 12. To see whether reducing sampling frequency for the monocular videogrammetric technique would affect the accuracy of acceleration-displacement fusion, the video displacement time history was also down-sampled at 15 and 3 Hz prior to the fusion. Results shown in Fig. 12 indicate that reducing the sampling frequency for the monocular videogrammetric technique does not significantly affect the displacement obtained from fusing the acceleration and the video displacement. Their PSD plots also indicate that the reduced sampling of the monocular videogrammetric technique only has a slight effect in the low frequency region and leads to a very mild drift in the fused displacement time history. These results suggest that a lower sampling frequency can be used for the monocular videogrammetric technique without sacrificing the measurement accuracy if acceleration is available for fusion.

Fig. 13 shows the velocity time histories obtained from fusing the measured acceleration with the video displacement sampled at 30, 15 and 3 Hz, respectively. Also shown in Fig. 13 are their corresponding PSD's. It is seen that all three velocity time histories as well as their PSD's match quite well. This indicates that reducing the videogrammetric sampling frequency does not affect the accuracy of velocity obtained from the Kalman filter fusion.

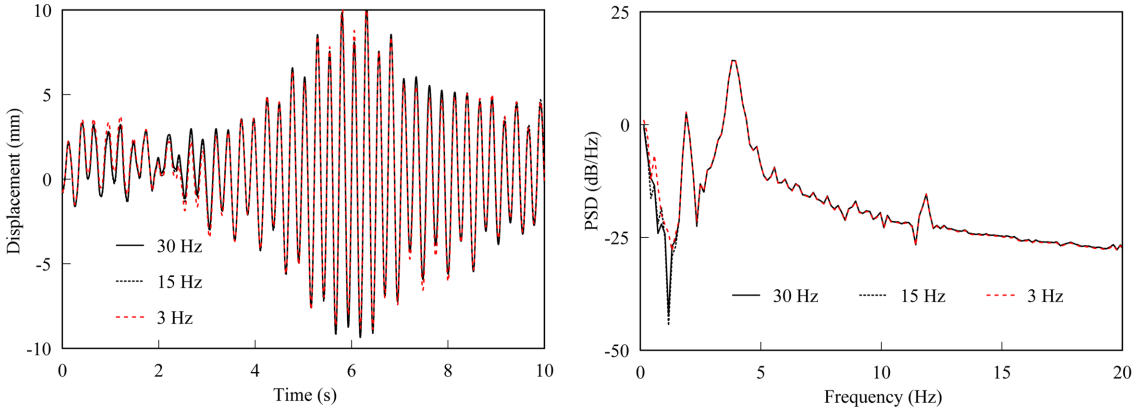


Fig. 12 Fused displacement time histories (with video displacement sampled at 30, 15 and 3 Hz, respectively) and their PSD's

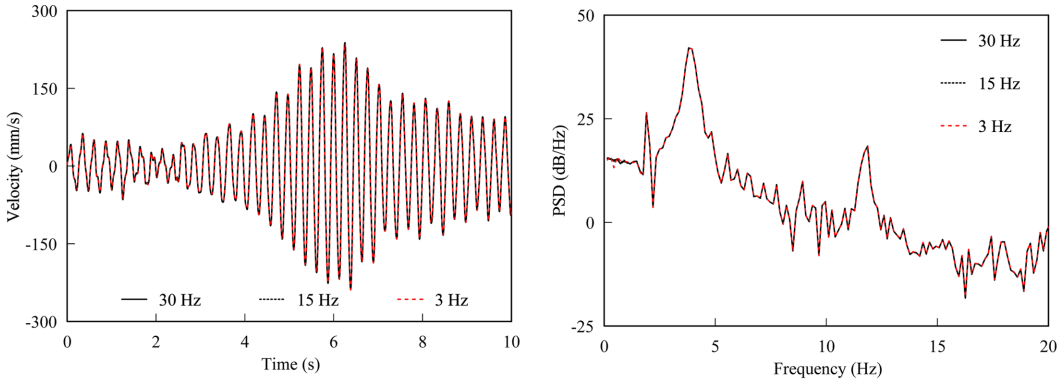


Fig. 13 Fused velocity time histories (with video displacement sampled at 30, 15 and 3 Hz, respectively) and their PSD's

5. Conclusions

With the rapid development in optical, electronics and computer technology, commercial digital cameras are now equipped with very high image resolution at a low cost. These cameras provide a good alternative image-based approach for measuring displacement response of flexible civil engineering structures in a remote non-contact fashion. Despite their low cost and high flexibility, these image-based techniques however suffer from low sampling rate and high measurement noise. In this study, an integrated visual-inertial method that combined a monocular videogrammetric displacement measurement technique and a collocated accelerometer was proposed for displacement and velocity measurement of civil engineering structures. The monocular videogrammetric technique could extract three-dimensional translation and rotation of a planar target from an image sequence recorded by one camera. The obtained displacement was then fused with acceleration obtained from a collocated accelerometer using a multi-rate Kalman filter with smoothing technique. The proposed measurement technique was illustrated by a shake table test and a pedestrian bridge test. Results show that the displacement measured from the monocular videogrammetric technique although can track the linear drift of the prescribed movement of the shake table but has a fairly narrow frequency bandwidth and is contaminated with noise. After fusing with acceleration, it is seen that both the accuracy and frequency bandwidth of the measured displacement can be significantly improved. In addition, the Kalman filter also could simultaneously provide accurate velocity estimates when fusing the displacement with acceleration measurement. The effect of reducing sampling frequency for the monocular videogrammetric technique was also studied. Results show that the accuracy of fused displacements is mildly affected when the sampling frequency is reduced from 30 to 3 Hz. It is seen however that the power spectral densities of the fused displacements remained pretty much the same despite that the sampling frequency is reduced to one tenth of its regular value. It is seen that the proposed visual-inertia technique is a viable and effective alternative for measuring displacement and velocity of civil engineering structures.

Acknowledgments

This study is supported by the Hong Kong Research Grants Council Competitive Earmarked Research Grant 611409.

References

Caetano, E., Silva, S. and Bateira, J. (2007), "Application of a vision system to the monitoring of cable structures", *Proceedings of the 7th International Conference on Cable Dynamics*, Wien, Austria.

Çelebi, M. and Sanli, A. (2002), "GPS in pioneering dynamic monitoring of long-period structures", *Earthq. Spectra*, **18**(1), 47-61.

Chan, W.S., Xu, Y.L., Ding, X.L. and Dai, W.J. (2006), "An integrated GPS-accelerometer data processing technique for structural deformation monitoring", *J. Geodesy*, **80**(12), 705-719.

Chang, C.C. and Ji, Y.F. (2007), "Flexible videogrammetric technique for three-dimensional structural vibration measurement", *J. Eng. Mech.-ASCE*, **133**(6), 656-664.

Chang, C.C. and Xiao, X.H. (2010), "Three-dimensional structural translation and rotation measurement using monocular videogrammetry", *J. Eng. Mech.-ASCE*, **136**(7), 840-848.

Grejner-Brzezinska, D.A., Toth, C., Sun, H., Wang, X. and Rizos, C. (2008), "GPS/INS/PL/TLS integration supporting navigation of geophysical sensors for unexploded ordnance detection and discrimination", *Proceedings of the 13th FIG International Symposium on Deformation Measurements and Analysis, 4th IAG Symposium on Geodesy for Geotechnical and Structural Engineering*, Lisbon, Portugal, May.

Harris, C. and Stephens, M. (1988), "A combined corner and edge detector", *Proceedings of The Fourth Alvey Vision Conference (AVC '88)*, Manchester, UK, August-September.

Hartley, R. and Zisserman, A. (2003), *Multiple view geometry in computer vision,* 2nd Ed., Cambridge University Press, UK.

Hudson, D.E. (1979), *Reading and interpreting strong motion accelerograms*, Earthquake Engineering Research Institute, Berkeley, Calif.

Ji, Y.F. and Chang, C.C. (2008a), "Nontarget image-based technique for small cable vibration measurement", *J. Bridge Eng.*, **13**(1), 34-42.

Ji, Y.F. and Chang, C.C. (2008b), "Nontarget stereo vision technique for spatiotemporal response measurement of line-like structures", *J. Eng. Mech.-ASCE*, **134**(6), 466-474.

Lepetit, V. and Fua, P. (2005), "Monocular model-based 3D tracking of rigid objects: a survey", *Found Trends Comput. Graph. Vis.*, **1**(1), 1-89.

Lewis, F.L., Xie, L. and Popa, D. (2008), *Optimal and robust estimation: with an introduction to stochastic control theory*, 2nd Ed., CRC Press, Boca Raton, FL, USA.

Li, X.J., Ge, L.L., Ambikairajah, E., Rizos, C., Tamura, Y. and Yoshida, A. (2006), "Full-scale structural monitoring using an integrated GPS and accelerometer system", *GPS Solut.*, **10**(4), 233-247.

Nickitopoulou, A., Protopsalti, K. and Stiros, S. (2006), "Monitoring dynamic and quasi-static deformations of large flexible engineering structures with GPS: Accuracy, limitations and promises", *Eng. Struct.*, **28**(10), 1471-1482.

Olaszek, P. (1999), "Investigation of the dynamic characteristic of bridge structures using a computer vision method", *Measurement*, **25**(3), 227-236.

Patsias, S. and Staszewski, W.J. (2002), "Damage detection using optical measurements and wavelets", *Struct. Health Monit.*, **1**(1), 5-22.

Psimoulis, P.A., Pytharouli, S., Karambalis, D. and Stiros, S.C. (2008), "Potential of global positioning system(GPS) to measure frequencies of oscillations of engineering structures", *J. Sound Vib.*, **318**(3), 606-623.

Psimoulis, P.A. and Stiros, S.C. (2008), "Experimental assessment of the accuracy of GPS and RTS for the determination of the parameters of oscillation of major structures", *Comput.-aided Civ. Inf.*, **23**(5), 389-403.

Roberts, G.W., Meng, X.L. and Dodson, A.H. (2004), "Integrating a global positioning system and accelerometers to monitor the deflection of bridges", *J. Surv. Eng.-ASCE*, **130**(2), 65-72.

Schmid, C., Mohr, R. and Bauckhage, C. (2000), "Evaluation of interest point detectors", *Int. J. Comput. Vision*, **37**(2), 151-172.

Smyth, A. and Wu, M.L. (2007), "Multi-rate Kalman filtering for the data fusion of displacement and acceleration response measurements in dynamic system monitoring", *Mech. Syst. Signal Pr.*, **21**(2), 706-723.