

ROBUST 4D HYPERSPECTRAL IMAGING AND ITS APPLICATIONS

by

JIAQIONG LI

(Under the Direction of Beiwen Li)

ABSTRACT

Over the past decades, 3D optical measurement techniques, including passive and active approaches, have been extensively studied across diverse fields. Among these, fringe projection profilometry (FPP), an active triangulation-based 3D measurement technique, is widely recognized for its speed, accuracy, and adaptability. This dissertation aims to improve FPP's measurement accuracy and completeness further and expand the data modality to broaden its applicability.

To improve accuracy on objects with sharp depth changes, a circular fringe projection technique is introduced, along with system calibration and 3D reconstruction. Its performance is compared to traditional linear FPP using both standard spheres and complex electronic samples, with results showing superior reconstruction accuracy. Factors affecting the calibration method's accuracy are also explored through sensitivity and phase error analyses.

To mitigate the influence of occlusion, where abrupt geometric changes prevent projector illumination and result in incomplete 3D reconstructions, this dissertation proposes texture-guided phase-to-depth networks that integrate texture maps with phase information, reducing shadow-induced errors. Experimental results from simulations and real-world scans validate the networks' effectiveness.

Expanding FPP's applicability, this dissertation introduces a 4D line-scan hyperspectral imaging system that captures both 3D spatial data and spectral information in the visible and near-infrared (VNIR) range, eliminating post-processing registration. Additionally, the work extends to a 4D Vis-SWIR line-scan hyperspectral imaging system, capturing 3D data and spectral images across the visible to shortwave infrared (Vis-SWIR) range. This dissertation proposes a line-scan homography

method to align VNIR and SWIR hyperspectral images and a geometric transformation technique to register hyperspectral images with 3D data. Results confirm registration accuracy and demonstrate applicability in food quality evaluation.

In conclusion, this dissertation contributes to the field of 3D optical measurement by addressing critical challenges in FPP and introducing innovative techniques that enhance accuracy and expand the dimensionality of data acquisition. These advancements hold considerable potential for both academic research and industrial applications, providing deeper insights into a range of 3D optical measurement technologies.

INDEX WORDS: 3D optical sensing, Fringe projection profilometry, Circular fringe projection profilometry, Occlusion mitigation, 4D hyperspectral imaging, Deep learning

ROBUST 4D HYPERSPECTRAL IMAGING AND ITS APPLICATIONS

by

JIAQIONG LI

B.S., University of Electronic Science and Technology of China, China, 2015

M.S., Shanghai Jiao Tong University, China, 2018

A Dissertation Submitted to the Graduate Faculty of The
University of Georgia in Partial Fulfillment of the Requirements for the Degree.

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2024

©2024

Jiaqiong Li

All Rights Reserved

ROBUST 4D HYPERSPECTRAL IMAGING AND ITS APPLICATIONS

by

JIAQIONG LI

Major Professor: Beiwen Li

Committee: Hongyue Sun
Beshoy Morkos
Xianqiao Wang
Peter Kner

Electronic Version Approved:

Ron Walcott

Dean of the Graduate School

The University of Georgia

December 2024

ACKNOWLEDGEMENTS

I would like to take this opportunity to show my best sincere appreciation to my advisor, Dr. Beiwen Li. Dr. Li provided me with the valuable opportunity to pursue my doctoral degree when I was still working for my previous company. During the challenging study journey, Dr. Li always supports me in exploring the uncertainty of all projects. Whenever I encounter difficulties, Dr. Li is always willing to take the time to discuss them with me. In each discussion, his extensive knowledge and experience have broadened my mind and fueled my passion for pursuing challenging research projects. Besides, I also gained valuable experience in leading a research team and managing projects from him. I feel so fortunate to have Dr. Beiwen Li as my advisor.

Besides, I would like to thank my advisory committee members, Dr. Hongyue Sun, Dr. Beshoy Morkos, Dr. Xianqiao Wang, and Dr. Peter Kner. Without their help and support, I could not implement my graduation plan smoothly. I also want to sincerely thank Dr. Xianqiao Wang for helping me with the transfer process from Iowa State University to the University of Georgia. In addition, I want to thank Dr. Hongyue Sun for his valuable suggestions in my job searching. My gratitude also goes to the previous committee members at Iowa State University, Dr. Meng Lu, Dr. Soumik Sarkar, Dr. Eliot Winer, and Dr. Yiliang Liao.

I also want to thank all my wonderful lab mates and previous colleagues: Badrinath Balasubramaniam, Yashwanth Maddipatla, Li-Hsin Yeh, Micah Mundy, Anush Lakshman, Adam Haroon, Dr. Yi Zheng and Dr. Vignesh Suresh. All my lab mates are so smart and reliable that it makes me feel pleased to work with them. Besides, I also want to express my sincere thanks to all my friends for their kindness and help.

Last but not least, I want to thank my husband, Dr. Weibing Gong, who always encourages, supports, and cares about me and tries his best to make me happy. I also want to thank my small cat,

Keegan, who is lying down at my side when I am writing this acknowledge. I also want to thank my parents for their understanding.

Thank each person who helped me, supported me, and cared about me.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
LIST OF FIGURES	viii
LIST OF TABLES	xvi
1 INTRODUCTION	1
1.1 Motivations	1
1.2 Related works	4
1.3 Challenges and objectives	22
1.4 Dissertation organization	24
2 COMPARATIVE ANALYSIS OF CIRCULAR AND LINEAR FRINGE PROJECTION PROFILOMETRY: FROM CALIBRATION TO 3D RECONSTRUCTION¹	25
2.1 Abstract	26
2.2 Introduction	26
2.3 Principle of the method	28
2.4 Experimental results	36
2.5 Discussion	41
2.6 Conclusions	44
3 TPDNET: TEXTURE-GUIDED PHASE-TO-DEPTH NETWORKS TO REPAIR SHADOW-INDUCED ERRORS FOR FRINGE PROJECTION PROFILOMETRY¹	46
3.1 Abstract	47
3.2 Introduction	47

3.3	Principles	51
3.4	Proposed model	53
3.5	Experiments	56
3.6	Discussion	60
3.7	Conclusion	63
4	4D LINE-SCAN HYPERSPECTRAL IMAGING¹	66
4.1	Abstract	67
4.2	Introduction	67
4.3	Methods	71
4.4	Experiments	77
4.5	Conclusion	85
5	4D VIS-SWIR LINE-SCAN HYPERSPECTRAL IMAGING¹	87
5.1	Abstract	88
5.2	Introduction	88
5.3	Methods	93
5.4	Experimental Evaluation	100
5.5	Two study cases	106
5.6	Conclusions	110
6	SUMMARY AND FUTURE WORK	112
6.1	Summary of research contributions	112
6.2	Future work	115
	Bibliography	118

LIST OF FIGURES

1.1	Epipolar geometry of a stereo vision system (adapted from (Hartley & Zisserman, 2003; S. Zhang, 2013)).	6
1.2	Performance of typical active optical 3D techniques (adapted from (Beraldin et al., 2012; Drouin & Beraldin, 2020)).	7
1.3	A schematic of a phase-shifting interferometer (S. Yang & Zhang, 2018). (a). Setup. (b) Phase shift interferogram.	8
1.4	A schematic of a typical coherence scanning interferometer	10
1.5	ToF depth measurement(adapted from (Foix et al., 2011)). (a) ToF based on pulse time delay. (b) ToF based on phase change.	11
1.6	A typical stripe scanner	13
1.7	A schematic diagram of a fringe projection profilometry system (adapted from (B. Li & Zhang, 2015)).	14
1.8	An example of simple binary coding and Gray coding (adapted from (S. Zhang, 2018)). (a) Simple binary coding. (b) Gray coding.	14
1.9	An example of the three-step phase-shifting method. (a)-(c) Sinusoidal fringe patterns used in the three-step phase-shifting method. (d)-(f) An illuminated cube.	16
1.10	An example of phase unwrapping. (a) One of the 18-step sinusoidal fringe images. (b) Wrapped phase. (c) Unwrapped phase. (d). Cross section of wrapped phase, unwrapped phase, and fringe order to show phase unwrapping using the fringe order.	17
1.11	Schematic of how the projector detects feature points using the phase cue (adapted from (S. Feng et al., 2021)).	19
1.12	Comparison of RGB, multispectral, and hyperspectral features. (a) RGB imaging feature. (b) Multispectral imaging feature. (c) Hyperspectral imaging feature.	20

1.13	Typical hyperspectral imaging approaches (Lucasbosch, 2021).	22
2.1	Pinhole imaging model. The figure is reprinted from (B. Li et al., 2014).	28
2.2	Camera calibration and target 3D estimation. (a) asymmetric circle grid calibration target, whose size is $(5(row) \times 9(column))$ with 20 mm center-to-center distance; (b) an image of the target at a pose with extracted circle centers; (c) the estimated 3D target orientations.	30
2.3	An example of fringe patterns for system calibration. (a)-(b). A circular and radial fringe pattern centered at (u_s, v_s) . $r(u_p, v_p)$ is the distance between the point $p(u_p, v_p)$ and the center $O^s(u_s, v_s)$, and $\theta(u_p, v_p)$ is the pixel angle. (c)-(d). A horizontal and vertical fringe pattern employed in Zhang and Huang's method(S. Zhang & Huang, 2006).	31
2.4	Experiment setup.	36
2.5	Evaluation of triangulation error. (a) Overlay of the target points $(x^{wtg}, y^{wtg}, z^{wtg})$ (red points) and triangulation points (x^w, y^w, z^w) (corners of mesh) using Zhang and Huang's method; (b) Similar overlay of target points and triangulation points using the proposed method; (c) Error between the target points and triangulation points using Zhang and Huang's method, and RMS error for X, Y and Z are 0.0024 mm, 0.0016 mm and 0.2413 mm; (d) Similar error using the proposed method, and RMS error for X, Y, and Z are 0.0033 mm, 0.0019 mm and 0.3401 mm.	38
2.6	Evaluating measurement accuracy of a standard sphere (radius: 49.975 mm): comparative overlays and error analysis. (a)-(d) comparative overlays of the standard sphere and 3D reconstructions: (a) with our proposed calibration and circular FPP; (b) Zhang and Huang's calibration method with our proposed circular FPP; (c) our proposed calibration method and linear FPP; (d) Zhang and Huang's calibration method and linear FPP. (e)-(h) error maps corresponding to (a)-(e), showcasing the deviation from the standard sphere. Mean absolute errors are (e) 0.2056mm, (f) 0.1562mm, (g) 0.2646mm, and (h) 0.2402mm. RMS errors are (e) 0.2965mm, (f) 0.2158mm, (g) 0.6663mm, and (h) 0.6525mm.	39

2.7	3D reconstruction results of a GPU board with the fan. (a) the GPU board with fan; (b) a circular fringe image of the proposed circular FPP; (c)-(d) 3D reconstruction results using the proposed circular FPP with either the proposed calibration or Zhang and Huang’s method; (e) a horizontal fringe image of linear FPP; (f)-(g) 3D reconstruction results using linear FPP with either the proposed calibration or Zhang and Huang’s method.	40
2.8	3D reconstruction results of a GPU cooler. (a) the GPU cooler; (b) a circular fringe image of the proposed circular FPP; (c)-(d) 3D reconstruction results using the proposed circular FPP with either the proposed calibration or Zhang and Huang’s method; (e) a horizontal fringe image of linear FPP; (f)-(g) 3D reconstruction results using linear FPP with either the proposed calibration or Zhang and Huang’s method.	40
2.9	Feature point mapping schematic: camera to projector.	43
3.1	Illustration of the formation of shadows and shadow-induced errors in FPP. (a) The formation of shadows; (b) Shadow-induced errors in the wrapped phase map; (c) 3D reconstructed geometry with holes.	48
3.2	Schematic of the proposed model	54
3.3	One sample in the dataset. (a)-(c) Three phase-shifting fringe patterns projected on an object; (d)-(i) Six gray code images projected on an object; (j) The corresponding texture image; (k)-(l) The corresponding wrapped phase map and unwrapped phase map; (m) The mask generated by a depth threshold; (n) The corresponding ground truth (the depth map directly exported from the virtual system).	57
3.4	Results of an object from the virtual scanning. (a)-(c) The input of the texture image, the mask, and the phase map; (d)-(f) 3D geometry from FPP, the ground truth, and the prediction; (g) The ground truth of the depth map; (h) The predicted depth map; (i) The absolute error map with RMSE $2.6485mm$	59
3.5	Effectiveness evaluation of the proposed model for scanning a single object in the real-world FPP system. (a) Texture image; (b) Unwrapped phase map; (c) 3D reconstructed geometry from FPP; (d) The depth map from FPP; (e) The error map for the non-shadow covered region with RMSE $1.5031mm$; (f) 3D reconstructed geometry from the predicted depth map.	60

3.6	Effectiveness evaluation of the proposed model for scanning two objects in the real-world FPP system. (a) Texture image; (b) Unwrapped phase map; (c) 3D reconstructed geometry from FPP; (d) The depth map from FPP; (e) The error map for the non-shadow covered region with RMSE $2.3579mm$; (f) 3D reconstructed geometry from the predicted depth map.	61
3.7	Predictions under different intensity levels of ambient light strength. (a) The texture image obtained without ambient light; (b) The texture image obtained with the same ambient light strength as the training data; (c) The texture image obtained with strong ambient light to make it close to saturation; (d)-(f) 3D reconstructed geometries by using the FPP method for the three cases; (g)-(i) Absolute error maps for the three cases by comparing ground truths and the predicted depth maps, with RMSE $12.5076mm$, $4.7793mm$, and $5.2146mm$, respectively; (j)-(l) 3D reconstructed geometries from the predicted depth maps for the three cases.	62
3.8	Comparing functions of the texture-dominant branch and phase-dominant branch. (a)-(d) Depth maps of the ground truth, texture-dominant branch, phase-dominant branch, and the output prediction; (e) Texture image; (f)-(h) Absolute error maps of (b)-(d) with RMSE $4.7078mm$, $2.4269mm$ and $2.6485mm$, respectively; (i)-(l) 3D geometries from the ground truth, the texture-dominant branch's prediction, the phase-dominant branch's prediction, and the output prediction.	64
4.1	Data processing pipeline of the proposed system.	71
4.2	Schematic diagram of the 4D line-scan hyperspectral structured light imager combining fringe patterns projection, stage moving, and slit images acquisition	72
4.3	Model of the line-scan structured light system. (a) Projector illumination and camera acquisition process; (b) Triangulations for obtaining the relationship between phase differences and depth differences.	74
4.4	At different depths, the system has different scaling factors in the X^W direction. The planes at Z_1^W and Z_2^W are the known planes for calibration. The scaling factor of any plane at arbitrary depth, ΔZ , is to be determined.	76

4.5	Process of spectral calibration. (a) Spectral image of Mercury Argon Calibration Source; (b) Intensity of the spectral image (a); (c) Third-order polynomial relationship between wavelength and pixel index.	78
4.6	Composition of 4D line-scan hyperspectral structured light imager	79
4.7	A sample evaluation of geometrical accuracy using the proposed method. (a) Overlay of the reconstructed sphere and the ideal sphere with a 20.1170 mm radius; (b) error maps (RMSE: 0.0895 mm); (c) Cross-section at X-axis middle position.	81
4.8	Radiance correction using surface normal estimation. (a) and (b) are the image intensities of the plastic ball before and after corrections, respectively.	81
4.9	Hyperspectral 3D measurement of a tortoise statue and associated video (Visualization 2). (a) RGB image of the tortoise statue; (b) Reconstructed tortoise statue at red light wavelength 640 nm; (c) the reconstructed tortoise statue at green light wavelength 528 nm; (d) Reconstructed tortoise statue at blue light wavelength 465 nm; (e) Intensity change of three different colored parts.	82
4.10	Reflectance and NDVI of spinach leaves at different fresh levels. (a) RGB images; (b) Reflectance; (c) NDVI images.	84
4.11	Defect detection on spinach leaves. (a) RGB image of leaf B with a defect; (b) Reconstructed 3D geometry of leaf B; (c) Depth gradient of leaf B; (d) Profile and vein of leaf B; (e) NDVI image (f) Reconstructed 3D geometry with NDVI value rendering; (g) Results of KNN method; (h) Marked defect.	85
5.1	Data processing pipeline of the proposed system to acquire hyperspectral 3D data.	92
5.2	Schematic diagram of the line-scan hyperspectral imaging (HSI) subsystem, including two synchronized line-scan HSI cameras and a conveyor belt. (a). A typical line-scan HSI camera; (b). Two line-scan HSI cameras are used in the proposed system; the left is for the VNIR wavelength range (400 - 1000 nm), and the right is for the SWIR wavelength range (900 - 1700 nm). In practice, the view planes of the two cameras are not aligned perfectly, and the 3D point P_1 in the right camera's view plane will translate to P_2 to be captured by the left camera.	94

5.3	Schematic diagram of the scanning process: the line-scan camera captures the object along the X-direction as it moves in the Y-direction; after traveling a distance, the object stops, and the frame-scan camera starts to scan it.	97
5.4	Experimental setup: The hyperspectral imaging subsystem consists of an InGaAs camera paired with an ImSpector N17E and a CMOS camera paired with an ImSpector V10, both synchronized; the 3D scanning subsystem includes a DLP projector and a CMOS camera (labeled as 5); a conveyor belt is utilized to translate samples constantly, and a halogen lamp acts as an extra light source. (a) Front view; (b) Side view.	101
5.5	Registration results of spectral images of an asymmetric circular grid. (a) and (e) spectral images at 700 nm. (b) and (f) Spectral images at 1347 nm. (c) Proposed method on (a) and (b) with RMSE 0.8332 pixels, MAE 0.4919 pixels. (d) Conventional homography (Hartley & Zisserman, 2003) on (a) and (b) with RMSE 1.3254 pixels, MAE 0.8414 pixels. (g) Proposed method on (e) and (f) with RMSE 0.8258 pixels, MAE 0.5336 pixels. (h) Conventional homography on (e) and (f) with RMSE 3.4614 pixels, MAE 2.2545 pixels.	102
5.6	Evaluation of geometric accuracy. (a) Overlay of the reconstructed sphere (shown in yellow) and a standard sphere (shown in blue) with a 20.0512mm radius. (b) Error map between the reconstructed sphere and standard sphere (RMSE: 0.1321mm, MAE: 0.1066mm). (c) Cross-section of the error map at the X-axis middle position. 103	
5.7	Registration results of 3D data and VNIR spectral images. (a), (e), and (i) Spectral images at 700 nm of an asymmetric circular grid in different poses. (b), (f), and (j) Corresponding 3D data. (c), (g), and (k) 3D data rendered with aligned spectral images at 700 nm. (d), (h), and (l) Differences between aligned spectral and FPP grayscale images, with RMSE and MAE values of 0.7566 and 0.6917 pixels, 0.5086 and 0.4304 pixels, and 0.7092 and 0.4132 pixels.	104

5.8	Registration results of 3D data and SWIR spectral images. (a), (e) and (i) Spectral images at 1347 nm of an asymmetric circular grid in different poses. (b), (f), and (j): Corresponding 3D data. (c), (g), and (k) 3D data rendered with aligned spectral images at 1347 nm. (d) and (h) Differences between aligned spectral and FPP grayscale images, with RMSE and MAE values of 0.6302 and 0.4529 pixels, 0.6714 and 0.5417 pixels, and 1.1951 and 1.035 pixels.	105
5.9	Spectral calibration. (a) Spectral image of the Mercury Argon wavelength calibration light source. (b) Spectral image of the Argon wavelength calibration light source. (c) Third-order polynomial relationship between wavelength and pixel with 95% prediction interval. The blue line indicates the nearly linear relationship for the VNIR wavelength range and the red line for the SWIR wavelength range.	107
5.10	Hyperspectral 3D scanning of a UNO box. (a) RGB image. (b)-(e) 3D renderings with spectral images at 436 nm, 547 nm, 762 nm, and 911 nm in the VNIR range. (f) Spectral image intensity changes in areas A, B, C, and D from 400-900 nm. (g)-(j) 3D renderings with spectral images at 1125 nm, 1376 nm, 1476 nm, and 1675 nm in the SWIR range. (k) Spectral image intensity changes in areas A, B, C, and D from 1000-1600 nm.	107
5.11	Reflectance changes of potato before and after sprouting. (a) RGB image of a potato before sprout. (b) - (d) Reflectance at 580 nm, 772 nm, and 874 nm. (e) RGB image of the same potato after sprout. The red A, blue B, and green C areas are the sprout, normal area, and sprout eye. (f) - (h) Reflectance of the potato with sprouts at 580 nm, 772 nm, and 874 nm. (i) Comparison of reflectance changes before and after sprout in the VNIR range. (j) Comparison of reflectances before and after sprout in the SWIR range.	109
5.12	Sprout detection based on reflectance at 580 nm and depth map. (a) A potato with sprouts. (b) Reflectance at 580 nm. (c) classified reflectance using the K-means method. (d) Sprout candidates with reflectance. (e) Depth map. (f) and (g) Depth derivatives along the x-axis and y-axis. (h) and (i) classified derivatives along the x-axis and y-axis using the K-means method. (j) Sprout candidates with depth. (k) Detected sprouts.	110

5.13 Sprout detection based on reflectance at 772 nm and depth map, yielding similar results. (a) The same potato with sprouts. (b) Reflectance at 772 nm. (c) classified reflectance using the K-means method. (d) Sprout candidates with reflectance. (e) Depth map. (f) and (g) Depth derivatives along the x-axis and y-axis. (h) and (i) classified derivatives along the x-axis and y-axis using the K-means method. (j) Sprout candidates with depth. (k) Detected sprouts. 111

LIST OF TABLES

3.1	Parameters of the virtual FPP system	53
5.1	Fitness of the third-order polynomial relationship wavelength vs pixel index	106

CHAPTER 1

INTRODUCTION

In recent years, 3D optical technologies, especially 3D structured light techniques, have made significant strides, reshaping numerous industries with their innovative applications, such as manufacturing, entertainment, robotics, etc. Among these techniques, fringe projection profilometry has stood out due to its accuracy, flexibility, and robustness. It is widely used to obtain an object's 3D geometry, notably in reverse engineering applications. Despite its advantages, fringe projection profilometry still faces several challenges, including accuracy in complex scenarios with sharp depth change and occlusion. This dissertation focuses on addressing three critical issues: enhancing measurement accuracy through circular fringe patterns; solving occlusion-induced shadow problems, which cause errors and voids in objects with complex geometries; and expanding data modalities by integrating 3D spatial data with hyperspectral images to provide a more comprehensive understanding of objects. This chapter provides a background of the dissertation research. Section 1.1 delves into the motivations behind this study, highlighting the need for enhanced 3D imaging techniques. Section 1.2 presents a review of related work, offering context and insights into the current state of the field. Section 1.3 outlines the specific objectives of this research. Finally, Section 1.4 introduces the structure of the dissertation.

1.1 Motivations

This chapter discusses the motivations behind the dissertation research, and we take three applications as examples: entertainment, biometric authentication, and precision agriculture.

1.1.1 Entertainment

3D optical measurement technology is employed to enhance interactions, making them more natural, intuitive, and immersive between users, entertainment devices, and the physical environment. Apple Vision Pro, Apple's first spatial computer, is an example of the advantages of 3D optical measurement in bridging the gap between users and the digital and physical realms. Equipped with twelve advanced cameras, a LiDAR sensor, and a depth projector, this device excels in various tracking functionalities, including precise headset positional tracking, user hand tracking, and eye tracking. Users can interact with the device effortlessly by natural gestures and movements without controllers. Additionally, the device's capability to track facial expressions in real-time significantly enhances applications like FaceTime, allowing for creating dynamic, persona-driven experiences. The implementation of 3D optical measurement in Apple Vision Pro fundamentally alters the user interaction experience, blurring the line between the digital and the physical world.

1.1.2 Biometric Authentication

Biometric characteristics, such as face and iris, captured by 3D optical technology are increasingly utilized for authentication and security purposes in train stations, airports, banks, homes, etc. This is attributed to the inherent stability and uniqueness of biometrics. For example, facial recognition based on 3D face data has gained popularity in applications like vending machines and iPhone Face ID. Compared to traditional 2D face imaging, 3D facial recognition is more reliable and effective due to the comprehensive representation of facial geometry (Abate et al., 2007). Besides, as a contactless method, facial recognition is more acceptable, particularly during events like the COVID-19 pandemic, where non-contact procedures are preferred over contact-based methods like fingerprint verification for hygiene and safety reasons.

Iris recognition is another form of contactless biometric authentication that has been applied in many fields, producing no false matches in several million comparison tests (Daugman, 2009). However, traditional 2D iris recognition scanners can be fooled by customized contact lenses, which threatens the safety of the authentication system. To address this, research has focused on 3D iris recognition employing optical technology. Connell et al. (Connell et al., 2013) introduced a structured light scanning system to capture the 3D geometry of human eyes, differentiating between

eyes with and without contact lenses. Daniel P et al. (Benalcazar et al., 2019) designed a 3D iris scanner from multiple 2D visible images and reconstructed dense 3D point clouds of iris. In addition to these developments, researchers are exploring multimodal biometric systems that combine facial and iris recognition, offering enhanced security compared to unimodal systems (De Marsico et al., 2014; Y. Wang et al., 2003; Z. Wang et al., 2011).

1.1.3 Precision agriculture

Precision agriculture is increasingly critical for efficiently and sustainably producing a diverse range of high-quality products to meet the needs of a growing population and their rising living standards. To achieve precision agriculture, various sensors, including 3D optical sensors, are utilized to obtain spatial variability within fields (Lee et al., 2010). For instance, the structured light sensor Kinect for Windows version 1 was utilized with other sensors for 3D plant modeling, plant organ classification, and weed detection and classification (Rosell-Polo et al., 2015). Nguyen et al. developed a structured light-based 3D reconstruction system to obtain accurate 3D models of plants (T. T. Nguyen et al., 2015). Chéné et al. utilized Kinect sensors to capture depth maps of plants and analyzed plant phenotyping with 3D measurements (Chéné et al., 2012). In addition to 3D optical sensors, other types of sensors, such as ultrasound, thermal cameras, and spectral and hyperspectral cameras, are also employed (Narvaez et al., 2017). These sensors not only capture 3D spatial geometry but also other important information like temperature, water stress, nitrogen deficiencies, etc. However, it's challenging to collect, organize, and integrate multiple data modalities from different sensors. In response, researchers are exploring the potential of high-dimensional optical sensors, hoping to capture extensive data more efficiently and simply with advanced devices. Based on structured light, Heist et al. (Heist et al., 2018a) developed a 5D hyperspectral imaging system to measure the shape (coordinates x, y, z) and spectral characteristics ($R(\lambda_i)$) over time (t), and applied the system to determine leaf water content, demonstrating its potential in precision agriculture. Similarly, Luo et al. (Luo, Lin, et al., 2022) designed a portable 4D snapshot hyperspectral imager for fast-spectral and surface morphology measurement and tested the proposed system on broccolis to show the promising application in plant growth monitoring. Overall, current research demonstrates that 3D optical sensors, especially those with high-dimensional capabilities, have the potential to revolutionize precision agriculture.

1.2 Related works

Generally, 3D optical techniques can be grouped into two main categories: active and passive metrology, based on whether controlled energy is emitted to benefit measurement (Mada et al., 2003). Active optical metrology involves the emission of controlled energy, such as a scanning laser or a projected pattern of light, which actively interacts with the object using the emitted energy. The emitted energy is critical for precise and accurate measurements, especially in environments where natural lighting conditions are not ideal for measurement. In contrast, passive optical metrology only relies on ambient or natural light sources to capture information about an object, which typically captures images from multiple perspectives and uses algorithms to reconstruct a 3D object model. This section will introduce commonly used passive and active optical 3D techniques, as well as hyperspectral imaging related to this dissertation.

1.2.1 Passive optical 3D imaging techniques

Over the past decades, passive optical 3D techniques have been extensively studied. Essentially, a passive 3D imaging system can recover 3D scene information without extra light projected onto the scene, but visible object features are necessary for passive metrologies. Passive 3D vision techniques can be categorized as single-view and multiple-view approaches (Se & Pears, 2020). Several single-view approaches, including shape from shading, shape from texture, and shape from focus, are briefly introduced in the following section. For multiple-view approaches, we focus on stereo vision.

Single-view approaches

Single-view approaches, also known as shape from X methods (Se & Pears, 2020), infer three-dimensional structures from a single viewpoint using cues like shading, texture, and focus (Se & Pears, 2020). Shape from shading (R. Zhang et al., 1999) infers surface shape from grayscale image shades by linking image intensity to surface orientation, leading to depth map creation through surface normal integration and regularized fitting. Shape from texture (Gårding, 1992) analyzes texture distortion caused by the imaging process to estimate surface shape. Meanwhile, shape from focus (Nayar & Nakagawa, 1994) utilizes blur in images taken from the same viewpoint but

at different depths of field to gauge object depth, constructing relative depth from the degree of defocus quantified through squared gradient values. Moreover, deep learning is increasingly applied to estimate depth maps from single-view images, which is also termed monocular depth estimation (Ming et al., 2021). Despite their potential, these single-view methods often face challenges and limitations in robustness and speed for 3D reconstruction, which leaves a gap for multiple-view approaches.

Stereo vision

As the most widely used multiple-view approach, stereo vision is a fascinating and flourishing technology that has been widely applied in various fields, such as robotics, autonomous vehicles, virtual reality (VR), augmented Reality (AR), etc. Inspired by the human vision system, a typical stereo-vision system consists of a pair of cameras to capture the scene from different perspectives. By matching the correspondence pixels in the captured images, disparity maps can be determined, and 3D geometry can be reconstructed (S. Zhang, 2013). The main tasks in stereo vision include system calibration, rectification, matching, depth map generation, and 3D reconstruction.

Fig. 1.1 presents the epipolar geometry of a typical stereo vision system, which provides geometry constraints to search stereo correspondence. In Fig. 1.1, an object point W in the world coordinate system $O_w x_w y_w z_w$ is imaged in two image planes, p_1 at the left camera image plane π_1 and p_2 at the right camera image plane π_2 . In other words, p_1 and p_2 are the intersections of the light ray O_1W and O_2W with their corresponding image planes, respectively. O_1 and O_2 are the optical centers of the left and right camera, and the line between O_1 and O_2 is the baseline. W , O_1 , and O_2 form the epipolar plane containing the baseline. e_1 and e_2 are the intersections of the baseline with image planes, which are named epipoles and are also the vanishing points of the baseline direction. Suppose we only have images from the left camera. As illustrated in Fig. 1.1, it's challenging to back-project 2D point p_1 to 3D point W , as the points on the light ray O_1W (i.e., W_1 and W_2) are all candidates. Similar to the human vision system, depth ambiguity can be avoided by pairs of 2D image points. However, searching for matching points usually brings a high computational cost. Epipolar lines significantly simplify the searching process and reduce computational complexity by constraining the search to the epipolar lines.

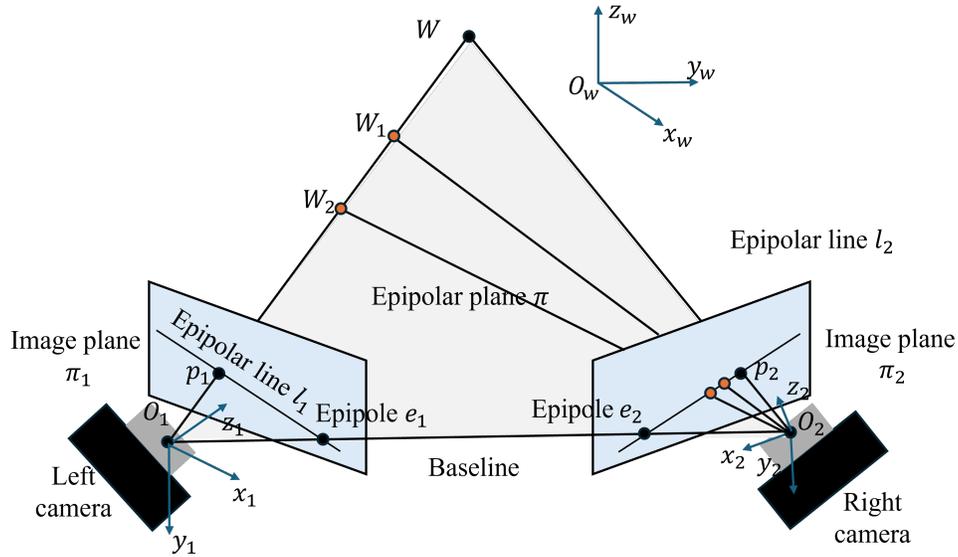


Figure 1.1: Epipolar geometry of a stereo vision system (adapted from (Hartley & Zisserman, 2003; S. Zhang, 2013)).

Stereo vision systems require calibration to establish the geometrical relationship between their two cameras. One of the widely used camera models in stereo calibration is the pinhole camera model. As presented in Fig. 1.1, the pinhole cameras (i.e., the left and right camera) are modeled by their optical centers and image planes, which can be calibrated using the widely acknowledged method proposed by Zhang et al. (Z. Zhang, 2000) to obtain its intrinsic parameters. After calibrating the individual cameras, stereo calibration can be treated as a general camera calibration with respect to a fixed reference coordinate system. This process ultimately determines the transformation between the two camera coordinate systems.

After establishing the geometrical relationship between the two cameras, the next step is to align the captured images to a common image plane, which is known as stereo rectification. The primary goal of stereo rectification is to align the corresponding points in each image to lie on the same epipolar lines, and it simplifies the following task of finding matching points between images. Based on the rectified image pairs, stereo-matching methods can be applied to find correspondence points in the same horizontal lines. Generally, the matching methods are categorized into global and local methods according to how they assign disparities to pixel (Lazaros et al., 2008a). Global methods,

like the scan-line optimization method (Mattochia et al., 2007), assign disparities according to the information perceived from the whole image, while local methods (i.e., the sum of absolute differences method (Mühlmann et al., 2002)) only rely on each pixel’s neighbors to determine disparities. Compared to global methods, local methods are usually faster. When the stereo system is calibrated, and correspondence pairs are determined, disparities can be obtained based on the triangulation relationship between the cameras and object.

In conclusion, despite inherent challenges and limitations associated with the complexity of processing and interpreting visual data from dual viewpoints, stereo vision is one of the most used 3D optical techniques and is applied in many fields, from robotics to virtual reality. And the development of high-performance graphic processing units is progressively mitigating its computational challenges, further broadening its application scope.

1.2.2 Active optical 3D techniques

This subsection will introduce three widely used active optical 3D techniques: interferometry-based, time-of-flight-based, and triangulation-based metrology. Fig. 1.2 summarizes the typical accuracy of the widely used active optical 3D imaging technology. It can be observed that each optical technique is only suited for a specific range of accuracies and object sizes. Therefore, the selection of optical sensors should be tailored to the intended application environments.

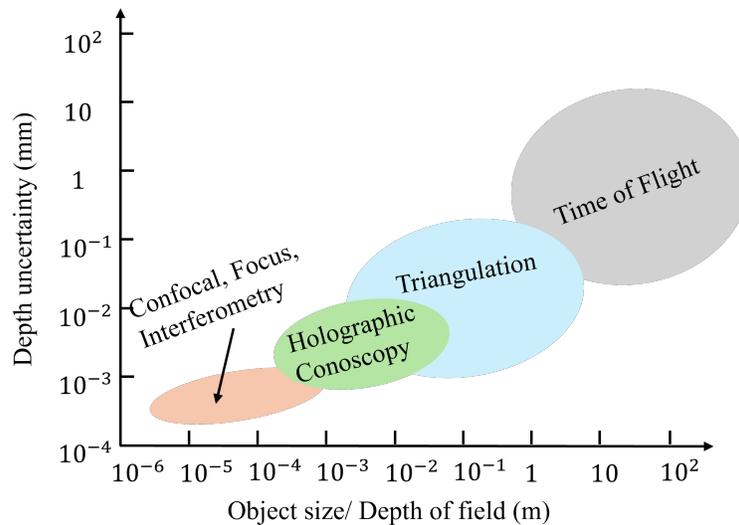


Figure 1.2: Performance of typical active optical 3D techniques (adapted from (Beraldin et al., 2012; Drouin & Beraldin, 2020)).

Interferometry-based metrology

Interferometry is the most accurate 3D optical metrology method, which can achieve nanometer, subnanometer, and even finer resolutions (S. Yang & Zhang, 2018). Due to high precision and non-contact advantages, it is widely used for micro-scale measurements, such as measuring cell structure and dynamics. Given interference patterns, there are primarily two types of interferometry to perform 3D surface measurement: phase-shifting interferometry (PSI) and coherence scanning interferometry (CSI).

Phase-shifting interferometry. Since temporal phase-shift measurement technology was brought into interferometry (Bruning et al., 1974; Carré, 1966; J. Wyant, 1973), phase-shifting interferometry has been widely studied in the past decades. Due to the high accuracy, PSI has been adapted to a variety of interferometer types, including Twyman–Green, Fizeau, Mach–Zehnder, and common path configurations like point diffraction and lateral shearing interferometers (Meddecki et al., 1996; Schreiber & Bruning, 2007; S. Yang & Zhang, 2018). Fig.1.3 depicts a schematic of a phase-shifting interferometer. In PSI, the laser usually acts as the light source. The phase shift interferograms are controlled through the mechanical motion of a piezoelectric ceramic transducer (PZT). When the PZT translates the reference surface, the optical distance difference (OPD) between the target light and the reference light will change, bringing in phase shifting. In addition to the motion of a PZT, phase shifting can also be achieved by the modulation of the light frequency. For an N -step phase-shifting algorithm, the phase can be retrieved by ((Marrugo et al., 2020a))

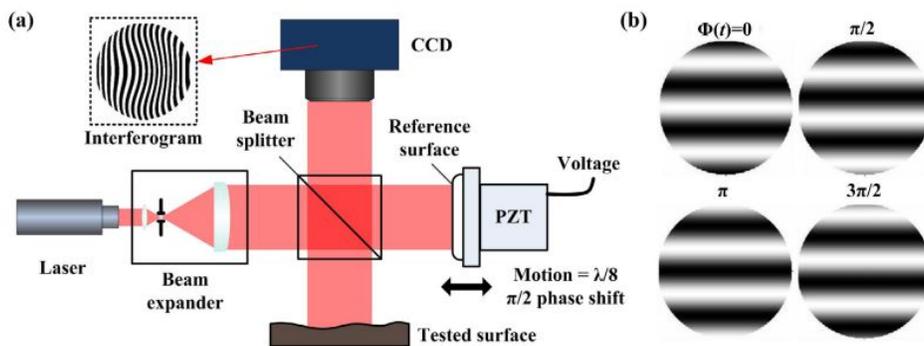


Figure 1.3: A schematic of a phase-shifting interferometer (S. Yang & Zhang, 2018). (a). Setup. (b) Phase shift interferogram.

$$\phi(x, y) = -\tan^{-1} \left[\frac{\sum_{k=1}^N I_k(x, y) \sin 2\pi k/N}{\sum_{k=1}^N I_k(x, y) \cos 2\pi k/N} \right], \quad (1.1)$$

where

$$I_k(x, y) = I'(x, y) + I''(x, y) \cos [\phi(x, y) + 2\pi k/N]. \quad (1.2)$$

In the above two equations, $I'(x, y)$ and $I''(x, y)$ denote the average intensity and the intensity modulation, respectively. k is the k -th interference pattern of the N -step. Due to the inverse tangent function, wrapped phase $\phi(x, y)$ has 2π discontinuities, which makes phase unwrapping necessary. Once the unwrapped phase $\Phi(x, y)$ is obtained, the height $h(x, y)$ can be calculated as follows (J. C. Wyant, 2013),

$$h(x, y) = \frac{\lambda}{4\pi} \Phi(x, y) \quad (1.3)$$

Here, λ denotes the wavelength. PSI is usually used to analyze very smooth surfaces due to its subnanometer resolution (Hutter et al., 2013), which limits its range of application.

Coherence scanning interferometry. Different from PSI, white light or broadband light acts as the light source to illuminate the object and reference surfaces in coherence scanning interferometry (Schmit et al., 2007). CSI interference fringes can only be observed over a narrow surface height range due to the limited coherence length of the white light source (de Groot, 2011; Marrugo et al., 2020a). Fig. 1.4 illustrates a schematic of a typical coherence scanning interferometer. When the Mirau interference objective is scanned vertically in the figure along its optical axis, the interference fringes will change, and the camera will record the set of changing interferograms. For this reason, CSI is also known as vertical scanning interferometry, coherence radar, or white-light scanning interferometry. In Fig.1.4, the Mirau interference objective is translated up. In other configurations, the object, the reference mirror, or the interferometer beam splitter can be moved up (de Groot, 2011). Generally, the scan motion is along the optical axis of the objective and perpendicular to the object's surface. When each point on the surface being measured is focused, the fringe contrast at that point will reach its peak. By noting the peak contrast location, the relationship between the height and the maximal fringe contrast will be established, yielding a highly sensitive surface topography image with subnanometer vertical resolution. However, CSI is also sensitive

to vibrations and environmental disturbances, which is challenging for some applications, such as in-situ or in-line measurement. The high precision and limitations make it typically utilized in fields like material science, semiconductor inspection, and precision engineering with a decent controlled environment.

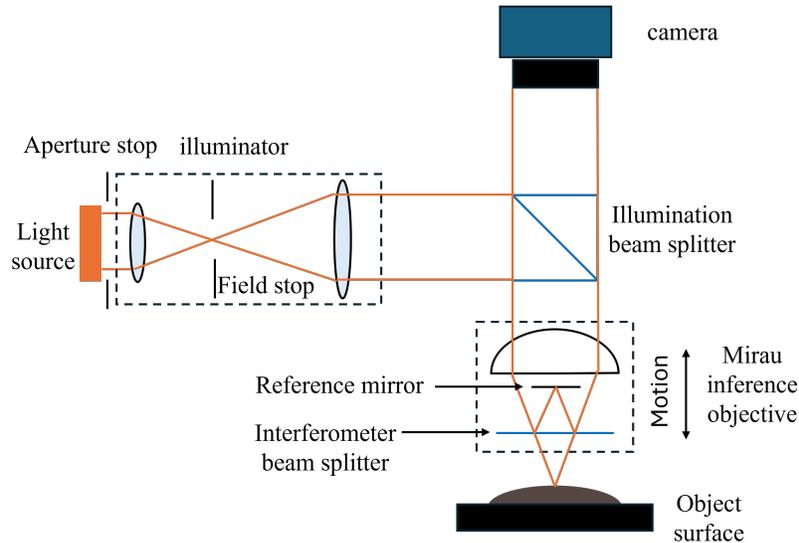


Figure 1.4: A schematic of a typical coherence scanning interferometer

Time-of-flight-based metrology

As a ranging technique with macro accuracy, time of flight (ToF) is applied in many fields, such as robotics (Alenyà et al., 2014), autonomous driving (Niskanen et al., 2021), agriculture (Vázquez-Arellano et al., 2016), etc. ToF works on the principle of measuring the time taken for an emitted signal, usually light or a laser, to reflect back to the sensor after hitting an object. The accurate measurement of the round-trip time τ can usually be obtained in two methods: direct methods that either measure the time delay τ by pulsed light or phase change $\Delta\varphi$ by continuous-wave (CW) modulation, and indirect methods that derive time delay or phase change from time-gated measurements of the signal at the receiver (Marrugo et al., 2020a).

Fig. 1.5 illustrates the principles of ToF based on pulse time delay and phase change. In Fig. 1.5(a), the pulse light signal is emitted from the emitter and then returns back to the receiver after the object surface reflects it with a short time delay. Thus, the distance can be determined by

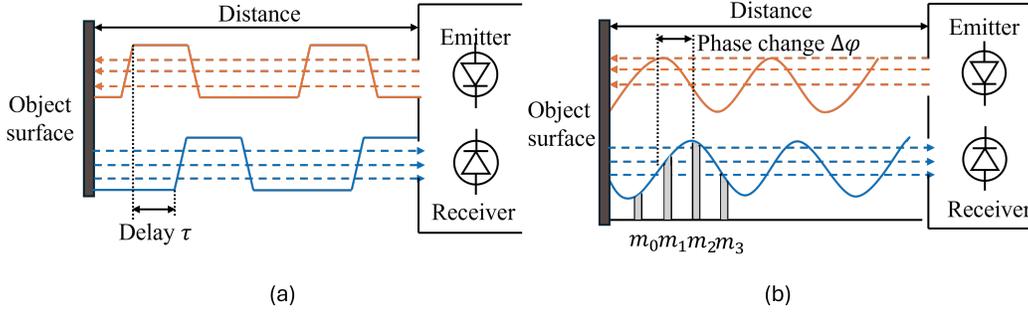


Figure 1.5: ToF depth measurement(adapted from (Foix et al., 2011)). (a) ToF based on pulse time delay. (b) ToF based on phase change.

$$d = \frac{c \cdot \tau}{2}. \quad (1.4)$$

Here, c is the speed of light. Though the above equation is quite simple, detecting a short time delay is challenging due to the extremely fast speed of light, which influences the measurement accuracy of pulse-based ToF and limits its application. Different from directly measuring the time delay in pulse-based ToF, phase-based ToF relies on the phase change of the emitted signal to determine the distance. As shown in Fig. 1.5(b), continuous waves like sinusoidal or other periodic signals are utilized to analyze the phase change between the emitted light and the detected reflected light. Every sensor pixel samples the reflected light four times at equal intervals in each period m_0, \dots, m_3 , and the phase change $\Delta\phi$ can be given by (Foix et al., 2011)

$$\Delta\phi = \arctan\left(\frac{m_3 - m_1}{m_0 - m_2}\right). \quad (1.5)$$

Based on the obtained phase change, the target distance can easily be determined by (Foix et al., 2011)

$$d = \frac{c \cdot \Delta\phi}{4\pi \cdot f_m}, \quad (1.6)$$

where f_m is the modulation frequency.

ToF technology is typically employed in large scenarios ranging from 1 m to 100 m (Beraldin et al., 2012). Besides, its sensors are often compact and portable, leading to widespread use in

commercial products, such as Microsoft Kinect. However, ToF has its limitations. It has lower measurement accuracy compared with other 3D optical methodologies like interferometry and triangulation-based metrology. Furthermore, ToF suffers both non-systematic errors, including multipath interference and boundary ambiguity, and systematic errors like motion error (Hansard et al., 2012).

Triangulation-based metrology

The active triangulation-based 3D imaging system is similar to the stereo vision system, which replaces one camera with a projection device like a digital video projector or a laser. Based on the scanning manner, active triangulation-based 3D scanners can be categorized into spot scanners, stripe scanners, and area-based scanners (Drouin & Beraldin, 2020). The following section will briefly introduce spot scanners and stripe scanners and then focus on area-based scanners, especially fringe projection profilometry systems.

Spot scanner. The spot scanner is also known as the point-based scanner, including a linear spot detector and a laser projector. The laser beam illuminates a very small spot part of the scene for each image frame (Drouin & Beraldin, 2020). To capture a whole image of the scene, the detector usually moves with the mounted 2D translation stage, or the laser rotates around two axes, which increases hardware complexity. Besides, As it takes extra time to move the camera or rotate the laser, it is unsuitable for dynamic scenes.

Strip scanner. Fig. 1.6 presents a typical stripe scanner, which consists of a camera, a laser line projector, and a controlled turntable. The illuminated 3D point p is projected to the 2D point q , which means the back-projected light ray intersects with the light plane at 3D point p and with the camera image plane at 2D point q . For each captured frame, the stripe scanner only recovers a line of points. Thus, the object rotates with the turntable to incrementally reconstruct the 3D geometry of the scene. In other setups, the object may move with a linear translation stage or a conveyer belt. The stripe scanner shares similar limitations with the spot scanner, as they both require the object to be rotated or moved incrementally, which increases additional hardware complexity and time consumption for scanning.

Fringe projection profilometry. As an area-based 3D scanning technique, the translation stage or the turntable is not necessary for fringe projection profilometry. Fig. 1.7 presents a typical fringe

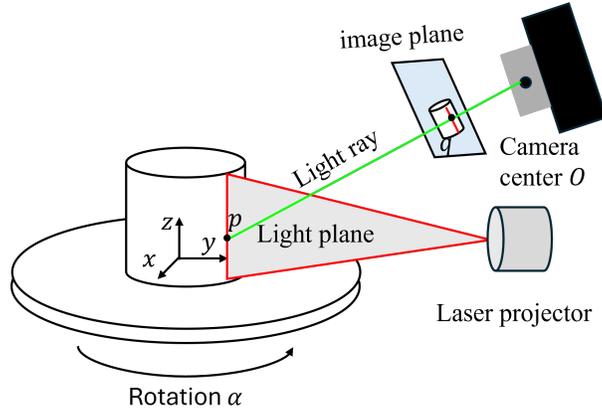


Figure 1.6: A typical stripe scanner

projection profilometry (FPP) system, which is composed of a camera and a digital video projector. The projector projects fringe pattern images that are distorted by the object's surface. The camera simultaneously captures the distorted fringe pattern images. Therefore, the distortion of fringe patterns is related to the object profile, and phase information of the distorted fringe patterns builds the correspondence relationship between the camera and projector.

There are many coding strategies for fringe patterns to build a one-to-one mapping between the camera and the projector, and two of the most widely used strategies are sinusoidal phase coding and binary coding. For binary coding, each horizontal or vertical line is assigned a unique value, which refers to a unique codeword. Binary coding includes simple coding and Gray coding. Fig. 1.8 presents an example of simple binary coding and Gray coding to present eight stripes with three fringe pattern images. In this figure, three fringe pattern images are utilized to present 3-bit coding, and each stripe in the fringe pattern images has a unique codeword. Compared to simple binary code, only one stripe changes in adjacent stripes in Gray coding, which makes Gray coding more robust to sampling errors (S. Zhang, 2018). Sampling errors are more likely to occur on fringe transitions, where the pixel changes from white to black or black to white. As fewer fringe transitions are on the Gray code images, there is a lower possibility of bringing sampling errors. However, the width of the stripe limits the smallest resolution, which makes it challenging to achieve pixel-level spatial resolution at a very high speed. Besides, as binary coding relies on the image intensity affected

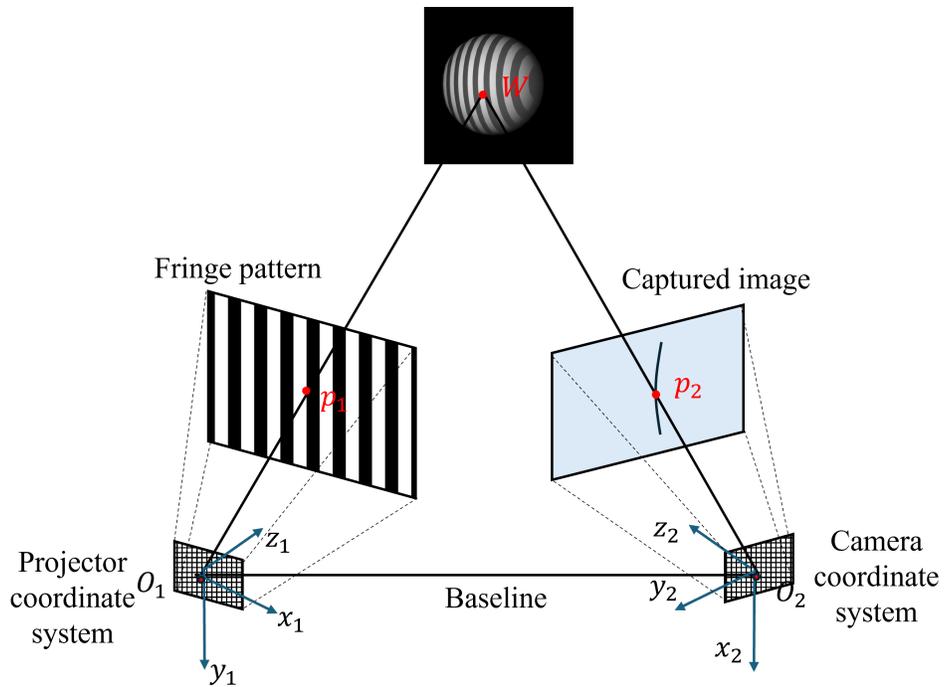


Figure 1.7: A schematic diagram of a fringe projection profilometry system (adapted from (B. Li & Zhang, 2015)).

by surface texture, the measurement accuracy varies for objects with different textures (S. Zhang, 2018).

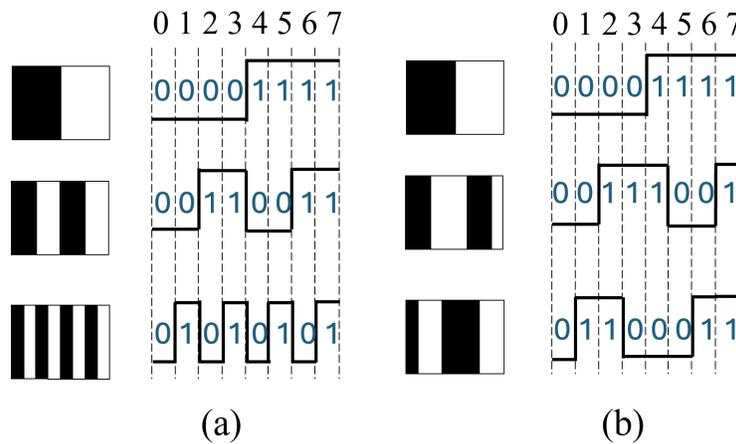


Figure 1.8: An example of simple binary coding and Gray coding (adapted from (S. Zhang, 2018)). (a) Simple binary coding. (b) Gray coding.

Compared with binary coding, sinusoidal fringe pattern images contain smooth and periodic variations in grayscale. Namely, the intensity of the image follows a sinusoidal function of the phase information and changes continuously and periodically from 0 to 255. The sinusoidal fringe pattern can be determined by

$$I(x, y) = I'(x, y) + I''(x, y) \cos [\phi(x, y)], \quad (1.7)$$

where $I(x, y)$, $I'(x, y)$ and $I''(x, y)$ are the intensity, the average intensity and intensity modulation at the pixel point (x, y) , and $\phi(x, y)$ is the phase. Generally, there are two main methods to recover the phase: the Fourier transform and derived methods and phase-shifting methods.

The 3D measurement technique based on the Fourier transform is named Fourier transform profilometry (FTP) (Takeda & Mutoh, 1983). For FTP, a single fringe pattern image is sufficient to recover the phase. The above equation can also be rewritten as

$$I(x, y) = I'(x, y) + \frac{I''(x, y)}{2} [e^{j\phi(x, y)} + e^{-j\phi(x, y)}], \quad (1.8)$$

where $e^{j\phi(x, y)}$ and $e^{-j\phi(x, y)}$ are two conjugate components. Usually, we can utilize a band-pass filter such as a Hanning window in the frequency domain to remove $I'(x, y)$. After filtering, the image will be

$$\tilde{I}(x, y) = \frac{I''(x, y)}{2} [e^{j\phi(x, y)} + e^{-j\phi(x, y)}]. \quad (1.9)$$

Then, the phase can be recovered from

$$\phi(x, y) = \tan^{-1} \left\{ \frac{Im [\tilde{I}(x, y)]}{Re [\tilde{I}(x, y)]} \right\} \quad (1.10)$$

Here, $Im [\tilde{I}(x, y)]$ and $Re [\tilde{I}(x, y)]$ are the imaginary and real part of the filtered image $\tilde{I}(x, y)$. As a single shot of the fringe image is sufficient to recover the phase, FTP is widely applied in high-speed 3D measurement. However, conventional FTP methods face three significant challenges: sensitivity to noise, strict requirements for surface optical properties, such as minimal texture, and difficulties in accurately measuring complex surface geometries (S. Zhang, 2018).

Another widely used method to recover the phase is the phase-shifting method (Malacara, 2007a). Unlike FTP, which only requires a single fringe pattern image, the conventional phase-shifting

method needs a set of phase-shifted sinusoidal fringe pattern images to recover the phase. The standard N -step phase shifting method can be described by (Zuo et al., 2018)

$$I(x, y) = I'(x, y) + I''(x, y) \cos [\phi(x, y) + 2\pi n/N], \quad (1.11)$$

where n is the phase-shifting index $n = 0, 1, 2, \dots, N - 1$. Then, the phase can be calculated by

$$\phi(x, y) = \tan^{-1} \left[\frac{\sum_{n=0}^{N-1} I_n(x, y) \sin 2\pi n/N}{\sum_{n=0}^{N-1} I_n(x, y) \cos 2\pi n/N} \right], \quad (1.12)$$

The standard phase-shifting method requires at least three fringe pattern images for implementation, as illustrated in Fig. 1.9 with an example of a cube illuminated by a three-step phase-shifting sequence. Generally, the more steps used in the phase shifting, the higher measurement accuracy it can achieve. This method, compared to Fourier Transform Profilometry (FTP), requires more fringe patterns, adversely affecting measurement speed, particularly in high-speed measurement applications. To enhance measurement speed, accuracy, and range, various adaptations of FTP and phase-shifting techniques have been developed, such as trapezoidal phase shifting, triangular phase shifting, π -shift Fourier transform profilometry, background-normalized Fourier transform profilometry, etc.(Zuo et al., 2018).

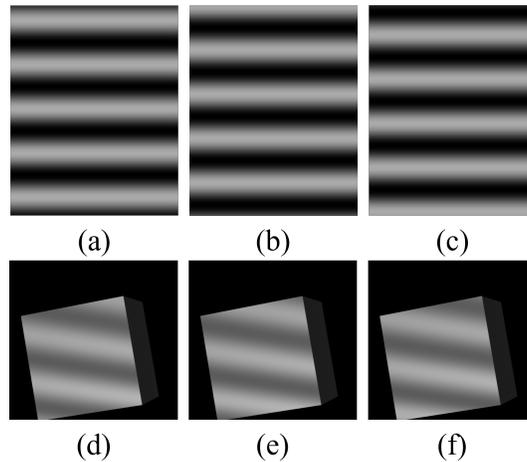


Figure 1.9: An example of the three-step phase-shifting method. (a)-(c) Sinusoidal fringe patterns used in the three-step phase-shifting method. (d)-(f) An illuminated cube.

Phase in Eq. 1.10 and Eq. 5.12 is at the range of $[-\pi, \pi)$, which is not continuous. Thus, the phase is also regarded as the wrapped phase that needs phase unwrapping to remove the phase jumps. After unwrapping, the phase without 2π discontinuity is named the unwrapped phase. Various unwrapping strategies have been investigated, and one of the straightforward ideas is to utilize Gray coding images (D. Zheng et al., 2017) to denote the fringe order. A set of Gray coding images is generated within the same period of the sinusoidal fringe patterns, and then both the sinusoidal fringe pattern and Gray coding pattern are projected onto the object. Fig. 1.10 illustrates an example of phase unwrapping using Gray coding. The fringe order can be identified by analyzing the intensity transitions of the captured Gray coding images. As the period of the wrapped phase is the same as that of the fringe order, the phase jump points are where the fringe order changes, as illustrated in Fig. 1.10 (e). Thus, the unwrapped phase can be recovered by increasingly adding times of 2π at the locations where the fringe order changes, which is mathematically described as

$$\Phi(x, y) = \phi(x, y) + 2\pi k(x, y), \quad (1.13)$$

where $k(x, y)$ is the fringe order. Besides Gray coding, other phase unwrapping methods, including multi-frequency phase unwrapping (Cheng & Wyant, 1985), spatial phase unwrapping (M. Zhao et al., 2011), geometric-constraint-based phase unwrapping (An et al., 2016), deep-learning-based unwrapping (W. Li et al., 2021) and so forth, have been explored to retrieve the unwrapped phase.

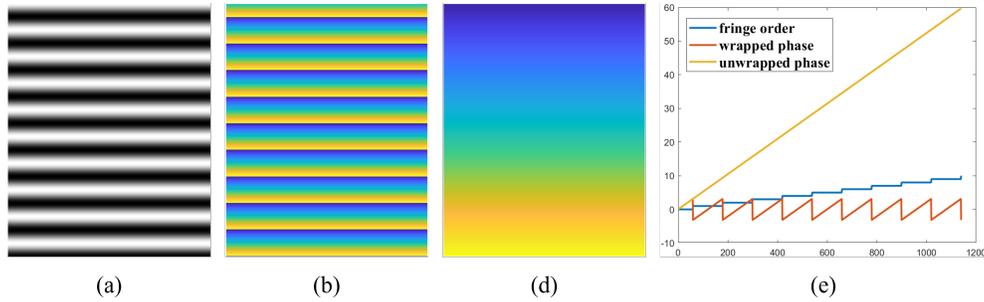


Figure 1.10: An example of phase unwrapping. (a) One of the 18-step sinusoidal fringe images. (b) Wrapped phase. (c) Unwrapped phase. (d). Cross section of wrapped phase, unwrapped phase, and fringe order to show phase unwrapping using the fringe order.

After obtaining the phase, system calibration is necessary to build the relationship between the phase and height. Generally, most of the widely used system calibration methods can be categorized into two groups: phase-height models and triangular stereo models (S. Feng et al., 2021). The

phase-height models directly build the relationship through linear fitting, linear inverse fitting, polynomial fitting, etc. In these methods, it is unnecessary to obtain the intrinsic and extrinsic parameters of the camera and the projector. Triangular stereo models are similar to conventional stereo vision models, which treat the projector as an inverse camera. Therefore, the camera and projector need calibration, and the pinhole camera model is widely used to depict them. For the camera, it can be described as

$$s^c \begin{bmatrix} u^c \\ v^c \\ 1 \end{bmatrix} = \begin{bmatrix} f_u^c & \gamma^c & u_0^c \\ 0 & f_v^c & v_0^c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11}^c & r_{12}^c & r_{13}^c & t_1^c \\ r_{21}^c & r_{22}^c & r_{23}^c & t_2^c \\ r_{31}^c & r_{32}^c & r_{33}^c & t_3^c \end{bmatrix} \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}. \quad (1.14)$$

In the above equation, rotation matrix $R^c = [r_{ij}^c]_{3 \times 3}$ and translation vector $\vec{t}^c = (t_1^c, t_2^c, t_3^c)^\top$ transform a point $p^w (x^w, y^w, z^w)$ in the world coordinate system to the point $p_1(x^c, y^c, z^c)$ in the camera coordinate system. The translation vector and rotation matrix consist of the extrinsic parameters. To simplify the calibration process, we usually align the world coordinate system with the camera coordinate system. Thus, the camera's rotation matrix is an identity matrix, and the translation vector is $\vec{t}^c = (0, 0, 0)^\top$. Following the transformation, the point p_1 is projected to 2D points (u^c, v^c) on the camera image plane through the intrinsic parameters, f_u^c , f_v^c , γ^c , and (u_0^c, v_0^c) . f_u^c and f_v^c are the focal lengths of the camera along u and v directions, γ^c is the skew factor, and (u_0^c, v_0^c) is the principal point on the 2D pixel coordinate. As the projector is regarded as an inverse camera, it can be similarly defined as

$$s^p \begin{bmatrix} u^p \\ v^p \\ 1 \end{bmatrix} = \begin{bmatrix} f_u^p & \gamma^p & u_0^p \\ 0 & f_v^p & v_0^p \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11}^p & r_{12}^p & r_{13}^p & t_1^p \\ r_{21}^p & r_{22}^p & r_{23}^p & t_2^p \\ r_{31}^p & r_{32}^p & r_{33}^p & t_3^p \end{bmatrix} \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}. \quad (1.15)$$

The camera calibration utilizes the widely recognized method by Zhang et al. (Z. Zhang, 2000). Unlike the camera, the projector cannot directly capture calibration features due to its different nature. The key to the system calibration using triangular stereo models hinges on making the

projector detect features like the camera. One widely used method is mapping the detected features in captured fringe images to the projected fringe patterns through one-to-one phase mapping (S. Zhang & Huang, 2006). Fig. 1.11 presents how the projector detects feature points using the phase cue. The calibrated camera detects the feature point $p(u^c, v^c)$, and the feature point's phase $\Phi_v(u^c, v^c)$ and $\Phi_h(u^c, v^c)$ can also be obtained through phase analysis. Normally, in the calibration process, we assume that the phase of the feature point $p(u^c, v^c)$ is equal to the phase of the correspondence feature point in the projected fringe patterns. The phase associated with the pixel index in fringe pattern images allows for the inference of the pixel index on the projector's fringe image plane. Consequently, this enables the projector to detect feature points, effectively calibrating the FPP system akin to a conventional stereo vision system. After acquiring the phase information and calibrating the FPP system, the object's 3D profile can be reconstructed.

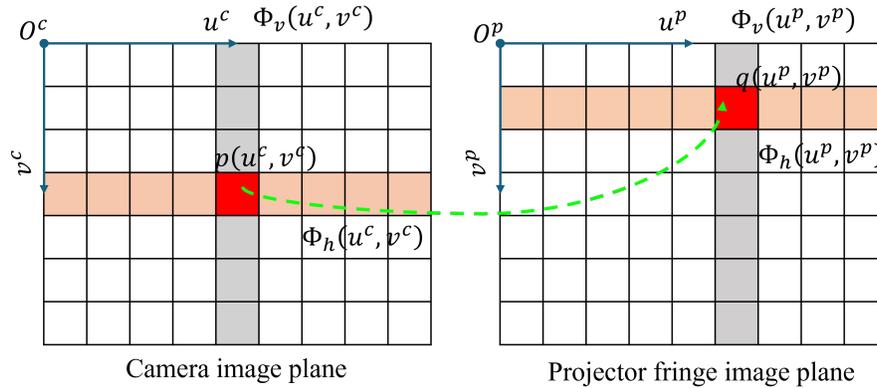


Figure 1.11: Schematic of how the projector detects feature points using the phase cue (adapted from (S. Feng et al., 2021)).

Generally, interference-based methods are primarily used for high-precision and microscale measurements, while Time-of-Flight techniques suit low-accuracy and large-scale measurements. FPP systems bridge these approaches, offering a balanced solution. However, FPP techniques are not off limitations and challenges. As an active measurement technique, occlusion by complex object surfaces brings shadows in captured fringe images, which causes the reconstruction to be incomplete and reduces measurement accuracy. The design of the fringe pattern also influences its measurement speed and accuracy. In addition, integrating 3D geometry with other data modalities can better reveal an object's characterization and broaden the application range. The subsequent

section introduces the principle of hyperspectral imaging, showcasing its potential integration with FPP to unveil the intrinsic properties of objects.

1.2.3 Hyperspectral imaging

Hyperspectral imaging (HSI) is a technique that generates a spatial map of spectral variation, making it widely applied in many fields, such as biomedical engineering (Fei, 2019), food quality assessment and control (ElMasry & Sun, 2010), agriculture (B. Lu et al., 2020), conservation of cultural relics (Pei et al., 2023), etc. In the wavelength domain, HSI is related to RGB imaging and multispectral imaging. Fig. 1.12 compares RGB, multispectral and hyperspectral features. Conventional RGB color imaging rarely reveals the physical and chemical properties. Multispectral imaging generally collects data in limited and relatively noncontiguous wide spectral bands, typically measured in micrometers or tens of micrometers. While HSI can capture images in hundreds of contiguous spectral bands with high spectral resolution, usually 10 nm or less. Besides, HSI techniques can acquire reflectance, absorption, or fluorescence spectrum for each pixel in the image (Q. Li et al., 2013). The hyperspectral images are usually presented in the form of a 3D data cube, where the cube faces show spatial information (spatial coordinates x and y), and the depth presents spectral information (wavelength λ).

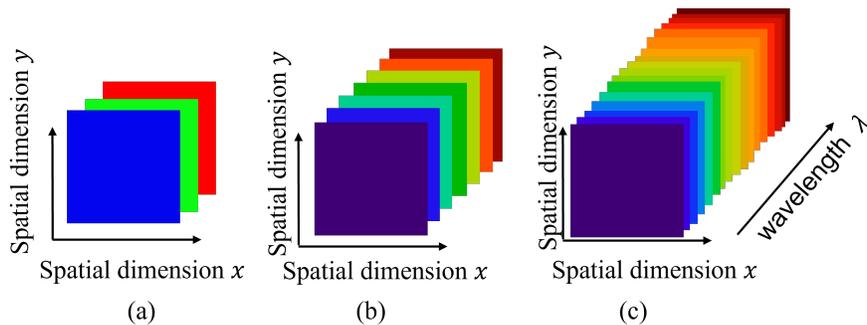


Figure 1.12: Comparison of RGB, multispectral, and hyperspectral features. (a) RGB imaging feature. (b) Multispectral imaging feature. (c) Hyperspectral imaging feature.

Fig.1.13 illustrates four typical hyperspectral approaches: the point scan, line scan, spectral scan, and snapshot. In point scan mode, a single spot of the object surface is scanned in each frame, and the reflected light is dispersed by a prism and captured by a linear detector to obtain spectral information of this spot. To scan the whole object and obtain the image data cube (x, y, λ) , the object or the

detector moves along the two spatial dimensions (x and y). Therefore, additional optomechanical devices are necessary, increasing the complexity of hardware configuration. Besides, the motion of the object or the detector makes the scanning time-consuming.

Fig. 1.13 (b) shows how the line-scan hyperspectral imaging works. Different from the point-scan mode, this method captures a 2D spectral image of an object's surface slit using a 2D matrix detector. During scanning, the detector or sample simultaneously moves perpendicular to the slit direction to collect the data cube. To ensure no distortion on the captured image, the synchronization between the object's or detector's motion and the detector's frame rate is crucial, increasing the complexity of implementation. Compared to the point-scan mode, the line-scan hyperspectral imager can capture more light, providing a longer exposure to the detector (Q. Li et al., 2013).

The point-scan and the line-scan HSI only capture a point or a slit of the object's surface. Unlike the mentioned point-scan and line-scan approaches, the spectral-scan HSI acquires a 2D image of the full object surface at a single spectral band using filters in each frame, as illustrated in Fig. 1.13 (c). By adjusting the filters, this technique can collect the hyperspectral data cube. Therefore, the spectral measurement range can be dynamic in the spectral-scan HSI, depending on how the filters are adjusted.

Fig. 1.13 (d) presents a schematic of the snapshot HSI, which is also named single-shot HSI. As indicated by the name, it can record both spatial and spectral information with one shot. Unlike the other three approaches that need to scan either in the spatial or spectral dimensions, the snapshot mode is an imaging technique without scanning. It can acquire a complete hyperspectral data cube by directly imaging the remapped and dispersed image zones onto a detector simultaneously, which enables direct data acquisition with minimal postprocessing for constructing a 3-D data cube (Q. Li et al., 2013). However, its spatial and spectral resolutions are constrained because the maximum voxels it can capture are limited by the CCD camera's pixel count. Consequently, increasing spatial sampling reduces spectral sampling and vice versa, depending on the CCD's capacity (L. Gao et al., 2009).

The richness and high dimensionality of spectral information obtained from hyperspectral imaging make it widely used in many applications (Boldrini et al., 2012). However, hyperspectral imagers are typically more costly and require more time to capture a sample's complete spectral and spatial data than conventional imaging systems.

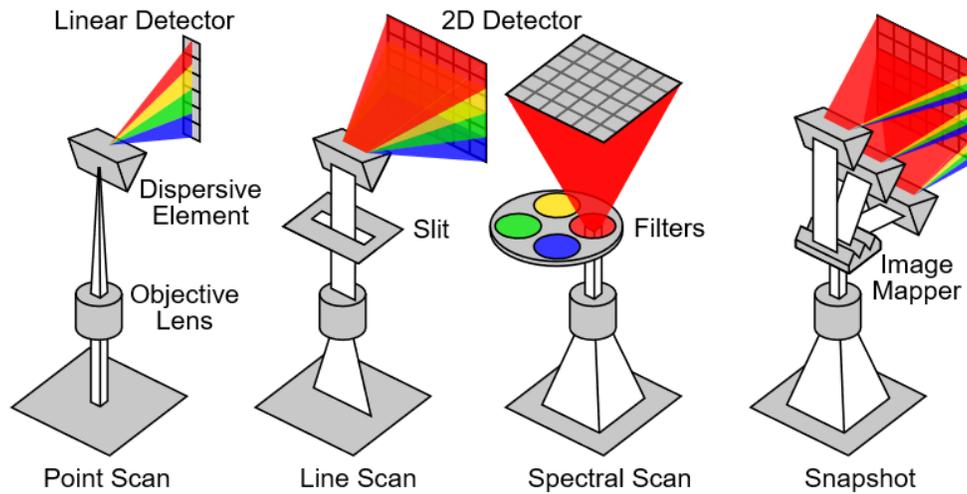


Figure 1.13: Typical hyperspectral imaging approaches (Lucasbosch, 2021).

1.3 Challenges and objectives

In the preceding sections, we have delineated the principal methodology of 3D reconstruction via fringe projection profilometry while also providing an overview of associated techniques for acquiring the 3D geometry of a sample. As an active triangulation-based method, fringe projection profilometry (FPP) exhibits robustness against noise, offering accuracy levels from millimeters to micrometers, and is apt for medium-scale applications (ranging from 0.01 m to 1 m). Despite its advantages, the technique faces numerous challenges. This dissertation is dedicated to addressing these challenges with the aim of enhancing the method's robustness and broadening its applications.

Fringe pattern direction and system calibration influence FPP's accuracy. The fringe pattern direction and system calibration are critical factors influencing the accuracy of FPP. The orientation of the fringe pattern directly affects the sensitivity of the 3D reconstruction process, as it determines the distribution and density of the fringes over the object's surface. On the other hand, system calibration ensures the accurate mapping of phase to height, which is a prerequisite for precise 3D reconstruction. Circular fringe patterns, as a new type of pattern design, have started to show some promise in FPP systems, as opposed to conventional linear horizontal or vertical patterns, particularly coaxial FPP systems. This dissertation introduces a novel circular fringe projection profilometry, including system calibration and 3D reconstruction techniques. It also evaluates the

performance of the proposed method against traditional linear FPP approaches. Details on the metrology will be discussed in Chapter 2.

Occlusion by complex object’s surface brings shadow-induced errors. In a typical FPP system, the optical axes of the projector and camera intersect, leading to unavoidable occlusions and shadows, especially in scenes with sharp objects’s surface changes. Phase information, the only cue for 3D reconstruction, cannot be recovered in the shadow areas, which leaves holes and failures in the final reconstructed 3D geometry. However, in the shadow area, texture images extracted from the captured fringe pattern images can still depict an object’s 3D shape, as demonstrated by the shape from the texture technique. Thus, texture images offer a potential avenue for enhancing 3D reconstructions in shadow areas. Recent advancements in deep learning have sparked interest in its application to FPP challenges, yielding considerable successes. This dissertation proposes a novel texture-guided phase-to-depth network designed to repair shadow-related errors in FPP by integrating texture and phase information to recover object depth accurately. The methodology of this research is elaborated in Chapter 2.

The integration of 3D and spectral imaging techniques better reveals an object’s characteristics. 3D spatial data and spectral images are two modalities that offer complementary insights into an object’s characteristics across different domains. 3D imaging depicts an object’s spatial attributes, including position, shape, size, surface texture, and roughness, while spectral imaging provides detailed information on the object’s physical and chemical properties, such as reflectance, absorption, and fluorescence. The integration of 3D and hyperspectral imaging techniques enhances the overall understanding of an object’s characteristics, thereby expanding their applicability in various fields. This dissertation introduces a novel 4D line-scan hyperspectral imaging system that captures spatial and spectral data simultaneously without data registration. The methodology of this research is further discussed in Chapter 4.

Decoupled system design enhances flexibility and performance in 4D hyperspectral imaging. Chapter 4 develops a 4D line-scan hyperspectral imaging system where 3D spatial data and hyperspectral images are captured simultaneously by a single camera. However, this coupled system design imposes limitations on the spectral range, restricting it to visible and near-infrared wavelengths, and constrains the scanning speed to 3 mm/min. By adopting a decoupled system design, where hyperspectral imaging and 3D scanning are conducted independently, we can significantly

improve system flexibility. This approach not only increases scanning speeds but also broadens the spectral range to visible-to-shortwave infrared (Vis-SWIR), which enhances the overall capabilities and applicability of 4D hyperspectral imaging systems. The methodology of the decoupled 4D Vis-SWIR line-scan hyperspectral imaging is introduced in Chapter 5.

1.4 Dissertation organization

Chapter 2 introduces a comparative analysis between circular and linear fringe projection profilometry. This chapter proposes a novel approach to circular fringe projection profilometry, providing an in-depth exploration of its performance compared to the conventional linear fringe projection profilometry (FPP). Chapter 3 delves into developing an innovative texture-guided phase-to-depth network to rectify shadow-induced errors commonly encountered in FPP. Chapter 4 explores 4D line-scan hyperspectral imaging covering the visible and near-infrared wavelength range. This chapter outlines the proposed system to obtain 3D geometry and hyperspectral images without data registration. Based on Chapter 4, Chapter 5 further develops a 4D visible-to-shortwave infrared line-scan hyperspectral imaging system, which has the ability to capture 3D spatial data and hyperspectral images with wavelength spinning from visible to shortwave infrared wavelength range. Chapter 6 summarizes the research contributions of this dissertation and simply introduces some valuable research directions to be investigated in the future.

CHAPTER 2

COMPARATIVE ANALYSIS OF CIRCULAR AND LINEAR FRINGE PROJECTION PROFILOMETRY: FROM CALIBRATION TO 3D RECONSTRUCTION¹

¹Li, Jiaqiong, and Beiwen Li. 2024. *Optics Continuum*, 3.3 (2024): 468-483.

Reprinted here with permission of the publisher.

2.1 Abstract

This study compares the accuracy of circular and linear fringe projection profilometry in the aspects of system calibration and 3D reconstruction. We introduce a novel calibration method and 3D reconstruction technique using circular and radial fringe patterns. Our approach is compared with the traditional linear phase-shifting method through several 2×2 experimental setups. Results indicate that our 3D reconstruction method surpasses the linear phase-shifting approach in performance, although calibration efficiency does not present a superior performance. Further analysis reveals that sensitivity and estimated phase error contribute to the relative underperformance in calibration. This paper offers insights into the potentials and limitations of circular fringe projection profilometry.

2.2 Introduction

Fringe projection profilometry (FPP), recognized for its high resolution, high accuracy, and robustness, has been widely applied in various fields for 3D measurement, including entertainment, manufacturing, robotics, and so on (X. Su & Zhang, 2010; J. Xu & Zhang, 2020b; S. Zhang, 2018; Zuo et al., 2018). The measurement accuracy of FPP is largely influenced by system calibration (S. Feng et al., 2021) and fringe pattern direction (Yu et al., 2020; R. Zhang et al., 2016).

The calibration methods for a typical fringe projection system consisting of a single camera and projector generally fall into phase-height mapping (Z. Zhang et al., 2011, 2013; W.-S. Zhou & Su, 1994), least-squares approach (L. Huang et al., 2010; Vo et al., 2012), and triangular stereo models (Z. Li et al., 2008; S. Zhang & Huang, 2006). This research primarily concentrates on triangular stereo models. A critical step of these models is enabling the projector to identify feature points like a camera. A prevalent approach is to map feature points detected by the camera to the projector with the assistance of phase. Up to now, many triangular stereo calibration methods (Z. Huang et al., 2015; S. Zhang & Huang, 2006; W. Zhang et al., 2017) have been investigated to obtain more precise calibration parameters. However, most of these methods rely on horizontal and vertical fringe patterns for phase acquisition and mapping of camera feature points onto the projector. Different from conventional horizontal and vertical fringe patterns for calibration, Li. et al. (B. Li & Zhang, 2014) recommended calibrating the system using the optimal fringe angle to

increase measurement accuracy. Given this, our research hypothesizes that calibrations employing circular and radial fringe patterns in circular FPP could yield superior outcomes.

Additionally, fringe pattern direction is another factor that impacts the measurement accuracy of FPP. Building upon epipolar geometry, Zhang. et al. (R. Zhang et al., 2016) formulated phase sensitivity as a function of the angle between the fringe direction and the epipolar line. They suggested orienting fringe patterns perpendicular to epipolar lines to maximize phase sensitivity and achieve better measurement accuracy. Further advancing this field, Zhang et al. (G. Zhang et al., 2023), utilizing extended epipolar geometry, introduced a novel approach of circular fringe projection profilometry and analyzed its 3D sensitivity. This method effectively circumvents the need to double pattern numbers while enabling real-time operations. However, it requires the generation of four distinct look-up tables and does not account for the effects of projector calibration. Mandapalli et al. (Mandapalli et al., 2021) introduced a radially symmetric circular fringe pattern for accurate unambiguous surface profiling of sudden height-discontinuous objects. This approach offered the capability to profile objects with a fourfold increase in dynamic range and at considerably lower fringe frequencies. The existing research demonstrates that circular fringe patterns have their advantages in phase sensitivity and suitability for profiling objects with abrupt depth changes, compared to conventional linear fringe patterns.

Furthermore, circular fringe patterns have been applied in uniaxial or parallel structured light systems using telecentric lenses. Zhao et al. (H. Zhao, Zhang, et al., 2016) proposed a CFPP method to determine the height of a point by calculating its distance from the optical center of a projector along the optical axis. However, this approach struggles with reconstructing the height at the nominal zero-phase point and has lower accuracy near this point. Addressing this challenge, Zhang et al. (C. Zhang et al., 2019) enhanced CFPP by employing a 2D ruler-based or plane constraint-based method to detect the zero-phase point, though this increases the method's complexity. Subsequently, Zhang et al. (J. Zhang et al., 2021) proposed a simpler approach that mitigates the zero-phase point issue and does not require calibrating the camera and projector in the system.

Given the advantages and flexibility of circular fringe patterns, this paper aims to develop and evaluate a novel CFPP method incorporating system calibration. First, we calibrate a structured light system using circular and radial fringe patterns, then formulate a corresponding 3D reconstruction method using circular fringe patterns. We compare our method's performance with Zhang and

Huang’s method (S. Zhang & Huang, 2006) and the linear phase-shifting method, conducting several 2×2 experiments on diverse samples for visual assessment. Our findings indicate that circular FPP outperforms traditional linear FPP, while our proposed calibration method exhibits slightly inferior performance compared to Zhang and Huang’s. In the discussion section, we delve into an in-depth analysis to understand the factors leading to this reduced accuracy in our calibration method. Through this analysis, we identify and elaborate upon the increasing trends in both sensitivity and estimated phase error. These factors are recognized as the key contributors to the diminished accuracy of our calibration method, providing vital insights for future enhancements in CFPP calibration techniques.

2.3 Principle of the method

This section will first explain how to calibrate an FPP system using circular and radial fringe patterns. Then, we will introduce the proposed circular fringe projection profilometry for 3D reconstruction.

2.3.1 System calibration

Pinhole imaging model

In a structured light system, both the camera and projector can be effectively characterized using a conventional pinhole imaging model, as depicted in Fig.2.1. The mathematical model of the pinhole imaging model can be explained as:

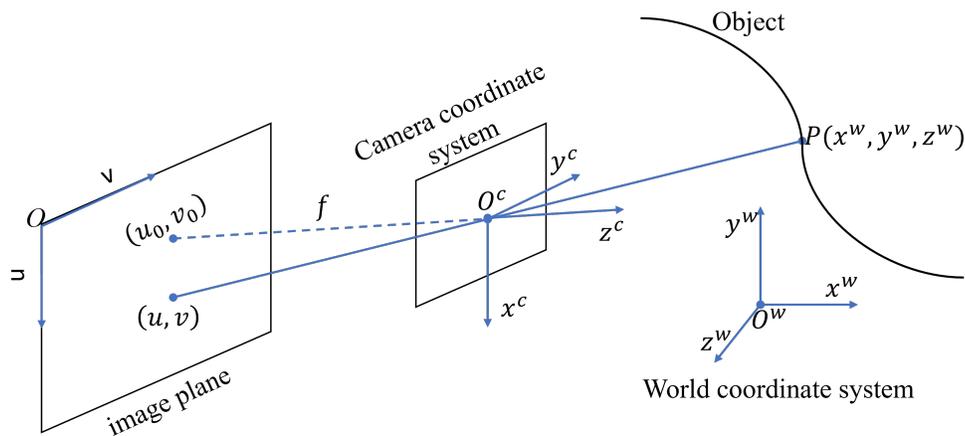


Figure 2.1: Pinhole imaging model. The figure is reprinted from (B. Li et al., 2014).

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & \gamma & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}. \quad (2.1)$$

In the model, rotation matrix $\mathbf{R} = [r_{ij}]_{3 \times 3}$ and translation vector $\vec{t} = (t_1, t_2, t_3)^\top$ transform a point $p^w (x^w, y^w, z^w)$ in the world coordinate system to a point $p^c (x^c, y^c, z^c)$ in the camera coordinate system, which constitutes the extrinsic parameters of the system. Following the transformation, the point $p^c (x^c, y^c, z^c)$ is projected to a 2D point (u, v) on the image plane within the intrinsic parameters, f_u, f_v, γ , and (u_0, v_0) . f_u and f_v are the focal lengths of the camera along u and v directions, γ is the skew factor, and (u_0, v_0) is the principal point on the 2D pixel coordinate. The intrinsic and extrinsic parameters can be further represented with a projection matrix \mathbf{M} ,

$$\mathbf{M} = \begin{bmatrix} f_u & \gamma & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_1 \\ m_{21} & m_{22} & m_{23} & m_2 \\ m_{31} & m_{32} & m_{33} & m_3 \end{bmatrix}. \quad (2.2)$$

The model of the camera can be described as

$$s^c \begin{bmatrix} u^c \\ v^c \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^c & m_{12}^c & m_{13}^c & m_{14}^c \\ m_{21}^c & m_{22}^c & m_{23}^c & m_{24}^c \\ m_{31}^c & m_{32}^c & m_{33}^c & m_{34}^c \end{bmatrix} \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}. \quad (2.3)$$

Similarly, we consider the projector in a structured light system as an inverse-imaging camera (S. Zhang & Huang, 2006), whose mathematical formulation is described as

$$s^p \begin{bmatrix} u^p \\ v^p \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^p & m_{12}^p & m_{13}^p & m_{14}^p \\ m_{21}^p & m_{22}^p & m_{23}^p & m_{24}^p \\ m_{31}^p & m_{32}^p & m_{33}^p & m_{34}^p \end{bmatrix} \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}. \quad (2.4)$$

Camera calibration

In this research, we employed Zhang’s renowned calibration method (Z. Zhang, 2000) in conjunction with the OpenCV camera calibration software toolbox (Bradski, 2000) to calibrate the camera. Fig.2.2(a) illustrates the design of our calibration target, where the centers of the circles act as the feature points. Camera calibration involves two primary components: intrinsic and extrinsic calibrations. The intrinsic calibration of the camera aims to estimate its internal parameters, f_u , f_v , γ , and (u_0, v_0) . This process involves capturing images of the calibration target in various poses (an example is depicted in Fig.2.2(b)). For each image, feature points (the centers of circles) are extracted and then used in an iterative optimization process to refine the estimation of the camera’s intrinsic parameters, utilizing the OpenCV camera calibration toolbox.

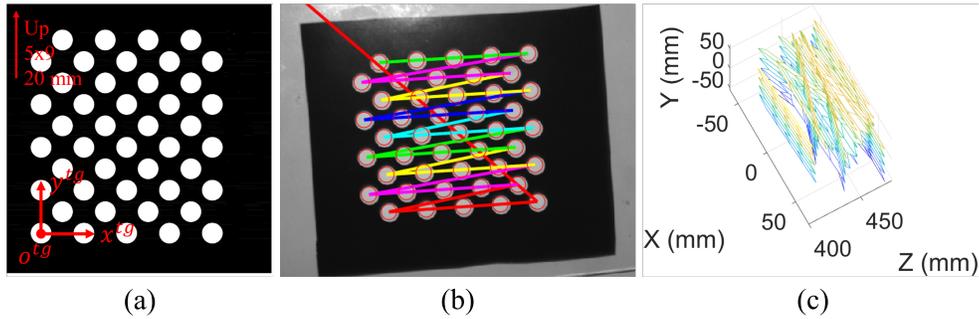


Figure 2.2: Camera calibration and target 3D estimation. (a) asymmetric circle grid calibration target, whose size is $(5(\text{row}) \times 9(\text{column}))$ with 20 mm center-to-center distance); (b) an image of the target at a pose with extracted circle centers; (c) the estimated 3D target orientations.

Extrinsic calibration primarily estimates the rotation (r_{ij}^{tg}) and translation (t_i^{tg}) parameters, which transform points from the planar target coordinate $(x^{tg}, y^{tg}, 0)$ to the camera coordinate (x^c, y^c, z^c) . Fig.2.2(a) defines the planar target coordinate, with the principal point being at the bottom left circle center. Having aligned the world coordinate with the camera coordinate, the rotation (r_{ij}^{tg}) and translation (t_i^{tg}) parameters convert the planar target points $(x^{tg}, y^{tg}, 0)$ into three-dimensional points $(x^{wtg}, y^{wtg}, z^{wtg})$ within the world coordinate system, as detailed in Equation 2.5. Additionally, Figure 2.2(c) illustrates the estimated 3D orientations of the target in various poses.

$$\begin{bmatrix} x^{wtg} \\ y^{wtg} \\ z^{wtg} \end{bmatrix} = \begin{bmatrix} r_{11}^{tg} & r_{12}^{tg} & r_{13}^{tg} & t_1^{tg} \\ r_{21}^{tg} & r_{22}^{tg} & r_{23}^{tg} & t_2^{tg} \\ r_{31}^{tg} & r_{32}^{tg} & r_{33}^{tg} & t_3^{tg} \end{bmatrix} \begin{bmatrix} x^{tg} \\ y^{tg} \\ 0 \\ 1 \end{bmatrix}. \quad (2.5)$$

Circular and radial fringe patterns for projector calibration

In calibrating a structured light system, which consists of a camera and a projector, the latter can be treated as an inverse camera. Thus, system calibration hinges on enabling the projector to capture feature points in a manner like a camera. To this end, phase is employed to establish a correspondence between feature points captured by the camera and those projected (S. Feng et al., 2021). Zhang and Huang's method (S. Zhang & Huang, 2006) utilized horizontal and vertical fringe patterns for correspondence mapping. Different from this method, we utilized circular and radial fringe patterns with their center coordinated at the specific point $O^s (u_s, v_s)$ to build the correspondence. An example of circular and radial fringe patterns is illuminated in Fig.2.3. The circular and radial fringe patterns can be generated based on Eq.2.6 and Eq.2.8.

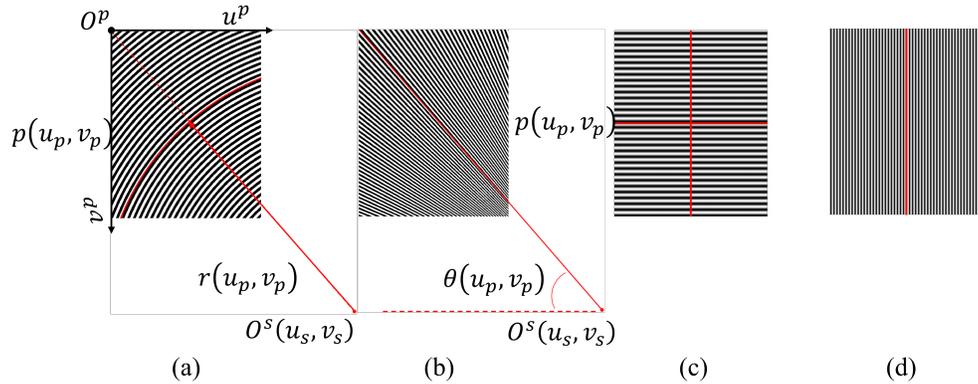


Figure 2.3: An example of fringe patterns for system calibration. (a)-(b). A circular and radial fringe pattern centered at (u_s, v_s) . $r(u_p, v_p)$ is the distance between the point $p(u_p, v_p)$ and the center $O^s(u_s, v_s)$, and $\theta(u_p, v_p)$ is the pixel angle. (c)-(d). A horizontal and vertical fringe pattern employed in Zhang and Huang's method(S. Zhang & Huang, 2006).

$$I_i^c(u_p, v_p) = A + B \cos\left(\frac{2\pi}{T}r(u_p, v_p) + 2\pi\frac{i-1}{N}\right), \quad (2.6)$$

where I_i^c is the grayscale intensity of the pixel in i^{th} circular fringe patterns at a coordinate (u_p, v_p) in the projector image plane, A and B are two constants set to 127.5 for 8-bit gray depth, T is the spatial period in the unit of pixel, i and N are the index and total number of the phase shifts, and $r(u_p, v_p)$ is the radial distance between pixel (u_p, v_p) and center (u_s, v_s) , as illuminated in Eq.2.7,

$$r(u_p, v_p) = \sqrt{(u_p - u_s)^2 + (v_p - v_s)^2}. \quad (2.7)$$

Similarly, we generated radial fringe patterns as described in Eq.2.8,

$$I_i^r(u_p, v_p) = A + B \cos\left(\frac{2\pi}{\Delta\theta}\theta(u_p, v_p) + 2\pi\frac{i-1}{N}\right), \quad (2.8)$$

where $\Delta\theta$ is the angle change per spatial period in the unit of rad, and $\theta(u_p, v_p)$ is the pixel angle between pixel (u_p, v_p) and center (u_s, v_s) , as given by

$$\theta(u_p, v_p) = \begin{cases} \arctan\left(\frac{v_p - v_s}{u_p - u_s}\right) & \text{if } u_p < u_s, \\ \frac{\pi}{2} & \text{if } u_p = u_s, \\ \pi - \arctan\left(\frac{v_p - v_s}{u_p - u_s}\right) & \text{otherwise.} \end{cases} \quad (2.9)$$

Besides, we binarized the circular and radial fringe patterns based on the threshold of average grayscale intensity, as shown in Eq.2.10,

$$I_i^b(u_p, v_p) = \begin{cases} 255 & \text{if } I_i(u_p, v_p) > I_{avg}, \\ 0 & \text{otherwise.} \end{cases} \quad (2.10)$$

For circular and radial fringe patterns, the relationship of the circular phase Φ_c and radial phase Φ_r with radial distance $r(u_p, v_p)$ and pixel angle $\theta(u_p, v_p)$ can be described as

$$\begin{cases} r(u_p, v_p) = \frac{T}{2\pi}\Phi_c, \\ \theta(u_p, v_p) = \frac{\Delta\theta}{2\pi}\Phi_r. \end{cases} \quad (2.11)$$

The phase of circular and radial fringe patterns Φ_c and Φ_r can be obtained using the phase-shift method (Malacara, 2007a) and Gray coding unwrapping technique (Y. Wang et al., 2011a) based on the captured fringe images from the camera. Thus, we can identify the corresponding feature points on the projector's image plane through Eq.2.12,

$$\begin{cases} u^p = -\frac{T}{2\pi}\Phi_c \cos\left(\frac{\Delta\theta}{2\pi}\Phi_r\right) + u_s, \\ v^p = -\frac{T}{2\pi}\Phi_c \sin\left(\frac{\Delta\theta}{2\pi}\Phi_r\right) + v_s. \end{cases} \quad (2.12)$$

Once we identify the projector feature points, we will build the correspondence between feature points on the projector image plane and the camera plane. Then, we utilize OpenCV library (Bradski, 2000) for stereo calibration to obtain intrinsic parameters of the camera and projector and the rotation matrix and translation vector of the projector to the camera.

2.3.2 Circular FPP for 3D reconstruction

In reconstructing the 3D geometry of samples, we only utilized binarized circular fringe patterns. Combing Eq.2.4, Eq.2.11 and Eq.2.12, we get the relationship between phase Φ_c and projection parameters $\mathbf{M} = [m_{ij}]_{3 \times 4}$,

$$\begin{cases} u^p = -\frac{T}{2\pi}\Phi_c \cos\theta + u_s = \frac{m_{11}^p x^w + m_{12}^p y^w + m_{13}^p z^w + m_{14}^p}{m_{31}^p x^w + m_{32}^p y^w + m_{33}^p z^w + m_{34}^p}, \\ v^p = -\frac{T}{2\pi}\Phi_c \sin\theta + v_s = \frac{m_{21}^p x^w + m_{22}^p y^w + m_{23}^p z^w + m_{24}^p}{m_{31}^p x^w + m_{32}^p y^w + m_{33}^p z^w + m_{34}^p}, \end{cases} \quad (2.13)$$

We simplify the above equation and rewrite it as

$$\begin{aligned} \left(\frac{T}{2\pi}\right)^2 \Phi_c^2 \left(m_{31}^p x^w + m_{32}^p y^w + m_{33}^p z^w + m_{34}^p\right)^2 &= k_1 x^{w^2} + k_2 y^{w^2} + k_3 z^{w^2} \\ &+ k_4 + k_5 x^w y^w + k_6 x^w z^w \\ &+ k_7 y^w z^w + k_8 x^w + k_9 y^w + k_{10} z^w, \end{aligned} \quad (2.14)$$

where parameters k_1, \dots, k_{10} are listed below that can be directly obtained based on the projector's calibration parameters. Though the expressions of parameters k_i are complex, they are actually constants once the projector is calibrated and the center of the circular fringe patterns is fixed.

$$\left\{ \begin{array}{l} k_1 = (m_{11}^p - u_s m_{31}^p)^2 + (m_{21}^p - v_s m_{31}^p)^2, \\ k_2 = (m_{12}^p - u_s m_{32}^p)^2 + (m_{22}^p - v_s m_{32}^p)^2, \\ k_3 = (m_{13}^p - u_s m_{33}^p)^2 + (m_{23}^p - v_s m_{33}^p)^2, \\ k_4 = (m_{14}^p - u_s m_{34}^p)^2 + (m_{24}^p - v_s m_{34}^p)^2, \\ k_5 = 2(m_{11}^p - u_s m_{31}^p)(m_{12}^p - u_s m_{32}^p) + 2(m_{21}^p - v_s m_{31}^p)(m_{22}^p - v_s m_{32}^p), \\ k_6 = 2(m_{11}^p - u_s m_{31}^p)(m_{13}^p - u_s m_{33}^p) + 2(m_{21}^p - v_s m_{31}^p)(m_{23}^p - v_s m_{33}^p), \\ k_7 = 2(m_{12}^p - u_s m_{32}^p)(m_{13}^p - u_s m_{33}^p) + 2(m_{22}^p - v_s m_{32}^p)(m_{23}^p - v_s m_{33}^p), \\ k_8 = 2(m_{11}^p - u_s m_{31}^p)(m_{14}^p - u_s m_{34}^p) + 2(m_{21}^p - v_s m_{31}^p)(m_{24}^p - v_s m_{34}^p), \\ k_9 = 2(m_{12}^p - u_s m_{32}^p)(m_{14}^p - u_s m_{34}^p) + 2(m_{22}^p - v_s m_{32}^p)(m_{24}^p - v_s m_{34}^p), \\ k_{10} = 2(m_{13}^p - u_s m_{33}^p)(m_{14}^p - u_s m_{34}^p) + 2(m_{23}^p - v_s m_{33}^p)(m_{24}^p - v_s m_{34}^p). \end{array} \right. \quad (2.15)$$

Eq.2.14 can be further rewritten as,

$$\vec{l}\vec{x} = 1, \quad (2.16)$$

where \vec{x} is

$$\vec{x} = [\Phi_c^2 x^2, \Phi_c^2 y^2, \Phi_c^2 z^2, \Phi_c^2, \Phi_c^2 xy, \Phi_c^2 xz, \Phi_c^2 yz, \Phi_c^2 x, \Phi_c^2 y, \Phi_c^2 z, x^2, y^2, z^2, xy, xz, yz, x, y, z]^\top, \quad (2.17)$$

and \vec{l} is

$$\vec{l} = [l_1, l_2, l_3, \dots, l_{19}], \quad (2.18)$$

$$\left\{ \begin{array}{llll} l_1 = \left(\frac{T}{2\pi}\right)^2 \frac{m_{31}^p{}^2}{k_4} & l_2 = \left(\frac{T}{2\pi}\right)^2 \frac{m_{32}^p{}^2}{k_4} & l_3 = \left(\frac{T}{2\pi}\right)^2 \frac{m_{33}^p{}^2}{k_4} & l_4 = \left(\frac{T}{2\pi}\right)^2 \frac{m_{34}^p{}^2}{k_4}, \\ l_5 = 2 \left(\frac{T}{2\pi}\right)^2 \frac{m_{31}^p m_{32}^p}{k_4} & l_6 = 2 \left(\frac{T}{2\pi}\right)^2 \frac{m_{31}^p m_{33}^p}{k_4} & l_7 = 2 \left(\frac{T}{2\pi}\right)^2 \frac{m_{32}^p m_{33}^p}{k_4}, & \\ l_8 = 2 \left(\frac{T}{2\pi}\right)^2 \frac{m_{31}^p m_{34}^p}{k_4} & l_9 = 2 \left(\frac{T}{2\pi}\right)^2 \frac{m_{32}^p m_{34}^p}{k_4} & l_{10} = 2 \left(\frac{T}{2\pi}\right)^2 \frac{m_{33}^p m_{34}^p}{k_4}, & \\ l_{11} = -\frac{k_1}{k_4} & l_{12} = -\frac{k_2}{k_4} & l_{13} = -\frac{k_3}{k_4}, & \\ l_{14} = -\frac{k_5}{k_4} & l_{15} = -\frac{k_6}{k_4} & l_{16} = -\frac{k_7}{k_4}, & \\ l_{17} = -\frac{k_8}{k_4} & l_{18} = -\frac{k_9}{k_4} & l_{19} = -\frac{k_{10}}{k_4}. & \end{array} \right. \quad (2.19)$$

One point to stress is that the parameters l_1, \dots, l_{19} are all constants once we calibrate the projector, though their expressions are complicated. After calculating the vector $\vec{l} = [l_1, \dots, l_{19}]$, we can formulate the 3D reconstruction by combing Eq.2.3 as

$$\mathbf{A}\vec{x}' = \vec{b}, \quad (2.20)$$

where the reconstruction matrix \mathbf{A} consists of

$$\mathbf{A} = [\vec{a}_1, \vec{a}_2, \vec{a}_3], \quad (2.21)$$

$$\vec{a}_1 = [m_{11}^c - u^c m_{31}^c, m_{12}^c - u^c m_{32}^c, m_{13}^c - u^c m_{33}^c, 0, \dots, 0], \quad (2.22)$$

$$\vec{a}_2 = [m_{21}^c - v^c m_{31}^c, m_{22}^c - v^c m_{32}^c, m_{23}^c - v^c m_{33}^c, 0, \dots, 0], \quad (2.23)$$

$$\begin{aligned} \vec{a}_3 = & \left[l_8 \Phi_c^2 + l_{17}, l_9 \Phi_c^2 + l_{18}, l_{10} \Phi_c^2 + l_{19}, \right. \\ & l_1 \Phi_c^2 + l_{11}, l_2 \Phi_c^2 + l_{12}, l_3 \Phi_c^2 + l_{13}, \\ & \left. l_5 \Phi_c^2 + l_{14}, l_6 \Phi_c^2 + l_{15}, l_7 \Phi_c^2 + l_{16} \right], \end{aligned} \quad (2.24)$$

The target \vec{x}' is rewritten as

$$\vec{x}' = \left[x^w, y^w, z^w, x^{w2}, y^{w2}, z^{w2}, x^w y^w, x^w z^w, y^w z^w \right]^T, \quad (2.25)$$

and

$$\vec{b} = \left[u^c m_{34}^c - m_{14}^c, v^c m_{34}^c - m_{24}^c, 1 - l_4 \Phi_c^2 \right]^T. \quad (2.26)$$

To solve Eq. 2.20 and obtain $[x^w, y^w, z^w]$, we first utilize MATLAB Symbolic Toolbox(The MathWorks, 2019) to get the symbolic expression and then calculate the numeric solution based on the obtained phase Φ_c , vector \vec{l} , and projection parameters.

2.4 Experimental results

2.4.1 Experimental setup

We set up a structured light system as illuminated in Fig.2.4. A digital complementary-metal-oxide-semiconductor (CMOS) camera (model: FLIR, Grasshopper3 GS3-U3-41C6C-C) and a digital-light-processing (DLP) projector (model: Texas Instruments, DLP Lightcrafter 4500) are employed to capture and project fringe images. The projector resolution is set as 1140×912 pixels, while the camera resolution is 960×1280 pixels. The camera is attached with a lens of 8 mm focal length (model: Computar M0814-MP2).

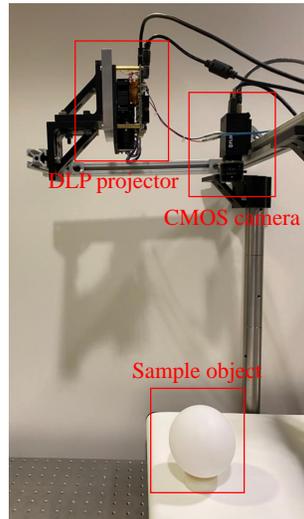


Figure 2.4: Experiment setup.

2.4.2 Comparison of measurement accuracy

To compare the performance of the proposed method with Zhang and Huang's calibration, we calibrate the structured light system by employing conventional horizontal and vertical fringe patterns. Fig.2.3 has already depicted fringe patterns employed in our calibration process: circular, radial, horizontal, and vertical fringe patterns. We both utilize 18-step phase-shifting with seven gray-coding for the two calibration methods to get phase information, and the spatial period of circular, horizontal, and vertical fringe patterns are 32, 36, and 18, respectively. $\Delta\theta$ of the radial fringe pattern is set at $2\pi/500$. The triangulation errors are estimated by comparing the target points $(x^{wtg}, y^{wtg}, z^{wtg})$ and the triangulation points using Zhang and Huang's method and the proposed method, and we also overlay the target points and triangulation points. The details of the results are illuminated in Fig.2.5. The root-mean-square (RMS) error for X, Y, and Z are $0.0024mm$, $0.0016mm$ and $0.2413mm$ for Zhang and Huang's method, and $0.0033mm$, $0.0019mm$ and $0.3401mm$ for our proposed method. Considering the overall calibration volumes ($100mm(X) \times 100mm(Y) \times 40mm(Z)$), the triangulation errors of the two methods are less than 1%, though the triangulation errors of our proposed method are a little bit larger than that of Zhang and Huang's method. The result of the proposed method being a little bit worse than Zhang and Huang's method will be discussed in the later discussion section. In this research, we also evaluate the measurement accuracies of our proposed calibration and 3D reconstruction method using a standard sphere with a radius of 49.975 mm. To compare these accuracies with Zhang and Huang's calibration method and linear FPP, we conducted a 2×2 experimental setup. This involved reconstructing the 3D profile of the standard sphere using either the proposed circular FPP or linear FPP, combined with calibration parameters derived from either our proposed calibration method or Zhang and Huang's method. Fig.2.6(a)-(d) displays the overlays of the standard sphere with the 3D reconstructions. Specifically, (a) represents the results using our proposed calibration and circular FPP; (b) shows the results when the proposed circular FPP method is paired with calibration parameters from Zhang and Huang's method; (c) illustrates the results of linear FPP combined with calibration parameters from our proposed method; and (d) depicts the results of linear FPP with calibration parameters from Zhang and Huang's method. Error maps corresponding to Fig.2.6 (a)-(e) are presented in Fig.2.6 (e)-(f), showing the absolute difference between the standard sphere and 3D reconstructions. The mean-absolute error (MAE) of the error maps are

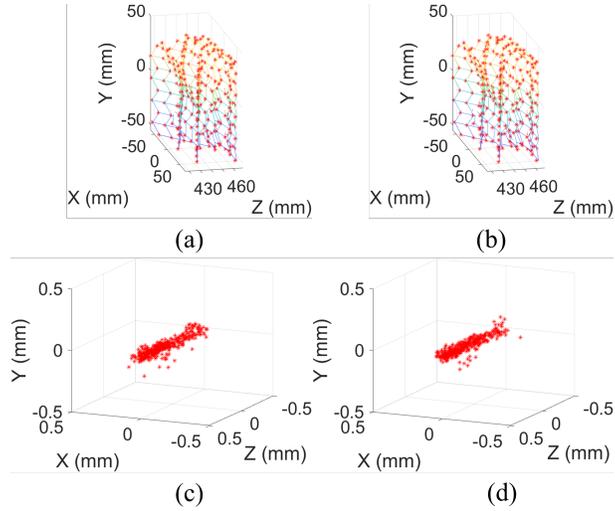


Figure 2.5: Evaluation of triangulation error. (a) Overlay of the target points (x^{wtg} , y^{wtg} , z^{wtg}) (red points) and triangulation points (x^w , y^w , z^w) (corners of mesh) using Zhang and Huang's method; (b) Similar overlay of target points and triangulation points using the proposed method; (c) Error between the target points and triangulation points using Zhang and Huang's method, and RMS error for X, Y and Z are 0.0024 mm, 0.0016 mm and 0.2413 mm; (d) Similar error using the proposed method, and RMS error for X, Y, and Z are 0.0033 mm, 0.0019 mm and 0.3401 mm.

0.2056mm, 0.1562mm, 0.2646mm, and 0.2402mm, respectively, with RMS errors of 0.2965mm, 0.2158mm, 0.6663mm, and 0.6525mm. Notably, when the proposed 3D reconstruction method is used with Zhang and Huang's calibration parameters, the MAE and RMS errors are the lowest among the results. Comparisons between Fig.2.6(e)-(h) reveal that when the calibration method is fixed, the proposed circular FPP yields lower RMS errors and MAEs than linear FPP. Conversely, using the same 3D reconstruction method with different calibration parameters shows that Zhang and Huang's method results in smaller RMS errors and MAEs compared to our proposed method, as evidenced by Fig.2.6(c), (f), (g), and (h). The reasons for Zhang and Huang's method outperforming the proposed calibration method will be discussed in the discussion section.

2.4.3 Test cases

We initially applied a similar 2x2 experimental design to a sample of a GPU board with a fan, with the 3D reconstruction results shown in Figure 2.7. The subfigures Fig.2.7(c)-(d) and (f)-(g) represent the outcomes obtained using various combinations of calibration and reconstruction methods: (c)

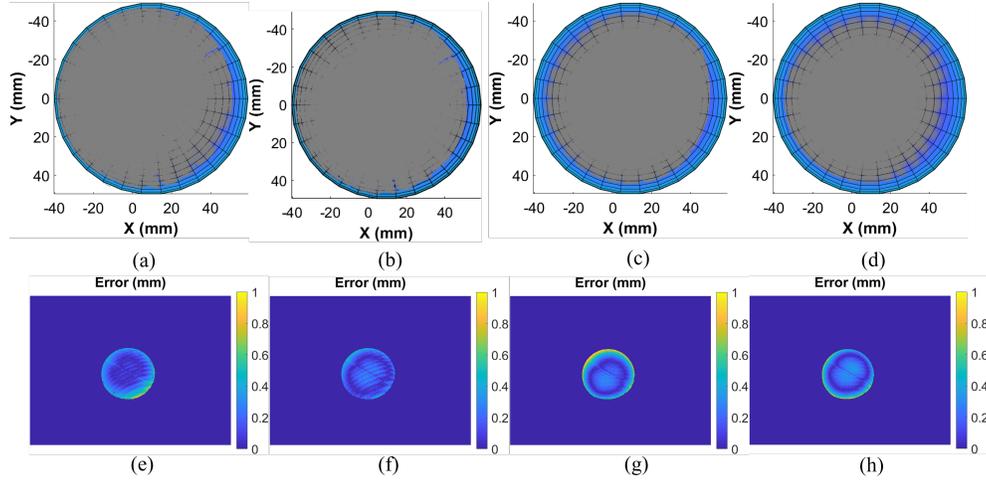


Figure 2.6: Evaluating measurement accuracy of a standard sphere (radius: 49.975 mm): comparative overlays and error analysis. (a)-(d) comparative overlays of the standard sphere and 3D reconstructions: (a) with our proposed calibration and circular FPP; (b) Zhang and Huang's calibration method with our proposed circular FPP; (c) our proposed calibration method and linear FPP; (d) Zhang and Huang's calibration method and linear FPP. (e)-(h) error maps corresponding to (a)-(d), showcasing the deviation from the standard sphere. Mean absolute errors are (e) 0.2056mm , (f) 0.1562mm , (g) 0.2646mm , and (h) 0.2402mm . RMS errors are (e) 0.2965mm , (f) 0.2158mm , (g) 0.6663mm , and (h) 0.6525mm .

employs our proposed calibration and circular FPP; (d) combines the proposed circular FPP with Zhang and Huang's calibration; (e) uses linear FPP with our proposed calibration; and (f) applies linear FPP with Zhang and Huang's calibration. Visually, we observed no significant differences among these 3D reconstruction results.

Additionally, we extended the 2×2 experimental setup to a sample of a GPU cooler with abrupt depth changes in metal fins. The 3D reconstruction results are presented in Figure 2.8. Similarly, Fig.2.8 (c)-(d) and (f)-(g) are results by utilizing the proposed calibration and circular FPP, the proposed circular FPP with Zhang and Huang's calibration, linear FPP with the proposed calibration, and linear FPP with Zhang and Huang's calibration, respectively. One to stress is that same-size median filters are deployed to remove noises in the process of phase unwrapping. A notable observation from Fig.2.8 (c)-(d) is that the results using our proposed circular FPP exhibit fewer noises in the fin area compared to those obtained using linear FPP (Fig.2.8 (f)-(g)). The

reconstruction outcomes of the GPU cooler sample demonstrate that our proposed method can perform on par with or even surpass linear FPP in certain aspects.

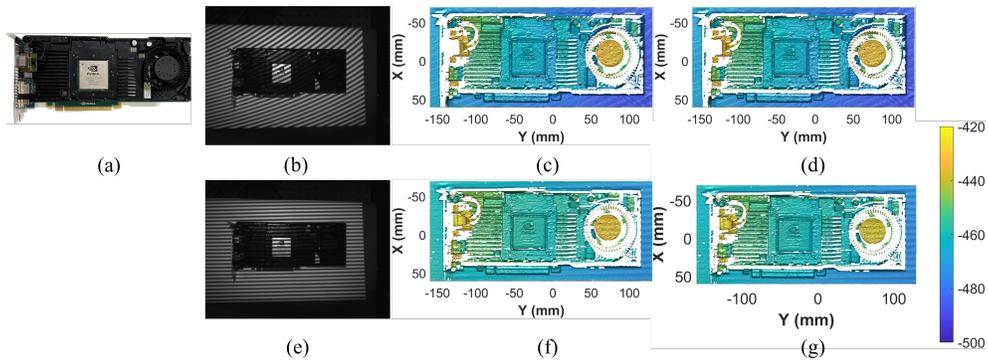


Figure 2.7: 3D reconstruction results of a GPU board with the fan. (a) the GPU board with fan; (b) a circular fringe image of the proposed circular FPP; (c)-(d) 3D reconstruction results using the proposed circular FPP with either the proposed calibration or Zhang and Huang’s method; (e) a horizontal fringe image of linear FPP; (f)-(g) 3D reconstruction results using linear FPP with either the proposed calibration or Zhang and Huang’s method.

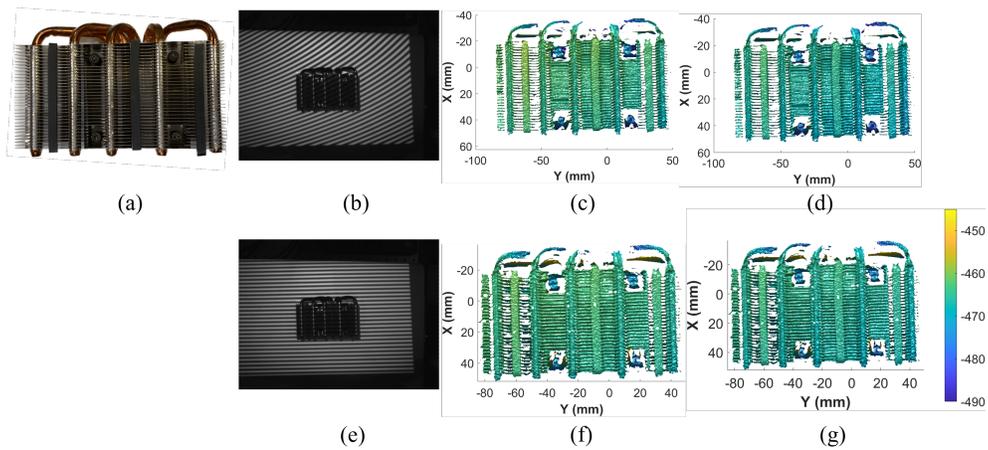


Figure 2.8: 3D reconstruction results of a GPU cooler. (a) the GPU cooler; (b) a circular fringe image of the proposed circular FPP; (c)-(d) 3D reconstruction results using the proposed circular FPP with either the proposed calibration or Zhang and Huang’s method; (e) a horizontal fringe image of linear FPP; (f)-(g) 3D reconstruction results using linear FPP with either the proposed calibration or Zhang and Huang’s method.

2.5 Discussion

When comparing the triangular error and the standard sphere error, the performance of our proposed calibration method slightly lags behind that of Zhang and Huang's approach. We will explore the underlying reasons for this, focusing on sensitivity and error analysis.

2.5.1 Sensitivity analysis of the projector calibration

The fringe pattern is the only difference between our proposed calibration method and Zhang and Huang's approach, bringing the difference in the mapping relationship between camera and projector feature points. Based on the relationship of phase and projector feature points of circular and fringe patterns (Eq.2.12), we define the sensitivity of the projector feature points relative to the phase. For the sensitivity along the u axis, denoted as $\mathbf{S}_{\mathbf{u},\text{proposed}}^{\mathbf{P}}$, we define it as:

$$\mathbf{S}_{\mathbf{u},\text{proposed}}^{\mathbf{P}} = \left(\frac{\partial u^{\mathbf{P}}}{\partial \Phi_c}, \frac{\partial u^{\mathbf{P}}}{\partial \Phi_r} \right) = \frac{T}{2\pi} \left(-\cos\left(\frac{\Delta\theta}{2\pi}\Phi_r\right), \frac{\Delta\theta}{2\pi}\Phi_c \sin\left(\frac{\Delta\theta}{2\pi}\Phi_r\right) \right). \quad (2.27)$$

where $\frac{\partial u^{\mathbf{P}}}{\partial \Phi_c}$ and $\frac{\partial u^{\mathbf{P}}}{\partial \Phi_r}$ represent the partial derivatives of the projector feature points in the u direction with respect to the circular (Φ_c) and radial (Φ_r) phases, respectively. Similarly, for the sensitivity along the v axis, denoted as $\mathbf{S}_{\mathbf{v},\text{proposed}}^{\mathbf{P}}$, it is given by:

$$\mathbf{S}_{\mathbf{v},\text{proposed}}^{\mathbf{P}} = \left(\frac{\partial v^{\mathbf{P}}}{\partial \Phi_c}, \frac{\partial v^{\mathbf{P}}}{\partial \Phi_r} \right) = \frac{T}{2\pi} \left(-\sin\left(\frac{\Delta\theta}{2\pi}\Phi_r\right), -\frac{\Delta\theta}{2\pi}\Phi_c \cos\left(\frac{\Delta\theta}{2\pi}\Phi_r\right) \right). \quad (2.28)$$

$\mathbf{S}_{\mathbf{u},\text{proposed}}^{\mathbf{P}}$ and $\mathbf{S}_{\mathbf{v},\text{proposed}}^{\mathbf{P}}$ are vectors, and we can have:

$$\begin{cases} \|\mathbf{S}_{\mathbf{u},\text{proposed}}^{\mathbf{P}}\| = \frac{T}{2\pi} \sqrt{\cos^2\left(\frac{\Delta\theta}{2\pi}\Phi_r\right) + \left(\frac{\Delta\theta}{2\pi}\Phi_c \sin\left(\frac{\Delta\theta}{2\pi}\Phi_r\right)\right)^2}, \\ \|\mathbf{S}_{\mathbf{v},\text{proposed}}^{\mathbf{P}}\| = \frac{T}{2\pi} \sqrt{\sin^2\left(\frac{\Delta\theta}{2\pi}\Phi_r\right) + \left(\frac{\Delta\theta}{2\pi}\Phi_c \cos\left(\frac{\Delta\theta}{2\pi}\Phi_r\right)\right)^2}. \end{cases} \quad (2.29)$$

In circular fringe patterns, the phase Φ_c increases progressively in the radial direction. Consequently, this increase in Φ_c directly influences the magnitudes of the sensitivity vectors $\|\mathbf{S}_{\mathbf{u},\text{proposed}}^{\mathbf{P}}\|$ and $\|\mathbf{S}_{\mathbf{v},\text{proposed}}^{\mathbf{P}}\|$ as a pixel moves outward from the center of the pattern. This radial increase in

sensitivity reflects that deviations in projector feature points are more pronounced with increasing distance from the central point.

For Zhang and Huang's method, horizontal and vertical fringe patterns are employed for calibration, so the phase is linear to pixel index (u, v) , and a similar definition of sensitivity in relation to phase can be given by:

$$\begin{cases} \mathbf{S}_{\mathbf{u},\text{Zhang}}^{\mathbf{p}} = \frac{du^{\mathbf{p}}}{d\Phi_v} = \frac{d}{d\Phi_v} \left(\frac{T_v}{2\pi} \Phi_v \right) = \frac{T_v}{2\pi}, \\ \mathbf{S}_{\mathbf{v},\text{Zhang}}^{\mathbf{p}} = \frac{dv^{\mathbf{p}}}{d\Phi_h} = \frac{d}{d\Phi_h} \left(\frac{T_h}{2\pi} \Phi_h \right) = \frac{T_h}{2\pi}, \end{cases} \quad (2.30)$$

In Zhang and Huang's method, it is observed that the sensitivities of projector feature points remain constant, and the constancy is independent of both the phase and the pixel index. Consequently, it implies that any deviation in the projector feature points remains consistent across different pixel indices.

2.5.2 Error analysis of the projector calibration

When calibrating the projector in a structured light system, the crucial step is identifying the projector feature points corresponding to the camera's observed feature points. As depicted in Fig. 2.9, the phases of circular and radial fringe patterns (Φ_c and Φ_r) establish the mapping relationship between camera and projector feature points, which is also illuminated in Eq. 2.12.

For a camera feature point $p(u, v)$ (e.g., circle centers in this research), where u and v are typically non-integer values, integration operation is needed to determine the phase from pixel indices. To reduce the influence of integration operations, bilinear interpolation is utilized for phase estimation around a point based on its four neighboring points (highlighted area in Fig. 2.12). These neighbors are denoted as $p_{11}(u_1^c, v_1^c)$, $p_{12}(u_1^c, v_2^c)$, $p_{21}(u_2^c, v_1^c)$, and $p_{22}(u_2^c, v_2^c)$, respectively. Here, u_1^c and v_1^c represent floor values of u and v , and u_2^c and v_2^c are their ceiling values. Then, the bilinear interpolation of phases can be given by

$$\Phi_c(u, v) = \begin{bmatrix} u_2^c - u \\ u - u_1^c \end{bmatrix}^T \begin{bmatrix} \Phi_c(u_1^c, v_1^c) & \Phi_c(u_1^c, v_2^c) \\ \Phi_c(u_2^c, v_1^c) & \Phi_c(u_2^c, v_2^c) \end{bmatrix} \begin{bmatrix} v_2^c - v \\ v - v_1^c \end{bmatrix}, \quad (2.31)$$

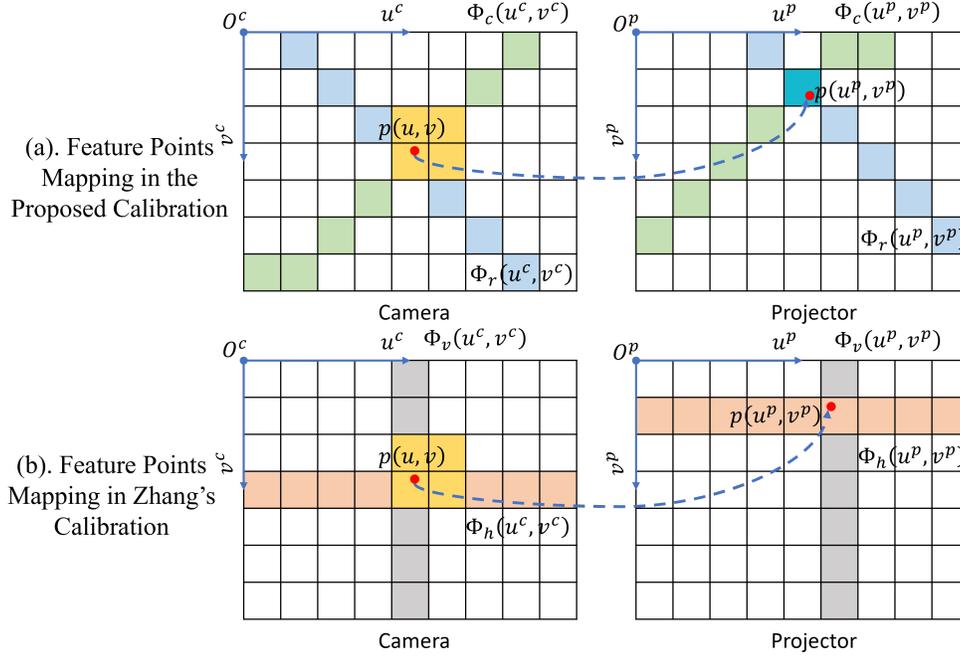


Figure 2.9: Feature point mapping schematic: camera to projector.

$$\Phi_r(u, v) = \begin{bmatrix} u_2^c - u \\ u - u_1^c \end{bmatrix}^T \begin{bmatrix} \Phi_r(u_1^c, v_1^c) & \Phi_r(u_1^c, v_2^c) \\ \Phi_r(u_2^c, v_1^c) & \Phi_r(u_2^c, v_2^c) \end{bmatrix} \begin{bmatrix} v_2^c - v \\ v - v_1^c \end{bmatrix}. \quad (2.32)$$

where $\Phi_c(u, v)$ and $\Phi_r(u, v)$ are estimated phases using bilinear interpolation. For bilinear interpolation, the maximal phase estimation error is the phase difference caused by a mismatched pixel index, which can be given by

$$\begin{cases} \Delta\Phi_c(u, v) = \frac{2\pi}{T} (r(u, v) - r(u-1, v-1)), \\ \Delta\Phi_r(u, v) = \frac{2\pi}{\Delta\theta} (\theta(u, v) - \theta(u-1, v-1)), \end{cases} \quad (2.33)$$

where $r(u, v)$ is the distance between the center of circular and radial fringe patterns and the point $p(u, v)$, $\theta(u, v)$ is the point's pixel angle. In this research, the center of circular and radial fringe patterns is off the camera image plane and partial derivatives $\frac{\partial}{\partial u}\Delta\Phi_c(u, v)$ and $\frac{\partial}{\partial v}\Delta\Phi_c(u, v)$ are greater than zero. Similarly, partial derivatives $\frac{\partial}{\partial u}\Delta\Phi_r(u, v)$ and $\frac{\partial}{\partial v}\Delta\Phi_r(u, v)$ also exceed zero. Therefore, $\Delta\Phi_c(u, v)$ and $\Delta\Phi_r(u, v)$ increase with respect to u and v , and maximal estimated phase

errors are exhibited in the bottom right corner. The subsequent steps of the projector calibration, which rely on the phase values, are thus influenced by the estimated phase errors. The accumulation of estimated phase errors inherently compromises the accuracy of the calibration method.

For Zhang and Hunag’s calibration method, bilinear interpolation is also utilized when mapping feature points from the camera to the projector, as depicted in Fig. 2.9. Horizontal and vertical fringe patterns are employed in this method, and phases $\Phi_h(u, v)$ and $\Phi_v(u^c, v^c)$ are linear to pixel (u^c, v^c) . As a result of the linear relationship, the bilinear interpolation is free from the introduction of estimated phase errors. The linear nature of the phase-to-pixel correspondence ensures a high level of accuracy in the mapping, thereby enhancing the overall accuracy of the calibration method.

In the analysis of sensitivity and error in projector calibration, it has been observed that the magnitudes of the sensitivity vectors $\|\mathbf{S}_{\mathbf{u}, \text{proposed}}^{\mathbf{p}}\|$ and $\|\mathbf{S}_{\mathbf{v}, \text{proposed}}^{\mathbf{p}}\|$ demonstrate an increasing trend in relation to the distance from a point to the center of circular and radial fringe patterns. Additionally, $\Delta\Phi_c(u, v)$ and $\Delta\Phi_r(u, v)$ also exhibit a rising tendency with respect to the pixel indices u and v . Given these findings, it is advisable to position the circle grid calibration target within the central region of the camera and projector’s field of view to mitigate the impact of the growing sensitivity and estimated phase errors. Furthermore, due to the increasing trend of both sensitivity and estimated phase error, the proposed calibration method reveals limitations compared to Zhang and Huang’s method. The latter approach, characterized by its linear phase-to-pixel relationship and absence of estimated phase errors in the mapping process, demonstrates higher accuracy and reliability.

2.6 Conclusions

In this research, we introduced a novel method utilizing circular and radial fringe patterns for the calibration and 3D reconstruction of structured light systems. Our approach was rigorously evaluated through a series of 2×2 experiments, allowing us to draw comprehensive comparisons between our proposed method, Zhang and Huang’s calibration method, and linear FPP. Our findings indicate that the proposed circular FPP surpasses that of linear FPP techniques. However, when it comes to calibration performance, our method exhibits a slight disadvantage compared to Zhang and Huang’s method. To reveal the reason for the inferior performance, we conducted a detailed analysis of the sensitivity and error in projector calibration. This investigation revealed that the relatively lower performance of our proposed calibration method is primarily attributed to the increased sensitivity

and estimated phase errors. This research not only stresses the strengths of our circular FPP method but also sheds light on its limitations, providing insights for future advancements in circular fringe projection profilometry.

CHAPTER 3

TPDNET: TEXTURE-GUIDED PHASE-TO-DEPTH NETWORKS TO REPAIR SHADOW-INDUCED ERRORS FOR FRINGE PROJECTION PROFILOMETRY¹

¹Li Jiaqiong, and Beiwen Li. 2024. *Photonics*, Vol. 10. No. 3. MDPI, 2023.

Reprinted here with permission of the publisher.

3.1 Abstract

This paper proposes a phase-to-depth deep learning model to repair shadow-induced errors for fringe projection profilometry (FPP). The model comprises two hourglass branches that extract information from texture images and phase maps and fuse the information from the two branches by concatenation and weights. The input of the proposed model contains texture images, masks, and unwrapped phase maps, and the ground truth is the depth map from CAD models. A loss function was chosen to consider image details and structural similarity. The training data contains 1,200 samples in the verified virtual FPP system. After training, we conducted experiments on the virtual and real-world scanning data, and the results support the model's effectiveness. The mean absolute error and the root mean squared error are $1.0279mm$ and $1.1898mm$ on the validation dataset. In addition, we analyze the influence of ambient light intensity on the model's performance. Low ambient light limits the model's performance as the model cannot extract valid information from the completely dark shadow regions in texture images. The contribution of each branch network is also investigated. Features from the texture-dominant branch are leveraged as guidance to remedy shadow-induced errors. Information from the phase-dominant branch network makes accurate predictions for the whole object. Our model provides a good reference for repairing shadow-induced errors in the FPP system.

3.2 Introduction

Fringe projection profilometry (FPP) has been widely applied in many fields, such as entertainment, remote environment reconstruction, manufacturing, medicine, and biology, due to its high accuracy and high-speed (J. Xu & Zhang, 2020b; Zuo et al., 2018). A typical FPP system comprises a projector for fringe pattern projection and a camera for image acquisition. The projector projects fringe pattern images onto the object surface, and the camera captures the fringe pattern images encoded by the object surface from a different perspective. Due to the triangulation of the projector, the camera, and the object, the light emitted from the projector is blocked by the adjacent surface of the object, leaving shadows in the captured images. As shown in Fig.3.1, shadows exist in the acquired fringe images, causing errors in the same areas of the wrapped phase map and unwrapped phase map. As the phase information of shadow-covered areas is lost, 3D reconstructed geometry in

shadow areas cannot be obtained based on phase maps, which reduces the accuracy of FPP methods.

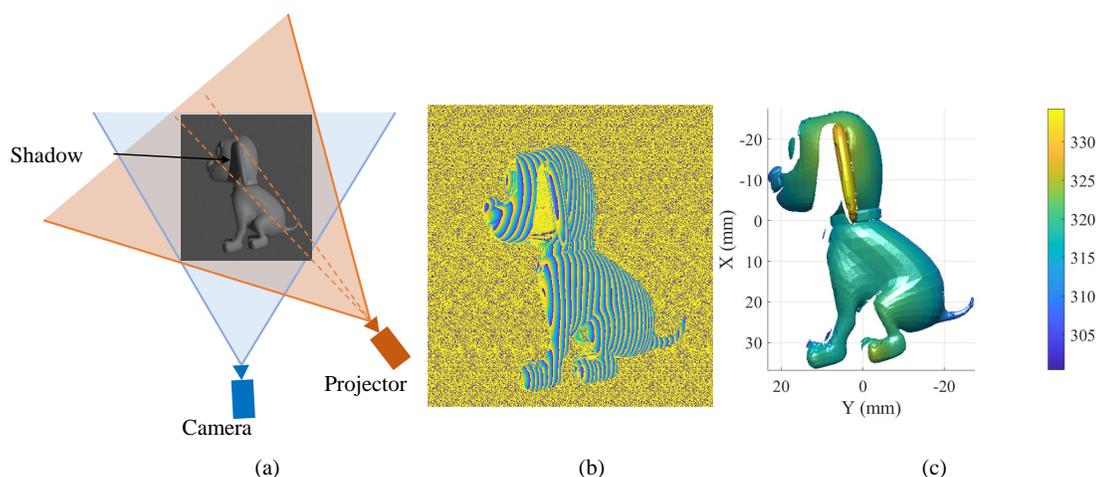


Figure 3.1: Illustration of the formation of shadows and shadow-induced errors in FPP. (a) The formation of shadows; (b) Shadow-induced errors in the wrapped phase map; (c) 3D reconstructed geometry with holes.

The simplest idea to remove shadow-induced errors is to avoid shadow generation in fringe pattern images. As shadow errors are due to triangulation of the camera, the projector, and the object, some researchers proposed uniaxial techniques (Jing et al., 2017; Meng et al., 2022; Y. Zheng & Li, 2020) to make the projector and the camera share the same optical axis and have the same region of projection and imaging. In uniaxial techniques, the retrieved 3D geometry is based on the relationship between depth and phase (or intensity or level of defocus), and beam splitters are commonly used. The projector and camera's optical axis and the beam splitter must be fixed at a specific angle to ensure a coaxial configuration, which increases system complexity and limits its application. Another method to reduce shadow errors is to use multiple devices. In the work of Skydan et al. (Skydan et al., 2005), three video projectors were set up at different angles and exclusively projected three different colored lighting patterns. Shadow areas that were not illuminated by one projector would be covered by another projector with different colored lighting patterns. And this method combined information from these two colored lighting patterns to reduce shadows. However, potential errors were brought as extra masks had to be produced to separate the valid areas and shadows. Meanwhile, the system complexity and cost increase with the number of projectors. Weinmann et al. (Weinmann et al., 2011) proposed a super-resolution structured light

system with multiple cameras and projectors to cover the full shape of objects with super-resolution. Though the proposed system could get full 3D reconstruction with high accuracy, complicated post-processing was needed. For the above methods, the hardware cost is higher than a typical FPP system with one projector and one camera, and extra post-processing or pre-processing cannot be avoided.

Besides avoiding shadow generation in fringe pattern images, researchers also tried to remove shadow-induced errors at the phase retrieval stage. Different frameworks and methods (L. Huang & Asundi, 2011; L. Lu et al., 2015; Shi et al., 2022) were used to identify and delete the invalid points in phase maps. Though no extra errors were introduced, holes were left in the final 3D reconstruction results. Based on phase monotonicity, Zhang (S. Zhang, 2009) identified and removed incorrect unwrapped points from low-illuminated object surface areas and filled holes in the phase map. Similarly, Chen et al. (F. Chen et al., 2010) used the standard deviation of the least-squares fitting to detect and remove the invalid phase points and recalculated phase values with two-dimensional linear interpolation. Interpolation or similar methods are inaccurate in estimating the lost phase information, especially for objects with an abruptly changed surface profile. Apart from interpolation methods, Sun et al. (Sun et al., 2020) proposed a discriminative approach, which classified the shadows into valid and invalid shadow areas and discriminately smoothed the shadow-induced errors with different methods. Though this method was resistant to large-scale shadows, its effectiveness was affected by incorrect shadow classification, and its performance decreased when the unwrapped phase map was obtained by the spatial phase unwrapping method.

Over the past few years, tremendous research has been developed in computer vision based on deep learning, and a series of remarkable achievements have been made, which also attracted researchers' interests in applying deep learning to three main steps of optical metrology, including pre-processing, phase analysis, and post-processing (Zuo et al., 2022). For example, Yan et al. (Yan et al., 2019) proposed a deep convolutional neural network to reduce the noise of the fringe pattern images, which obtained high-quality fringe pattern images at the pre-processing stage. In addition, deep learning is also leveraged to solve the tasks of phase analysis, including phase demodulation (Y. Li et al., 2020; T. Yang et al., 2020; Q. Zhang et al., 2020), phase unwrapping (Spoorthi et al., 2018; K. Wang et al., 2019), phase map denoising (Montresor et al., 2020; Yan et al., 2020), and so on. Meanwhile, Li et al. (Z.-w. Li et al., 2009) used a network to make an accurate phase-to-height

mapping. Except for networks targeted at one or two stages of image processing in optical metrology, end-to-end neural networks were developed to directly convert fringe pattern images to depth maps or even 3D shapes (H. Nguyen et al., 2020; Van der Jeught & Dirckx, 2019; Y. Zheng, Wang, et al., 2020).

Repairing shadow-induced holes in phase maps can be regarded as a type of image completion problem, which has been the subject of much research (Elharrouss et al., 2020). But to our knowledge, few neural networks have been developed to repair the shadow-induced errors in the FPP system, likely due to the difficulty of generating sufficient training data. However, FPP systems can be virtualized through 3D graphics to rapidly generate training data (Y. Zheng, Wang, et al., 2020), laying the foundation for repairing shadow-induced errors based on deep learning. Utilizing a virtual system built with the graphic software Blender (Team, 2018), Wang et al. (C. Wang & Pang, 2022) proposed a deep learning method to detect and repair shadow-induced errors. Their method can be conducted by only one fringe image, which makes it flexible to many FPP systems. But the effectiveness of their model is influenced by the reliability of shadow detection. Thus, methods that can avoid shadow detection and segmentation are worth trying.

To complement these methods, we propose a phase-to-depth deep learning model to repair shadow-induced errors. The proposed model contains two hourglass branch networks: a texture-dominant branch network and a phase-dominant branch network. The two combined branch networks extract information from texture images and phase maps, predict depth maps, and repair errors caused by shadows. The proposed model fuses information from texture images and phase maps in the following locations:

1. Each layer in the encoder stage of the phase-dominant branch is fused with a layer from the decoder stage of the other branch via concatenation and addition
2. At the end of the decoder stage, depth maps from two branches are fused by multiplying the respective weights.

To generate enough training data, we build and evaluate the virtual system of a real-world FPP system based on computer graphics. After training the proposed model on the virtual scanning data, the mean absolute error (MAE) and root mean squared error (RMSE) are $1.0279mm$ and $1.1898mm$, respectively. Experimental results on the real-world scanning data support the model's effectiveness.

In addition, we analyzed the influence of ambient light on prediction and found that low ambient light decreases the performance of our model as valid information in texture images decreases. And properly increasing the ambient light will improve the performance of our model. We also tested the functions of the texture-dominant branch and phase-dominant branch. The results show that information from texture images contributes to remedying the holes in 3D geometry caused by shadows, and the phase-dominant branch network makes accurate predictions for the major object geometries.

3.3 Principles

This section will introduce the theoretical foundations of a real-world FPP system and the corresponding virtual FPP system for generating training data.

3.3.1 FPP model

We adopted the three-step phase-shifting algorithm (Malacara, 2007a) to analyze fringe images, and fringe images can be described as

$$I_k(x, y) = I'(x, y) + I''(x, y) \cos[\phi(x, y) + \delta_k], \quad (3.1)$$

$$\delta_k = \frac{2k\pi}{N}; N = 3, k = 1, 2, 3, \quad (3.2)$$

where $I_k(x, y)$ denotes the image intensity at pixel (x, y) for the k th fringe image, and $I'(x, y)$ and $I''(x, y)$ presents the average intensity and the intensity modulation, respectively. In our work, we treat the gray images with average intensity as texture images and input them into the proposed model. Phase $\phi(x, y)$ can be calculated by,

$$\phi(x, y) = \tan^{-1} \left[\frac{\sqrt{3}(I_1 - I_3)}{2I_2 - I_1 - I_3} \right]. \quad (3.3)$$

Given the discontinuity of the arctangent function, phase $\phi(x, y)$ ranges in $[-\pi/2, \pi/2)$, which is called as wrapped phase. The absolute phase (unwrapped phase) can be obtained by

$$\Phi(x, y) = \phi(x, y) + 2\pi k(x, y). \quad (3.4)$$

Here, $\Phi(x, y)$ denotes the unwrapped absolute phase without phase jumps, and $k(x, y)$ is the fringe order. To remove the discontinuity in wrapped phase maps, we encoded the fringe order $k(x, y)$ by projecting additional binary images using a gray-coding technique (Y. Wang et al., 2011b). The unwrapped phase maps are also part of the input for the proposed model.

3.3.2 Virtual FPP model

As deep learning is a data-driven method and generating a large training dataset in a real-world FPP system is quite expensive, we replace the real-world FPP system with its virtual system for an easier generation of training data. Based on computer graphics and the open-source 3D software Blender(Team, 2018), we utilized the same virtual FPP system proposed by Zheng et al. (Y. Zheng, Wang, et al., 2020). We fixed the virtual camera at the origin point to simplify the system and adopted a spotlight to mimic a real projector. The power of the spotlight is set to 30W, and the spot size was 180° . The resolutions of the virtual camera and the virtual projector are 514×544 and 912×1140 , respectively. The focal length of the virtual camera was $12.03mm$, and (Shift X, Shift Y) was $(0.018, -0.021)$. Table 3.1 presents other parameters of the virtual camera and the virtual projector. Ambient light intensity and exposure were set to ensure that no underexposure or overexposure is present on the model in any captured fringe images. (The effects of ambient light intensity on the model’s performance are discussed in Sec. 3.6.1.)

To verify the established virtual FPP system, we compared the scanning results of the real-world FPP system and its corresponding virtual system. Utilizing the iterative closest point (ICP) algorithm and K-nearest neighbor (KNN) search (Y. Zheng, Zhang, et al., 2020) to make point-to-point registration of the two scanning results, the average and root-mean-square of the 3D Euclidean differences are $0.083mm$ and $0.088mm$. Based on the similarity of the scanning results from the virtual FPP system and the real-world FPP system, the captured images from the virtual FPP system can be utilized as training data.

Table 3.1: Parameters of the virtual FPP system

Parameters	Virtual camera	Virtual projector (spotlight)
Extrinsic matrix Location	(0, 0, 0)	(0.0227m, 0.0885m, -0.1692m)
Rotation	(0°, 0°, 180°)	(-14.07°, 0.46°, 1.44°)
Scale	(1.0, 1.0, 1.0)	(0.2, 0.2, 1.0)

3.4 Proposed model

In this work, we propose an hourglass deep learning model to remove shadow-induced errors and produce depth map predictions with texture images and phase maps, and combine the structural similarity index and a Laplacian pyramid loss as the loss function to optimize the training process.

3.4.1 Network architecture

Inspired by the work of Hu et al. (Hu et al., 2021), we propose a model with a texture-dominant branch and a phase-dominant branch to produce depth map predictions and then fuse depth predictions from two branches by multiplying their weight maps. The network architecture is illustrated in Fig. 3.2.

The texture-dominant branch is similar to a typical Unet (Ronneberger et al., 2015), which is a symmetric encoder-decoder network with skip connections to produce depth map predictions and offer guidance to the phase-dominant branch. The difference is that we used the ResBlock (He et al., 2016) rather than the normal 2D-convolution layer as the backbone. The downsample layer contains a maxpool layer and a ResBlock layer. The upsample layer is similar to the downsample layer, which replaces the maxpool layer with a ConvTranspose2d layer. Finally, the two-channel output layer in the texture-dominant branch is split into a depth map and a weight map, which will be utilized for depth map fusion.

The phase-dominant branch also has a symmetric encoder-decoder network with skip connections, which is similar to the texture-dominant branch. The only difference is that the phase-dominant branch fuses the feature from the texture-dominant branch at the encoder stage by concatenation and a merging layer. The merging layer is composed of two 2D-convolution layers followed by a

batch normalization layer and a ReLU activation. Thus, the lost phase information in the phase map can be recovered through features from the texture-dominant branch.

The depth maps from the two branches are fused using the method proposed by Gansbeke et al. (Van Gansbeke et al., 2019). Both branches produce a weight map, and the weight maps are rescaled in the range of (0, 1) through a softmax layer. The weight map of each branch multiplies its corresponding depth map to get the weighted depth map. And then, the weighted depth maps are added together to get the final output. Higher values in weight maps mean the corresponding depth map contributes more to the final prediction. Ideally, the model would place a higher weight on the phase-dominant branch in areas where this data is available (in areas illuminated by the projector), and the texture-dominant branch where that data is unavailable, i.e., in shadow areas. This behavior would enable the model to fill the holes caused by shadows and make accurate predictions for non-shadow regions simultaneously. The effectiveness of fusion is supported by the results in Sec. 3.5.3, and we discuss the effects of the two branches in Sec. 3.6.2.

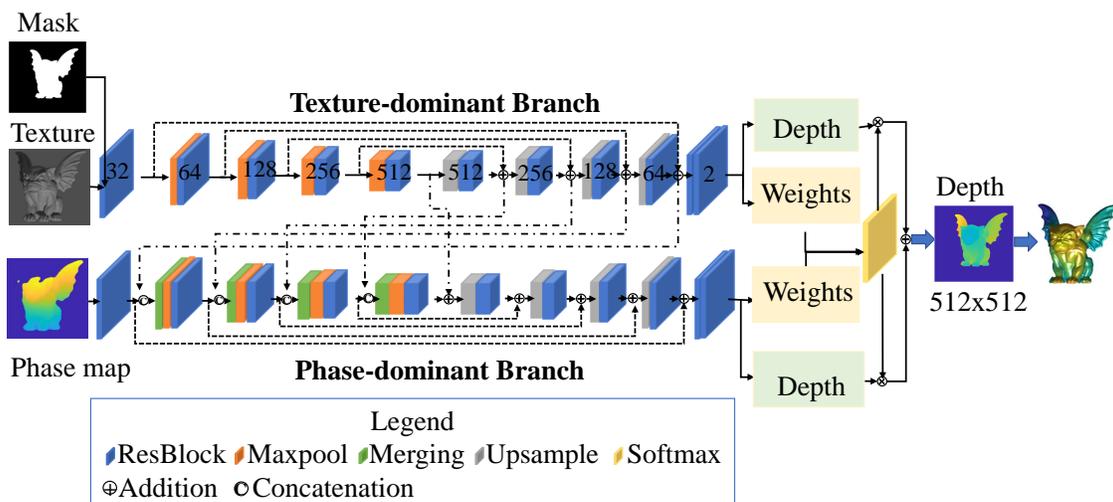


Figure 3.2: Schematic of the proposed model

3.4.2 Loss function

One of the most commonly used loss functions in computer vision is the mean absolute error (l_1 loss) or the mean square error (l_2 loss). However, the limitations of l_2 loss and l_1 loss are obvious. l_2 loss assumes that noise is independent of the local characteristics of the image, and the residuals

between predictions and true values follow a Gaussian distribution(H. Zhao, Gallo, et al., 2016). And l_1 loss assumes that the residuals follow the Laplace distribution. These assumptions are not valid in the current situation, as shadow-induced errors are related to local information of phase maps. Furthermore, both l_1 loss and l_2 loss are the metrics for evaluating average errors, which blur the edges of prediction images. Considering these limitations, our loss function contains two parts: the structural similarity index (SSIM) and the Laplacian pyramid loss. SSIM was first introduced by Wang et al.(Z. Wang et al., 2004), which is a metric to compare the differences between two images from luminance, contrast, and structure, and it can be described as

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \quad (3.5)$$

In Eq. 3.5, μ_x and μ_y defines the mean intensity of image x and image y , σ_x and σ_y are the standard deviations, σ_{xy} is the corresponding covariance, and C_1 and C_2 are two constants. SSIM ranges from 0 to 1. When a predicted image is more similar to the ground truth, SSIM value will be closer to 1. Thus, SSIM loss can be written as

$$L_{SSIM} = 1 - SSIM(P, GT), \quad (3.6)$$

where P is the prediction image and GT is the ground truth. As SSIM loss focuses on the overall similarity based on the human visual system, we still need an extra metric to evaluate the details at the edges. The Laplacian pyramid (Lap_1) loss (Bojanowski et al., 2017) can describe finer details, and it is mathematically expressed as

$$Lap_1(x, y) = \sum_{j=0}^2 2^{2j} \|L^j(x) - L^j(y)\|_1, \quad (3.7)$$

where $L^j(x)$ defines the j -th level of the Laplacian pyramid representation of image x . We combine SSIM loss and Laplacian pyramid loss to evaluate prediction from the overall structural similarity and details at edges. The loss function for the predicted depth map is

$$L = \lambda_1 L_{SSIM} + \lambda_2 Lap_1, \quad (3.8)$$

where λ_1 and λ_2 are hyper-parameters, given as 100 and 10 empirically in the following experiments. When training the model, depth maps from the two branches and the final predicted depth map are all considered. Thus, the training loss is

$$L_{training} = k_1 L_{out} + k_2 L_T + k_3 L_P, \quad (3.9)$$

where L_{out} is the loss between the ground truth and the output depth map, L_T is the loss between the ground truth and the depth map from the texture-dominant branch, and L_P is the loss between the ground truth and the depth map from the phase-dominant branch. Here, k_1 , k_2 , and k_3 are constants, empirically given as 1, 0.1, and 0.1.

3.5 Experiments

In this section, we first introduce the process to generate the training data in the verified virtual FPP system, followed by the training setup for the proposed model. Then, we implement experiments on virtual scanning data and real-world scanning data to validate the effectiveness of our model.

3.5.1 Dataset preparation

Before scanning objects in the virtual FPP system, we collected 48 CAD models from public CAD model resources, including Thingi10K(Q. Zhou & Jacobson, 2016) and Free3D(“Free 3D Models”, n.d.), mostly animals and statues. Based on the bounding box of an object and the focal length and resolution of the virtual camera, we calculate the coordinates of the object’s centroid and dimensions. Then the object is imported into the virtual system and positioned to fit the virtual camera’s field of view. To augment the dataset, we rotate each CAD model around the y-axis five times with 72° each time and then do the same thing around the z-axis. Thus, we scan each CAD model from 25 different perspectives, producing 1,200 3D scenes. For each 3D scene, three phase-shifting fringe patterns and six gray code images are projected on the object and captured by the virtual camera as shown in Fig.3.3(a)-(i). In addition, we directly export a depth map from the virtual system as illustrated in Fig.3.3(n), which represents the distance between the plane ($Z = 0$) and the object’s surface. Note that the depth map is different from the height map of the object. As nine images and one depth map are required for each 3D scene, $9 \times 1,200 + 1,200 = 12,000$ images need to be

rendered totally, which is computationally expensive for a personal computer. Therefore, we utilize a workstation with an Intel Xeon processor and eight Quadro RTX 5000 graphics cards to render images in parallel.

After obtaining the fringe images and gray code images, we get the corresponding texture images, wrapped phase maps, and unwrapped phase maps based on the method mentioned in Sec.3.3.1, which are presented in Fig.3.3(j)-(l). In the wrapped phase map, the shadow caused by occlusion from the surrounding protrusion surface makes part of the phase information invalid and brings plenty of noise. By setting light intensity threshold and modulation, we filter out noises and invalid phase points and leave holes in the unwrapped phase map. Surfaces where the fringe patterns were occluded are still visible in the texture image, so we combine the texture image and unwrapped phase map as input data and the depth map as the ground truth to train the model. By setting a suitable depth threshold, we generate the mask (Fig.3.3(m)) to remove unnecessary pixels in the texture image. Note that the related shadow areas are kept in texture images as masks are obtained by depth threshold, not light intensity threshold. Additionally, all the input images are cropped to 512×512 before being imported into the proposed model.

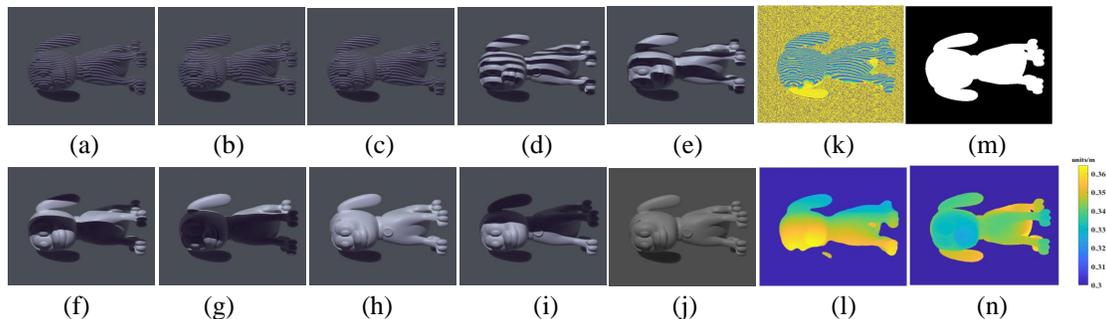


Figure 3.3: One sample in the dataset. (a)-(c) Three phase-shifting fringe patterns projected on an object; (d)-(i) Six gray code images projected on an object; (j) The corresponding texture image; (k)-(l) The corresponding wrapped phase map and unwrapped phase map; (m) The mask generated by a depth threshold; (n) The corresponding ground truth (the depth map directly exported from the virtual system).

3.5.2 Training

This subsection describes the setup for the model training process. We train the model on a workstation with an Intel Xeon processor, 128 GB RAM, and a Quadro RTX 5000 graphics card

using the PyTorch library(Paszke et al., 2019). The training data (1200 data samples) from the virtual FPP system is split into 90% and 10% as a training dataset and validation dataset, respectively. The batch size is six, and the maximum epoch is 1, 200. We use the SGD optimizer(Robbins & Monro, 1951) with $lr = 0.0003$, $momentum = 0.9$ and $weightdecay = 0.0001$. We reduce the learning rate using the scheduler class ReduceLROnPlateau in PyTorch, and the related hyper-parameters are $mode = min$, $factor = 0.9$, $patience = 2$, $threshold = 0.01$, $min_lr = 1e - 6$.

3.5.3 Results with data from the virtual system

We train and evaluate our model after setting up hyper-parameters. The mean absolute error (MAE) and root mean squared error (RMSE) are $1.0279mm$ and $1.1898mm$ for the validation dataset with 120 data samples. The results of one object are shown in Fig. 3.4. Shadow-induced errors cause part of phase information to be lost, and related 3D geometry cannot be retrieved directly from the phase map as presented in Fig. 3.4(d). Specifically, the whole left wing of the object is lost, and there is an abrupt depth change between the lost wing and its surroundings. The predicted depth map and retrieved 3D geometries are presented in Fig. 3.4(f) and (j), where the model repairs the lost part of the object. Besides, the predicted result contains the details presented in the ground truth. The rear wing of this gargoyle model represents a worst-case scenario, where the fringes have been occluded, and there is an abrupt change in the depth map between this feature and the rest of the 3D model. These combined factors increase the overall RMSE for this object above the average for the validation set at $2.6485mm$.

3.5.4 Results with data from the real-world system

We deployed the pre-trained model to the real-world scanning data to test the model’s effectiveness, i.e. that the model can remedy the shadow-induced errors and make good predictions for the non-shadow regions. We first scanned a single object which did not appear in the training data. The texture image and unwrapped phase map were generated as illustrated in Fig.3.5(a) and Fig.3.5(b). We manually labeled the texture image to get the mask. Importing the texture image, mask, and phase map into the pre-trained model, we got the predicted depth map and obtained 3D geometry as presented in Fig.3.5(f). For the non-shadow covered region, we calculated the absolute error map as shown in Fig.3.5(e) by comparing the absolute difference between the predicted depth map

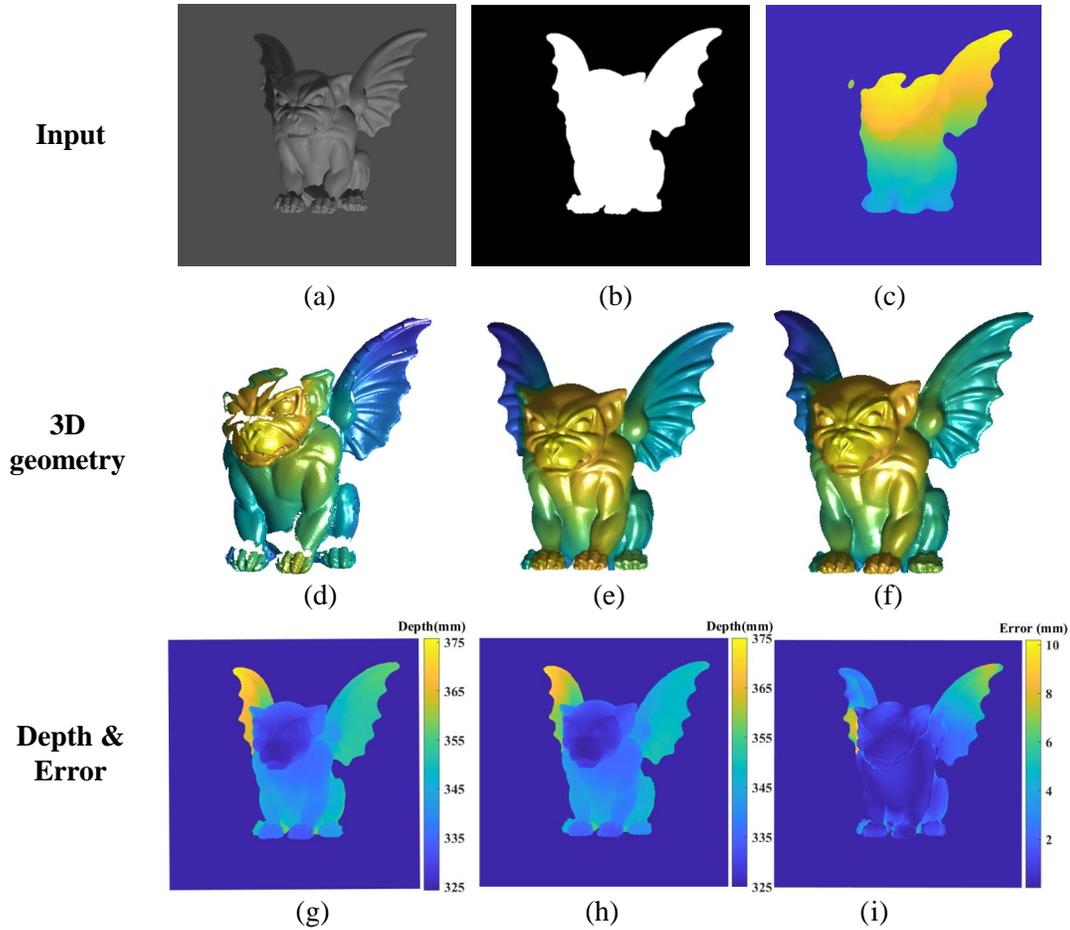


Figure 3.4: Results of an object from the virtual scanning. (a)-(c) The input of the texture image, the mask, and the phase map; (d)-(f) 3D geometry from FPP, the ground truth, and the prediction; (g) The ground truth of the depth map; (h) The predicted depth map; (i) The absolute error map with RMSE $2.6485mm$.

and the depth map from FPP, and RMSE for the non-shadow covered region was $1.5031mm$. All holes on the predicted 3D geometry were filled except for a minor discontinuity that occurred in the highlighted region. We assume that the low light intensity in the texture image's shadow areas causes the model to not extract enough valid information to fill the holes, and a detailed investigation in this regard is conducted in Sec. 3.6.1.

Similarly, we scanned the two separate objects, as shown in Fig.3.6(a). The two-object scene was not included in the training data. We also recovered the 3D geometry for the two objects, as

presented in Fig.3.6(f). The RMSE of these two objects in the non-shadow region was $2.3579mm$, which kept the same level of accuracy as the prediction for a single object. Therefore, results from real-world experiments validate the effectiveness of our model.

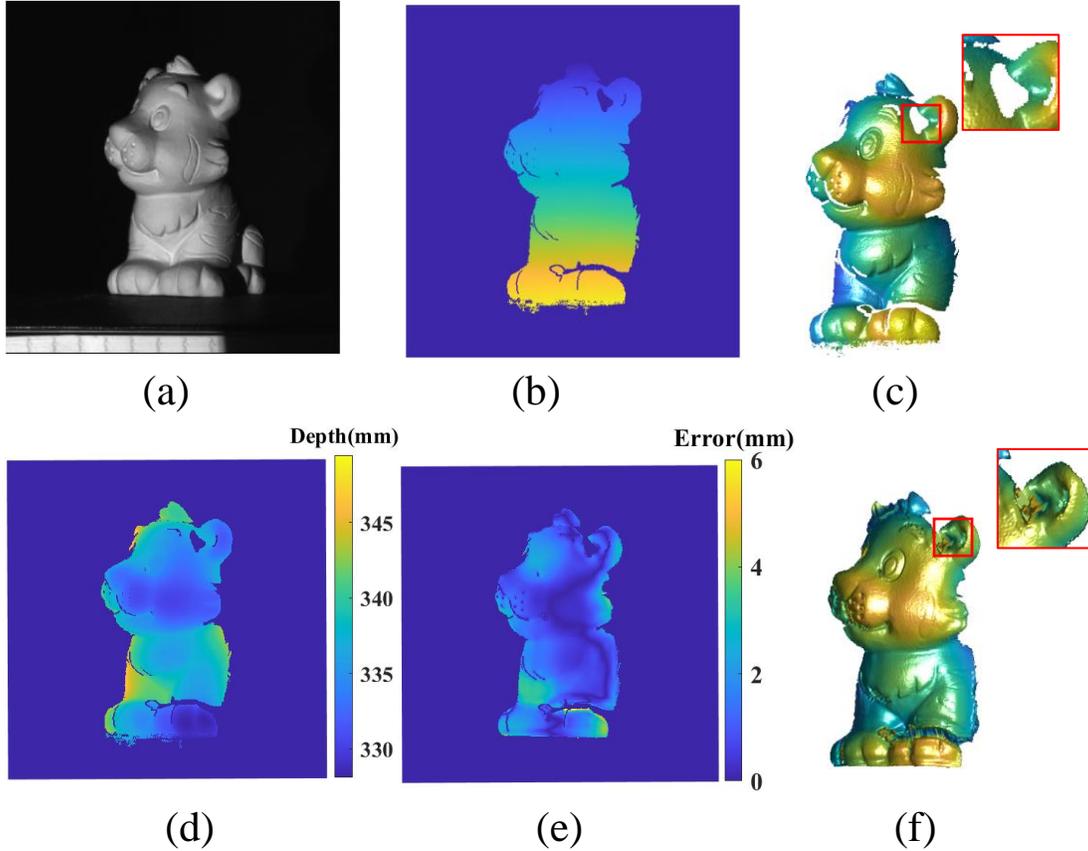


Figure 3.5: Effectiveness evaluation of the proposed model for scanning a single object in the real-world FPP system. (a) Texture image; (b) Unwrapped phase map; (c) 3D reconstructed geometry from FPP; (d) The depth map from FPP; (e) The error map for the non-shadow covered region with RMSE $1.5031mm$; (f) 3D reconstructed geometry from the predicted depth map.

3.6 Discussion

3.6.1 Influence of environment light on prediction

When we tested the model with data from the real-world FPP system, we found the filled surface was not smooth, and we assumed the low ambient light makes the shadow area of the texture image completely dark so that the contribution from the texture-dominant branch is limited. To

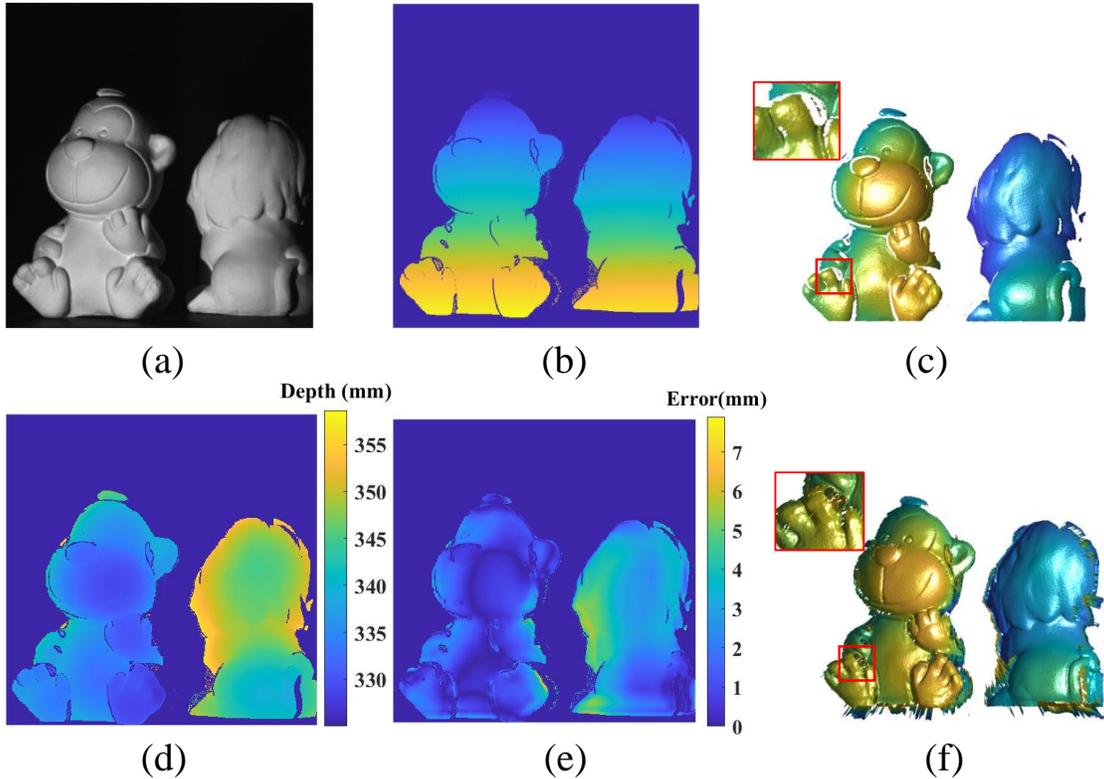


Figure 3.6: Effectiveness evaluation of the proposed model for scanning two objects in the real-world FPP system. (a) Texture image; (b) Unwrapped phase map; (c) 3D reconstructed geometry from FPP; (d) The depth map from FPP; (e) The error map for the non-shadow covered region with RMSE $2.3579mm$; (f) 3D reconstructed geometry from the predicted depth map.

test this idea, we deployed the same pre-trained model to the data captured at the same virtual FPP system with different ambient light intensities. We set the ambient light strength of the virtual FPP system at three different levels: without ambient light, the same ambient light strength utilized to generate the training data, and the strong ambient light that nearly saturates the texture image. The texture images at the three different ambient light intensities are shown in Fig.3.7(a)-(c). As shown in Fig.3.7(a), the shadow area is dark, and little valid information exists, as almost all lights are blocked. From Fig.3.7(a) to Fig.3.7(c), the shadow area becomes lighter, and global image intensity increases with the increase of environment light. However, despite this variation, the phase maps remain identical, and the 3D reconstructed geometries by using the FPP method are presented in Fig.3.7(d)-(f)). Masks are generated based on the depth threshold, which is not influenced by the

light intensity, so the mask and the phase map are unchanged among the input data, and the ambient light strength change impacts only the texture image. The 3D reconstructed geometries from the predicted phase maps are illustrated in Fig.3.7(j)-(l). Compared to the 3D geometry in Fig.3.7(d), our model remedies the holes in the foot area but leaves the holes in the face area of the object, and the filled surface is not smooth. When the ambient light strength increases and the texture image can provide more information about the shadow area to our model, the holes in the face area are filled, and the filled surface is smoother. The RMSEs of the predicted depth maps for the three cases are $12.5076mm$, $4.7793mm$, and $5.2146mm$, respectively, which also supports that appropriately increasing ambient light strength can improve the model’s performance. The RMSE of the predicted depth map in strong ambient light is higher than that in normal ambient light, indicating that high ambient light intensity also limits the model’s performance.

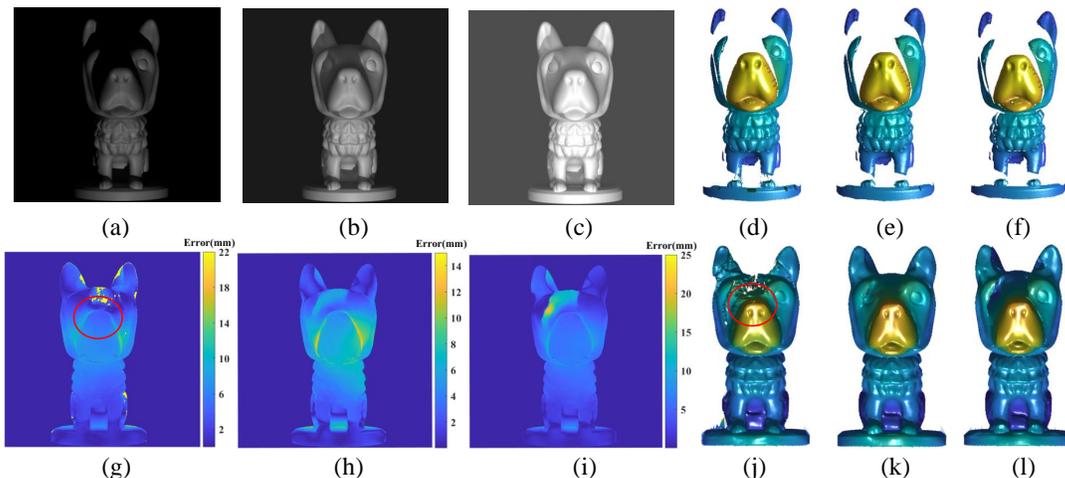


Figure 3.7: Predictions under different intensity levels of ambient light strength. (a) The texture image obtained without ambient light; (b) The texture image obtained with the same ambient light strength as the training data; (c) The texture image obtained with strong ambient light to make it close to saturation; (d)-(f) 3D reconstructed geometries by using the FPP method for the three cases; (g)-(i) Absolute error maps for the three cases by comparing ground truths and the predicted depth maps, with RMSE $12.5076mm$, $4.7793mm$, and $5.2146mm$, respectively; (j)-(l) 3D reconstructed geometries from the predicted depth maps for the three cases.

3.6.2 Function comparison of the two branch networks

As presented in experiments, our model can convert phase maps to depth maps and reduce shadow-induced errors by combing the texture-dominant branch network and the phase-dominant branch

network. We are curious about how each branch network contributes to the output predictions. Thus, we compared the results from each branch as illustrated in Fig.3.8. Fig. 3.8(f) and (g) depict the absolute error of the results from the texture-dominant branch and phase-dominant branch, respectively. The texture-dominant branch network can produce detailed predictions for the whole object, even for the shadow-covered area. The RMSE of the texture-dominant branch is $4.7078mm$, with most of the discrepancies concentrated on the gargoyle's wings. Compared to the texture-dominant branch network, the RMSE of the depth map from the phase-dominant branch network is much lower at $2.4269mm$, and the absolute error on the wings of the object in Fig.3.8(g) is lower than in Fig.3.8(f). Thus, the phase-dominant branch network can make more accurate predictions. Based on the investigation in Sec.3.6.1, the texture-dominant branch network can extract features in shadow-covered areas to remedy the shadow-induced errors in predictions. As we use two weights maps to combine the two depth maps and generate the final output prediction, the RMSE of the final output is $RMSE : 2.6485mm$, which is higher than the result from the phase-dominant branch network and much lower than the result from the texture-dominant branch network. Therefore, the texture-dominant branch network can extract features from the texture images to repair shadow-induced errors but make a lower-quality prediction of the object. The phase-dominant branch can produce relatively accurate predictions, and it needs features from the texture-dominant branch network to make predictions for the shadow-covered area.

3.7 Conclusion

Shadow-induced errors bring invalid information to phase maps and cause inaccurate 3D reconstruction in measurements taken with FPP. Existing solutions include adding extra devices or making the camera and the projector uniaxial to avoid shadow generation, removing the shadow-induced errors directly from phase maps and leaving holes in 3D reconstruction, and smoothing the errors by interpolation or similar methods. Rather than implementing these methods, we proposed a deep learning model to convert phase maps to depth maps and repair shadow-induced errors, and the effectiveness of our model was verified on both the virtual scanning data and real-world experiments. The contributions of this paper are:

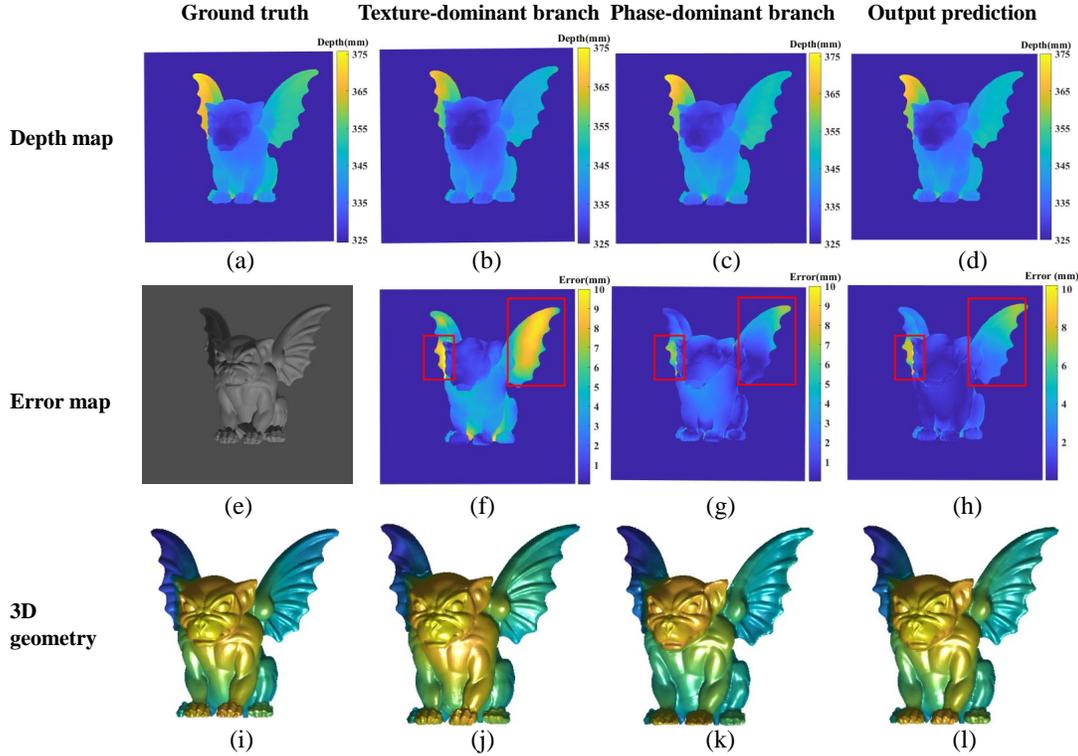


Figure 3.8: Comparing functions of the texture-dominant branch and phase-dominant branch. (a)-(d) Depth maps of the ground truth, texture-dominant branch, phase-dominant branch, and the output prediction; (e) Texture image; (f)-(h) Absolute error maps of (b)-(d) with RMSE $4.7078mm$, $2.4269mm$ and $2.6485mm$, respectively; (i)-(l) 3D geometries from the ground truth, the texture-dominant branch’s prediction, the phase-dominant branch’s prediction, and the output prediction.

- Texture images are leveraged as guidance for shadow-induced error removal, and information from phase maps and texture images are combined at two stages.
- A specified loss function that combines image edge details and structural similarity is designed to better train the model.

The proposed model has several limitations. The model’s performance is influenced by the ambient light strength. When the FPP system is without ambient light or with highly intense ambient light, the model cannot extract valid information from the texture images. Proper exposure can maximize the model’s accuracy. Another limitation is that the proposed model’s accuracy is lower than FPP in the object’s non-shadow region, suggesting it does not directly use the information from

the phase branch. Yet, an advantage of this proposed model is that it makes predictions for the whole object rather than the only shadow-covered regions, meaning that the entire process does not require any extra preprocessing, such as shadow detection or image semantic segmentation. Besides, limited by time and resources, we focus on repairing shadow-induced errors and do not further consider the generalization ability of the proposed model in the case that the virtual FPP system does not match its real-world system. For future work, researchers can design end-to-end deep learning models to get depth maps from captured images and repair shadow-induced errors. Researchers can also combine the FPP and deep learning methods to repair shadow-induced errors without sacrificing accuracy in non-shadow regions. To increase the network's generalization ability, researchers can utilize the data from diverse virtual systems (F. Wang et al., 2021) or take the camera-projector system's parameters into consideration when designing the network. Our model provides a good reference for this and other related tasks for the FPP system.

CHAPTER 4

4D LINE-SCAN HYPERSPECTRAL IMAGING¹

¹Li, Jiaqiong, et al. *Optics Express* 29, 34835-34849 (2021)

Reprinted here with permission of the publisher.

4.1 Abstract

This paper proposes a 4D line-scan hyperspectral imager that combines 3D geometrical measurement and spectral detection with high spectral resolution and spatial accuracy. We investigated the geometrical optical model of a camera attached to a spectrograph, theoretically explored the mathematical model for line-scan fringe projection profilometry, and established the 3D reconstruction and calibration methods under this proposed line-scan high-dimensional imaging system. The spectral resolution of the system is 2.8 nm, and the spatial root-mean-square error is 0.0895 mm when measuring a standard sphere with a diameter of 40.234 mm. We measure a colored statue to showcase the intensity change along the dimension of wavelength. In addition, the quality and defect of spinach leaves are inspected based on spectral data and depth data, which demonstrates the potential application of the system in the food industry.

4.2 Introduction

The advancement of optical sensing technology has greatly facilitated machine vision systems in acquiring abundant and valuable information. Compared with an RGB camera that only has three color channels, a hyperspectral imager can provide the optical information of the scene at dozens to thousands of wavelength bands. With such high capacity, hyperspectral imaging has been contributing significantly to various fields such as food quality inspections (Ariana & Lu, 2008; Kim et al., 2001; Siripatrawan et al., 2011), geologic remote sensing (Ham et al., 2005; Pearlman et al., 2003; Zhong et al., 2018), and plant phenotyping (Leucker et al., 2016; Wahabzada et al., 2016). Geometrical three-dimensional (3D) machine vision is another imaging technique that can provide depth information of the scene. Various 3D vision techniques are based on different optical principles, including time of flight (Y. Cui et al., 2010), stereo vision (Lazaros et al., 2008b), depth from defocus (Chaudhuri & Rajagopalan, 2012), laser triangulation (França et al., 2005), structured light (J. Xu & Zhang, 2020a), etc. These techniques are advantageous at different resolution and field-of-view levels, contributing to diverse applications such as robotics, manufacturing, and entertainment (Marrugo et al., 2020b; S. Zhang, 2013).

Given that hyperspectral imaging and 3D imaging provide two unique sensing modalities that are complementary to each other, researchers started to investigate and develop systems that can

simultaneously retrieve the spectral and depth information, which is known as four-dimensional (4D) imaging (x , y , z , and spectrum). One straightforward idea is to directly integrate a hyperspectral imager and a depth-sensing imager in one system and fuse the data during post-processing. For the data fusion-based methods, identifying the reference information to match the hyperspectral and depth data is critical for success. This reference information can be the geographical locations in remote sensing, which can be directly identified in hyperspectral and depth data. Tolt et al. (Tolt et al., 2011a) combined 24-band hyperspectral images with the corresponding LiDAR data and managed to detect the shadows in the remote sensing images. Ghosh et al. (Ghosh et al., 2014) also adopted the strategy of using hyperspectral image plus LiDAR data to map tree species. However, geographical information is not a universal solution since it is not always available in all measurement ranges. In close-range scanning, researchers added additional reference targets into the scene, which can be recognized by both the hyperspectral imager and depth-sensing imager as cues to merge the two sets of data. Huang et al. (P. Huang et al., 2018a) combined a line-scan hyperspectral imager with Microsoft's Kinect V2 for phenotyping. Using a line-scan hyperspectral imager enables the system to reach a high spectral resolution of 7 nm with a wide spectral range of 400 to 1000 nm. In this work, they utilized a reference gauge where 8 points at two depth levels were selected for data registration. With a similar idea, Buckley et al. (Buckley, Kurz, Schneider, & Sensing, 2012) integrated a laser scanning system with a hyperspectral imager, using several retro-reflective targets placed in the scene.

Although additional reference targets can facilitate the registration of hyperspectral and depth data, they increase the complexity and reduce the flexibility of the scanning process. On the other hand, feature detection-based registration (Monteiro et al., 2013a; Nieto et al., 2010; Sima & Buckley, 2013) can avoid using extra reference targets. However, its reliability and accuracy need to be investigated and validated, which is not trivial since hyperspectral image and depth image are distinct data modalities. To avoid reference target-based or feature detection-based matching, co-calibration of the 3D imager and the hyperspectral imager is proposed (Rogelj et al., 2019a, 2019b). Once the 3D geometry is obtained, its projection on the hyperspectral imaging plane can be retrieved according to the co-calibration information. However, since hyperspectral imaging and 3D imaging are conducted on distinct imaging sensors with different modalities, their spatial resolutions are typically different. Performing a co-calibration under such a scenario would be a

non-trivial problem given that additional data processing is typically required to avoid introducing pixel mismatch between the dual-resolution imagers. Recently, Luo et al. (Luo et al., 2021) proposed to integrate a line-scan hyperspectral imager with fringe projection profilometry, whose inputs came from the same beamsplitter. Therefore, the data can be merged by a homographic transformation. However, the line-scan hyperspectral imager is implemented by a galvanometer mirror, which may result in limited spatial resolution (X. Liu et al., 2020) on the hyperspectral imaging plane. Also, data fusion still cannot be avoided since the image for 3D reconstruction is collected on one imaging plane, whereas the hyperspectral information is collected on another, which introduces additional complexities and errors.

Unlike the aforementioned methods, integrated systems that use stereo vision as the depth-sensing module can bypass the data fusion problem, as stereo vision inherently includes a pixel-correspondence identification process. Thus it can find the corresponding pixels between the hyperspectral imaging plane and the normal imaging plane (or only between hyperspectral imaging planes) even before the 3D geometry is recovered. Wang et al. (L. Wang et al., 2016a) used a coded aperture and dispersive prism-based hyperspectral imager in a stereo vision system, and Yao et al. (Yao et al., 2019) placed a bandpass filter in front of one of the two cameras. Behmann et al. (Behmann et al., 2015) replaced the two cameras with two line-scan hyperspectral imagers, thus achieving higher spectral resolution and broader spectral range. However, as stereo vision is essentially a passive method, 3D reconstruction largely depends on the scene's texture. If the scene's texture is uniform (e.g., a pure whiteboard) or repetitive (e.g., a chessboard), the pixel correspondence may not be found, and the method may fail. Also, Zia et al. (Liang et al., 2013) suggested that depth reconstruction calculated from different spectral bands may differ. Thus a post-process such as point-cloud registration must unify the reconstruction results, which further increases the complexity. Unlike the stereo vision method using only cameras, Heist et al. (Heist et al., 2018b) added a structured light source to provide the clue for solving the pixel correspondence problem, which eliminates the dependence on scene texture. Also, the depth can be reconstructed using the spectral band of the structured light source, which has the best illumination quality. However, they used a snapshot hyperspectral camera in their proposed high-dimensional imaging system, which is simple for calibration and depth reconstruction but limits the spatial accuracy, spectral resolution, and spectral range. Besides stereo visions, modulating the spectrums of the

projected patterns is another approach to avoid data fusion (Y. Xu et al., 2020a). However, such a system requires an accurate diffraction and recombination process, which increases the complexity of the optical configuration, and the dependence on the digital micromirror device (DMD) for spectral modulation limits the spectral resolution and the number of channels.

Throughout the above discussion, we can notice that current active approaches (e.g., LiDAR, fringe projection) suffer from the complexities and errors brought by data fusion. On the other hand, stereo hyperspectral visions can bypass the data fusion process, but the success of the 3D reconstruction relies on the texture of the scene, given the nature of passive depth sensing. Another observation is that line-scan hyperspectral imaging systems generally outperform snapshot hyperspectral cameras in terms of accuracy and measuring range. Therefore, in this paper, we propose an active 4D line-scan hyperspectral imaging system that simultaneously achieves high-accuracy 3D geometry and hyperspectral imaging without data fusion. The data processing pipeline is illustrated in Fig. 4.1. A sequence of fringe patterns is projected on the object by the projector. As the object moves along the linear guide rail, the camera captures the raw slit images of the whole object. After stitching all the slit images at specific pixel indexes, we get the illuminated fringe images at different pixel indexes. Then we calculate the average intensity of these images and obtain spectral data at different wavelengths. Besides, we calculate 3D geometry from illuminated fringe images with the highest intensity contrast. Now, at all pixel indices, both 3D geometry information and hyperspectral data are obtained, thus, 4D imaging is achieved. We can notice that since only one camera imaging plane is utilized, there is no need to perform data fusion. To verify the geometrical accuracy, we first measured a standard spherical object with a diameter of 40.234 mm, and the experimental root-mean-square error (RMSE) is 0.0895 mm. Besides, the spectral resolution can achieve 2.8 nm. Then a colored tortoise statue was measured to show that the intensity changed along wavelength. Finally, we conducted a case study on quality inspections and defect detection of spinach leaves, which demonstrated its potential application in nondestructive food and agriculture evaluation.

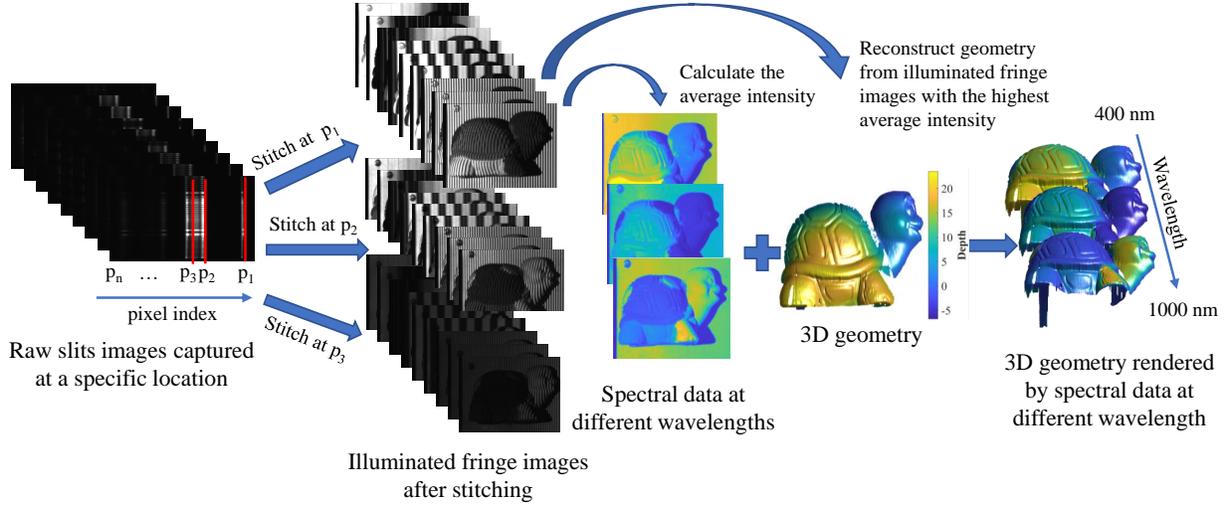


Figure 4.1: Data processing pipeline of the proposed system.

4.3 Methods

4.3.1 Modeling the line-scan hyperspectral structured light imager

Figure 4.2 shows the schematic diagram of the proposed 4D line-scan hyperspectral structured light imager. A projector projects multiple fringe images to illuminate the object's surface, and a camera attached with a line-scan spectrograph observes the scene. Note that the imaging planes of the projector and the camera are supposed to be set parallel, and they are not drawn parallel in Fig.4.2 because of simplification and for the purpose of better visualization. The light can only pass through a narrow slit inside the spectrograph. Therefore, the camera can only acquire a small line of the scene at a time. To address this issue, we incorporated a linear translation stage to hold samples, and it moves in a direction that is perpendicular to the slit of the spectrograph. In this way, the camera can acquire the entire scene in a line-scanning manner. In each line scan, the light passes through the slit and then gets scattered onto the camera imaging plane (Complementary metal-oxide-semiconductor, CMOS), and each row of the CMOS can receive the intensity of the light at a specific wavelength.

We adopted a three-step phase-shifting algorithm (Malacara, 2007b) which can facilitate 3D geometric reconstruction. The fringe patterns are projected on objects by the projector, and the

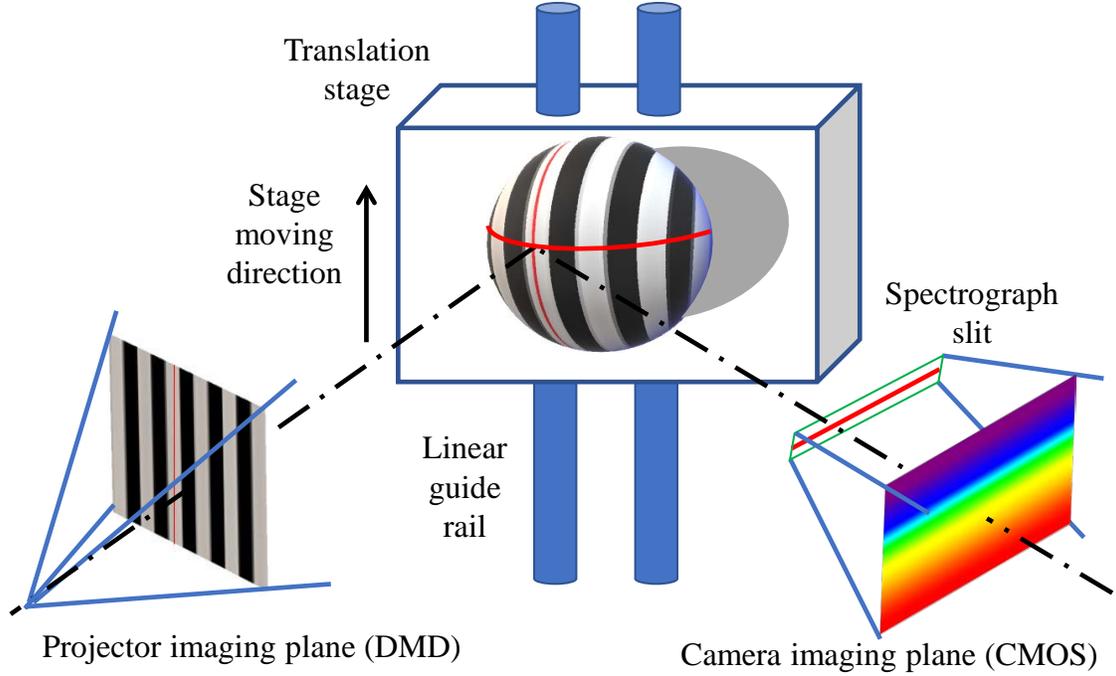


Figure 4.2: Schematic diagram of the 4D line-scan hyperspectral structured light imager combining fringe patterns projection, stage moving, and slit images acquisition

camera captures these illuminated fringe images. As we use three temporal phase-shifted intensity patterns, and phase shift between each pattern is equally set as $2\pi/3$, the camera captured image intensities at pixel location (u, v) can be described as

$$I_1(u, v) = I'(u, v) + I''(u, v) \cos[\phi(u, v) - 2\pi/3], \quad (4.1)$$

$$I_2(u, v) = I'(u, v) + I''(u, v) \cos[\phi(u, v)], \quad (4.2)$$

$$I_3(u, v) = I'(u, v) + I''(u, v) \cos[\phi(u, v) + 2\pi/3], \quad (4.3)$$

where $I_1(u, v)$, $I_2(u, v)$, and $I_3(u, v)$ are the captured image intensities at pixel location (u, v) in three different illuminated fringe images, $I'(u, v)$ denotes the average intensity that represents the ambient light, and $I''(u, v)$ represents the intensity modulation induced by sinusoidal fringe projection. By simultaneously solving these three equations, the wrapped phase $\phi(u, v)$ which is

determined by the profile of subjects can be calculated as

$$\phi(u, v) = \tan^{-1} \left[\frac{\sqrt{3}(I_1 - I_3)}{2I_2 - I_1 - I_3} \right]. \quad (4.4)$$

Therefore, given the camera captured images $I_1(u, v)$, $I_2(u, v)$, and $I_3(u, v)$, the wrapped phase $\phi(u, v)$ can be uniquely solved. The wrapped phase information essentially contains the information regarding the distortion of phase lines by geometric variations. Retrieving such information is crucial for 3D geometric reconstruction. However, due to the discontinuities of the arctangent function, $\phi(u, v)$ can only range in $[-\pi, \pi)$, which can cause phase jumps in the phase map. Such $\phi(u, v)$ is known as wrapped phase. To unwrap the phase, we adopted a three-frequencies (including low, middle, and high frequencies) phase-unwrapping algorithm (Towers et al., 2003), which can remove the phase jumps:

$$\Phi(u, v) = \phi(u, v) + 2\pi k(u, v). \quad (4.5)$$

Here, $\Phi(u, v)$ denotes the unwrapped absolute phase without phase jumps, and $k(u, v)$ is the fringe order. The absolute phase map can be easily translated to the column index of the DLP-projector(digital micromirror device, DMD):

$$u^P = T\Phi/2\pi, \quad (4.6)$$

where T is the period of the high-frequency fringe pattern.

We investigated the model and calibration method of the line-scan structured light system. Fig. 4.3(a) shows that the imaging planes of the projector and camera are parallel to each other, and the phase direction (u^P) of the projector is perpendicular to the spectrograph slit and is parallel to the direction of the translation stage.

To calibrate the system, we first need to scan a horizontal reference plane and then calculate its phase information. In order to better illustrate the geometrical relationships, we moved the geometries to Fig.4.3(b) and added auxiliary lines. As shown in Fig.4.3(b), when a vertical line ($\overline{O^P B}$) passes through O^P , the intersection of the vertical line and the reference plane is defined as the origin of the world coordinates, O^W . Therefore, the reference plane is also the $X^W Y^W$ plane of the world coordinate, meaning that all the points on this plane have Z^W as 0. $P_1(X_1, Y_1, Z_1)$ can be

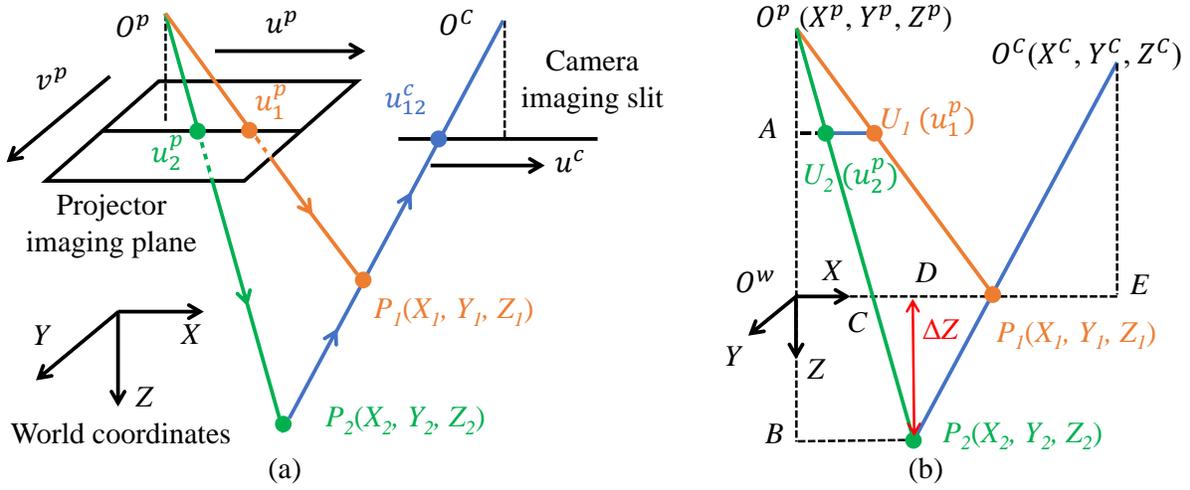


Figure 4.3: Model of the line-scan structured light system. (a) Projector illumination and camera acquisition process; (b) Triangulations for obtaining the relationship between phase differences and depth differences.

any point on the reference plane, and $Z_1 = 0$. $P_2(X_2, Y_2, Z_2)$ is another point at a different height, and both P_1 and P_2 are projected onto the same point of column u_{12}^c on the camera imaging slit. The projections of P_1 and P_2 on the projector imaging plane are U_1 and U_2 , respectively and their corresponding column indexes are u_1^p and u_2^p , which can be calculated from Eq.4.6:

$$u_1^p = T\Phi_1/2\pi, u_2^p = T\Phi_2/2\pi, \quad (4.7)$$

where Φ_1 and Φ_2 are the absolute phase values of P_1 and P_2 , respectively. Now, the ultimate question for depth reconstruction is: given (1) the phase values of Φ_1 and Φ_2 and (2) P_1 is on the reference plane (i.e., $Z_1 = 0$), what is the depth of Z_2 ? This question is also equivalent to how to obtain the ΔZ in Fig. 4.3(b).

In Fig. 4.3(b), ΔDP_2P_1 is similar to ΔP_1EO^c , which gives us

$$\frac{\Delta Z}{Z^c} = \frac{X_1 - X_2}{X^c - X_1}. \quad (4.8)$$

Also, since $\Delta O^P A U_2$ and $\Delta O^P A U_1$ are similar to $\Delta O^P B P_2$ and $\Delta O^P O^w P_1$, respectively, we have

$$\frac{X_1}{\overline{AU_1}} = \frac{Z^P}{O^P A}, \frac{X_2}{\overline{AU_2}} = \frac{Z^P + \Delta Z}{O^P A}. \quad (4.9)$$

By simultaneously solving Eq. 4.8 and Eq. 4.9, we can obtain

$$\Delta Z = (\overline{AU_1} - \overline{AU_2}) \frac{Z^P}{O^P A} \cdot \left[\left(\frac{1}{Z^C} + \frac{\overline{AU_2}}{O^P A \cdot X^C - \overline{AU_1} \cdot Z^P} \right) \cdot \left(X^C - \frac{\overline{AU_1} \cdot Z^P}{O^P A} \right) \right]^{-1} \quad (4.10)$$

Since $\overline{AU_1}$ and $\overline{AU_2}$ are the line segments exists on DMD, they are much smaller than the other values in Eq. 4.10, so Eq. 4.10 can be simplified as

$$\Delta Z = (\overline{AU_1} - \overline{AU_2}) \frac{Z^P}{O^P A} \cdot \frac{Z^C}{X^C}. \quad (4.11)$$

$\overline{AU_1}$ and $\overline{AU_2}$ can be calculated as

$$\overline{AU_1} = l \cdot u_1^p = l \cdot T \Phi_1 / 2\pi, \overline{AU_2} = l \cdot u_2^p = l \cdot T \Phi_2 / 2\pi, \quad (4.12)$$

where l is a scaling factor that transforms column index on DMD to real-world length, having a dimension of mm/pixel. Finally, substituting Eq. 4.12 into Eq. 4.11, we have

$$\Delta Z = (\Phi_1 - \Phi_2) \frac{lT}{2\pi} \cdot \frac{Z^P}{O^P A} \cdot \frac{Z^C}{X^C}. \quad (4.13)$$

Clearly, for any given point P_1 on the reference plane, l , T , Z^P , $O^P A$, Z^C , and X^C remain constants, so Eq. 4.13 can be simplified as

$$\Delta Z = k_z \cdot (\Phi_1 - \Phi_2). \quad (4.14)$$

As a result, we can measure at least two horizontal planes with two known depths and calculate the corresponding phase maps, and then we can determine k_z . The two known horizontal planes can also be used to calibrate X^W and Y^W coordinates.

For the X^W direction, as shown in Fig. 4.4, when we place the two calibration planes at depths Z_1^W and Z_2^W , the scaling factors of the camera at the corresponding depths, s_1 and s_2 , can be

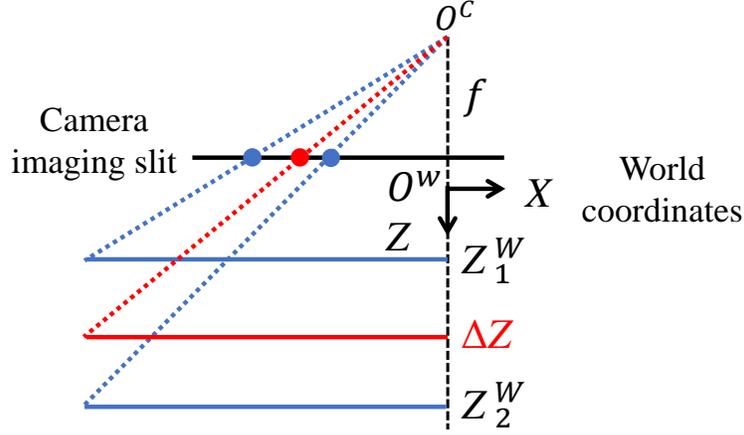


Figure 4.4: At different depths, the system has different scaling factors in the X^W direction. The planes at Z_1^W and Z_2^W are the known planes for calibration. The scaling factor of any plane at arbitrary depth, ΔZ , is to be determined.

measured given known dimensions on the planes and the corresponding number of pixels on the imaging plane. Also, the scaling factors equal to the depth ratios:

$$s_1 = \frac{f}{O^c O^w + Z_1^w}, s_2 = \frac{f}{O^c O^w + Z_2^w}, \quad (4.15)$$

where f is the focal length of the camera. Similarly, at an arbitrary depth, ΔZ , the scaling factor is

$$s = \frac{f}{O^c O^w + \Delta Z}. \quad (4.16)$$

Substituting Eq. 4.15 into Eq. 4.16, the scaling factor can be calculated as

$$s = \frac{-s_1 s_2 (Z_1^w - Z_2^w)}{-s_1 Z_1^w + s_2 Z_2^w + \Delta Z (s_1 - s_2)} = \frac{k_{x_1}}{k_{x_2} + k_{x_3} \Delta Z}. \quad (4.17)$$

For the Y^W direction, the scaling factor, k_y , is determined by the camera frame rate and the speed of the translation stage, and it is independent of ΔZ (P. Zhang et al., 2015). Thus, at least one plane needs to be scanned to determine it. As a summary of the model, given a fringe illuminated image, its image center (u_o^c, v_o^c) and phase differences map $\Delta\Phi$ (corresponding to the reference plane), the

3D geometrical information of any pixel (u^c, v^c) can be recovered by:

$$\begin{cases} Z = k_z \Delta \Phi \\ Y = k_y (v^c - v_o^c) \\ X = \frac{k_{x1}}{k_{x2} + k_{x3} Z} (u^c - u_o^c) \end{cases} \quad (4.18)$$

Note that this 3D reconstruction model ignores the lens distortion because the line scanning camera sees through only the middle part of the lens, which has fewer geometric distortions (P. Zhang et al., 2015).

4.3.2 Spectral Calibration

Before getting the spectral data, we should get the relationship pixel index p and wavelength λ and complete the spectral calibration. A Mercury Argon wavelength calibration light source (model: Ocean Insight HG-2) is equipped with the spectrograph and gets the image shown in Fig.5.9(a). The calibration light source emits strong and narrow spectral lines, where pixel index can be acquired from Fig.5.9(b). The specifications of the calibration light source offer the wavelength at these pixel indexes. The relationship between pixel index and wavelength is a third-order polynomial according to the specifications.

$$\lambda = I + C_1 p + C_2 p^2 + C_3 p^3, \quad (4.19)$$

where λ is the calibrated wavelength at pixel index p , I , C_1 , C_2 and C_3 are coefficients. In Fig. 5.9(b), Several spectral lines located at indexes 605, 646, 785, 823, 966, 1044, 1101, 1137, 1219, and their corresponding wavelengths 404.656 nm, 435.833 nm, 546.074 nm, 579.066 nm, 696.543 nm, 763.511 nm, 811.531 nm, 842.465 nm, 912.297 nm, are used to calculate these parameters. The results are $[I, C_1, C_2, C_3] = [23.3303, 0.4581, 3.4464 \cdot 10^{-4}, -1.0027 \cdot 10^{-7}]$ with RMSE 0.5249 nm. According to the spectrograph specifications, the spectral resolution is 2.8 nm.

4.4 Experiments

In this section, a spherical object was first measured to verify the geometrical accuracy of the system. Then we measured a colored tortoise statue to present the intensity change along with

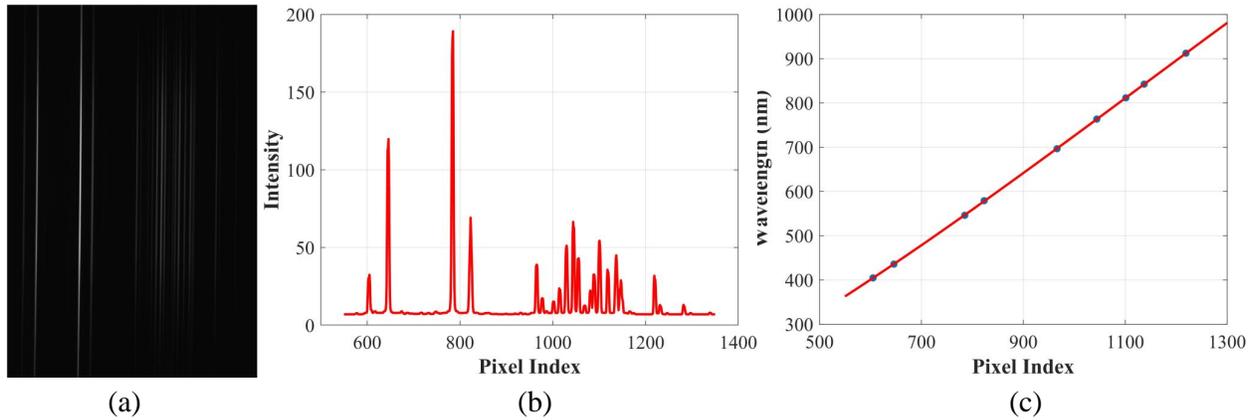


Figure 4.5: Process of spectral calibration. (a) Spectral image of Mercury Argon Calibration Source; (b) Intensity of the spectral image (a); (c) Third-order polynomial relationship between wavelength and pixel index.

the wavelengths. Lastly, we showed the potential application of our innovative high-dimensional imaging system in the food industry by evaluating quality and detecting defects on spinach leaves.

4.4.1 System setup

Figure 4.6 presents the system that contains two parts, the moving part and the static part. In the static part, a CMOS camera (model: FLIR Grasshopper3 GS3-U3-41C6C-C), a visible and near-infrared (VNIR) spectrograph (model: Specim ImSpector V10H), and a 12 mm focal length lens are connected end by the end to capture images. For the camera, the exposure time is 40 ms, and the resolution is 1920×1200 pixels. The spectral range of the spectrograph is 400 - 1000 nm, and the spectral resolution is 2.8 nm with a 30 μm slit. Besides, a DLP developmental kit (model: DLP Lightcrafter 4500) is used for pattern projection, whose resolution is 912×1140 pixels. The projector period is 160 ms, 159.7 ms for pattern projection, and 0.3 ms for intervals. The projector and camera start working simultaneously to ensure the camera can capture line-scan images with the illumination of fringe patterns. In addition, a halogen lamp (model: Sunlite 41319-SU MR16 FXL) acts as the light source. The moving part is composed of a stepper motor (model: NEMA 23), a linear rail (model: TFCFL Linear Rail 100 mm), and a stepper motor driver (model: MYSWEETY

TB6600). The step angle of the stepper motor is 1.8° , and the lead of the ballscrew in linear rail is 4 mm. The motor period is set up as 400 ms. Thus, the uniform speed of the stepper motor is $3\text{mm}/\text{min}$. Besides the two parts, a microcontroller (model: Arduino Uno) controls the specimen movement, pattern projection, and image acquisition to achieve automation.

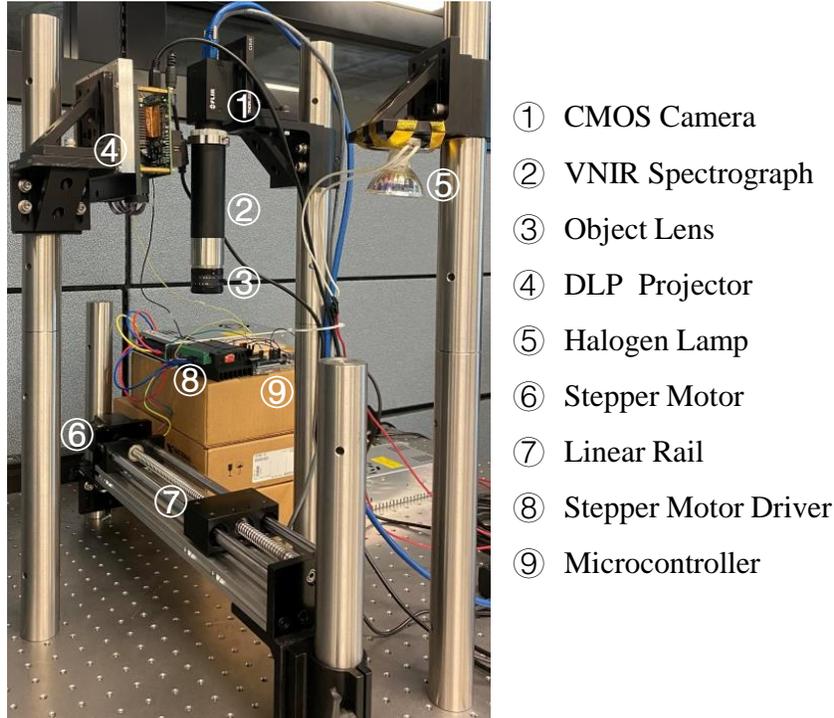


Figure 4.6: Composition of 4D line-scan hyperspectral structured light imager

4.4.2 Geometrical accuracy verification

To perform accurate 3D geometric reconstruction, scaling factors in Eq. 4.18 should be calculated. We firstly scanned two reference planes attached with a printed chessboard, and recorded the image center u_o^c . The height difference between the two planes is known as ΔZ . Based on Eq. 4.44.5, The unwrapped phase of the two reference planes are Φ_1 and Φ_2 , respectively. Then the phase difference of these two planes is $\Delta\Phi$. Thus, the scaling factor k_z is obtained by identifying the ratio between ΔZ and $\Delta\Phi$. The width of the square of the chessboard is known as w , and pixel numbers of a square in X-axis and Y-axis direction are p_x and p_y , respectively. Then scaling factors k_y was

computed by w/p_y . Analogously, we also got the pixel-length scaling factors s_1 and s_2 for plane P_1 and P_2 . The scaling factor in X-axis direction was calculated based on Eq. 4.18. Following these procedures, all the key scaling factors for 3D spatial measurement were obtained.

A plastic spherical object was measured to verify the accuracy of the proposed method and the system. The actual diameter of this sphere is 40.234 ± 0.012 mm which is obtained by six consecutive measurements with a digital caliper. Then, we measured the spherical object again using the proposed method. Fig.4.7(a) demonstrates the overlay of reconstructed 3D geometry and an ideal sphere with a radius of 20.1170 mm. The radius of the reconstructed sphere is 20.0901 mm, whose measurement error is 0.1335%. Fig.4.7(b) presents the error map between the reconstructed geometry and the standard sphere, and its RMSE is 0.0895 mm. To better visualize the results of the error map, we made a cross-section at the X-axis shown as Fig.4.7(c). The overall error of the reconstructed sphere is below 0.1 mm. However, the error at the edge is higher than those in the middle area due to shadow. Therefore, this experiment demonstrates that the 4D line-scan hyperspectral imaging system can achieve high spatial accuracy. Also, the reconstructed 3D geometry can be used for radiance correction. As shown in Fig.4.8(a), the image intensity of the plastic ball is not uniform, and the outer regions that are not pointing to the irradiation direction have a lower intensity. On the other hand, the regions that point to the irradiation direction have higher intensity. This measurement is not correct, because the plastic ball has a diffuse surface with uniform radiance. To correct the aberration, after we reconstructed 3D geometry, we estimated a normal map $N(u, v)$ which contains all the normal vectors of the plastic ball. Then, the angles of all the normal vectors, $\theta(u, v)$, can be calculated given the irradiation direction. Finally, the image intensity can be corrected as $I_{original}/\cos(\theta)$, as shown in Fig.4.8(b). Now, we can observe that the surface has a uniform radiance distribution. It should be noted that due to the estimation errors of the normal vectors, the correction process can introduce image noises.

4.4.3 Measurement of a tortoise statue

We measured a colored tortoise statue to show intensity change along different wavelengths, and the associated video of a sample data acquisition process is shown in [Visualization 1](#). As shown in Fig. 4.9(a), parts A, B, and C of the tortoise were printed with red, green, and blue colors. We calculated the mean intensity of these parts and got the results as Fig.4.9(e). It shows that the

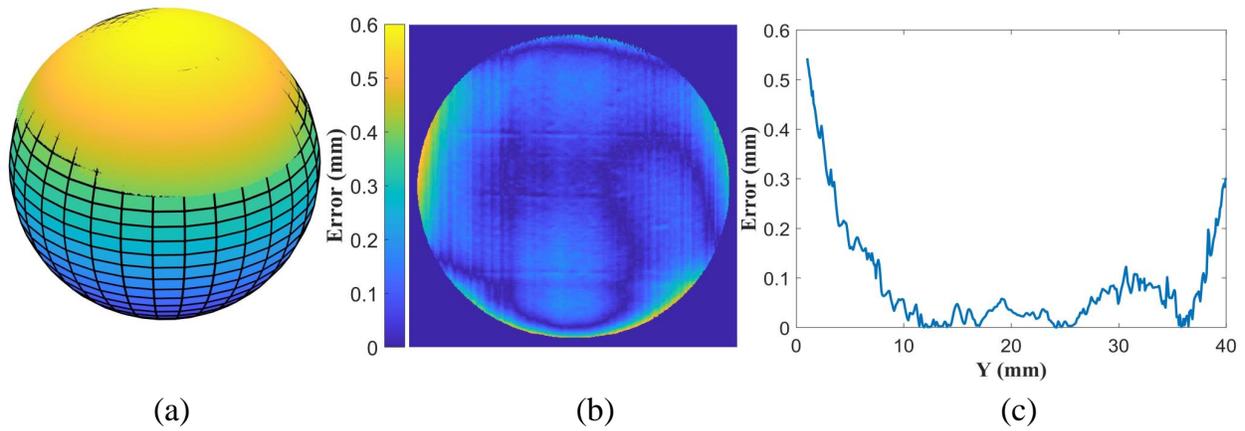


Figure 4.7: A sample evaluation of geometrical accuracy using the proposed method. (a) Overlay of the reconstructed sphere and the ideal sphere with a 20.1170 mm radius; (b) error maps (RMSE: 0.0895 mm); (c) Cross-section at X-axis middle position.

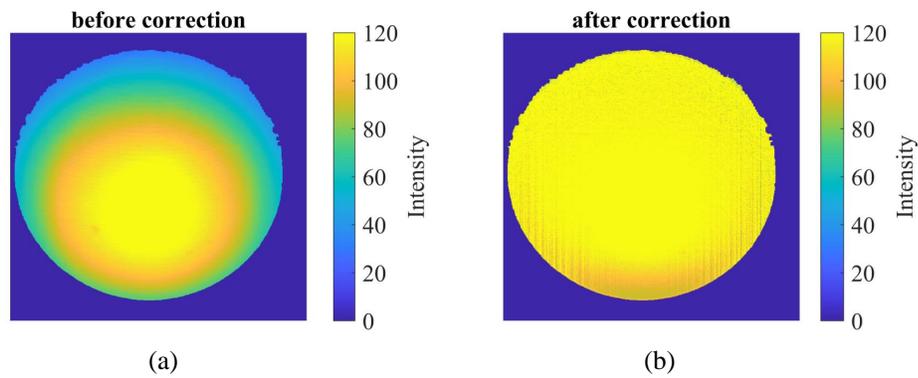


Figure 4.8: Radiance correction using surface normal estimation. (a) and (b) are the image intensities of the plastic ball before and after corrections, respectively.

intensity of a specific colored part is the highest at its corresponding wavelengths (red at 640 nm, green at 528 nm, blue at 465 nm). For example, the intensity of the red part A achieves the highest when the wavelength arrives at 640 nm (Fig.4.9(b)). Besides, the overall intensity at a wavelength of 465 nm is the highest, caused by nonhomogeneous light sources. Given this, we reconstructed the object at the wavelength of 465 nm to get the highest signal-to-noise ratio and best results. Despite a small hole at the rearfoot region caused by saturation, the overall 3D reconstruction quality is high,

as shown in Fig.4.9(b) - (d) and its associated video ([Visualization 2](#)), which further demonstrates the success of the proposed method.

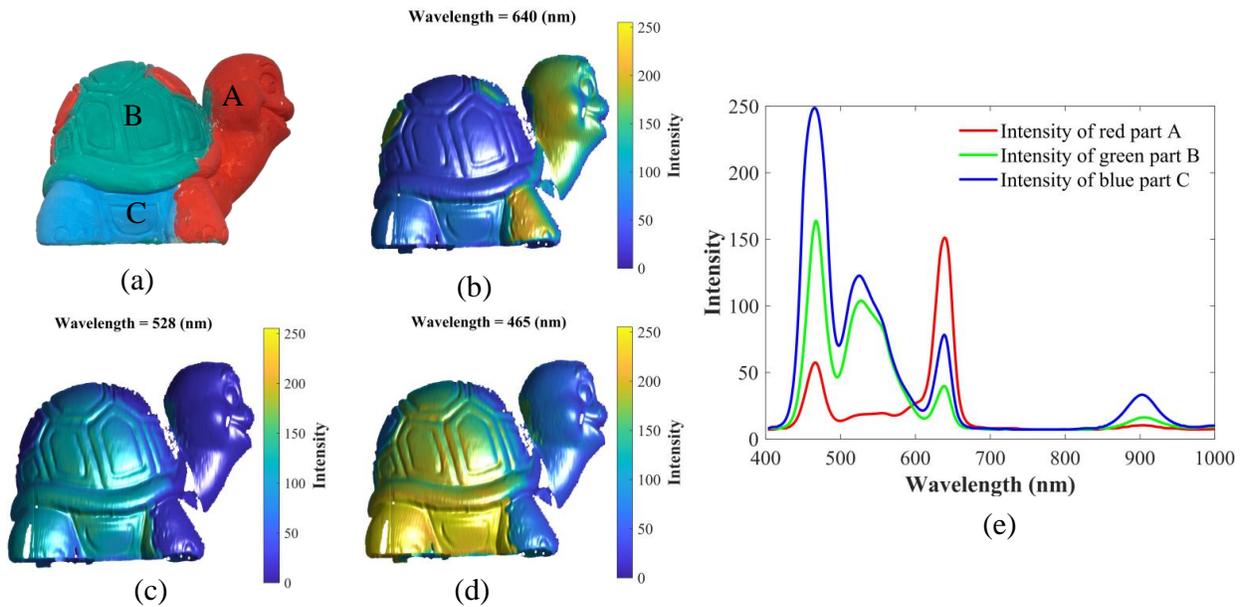


Figure 4.9: Hyperspectral 3D measurement of a tortoise statue and associated video ([Visualization 2](#)). (a) RGB image of the tortoise statue; (b) Reconstructed tortoise statue at red light wavelength 640 nm; (c) the reconstructed tortoise statue at green light wavelength 528 nm; (d) Reconstructed tortoise statue at blue light wavelength 465 nm; (e) Intensity change of three different colored parts.

4.4.4 Case study: evaluation of vegetation production quality

As a non-destructive detection tool, hyperspectral imaging has been widely explored for food quality and safety evaluation and inspection in past years (Y. Lu et al., 2020). Based on hyperspectral imaging, various physicochemical properties (such as reflectance, transmittance, fluorescence, and chemical composition) can be acquired to evaluate internal and external quality. As a complementary sensing modality of hyperspectral imaging, 3D geometrical measurement can provide the 3D geometry of products. Therefore, we present a case study of evaluating the quality and detecting defects of spinach leaves using the proposed 4D line-scan hyperspectral imaging system to demonstrate its potential applications in the food industry. In the system, a halogen lamp illuminated the system to enhance image intensity and reduce the impact of noises. Besides, we set up the blue LED for fringe patterns projection. Fig. 4.10(a) presents spinach leaves at the different fresh levels. The fresh level of leaves decreases from leaves A to E. Leaves A and B are at optimal quality, while leaves C, D,

and E are non-fresh and slightly yellow. Reflectance of leaves can be calculated as Eq.4.20 (H. Liu et al., 2020).

$$r_p(\lambda, x, y) = \frac{i_p(\lambda, x, y) - i_d(\lambda, x, y)}{i_w(\lambda, x, y) - i_d(\lambda, x, y)}, \quad (4.20)$$

where r_p is the reflectance at the location (x, y) and the wavelength λ . i_p , i_d and i_w denote image intensity values of the leaves, dark and white reference, respectively. Before calculating reflectance, we built masks to remove background based on spectral images at 840 nm. Then, the average reflectance value of whole leaves was calculated. In Fig. 4.10(b), the reflectance values present a general increase from leaf A to leaf E, corresponding to leaf quality deteriorating gradually. In addition, reflectance values of leaf D and E are higher than the values of other leaves in the range from 470 nm to 700 nm. The increasing trend of reflectance value is due to the fact that leaves turn yellow and chlorophyll content decreases. Merzlyak et al. (Merzlyak et al., 2003) found the same relationship that reflectance value increases sharply with a decrease of chlorophyll content, which supports our findings.

The normalized difference vegetation index (NDVI, (D. Cui et al., 2009)) is a widely used indicator in remote sensing to describe vegetation distribution. Leaf cells can absorb light in the red region for photosynthesis and reflect near-infrared (NIR) light. Thus the more green leaves a plant has, the more red light can be absorbed, and NIR light can be reflected. NDVI is defined in Eq.4.21.

$$NDVI = \frac{r_{NIR} - r_{Red}}{r_{NIR} + r_{Red}}, \quad (4.21)$$

where r_{NIR} and r_{Red} represent reflectance values acquired in the NIR and red regions, respectively. In detail, NDVI images are obtained based on reflectance values at wavelengths 800 nm and 660 nm. According to the definition, NDVI varies between -1 and 1, and a higher value stands for higher chlorophyll content. Fig. 4.10(c) presents that NDVI value decreases gradually from leaf A to leaf E, which indicates the decline of chlorophyll content. The declining trend of NDVI value is corresponding to an increase of reflectance from leaf A to leaf E, as both of the two indicators are related to chlorophyll content. As a result, we verified Cui's claim (D. Cui et al., 2009) that NDVI can be used as a chlorophyll content indicator, and the proposed system has the potential to acquire NDVI data to assess chlorophyll content.

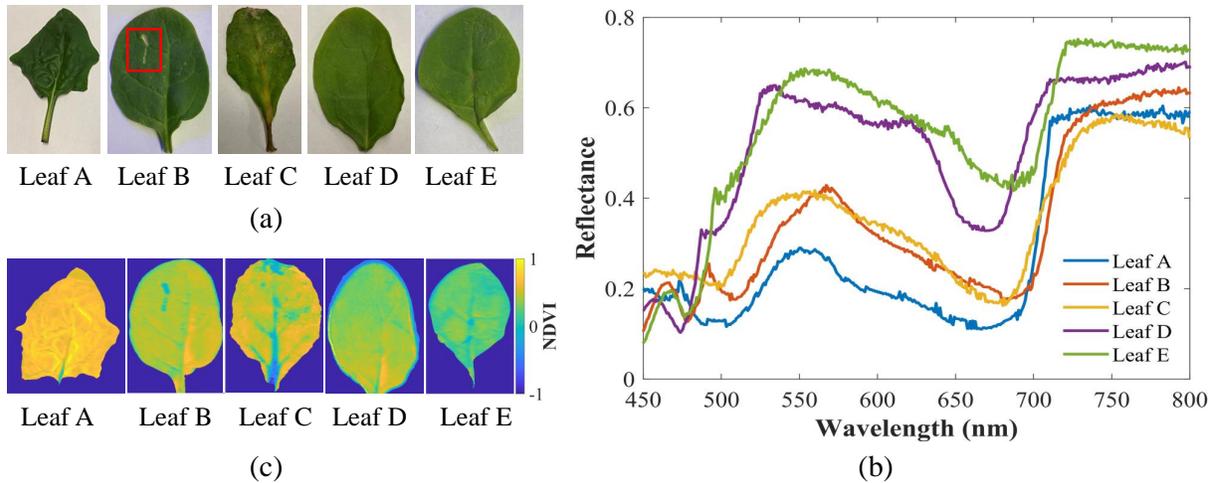


Figure 4.10: Reflectance and NDVI of spinach leaves at different fresh levels. (a) RGB images; (b) Reflectance; (c) NDVI images.

Besides spectral data for evaluating the average or overall leaves quality, spatial data might benefit local defect detection on leaves. In Fig. 4.11(a), there is an obvious defect on leaf B. As chlorophyll content in the defect region is less than the normal region, the defect region has a darker presentation in the NDVI image (Fig. 4.11(e)). As presented in Fig. 4.11(f), we cannot simply apply the NDVI image to detect the defect region, as the NDVI values at the vein and edge region are at the same level as the defect region. Besides the NDVI image, we noticed that depth change at the vein region and the edge is quite obvious in Fig. 4.11(b). Therefore, we can highlight the defect, edge, and vein region based on the NDVI image. Then based on depth data, the edge and vein region can be ruled out to leave only the defect on the highlighted image. In Fig. 4.11(g), we employed the k-nearest neighbors (KNN) algorithm (Cover & Hart, 1967) to classify leaf B into two groups, the dark parts for the healthy regions and the highlighted parts as potential candidates for local defects. Then depth gradient of leaf B was calculated and shown in Fig. 4.11(c). By setting up gradient threshold, image opening and closing operations, leaf vein and profile are shown as Fig. 4.11(d). Lastly, the highlighted vein and edge region were masked out from Fig. 4.11(g) by subtracting Fig. 4.11(d) and the defect region was finally identified as shown in Fig. 4.11(h). Essentially, reflectance value and NDVI image based on spectral data support us to filter non-fresh leaves, and 3D geometry is utilized to conduct regional analysis to distinguish defects from healthy

regions on leaves. As an initial exploration, only several spinach leaves were analyzed, but it shows that the 4D imaging technique could be a potential defect detection method for the food industry. In the future, we will investigate the generalization ability of the proposed method by testing on a higher volume of samples.

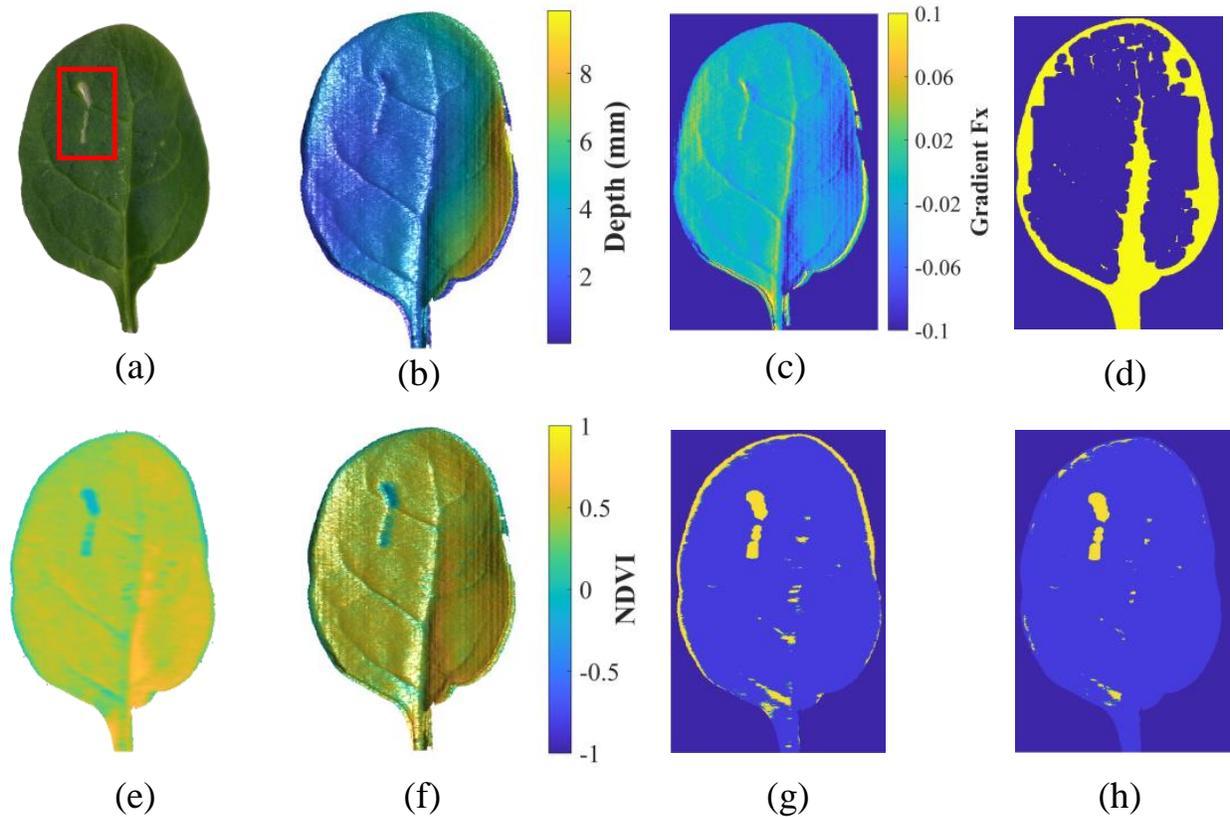


Figure 4.11: Defect detection on spinach leaves. (a) RGB image of leaf B with a defect; (b) Reconstructed 3D geometry of leaf B; (c) Depth gradient of leaf B; (d) Profile and vein of leaf B; (e) NDVI image (f) Reconstructed 3D geometry with NDVI value rendering; (g) Results of KNN method; (h) Marked defect.

4.5 Conclusion

We presented a 4D line-scan hyperspectral imaging system to simultaneously acquire 3D geometrical and spectral data of specimens with high spatial accuracy and spectral resolution. The system contains two parts, the moving part for specimen shifting and the static part for pattern projection

and image acquisition. Under the control of a microcontroller, the illuminated fringe images were acquired automatically. Then by using the proposed methods, we obtained 3D geometry and spectral information without data fusion. To verify 3D geometrical accuracy, we measured a standard sphere with a diameter of 40.234 ± 0.012 mm. The radius of the reconstructed sphere was 20.0901 mm, and its corresponding RMSE was 0.0895 mm compared with the actual radius. The measurement of the standard spherical object demonstrated that the system achieved high 3D geometrical accuracy. Besides, we measured a colored tortoise statue to present the intensity change along wavelength from 400 nm to 1000 nm. The intensity of a specific colored part achieved the highest at its corresponding wavelength region. In addition, several leaves with diverse fresh levels were measured to demonstrate the potential application of the system in the food industry. The average reflectance value and NDVI image based on spectral data helped generically identify non-fresh leaves, and the spatial data benefited local defect detection. Briefly, the system shows the potential to be employed in all green vegetation products.

Compared with published research, hyperspectral-depth data fusion was circumvented in our system, reducing computational complexities and avoiding fusion errors. Furthermore, precise 3D geometry was acquired based on the proposed methods through a calibration procedure developed. Except for these advantages, possible improvements can be made in the future to achieve higher measurement speed and better illumination. The current measurement speed is 3 mm/min due to the low-cost linear guide rail, which is much slower than the practical production line speed. If a more advanced translation stage is available, the scanning rate will be enhanced. With a higher rate, more specimens can be measured in a relatively short period of time. Thus, the system has the potential to be employed in a production line. Besides, in some wavelengths, the intensity of illuminated pattern images was relatively low, which decreased the signal-to-noise ratio of spectral images. With a better illumination source, spectral images with higher image quality can be obtained.

CHAPTER 5

4D VIS-SWIR LINE-SCAN HYPERSPECTRAL IMAGING¹

¹Li, Jiaqiong, et al. Accepted by *Optics Express*, October 11, 2024.

Reprinted here with permission of the publisher.

5.1 Abstract

This paper proposes a four-dimensional (4D) line-scan hyperspectral imaging system to acquire 3D spatial data and hyperspectral images covering from visible to short-wave infrared (Vis-SWIR) wavelength range. The system captures visible and near-infrared (VNIR) and SWIR hyperspectral images using two line-scan sensors, while 3D spatial data is acquired via a fringe projection profilometry subsystem. To align the VNIR and SWIR hyperspectral images, we utilize a line-scan homography method and propose a transformation method to register 3D spatial data with hyperspectral images. The mean reprojection error for hyperspectral image registration is 0.5396 pixels, and the registration of 3D spatial data with hyperspectral images achieves subpixel accuracy. Spatial accuracy is demonstrated with a root mean squared error (RMSE) of 0.1321 mm and a mean absolute error (MAE) of 0.1066 mm by measuring a standard sphere with a 20.0512 mm radius. The spectral resolutions are 11.2 nm in the VNIR range and 5 nm in the SWIR range. Two case studies were conducted: one involving a colorful object with rich features and colors, and another involving a potato before and after sprouting. Results from the measurement of a colorful object demonstrate the proposed system's registration accuracy and image intensity variation across wavelengths, while the potato study highlights the system's potential applications in the food industry.

5.2 Introduction

Hyperspectral imaging (HSI) captures and processes hundreds of continuous spectral bands across a wide electromagnetic spectrum, allowing it to acquire detailed information related to the material composition and chemical properties of objects (Chang, 2007). This capability has led to its application in various fields, including precision agriculture (Dale et al., 2013; B. Lu et al., 2020; Sethy et al., 2022), food quality and safety control (Y.-Z. Feng & Sun, 2012), medical diagnostics (Fei, 2019; G. Lu & Fei, 2014), and environmental monitoring (Stuart et al., 2019). In parallel, three-dimensional (3D) measurement techniques provide detailed geometrical information, depicting an object's size, shape, and surface topography. Techniques such as time of flight (Horaud et al., 2016), stereo vision (Hartley & Zisserman, 2003), laser triangulation (Dorsch et al., 1994), and structured light (Malacara, 2007a) are commonly used for 3D measurements. In addition to conventional 3D techniques, researchers have explored other approaches from spectral clues, such as a 3D camera

based on active laser light absorption of atmospheric oxygen (Kääriäinen & Seppä, 2024) and passive absorption-based hyperspectral thermal ranging (Dorken Gallastegi et al., 2023). However, these approaches still face limitations in speed or accuracy compared to conventional techniques. Combining HSI with 3D shape measurements, referred to as 4D hyperspectral imaging (incorporating spatial dimensions x , y , and z , along with the spectral dimension λ), offers significant advantages by integrating these complementary data modalities (H. Liu et al., 2020). This fusion allows for a more comprehensive analysis, capturing both material properties and geometric characteristics, and expands the potential for applications across a wide range of fields. Thus, this study will focus on 4D hyperspectral imaging.

Various 4D hyperspectral imaging systems have been developed to acquire both spectral images and 3D spatial data. These systems can be categorized into independent or dependent designs based on whether the sensors for the two data modalities interact or influence each other during data acquisition. Basically, dependent systems, where two data modalities are acquired by a single sensor or multiple sensors sharing the same optical path, feature a simple optical path. This configuration ensures that spectral images and 3D spatial data are conjugated without registration errors. However, dependent systems often use filters, prisms, or coded apertures to perform coupled acquisition of both spectral images and 3D geometric imaging. Such optical mechanisms usually have limitations in spectral resolution and range that are often restricted to the visible and near-infrared (VNIR) or short-wave infrared (SWIR) wavelength ranges. For example, Heist et al. (Heist et al., 2018a) proposed a 5D hyperspectral imaging system, capturing spatial coordinates (x , y , z), wavelength-dependent reflectance ($R(\lambda_i)$), and time (t). This system employs two hyperspectral snapshot cameras and a broadband high-speed pattern projector to acquire hyperspectral images in the VNIR range and 3D geometry over time. Due to the spectral resolution constraints of the hyperspectral snapshot cameras, the system is limited to 25 spectral bands. Similarly, Zhao et al. (H. Zhao et al., 2018) proposed a dual-path 4D hyperspectral imaging system based on an acousto-optical tunable filter (AOTF) and laser triangulation ranging to capture spectral images, with the spectral range constrained to 440 - 760 nm. Additionally, a coded aperture and dispersive prisms were utilized in a cross-modal stereo system to simultaneously capture spectral and spatial images (L. Wang et al., 2016b). Our research group previously proposed a 4D hyperspectral imaging based on a spectrograph and fringe projection profilometry without data registration error (J. Li et al., 2021). Though the spectral resolution is 2.8

nm, the spectral range is still restricted to the VNIR range due to the spectrograph. Similar to our previous system, Qi et al. (Qi et al., 2023) proposed 4D multiframe co-encoded spectral imaging based on fringe projection profilometry and an Amici prism, with the spectral range constrained to 420-660 nm. Luo et al. (Luo, Forsberg, & He, 2022) proposed a 5D-fusing imaging system that can capture objects' surface shape, spectral characteristics, and polarization states based on a polarized camera and a fringe projection profilometry (FPP) system with a liquid-crystal tunable filter. Limited by the liquid-crystal tunable filter, the proposed system can only achieve 10 nm spectral resolution in the visible light range. Utilizing a commercial time-of-flight sensor and Amici prism, Rueda et al. (Rueda et al., 2017) proposed a single aperture system that can acquire depth and hyperspectral images with a limited number of 16 spectral bands. Different from conventional methods that use filters and prisms within the optical path of spectral sensors, Xu et al. (Y. Xu et al., 2020b) proposed a hyperspectral projector to simultaneously acquire 3D spatial and spectral images by structural illumination based on a diffraction grating. However, their proposed method needs complex spectral and spatial encoding and reconstruction, and structural illumination stripes limit the spectral channels. In addition to limitations in spectral resolution and range, the scanning speed may also be affected in some configurations, as the scanning speed of one modality can directly influence the other. For example, in our previous work (J. Li et al., 2021), the scanning speed was constrained to 3 mm/min, limiting its practicality. This restriction was due to a single camera working in line-scan mode to capture hyperspectral images and 3D data. Besides, the application of fringe projection profilometry required multiple fringe patterns captured in each movement. We also had to limit the scanning speed to avoid motion blur in acquired 3D spatial data. Similar speed constraints due to the need for multiple frames in 3D reconstruction have been observed in other configurations, such as those described in (Qi et al., 2023) and (Luo, Forsberg, & He, 2022), where multiple frames were required for accurate 3D reconstruction.

Unlike systems that use a single sensor or sensors sharing the same optical path, independent systems typically acquire spectral images and spatial data using independent sensors, making data registration necessary. Since the two data modalities come from different sources, the spectral resolution, spectral range, and spatial accuracy depend on the respective sensors. Additionally, 3D spatial data can be obtained from various sources, including LiDAR sensors, Microsoft's Kinect, stereo vision systems, and structured light, further adding to the flexibility of the system. For instance,

Tolt et al. (Tolt et al., 2011b) combined 24-band hyperspectral images from a hyperspectral push-broom sensor with corresponding LiDAR data by detecting shadows in remote sensing. Monteiro et al. (Monteiro et al., 2013b) used SIFT and SURF feature descriptors to register hyperspectral images with LiDAR point clouds. Terrestrial LiDAR has been used to register 3D point clouds with hyperspectral images, aiding applications that require linked geometric and chemical information (Buckley, Kurz, & Schneider, 2012). Microsoft's Kinect V2 has also been utilized to capture 3D spatial data and register it with hyperspectral images (P. Huang et al., 2018b; L. Zhang et al., 2020). Additionally, Luo et al. (Luo, Forsberg, Fu, & He, 2022) employed a line structured light stereo vision system, a line-scan hyperspectral imager, and a line laser to acquire 3D spatial data and fluorescence spectral images. They also developed a 4D dual-mode staring hyperspectral-depth imager that can acquire reflectance and fluorescence spectra in the VNIR range and 3D spatial data by registering 3D spatial data from stereo vision with spectral images from a hyperspectral staring scanner (Luo, Forsberg, Fu, Xing, et al., 2022). Beyond 3D sensors or stereo vision systems, Chen et al. (X. Chen et al., 2023) designed a light-sheet profilometer based on the Scheimpflug principle to estimate the distance between sensors and the object, registering spectral data and 3D point clouds based on the mapping relationship between the two sensors. Generally, independent systems require the registration of 3D spatial data with spectral images, which can introduce mismatching errors. However, the independence of different sensors allows these systems to achieve a broader spectral range, higher spectral resolution, and faster scanning speeds.

Adopting an independent design, we propose a 4D line-scan hyperspectral imaging system that acquires 3D spatial data and hyperspectral images that can cover a wide spectrum range from visible to short-wave infrared (Vis-SWIR) range. Raw slit images are captured by two separate line-scan hyperspectral imagers as the object moves along a conveyor belt. After a fixed-distance movement, the object stops, and a fringe projection profilometry system scans it for 3D data acquisition. The data processing has been illustrated in Fig. 5.1. First, the raw slit images are stitched together. Using the calibrated relationship between pixel index and wavelength, we separately obtain hyperspectral images in the VNIR (400-1000 nm) and SWIR (1000-1700 nm) wavelength ranges. Then, to align the VNIR and SWIR hyperspectral images, we employ a line-scan homography method. On the other side, 3D spatial data is obtained through fringe projection profilometry. After acquiring both the 3D spatial data and hyperspectral images, we apply the proposed transformation method to

register them to obtain hyperspectral 3D data. To verify the registration accuracy, we scanned an asymmetric circular grid at different poses, using the circular centers as feature points. The mean reprojection error for mapping VNIR and SWIR hyperspectral image registration is 0.5396 pixels, outperforming conventional homography with a mean reprojection error of 3.4975 pixels. The mean absolute error (MAE) and root mean square error (RMSE) for the registration of 3D spatial data with VNIR hyperspectral images are 0.6716 pixels and 0.4929 pixels, respectively. The registration of 3D spatial data with SWIR hyperspectral images also achieves subpixel accuracy. Additionally, the RMSE and MAE for the 3D reconstruction of a standard sphere with a 20.0512 mm radius are 0.1321 mm and 0.1066 mm, respectively. The spectral resolutions are 11.2 nm in the VNIR range and 5 nm in the SWIR range. After evaluating the proposed system, we conducted an experiment on a UNO box with rich features and colors to show the proposed system’s performance. We also implemented the proposed system to detect sprouts in a potato, highlighting its potential in the food industry.

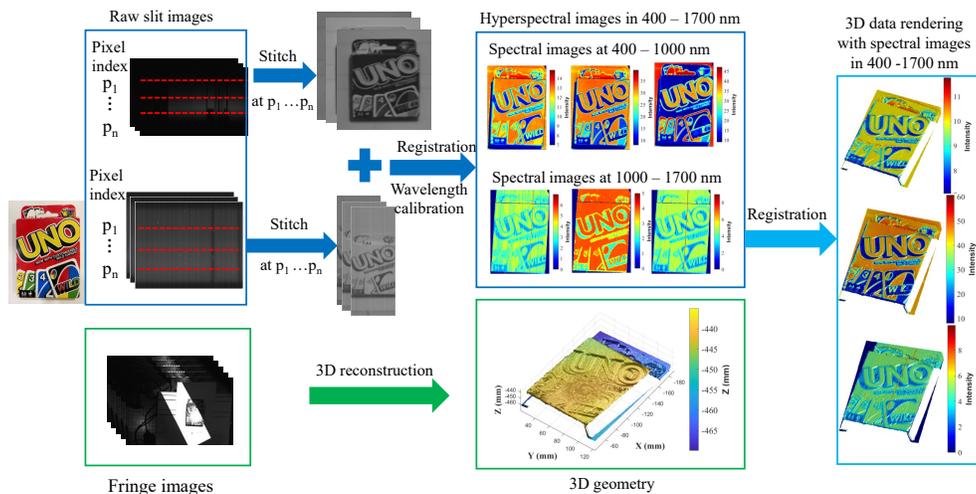


Figure 5.1: Data processing pipeline of the proposed system to acquire hyperspectral 3D data.

The paper is organized as follows: Section 5.3 introduces the proposed method of 4D Vis-SWIR line-scan hyperspectral imaging, covering the line-scan hyperspectral imaging model, registration of VNIR and SWIR spectral images, 3D measurement using fringe projection profilometry, and the registration of 3D data with hyperspectral images. In Section 5.4, we first assess the spatial accuracy, and then quantitatively evaluate the accuracy of the proposed registration method. We also conduct the spectral calibration. In Section 5.5, we conducted two study cases: a UNO box with

multiple colors and rich features and a potato before and after scanning. Section 5.6 summarizes the contributions and limitations of the proposed methods.

5.3 Methods

5.3.1 Line-scan hyperspectral imaging

Fig. 5.2 illustrates the schematic diagram of the proposed line-scan hyperspectral imaging (HSI) subsystem. The subsystem comprises two synchronized line-scan HSI cameras covering the VNIR and SWIR wavelength ranges, respectively. Each is positioned with the optical axis perpendicular to the conveyor belt plane. As the conveyor belt moves at a constant velocity, perpendicular to the direction of the slit inside the ImSpector attached to the cameras, the cameras capture complete scenes of the samples passing through their fields of view. We set up the line-scan camera coordinate system as depicted in Fig. 5.2 (a). A 3D point $P(X, Y, Z)^T$ in the camera coordinate system is projected to $p(u, v)^T$ in the image plane coordinate system. The line-scan camera follows perspective projection (Draréni et al., 2011) along the slit direction, which can be depicted as

$$u = \frac{fX + p_u Z}{Z}, \quad (5.1)$$

where f and p_u are the focal length and the principal point along the slit direction, respectively. As the conveyor belt moves at a constant speed, the v coordinate is related to Y by a scaling factor S_v that depends on the conveyor belt's speed and the camera's acquisition frequency (Draréni et al., 2011), illustrated as

$$v = S_v Y. \quad (5.2)$$

5.3.2 Registration of VNIR and SWIR hyperspectral images

We use two synchronized line-scan cameras to capture HSI in the VNIR (400-1000 nm) and SWIR (900-1700 nm) wavelength ranges. As shown in Fig. 5.2 (b), the view planes of the two cameras are not aligned perfectly that the 3D point $P_1(X_1, Y_1, Z_1)^T$ captured by the SWIR line-scan camera has to move to $P_2(X_2, Y_2, Z_2)^T$ to be scanned by the other camera. Thus, the corresponding 2D projection $p_1(u_1, v_1)^T$ and $p_2(u_2, v_2)^T$ in image plane will be different. To obtain hyperspectral

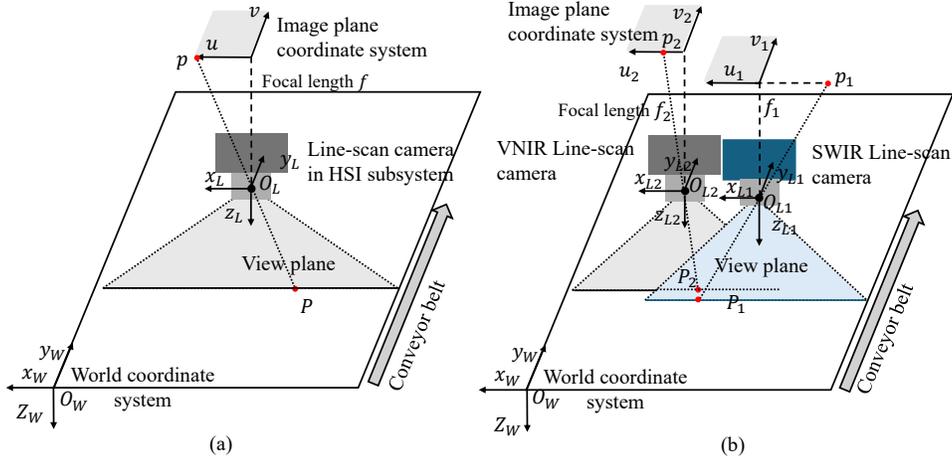


Figure 5.2: Schematic diagram of the line-scan hyperspectral imaging (HSI) subsystem, including two synchronized line-scan HSI cameras and a conveyor belt. (a). A typical line-scan HSI camera; (b). Two line-scan HSI cameras are used in the proposed system; the left is for the VNIR wavelength range (400 - 1000 nm), and the right is for the SWIR wavelength range (900 - 1700 nm). In practice, the view planes of the two cameras are not aligned perfectly, and the 3D point P_1 in the right camera's view plane will translate to P_2 to be captured by the left camera.

images from 400 nm to 1700 nm, registration of HSI images from the two cameras is necessary, and geometric transformation between the projections p_1 and p_2 needs to be explored. The relationship between points P_1 and P_2 can be given by

$$\mathbf{P}_1 = \mathbf{R}_{21}\mathbf{P}_2 + \mathbf{t}_{21}, \quad (5.3)$$

where \mathbf{R}_{21} and \mathbf{t}_{21} are the rotation matrix and translation vector from point P_2 to P_1 . In the following steps, we will align the world coordinate system with the VNIR line-scan camera's coordinate system to simplify analysis. By combining Eq. 5.1 and Eq. 5.2, we can represent the VNIR line-scan camera using:

$$\mathbf{K}_2 = \begin{pmatrix} f_2 & 0 & p_2 & 0 \\ 0 & S_v & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad (5.4)$$

$$\begin{pmatrix} Z_2 u_2 \\ v_2 \\ Z_2 \\ 1 \end{pmatrix} = \mathbf{K}_2 \begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \\ 1 \end{pmatrix}, \quad (5.5)$$

where \mathbf{K}_2 represents the intrinsic matrix of the VNIR line-scan camera. It is important to note that the left side of Eq. (5.5) is not homogeneous because the coordinate v_2 does not conform to the projective projection, differing from a typical area-scan camera model. The above equation can also be rewritten as

$$\begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{Z_2}{f_2} & 0 & -\frac{Z_2 p_2}{f_2} \\ 0 & \frac{1}{S_v} & 0 \\ 0 & 0 & Z_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_2 \\ v_2 \\ 1 \end{pmatrix}. \quad (5.6)$$

Similarly, the SWIR line-scan camera can be depicted as

$$\mathbf{K}_1 = \begin{pmatrix} f_1 & 0 & p_1 & 0 \\ 0 & S_v & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (5.7)$$

$$\begin{pmatrix} S u_1 \\ v_1 \\ S \\ 1 \end{pmatrix} = \mathbf{K}_1 \begin{pmatrix} \mathbf{R}_1 & \mathbf{t}_1 \\ \mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{pmatrix}. \quad (5.8)$$

In the above equation, S is a scale factor, and \mathbf{R}_1 and \mathbf{t}_1 are the rotation matrix and translation vector from the world coordinate system (VNIR line-scan camera coordinate system) to the SWIR line-scan camera coordinate system. As the two cameras are synchronized, the scaling factor of v coordinate S_v is the same for the two cameras. Combing Eq. 5.3 with the above equation, we can get the relationship between the point P_2 and the projection p_1 in the SIWR line-scan camera

image plane:

$$\begin{pmatrix} Su_1 \\ v_1 \\ S \end{pmatrix} = \mathbf{K}_1 \begin{pmatrix} \mathbf{R}_1 & \mathbf{t}_1 \\ \mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R}_{21} & \mathbf{t}_{21} \\ \mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \\ 1 \end{pmatrix}. \quad (5.9)$$

Substituting Eq.5.6 into the above equation, we can obtain the relationship between the projections (Fang et al., 2022):

$$\begin{pmatrix} Su_1 \\ v_1 \\ S \end{pmatrix} = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{pmatrix} \begin{pmatrix} u_2 \\ v_2 \\ 1 \end{pmatrix} \quad (5.10)$$

Here, We emphasize that registering images captured by line-scan cameras differs from conventional image registration using homography (Hartley & Zisserman, 2003) because the left side term in the above equation is not homogeneous. The Matrix $\mathbf{H} = \{h_i\}$ can also be treated as the line-scan homography that registers VNIR hyperspectral images with SWIR hyperspectral images.

5.3.3 FPP for 3D data acquisition

To acquire 3D spatial data of samples, we build a fringe projection profilometry (FPP) subsystem consisting of a digital projector and a camera. The eighteen-step phase-shifting algorithm (Malacara, 2007a) is utilized, and the captured image intensity at the pixel (x, y) can be illustrated as

$$I_i(x, y) = I_{amb}(x, y) + I_{mod}(x, y) \cos [\phi(x, y) - 2\pi i/N], \quad (5.11)$$

where $I_i(x, y)$, $I_{amb}(x, y)$, and $I_{mod}(x, y)$ are image intensity, ambient light intensity, and intensity modulation at the pixel (x, y) , respectively. Here, N is the phase step, equal to 18, and i notes the i -th step. $\phi(x, y)$ is the wrapped phase, which can be computed by

$$\phi(x, y) = \tan^{-1} \left[\frac{\sum_{i=0}^{N-1} I_i(x, y) \sin 2\pi i/N}{\sum_{i=0}^{N-1} I_i(x, y) \cos 2\pi i/N} \right], \quad (5.12)$$

The wrapped phase essentially depicts the object's 3D geometry through phase distortion determined by the object's 3D profile. Therefore, phase is crucial to retrieve the object's 3D geometry. However,

the wrapped phase only ranges in $[-\pi, \pi)$, containing discontinuities due to the arctangent function. To remove the discontinuities, we utilize the Gray-code unwrapping method (Sansoni et al., 1999) to acquire the absolute phase:

$$\Phi(x, y) = \phi(x, y) + 2\pi k(x, y). \quad (5.13)$$

Here, $\Phi(x, y)$ represents the continuous absolute phase, free from any discontinuities, while $k(x, y)$ denotes the fringe order. Once the absolute phase is determined, the 3D geometry of the object (X_w, Y_w, Z_w) can be reconstructed with the calibrated FPP system (S. Zhang & Huang, 2006).

5.3.4 Registration of 3D geometry and hyperspectral images

In the proposed system, an object translates at a constant velocity along the conveyor belt and stops after a distance to be captured by the frame-scan camera in the FPP subsystem, as illustrated in Fig. 5.3. To obtain 4D (x, y, z, λ) data of an object, we need to register 3D spatial data with 2D HSI images.

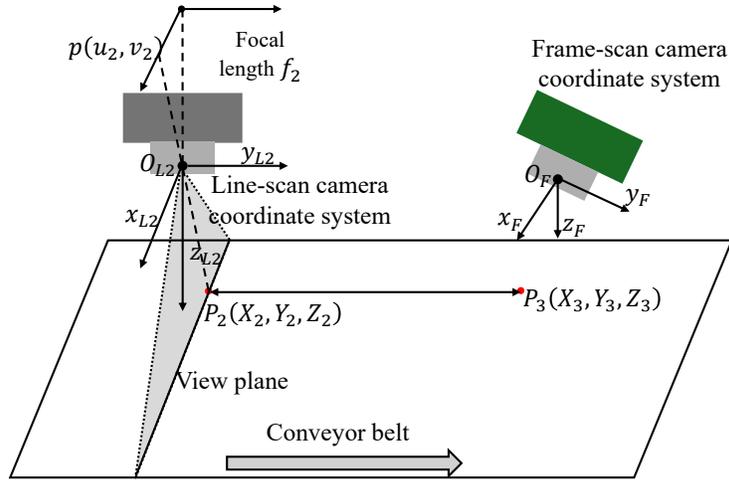


Figure 5.3: Schematic diagram of the scanning process: the line-scan camera captures the object along the X-direction as it moves in the Y-direction; after traveling a distance, the object stops, and the frame-scan camera starts to scan it.

Following the above setup, we align the world coordinate system to the VNIR line-scan camera coordinate. Combining Eq. 5.1 and Eq. 5.2, we can get the relationship between the homogenous pixel point $p(u_2, v_2, 1)^T$ and 3D point $P_2(X_2, Y_2, Z_2)^T$ as

$$\begin{pmatrix} u_2 \\ v_2 \\ 1 \end{pmatrix} = \frac{1}{Z_2} \begin{pmatrix} f_2 X_2 + p_2 Z_2 \\ s Y_2 Z_2 \\ Z_2 \end{pmatrix} \sim \begin{pmatrix} f_2 & 0 & p_2 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_2 \\ Y_2 Z_2 \\ Z_2 \end{pmatrix}. \quad (5.14)$$

Here, we can notice that the above equation contains a quadratic cross term $Y_2 Z_2$ because the v coordinate of a line-scan camera does not follow the perspective projection. The 3D point $P_2(X_2, Y_2, Z_2)^T$ in the line-scan camera view plane translates to the point $P_3(X_3, Y_3, Z_3)^T$. Thus, we can have:

$$\begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \end{pmatrix} = \mathbf{R}_{\mathbf{FL}} \begin{pmatrix} X_3 \\ Y_3 \\ Z_3 \end{pmatrix} + \mathbf{t}_{\mathbf{FL}} = \begin{pmatrix} r_{11} X_3 + r_{12} Y_3 + r_{13} Z_3 + t_x \\ r_{21} X_3 + r_{22} Y_3 + r_{23} Z_3 + t_y \\ r_{31} X_3 + r_{32} Y_3 + r_{33} Z_3 + t_z \end{pmatrix}. \quad (5.15)$$

In the above equation, the 3D point $P_3(X_3, Y_3, Z_3)^T$ is in the frame-scan camera image plane and is captured for the 3D reconstruction. The rotation matrix $\mathbf{R}_{\mathbf{FL}}$ and translation vector $\mathbf{t}_{\mathbf{FL}}$ depict the transformation from the frame-scan camera coordinate system to the VNIR line-scan camera coordinate system. Substituting Eq. 5.14 to Eq. 5.15, we will get the relationship between the pixel point $(u_2, v_2, 1)^T$ in the hyperspectral images and 3D point $(X_3, Y_3, Z_3)^T$ as

$$\begin{pmatrix} u_2 \\ v_2 \\ 1 \end{pmatrix} \sim \begin{pmatrix} m_1 & m_2 & m_3 & m_4 & 0 & \dots & 0 \\ m_5 & m_6 & m_7 & m_8 & m_9 & \dots & m_{14} \\ m_{15} & m_{16} & m_{17} & m_{18} & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} X_3, Y_3, Z_3, 1, X_3^2, Y_3^2, Z_3^2, X_3 Y_3, X_3 Z_3, Y_3 Z_3 \end{pmatrix}^T. \quad (5.16)$$

$\mathbf{M}_{3 \times 10}$

In the equation above, the geometric transformation matrix \mathbf{M} establishes the relationship between a 3D point and its projection in the hyperspectral image. This matrix is determined by the VNIR line-scan camera's intrinsic matrix \mathbf{K}_2 , rotation matrix $\mathbf{R}_{\mathbf{FL}}$, and translation vector $\mathbf{t}_{\mathbf{FL}}$. Unlike the homography described in (Z. Zhang, 2000), the transformation matrix here is 3×10 because the right side expression includes quadratic and cross-term components, represented as $(X_3, Y_3, Z_3, \dots, Y_3 Z_3)^T$. To register 3D data with hyperspectral images, we need to get the solution of the transformation matrix \mathbf{M} . By multiplying Eq. 5.16 with the cross-product skew matrix $(u_3, v_3, 1)^T$, we will get:

$$\mathbf{A}_{3 \times 18} \mathbf{m} = \mathbf{0}, \quad (5.17)$$

where $\mathbf{m}^T = (m_1, m_2, \dots, m_{18})$ consists of eighteen non-zero elements of \mathbf{M} , and \mathbf{A} is a 3×18 matrix that can be defined by its submatrices \mathbf{A}_1 , \mathbf{A}_2 , and \mathbf{A}_3 as follows:

$$\begin{aligned}\mathbf{A}_1 &= \begin{pmatrix} 0, 0, 0, 0, -X_3, -Y_3, -Z_3, -1, -X_3^2, -Y_3^2, -Z_3^2 \\ -X_3Y_3, -X_3Z_3, -Y_3Z_3, X_3v_2, Y_3v_2, Z_3v_2, v_2 \end{pmatrix}^T, \\ \mathbf{A}_2 &= \begin{pmatrix} X_3, Y_3, Z_3, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -X_3u_2, -Y_3u_2, -Z_3u_2, -u_2 \end{pmatrix}^T, \\ \mathbf{A}_3 &= \begin{pmatrix} -X_3v_2, -Y_3v_2, -Z_3v_2, -v_2, X_3u_2, Y_3u_2, Z_3u_2, u_2 \\ X_3^2u_2, Y_3^2u_2, Z_3^2u_2, X_3Y_3u_2, X_3Z_3u_2, Y_3Z_3u_2, 0, 0, 0, 0 \end{pmatrix}^T.\end{aligned}$$

Based on the above equation, \mathbf{A} can be directly computed based on $P_3(X_3, Y_3, Z_3)^T$ and $(u_2, v_2, 1)^T$. With the known matrix \mathbf{A} , \mathbf{M} can be estimated using the singular value decomposition (SVD) algorithm. Affected by noises in 3D reconstruction and the projection in hyperspectral images, the solution from SVD may not be accurate, and we utilize the Levenberg-Marquardt algorithm (Moré, 2006) to optimize it. Based on Eq. 5.16, the reprojected 2D pixel point $\tilde{p}_2(\tilde{u}_2, \tilde{v}_2)$ can be calculated by

$$\begin{cases} \tilde{u}_2 = \frac{m_1X_3 + m_2Y_3 + m_3Z_3 + m_4}{m_{15}X_3 + m_{16}Y_3 + m_{17}Z_3 + m_{18}}, \\ \tilde{v}_2 = \frac{m_5X_3 + m_6Y_3 + m_7Z_3 + \dots + m_{14}Y_3Z_3}{m_{15}X_3 + m_{16}Y_3 + m_{17}Z_3 + m_{18}}. \end{cases} \quad (5.18)$$

The nonlinear optimization objective function to minimize the reprojection error is as follows:

$$e(\mathbf{M}) = \min \left(\sum_{i=1}^m \sum_{j=1}^n \text{Dist}(p_{2,ij}, \tilde{p}_{2,ij}) \right), \quad (5.19)$$

where $p_{2,ij}$ is a feature point captured by the hyperspectral line-scan camera at (i, j) , $\tilde{p}_{2,ij}$ is the corresponding reprojected feature point, and the operation $\text{Dist}(\cdot)$ is to compute the Euclidean distance. The Jacobian matrix for the implementation of the Levenberg-Marquardt algorithm is described as

$$\mathbf{J}_{2 \times 18} = \begin{pmatrix} \frac{\partial u_2}{\partial m_i} \\ \frac{\partial v_2}{\partial m_i} \end{pmatrix}. \quad (5.20)$$

Here, we present the complete method for the mentioned image registration. From n scans of an asymmetric circular calibration grid:

- Extract the feature points (circular centers) in the VNIR and SWIR hyperspectral images.
- Estimate the line-scan homography \mathbf{H} and S in Eq. 5.10 for all n scans using the direct linear transform algorithm (Hartley & Zisserman, 2003).
- Retrieve the 3D data of all scans and extract the feature points in the images captured by the camera in the FPP subsystem.
- Estimate the transformation matrix \mathbf{M} using SVD and optimize the solution using the Levenberg-Marquardt algorithm.

5.4 Experimental Evaluation

5.4.1 system setup

The experimental setup, depicted in Fig. 5.4, consists of two separate subsystems: the HSI and the 3D scanning subsystems. In the HSI subsystem, an indium gallium arsenide (InGaAs) camera (Xenics, Xeva-1.7-320-TE1-USB-100) is synchronized with a complementary metal-oxide semiconductor (CMOS) camera (FLIR, Grasshopper3 GS3-U3-41C6C-C), both operating at a scanning frequency of 10Hz . The InGaAs camera has an image resolution of 320×256 pixels, while the CMOS camera's resolution is set as 1920×1200 pixels. The InGaAs camera, equipped with an ImSpector N17E (Specim, ImSpector N17E), captures hyperspectral images in the 900 - 1700 nm range. Meanwhile, the CMOS camera, paired with an ImSpector V10 (Specim, ImSpector V10), captures hyperspectral images in the 400 - 1000 nm range. Besides, a halogen lamp (model: Philips LED, Philips 415604) is used to provide sufficient illumination. In the 3D scanning subsystem, a CMOS camera (model: FLIR Grasshopper3 GS3-U3-41C6C-C) and a digital light processing (DLP) projector (model: DLP Lightcrafter 4500) form the fringe projection profilometry system for 3D data acquisition. Additionally, a conveyor belt (Dorner, 22EDM16-026320A010101), driven by a Nema 34 step motor, moves objects from the HSI subsystem to the 3D scanning subsystem at a constant velocity, stopping after a distance for 3D data acquisition.

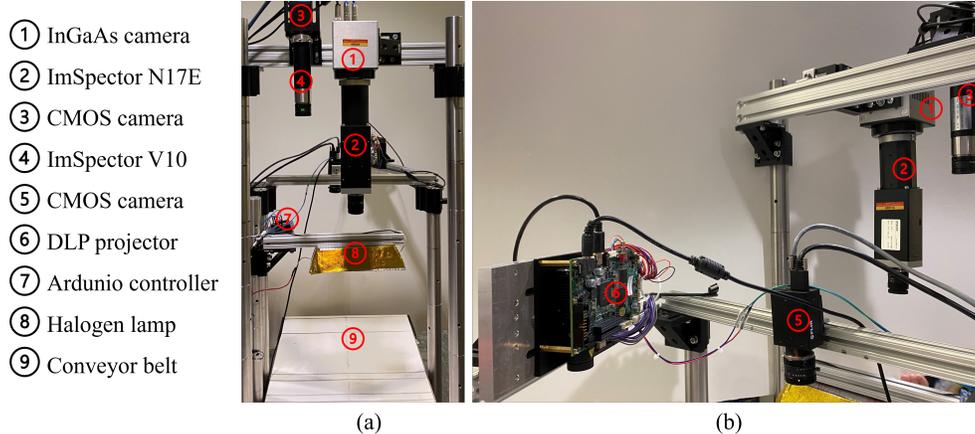


Figure 5.4: Experimental setup: The hyperspectral imaging subsystem consists of an InGaAs camera paired with an ImSpector N17E and a CMOS camera paired with an ImSpector V10, both synchronized; the 3D scanning subsystem includes a DLP projector and a CMOS camera (labeled as 5); a conveyor belt is utilized to translate samples constantly, and a halogen lamp acts as an extra light source. (a) Front view; (b) Side view.

5.4.2 Evaluation of hyperspectral image registration

The VNIR and SWIR hyperspectral images are acquired by two separate cameras, which makes it necessary to register them. We scanned an asymmetric circular grid, and the circular centers were detected using the toolbox OpenCV(Bradski, 2000) as feature points for this purpose. Fig. 5.5 (a) and (e) show spectral images captured by the VNIR line-scan camera at a wavelength of 700 nm for two different poses, while Fig. 5.5 (b) and (f) show spectral images captured by the SWIR line-scan camera at a wavelength of 1,347 nm simultaneously. For better visualization, we increased the overall image intensity of both spectral images. Using the proposed method, we aligned the SWIR HSI images to the VNIR HSI images, with results presented in Fig. 5.5 (c) and (g). Here, we present the registration results for spectral images at 700 nm and 1,347 nm as representative examples, as the line-scan spectral image registration at other wavelengths is the same. The brighter areas in Fig. 5.5 (c)-(d) and (g)-(h) indicate the aligned spectral images at 1,347 nm. Red dots indicate the feature points detected in the VNIR spectral images at 700 nm, while blue dots note feature points from the aligned SWIR image at 1,347 nm. In Fig.5.5 (c) and (g), most of all feature points are well-aligned without visible differences. The root mean square error (RMSE) and mean absolute error (MAE)

metrics, which represent the pixel distance between the detected circular centers in the VNIR and SWIR images, are used to quantitatively assess the registration performance. RMSE and MAE for Fig. 5.5 (c) are 0.83323 and 0.4919 pixels, and 0.8258 and 0.5336 pixels for Fig.5.5 (g). For comparison, we also registered the same spectral images using conventional homography (Hartley & Zisserman, 2003), and results are shown in Fig. 5.5 (d) and (h) where mismatches between feature points are visible, especially in Fig. 5.5 (h). RMSE and MAE are 1.3254 and 0.8414 pixels for Fig. 5.5 (d), and 3.4614 and 2.2545 pixels for Fig. 5.5 (h). For three scans of the asymmetric circular grid, the mean reprojection error is 0.5396 pixels using the proposed method, while 3.4975 pixels using conventional homography (Hartley & Zisserman, 2003). The results presented in Fig. 5.5 demonstrate that the proposed method can register hyperspectral images well and outperforms conventional homography (Hartley & Zisserman, 2003). This improvement can be attributed to the proposed method's ability to account for the line-scan nature of the sensors used in this study, whereas conventional homography is more suited for area-scan sensors.

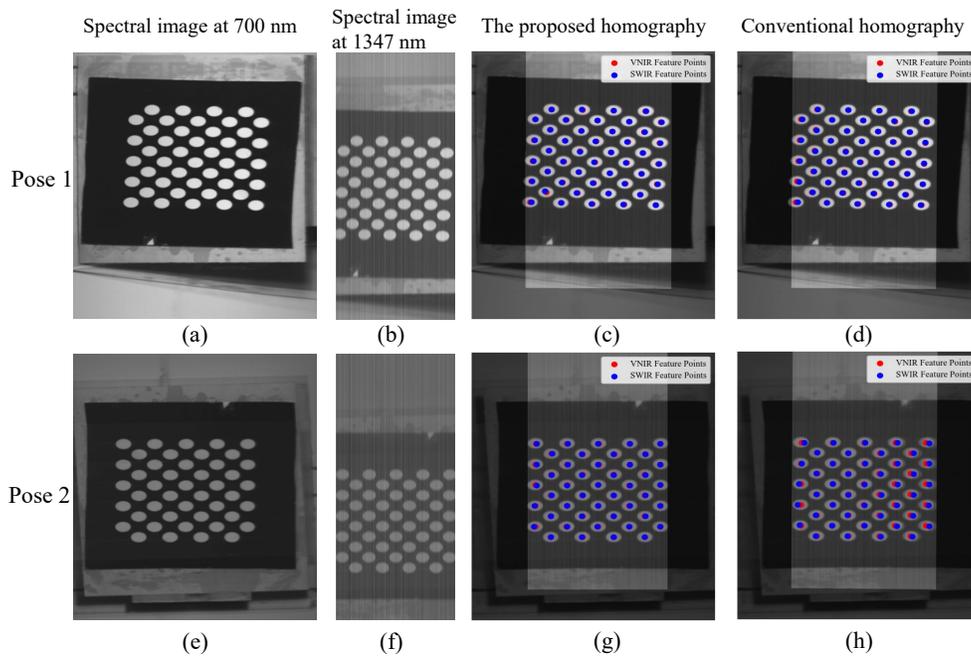


Figure 5.5: Registration results of spectral images of an asymmetric circular grid. (a) and (e) spectral images at 700 nm. (b) and (f) Spectral images at 1347 nm. (c) Proposed method on (a) and (b) with RMSE 0.8332 pixels, MAE 0.4919 pixels. (d) Conventional homography (Hartley & Zisserman, 2003) on (a) and (b) with RMSE 1.3254 pixels, MAE 0.8414 pixels. (g) Proposed method on (e) and (f) with RMSE 0.8258 pixels, MAE 0.5336 pixels. (h) Conventional homography on (e) and (f) with RMSE 3.4614 pixels, MAE 2.2545 pixels.

5.4.3 Spatial accuracy

A plastic standard sphere was measured to verify the geometric accuracy of the proposed method. The actual diameter of the sphere, measured using a digital caliper with eight consecutive measurements, is 40.1025 ± 0.0900 mm. We then scanned the sphere and acquired its 3D data using FPP. Fig. 5.6(a) shows the overlay of the reconstructed 3D geometry and a standard sphere, represented in yellow and blue, respectively. The reconstructed sphere's radius is 20.0046 mm, with a measurement error of 0.2326% . Fig. 5.6(b) displays the error map between the reconstructed geometry and the standard sphere, with an RMSE of 0.1321 mm and an MAE of 0.1066 mm. To better visualize the error map, a cross-section along the X-axis is shown in Fig. 5.6(c). This experiment results demonstrate the proposed system's high geometric accuracy, which is also essential for registering 3D spatial data with hyperspectral images.

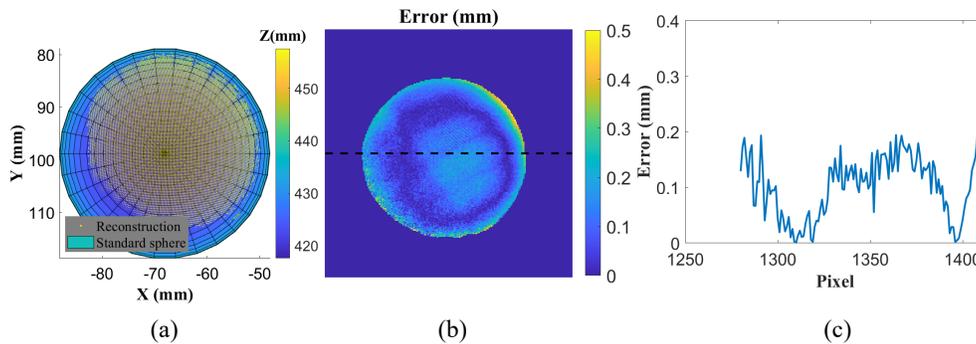


Figure 5.6: Evaluation of geometric accuracy. (a) Overlay of the reconstructed sphere (shown in yellow) and a standard sphere (shown in blue) with a 20.0512 mm radius. (b) Error map between the reconstructed sphere and standard sphere (RMSE: 0.1321 mm, MAE: 0.1066 mm). (c) Cross-section of the error map at the X-axis middle position.

5.4.4 Evaluation of the registration of 3D data and hyperspectral images

To obtain hyperspectral 3D data (spatial coordinates (x, y, z) and spectral dimension λ), we registered 3D spatial data with hyperspectral images. We used the same asymmetric circular grid as the scanned target in this process. We scanned the grid at seven different poses with one scan per pose, and VNIR hyperspectral images are registered with 3D data for each pose using the proposed method. Our method achieved an average MAE and RMSE of 0.6716 pixels and 0.4929 pixels in this process. Three registration results are presented in Fig. 5.7 to visualize the proposed method performance.

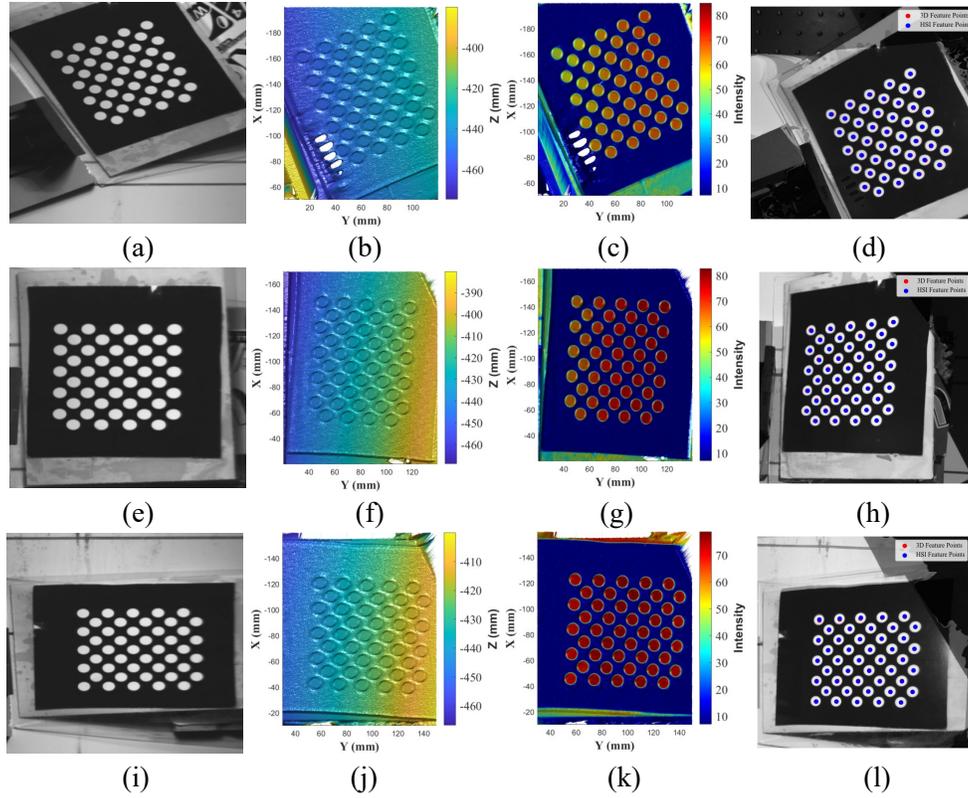


Figure 5.7: Registration results of 3D data and VNIR spectral images. (a), (e), and (i) Spectral images at 700 nm of an asymmetric circular grid in different poses. (b), (f), and (j) Corresponding 3D data. (c), (g), and (k) 3D data rendered with aligned spectral images at 700 nm. (d), (h), and (l) Differences between aligned spectral and FPP grayscale images, with RMSE and MAE values of 0.7566 and 0.6917 pixels, 0.5086 and 0.4304 pixels, and 0.7092 and 0.4132 pixels.

Similarly, we also registered 3D spatial data with SWIR hyperspectral images using the proposed method. Limited by the InGaAs camera's field view, we only scanned the same asymmetric circular grid at three poses, and the registration results are illustrated in Fig. 5.8. The registration results showed RMSE and MAE values of 0.6302 and 0.4529 pixels, 0.6714 and 0.5417 pixels, and 1.1951 and 1.035 pixels, respectively. Besides, we also increased the SWIR spectral image intensity here to enhance feature point detection and visualization. Overall, the registration error for both 3D and VNIR spectral images, as well as 3D and SWIR spectral images, was able to achieve a sub-pixel level, demonstrating the method's effectiveness.

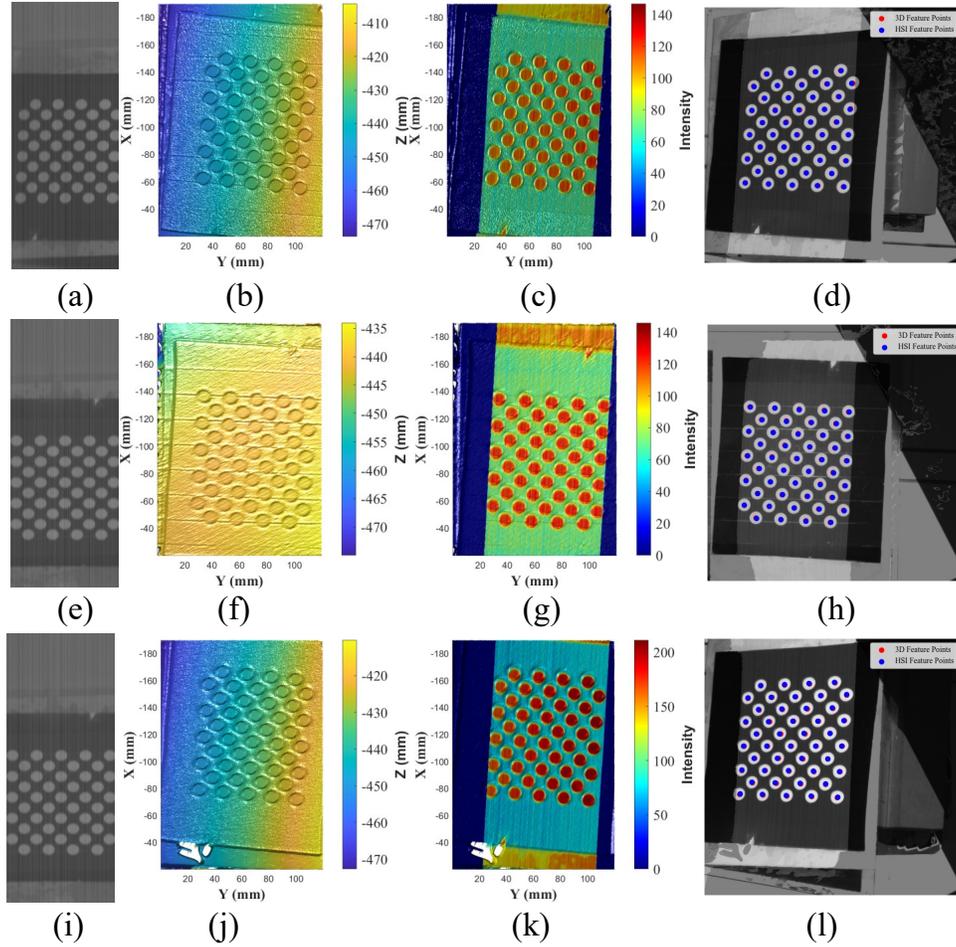


Figure 5.8: Registration results of 3D data and SWIR spectral images. (a), (e) and (i) Spectral images at 1347 nm of an asymmetric circular grid in different poses. (b), (f), and (j): Corresponding 3D data. (c), (g), and (k) 3D data rendered with aligned spectral images at 1347 nm. (d) and (h) Differences between aligned spectral and FPP grayscale images, with RMSE and MAE values of 0.6302 and 0.4529 pixels, 0.6714 and 0.5417 pixels, and 1.1951 and 1.035 pixels.

5.4.5 Spectral Calibration

We conduct the spectral calibration to determine the wavelength λ based on the pixel index p . According to the specifications of ImSpector V10 and N17E, the relationship between pixel index and wavelength is a third-order polynomial:

$$\lambda = I + C_1 p + C_2 p^2 + C_3 p^3. \quad (5.21)$$

To determine the constant parameters I , C_1 , C_2 , and C_3 for the two ImSpectors, we utilize specific wavelength calibration light sources. The ImSpector V10 is paired with a Mercury Argon light source (model: Ocean Optics, HG-2) to acquire the spectral image shown in Fig. 5.9(a). Similarly, the ImSpector N17E is paired with an Argon light source (model: Ocean Optics, AR-2) to acquire the spectral image shown in Fig. 5.9 (b). These calibration sources emit strong, narrow spectral lines, allowing us to determine the pixel index. Using the wavelength information provided in the specifications, we estimated the constant parameters in the third-order polynomial, as shown in the Table. 5.1. For the VNIR wavelength range, the root mean square error (RMSE) and mean absolute error (MAE) of fitting are 1.9422 nm and 1.3211 nm, less than the ImSpector V10 spectral resolution, 11.2 nm. For the SWIR wavelength range, RMSE and MAE are 4.1439 nm and 3.2900 nm, which are also below the spectral resolution of 5 nm for the ImSpector N17E. Besides, the R^2 values for both fits are over 0.9, indicating good fitness of the third-order polynomial. Because the constant parameters C_2 and C_3 are much smaller than C_1 and C_2 , the relationship between the wavelength and pixel index is nearly linear in VNIR and SWIR wavelength ranges, as illustrated in Fig. 5.9 (c).

Table 5.1: Fitness of the third-order polynomial relationship wavelength vs pixel index

	I	C₁	C₂	C₃	RMSE (nm)	MAE (nm)	R²
VNIR	228.4341	0.7158	$1.4895e-4$	$-5.2393e-8$	1.9422	1.3211	0.9999
SWIR	903.5601	4.8128	-0.0043	$4.9844e-6$	4.1439	3.2900	0.9996

5.5 Two study cases

5.5.1 Hyperspectral 3D scanning of a UNO box

We scanned a UNO box with multiple colors and rich features and obtained its hyperspectral 3D data to show spectral image intensity change along wavelengths and registration accuracy. A, B, C, and D areas on the UNO box are in red, yellow, green and blue. The mean image intensity of these areas change along wavelengths in the VNIR and SWIR ranges are illustrated in Fig. 5.10 (f) and (k). Besides, the spectral image intensity in the SWIR range is much lower than that in

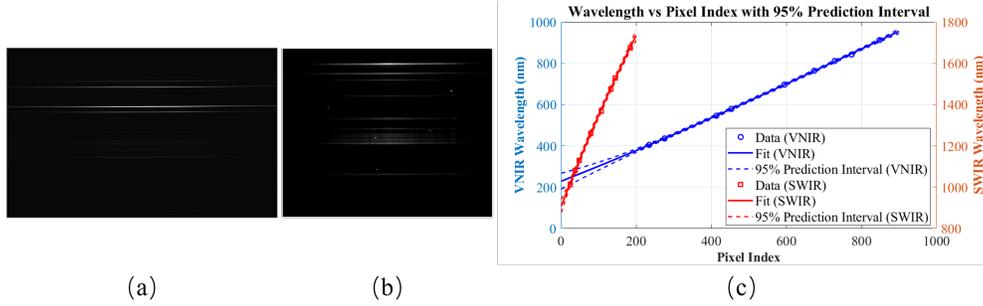


Figure 5.9: Spectral calibration. (a) Spectral image of the Mercury Argon wavelength calibration light source. (b) Spectral image of the Argon wavelength calibration light source. (c) Third-order polynomial relationship between wavelength and pixel with 95% prediction interval. The blue line indicates the nearly linear relationship for the VNIR wavelength range and the red line for the SWIR wavelength range.

the VNIR range, and some vertical line noises exist in the SWIR images due to inhomogeneous illumination and different image sensors and optical lenses. As presented in Fig. 5.10 (b) - (e) and (g)-(j), features (letters, numbers, and lines) in spectral images are aligned well with corresponding 3D reconstructions without visible mismatches, which supports the registration accuracy of the proposed method.

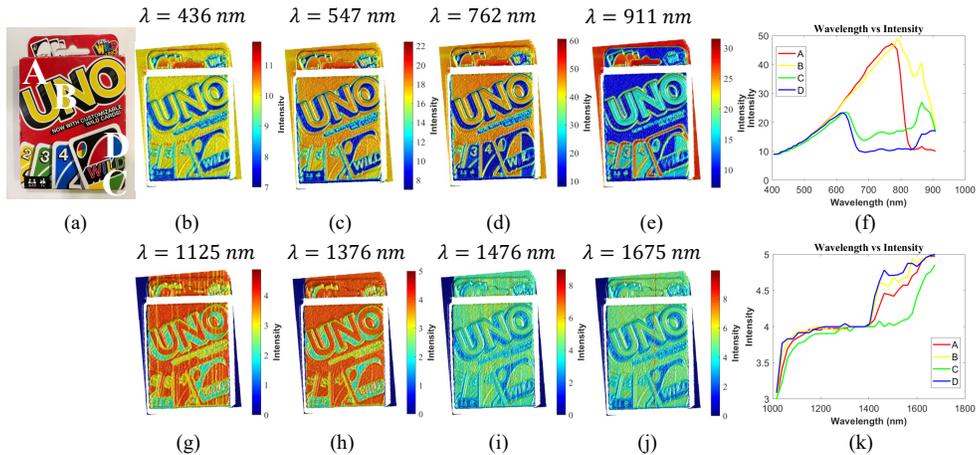


Figure 5.10: Hyperspectral 3D scanning of a UNO box. (a) RGB image. (b)-(e) 3D renderings with spectral images at 436 nm, 547 nm, 762 nm, and 911 nm in the VNIR range. (f) Spectral image intensity changes in areas A, B, C, and D from 400-900 nm. (g)-(j) 3D renderings with spectral images at 1125 nm, 1376 nm, 1476 nm, and 1675 nm in the SWIR range. (k) Spectral image intensity changes in areas A, B, C, and D from 1000-1600 nm.

5.5.2 Detecting potato sprouts

Hyperspectral imaging has been extensively explored in the food industry for measuring various physicochemical properties, such as reflectance, transmittance, and fluorescence, to evaluate food quality (Zhu et al., 2020). Complementary to HSI, 3D geometrical measurement can also enhance food quality assessment with additional 3D data. This study presents a case of detecting potato sprouts using the proposed 4D Vis-SWIR hyperspectral line-scan system, demonstrating its potential in the food industry. We scanned a potato and acquired its hyperspectral 3D data before and after it sprouted. Using the hyperspectral images, we computed the relative reflectance as follows:

$$I_{ref} = \frac{I_{raw} - I_d}{I_w - I_d}, \quad (5.22)$$

where I_{ref} , I_{raw} , I_d , and I_w are reflectance, raw spectral image intensity, spectral image intensity of a dark reference, and spectral image intensity of a white reference, respectively. The comparison of reflectance change before and after the sprout is shown in Fig. 5.11, where the first row and the second row are the captured spectral data before and after sprouting, respectively. The red A, blue B, and green C areas are the sprout, normal area, and sprout eye, respectively. Comparing the reflectance before and after sprouting, the reflectance of area A has visible changes in the VNIR and SWIR ranges, while the reflectance of areas B and C remains relatively unchanged. This observation is further supported by the two plots of reflectance versus wavelength before and after sprouting presented in Fig. 5.11 (i) and (j). These findings indicate that hyperspectral images are effective in illustrating the reflectance change of potatoes before and after sprouting.

Hyperspectral imaging (Y. Gao et al., 2018; Peraza-Alemán et al., 2024) and 3D imaging (Q. Su et al., 2017, 2020) have each been employed independently to assess potato quality. In this work, we present a case to support that the integration of hyperspectral imaging with 3D imaging provides a more effective approach for detecting potato sprouts. Specifically, we used the combination of reflectance at 580 nm and depth data as a representative example. As shown in Fig. 5.12 (b) and (c), the reflectance map at 580 nm was classified into seven groups using the k-means algorithm (Lloyd, 1982), with the eighth label representing the image background. Sprout candidates were identified based on a threshold, which was set to 5 in this study, as shown in Fig. 5.12 (d). However, directly detecting sprouts from the reflectance at 580 nm proves to be challenging, as the reflectance

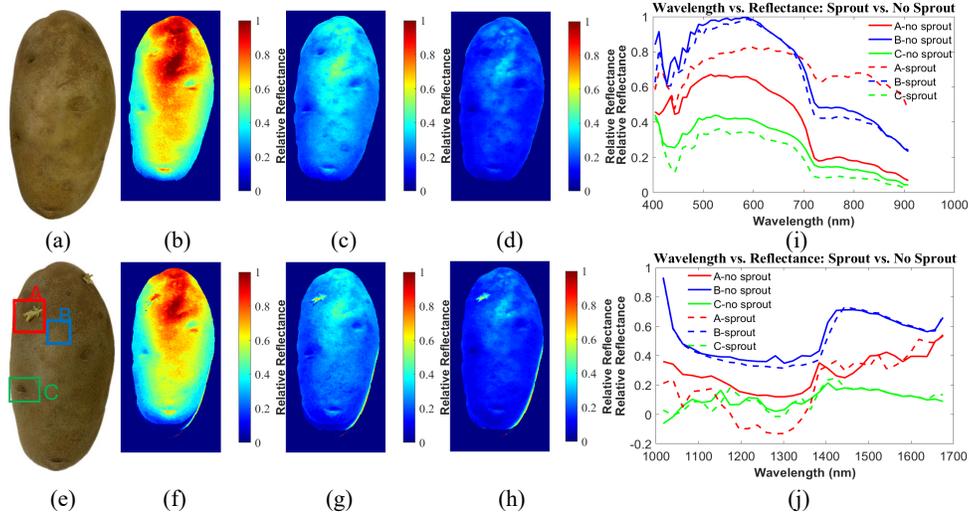


Figure 5.11: Reflectance changes of potato before and after sprouting. (a) RGB image of a potato before sprout. (b) - (d) Reflectance at 580 nm, 772 nm, and 874 nm. (e) RGB image of the same potato after sprout. The red A, blue B, and green C areas are the sprout, normal area, and sprout eye. (f) - (h) Reflectance of the potato with sprouts at 580 nm, 772 nm, and 874 nm. (i) Comparison of reflectance changes before and after sprout in the VNIR range. (j) Comparison of reflectances before and after sprout in the SWIR range.

values of the sprouts are similar to those of surrounding areas. Given that sprouts cause local abrupt depth changes, we computed depth derivatives along the x-axis and y-axis, as illustrated in Fig. 5.12 (e) and (f). In these figures, the depth derivatives in the sprouting area are distinguishable and higher compared to neighboring areas. Significant depth changes are also observed at the potato edges. Similarly, we classified the depth derivatives along the x-axis and y-axis into seven groups and sorted the groups in descending order according to the derivative values, as depicted in Fig. 5.12 (h) and (i). Based on the prior knowledge that the depth of the sprout area varies dramatically, we identified areas where the number of groups exceeds the threshold in both x-axis and y-axis classifications as sprout candidates with depth, as shown in Fig. 5.12. These candidates are mainly concentrated in the sprouting area and at the edges. By combining sprout candidates from the reflectance and depth (using an "AND" operation between Fig. 5.12 (d) and (j)), we successfully localized the potato sprouts while mitigating the impact of edges and regions with similar reflectance values in the surrounding areas. In addition, we tested sprout detection using reflectance at 772 nm along with the depth map, which produced comparable results (shown in Fig. 5.13), further

demonstrating the robustness of the integrated approach. Besides, the depth map is used to segment the potato from the background. The sprout detection result is shown in Fig. 5.12 (k). Although only one sprouted potato was analyzed and the basic K-means clustering method was utilized in this initial exploration, the results indicate that the proposed system has the potential to be an effective method for evaluating food quality. Future work will focus on testing the generalization ability of the proposed method with a larger volume of samples. With the availability of high-volume hyperspectral and 3D imaging data, we can develop specialized neural networks tailored for more comprehensive and robust food quality assessments.

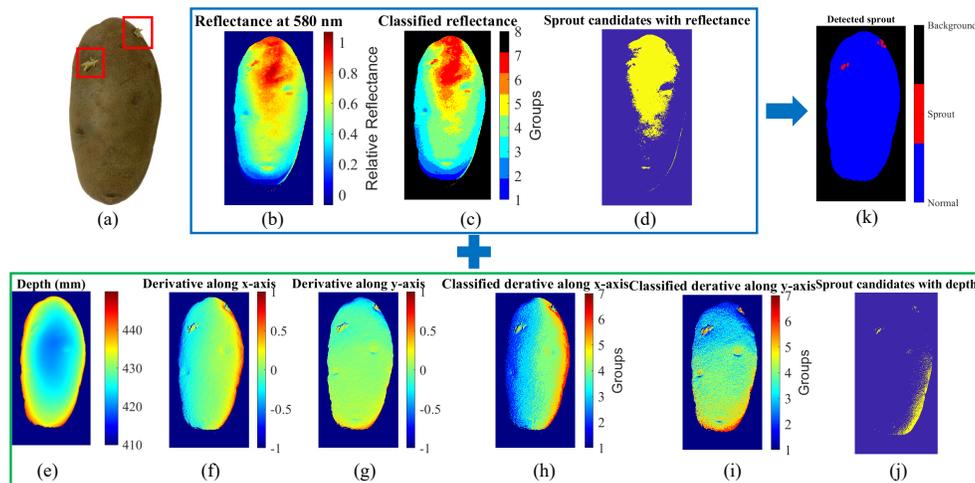


Figure 5.12: Sprout detection based on reflectance at 580 nm and depth map. (a) A potato with sprouts. (b) Reflectance at 580 nm. (c) classified reflectance using the K-means method. (d) Sprout candidates with reflectance. (e) Depth map. (f) and (g) Depth derivatives along the x-axis and y-axis. (h) and (i) classified derivatives along the x-axis and y-axis using the K-means method. (j) Sprout candidates with depth. (k) Detected sprouts.

5.6 Conclusions

We presented a 4D Vis-SWIR line-scan hyperspectral imaging system designed to acquire 3D spatial and spectral data of objects with high spatial accuracy and spectral resolution. Key innovations include the line-scan homography method for registering VNIR and SWIR images and a transformation method for aligning 3D spatial data with hyperspectral images from line-scan sensors. Our system registers hyperspectral images in a wide range of 400-1700 nm leveraging the spectral data acquired from two separate cameras. The registration of 3D spatial data with hyperspectral images

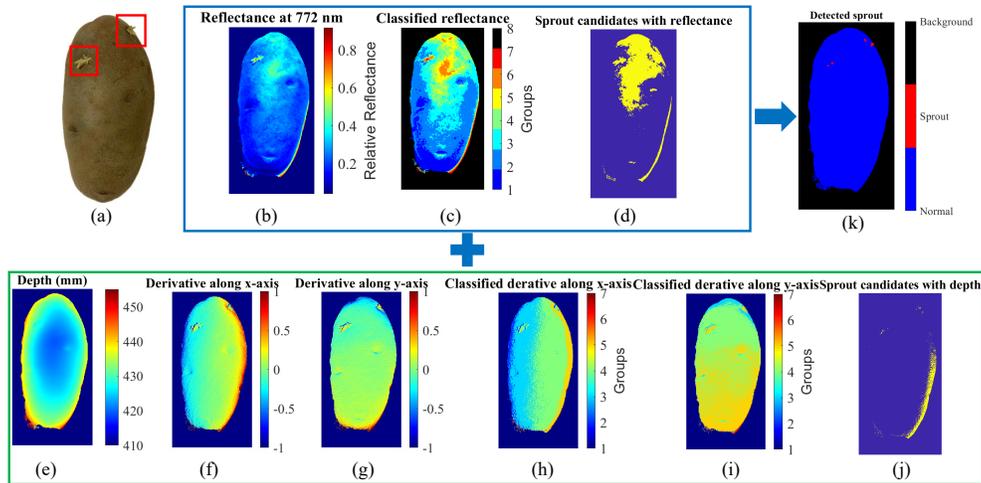


Figure 5.13: Sprout detection based on reflectance at 772 nm and depth map, yielding similar results. (a) The same potato with sprouts. (b) Reflectance at 772 nm. (c) classified reflectance using the K-means method. (d) Sprout candidates with reflectance. (e) Depth map. (f) and (g) Depth derivatives along the x-axis and y-axis. (h) and (i) classified derivatives along the x-axis and y-axis using the K-means method. (j) Sprout candidates with depth. (k) Detected sprouts.

achieves subpixel accuracy. Additionally, the spectral resolution is 11.2 nm in the VNIR range and 5 nm in the SWIR range. The spatial accuracy for measuring a standard sphere with a 20.0512 mm radius is RMSE 0.1321 mm and MAE 0.1066 mm. Our experimental results demonstrate the system's performance on objects with rich features and colors and its potential in the food industry.

Compared with existing research, the proposed system covers a broader spectral range from 400 nm to 1700 nm, expanding its potential applications in fields such as agriculture and food quality assessment. Additionally, the ability to adjust the scanning speed of each subsystem independently enhances its flexibility for practical applications. However, the field of view is limited by the overlap of the two line-scan cameras, and registration errors are unavoidable. Furthermore, at certain wavelengths, the low intensity of spectral images reduces the signal-to-noise ratio. Improved illumination could enhance the image quality of spectral images. In conclusion, the proposed 4D hyperspectral imaging system offers significant advancements in spectral range and resolution, spatial accuracy, and practical applicability despite some limitations. Future work should focus on enhancing illumination and reducing registration errors to further improve system performance.

CHAPTER 6

SUMMARY AND FUTURE WORK

6.1 Summary of research contributions

This dissertation focuses on robust 4D hyperspectral imaging and its applications, and the contributions include robust 3D measurement with fringe projection profilometry and the integration of 3D spatial data with hyperspectral images.

- **Developed circular fringe projection profilometry for the system calibration and 3D reconstruction to enhance measurement accuracy:** In fringe projection profilometry, fringe pattern direction and system calibration play crucial roles in measurement accuracy. This dissertation introduces a novel calibration method utilizing circular and radial fringe patterns and a 3D reconstruction technique only with circular fringe patterns. The key contributions include demonstrating that the proposed circular fringe projection profilometry (CFPP) method enhances 3D reconstruction accuracy compared to traditional linear FPP, particularly in capturing complex geometries with abrupt depth changes. Through a series of rigorous 2×2 experiments, we provided comprehensive comparisons that highlight the strengths of our method. Additionally, the sensitivity and error analysis offers valuable insights into the factors affecting calibration performance, specifically the roles of increased sensitivity and estimated phase errors. This work has been published in the journal of *Optics Continuum* (J. Li & Li, 2024), and the details have been discussed in Chapter 2.
- **Investigated texture-guided phase-to-depth networks to repair shadow-induced errors in FPP:** Occlusions caused by complex object surfaces present a significant challenge in fringe projection profilometry (FPP). The intrinsic triangular configuration between the camera, projector, and object often results in shadow regions with invalid phase information,

leading to holes in the reconstructed 3D geometry. Both the texture images and phase information can provide critical clues that aid in estimating the 3D profile. Building on this prior knowledge, this dissertation proposes texture-guided phase-to-depth networks to address the shadow-induced errors in FPP. The proposed method integrates phase maps and texture images at two stages to enhance depth prediction accuracy. A specified loss function that considers image edge details and structural similarity is formulated to optimize model training. Besides, the digital twin technique is utilized to generate sufficient simulated data to train the proposed network. Unlike other approaches that employ multiple cameras or projectors to alleviate shadow effects, this method utilizes only a single camera-projector setup, minimizing additional hardware costs. Besides, by preserving the triangular configuration of the camera and projector, the proposed method maintains the flexibility of the FPP system and avoids geometric constraints associated with uniaxial configuration. The proposed method enhances the robustness of 3D reconstruction in complex scenes, particularly in regions affected by severe occlusions, by effectively utilizing texture cues to compensate for missing phase information. This work has been published in the journal of *Photonics* (J. Li & Li, 2023), and the details have been elaborated further in Chapter 3.

- **Proposed a 4D VNIR line-scan hyperspectral imaging system for simultaneous acquisition of hyperspectral images and 3D spatial data without registration.** 3D spatial data and spectral images are two data modalities that provide complementary insights into an object's characteristics across different domains. 3D imaging shows an object's spatial attributes, including position, shape, size, surface texture, and roughness, while spectral imaging furnishes detailed information on the object's physical and chemical properties, such as reflectance, absorption, and fluorescence. Extending the data representation from three dimensions to four dimensions (x , y , z , and wavelength λ) enhances the capability of capturing both spatial and spectral characteristics simultaneously, thereby broadening the applicability in diverse fields, including quality assessment in the food industry, agricultural monitoring, and environmental remote sensing. This dissertation proposes a novel 4D line-scan hyperspectral imaging system that can simultaneously capture 3D spatial data and hyperspectral images in the visible-to-infrared (VNIR) wavelength range. By utilizing a single camera for the acquisition of hyperspectral images and 3D scanning, the system inherently eliminates

the need for complex data alignment. Therefore, it avoids potential mismatching errors and preserves data integrity. The proposed system is rigorously evaluated and demonstrates high spatial accuracy along with superior spectral resolution. Several case studies, such as the detection of defects in spinach leaves, are conducted to illustrate the system's potential for practical applications, showcasing its versatility and effectiveness in real-world scenarios. The proposed method has been published in the journal of *Optics Express* (J. Li et al., 2021), and the details have been discussed in Chapter 4.

- **Proposed a 4D Vis-SWIR line-scan hyperspectral imaging system with a decoupled system architecture to extend the spectral range and enhance scanning speed:** Shortwave infrared (SWIR) hyperspectral imaging plays a critical role in numerous applications due to its ability to capture valuable spectral information beyond the VNIR spectrum. In light of this, this dissertation proposes a system that captures hyperspectral images over a broader wavelength range. In Chapter 4, a 4D visible and near-infrared (VNIR) line-scan hyperspectral imaging system is developed, which acquires both hyperspectral images and 3D spatial data using a single camera. Although this configuration enables simultaneous capture and avoids complex data registration, it limits the spectral range to the visible and near-infrared (VNIR) wavelengths and the scanning speed to 3 mm/min. To overcome these limitations, this dissertation introduces a 4D Vis-SWIR line-scan hyperspectral imaging system that employs a decoupled system architecture, wherein hyperspectral imaging and 3D spatial data acquisition are performed independently through two subsystems. The hyperspectral imaging subsystem with two line-scan imagers can capture hyperspectral images covering from visible to shortwave infrared wavelength range. Meanwhile, the 3D imaging subsystem scans objects for 3D reconstruction. The key innovations of this method include a novel line-scan homography method for precise registration of VNIR and SWIR hyperspectral images and a transformation algorithm for accurate alignment of 3D spatial data with hyperspectral images. The integration of 3D spatial data with hyperspectral images is achieved with subpixel accuracy, ensuring high-quality data fusion. In addition, the decoupled system architecture allows the system to cover a spectral range from 400 nm to 1700 nm, which provides a broader scope than existing systems, and it enables the system to adjust the scanning speed of each subsystem independently, further increasing the system's practicality. This advanced

methodology is already accepted by the journal of *Optics Express* and is elaborated upon in Chapter 5.

6.2 Future work

Based on the proposed methods in this dissertation, we can further explore the applications of 4D hyperspectral imaging in various fields. Besides, as the proposed 4D hyperspectral imaging systems are aimed at indoor applications, there is a necessity to investigate a 4D hyperspectral imaging system for outdoor scenarios.

6.2.1 Application of 4D hyperspectral imaging

In addition to the simple cases presented in this dissertation, the proposed 4D hyperspectral imaging systems have potential applications in various other fields, such as cultural heritage preservation, industrial quality control, and household waste classification.

In the field of cultural heritage, accurate documentation and analysis of artifacts such as colorful statues and sculptures are crucial for preservation and restoration efforts. The 3D data acquired by the system can capture the precise geometry details of these artifacts, providing valuable information about their structural integrity and any physical damage they may have sustained. Simultaneously, the hyperspectral images can be utilized to analyze the pigments and materials of the artifacts, which allows us to identify the original pigments, detect overpainting, and understand the degradation processes. Such detailed spatial and spectral information can assist conservators in making informed decisions regarding restoration techniques and preservation strategies. In industrial settings, hyperspectral imaging can be employed to classify the levels of metal corrosion by detecting subtle spectral differences associated with different stages of oxidation and material degradation. on the other side, the 3D data can be utilized to analyze the surface roughness and topography of metal components. By integrating spatial and spectral data, manufacturers can achieve a more thorough evaluation of product quality, leading to improved reliability and performance of industrial components. In household waste classification, hyperspectral imaging can identify and categorize different types of household waste based on their spectral uniqueness, distinguishing materials such as plastics, metals, paper, organic matter, and glass. Additionally, the 3D spatial data can be used to estimate the volume and shape of waste items, providing insights into the volume ratio

of each waste type. Therefore, the integration of hyperspectral and 3D imaging is also beneficial in this field.

Furthermore, the extensive spectral data acquired by 4D hyperspectral imaging systems enable the application of deep learning-based methods. The abundance of spectral images provides a rich dataset for developing and training specific deep-learning models tailored to various tasks. By leveraging these models, it is possible to enhance the analysis and interpretation of both spatial and spectral information, thereby advancing the capabilities of 4D hyperspectral imaging in diverse applications.

6.2.2 4D hyperspectral imaging system for outdoor applications with enhanced robustness and portability

Outdoor applications, such as intelligent crop monitoring and environmental surveillance, increasingly demand 4D hyperspectral imaging systems that can capture both spatial and spectral data under complex and dynamic field conditions. In outdoor environments, factors such as fluctuating weather, varying illumination, and environmental noise bring significant challenges in acquiring robust and accurate 4D hyperspectral data. Besides, in outdoor use, a 4D hyperspectral imaging system is usually mounted on carrier platforms like drones or ground-based robots, which requires the system design to be compact and portable. The previously proposed 4D hyperspectral imaging systems are primarily designed for indoor applications, where experimental conditions are relatively controlled and stable. The bulky system setup of these two systems also makes them unsuitable for mobile outdoor applications. Furthermore, the fringe projection profilometry utilized in these systems is inherently better suited for indoor environments due to its measurement range and sensitivity to object surface reflectance conditions. Therefore, to meet the requirements of outdoor scenarios, a novel 4D hyperspectral imaging system tailored for outdoor scenarios with enhanced robustness and portability is necessary.

Several existing methods aim to generate 4D hyperspectral data in outdoor scenarios by combining spatial and spectral imaging techniques. One common approach involves integrating 3D spatial data from Time-of-Flight (ToF) cameras, stereo vision systems, LiDAR, or laser scanners with spectral images captured by snapshot hyperspectral cameras or line-scan methods. For instance, LiDAR provides high-precision depth information, which, when fused with spectral data,

enhances the accuracy of scene reconstruction in outdoor environments (Kahraman & Bacher, 2021). Additionally, spectral information itself can be utilized for range measurement. For instance, by modeling the wavelength-dependent thermal attenuation by the air, researchers developed a passive, texture-independent depth measurement method with far range (Nagase et al., 2022). Learning-based methods have also emerged as promising tools for generating 4D hyperspectral data. Approaches like Neural Radiance Fields (G. Chen et al., 2024) and 3D Gaussian splatting (Di et al., 2024; Sinha et al., 2024) can render complex scenarios utilizing spectral images. However, despite these advancements, challenges remain. Data registration-based methods inherently suffer from mismatches between 3D spatial data and hyperspectral images. The integration of multiple sensors can lead to increased system complexity and reduced portability, making them less suitable for drone or robot-mounted applications. Learning-based methods often require substantial computational resources and extensive training data, which may not be practical in field conditions.

In the proposed 4D Vis-SWIR hyperspectral imaging system, we utilized two line-scan cameras to capture hyperspectral images spanning wavelengths from 400 to 1700 nm. Actually, the two line-scan cameras can consist of a stereo vision system to acquire 3D spatial data and hyperspectral images at the same time. By utilizing the stereo vision method for 3D data acquisition, the new system will become more suitable for outdoor applications, as it alleviates the depth range limitations inherent in fringe projection profilometry. Besides, the elimination of the bulky 3D imaging subsystem used in the previously proposed method reduces the system's complexity and enhances its compactness, making it more conducive to integration with carrier platforms. However, implementing this method is challenging. Generating disparity maps and estimating accurate depth from the two line-scan cameras are difficult due to the limited overlap of their fields of view. Besides, system calibration is not straightforward, especially in outdoor scenarios where feature points with known geometric information are scarce.

Bibliography

- Abate, A. F., Nappi, M., Riccio, D., & Sabatino, G. (2007). 2d and 3d face recognition: A survey. *Pattern recognition letters*, 28(14), 1885–1906.
- Alenyà, G., Foix, S., & Torras, C. (2014). Tof cameras for active vision in robotics. *Sensors and Actuators A: Physical*, 218, 10–22.
- An, Y., Hyun, J.-S., & Zhang, S. (2016). Pixel-wise absolute phase unwrapping using geometric constraints of structured light system. *Optics express*, 24(16), 18445–18459.
- Ariana, D. P., & Lu, R. (2008). Quality evaluation of pickling cucumbers using hyperspectral reflectance and transmittance imaging: Part i. development of a prototype. *Sensing and Instrumentation for Food Quality and Safety*, 2(3), 144–151.
- Behmann, J., Mahlein, A.-K., Paulus, S., Kuhlmann, H., Oerke, E.-C., & Plümer, L. (2015). Calibration of hyperspectral close-range pushbroom cameras for plant phenotyping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 106, 172–182.
- Benalcazar, D. P., Bastias, D., Perez, C. A., & Bowyer, K. W. (2019). A 3d iris scanner from multiple 2d visible light images. *IEEE Access*, 7, 61461–61472. <https://doi.org/10.1109/ACCESS.2019.2915786>
- Beraldin, J.-A., Carrier, B., MacKinnon, D., & Cournoyer, L. (2012). Characterization of triangulation-based 3d imaging systems using certified artifacts. *NCSLI Measure*, 7(4), 50–60.
- Bojanowski, P., Joulin, A., Lopez-Paz, D., & Szlam, A. (2017). Optimizing the latent space of generative networks. *arXiv preprint arXiv:1707.05776*.
- Boldrini, B., Kessler, W., Rebner, K., & Kessler, R. W. (2012). Hyperspectral imaging: A review of best practice, performance and pitfalls for in-line and on-line applications. *Journal of near infrared spectroscopy*, 20(5), 483–508.

- Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.
- Bruning, J. H., Herriott, D. R., Gallagher, J. E., Rosenfeld, D. P., White, A. D., & Brangaccio, D. J. (1974). Digital wavefront measuring interferometer for testing optical surfaces and lenses. *Applied optics*, 13(11), 2693–2703.
- Buckley, S., Kurz, T., & Schneider, D. (2012). The benefits of terrestrial laser scanning and hyperspectral data fusion products. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 39, 541–546.
- Buckley, S., Kurz, T., Schneider, D., & Sensing, R. (2012). The benefits of terrestrial laser scanning and hyperspectral data fusion products. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 39(B7), 541–546.
- Carré, P. (1966). Installation et utilisation du comparateur photoélectrique et interférentiel du bureau international des poids et mesures. *Metrologia*, 2(1), 13.
- Chang, C.-I. (2007). *Hyperspectral data exploitation: Theory and applications*. John Wiley & Sons.
- Chaudhuri, S., & Rajagopalan, A. N. (2012). *Depth from defocus: A real aperture imaging approach*. Springer Science & Business Media.
- Chen, F., Su, X., & Xiang, L. (2010). Analysis and identification of phase error in phase measuring profilometry. *Optics express*, 18(11), 11300–11307.
- Chen, G., Narayanan, S. K., Ottou, T. G., Missaoui, B., Muriki, H., Pradalier, C., & Chen, Y. (2024). Hyperspectral neural radiance fields. *arXiv preprint arXiv:2403.14839*.
- Chen, X., Huang, X., & He, S. (2023). 4d hyperspectral surface topography measurement system based on the scheimpflug principle and hyperspectral imaging. *Applied Optics*, 62(33), 8855–8868.
- Chéné, Y., Rousseau, D., Lucidarme, P., Bertheloot, J., Caffier, V., Morel, P., Belin, É., & Chapeau-Blondeau, F. (2012). On the use of depth camera for 3d phenotyping of entire plants. *Computers and Electronics in Agriculture*, 82, 122–127.
- Cheng, Y.-Y., & Wyant, J. C. (1985). Multiple-wavelength phase-shifting interferometry. *Applied optics*, 24(6), 804–807.
- Connell, J., Ratha, N., Gentile, J., & Bolle, R. (2013). Fake iris detection using structured light. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 8692–8696.

- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1), 21–27.
- Cui, D., Li, M., & Zhang, Q. (2009). Development of an optical sensor for crop leaf chlorophyll content detection. *Computers and Electronics in Agriculture*, 69(2), 171–176.
- Cui, Y., Schuon, S., Chan, D., Thrun, S., & Theobalt, C. (2010). 3d shape scanning with a time-of-flight camera. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1173–1180.
- Dale, L. M., Thewis, A., Boudry, C., Rotar, I., Dardenne, P., Baeten, V., & Pierna, J. A. F. (2013). Hyperspectral imaging applications in agriculture and agro-food product quality and safety control: A review. *Applied Spectroscopy Reviews*, 48(2), 142–159.
- Daugman, J. (2009). How iris recognition works. In *The essential guide to image processing* (pp. 715–739). Elsevier.
- De Marsico, M., Galdi, C., Nappi, M., & Riccio, D. (2014). Firme: Face and iris recognition for mobile engagement. *Image and Vision Computing*, 32(12), 1161–1172.
- de Groot, P. (2011). Coherence scanning interferometry. In *Optical measurement of surface topography* (pp. 187–208). Springer.
- Di, D., Yang, J., Luo, C., Xue, Z., Chen, W., Yang, X., & Gao, Y. (2024). Hyper-3dg: Text-to-3d gaussian generation via hypergraph. *arXiv preprint arXiv:2403.09236*.
- Dorken Gallastegi, U., Rueda-Chacón, H., Stevens, M. J., & Goyal, V. K. (2023). Absorption-based hyperspectral thermal ranging: Performance analyses, optimization, and simulations. *Optics Express*, 32(1), 151–166.
- Dorsch, R. G., Häusler, G., & Herrmann, J. M. (1994). Laser triangulation: Fundamental uncertainty in distance measurement. *Applied optics*, 33(7), 1306–1314.
- Draréni, J., Roy, S., & Sturm, P. (2011). Plane-based calibration for linear cameras. *International Journal of Computer Vision*, 91, 146–156.
- Drouin, M.-A., & Beraldin, J.-A. (2020). Active triangulation 3d imaging systems for industrial inspection. *3D Imaging, Analysis and Applications*, 109–165.
- Elharrouss, O., Almaadeed, N., Al-Maadeed, S., & Akbari, Y. (2020). Image inpainting: A review. *Neural Processing Letters*, 51(2), 2007–2028.

- ElMasry, G., & Sun, D.-W. (2010). Principles of hyperspectral imaging technology. In *Hyperspectral imaging for food quality analysis and control* (pp. 3–43). Elsevier.
- Fang, L., Shi, Z., Li, C., Liu, Y., & Zhao, E. (2022). Geometric transformation modeling for line-scan images under different camera poses. *Optical Engineering*, *61*(10), 103103–103103.
- Fei, B. (2019). Hyperspectral imaging in medical applications. In *Data handling in science and technology* (pp. 523–565, Vol. 32). Elsevier.
- Feng, S., Zuo, C., Zhang, L., Tao, T., Hu, Y., Yin, W., Qian, J., & Chen, Q. (2021). Calibration of fringe projection profilometry: A comparative review. *Optics and lasers in engineering*, *143*, 106622.
- Feng, Y.-Z., & Sun, D.-W. (2012). Application of hyperspectral imaging in food safety inspection and control: A review. *Critical reviews in food science and nutrition*, *52*(11), 1039–1058.
- Foix, S., Alenya, G., & Torras, C. (2011). Lock-in time-of-flight (tof) cameras: A survey. *IEEE Sensors Journal*, *11*(9), 1917–1926.
- França, J. G. D., Gazziro, M. A., Ide, A. N., & Saito, J. H. (2005). A 3d scanning system based on laser triangulation and variable field of view. *IEEE International Conference on Image Processing 2005*, *1*, I–425.
- Free 3d models [Accessed: May, 2022]. (n.d.).
- Gao, L., Kester, R. T., & Tkaczyk, T. S. (2009). Compact image slicing spectrometer (iss) for hyperspectral fluorescence microscopy. *Optics express*, *17*(15), 12293–12308.
- Gao, Y., Li, Q., Rao, X., & Ying, Y. (2018). Precautionary analysis of sprouting potato eyes using hyperspectral imaging technology. *International Journal of Agricultural and Biological Engineering*, *11*(2), 153–157.
- Gårding, J. (1992). Shape from texture for smooth curved surfaces in perspective projection. *Journal of Mathematical Imaging and Vision*, *2*, 327–350.
- Ghosh, A., Fassnacht, F. E., Joshi, P. K., & Koch, B. (2014). A framework for mapping tree species combining hyperspectral and lidar data: Role of selected classifiers and sensor across three spatial scales. *International Journal of Applied Earth Observation and Geoinformation*, *26*, 49–63.

- Ham, J., Chen, Y., Crawford, M. M., & Ghosh, J. (2005). Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3), 492–501.
- Hansard, M., Lee, S., Choi, O., & Horaud, R. P. (2012). *Time-of-flight cameras: Principles, methods and applications*. Springer Science & Business Media.
- Hartley, R., & Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge university press.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Heist, S., Zhang, C., Reichwald, K., Kühmstedt, P., Notni, G., & Tünnermann, A. (2018a). 5d hyperspectral imaging: Fast and accurate measurement of surface shape and spectral characteristics using structured light. *Optics express*, 26(18), 23366–23379.
- Heist, S., Zhang, C., Reichwald, K., Kühmstedt, P., Notni, G., & Tünnermann, A. (2018b). 5d hyperspectral imaging: Fast and accurate measurement of surface shape and spectral characteristics using structured light. *Optics express*, 26(18), 23366–23379.
- Horaud, R., Hansard, M., Evangelidis, G., & Ménier, C. (2016). An overview of depth cameras and range scanners based on time-of-flight technologies. *Machine vision and applications*, 27(7), 1005–1020.
- Hu, M., Wang, S., Li, B., Ning, S., Fan, L., & Gong, X. (2021). Penet: Towards precise and efficient image guided depth completion. *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 13656–13662.
- Huang, L., & Asundi, A. K. (2011). Phase invalidity identification framework with the temporal phase unwrapping method. *Measurement Science and Technology*, 22(3), 035304.
- Huang, L., Chua, P. S., & Asundi, A. (2010). Least-squares calibration method for fringe projection profilometry considering camera lens distortion. *Applied optics*, 49(9), 1539–1548.
- Huang, P., Luo, X., Jin, J., Wang, L., Zhang, L., Liu, J., & Zhang, Z. (2018a). Improving high-throughput phenotyping using fusion of close-range hyperspectral camera and low-cost depth sensor. *Sensors*, 18(8), 2711.

- Huang, P., Luo, X., Jin, J., Wang, L., Zhang, L., Liu, J., & Zhang, Z. (2018b). Improving high-throughput phenotyping using fusion of close-range hyperspectral camera and low-cost depth sensor. *Sensors*, *18*(8), 2711.
- Huang, Z., Xi, J., Yu, Y., & Guo, Q. (2015). Accurate projector calibration based on a new point-to-point mapping relationship between the camera and projector images. *Applied Optics*, *54*(3), 347–356.
- Hutter, T., Elliott, S. R., & Ruschin, S. (2013). Dynamic range enhancement and phase-ambiguity elimination in wavelength-interrogated interferometric sensor. *Sensors and Actuators B: Chemical*, *178*, 593–597.
- Jing, H., Su, X., & You, Z. (2017). Uniaxial three-dimensional shape measurement with multioperation modes for different modulation algorithms. *Optical Engineering*, *56*(3), 034115.
- Kääriäinen, T., & Seppä, J. (2024). 3d camera based on laser light absorption by atmospheric oxygen at 761 nm. *Optics Express*, *32*(4), 6342–6349.
- Kahraman, S., & Bacher, R. (2021). A comprehensive review of hyperspectral data fusion with lidar and sar data. *Annual Reviews in Control*, *51*, 236–253.
- Kim, M. S., Chen, Y., & Mehl, P. (2001). Hyperspectral reflectance and fluorescence imaging system for food quality and safety. *Transactions of the ASAE*, *44*(3), 721.
- Lazaros, N., Sirakoulis, G. C., & Gasteratos, A. (2008a). Review of stereo vision algorithms: From software to hardware. *International Journal of Optomechatronics*, *2*(4), 435–462.
- Lazaros, N., Sirakoulis, G. C., & Gasteratos, A. (2008b). Review of stereo vision algorithms: From software to hardware. *International Journal of Optomechatronics*, *2*(4), 435–462.
- Lee, W.-S., Alchanatis, V., Yang, C., Hirafuji, M., Moshou, D., & Li, C. (2010). Sensing technologies for precision specialty crop production. *Computers and electronics in agriculture*, *74*(1), 2–33.
- Leucker, M., Mahlein, A.-K., Steiner, U., & Oerke, E.-C. (2016). Improvement of lesion phenotyping in cercospora beticola–sugar beet interaction by hyperspectral imaging. *Phytopathology*, *106*(2), 177–184.
- Li, B., Karpinsky, N., & Zhang, S. (2014). Novel calibration method for structured-light system with an out-of-focus projector. *Applied optics*, *53*(16), 3415–3426.

- Li, B., & Zhang, S. (2014). Structured light system calibration method with optimal fringe angle. *Applied optics*, 53(33), 7942–7950.
- Li, B., & Zhang, S. (2015). Flexible calibration method for microscopic structured light system using telecentric lens. *Optics express*, 23(20), 25795–25803.
- Li, J., & Li, B. (2023). Tpdnet: Texture-guided phase-to-depth networks to repair shadow-induced errors for fringe projection profilometry. *Photonics*, 10(3). <https://doi.org/10.3390/photonics10030246>
- Li, J., & Li, B. (2024). Comparative analysis of circular and linear fringe projection profilometry: From calibration to 3d reconstruction. *Opt. Continuum*, 3(3), 468–483. <https://doi.org/10.1364/OPTCON.520433>
- Li, J., Zheng, Y., Liu, L., & Li, B. (2021). 4d line-scan hyperspectral imaging. *Optics Express*, 29(21), 34835–34849.
- Li, Q., He, X., Wang, Y., Liu, H., Xu, D., & Guo, F. (2013). Review of spectral imaging technology in biomedical engineering: Achievements and challenges. *Journal of biomedical optics*, 18(10), 100901–100901.
- Li, W., Yu, J., Gai, S., & Da, F. (2021). Absolute phase retrieval for a single-shot fringe projection profilometry based on deep learning. *Optical Engineering*, 60(6), 064104–064104.
- Li, Y., Qian, J., Feng, S., Chen, Q., & Zuo, C. (2020). Single-shot spatial frequency multiplex fringe pattern for phase unwrapping using deep learning. *Optics Frontier Online 2020: Optics Imaging and Display*, 11571, 314–319.
- Li, Z.-w., Shi, Y.-s., Wang, C.-j., Qin, D.-h., & Huang, K. (2009). Complex object 3d measurement based on phase-shifting and a neural network. *Optics Communications*, 282(14), 2699–2706.
- Li, Z., Shi, Y., Wang, C., & Wang, Y. (2008). Accurate calibration method for a structured light system. *Optical Engineering*, 47(5), 053604–053604.
- Liang, J., Zia, A., Zhou, J., & Sirault, X. (2013). 3d plant modelling via hyperspectral imaging. *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 172–177.
- Liu, H., Bruning, B., Garnett, T., & Berger, B. (2020). Hyperspectral imaging and 3d technologies for plant phenotyping: From satellite to close-range sensing. *Computers and Electronics in Agriculture*, 175, 105621.

- Liu, X., Jiang, Z., Wang, T., Cai, F., & Wang, D. (2020). Fast hyperspectral imager driven by a low-cost and compact galvo-mirror. *Optik*, *224*, 165716.
- Lloyd, S. (1982). Least squares quantization in pcm. *IEEE transactions on information theory*, *28*(2), 129–137.
- Lu, B., Dao, P. D., Liu, J., He, Y., & Shang, J. (2020). Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sensing*, *12*(16), 2659.
- Lu, G., & Fei, B. (2014). Medical hyperspectral imaging: A review. *Journal of biomedical optics*, *19*(1), 010901–010901.
- Lu, L., Xi, J., Yu, Y., Guo, Q., Yin, Y., & Song, L. (2015). Shadow removal method for phase-shifting profilometry. *Applied optics*, *54*(19), 6059–6064.
- Lu, Y., Saeys, W., Kim, M., Peng, Y., & Lu, R. (2020). Hyperspectral imaging technology for quality and safety evaluation of horticultural products: A review and celebration of the past 20-year progress. *Postharvest Biology and Technology*, *170*, 111318.
- Lucasbosch. (2021, February). File:multispectral imaging approaches.svg. https://commons.wikimedia.org/wiki/File:Multispectral_imaging_approaches.svg
- Luo, J., Forsberg, E., Fu, S., & He, S. (2022). High-precision four-dimensional hyperspectral imager integrating fluorescence spectral detection and 3d surface shape measurement. *Applied Optics*, *61*(10), 2542–2551.
- Luo, J., Forsberg, E., Fu, S., Xing, Y., Liao, J., Jiang, J., Zheng, Y., & He, S. (2022). 4d dual-mode staring hyperspectral-depth imager for simultaneous spectral sensing and surface shape measurement. *Optics Express*, *30*(14), 24804–24821.
- Luo, J., Forsberg, E., & He, S. (2022). 5d-fusion imaging for surface shape, polarization, and hyperspectral measurement. *Applied Optics*, *61*(26), 7776–7785.
- Luo, J., Li, S., Forsberg, E., & He, S. (2021). 4d surface shape measurement system with high spectral resolution and great depth accuracy. *Optics Express*, *29*(9), 13048–13070.
- Luo, J., Lin, Z., Xing, Y., Forsberg, E., Wu, C., Zhu, X., Guo, T., Wang, G., Bian, B., Wu, D., et al. (2022). Portable 4d snapshot hyperspectral imager for fast spectral and surface morphology measurements. *Progress In Electromagnetics Research*, *173*, 25–36.

- Mada, S. K., Smith, M. L., Smith, L. N., & Midha, P. S. (2003). Overview of passive and active vision techniques for hand-held 3d data acquisition. *Opto-Ireland 2002: Optical Metrology, Imaging, and Machine Vision*, 4877, 16–27.
- Malacara, D. (2007a). *Optical shop testing* (Vol. 59). John Wiley & Sons.
- Malacara, D. (2007b). *Optical shop testing* (Vol. 59). John Wiley & Sons.
- Mandapalli, J. K., Ravi, V., Gorthi, S. S., Gorthi, S., & Gorthi, R. K. (2021). Single-shot circular fringe projection for the profiling of objects having surface discontinuities. *JOSA A*, 38(10), 1471–1482.
- Marrugo, A. G., Gao, F., & Zhang, S. (2020a). State-of-the-art active optical techniques for three-dimensional surface metrology: A review. *JOSA A*, 37(9), B60–B77.
- Marrugo, A. G., Gao, F., & Zhang, S. (2020b). State-of-the-art active optical techniques for three-dimensional surface metrology: A review. *JOSA A*, 37(9), B60–B77.
- Mattoccia, S., Tombari, F., & Di Stefano, L. (2007). Stereo vision enabling precise border localization within a scanline optimization framework. *Computer Vision–ACCV 2007: 8th Asian Conference on Computer Vision, Tokyo, Japan, November 18-22, 2007, Proceedings, Part II* 8, 517–527.
- Meddecki, H., Tejnil, E., Goldberg, K., & Bokor, J. (1996). Phase-shifting point diffraction interferometer. *Optics letters*, 21(19), 1526–1528.
- Meng, W., Quanyao, H., Yongkai, Y., Yang, Y., Qijian, T., Xiang, P., & Xiaoli, L. (2022). Large dof microscopic fringe projection profilometry with a coaxial light-field structure. *Optics Express*, 30(5), 8015–8026.
- Merzlyak, M. N., Solovchenko, A. E., & Gitelson, A. A. (2003). Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit. *Postharvest biology and technology*, 27(2), 197–211.
- Ming, Y., Meng, X., Fan, C., & Yu, H. (2021). Deep learning for monocular depth estimation: A review. *Neurocomputing*, 438, 14–33.
- Monteiro, S. T., Nieto, J., Murphy, R., Ramakrishnan, R., & Taylor, Z. (2013a). Combining strong features for registration of hyperspectral and lidar data from field-based platforms. *2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS*, 1210–1213.

- Monteiro, S. T., Nieto, J., Murphy, R., Ramakrishnan, R., & Taylor, Z. (2013b). Combining strong features for registration of hyperspectral and lidar data from field-based platforms. *2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS*, 1210–1213.
- Montresor, S., Tahon, M., Laurent, A., & Picart, P. (2020). Computational de-noising based on deep learning for phase data in digital holographic interferometry. *APL Photonics*, 5(3), 030802.
- Moré, J. J. (2006). The levenberg-marquardt algorithm: Implementation and theory. *Numerical analysis: proceedings of the biennial Conference held at Dundee, June 28–July 1, 1977*, 105–116.
- Mühlmann, K., Maier, D., Hesser, J., & Männer, R. (2002). Calculating dense disparity maps from color stereo images, an efficient implementation. *International Journal of Computer Vision*, 47, 79–88.
- Nagase, Y., Kushida, T., Tanaka, K., Funatomi, T., & Mukaigawa, Y. (2022). Shape from thermal radiation: Passive ranging using multi-spectral lwir measurements. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12661–12671.
- Narvaez, F. Y., Reina, G., Torres-Torriti, M., Kantor, G., & Cheein, F. A. (2017). A survey of ranging and imaging techniques for precision agriculture phenotyping. *IEEE/ASME Transactions on Mechatronics*, 22(6), 2428–2439.
- Nayar, S. K., & Nakagawa, Y. (1994). Shape from focus. *IEEE Transactions on Pattern analysis and machine intelligence*, 16(8), 824–831.
- Nguyen, H., Wang, Y., & Wang, Z. (2020). Single-shot 3d shape reconstruction using structured light and deep convolutional neural networks. *Sensors*, 20(13), 3718.
- Nguyen, T. T., Slaughter, D. C., Max, N., Maloof, J. N., & Sinha, N. (2015). Structured light-based 3d reconstruction system for plants. *Sensors*, 15(8), 18587–18612.
- Nieto, J. I., Monteiro, S. T., & Viejo, D. (2010). 3d geological modelling using laser and hyperspectral data. *2010 IEEE International Geoscience and Remote Sensing Symposium*, 4568–4571.
- Niskanen, I., Immonen, M., Hallman, L., Yamamuchi, G., Mikkonen, M., Hashimoto, T., Nitta, Y., Keränen, P., Kostamovaara, J., & Heikkilä, R. (2021). Time-of-flight sensor for getting

- shape model of automobiles toward digital 3d imaging approach of autonomous driving. *Automation in Construction*, 121, 103429.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.
- Pearlman, J. S., Barry, P. S., Segal, C. C., Shepanski, J., Beiso, D., & Carman, S. L. (2003). Hyperion, a space-based imaging spectrometer. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6), 1160–1173.
- Pei, Z., Huang, Y. M., & Zhou, T. (2023). Review on analysis methods enabled by hyperspectral imaging for cultural relic conservation. *Photonics*, 10(10), 1104.
- Peraza-Alemán, C. M., López-Maestresalas, A., Jarén, C., Rubio-Padilla, N., & Arazuri, S. (2024). A systematized review on the applications of hyperspectral imaging for quality control of potatoes. *Potato Research*, 1–23.
- Qi, H., Lyu, N., Yu, H., Zheng, D., & Han, J. (2023). 4-d multiframe co-encoded spectral imaging system. *Optics and Lasers in Engineering*, 169, 107697.
- Robbins, H., & Monro, S. (1951). A stochastic approximation method. *The annals of mathematical statistics*, 400–407.
- Rogelj, L., Pavlovčič, U., Stergar, J., Jezeršek, M., Simončič, U., & Milanič, M. (2019a). Curvature and height corrections of hyperspectral images using built-in 3d laser profilometry. *Applied optics*, 58(32), 9002–9012.
- Rogelj, L., Pavlovčič, U., Stergar, J., Jezeršek, M., Simončič, U., & Milanič, M. (2019b). Curvature and height corrections of hyperspectral images using built-in 3d laser profilometry. *Applied Optics*, 58(32), 9002–9012.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, 234–241.
- Rosell-Polo, J. R., Cheein, F. A., Gregorio, E., Andújar, D., Puigdomènech, L., Masip, J., & Escolà, A. (2015). Advances in structured light sensors applications in precision agriculture and livestock farming. *Advances in agronomy*, 133, 71–112.

- Rueda, H., Fu, C., Lau, D. L., & Arce, G. R. (2017). Single aperture spectral+ tof compressive camera: Toward hyperspectral+ depth imagery. *IEEE Journal of Selected Topics in Signal Processing*, *11*(7), 992–1003.
- Sansoni, G., Carocci, M., & Rodella, R. (1999). 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.
- Schmit, J., Creath, K., & Wyant, J. (2007). Surface profilers, multiple wavelength, and white light interferometry. *Optical shop testing*, 667755.
- Schreiber, H., & Bruning, J. H. (2007). Phase shifting interferometry. *Optical shop testing*, 547–666.
- Se, S., & Pears, N. (2020). Passive 3d imaging. *3D Imaging, Analysis and Applications*, 39–107.
- Sethy, P. K., Pandey, C., Sahu, Y. K., & Behera, S. K. (2022). Hyperspectral imagery applications for precision agriculture-a systemic survey. *Multimedia Tools and Applications*, *81*(2), 3005–3038.
- Shi, B., Ma, Z., Liu, J., Ni, X., Xiao, W., & Liu, H. (2022). Shadow extraction method based on multi-information fusion and discrete wavelet transform. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–15.
- Sima, A. A., & Buckley, S. J. (2013). Optimizing sift for matching of short wave infrared and visible wavelength images. *Remote Sensing*, *5*(5), 2037–2056.
- Sinha, S. N., Graf, H., & Weinmann, M. (2024). Spectralgaussians: Semantic, spectral 3d gaussian splatting for multi-spectral scene representation, visualization and analysis. *arXiv preprint arXiv:2408.06975*.
- Siripatrawan, U., Makino, Y., Kawagoe, Y., & Oshita, S. (2011). Rapid detection of escherichia coli contamination in packaged fresh spinach using hyperspectral imaging. *Talanta*, *85*(1), 276–281.
- Skydan, O. A., Lalor, M. J., & Burton, D. R. (2005). Using coloured structured light in 3-d surface measurement. *Optics and Lasers in Engineering*, *43*(7), 801–814.
- Spoorthi, G., Gorthi, S., & Gorthi, R. K. S. S. (2018). Phasenet: A deep convolutional neural network for two-dimensional phase unwrapping. *IEEE Signal Processing Letters*, *26*(1), 54–58.

- Stuart, M. B., McGonigle, A. J., & Willmott, J. R. (2019). Hyperspectral imaging in environmental monitoring: A review of recent developments and technological advances in compact field deployable systems. *Sensors*, *19*(14), 3071.
- Su, Q., Kondo, N., Al Riza, D. F., & Habaragamuwa, H. (2020). Potato quality grading based on depth imaging and convolutional neural network. *Journal of Food Quality*, *2020*(1), 8815896.
- Su, Q., Kondo, N., Li, M., Sun, H., & Al Riza, D. F. (2017). Potato feature prediction based on machine vision and 3d model rebuilding. *Computers and electronics in agriculture*, *137*, 41–51.
- Su, X., & Zhang, Q. (2010). Dynamic 3-d shape measurement method: A review. *Optics and Lasers in Engineering*, *48*(2), 191–204.
- Sun, Z., Jin, Y., Duan, M., Kan, Y., Zhu, C., & Chen, E. (2020). Discriminative repair approach to remove shadow-induced error for typical digital fringe projection. *Optics Express*, *28*(18), 26076–26090.
- Takeda, M., & Mutoh, K. (1983). Fourier transform profilometry for the automatic measurement of 3-d object shapes. *Applied optics*, *22*(24), 3977–3982.
- Team, B. D. (2018). *Blender - a 3d modelling and rendering package*. Blender Foundation. Stichting Blender Foundation, Amsterdam. <http://www.blender.org>
- The MathWorks, I. (2019). *Symbolic math toolbox*. Natick, Massachusetts, United State. <https://www.mathworks.com/help/symbolic/>
- Tolt, G., Shimoni, M., & Ahlberg, J. (2011a). A shadow detection method for remote sensing images using vhr hyperspectral and lidar data. *2011 IEEE international geoscience and remote sensing symposium*, 4423–4426.
- Tolt, G., Shimoni, M., & Ahlberg, J. (2011b). A shadow detection method for remote sensing images using vhr hyperspectral and lidar data. *2011 IEEE international geoscience and remote sensing symposium*, 4423–4426.
- Towers, C. E., Towers, D. P., & Jones, J. D. (2003). Optimum frequency selection in multifrequency interferometry. *Optics letters*, *28*(11), 887–889.
- Van der Jeught, S., & Dirckx, J. J. (2019). Deep neural networks for single shot structured light profilometry. *Optics express*, *27*(12), 17091–17101.

- Van Gansbeke, W., Neven, D., De Brabandere, B., & Van Gool, L. (2019). Sparse and noisy lidar completion with rgb guidance and uncertainty. *2019 16th international conference on machine vision applications (MVA)*, 1–6.
- Vázquez-Arellano, M., Griepentrog, H. W., Reiser, D., & Paraforos, D. S. (2016). 3-d imaging systems for agricultural applications—a review. *Sensors*, *16*(5), 618.
- Vo, M., Wang, Z., Pan, B., & Pan, T. (2012). Hyper-accurate flexible calibration technique for fringe-projection-based three-dimensional imaging. *Optics Express*, *20*(15), 16926–16941.
- Wahabzada, M., Mahlein, A.-K., Bauckhage, C., Steiner, U., Oerke, E.-C., & Kersting, K. (2016). Plant phenotyping using probabilistic topic models: Uncovering the hyperspectral language of plants. *Scientific reports*, *6*(1), 1–11.
- Wang, C., & Pang, Q. (2022). The elimination of errors caused by shadow in fringe projection profilometry based on deep learning. *Optics and Lasers in Engineering*, *159*, 107203.
- Wang, F., Wang, C., & Guan, Q. (2021). Single-shot fringe projection profilometry based on deep learning and computer graphics. *Optics Express*, *29*(6), 8024–8040.
- Wang, K., Li, Y., Kemao, Q., Di, J., & Zhao, J. (2019). One-step robust deep learning phase unwrapping. *Optics express*, *27*(10), 15100–15115.
- Wang, L., Xiong, Z., Shi, G., Zeng, W., & Wu, F. (2016a). Simultaneous depth and spectral imaging with a cross-modal stereo system. *IEEE Transactions on Circuits and Systems for Video Technology*, *28*(3), 812–817.
- Wang, L., Xiong, Z., Shi, G., Zeng, W., & Wu, F. (2016b). Simultaneous depth and spectral imaging with a cross-modal stereo system. *IEEE Transactions on Circuits and Systems for Video Technology*, *28*(3), 812–817.
- Wang, Y., Zhang, S., & Oliver, J. H. (2011a). 3d shape measurement technique for multiple rapidly moving objects. *Optics express*, *19*(9), 8539–8545.
- Wang, Y., Zhang, S., & Oliver, J. H. (2011b). 3d shape measurement technique for multiple rapidly moving objects. *Optics express*, *19*(9), 8539–8545.
- Wang, Y., Tan, T., & Jain, A. K. (2003). Combining face and iris biometrics for identity verification. *Audio-and Video-Based Biometric Person Authentication: 4th International Conference, AVBPA 2003 Guildford, UK, June 9–11, 2003 Proceedings 4*, 805–813.

- Wang, Z., Wang, E., Wang, S., & Ding, Q. (2011). Multimodal biometric system using face-iris fusion feature. *J. Comput.*, 6(5), 931–938.
- Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE transactions on image processing*, 13(4), 600–612.
- Weinmann, M., Schwartz, C., Ruiters, R., & Klein, R. (2011). A multi-camera, multi-projector super-resolution framework for structured light. *2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, 397–404.
- Wyant, J. C. (2013). Computerized interferometric surface measurements. *Applied optics*, 52(1), 1–8.
- Wyant, J. (1973). Double frequency grating lateral shear interferometer. *Applied optics*, 12(9), 2057–2060.
- Xu, J., & Zhang, S. (2020a). Status, challenges, and future perspectives of fringe projection profilometry. *Optics and Lasers in Engineering*, 106193.
- Xu, J., & Zhang, S. (2020b). Status, challenges, and future perspectives of fringe projection profilometry. *Optics and Lasers in Engineering*, 135, 106193.
- Xu, Y., Giljum, A., & Kelly, K. F. (2020a). A hyperspectral projector for simultaneous 3d spatial and hyperspectral imaging via structured illumination. *Optics Express*, 28(20), 29740–29755.
- Xu, Y., Giljum, A., & Kelly, K. F. (2020b). A hyperspectral projector for simultaneous 3d spatial and hyperspectral imaging via structured illumination. *Optics Express*, 28(20), 29740–29755.
- Yan, K., Yu, Y., Huang, C., Sui, L., Qian, K., & Asundi, A. (2019). Fringe pattern denoising based on deep learning. *Optics Communications*, 437, 148–152.
- Yan, K., Yu, Y., Sun, T., Asundi, A., & Kemaq, Q. (2020). Wrapped phase denoising using convolutional neural networks. *Optics and Lasers in Engineering*, 128, 105999.
- Yang, S., & Zhang, G. (2018). A review of interferometry for geometric measurement. *Measurement Science and Technology*, 29(10), 102001.
- Yang, T., Zhang, Z., Li, H., Li, X., & Zhou, X. (2020). Single-shot phase extraction for fringe projection profilometry using deep convolutional generative adversarial network. *Measurement Science and Technology*, 32(1), 015007.

- Yao, M., Xiong, Z., Wang, L., Liu, D., & Chen, X. (2019). Computational spectral-depth imaging with a compact system. *SIGGRAPH Asia 2019 Posters*. <https://doi.org/10.1145/3355056.3364570>
- Yu, J., Gao, N., Zhang, Z., & Meng, Z. (2020). High sensitivity fringe projection profilometry combining optimal fringe frequency and optimal fringe direction. *Optics and Lasers in Engineering*, *129*, 106068.
- Zhang, C., Zhao, H., Qiao, J., Zhou, C., Zhang, L., Hu, G., & Geng, H. (2019). Three-dimensional measurement based on optimized circular fringe projection technique. *Optics express*, *27*(3), 2465–2477.
- Zhang, G., Lau, D. L., Xu, B., Zhu, C., & Liu, K. (2023). Circular fringe projection profilometry and 3d sensitivity analysis based on extended epipolar geometry. *Optics and Lasers in Engineering*, *162*, 107403.
- Zhang, J., Luo, B., Su, X., Li, L., Li, B., Zhang, S., & Wang, Y. (2021). A convenient 3d reconstruction model based on parallel-axis structured light system. *Optics and Lasers in Engineering*, *138*, 106366.
- Zhang, L., Jin, J., Wang, L., Huang, P., & Ma, D. (2020). A 3d white referencing method for soybean leaves based on fusion of hyperspectral images and 3d point clouds. *Precision Agriculture*, *21*, 1173–1186.
- Zhang, P., Arre, T. J., & Ide-Ektessabi, A. (2015). A line scan camera-based structure from motion for high-resolution 3d reconstruction. *Journal of Cultural Heritage*, *16*(5), 656–663.
- Zhang, Q., Lu, S., Li, J., Li, W., Li, D., Lu, X., Zhong, L., & Tian, J. (2020). Deep phase shifter for quantitative phase imaging. *arXiv preprint arXiv:2003.03027*.
- Zhang, R., Guo, H., & Asundi, A. K. (2016). Geometric analysis of influence of fringe directions on phase sensitivities in fringe projection profilometry. *Applied optics*, *55*(27), 7675–7687.
- Zhang, R., Tsai, P.-S., Cryer, J. E., & Shah, M. (1999). Shape-from-shading: A survey. *IEEE transactions on pattern analysis and machine intelligence*, *21*(8), 690–706.
- Zhang, S. (2009). Phase unwrapping error reduction framework for a multiple-wavelength phase-shifting algorithm. *Optical Engineering*, *48*(10), 105601.
- Zhang, S. (2013). *Handbook of 3d machine vision: Optical metrology and imaging*. CRC press.

- Zhang, S. (2018). High-speed 3d shape measurement with structured light methods: A review. *Optics and lasers in engineering*, 106, 119–131.
- Zhang, S., & Huang, P. S. (2006). Novel method for structured light system calibration. *Optical Engineering*, 45(8), 083601–083601.
- Zhang, W., Li, W., Yu, L., Luo, H., Zhao, H., & Xia, H. (2017). Sub-pixel projector calibration method for fringe projection profilometry. *Optics express*, 25(16), 19158–19169.
- Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11), 1330–1334.
- Zhang, Z., Huang, S., Meng, S., Gao, F., & Jiang, X. (2013). A simple, flexible and automatic 3d calibration method for a phase calculation-based fringe projection imaging system. *Optics express*, 21(10), 12218–12227.
- Zhang, Z., Ma, H., Guo, T., Zhang, S., & Chen, J. (2011). Simple, flexible calibration of phase calculation-based three-dimensional imaging system. *Optics letters*, 36(7), 1257–1259.
- Zhao, H., Gallo, O., Frosio, I., & Kautz, J. (2016). Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging*, 3(1), 47–57.
- Zhao, H., Zhang, C., Zhou, C., Jiang, K., & Fang, M. (2016). Circular fringe projection profilometry. *Optics letters*, 41(21), 4951–4954.
- Zhao, H., Wang, Z., Jia, G., Li, X., & Zhang, Y. (2018). Field imaging system for hyperspectral data, 3d structural data and panchromatic image data measurement based on acousto-optic tunable filter. *Optics express*, 26(13), 17717–17730.
- Zhao, M., Huang, L., Zhang, Q., Su, X., Asundi, A., & Kemao, Q. (2011). Quality-guided phase unwrapping technique: Comparison of quality maps and guiding strategies. *Applied optics*, 50(33), 6214–6224.
- Zheng, D., Da, F., Kemao, Q., & Seah, H. S. (2017). Phase-shifting profilometry combined with gray-code patterns projection: Unwrapping error removal by an adaptive median filter. *Optics express*, 25(5), 4700–4713.
- Zheng, Y., & Li, B. (2020). Uniaxial high-speed microscale three-dimensional surface topographical measurements using fringe projection. *Journal of Micro and Nano-Manufacturing*, 8(4).
- Zheng, Y., Wang, S., Li, Q., & Li, B. (2020). Fringe projection profilometry by conducting deep learning from its digital twin. *Optics Express*, 28(24), 36568–36583.

- Zheng, Y., Zhang, X., Wang, S., Li, Q., Qin, H., & Li, B. (2020). Similarity evaluation of topography measurement results by different optical metrology technologies for additive manufactured parts. *Optics and Lasers in Engineering*, 126, 105920.
- Zhong, Y., Wang, X., Xu, Y., Wang, S., Jia, T., Hu, X., Zhao, J., Wei, L., & Zhang, L. (2018). Mini-uav-borne hyperspectral remote sensing: From observation and processing to applications. *IEEE Geoscience and Remote Sensing Magazine*, 6(4), 46–62.
- Zhou, Q., & Jacobson, A. (2016). Thingi10k: A dataset of 10,000 3d-printing models. *arXiv preprint arXiv:1605.04797*.
- Zhou, W.-S., & Su, X.-Y. (1994). A direct mapping algorithm for phase-measuring profilometry. *Journal of modern optics*, 41(1), 89–94.
- Zhu, M., Huang, D., Hu, X.-J., Tong, W.-H., Han, B.-L., Tian, J.-P., & Luo, H.-B. (2020). Application of hyperspectral technology in detection of agricultural products and food: A review. *Food science & nutrition*, 8(10), 5206–5214.
- Zuo, C., Feng, S., Huang, L., Tao, T., Yin, W., & Chen, Q. (2018). Phase shifting algorithms for fringe projection profilometry: A review. *Optics and lasers in engineering*, 109, 23–59.
- Zuo, C., Qian, J., Feng, S., Yin, W., Li, Y., Fan, P., Han, J., Qian, K., & Chen, Q. (2022). Deep learning in optical metrology: A review. *Light: Science & Applications*, 11(1), 1–54.