

PAPER

3D dynamic displacement-field measurement for structural health monitoring using inexpensive RGB-D based sensor

To cite this article: Mohamed Abdelbarr *et al* 2017 *Smart Mater. Struct.* **26** 125016

View the [article online](#) for updates and enhancements.

3D dynamic displacement-field measurement for structural health monitoring using inexpensive RGB-D based sensor

Mohamed Abdelbarr¹ , Yulu Luke Chen², Mohammad R Jahanshahi³ , Sami F Masri¹, Wei-Men Shen⁴ and Uvais A Qidwai⁵

¹ Sonny Astani Department of Civil and Environmental Engineering, University of Southern California, United States of America

² Department of Astronautical Engineering, University of Southern California, United States of America

³ Lyles School of Civil Engineering, Purdue University, United States of America

⁴ Department of Computer Science, University of Southern California, United States of America

⁵ Department of Computer Science and Engineering, Qatar University, Qatar

E-mail: abdelbar@usc.edu

Received 7 August 2017, revised 16 October 2017

Accepted for publication 18 October 2017

Published 8 November 2017



CrossMark

Abstract

The advent of inexpensive digital cameras with depth sensing capabilities (RGB-D cameras) has opened the door to numerous useful applications that need quantitative measures of dynamic fields whose simultaneous time history quantification (at many points as dictated by the resolution of the camera) provides capabilities that were previously accessible only through expensive sensors (e.g., laser scanners). This paper presents a comprehensive experimental and computational study to evaluate the performance envelope of a representative RGB-D sensor (the first generation of Kinect sensor) with the aim of assessing its suitability for the class of problems encountered in the structural dynamics field, where reasonably accurate information of evolving displacement *fields* (as opposed to few discrete locations) that have simultaneous dynamic planar translational motion with significant rotational (torsional) components. This study investigated the influence of key system parameters of concern in selecting an appropriate sensor for such structural dynamic applications, such as amplitude range, spectral content of the dynamic displacements, location and orientation of sensors relative to target structure, fusing of measurements from multiple sensors, sensor noise effects, rolling-shutter effects, etc. The calibration results show that if the observed displacement field generates discrete (pixel) sensor measurements with sufficient resolution (observed displacements more than 10 mm) beyond the sensor noise floor, then the subject sensors can typically provide reasonable accuracy for translational motion (about 5%) when the frequency range of the evolving field is within about 10 Hz. However, the expected error for torsional measurements is around 6% for static motion and 10% for dynamic rotation for measurements greater than 5°.

Keywords: 3D displacement field, RGB-D sensor, data fusion, non-contact measurements, vibration

(Some figures may appear in colour only in the online journal)

1. Introduction

1.1. Background and motivation

The quantitative measurement of the multicomponent displacement field of *civil infrastructures* [1–5] is an important and challenging problem. Civil infrastructures are continuously deteriorating due to misuse, material aging, and absence of sufficient maintenance. Additionally, the increase of load demands and environmental changes can cause overstressing that leads to the failure of infrastructures [6]. Acquiring the structural displacements is an important indicator in assessing structural safety [7], and identifying changes or damage that might occur in a structure, and hence improve the maintenance options. Furthermore, acquired response data are used as the basis for the performance analysis of the structure via several system identification methods [8], model updating [9, 10], and various structural control systems [11]. The structural response civil infrastructure systems under dynamic loads is important to accurately evaluate the load-carrying capacity of a flexible structure, and estimate its physical parameters in a state-space model identified from the displacement response data [12–16]. Dynamic displacement measurements are also important for strain quantification as they can be directly calculated from displacement measurements, especially for large strain applications [17].

However, it is challenging to perform quantitative measurements using conventional contact-type sensors to acquire the time history of the multi-component displacement and torsional/rotational field of a distributed system undergoing dynamic response. The development and availability of consumer grade vision-based systems, at affordable cost, provides the potential for developing monitoring systems that can enhance the capabilities and increase the fidelity to monitor displacement fields (as opposed to a specific discrete point) of the distributed systems that are undergoing complex three-dimensional (3D) displacement.

Conventional sensors such as accelerometers, linear variable-differential transformers (LVDTs), inclinometers, gyroscopes and global positioning systems (GPS) have various practical limitations in obtaining accurate displacement field measurements. However, vision-based methods have shown promising results in both laboratory and field experiments. Vision-based approaches have the advantage of both high spatial and temporal resolution. Also, for full-field dynamic displacement measurements, vision-based methods have several advantages: (1) contactless sensing; (2) full-field displacement measurements; (3) multi-component measurements (torsion effects); (4) 3D measurements; (5) cost-effective data acquisition systems; and (6) easy operation. These features make the vision-based sensors with depth capabilities a promising technology to track and quantify evolving displacement fields.

1.2. Literature review

Conventional contact-type sensors have been deployed for continuous monitoring of the dynamic displacement and

rotations directly for high-rise buildings and long span suspension bridges. Conventional sensors include LVDTs, linear potentiometers, GPS [18–22], accelerometers [23, 24], inclinometers [25], and micro-electro-mechanical systems (MEMS) gyroscopes [26]. One of the main drawback of contact-type sensors is their operation requirement of a fixed platform as a reference point to measure the relative displacement between the reference point and a point on the structure; however, this is a major disadvantage of position sensors since it is not easy to find a rigid platform close to the structure in many cases.

Furthermore, each sensor has its practical limitations. For example, the main limitation of *GPS* technology is that the GPS signals cannot be received indoors. In addition, the multipath interference caused by reflective GPS signals of surrounding surfaces such as water, metal, and glass, adds error on true GPS signals [27]. For indirect measurement system such as *accelerometers*, the displacements are computed from recorded acceleration data by implementing numerical integration twice to obtain velocity first, and then displacement. During the numerical integration process, the drift error and direct current (DC) bias will be amplified and the displacement estimate significantly differ from the actual displacements [28, 29]. *Inclinometers* are highly sensitive to linear acceleration, environmentally induced magnetic fields, require a long response time, and only measure rotations about axes perpendicular to gravity [30]. In order to obtain the values for rotational response from *MEMS gyroscopes*, it is required to integrate the angular speed with respect to time. However, the integration process can produce large drift over time due to the presence of bias and noise in the angular rate signal [31, 32]. These effects are more severe for low cost MEMS gyroscopes.

Although the *contact-type sensors* are already well-developed and can acquire precise data with high sampling rate, they can only obtain displacement/rotation for one single spot on a structure. If full-field displacements need to be measured, this requires one or several sensor networks consisting of large number of sensor nodes. This will raise the complexity of the corresponding cabling, power consumption, data communication, synchronization, and computation needs. In some applications, the typical sensor networks add mass on the target structure, and may change the structure's dynamic characteristics. If the tested structure undergoes damage during its monitoring phase, it could damage contact-type sensors and degrade the performance of the monitoring system. To overcome these problems, the technology for displacement measurement is trending toward developing cost-effective contactless sensors that can be quickly deployed on remote sites to measure the full-field displacement dynamic, in three dimensions.

The terrestrial laser scanner (TLS) is an important contactless technique for displacement measurement based on optical methods. TLS has been widely used for scanning large and complicated scenes that consist of complex objects and shapes with greater precision and resolution in a 3D space by generating precise 3D point cloud data [33, 34]. TLS applications include the measurement of static displacement,

displacement and dimensions of structural components [35–38]. TLS systems are limited for static displacement and displacement, since they are programmed to measure one point at a time. Recently, Kim *et al* [39] used a specific model time-of-flight TLS RIEGL VZ-400 to measure the 2D dynamic displacement of a cantilever beam with line scan mode (i.e., by repetitively moving the laser beam along a line); however, only a few commercial models have this function. Overall, the laser scanners are infeasible for dynamic field displacement measurements due to their hardware limitations and exorbitant cost.

Digital cameras are alternative optical methods. In recent years, the technology of digital cameras has been improved and advanced significantly to provide high-quality images or videos with light-weight and compact-size bodies. The techniques using digital cameras to measure full-field 3D dynamic displacements are diverse, including close-range photogrammetry, 2D/3D digital image correlation, blind identification and motion magnification have been proposed [40–49]. Although digital image sensors and digital image processing can provide diverse techniques for contactless displacement measurements, target illumination is a big issue, especially when a shadow on the object confuses the image processing software.

Vision-based systems have been developed to measure rotational angle of large civil structures. Jeon *et al* [50] proposed 6-DOF displacement static measurement system using a paired structured light system that included lasers, cameras, and screens. Lee *et al* [51] introduced multiple camcorders and image processing techniques to measure the dynamic rotational angle of a structure. Park *et al* [52] utilized a system consisting of a laser source, a frame grabber, and a commercially available home video camcorder to measure rotational angles for bridge supports. However, vision-based systems are easily affected by external changes such as weather and illumination, and only work properly if installed in the correct line of sight. In addition, measurement accuracy in these systems is affected by the distance between the camera and the target.

Recently, the rapid development of low-cost, off-the-shelf, RGB-D sensors using time-of-flight or structured light technologies has attracted researchers' attention in various science and engineering fields although some of the RGB-D sensors were originally designed to be a natural user interface for entertainment purposes. RGB-D sensor consists of RGB color sensor that capture pixel color information, alongside with depth sensor that acquire depth data. In 2010, Microsoft released the first-generation Kinect using a patented technology based on structured light that could acquire 640×480 depth images at 30 frames per second (fps). In 2014, the second-generation Kinect was launched that used time-of-flight technology to obtain 512×424 depth images at 30 fps. Although both sensors have a relatively low price (currently under \$100), they provide reasonable performance and depth accuracy. The 3D measurement features make the inexpensive Kinect sensors have wide applications, including gesture recognition [53–55], augmented reality [56, 57], robotic

navigation [58, 59], medical applications [60, 61], earth science [62] and structural engineering [63].

Most inexpensive off-the-shelf RGB-D sensors are based on infrared technology and they are subject to interference by direct sun light. This restricts their outdoor applications. Several in-depth evaluations of noise sources and system limitations, including ambient background light, semi-transparent and scattering media, dynamic scenery, multi-sensor interference, etc., for RGB-D sensors are reported by [64–66]. The noise sources investigated in these studies could create invalid depth values or 'holes' in the depth image. The RGB-D sensing technology is still advancing to overcome the system limitations [67–69] to manufacture affordable 3D hand-held sensors to provide high quality data for indoor and outdoor applications.

1.3. Objectives

This study is focused on using a class of inexpensive RGB-D sensors that are equipped with an RGB camera and an active depth sensor to monitor evolving multi-component displacement fields of flexible structures under dynamic loads. Kinect v1 is a representative of this class of sensors that are designed for video games, where the requirements for gaming is quite different than using it as an accurate scientific sensors. It is important to mention that the second-generation Kinect (v2), which was released when the tests of this study was towards the end. Our previous work [70] focused on the Kinect calibration and performance evaluation for 1D dynamic translational motion. The goal of this study is to complete the evaluation of the Kinect v1 performance envelope for dynamical displacement field measurements, including 3D translational motion with simultaneous rotational/torsional motion components. Performance characteristics that are evaluated include linearity of measurements, distortion, noise effects, displacement and amplitude bounds, displacement accuracy relative to motion frequencies, displacement accuracy relative to direction of motion of target with respect to sensor, influence of lighting conditions, and the effects of distance between sensor and target.

1.4. Scope

The remainder of this paper is organized as follows. Section 2 describes the data extraction approach to obtain the Kinect data. Section 3 proposes the methodology used to compute the time history records of dynamic translation and rotation. Section 4 presents the experimental setup for RGB-D data acquisition to measure dynamic displacement of a flexible structure. illustrates the different type of sensors and data acquisition system used in this study. Section 5 explains the data calibration and processing procedures for different sensors used in this study. Sections 6 and 7 discuss the experimental results and data analysis for Kinect v1 sensor performance under different types of dynamic excitation. The conclusions and discussion of future work are provided in section 8.

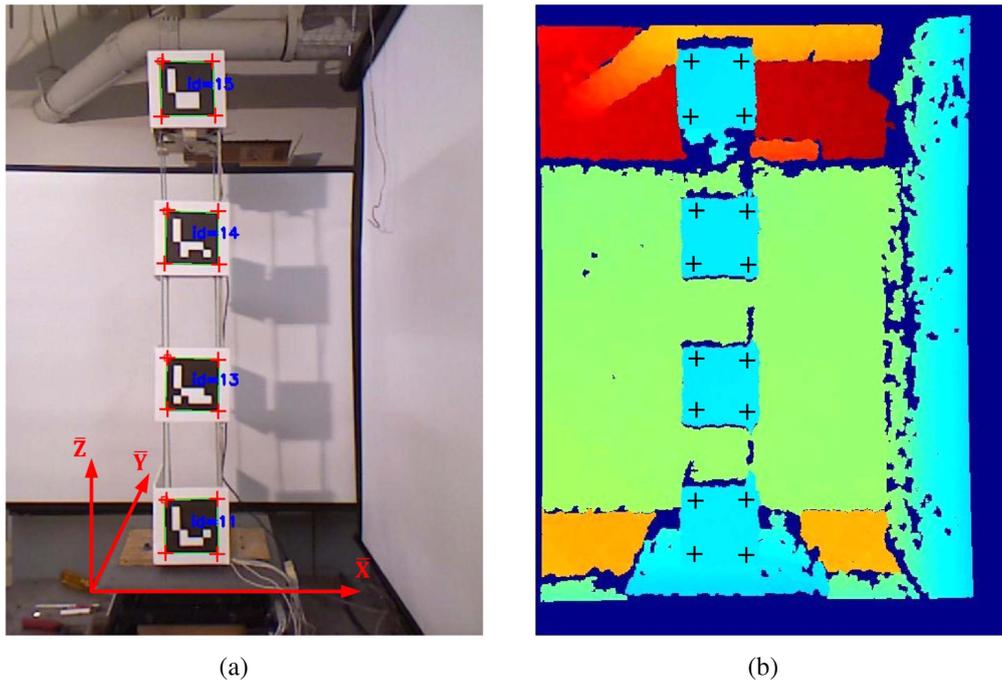


Figure 1. Extracting points of interest from ArUco markers (a) in a color image and mapping to (b) a depth image where the axis of motion parallel is to \bar{X} -direction.

2. Kinect data extraction

2.1. RGB-D sensor calibration and data registration

Before computing displacement and rotation time histories, Kinect camera calibration was performed to compute intrinsic and extrinsic parameters for the Kinect's color camera and infrared (IR) camera. The camera intrinsic parameters are focal length, principal point, and distortion coefficients. The extrinsic parameters include rotation and translation matrices. The Camera Calibration Toolbox for MATLAB was used to estimate the intrinsic and extrinsic parameters [71]. Furthermore, the Kinect's RGB camera and IR camera are located side by side. However, the field-of-views are different between the two cameras, which leads to a shift in the location of the pixel coordinates on the color image compared to the depth image that are taken simultaneously. The Kinect's development software kits (such as OpenNI and Windows SDK) provide programming routines using predefined calibration parameters, that were deployed to align color and depth images. More details about the Kinect's cameras calibration process, and RGB and depth data alignment can be found in [72].

2.2. Target detection and tracking

The Kinect recorded files contain the data that can be used to quantify the dynamic displacement field of the structure. ArUco markers were used to track the points of interest in the color image. The robust marker detection algorithm developed by [73] was used to extract and locate the markers' corners and determine their pixel coordinates (u_C, v_C) in each color image. When the color and depth images are aligned, the pixel coordinates of a depth map are identical to the pixel

coordinates of a color image. Figure 1 shows the ArUco markers detected in a color image figure 1(a) and mapped to the corresponding depth image figure 1(b). It should be mentioned that there is a small time shift between a pair of color and depth frames since Kinect v1 does not support hardware synchronization. Many programming libraries (e.g., Kinect SDK, OpenNI, etc) minimize the time shift to a few milliseconds. The small time shift can be considered as part of the measurement error.

3. Theory and methods

3.1. Displacement calculation

For a pair of aligned color and depth images, the depth value Z_i of the i th target point P_i in the color image can be directly obtained from the depth image using its pixel coordinates: $Z_i(u_C^i, v_C^i)$ where u_C^i and v_C^i are the coordinates of P_i in the color image plane. Subsequently, the world coordinates (X_i, Y_i, Z_i) of the point P_i with respect to the color camera of the Kinect sensor can be computed using the following equations that are derived from a pinhole camera model [74]:

$$X_i = \frac{(u_C^i - c_x) \times Z_i}{f_x}, \quad (1a)$$

$$Y_i = \frac{(v_C^i - c_y) \times Z_i}{f_y}, \quad (1b)$$

$$Z_i = Z_i(u_C^i, v_C^i), \quad (1c)$$

where (c_x, c_y) are the principal points, and f_x and f_y are the focal lengths of the color camera in x and y directions on the

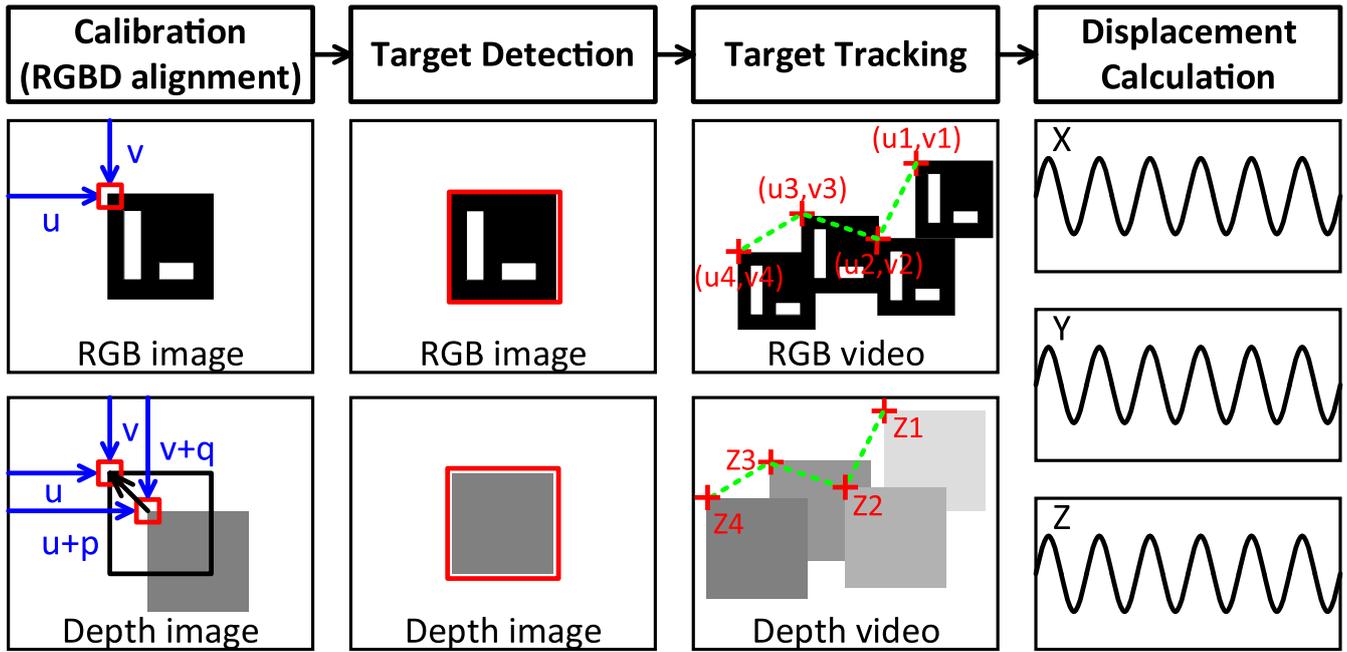


Figure 2. Overview of measuring dynamic displacements using the Kinect sensor. The process of displacement calculation consists of stereo calibration for color and depth images, target (ArUco markers) detection, tracking and displacement calculation using equation (1).

Table 1. Storage arrangement for time history for world coordinates of target points of ArUco marker. Note that for each ArUco marker, the coordinates for its four corner points are stored.

Time	Point 1			Point 2			...	Point 4		
t_1	X_{11}	Y_{11}	Z_{11}	X_{12}	Y_{12}	Z_{12}	...	X_{14}	Y_{14}	Z_{14}
t_2	X_{21}	Y_{21}	Z_{21}	X_{22}	Y_{22}	Z_{22}	...	X_{24}	Y_{24}	Z_{24}
\vdots										
t_m	X_{m1}	Y_{m1}	Z_{m1}	X_{m2}	Y_{m2}	Z_{m2}	...	X_{m4}	Y_{m4}	Z_{m4}

image plane respectively. The intrinsic parameters $f_x, f_y, c_x,$ and c_y are estimated through the camera calibration as discussed in section 2.1. The measurement of dynamic displacement between two points on two sequential frames is computed using the pinhole camera model shown in equation (1). ArUco marker detections, frame timestamp readings, and target points detection and localization in the color image were performed in C++. MATLAB was used to obtain the depth values, and compute the world coordinates with respect to the color camera of the Kinect sensor. The time histories of the positions for the four corners points of the ArUco markers were stored in accordance with the format shown in table 1. The displacements in X, Y, or Z direction of a target point can be plotted according to the time history record. Figure 2 summarizes the process of displacements calculation using the Kinect sensor.

3.2. Rotation calculation

After computing the time history for 3D coordinates of each marker, the rotation angle time history of each marker can be computed based on the geometrical relationships shown in

figure 3, using following equations:

$$\beta_j = \tan^{-1}\left(\frac{Y_{j1} - Y_{j2}}{X_{j1} - X_{j2}}\right), \quad (2a)$$

$$\theta_j = \tan^{-1}\left(\frac{Z_{j1} - Z_{j2}}{X_{j1} - X_{j2}}\right), \quad (2b)$$

where β_j is the estimated rotation angle when the axis of rotation is parallel to the Kinect's Z-axis (i.e., depth axis) at time step t_j based on world coordinates computed from Kinect data, θ_j is the estimated rotation angle when the axis of rotation is parallel to the Y-axis (i.e., pixel axis) at time step t_j based on world coordinates computed from Kinect data. $X_{j1}, Y_{j1}, Z_{j1}, X_{j2}, Y_{j2},$ and Z_{j2} are the computed world coordinates based on Kinect data at time step t_j for marker corner points 1 and 2, respectively. As discussed in section 7.1, when the computed rotation angle is based on world coordinates X and Y, it is tagged as 'pixel-based' data. However, when the computed angle is based on world coordinates X and Z, it is tagged as 'depth-and-pixel-based' data.

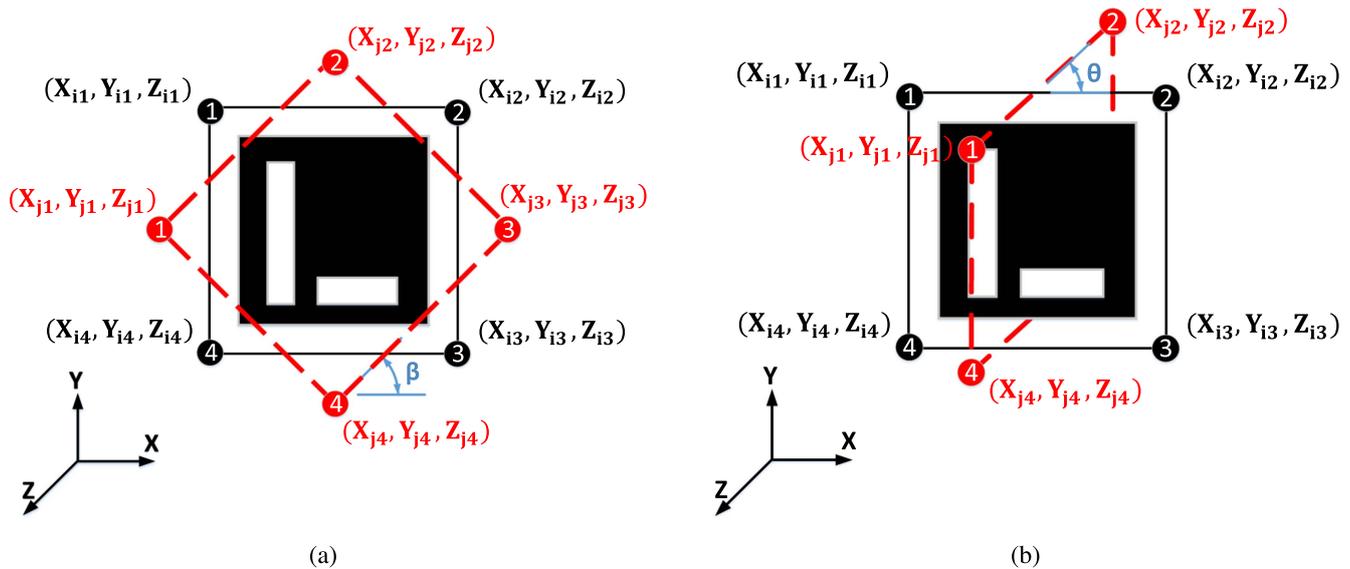


Figure 3. Computed angle using Kinect data (a) *pixel-based* data; refer to the computed rotation angle based on Kinect data when the axis of rotation is parallel to the Kinect's Z-axis (i.e., the depth axis), and (b) *depth-and-pixel-based* data; refer to the computed rotation angle based on Kinect's data when the axis of rotation is parallel to the Kinect Y-axis.

4. Experimental setup

4.1. Testbed structure

To perform the dynamic tests, a flexible testbed shown in figure 4 was designed and built. The testbed was equipped with two sensor configurations: (1) the contact-type sensors setup that was used for calibration purposes (see figure 4(a)), and (2) the non-contact-type sensor setup (i.e., Kinects and target markers, see figure 4(b)). In designing the testbed structure, some desirable objectives were incorporated in the design plan. The designed structure can simulate a frame, or a wing-like structure, and to have the natural frequency of the structure in the range of 1–5 Hz. This requirement was due to the fact that a high-amplitude response is needed to sense and track the response. Furthermore, the above frequency range covers the bandwidth of many typical civil structures. In addition, the testbed was designed to keep the second and third modes of the testbed response below 10 Hz, so as to result in modal interaction (i.e., couple translational and torsional modes for the random input studies).

The final design consisted of three modular sections each with dimensions $205 \times 140 \times 300$ mm as shown in figure 5. As can be seen, each module has a wooden slab with 6.4 mm thickness, and four steel columns with 3.2 mm diameter. A mass of 2 Kg was attached to the second module to adjust the mass distribution in order to keep the torsional mode of the testbed response below 10 Hz. The shaker was connected to a rotating round wooden table with a diameter of 134 mm via a 407 mm aluminum arm. The shaker and round table form a simple slider-crank mechanism to convert the linear motion (shaker) to circular motion (round table). The structure was configured in two different ways: (1) the testbed was mounted on an electromagnetic long-stroke shaker (see figure 5(a)), and (2) the testbed was mounted on a round rotating wooden table (see figure 5(b)). The exciter could be modified to

generate arbitrary dynamic scenarios such as harmonic, swept-sine and random excitation forces.

The frequency of the first three lateral mode shapes were estimated through the fast Fourier transform (FFT) analysis of the recorded acceleration from sensor Acc 1 and Acc 3, as shown in figure 6. The corresponding mode shapes from finite element analysis (FEA) model are displayed in figure 7. The first mode is primarily a bending mode as expected. The second mode is a pure torsional mode as shown in figure 7(b). The frequency of the structure shown in figure 7(c) is 7.3 Hz, which is at the lower end of the target range, since the model includes a concentrated mass at the second module. The frequency values for the first three mode shapes from the experimental data and FEA are listed in table 2.

4.2. Consumer-grade RGB-D camera

The Kinect v1 shown in figure 8(f) was used in this study as the non-contact type sensor. The Kinect main components are: a color complementary metal-oxide-semiconductor (CMOS) sensor, an infrared CMOS sensor, and an infrared projector, to capture 640×480 pixel RGB images and generate 640×480 pixel depth maps at 30 fps using PrimeSense's light coding technology. The Kinect v1 can run on different platforms (i.e., Windows, Linux, Mac, etc), and has various choices of programming tools (i.e., Microsoft SDK, OpenNI, OpenKinect, etc).

Three Kinects were mounted on adjoining structures. According to [75], the error in depth measurements increases quadratically as the sensor-object distance increases. So, the Kinects were placed about 1000 mm from the testbed structure to minimize the depth error that close to the Kinect's minimum range of 800 mm. To reduce the motion blur [72] in recorded Kinect data under dynamic motion, two lighting panels were mounted in front and on the side of the testbed structure.

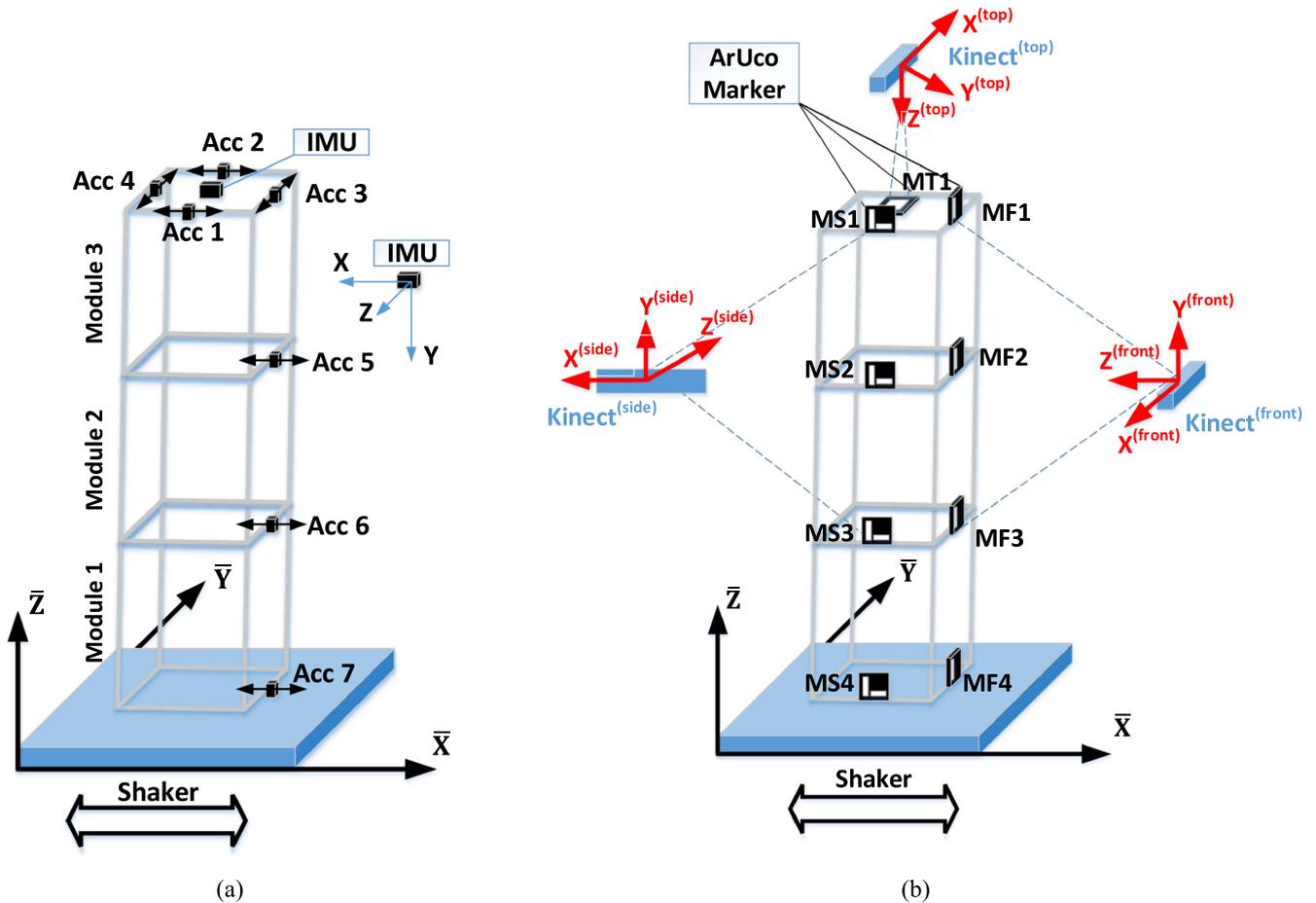


Figure 4. Sensors locations (a) accelerometers Acc 1, Acc 2, Acc 5, Acc 6, Acc 7 are used to measure acceleration in the \bar{X} -direction; accelerometers Acc 3, Acc 4 are used to measure acceleration in the \bar{Y} -direction; and an IMU is used to measure rotation (torsion) in module 3 around the IMU's y-axis that coincide with the structure's axis of symmetry, and (b) ArUco markers deployed on test-bed structure and the Microsoft Kinects used; Kinect^{side}, Kinect^{front}, Kinect^{top} (note that the depth axis for Kinect^{side} is perpendicular to the shaker direction of motion, the depth axis for Kinect^{front} is parallel to the shaker direction of motion, and the depth axis for Kinect^{top} is perpendicular to the plane of motion (i.e., $\bar{X}\bar{Y}$ plane).

The ArUco markers of size (100 mm \times 100 mm) were attached on each floor of the testbed structure. The size of markers was chosen relatively large compare to the structure dimensions for calibration process. The ArUco module is based on the ArUco library, a popular library for detection of square fiducial markers developed by [73]. The outer corners of the marker represented the points of interest to be detected. One of the main advantages of the chosen markers is that each ArUco marker has a unique pattern, that facilitates the marker detection process.

The approximate location of the Kinect sensors, ArUco markers as well as the markers nomenclature used throughout this study are shown in figures 4(b), 8(a)–(c). The Kinect^{front} was used to detect markers MF1, MF2, MF3, and MF4, the Kinect^{side} was used to detect markers MS1, MS2, MS3, and MS4, and the Kinect^{top} for marker MT1.

4.3. Contact-type sensors

Three contact-type sensors were deployed to serve as ground truth to validate the accuracy of the Kinects' data. An LVDT transducer, SchaevitzTM Sensors HR 4000 was used to record the displacement of the shaker stroke. Seven Endevco 7290E

variable capacitance accelerometers were used to measure acceleration at different locations of the testbed structure. The sensors had a full scale range of ± 10 g with a frequency bandwidth of 0–500 Hz. In addition, an inertial measurement unit (IMU), Yost Labs 3-SpaceTM Sensor Micro USB, was deployed at the third module. The approximate location of the accelerometers, IMU, as well as the sensor numbering used throughout this study are displayed in figures 4(a), 8(d), and (e).

4.4. Data acquisition systems

This study used three data acquisition systems. The first one was for controlling the shaker, collecting accelerometers and LVDT measurements using NI LabVIEW. The second one was for acquiring IMU data (accelerometer and gyroscope data) using the YEI 3-Space Sensor Software Suite. The third one was for Kinect data acquisition.

The Kinect data acquisition software was developed in C/C++ language using OpenNI (ONI), OpenCV and ArUco libraries, and operated on a Microsoft Windows 7 platform. The color and depth images were aligned with reasonable accuracy using the OpenNI library. The Kinect data acquisition

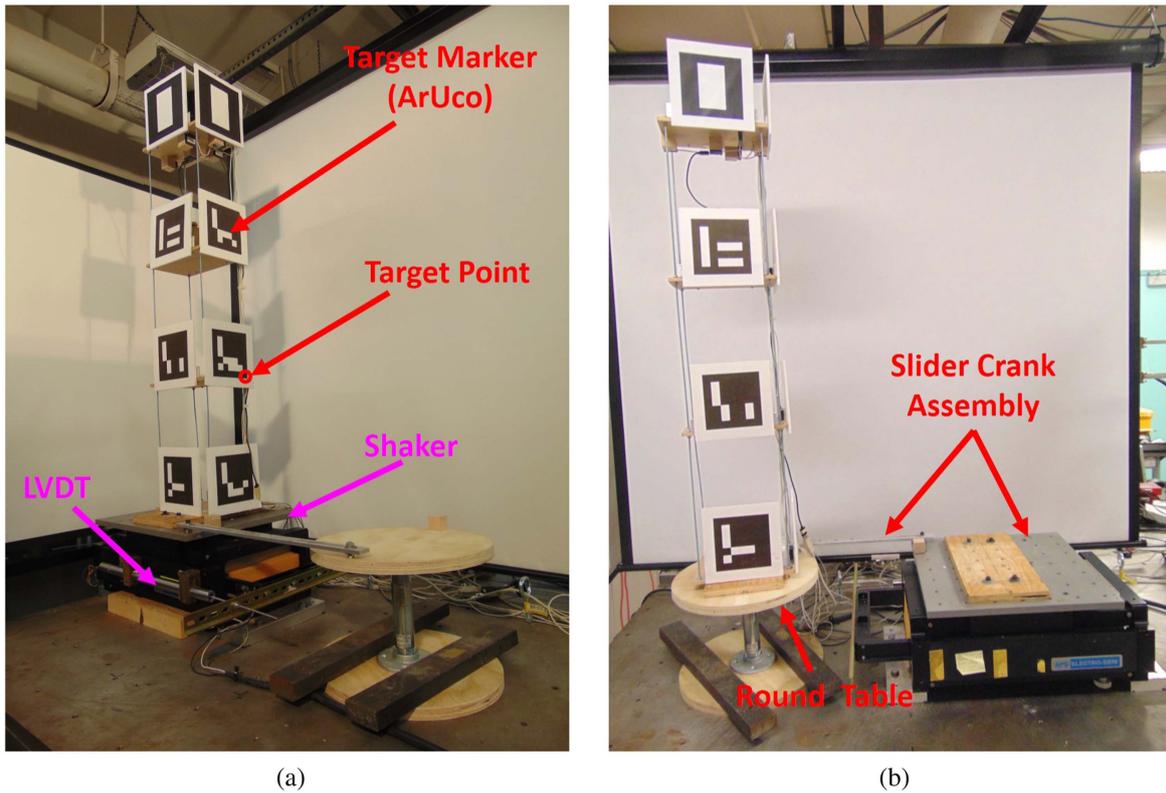


Figure 5. Overview of the experimental setups (a) the experimental setup used for the evaluation of harmonic translation, harmonic rotation, static rotation and random translation, and (b) the experimental setup used for the evaluation of random rotation, including slider-crank assembly (uniaxial shaker connected to round rotating table by rigid arm to convert linear motion to circular motion).

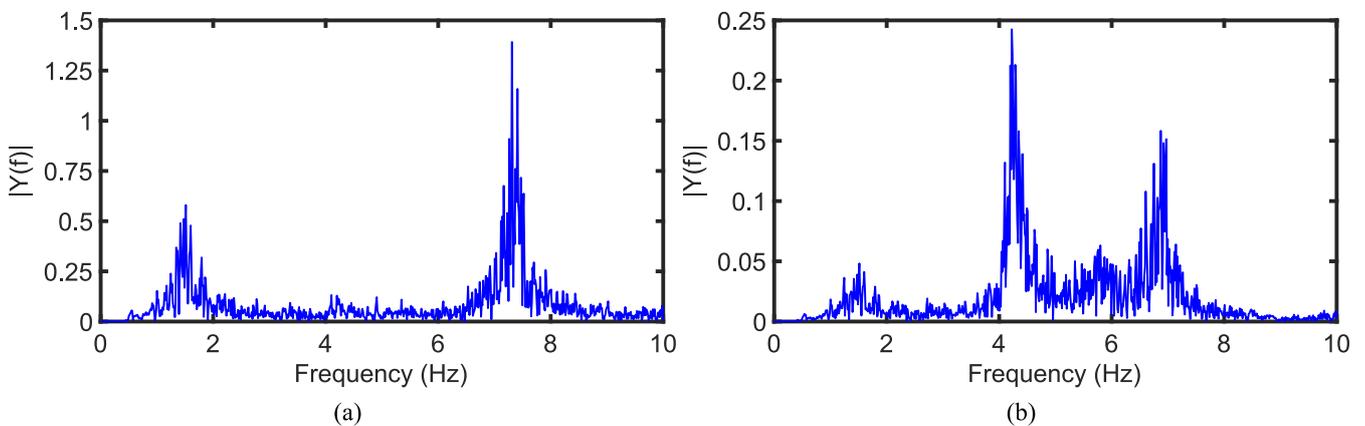


Figure 6. Testbed frequency response based on acceleration data acquired from accelerometer sensors (a) Acc 1 located at module 3, and (b) Acc 3 located at module 3 (note that the amplitude scales are significantly different).

software saved aligned color and depth frames together into a video file using the OpenNI file format. The frame numbers and timestamps were also saved into a text file.

5. Data preprocessing and verification

5.1. Kinect

After extracting Kinect raw data from the recorded ONI video files, the data was processed through two steps. The Kinect

data was filtered to remove DC bias, reduce noise, and smooth the signal. Two FFT digital filters were applied to the raw data: a high-pass FFT filter was used to remove the DC bias, and a low-pass FFT filter was used to clean high-frequency noise.

5.2. Accelerometers

The computed displacements from acceleration data were used as ground truth for the evaluation of Kinect displacement data at each floor. In order to obtain the displacement and velocity time histories, the acceleration records were

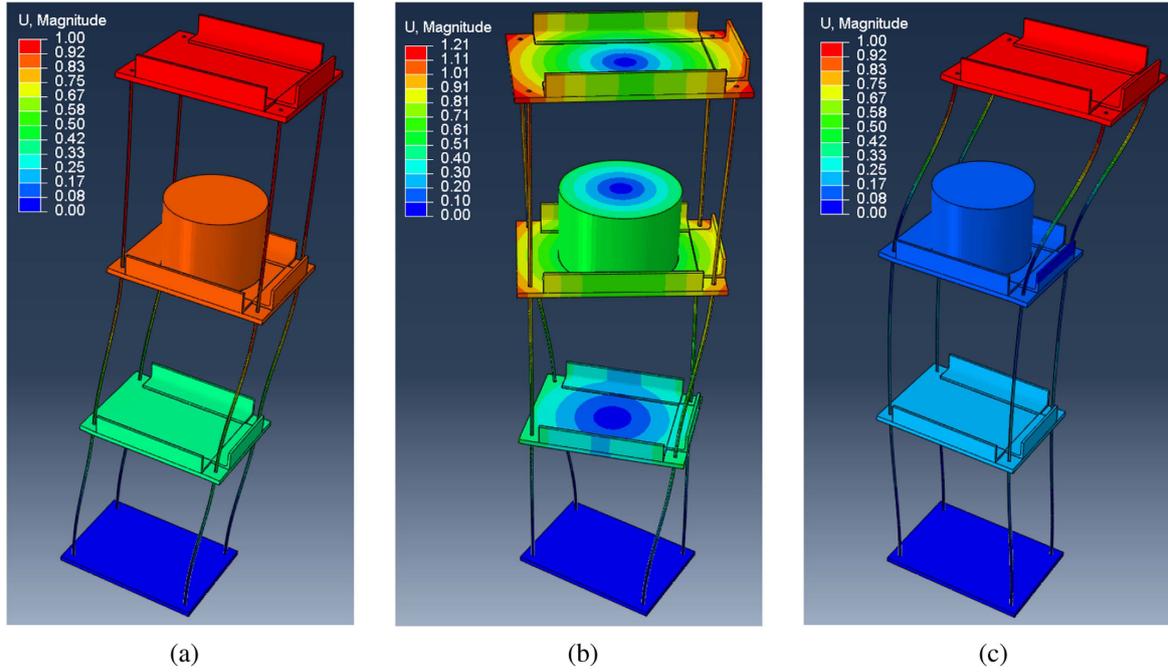


Figure 7. Mode shapes from FEA model (a) the first mode of response at 1.3 Hz, (b) the second mode of response at 4.3 Hz, and (c) the third mode of response at 7.3 Hz. Note that the Abaqus FEA software was used to generate the model.

Table 2. Comparison of natural frequencies from FEA model and experimental measurements.

Mode	ω (Hz)		
	Experimental	FEM	$\Delta\omega$ (% Difference)
1	1.54	1.46	5.48
2	4.25	4.10	3.66
3	7.30	7.68	4.95

windowed, detrended, band-pass filtered and numerically integrated for each accelerometer. Figure 9 shows a typical time history of measured accelerations and computed displacements for an accelerometer. The computed displacements were calibrated with direct displacement measurements from the LVDT attached at the base to evaluate the effects of bias and drift due to double integration [28, 29]. Figure 9(b) illustrates the aligned computed displacement and LVDT measurements. It can be seen that the two signals match well and the computed root mean square (rms) error was less than 10% for general random signal. However, for harmonic signal, the rms error was less than 2% as it is discussed in section 6.1.

5.3. Inertial measurement unit (IMU)

MEMS IMU consists of tri-axial accelerometers, gyroscopes, and magnetometers. MEMS gyroscopes are typically angular rate gyroscopes that are designed to measure the angular rate. In order to get the rotation angle using a MEMS rate gyroscope, it is required to integrate the measured angular rate with respect to time. This integration procedure was calibrated to evaluate the error in computed angular time-histories

due to bias and noise in the acquired angular rate signal [31, 76]. The angular rate records were windowed, detrended, band-pass filtered and numerically integrated in order to obtain the angular time-histories and to reduce the drift and signal noise.

The experimental slider-crank setup shown in figure 5(a) was used to calibrate the gyroscope. Figure 10(a) illustrates the slider-crank mechanism. The angle of rotation of the slider-crank was derived from basic geometry concepts defined as follows:

$$m = L - x, \quad (3a)$$

$$p = \sqrt{m^2 + R^2}, \quad (3b)$$

$$\alpha = \text{Tan}^{-1}\left(\frac{m}{R}\right), \quad (3c)$$

$$C = \text{Cos}^{-1}\left(\frac{p^2 + R^2 - L^2}{2pR}\right), \quad (3d)$$

$$\theta_{\text{slider-crank}} = C - \alpha. \quad (3e)$$

Upon substituting equations (3a)–(3d) into (3e), the angle of rotation of the slider-crank can be computed using the following equation:

$$\theta_{\text{slider-crank}} = \text{Cos}^{-1}\left(\frac{(L-x)^2 + 2R^2 - L^2}{2R\sqrt{(L-x)^2 + R^2}}\right) - \text{Tan}^{-1}\left(\frac{L-x}{R}\right), \quad (3f)$$

where $\theta_{\text{slide crank}}$ is the angle of rotation of the crank, L is the length of the rod, and x is the shaker displacement measured by the LVDT. The IMU was fixed on wooden cube attached to the round table. Two sets of tests were performed based on

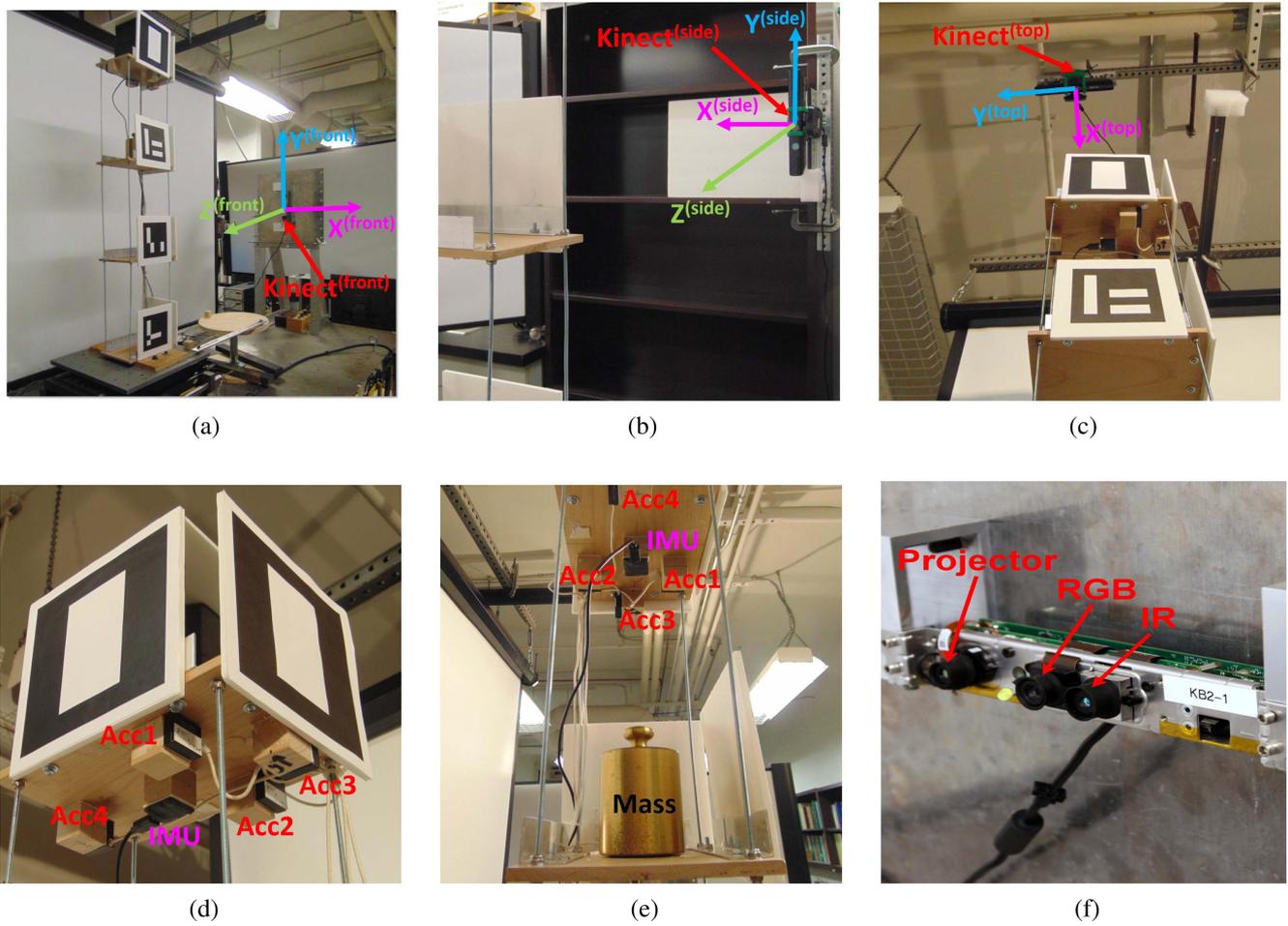


Figure 8. Test apparatus and components (a) the Kinect^{front} fixture, (b) the Kinect^{side} fixture, (c) the Kinect^{top} fixture, (d) the 4 accelerometers and IMU deployed on module 3, (e) the 2 Kg mass attached to module 2, and (f) the Microsoft Kinect sensor. Note that the size of markers (100 mm × 100 mm) are relatively large for calibration process, while smaller markers can be used.

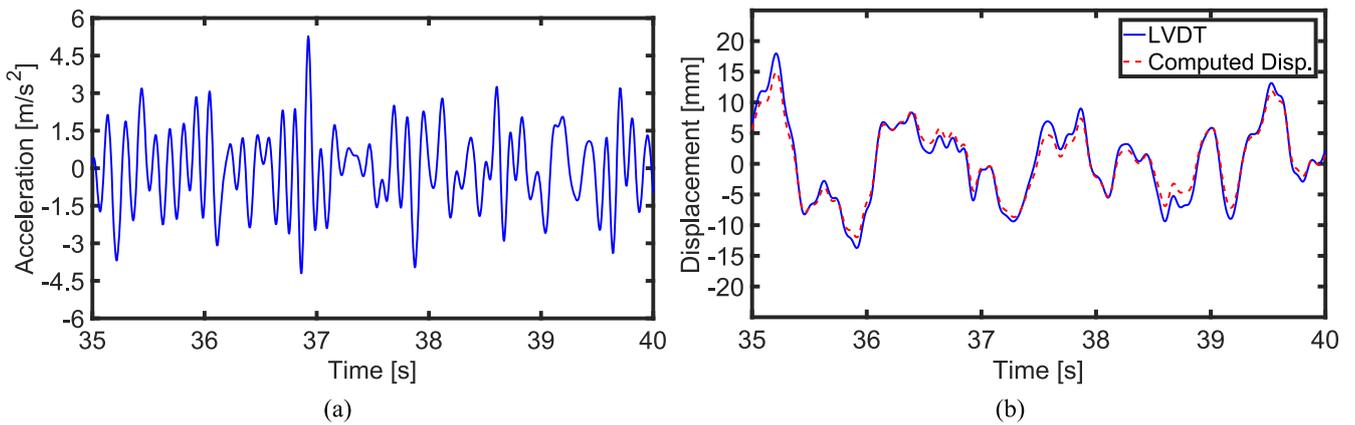


Figure 9. Sample of (a) recorded data from an accelerometer (Acc 7), and (b) corresponding computed displacement time history and LVDT measurements.

the excitation type. For the first set, the shaker was excited with 2 Hz harmonic excitation, repeated exactly three times to calibrate the gyroscope rotation time history around the IMU local axes (see figure 10(b)). Figure 11 displays a comparison between the computed angle of rotation and the ground truth (slider-crank). The error analysis based on the harmonic

excitation is summarized in table 3. For the x and z directions, the normalized errors are approximately 5%. However, for the y direction measurements, the error is less than 3%.

The second set of tests were conducted to measure low-frequency random vibrations using the IMU gyroscope. Three distinct input random signals with relatively low rms levels

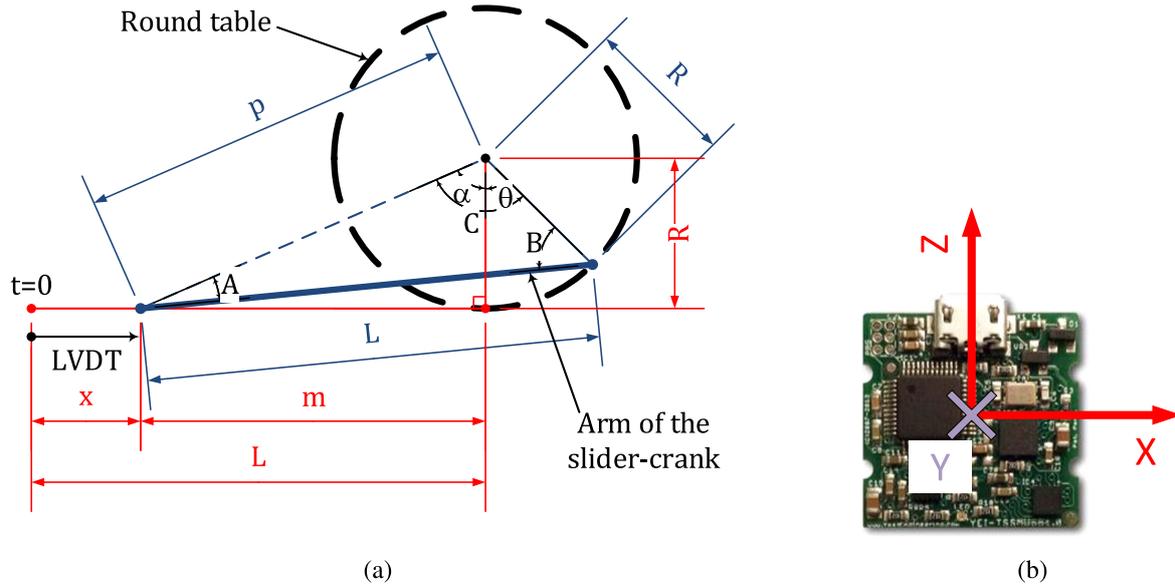


Figure 10. Sketch for (a) slider-crank geometry, and (b) IMU axis.

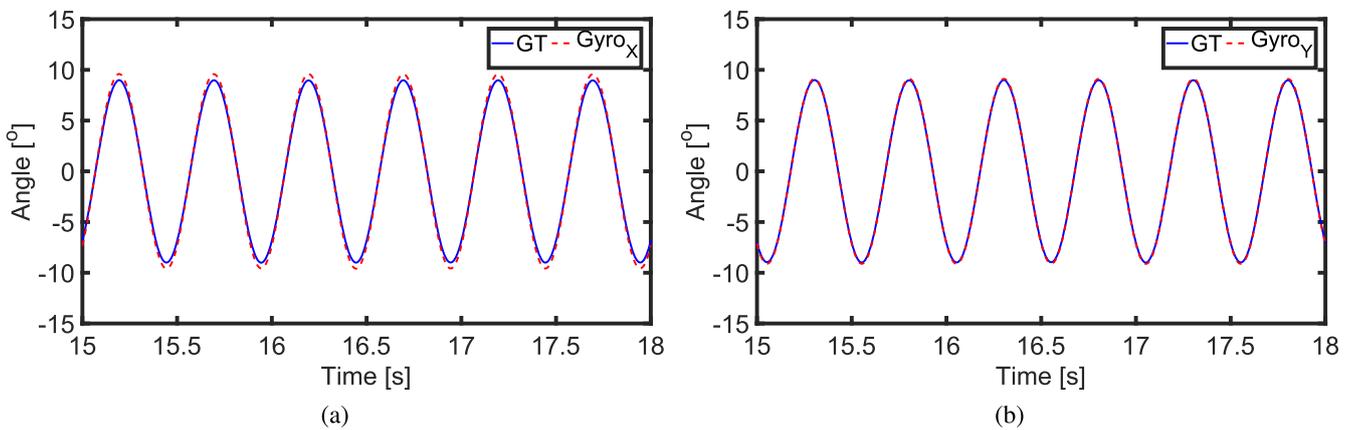


Figure 11. Sample aligned IMU-Gyroscope and ground truth (GT) data under harmonic excitation (2.0 Hz, 16.0°): (a) x-direction, and (b) y-direction. The GT is the computed angle using the slider-crank mechanism.

Table 3. Comparison of normalized error, mean error, and standard deviation error under harmonic and random measurements of angle of rotation obtained from IMU.

	IMU-x-direction		IMU-y-direction		IMU-z-direction	
	Harmonic	Random (rms = 2.35°)	Harmonic	Random (rms = 2.00°)	Harmonic	Random (rms = 2.83°)
Normalized error (%)	5.81	9.82	1.98	10.28	5.08	6.10
Mean error (%)	...	1.52	...	2.2	...	0.90
Std error (%)	...	1.24	...	1.63	...	0.91

were generated to excite the shaker (approximately 2.3°). The three input signals were sampled from a Gaussian distribution and filtered by a low-pass filter to remove high-frequency components above 10 Hz. The random rotation for the slider-crank and the gyroscope were computed. Figure 12 illustrates the computed random rotation for the slider-crank and the gyroscope for the IMU's x and y directions. Figure 13 shows

the comparisons of the probability density functions (PDFs) estimated based on the computed angles from the slide-crank (ground truth) and gyroscope by integration. For instance, there is a good correlation between the two computed quantities. Three metrics were deployed to quantify the accuracy of random rotation measurements for the IMU-gyroscope sensor with respect to the ground truth data: the normalized

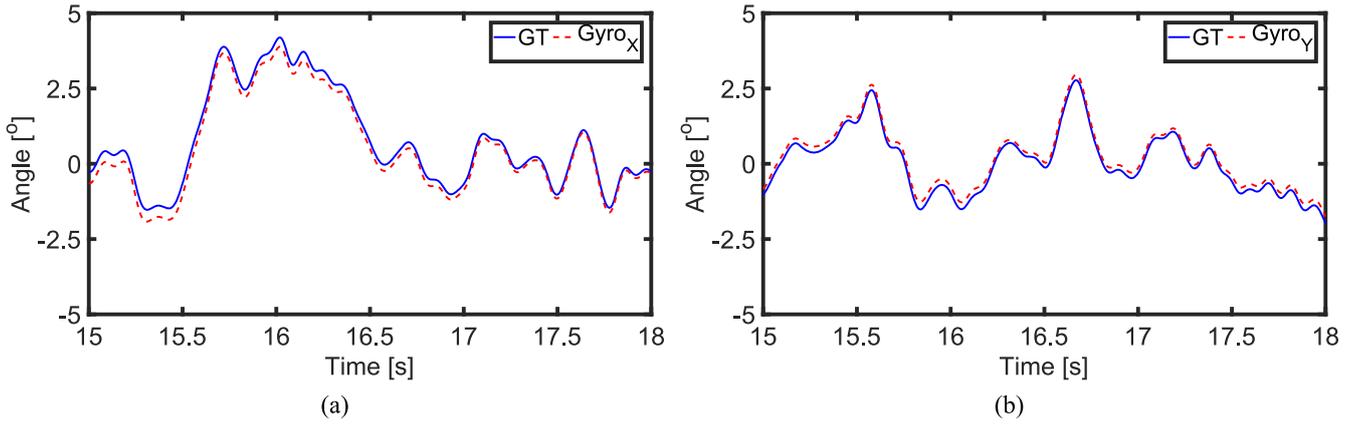


Figure 12. Sample aligned IMU-Gyroscope and GT data under random excitation (a) x -direction (rms of the GT signal = 2.35°), and (b) y -direction (rms of the GT signal = 2.00°). The GT is the computed angle using the slider-crank mechanism.

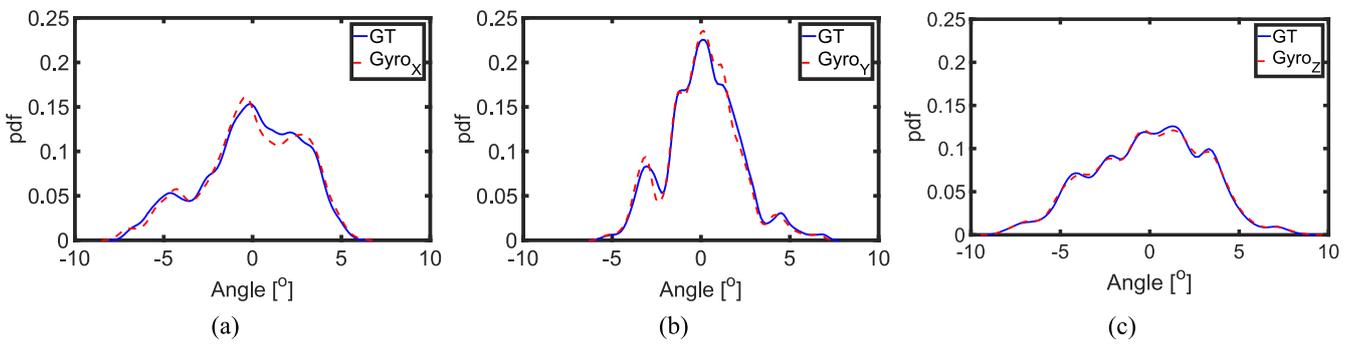


Figure 13. PDFs of the IMU-Gyroscope measurements for the three directions (a) IMU- x -direction (rms of the GT signal = 2.35°), (b) IMU- y -direction (rms of the GT signal = 2.00°), and (c) IMU- z -direction (rms of the GT signal = 2.83°). The GT is the computed angle using the slider-crank mechanism.

error, the mean error for the two PDFs, and standard deviation error for the two PDFs, which are defined as follows:

$$\text{Normalized error} = \frac{\|X_{GT} - X_{IMU}\|_2}{\|X_{GT}\|_2}, \quad (4)$$

$$\text{Mean error} = \frac{\mu_{GT} - \mu_{IMU}}{\mu_{GT}}, \quad (5)$$

$$\text{Standard deviation error} = \frac{\sigma_{GT} - \sigma_{IMU}}{\sigma_{GT}}, \quad (6)$$

where $\|\cdot\|_2$ is the Euclidean norm, X_{GT} is the time history of the ground truth rotation, X_{IMU} is the time history of the computed gyroscope rotation measurements, μ_{GT} and σ_{GT} are the mean and standard deviation for the PDFs of the ground truth data, and μ_{IMU} and σ_{IMU} are the mean and standard deviation for the PDFs of the computed gyroscope rotation measurements, respectively. The results of the quantitative error analyses are summarized in table 3. On the whole, the normalized error, mean error, and standard deviation error decrease when the rms amplitude levels (2.00° , 2.35° , 2.86°) increase. For the random vibration tests, the normalized error analyses show relatively acceptable performance as the rotation angle increases. This can be explained by noting that as the signal rms increases, the bias and noise effects decrease.

Overall, the error analyses of the random tests are acceptable for the problem of interest in this study.

5.4. Data synchronization and alignment

Since different computers were used to record the Kinect videos and ground truth raw data (i.e., acceleration or rotation rate), the Kinect video recording and ground truth data acquisition started at different times. In order to compare the two data measurements, a data alignment procedure was performed. The measurements collected from the Kinect sensor and the ground truth transducer had similar waveforms; hence, a cross-correlation technique can be used to estimate the time delay between the two signals by measuring their similarity. Subsequently, the two signals were synchronized by shifting one of them according to the time delay. It should be noted that the Kinect, accelerometers and IMU data were acquired at different sampling rates. The Kinect mean sampling rate was 25 Hz, however for the accelerometers and IMU the sampling rate was 200 Hz. To enhance accuracy of the synchronization procedure using the cross-correlation approach, the two signals were resampled at a higher sampling rate of 1000 Hz with spline interpolation before data alignment, which also smoothed the peaks to obtain better alignment results. Figure 14 illustrates an example of data alignment after performing cross-correlation.

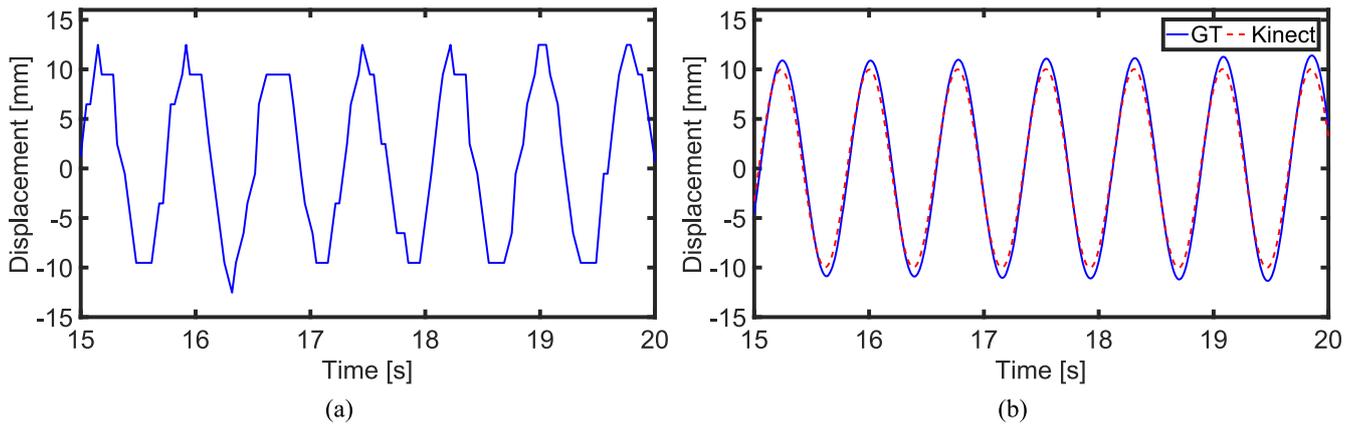


Figure 14. Kinect data post-processing (a) raw depth data, and (b) aligned Kinect and GT data after applying the Kinect post-processing approach that consists of Kinect data filtering, smoothing and alignment.

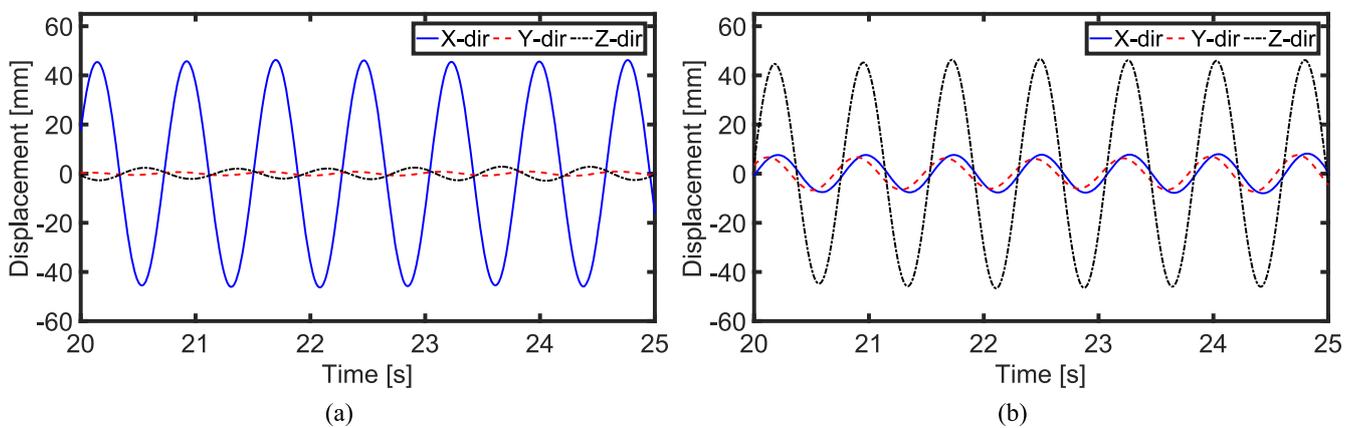


Figure 15. Sample measurements for the structure floors (1.3 Hz) (a) *pixel-based* measurements (Kinect^{side}) based on marker MS1, and (b) *depth-based* measurements (Kinect^{front}) based on marker MF1. Note that x , y , z directions refer to the Kinects' axes.

6. Experimental validation for translational motion

6.1. Harmonic test

The RGB-D camera (Kinect v1) was used to acquire harmonic translational motion with different combinations of frequency and peak amplitudes. The excitation frequency was chosen based on the finite element model results shown in figure 7. In order to get a pure displacement motion, the testbed was excited by a sinusoidal signal with 1.3 Hz frequency corresponding to a cantilever mode shape (i.e., first mode). The harmonic motion of the testbed structure was recorded by the three Kinect sensors. Therefore, there were two cases: (1) the direction of motion was *perpendicular* to the depth axis (i.e., z^{side} -axis) of the Kinect^{side}; (2) the direction of motion was parallel to the depth axis (i.e., z^{front} -axis) of the Kinect^{front} (see figure 4(b)).

For the first case, the z -direction (z^{side}) displacements remained steady while the x -direction (x^{side}) displacements were variable (see figure 15(a)). For the second case, the z -direction (z^{front}) displacements were variable while the x -direction (x^{front}) displacements were constant (see figure 15(b)). Since the structure had no vertical displacement, all of the y -direction measurements for Kinect^{side} and Kinect^{front} were nearly constant. To differentiate between the

two scenarios, the first case was called '*depth-based*' measurement since it was dominated by the depth value. Similarly, the second case was named '*pixel-based*' measurement for it was governed by the color pixel coordinates. Figure 16(a) illustrates the *pixel-based* measurements for all floors. Figure 16(b) shows the *depth-based* measurements for all floors. It can be seen that the pixel and depth based measurements level are almost similar.

The ground truth measurement in this test was the computed displacement from recorded acceleration data. This process was calibrated first using the LVDT attached to the small shaker. The rms error between the processed acceleration and the LVDT was less than 2%. Figure 17(a) illustrates the superimposed direct displacement measurements (LVDT) versus the indirect measurements (processed acceleration), Kinect depth data (the front marker (MF4)), and Kinect pixel data (side marker (MS1)) at the first floor. It can be seen that the LVDT, double integrated acceleration, and the Kinect data match perfectly.

Figure 17(b) compares the displacement values for the fourth floor based on the direct measurement of the displacement of the three markers (MF1, MS1, MT1), and the ground truth obtained from acceleration. It can be seen that the displacement measurements from the three Kinects and

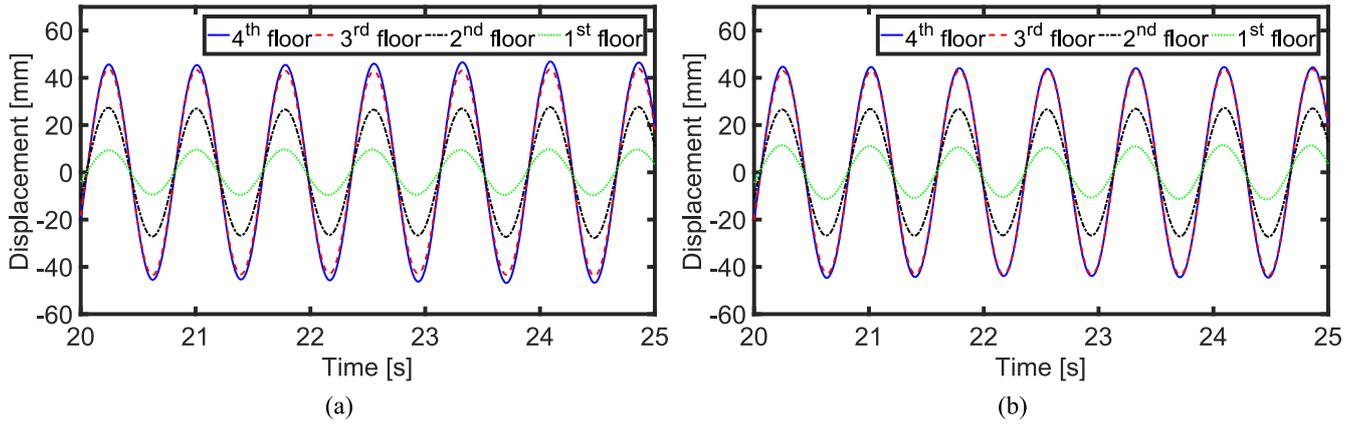


Figure 16. Sample measurements for the structure floors (1.3 Hz) (a) *pixel*-based measurements (Kinect^{side}) based on markers MS1, MS2, MS3, MS4 attached at 4th floor, 3rd floor, 2nd floor, and 1st floor, respectively; and (b) *depth*-based measurements (Kinect^{front}) based on markers MF1, MF2, MF3, MF4 attached at 4th floor, 3rd floor, 2nd floor, and 1st floor, respectively. Note that the response for the 3rd and 4th floors are almost similar due to the mass attached at the 3rd floor.

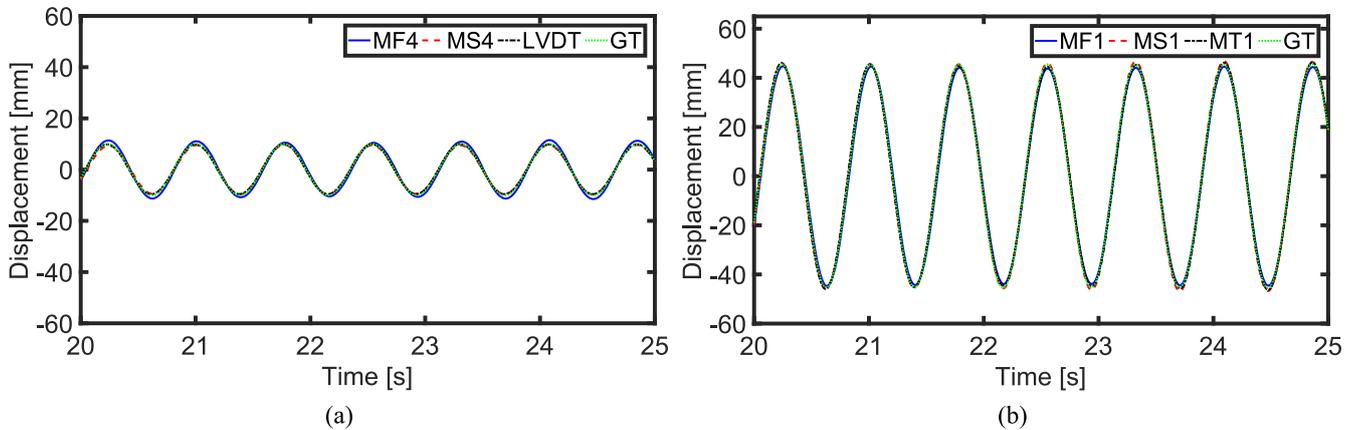


Figure 17. Superimposed plots for the Kinects data versus GT measurements (1.3 Hz) (a) *first* floor, and (b) *fourth* floor. The GT is the computed displacement from acceleration records.

the ground truth measurement completely match, with an rms error less than 4% with respect to the ground truth data.

For each test, the *normalized error* between the Kinect and ground truth data (computed displacement via double integration) were computed for the four corner points of the target markers pattern using the following equation:

$$\text{Normalized error} = \frac{\|X_{GT} - X_{Kinect}\|_2}{\|X_{GT}\|_2}, \quad (7)$$

where $\|\cdot\|_2$ is the Euclidean norm, X_{GT} is the array corresponding to the time history of the sampled ground truth displacement, and X_{Kinect} is the processed displacement time history for the Kinect sensor. Since the estimated normalized errors are similar for the four target points of each ArUco marker, the mean results for each marker are shown in figure 18, that shows one set of fitting curves representing the accuracy of the displacement measurements for different displacement values for different floor, that were estimated using equation (7) under the scenarios discussed below.

For the *Depth*-based measurements, the normalized errors remain at the same level of about 5% that is within the

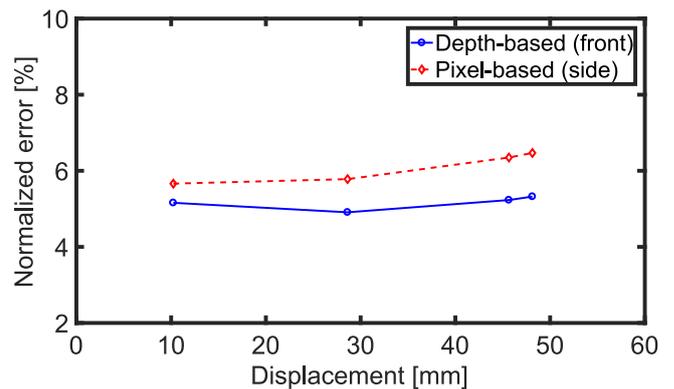


Figure 18. *Depth*- and *pixel*-based measurements errors (1.3 Hz).

expected range based on [70]. However, for the *Pixel*-based measurements, the image sensors when capturing images of moving objects [60, 77]. When the displacement amplitude increases under constant frequency (i.e., average speed increased), severer image distortion decreases the accuracy of horizontal (*x*-axis) displacement measurements [70].

6.2. Random test

Since in practical applications the excitation of interest is not simply harmonic, additional tests were conducted to measure low-frequency random response using the Kinect sensor. The structure was excited by a random signal having 6.81 mm rms level. The signal was sampled from Gaussian distribution, and was filtered by a low-pass filter to remove high-frequency components above 10 Hz. Based on the ground truth, the rms of the structure displacement response was 16.42 mm, 15.96 mm, 10.01 mm, and 6.81 mm for the fourth, third, second, and first floor, respectively. The ground truth measurements in this experiment were the calculated displacements from the recorded acceleration data. The experimental data was classified into three groups according to the response rms levels.

The random motions of the structure were captured by the three Kinect sensors. The data collected by the Kinect^{side} and Kinect^{top} were the *pixel*-based measurement. However, the data collected by the Kinect^{front} was the *depth*-based measurement. Figures 19(a), (c), and (e) show the pixel-based measurements for the three different rms levels, and figures 19(b), (d), and (f) illustrate the depth-based measurements for the three different rms levels.

The differences between sampled Kinect and ground truth measurements were compared using PDFs for three different rms amplitude levels (16.42, 10.01, and 6.81 mm). The PDFs were computed based on the normal kernel density estimation. Figure 20 illustrates the comparisons of the PDFs computed with the Kinect and ground truth data for pixel-based measurement. Figure 21 shows the comparisons of the PDFs estimated with the Kinect and ground truth data for depth-based measurement. Three indices were used to quantify the accuracy of random displacement estimations for the Kinect sensor with respect to the ground truth data: the normalized error, the mean error for the two PDFs, and standard deviation error for the two PDFs that are defined as follows:

$$\text{Normalized error} = \frac{\|X_{GT} - X_{Kinect}\|_2}{\|X_{GT}\|_2}, \quad (8)$$

$$\begin{aligned} \text{Mean error (PDFs of displacement or rotation)} \\ = \frac{\mu_{GT} - \mu_{Kinect}}{\mu_{GT}}, \end{aligned} \quad (9)$$

$$\begin{aligned} \text{Standard deviation error (PDFs of displacement} \\ \text{or rotation)} = \frac{\sigma_{GT} - \sigma_{Kinect}}{\sigma_{GT}}, \end{aligned} \quad (10)$$

where $\|\cdot\|_2$ is the Euclidean norm, X_{GT} is the time history of the ground truth displacements, X_{Kinect} is the time history of the Kinect measurements, μ_{GT} and σ_{GT} are the mean and standard deviation for the PDFs of the ground truth data, and μ_{Kinect} and σ_{Kinect} are the mean and standard deviation for the PDFs of the Kinect measurements, respectively. The results of the quantitative error analysis are summarized in table 4.

Overall, the normalized error, mean error, and standard deviation errors decrease when the rms amplitude levels (6.81, 10.01, and 16.42 mm) increase. For the random vibration tests, the normalized error analysis shows relatively

similar performance for the '*pixel*-based' and '*depth*-based' measurements. However, for mean error shows relatively low performance for the depth-based measurements, especially for the small rms excitation level (6.811 mm). As mentioned earlier, the depth measurement error of Kinect v1 is approximately 3 mm for the distance of 1 m [75, 78]. Consequently, when a 6.81 mm rms signal is compared to a larger rms signal, the signal-to-noise ratio of the former one is smaller than the latter one, leading to larger error values. Overall, the error analyses of the random tests are in agreement with the findings in section 6.1 for the harmonic test and [70].

7. Experimental validation for rotational motion

7.1. Static test

The RGB-D camera (Kinect v1) was deployed to quantify the structure rotation under very low frequency, almost (0 Hz) of the testbed structure around the global axis \bar{Z} . To perform the test, the structure was rotated through different angles of rotation (5° , 10° , 15° , 20° , 25° , 30° , 35° , 40° , and 45°) around the global axis \bar{Z} .

The angle of rotation of the testbed structure was measured by the Kinect sensors in three different directions (see figure 4(b)). Since the structure had no vertical motions, for the Kinect^{top} the z -direction motion is steady, while the x - and y -direction of motion varied according to the structure's rotation. This case was called '*pixel*-based' rotational measurements since it was dominated by color pixel coordinates as illustrated in figure 3(a). For the Kinect^{side} and Kinect^{front}, the y -direction measurements with respect to Kinect coordinates are almost constant, and the angle of rotation evaluation was dominated by the x - and z -direction of motions. This case is referred to as '*depth-and-pixel*-based' rotational measurements as shown in figure 3(b).

The angles of rotation estimated based on the Kinect sensor data were compared to the corresponding ground truth measurement. The structure was rotated around its axis of symmetry manually with the use of a protractor. For each test, the *normalized error* between Kinect and ground truth data were computed for the nine markers attached to the structure using the following equation:

$$\text{Normalized error} = \frac{\|\theta_{GT} - \theta_{Kinect}\|_2}{\|\theta_{GT}\|_2}, \quad (11)$$

where $\|\cdot\|_2$ is the Euclidean norm, θ_{GT} is the array corresponding to the ground truth measurements, θ_{Kinect} is the processed structure rotation time history for the Kinect sensor. Since the estimated normalized errors are similar for the front markers (MF1, MF2, MF3, MF4) and side markers (MS1, MS2, MS3, MS4), the mean normalized error for each set of markers are evaluated. Figures 22(a) and (b) show two sets of fitting curves representing the relationship between the accuracy of the structure rotation and angle of rotation for different Kinects.

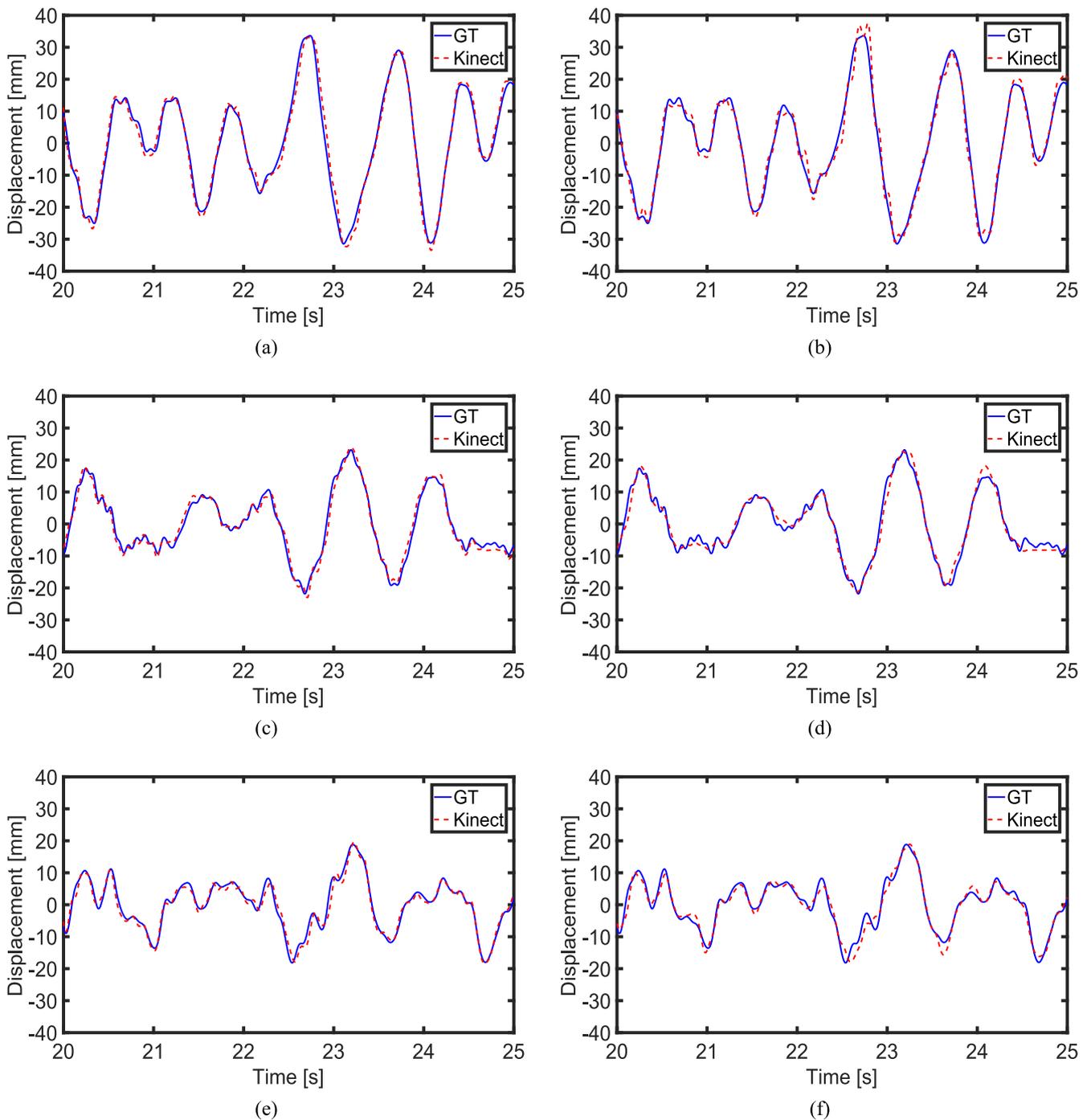


Figure 19. Depth- and pixel-based measurements for three rms levels (a) pixel-based 4th floor (marker MS1; rms = 16.42 mm), (b) depth-based 4th floor (marker MF1; rms = 16.42 mm), (c) pixel-based 2nd floor (marker MS3; rms = 10.01 mm), (d) depth-based 2nd floor (marker MF3; rms = 10.01 mm), (e) pixel-based 1st floor (marker MS4; rms = 6.81 mm), and (f) depth-based 1st floor (marker MF1; rms = 6.81 mm). The GT is the computed displacement from acceleration records.

Figure 22(a) illustrates the error analysis results for the angle of rotation based on *depth-and-pixel* based rotational measurements. For the front and side markers sets, the error analysis is showing the same behavior and trend. For angles of rotation greater than 5° , the normalized errors are approximately 5%. However, for small angles (less than 5°) the normalized error increases due to the precision of Kinect measurements. According to [75], the depth measurement error

of Kinect v1 is approximately 3 mm for a distance of 1 m. This corresponds to an approximate error of a 1.5° for 100 mm marker. Figure 22(b) shows the fitting curves of the mean normalized errors for *depth-and-pixel*-based rotational measurements (mean error for front and side markers sets), and normalized errors for *pixel*-based rotational measurements (marker MT1). For the angle of rotation estimations based on pixel measurements, the normalized errors stay below 3%.

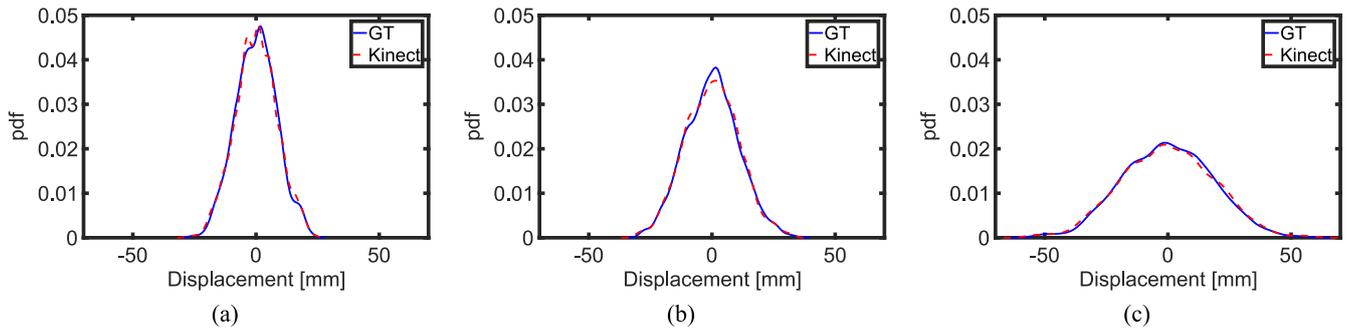


Figure 20. PDFs of the estimated Kinect displacements based on *pixel*-based measurement versus GT (a) rms = 6.81 mm (marker MS4), (b) rms = 10.01 mm (marker MS3), and (c) rms = 16.42 mm (marker MS1). The GT is the computed displacement from acceleration records.

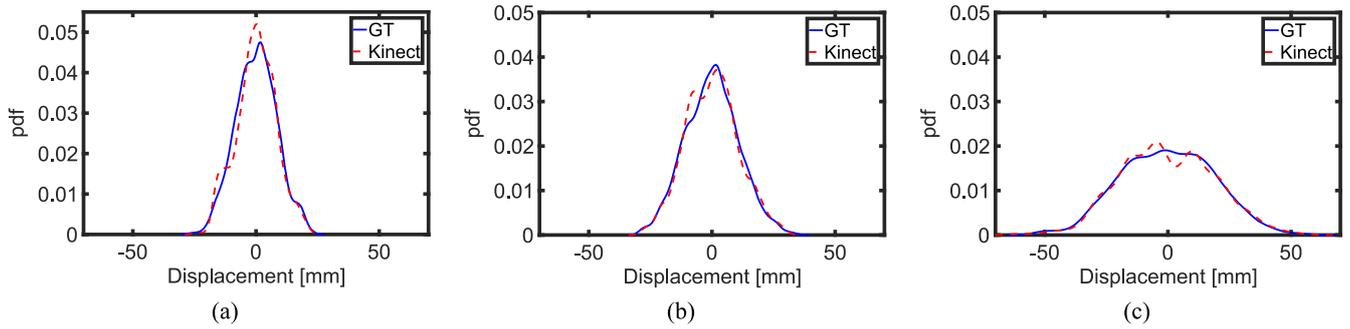


Figure 21. PDFs of the estimated Kinect displacements based on *depth*-based measurement versus GT (a) rms = 6.81 mm (marker MF4), (b) rms = 10.01 mm (marker MF3), and (c) rms = 16.42 mm (marker MF1). The GT is the computed displacement from acceleration records.

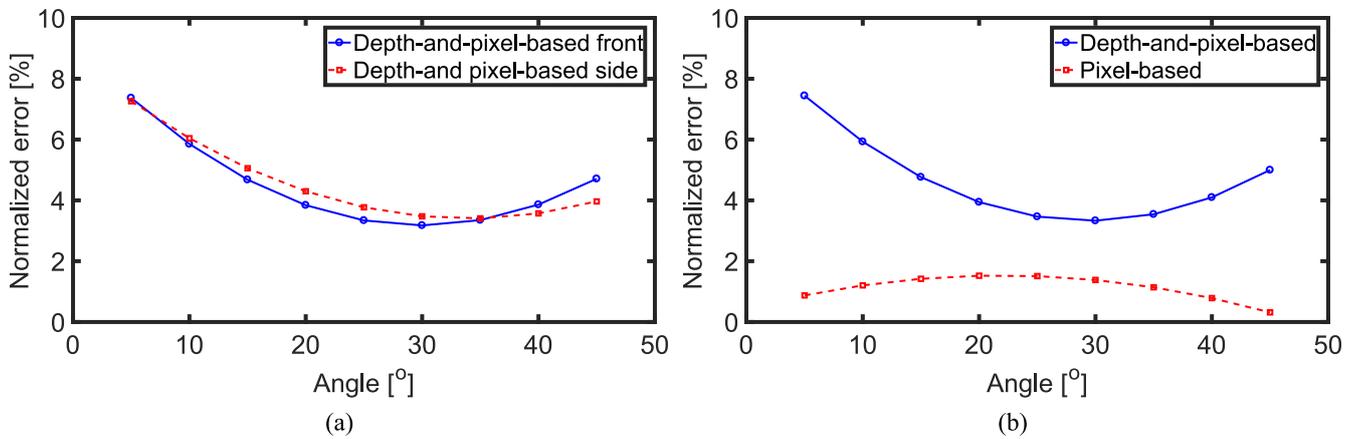


Figure 22. Normalized estimation errors (a) *depth*-and-*pixel*-based rotational measurements based on the data acquired from the front markers set (MF1, MF2, MF3, MF4) and the side markers set (MS1, MS2, MS3, MS4), and (b) *pixel*-based rotational measurements (marker MT1), *depth*-and-*pixel*-based rotational measurements (mean of the side and front markers sets).

Table 4. Comparison of normalized error, mean error, and standard deviation error under depth-based and pixel-based measurements.

	Rms = 6.81 mm		Rms = 10.01 mm		Rms = 16.42 mm	
	Depth (z)	Pixel (x)	Depth (z)	Pixel (x)	Depth (z)	Pixel (x)
Normalized error (%)	21.80	18.74	15.10	15.50	14.46	15.02
Mean error (%)	18.00	5.05	10.89	5.50	5.71	5.02
Std error (%)	1.77	2.02	0.5	1.06	0.80	3.50

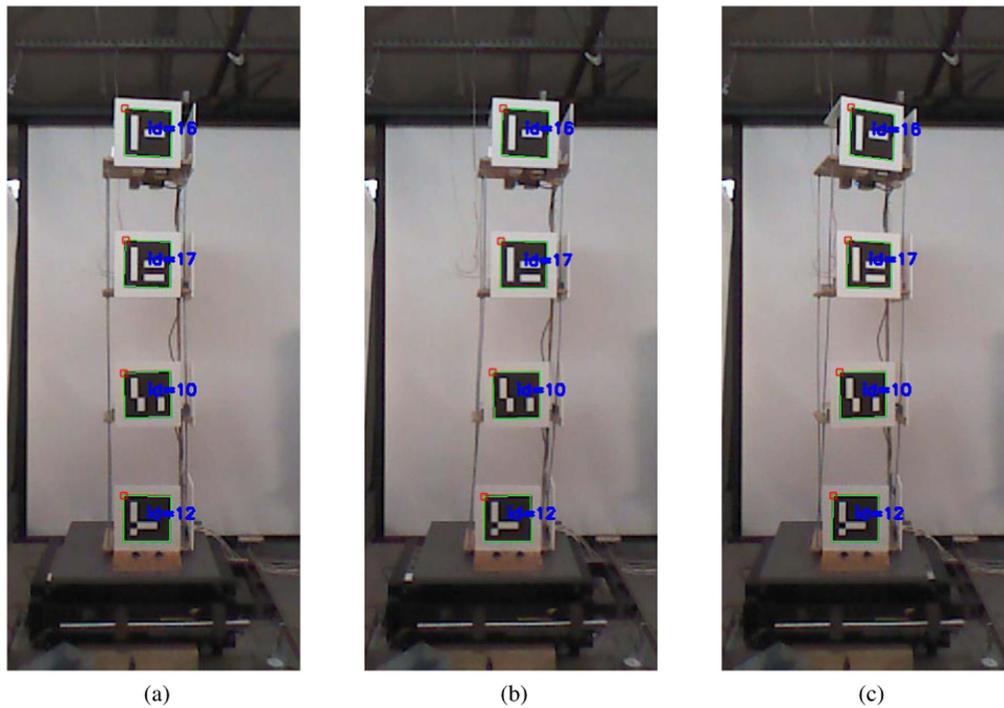


Figure 23. Testbed mode shapes (a) the initial condition, (b) the first mode of response at 1.3 Hz, and (c) the second mode of response at 4.3 Hz. It can be seen that the first mode of response is pure translational and the second mode is a pure torsional.

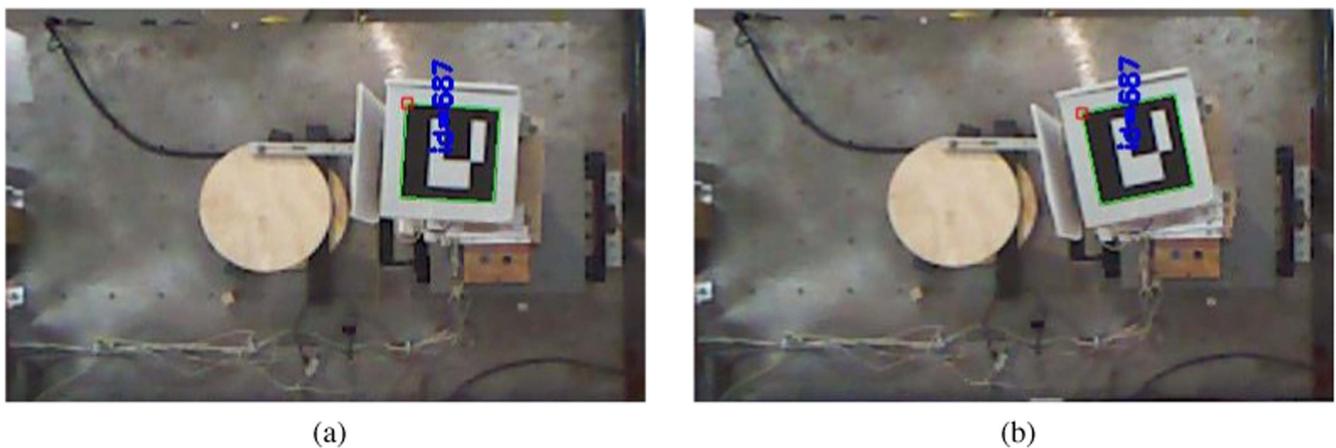


Figure 24. Tracking of top marker (MT1) during torsional test (a) the initial position, (b) the maximum torsion.

7.2. Harmonic test

For the second test, the structure was vibrated through a harmonic excitation with 4.3 Hz frequency and amplitude 13 mm. The test frequency and amplitude were chosen in order to ensure a pure torsional motion for the structure that matched the second mode shape with a maximum angle between (5° and 10°) based on the results in section 7.1. Figure 23 illustrates the difference in motion between the two harmonic tests performed at the designed frequencies 1.3 and 4.3 Hz for almost the same base excitation amplitude of 13 mm. Figure 24 shows two frames extracted from the ONI file for the top marker (MT1) during torsional dynamic test showing to the initial and maximum rotations of the fourth floor.

The torsional deformation field of the testbed structure was recorded by the three Kinect sensors. Since one IMU was used and attached to the fourth floor (location of maximum torsional angle), the emphasis was on the data collected by the three Kinects for the three markers (MF1, MS1, MT1) for the fourth floor (see figure 4). Figure 25 illustrates the difference between the pure translation corresponding to the first mode shape and pure torsional motion reflecting the second mode shape. Thus by tracking marker MS1, based on data collected from Kinect^{side} under 1.3 Hz test, and marker MT1, based on data collected from Kinect^{top} under 4.3 Hz test. The black dots represents the center of the marker.

The ground truth measurement in this test was the calculated angle of rotation from the recorded gyroscope data. To this end, numerical integration was used to obtain the

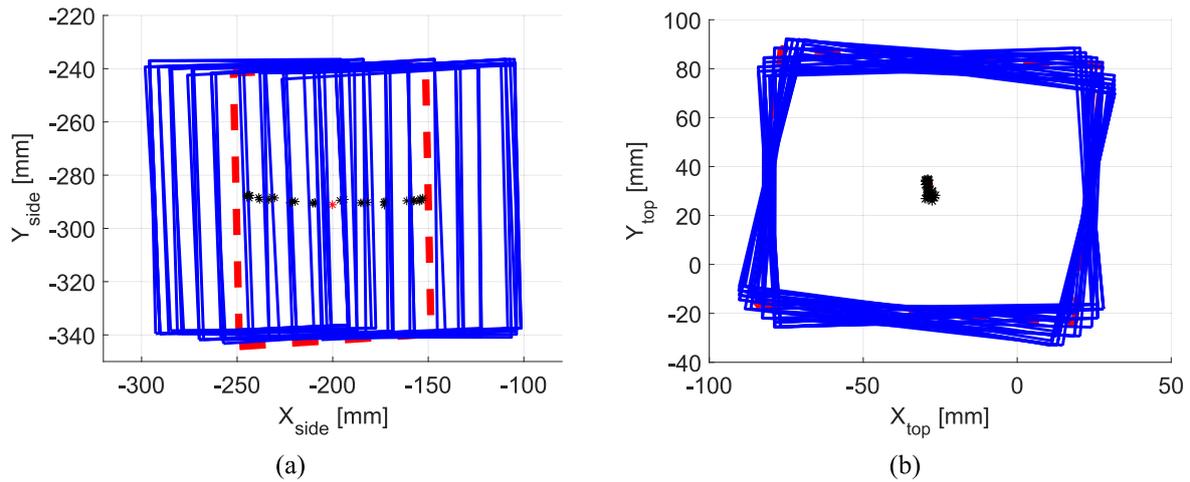


Figure 25. Markers tracking comparison for translational and rotational motion (a) translational motion (marker MS1, 1.3 Hz test), and (b) rotational motion (marker MT1, 4.3 Hz test).

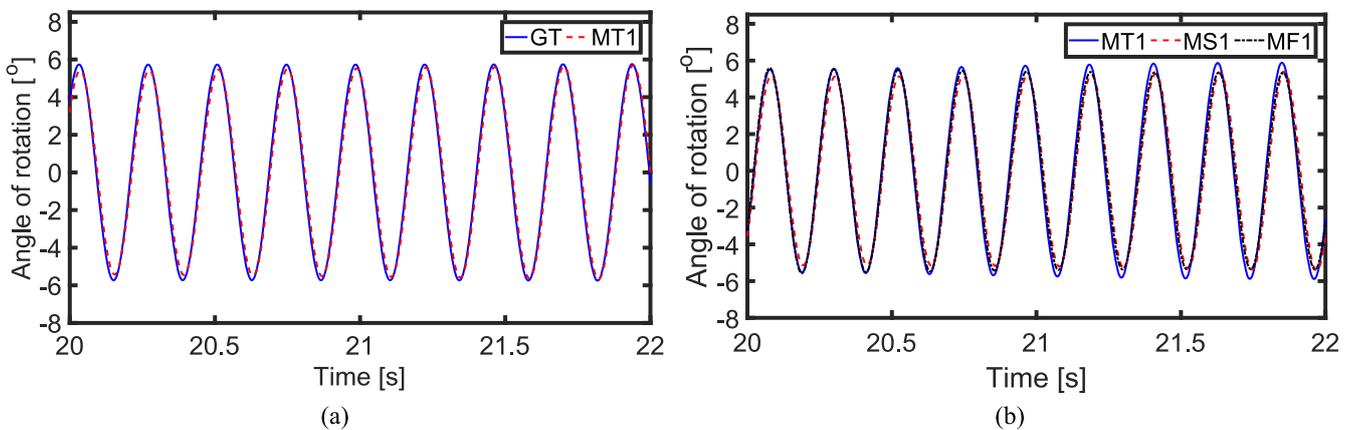


Figure 26. Sample torsional measurements for the structure *fourth* floor (4.3 Hz) (a) marker MT1 (Kinect^{top}), and (b) marker MT1 (Kinect^{top}), MS1 (Kinect^{side}), MF1 (Kinect^{front}); GT is the computed angle of rotation from IMU data

torsional angle time history for the recorded angular velocity. Figure 26(a) illustrates the torsional measurement based on *pixel*-based data (marker MT1) versus the gyroscope data. It can be seen that the two signals almost match where the rms error is 9.71%. The dynamic torsion compared to static rotation has greater error due to the rolling shutter effect explained in section 6. However, for *depth-and-pixel*-based rotational measurements (markers MS1, MF1), the rms error was found to be 15.71% for the discussed case. The *depth-and-pixel*-based data introduces higher rms error since the measurements combine the effect of rolling shutter from *pixel*-based data and the noise effect from the *depth*-based measurement. Overall, the error analyses of the harmonic test agree with what have been discussed in section 7.1 for the static tests. Figure 26(b) shows the dynamic torsional measurements obtained from the three Kinects for the structure's fourth floor.

7.3. Random test

The experimental setup shown in figure 5(b) was used to quantify the structure's rotation around \bar{Z} axis under random

rotational excitation. A separate test was performed for rotation to ensure the possibility of applying different random excitations with different rms amplitudes. In order to get a pure rotational motion, the structure was attached to a rotating table linked to the linear exciter (shaker). The structure was excited by two random signals with two rms levels similar to the displacement test. The rms of structure rotation response was 5.46° and 1.85°, respectively for the fourth floor based on the ground truth data. The ground truth measurement in this test was the computed angle of rotation time history from the recorded angle rate by the IMU. The experimental data could be classified into two groups according to rms levels. The random motions of the structure were captured by the three Kinect sensors in three different orientations. The data collected by Kinect^{side} and Kinect^{front} were *pixel-and-depth*-based measurement. However, the data collected by the Kinect^{top} was the *pixel*-based measurement. Figure 27 illustrates the *pixel-and-depth*-based measurements under the two different rms levels while figure 28 shows the *pixel*-based measurements under the two different rms levels for the testbed's fourth floor.

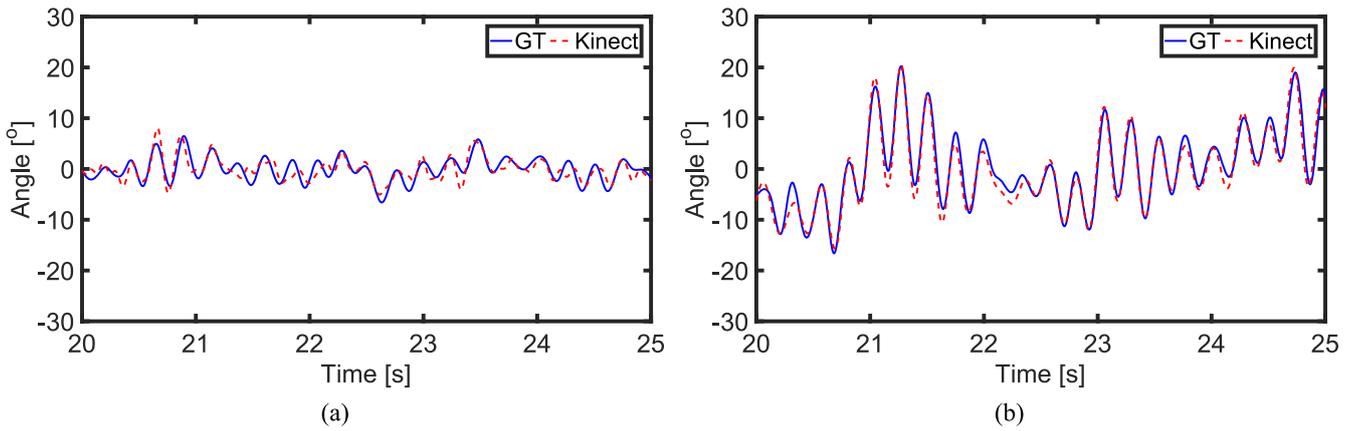


Figure 27. Sample torsional measurements for the structure's *fourth* floor based on marker MF1 (Kinect^{front}) for two rms levels (a) rms = 1.85°, and (b) rms = 5.46°; GT is the computed angle of rotation from IMU data.

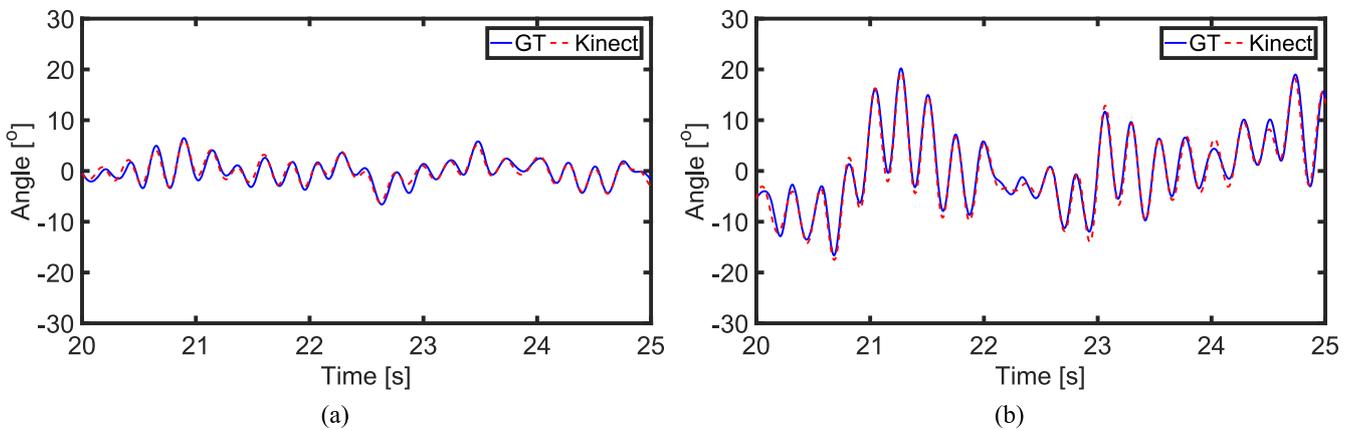


Figure 28. PDFs of the estimated Kinect torsional measurements for the structure's *fourth* floor based on marker MF1 (Kinect^{front}) for two rms levels (a) rms = 1.85°, and (b) rms = 5.46°; GT is the computed angle of rotation from IMU data.

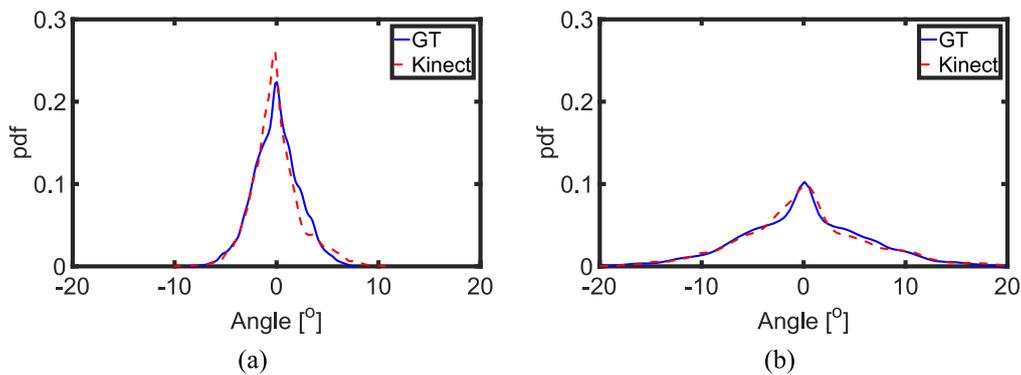


Figure 29. PDFs of the estimated Kinect torsional measurements for the structure's *fourth* floor based on marker MT1 (Kinect^{front}) for two rms levels (a) rms = 1.85°, and (b) rms = 5.46°; GT is the computed angle of rotation from IMU data.

The differences between sampled Kinect and ground truth measurements were compared using PDFs for the two different rms amplitude levels (5.46° and 1.85°). The PDFs were computed based on the normal kernel density estimation. Figure 29 shows the comparisons of the PDFs estimated from the Kinect and ground truth data for *pixel-and-depth*-based measurements. Figure 30 illustrates the comparisons of the PDFs computed with the Kinect and ground truth data for *pixel*-based measurements. Three indices were used to quantify the accuracy of the random

displacement measurements for the Kinect sensor with respect to the ground truth data: the normalized error, the mean error for the two PDFs, and standard deviation error for the two PDFs, which are defined according to equations (8)–(10). The results of the error analysis are summarized in table 5. The normalized error, mean error, and standard deviation error decrease when the rms amplitude levels (1.85° and 5.46°) increase. For the random vibration tests, the normalized error analysis shows relatively poor performance for the *pixel-depth*-based rotational

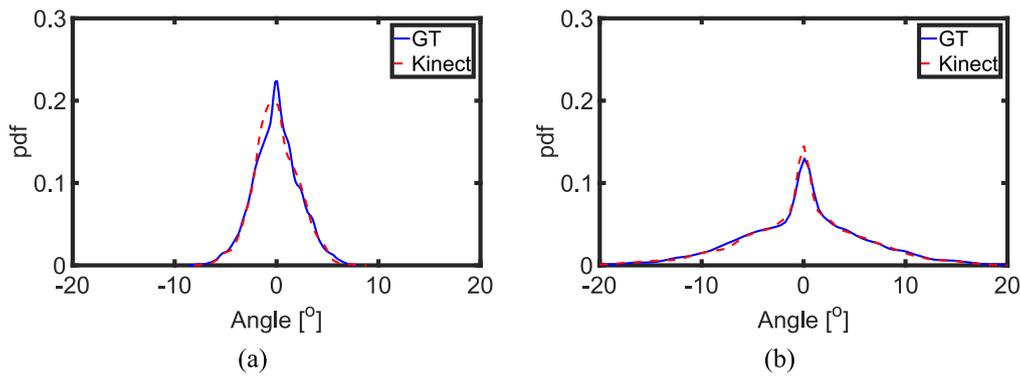


Figure 30. Sample torsional measurements for the structure's *fourth* floor based on marker MT1 (Kinect^{OP}) for two rms levels (a) rms = 1.85°, and (b) rms = 5.46°; GT is the computed angle of rotation from IMU data.

Table 5. Comparison of normalized error, mean error, and standard deviation error under pixel-based and depth-pixel-based torsion measurements.

	Rms = 1.85°		Rms = 5.46°	
	Pixel (x, y)	Pixel-depth (x, z)	Pixel (x, y)	Pixel-depth (x, z)
Normalized error (%)	30.11	65.21	10.45	15.94
Mean error (%)	10.03	25.60	0.75	4.21
Std error (%)	3.55	8.28	1.66	0.75

measurements, especially for the small rms level (error of 65.21% for rms = 1.85°). The error could be mainly caused by the rolling shutter distortion and depth measurements noise, which is mentioned in section 7.1. Comparatively, the Kinect sensor has better performance for the *pixel*-based rotation measurements, particularly for large rms levels (error of 10.45% for rms = 5.46°). Overall, the error analyses of the random test are in agreement with what have been discussed in sections 7.1 and 7.2 for the static and harmonic tests.

8. Summary and conclusions

The comprehensive experimental study reported in this paper presented the feasibility of a vision-based approach for obtaining direct measurements of the absolute displacement and rotational time history at selectable locations of dispersed testbed structure. The measurements were obtained using an inexpensive RGB-D camera (the first generation Kinect as being a representative one). The performance characteristics of the Kinect were evaluated for a class of structural dynamics problems. The Kinect sensor provides the potential of quantifying multi-component displacement fields, including displacement and rotational fields. The field measurements could be deployed in different fields of structural health monitoring applications in a cost-effective way.

The results of the performed calibration studies show that the Kinect sensor is capable of measuring displacements and rotations under different situation involving different amplitude levels, various frequency ranges, and different relative motions between the sensor and the target structure locations. The analysis results indicate that for displacements larger than 10 mm, the estimations have an error of about 5% if the

structure's vibration is parallel to Kinect's depth axis (*z*-axis). However, when the motion is parallel to Kinect's *x*-axis, larger measurement errors were observed due to the rolling shutter distortion of the CMOS sensors that are used by the Kinect's RGB and IR cameras. For rotation angles larger than 5°, if the structure's rotation axis is parallel to Kinect's depth axis (*z*-axis), the measurements have an error of about 5% when the rotation axis is parallel to Kinect's *y*-axis, larger measurement errors were observed, due the combination of noise in the depth data and rolling shutter error in pixel data.

Overall, this study showed that the Kinect sensor is a convenient, feasible, and cost-effective tool to measure the evolving multi-component displacement and rotational fields for structural dynamics problems. However, there are still additional research studies that are needed to extract RGB features without the need to have markers attached to the surface of the target structure. In addition, there is a need to evaluate data fusion concepts regarding the acquired data from different Kinects as one single sensor unit.

Acknowledgments

This research was supported in part by a grant from Qatar University and Qatar Foundation.

ORCID iDs

Mohamed Abdelbarr  <https://orcid.org/0000-0003-4558-7119>

Mohammad R Jahanshahi  <https://orcid.org/0000-0001-6583-3087>

References

- [1] Kao C-Y and Loh C-H 2013 Monitoring of long-term static deformation data of Fei-Tsui arch dam using artificial neural network-based approaches *Struct. Control Health Monit.* **20** 282–303
- [2] Scaioni M, Barazzetti L, Giussani A, Previtali M, Roncoroni F and Alba M I 2014 Photogrammetric techniques for monitoring tunnel deformation *Earth Sci. Inf.* **7** 83–95
- [3] Nayeri R D, Masri S F, Ghanem R G and Nigbor R L 2008 A novel approach for the structural identification and monitoring of a full-scale 17-story building based on ambient vibration measurements *Smart Mater. Struct.* **17** 025006
- [4] Jiang R, Jáuregui D V and White K R 2008 Close-range photogrammetry applications in bridge measurement: literature review *Measurement* **41** 823–34
- [5] Baqersad J, Niezrecki C and Avitabile P 2015 Extracting full-field dynamic strain on a wind turbine rotor subjected to arbitrary excitations using 3D point tracking and a modal expansion technique *J. Sound Vib.* **352** 16–29
- [6] Mazen Wahbeh A, Caffrey J P and Masri S F 2003 A vision-based approach for the direct measurement of displacements in vibrating systems *Smart Mater. Struct.* **12** 785
- [7] Park J-W, Lee J-J, Jung H-J and Myung H 2010 Vision-based displacement measurement method for high-rise building structures using partitioning approach *NDT&E Int.* **43** 642–7
- [8] Kim J-B, Lee E-T, Rahmatalla S and Eun H-C 2013 Non-baseline damage detection based on the deviation of displacement mode shape data *J. Nondestruct. Eval.* **32** 14–24
- [9] Simoen E, Roeck G D and Lombaert G 2015 Dealing with uncertainty in model updating for damage assessment: a review *Mech. Syst. Signal Process.* **56** 123–49
- [10] Friswell M and Mottershead J E 2013 *Finite Element Model Updating in Structural Dynamics* vol 38 (Netherlands: Springer)
- [11] Gkatzogias K I and Kappos A J 2016 Semi-active control systems in bridge engineering: a review of the current state of practice *Struct. Eng. Int.* **26** 290–300
- [12] Doebling S W, Farrar C R, Prime M B and Shevitz D W 1996 Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review *Report LA-13070-MS* Los Alamos National Laboratory, USA
- [13] Farrar C R and Doebling S W 1997 An overview of modal-based damage identification methods *Proc. DAMAS Conf.* pp 269–78
- [14] Soon H, Farrar C R, Hemez F M, Shunk D D, Stinemas D W and Nadler B R 2003 A review of structural health monitoring literature *Report LA-13976-MS* Los Alamos National Laboratory
- [15] Farrar C R and Doebling S W 1999 Damage detection and evaluation II *Modal Analysis and Testing* (Dordrecht: Springer) pp 345–78
- [16] Philips Adewuyi A, Wu Z and Kammrjuman Serker N H M 2009 Assessment of vibration-based damage identification methods using displacement and distributed strain measurements *Struct. Health Monit.* **8** 443–61
- [17] Aktan A E, Catbas F N, Grimmelsman K A and Tsikos C J 2000 Issues in infrastructure health monitoring for management *J. Eng. Mech.* **126** 711–24
- [18] Breuer P, Chmielewski T, Górski P, Konopka E and Tarczyński L 2015 Monitoring horizontal displacements in a vertical profile of a tall industrial chimney using global positioning system technology for detecting dynamic characteristics *Struct. Control Health Monit.* **22** 1002–23
- [19] Yi T-H, Li H-N and Gu M 2013 Recent research and applications of GPS-based monitoring technology for high-rise structures *Struct. Control Health Monit.* **20** 649–70
- [20] Im S B, Hurlbauss S and Kang Y J 2011 Summary review of GPS technology for structural health monitoring *J. Struct. Eng.* **139** 1653–64
- [21] Çelebi M and Sanli A 2002 GPS in pioneering dynamic monitoring of long-period structures *Earthq. Spectra* **18** 47–61
- [22] Breuer P, Chmielewski T, Górski P and Konopka E 2002 Application of GPS technology to measurements of displacements of high-rise structures due to weak winds *J. Wind Eng. Ind. Aerodyn.* **90** 223–30
- [23] Lynch J P and Loh K J 2006 A summary review of wireless sensors and sensor networks for structural health monitoring *Shock Vib. Digest* **38** 91–130
- [24] Park J W, Sim S H, Jung H J and Billie S 2013 Development of a wireless displacement measurement system using acceleration responses *Sensors* **13** 8377–92
- [25] Painter C C and Shkel A M 2003 Detection of orientation and predicted performance of a mems absolute angle measuring gyroscope *Proc. 4th Int. Workshop on Structural Health Monitoring* vol 1, pp 1011–8
- [26] Stančin S and Tomažič S 2011 Angle estimation of simultaneous orthogonal rotations from 3D gyroscope measurements *Sensors* **11** 8536–49
- [27] Kijewski-Correa T and Kochly M 2007 Monitoring the wind-induced response of tall buildings: GPS performance and the issue of multipath effects *J. Wind Eng. Ind. Aerodyn.* **95** 1176–98
- [28] Smyth A and Pei J 2000 Integration of measured response signals for nonlinear structural health monitoring *Proc. 3rd US–Japan Workshop on Nonlinear System Identification and Health Monitoring*
- [29] Smyth A 2004 The potential of GPS and other displacement sensing for enhancing acceleration sensor monitoring array data by solving low frequency integration problems *Proc. 2nd Int. Conf. on Bridge Maintenance, Safety, Management and Cost, IABMAS04 (Kyoto)* 533
- [30] Hou X, Yang X and Huang Q 2005 Using inclinometers to measure bridge deflection *J. Bridge Eng.* **10** 564–9
- [31] Park S, Horowitz R and Tan C-W 2008 Dynamics and control of a MEMS angle measuring gyroscope *Sensors Actuators A* **144** 56–63
- [32] Park S 2010 Adaptive control of a vibratory angle measuring gyroscope *Sensors* **10** 8478–90
- [33] Petrie G and Toth C K 2009 Terrestrial laser scanners *Topographic Laser Ranging and Scanning Principles and Processing* pp 87–128
- [34] OuYang D and Feng H-Y 2005 On the normal vector estimation for point cloud data from smooth surfaces *Comput.-Aided Des.* **37** 1071–79
- [35] Kim J, Kim K and Sohn H 2014 Autonomous dynamic displacement estimation from data fusion of acceleration and intermittent displacement measurements *Mech. Syst. Signal Process.* **42** 194–205
- [36] Olsen M J, Kuester F, Chang B J and Hutchinson T C 2009 Terrestrial laser scanning-based structural damage assessment *J. Comput. Civil Eng.* **24** 264–72
- [37] Park H S, Lee H M, Adeli H and Lee I 2007 A new approach for health monitoring of structures: terrestrial laser scanning *Comput.-Aided Civil Infrastruct. Eng.* **22** 19–30
- [38] Gordon S J and Lichti D D 2007 Modeling terrestrial laser scanner data for precise structural deformation measurement *J. Surv. Eng.* **133** 72–80
- [39] Kim K and Kim J 2015 Dynamic displacement measurement of a vibratory object using a terrestrial laser scanner *Meas. Sci. Technol.* **26** 045002
- [40] Yang Y, Dorn C, Mancini T, Talken Z, Kenyon G, Farrar C and Mascareñas D 2017 Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification *Mech. Syst. Signal Process.* **85** 567–90

- [41] Yang Y, Dorn C, Mancini T, Talken Z, Nagarajaiah S, Kenyon G, Farrar C and Mascareñas D 2017 Blind identification of full-field vibration modes of output-only structures from uniformly-sampled, possibly temporally-aliased (sub-nyquist), video measurements *J. Sound Vib.* **390** 232–56
- [42] Chen J G, Wadhwa N, Cha Y-J, Durand F, Freeman W T and Buyukozturk O 2015 Modal identification of simple structures with high-speed video using motion magnification *J. Sound Vib.* **345** 58–71
- [43] Bartilson D T, Wiegand K T and Hurlebaus S 2015 Targetless computer vision for traffic signal structure vibration studies *Mech. Syst. Signal Process.* **60–61** 571–82
- [44] Chen J G, Wadhwa N, Y-j Cha, Durand F and Freeman W T 2015 Modal identification of simple structures with high-speed video using motion magnification *J. Sound Vib.* **345** 58–71
- [45] Feng D and Feng M Q 2016 Vision-based multipoint displacement measurement for structural health monitoring *Struct. Control Health Monit.* **23** 876–90
- [46] Rajaram S et al 2017 Full-field deformation measurements during seismic loading of masonry buildings *Struct. Control Health Monit.* **24** e1903
- [47] Jeon H, Myeong W, Shin J U, Park J W, Jung H J and Myung H 2014 Experimental validation of visually servoed paired structured light system (ViSP) for structural displacement monitoring *IEEE/ASME Trans. Mechatronics* **19** 1603–11
- [48] Wu L-J, Casciati F and Casciati S 2014 Dynamic testing of a laboratory model via vision-based sensing *Eng. Struct.* **60** 113–25
- [49] Detchev I, Habib A and El-Badry M 2013 Dynamic beam deformation measurements with off-the-shelf digital cameras *J. Appl. Geodesy* **7** 147–57
- [50] Jeon H, Bang Y and Myung H 2011 A paired visual servoing system for 6-DOF displacement measurement of structures *Smart Mater. Struct.* **20** 1–16
- [51] Lee J-H, Ho H-N, Shinozuka M and Lee J-J 2012 An advanced vision-based system for real-time displacement measurement of high-rise buildings *Smart Mater. Struct.* **21** 125019
- [52] Park S W, Park H S, Kim J H and Adeli H 2015 3d displacement measurement model for health monitoring of structures using a motion capture system *Measurement* **59** 352–62
- [53] Ren Z, Yuan J, Meng J and Zhang Z 2013 Robust part-based hand gesture recognition using kinect sensor *IEEE Trans. Multimedia* **15** 1110–20
- [54] Patsadu O, Nukoolkit C and Watanapa B 2012 Human gesture recognition using kinect camera *Proc. 2012 9th Int. Joint Conf. on Computer Science and Software Engineering (JCSSE)* pp 28–32
- [55] Biswas K K and Basu S K 2011 Gesture recognition using microsoft kinect® *Proc. 2011 5th Int. Conf. on Automation, Robotics and Applications (ICARA)* pp 100–3
- [56] Meng M, Fallavollita P, Blum T, Eck U, Sandor C, Weidert S, Waschke J and Navab N 2013 Kinect for interactive AR anatomy learning *Proc. 2013 IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR)* pp 277–8
- [57] Vera L, Gimeno J, Coma I and Fernández M 2011 Augmented mirror: interactive augmented reality system based on Kinect *Proc. 13th IFIP TC 13 Int. Conf. on Human-Computer Interaction-INTERACT 2011* pp 483–6
- [58] Correa D S O, Sciotti D F, Prado M G, Sales D O, Wolf D F and Osorio F S 2012 Mobile robots navigation in indoor environments using Kinect sensor *Proc. 2012 2nd Brazilian Conf. on Critical Embedded Systems (CBSEC)* pp 36–41
- [59] El-laithy R A, Huang J and Yeh M 2012 Study on the use of Microsoft Kinect for robotics applications *Proc. 2012 IEEE/ION Position, Location and Navigation Symposium* pp 1280–8
- [60] Tahavori F, Alnowami M, Jones J, Elangovan P, Donovan E and Wells K 2013 Assessment of Microsoft Kinect technology (Kinect for Xbox and Kinect for Windows) for patient monitoring during external beam radiotherapy *Proc. 2013 IEEE Nuclear Science Symp. and Medical Imaging Conference (NSS/MIC)* pp 1–5
- [61] Alnowami M, Alnowami B, Tahavori F, Copland M and Wells K 2012 A quantitative assessment of using the Kinect for Xbox360 for respiratory surface motion tracking *Proc. SPIE* **8316** 83161T
- [62] Tortini R, Bonali F L, Corazzato C, Carn S A and Tibaldi A 2014 An innovative application of the Kinect in Earth sciences: quantifying deformation in analogue modelling of volcanoes *Terra Nova* **26** 273–81
- [63] Qi X, Lichti D, El-Badry M, Chow J and Ang K 2014 Vertical dynamic deflection measurement in concrete beams with the Microsoft Kinect *Sensors* **14** 3293–307
- [64] Mallick T, Das P P and Majumdar A K 2014 Characterizations of noise in Kinect depth images: a review *IEEE Sens. J.* **14** 1731–40
- [65] Kadambi A, Bhandari A and Raskar R 2014 3D depth cameras in vision: benefits and limitations of the hardware *Computer Vision and Machine Learning with RGB-D Sensors* pp 3–26
- [66] Sarbolandi H, Lefloch D and Kolb A 2015 Kinect range sensing: structured-light versus time-of-flight Kinect *Comput. Vis. Image Underst.* **139** 1–20
- [67] Moss R, Yuan P, Bai X, Quesada E, Sudharsanan R, Stann B L, Dammann J F, Giza M M and Lawler W B 2012 Low-cost compact MEMS scanning Ladar system for robotic applications *Proc. SPIE* **8379** 837–903
- [68] Matsuda N, Cossairt O and Gupta M 2015 MC3D: Motion contrast 3D scanning *Proc. 2015 IEEE Int. Conf. on Computational Photography (ICCP)* pp 1–10
- [69] Kadambi A, Taamazyan V, Shi B and Raskar R 2015 Polarized 3D: high-quality depth sensing with polarization cues *Proc. IEEE Int. Conf. on Computer Vision* pp 3370–8
- [70] Chen Y L, Abdelbarr M, Jahanshahi M R and Masri S F 2017 Color and depth data fusion using an RGB-D sensor for inexpensive and contactless dynamic displacement-field measurement *Struct. Control Health Monit.* **24** e2000
- [71] Bouguet J-Y 2015 Camera calibration toolbox for Matlab (Accessed: 12 October 2017)
- [72] Chen Y L, Jahanshahi M R, Manjunatha P, Gan W, Abdelbarr M, Masri S F, Becerik-Gerber B and Caffrey J P 2016 Inexpensive multimodal sensor fusion system for autonomous data acquisition of road surface conditions *IEEE Sens. J.* **16** 7731–43
- [73] Garrido-Jurado S, Muñoz-Salinas R, Madrid-Cuevas F J and Marín-Jiménez M J 2014 Automatic generation and detection of highly reliable fiducial markers under occlusion *Pattern Recognit.* **47** 2280–92
- [74] Bradski G and Kaehler A 2008 *Learning OpenCV: Computer Vision with the OpenCV Library* (Sebastopol, CA: O'Reilly Media, Inc.)
- [75] Khoshelham K and Elberink S O 2012 Accuracy and resolution of Kinect depth data for indoor mapping applications *Sensors* **12** 1437–54
- [76] Olivares A, Olivares G, Gorriiz J M and Ramirez J 2009 High-efficiency low-cost accelerometer-aided gyroscope calibration *Proc. Int. Conf. on Test and Measurement, 2009. ICTM'09* vol 1, pp 354–60
- [77] Ait-Aider O, Andreff N, Lavest J M and Martinet P 2006 Simultaneous object pose and velocity computation using a single view from a rolling shutter camera *Proc. European Conf. on Computer Vision-ECCV 2006* vol 2, pp 56–68
- [78] Smisek J, Jancosek M and Pajdla T *Consumer Depth Cameras for Computer Vision* ch 1 (London: Springer) 167