MULTI-SCALE, MULTI-MODAL, HIGH-SPEED 3D SHAPE MEASUREMENT

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Yatong An

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2019

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Prof. Song Zhang, Chair

School of Mechanical Engineering

Prof. Xinyan Deng

School of Mechanical Engineering

Prof. David Cappelleri

School of Mechanical Engineering

Prof. Gary Cheng

School of Industrial Engineering

Approved by:

Prof. Jay Gore

Head of the School of Mechanical Engineering Graduate Program

ACKNOWLEDGMENTS

I would like to take this opportunity to thank everyone who has helped me in this dissertation and throughout my Ph.D. study. No words can express how grateful I am for their generous support and tremendous help along the way of my research life at Purdue University.

First, I would like to thank my advisor, Prof. Song Zhang, who provided me this opportunity to do this innovative and exciting research. Without this opportunity offered by Prof. Zhang, I could never be able to do research in this amazing field of optical 3D measurement. Also, Prof. Zhang's close guidance and invaluable instructions have influenced me a lot in the way of my research. Because of Prof. Zhang, I was able to broaden my horizons, to fulfill new potential and to push my limits in my research.

Meanwhile, I would like to thank my Ph.D. advisory committee members, Prof. Xinyan Deng, Prof. David Cappelleri and Prof. Gary Cheng, for their tremendous help and generous support. They provided a lot of great suggestions on this research and gave me many invaluable insights during this dissertation work. Their help has benefited me very much and advanced this dissertation drastically. I am very thankful to their generosity and willingness to help me in this dissertation research.

Next, I would like to thank all the past and present team members I have had the privilege of working with: Prof. Tyler Bell, Prof. Junfei Dai, Prof. Beiwen Li, Prof. Min Zhong, Michael Crawford, Michael Feller, Jae-Sang Hyun, Xiaowei Hu, Chufan Jiang, Ziping Liu, Betrice Lim, Bogdan Vlahov, Duo Wang, and Peter Zibley. They were always very kind and patient to discuss problems and share ideas. It was an amazing experience working with them together. Without their tremendous help, it could take me longer to finish this dissertation.

Last but not least, I would like to thank every friend who supported and encouraged me. Moreover, I would like to give special thanks to my parents and sister, who always back me up and cheer me up.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

ABSTRACT

An, Yatong Ph.D., Purdue University, May 2019. Multi-Scale, Multi-Modal, High-Speed 3D Shape Measurement. Major Professor: Prof. Song Zhang, School of Mechanical Engineering.

With robots expanding their applications in more and more scenarios, practical problems from different scenarios are challenging current 3D measurement techniques. For instance, infrastructure inspection robots need large-scale and high-spatial-resolution 3D data for crack and defect detection, medical robots need 3D data well registered with temperature information, and warehouse robots need multi-resolution 3D shape measurement to adapt to different tasks. In the past decades, a lot of progress has been made in improving the performance of 3D shape measurement methods. Yet, measurement scale and speed and the fusion of multiple modalities of 3D shape measurement techniques remain vital aspects to be improved for robots to have a more complete perception of the real scene. In this dissertation, we will focus on the digital fringe projection technique, which usually can achieve high-accuracy 3D data, and expand the capability of that technique to complicated robot applications by 1) extending the measurement scale, 2) registering with multi-modal information, and 3) improving the measurement speed of the digital fringe projection technique.

The measurement scale of the digital fringe projection technique mainly focused on a small scale, from several centimeters to tens of centimeters, due to the lack of a flexible and convenient calibration method for a large-scale digital fringe projection system. In this study, we first developed a flexible and convenient large-scale calibration method and then extended the measurement scale of the digital fringe projection technique to several meters. The meter scale is needed in many large-scale robot applications, including large infrastructure inspection. Our proposed method includes two steps: 1) accurately calibrate intrinsics (i.e., focal lengths and principal points)

with a small calibration board at close 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-scale 3D sensor (e.g., Microsoft Kinect). The two-step strategy avoids fabricating a large and accurate calibration target, which is usually expensive and inconvenient for doing pose adjustments. With a small calibration board and a low-cost 3D sensor, we calibrated a large-scale 3D shape measurement system with a FOV of $(1120 \times 1900 \times 1000)mm^3$ and verified the correctness of our method.

Multi-modal information is required in applications such as medical robots, which may need both to capture the 3D geometry of objects and to monitor their temperature. To allow robots to have a more complete perception of the scene, we further developed a hardware system that can achieve real-time 3D geometry and temperature measurement. Specifically, we proposed a holistic approach to calibrate both a structured light system and a thermal camera under exactly the same world coordinate system, even though these two sensors do not share the same wavelength; and a computational framework to determine the sub-pixel corresponding temperature for each 3D point, as well as to discard those occluded points. Since the thermal 2D imaging and 3D visible imaging systems do not share the same spectrum of light, they can perform sensing simultaneously in real time. We developed a hardware system that achieved real-time 3D geometry and temperature measurement at 26Hz with 768×960 points per frame.

In dynamic applications, where the measured object or the 3D sensor could be in motion, the measurement speed will become an important factor to be considered. Previously, people projected additional fringe patterns for absolute phase unwrapping, which slowed down the measurement speed. To achieve higher measurement speed, we developed a method to unwrap a phase pixel by pixel by solely using geometric constraints of the structured light system without requiring additional image acquisition. Specifically, an artificial absolute phase map Φ_{min}, at a given virtual depth plane $z = z_{min}$, is created from geometric constraints of the calibrated structured light system, such that the wrapped phase can be pixel-by-pixel unwrapped by referring to Φ_{min}. Since Φ_{min} is defined in the projector space, the unwrapped phase obtained from this method is an absolute phase for each pixel. Experimental results demonstrate the success of this proposed novel absolute-phase unwrapping method. However, the geometric constraint-based phase unwrapping method using a virtual plane is constrained in a certain depth range. The depth range limitations cause difficulties in two measurement scenarios: measuring an object with larger depth variation, and measuring a dynamic object that could move beyond the depth range. To address the problem of depth limitation, we further propose to take advantage of an additional 3D scanner and use additional external information to extend the maximum measurement range of the pixel-wise phase unwrapping method. The additional 3D scanner can provide a more detailed reference phase map Φ_{ref} to assist us to do absolute phase unwrapping without the depth constraint. Experiments demonstrate that our method, assisted by an additional 3D scanner, can work for a large depth range, and the maximum speed of the low-cost 3D scanner is not necessarily an upper bound of the speed of the structured light system. Assisted by Kinect V2, our structured light system achieved 53Hz with a resolution 1600×1000 pixels when we measured dynamic objects that were moving in a large depth range.

In summary, we significantly advanced the 3D shape measurement technology for robots to have a more complete perception of the scene by enhancing the digital fringe projection technique in measurement scale (space domain), speed (time domain), and fusion with other modality information. This research can potentially enable robots to have a better understanding of the scene for more complicated tasks, and broadly impact many other academic studies and industrial practices.

1. INTRODUCTION

3D shape information can be applied in many robot problems, such as inspection, picking, and placing. With robot applications expanding into more and more complicated scenes, a more complete perception of the scene is needed for robots to finish more challenging tasks with high autonomy and efficiency. For instance, infrastructure inspection robots may need to measure with both a large scale and high spatial resolution to improve inspection efficiency and accuracy. Surgery robots need accurate 3D shape information that is well registered with temperature information to handle complex surgeries and to avoid potential thermal injury. Over the past decades, different types of 3D shape measurement methods, such as time of flight (TOF) [1, 2], laser triangulation [3], stereo vision [4, 5], and structured light [6], have been developed. Yet none of them has become capable enough to satisfy all various requirements for all robot applications. The performance of current 3D shape measurement methods still needs to be improved to face more and more complicated challenges from robot applications. In this dissertation, we will focus on the digital fringe projection technique, which can usually achieve highly accurate 3D measurement, and we will extend the capability of this technique in three aspects: 1) measurement scale (space domain); 2) speed (time domain); and 3) registration with multi-modal information, such that robots can have a more complete perception of the scene for complicated tasks.

In this chapter, we will provide an overview of the dissertation. Specifically, Section 1.1 introduces the motivations behind our research, and Section 1.2 summarizes our objectives for this dissertation. Lastly, the overall organization of this dissertation is introduced in Section 1.3.

1.1 Motivations

In this section, we will introduce three robotics applications: inspection robots, medical robots, and warehouse robots. In each application, we will explain the challenges faced by current 3D shape measurement methods, which motivate our research in this dissertation.

1.1.1 Inspection robots

Flaws, such as cracks, can happen on the surface of infrastructures. Early detection and subsequent repair can prevent infrastructure flaws from propagating internally and causing further degradation. Manual inspection is time-consuming and costly, and it can be dangerous to humans. Nowadays, robots are more and more widely adopted for infrastructure detection and inspection tasks. For instance, a quadcopter was designed for vertical infrastructure inspection [7], and mobile robots were adopted for tunnel inspection [8] and other indoor and outdoor tasks. 3D sensors can be mounted on robots when doing inspections, such as Xbox Kinect for road pothole and crack inspection [9]. However, the accuracy of current commercial sensors is limited. For instance, the accuracy of the Kinect V2 is at the millimeter level, and the noise of the Kinect V2 can be larger than 10mm when tested on a planar object positioned at 1m from the device [10]. According to the Federal Highway Administration, minor and moderate cracks are less than 3.2mm, and severe cracks are over 3.2mm in scale [11]. Large noise and low accuracy make current commercial sensors, such as Kinect, difficult to apply in minor and moderate crack detection. Although laser scanning techniques have high accuracy, they usually provide a stack of cross-section measurements, which could neglect vertical or horizontal cracks. Structured light techniques based on digital fringe projection could measure an area with sub-millimeter accuracy to meet requirements for minor or moderate crack inspection. However, they are typically adopted in small or regular scales (a workspace size of several millimeters to tens of centimeters) due to the lack of flexible system calibration methods for larger-scale workspaces. Developing a large-scale system calibration method and extending the measurement area of structured light techniques are necessary for minor and moderate crack detection in infrastructure inspection applications.

1.1.2 Medical robots

Due to the involvement of robots, robotic surgery could be done with precision, miniaturization, and smaller incisions. For instance, Da Vinci surgical systems have been applied in many surgery applications, including cardiac surgery, thoracic surgery, and craniotomy. Bone drilling is needed in some surgery applications, such as craniotomy. During bone drilling, the position and angle of the drill need to be adjusted to avoid bleeding and brain penetration [12]. 3D shape measurement can provide the robot with feedback to adjust the position and angle of the drill. Meanwhile, heat generated during drilling could cause thermal injury to the cortex and thermal necrosis. It is reported that the critical level at which thermal necrosis appears is when the bone is exposed to a temperature of around 56°C over a time of 10 seconds [13]. Therefore, temperature needs to be monitored during the surgery process. By mapping the thermal image onto a 3D shape measurement result, we can get a quantitative analysis of the thermal distribution, such as the heat flux per square inch. The quantitative analysis can provide feedback to control the speed, angle, feed, and depth of the drill and make it form a closed loop system. Therefore, developing a system that can measure both 3D and temperature information will be of interest in those robotic surgery applications.

1.1.3 Warehouse robots

Warehouse robots have been used to tackle monotonous tasks such as picking orders from shelves, picking from bins, and loading parts. For instance, there was the

Amazon robot picking challenge[1] to pick daily supplies from shelves in a warehouse, IAM robots were designed for complex picking at human-level speed[2], and Midwest Engineered Systems developed a robotic system for random bin picking and part loading to move heavy parts from multiple bins and place them onto a conveyor to begin a heat-treating operation[3].

In those warehouse robots, 3D scanners could be mounted on moving robot arms for object detection and positioning. For instance, people can mount commercial 3D sensors, such as Kinect and Real Sense, on robot arms for picking applications [14, 15]. To accurately segment densely crowded objects (such as crowded shelves in a daily supply warehouse), structured light 3D sensors using digital fringe projection techniques could achieve better performance than current commercial sensors due to their high resolution and high accuracy. However, digital fringe projection systems of comparable prices with current commercial 3D scanners usually have constraints on measurement speed, which would decrease the efficiency of warehouse robots. To increase 3D measurement efficiency using low-cost hardware, a new, faster 3D measurement algorithm is required.

1.2 Objectives

Facing challenges and requirements from the above motivation stories, we set the objectives of this dissertation to be the following:

- **Develop a method for large-scale structured light system calibration.** Calibration is one of the important reasons that structured light systems are not widely used for large-scale 3D shape measurement. To achieve accurate 3D shape measurements, the system first needs to be properly calibrated. The most extensively adopted calibration method involves capturing a planar calibration target with predefined features at different poses. This method can achieve good

[1] `https://www.amazonrobotics.com/#/pickingchallenge`

[2] `https://www.iamrobotics.com/`

[3] `http://mwes.com/portfolio/random-bin-picking-and-part-loading-system/`

calibration accuracy, yet it is primarily applied in a small-scale 3D system. To achieve good calibration accuracy for a large-scale 3D system, the calibration target is preferably similar in size to the system's field of view (FOV), which is challenging since a large and accurate calibration target is expensive and difficult to fabricate and inconvenient to use for pose adjustment. In this research, we aim to develop a flexible and convenient calibration method for a large-scale 3D measurement system, and to use this method to calibrate a 3D measurement system to do a large-scale shape measurement with high accuracy. The details of this research will be introduced in Chapter 3.

- **Develop a method for simultaneous 3D surface geometry measurement and temperature measurement in real time.** As mentioned in Sec. 1.1.2, in some medical robot applications, both 3D information and thermal information are important. A thermal map well registered with 3D information could provide quantitative analysis of thermal distribution for feedback control. However, a thermal camera and a regular camera have different modalities in terms of sensing spectra, image resolutions, and lens distortions. Those differences will create challenges for calibration between the thermal camera and the 3D sensor and make it difficult to accurately map the thermal information onto the 3D geometry. In this dissertation, we aim to develop a holistic approach to calibrating both the thermal camera and the 3D sensor under exactly the same world coordinate system, and to develop a computational framework to determine the sub-pixel corresponding temperature for each 3D point as well as discard occluded points caused by different perspectives between the thermal camera and the 3D sensor. Finally, we would like to build a hardware system to achieve real-time 3D geometry measurement and temperature measurement simultaneously and to register the two types of information together. More details of this research will be presented in Chapter 4.

- **Increase the speed of digital fringe projection techniques through novel shape measurement algorithms.** Current extensively applied fringe analysis methods usually only provide a wrapped phase value ranging from $-\pi$ to π, and a phase unwrapping algorithm is needed to remove those 2π ambiguities to obtain a continuous unwrapped phase map. To obtain an absolute unwrapped phase map, additional fringe patterns are usually needed, such as in multiple-wavelength phase unwrapping algorithms and gray coding phase unwrapping algorithms. However, projecting additional fringe patterns will slow down the measurement speed. To increase the measurement speed, one possible way is to take advantage of some inherent geometric constraints between a projector and a camera in a digital fringe projection (DFP) system for absolute phase unwrapping, such that additional patterns can be avoided. In this dissertation, we will explore the possibility to unwrap a phase pixel by pixel solely using geometric constraints of the structured light system, such that the shape measurement speed can be improved. Details on using geometric constraints for phase unwrapping will be explained in Chapter 5.

- **Enhance the performance of digital fringe projection techniques in a multi-resolution shape measurement system.** Nowadays, robots may have multiple sensors to adapt to different environments and tasks, such as a multi-resolution 3D measurement system for overall 3D model generation and small detail scanning in a crime scene. Those multiple sensors could mutually benefit each other to achieve better performance. Therefore, in a departure from the previous objectives of enlarging the measurement range and increasing the speed solely via algorithms without external hardware assistance, we would like to explore the possibility of taking advantage of another 3D sensor in a multi-resolution system to enhance the performance of our digital fringe projection techniques in terms of the measurement range and the speed. Particularly, this dissertation aims to develop a multi-resolution 3D shape measurement system and enhance the measurement range and speed of the DFP system with assistance of another 3D sensor. We will illustrate the details of this research in Chapter 6.

1.3 Dissertation organization

In Chapter 2, we will do a literature review on 3D shape measurement methods and summarize state-of-the-art approaches for enlarging the measurement scale, increasing the measurement speed, and fusing 3D data with other modality information. Chapter 3 will introduce our developed large-scale calibration framework and present our large-scale 3D measurement results. The high-resolution, real-time simultaneous 3D geometry and temperature measurement will be illustrated in Chapter 4, along with a holistic method of calibration between a thermal and regular camera, and the results of dynamic 3D geometries well registered with temperature information. Chapter 5 will introduce our pixel-wise absolute phase unwrapping method without additional patterns, which increases the measurement speed. In Chapter 6, we will show our developed multi-resolution 3D shape measurement system and explain our work to enhance the performance of digital fringe projection techniques by taking advantage of another 3D scanner in the multi-resolution system. Finally, Chapter 7 will summarize the contributions of this dissertation and provide insight about future research directions.

2. LITERATURE REVIEW

2.1 Relevant literature on 3D shape measurement techniques

Depending on whether active illumination is needed or not, optical 3D shape measurement methods can be classified into passive and active methods. Passive methods do not need active illumination, and they include depth from defocus [16] and stereo vision [4, 5]. Active methods take advantage of active illumination, and they include time of flight [1, 2], laser triangulation [3], and structured light [6].

2.1.1 Passive methods

Passive methods usually work well if an object has rich texture, but their accuracy will be compromised for objects with uniform texture or low texture variations. In this section, we will mainly introduce the stereo vision technique, which is more related to our research.

Stereo vision

Stereo vision [4] imitates the human vision system of two eyes by using two cameras. Figure 2.1 schematically shows a stereo vision system that captures a real scene from two different perspectives. The real-world 3D point P and its projection onto two camera images P^l, P^r form a triangle, and this geometric relationship can be used to calculate 3D coordinates of P combined with system calibration results [5].

Given two images from the two cameras in a stereo vision system, correspondence searching between two camera images needs to be done before 3D coordinates are calculated. The mathematical theory for correspondence searching is epipolar geometry, which constrains the corresponding pixels of one camera image pixel to be a line on the other camera image. Figure 2.1 illustrates the *epipolar geometry*. In Figure 2.1, O^l and O^r are the focal points of the left camera and right camera, respectively. E^l and E^r, which are the two intersection points between the line O^lO^r and two camera images, are called *epipoles* in the stereo vision system. For a pixel P^l on a left camera image, it can have multiple points P_1,P_2,P_3,P_4 in a 3D space, and all those 3D points will fall on the same line L^r on the other camera image. Similarly, all pixels on the line L^l can only correspond to points on the line L^r on the other camera image. The lines L^l and L^r are called *epipolar lines* in the stereo vision system. The main advantage of using epipolar geometry is to improve computational efficiency when searching for correspondence between two images. With the epipolar geometry constraint, the correspondence searching problem can be reduced from the original 2D image search to a 1D problem.

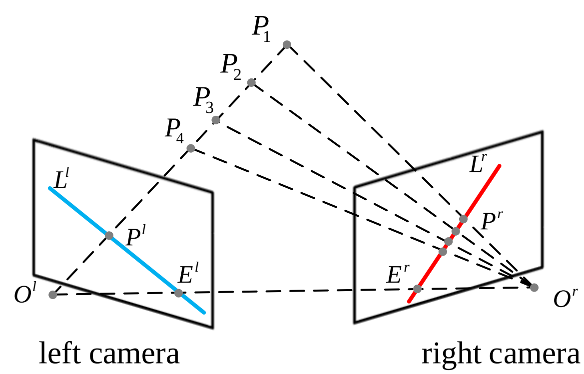


Figure 2.1. Epipolar geometry constraint: one pixel on one image can only correspond to one line (called an epipolar line) on the other image.

Based on epipolar geometry, numerous works are further done to precisely locate a correspondence pixel, and these methods can be classified into local-based algorithms and global optimization algorithms. Local-based algorithms usually have higher speed and lower accuracy, and global optimization algorithms are usually more time-consuming but more accurate [17, 18]. With the availability of open data sets, such as KITTI [19] and Middleburry [20], machine learning methods [21] now play an important role in correspondence matching.

Because of easy system setup, stereo vision has been intensively applied in mobile robots [22, 23], urban 3D semantic modeling [24], and 3D face recognition [25]. Yet, since stereo vision relies on an object's natural texture to find correspondence, it usually does not work for objects with uniform texture and homogeneous object regions, such as a white wall.

2.1.2 Active methods

Active methods are usually less sensitive to object surface properties, since they are mainly based on the active illumination from an emitter. In this section, we will introduce time of flight, laser triangulation and structured light techniques respectively.

Time of flight

Time of flight technologies imitate bats' echolocation system, in which bats make an ultrasonic call and listen to the returning echoes to estimate the distance. Similarly, in TOF technologies, a light pulse is emitted from a laser source, then received by a detector. The distance can be calculated as $c \times \Delta t/2$, where c is the light speed, Δt is the time difference between light pulse emitting and receiving, and division by 2 is because the overall process is a round trip. Time of flight technologies can be divided into two categories: optical shutter approach and intensity modulation approach. Optical shutter methods use the time difference between emitting and receiving directly, and intensity modulation methods use the phase difference between emitting and receiving light [1].

The main advantage of the time of flight sensor is its compact design, since the emitter and detector could have the same view angle [26]. Yet, because the speed of light is extremely fast (about 3×10^8 m/s), time of flight sensors have difficulty achieving very high depth resolution [27]. Accuracies of current commercial 3D sensors, such as Kinect V2 [28] and Swiss Ranger SR4000 [29], are usually at the millimeter level, even though these sensors find their applications in many fields. For instance, Kinect V2 is usually applied in human computer interaction (HCI) applications [30] and somatosensory games [31]. Light detection and ranging (LIDAR) devices [32], which also usually use time of flight principles, can be found in applications including autonomous ground vehicles [33], aerial flights [34], and outdoor robotics [35].

Laser triangulation

Laser triangulation [36] usually consists of a laser light source, a detector, and a lens that focuses the emitted light to the detector. To perform the measurement, laser triangulation systems first need to be calibrated. Then, a laser point or line is emitted onto the object surface. The detector, such as a CCD array [37] or CMOS array [38], will capture the scene, and the laser point or line will be extracted from images. Since the laser source, object, and detector form a triangle, the depth of the object can be retrieved by triangulation combined with the calibration information [39].

Laser triangulation can usually achieve high-depth resolution, and they can do surface measurement for manufacturing inspections [40] and closed-loop feedback control systems, where high accuracy is required. However, laser triangulation techniques usually sample a scene line by line or point by point, which slows down the measurement speed and makes it difficult to measure dynamically moving or deformable objects.

Structured light

Structured light techniques project some patterns onto objects and use those patterns for correspondence searching. The structured light system is very similar to a stereo vision system, except that structured light usually uses a projector to project patterns. Instead of relying on the natural texture of objects, structured light systems usually find correspondence by analyzing the projected patterns, which could be more robust and more reliable.

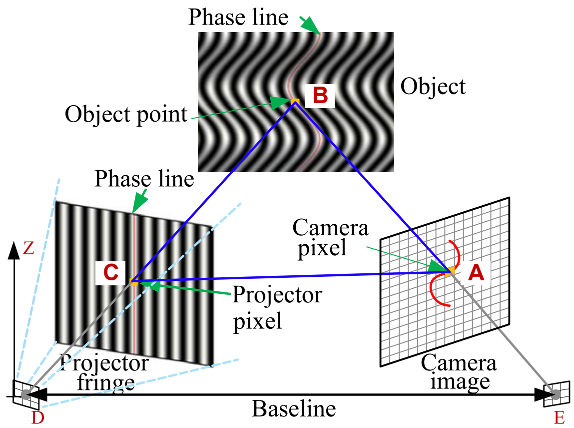


Figure 2.2. Principles of the digital fringe projection technique [41]. The system consists of a projector and a camera. The projector will project stripe patterns onto the object, and patterns will be distorted because of the object's geometry. By analyzing the distortion information of the stripes, the 3D geometry of the object can be recovered.

Digital fringe projection is one commonly used structured light technique. Figure 2.2 is an example of a digital fringe projection system. By projecting some stripe patterns onto an object, the correspondence between the projector and camera images can be established through fringe analysis. There are two main sets of methods to do fringe analysis: Fourier Transform (FT)-based methods [42, 43] and phase-shifting-based methods [44]. FT-based methods can usually achieve high speed, since only one fringe image is needed. However, these are usually adopted to measure relatively smooth surfaces without too much texture variations. Phase-shifting methods are more robust to texture variations and ambient illumination change because of multiple fringe images.

In N-step phase shifting algorithms, the intensities of the kth fringe image can be described as:

$$I_k(x, y) = I'(x, y) + I''(x, y)\cos(\phi + 2k\pi/N), \tag{2.1}$$

where I' is the average intensity, I'' is the intensity modulation, and ϕ is the phase to be solved for. Using a least-square method, we can obtain

$$\phi = -\tan^{-1}\left[\frac{\sum_{k=1}^{N} I_k \sin(2k\pi/N)}{\sum_{k=1}^{N} I_k \cos(2k\pi/N)}\right]. \tag{2.2}$$

Since an inverse tangent function is used, the phase values obtained from this equation only vary from $-\pi$ to $+\pi$. The phase with 2π modulus should be unwrapped to obtain a continuous phase map, and this process is called *phase unwrapping*. Essentially, phase unwrapping algorithms are trying to determine an integer number $k(x, y)$ for each pixel and unwrap the phase by adding $k(x, y)$ multiples of 2π, and $k(x, y)$ is also called *fringe order*. There are numerous spatial [45] or temporal [46] phase unwrapping algorithms. The spatial phase unwrapping typically only provides phase values relative to a point on the phase map, and this unwrapped phase map is called a relative phase map. Temporal phase unwrapping can provide absolute phase information by coding the fringe order information in additional patterns.

In temporal phase unwrapping methods, additional patterns need to be projected to do phase unwrapping, such as simple binary coding or gray coding. Figure 2.3 is an illustration of the simple binary coding method. For each continuous stripe region on the wrapped phase map, a unique code can be retrieved by binarizing the additional fringe images, such as 000, 001, 010, shown in Figure 2.3. These codes encode a fringe order from which we can determine how many 2π's we should add to the wrapped phase value to obtain an absolute phase map. Gray coding [47, 48], shown in Figure 2.3(b), is a more robust binary coding method compared with simple binary coding, shown in Figure 2.3(a), since there is only one code change in the 2π discontinuity locations on the wrapped phase map. In the red dashed boxes, simple binary coding changes three times, but gray coding changes only once.

However, the more additional patterns projected, the slower the measurement speed. To reduce the number of additional patterns needed, more advanced algorithms are developed. Li et al. [49] proposed using a single additional image consisting of six types of slits to do phase unwrapping. Those slits can form a pseudorandom

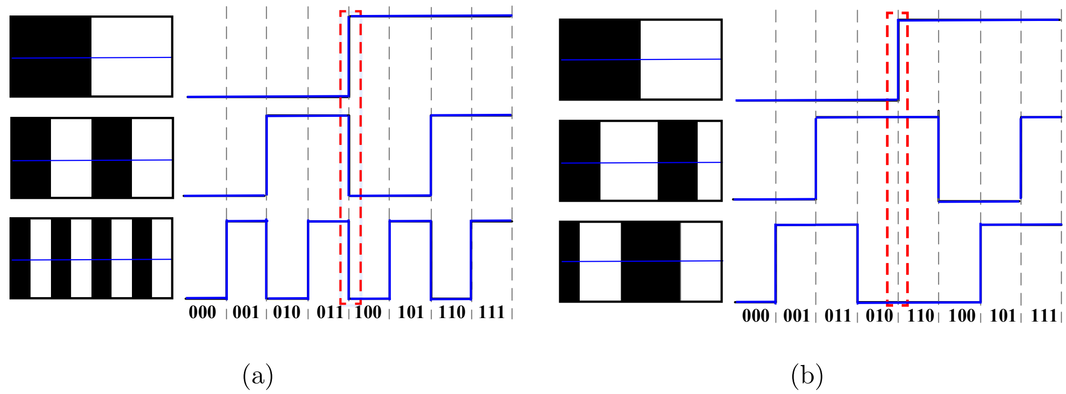


Figure 2.3. Example of using binary coding for phase unwrapping. (a) Example of simple binary coding; (b) Example of gray coding.

sequence, and the fringe order can be retrieved by checking the position of a subsequence of those slits from a whole sequence. Zhang [50] attempted to use a single stair image in which the stair changes at the position of the 2π jumps, such that fringe orders can be retrieved from the single stair image. Zuo et al. [51] encoded the wrapped phase and base phase into four fringe patterns to reduce the total number of patterns. An and Zhang [52] proposed to use one additional random pattern and combine the geometric constraints between camera and projector for absolute phase unwrapping.

2.2 Relevant literature on large-scale measurement

Methods to do large-scale measurement can be classified into software-based methods and hardware-based methods. Software-based methods try to develop some algorithms without changing the hardware. In contrast, hardware-based methods try to improve the hardware system to measure a large-scale area.

Software-based methods include stitching algorithms. In stitching algorithms, researchers try to capture or scan small patches of an object and stitch them together. To stitch two images, an error metric is usually chosen to compare pixels between the two images. Based on the error metric, a *search* technique will be applied to find the corresponding pixel in the other image, and that correspondence can be used to estimate motion parameters, which will be used to transform one image to the same coordinate system as the other one [53]. 2D image stitching methods can be classified into direct (pixel-based) alignment [54], feature-based registration [55], and global registration methods [56].

Stitching methods evolve from 2D images to 3D geometries. Depending on whether markers are necessary or not, 3D stitching methods can be classified into marker-based and markerless methods. Marker-based algorithms can generally achieve better accuracy and are usually used in applications such as manufacturing inspection and quality control [57, 58]. Markerless methods usually rely on natural features of the object to merge 3D data, and they can be applied to reconstruct a building or holistic large statue [59, 60].

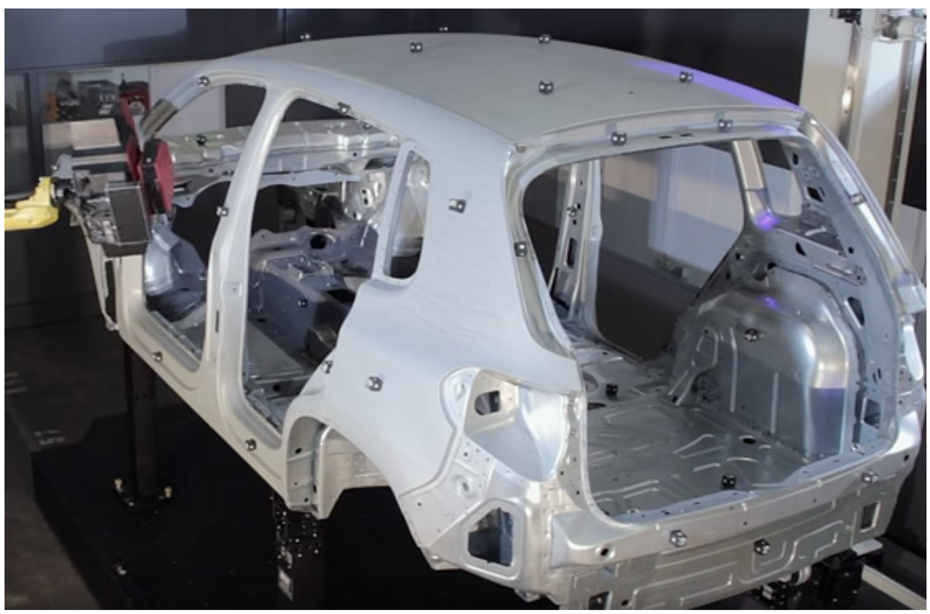


Figure 2.4. Example of 3D stitching with markers in a manufacturing application [61].

Hardware-based algorithms try to improve the hardware system to sense a larger area. For instance, in the 2D imaging field, wide-angle cameras were designed and are used in automotive applications [62]. One advantage of hardware-based algorithms is that they do not need to do post-processing, since the they can measure a large scene directly in one frame.

For 3D shape measurement using structured light techniques, people have attempted to use stitching algorithms to do a large-scale measurement from the software side, as shown in Figure 2.4. Yet for some applications, capturing a large scene directly in one frame is required, such that we also need to improve the structured light techniques to allow the system to measure a large scene directly in one frame, instead of stitching patches together.

2.3 Relevant literature on multi-modal information fusion

To do multi-modal fusion, usually we can use two categories of methods. The first category is based on image registration techniques. Given two images of different modalities, these methods will do registration between the two images to find a mapping relationship between them. The second category of methods are based on calibration techniques. These methods try to calibrate the sensors in advance, such that they can know the mapping relationship between images from different sensors.

Methods based on multi-modal image registration usually rely on feature extraction and feature matching between different modality images. It is very crucial for these registration methods to choose the right similarity criteria [63]. This category of methods was previously adopted in the medical field [64].

Methods based on calibration try to calibrate different sensors in advance. The main idea when calibrating sensors of multiple modalities is usually to let different sensors sense some common land markers or feature points. Based on pair-wise data, iterative registration and optimization algorithms are adopted to compute the extrinsic parameters between these sensors [65]. Based on this idea, people have calibrated laser rangefinder and gray cameras [66,67], depth and color cameras [68], and thermal and infrared cameras [69]. Figure 2.5 is an example of a project that tries to calibrate a laser scanning device and thermal camera for data fusion. Also, different types of targets are designed to calibrate 3D sensor and thermal images [70–73]. Examples of calibration targets are shown in Figure 2.6.

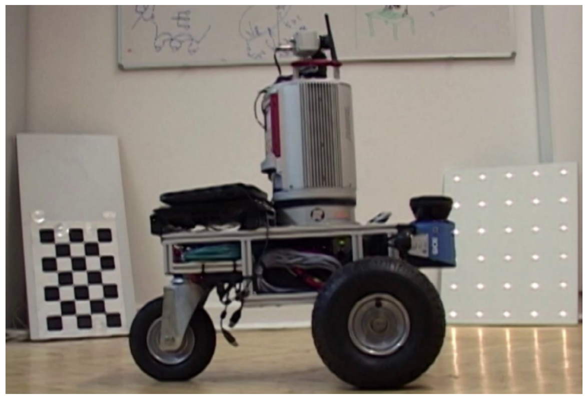


Figure 2.5. A project that tries to calibrate a laser scanning device and thermal camera for data fusion [74].

(a) (b) (c)

Figure 2.6. Different calibration boards for thermal camera calibration. (a) Calibration board using different materials [70]; (b) Calibration board with holes [71]; (c) Calibration board with heat sources [75].

Yet, there still lacks a method to calibrate the structured light system and thermal cameras for 3D and temperature data fusion. Among various 3D shape measurement methods, structured light usually can achieve high accuracy, high resolution, and high speed. Fusing the 3D from a structured light system with temperature information could have a lot of applications, including medical applications, as discussed in Chapter 1.

2.4 Relevant literature on high-speed 3D shape measurement

To achieve higher speed, there are usually two options: to use better hardware, and to improve the 3D measurement algorithm.

A straightforward method for high-speed measurement is to use better hardware, such as high-speed cameras. However, better hardware means increasing cost. For instance, the price of high-speed cameras can reach tens of thousands of dollars (such as Phantom VEO cameras with thousands of frames per second), while regular-speed cameras cost several hundred dollars or even less. Similarly, high-speed projectors could cost thousands of dollars, and regular commercial projectors cost several hundred dollars or less.

Developing more efficient algorithms could increase the measurement speed without increasing any hardware costs. For instance, people developed more efficient stereo matching algorithms [76] for stereo vision systems. Lei and Zhang [77] developed binary defocusing techniques for high-speed 3D measurement using digital fringe projection techniques.

Among 3D shape measurement methods discussed in Section 2.1, structured light using digital fringe projection techniques can usually achieve high accuracy. However, structured light devices are usually very expensive. For instance, the cost of a Gom Atom core is as high as $35,000[1]. Thus, decreasing the hardware cost and developing fast measurement algorithms are of great interests in this field.

[1] `https://www.atos-core.com/index.php`

2.5 Conclusion

In this chapter, we introduced state-of-the-art 3D shape measurement methods including both passive (stereo vision) and active (TOF, laser triangulation and structured light) methods. Active methods utilize active illumination, and they are less sensitive to object surface properties. Among those active methods, structured light that uses digital fringe projection devices has gained a lot of attraction because of its flexibility and accuracy. Also, we reviewed and summarized current methods for enlarging the measurement area of 3D measurement systems, and did a literature review on multi-modal systems that can measure 3D and temperature simultaneously. In addition, relevant literature on high-speed 3D shape measurement was summarized in this chapter. Although a lot of progress has been achieved in 3D shape measurement technology, with the increasing demand and expectations in the robotics field, a large-scale 3D measurement system with multi-modal information in high speed is needed, and our progress toward that system will be introduced in the following from Chapter 3 to Chapter 6.

3. METHOD FOR LARGE-SCALE STRUCTURED LIGHT SYSTEM CALIBRATION

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-scale structured light system calibration since fabricating large calibration targets is difficult and expensive. This chapter presents a large-scale 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-scale three dimensional (3D) sensor (e.g., Microsoft Kinect). We have developed a large-scale 3D shape measurement system with a FOV of 1120mm$\times$1900mm$\times$1000mm. Experiments demonstrate our system can achieve measurement accuracy as high as 0.07mm with a standard deviation of 0.80mm by measuring a 304.8mm diameter sphere. As a comparison, Kinect V2 only achieved mean error of 0.80mm with a standard deviation of 3.41mm for the FOV of measurement. The major content of this chapter was originally published in *Applied Optics* [78] (also listed as journal article [J3] in "LIST OF PUBLICATIONS").

3.1 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-scale 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-scale 3D shape measurement is due to a lack of an accurate yet flexible calibration method for such a scale.

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 [79, 80]. In the 1980s, Tsai [81] 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 [82] 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 [83–86]. Furthermore, active digital displays, such as liquid crystal display (LCD), have also been employed for accurate camera calibration [87, 88].

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 [89–91]. 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 [92–94]. Another approach is to estimate the relationship between the depth and encoded information (e.g., phase) through optimization [95–98].

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. [99] 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 [100] 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 [101], bundle adjustment [102], or residual error compensation with planar constraints [103].

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-scale 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 scale measurement applications.

This chapter presents a calibration method that does not require an equivalent size calibration target to the sensing FOV but rather uses a large-scale 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-scale, 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) [104], 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 [105]. 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 largescale 3D shape measurement can measure a FOV of 1120mm$\times$1900mm$\times$1000mm. Experiments demonstrate our system can achieve measurement accuracy as high as 0.07mm with a standard deviation of 0.80mm by measuring a 304.8mm diameter sphere. As a comparison, Kinect V2 only achieved a mean error of 0.80mm with a standard deviation of 3.41mm for the FOV of measurement.

Section 6.2 explains the principles of the proposed calibration method. Section 6.3 presents experimental results to further validate the proposed method. Section 3.4 discusses the advantages and possible limitations of the proposed calibration method. Finally, Section 5.5 summarizes this chapter.

3.2 Principle

This section thoroughly explains the principles of the proposed large-scale 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-scale structured light system calibration.

3.2.1 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

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & \gamma & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R}, & \mathbf{T} \end{bmatrix} \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}, \tag{3.1}$$

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:

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}. \tag{3.2}$$

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:

$$\mathbf{D} = \begin{bmatrix} k_1 & k_2 & p_1 & p_2 & k_3 \end{bmatrix}^T, \tag{3.3}$$

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:

$$\begin{aligned} u' &= u(1 + k_1 r^2 + k_2 r^4 + k_3 r^6), \\ v' &= v(1 + k_1 r^2 + k_2 r^4 + k_3 r^6), \end{aligned} \tag{3.4}$$

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:

$$\begin{aligned} u' &= u + [2p_1 uv + p_2(r^2 + 2u^2)], \\ v' &= v + [p_1(r^2 + 2v^2) + 2p_2 uv]. \end{aligned} \tag{3.5}$$

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.

3.2.2 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 kth fringe image I_k can be described as

$$I_k(x, y) = I'(x, y) + I''(x, y)\cos(\phi + 2i\pi/N), \tag{3.6}$$

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

$$\phi(x, y) = -\tan^{-1}\left[\frac{\sum_{i=1}^{N} I_k \sin(2i\pi/N)}{\sum_{i=1}^{N} I_k \cos(2i\pi/N)}\right]. \tag{3.7}$$

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:

$$\Phi(x, y) = \phi(x, y) + K(x, y) \times 2\pi. \tag{3.8}$$

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.

3.2.3 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 [100]. But for a large-scale structured light system, fabricating a very large calibration board at the projectors focus range and to fit the projectors FOV is both difficult and expensive. As in Fig. 3.1, 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. [105] 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-scale structured light system in its defocused area. As shown in Fig. 3.1, 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 [77]. 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:

$$v^p = \phi_v^c(u^c, v^c) \times T^h/(2\pi). \tag{3.9}$$

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:

$$u^p = \phi_h^c(u^c, v^c) \times T^v/(2\pi). \tag{3.10}$$

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.

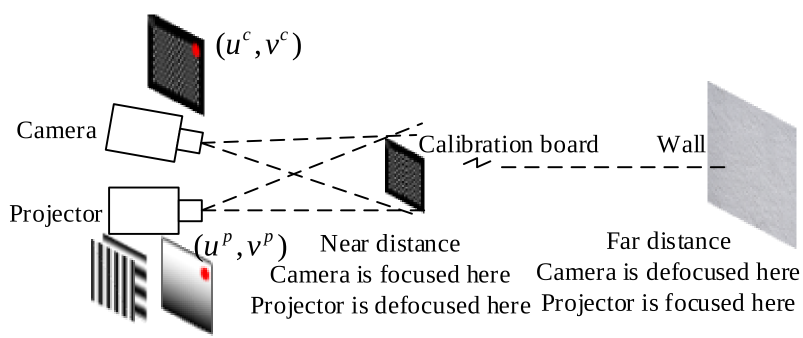


Figure 3.1. Out-of-focus projector calibration. Since fabricating a very large calibration board at the projector's focus range (wall) is both difficult and expensive, we instead calibrate the projector at its defocus range. In the near defocus range, we can use a regular-size calibration board to calibrate the projector.

3.2.4 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 scale, we next adjust the cameras focus and angle with respect to the projector. The cameras 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. 3.2, 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 [104]. 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.

3.2.5 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

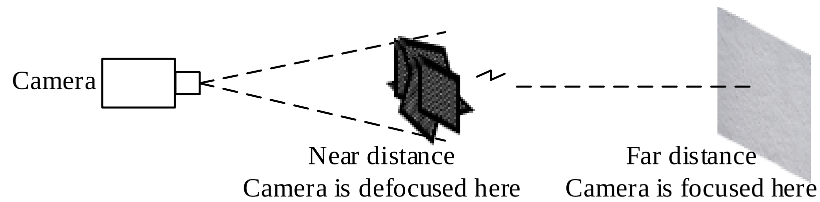


Figure 3.2. Camera intrinsic calibration. For a near focused lens, even when the lens is focused at infinity, the level of camera lens defocusing is still not enough to fail a conventional camera calibration approach. Given this, the camera can be calibrated directly by capturing different poses of a calibration board at its near defocus range.

works well for small scale structured light systems, yet the calibration board will be too small for the large-scale structured light system, as shown in Fig. 3.3. 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.

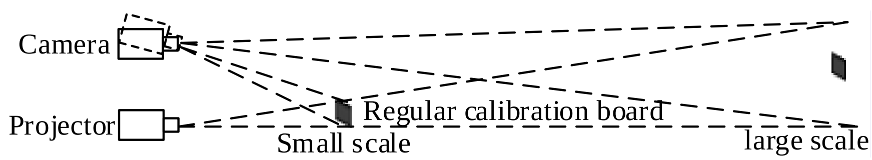


Figure 3.3. Extrinsic calibration explanation. For extrinsic calibration in the small scale, usually we can put a regular calibration board in the working zone (the eventual capture area of the large-scale system). For large-scale extrinsic calibration, the regular calibration board will be too small, and it is not practical to fabricate a very large calibration board, let alone, use it for flexible calibration.

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-scale 3D sensor (e.g., Microsoft Kinect V2). As shown in Fig. 3.4, 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 Kinects 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 OpenCVs Calibration Toolbox.

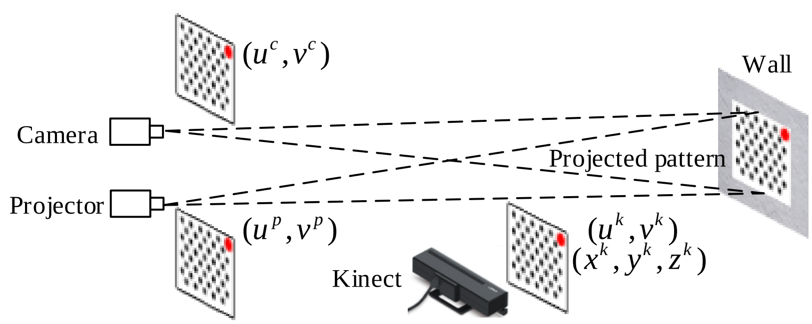


Figure 3.4. Extrinsic calibration principle. We project some markers onto a real 3D scene, then we use the camera and Kinect to capture it at the same time. So for each marker, we can get its position in the camera image coordinate system (u^c, v^c), its position in the projector image coordinate system (u^p, v^p), and its 3D coordinate (x^k, y^k, z^k) in the Kinect space.

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

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

3.2.6 Overall framework of large-scale structured light system calibration

Here we summarize the entire framework of our proposed large-scale calibration method. Briefly speaking, we split the whole traditional calibration problem into two stages to make it adaptable to a large-scale 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-scale structured light system calibration is shown in Fig. 3.5.

- *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-scale 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.

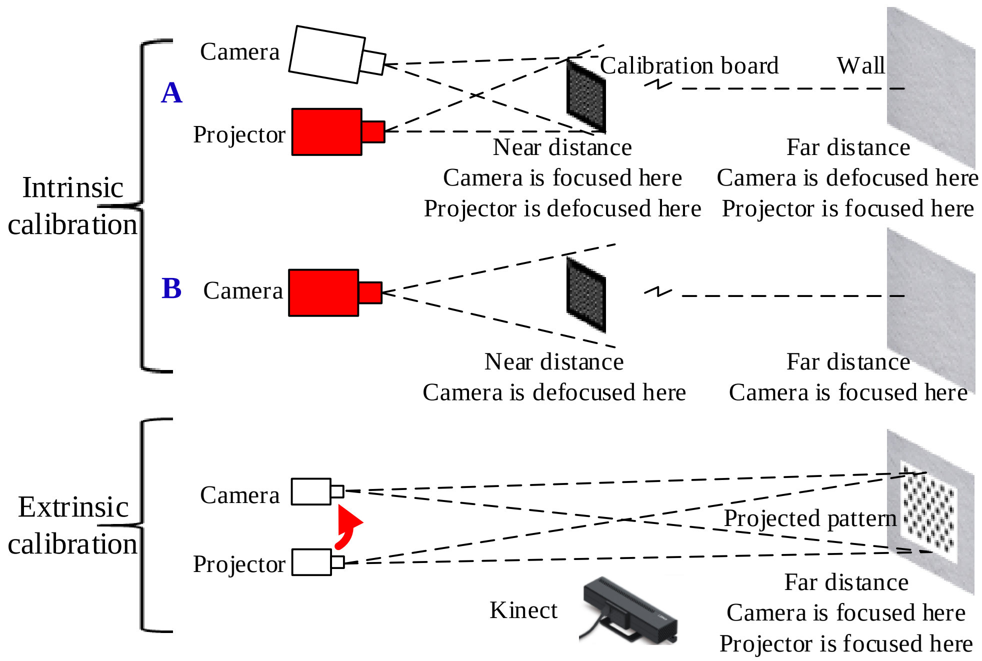


Figure 3.5. Overall framework for large-scale structured light system calibration. The proposed method includes two stages: 1) accurately calibrate intrinsics (i.e., focal lengths, principle points) at a near range where both camera and projector are out of focus; and 2) calibrate the extrinsic parameters (translation and rotation) from camera to projector with the assistance of a low-accuracy large-scale 3D sensor (e.g., Microsoft Kinect).

1-B: Camera intrinsic calibration. Now adjust the camera focus to be at the far distance (the eventual capture area of the large-scale 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 [104].

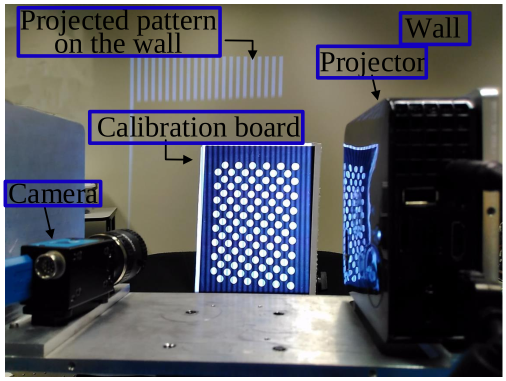
- *Stage 2: Extrinsic calibration.* Set the system properly (e.g., changing the distance and angle between the projector and camera) for large-scale 3D shape measurement. Project some specific patterns with feature points (u^p, v^p) onto a large 3D scene with the projector. Let the camera and Kinect 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 Kinects 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.

3.3 Experiment

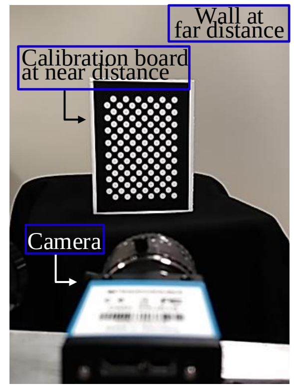
To verify the performance of the proposed method, we developed a structured light system that includes a complementary metal-oxide-semiconductor camera (Model: DMK23UX174) with a 12mm focal length lens (Model: Computar M1214-MP2). The resolution of the camera is set to be 1600×1200 pixels. The projector 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 0.5m∼4.5m.

We followed the framework proposed in Section 3.2.6 to calibrate the system. Figure 3.6 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. 3.6(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 3.6(b) shows the setup for camera intrinsic calibration.

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



(a)



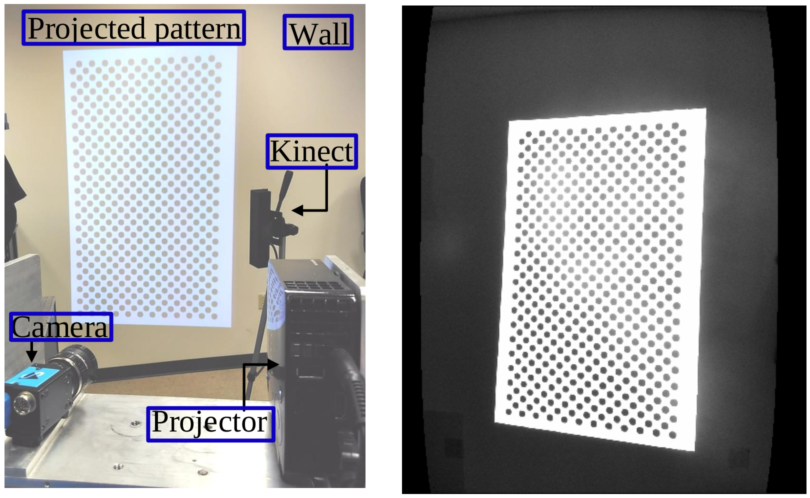
(b)

Figure 3.6. 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; (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.

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. 3.7(a). Figure 3.7(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-scale structured light system setup is as in Fig. 3.7(a), excluding the Kinect. It consists of one camera and one projector. The baseline between the projector and the camera is ∼286mm. The specific calibration parameters are as follows:

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



(a) (b)

Figure 3.7. 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-scale structured light system is exactly the same excluding the Kinect; (b) Depth image captured by Kinect V2. In this picture, color image is projected onto the depth one.

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

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

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

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.

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 scale of the scene is about 1120mm$\times$1900mm$\times$1000mm. The photograph of the scene layout is shown in Fig. 3.8(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. 3.8(b).

(a) (b)

Figure 3.8. Measurement on a large scene. The scene is about $1120 \times 1900 \times 1000\text{mm}^3$. (a) Layout of the measured scene; (b) Unwrapped phase.

The 3D reconstruction result is shown in Fig. 3.9. 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. 3.9(a) without any filter, and the result of the Kinect is shown in Fig. 3.9(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. 3.9(c).

In this experiment, to make the structured light system and Kinect have similar FOV, we put our system at a distance of about 1.8m from the object, while the Kinect was about 0.6m from the object, close to the nearest distance it can measure. The specific distance and dimensions of the scene are shown in Figs. 3.9(b) and 3.9(d). As we can visually see in Figs. 3.9(a) and 3.9(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 12in. (304.8mm) diameter (Vickerman, City of Norwood Young America, Minnesota, USA). An example of the fringe image is shown in Fig. 3.10(a), and the wrapped phase computed is shown in Fig. 3.10(b). The zoomed-in 3D reconstruction result from our structured light system is shown in Fig. 3.10(c) without any filter. The zoomed-in eight-time-averaged 3D reconstruction result of the Kinect is shown in Fig. 3.10(d).

We analyzed the sphere to verify the accuracy of our system. Shown in Fig. 3.11, we fit a sphere using the measured 3D data with the preknown diameter of 12in. (304.8mm). The sphere fitting results are shown in Figs. 3.11(a) and 3.11(b), respectively. Then we computed the error of each 3D point (x_i, y_i, z_i) in the radial direction, that is,

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

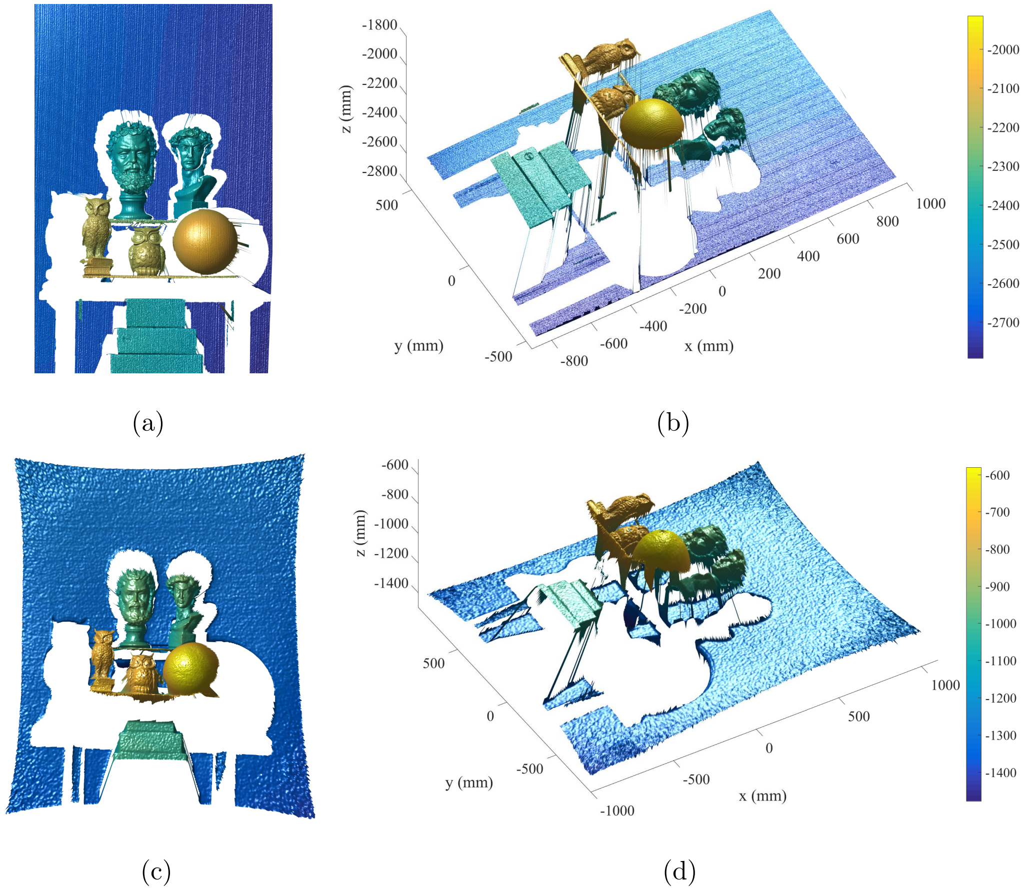
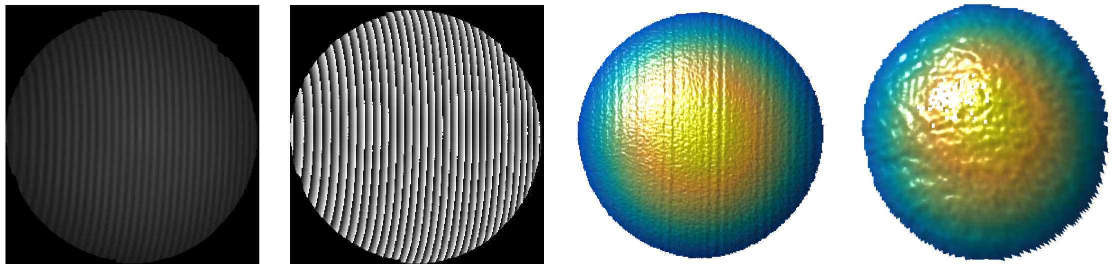


Figure 3.9. Measurement result of a large scene. The scene is about 1120×1900 mm^2. (a) 3D reconstruction result of the structured light system; (b) Three dimensions of the scene captured by the structured light system; (c) 3D reconstruction result by Kinect V2; (d) Three dimensions of the scene captured by Kinect V2.

where (x_0, y_0, z_0) is the sphere center obtained through fitting, and R is the preknown radius (e.g., 152.4mm). The error maps are shown in Figs. 3.11(c) and 3.11(d). The mean measurement error of our system is 0.07mm, and the standard deviation is 0.80mm. As a comparison, the mean error for the data obtained from the Kinect V2 is 0.80mm, and its standard deviation is 3.41mm. 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



(a) (b) (c) (d)

Figure 3.10. Further analysis on the sphere. (a) The fringe image captured by the camera; (b) The wrapped phase of the sphere; (c) 3D reconstructed result without any filter of the structured light system; (d) Averaged 3D reconstructed result of Kinect.

maps are shown in Figs. 3.11(e) and 3.11(f). One may notice that both measurement results still have gross profiles that could be a result of an inaccurate radius (152.4mm) 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 3.12 shows the different results obtained when performing downsampling. Figure 3.12(a) is our original 3D data. Figure 3.12(b) is the result of 1/2 sampling (we skip one pixel for each two in both horizontal and vertical directions). Figure 3.12(c) is the result of 1/3 sampling (we skip two pixels for each three in both horizontal and vertical directions). Figure 3.12(d) is the result of Kinect.

In the process of downsampling from Figs. 3.12(a) to 3.12(c), we found that the result is blurred gradually and some details are lost. Figure 3.12(c) has similar spatial resolution as the Kinect. By comparing the nose and beard in Figs. 3.12(c) and 3.12(d), we can find that Fig. 3.12(c) conserves more details and is more accurate than Fig. 3.12(d). By comparing the forehead and face, we can find that Fig. 3.12(c) is more smooth. The noise and bumps are much less in Fig. 3.12(c) than in Fig. 3.12(d).

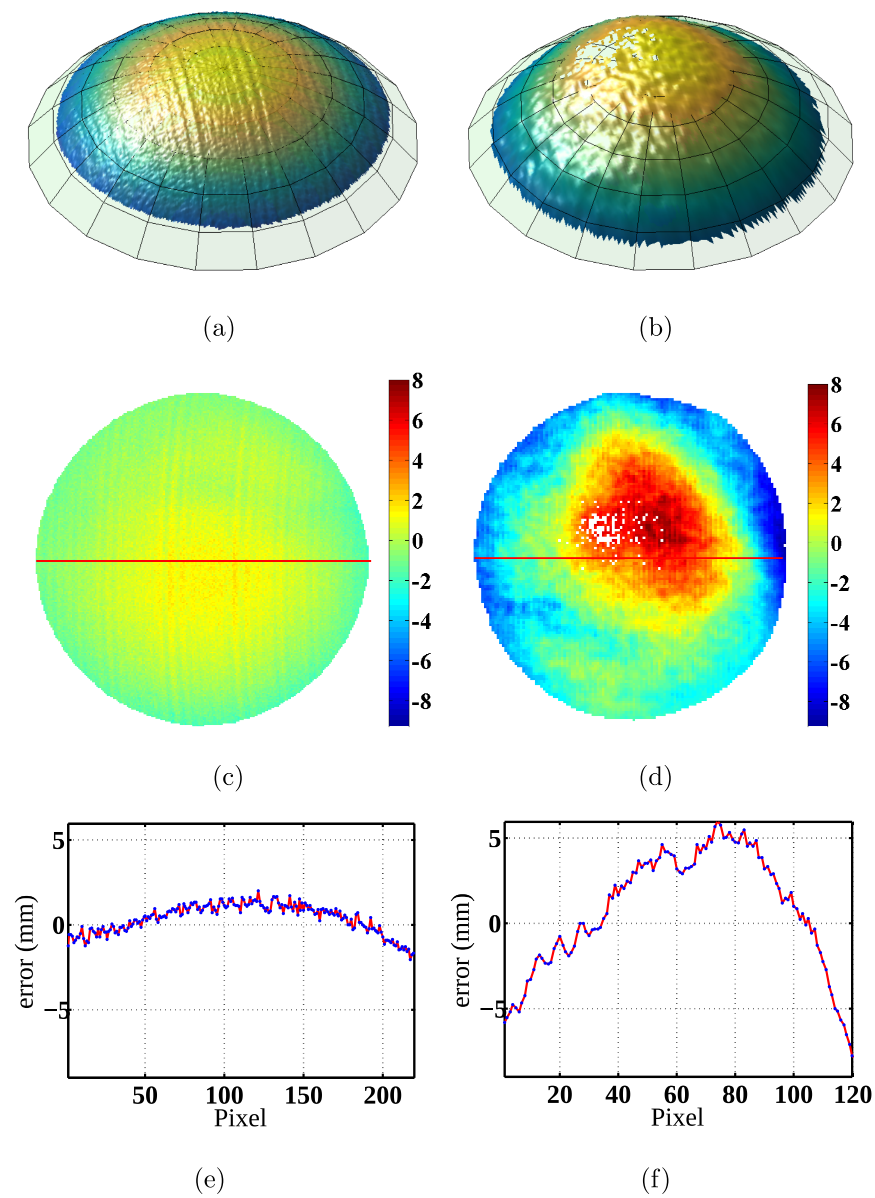


Figure 3.11. Error analysis on the sphere we measured. (a) Fitted sphere with a radius of 152.4mm overlaying the 3D data points measured by our system; (b) Fitted sphere with a radius of 152.4mm overlaying the 3D data points measured by Kinect; (c) Error maps of 3D data measured by our system (mean error 0.07mm, standard deviation 0.80mm); (d) Error map of 3D data acquired by Kinect (mean error 0.80mm, standard deviation 3.41mm); (e) Cross section of error map from our system; (f) Cross section of the error map from Kinect.

That demonstrates the accuracy of our system and validates the proposed calibration framework.

(a) (b) (c) (d)

Figure 3.12. 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.

3.4 Discussion

Structured light technologies are intensively used in the measurement field; however, they are mostly adopted in close and small scale applications due to a lack of accurate yet flexible methods for large-scale system calibration. This chapter proposed a novel calibration framework for the structured light system so that it can be adopted within long and large-scale 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 large-scale 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-scale 3D sensor (e.g., Microsoft Kinect V2), instead of a large calibration board, to aid in the calibration of the large-scale 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-scale, 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. [83], 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 1120mm$\times$1900mm$\times$1000mm. Experiments demonstrate our system can achieve measurement accuracy as high as 0.07mm with a standard deviation of 0.80mm by measuring a 304.8mm diameter sphere. As a comparison, Kinect V2 only achieved a mean error of 0.80mm with a standard deviation of 3.41mm for the FOV of measurement.

As we mentioned in Section 6.2, 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 [104], 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).

3.5 Summary

This chapter has presented a calibration framework for the large-scale 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 used 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-scale 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 was about 1120mm×1900mm×1000mm. By measuring a 304.8mm diameter sphere, it showed that our system could achieve accuracy as high as 0.07mm with a standard deviation of 0.80mm. As a comparison, Kinect V2 only achieved a mean error of 0.80mm with a standard deviation of 3.41mm.

4. HIGH-RESOLUTION, REAL-TIME SIMULTANEOUS 3D SURFACE GEOMETRY AND TEMPERATURE MEASUREMENT

The previous chapter introduced our developed method for large-scale structured light system calibration. In some robotics applications, such as medical robot surgery, not only 3D shape information, but also temperature information are important. This chapter will present a method to simultaneously measure three dimensional (3D) surface geometry and temperature in real time. Specifically, we developed 1) a holistic approach to calibrate both a structured light system and a thermal camera under exactly the same world coordinate system even though these two sensors do not share the same wavelength; and 2) a computational framework to determine the sub-pixel corresponding temperature for each 3D point as well as discard those occluded points. Since the thermal 2D imaging and 3D visible imaging systems do not share the same spectrum of light, they can perform sensing simultaneously in real time: we developed a hardware system that can achieve real-time 3D geometry and temperature measurement at 26Hz with 768×960 points per frame. The majority of this chapter was originally published in *Optics Express* [106] (also listed as journal article [J1] in "LIST OF PUBLICATIONS").

4.1 Introduction

Real-time measurement of a 3D geometric shape is vital for numerous applications including manufacturing, medical practices, and more [107]; temperature sensing using a thermal imaging camera is also of great interest to benefit both scientific research and industrial practices [71, 74, 108, 109]. We believe that the combination of these two sensing modalities can substantially increase applications.

Static and real-time 3D shape measurement have been extensively studied over the past decades. 3D shape measurement techniques use different principles to achieve different capabilities. In general, 3D shape measurement techniques include stereo vision [5], laser triangulation [3], time of flight [1, 2] (e.g., Microsoft Kinect 2), structured light [6] (e.g., Intel RealSense and Microsoft Kinect 1), as well as shape from focus/defocus [16]. Among these methods, stereo vision and shape from focus/defocus do not require active illumination, and thus are regarded as passive methods. The passive methods can work well if an object surface has rich texture information, yet their accuracy will be compromised if a surface is uniform or has low texture variations. In contrast, those methods requiring active illumination are less sensitive to surface properties since 3D reconstruction is mainly based on the emission sent out from the emitter. Among those active methods, structured light techniques that use digital video projection devices to project computer generated structured patterns are popular because of their flexibility and accuracy.

The active structured light method can work well for both visible and near infrared wavelengths, yet they cannot work for longer wavelengths at which the silicon-based sensing devices fail to operate (e.g., thermal spectrum). Therefore, to the best of our knowledge, there are no systems that can simultaneously measure 3D geometric shape and surface temperature in real time. One major challenge is that the regular camera and the thermal camera do not *see* the same wavelength, thus it is difficult to calibrate these two types of cameras under the same coordinate system. Furthermore, commercially available, relatively inexpensive thermal cameras have low resolutions and large distortions, making the mapping between a thermal camera and a regular camera challenging.

This chapter proposes a method to address the aforementioned challenges. To conquer these challenges, the calibration of these two cameras has to be carried out under the same world coordinate system and preferably uses the same calibration target. We propose a method that allows these two types of cameras to *see* the same object features. The basic idea is that we use a heat lamp to shine thermal energy on a black/white circle pattern calibration target. Due to the emissivity differences of black and white areas, the thermal camera and the regular camera can see the same calibration features (e.g., circles). By this means, these two cameras can be calibrated under the same world coordinate system using the same calibration target. Since the regular camera and the projector share the same spectrum, the structured light system can be calibrated using the same circle patterns. By coinciding the world coordinate system with the regular camera lens coordinate system, the whole system including the thermal camera is calibrated under the same world coordinate system. Since thermal cameras usually have a much lower resolution and larger distortion than a regular camera, we developed a computational framework to achieve sub-pixel corresponding temperature mapping point for each 3D point, and discard those occluded 3D points that are not visible to the thermal camera. Two different hardware systems have been developed to verify the performance of the proposed method: 1) a static system that has a resolution of 1280×1024 points per frame; and 2) a real-time system that can achieve simultaneous 3D geometric shape and surface temperature measurement at 26Hz with a resolution of 768×960 pixels per frame.

4.2 Principle

4.2.1 Least-square phase-shifting algorithm

Phase-shifting algorithms are extensively used in optical metrology because of their speed and accuracy. There are many phase-shifting algorithms such as three-step, four-step, five-step, etc. Generally, the more steps used, higher accuracy phase is obtained due to the averaging effect. For an N-step phase-shifting algorithm with equal phase shifts, the intensities of the kth fringe image can be described as:

$$I_k(x, y) = I'(x, y) + I''(x, y)\cos(\phi + 2k\pi/N), \tag{4.1}$$

where I' is the average intensity, I'' is the intensity modulation, and ϕ is the phase to be solved for. Using a least-square method, we can get

$$\phi = -\tan^{-1}\left[\frac{\sum_{k=1}^{N} I_k \sin(2k\pi/N)}{\sum_{k=1}^{N} I_k \cos(2k\pi/N)}\right]. \tag{4.2}$$

Since an inverse tangent function is used, the phase values obtained from this equation only vary from $-\pi$ to $+\pi$. The phase with 2π modulus should be unwrapped to obtain a continuous phase map for 3D shape reconstruction. There are numerous spatial or temporal phase unwrapping algorithms. Essentially those algorithms are trying to determine a fringe order $n(x, y)$ for each pixel and unwrap the phase by adding $2n\pi$. The spatial phase unwrapping typically only provides phase values relative to a point on the phase map, while the temporal phase unwrapping can provide absolute phase information that can be pre-defined. Once the absolute phase is obtained, the phase can be converted to 3D coordinates with a calibrated system, or to carry on unique information for other analysis, e.g., establish mapping between a camera and a projector for system calibration.

4.2.2 Pinhole camera model

In this research, we use a well-established pinhole model for the regular camera, the thermal camera, as well as the projector. The pinhole model essentially establishes the relationship between a point (x^w, y^w, z^w) in the world coordinate system, (x^c, y^c, z^c) in the camera lens coordinate system, and its imaging point (u, v) on the camera sensor. The linear pinhole model can be written as,

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{A}\left[\mathbf{R}, \mathbf{T}\right]\begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}, \tag{4.3}$$

where s is a scaling factor indicating the depth, $\mathbf{R}$ and $\mathbf{T}$ are the rotation matrix and the translation vector that represent the transformation from the world coordinate system to the camera lens coordinate system; and $\mathbf{A}$ is the intrinsic matrix of the camera describing the projection from the lens coordinate system to the 2D imaging plane. These matrices are usually in the following forms

$$\mathbf{A} = \begin{bmatrix} f_u & \gamma & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \mathbf{T} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}, \tag{4.4}$$

where f_u and f_v are the effective focal lengths of the camera lens; (u_0, v_0) is the location of principle point; and γ is the skew factor of u and v axes, which is usually 0 for modern cameras. r_{ij} and t_i represent the rotation and translation from the world coordinate system to the camera lens coordinate system.

The linear model works well for perfectly designed and fabricated lenses, yet most lenses have distortion that the linear model does not represent. Among different kinds of distortions, radial and tangential distortions are the most severe and common. Typically, five coefficients are used to describe radial and tangential distortions as

$$\mathbf{Dist} = [k_1, k_2, p_1, p_2, k_3], \tag{4.5}$$

where k_1, k_2, and k_3 describe radial distortions, and p_1 and p_2 describe tangential distortions. Radial distortions can be modeled as,

$$\begin{aligned} u' &= u(1 + k_1 r^2 + k_2 r^4 + k_3 r^6), \\ v' &= v(1 + k_1 r^2 + k_2 r^4 + k_3 r^6), \end{aligned} \tag{4.6}$$

where $r = \sqrt{u^2 + v^2}$, and (u', v') is the location of pixel (u, v) after radial distortion. Similarly, tangential distortions can be modeled as,

$$\begin{aligned} u' &= u + [2p_1 uv + p_2(r^2 + 2u^2)], \\ v' &= v + [p_1(r^2 + 2v^2) + 2p_2 uv]. \end{aligned} \tag{4.7}$$

4.2.3 3D structured light system calibration

System calibration is intended to estimate the intrinsic and extrinsic matrix of the camera and the projector as well as the geometric relationship between them.

Structured light system calibration follows the well-established method described in Reference [?]. In brief, a flat circle pattern board shown in Fig. 4.1(a) is used. It is placed at different orientations within the calibration volume to capture those 2D images. For each pose, both horizontal and vertical fringe patterns are projected to capture absolute horizontal and vertical absolute phase maps. These phase maps are then used to establish a one-to-one mapping between the camera and the projector [?] and to determine the corresponding points for each feature point on the camera. In this case, the feature points are those circle centers on the calibration board.

We used OpenCV camera calibration toolbox to detect those circle centers for the camera images and then found those corresponding center points for the projector by building the one-to-one mapping through phases. Once those center points are detected, the intrinsic parameters for the camera ($\mathbf{A}^r$) and the projector ($\mathbf{A}^p$) are estimated. We then use the stereo-calibration toolbox provided by the OpenCV camera calibration library to estimate the extrinsic parameters: $\mathbf{R}^r$ the rotation matrix for the camera, $\mathbf{T}^r$ the translation vector for the camera, $\mathbf{R}^p$ the ration matrix for the projector, and $\mathbf{T}^p$ the translation vector for the projector. In this research, we coincide the world coordinate system with the camera lens coordinate system, and thus $\mathbf{R}^r$ is an identity matrix, and $\mathbf{T}^r$ is a zero vector.

As discussed in Reference [100], once the system is calibrated, 3D coordinates (x^w, y^w, z^w) in the world coordinate system can be computed for each camera point by solving the linear equations

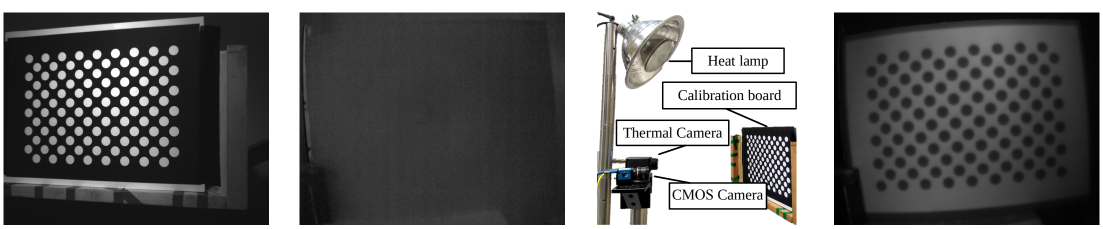
$$s^r \begin{bmatrix} u^r \\ v^r \\ 1 \end{bmatrix} = \mathbf{A}^r \left[\mathbf{R}^r, \mathbf{T}^r\right] \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}, \tag{4.8}$$

$$s^p \begin{bmatrix} u^p \\ v^p \\ 1 \end{bmatrix} = \mathbf{A}^p \left[\mathbf{R}^p, \mathbf{T}^p\right] \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}, \tag{4.9}$$

combined with the absolute phase constraint. Here (u^r, v^r) is the camera image coordinates, and (u^p, v^p) is the projector image coordinate. We only use the linear calibration model for our structured light system because we found such a model can achieve sufficient good accuracy.

4.2.4 Thermal camera calibration

Since a thermal camera is only sensitive to temperature variations, it cannot *see* the color difference in visible images. For example, Figure 4.1(a) shows the regular image of the circle pattern we used for system calibration, and Figure 4.1(b) shows image from the thermal camera. To solve this problem, we used a heat lamp (Model: GE Lighting 48037), as shown in Fig. 4.1(c), to shine thermal energy to the calibration board. Due to different emissivity of black and white areas, the thermal camera can capture circle patterns that are used for structured light system calibration. Figure 4.1(b) shows the thermal image of the circle patterns captured by a thermal camera without the heat lamp. Figure 4.1(d) shows the thermal image after turning on the heat lamp. Once the thermal camera can capture circle pattern images, its calibration becomes the well-established regular camera calibration problem.



(a) (b) (c) (d)

Figure 4.1. Stereo calibration between the regular and thermal camera. (a) Calibration board used to calibrate the whole system (image was captured by the regular camera); (b) Image captured by a thermal camera before turning on the heat lamp; (c) System setup to calibrate the thermal and regular camera; (d) Image captured by a thermal camera after turning on the heat lamp.

However, as can be seen in Fig. 4.1(d), the thermal image has serious distortions. Therefore, a linear calibration model is no longer sufficient for thermal imaging camera calibration, and the nonlinear distortion coefficients **Dist** are considered in our research. Similarly, after capturing a sequence of circle pattern images under different poses, the intrinsic parameter matrix $\mathbf{A}^t$ can be estimated.

Because the thermal camera calibration can use the same calibration target as the regular camera, the stereo calibration can also be carried for the regular camera and thermal camera pair to establish the geometric relationship between these two cameras. Again, we coincide the world coordinate system with the regular camera lens coordinate system and then estimate the rotation matrix $\mathbf{R}^t$, and the translation vector $\mathbf{T}^t$ for the thermal camera.

4.2.5 Sub-pixel mapping between structured light system and thermal camera

Since the world coordinate system coincides with the regular camera for both the structured light system calibration and the thermal camera calibration, all these device calibrations are under exactly the same world coordinate system. That is, $\mathbf{R}^p$ and $\mathbf{R}^t$ respectively describe the rotation from the projector coordinate system to the regular camera lens coordinate system, and the rotation from the thermal camera coordinate system to the regular lens coordinate system; $\mathbf{T}^p$ and $\mathbf{T}^t$ respectively describe the translation from the projector coordinate system to the regular camera lens coordinate system, and the translation from the thermal camera coordinate system to the regular lens coordinate system. Therefore, it is straightforward to find the corresponding (u^t, v^t) point for a given 3D point $\mathbf{P}^w = (x^w, y^w, z^w)$ recovered from the structured light system. Mathematically, one can solve the following equation to find the mapped point

$$s^t \begin{bmatrix} u^t \\ v^t \\ 1 \end{bmatrix} = \mathbf{A}^t \left[\mathbf{R}^t, \mathbf{T}^t\right] \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}, \tag{4.10}$$

assuming a linear calibration model is used, where s^t is a scaling factor.

However, as discussed above, although the structured light system uses a linear model, the thermal camera has to use the nonlinear model to represent its severe distortions. Therefore, the thermal image has to be rectified using the distortion coefficients **Dist** before mapping. In other words, (u^t, v^t) obtained directly from Eq. (4.10) corresponds to the rectified thermal image point, not the actual camera image point with distortions.

Even after rectification, the mapped point does not correspond to the thermal camera image pixel since the thermal cameras resolution is much lower than the regular cameras resolution. Hence, we propose to use a 2D Gaussian model to find the actual sub-pixel mapped thermal image point (or the actual temperature corresponding to that point).

Assume a 3D point is mapped to (u_0, v_0) based on Eq. (4.10). The Gaussian model provides the weighted average on all neighboring pixel values. Weights are related to the distance between the neighboring pixel and the mapped (u_0, v_0) point, which can be described as

$$f(i,j) = \exp^{-\frac{(u_{ij}-u_0)^2+(v_{ij}-v_0)^2}{2\sigma^2}}, \tag{4.11}$$

where $f(i,j)$ is the weight function without normalization for a pixel at $[u_{ij}, v_{ij}]$, and σ is the standard deviation. Suppose the window size is $2L \times 2L$, the normalized weight is

$$w(i,j) = \frac{f(i,j)}{\sum_{floor(u_0)-L+1}^{floor(u_0)+L} \sum_{floor(v_0)-L+1}^{floor(v_0)+L} f(i,j)}, \tag{4.12}$$

where *floor* takes the nearest integer less than or equal to that element.

Finally, the temperature T corresponding to the mapped point (u0, v0) is computed using

$$\mathcal{T}(u_0, v_0) = \sum_{floor(u_0)-L+1}^{floor(u_0)+L} \sum_{floor(v_0)-L+1}^{floor(v_0)+L} w(i,j)\mathcal{T}(i,j). \tag{4.13}$$

4.2.6 Invisible 3D point culling

Since two cameras have different perspectives, there are areas that can be seen by only one camera but not the other. In other words, some part of the object will be occluded in a specific viewpoint. Figure 4.2 illustrates one scenario, Curve $\widehat{ABC}$ d can be seen by the thermal camera, but not the part of Curve $\widehat{DEF}$. Therefore, those invisible points to the thermal camera should be properly handled, otherwise, incorrect temperature mapping can be generated for 3D data points. To accurately detect those areas, we employed both the *occlusion culling* method [110] and the *back-face culling* culling method [111].

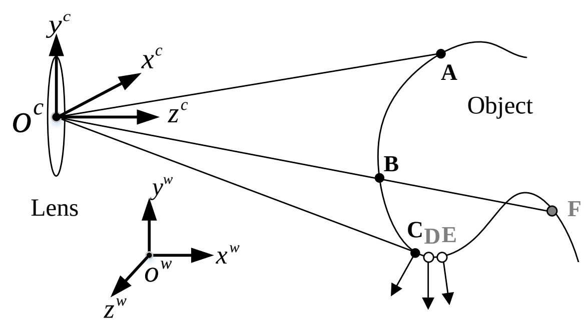


Figure 4.2. Illustration of culling. Curve $\widehat{ABC}$ can be seen in the view point of $\mathbf{O}^c$, but not the part of Curve $\widehat{DEF}$. Generally speaking, $\widehat{EF}$ can be detected by the occlusion culling algorithm since they are obviously hidden by some other parts, while $\widehat{DE}$ can be detected by the backface culling algorithm since they are on the edge of visible and invisible parts. For better culling results, we combine both occlusion culling and back-face culling methods.

Occlusion culling method finds occluded areas by finding the projected depth information of 3D points to the camera: if two points are corresponding to the same point, the point further away cannot be *seen* by the camera and thus should be regarded as an occluded point and discarded. For example, the B and F illustrated in Fig. 4.2 correspond to the same point on camera $\mathbf{O}^c$; since F is further away from the camera, it should be regarded as occluded and thus discarded.

The occlusion culling method can be easily executed. We use Eq. (4.10) to map all points $\mathbf{P}^w = (x^w, y^w, z^w)$ on a 3D surface to the thermal image sensor. To quickly locate those occluded points, we create a vector $\mathbf{S}_{ij}$ map for each pixel (i, j) on the thermal image to store projected depth z values and the corresponding 3D point.

$$\mathbf{S}_{ij} = \{z_{p_1}, z_{p_2}, \ldots, z_{p_{n_{ij}}}\}, \tag{4.14}$$

where $z_{p_1}, z_{p_2}, \ldots, z_{p_{n_{ij}}}$ are z-values of 3D points that are mapped to (i, j) on the thermal image. We then find the smallest element $z_{ij}^{\min}$ in $\mathbf{S}_{ij}$,

$$z_{ij}^{\min} = \min\{\mathbf{S}_{ij}\} \tag{4.15}$$

and discard all 3D points p_k in $\mathbf{S}_{ij}$ whose z-values satisfy

$$z_{p_k} > z_{ij}^{\min} + th. \tag{4.16}$$

where th is the predefined threshold. In other words, we discard those points whose z-values are larger than the smallest one by a threshold th. The threshold value is determined from the prior knowledge of the hardware system and the type of object to be measured.

Practically, since the resolutions of two cameras are different, we can set a virtual camera with higher or lower resolution for more accurate culling. Suppose the resolution of a virtual camera is N times of the resolution of a real one. (u^t, v^t) determined from Eq. (4.10) need to be scaled up by factor of N, i.e.,

$$u_{new} = floor(u^t \times N), \tag{4.17}$$

$$v_{new} = floor(v^t \times N). \tag{4.18}$$

Instead of creating a vector for each (u, v), we create a vector for each (u_{new}, v_{new}), and the conditions to discard a 3D point is the same as Eq. (4.16).

If the occluded points are far away from the front point, such as those points between $\widehat{EF}$ on Fig. 4.2, they can be easily detected by the occlusion culling method and discarded. However, this method does not work well for those points that are close to an edge of an object viewed from $\mathbf{O}^c$, such as those points between $\widehat{DE}$. This is because the occlusion culling method solely relies on the depth difference to determine which points to be discarded, and if the difference is very small (within the predefined threshold), those occluded points will fail to be detected. To handle such a condition, we propose to use *back-face culling* method.

Back-face culling method detects occluded points by its surface normal direction. If the point normals ($\mathbf{n}_p$) direction has a positive projection component to a cameras view direction, the point is regarded as a back-face point and thus should be discarded. The hollow circled dots D and E on Fig. 4.2 are regarded as back-face points and should be discarded. To implement the back-face culling method, the normal for each point of the point cloud data generated by the structured light system should be computed. Fortunately, since the point cloud data coming out of a structured light system are naturally aligned with the camera pixel grid, the point normal computation is straightforward: the averaged normals of the triangles formed by combining the surrounding pixels. We compute a point normal by considering the 3×3 neighborhood points of point $\mathbf{P}$, and normal $\mathbf{n}_P$ is the average of $\mathbf{n}_1, \cdots, \mathbf{n}_8$.

From the discussion in Sec. 4.2.5, the point cloud data coming out of the structured light system is in the world coordinate system that is perfectly aligned with the regular camera lens coordinate system, and thus all point normals should direct towards the regular camera. Since the thermal camera also has the same world coordinate system, the back-face point can be defined as

$$\mathbf{n}_P \cdot (\mathbf{P} - \mathbf{O}^c) > 0. \tag{4.19}$$

Here $\mathbf{P}$ is the (x^w, y^w, z^w) coordinates of an arbitrary point P on the surface, and $\mathbf{O}^c$ is the 3D coordinates of the second camera lens origin in the world coordinate system.

4.3 Experiment

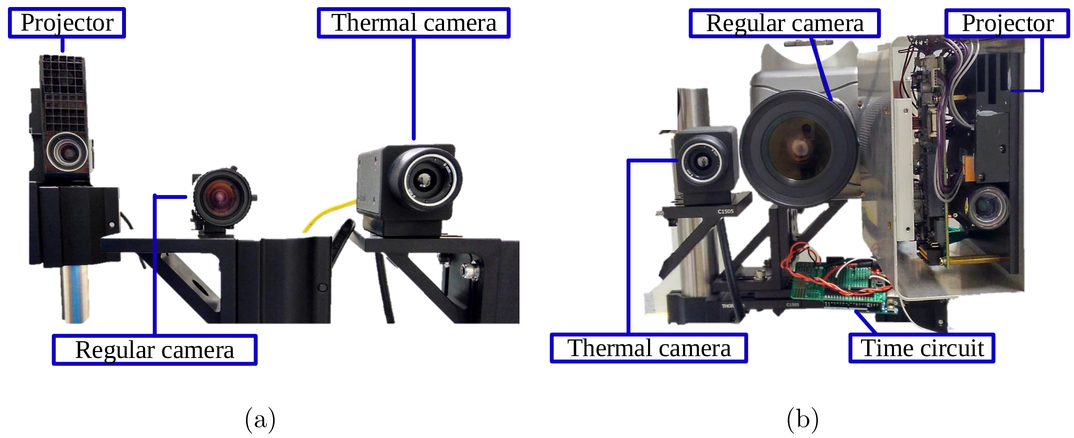
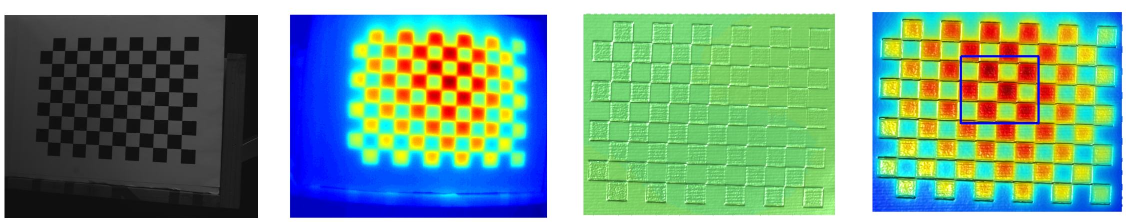


Figure 4.3. Experimental system setups. (a) Static object measurement system consists of a DLP projector (DELL M115HD), a CMOS camera (Imaging Source 23UX174) and a thermal camera (FLIR A35); (b) Real-time measurement system consists of a thermal camera (FLIR A35), a high-speed DLP projector (LightCrafer 4500), a high-speed CMOS camera (Vision Research Phantom V9.1), and an external timing generator (Arduino UNO R3).

We developed a hardware system to verify the performance of the proposed method. Figure 4.3(a) shows the hardware system we developed. The overall system includes a digital-light-processing (DLP) projector (Model: DELL M115HD), a complementary metal-oxide semiconductor, (CMOS) camera (Model: Imaging Source 23UX174) and a thermal camera (Model: FLIR A35). The resolution of the projector is 1280×800, the resolution of the CMOS camera is 1280×1024, and the resolution of the thermal camera is 320×256. The CMOS camera is attached with 8 mm focal length lens (Model: Computar M0814-MP2). For all 3D shape measurement experiments carried out with this system, we used $N = 9$ phase-shifted fringe patterns with a fringe period of 18 pixels to obtain wrapped phase map. The wrapped phase map is then unwrapped by projecting 7 binary coded patterns to uniquely determine fringe orders for each pixel. The absolute phase is further converted to 3D geometry pixel by pixel.

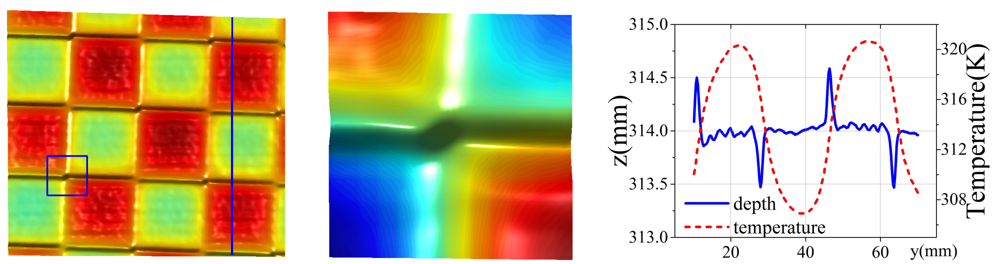
We first measured a black/white checkerboard heated up by the thermal lamp that was used for system calibration to verify the mapping accuracy. Figure 4.4 shows the results. Figure 4.4(a) shows the checkerboard image from the regular CMOS camera; and Figure 4.4(b) shows the image captured by the thermal camera at the same time. Color shows temperature ranging from 290 to 323 Kelvin (K). We measured the 3D geometry of the checkerboard using the structured light system and then mapped the thermal temperature image onto the 3D geometry. Figure 4.4(c) shows the 3D geometry of the checkerboard rendered in shaded mode, and Figure 4.4(d) shows the mapping result. This figure shows that the temperature difference between black and white blocks. Temperature in black blocks are higher than that in white blocks since black is of higher emissivity. This figure also shows that the boundary between black and white blocks are also very clear. Therefore, the mapping was fairly accurate, at least visually.



(a) (b) (c) (d)

Figure 4.4. Mapping example of a cheeseboard. (a) Image of the cheeseboard captured by the CMOS camera. Its resolution is 1280 × 1024; (b) Image captured by the thermal camera before rectification. Its resolution is 320 × 256;(c) 3D reconstructed geometry; (d) Mapping result. Color represents temperature ranging from 290 to 323 K in both (b) and (d).

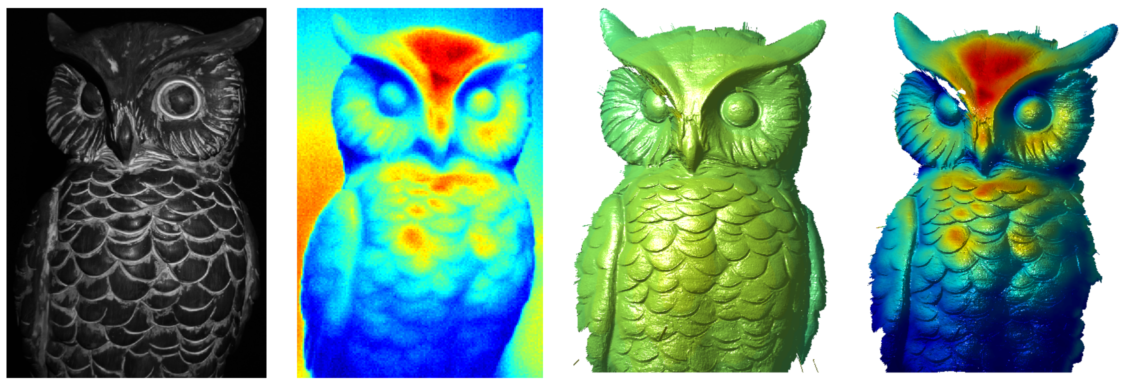
To better visualize the mapping quality, we showed close-up views of checker squares in Fig. 4.5(a), and area around the corners in Fig. 4.5(b). Again the mapping quality is pretty good. We further analyzed the mapping quality by plotting a vertical slice of the result, as shown in Fig. 4.5(c). Due to the large contrast of the checkerboard, the 3D shape measurement system created border artifacts (transition from the black to white or from white to black is not smooth). For better comparison, we detrend the depth values using a linear model and shifted them by adding 314 mm. This figure shows that these borders are perfectly aligned with the middle of the temperature changes. In summary, the sub pixel mapping method developed in this research is very accurate even though the thermal camera has a much lower resolution than the regular camera.



(a) (b) (c)

Figure 4.5. Zoom-in analysis. (a) Zoomed-in result of the blue rectangle part of Fig. 4.4(d); (b) Further zoomed-in result of the corner part in (a); (c) Temperature and depth of the cross section in (a). For better comparison, we detrend the depth values using a linear model and shifted them by adding 314 mm. Color represents temperature ranging from 290 to 323 K in (a) and (b).

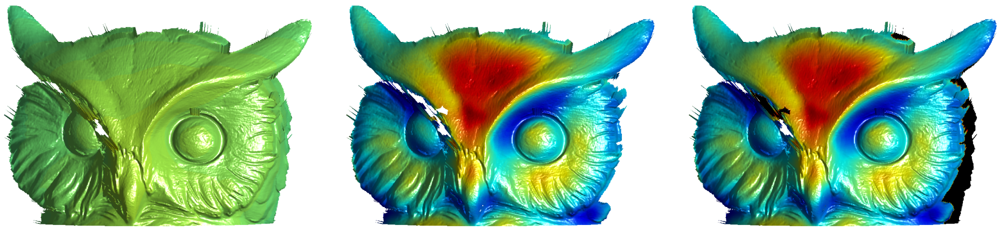
Since the checkerboard we used for previous measurement is flat, the occlusion problem is not obvious. We then measured a complex shape 3D statue to further verify the performance of the mapping method and validate the performance of the culling method. Figure 4.6(a) shows the photograph of the statue captured by the regular camera of the structured light system. Again, the statue was heated up by the thermal lamp, and we captured a temperature image by the thermal camera, as shown in Fig. 4.6(b). Figure 4.6(c) shows the 3D reconstruction from the structured light system. Figure 4.6(a) and Fig. 4.6(b) show that these two images are of different poses, which is caused by different viewpoints of the CMOS and thermal cameras. Therefore, we have to properly remove those occlusion 3D points from the thermal camera in order to generate the correct temperature map. Figure 4.6(d) shows the temperature mapped onto the recovered 3D geometry after applying the culling methods discussed in this chapter. Clearly a lot areas are culled out since they cannot be been seen by the thermal camera. For those points that can be seen by the thermal camera, the temperature mapping is fairly accurate.



(a) (b) (c) (d)

Figure 4.6. Mapping example of a 3D object. (a) Photography of the measured object; (b) Image captured by the thermal camera before rectification; (c) 3D reconstructed geometry; (d) Temperal mapping result. Color represents temperature ranging from 292 to 297.5 K in both (b) and (d).

To better visualize the culling effects, Figure 4.7(a) and Fig. 4.7(b) respectively shows the zoomed-in view of the 3D geometry and that of the 3D geometry with temperature mapping of the head of the statue. Comparing these two images, we can clearly see that a lot points on the 3D geometry are culled out because the thermal camera cannot see them. To clearly mark those points that are culled out, Fig. 4.7(c) highlights those culled out points as black. These experiments demonstrated that our proposed mapping and culling methods both perform satisfactorily.



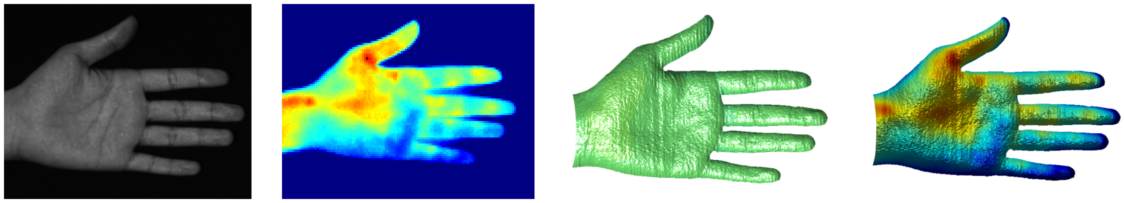
(a) (b) (c)

Figure 4.7. Zoom-in view of the object showed earlier. (a) The top part of the original 3D geometry; (b) Temperal mapping result; (c) Highlighted points that are culled out as black.

Since the thermal camera and the regular CMOS camera do not see the same spectrum of light, and 3D shape measurement and surface temperature measurement can be done at the same time, making real-time applications possible. To demonstrate this capability, we developed a system that uses the same thermal camera, a high-speed DLP projector (Model: Texas Instrument LightCrafer 4500), and the high-speed CMOS camera (Model: Vision Research Phantom V9.1). Three devices are synchronized by using an external timing generator (Model: Arduino UNO R3 board with DIP ATmega328P). The whole system is shown in Fig. 4.3(b). The resolution of the camera is 768×960 and it is fitted with a 24 mm focal length lens (Model: Sigma 24mm /1.8 EX DG Aspherical Macro). For this experiment, the projector projects 912×1140 resolution binary dithered patterns at 780 Hz, and the thermal camera captures at 26 Hz. We used an enhanced two-frequency temporal phase unwrapping algorithm [112] to obtain absolute phases that are further converted to 3D geometry. Since it requires 6 fringe patterns to recover one 3D geometry, the 3D data acquisition speed is actually 130 Hz, which is 5 times of the thermal camera acquisition speed. Then we just pick one in every five frames to do mapping.

To demonstrate the real-time capacity, we measured both hands and human facial expressions using such a system. Figure 4.8 and Visualization 4.1, Visualization 4.2, Visualization 4.3, Visualization 4.4 and Visualization 4.5 are the results of the hand.

Figure 4.8(a) shows one of the fringe images of the hand from the CMOS camera (associated with Visualization 4.1); and Figure 4.8(b) shows the image captured by the thermal camera at the same time (associated with Visualization 4.2). For real-time experiments, we employed the enhanced two-wavelength phase shifting method [112] for 3D reconstruction, and mapped the temperature onto the 3D geometry simultaneously. Figure 4.8(c) shows one frame of the 3D reconstructed geometry (associated with Visualization 4.3); and Figure 4.8(d) shows the same frame with temperature mapping (associated with Visualization 4.4). In both Fig. 4.8(b) and Fig. 4.8(d), color represents temperature ranging from 296 to 303 K.

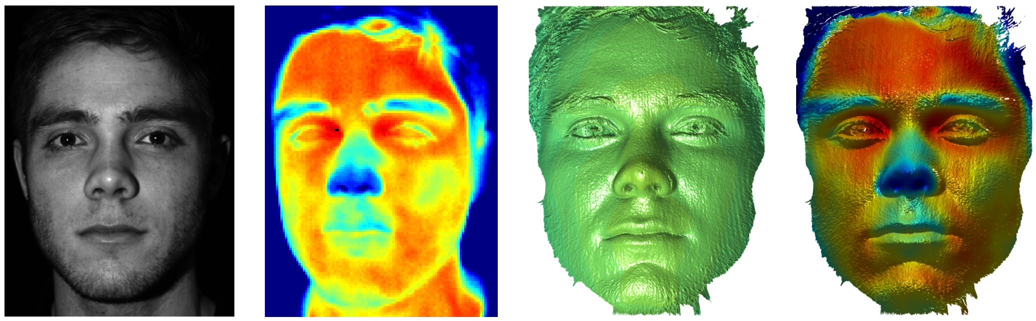


(a) (b) (c) (d)

Figure 4.8. Example of real-time mapping of hand (Visualization 4.1-4.5). (a) Photography of the hand to be measured captured by the CMOS camera (associated with Visualization 4.1); (b) Image captured by the thermal camera at the same time (associated with Visualization 4.2); (c) One frame of the 3D reconstructed geometry (associated with Visualization 4.3); (d) The same frame of the temperature mapping result (associated with Visualization 4.4). Color represents temperature ranging from 296 to 303K in both (b) and (d).

We also measured human facial expressions. Figure 4.9 and Visualization 4.6, Visualization 4.7, Visualization 4.8, Visualization 4.9 and Visualization 4.10 show the results. Fig. 4.9(a) shows the human face from the CMOS camera (associated with Visualization 4.6); and Figure 4.9(b) is the image captured by the thermal camera at the same time (associated with Visualization 4.7). Figure 4.9(c) shows one frame of the real-time 3D reconstruction result (associated with Visualization 4.8); and Figure 4.9(d) shows the same frame with temperature mapping (associated with

Visualization 4.9). In both Fig. 4.9(b) and Fig. 4.9(d), color represents temperature ranging from 297 to 305 K. This experiment verifies our algorithms capacity for real-time 3D geometric shape measurement and temperature mapping.



(a) (b) (c) (d)

Figure 4.9. Example of real-time mapping of human face (Visualization 4.6-4.10). (a) Photography of the human face captured by the CMOS camera (associated with Visualization 4.6); (b) Thermal image captured by thermal camera at the same time (associated with Visualization 4.7); (c) One frame of the 3D reconstructed geometry (associated with Visualization 4.8); (d) The same frame of temperature mapping result (associated with Visualization 4.9). Color represents temperature ranging from 297 to 305 K in both (b) and (d).

4.4 Summary

This chapter has presented a high-resolution, real-time simultaneous 3D geometric shape and temperature measurement method. We developed a holistic approach to calibrate both structured light system and thermal camera under exactly the same world coordinate system even though these two sensors do not share the same wavelength; and a computational framework to determine the sub-pixel corresponding temperature for each 3D point as well as discard those occluded points. Experiments verified the accuracy of our algorithm, and we demonstrated that the proposed method can be applied in real-time applications.

5. PIXEL-WISE ABSOLUTE PHASE UNWRAPPING USING GEOMETRIC CONSTRAINTS OF STRUCTURED LIGHT SYSTEM

In the previous two chapters, we projected additional patterns for absolute phase unwrapping. Projecting additional patterns usually slows down the measurement speed. To address that problem, this chapter will present a method to unwrap phase pixel by pixel by solely using geometric constraints of the structured light system without requiring additional image acquisition or another camera. Specifically, an artificial absolute phase map, Φ_{min}, at a given *virtual* depth plane $z = z_{min}$, is created from geometric constraints of the calibrated structured light system; the wrapped phase is pixel-by-pixel unwrapped by referring to Φ_{min}. Since Φ_{min} is defined in the projector space, the unwrapped phase obtained from this method is absolute for each pixel. Experimental results demonstrate the success of this proposed novel absolute phase unwrapping method. The majority of this chapter was originally published in *Optics Express* [113] (also listed as journal article [J2] in "LIST OF PUBLICATIONS").

5.1 Introduction

Three dimensional (3D) shape measurement has numerous applications including in-situ quality control in manufacturing and disease diagnoses in medical practices.

Among all 3D shape measurement techniques developed, using phase instead of intensity has the merits of robustness to sensor noise, robustness to surface reflectivity variations, and being able to achieve high spatial and/or temporal resolutions [41]. Over the years, numerous phase retrieval methods have been developed including the Fourier method [42], the Windowed Fourier method [114], the phase-shifting methods [44]. Overall, a typical fringe analysis method only provides phase values ranging from $-\pi$ to $+\pi$ with a modulus of 2π, and thus a phase unwrapping algorithm has to be employed to obtain the continuous phase map before 3D reconstruction.

Conventionally, there are two types of phase unwrapping methods: spatial phase unwrapping and temporal phase unwrapping. The spatial phase unwrapping detects 2π discontinuities from the phase map itself and removes them by adding or subtracting multiple $K(x, y)$ of 2π accordingly. The integer number $K(x, y)$ is often referred as fringe order. The book edited by Ghiglia and Pritt [115] summarizes numerous phase unwrapping algorithms with some being faster yet less robust and some being more robust yet slower; the review paper written by Su and Chen [116] covers a wide range of reliability-guided phase unwrapping algorithms. Regardless of the robustness and speed of a spatial phase unwrapping algorithm, it typically only generates a *relative phase* phase map: a phase map that is relative to a point on the phase map itself within a connected component; thus it is difficult for any spatial phase unwrapping method to be employed if multiple isolated objects are to be simultaneously measured in the *absolute* sense. Furthermore, the majority of spatial phase unwrapping algorithms fail if abrupt surface geometric shape changes introduce more than 2π phase changes from one pixel to its neighboring pixels.

Temporal phase unwrapping, in contrast, tries to fundamentally eliminate the problems associated with the spatial phase unwrapping by acquiring more information. In essence, instead of finding the number of 2π, or fringe order $K(x, y)$, to be added to each pixel from phase values surrounding that pixel, temporal phase unwrapping finds fringe order $K(x, y)$ by referring to additional captured information, such as more fringe patterns. In other words, temporal phase unwrapping looks for information acquired temporally instead of spatially. Over the years, numerous temporal phase unwrapping methods have been developed including two- or multifrequency (or -wavelength) phase-shifting techniques [117–119], gray-coding plus phase-shifting methods [47, 48], spatial-coding plus phase-shifting method [120], and phase-coding plus phase-shifting methods [121–123]. Temporal phase unwrapping can provide absolute phase since the phase is unwrapped by referring to pre-defined information. The aforementioned temporal phase unwrapping methods work well to retrieve absolute phase, yet they require capture additional images for fringe order $K(x,y)$ determination. Since more images are acquired, temporal phase unwrapping slows down measurement speeds, which is not desirable for high-speed applications.

To address the reduced acquisition speed limitation of conventional temporal phase unwrapping approaches, researchers attempted to add the second camera to a standard single-camera, single projector structured light system for absolute phase unwrapping [124–127]. Because the second camera is available to capture images from another perspective, stereo geometric constraints and epipolar geometry can be used for fringe order $K(x,y)$ determination without using conventional spatial or temporal phase unwrapping. Furthermore, because the projector projects encoded structured patterns on it, the phase information can be used to ease the stereo matching problem of a traditional dual camera stereo technique. Basically, a point on the left camera is constrained to match points on the right camera with the *same phase* value. Since the wrapped phase map is periodical and contains stripes, the possible candidates on the right camera are not unique. By applying the epipolar geometric constraint of the stereo vision cameras, the corresponding points are limited to a few points on an epipolar line (only one point per fringe period). Finally, the correct corresponding point can be determined by verifying with the second camera image, the calibration volume, along with other techniques. This approach has been proven successful for absolute complex geometry capture. However, it usually requires global backward and forward checking to select the correct corresponding point out of many candidate points. Because a global searching is required, its computation speed is slow, and it is difficult to measure objects with sharp changing surface geometries. Furthermore, such a system requires accurately calibrating three sensors (two cameras and one projector), which is usually nontrivial.

To overcome limitations of the approach that requires global backward and forward searching, Lohry et al. [128] developed a method that combines with the conventional stereo approach to speed up the whole process. The proposed method includes two stages: 1) using a stereo matching algorithm to obtain the coarse disparity map to avoid global searching and checking; and 2) using local wrapped phase to further refine the coarse disparity to achieve higher measurement accuracy. To obtain more accurate disparity maps but not increasing the number of images used, the approach proposed by Lohry et al. [128] embedded a statistical pattern into the regular fringe pattern. This method does not require any geometric constraint imposed by the projector, and thus no projector calibration is required, further simplifying system development. However, due to the pixel-by-pixel disparity refinement, the processing speed is still limited. In general, it is still difficult for any of these methods to achieve real-time processing without significant hardware level program implementation and optimization. And because of the use of a second camera, they all increase hardware cost and algorithm complexity.

This chapter proposes a novel absolute phase unwrapping method that determines absolute phase solely through geometric constraints of the structured light system without requiring another camera, more fringe patterns, or global search. Since no additional images are required, the measurement speeds are not compromised for 3D shape measurement; and because no global searching is required, the processing speed can be high. In brief, an artificial absolute phase map, Φ_{min}, at a given depth $z = z_{min}$ is created from geometric constraints of the structured light system. For the proposed method, the wrapped phase is unwrapped pixel by pixel through referring to the artificially created phase map Φ_{min}. Since Φ_{min} is defined in the projector space, the unwrapped phase obtained from this method is absolute. Experimental results demonstrate the success of this proposed novel absolute phase unwrapping method, despite its limited working depth range.

Section 6.2 explains the principles of the proposed absolute phase unwrapping method. Section 6.3 presents experimental results to validate the proposed method and illustrate its limitations. Section 5.4 discusses the merits and limitations of the proposed absolute phase unwrapping method, and finally, Section 5.5 summarizes the chapter.

5.2 Principle

This section thoroughly explains the principle of the proposed method. Specifically, we will present the standard pinhole camera model, and then detail the proposed pixel-by-pixel absolute phase unwrapping method through theoretical derivations and graphical illustrations.

5.2.1 Three-step phase-shifting algorithm

Using phase instead of intensity for 3D optical metrology is advantageous since it is more robust to noise and surface reflectivity variations. Over the years, many fringe analysis techniques were developed to retrieve phase information including Fourier method and various phase-shifting methods [44]. Compared to other phase retrieval methods (e.g., Fourier or Windowed Fourier), phase-shifting methods have the advantage of measurement accuracy and robustness. Without loss of generality, this research uses a three-step phase-shifting algorithm for phase retrieval as an example to verify the performance of our proposed absolute phase unwrapping algorithm. Three phase-shifted fringe images with equal phase shifts can be mathematically written as

$$I_1(x,y) = I'(x,y) + I''(x,y)\cos(\phi - 2\pi/3), \tag{5.1}$$

$$I_2(x,y) = I'(x,y) + I''(x,y)\cos(\phi), \tag{5.2}$$

$$I_3(x,y) = I'(x,y) + I''(x,y)\cos(\phi + 2\pi/3). \tag{5.3}$$

Where $I'(x,y)$ is the average intensity, $I''(x,y)$ is intensity modulation, and ϕ is the phase to be solved for. Solving Eqs.(6.1)–(6.3) simultaneously leads to

$$\phi(x,y) = \tan^{-1}\left[\frac{\sqrt{3}(I_1 - I_3)}{2I_2 - I_1 - I_3}\right]. \tag{5.4}$$

The phase obtained from Eq.(6.5) ranges from $-\pi$ to π with 2π discontinuities. To remove 2π discontinuities, a spatial or temporal phase unwrapping algorithm can be used. Phase unwrapping essentially determines integer number $K(x, y)$ for each point such that the unwrapped phase can be obtained using the following equation

$$\Phi(x, y) = \phi(x, y) + 2\pi \times K(x, y). \tag{5.5}$$

Here $K(x, y)$ is often referred as fringe order. If $K(x, y)$ is pre-defined in an absolute sense (such as those obtained from a temporal phase unwrapping algorithm), the unwrapped phase $\Phi(x, y)$ is *absolute phase*. A spatial phase unwrapping typically yields $K(x, y)$ that is relative to one point on the wrapped phase map, and thus the spatial phase unwrapping can only generate *relative phase*. It is important to note that we denote $\Phi(x, y)$ as the unwrapped phase of $\phi(x, y)$ for this entire chapter.

Instead of using a conventional temporal phase unwrapping method to obtain the absolute phase map by capturing more fringe images, we propose a new method to obtain the absolute phase map pixel by pixel solely by using geometric constraints of the structured light system without requiring any additional image acquisition or the second camera.

5.2.2 Structured light system model

We first discuss the modeling of structured light system since it is critical to understanding the proposed method on how to use geometric constraints for pixel-by-pixel absolute phase unwrapping. We use a well-known pinhole model to describe the imaging system. This model essentially describes the projection from 3D world coordinates (x^w, y^w, z^w) to 2D imaging coordinates (u, v). The linear pinhole model can be mathematically represented as,

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & \gamma & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}. \tag{5.6}$$

Where r_{ij} and t_i respectively represents the rotation and the translation from the world coordinate system to the lens coordinate system; s is a scaling factor; f_u and f_v

respectively describes the effective focal lengths; γ is the skew factor of u and v axes; (u_0, v_0) is the principle point, the intersection of the optical axis with the imaging plane.

To simplify mathematical representation, we define the projection matrix $\mathbf{P}$ as

$$\mathbf{P} = \begin{bmatrix} f_u & \gamma & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}, \tag{5.7}$$

$$= \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}. \tag{5.8}$$

Projection matrix $\mathbf{P}$ can be estimated through a well-established camera calibration approach.

The same lens model for the camera is applicable to the projector since the projector can be treated as the inverse of a camera [100]. If the camera and the projector calibration is performed under the same world coordinate system, i.e., define the same world coordinate system, the projection matrix for the camera and the projector will be physically correlated. For simplicity, we typically coincide the world coordinate system with the camera lens coordinate system or the projector lens coordinate system. Therefore, we will have two sets of equations with one for the camera and the other for the projector lens

$$s^c \begin{bmatrix} u^c & v^c & 1 \end{bmatrix}^t = \mathbf{P^c} \begin{bmatrix} x^w & y^w & z^w & 1 \end{bmatrix}^t, \tag{5.9}$$

$$s^p \begin{bmatrix} u^p & v^p & 1 \end{bmatrix}^t = \mathbf{P^p} \begin{bmatrix} x^w & y^w & z^w & 1 \end{bmatrix}^t. \tag{5.10}$$

Here superscript $^{\mathbf{p}}$ represents projector, superscript $^{\mathbf{c}}$ presents camera, and t denotes the transpose operation of a matrix.

After structured light system calibration, the projection matrices, $\mathbf{P^c}$ and $\mathbf{P^p}$, are known. Equations (5.9)-(5.10) provide 6 equations with 7 unknowns $(s^c, s^p, x^w, y^w, z^w, u^p, v^p)$ for each camera pixel (u^c, v^c), and one additional constraint equation is required to solve all unknowns uniquely. For example, to recover (x^w, y^w, z^w) coordinates for a 3D shape measurement system, the absolute phase can be used for a phase-shifting method [100]. The absolute phase, $\Phi(x, y)$, essentially creates a one-to-many mapping constraint equation that maps one point on the camera image plane (u^c, v^c) to a line, u^p or v^p, on the projector image plane with exactly the same phase value.

Assume that fringe patterns vary sinusoidally along u^p direction and remain constant along v^p direction. If absolute phase Φ is known for any given point, u^p can be solved as

$$u^p = \Phi \times T/(2\pi), \tag{5.11}$$

assuming the absolute phase starts with 0 at $u^p = 0$ and increases with u^p. Here, T is the fringe period in pixels.

5.2.3 Absolute phase unwrapping using minimum phase map

Figure 5.1 graphically illustrates that using simple geometric optics and pinhole models of the lenses, the camera sensor plane can be mapped to the projector sensor plane if the object plane is a flat surface that is precisely placed at $z^w = z_{min}$. Once the mapped region is found on the projector sensor plane, the corresponding phase map can be pre-defined. Therefore, for the virtually defined z_{min} plane, the corresponding phase Φ_{min} can be precisely created. In this chapter, we propose to use the artificially created phase map Φ_{min} for absolute phase unwrapping.

Mathematically, for a given camera pixel (u^c, v^c), if we know z^w value, all seven unknowns including (u^p, v^p) can be uniquely solved using Eqs. (5.9)-(5.10). If (u^p, v^p) is known, the corresponding absolute phase value for that camera pixel (u^c, v^c) can be uniquely defined as

$$\Phi = 2\pi \times u^p/T \tag{5.12}$$

on the projector space. Here we assume the projector uses a fringe period of T pixels, and the fringe patterns vary along u^p direction sinusoidally.

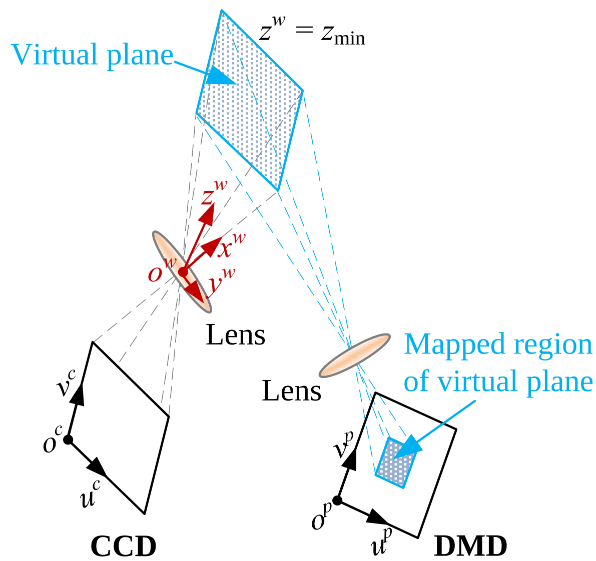


Figure 5.1. By using geometric constraint of a structured light system, one can establish the mapping between the camera image sensor (e.g., charge-coupled device, or CCD) and the corresponding region on the projector sensor (e.g., digital micro-mirror device, or DMD) for a virtual z_{min} plane.

Therefore, for a virtual measurement plane at $z^w = z_0$, one artificial absolute phase map can be defined pixel by pixel. If $z^w = z_0 = z_{min}$ is the closest depth of interest, we define this artificially created phase map as the minimum phase map Φ_{min}, which apparently is a function of z_{min}, fringe period T, and projection matrices, i.e.,

$$\Phi_{min}(u^c, v^c) = f(z_{min}, T, \mathbf{P}^c, \mathbf{P}^p). \tag{5.13}$$

As aforementioned, once a structured light system is calibrated under the same world coordinate system, the projection matrices $\mathbf{P}^c$ and $\mathbf{P}^p$ are known. Given z_{min}, we can solve for the corresponding x^w and y^w for each camera pixel (u^c, v^c) by simultaneously solving Eqs. (5.9)-(5.10),

$$\begin{bmatrix} x^w \\ y^w \end{bmatrix} = \mathbf{M}^{-1}b, \tag{5.14}$$

where

$$\mathbf{M} = \begin{bmatrix} p^c_{31}u^c - p^c_{11} & p^c_{32}u^c - p^c_{12} \\ p^c_{31}v^c - p^c_{21} & p^c_{32}v^c - p^c_{22} \end{bmatrix}, \tag{5.15}$$

$$b = \begin{bmatrix} p^c_{14} - p^c_{34}u^c - (p^c_{33}u^c - p^c_{13})z_{\min} \\ p^c_{24} - p^c_{34}v^c - (p^c_{33}v^c - p^c_{23})z_{\min} \end{bmatrix}. \tag{5.16}$$

Here p^c_{ij} denotes the matrix parameters of $\mathbf{P}^c$ in i-th row and j-th column. With known (x^w, y^w), Eq. (5.10) yields the corresponding (u^p, v^p) for each camera pixel

$$s^p \begin{bmatrix} u^p & v^p & 1 \end{bmatrix}^t = \mathbf{P}^p [\ x^w \quad y^w \quad z_{min} \quad 1\ ]^t. \tag{5.17}$$

Once (u^p, v^p) is calculated, we can determine the absolute phase value $\Phi_{min}(u^c, v^c)$ corresponding to z_{min} for that pixel using Eq. (5.12). Because $\Phi_{min}(u^c, v^c)$ is created pixel to pixel on the camera imaging sensor, such a phase map can be used to unwrap the phase map pixel by pixel. And since this phase is defined on the projector space, the obtained unwrapped phase by referring to $\Phi_{min}(u^c, v^c)$ is absolute.

Figure 5.2 illustrates the basic concept of using the minimum phase to correct 2π discontinuities. Assume the region on the projector that a camera captures at $z = z_{min}$ is shown in the red dashed window, the wrapped phase, ϕ_1, directly obtained from three phase-shifted fringe patterns has one 2π discontinuities, as shown in Fig. 5.2(a). The corresponding Φ_{min} is the continuous phase (or unwrapped phase) on the projector space, as shown in Fig. 5.2(b). The cross sections of the phase maps are shown in Fig. 5.2(c). This example shows that if the camera phase is below Φ_{min}, 2π should be added to the camera wrapped phase for phase unwrapping. And if the wrapped phase ϕ is captured at $z > z_{min}$ as illustrated in the solid blue windowed region, 2π should also be added to unwrap the phase if the wrapped phase is below Φ_{min}.

Figure 5.3 illustrates the cases to unwrap 3 and 4 periods camera captured phase maps. Figure 5.3(a) shows a case where there are two 2π discontinuous locations, Point A and Point B. Between Point A and Point B, the phase difference $\Phi_{min} - \phi$ is larger than 0 but less than 2π; and on the right of Point B, the phase difference is

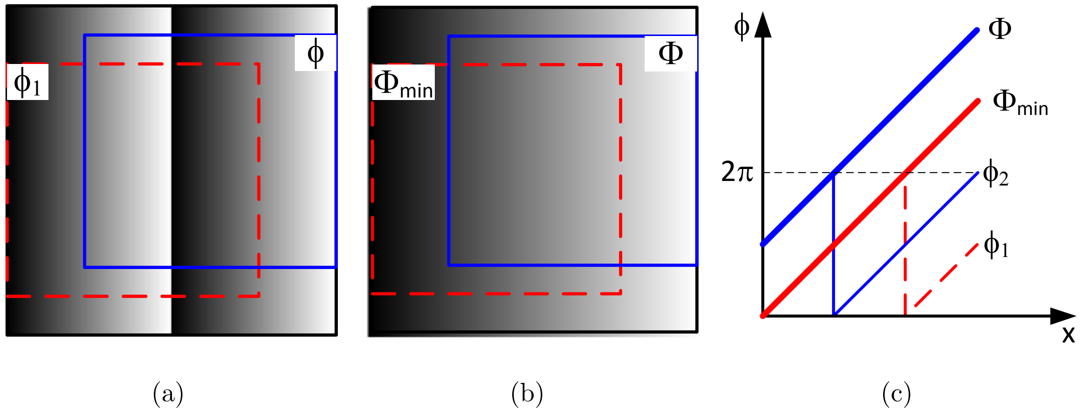


Figure 5.2. Conceptual idea of removing 2π jump of low-frequency phase map by using the minimum phase map determined from geometric constraints. (a) Windowed regions shows phase map that is acquired by the camera at different depths z: the red dashed window shows z_{min} and the solid blue window shows $z > z_{min}$; (b) Corresponding Φ_{min} and Φ defined on the projector; (c) Cross sections of Φ_{min} and Φ and the phase maps with 2π discontinuities.

larger than 2π. Therefore, 2π should be added to unwrap the point between Point A and Point B, and 4π should be added on the right side of Point B.

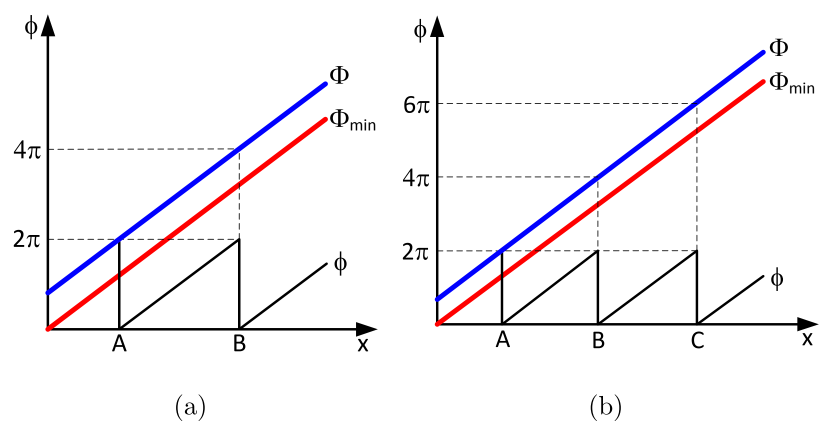


Figure 5.3. Determination of fringe order K, for multiple periods of fringe patterns. (a) Example of having three periods of fringe patterns; (b) Example of having four periods of fringe patterns.

For cases with 4 fringe periods, as shown in Fig. 5.3(b), if $0 < \Phi_{min} - \phi < 2\pi$ (i.e., between Point A and Point B), 2π should be added; $2\pi < \Phi_{min} - \phi < 4\pi$ (i.e., between Point B and Point C), 4π should be added; and $4\pi < \Phi_{min} - \phi < 6\pi$ (i.e., beyond C), 6π should be added.

In general, the fringe order K for each pixel must satisfy the following condition

$$2\pi \times (K-1) < \Phi_{min} - \phi < 2\pi \times K. \tag{5.18}$$

In other words, fringe order K can be determined as

$$K(x,y) = ceil\left[\frac{\Phi_{min} - \phi}{2\pi}\right]. \tag{5.19}$$

Here, $ceil[]$ is the ceiling operator that gives the nearest upper integer number.

5.3 Experiment

To verify the performance of the proposed temporal phase unwrapping method, we developed a structured light system, shown in Fig. 5.4, that includes one single CCD camera (Model: The Imaging Source DMK 23U618) with an 8 mm focal length lens (Model: Computar M0814-MP2) and one digital light processing (DLP) projector (Model: Dell M115HD). The camera resolution is 640×480. The lens is a 2/3-inch lens with an aperture of F/1.4. The projectors native resolution is 1280×800 with a focal length of 14.95 mm fixed lens having an aperture of F/2.0. The projection distance ranges from 0.97 m to 2.58 m. The system was calibrated using the method developed by Li et al. [105] and the camera lens coordinate system was chosen as the world coordinate system for both the camera and the projector.

We tested the proposed absolute phase unwrapping method by measuring a single object. Figure 5.5 shows the results. In this and all following experiments, the fringe period used is 20 pixels, and three equally phase-shifted fringe patterns are captured. Figure 5.5(a) shows the photograph of the object to be measured, indicating complex 3D geometric structures. Figure 5.5(b) shows one of three captured fringe patterns. From three phase-shifted fringe patterns, the wrapped phase is then computed, as

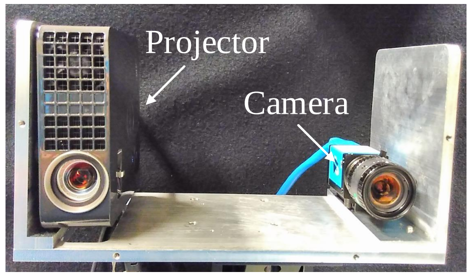


Figure 5.4. Photograph of the experimental system. The experimental system only uses one single projector and one single camera that is the same as a typical structured light system.

shown in Fig. 5.5(c). The phase map contains many periods of fringe patterns and thus has to be unwrapped before 3D reconstruction. We then generated the minimum phase map Φ_{min} at depth $z_{min} = 880$ mm, as shown in Fig. 5.5(d). Using the minimum phase map, we can determine fringe order for the wrapped phase map shown in Fig. 5.5(c), from which the unwrapped phase can be obtained. Figure 5.5(e) shows the unwrapped phase map. Since the unwrapped phase is absolute phase, we can use the calibration data to reconstruct 3D geometry using the method discussed by Zhang and Huang [100]. Figure 5.5(f) shows the recovered 3D geometry, which is continuous and smooth, suggesting the proposed absolute phase unwrapping works well for single 3D object measurement.

Since the proposed phase unwrapping method can obtain absolute phase, it should be possible to simultaneously measure multiple isolated objects. To verify this capability, we measured two isolated 3D objects shown in Fig. 5.6(a). Figure 5.6(b) shows one fringe pattern, and Figure 5.6(c) shows the wrapped phase map. Using the same minimum phase map shown in Fig. 5.5(d), we generated the unwrapped phase as shown in Fig. 5.6(d). Finally, 3D geometry can be recovered as shown in Fig. 5.6(e). Clearly, both objects are properly reconstructed. This experiment demonstrates that two isolated complex objects can indeed be properly measured using the proposed

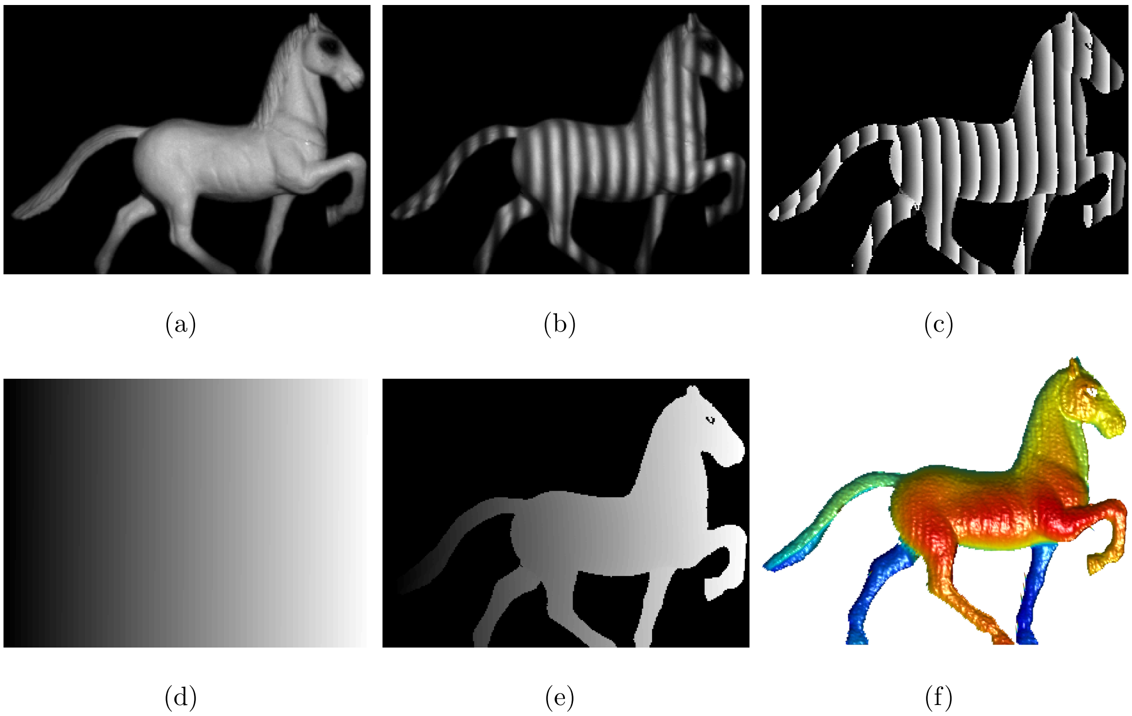
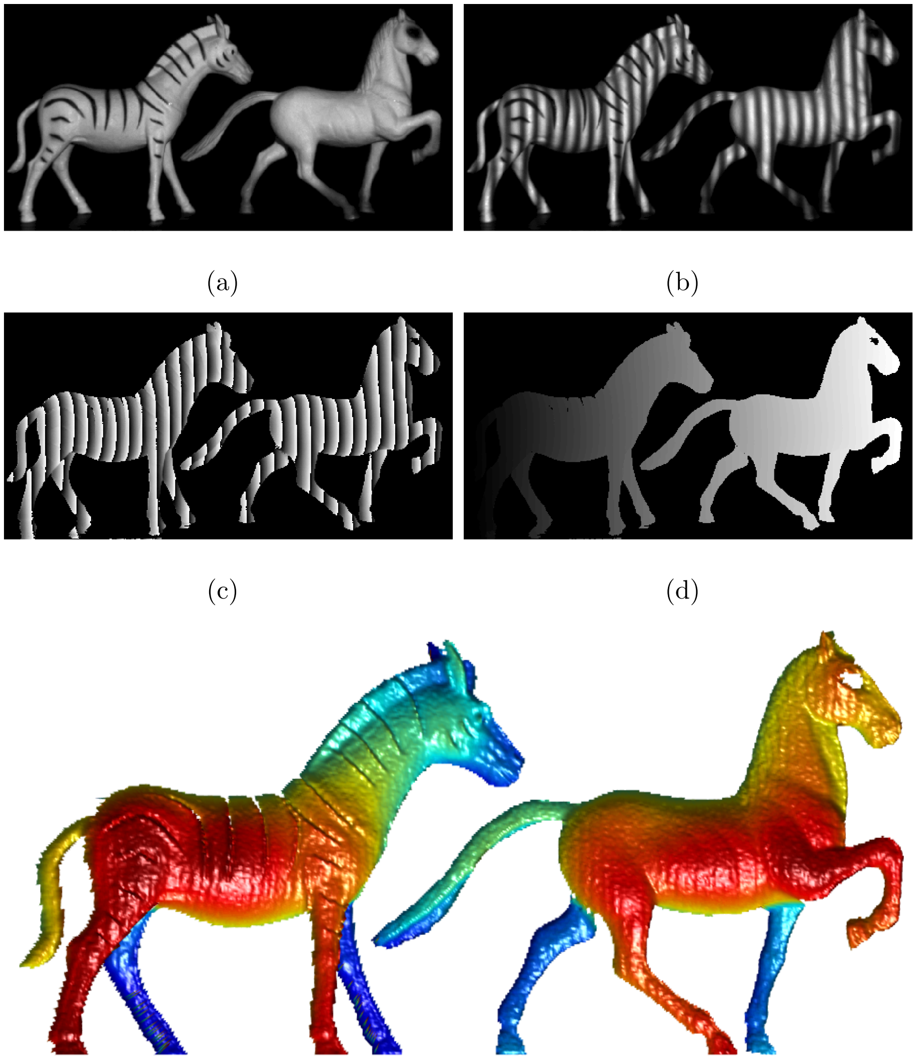


Figure 5.5. Measurement result of a single 3D object. (a) Photograph of the measured object; (b) One of three phase-shifted fringe patterns; (c) Wrapped phase map ϕ; (d) Artificially generated minimum phase map, Φ_{min}, using geometric constraints of the structured light system; (e) Unwrapped phase map Φ; (f) Reconstructed 3D geometry.

method, confirming that the proposed phase unwrapping method can perform pixel-by-pixel phase unwrapping.



(a) (b) (c) (d) (e)

Figure 5.6. Measurement result of two separate 3D objects. (a) Photograph of the objects; (b) One of the three phase-shifted fringe patterns; (c) Wrapped phase map ϕ; (d) Unwrapped phase map Φ; (e) Reconstructed 3D geometry.

We also experimentally compared our proposed absolute phase unwrapping method with a conventional temporal phase unwrapping method. Figures 5.7-5.8 show the results. In this experiment, we used 7 binary patterns to determine fringe order $K(x, y)$ that were used to temporarily unwrap the phase obtained from three phase-shifted fringe patterns [129]. Figure 5.7(a) shows the experimental object photograph. Again, we used two isolated 3D objects. Figure 5.7(b) shows the wrapped phase map from these phase-shifted fringe patterns. Figure 5.7(c) shows the unwrapped phase map by applying the conventional temporal phase unwrapping method. Since the system is calibrated, 3D shape was further reconstructed from the unwrapped phase map. Figure 5.7(d) shows the 3D result rendered in shaded mode. It is obvious that there are phase unwrapping artifacts (i.e., spikes) if no filtering is applied. This is a very common problem associated with any temporal phase unwrapping approach due to sampling error and camera noise [130]. In this research, we simply apply a median filter to locate those incorrectly unwrapped phase points and adjust them using the approach detailed by Karpinsky et al. [131]. Figure 5.7(e) shows the unwrapped phase, and Fig. 5.7(f) shows the final 3D reconstruction after applying a 11×11 median filter. As anticipated, the spiky noisy points are effectively reduced.

We then used our proposed approach to unwrap the phase map shown in Fig. 5.7(b) with the minimum phase map shown in Fig. 5.8(a). The unwrapped phase and 3D reconstruction are shown in Fig. 5.8(b)-5.8(c). It should be noted that no filtering was applied, and the result shows no spiky noise. This experiment demonstrated that our proposed method is actually more robust than temporal phase unwrapping. This is because the proposed method determines fringe order by referring to an artificially generated ideal and noise-free phase map Φ_{min}. In contrast, the conventional temporal phase unwrapping method determines fringe order by referring to other camera captured information that inherently contains noise.

To further visualize the difference between the unwrapped phase using our proposed method and the conventional temporal phase unwrapping method, the same cross section of two unwrapped phase maps shown in Figs. 5.7(e) and 5.8(b) are plotted in Fig. 5.8(d). They overlap well with each other on the object surface, further verifying that the phase obtained from our proposed phase unwrapping method is absolute.

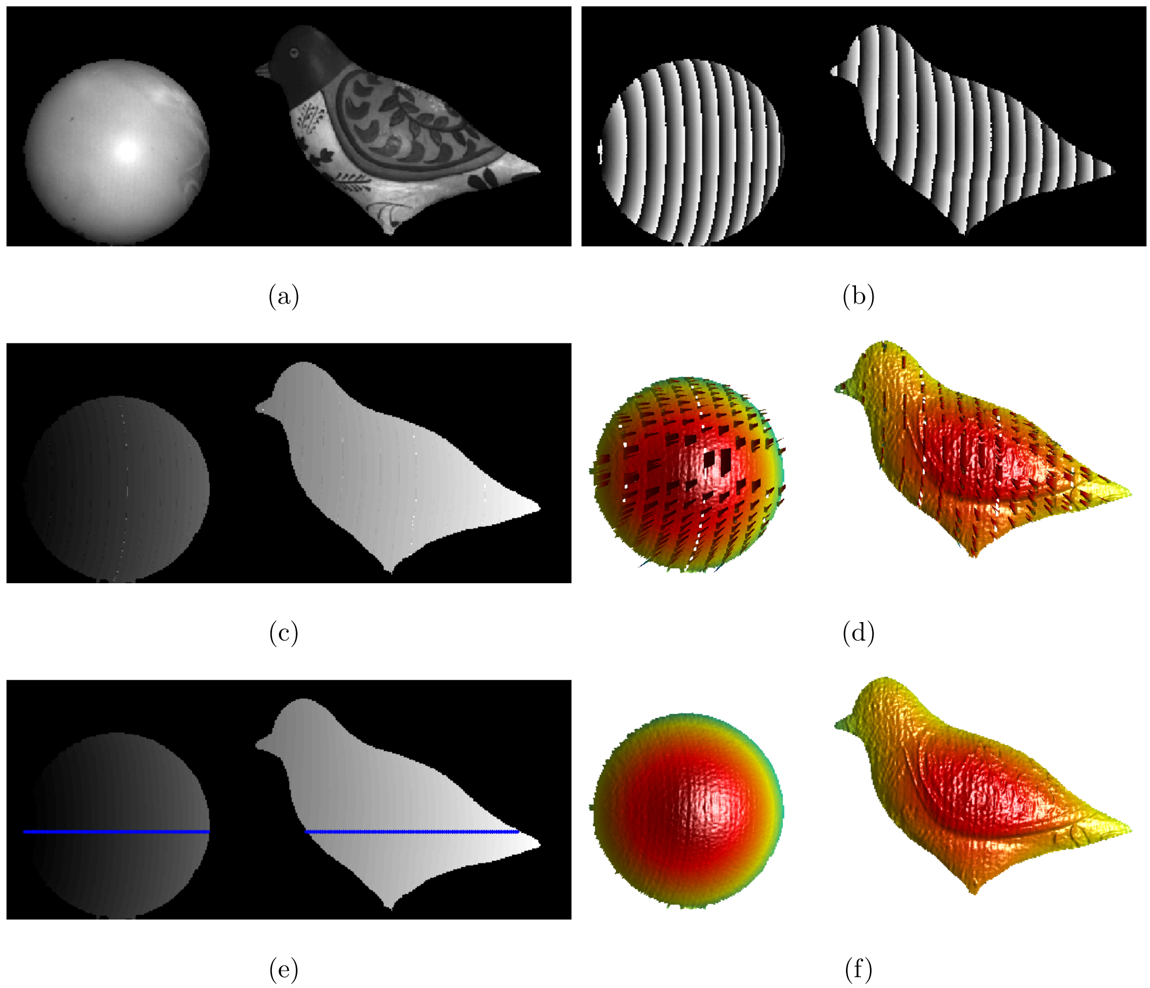


Figure 5.7. Measurement results using the conventional temporal phase unwrapping approach. (a) Photograph of the measured objects; (b) Wrapped phase map ϕ; (c) Unwrapped phase map by applying the conventional temporal phase unwrapping method; (d) Reconstructed 3D geometry by the conventional temporal phase unwrapping method without filter; (e) Unwrapped phase map using the conventional temporal phase unwrapping method after applying a 11×11 median filter; (f) Reconstructed 3D geometry by the conventional temporal phase unwrapping method with filter.

To validate the robustness of the proposed phase unwrapping method, we measured a puppy toy. Figure 5.9(a) shows one photograph of puppy toy that has black hair and yellow legs. Apparently, this type of object is very difficult to measure

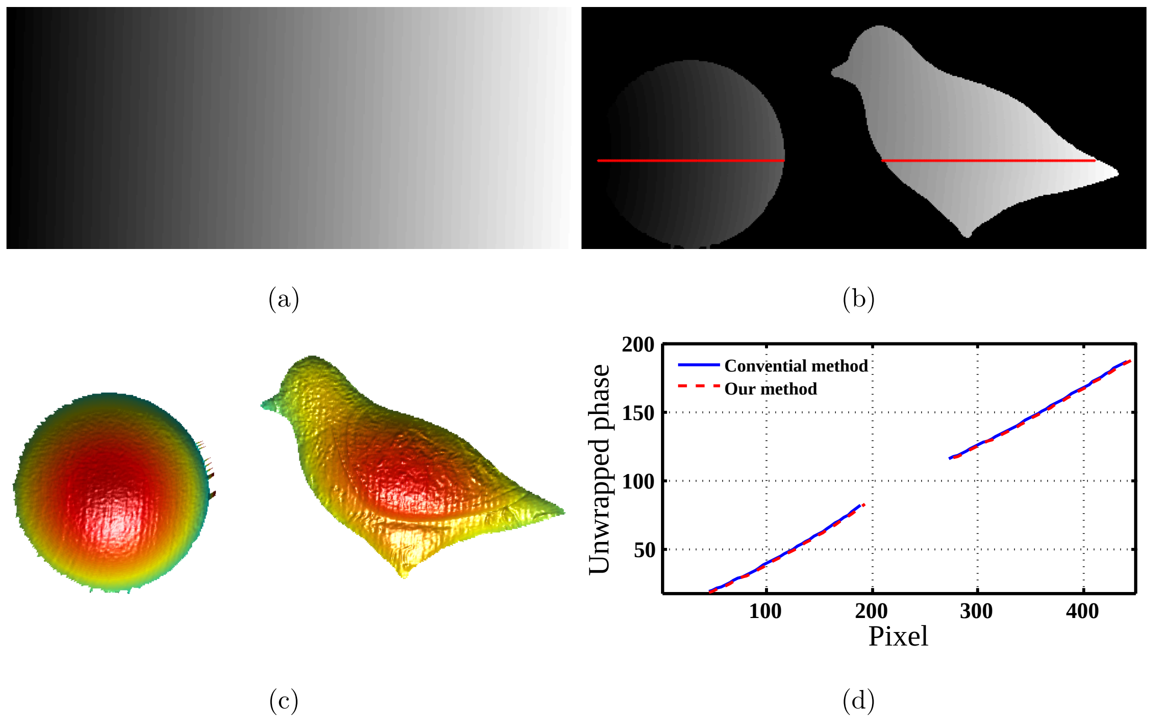
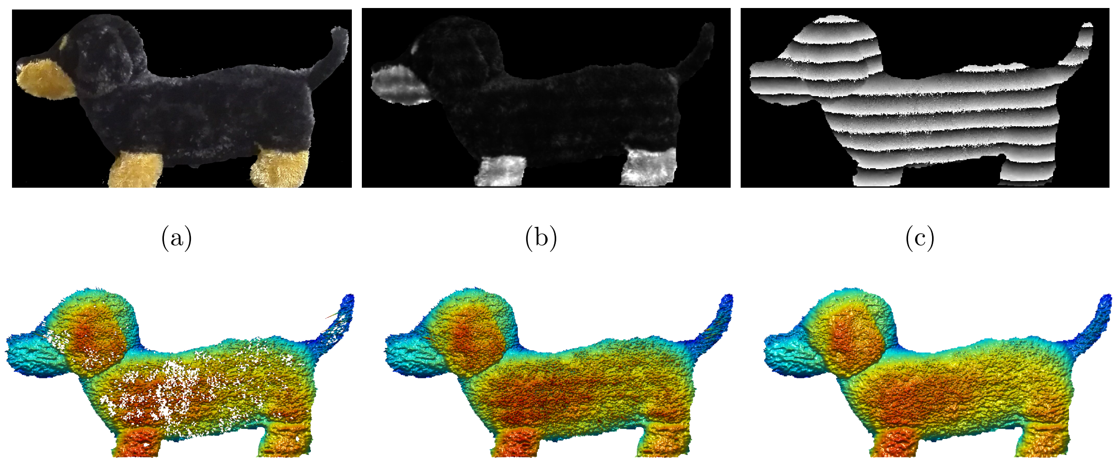


Figure 5.8. Measurement result by our proposed method. (a) Artificially generated minimum phase map, Φ_{min}, using geometric constraints of the structured light system; (b) Unwrapped phase map Φ by our proposed method; (c) Reconstructed 3D geometry; (d) Unwrapped phase comparison in a cross section between our proposed method and the conventional phase unwrapping one.

especially using a three-step phase-shifting algorithm due to complex structure and low contrast. Figure 5.9(b) shows one of three phase-shifted fringe patterns, demonstrating that the fringe contrast is very low (i.e., low SNR). Applying the three-step phase shifting algorithm yielded the wrapped phase map shown in Fig. 5.9(c). Again, large noise is apparent on the phase map. We then applied a three-frequency phase-unwrapping method with three fringe periods of 36, 162, 800 pixels to temporally unwrap the phase map. Figure 5.9(e) shows 3D reconstruction using the temporally unwrapped phase map: numerous points are not properly measured (holes on the geometry). We then applied our proposed phase unwrapping method to unwrap the wrapped phase map shown in Fig. 5.9(c), and 5.9(d) shows the corresponding 3D reconstruction. This experiment indicates that our proposed method can successfully reconstruct the entire geometry without apparent spiky noise. As a comparison, we also measured the same object with more-step phase-shifting algorithm to obtain *Ground Truth*. Figure 5.9(f) shows the result. In this experiment, we used a ten-frequency phase-shifting algorithm for temporal phase unwrapping and captured 18 equally phase-shifted fringe patterns for each frequency to reduce noise impact.

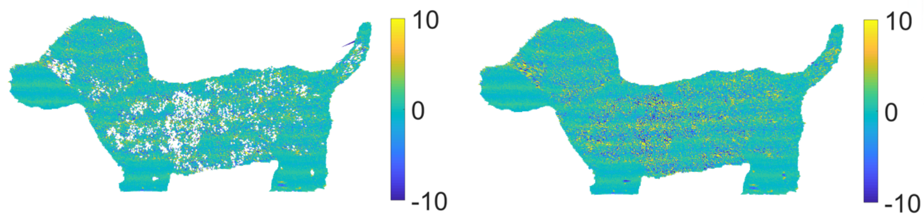


(a) (b) (c)

(d) (e) (f)

Figure 5.9. Phase unwrapping comparison between conventional method and our proposed method. (a) Photography of the testing object; (b) One of the three phase-shifted fringe images; (c) Wrapped phase; (d) 3D geometry reconstructed from the phase obtained by a three-frequency phase-shifting method; (e) 3D geometry reconstructed by the phase obtained from geometric-constraints based phase unwrapping method; (f) *Ground-truth* 3D geometry.

Meanwhile, we compared the 3D reconstruction difference between the three-frequency unwrapping algorithm result and the ground truth, and between our result and the ground truth. Figure 5.10(a) is the 3D reconstruction difference between the reconstruction result using three-frequency phase unwrapping algorithm and the ground truth, and Fig. 5.10(b) is the 3D reconstruction difference between our 3D result and the ground truth. There are mainly two types of errors in the comparing results. One is caused by phase unwrapping error, which will lead to holes in the comparing result. The other one is caused by the inaccurate phase calculation (both the three-frequency algorithm and our method used three-step phase shifted patterns, while the ground truth used eighteen-step phase shifted patterns), which will lead to the depth difference. By comparing Fig. 5.10(a) and Fig. 5.10(b), we can see that our result have less holes, which demonstrating the robustness of our method.



(a) (b)

Figure 5.10. 3D reconstruction error. (a) 3D reconstruction error between the three-frequency phase unwrapping result and the ground truth; (b) 3D reconstruction error between our result and the ground truth.

Finally, we measured a large depth range sphere to compare the difference between our approach and the conventional temporal phase unwrapping approach. Figure 5.11 shows the results. For a large depth range measurement, the proposed method fails to correctly measure the overall object surface, shown in Fig. 5.11(a) and Fig. 5.11(c); yet the conventional temporal phase unwrapping method works well, shown in Fig. 5.11(b) and Fig. 5.11(d), indicating that the proposed method does not have the same measurement capacities as the conventional temporal phase unwrapping algorithm.

To understand the depth range limitation of the proposed method, we need understand how the phase is unwrapped if the object surface point is far away from the z_{min} plane. Figure 5.12 illustrates the maximum depth range, Δz_{max} , that the proposed method can handle. Point A on the z_{min} plane and Point B on the object plane are imaged to the same point by the camera, yet they are projected from different points

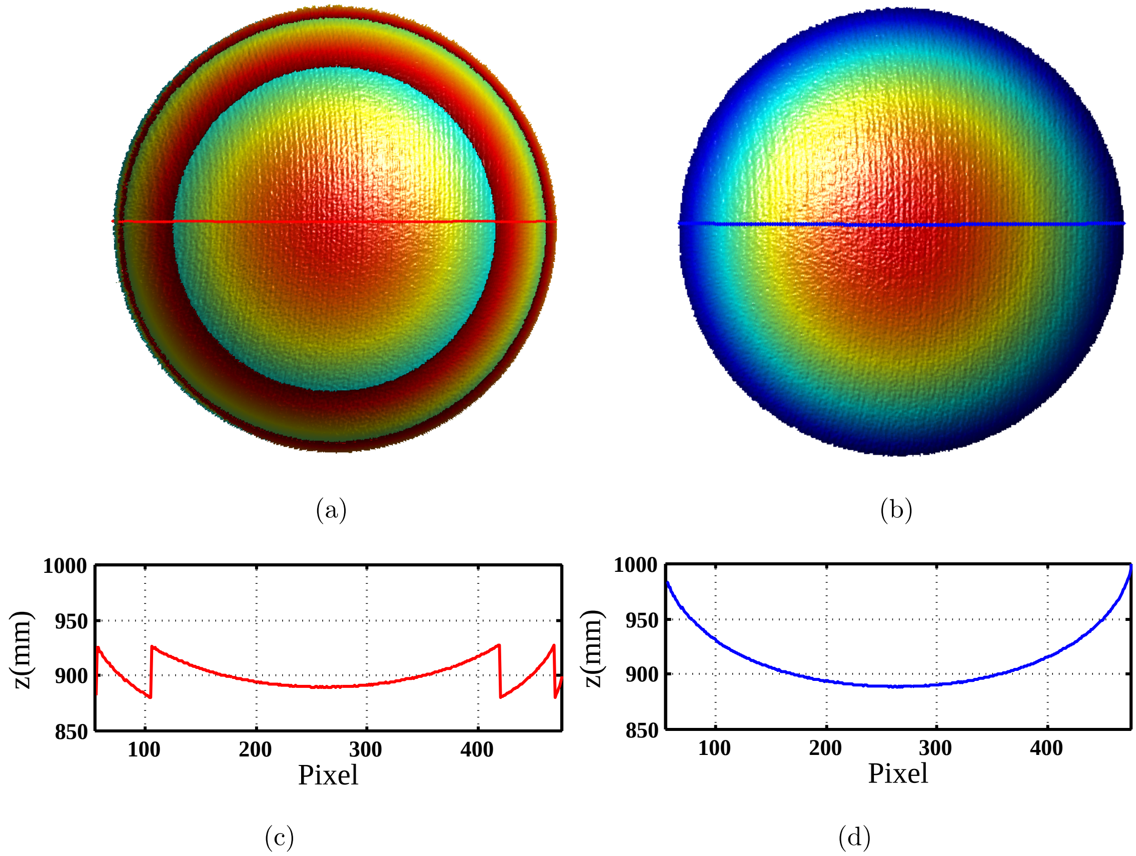


Figure 5.11. Measurement result of a large sphere. For this large depth range sphere, our proposed method fails, while the conventional temporal phase unwrapping approach can work well. (a) Reconstructed 3D geometry by our proposed method; (b) Reconstructed 3D geometry by the conventional temporal phase unwrapping approach; (c) Cross section of the 3D geometry reconstructed by our proposed method; (d) Cross section of the 3D geometry reconstructed by the conventional temporal phase unwrapping approach.

on the projector. If Point A and Point B have more than 2π phase difference from projected patterns, the proposed method fails to determine correct fringe order.

Assuming the angle between projection direction and camera capture direction is θ, and the spatial span of one projected fringe period is Δy, from simple trigono-

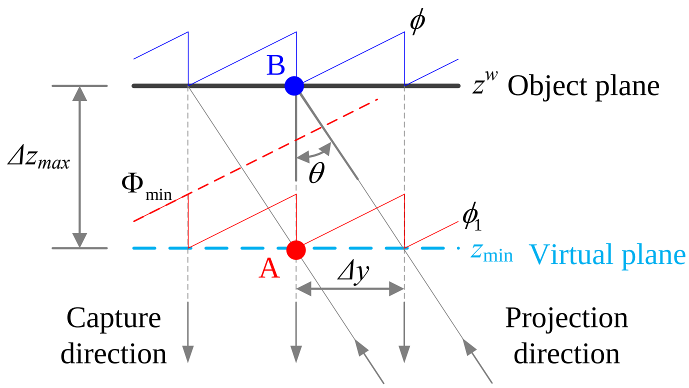


Figure 5.12. The maximum depth range that the proposed absolute phase unwrapping method can handle is defined by the angle between the projector and the camera, the projection matrices for the camera and projector, as well as the projected fringe periods in space.

metrical derivations, we can find that the maximum depth range that our proposed method can handle is

$$\Delta z_{max} = \Delta y / \tan\theta. \tag{5.20}$$

This strong limitation is practically reasonable. For example, considering the experimental system we used for all our experiments, the angle between the projector optical axis and the camera optical axis is approximately $\theta = 10°$. If we project horizontal fringe patterns with a fringe period of 20 pixels, which is approximately $\Delta y = 20/800 = 0.025 = 2.5\%$ of the overall range of the projection area along y or vertical direction. Here 800 is the overall height of projector sensor in pixels. For this case, the depth range is limited to $\Delta z_{max} = \Delta y / \tan\theta = 0.14 = 14\%$. Furthermore, since our camera only captures approximately 3/4 of the projectors projection area, the overall maximum depth range is approximately $0.14 \times 4/3 = 0.19 = 19\%$ of sensing range of the camera, which is pretty good. If the camera is sensing 300 mm along y axis, the overall depth range the proposed method is approximately 58 mm, which is reasonable for many applications. To further increase the maximum depth range, one can increase fringe period, or decrease the angle between the projector and the camera.

5.4 Discussion

This proposed pixel-wise absolute phase unwrapping method has the following advantages:

- *High-speed 3D shape measurement.* Unlike traditional temporal phase unwrapping method, the proposed absolute phase unwrapping method does not require any additional image acquisition, and thus it is more suitable for high-speed applications.

- *High-speed processing.* The proposed method is inherently a pixel operation that does not refer to neighboring pixels or using any filters; the processing speed is fast especially if it is implemented on a parallel processor (e.g., graphics processing unit, GPU).

- *Simple system setup.* Unlike those state-of-art methods using one more camera without requiring more image acquisition, the proposed method does not change the single-projector and single-camera structured light system set up, and thus it can be directly employed by any conventional structured light system.

- *Simultaneous multiple objects measurement.* Similar to temporal phase unwrapping method, the proposed absolute phase unwrapping is pixel by pixel, and thus can be used to measure multiple objects at exactly the same time, as demonstrated by the experimental data in Sec. 6.3.

- *Robustness in fringe order determination.* The phase unwrapping artifacts (i.e. spikes) are minimum without any filtering, indicating that fringe order determination is very robust. This is because the proposed method determines fringe order by referring to an artificially generated ideal absolute phase map Φ_{min}

without any noise. In comparison, the conventional temporal phase unwrapping method determines fringe order by referring to other camera captured information that contains noise.

However, this proposed absolute phase unwrapping method is not trouble free, as demonstrated in our experimental data (Fig. 9). The major limitations are:

- *Confined measurement depth range.* As mentioned above, the maximum measurement depth range that the proposed approach can handle is within 2π changes in phase domain from the object plane to the minimum phase generation plane. In other words, any point on the object surface should not be too far away from z_{min} such that it will cause more than 2π changes. This is practically reasonable since the overall maximum depth range for our measurement system is approximately 19% of the cameras overall sensing range.

- *Good z_{min} estimation.* Since the maximum depth range is limited by the distance from z_{min} plane to object plane, more accurate use of z_{min} plane leads to larger depth measurement range; and incorrect use of z_{min} plane could lead to incorrect phase unwrapping. In our research, we coincide the world coordinate system with the camera lens coordinate. By doing so, z_{min} plane has the minimum z^w value for 3D reconstruction. By doing so, one can estimate z_{min} of interest by a variety of means, one of which being the use of a ruler to measure the distance from the closet object point to the camera lens.

Even with these limitations, the proposed pixel-by-pixel absolute phase unwrapping without the use of any additional image or hardware can substantially benefit the optical metrology field, especially for applications where high-speed absolute 3D shape measurement is required.

5.5 Summary

This chapter has presented a method to unwrap phase pixel by pixel by referring to the artificial minimum phase map created solely using geometric constraints of the structured light system. Unlike conventional temporal phase unwrapping algorithms that require one to capture more images, the proposed absolute phase unwrapping method requires no additional image acquisition. Compared with those absolute phase measurement methods that use one additional camera, the proposed method does not require any additional camera to obtain absolute phase. Since it does not require any additional image acquisition or another camera, the proposed method has the advantage of measurement speed without increasing system complexity or cost. Experimental results demonstrated the success of our proposed pixel-by-pixel absolute phase unwrapping method. Despite its confined depth range, the proposed method is of significance to applications where high-speed 3D absolute shape measurement is necessary.

6. PIXEL-BY-PIXEL ABSOLUTE PHASE RETRIEVAL ASSISTED BY AN ADDITIONAL THREE-DIMENSIONAL SCANNER

The geometric constraint-based phase unwrapping method introduced in the previous chapter is confined in a limited depth range. To address the depth limitation problem, this chapter will present a novel absolute phase unwrapping method assisted by a low-cost three-dimensional (3D) scanner. The proposed absolute phase unwrapping method leverages a low-cost 3D scanner to capture rough 3D data of the scene and transforms the rough 3D data to the world coordinate system to generate an artificial reference phase map Φ_{ref} . By referring to Φ_{ref}, we can do absolute phase unwrapping directly without projecting any additional patterns, such that the digital fringe projection (DFP) system can achieve higher measurement speed. We develop a multiresolution system consisting of a DFP system and a Kinect V2 to validate our method. Experiments demonstrate that our method works for a large depth range, and the speed of the low-cost 3D scanner is not necessarily the maximum speed of our proposed method. Assisted by the Kinect V2, whose maximum speed is only 30Hz, our DFP system achieves 53Hz with a resolution 1600×1000 pixels when we measure dynamic objects that are moving in a large depth range of 400mm. The majority of this chapter was originally published in *Applied Optics* [132] (also listed as journal article [J9] in "LIST OF PUBLICATIONS").

6.1 Introduction

Three-dimensional (3D) shape measurement has applications in various fields including online inspection, disease diagnosis, and entertainment. Among various 3D shape measurement methods, digital fringe projection (DFP) usually can achieve high resolution and high accuracy. Yet, generally, fringe analysis methods, such as Fourier transform methods [42,114], demodulation and convolution methods [133], and phase-shifting methods [44], only provide phase values ranging from $-\pi$ to π with a modulus of 2π. The 2π jumps need to be removed to obtain a continuous unwrapped phase map to do 3D shape measurement.

Numerous phase unwrapping methods have been developed over the past decades. In general, phase unwrapping methods can be classified into two categories: spatial phase unwrapping and temporal phase unwrapping. Spatial phase unwrapping methods detect 2π jumps from the wrapped phase map itself and add a multiple integer $k(x, y)$of 2π's to each pixel according to the phase values of neighboring pixels. The multiple integer $k(x, y)$is also referred to as the *fringe order*. Ghiglia and Pritt [134] summarized various spatial phase unwrapping methods. Later, Su and Chen [116] reviewed many reliability-guided spatial phase unwrapping algorithms, and quality-guided spatial phase unwrapping algorithms were compared by Zhao et al. [135]. Although numerous improvements have been made, spatial phase unwrapping algorithms typically only provide a relative unwrapped phase map since phase values are usually relative to a specific point on the phase map itself. Consequently, spatial phase unwrapping methods are fundamentally limited to measuring smooth surfaces (the object has to be smooth in at least one path such that the geometry will not lead to more than π phase change in two successive points). It is challenging to apply spatial phase unwrapping methods to measure objects with abrupt depth changes and simultaneously measure multiple isolated objects.

Different from spatial phase unwrapping methods, temporal phase unwrapping methods usually encode fringe orders into additional patterns. For instance, gray coding methods project additional gray coded patterns to retrieve a fringe order for each pixel [47]. Two- or multiwavelength phase shifting algorithms [117, 118] leverage additional patterns with different fringe periods. The unwrapped phase map provided by temporal phase unwrapping methods is an absolute unwrapped phase map which can be used to calculate absolute 3D world coordinates directly. Whereas, since additional patterns need to be projected, temporal phase unwrapping methods inherently slow down the measurement speed and are not suitable for high-speed applications.

In order to solve the speed problem of temporal phase unwrapping algorithms, researchers attempted to add one or several more cameras to the standard single camera single projector DFP system for phase unwrapping. Typically, those methods [124, 126, 136, 137] can be categorized as stereo phase unwrapping because they are usually based on the stereo geometric constraints between two or multiple cameras. Basically, a pixel on one camera image is constrained to correspond to several candidate pixels with the same phase value on an epipolar line on another camera. The final correct corresponding pixel can be determined with the combination of predefined measurement volume, leftright consistency, and other techniques. Over recent years, many improvements have been achieved, such as calibration improvement [126] and adaptive depth constraint [137], to make the stereo phase unwrapping methods more powerful. However, stereo phase unwrapping methods typically need a global backward and forward consistency checking to determine a unique corresponding point, and the consistency checking slows down the computational speed. Also, adding one or several more cameras only for phase unwrapping increases the hardware cost.

To address problems of the aforementioned phase unwrapping methods, An et al. [113] proposed a geometric constraint based absolute phase unwrapping method for the single camera single projector system without projecting any additional patterns. The geometric constraintbased phase unwrapping method sets a virtual plane z_{min} at the nearest point on the object away from the system. Then the geometric constraints between the projector and the camera are utilized to generate an artificial phase map Φ_{min}, and absolute phase unwrapping can be done by referring to the artificial phase map Φ_{min} pixel by pixel. Details of the geometric constraintbased phase unwrapping method will be introduced in Section 6.2.2 in this chapter. However, the geometric constraintbased phase unwrapping method is limited to measuring objects that are within a certain depth range Δz away from the virtual plane z_{min}, making it difficult to measure objects with large depth variances or objects moving in a large depth range.

Instead of using a virtual plane, we can use a rough 3D measurement data, away from which the real object could be always within Δz, such that the depth range constraint of the geometric constraintbased phase unwrapping method could be solved. Nowadays, multiresolution shape measurement systems are used in many applications, such as defect inspection [138] and criminal scene documentation [139]. A multiresolution shape measurement system usually can provide such a rough 3D measurement to generate a new artificial phase map for absolute phase unwrapping.

Based on the above idea, we propose a novel absolute phase unwrapping method that significantly enhances the geometric constraintbased phase unwrapping algorithm to relax the constraint of depth range. Conceptually, we take advantage of an additional 3D scanner in a multiresolution system to obtain rough 3D data of objects, and then transform the rough 3D to the DFP coordinate system. With the transformed rough 3D data, we can generate a novel artificial phase map Φ_{ref} , and then use Φ_{ref} to do phase unwrapping pixel by pixel. Comparing with a virtual plane z_{min}, the rough 3D is closer to the real object such that all parts of the real object are within a small distance away from the rough 3D. Therefore, our proposed method using Φ_{ref} generated from rough 3D data can address the depth limitation problem of the geometric constraintbased method that uses Φ_{min}. We develop a multiresolution system consisting of a DFP system and a Kinect V2 to validate our proposed method. We have done experiments on objects with large depth variances, on multiple isolated objects at different depths, and on dynamic objects moving along the depth direction. Experiments demonstrate the success of our proposed method. Besides, though our phase unwrapping method is assisted by a low-cost 3D scanner, we show that the speed of the additional 3D scanner is not necessarily the maximum speed of our proposed method. The maximum speed of the Kinect is only 30Hz, but our DFP system still achieves 53Hz with a resolution 1600×1000 pixels when we measure dynamic objects that are moving in a large depth range of 400mm.

6.2 Principle

In this section, we will thoroughly explain the principles behind our proposed method. Specifically, we will introduce the three-step phase shifting algorithm and the geometric constraintbased phase unwrapping algorithm, and then elucidate the framework of our proposed method.

6.2.1 Three-step phase shifting algorithm

Phase shifting algorithms have been extensively applied over the past decades due to their robustness and accuracy. Particularly, the three-step phase shifting algorithm is desired for high-speed applications since it needs the least number of fringe patterns for phase retrieval. The three fringe images can be described using the following equations in mathematics:

$$I_1(x, y) = I'(x, y) + I''(x, y) \cos[\phi(x, y) - 2\pi/3], \tag{6.1}$$

$$I_2(x, y) = I'(x, y) + I''(x, y) \cos[\phi(x, y)], \tag{6.2}$$

$$I_3(x, y) = I'(x, y) + I''(x, y) \cos[\phi(x, y) + 2\pi/3], \tag{6.3}$$

where $I'(x, y)$ is the average intensity, $I''(x, y)$ is the intensity modulation, and $\phi(x, y)$ is the phase that needs to be calculated. By solving the equation set (6.1)-(6.3), we can obtain $I'(x, y)$ and $\phi(x, y)$ as:

$$I'(x, y) = [I_1(x, y) + I_2(x, y) + I_3(x, y)]/3, \tag{6.4}$$

$$\phi(x, y) = \tan^{-1}\left\{\frac{\sqrt{3}[I_1(x, y) - I_3(x, y)]}{2I_2(x, y) - I_1(x, y) - I_3(x, y)}\right\}. \tag{6.5}$$

The phase map $\phi(x, y)$ is a wrapped one ranging from $-\pi$ to π with a modulus of 2π due to the arctangent function in (6.5). To obtain a continuous phase map without 2π jumps, we need a phase unwrapping algorithm. Essentially, phase unwrapping algorithms try to add an integer multiple of 2π's to each pixel as the following equation:

$$\Phi(x, y) = \phi(x, y) + 2\pi \times k(x, y), \tag{6.6}$$

where $\Phi(x, y)$ is a continuous unwrapped phase map, and $k(x, y)$ is an integer number that is also referred to as fringe order. $\Phi(x, y)$ is an absolute unwrapped phase map that can be directly used to recover absolute 3D world coordinates of each pixel if the fringe order can be uniquely determined based on a pre-defined value. Numerous spatial and temporal phase unwrapping methods have been developed over the past decades. Yet, as discussed in Sec. 6.1, spatial phase unwrapping methods only provide a relative phase map that is referred to a specific point on the phase map. Temporal phase unwrapping methods can provide an absolute unwrapped phase map, but they usually require additional patterns, thus slowing down the measurement speed.

6.2.2 Geometric constraint-based phase unwrapping algorithm

An et al. [113] developed a geometric constraint-based absolute phase unwrapping algorithm that needs no additional patterns. The geometric constraint-based phase unwrapping algorithm leverages the inherent geometric constraint of the DFP system that consists of one camera and one projector to generate an artificial phase map, and refers to the artificial phase map to do absolute phase unwrapping.

In mathematics, both the projector and the camera can be described as a pinhole model to establish a relationship between 3D world coordinates (x^w, y^w, z^w) and corresponding image coordinates (u, v) as follows:

$$s^c \begin{bmatrix} u^c & v^c & 1 \end{bmatrix}^t = \mathbf{A}^c \begin{bmatrix} \mathbf{R}^c & \mathbf{T}^c \end{bmatrix} \begin{bmatrix} x^w & y^w & z^w & 1 \end{bmatrix}^t, \tag{6.7}$$

$$s^p \begin{bmatrix} u^p & v^p & 1 \end{bmatrix}^t = \mathbf{A}^p \begin{bmatrix} \mathbf{R}^p & \mathbf{T}^p \end{bmatrix} \begin{bmatrix} x^w & y^w & z^w & 1 \end{bmatrix}^t, \tag{6.8}$$

where the superscript ^{c} stands for camera; the superscript ^{p} stands for projector; the superscript ^{t} represents the transpose operation of a matrix; $\mathbf{A}$ is a 3×3 intrinsic matrix; $\mathbf{R}$ and $\mathbf{T}$ are a 3×3 rotation matrix and a 3×1 translation vector respectively. $\mathbf{A}$, $\mathbf{R}$ and $\mathbf{T}$ can be estimated using a DFP system calibration method [78, 100].

For a camera pixel (u^c, v^c), (6.7)-(6.8) only provide 6 equations, yet there are 7 unknowns $(x^w, y^w, z^w, u^p, v^p, s^c, s^p)$. To solve the 3D coordinates (x^w, y^w, z^w) for a camera pixel (u^c, v^c), one additional equation is necessary. The absolute phase map Φ provides that additional necessary equation.

Alternatively, if we set a virtual plane at $z = z_{min}$, a unique projector pixel (u^p, v^p) can be determined for each camera pixel (u^c, v^c) through (6.7)-(6.8). Once the corresponding projector pixel (u^p, v^p) is determined, a virtual absolute phase value $\Phi_{min}(u^c, v^c)$ can be calculated. In mathematics, the process of generating an artificial phase value $\Phi_{min}(u^c, v^c)$ can be described as following.

For notation simplicity, we define a camera projection matrix $\mathbf{P}^c$ and a projector projection matrix $\mathbf{P}^p$ as $\mathbf{P}^c = \mathbf{A}^c[\mathbf{R}^c\ \mathbf{T}^c]$, $\mathbf{P}^p = \mathbf{A}^p[\mathbf{R}^p\ \mathbf{T}^p]$. Both $\mathbf{P}^c$ and $\mathbf{P}^p$ are of dimension 3×4, and we use p^c_{ij} to denote the i^{th} row and j^{th} column element in $\mathbf{P}^c$, p^p_{ij} to denote the i^{th} row and j^{th} column element in $\mathbf{P}^p$. Then, for each camera pixel (u^c, v^c) with a known depth value z_{min}, its corresponding coordinates in x and y axes (x_{min}, y_{min}) can be computed as:

$$\begin{bmatrix} x_{min} \\ y_{min} \end{bmatrix} = \mathbf{M}^{-1}\mathbf{b}, \tag{6.9}$$

where

$$\begin{aligned} \mathbf{M} &= \begin{bmatrix} p^c_{31}u^c - p^c_{11} & p^c_{32}u^c - p^c_{12} \\ p^c_{31}v^c - p^c_{21} & p^c_{32}v^c - p^c_{22} \end{bmatrix}, \\ \mathbf{b} &= \begin{bmatrix} p^c_{14} - p^c_{34}u^c - (p^c_{33}u^c - p^c_{13})z_{min} \\ p^c_{24} - p^c_{34}v^c - (p^c_{33}v^c - p^c_{23})z_{min} \end{bmatrix}. \end{aligned} \tag{6.10}$$

Then the corresponding projector pixel (u^p, v^p) of the camera pixel (u^c, v^c) can be determined using the pinhole model as:

$$s^p \begin{bmatrix} u^p & v^p & 1 \end{bmatrix}^t = \mathbf{P}^p \begin{bmatrix} x_{min} & y_{min} & z_{min} & 1 \end{bmatrix}^t. \tag{6.11}$$

Suppose the fringe patterns vary along u^p direction with a fringe period of T^p pixels. Then the artificial phase value for the camera pixel (u^c, v^c) is:

$$\Phi_{min}(u^c, v^c) = u^p \times 2\pi/T^p. \tag{6.12}$$

By referring to the artificial phase map Φ_{min}, An et al. [113] proved that the fringe order of camera pixel (u^c, v^c) can be obtained using the following equation:

$$k(u^c, v^c) = ceil\left[\frac{\Phi_{min}(u^c, v^c) - \phi(u^c, v^c)}{2\pi}\right]. \tag{6.13}$$

6.2.3 Proposed absolute 3D shape measurement method

However, the above geometric constraint-based phase unwrapping method has a confined measurement depth range. The maximum depth range it can measure is:

$$\Delta z \approx T_s/\tan(\theta), \tag{6.14}$$

where T_s is the spatial span of one fringe period, θ is the angle between projector direction and camera direction, Δz is the distance from the virtual plane z_{min}. In another word, the object has to be within Δz from the virtual plane z_{min} for correct geometry reconstruction. This constraint makes the method be challenging to measure objects with large depth variances and to measure moving objects which could move beyond the depth range.

Instead of using a virtual plane z_{min}, we propose to take assistance of an additional 3D scanner. Figure 6.1 shows the conceptual idea of our proposed method. We use an additional 3D scanner to obtain a rough measurement of the object, and transform the rough 3D to the DFP coordinate system. Then we use the transformed rough 3D to generate a new artificial phase map Φ_{ref}. Since all parts of the real object are close to the captured rough 3D, the new artificial phase map Φ_{ref} generated from the rough 3D could do phase unwrapping without depth range constraint of the virtual plane z_{min} method.

For instance, we can use Kinect V2 as the additional 3D scanner. The whole system needs to be calibrated first. We can use the method proposed in [78] to do

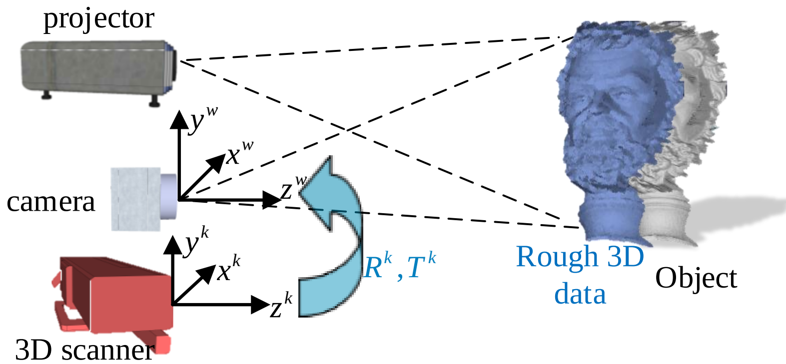


Figure 6.1. Conceptual idea of the proposed method. We leverage an additional 3D scanner to capture a rough 3D of the object, and transform the rough 3D to the world coordinate system. The transformed rough 3D is used to generate a novel artificial phase map Φ_{ref} to assist us to do phase unwrapping.

calibration, and obtain intrinsic parameters of the camera and the projector $\mathbf{A}^c$, $\mathbf{A}^p$; extrinsic parameters $\mathbf{R}^c$, $\mathbf{T}^c$, $\mathbf{R}^p$, $\mathbf{T}^p$; and a rotation matrix $\mathbf{R}^k$ and a translation vector $\mathbf{T}^k$ that transform 3D data from the Kinect coordinate system to the world coordinate system. The world coordinate system is set to be the same as the DFP coordinate system in this chapter. Details of the calibration process can be referred to [78].

Then we use the additional 3D sensor to capture a rough 3D of the measured scene, and transform the 3D data to the DFP coordinate system based on $\mathbf{R}^k$ and $\mathbf{T}^k$. Mathematically,

$$\begin{bmatrix} x^{kt} & y^{kt} & z^{kt} \end{bmatrix}^t = \begin{bmatrix} \mathbf{R}^k & \mathbf{T}^k \end{bmatrix} \begin{bmatrix} x^k & y^k & z^k & 1 \end{bmatrix}^t, \tag{6.15}$$

where (x^k, y^k, z^k) is the 3D data provided by the additional 3D sensor, (x^{kt}, y^{kt}, z^{kt}) is the same 3D data in the DFP coordinate system.

(x^{kt}, y^{kt}, z^{kt}) can serve as the similar function as a virtual plane $(x_{min}, y_{min}, z_{min})$ in the geometric constraint-based phase unwrapping method. Yet, because of noise of the additional 3D sensor and calibration accuracy, the transformed 3D points (x^{kt}, y^{kt}, z^{kt}) may not lie in front of the actual object. We move forward the transformed 3D points by deducting a value z^{tol} on z^{kt}, such that the transformed 3D is in front of the real object as:

$$z^{true} - \Delta z < z^{kt} - z^{tol} < z^{true}, \tag{6.16}$$

where z^{true} is the actual depth of the real object. The value of z^{tol} can be determined according to the noise of the additional 3D scanner and the calibration accuracy of the overall system. Based on the 3D data $(x^{kt}, y^{kt}, z^{kt} - z^{tol})$, we can establish a correspondence relationship between camera and projector pixels through the following equations:

$$s^c \begin{bmatrix} u^c & v^c & 1 \end{bmatrix}^t = \mathbf{A}^c \begin{bmatrix} \mathbf{R}^c & \mathbf{T}^c \end{bmatrix} \begin{bmatrix} x^{kt} & y^{kt} & z^{kt} - z^{tol} & 1 \end{bmatrix}^t, \tag{6.17}$$

$$s^p \begin{bmatrix} u^p & v^p & 1 \end{bmatrix}^t = \mathbf{A}^p \begin{bmatrix} \mathbf{R}^p & \mathbf{T}^p \end{bmatrix} \begin{bmatrix} x^{kt} & y^{kt} & z^{kt} - z^{tol} & 1 \end{bmatrix}^t. \tag{6.18}$$

Similarly as Φ_{min} generation method, if we suppose fringe patterns vary along u^p direction with a fringe period T^p, a new reference artificial phase value can be obtained as:

$$\Phi_{ref}(u^c, v^c) = u^p \times 2\pi / T^p. \tag{6.19}$$

The fringe order $k(u^c, v^c)$ can be determined similarly as (6.13) as well. Mathematically,

$$k(u^c, v^c) = ceil\left[\frac{\Phi_{ref}(u^c, v^c) - \phi(u^c, v^c)}{2\pi}\right]. \tag{6.20}$$

Appropriate fringe density needs to be selected in this method. Usually, 3D data from the low-cost 3D scanner after coordinate system transformation are around the real object, that is $z^{kt}(u^c, v^c) = z^{true}(u^c, v^c) + \epsilon(u^c, v^c)$ where ϵ is the alignment difference caused by the noise of the low-cost 3D scanner and calibration accuracy, and z^{true} is the real object. The range of ϵ in the whole measurement volume can usually be obtained during the system calibration process. For correct phase unwrapping, the reference 3D should be within Δz in front of the object for correct phase unwrapping. When the fringe period decreases (fringe density increases), Δz will become smaller, such that the range of reference 3D become smaller. However, Δz should be always larger than the range of alignment difference ϵ for robust phase unwrapping, from which the smallest Δz can be determined. Then through (6.14), we can obtain the smallest fringe period (maximum fringe density), and appropriate fringe patterns should have larger fringe period than that.

In practice, the additional 3D scanner usually has a much lower resolution than the DFP system. To obtain a reference 3D coordinate for each camera pixel, we need to up-sample 3D data captured by the additional 3D scanner. Simple interpolation can be used to do up-sampling. For instance, we can use the bilinear interpolation. Suppose we want to increase the resolution of the additional 3D scanner data by a factor s in both the row and column direction. Then, for each grid square consisting of (u_i^k, v_i^k), (u_i^k+1, v_i^k), (u_i^k, v_i^k+1) and (u_i^k+1, v_i^k+1), we interpolate 3D coordinates for $(u_i^k + s_i, v_i^k + s_j)$ as:

$$\begin{aligned}
x(u_i^k + s_i, v_i^k + s_j) =&(1 - s_i)[(1 - s_j)x(u_i^k, v_i^k) + s_j x(u_i^k, v_i^k + 1)]+\\
&s_i[(1 - s_j)x(u_i^k + 1, v_i^k) + s_j x(u_i^k + 1, v_i^k + 1)],\\
y(u_i^k + s_i, v_i^k + s_j) =&(1 - s_i)[(1 - s_j)y(u_i^k, v_i^k) + s_j y(u_i^k, v_i^k + 1)]+\\
&s_i[(1 - s_j)y(u_i^k + 1, v_i^k) + s_j y(u_i^k + 1, v_i^k + 1)],\\
z(u_i^k + s_i, v_i^k + s_j) =&(1 - s_i)[(1 - s_j)z(u_i^k, v_i^k) + s_j z(u_i^k, v_i^k + 1)]+\\
&s_i[(1 - s_j)z(u_i^k + 1, v_i^k) + s_j z(u_i^k + 1, v_i^k + 1)],
\end{aligned} \tag{6.21}$$

where $s_i = 0, 1/s, 2/s, \ldots, 1$ and $s_j = 0, 1/s, 2/s, \ldots, 1$.

Also, perspective difference between the DFP system and the additional 3D scanner needs to be considered. First, perspective difference could cause multiple 3D points projected to a same camera pixel. As shown in Fig. 6.2, both points E' and F' are projected to a same camera pixel. To solve this problem of multiple corresponding rough 3D points, we can only store the point with the smallest depth value for the camera pixel since other 3D points with larger depth values will be occluded by a nearer one in the camera's view. By doing this, we can determine a correct 3D point for the camera pixels and prevent wrong artificial phase values later.

Meanwhile, perspective difference can cause artifacts, such as holes and missing boundaries in the reference phase map. For instance, curve $\widehat{AB}$ in Fig. 6.2 can be

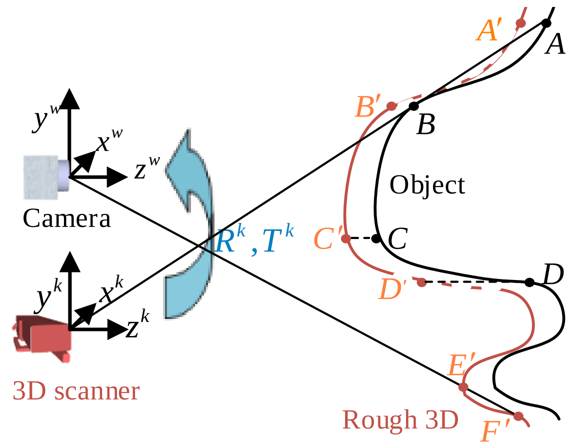


Figure 6.2. Problems caused by view difference and calibration accuracy. Curve $\widehat{AB}$ can be seen by the camera, but not the 3D scanner, leading to missing reference 3D coordinates in the reference phase map Φ_{ref}. Also, multiple 3D points could be projected to a same camera pixel as E' and F', and we will only store the point with the smallest depth value for that camera pixel. Generally, 3D points with large depth gradient could have a large alignment difference from the real object as the comparison example between C' and D'. To guarantee each camera pixel's reference 3D coordinates are reliable, we will remove 3D points with large depth gradients.

seen by the camera, but not the additional 3D scanner. Therefore, no reliable 3D coordinates can be directly assigned to corresponding camera pixels, leading to holes in the reference phase map Φ_{ref}. Also, due to the calibration accuracy of $\mathbf{R}^k$ and $\mathbf{T}^k$, the transformed rough 3D data in the world coordinate system may not align well with the real object, especially 3D points with a large depth gradient (or depth change) could have a large alignment difference from the real object. As in Fig. 6.2, Point D' has a much larger alignment difference than C' and simultaneously D' has a much larger depth gradient than C'. Thus, we will remove points with large depth gradients to increase the reliability of each camera pixel's reference 3D coordinates. Removing points with large depth gradients can also cause holes in the reference phase map Φ_{ref}. If the above hole problems caused by perspective difference and calibration accuracy occur on boundaries, we usually call the problem missing boundaries instead of holes. To alleviate these artifacts of holes and missing boundary problems, we used the method proposed in [52] to do hole filling and bring back the missing boundary information to refine our phase unwrapping result. Briefly, holes are filled using interpolation and the phase constraint and boundaries are filled using extrapolation and the phase constraint. Details of the refinement algorithm can be referred to [52].

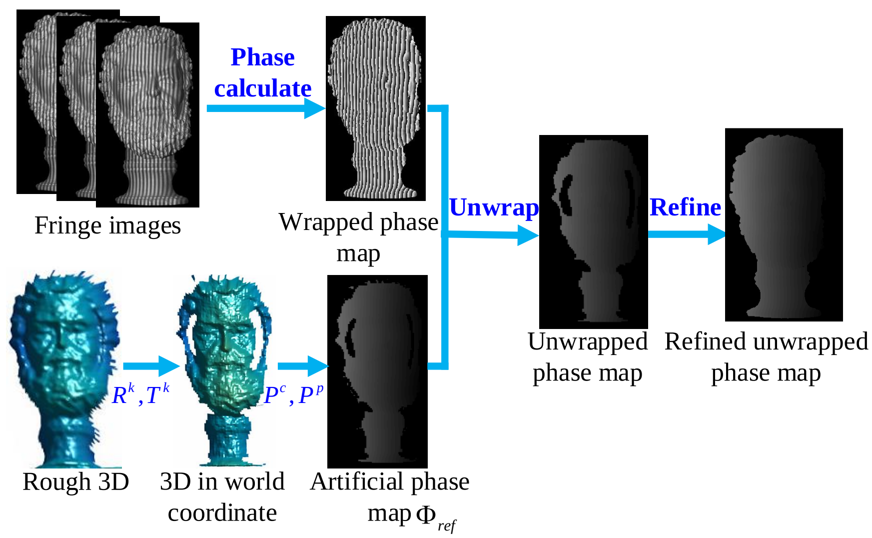


Figure 6.3. Procedures of the proposed phase unwrapping method. From fringe images, we do phase calculation and obtain a wrapped phase map. We use an additional 3D scanner to capture a rough 3D of the object, and transform the rough 3D to the world coordinate system. Due to view difference, holes and boundary artifacts could happen when transforming the coordinate system. Then we project the transformed 3D to the camera and project image planes to find the correspondence between projector and camera pixels, and this correspondence can be leveraged to build a novel artificial phase map Φ_{ref}. Φ_{ref} is used to do phase unwrapping. In the end, we add a refinement stage to alleviate artifacts, such as holes and missing boundaries.

In a summary, the complete phase unwrapping procedures are shown in Fig. 6.3. From the fringe images, we can calculate the wrapped phase based on (6.5). An additional 3D scanner can provide rough 3D data of the object. Then we transform the rough 3D to the DFP coordinate system through $\mathbf{R}^k$, $\mathbf{T}^k$. Due to the quality of the additional 3D scanner, up-sampling may be necessary. Also, because of perspective difference and calibration accuracy, holes and missing boundaries could happen on the transformed 3D data. Then we project the transformed 3D data to the camera image coordinate through $\mathbf{P}^c, \mathbf{P}^p$ using (6.17) and (6.18), and generate a new artificial phase map Φ_{ref} through (6.19). Next, (6.20) is adopted to do phase unwrapping pixel by pixel. We do refinement on the unwrapped phase map using interpolation and extrapolation to fill holes and bring back missing boundary information as the last step. In this framework, the artificial phase map Φ_{ref} is generated based on captured 3D data, instead of a virtual plane. The captured 3D data can be very close to the real 3D object. Therefore, our proposed method can address the confined depth range problem of virtual plane method. As a result, our proposed method can work for objects with large depth variances, and dynamic objects moving in a large depth range.

6.3 Experiment

To verify our proposed method, we set up a multi-resolution system as shown in Fig. 6.4. We used Kinect V2 as the additional 3D scanner whose resolution is 512×424 pixels. The working distance of Kinect V2 is 0.5m $\sim$ 4.5m. The DFP system consists of a digital-light-processing (DLP) projector (Model: LightCrafter 4500) whose resolution is 1140×912 pixels, a CMOS camera (Model: PointGrey Grasshopper 3 GS3-U3-23S6M-C) attached with a 12mm focal length lens (Model: Computar M1214-MP2). The resolution of the camera is set to be 1600×1000 pixels. The baseline between the projector and the camera is ∼190mm. To reduce the interference of Kinect light to the DFP system, we put a filter (Model: Hoya UV&IR Cut) in front of the camera.

We calibrated the multi-resolution system using the method proposed in [78]. Also, from the calibration process, we found the range of alignment difference in the measurement volume caused by the noise of Kinect V2 and calibration accuracy.

We chose fringe period as 18 pixels with which the maximum depth range Δz is about 57mm (10/atan(190/1070)≈57mm) based on (6.14) and Δz is larger than the alignment difference range. Then, we use z^{tol} to adjust the reference 3D data from Kinect such that the reference 3D is within 57mm in front of the measuring object as required by (6.16), and we set z^{tol} to be 30mm as an example in our experiments.

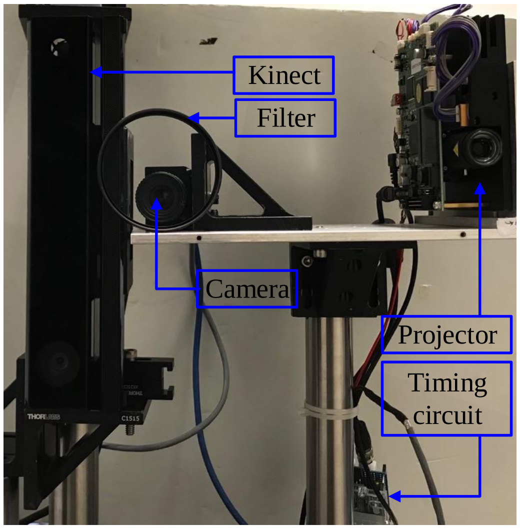
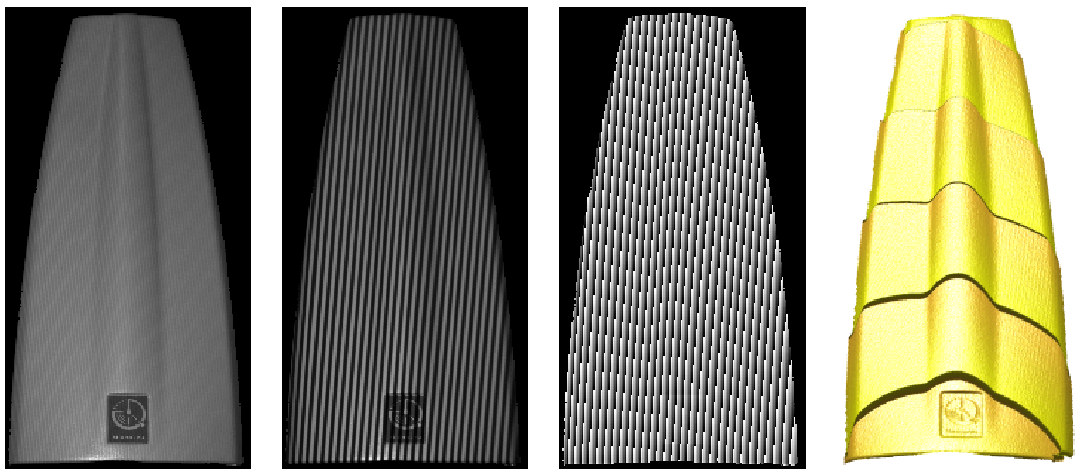


Figure 6.4. System setup. The system consists of a DLP projector, a CMOS camera and an auxiliary 3D sensor Kinect V2. To reduce the interference of Kinect light to the structured light system, we put a filter in front of the camera.

Firstly we measured a single object whose depth variation is about 260mm (from 1070mm to 1330mm), which is larger than the maximum depth range (57mm) of the virtual plane method. Figure 6.5(a) shows the texture image. Figures 6.5(b)-6.5(c) show the fringe image and the wrapped phase map respectively. The 3D measurement result using the virtual plane method is shown in Fig. 6.5(d), from which we can see the object breaks into several pieces due to the measurement depth limitation of the virtual plane method.

To address the depth limitation problem, we take advantage of the additional 3D scanner, Kinect V2, to capture a rough 3D first. The captured rough 3D is shown in Fig. 6.6(a). We up-sample the Kinect 3D data by 10 times, and transform the up-sampled 3D data from Kinect coordinate system to the DFP coordinate system, and the transformed 3D is shown in Fig. 6.6(b). Essentially, the transformed Kinect 3D data provide a reference depth value for each camera pixel. Based on the Kinect 3D data and (6.19), we generate an artificial phase map Φ_{ref} shown in Fig. 6.6(c). Then, we use (6.20) to do phase unwrapping, and obtain the unwrapped phase map shown in Fig. 6.6(d).



(a) (b) (c) (d)

Figure 6.5. Experiment on a single sculpture. (a) Texture image; (b) One of three fringe images; (c) Wrapped phase calculated from the three fringe images; (d) 3D reconstruction using the virtual plane method.

Due to the perspective difference between Kinect and the structured light system, some part of wrapped phase map cannot be unwrapped properly. For instance, the left boundary of Fig. 6.6(d) is sawtooth because left boundary of the object cannot be seen by Kinect. To deal with the problem of perspective difference, we refine the unwrapped phase map of Fig. 6.6(d) by extrapolation and phase constraint, as mentioned in Sec. 6.2.3. The result is shown in Fig. 6.6(e), and we call Fig. 6.6(e) a refined unwrapped phase map. Based on the refined unwrapped phase map, we do 3D reconstruction and the result is shown in Fig. 6.6(f).

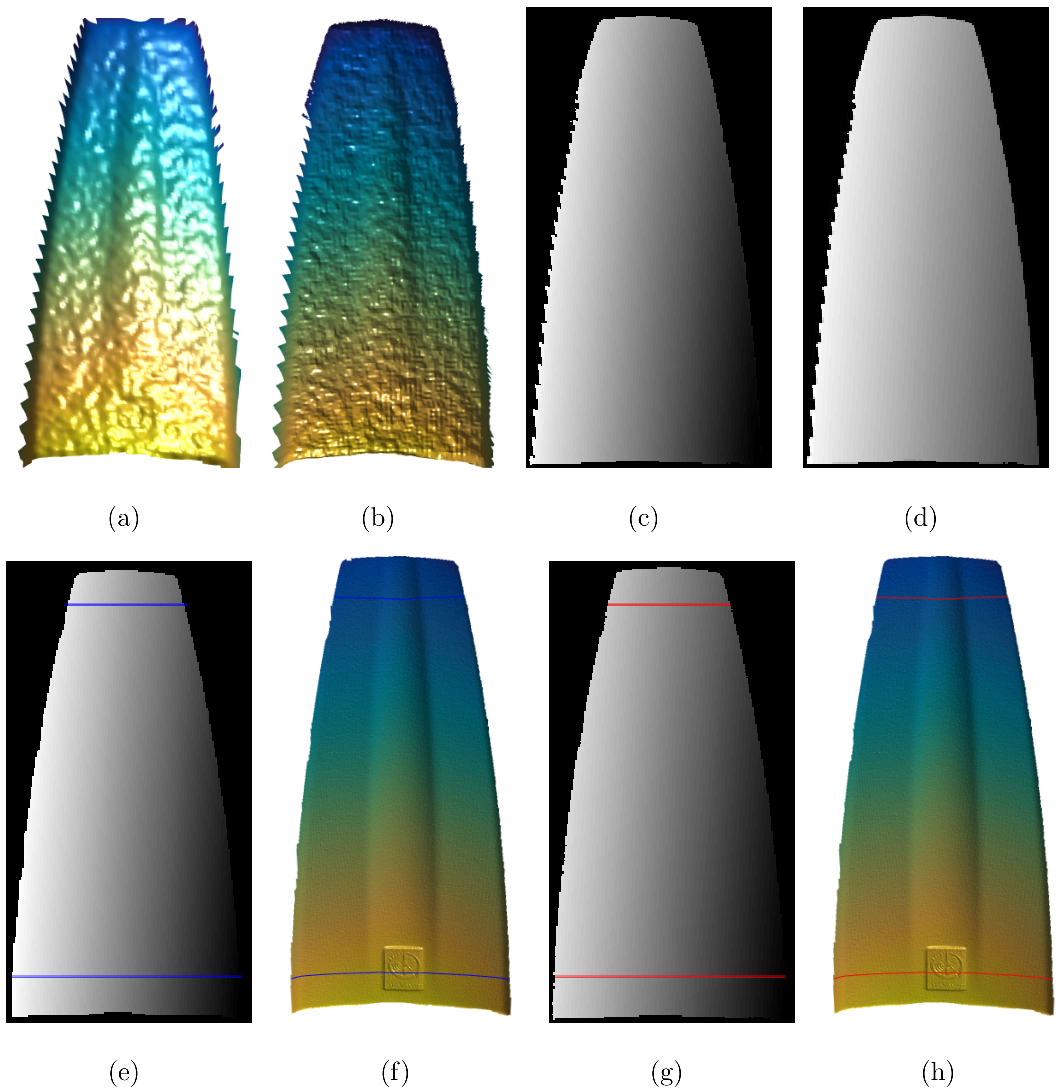
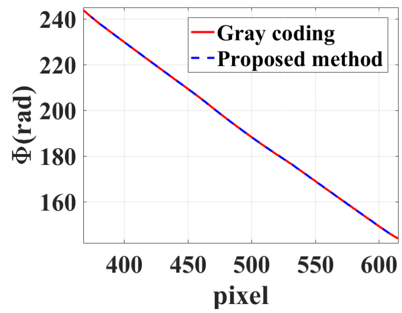


Figure 6.6. 3D shape measurement results on a single object. (a) Rough 3D data captured by Kinect; (b) Rough 3D data in the DFP coordinate system, parts of left boundary are missing due to the viewpoint of Kinect; (c) Φ_{ref} generated from (b); (d) Phase unwrapping results using (b); (e) Refined unwrapped phase map; (f) 3D reconstruction result using (e); (g) Unwrapped phase using gray coding algorithm; (h) 3D reconstruction result using (g).

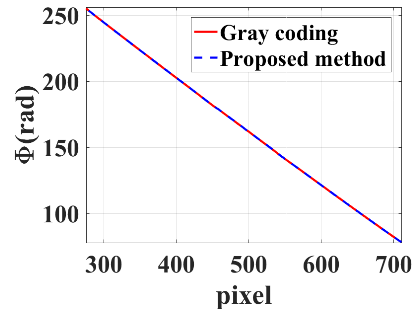
To demonstrate the correctness of our proposed algorithm, we compare our phase unwrapping result with the conventional gray coding phase unwrapping algorithm [47]. Seven additional gray coded patterns are projected onto the same object. Figure 6.6(g) shows the gray coding phase unwrapping result, and the 3D reconstruction result using the gray coding algorithm is shown in Fig. 6.6(h).

We take two cross-sections on the unwrapped phase map and the 3D geometry to compare our proposed method with the gray coding algorithm. One cross-section is at a near distance, and the other is at a far distance. Figure. 6.7(a) shows the comparison between our phase unwrapping result and gray coding result at a far cross-section, and Fig.6.7(b) shows the comparison between our phase unwrapping result and gray coding result at a near cross-section. The comparison between our method and gray coding on a far 3D cross-section is shown in Figs. 6.7(c), and the comparison between our method and gray coding on a near 3D cross-section is shown in 6.7(d). In all Figs. 6.7(a)-6.7(d), we can see that our results well overlap with the results using the gray coding algorithm. The depth difference on the two cross sections are shown in Fig.6.8(a) and Fig. 6.8(b). The comparison results verify the correctness of our proposed method and demonstrate the capability of our method to measure an object with a large depth variation. The depth range of our phase unwrapping method achieved at least 260mm, which is much larger than that of the virtual plane method ∼57mm.

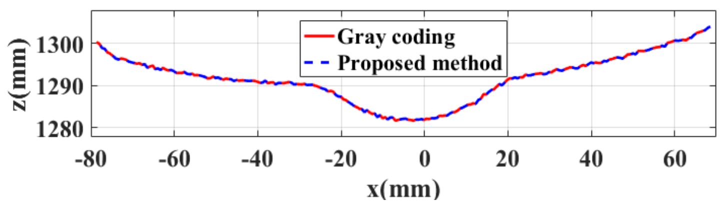
Since our algorithm can provide an absolute phase map, it can be used to measure multiple isolated objects simultaneously. To demonstrate the capability of simultaneous multiple object measurement, we lay a scene consisting of multiple objects. Those objects are put in different depths. Two objects are put at ∼1000mm away from the system, and the other two are put ∼1200mm away from the system. Figure 6.9(a) is a photo of the scene. We use Kinect to obtain a rough 3D measurement, and transform Kinect 3D data to the DFP coordinate system. The transformed Kinect 3D data are shown in Fig. 6.9(b). Based on the transformed Kinect 3D data and (6.19), we generate an artificial phase map Φ_{ref} shown in Fig. 6.9(c). As you can see in Fig. 6.9(c), there are hole and boundary artifacts due to the perspective difference and alignment difference explained in Sec. 6.2.3. The phase unwrapping result of our method using the artificial phase map Φ_{ref} is shown in Fig. 6.9(d). To obtain a



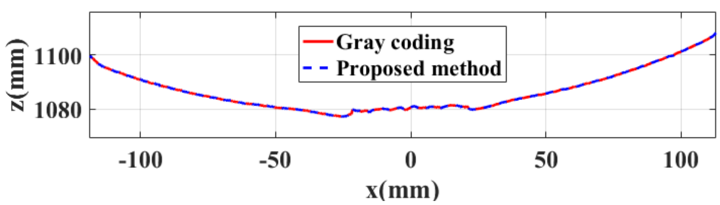
(a)



(b)



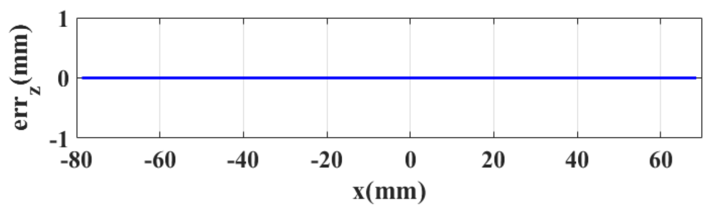
(c)



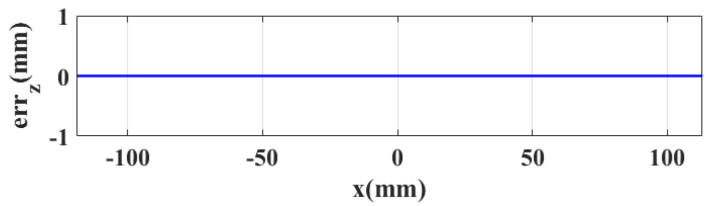
(d)

Figure 6.7. Comparison results on the phase unwrapping results and 3D measurement results between our proposed method and the gray coding algorithm. (a) Phase unwrapping result comparison on a far distance; (b) Phase unwrapping result comparison on a near distance; (c) 3D result comparison on a far distance; (d) 3D result comparison on a near distance.

complete unwrapped phase map, we again use interpolation and phase constraint to fill the holes in the unwrapped phase map of Fig. 6.9(d), and then use extrapolation and phase constraint to unwrap the missed boundary parts. The refined phase map is shown in Fig. 6.9(e). Based on the refined phase map, we do 3D reconstruction, and the final result is shown in Fig. 6.9(f).



(a)



(b)

Figure 6.8. Difference between our 3D reconstruction result and the gray coding result on the two cross sections. (a) 3D difference on a far cross section; (b) 3D difference on a near cross section.

Meanwhile, we adopted the gray coding algorithm to do phase unwrapping as a comparison. Figure 6.10(a) and Fig. 6.10(b) show the unwrapped phase map and the 3D reconstruction result using the gray coding algorithm.

The cross-section analysis results are shown in Fig. 6.11. We take two cross-sections on the unwrapped phase maps and the reconstructed geometries. One cross-section is at a near distance and the other one is at a far distance. The unwrapped phase map comparison on a far cross-section is shown in Fig. 6.11(a), and the unwrapped phase map comparison on a near cross-section is shown in Fig. 6.11(b). The geometry comparison results on a far and near cross-section are shown in Figs. 6.11(c)-6.11(d) respectively. From Figs. 6.11(a)-6.11(d), we can see that our results well overlap with the results using the gray coding algorithm, which demonstrates that our unwrapped phase is an absolute one, and demonstrates the capability of our algorithm to measure complex objects and multiple isolated objects. In this experiment, our

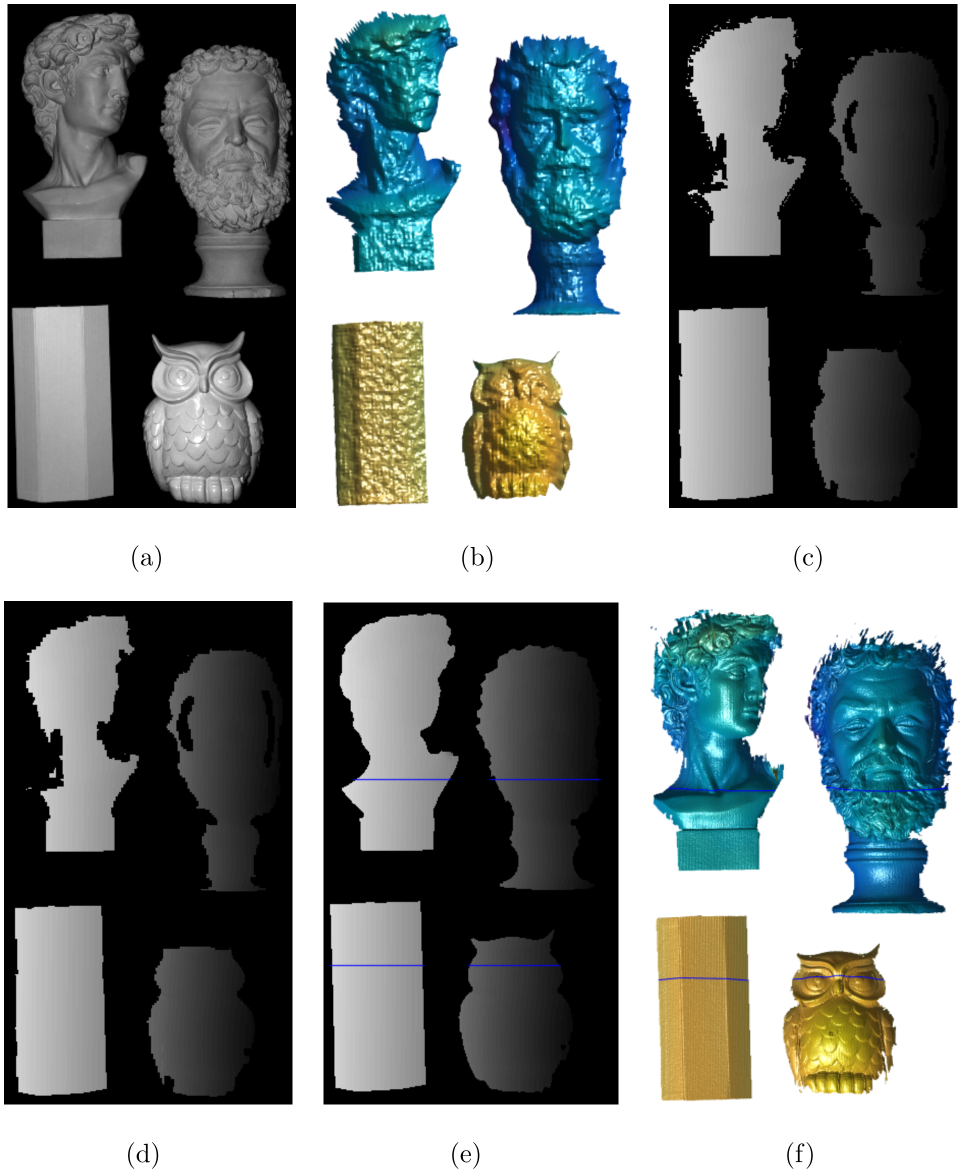
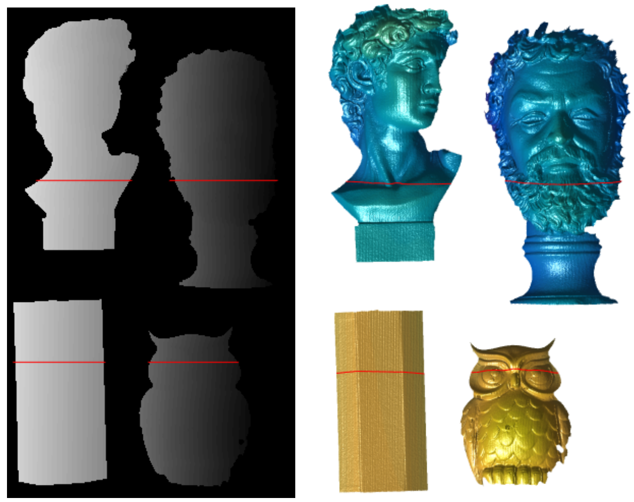


Figure 6.9. Experiment result on a complex scene. (a) Texture of the complex scene that is consisting of multiple sculptures at different depths; (b) Up-sampled Kinect 3D data in the DFP coordinate system; (c) Artificial phase map Φ_{ref} generated based on (b); (d) Phase unwrapped result assisted by (c); (e) Refined unwrapped phase map; (f) 3D geometry recovered from (e).

proposed phase unwrapping method successfully works for a depth range of ~200mm, which is much larger than that of the virtual plane method 57mm again.



(a) (b)

Figure 6.10. Experimental results using the gray coding algorithm. (a) Unwrapped phase map using the gray coding algorithm; (b) 3D reconstruction based on (a).

Though we use Kinect to assist our phase unwrapping, the speed of Kinect is not necessarily an upper bound of the DFP system's speed in our proposed method. In our multi-resolution system, the speed of the camera and the projector is set to be 160Hz which is the camera's fastest speed at the resolution of 1600×1000. The camera and projector are synchronized by a timing circuit. Since we only need three equally phase shifted fringe images for absolute 3D shape measurement, the DFP system's theoretical maximum 3D frame rate is $160/3 \approx 53$Hz. We use software trigger to synchronize the DFP system and Kinect. Though the maximum speed of Kinect is only 30Hz, we propose to do interpolation on a time series of Kinect 3D frames to assist phase unwrapping for the higher speed DFP system. Suppose the Kinect 3D frames are $X(t_1), X(t_2), \ldots, X(t_n)$ where t_i are the time indices of those Kinect 3D frames, we can do time interpolation to estimate the intermediate 3D between two consecutive Kinect 3D frames, such as $X(t_s) = \frac{t_{i+1}-t_s}{t_{i+1}-t_i}X(t_i) + \frac{t_s-t_i}{t_{i+1}-t_i}X(t_{i+1})$ where $t_s \in (t_i, t_{i+1})$ assuming a constant speed of each 3D point between two consecutive 3D

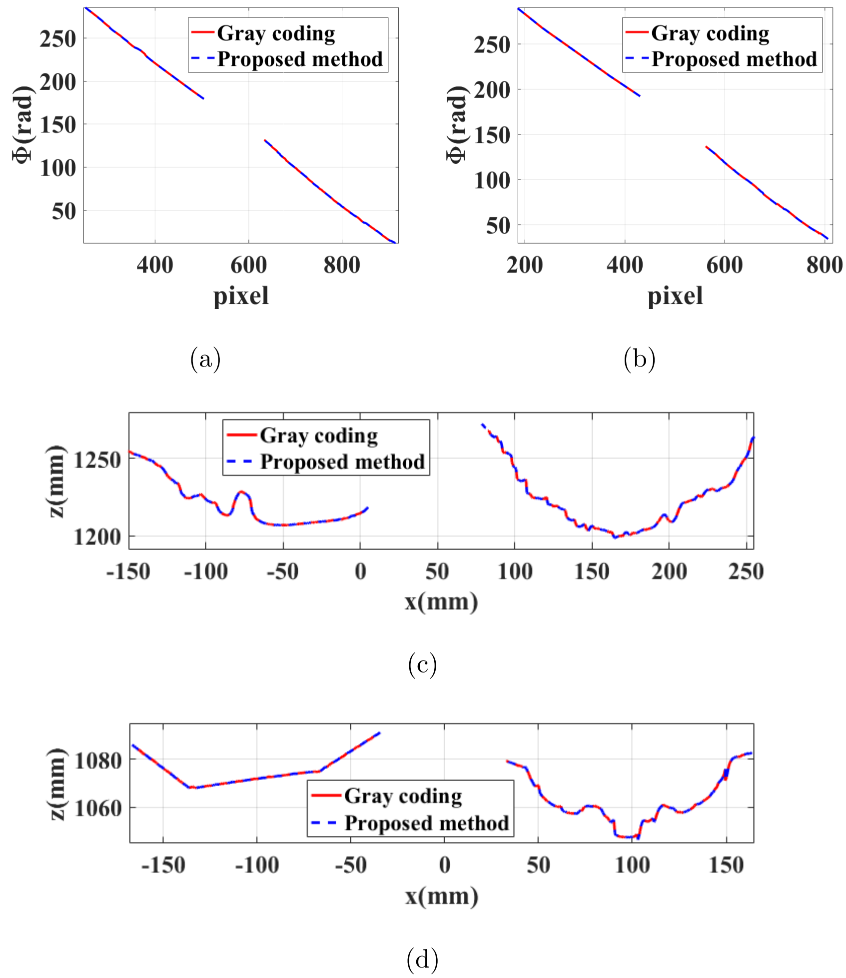


Figure 6.11. Comparison results on the phase unwrapping results and 3D measurement results between our proposed method and the gray coding algorithm. (a) Phase unwrapping result comparison on a far distance; (b) Phase unwrapping result comparison on a near distance; (c) 3D result comparison on a far distance; (d) 3D result comparison on a near distance.

frames. Therefore, we can obtain more reference 3D frames to do phase unwrapping for a higher speed DFP system.

We do the following dynamic experiments to verify our time interpolation strategy. In our experiments, we would like to achieve the DFP system's theoretical maximum speed 53Hz which is faster than the Kinect's maximum speed 30Hz. To achieve that, we will capture two sets of three equally phase shifted patterns using the DPF system and one 3D frame from Kinect for reference each time, such that a 26.5Hz of Kinect speed is enough for the DFP to achieve 53Hz. Suppose the DFP system captures fringe images at $\{t_1, t_2, t_3, t_4, t_5, \ldots\}$ and Kinect captures 3D frames at $\{t_1, t_3, t_5, \ldots\}$, we average two consecutive Kinect 3D frames to obtain a reference for intermediate times $\{t_2, t_4, \ldots\}$. For instance, we average two Kinect 3D frames at t_1, t_3 to obtain a Kinect reference 3D frame for t_2. Based on the above idea, we measured two isolated objects that are in motion. Again, the two objects are put in different depths. They move in a depth range from 1200mm to 1550mm. Figure 6.12(a) shows one example frame of the texture captured by the camera (associated with Visualization 6.1). Figure 6.12(b) shows the corresponding 3D data captured by Kinect (associated with Visualization 6.2). The Kinect 3D data are transformed to the DFP coordinate system for better comparison with the DFP 3D data. Figure 6.12(c) shows our 3D result (associated with Visualization 6.3), and Fig. 6.12(d) is another perspective of our shape measurement result (associated with Visualization 6.4). Our result is in ∼53Hz and Kinect 3D is in ∼27Hz. This dynamic experiment demonstrates the success of our proposed method to measure dynamic objects moving in a large depth range of at least 350mm (1200mm-1550mm), which is much larger than that of the virtual plane method 57mm. Also, this experiment demonstrates that the maximum speed of the additional 3D scanner is not necessarily an upper bound of the speed in our method.

A dynamic human body is interested in many applications. Therefore, we also measure a dynamic human body using our system. The body moves in a depth range from 1150mm to 1550mm. Figure 6.13(a) shows one example frame of the texture captured by the camera (associated with Visualization 6.5). Figure 6.13(b) shows the corresponding 3D data captured by Kinect (associated with Visualization 6.6). The Kinect 3D data in Fig. 6.13(b) are transformed to the DFP coordinate system for better comparison with the DPF 3D data. Figure 6.13(c) shows our 3D measurement result (associated with Visualization 6.7), and Fig. 6.13(d) is another perspective of our shape measurement result (associated with Visualization 6.8). In

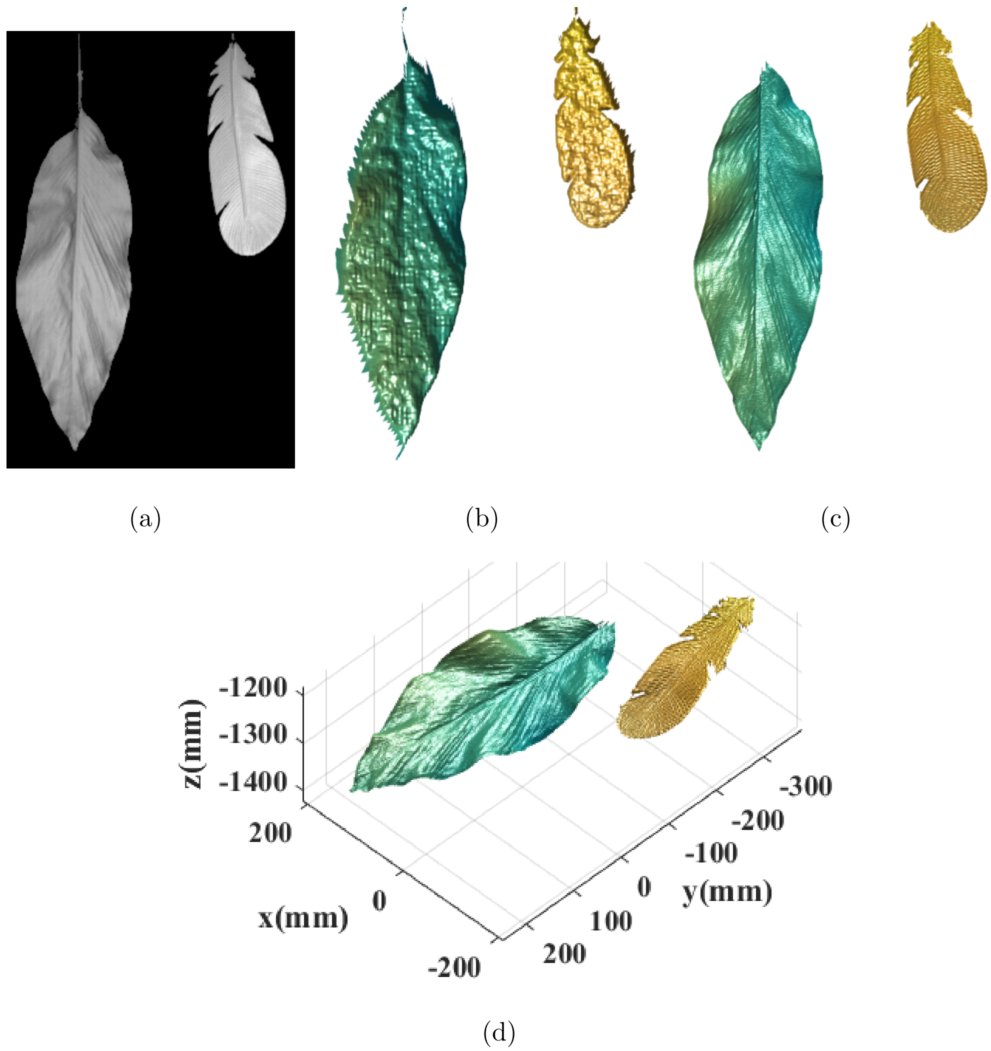
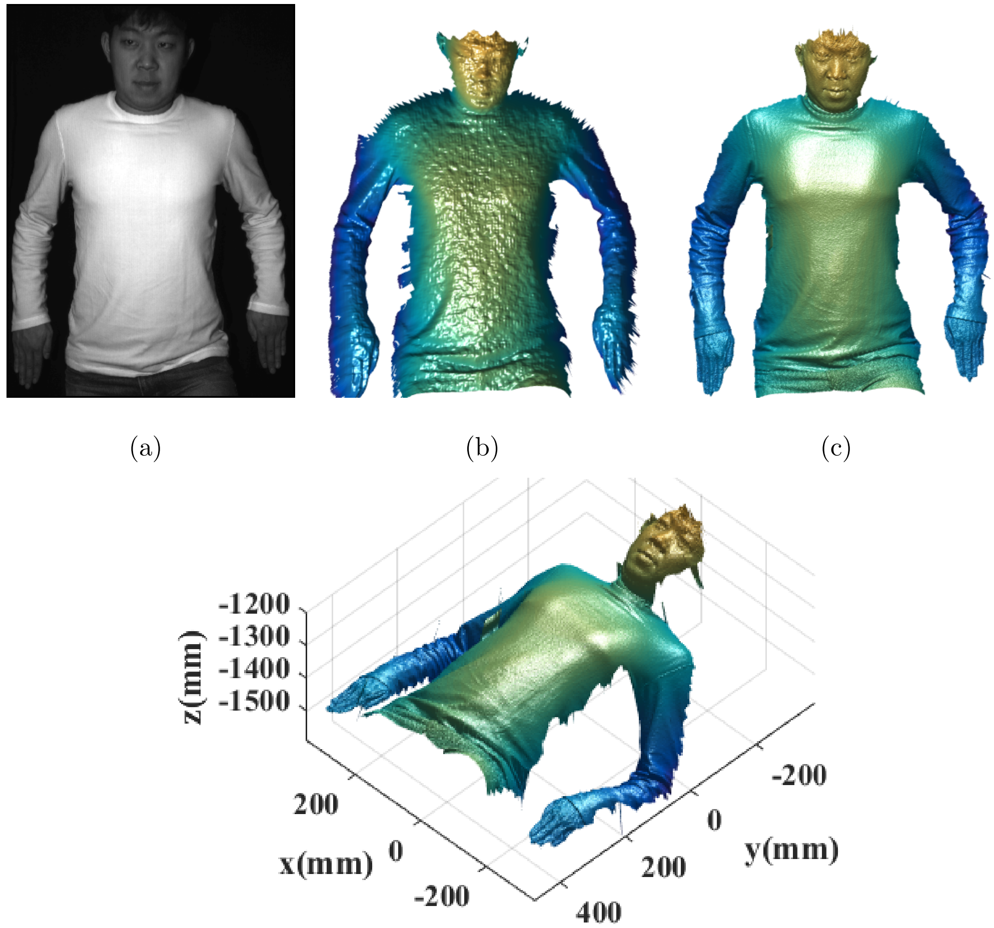


Figure 6.12. Dynamic experiment on two isolated objects. (a) Texture of the two isolated objects (associated with Visualization 6.1); (b) Kinect 3D in the DFP coordinate system (associated with Visualization 6.2); (c) 3D shape measurement result using our proposed method (associated with Visualization 6.3); (d) Another perspective of our reconstructed 3D (associated with Visualization 6.4).

this experiment, we further extend the measurement depth range to 400mm using our proposed method, and once again, our dynamic experiment result is 53Hz, which is faster than the maximum speed of Kinect V2 30Hz.



(a) (b) (c)

(d)

Figure 6.13. Dynamic experiment on a human body. (a) Texture of the human body (associated with Visualization 6.5); (b) Kinect 3D in the DFP coordinate system (associated with Visualization 6.6); (c) 3D shape measurement result using our proposed method (associated with Visualization 6.7); (d) Another perspective of our reconstructed 3D (associated with Visualization 6.8).

6.4 Summary

This chapter presented a novel method for absolute phase unwrapping that needs no additional patterns. Specifically, we leveraged a low-cost 3D scanner to capture rough 3D of the scene. The rough 3D data can be transformed to the structured light coordinate system to generate an artificial reference phase map Φ_{ref}. By referring to Φ_{ref}, we did absolute phase unwrapping directly without projecting any additional patterns. We developed a multi-resolution system that consists of a DFP system and Kinect V2 to validate our algorithm. Experimental results showed the success of our propose method that uses only three phase shifted fringe patterns to do absolute shape measurement of objects, and demonstrated the capability of our method to measure objects with large depth variations and dynamic objects moving in a large depth range. Besides, we showed that the speed of the additional 3D scanner is not necessarily a speed limit of our proposed method. Though the maximum speed of Kinect is only 30Hz, our DFP system achieved 53Hz with a resolution 1600×1000 pixels to measure dynamic objects moving in a large depth range.

7. SUMMARY AND FUTURE PROSPECTS

7.1 Summary of contributions

For robots to have a more complete perception of the scene, we advanced the capability of the digital fringe projection technique in measurement scale (space domain), speed (time domain), and fusion with other modality information. Specifically, the contributions of this dissertation are as follows.

- **Developed a flexible method for large-scale structured light system calibration.** Conventional structured light system calibration often requires the usage of a calibration target with a similar size as the field of view (FOV). This brings challenges to large-scale structured light system calibration because a large calibration target is difficult and expensive to fabricate and inconvenient to use for pose adjustment. We have developed a new large-scale system calibration method that does not need a large calibration target, and the new method is more flexible and more convenient. Our proposed method uses a regular-sized calibration board to perform intrinsic calibration at a near range, then calibrates the extrinsic parameters between camera and projector with the assistance of a low-accuracy, large-scale 3D sensor (e.g., Microsoft Kinect). We applied our proposed method to calibrate a large-scale 3D shape measurement system with a FOV of $(1120 \times 1900 \times 1000)mm^3$, and the large-scale system achieved measurement accuracy as high as 0.07mm with a standard deviation of 0.80mm by measuring a 304.8mm-diameter sphere. As a comparison, Kinect V2 only achieved mean error of 0.80mm with a standard deviation of 3.41mm for the FOV of measurement. We have published this research work in the journal of *Applied Optics*, and the details of this research work were described in Chapter 3.

- **Developed a high-resolution, real-time simultaneous 3D surface geometry and temperature measurement method.** We prototyped a high-resolution, real-time simultaneous 3D geometric shape and temperature measurement system. The developed system can fuse the 3D geometry and temperature information together for potential quantitative analysis in medical robot surgery applications. The contributions in this research include a holistic approach to calibrating both the structured light system and the thermal camera under exactly the same world coordinate system even though these two sensors do not share the same wavelength, and a computational framework to determine the sub-pixel corresponding temperature for each 3D point as well as to discard those occluded points. Since the thermal 2D imaging and 3D visible imaging systems do not share the same spectrum of light, they can perform sensing simultaneously in real time. Therefore, we developed a hardware system that can achieve real-time 3D geometry and temperature measurement at 26Hz with 768×960 points per frame. This research was published in the journal of *Optics Express*, and details were described in Chapter 4.

- **Developed a pixel-wise absolute phase unwrapping method using geometric constraints of structured light system.** To achieve an absolute phase map, additional fringe patterns are usually needed, such as in multiple-wavelength phase unwrapping methods and gray coding phase unwrapping methods. However, projecting additional fringe patterns slows down the measurement speed. In this dissertation, we developed a novel method to unwrap a phase pixel by pixel solely using geometric constraints of the structured light system without requiring additional image acquisition. Specifically, an artificial absolute phase map Φ_{min}, at a given virtual depth plane $z = z_{min}$, is created from geometric constraints of the calibrated structured light system; the wrapped phase is pixel-by-pixel unwrapped by referring to Φ_{min}. Our proposed method has advantages for high-speed 3D shape measurement, parallel data processing, simultaneous multiple object measurement, fringe order determination and simple system setup. This work has been published in the journal of *Optics Express*, and its details were introduced in Chapter 5.

- **Developed a pixel-wise absolute phase unwrapping method assisted by an additional 3D scanner.** In a multi-resolution 3D measurement system, those multiple sensors could mutually benefit each other. Inspired by that idea, we explored a novel absolute phase unwrapping method assisted by a low-cost three-dimensional (3D) scanner. The proposed absolute phase unwrapping method leverages a low-cost 3D scanner to capture rough 3D data of the scene, and transforms the rough 3D data to the world coordinate system to generate an artificial reference phase map Φ_{ref}. By referring to Φ_{ref}, we can do absolute phase unwrapping directly without projecting any additional patterns, such that the DFP system can achieve higher measurement speed. Since the captured 3D data are close to the measured object, this proposed method can also work for a large depth range, which expands its possibility to measure objects with large depth variations and objects moving in a large depth range. We further proved that the speed of the maximum speed of the low-cost 3D scanner is not necessarily the upper bound of our DFP measurement speed. Finally, we prototyped a multi-resolution system consisting a DFP system and a Kinect V2, and the DFP system achieved 53Hz with a resolution 1600×1000 pixels when we measured dynamic objects that were moving in a large depth range of 400mm. We have published this research work in the journal of *Applied Optics*, and its details were described in Chapter 6.

7.2 Future prospects

In this dissertation research, we have developed multi-scale, multi-modal, high-speed 3D shape measurement methods. Based on the achievements in this dissertation, some works could be further studied in the future. In this section, we will briefly introduce several examples in which our technologies could be applied in the future to show the potential of our developed multi-scale, multi-modal, high-speed 3D shape measurement.

7.2.1 Full field 360° 3D measurement

Full field 360° 3D measurement could be further studied based on this dissertation. Currently, the structured light systems using digital fringe projection technologies are mainly focused on the surface 3D reconstruction from one perspective. In the future, through using multiple structure light systems and merging 3D results from different systems, full field 360° 3D measurement could become possible.

The full field 360° 3D measurement through structured light technologies could be applied in the entertainment filed. Though some commercial sensors, such as Kinect and RealSense, have been applied in the entertainment field, their accuracy and speed are seriously compromised, making those commercial sensors difficult to be applied in highly accurate and dynamic 3D measurements. For instance, in many interactive video sports games (e.g., somatosensory tennis, volleyball and ping-pong) and video dancing games, it is challenging for commercial 3D sensors to accurately measure the position, gesture or speed of the player's motion. The inaccurate measurement could lead to a wrong evaluation of the player's performance, or even mislead players in those games.

The technologies developed in this dissertation could make it possible to extend structured light technologies for full field 360° 3D measurement in the entertainment field. First, our proposed large-scale structured light calibration method extends the application of structured light technologies to a large scale, such that the player can move naturally and freely without space restriction. Second, our proposed phase unwrapping methods can help to achieve higher speed measurement of the dynamic scene. Thirdly, our multi-modal project, introduced in Chapter 4, can help us to integrate a thermal camera to monitor the player's heartbeat and respiration, which can be used to detect the player's health status and fatigue strength. With a temperature map registered with a 3D model, the heartbeat and respiration information could be more accurate than with previous image-based methods, since the player's distance and head pose can be taken into consideration. Meanwhile, the better-quality 3D data based on our technologies could make pose estimation and gesture detection easier and more accurate. Computers could achieve a better analysis of the player's gestures and make a better evaluation of the player's performance. Therefore, video sports games could become more realistic and more scientific.

Taken one step further, beyond interactive sports games, our technologies could even help in some sports training programs. Using full field 360° 3D measurement technologies, we can obtain a more accurate measurement of the player's biometrics (e.g., height, mass). With a better estimation of the player's biometrics, it will be possible for computers to design a more scientifically customized training process, to track the player's training progress, and to correct the sports player's pose, gesture, and speed along the training process.

7.2.2 Enhanced low-cost 3D measurement

Currently structured light systems using digital fringe projection techniques are still quite expensive, limiting their popularity for the mass consumer. In the future, using low-cost devices but maintain similar quality 3D data through enhancement algorithms could be further studied, such that the cost of high quality structured light systems could be decreased.

The enhanced low-cost 3D measurement could be applied in 3D teleconference. It is predicted that offices will become obsolete and virtual workspaces will become reality[1] with the prospering of AR/VR technologies. Particularly, 3D teleconferencing is becoming attractive for collaboration among geographically distributed teams by avoiding travel and increasing flexibility. Also, it allows users to share the same 3D world and offers realistic and natural interaction experiences for users, such as

[1] https://venturebeat.com/2018/06/02/10-ambitious-predictions-for-how-vr-ar-will-shape-our-world/

maintaining appropriate eye contact with multiple speakers and reading each other's body languages, including subtle movement and gaze direction.

The technologies developed in this dissertation could potentially benefit the enhanced low cost 3D measurement for the teleconference field in the following aspects. First, the technologies enable us to capture large-scale, high-speed, and high-accuracy 3D data to make the 3D teleconference more realistic and more natural. The large-scale calibration method allows whole human bodies to be captured instead of only heads or upper bodies. The high-speed and high-accuracy data can better deliver subtle movements to other users. Second, our system may have an advantage in following data processing algorithms such as multi-view integration, texture mapping, and object segmentation. As a comparison, Kinect data still suffers from depth-color image registration, which can confuse and distract teleconference users. With better post-processing results, our 3D data can also be easily used for various virtual reality, augmented reality, and mixed reality tasks. Third, our technologies allow us to integrate with other modality information for additional functions, such as temperature information in telemedicine.

REFERENCES

REFERENCES

[1] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-flight cameras in computer graphics. In *Computer Graphics Forum*, volume 29, pages 141–159. Wiley Online Library, 2010.

[2] F. Chiabrando, R. Chiabrando, D. Piatti, and F. Rinaudo. Sensors for 3d imaging: Metric evaluation and calibration of a ccd/cmos time-of-flight camera. *Sensors*, 9(12):10080–10096, 2009.

[3] C. P. Keferstein and M. Marxer. Testing bench for laser triangulation sensors. *Sensor Review*, 18(3):183–187, 1998.

[4] B. D. Lucas, T. Kanade, et al. An iterative image registration technique with an application to stereo vision. 1981.

[5] U. R. Dhond and J. K. Aggarwal. Structure from stereo-a review. *IEEE transactions on systems, man, and cybernetics*, 19(6):1489–1510, 1989.

[6] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado. A state of the art in structured light patterns for surface profilometry. *Pattern recognition*, 43(8):2666–2680, 2010.

[7] I. Sa and P. Corke. Vertical infrastructure inspection using a quadcopter and shared autonomy control. In *Field and Service Robotics*, pages 219–232. Springer, 2014.

[8] W. Wang, W. Zhao, L. Huang, V. Vimarlund, and Z. Wang. Applications of terrestrial laser scanning for tunnels: a review. *Journal of Traffic and Transportation Engineering (English Edition)*, 1(5):325–337, 2014.

[9] Y. Chin, T. Ogitsu, and H. Mizoguchi. Road surface mapping using kinect sensor for road maintenance. *Journal of the Institute of Industrial Applications Engineers*, 4(4):153–158, 2016.

[10] L. Yang, L. Zhang, H. Dong, A. Alelaiwi, and A. El Saddik. Evaluating and improving the depth accuracy of kinect for windows v2. *IEEE Sensors Journal*, 15(8):4275–4285, 2015.

[11] F. H. A. (FHWA) and F. T. Admistration. Highway and rail transit tunnel inspection manual. 2005.

[12] K. S. H. T. M. Helezon and M. Tekniği. Modified twist drill technique in the management of chronic subdural hematoma. *Turkish neurosurgery*, 23(1):50–54, 2013.

[13] G. Pearce, C. Bainbridge, J. Patrick, M. Lenz, G. Jones, et al. An investigation into thermal necrosis of bone associated with surgical procedures. *WIT Transactions on Biomedicine and Health*, 8, 2005.

[14] A. Zeng, K.-T. Yu, S. Song, D. Suo, E. Walker, A. Rodriguez, and J. Xiao. Multi-view self-supervised deep learning for 6d pose estimation in the amazon picking challenge. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pages 1386–1383. IEEE, 2017.

[15] S. Kumar, A. Majumder, S. Dutta, R. Raja, S. Jotawar, A. Kumar, M. Soni, V. Raju, O. Kundu, E. H. L. Behera, et al. Design and development of an automated robotic pick & stow system for an e-commerce warehouse. *arXiv preprint arXiv:1703.02340*, 2017.

[16] M. Subbarao and G. Surya. Depth from defocus: a spatial domain approach. *International Journal of Computer Vision*, 13(3):271–294, 1994.

[17] N. Lazaros, G. C. Sirakoulis, and A. Gasteratos. Review of stereo vision algorithms: from software to hardware. *International Journal of Optomechatronics*, 2(4):435–462, 2008.

[18] B. Tippetts, D. J. Lee, K. Lillywhite, and J. Archibald. Review of stereo vision algorithms and their suitability for resource-limited systems. *Journal of Real-Time Image Processing*, 11(1):5–25, 2016.

[19] M. Menze and A. Geiger. Object scene flow for autonomous vehicles. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[20] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *German Conference on Pattern Recognition*, pages 31–42. Springer, 2014.

[21] W. Luo, A. G. Schwing, and R. Urtasun. Efficient deep learning for stereo matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5695–5703, 2016.

[22] A. A. Shetty, C. G. Nayak, and V. George. Stereo vision for mobile robots: A review. *International Journal of Engineering & Technology*, 2(11):25–29, 2015.

[23] L. Nalpantidis and A. Gasteratos. Stereo vision for robotic applications in the presence of non-ideal lighting conditions. *Image and Vision Computing*, 28(6):940–951, 2010.

[24] S. Sengupta, E. Greveson, A. Shahrokni, and P. H. Torr. Urban 3d semantic modelling using stereo vision. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 580–585. IEEE, 2013.

[25] N. Uchida, T. Shibahara, T. Aoki, H. Nakajima, and K. Kobayashi. 3d face recognition using passive stereo vision. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 2, pages II–950. IEEE, 2005.

[26] B. Li. Superfast 3d shape measurement with application to flapping wing mechanics analysis. 2017.

[27] S. Zhang. *High-Speed 3D imaging with digital fringe projection techniques*, volume 5. CRC Press, 2016.

[28] E. Lachat, H. Macher, M. Mittet, T. Landes, and P. Grussenmeyer. First experiences with kinect v2 sensor for close range 3d modelling. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(5):93, 2015.

[29] C. Ye and M. Bruch. A visual odometry method based on the swissranger sr4000. Technical report, ARKANSAS UNIV AT LITTLE ROCK, 2010.

[30] H. Cook, Q. V. Nguyen, S. Simoff, T. Trescak, and D. Preston. A close-range gesture interaction with kinect. In *Big Data Visual Analytics (BDVA), 2015*, pages 1–8. IEEE, 2015.

[31] Q. Liu, S. Yu, Y. Wang, H. Le, and Y. Yuan. A hand-waving dance teaching system based on kinect. In *International Conference on Blended Learning*, pages 354–365. Springer, 2017.

[32] A. G. Kashani, M. J. Olsen, C. E. Parrish, and N. Wilson. A review of lidar radiometric processing: From ad hoc intensity correction to rigorous radiometric calibration. *Sensors*, 15(11):28099–28128, 2015.

[33] D. Watzenig and M. Horn. Introduction to automated driving. In *Automated Driving*, pages 3–16. Springer, 2017.

[34] G. Hemann, S. Singh, and M. Kaess. Long-range gps-denied aerial inertial navigation with lidar localization. In *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, pages 1659–1666. IEEE, 2016.

[35] B. Suger, B. Steder, and W. Burgard. Traversability analysis for mobile robots in outdoor environments: A semi-supervised learning approach based on 3d-lidar data. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pages 3941–3946. IEEE, 2015.

[36] J. G. D. França, M. A. Gazziro, A. N. Ide, and J. H. Saito. A 3d scanning system based on laser triangulation and variable field of view. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 1, pages I–425. IEEE, 2005.

[37] F.-J. Shiou and M.-X. Liu. Development of a novel scattered triangulation laser probe with six linear charge-coupled devices (ccds). *Optics and Lasers in Engineering*, 47(1):7–18, 2009.

[38] J.-H. Wu, R.-S. Chang, and J.-A. Jiang. A novel pulse measurement system by using laser triangulation and a cmos image sensor. *Sensors*, 7(12):3366–3385, 2007.

[39] J. G. D. Franca, M. A. Gazziro, A. N. Ide, and J. H. Saito. A 3d scanning system based on laser triangulation and variable field of view. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 1, pages I–425. IEEE, 2005.

[40] M. F. M. Costa. Surface inspection by an optical triangulation method. *Optical Engineering*, 35(9):2743–2748, 1996.

[41] S. Zhang. Recent progresses on real-time 3d shape measurement using digital fringe projection techniques. *Optics and lasers in engineering*, 48(2):149–158, 2010.

[42] M. Takeda and K. Mutoh. Fourier transform profilometry for the automatic measurement of 3-d object shapes. *Applied optics*, 22(24):3977–3982, 1983.

[43] X. Su and Q. Zhang. Dynamic 3-d shape measurement method: a review. *Optics and Lasers in Engineering*, 48(2):191–204, 2010.

[44] D. Malacara. *Optical shop testing*, volume 59. John Wiley & Sons, 2007.

[45] D. C. Ghiglia and M. D. Pritt. *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software*. John Wiley and Sons, Inc, New York, NY, 1998.

[46] H. O. Saldner and J. M. Huntley. Temporal phase unwrapping: application to surface profiling of discontinuous objects. *Applied optics*, 36(13):2770–2775, 1997.

[47] G. Sansoni, M. Carocci, and R. Rodella. Three-dimensional vision based on a combination of gray-code and phase-shift light projection: analysis and compensation of the systematic errors. *Applied Optics*, 38(31):6565–6573, 1999.

[48] Q. Zhang, X. Su, L. Xiang, and X. Sun. 3-d shape measurement based on complementary gray-code light. *Optics and Lasers in Engineering*, 50(4):574–579, 2012.

[49] Y. Li, C. Zhao, Y. Qian, H. Wang, and H. Jin. High-speed and dense three-dimensional surface acquisition using defocused binary patterns for spatially isolated objects. *Optics Express*, 18(21):21628–21635, 2010.

[50] S. Zhang. Composite phase-shifting algorithm for absolute phase measurement. *Optics and Lasers in Engineering*, 50(11):1538–1541, 2012.

[51] C. Zuo, Q. Chen, G. Gu, S. Feng, and F. Feng. High-speed three-dimensional profilometry for multiple objects with complex shapes. *Optics express*, 20(17):19493–19510, 2012.

[52] Y. An and S. Zhang. Three-dimensional absolute shape measurement by combining binary statistical pattern matching with phase-shifting methods. *Applied optics*, 56(19):5418–5426, 2017.

[53] R. Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 2(1):1–104, 2006.

[54] A. Bab-Hadiashar and D. Suter. Robust total least squares based optic flow computation. In *Asian Conference on Computer Vision*, pages 566–573. Springer, 1998.

[55] I. Zoghlami, O. Faugeras, and R. Deriche. Using geometric corners to build a 2d mosaic from a set of images. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 420–425. IEEE, 1997.

[56] H.-Y. Shum and R. Szeliski. Construction of panoramic image mosaics with global and local alignment. In *Panoramic vision*, pages 227–268. Springer, 2001.

[57] G. Sansoni, M. Trebeschi, and F. Docchio. State-of-the-art and applications of 3d imaging sensors in industry, cultural heritage, medicine, and criminal investigation. *Sensors*, 9(1):568–601, 2009.

[58] G. Sansoni and F. Docchio. Three-dimensional optical measurements and reverse engineering for automotive applications. *Robotics and Computer-Integrated Manufacturing*, 20(5):359–367, 2004.

[59] B. Micusik and J. Kosecka. Piecewise planar city 3d modeling from street view panoramic sequences. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2906–2912. IEEE, 2009.

[60] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski. Building rome in a day. *Communications of the ACM*, 54(10):105–112, 2011.

[61] G. Metrology. Atos scanbox series 8, 2016.

[62] C. Hughes, M. Glavin, E. Jones, and P. Denny. Wide-angle camera technology for automotive applications: a review. *IET Intelligent Transport Systems*, 3(1):19–31, 2009.

[63] M. Simonovsky, B. Gutiérrez-Becker, D. Mateus, N. Navab, and N. Komodakis. A deep metric for multimodal registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 10–18. Springer, 2016.

[64] A. P. James and B. V. Dasarathy. Medical image fusion: A survey of the state of the art. *Information Fusion*, 19:4–19, 2014.

[65] Y. Huang, X. Qian, and S. Chen. Multi-sensor calibration through iterative registration and fusion. *Computer-Aided Design*, 41(4):240–255, 2009.

[66] S. Wasielewski and O. Strauss. Calibration of a multi-sensor system laser rangefinder/camera. In *Intelligent Vehicles' 95 Symposium., Proceedings of the*, pages 472–477. IEEE, 1995.

[67] V. Fremont, P. Bonnifait, et al. Extrinsic calibration between a multi-layer lidar and a camera. In *Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on*, pages 214–219. IEEE, 2008.

[68] C. Zhang and Z. Zhang. Calibration between depth and color sensors for commodity depth cameras. In *Computer vision and machine learning with RGB-D sensors*, pages 47–64. Springer, 2014.

[69] S. Vidas, R. Lakemond, S. Denman, C. Fookes, S. Sridharan, and T. Wark. A mask-based approach for the geometric calibration of thermal-infrared cameras. *IEEE Transactions on Instrumentation and Measurement*, 61(6):1625–1635, 2012.

[70] F. B. Djupkep Dizeu, X. Maldague, and A. Bendada. Mapping of the indoor conditions by infrared thermography. *Journal of Imaging*, 2(2):10, 2016.

[71] K. Skala, T. Lipić, I. Sović, L. Gjenero, and I. Grubišić. 4d thermal imaging system for medical applications. *Periodicum biologorum*, 113(4):407–416, 2011.

[72] Y. Chen and W. Warren. 3d fusion of infrared images with dense rgb reconstruction from multiple views-with application to fire-fighting robots. *Recuperado a partir de https://pdfs. semanticscholar. org/9702/31551cbf43023bb77d0a7c233b5311c31997. pdf*, 2013.

[73] M. Gschwandtner, R. Kwitt, A. Uhl, and W. Pree. Infrared camera calibration for dense depth map construction. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 857–862. IEEE, 2011.

[74] D. Borrmann, A. Nüchter, M. DHakulović, I. Maurović, I. Petrović, D. Osmanković, and J. Velagić. A mobile robot based system for fully automated thermal 3d mapping. *Advanced Engineering Informatics*, 28(4):425–440, 2014.

[75] G. Cardone, A. Ianiro, G. Dello Ioio, and A. Passaro. Temperature maps measurements on 3d surfaces with infrared thermography. *Experiments in fluids*, 52(2):375–385, 2012.

[76] A. Geiger, M. Roser, and R. Urtasun. Efficient large-scale stereo matching. In *Asian conference on computer vision*, pages 25–38. Springer, 2010.

[77] S. Lei and S. Zhang. Flexible 3-d shape measurement using projector defocusing. *Optics letters*, 34(20):3080–3082, 2009.

[78] Y. An, T. Bell, B. Li, J. Xu, and S. Zhang. Method for large-range structured light system calibration. *Applied optics*, 55(33):9563–9572, 2016.

[79] C. B. Duane. Close-range camera calibration. *Photogramm. Eng*, 37(8):855–866, 1971.

[80] I. Sobel. On calibrating computer controlled cameras for perceiving 3-d scenes. *Artificial Intelligence*, 5(2):185–198, 1974.

[81] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal on Robotics and Automation*, 3(4):323–344, 1987.

[82] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11):1330–1334, 2000.

[83] J.-M. Lavest, M. Viala, and M. Dhome. Do we really need an accurate calibration pattern to achieve a reliable camera calibration? In *European Conference on Computer Vision*, pages 158–174. Springer, 1998.

[84] A. Albarelli, E. Rodolà, and A. Torsello. Robust camera calibration using inaccurate targets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31:376–383, 2009.

[85] K. H. Strobl and G. Hirzinger. More accurate pinhole camera calibration with imperfect planar target. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 1068–1075. IEEE, 2011.

[86] L. Huang, Q. Zhang, and A. Asundi. Flexible camera calibration using not-measured imperfect target. *Applied optics*, 52(25):6278–6286, 2013.

[87] C. Schmalz, F. Forster, and E. Angelopoulou. Camera calibration: active versus passive targets. *Optical Engineering*, 50(11):113601–113601, 2011.

[88] L. Huang, Q. Zhang, and A. Asundi. Camera calibration with active phase target: improvement on feature detection and optimization. *Optics letters*, 38(9):1446–1448, 2013.

[89] Y. Wen, S. Li, H. Cheng, X. Su, and Q. Zhang. Universal calculation formula and calibration method in fourier transform profilometry. *Appl. Opt.*, 49(34):6563–6569, 2010.

[90] Y. Xiao, Y. Cao, and Y. Wu. Improved algorithm for phase-to-height mapping in phase measuring profilometry. *Appl. Opt.*, 51(8):1149–1155, 2012.

[91] 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(2):256–261, 2012.

[92] Q. Hu, P. S. Huang, Q. Fu, and F.-P. Chiang. Calibration of a three-dimensional shape measurement system. *Optical Engineering*, 42(2):487–493, 2003.

[93] X. Mao, W. Chen, and X. Su. Improved fourier-transform profilometry. *Applied optics*, 46(5):664–668, 2007.

[94] E. Zappa et al. Fourier-transform profilometry calibration based on an exhaustive geometric model of the system. *Optics and Lasers in Engineering*, 47(7):754–767, 2009.

[95] H. Guo, M. Chen, and P. Zheng. Least-squares fitting of carrier phase distribution by using a rational function in fringe projection profilometry. *Optics letters*, 31(24):3588–3590, 2006.

[96] H. Du and Z. Wang. Three-dimensional shape measurement with an arbitrarily arranged fringe projection profilometry system. *Optics letters*, 32(16):2438–2440, 2007.

[97] L. Huang, P. S. Chua, and A. Asundi. Least-squares calibration method for fringe projection profilometry considering camera lens distortion. *Applied optics*, 49(9):1539–1548, 2010.

[98] M. Vo, Z. Wang, B. Pan, and T. Pan. Hyper-accurate flexible calibration technique for fringe-projection-based three-dimensional imaging. *Optics Express*, 20(15):16926–16941, 2012.

[99] R. Legarda-Sa, T. Bothe, W. P. Ju, et al. Accurate procedure for the calibration of a structured light system. *Optical Engineering*, 43(2):464–471, 2004.

[100] S. Zhang and P. S. Huang. Novel method for structured light system calibration. *Optical Engineering*, 45(8):083601, 2006.

[101] Z. Li, Y. Shi, C. Wang, and Y. Wang. Accurate calibration method for a structured light system. *Optical Engineering*, 47(5):053604–053604, 2008.

[102] Y. Yin, X. Peng, A. Li, X. Liu, and B. Z. Gao. Calibration of fringe projection profilometry with bundle adjustment strategy. *Optics Letters*, 37(4):542–544, 2012.

[103] D. Han, A. Chimienti, and G. Menga. Improving calibration accuracy of structured light systems using plane-based residual error compensation. *Optical Engineering*, 52(10):104106–104106, 2013.

[104] T. Bell, J. Xu, and S. Zhang. Method for out-of-focus camera calibration. *Applied optics*, 55(9):2346–2352, 2016.

[105] B. Li, N. Karpinsky, and S. Zhang. Novel calibration method for structured-light system with an out-of-focus projector. *Applied optics*, 53(16):3415–3426, 2014.

[106] Y. An and S. Zhang. High-resolution, real-time simultaneous 3d surface geometry and temperature measurement. *Optics express*, 24(13):14552–14563, 2016.

[107] S. Zhang. *High-Speed 3D imaging with digital fringe projection techniques*, volume 5. CRC Press, 2016.

[108] J. Rangel, S. Soldan, and A. Kroll. 3d thermal imaging: Fusion of thermography and depth cameras. In *International Conference on Quantitative InfraRed Thermography*, 2014.

[109] M. A. Akhloufi and B. Verney. Multimodal registration and fusion for 3d thermal imaging. *Mathematical Problems in Engineering*, 2015, 2015.

[110] S. Coorg and S. Teller. Real-time occlusion culling for models with large occluders. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 83–ff. ACM, 1997.

[111] G. Vaněkčkek. Back-face culling applied to collision detection of polyhedra. *The Journal of Visualization and Computer Animation*, 5(1):55–63, 1994.

[112] J.-S. Hyun and S. Zhang. Enhanced two-frequency phase-shifting method. *Applied optics*, 55(16):4395–4401, 2016.

[113] Y. An, J.-S. Hyun, and S. Zhang. Pixel-wise absolute phase unwrapping using geometric constraints of structured light system. *Optics express*, 24(16):18445–18459, 2016.

[114] Q. Kemao. Windowed fourier transform for fringe pattern analysis. *Applied Optics*, 43(13):2695–2702, 2004.

[115] M. D. Pritt and D. C. Ghiglia. *Two-dimensional phase unwrapping: theory, algorithms, and software*. Wiley, 1998.

[116] X. Su and W. Chen. Reliability-guided phase unwrapping algorithm: a review. *Optics and Lasers in Engineering*, 42(3):245–261, 2004.

[117] Y.-Y. Cheng and J. C. Wyant. Two-wavelength phase shifting interferometry. *Applied optics*, 23(24):4539–4543, 1984.

[118] Y.-Y. Cheng and J. C. Wyant. Multiple-wavelength phase-shifting interferometry. *Applied optics*, 24(6):804–807, 1985.

[119] C. E. Towers, D. P. Towers, and J. D. Jones. Optimum frequency selection in multifrequency interferometry. *Optics letters*, 28(11):887–889, 2003.

[120] Y. Li, H. Jin, and H. Wang. Three-dimensional shape measurement using binary spatio-temporal encoded illumination. *Journal of Optics A: Pure and Applied Optics*, 11(7):075502, 2009.

[121] Y. Wang and S. Zhang. Novel phase-coding method for absolute phase retrieval. *Optics letters*, 37(11):2067–2069, 2012.

[122] D. Zheng and F. Da. Phase coding method for absolute phase retrieval with a large number of codewords. *Optics express*, 20(22):24139–24150, 2012.

[123] C. Zhou, T. Liu, S. Si, J. Xu, Y. Liu, and Z. Lei. An improved stair phase encoding method for absolute phase retrieval. *Optics and Lasers in Engineering*, 66:269–278, 2015.

[124] Z. Li, K. Zhong, Y. F. Li, X. Zhou, and Y. Shi. Multiview phase shifting: a full-resolution and high-speed 3d measurement framework for arbitrary shape dynamic objects. *Optics letters*, 38(9):1389–1391, 2013.

[125] K. Zhong, Z. Li, Y. Shi, C. Wang, and Y. Lei. Fast phase measurement profilometry for arbitrary shape objects without phase unwrapping. *Optics and Lasers in Engineering*, 51(11):1213–1222, 2013.

[126] C. Bräuer-Burchardt, P. Kühmstedt, and G. Notni. Code minimization for fringe projection based 3d stereo sensors by calibration improvement. *arXiv preprint arXiv:1404.7298*, 2014.

[127] K. Song, S. Hu, X. Wen, and Y. Yan. Fast 3d shape measurement using fourier transform profilometry without phase unwrapping. *Optics and Lasers in Engineering*, 84:74–81, 2016.

[128] W. Lohry, V. Chen, and S. Zhang. Absolute three-dimensional shape measurement using coded fringe patterns without phase unwrapping or projector calibration. *Optics express*, 22(2):1287–1301, 2014.

[129] S. Zhang. Flexible 3d shape measurement using projector defocusing: extended measurement range. *Optics letters*, 35(7):934–936, 2010.

[130] C. Zuo, L. Huang, M. Zhang, Q. Chen, and A. Asundi. Temporal phase unwrapping algorithms for fringe projection profilometry: A comparative review. *Optics and Lasers in Engineering*, 85:84–103, 2016.

[131] N. L. Karpinsky, M. Hoke, V. Chen, and S. Zhang. High-resolution, real-time three-dimensional shape measurement on graphics processing unit. *Optical Engineering*, 53(2):024105, 2014.

[132] Y. An and S. Zhang. Pixel-by-pixel absolute phase retrieval assisted by an additional three-dimensional scanner. *Applied Optics*, 58(8):2033–2041, 2019.

[133] S. Tang and Y. Y. Hung. Fast profilometer for the automatic measurement of 3-d object shapes. *Applied Optics*, 29(20):3012–3018, 1990.

[134] D. C. Ghiglia and M. D. Pritt. *Two-dimensional phase unwrapping: theory, algorithms, and software*, volume 4. Wiley New York, 1998.

[135] M. Zhao, L. Huang, Q. Zhang, X. Su, A. Asundi, and Q. Kemao. Quality-guided phase unwrapping technique: comparison of quality maps and guiding strategies. *Applied optics*, 50(33):6214–6224, 2011.

[136] T. Tao, Q. Chen, J. Da, S. Feng, Y. Hu, and C. Zuo. Real-time 3-d shape measurement with composite phase-shifting fringes and multi-view system. *Optics express*, 24(18):20253–20269, 2016.

[137] T. Tao, Q. Chen, S. Feng, J. Qian, Y. Hu, L. Huang, and C. Zuo. High-speed real-time 3d shape measurement based on adaptive depth constraint. *Optics express*, 26(17):22440–22456, 2018.

[138] J. Vargas and J. A. Quiroga. Defect inspection by an active 3d multiresolution technique. In *Ninth International Symposium on Laser Metrology*, volume 7155, page 71550H. International Society for Optics and Photonics, 2008.

[139] E. Hołowko, K. Januszkiewicz, P. Bolewicki, R. Sitnik, and J. Michoński. Application of multi-resolution 3d techniques in crime scene documentation with bloodstain pattern analysis. *Forensic science international*, 267:218–227, 2016.

VITA

VITA

Yatong An is a Ph.D. candidate at the School of Mechanical Engineering in Purdue University. He received his M.Phil. degree from the Department of Computer Science in the University of Hong Kong and his B.S. degree from the Department of Control Science and Engineering in Zhejiang University, respectively. His research interests include optical measurement, 3D reconstruction, multi-modal sensing, image processing, computer vision, machine learning, robotics, control theory and bioinformatics. By the time of graduation, he has published 10 research articles, including one under review. Mr. An was awarded the Optics and Photonics Education Scholarship by SPIE.

LIST OF PUBLICATIONS

LIST OF PUBLICATIONS

Journal articles

[J9] Y. An and S. Zhang, "Pixel-by-pixel absolute phase retrieval assisted by an additional three-dimensional scanner," *Appl. Opt.*, 58(8):2033-2041, 2019.

[J8] Y. An and S. Zhang, "Three-dimensional absolute shape measurement by combining binary statistical pattern matching with phase-shifting methods," *Appl. Opt.*, 56(19): 5418-5426, 2017.

[J7] J. Dai, Y. An, and S. Zhang, "Absolute three-dimensional shape measurement with a known object," *Opt. Express*, 25(9):10384-10396, 2017.

[J6] B. Li, Y. An, D. Cappelleri, J. Xu and S. Zhang, "High-accuracy, high-speed 3D structured light imaging techniques and potential applications to intelligent robotics," *Int. J. Intell. Robot. Applic.*, 1(1):86103, 2017.

[J5] B. Li, Y. An and S. Zhang, "Single-shot absolute 3D shape measurement with Fourier transform profilometry," *Appl. Opt.*, 55(19):5219-5225, 2016.

[J4] Y. An, Z. Liu and S. Zhang, "Evaluation of pixel-wise geometric constraints based phase unwrapping method for low signal-to-noise-ratio (SNR) phase," *Advanced Optical Technologies*, 5(5-6):423432, 2016.

[J3] Y. An, T. Bell, B. Li, J. Xu and S. Zhang, "Method for large range structured light system calibration," *Appl. Opt.*, 55(33):9563-9572, 2016.

[J2] Y. An, J.S. Hyun and S. Zhang, "Pixel-wise absolute phase unwrapping using geometric constraints of structured light system," *Opt. Express*, 24(15):18445-18459, 2016.

[J1] Y. An, and S. Zhang, "High-resolution, real-time simultaneous 3D surface geometry and temperature measurement," *Opt. Express*, 24(13):14552-14563, 2016.

Conference papers

[C1] Y. An and S. Zhang, “High-speed, high-accuracy large range 3D measurement”, In *Proc. SPIE*, volume 10220, 2017.