applied optics

Method for large-range structured light system calibration

Yatong An,[1,†] Tyler Bell,[1,†] Beiwen Li,[1] Jing Xu,[2] and Song Zhang[1,*]

[1]School of Mechanical Engineering, Purdue University, West Lafayette, Indiana 47907, USA
[2]Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China
*Corresponding author: szhang15@purdue.edu

Received 15 July 2016; revised 24 October 2016; accepted 26 October 2016; posted 28 October 2016 (Doc. ID 270704); published 18 November 2016

Structured light system calibration often requires the usage of a calibration target with a similar size as the field of view (FOV), which brings challenges to a large-range structured light system calibration since fabricating large calibration targets is difficult and expensive. This paper presents a large-range system calibration method that does not need a large calibration target. The proposed method includes two stages: (1) accurately calibrate intrinsics (i.e., focal lengths and principle points) at a near range where both the camera and projector are out of focus, and (2) calibrate the extrinsic parameters (translation and rotation) from camera to projector with the assistance of a low-accuracy, large-range three-dimensional (3D) sensor (e.g., Microsoft Kinect). We have developed a large-scale 3D shape measurement system with a FOV of 1120 mm × 1900 mm × 1000 mm. Experiments demonstrate our system can achieve measurement accuracy as high as 0.07 mm with a standard deviation of 0.80 mm by measuring a 304.8 mm diameter sphere. As a comparison, Kinect V2 only achieved mean error of 0.80 mm with a standard deviation of 3.41 mm for the FOV of measurement.

OCIS codes: (120.0120) Instrumentation, measurement, and metrology; (120.2650) Fringe analysis; (100.5070) Phase retrieval.

http://doi.org/10.1364/AO.55.009563

1. INTRODUCTION

Optically measuring three-dimensional (3D) surface geometry plays an increasingly important role in numerous applications. High-accuracy 3D shape measurements are of great importance to medicine and manufacturing, as well as other applications. Structured light technologies are increasingly used for close- and small-range 3D shape measurements, yet they are not as popular for long- and large-range 3D shape measurement. It is well known that structured light system measurement accuracy heavily hinges on accurately calibrating the system. We believe one of the reasons why structured light technologies are not widely used for large-range 3D shape measurement is due to a lack of an accurate yet flexible calibration method for such a range.

Structured light system calibration starts and evolves with camera calibration. The evolution of camera calibration started with straightforward software algorithms. More sophisticated algorithms and expensively fabricated calibration targets came along next to improve calibration precision. Most recently, the focus has been on reducing the fabrication costs while improving the software algorithms. In the 1970s, researchers developed straightforward software algorithms for camera calibration yet used accurately fabricated 3D targets with precisely measured 3D feature points [1,2]. In the 1980s, Tsai [3] reduced the target complexity from 3D to 2D, employed a precision translation stage, and developed more sophisticated algorithms for camera calibration. In the 2000s, Zhang [4] developed an even more flexible calibration approach by allowing for 2D targets with flexible motion. Of course, the software algorithm behind the calibration was now more complex than before. Lately, researchers have been developing methods for camera calibration by using unknown feature points or even imperfect calibration targets [5–8]. Furthermore, active digital displays, such as liquid crystal display (LCD), have also been employed for accurate camera calibration [9,10].

Compared with camera calibration, structured light system calibration is more complex because it uses a projector that cannot physically capture images like a camera. Due to the difficulty of calibrating a projector, researchers in the optics community often use the simple reference-plane-based method [11–13]. The reference-plane-based method can work if telecentric lenses are used or the measurement depth range is not large. To overcome the limitations of the reference-plane-based calibration method, researchers have also developed numerous structured light system calibration approaches. One approach is to calibrate the positions and orientations of the camera and the

projector through a complicated and time-consuming calibration process [14–16]. Another approach is to estimate the relationship between the depth and encoded information (e.g., phase) through optimization [17–20].

By treating the projector as the inverse of a camera, researchers have developed some similar geometric calibration approaches for projector calibration. For example, Legarda-Sáenz *et al.* [21] proposed to use phase to establish corresponding points between the projector and the camera and to calibrate the projector with the calibrated camera parameters; Zhang and Huang [22] developed a method that allows the projector to *capture* images like a camera and to calibrate camera and projector independently so the calibration error of the camera does not affect the projector calibration, and vice versa. Lately, researchers also developed improved calibration methods by using linear interpolation [23], bundle adjustment [24], or residual error compensation with planar constraints [25].

All aforementioned camera, projector, and structured light system calibration methods require the use of the calibration target being similar in size to the field of view (FOV) of the device; such a typical requirement brings challenges for large-range structured light system calibration since precisely fabricating large calibration targets is often difficult and expensive. Due to this major challenge, structured light technologies are primarily used in close- and small-range measurement applications.

This paper presents a calibration method that does not require an equivalent size calibration target to the sensing FOV but rather uses a large-range and low-accuracy 3D sensor in addition to a regular sized calibration target. Geometric structured light system calibration includes estimating the intrinsics (i.e., focal lengths and principle points) of the camera and the projector, as well as estimating the extrinsics (i.e., translation and rotation between camera coordinate system and projector coordinate system). To our knowledge, the intrinsic parameter calibration is more difficult than extrinsic parameter calibration since accurately estimating focal lengths and principle points often requires many feature points within the FOV. In comparison, the extrinsic parameter calibration can use one single pose and fewer feature points to estimate the transformation from one coordinate system to another. The proposed method takes advantage of the different difficulty levels of intrinsic and extrinsic calibrations. Specifically, the proposed method is divided into two stages: the first stage is to accurately calibrate intrinsics at a close range using a more precisely fabricated calibration target even though both the camera and projector are out of focus at this close range, and the second stage is to calibrate the translation and rotation from camera to projector (i.e., extrinsic parameters) using a low-accuracy, yet large-range, 3D sensor (e.g., Microsoft Kinect). The proposed calibration method is built on foundations that we developed for out-of-focus camera and projector calibration. In particular, we found that the severely out-of-focus camera intrinsics can be accurately estimated directly or by using an active calibration target (e.g., LCD) [26], and the out-of-focus projector can be accurately calibrated by establishing a one-to-one mapping in the phase domain and using an in-focus camera to assist in feature point detection [27]. Once the intrinsics are estimated, the extrinsic parameters of the structured light system can be accurately estimated using a low-resolution and low-accuracy 3D sensor with many actively identified feature points of any object (e.g., a wall). The system we developed for large-scale 3D shape measurement can measure a FOV of 1120 mm × 1900 mm × 1000 mm. Experiments demonstrate our system can achieve measurement accuracy as high as 0.07 mm with a standard deviation of 0.80 mm by measuring a 304.8 mm diameter sphere. As a comparison, Kinect V2 only achieved a mean error of 0.80 mm with a standard deviation of 3.41 mm for the FOV of measurement.

Section 2 explains the principles of the proposed calibration method. Section 3 presents experimental results to further validate the proposed method. Section 4 discusses the advantages and possible limitations of the proposed calibration method. Finally, Section 5 summarizes the paper.

2. PRINCIPLE

This section thoroughly explains the principles of the proposed large-range structured light system calibration method. Specifically, we will present the standard pinhole camera model, phase-shifting algorithm, out-of-focus projector calibration, camera calibration, system extrinsic calibration, and overall framework of large-range structured light system calibration.

A. Camera/Projector Lens Model

To describe the relationship between the 3D world coordinates (x^w, y^w, z^w) and the 2D image coordinates (u, v), the most widely used model is the pinhole model. Mathematically, the pinhole model for a camera can be represented as

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & \gamma & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} [\mathbf{R} \quad \mathbf{t}] \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}, \tag{1}$$

where s is the scaling factor, f_u and f_v are the effective focal lengths along the u and v directions, γ is the skew factor of the u and v axes, and (u_0, v_0) is the principle point that is the intersecting point between the optical axis and the image plane. $\mathbf{R}$ and $\mathbf{t}$ describe the rotation matrix and the translation vector between the world coordinate system and the lens coordinate system. Usually, they are represented using the following forms:

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \qquad \mathbf{t} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}. \tag{2}$$

Distortion is a very common problem for lenses. Among the different kinds of distortions, radial and tangential distortions are the two most common. Mathematically, these two kinds of distortions can be modeled using the following five parameters:

$$\mathbf{D} = [k_1 \quad k_2 \quad p_1 \quad p_2 \quad k_3]^T, \tag{3}$$

where k_1, k_2, and k_3 are radial distortion coefficients, and p_1 and p_2 are the tangential distortion coefficients. Based on these coefficients, we can rectify the radial distortion using the following model:

$$u' = u(1 + k_1 r^2 + k_2 r^4 + k_3 r^6),$$
$$v' = v(1 + k_1 r^2 + k_2 r^4 + k_3 r^6), \quad (4)$$

where (u, v) is the pixel in the input image, (u', v') is the pixel coordinate after the radial distortion corrections, and $r = \sqrt{(u - u_0)^2 + (v - v_0)^2}$. Similarly, we can rectify the tangential distortion using the following model:

$$u' = u + [2p_1 uv + p_2(r^2 + 2u^2)],$$
$$v' = v + [p_1(r^2 + 2v^2) + 2p_2 uv]. \quad (5)$$

The projector has inverse optics from the camera; simply put, a projector projects an image instead of capturing an image. A projector and a camera share exactly the same mathematical model between their world coordinate systems and their image coordinate systems. Thus, we can also directly apply the given pinhole model to describe a projector model.

B. Phase-Shifting Algorithm

Using phase instead of intensity is more advantageous because phase is more accurate and robust to both noise and ambient lighting effects. There are many phase-shifting methods (three-step, four-step, etc.) and phase unwrapping methods, including spatial and temporal ones. Generally speaking, the more steps are used, the more accurate results we can get. For a number of N equally phase-shifted fringe patterns, mathematically the ith fringe image I^i can be described as

$$I^i(x, y) = I'(x, y) + I''(x, y)\cos(\phi + 2i\pi/N), \quad (6)$$

where $I'(x, y)$ is the average intensity, $I''(x, y)$ is the intensity modulation, $i = 1, 2, \ldots, N$, and $\phi(x, y)$ is the phase to be solved for. Using a least square method, we can get

$$\phi(x, y) = -\tan^{-1}\left[\frac{\sum_{i=1}^{N} I^i \sin(2i\pi/N)}{\sum_{i=1}^{N} I^i \cos(2i\pi/N)}\right]. \quad (7)$$

This equation can give a wrapped phase that ranges from $-\pi$ to $+\pi$. Next, we must adjust those 2π discontinuities. The process of adjusting 2π discontinuities is called phase unwrapping. Over many years, a variety of phase unwrapping methods have been developed. The two most popular categories are spatial unwrapping and temporal unwrapping methods. Essentially we want to find a fringe order $k(x, y)$ for each pixel; then the phase can be unwrapped using the following equation:

$$\Phi(x, y) = \phi(x, y) + k(x, y) \times 2\pi. \quad (8)$$

Fundamentally the difference between temporal unwrapping and spatial unwrapping methods is that for the temporal phase unwrapping, one can retrieve an *absolute* phase map; while spatial methods retrieve a *relative* phase map. The reason is that spatial phase unwrapping algorithms usually find $k(x, y)$ through analyzing the point to be processed and its neighboring pixels. Thus, the obtained phase using a spatial phase unwrapping method is relative to one point (i.e., a *relative* phase map). In contrast, temporal phase unwrapping methods uniquely compute the phase values for each pixel by projecting additional coded patterns. Thus, the retrieved phase map is an *absolute* one. An absolute phase is necessary for 3D reconstruction without ambiguity. Given this, we will use a temporal phase unwrapping method in this research. Specifically, we will use gray coded patterns for phase unwrapping in later experiments.

C. Out-of-Focus Projector Intrinsic Calibration

As previously mentioned, a projector has inverse optics with respect to a camera. The most popular way to calibrate a projector is the one proposed by Zhang and Huang [22]. But for a large-range structured light system, fabricating a very large calibration board at the projector's focus range and to fit the projector's FOV is both difficult and expensive. As in Fig. 1, the projector is focused at the wall, which is a far distance from the projector. It is not practical to design that kind of large calibration board that is the size of a wall. To solve this problem, it is desirable to calibrate the projector within its defocus range, which is near to the projector.

Li *et al.* [27] prove that out-of-focus projector can be calibrated accurately both theoretically and practically. This gives us the possibility to calibrate the projector of a large-range structured light system in its defocused area. As shown in Fig. 1, we can use a regular-size calibration board to calibrate the projector at its near defocus range.

The whole process of calibrating such an out-of-focus projector is similar to what might be done for a regular small-scale structured light system. We can use the projector to project both horizontal and vertical phase-shifted patterns. Since the projector is severely defocused at the position of the calibration board, we can set the projector to project binary patterns. Because of the effect of defocusing, binary patterns can approximate sinusoidal ones [28]. Next, a camera which is focused at the calibration board can be used to capture fringe images and do phase unwrapping. Theoretically, this can create a one-to-one mapping between a camera pixel and a projector pixel in the phase domain.

Take a circle grid calibration board as an example. For a specific circle center (u^c, v^c) in the camera image, we need to find the corresponding pixel (u^p, v^p) in the projector image coordinate system. If we project the horizontal patterns onto the calibration board with the smallest fringe period being T_h, we can compute phase and do phase unwrapping to retrieve the absolute phase ϕ_v^c in the vertical gradient direction. Then for each camera pixel (u^p, v^p), its phase value $\phi_v^c(u^p, v^p)$ maps to a projector pixel line v^p by the following linear constraint:

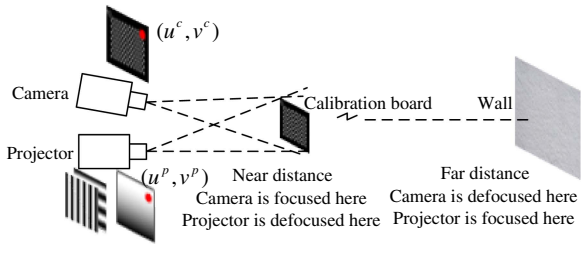


Fig. 1. Out-of-focus projector calibration. Since fabricating a very large calibration board at the projector's focus range (wall) is both difficult and expensive, we instead calibrate the projector at its defocus range. In the near defocus range, we can use a regular size calibration board to calibrate the projector.

$$v^p = \phi_v^c(u^c, v^c) \times T^h/(2\pi). \quad (9)$$

Similarly, when we project vertical patterns with the smallest fringe period being T^v, we can retrieve the absolute phase ϕ_h^c in the horizontal gradient direction, which maps to an orthogonal projector pixel line u^p determined by a similar linear constraint as follows:

$$u^p = \phi_h^c(u^c, v^c) \times T^v/(2\pi). \quad (10)$$

For each circle center pixel (u^c, v^c) in the camera image, we can find the corresponding pixel (u^p, v^p) in the projector image. Using this approach, the projector can *see* the circle grid patterns. By placing the calibration board in different spatial orientations and finding the circle grid of the projector image in each pose, finally we can calibrate the projector, similar to a camera, and get its intrinsic matrix.

D. Camera Intrinsic Calibration

Given that the projector is calibrated, it can now be fixed; however, the camera is still in focus at a near distance. So that the entire system can work well for a large sensing range, we next adjust the camera's focus and angle with respect to the projector. The camera's focus is set such that it is now in focus at the far distance, and its angle from the projector is set to achieve an optimal matching between the FOVs of each device.

Now that the camera is focused at a far distance, its calibration faces a similar problem as calibrating the projector. Namely, the problem is that using a very large calibration board to fill the FOV of the camera is not practical in either fabrication or economy. To address this, we take advantage of the idea again to calibrate the camera intrinsics at its defocus range.

If the lens used has a short focal length, as in the practical experiments, even when the lens is focused at infinity, the level of camera lens defocusing is still not enough to fail a conventional camera calibration approach. Given this, we can still directly calibrate the camera by capturing different poses of a calibration board. As in Fig. 2, we can capture different poses of the calibration board at a near distance, albeit the camera is out of focus, and then use the OpenCV calibration toolbox to get the intrinsic matrix of the camera. If a long focal length lens is used, one can adopt the out-of-focus camera calibration approach discussed by Bell and Zhang [26]. It uses a digital display (e.g., LCD monitor) to generate fringe patterns which encode feature points into the carrier phase; these feature points can be accurately recovered even if the fringe patterns are substantially blurred (i.e., the camera is substantially defocused). That method can be adopted here to make our algorithm more generic.

E. Structured Light System Extrinsic Calibration

In traditional methods of calibrating the extrinsic parameters of a structured light system, a regular sized calibration board is used within the FOV of each device. This works well for small-range structured light systems, yet the calibration board will be too small for the large-range structured light system, as shown in Fig. 3. Since the FOV is too large, it is neither practical nor economically efficient to fabricate a very large calibration board to fill the whole FOV of the structured light system.

To deal with this problem, we propose a novel method to calibrate the extrinsic parameters between the projector and the camera. Our proposed method uses the assistance of a low-accuracy, large-range 3D sensor (e.g., Microsoft Kinect V2). As shown in Fig. 4, we use the projector to project some markers (e.g., a circle grid) onto a real 3D scene (like a wall), where we obtain (u^p, v^p) of markers in our predesigned projector image. Then the camera can capture and detect the position (u^c, v^c) of markers in the camera image. Also, the Kinect can capture and detect the position (u^k, v^k) of markers in the Kinect color space. Simultaneously, the Kinect can capture the depth image and map the 3D coordinates (x^k, y^k, z^k) by its own built-in function into the color space. From here, we can get the 3D coordinate information of those markers in the Kinect's world space. To summarize, for each marker, we have

$$\begin{cases} (u^c, v^c), & \text{position in the camera image coordinate system} \\ (u^p, v^p), & \text{position in the projector image coordinate system} \\ (x^k, y^k, z^k), & \text{3D coordinates in the Kinect space.} \end{cases}$$

Using this information for each feature point, the extrinsic calibration is converted into a conventional stereo calibration problem. We can solve for the translation $\mathbf{T}^c$, $\mathbf{T}^p$ and rotation matrices $\mathbf{R}^c$, $\mathbf{R}^p$ of the projector and camera using one of the many well developed methods or software frameworks, such as the StereoCalibration method within OpenCV's Calibration Toolbox.

In general, the 3D scene for extrinsic calibration can be some complex environment, not necessarily a wall or a flat object. As long as we find the correspondence between camera pixel, projector pixel, and 3D coordinates, we can calibrate the extrinsic matrix between the projector and the camera. Further, this method can be extended by using horizontal and vertical phase-shifting patterns to encode feature points and establish correspondence. In this approach, we use the regular camera and Kinect to capture those fringe images and do phase computation and unwrapping simultaneously.

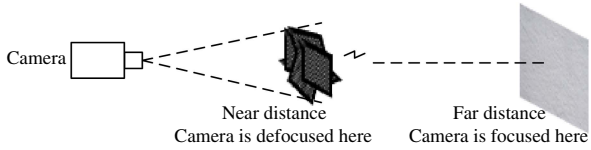


Fig. 2. Camera intrinsic calibration. For a near focused lens, even when the lens is focused at infinity, the level of camera lens defocusing is still not enough to fail a conventional camera calibration approach. Given this, the camera can be calibrated directly by capturing different poses of a calibration board at its near defocus range.

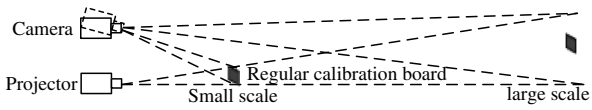


Fig. 3. Extrinsic calibration explanation. For extrinsic calibration in the small range, usually we can put a regular calibration board in the working zone (the eventual capture area of the large-range system). For large-range extrinsic calibration, the regular calibration board will be too small, and it is not practical to fabricate a very large calibration board, let alone use it for flexible calibration.

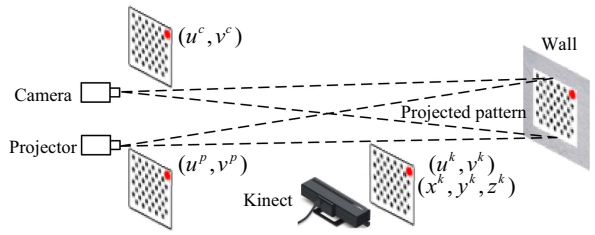


Fig. 4. Extrinsic calibration principle. We project some markers onto a real 3D scene, then we use the camera and Kinect to capture it at the same time. So for each marker, we can get its position in the camera image coordinate system (u^c, v^c), its position in the projector image coordinate system (u^p, v^p), and its 3D coordinate (x^k, y^k, z^k) in the Kinect space.

We can use phase to find the correspondence of (u^c, v^c), (u^p, v^p), and the pixels' 3D information (x^k, y^k, z^k) from Kinect.

With the assistance of a low-accuracy, large-range 3D sensor (e.g., Microsoft Kinect), the calibration process becomes much more flexible for the calibration of a large-range structured light system.

F. Overall Framework of Large-Range Structured Light System Calibration

Here we summarize the entire framework of our proposed large-range calibration method. Briefly speaking, we split the whole traditional calibration problem into two stages to make it adaptable to a large-range structured light system. The first stage is the intrinsic calibration process, and the second stage is the extrinsic calibration process. The overall framework for a large-range structured light system calibration is shown in Fig. 5.

- *Stage 1: Intrinsic calibration.*

1-A: Projector intrinsic calibration. Let the projector be focused at the far distance (the eventual capture area of the large-range system) and the camera be focused at the near distance. Put a calibration board in front of the system at a near distance. Let the projector project square binary patterns both horizontally and vertically, and the camera capture images simultaneously. Then unwrap the phase and find the absolute phase of the feature points on the calibration board. By feature point mapping in the phase domain, the projector can *see* the feature points (like circle centers). Place the calibration board at different poses and repeat the above process to get the feature map (u^p, v^p) for each pose. Then use some well-developed algorithms (like the OpenCV calibration toolbox) to compute the intrinsic matrix of the projector.

1-B: Camera intrinsic calibration. Now adjust the camera focus to be at the far distance (the eventual capture area of the large-range system). At a near distance, adopt conventional calibration methods (e.g., OpenCV camera calibration) to perform out-of-focus camera calibration. If a far focal length lens used, adopt the out-of-focus camera calibration approach discussed by Bell and Zhang [26].

- *Stage 2: Extrinsic calibration*. Set the system properly (e.g., changing the distance and angle between the projector and camera) for large-range 3D shape measurement. Project some specific patterns with feature points (u^p, v^p) onto a large 3D scene with the projector. Let the camera and Kinect capture the patterns directly, getting the pixel position of the feature points (u^c, v^c) in the camera image and corresponding 3D coordinates (x^k, y^k, z^k) in the Kinect's world space. It is worth noting that alternatively, instead of projecting feature points directly, one can also project phase-shifted patterns and find the correspondence between projector, camera, and Kinect images by phase. By repeating this process for different poses, we can build a correspondence map for each pose and do stereo calibration for the projector and camera to get their extrinsic parameters including rotation matrices and translation vectors.

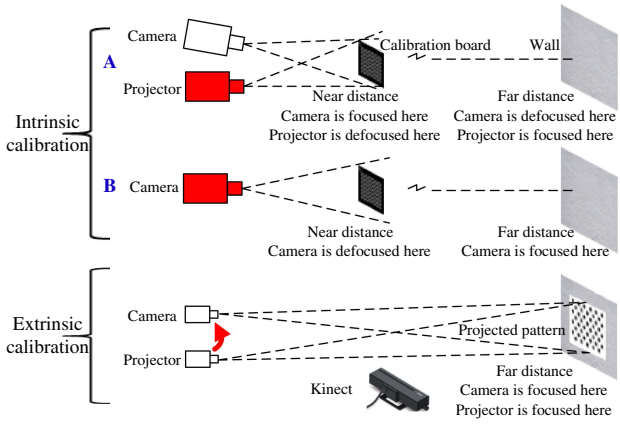
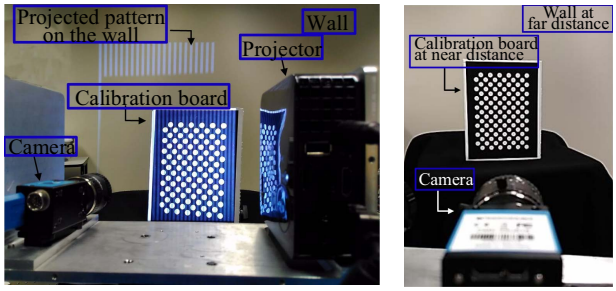


Fig. 5. Overall framework for large-range structured light system calibration. The proposed method includes two stages: (1) accurately calibrate intrinsics (i.e., focal lengths, principle points) at a near range where both camera and projector are out of focus, and (2) calibrate the extrinsic parameters (translation and rotation) from camera to projector with the assistance of a low-accuracy, large-range 3D sensor (e.g., Microsoft Kinect).

3. EXPERIMENT

To verify the performance of the proposed method, we developed a structured light system that includes a complementary metal-oxide-semiconductor camera (Model: DMK23UX174) with a 12 mm focal length lens (Model: Computar M1214-MP2). The resolution of the camera is set to be 1600 × 1200 pixels. The projector is a digital light processing one (Model: DELL M115HD) with a resolution of 1280 × 800 pixels. The auxiliary 3D sensor we used is a Kinect V2 with a depth map resolution of 512 × 424 pixels. The working distance of the Kinect V2 is 0.5 m ∼ 4.5 m.

We followed the framework proposed in Section 2.F to calibrate the system. Figure 6 shows the system setup for calibrating the intrinsic matrices of the projector and camera. Since the projector is substantially defocused at the position of the calibration board, we used square binary phase shifting patterns with fringe periods of $T^h = T^v = 36$ pixels to get a reasonable contrast when calibrating the projector. As shown in Fig. 6(a), the projected patterns have a sharp binary representation at the distance of the wall, yet a sinusoidal structure at the distance of the calibration board, due to the defocusing effect of the projector. Figure 6(b) shows the setup for camera intrinsic calibration.



(a) (b)

Fig. 6. Out-of-focus projector calibration and camera calibration. (a) System setup for out-of-focus projector calibration. The projector is focused at a far distance, and the camera is focused at a near distance where the calibration board is placed; (b) system setup for the out-of-focus camera calibration. The camera is focused at a far distance, like the wall. To calibrate it, we can just simply put a calibration board at a near distance, since the depth of view of the camera is large enough to see the specific patterns clearly.

To calibrate the extrinsic parameters between the camera and the projector, the additional 3D sensor we used was the Kinect V2. We designed the circle grid patterns as markers that can be projected by the projector, and they are used to find the correspondence between the Kinect 3D points (x^k, y^k, z^k), the camera image points (u^c, v^c), and the projector image points (u^p, v^p), like the setup shown in Fig. 7(a). Figure 7(b) shows an image captured by Kinect in which the RGB image in the color space is projected onto the depth image, from which we can decode the 3D coordinate information for the feature points directly.

The final large-range structured light system setup is as in Fig. 7(a), excluding the Kinect. It consists of one camera and one projector. The baseline between the projector and the camera is ∼286 mm. The specific calibration parameters are as follows:

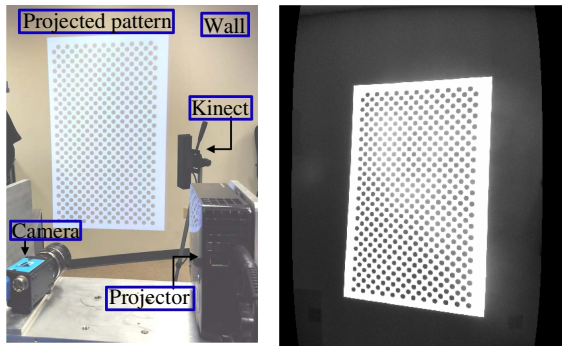
$$\mathbf{A}^c = \begin{bmatrix} 2064.897017 & 0.000000 & 784.018126 \\ 0.000000 & 2068.863052 & 579.626389 \\ 0.000000 & 0.000000 & 1.000000 \end{bmatrix},$$

$$\mathbf{A}^p = \begin{bmatrix} 1972.295665 & 0.000000 & 626.328053 \\ 0.000000 & 1970.495310 & 36.532982 \\ 0.000000 & 0.000000 & 1.000000 \end{bmatrix},$$

$$\mathbf{R}^p = \begin{bmatrix} 0.999797 & -0.017941 & -0.009124 \\ 0.018580 & 0.996963 & 0.075623 \\ 0.007740 & -0.075777 & 0.997095 \end{bmatrix},$$

$$\mathbf{T}^p = \begin{bmatrix} -7.481359 \\ 286.028125 \\ 14.460165 \end{bmatrix}, \quad \mathbf{R}^c = \mathbf{I}, \quad \mathbf{T}^c = \mathbf{0},$$

where $\mathbf{A}^c, \mathbf{A}^p$ are the intrinsic matrices of the camera and the projector, $\mathbf{R}^c, \mathbf{R}^p$ are the rotation matrices of the camera and the projector, and $\mathbf{T}^c, \mathbf{T}^p$ are the translation vectors of the camera and projector.

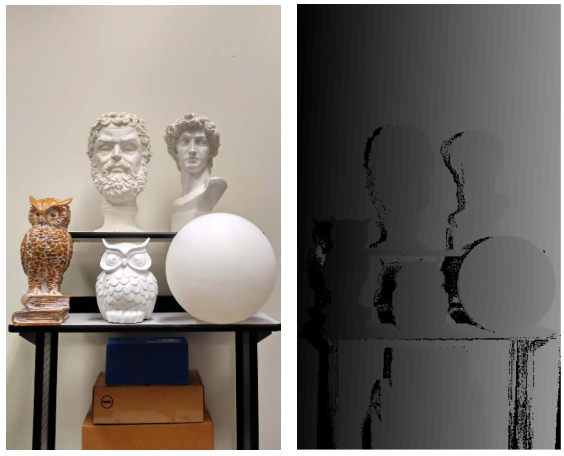


(a) (b)

Fig. 7. System setup and extrinsic calibration. We set both the camera and the projector focus at a far distance and project some circle markers on the wall. Then we can detect positions of circle centers in the camera image. With the help of another 3D scanner (Kinect), we can find the 3D coordinates of those circle centers. Then we can build the point mapping between the projector and the camera with 3D coordinate information. (a) Extrinsic calibration setup. The large-range structured light system is exactly the same excluding the Kinect. (b) Depth image captured by Kinect V2. In this picture, a color image is projected onto the depth one.

We measured a very large scene to test our proposed calibration framework. We lay a scene with multiple sculptures, like a museum. The overall sensing range of the scene is about 1120 mm × 1900 mm × 1000 mm. The photograph of the scene layout is shown in Fig. 8(a). We use square binary phase shifted patterns to do 3D measurement. Since the scene is far away from the projector lens, even a pattern with small fringe period can be quite wide on the imaged scene. Therefore, we use the fringe period $T = 8$ pixels for phase-shifted patterns, and the number of steps is $N = 8$. The unwrapped phase is shown in Fig. 8(b).

The 3D reconstruction result is shown in Fig. 9. For a better visualization and comparison of the accuracy, we captured the same scene using Kinect V2 as well. The reconstruction result of our system is shown in Fig. 9(a) without any filter, and the



(a) (b)

Fig. 8. Measurement on a large scene. The scene is about 1120 mm × 1900 mm × 1000 mm. (a) Layout of the measured scene. (b) Unwrapped phase.

result of the Kinect is shown in Fig. 9(c). Since we used $N = 8$ steps in our 3D reconstruction process, for fair comparison, we captured the same scene eight times using the Kinect and averaged over the eight Kinect 3D geometries to produce the final Kinect 3D result, as shown in Fig. 9(c).

In this experiment, to make the structured light system and Kinect have similar FOV, we put our system at a distance of about 1.8 m from the object, while the Kinect was about 0.6 m from the object, close to the nearest distance it can measure. The specific distance and dimensions of the scene are shown in Figs. 9(b) and 9(d). As we can visually see in Figs. 9(a) and 9(b), our 3D reconstruction produces a more smooth and accurate result. The measurement quality is much better than Kinect even though our system is set up much farther than Kinect V2. Lots of details are kept in our reconstruction result, even in this large sensing area with a long working distance.

To test the accuracy our system can achieve, we picked the sphere in the imaged scene and performed further analysis. It is a matte, white plastic sphere with a 12 in. (304.8 mm) diameter (Vickerman, City of Norwood Young America, Minnesota, USA). An example of the fringe image is shown in Fig. 10(a), and the wrapped phase computed is shown in Fig. 10(b). The zoomed-in 3D reconstruction result from our structured light system is shown in Fig. 10(c) without any filter. The zoomed-in eight-time-averaged 3D reconstruction result of the Kinect is shown in Fig. 10(d).

We analyzed the sphere to verify the accuracy of our system. Shown in Fig. 11, we fit a sphere using the measured 3D data with the preknown diameter of 12 in. (304.8 mm). The sphere fitting results are shown in Figs. 11(a) and 11(b), respectively. Then we computed the error of each 3D point (x_i, y_i, z_i) in the radial direction, that is,

$$e_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2} - R,$$

where (x_0, y_0, z_0) is the sphere center obtained through fitting, and R is the preknown radius (e.g., 152.4 mm). The error maps are shown in Figs. 11(c) and 11(d). The mean measurement error of our system is 0.07 mm, and the standard deviation is 0.80 mm. As a comparison, the mean error for the data obtained from the Kinect V2 is 0.80 mm, and its standard

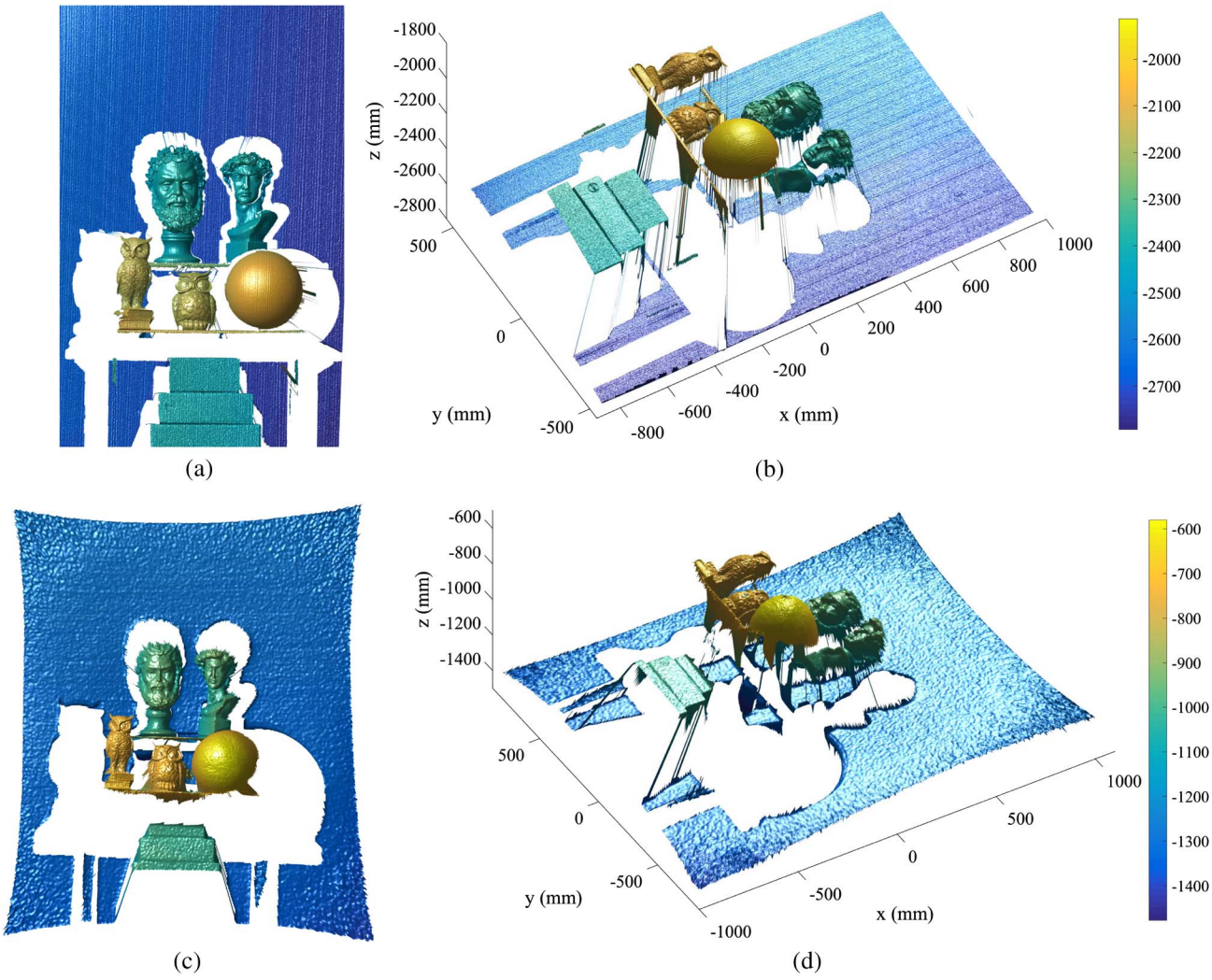


Fig. 9. Measurement result of a large scene. The scene is about 1120 mm × 1900 mm. (a) 3D reconstruction result of the structured light system. (b) Three dimensions of the scene captured by the structured light system. (c) 3D reconstruction result by Kinect V2. (d) Three dimensions of the scene captured by Kinect V2.

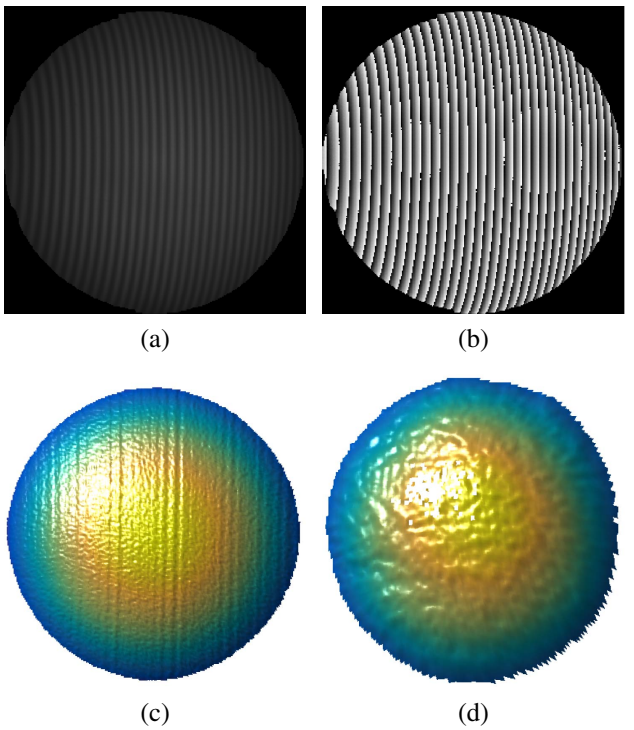


Fig. 10. Further analysis on the sphere. (a) Fringe image captured by the camera. (b) Wrapped phase of the sphere. (c) 3D reconstructed result without any filter of the structured light system. (d) Averaged 3D reconstructed result of Kinect.

deviation is 3.41 mm. These results clearly demonstrate that our system indeed can achieve much higher measurement accuracy than the Kinect V2, although we used it to calibrate our system. Cross sections of the error maps are shown in Figs. 11(e) and 11(f). One may notice that both measurement results still have gross profiles that could be a result of an inaccurate radius (152.4 mm) used for sphere fitting.

For a similar sensing area, the resolution of our camera is 825×1330, and the resolution of the Kinect is 270×512. Roughly, our resolution is about 2 times higher than the Kinect in both width and height directions. For fair comparison, we downsampled our 3D points to make the spatial resolution similar to the Kinect. Figure 12 shows the different results obtained when performing downsampling. Figure 12(a) is our original 3D data. Figure 12(b) is the result of 1/2 sampling (we skip one pixel for each two in both horizontal and vertical directions). Figure 12(c) is the result of 1/3 sampling (we skip two pixels for each three in both horizontal and vertical directions). Figure 12(d) is the result of Kinect.

In the process of downsampling from Figs. 12(a) to 12(c), we found that the result is blurred gradually and some details are lost. Figure 12(c) has similar spatial resolution as the Kinect. By comparing the nose and beard in Figs. 12(c) and 12(d), we can find that Fig. 12(c) conserves more details and is more accurate than Fig. 12(d). By comparing the forehead and face, we can find that Fig. 12(c) is more smooth. The noise and bumps are much less in Fig. 12(c) than in Fig. 12(d). That demonstrates the accuracy of our system and validates the proposed calibration framework.

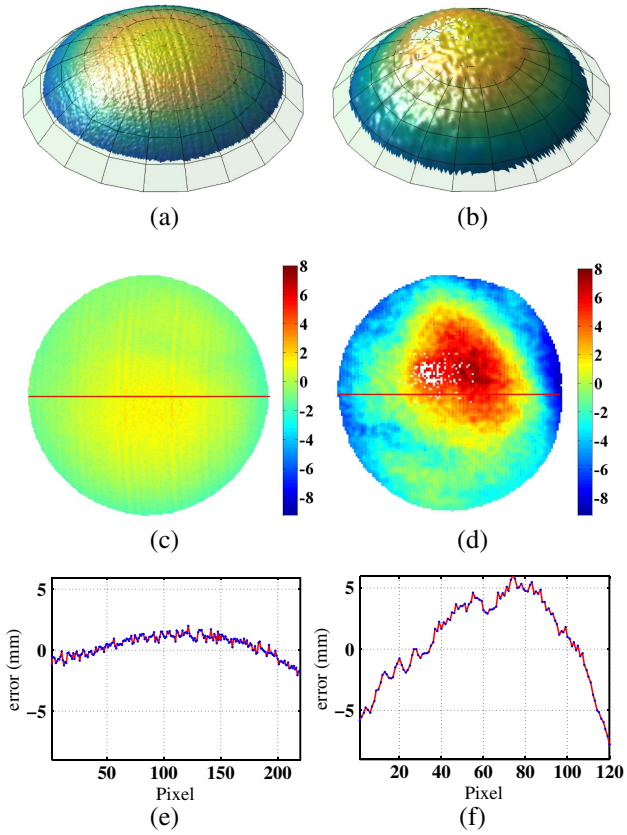


Fig. 11. Error analysis on the sphere we measured. (a) Fitted sphere with a radius of 152.4 mm overlaying the 3D data points measured by our system. (b) Fitted sphere with a radius of 152.4 mm overlaying the 3D data points measured by Kinect. (c) Error maps of 3D data measured by our system (mean error 0.07 mm, standard deviation 0.80 mm). (d) Error map of 3D data acquired by Kinect (mean error 0.80 mm, standard deviation 3.41 mm). (e) Cross section of error map from our system. (f) Cross section of the error map from Kinect.

4. DISCUSSION

Structured light technologies are intensively used in the measurement field; however, they are mostly adopted in close- and small-range applications due to a lack of accurate yet flexible methods for large-range system calibration. This paper proposed a novel calibration framework for the structured light system so that it can be adopted within long- and large-range 3D shape measurement application areas.

The novelties and contributions of the proposed calibration framework can be summarized as follows:

- *Split the whole traditional calibration problem into two stages*. The proposed method takes the advantages of the different difficulty levels of intrinsic and extrinsic calibrations. Specifically, we divided the problem into two stages: the first one is to accurately calibrate intrinsics at a close range; and the second stage is to calibrate the translation and rotation from camera to projector (i.e., extrinsic parameters).
- *Do not need a large calibration board*. Almost all cameras, projectors, and structured light system calibration methods require the use of a calibration target of similar size as the FOV of the device. Such a typical requirement brings challenges for

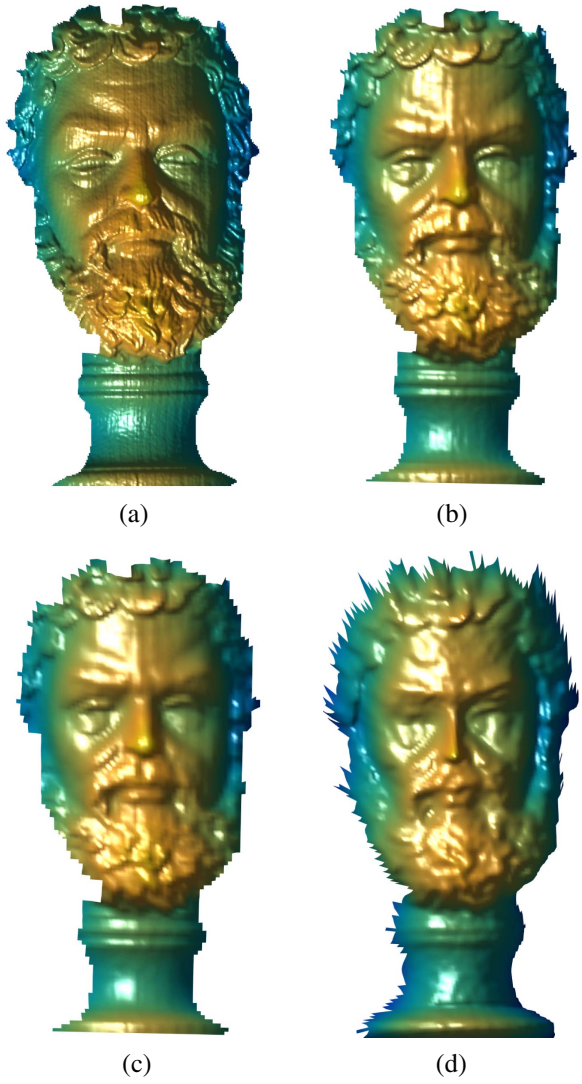


Fig. 12. Downsampling analysis. (a) Original data by the structured light system. (b) 1/2 downsampling based on (a). (c) 1/3 downsampling based on (a). (d) Original data by Kinect.

large-range structured light system calibration since precisely fabricating large calibration targets is often difficult and expensive. In our proposed calibration framework, we use a low-accuracy, large-range 3D sensor (e.g., Microsoft Kinect V2), instead of a large calibration board, to aid in the calibration of the large-range structured light system.

• *High accuracy.* We accurately calibrate intrinsic parameters of the out-of-focus projector and camera at a close range using a more precisely fabricated calibration target, which contributes significantly toward achieving highly accurate measurements. Though we use a low-accuracy, yet large-range, 3D sensor (Microsoft Kinect V2) to calibrate the extrinsic parameters, we are able to detect feature points with subpixel accuracy on the 2D image plane of the camera and projector; as demonstrated by Lavest *et al.* [5], our calibration method can tolerate the rather large measurement error of the 3D feature points provided by the Kinect V2. We have developed a large-scale 3D shape measurement system with a FOV of 1120 mm × 1900 mm × 1000 mm. Experiments demonstrate our system can achieve measurement accuracy as high as 0.07 mm with a standard deviation of 0.80 mm by measuring a 304.8 mm diameter sphere. As a comparison, Kinect V2 only achieved a mean error of 0.80 mm with a standard deviation of 3.41 mm for the FOV of measurement.

As we mentioned in Section 2, this framework can be more generic when the following works are combined:

• *Camera intrinsic calibration*. For some cases, a far focal length lens could be used. The conventional calibration method may fail in those cases. If a far focal length lens is used, one can use the out-of-focus camera calibration approach discussed by Bell and Zhang [26], which uses an LCD panel to calibrate the defocused camera in a structured light system.

• *Extrinsic calibration with a complex 3D scene*. As long as we find the correspondence between camera pixel, projector pixel, and 3D coordinates, we can calibrate the extrinsic matrix between the projector and the camera with standard stereo calibration toolboxes. For extrinsic calibration, the 3D scene can be some complex environment, not necessarily a wall or a flat object. A possible way for calibrating within a complex 3D scene is to project both horizontal and vertical phase-shifting patterns onto the scene. Then the regular camera and Kinect can be used to capture those fringe images and do phase computation and unwrapping simultaneously. We can use phase to find the correspondence of (u^c, v^c), (u^p, v^p) and the corresponding Kinect 3D information (x^k, y^k, z^k).

5. SUMMARY

This paper presents a calibration framework for the large-range structured light system. The calibration method does not need a large calibration target, which is both complicated and expensive in fabrication. Specifically, we split the whole system calibration into two stages. The first one is to accurately calibrate intrinsics of the camera and projector. We use a regular size calibration board to perform intrinsic calibration at a near range where both camera and projector are out of focus. The second stage is to calibrate the translation and rotation from camera to projector with the assistance of a low-accuracy, large-range 3D sensor (e.g., Microsoft Kinect). We did experiments on a large scene to demonstrate the accuracy and capacity of the proposed calibration framework. The large scene in our experiment is about 1120 mm × 1900 mm × 1000 mm. By measuring a 304.8 mm diameter sphere, it shows that our system can achieve accuracy as high as 0.07 mm with a standard deviation of 0.80 mm. As a comparison, Kinect V2 only achieved a mean error of 0.80 mm with a standard deviation of 3.41 mm.

Funding. Directorate for Engineering (ENG) (CMMI-1521048).

†These authors contributed equally to this work.

REFERENCES

1. C. B. Duane, "Close-range camera calibration," Photogramm. Eng. **37**, 855–866 (1971).
2. I. Sobel, "On calibrating computer controlled cameras for perceiving 3-D scenes," Artif. Intell. **5**, 185–198 (1974).

3. R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," IEEE J. Robot. Autom. **3**, 323–344 (1987).
4. Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1330–1334 (2000).
5. J. Lavest, M. Viala, and M. Dhome, "Do we really need an accurate calibration pattern to achieve a reliable camera calibration?" in *Computer Vision ECCV'98* (Springer, 1998), pp. 158–174.
6. A. Albarelli, E. Rodolà, and A. Torsello, "Robust camera calibration using inaccurate targets," Trans. Pattern Anal. Mach. Intell. **31**, 376–383 (2009).
7. K. H. Strobl and G. Hirzinger, "More accurate pinhole camera calibration with imperfect planar target," in *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)* (IEEE, 2011), pp. 1068–1075.
8. L. Huang, Q. Zhang, and A. Asundi, "Flexible camera calibration using not-measured imperfect target," Appl. Opt. **52**, 6278–6286 (2013).
9. C. Schmalz, F. Forster, and E. Angelopoulou, "Camera calibration: active versus passive targets," Opt. Eng. **50**, 113601 (2011).
10. L. Huang, Q. Zhang, and A. Asundi, "Camera calibration with active phase target: improvement on feature detection and optimization," Opt. Lett. **38**, 1446–1448 (2013).
11. Y. Wen, S. Li, H. Cheng, X. Su, and Q. Zhang, "Universal calculation formula and calibration method in Fourier transform profilometry," Appl. Opt. **49**, 6563–6569 (2010).
12. Y. Xiao, Y. Cao, and Y. Wu, "Improved algorithm for phase-to-height mapping in phase measuring profilometry," Appl. Opt. **51**, 1149–1155 (2012).
13. Y. Villa, M. Araiza, D. Alaniz, R. Ivanov, and M. Ortiz, "Transformation of phase to (x, y, z)-coordinates for the calibration of a fringe projection profilometer," Opt. Laser Eng. **50**, 256–261 (2012).
14. Q. Hu, P. S. Huang, Q. Fu, and F.-P. Chiang, "Calibration of a three-dimensional shape measurement system," Opt. Eng. **42**, 487–493 (2003).
15. X. Mao, W. Chen, and X. Su, "Improved Fourier-transform profilometry," Appl. Opt. **46**, 664–668 (2007).
16. E. Zappa and G. Busca, "Fourier-transform profilometry calibration based on an exhaustive geometric model of the system," Opt. Lasers Eng. **47**, 754–767 (2009).
17. H. Guo, M. Chen, and P. Zheng, "Least-squares fitting of carrier phase distribution by using a rational function in fringe projection profilometry," Opt. Lett. **31**, 3588–3590 (2006).
18. H. Du and Z. Wang, "Three-dimensional shape measurement with an arbitrarily arranged fringe projection profilometry system," Opt. Lett. **32**, 2438–2440 (2007).
19. L. Huang, P. S. Chua, and A. Asundi, "Least-squares calibration method for fringe projection profilometry considering camera lens distortion," Appl. Opt. **49**, 1539–1548 (2010).
20. M. Vo, Z. Wang, B. Pan, and T. Pan, "Hyper-accurate flexible calibration technique for fringe-projection-based three-dimensional imaging," Opt. Express **20**, 16926–16941 (2012).
21. R. Legarda-Sáenz, T. Bothe, and W. P. Jüptner, "Accurate procedure for the calibration of a structured light system," Opt. Eng. **43**, 464–471 (2004).
22. S. Zhang and P. S. Huang, "Novel method for structured light system calibration," Opt. Eng. **45**, 083601 (2006).
23. Z. Li, Y. Shi, C. Wang, and Y. Wang, "Accurate calibration method for a structured light system," Opt. Eng. **47**, 053604 (2008).
24. Y. Yin, X. Peng, A. Li, X. Liu, and B. Z. Gao, "Calibration of fringe projection profilometry with bundle adjustment strategy," Opt. Lett. **37**, 542–544 (2012).
25. D. Han, A. Chimienti, and G. Menga, "Improving calibration accuracy of structured light systems using plane-based residual error compensation," Opt. Eng. **52**, 104106 (2013).
26. T. Bell and S. Zhang, "Method for out-of-focus camera calibration," Appl. Opt. **55**, 2346–2352 (2016).
27. B. Li, N. Karpinsky, and S. Zhang, "Novel calibration method for structured light system with an out-of-focus projector," Appl. Opt. **53**, 3415–3426 (2014).
28. S. Lei and S. Zhang, "Flexible 3-D shape measurement using projector defocusing," Opt. Lett. **34**, 3080–3082 (2009).