applied optics

3D shape, deformation, and vibration measurements using infrared Kinect sensors and digital image correlation

Hieu Nguyen,[1] Zhaoyang Wang,[1,*] Patrick Jones,[1] and Bing Zhao[2]

[1]Department of Mechanical Engineering, The Catholic University of America, Washington, DC 20064, USA
[2]Optic Fringe Corp., 8 Cobblestone Way, North Billerica, Massachusetts 01862, USA
*Corresponding author: wangz@cua.edu

Received 6 June 2017; revised 25 September 2017; accepted 29 September 2017; posted 2 October 2017 (Doc. ID 297593); published 8 November 2017

Consumer-grade red-green-blue and depth (RGB-D) sensors, such as the Microsoft Kinect and the Asus Xtion, are attractive devices due to their low cost and robustness for real-time sensing of depth information. These devices provide the depth information by detecting the correspondences between the captured infrared (IR) image and the initial image sent to the IR projector, and their essential limitation is the low accuracy of 3D shape reconstruction. In this paper, an effective technique that employs the Kinect sensors for accurate 3D shape, deformation, and vibration measurements is introduced. The technique involves using the RGB-D sensors, an accurate camera calibration scheme, and area- and feature-based image-matching algorithms. The IR speckle pattern projected from the Kinect projector considerably facilitates the digital image correlation analysis in the regions of interest with enhanced accuracy. A number of experiments have been carried out to demonstrate the validity and effectiveness of the proposed technique and approach. It is shown that the technique can yield measurement accuracy at the 10 μm level for a typical field of view. The real-time capturing speed of 30 frames per second makes the proposed technique suitable for certain motion and vibration measurements, such as non-contact monitoring of respiration and heartbeat rates.

OCIS codes: (110.6880) Three-dimensional image acquisition; (150.1135) Algorithms; (150.6910) Three-dimensional sensing; (170.3010) Image reconstruction techniques; (150.0155) Machine vision optics.

https://doi.org/10.1364/AO.56.009030

1. INTRODUCTION

In the past decade, the consumer grade red-green-blue and depth (RGB-D) sensors have made a dramatic impact in several research areas, such as computer vision, robotics, computer gaming, surveillance, and forensics [1–5]. Microsoft Kinect, Asus Xtion, Intel RealSense, and Apple's TrueDepth camera on iPhone X are some recent and representative developments in consumer range-sensing technology, and all have received tremendous attention due to their low cost, flexibility of use, and fast speed of sensing.

The Microsoft Kinect sensor is a RGB-D sensor consisting of an infrared-radiation (IR) projector, an IR camera, and a regular RGB camera, which are used to capture the depth and color images at real-time speed [6]. During sensing, the IR projector emits a speckle pattern onto a scene in which objects of interest are positioned. The distorted IR pattern is then acquired by the IR camera and subsequently analyzed using a stereo imaging algorithm, which is based on a PrimeSense patent [7]. Technically, the structured speckle pattern helps establish correspondences between the captured image and the initial speckle image stored in the device, with which the depth information can be retrieved using an algorithm based on geometric triangulation. The Microsoft Kinect sensor was primarily designed for human-computer interaction in computer gaming and virtual reality applications. However, the characteristic raw data that can be collected from the sensor has attracted researchers from a variety of fields to use them for broader applications.

While the sensor is a great development in 3D imaging technology, there are several aspects in which the performance of the device is limited. Some of these aspects include: the limited working range, the inability to achieve desired outdoor performance, the incapability of handling specular and transparent objects, and the relatively low accuracy compared with many other 3D imaging techniques. Particularly, some studies [8–10] reveal that the measurement accuracy of the Kinect sensor is around 0.2–1 pixels, which is in practice related to many factors, such as the measurement distance. In recent years, many research

efforts have been made to cope with the aforementioned issues, with a focus on improving the 3D measurement accuracy and speed [9–12]. For example, one of the notable studies [13] describes treating the IR camera as a regular camera to capture images. It proposes a cross-model stereo-vision approach in which the IR projector is blocked in an attempt to address the problems associated with reflective and transparent objects. While attempting to correspond the RGB and IR images, the accuracy of the stereo matching decreases due to the intensity difference between the two images. Some other studies describe a RGB-D mapping system by taking advantage of combining the color and depth images [14,15]. The technique allows the feature points in the RGB image to be extracted and then located in the depth image to estimate their transformations for improving the 3D imaging reliability in robot localization and navigation. Recently, Alhwarin *et al.* [16] proposed a technique of using the IR cameras from two sets of the Kinect sensors, together with the use of known dot patterns, to enhance the image-matching process. Their method can attenuate the reflection problem of shiny objects, but the measurement accuracy remains unchanged.

This paper presents a simple yet effective technique to achieve accurate 3D imaging of objects based on using the IR camera and IR projector of the Microsoft Kinect sensor. For the hardware setup, the proposed technique uses two Kinect devices arranged in a stereo rig orientation. Instead of using the images provided by the RGB cameras, the system takes advantage of the narrow-band IR cameras. Figure 1 illustrates an example of the proposed experimental system and the images captured by the Kinect device with and without IR patterns from two different views. For accurate 3D measurements, every region of interest (ROI) in the captured 2D images must contain sufficient intensity variations to ensure that they can be uniquely and accurately identified for the detection of pixel correspondences. Such intensity variations are provided by the pattern of speckles, which are projected from the IR projector onto the scene containing objects. Because the technique needs only a single IR projector, and using two IR projectors can lead to undesired interference between the two IR patterns, one of the Kinect IR projectors is turned off. The IR pattern allows the proposed technique to require neither projecting visible light nor fabricating artificial speckle or texture patterns on the surfaces of objects. Particularly, the IR patterns can help accurately measure the depth information in the textureless regions where the ordinary RGB cameras will fail yielding faithful results. The real surface textures of the objects, which are acquired by the RGB camera(s), can be later projected on the reconstructed 3D model if necessary. In addition to taking advantage of the IR cameras and IR projector, the proposed technique also consists of two essential components: accurate camera calibration and advanced stereo matching schemes. The camera calibration process uses a sophisticated lens distortion model and a bundle adjustment scheme to precisely describe the relation between the 3D world coordinates of a point and its corresponding locations in the camera images. The stereo matching process involves both feature-based and area-based matching algorithms for practical applications in which automated full-field analysis in the presence of geometric discontinuities can be feasible.

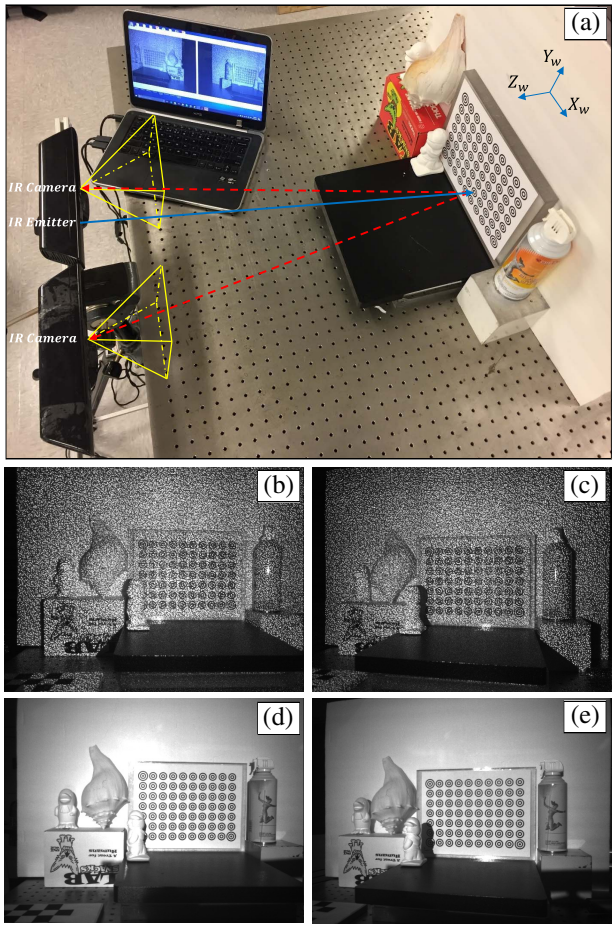

Fig. 1. Example of the proposed system and the images captured by the left and right IR cameras. (a) Experimental setup; (b) and (c) images captured with IR illumination; and (d) and (e) images captured without IR illumination.

The paper is organized as follows: Section 2 describes the fundamental concept of the camera calibration and how the 3D coordinates of points are determined; Section 3 presents the image matching algorithms that are adopted to find the same points in two different images; and Section 4 demonstrates a number of experiments to validate the proposed approach. A summary with a brief discussion is outlined in the last section.

2. GEOMETRICAL MODEL AND CAMERA CALIBRATION

Figure 2 is a schematic of the stereo-vision technique. In the figure, O_l and O_r are the optical centers of the left and right cameras, and an arbitrary physical point P is imaged as point P_l and point P_r in the image planes of the left and right cameras, respectively. Retrieving the 3D information of point P with relation to points P_l and P_r in a world coordinate system requires two primary crucial steps: (1) Calibrate the cameras in advance to get the camera parameters, which yield geometry information of the stereo-vision system. (2) Perform stereo matching of

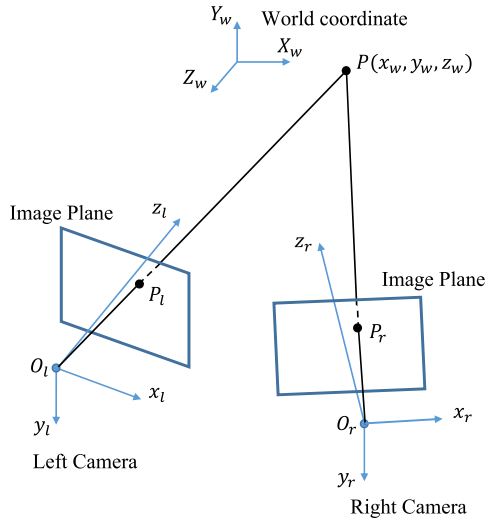

Fig. 2. Schematic of the stereo-vision imaging.

points to link the same physical points from two images, which are essential for the 3D coordinate determination.

The camera calibration involves a description of the relation between the 3D world coordinate of a point and its corresponding location in the image plane. In Fig. 2, an arbitrary point (x_w, y_w, z_w) in the world coordinate system can be expressed as (x_l, y_l, z_l) in a left-camera coordinate system by using the following equation:

$$\begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} = \begin{bmatrix} R^l_{11} & R^l_{12} & R^l_{13} & T^l_1 \\ R^l_{21} & R^l_{22} & R^l_{23} & T^l_2 \\ R^l_{31} & R^l_{32} & R^l_{33} & T^l_3 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = [\boldsymbol{R}^l \quad \boldsymbol{T}^l] \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}, \quad (1)$$

where $\boldsymbol{R}^l$ and $\boldsymbol{T}^l$ components indicate the rotation and translation parameters that transform the world coordinate system to the left camera coordinate system, and they are also called camera extrinsic parameters. In the imaging plane of the left camera, the pixel location (u_l, v_l) of the aforementioned point can be described from a pinhole model as

$$\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = \frac{1}{z_l} \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{ln} \\ y_{ln} \\ 1 \end{bmatrix}, \quad (2)$$

where α and β are the horizontal and vertical distances from the lens to the image plane in pixel unit, γ is a skew factor, and (u_0, v_0) are the coordinates of the principal point. These parameters are often called camera intrinsic parameters. In Eq. (2), $x_{ln} = x_l/z_l$, and $y_{ln} = y_l/z_l$.

Because the camera lens has an optical distortion effect, a lens distortion model is added to the camera calibration. The actual pixel location $(\tilde{u_l}, \tilde{v_l})$ in the captured digital image can be modeled as

$$\begin{bmatrix} \tilde{u}_l \\ \tilde{v}_l \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{x}_{ln} \\ \tilde{y}_{ln} \\ 1 \end{bmatrix}, \quad (3)$$

with

$$\begin{aligned} \tilde{x}_{ln} &= (1 + k_0 r^2 + k_1 r^4 + k_2 r^6 + k_3 r^8 + k_4 r^{10}) x_{ln} \\ &\quad + (k_5 + k_7 r^2) r^2 + (k_9 + k_{11} r^2)(r^2 + 2x_{ln}^2), \\ \tilde{y}_{ln} &= (1 + k_0 r^2 + k_1 r^4 + k_2 r^6 + k_3 r^8 + k_4 r^{10}) y_{ln} \\ &\quad + (k_6 + k_8 r^2) r^2 + (k_{10} + k_{12} r^2)(r^2 + 2y_{ln}^2), \\ r^2 &= x_{ln}^2 + y_{ln}^2, \end{aligned} \quad (4)$$

where $(k_0, \ldots, k_4)$, $(k_5, \ldots, k_8)$, and $(k_9, \ldots, k_{12})$ represent radial, prism, and tangential distortion coefficients, respectively.

By using a camera calibration target, which is typically a flat planar board with uniformly spaced checker or circle patterns, the camera intrinsic and extrinsic parameters as well as the lens distortion parameters can be determined from a bundle-adjustment-based camera calibration process. The relevant algorithms can be found in Refs. [17,18]. For the proposed system, it is noteworthy that the IR projector is turned off during camera calibration so that the speckle pattern will not appear in the captured calibration images. Instead, a conventional halogen lamp serves as the light source during camera calibration.

With a stereo-vision system that contains two separate cameras, Eq. (1) gives the following equation for a typical point (x_w, y_w, z_w) in the world coordinate system:

$$\begin{bmatrix} x_{ln} z_l \\ y_{ln} z_l \\ z_l \end{bmatrix} = \begin{bmatrix} R^l_{11} & R^l_{12} & R^l_{13} & T^l_1 \\ R^l_{21} & R^l_{22} & R^l_{23} & T^l_2 \\ R^l_{31} & R^l_{32} & R^l_{33} & T^l_3 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} x_{rn} z_r \\ y_{rn} z_r \\ z_r \end{bmatrix} = \begin{bmatrix} R^r_{11} & R^r_{12} & R^r_{13} & T^r_1 \\ R^r_{21} & R^r_{22} & R^r_{23} & T^r_2 \\ R^r_{31} & R^r_{32} & R^r_{33} & T^r_3 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}, \quad (5)$$

where the terms with a l and a r symbol in their super- and subscripts are associated with the left and right cameras, respectively. In Eq. (5), the extrinsic parameters R^l, T^l, R^r, and T^r are acquired from camera calibration in advance, and x_{ln} and y_{ln} (similarly x_{rn} and y_{rn}) can be obtained from the captured image using Eqs. (3) and (4). Consequently, there are totally six separate equations, as shown in Eq. (5) and five unknowns: x_w, y_w, z_w, z_l, and z_r. By eliminating z_l and z_r, Eq. (5) gives four separate equations and three unknowns: x_w, y_w, and z_w, which can be determined from a linear least-squares solution.

3. IMAGE MATCHING

A. Area-Based Matching: Digital Image Correlation

The stereo-vision-based technique requires finding the points in the left image with their corresponding positions in the right image with subpixel accuracies. This process is called image registration or matching in the computer vision field and digital image correlation (DIC) in the fields of optics and mechanics. To acquire accurate full-field results, the image matching algorithm generally uses an area-based matching scheme to detect the best matching between two sets of pixels (named subsets) centered at each feature pixel to be interrogated. To establish a metric for finding the similarities between the reference subset

in one image and the target subset in another image, the following cost function or correlation criterion can be used [19]:

$$C = \frac{1}{N^2}\sum_{i=1}^{N}[af(\tilde{u}_{li}, \tilde{v}_{li}) + b - g(\tilde{u}_{ri}, \tilde{v}_{ri})]^2, \quad (6)$$

where a is a scale factor, b is an offset of intensity, and $f(\tilde{u}_{li}, \tilde{v}_{li})$ and $g(\tilde{u}_{ri}, \tilde{v}_{ri})$ indicate the intensity values at the ith pixel in the reference subset and the potential matching pixel in the target subset, respectively. The basic principle of the correlation analysis is to minimize the cost coefficient C in Eq. (6). For a representative pixel $P_{l0}(\tilde{u}_{l0}, \tilde{v}_{l0})$ in the reference image to be matched, a square reference region named a subset with the size of $N = (2M + 1) \times (2M + 1)$ pixels centered at the interrogated point P_{l0} is chosen and then used to match with the corresponding subset in the target image, which is normally of irregular shape. Denoting (ξ, η) as the translation or shift amount between the centers P_{l0} and P_{r0} of the two matching subset patterns, a commonly used displacement mapping function for the reference and target subsets can be expressed as [20]

$$\begin{aligned}\tilde{u}_{ri} &= \tilde{u}_{li} + \xi + \xi_u\Delta_u + \xi_v\Delta_v + \xi_{uu}\Delta_u^2 + \xi_{vv}\Delta_v^2 + \xi_{uv}\Delta_u\Delta_v \\ \tilde{v}_{ri} &= \tilde{v}_{li} + \eta + \eta_u\Delta_u + \eta_v\Delta_v + \eta_{uu}\Delta_u^2 + \eta_{vv}\Delta_v^2 + \eta_{uv}\Delta_u\Delta_v,\end{aligned} \quad (7)$$

where $i = 1, 2, \ldots, N$; $\Delta_u = \tilde{u}_{li} - \tilde{u}_{l0}$; $\Delta_v = \tilde{v}_{li} - \tilde{v}_{l0}$; and ξ_u, ξ_v, ξ_{uu}, ξ_{vv}, ξ_{uv}, η_u, η_v, η_{uu}, η_{vv}, and η_{uv} are the coefficients of the displacement mapping function. To determine all the 12 unknowns in the mapping function [ξ, η, and the other 10 coefficients in Eq. (7)], as well as the scale and offset parameters (a and b) involved in Eq. (6), the matching technique often employs an iterative algorithm, such as the Newton–Raphson or the Levenberg–Marquardt [21] method to carry out the matching optimization. In the iteration process, an interpolation operation, such as B-spline interpolation, must be carried out to get the intensity values at subpixel locations [22]. Moreover, in order to speed up the processing speed in practice, the DIC matching can run on pixels at a step size larger than one, and the full-field matching results will then be obtained by data interpolation from those grid pixels. The step size is usually set to 2–5, and can be set to larger for a faster processing speed with reduced analysis accuracy.

B. Feature-Based Matching for Initial Guess

The iterative algorithm is capable of performing the image-matching process with high accuracy at a very fast speed upon a reasonably good initial guess for the unknown transformation parameters, which are mainly the low-order terms ξ, η, ξ_u, ξ_v, η_u, and η_v in Eq. (7). In the general image-matching process, such an initial guess can be carried out by using a manual way of selecting three pairs of matching points from each ROI in the reference and target images, or by using an automatic full-field scanning process in the case of small shape change of the target subset with respect to the reference subset. The manual initial-guess method hampers the automatic nature of the DIC process, and the conventional automatic initial-guess method is highly limited when dealing with geometric discontinuities. In order to cope with this problem, a feature-based matching scheme is employed to conduct the initial guess. The featured-based matching can extract many features of interest in the images, such as the edges of local patterns, and build a unique descriptor for each feature point. The feature points between the reference and target images can then be compared by using their descriptors to find the best matching pairs. The feature-based matching can detect matching over the entire images, so it can easily cope with the issue of geometric discontinuities, including occlusions and shadows.

The most well-known feature-based matching method is the scale-invariant feature transform (SIFT) algorithm. It is known to be very robust at handling variances in lighting, image rotation, translation, and scaling. The SIFT matching scheme, including the SIFT algorithm and the relevant random sample consensus (RANSAC) algorithm, is a very fast process; however, it detects matching only at tens or hundreds of discrete points and cannot produce dense and full-field matching. Figure 3 gives an example of using the SIFT method to extract feature points and detect matching between two images.

The SIFT matching scheme can be incorporated into the DIC analysis to find matches at sparse point locations. Although the number of detected matching pairs is limited, it is sufficient for the initial guess task [23,24]. Equation (7) shows that three pairs of matching points are required to solve for the six transformation parameters (ξ, ξ_u, ξ_v, η, η_u, and η_v) since the other parameters can be set to zeros. It is also noted that the scale factor a and intensity offset b in Eq. (6) can be set to one and zero, respectively, for the initial guess.

To start the DIC analysis, a randomly picked SIFT point in the reference image is selected as the seed point, and its two closest neighbor SIFT points together with the three matching points in the target image are chosen to determine the aforementioned six parameters. Iteration convergence and disparity limitation criteria can be used to determine whether the DIC analysis of the seed point is successful or not. Upon a successful analysis, the matching results will be propagated to its nearest neighboring pixels on the calculation grid as initial guess. The guideline of the subsequent matching computation is as follows: for all the grid points that have been processed, the one that has the smallest correlation coefficient is chosen as the next point to be propagated. This procedure means that the matching results (i.e., optimized parameters) of the previous point are used as the initial values of parameters for its non-processed grid neighbors. By following this procedure, the image matching process runs through a path guided by the best matching. The propagation of the DIC analysis stops when it reaches any bad region, or any region that cannot provide the correct results. If the DIC analysis of a SIFT seed point fails or the matching propagation stops, the process will go back to find another SIFT point in no particular order. If a grid point has

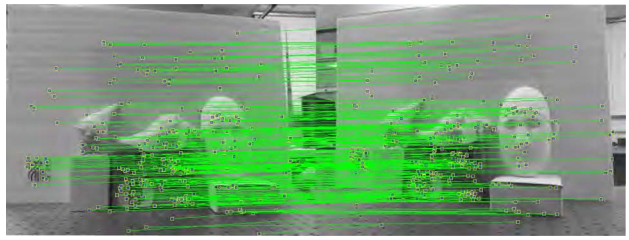

Fig. 3. Example of feature-point matching with the SIFT method.

been processed successfully in a previous computation, it will not be analyzed again in any subsequent matching process to avoid redundant processing. Otherwise, it will be processed again in subsequent analysis that originates from a different seed point.

4. EXPERIMENTS AND RESULTS

A number of experiments have been conducted to evaluate the performance and effectiveness of the proposed technique. In the experiments, two devices of the Kinect for Windows and Kinect SDK v1.8 are used, and the computer is a laptop with an Intel Core i7-3537U processor with 8 GB RAM. The image capture and analysis program is written with C++ language, and the images from the left and right IR cameras (with resolution of 640 × 480 pixels) are acquired simultaneously and saved as 8-bit grayscale images. The distances between the Kinect devices and the object(s) being measured are adjusted according to the field of measurement and the speckle density of the projected pattern in each experiment. For each different experimental setup, the IR cameras are calibrated by using the calibration technique described previously, which uses the frontal image concept and high-precision control point detection scheme.

A. Accuracy Test

The first experiment aims at testing the measurement accuracy of the proposed 3D experimental technique. In the experiment, a flat gage block (101.6 mm × 101.6 mm × 25.4 mm) is selected as the specimen, which stands on an optical translation stage driven by a differential adjuster with sub-micrometer resolution. The block has a white surface to facilitate the testing, and the distance from the Kinect devices to the experimental reference plane is around 0.7 m. In the analysis, the subset size is set to 21 × 21 pixels, and the calculation grid step is set to 2 pixels. The displacement results are calculated by subtracting the initial location from each new location.

Figure 4 shows a typical result of the out-of-plane coordinates plotted over the original image captured by the left camera. The average out-of-plane displacement at each position relative to the initial position is summarized in Table 1. The results indicate that the largest error of displacement measurement is 0.046 mm. Considering that the resolution of the IR cameras is low and the field width of view is about 710 mm, the measurement accuracy can be regarded as quite high. This experiment helps confirm the validity of the proposed 3D shape measurement approach.

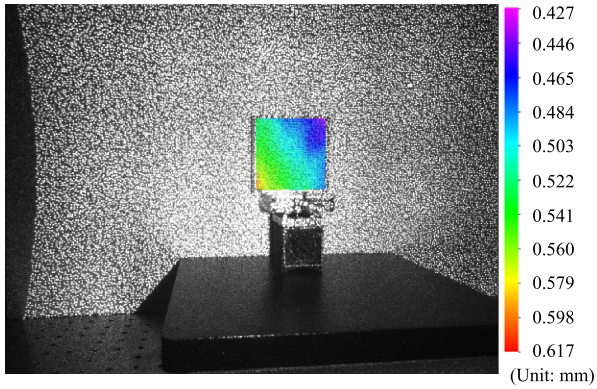

Fig. 4. Out-of-plane coordinate plot of the gage block over the left image.

Table 1. Actual and Measured Displacements of the Gage Block[a]

Position	Actual	Measured	Error
1	0.5	0.473	−0.027
2	1.0	1.026	0.026
3	1.5	1.543	0.043
4	2.0	2.046	0.046
5	2.5	2.503	0.003
6	3.0	2.981	−0.019

[a]unit: mm.

B. 3D Shape Measurement

The proposed technique has been employed to measure the 3D shapes and deformations of a variety of objects. Figure 5 shows some representative measurement results of four selected experiments that involve a throw pillow, a human upper body, an unpainted tribal mask, and a scene with three separate objects, respectively. The top two rows in Fig. 5 show the 3D reconstruction of the pillow and the human body where one of the captured IR images and a schematic color map of the depth are shown. The third row displays the 3D reconstructed tribal mask, which is generated by a stitching process to combine the 3D images acquired at three different positions. The bottom row in the figure exhibits the 3D reconstructed shapes of multiple separate objects. These experiments intend to demonstrate the abilities of the proposed approach for measurement of deformations, measurement of dark-color objects, measurement of objects with shiny or textureless surfaces, and measurement of multiple objects, as well as measurement at different scales.

C. Motion and Vibration Measurement

In order to verify the ability of the proposed technique for motion and vibration measurements, a controlled experiment has been implemented. The experimental setup is shown in Fig. 6(a) where the specimen is a sheet of white paper and the vibration source is a mechanical shaker. The shaker, oriented horizontally with a ring stand, has a small accelerometer attached to its end. The accelerometer is then attached with double-sided tape to the paper that is loosely fixed between two optical posts. The paper size is substantially larger than the accelerometer end to better demonstrate the vibration effect, but the vibration analysis will rely on the data in the region covered by the accelerometer end. In the experiment, the shaker is excited by a sinusoidal wave generated with a vibration frequency of 5 Hz and an amplitude of 0.5 V (voltage). An amplifier with a gain of 15 is deployed to amplify the wave signal for the shaker. The signal detected by the accelerometer is utilized to calculate and monitor the actual displacement and vibration frequency of the shaker, which are necessary for evaluating the displacement and vibration frequency extracted by the proposed 3D measurement technique.

The IR cameras have a capturing speed of 30 frames per second (fps), which is much higher than the vibration frequency of 5 Hz, so most of the details of the vibration can

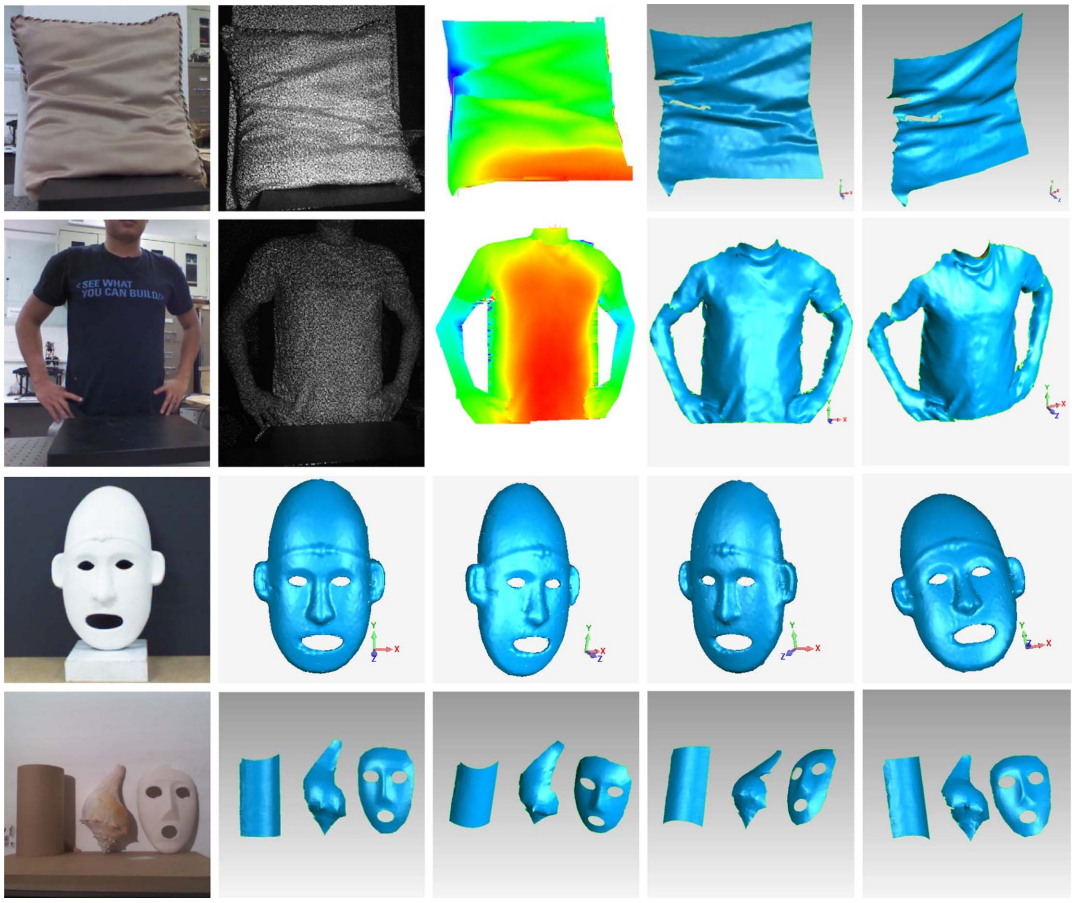

Fig. 5. 3D shape and deformation measurements of a variety of objects.

be captured in the experiment. Figure 6(b) shows a 2D-map illustration of the out-of-plane coordinates in a selected region where the small ROI is the region connected to the accelerometer and shaker. The large rectangular ROI is chosen for the purpose of vibration demonstration, whereas the small circular ROI is selected for the purpose of comparison with the results measured by the accelerometer. Figures 6(c) and 6(d) illustrate the vibration displacements detected by the accelerometer and the vibration frequency subsequently calculated using the fast Fourier transform (FFT) algorithm, respectively. Figures 6(e) and 6(f) plot the vibration displacements over the small ROI and the corresponding vibration frequency measured by the proposed technique, respectively. It is evident from the figures that the measurement results of the proposed technique match well with the ones provided by the accelerometer device. Since the proposed technique provides non-contact full-field measurements, it potentially has much broader applications than the accelerometer counterpart.

In the recent decade, there has been a surge of interest in exploring novel health monitoring systems. In the meantime, home automation has become more prevalent in homes and will continue to rise over the next decade and beyond. The proposed technique allows for the technology to advance towards non-contact health monitoring that will have a notable impact on the betterment of smart homes. The current home automation systems generally include surveillance cameras for security and other relevant purposes. Given its measurement accuracy, the proposed technique could be added to the existing surveillance systems to automatically detect harmonic motion and vibratory signals, such as those of the respiration and heartbeats of human beings [25].

The goal of the last two experiments is to demonstrate the potential application of the proposed technique to measuring the respiration rates and pulse waves of human beings. Considering that the typical respiration rate is in the range of 0.16–0.66 Hz and the typical pulse is in the range to 0.7–2 Hz [26–29], the 30-fps frame rate of the IR camera is adequate for an accurate detection of the human respiration and pulse rates. It should be pointed out that these two measurements rely on only the average displacement in the corresponding ROI for each pair of images. Therefore, the DIC matching process does not have to be applied to every pixel; instead, the DIC computation grid step can be set to five or larger (i.e., the DIC matching is conducted at every five or more pixels) to speed up the analysis. The full-field matching results are not required, but if desired, they can be obtained by using data interpolation.

Figure 7 shows the measurement results of the respiration rate testing, for which Figs. 7(a) and 7(b) are the out-of-plane coordinate maps at two consecutive lowest and highest positions, respectively. Figures 7(c) and 7(d) show the detected average out-of-plane displacements in the ROI and the

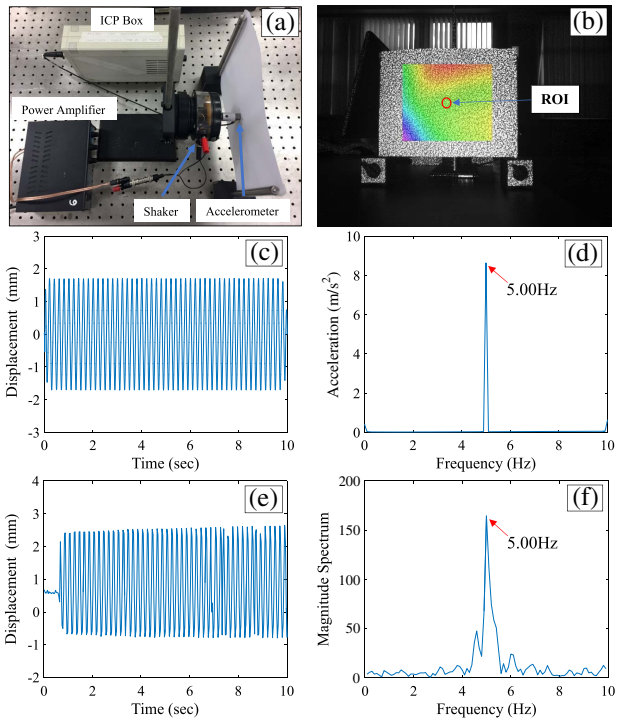

Fig. 6. Vibration test: (a) experimental setup; (b) color illustration of the out-of-plane coordinates (see **Visualization 1**); (c) and (d) displacements and vibration frequency acquired from the accelerometer; and (e) and (f) displacements and vibration frequency measured by the proposed technique.

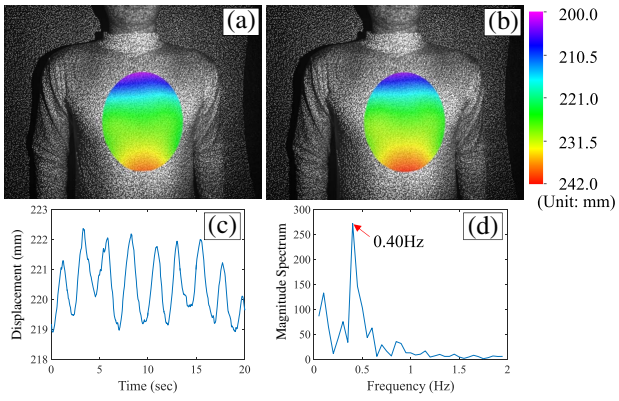

Fig. 7. Measurement of respiration rate: (a) and (b) out-of-plane coordinate maps at the lowest and highest positions; (c) displacement distribution; and (d) frequency spectrum (see **Visualization 2**).

corresponding frequency spectrum. Because the average displacements are used in the analysis, the shape and size of the ROI are not important as long as they are reasonable. The analysis reveals that the detected frequency is 0.40 Hz, which gives a respiration rate of 24 breaths per minute. This has been verified by manually counting the breaths from the captured video.

Similarly, Fig. 8 shows the results of the pulse wave measurement, for which Figs. 8(a) and 8(b) illustrate the out-of-plane

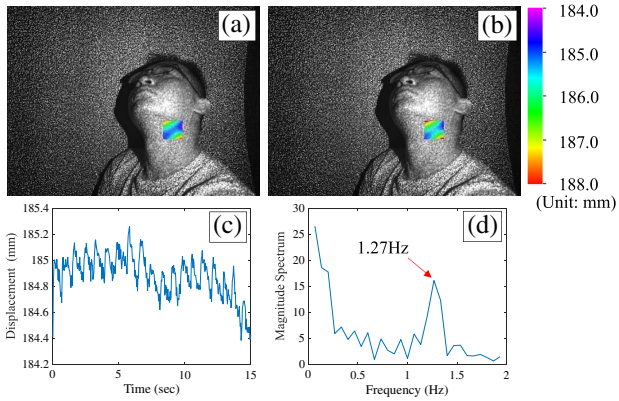

Fig. 8. Measurement of pulse wave: (a) and (b) out-of-plane coordinate maps at the lowest and highest positions; (c) displacement distribution; and (d) frequency spectrum (see **Visualization 3**).

coordinate maps at two consecutive lowest and highest positions, respectively. Figures 8(c) and 8(d) are the detected average out-of-plane displacements in the ROI and the corresponding frequency spectrum, respectively. It can be seen from the figures that the detected pulse frequency is 1.27 Hz, which gives a respiration rate of 76 beats per minute. This has also been verified by manually counting the pulse waves from the captured video.

These two preliminary experiments have the potential application of being used in smart homes for automatic noncontact monitoring of respiration and heartbeat rates.

5. CONCLUSIONS

In this paper, a robust technique to acquire accurate 3D shape, deformation, motion, and vibration measurements using the Kinect RGB-D sensors is presented. The proposed approach takes advantage of the IR camera and IR projector in the Kinect device and combines two key steps to achieve high-accuracy 3D coordinate measurements: one is an accurate camera calibration scheme using a sophisticated lens-distortion model, bundle-adjustment algorithm, and frontal-image concept; the other is an enhanced image-matching process using SIFT-RANSAC initial guess, area-based iterative correlation, correlation-coefficient-guided analysis, and sophisticated B-spline interpolation. Unlike the conventional stereo-vision technique, the proposed approach uses an invisible IR speckle pattern to facilitate image matching, which makes it capable of measuring scenes or objects (e.g., a textureless object) that cannot be fulfilled by the conventional technique. As demonstrated in the paper, the proposed technique provides accurate 3D measurement results for a variety of objects at different scales. In comparison with the popular 3D imaging technique based on structured light illumination [30], the IR projector and sensors can cope well with objects having dark or shiny surfaces. The experiments carried out in the paper demonstrated that the proposed technique and approach are effective and practical for the real-world applications.

In regard to the limitations of the technique, a key drawback originates from the low resolution of the Kinect IR camera

(640 × 480 pixels). Furthermore, because of the focusing issue of the IR speckle patterns, the desired working distance of the proposed system is limited to a range from 0.5 m to 2 m. A short working distance below 0.5 m introduces strong reflections from the IR projector, which leads to the missing of speckle pattern information. A long working distance above 2 m brings reduced accuracy to the measurement because of the limited resolution of the IR camera and the low density of speckle patterns. Despite these limitations, the study opens the door for future research and development, as well as applications. As hybrid sensors capable of providing enhanced specifications and performance available in the future, it will be very feasible to achieve 3D shape, deformation, motion, and vibration measurements with higher resolution, higher speed, and higher accuracy [31–34].

Acknowledgment. The authors sincerely thank Drs. J. Vignola and D. Turo for their help with the mechanical shaker and accelerometer.

REFERENCES

1. Z. Zhang, "Microsoft Kinect sensor and its effect," IEEE MultiMedia **19**, 4–10 (2012).
2. S. Zug, F. Penzlin, A. Dietrich, T. Nguyen, and S. Albert, "Are laser scanners replaceable by Kinect sensors in robotic application?" in *IEEE International Symposium on Robotic and Sensors Environments* (IEEE, 2012), pp. 144–149.
3. K. Lai, L. Bo, X. Ren, and D. Fox, "Sparse distance learning for object recognition combining RGB and depth information," in *IEEE International Conference on Robotics and Automation* (IEEE, 2011), pp. 4008–4013.
4. J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from a single depth images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2011), pp. 1297–1304.
5. J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with Microsoft Kinect sensor: a review," IEEE Trans. Cybernet. **43**, 1290–1303 (2013).
6. "Kinect for Xbox One," http://www.xbox.com/en-US/xbox-one/accessories/kinect.
7. "PrimeSense," https://en.wikipedia.org/wiki/PrimeSense.
8. K. Mankoff and T. Russo, "The Kinect: a low-cost, high-resolution, short-range 3D camera," Earth Surf. Process. Lardf. **38**, 926–936 (2013).
9. J. Smisek, M. Jancosek, and T. Pajdla, "3D with Kinect," in *Advances in Computer Vision and Pattern Recognition* (2013), pp. 3–25.
10. K. Khoshelham and S. Elberink, "Accuracy and resolution of Kinect depth data for indoor mapping applications," Sensors **12**, 1437–1454 (2012).
11. S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotto, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon, "KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera," in *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (ACM, 2011), pp. 559–568.
12. M. Duo, J. Taylor, H. Fuchs, A. Fitzgibbon, and S. Izadi, "3D scanning deformable objects with a single RGBD sensor," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2015), pp. 493–501.
13. W. C. Chiu, U. Blanke, and M. Fritz, "Improving the Kinect by cross-model stereo," in *British Machine Vision Conference (BMVC)* (2011), pp. 1–10.
14. P. Henry, M. Kraimin, E. Herbst, X. Ren, and D. Fox, "RGB-D mapping: using Kinect-style depth cameras for dense 3D modeling of indoor environments," Int. J. Robot. Res. **31**, 647–663 (2012).
15. F. Endres, J. Hess, J. Sturm, D. Cremers, and W. Burgard, "3D mapping with an RGB-D camera," IEEE Trans. Robot. **30**, 177–187 (2014).
16. F. Alhwarin, A. Ferrein, and I. Scholl, "IR stereo Kinect: improving depth images by combining structured light with IR stereo," in *Pacific Rim International Conference on Artificial Intelligence* (2014), Vol. **8862**, pp. 409–421.
17. Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1330–1334 (2000).
18. M. Vo, Z. Wang, L. Luu, and J. Ma, "Advanced geometric camera calibration for machine vision," Opt. Eng. **50**, 110503 (2011).
19. B. Pan, H. Xie, and Z. Wang, "Equivalence of digital image correlation criteria for pattern matching," Appl. Opt. **49**, 5501–5509 (2010).
20. B. Pan, K. Qian, H. Xie, and A. Asundi, "Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review," Meas. Sci. Technol. **20**, 062001 (2009).
21. M. L. A. Lourakis and A. A. Argyros, "Is Levenberg-Marquardt the most efficient optimization algorithm for implementing bundle adjustment?" in *Proceedings of the Tenth IEEE International Conference on Computer Vision* (IEEE, 2005), pp. 1526–1531.
22. L. Luu, Z. Wang, M. Vo, T. Hoang, and J. Ma, "Accuracy enhancement of digital image correlation with B-spline interpolation," Opt. Lett. **36**, 3070–3072 (2011).
23. Z. Wang, H. Kieu, H. Nguyen, and M. Le, "Digital image correlation in experimental mechanics and image registration in computer vision: similarities, differences and complements," Opt. Lasers Eng. **65**, 18–27 (2015).
24. H. Kieu, T. Pan, Z. Wang, M. Le, H. Nguyen, and M. Vo, "Accurate 3D shape measurement of multiple separate objects with stereo vision," Meas. Sci. Technol. **25**, 035401 (2014).
25. F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. Miller, "Smart homes that monitor breathing and heart rate," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (ACM, 2015), pp. 837–846.
26. Z. Chen, D. Lau, J. Teo, S. Ng, X. Yang, and P. Kei, "Simultaneous measurement of breathing rate and heart rate using a microbend multimode fiber optic sensor," J. Biomed. Opt. **19**, 057001 (2014).
27. J. Hernandez, D. McDuff, and R. W. Picard, "BioWatch: estimation of heart and breathing rates from wrist motions," in *Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare* (IEEE, 2015), pp. 169–176.
28. J. Wu, R. Chang, and J. Jiang, "A novel pulse measurement system by using laser triangulation and a CMOS image sensor," Sensors **7**, 3366–3385 (2007).
29. X. Shao, X. Dai, Z. Chen, and X. He, "Real-time 3D digital image correlation method and its application in human pulse monitoring," Appl. Opt. **55**, 696–704 (2016).
30. H. Nguyen, D. Nguyen, Z. Wang, H. Kieu, and M. Le, "Real-time, high-accuracy 3D imaging and shape measurement," Appl. Opt. **54**, A9–A17 (2015).
31. Z. Wang, H. Nguyen, and J. Quisberth, "Audio extraction from silent high-speed video using an optical technique," Opt. Eng. **53**, 110502 (2014).
32. R. Wu, Y. Chen, Y. Pan, Q. Wang, and D. Zhang, "Determination of three-dimensional movement for rotary blades using digital image correlation," Opt. Lasers Eng. **65**, 38–45 (2015).
33. J. Espinosa, J. Perez, B. Ferrer, and D. Mas, "Method for targetless tracking subpixel in-plane movements," Appl. Opt. **54**, 7760–7765 (2015).
34. T. Nguyen, G. Nehmetallah, D. Tran, A. Darudi, and P. Soltani, "Fully automated, high speed, tomographic phase object reconstruction using the transport of intensity equation in transmission and reflection configurations," Appl. Opt. **54**, 10443–10453 (2015).