Smart three-dimensional machine vision

Huang, Chongtian

2019

Huang, C. (2019). Smart three-dimensional machine vision. Master's thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/102839>

<https://doi.org/10.32657/10220/48588>

Smart Three-Dimensional Machine

Vision

HUANG CHONGTIAN

School of Mechanical & Aerospace Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirements for the degree of
Master of Engineering

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

5 May 2019

.....

Date

Huang Chongtian

.....

HUANG CHONGTIAN

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

5 May 2019

.....

Date

.....

Prof. Anand ASUNDI

Authorship Attribution Statement

This thesis **does not** contain any materials from papers published in peer-reviewed journals or from papers accepted at conferences in which I am listed as an author.

5 May 2019

.....

Date

Huang Chongtian

.....

HUANG CHONGTIAN

Acknowledgements

First and foremost, I wish to express my greatest gratitude to my advisor, *Professor Anand Asundi*, for his continuous guidance, support, assistance and immense knowledge throughout the program.

Secondly, I would like to thanks to *Professor Sui* Lianshenga from the Xian University of Technology and *Professor Qian Kemao* from the school of Computer Engineering for their suggestions and assistance. I would like to thank *Professor Zuo chao* and *Dr. Hu yan* from Nanjing Science and Technology University for their initial support.

Thirdly, Thanks to my fellow lab mates and colleagues in Center for Optics and Laser Engineering and Shanghai University: *Dr.SUN Jianfei*, *Mr.ZHANG Wei*, and *Mr. YAN ketao*, *Mr. CHENG Dingfu* for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last year. Also, I thank my colleagues *Dr.Achyut ADHIKARI*, *Dr. Rosmin Elsa MOHAN*, *Mr. ZHU kuanhui* and *Dr. Rachel ruiqi WANG* for enlightening me the first glance of research.

My sincere thanks also go to *Ms.YU Xiang* and *Ms. LIU Yixuan* from the School of Art, Design, and Media for her inspiration in my research.

The last but not least, I would like to thank my family and friends: my parents *Mr. HUANG Quanjiang* and *Ms. YANG Migui* for giving birth to me at the first place and supporting me spiritually throughout my life, and special thanks to *Ms. LIU Yixuan* for her support.

“If I had one hour to save the world, I would spend 55 minutes defining the problem and only five minutes finding the solution.”

—Einstein, Albert

To my dear family

Abstract

Three-dimensional inspection is increasingly being demanded by industry in various sectors and none more so than in wafer fabs. While there are many three-dimensional surface measuring instruments, there is still a gap to their adoption in the industry due to lack of methodologies for measuring large objects with micrometer axial resolution and at the same time to zoom into small regions with nanometer resolution. Apart from this, there is also a need to automate the inspection process. In this report, non-interferometric schemes are developed and improved for inspection of specular objects at different scales and a proposal for identifying defects with deep learning over this large field which can then be imaged in three dimensions using other novel systems with higher resolution.

Firstly, for wafer flatness and warpage measurement, a Phase Measuring Reflectometry (PMR) with a large field of view is enhanced to measure wafers as large as 300 mm with no-scanning or moving components. Compared with existing scanning-based systems, this provides full-field, high axial resolution (few micrometers) measurement capabilities of wafers. Wafer warpage during various stages of processing can be measured rapidly thus providing a basis to modify or improve the processing parameters. The system which uses dual camera approach for measurement using structured light also allows for re-positioning and registration of the sample. Furthermore, We can obtain the rectified wafer image and global surface profile by the inspection system. Next, from the processed high-resolution images, defects can be highlighted. Traditionally, this would be done using image processing methods which are generally slow and not precise. To improve the robustness and reliability of defect detection, we introduce deep-learning methods for two-dimensional defects localization and segmentation. This approach requires training the machine to learn different defect shapes and sizes which then can automatically identify defects on test samples. Application of this can be more widespread, and we have demonstrated the scratch and dig measurement of optics as well as for defect detection in the electronics industry.

Usually, the PMR system can resolve defects, but the axial resolution of a few micrometers is not sufficient to get accurate three-dimensional profiles. Furthermore, the wafer has many dies which need to be inspected at a higher resolution. Towards this end, We developed a novel multi-scale system by improving a commercial device built by a local company. This system can measure three-dimensional profiles in two ways: first, using the principle of the Transport of Intensity Equation (TIE), the system can obtain three-dimensional profiles of defects or dies with nanometer axial resolution. Furthermore, by attaching the system to an industrial microscope, high spatial resolution can be obtained as well. Secondly, if the height of objects is much larger compared to the wavelength of visible light, a deconvolution-based approach is used to extract the depth from a stack of images. In the current system, the focus scanning is in high speed with the assistant of a novel deep learning method thus improving the processing times over conventional confocal-type systems. Furthermore, the deconvolution-based depth retrieval method does not require the usage of a pinhole as used in confocal systems. With learning methods, the sample's shape can be reconstructed by computer without intensive iteration.

Finally, an aluminum sample inspection results are shown to prove the performances of the proposed system scheme. Compared with conventional methods such as point or line scanning methods, this scheme can obtain the whole surface profile without scanning or contact, and with the deep learning algorithm and the multi-scale measurement scheme, a single inspection system can inspect not only the global surface profile but also the details of the three-dimensional structure of defects plus devices. This methodology can be adopted in many three-dimensional machine vision applications for inspecting digs and scratches of optical components as well as wavefront analysis, in MEMS and micro-systems arena for wafer warpage and die inspection in three-dimensional, as well as other consumer applications.

Contents

Acknowledgements	ix
Abstract	xiii
List of Figures	xix
Symbols and Acronyms	xxiii
1 Introduction	1
1.1 Introduction	1
1.2 Objectives	3
1.3 Scope	3
1.4 Outline	4
2 Literature review	5
2.1 Background	5
2.2 Grating based methods	7
2.2.1 Fringe projection profilometry	9
2.2.1.1 Principle of fringe projection	10
2.2.1.2 Testing of 3D fringe profilometer from d’Optron	11
2.2.2 Phase Measuring Reflectometry (PMR)	12
2.2.2.1 Principle of PMR	13
2.2.2.2 Testing of monoscopic PMR setup	15
2.3 Depth from focus (DFF) and deconvolution	15
2.3.1 Introduction	15
2.3.2 Mathematical model and current solutions	18
2.4 Transport of Intensity Equation (TIE)	22
2.4.1 Introduction	22
2.4.2 Mathematical model and current solutions	24
2.4.3 Current experimental setup	26
2.5 Deep learning methods	28
2.5.1 Introduction	29
2.5.2 Deep learning in 2D machine vision	31

2.5.3	Deep learning in 3D machine vision and optical problem . . .	33
2.6	Conclusion	35
3	PMR for large objects with specular surface	39
3.1	System Configuration	39
3.2	System calibration	40
3.2.1	Camera calibration	40
3.2.2	Screen calibration	43
3.3	Shape reconstruction	45
3.3.1	Surface optimization	45
3.3.2	Shape fusion	47
3.4	Experiment results	48
3.4.1	The measurement of a plane mirror	49
3.4.2	The measurement of a wafer	51
3.4.3	The measurement of an aluminum foil	52
3.4.4	Conclusion	53
3.5	Summary	53
4	Multi-scale measurement subsystem based on DFF and TIE	55
4.1	System configuration and calibration	55
4.1.1	Telecentric lens subsystem	56
4.1.2	Illumination subsystem	58
4.1.3	Focus calibration	59
4.2	DFF based on Convolution LSTM	60
4.2.1	Concepts	60
4.2.2	Implementation	61
4.2.3	Results	63
4.3	Measurement method based on TIE	66
4.4	Summary	67
5	Smart metrology with the deep learning	69
5.1	Introduction	69
5.2	The neural network configuration	70
5.3	Smart measurement architecture	71
5.3.1	Concepts	71
5.3.2	Implements	72
5.4	Results	72
5.4.1	Training of mask perceptual neural network	72
5.4.2	Segmentation and classification	74
5.5	Summary	75
6	Conclusion	77
6.1	Summary of results	77
6.2	Outlook for future research	78

List of Author's Awards, Patents, and Publications	79
List of Author's Awards, Patents, and Publications	79
Bibliography	81

List of Figures

1.1	Shape and topography measurement applications	2
1.2	Three-dimensional metrology	2
1.3	two-dimensional Inspection System Architecture	3
2.1	3D Optical Metrology, we focused on the methods in the dash box .	6
2.2	Fringe projection profilometry system	9
2.3	Ideal geometry of Fringe Projection System	11
2.4	The hardware and software of the 3D profilometer	12
2.5	Left image is a 2D image of a mechanical component, the left image is the 3D profile of that component	12
2.6	The Relationship between the reflected ray angle change and the sample's slopes.	13
2.7	Incident Ray according to phase	14
2.8	Left image is the monoscopic PMR hardware setup, Right image is a lens measurement results	16
2.9	Formation of focused and defocused point	16
2.10	Schematic representation of the topography and texture retrieval algorithm.	19
2.11	point source image at different location along optical axis	19
2.12	Optical System used for TIE	27
2.13	Transmission TIE system(Courtesy of d'Optron)	28
2.14	Three images dz, 0 and -dz are captured by the TIE system from d'Optron, then 3D profile (Right image) of the biological sample can be reconstructed.	29
2.15	A Venn diagram showing the relationship between deep learning and representation learning.	30
2.16	Flow chart of a typical deep learning procedure	31
2.17	Various machine vision tasks(image credits: Slide 19 from cs231n 2017 lecture11)	32
2.18	Different 3D shape representations	34
3.1	The experimental PMD system configuration: left image is side view, right image is top view	40
3.2	3D position of the calibrated system	41
3.3	Those bigger dots in the dot grid indicate the orientation of the calibration board	41

3.4	Left points map shows the left camera calibration errors, right image shows the right camera calibration errors	43
3.5	The left image represents the y-direction phase results with points coordinates of circle marks, the right one shows the x-direction phase, both images are captured by left camera	44
3.6	Camera pixel errors of screen calibration	45
3.7	The procedure of screen calibration: Firstly, we obtain the accurate phase value, and screen coordinates can be represented by phase value. Eventually, we can obtain the transformation matrix between screen coordinates and camera coordinates	46
3.8	Initial surface searching	47
3.9	Surface profiles reconstructed from two cameras	48
3.10	Surface profiles fusion	48
3.11	Complete profile fusion workflow	49
3.12	The Packaged PMR system	50
3.13	Part of captured fringe images of plane mirror	50
3.14	Reconstruction results of different area in a 1/4 wave mirror	51
3.15	The left image shows the image of the wafer; the right images shows one of recorded fringe images.	51
3.16	The 4-inch wafer measurement result and the comparison of the line-profile between stylus(Taylorscan [®]) and PMD	52
3.17	The difference between two results	52
3.18	The aluminum foil and its surface texture	52
3.19	Part of captured fringe images of the aluminum foil	53
3.20	The surface topography of the aluminum foil and the comparison of two profiles	53
4.1	The schematic of the multi-scale measurement subsystem	56
4.2	Mechanical drawing of telecentric lens part(Ballon 2,4,9: adjustable lens tube)	56
4.3	LED characteristic parameter	58
4.4	Mechanical drawing of illumination part	58
4.5	the relation of optical power to current at 30°C The error bars show the standard deviation.	59
4.6	TIE optical system	59
4.7	Experimental results of the hardware system, the position along the optical axis is adjusted by a optical travel Translation Stages	60
4.8	The graph of LSTM DFF model	61
4.9	The Tensorboard graph of DFF RNN	64
4.10	a-d is the image simulated at different focus plane; e is the all-in-focus image; f is the depth map p ; x-y and x-z is the spatial section and the axial section of a PSF	64
4.11	The training loss and the validation loss of each training step	65
4.12	8 images taken along the optical axis	65

4.13	The left mesh graph shows the surface profile measured by the Taylorscan [®] stylus CMM; the right mesh shows the results reconstructed by the DFF RNN	66
4.14	The comparison between the result reconstructed from DFF RNN and the result from Taylorscan [®] . The profile plot is obtained by the red section plane	66
4.15	TIE phase retrieval algorithm flow chart based FFT Poisson solver .	67
4.16	Three images selected from the DFF image stack	67
4.17	The surface roughness obtained by TIE method	68
4.18	The comparison of roughness measurement between TIE method and Taylor Hobson stylus CMM	68
5.1	Coarase to fine smart measurement scheme	70
5.2	Neural network for mask perceptual (Courtesy of https:// skrish13.github.io/articles/2018-03/fair-cv-saga)	71
5.3	Smart measurement architecture	72
5.4	Measurement apparatus	73
5.5	Neural Network Data processing Stream	73
5.6	Global loss, bounding box loss and mask loss	74
5.7	The original image from the left camera and the rectified image obtained from PMR	74
5.8	The surface topography obtained from PMR	75
5.9	The inference result of a aluminum plate with defects	75

Symbols and Acronyms

Symbols

$\mathcal{R}^n$	the n -dimensional Euclidean space
$U(x, y)$	Complex wave
λ	Wavelength
$\vec{r}$	Vector in the detector plane
$\vec{\mathbf{n}}$	Normal vector
Z	The position along the optical axis
P	Focus power
i_t	Input gate of LSTM
f_t	Forget gate of LSTM
o_t	Output gate of LSTM
j	$\sqrt{-1}$
$h(x, y, z)$	Point Spread Function
$\sigma(\cdot)$	The Sigmoid function
σ_0, σ_1	The parameter of PSF
∇	Two-dimensional gradient operator
$\odot$	Element-wise multiplication operator
$*$	Convolution notation
$\langle \rangle$	Time-averaged operation in one period
J	Loss/cost function
ϕ	Phase information
$\mathbf{m}^T$	Hermitian transpose of $\mathbf{m}$
$\mathbf{m}^{-1}$	Inverse of matrix $\mathbf{m}$
$\mathcal{F}(f)$	Fourier transform of f
$\mathcal{F}(f)^{-1}$	Inverse Fourier transform of f
$\exp(x)$	The exponential function of x : e^x

x, y, z	Camera space coordinates
X, Y, Z	World space coordinates
u, v	Image(pixel) space coordinates
$\mathbf{W}$	Weight matrix of neural network
I_z	Image at position z

Acronyms

DOF	Depth Of Field
FOV	Field Of View
BS	Beam Splitter
PSF	Point Spread Function
SFF	Shape From Focus
DFP	Depth From Focus
LSTM	Long Short-Term Memory
ConvLSTM	Convolutional LSTM
RNN	Recurrent Neural Network
SML	Sum Modified Laplacian
CMM	Coordinate Measurement Machine
CCD	Charge-coupled Device
CMOS	Complementary Metal Oxide Semiconductor
CPU	Central Processing Unit
FPGA	FieldProgrammable Gate Array
GPU	Graphics Processing Unit
3D	Three-dimensional
2D	Two-dimensional
WTF	Windows Fourier Transform
FTP	Fourier-transform profilometry
FFT	Fast Fourier Transform
MEMS	Microelectromechanical Systems
LCD	Liquid Crystal Display
NG	Not Good
PDE	Partial differential equation
SLM	Spatial Light Modulator
VH	Volume hologram

FIS	Focused Image Surface
FISM	Focused Image Surface Method
AI	Artificial Intelligence
ROI	Regions of interest
CAD	Computer-aided Design
LED	Light Emitting Diode
TCP/IP	Transmission Control Protocol/Internet Protocol
SUT	Sample Under Test
CNN	Convolutional Neural Network
SNR	Signal noise ratio
TIE	Transport of Intensity Equation
TOF	Time of Flight
PMR	Phase Measuring Reflectometer

Chapter 1

Introduction

In this introductory chapter, the background and applications of optical metrology systems are discussed in Section 1.1. The objectives and major contributions are discussed in Section 1.2, followed by an outline of the thesis given in Section 1.3.

1.1 Introduction

With the development of manufacturing technology, highly accurate, non-destructive measurement methods and effective post-processing algorithms have become vitally important, especially the three-dimensional shape measurement of the wafer, MEMS or die. Generally, the machine vision has been used as a form of inspection. There are various application cases, and many researchers have done research works. Golnabi and Asadpour emphasized the importance of machine vision in the industrial process in [1]. Traditional machine learning methods have various successful examples in the two-dimensional inspection area, like ImageNet [2]. The two-dimensional photogrammetry technology based on machine vision technologies has been applied in industries widely, however, inspection techniques based on two-dimensional photogrammetry is not enough for more flexibility and more accurate measurement demands as refer in fig.1.1, three-dimensional shape or depth measurement needs to be involved.

FIGURE 1.1: Shape and topography measurement applications

FIGURE 1.2: Three-dimensional metrology

Due to the emerging of those applications mentioned above, many depth or three-dimensional shape measurement methods have been developed recently, the traditional contact three-dimensional shape measurement has a series of drawbacks, such as slow measurement speed and limited measurement ranges. Optical metrology has various advantages, such as non-contact, full-field, and high-resolution. These optical methods can be classified into 2 types, one is non-interferometric methods which includes confocal and deconvolution based methods, nano-scale measurement methods utilizing the TIE[3], TOF[4] and structure light methods[5], as illustrated in fig.1.2. The other is interferometric methods, such as digital holography methods and white light interferometer. These methods provide high accuracy from nanometer to millimeter scale. The architecture of three-dimensional vision system

comprises two parts, the first part is the optical subsystem, such as lens subsystem, diffraction optical elements, light source, and the other part is intensity transforming elements, which comprises CCD or CMOS sensors and digital processing units shown in fig.1.3. The measurement system can obtain crucial information, such as

FIGURE 1.3: two-dimensional Inspection System Architecture

the volume of defects or samples, the size of workpieces and the topography of a surface after appropriate processing. Those data is beneficial in various areas, such as products quality assessment, and in-situ re-manufacturing.

1.2 Objectives

The objective of this research is to develop a smart three-dimensional vision-based inspection system using novel 3D measurement technologies coupled with Deep Learning based analysis.

1.3 Scope

The scope of this project is limited to non-interferometric methods with deep learning for defects identification applied in the wafer or other electronics; such an inspection requires measuring large objects with micrometer axial resolution and zooming into the small region with a multi-scale resolution for inspecting smaller features.

For measuring larger specular surface, a PMR system is improved in this research. The method based on the TIE is used for measuring smaller feature with sub-micron resolution. Due to the limitation of TIE for large depth measurement, We proposed a depth from focus approach based on deep learning. For defect detection and mapping, deep learning for automatic classification and characterizing is proposed.

In summary, major ideas proposed in this thesis are:

- A new PMR setup for large field of view measurement.
- An efficient approach to extract the height information for an image stack based on deep neural work algorithm.
- A novel multi-scale metrology based on deconvolution methods and a novel reflective TIE approach.
- Based on the PMR and novel multi-scale metrology, introducing an inspection scheme which can provide the full-field measurement with micrometer resolution and the small region of interest measurement with nanometer resolution.
- Development of a computer vision system for on-line industrial inspection, focusing on deep learning based defect detection in two dimensions.

1.4 Outline

We provide a literature review that briefly gives definitions of important concepts in chapter 2, and an overview of past research works on three-dimensional shape measurement. Furthermore, the introduction of deep learning applications in machine vision is also included. We introduce our current works in the full-field measurement with PMR methods in chapter 3, and zoom to the measurement of smaller region with the TIE, and deconvolution methods in chapter 4. To combine these measurement techniques into a standard equipment for measuring large objects with micrometer axial resolution and at the same time to zoom into small regions with nanometer resolution, We introduce the the mask perceptual neural network in chapter 5. In chapter 6, we conclude the thesis with a discussion of main contribution of our research and possible direction for future studies.

Chapter 2

Literature review

2.1 Background

The three-dimensional (3D) information of the object or the scene provides intuitive information in various applications such as industrial inspection. The most intuitive information can assist in clearer comprehension of the characteristics of objects or scenes, and providing more accurate inspection results. Due to this reason, the 3D shape measurement has been studied for a long period of time in various area.

Instead of beginning with 3D optical metrology, researches started on the two-dimensional (2D) context to help the machine to sense the world. A mature architecture of machine vision proposed by D.Marr provides a theoretical basis for the development of the machine vision. The Machine vision in Marr's theory has three stages. The first stage is the primal sketch of the scene, in this stage, we mainly analyze the point, line and texture of an object, it is the 2D analysis in our context. The second stages is the 2.5D sketch, which mainly analyzes the surface patches with its depth and orientation. The 2.5D data is similar to the images people received through eyes, which contains recognizable information of the distance of the object, but the blocking parts remains unknown. Based on the 2.5D information, we got the three-dimensional model of the scene at the third stage.

In 3D optical measurement systems, we usually start from the 2D analysis. The only thing that can be obtained directly is the magnitude-square of the Fourier

FIGURE 2.1: 3D Optical Metrology, we focused on the methods in the dash box

transform of the underlying signal, which is the power spectral density. The depth information cannot be obtained directly, but we can use stereo vision approaches instead. With two cameras separated by a certain distance, we take two images from the same scene, as same as human eyes; the depth information can be calculated with the triangulation. Apart from the geometrical calculation based on the intensity distribution of the image, the phase information of the light wave also encodes enormously structural patterns of the image. These structural patterns can be used as an indicator for the shape of a surface in specific refraction index.

Currently, there are numerous optical three-dimensional measurement approaches as depicted in Fig.2.1. The coordinate measurement machine (CMM) is a traditional contact measurement methods based on point-by-point scanning on the sample surface. The CMM is slow and usually unbalanced between the speed and accuracy requirement in the industrial environment. For non-contact measurements, laser scanning methods were introduced, but they still require a point-by-point or line-by-line scanning. Some full-field inspection methods also have been studied, such as stereo vision methods, the time of flight (TOF) methods, grating based methods like PMR, and the TIE or deconvolution based methods. However, the TOF methods measure the distance by count the time of light flying, which only has millimeter resolution at most. Grating-based methods can measure the large objects like the 8 inches wafer in micrometer resolution without any scanning. With the sacrificing of the size of the field of the view, deconvolution-based methods can

extract the three-dimensional detail information with micrometer resolution both laterally and axially. However, it still cannot obtain nanometer resolution. There is a requirement to combine them with the nanometer resolution TIE based methods to implement a novel multi-scale measurement method to meet the increasingly demands from the industry. With the help of the deep learning algorithm, We put our efforts into the multi-scale measurement scheme to make the measurement techniques widely approachable. Mainly, there are two reasons for these efforts. One is to increase the measurement speed and accuracy, the other is automatically focusing and zooming into the smaller region in a large sample (e.g., defects on the aluminum plate and dies on the wafer) to analyze them at a multi-scale.

In this chapter, an introductory review on PMR, depth retrieval based on TIE, depth retrieval based on conventional deconvolution, and depth from focus methods are provided. In addition to metrology aspects, the introduction of deep learning methods, its application in metrology and inspection system is also provided. Although it is difficult to give an exhaustive and unambiguous description on the associated techniques, a descriptive summary is provided as a great help to define the background of the project which provides a focus for the thesis.

2.2 Grating based methods

Different from three-dimensional laser scanning methods, grating-based methods can obtain the depth information without any scanning component. Compared with TOF and stereo vision methods, grating based methods usually have higher resolution. One of the modern and mature grating-based methods is fringe projection profilometry, which is applied in the diffused object measurement. For measuring the specular or partially specular objects, a fringe reflection measurement technique called phase measuring reflectometry (PMR) was developed following the history of the fringe projection technique. Both of them are reviewed in this section.

The grating techniques include

- Grating pattern design: The grating pattern was utilized by Takeda, Mitsuho, Mutoh and Kazuhiro[6] in the 3D shape measurement firstly. In their

model, the grating pattern is expressed by the Fourier series expansion. Only one single spectrum of the fringe signal is used for Fourier transform profilometry. However, the spatial resolution of Fourier-transform profilometry (FTP) method is lower than fringe phase shift method[7]. In comparison with the FTP method, the hexagonal grating method can obtain better accuracy with single-shot only[8]. Another division of pattern design is to use a coded pattern to classify the certain pixels, once the actual pixel coordinates are determined at different camera views, the three-dimensional coordinates of the space point can be obtained by triangulating[9][10]. Also, the gray-code projection can also be used as a pattern[11]. Additionally, there are many new fringe patterns which were used to improve the synchronization between cameras and projectors, such as the SPMW grating fringe projection[12], Rochi fringe pattern defocusing[13].

- Grating analysis methods: Fourier transform methods proposed by Takeda's from the beginning[6], commonly, this type of methods can be classified as the Fourier Transform profilometry or FTP. The Fourier methods can prevent noise effects, but the global transform operation changes the whole spectrum results, which might be caused by the signal from other positions, such as the first harmonic superposes with other harmonics. The Gabor Transform and the Windows Fourier Transform (WFT) are possible solutions to reduce this error[14] [15]. Instead of using fixed windows size, the continuous wavelet transform based on wavelet theory can also be applied in the fringe analyzing[16]. Apart from these straightforward methods, a number of fuzzy methods such as neural network methods only require a single image to reconstruct the shape of the sample. In the low-resolution measurement, fuzzy methods appear to be superior to Fourier transform methods[17].
- Phase unwrapping techniques: It is difficult to obtain the height directly from the grating imaging. After grating analysis, we get the wrapped phase. Spatial phase unwrapping methods such as Goldstein's method[18] and quality-guided method[19] generally only need one wrapped phase image to obtain the unwrapped phase on each pixel based on the phase value of neighbor pixels. To unwrap the phase of the discontinuing surface, Saldner and Huntley proposed a temporal phase unwrapping method, which takes the advantage of the multiple wrapped phase to handle surface discontinuities.

2.2.1 Fringe projection profilometry

The fringe projection profilometry belongs to the current three-dimensional machine vision catalog. This method has been widely applied in various areas, from measuring the three-dimensional profile of MEMS[20] to the oral-dental measurement[21] and the large architecture deformation measurement for historical architecture conservation[22]. Fringe projection techniques were first proposed by Takeda to avoid the usage of two gratings in the Moire contouring technique[6]. With a much higher sensitivity than the Moire techniques, fringe projection methods have been studied by numerous researchers during the last three decades. Recently those techniques have been systematically reviewed by Sai Siva Gorthi and Pramod Rastogi in [5].

As illustrated in Fig.2.2, a fringe projection measurement system comprises of projection units, cameras, and image processing units. This setup cannot measure the specular sample, therefore the PMR system uses a special setup, which is illustrated in Fig.2.6. The pattern generated on LCD monitor illuminates the sample, and the camera records the image reflected off the sample distortions in pattern, which is proportional to the slope of the surface. The differentiation of the measured phase provides the curvature of this sample; the integration of slope phase provides the sample profile simultaneously.

FIGURE 2.2: Fringe projection profilometry system

2.2.1.1 Principle of fringe projection

The fringe projection takes advantage of the sinusoidal intensity distribution on diffused surface of the sample under the test, the projector projects the fringe pattern to the tested sample, which could deform the fringe intensity distribution. The mathematical model can be derived based on a simple geometrical approach as illustrated in Fig.2.3. In this model, we assume the projected light beam is collimated, which means the projected fringes are projected straightly to a sample and equally spaced, the fringe has a period of p_0 as seen on the reference plane X , and its intensity at the point C on the reference plane X is:

$$I_C = a(x, y) + b(x, y)\cos(2\pi OC/p_0) \quad (2.1)$$

a is the background intensity, b is defined as the fringe contrast, the point O is located at the intersection of the camera optical axis with the reference plane. In general, the term $2\pi OC/p_0$ of the cosine function in the Equ.2.1 represents the phase ϕ_R at the point C . The A_n on the image plane records the intensity at point D on the object plane or at C on the virtual reference plane X . The recorded intensity value is modified by the surface reflectivity $r(x, y)$ of the object, which can be represented as:

$$I_D = r(x, y)I_A \quad (2.2)$$

whose I_A is the intensity at the point A . The intensity at D is determined by the light beam AD , comparing with the phase on the reference plane X , the phase difference between point C and D is defined by ϕ_{CD} ,

$$\phi_{CD} = \frac{2\pi AC}{p_0} \quad (2.3)$$

from the Fig.2.3, the height BD is $BD = AC * \tan\theta_o(1 + \tan\theta_o/\tan\theta_n)$, the θ_n in many camera systems is nearly 90 degrees, hence the height $BD = AC\tan\theta_o$. A random point C_R on the sample surface has a phase value $\phi_R = 2\pi n + \phi'_R$, its intensity value is represented as:

$$I_R = a(x, y) + b(x, y)\cos(\phi_M + \phi'_R); \quad (2.4)$$

according to [23], we can obtain the $\tan\phi'_R$ by capturing the intensity N times ($N = 3$ or $N > 3$), i.e: $I_1, I_2, \dots, I_N$ with phase shifting $2\pi/N$ phase each time,

whose phase increment is $2\pi/N$. $\tan\phi'_R$ can be represented as:

$$\tan\phi'_R = \frac{\sum_1^N I_n \sin(2\pi n/N)}{\sum_1^N I_n \cos(2\pi n/N)} \quad (2.5)$$

We can obtain the height of the object followed by the phase unwrapping and phase-height mapping.

FIGURE 2.3: Ideal geometry of Fringe Projection System

2.2.1.2 Testing of 3D fringe profilometer from d'Optron

The validation of the fringe projection profilometry has been checked by a device from a local company called d'Optron. The setup is illustrated by Fig.2.4. We can use its customized software to control the device, the user interface is shown in Fig.2.4. The measurement results are depicted in Fig.2.5.

FIGURE 2.4: The hardware and software of the 3D profilometer

FIGURE 2.5: Left image is a 2D image of a mechanical component, the left image is the 3D profile of that component

2.2.2 Phase Measuring Reflectometry (PMR)

The surface characteristic of samples limits traditional fringe projection techniques. Only the diffused surface can be measured accurately, as the specular surface reflects fringes away from the camera. To solve this problem, the PMR can be used to measure the topography of the sample based on the principle of reflection light. The angle of reflection changes as the normal of the surface changes. Therefore the grating is deformed due to the variation of surface flatness. In the process of analysing the deformed image, we can obtain the topography of a reflective surface. We name this technique as Phase Measuring Reflectometry (PMR). Since we also displayed gratings to detect the deformed virtual image reflected by the reflective sample surface, It provided us with sufficient evidence to classify this technique into the category of grating techniques. The advantage of grating techniques is that it can provide the high-resolution and large field of view.

FIGURE 2.6: The Relationship between the reflected ray angle change and the sample's slopes.

2.2.2.1 Principle of PMR

The simple PMR setup and its geometrical light model is illustrated by Fig.2.6. From the camera side, the phase ϕ can be obtained in every pixel by phase shift methods as we mentioned in the last subsection. Using the geometrical model, the phase ϕ can be represented as:

$$\phi = d \cdot \tan 2\alpha \quad (2.6)$$

During the screen calibration, we can obtain a mapping function:

$$f : (\varphi_x, \varphi_y) \mapsto (x_s, y_s) \quad (2.7)$$

which maps the measured phase to the corresponding metric position of that point. According to Snell's law, the incident ray v from screen to camera can be determined as illustrated in the Fig.2.7. However, it is obviously that the reflect point is hard to determine if we only rely on one camera, the depth can be any point from depth z_1 to depth z_2 . To handle this problem, one more camera can be used to provide another incident light ray to form an intersection point. The normal vector $\vec{n} = (N_x, N_y, N_z)$ of that point can be calculated as:

$$\vec{n} = -\frac{\vec{w} + \vec{v}}{\|\vec{w} + \vec{v}\|} \quad (2.8)$$

FIGURE 2.7: Incident Ray according to phase

at that points, the slope are therefore calculated as:

$$\begin{cases} s_x = -\frac{N_x}{N_z} \\ s_y = -\frac{N_y}{N_z} \end{cases} \quad (2.9)$$

According to Southwell's configuration, the gradient is represented as:

$$\begin{cases} \frac{s_x^{p1} + s_x^{p2}}{2} = h^{p2} - h^{p1} \\ \frac{s_y^{p1} + s_y^{p3}}{2} = h^{p3} - h^{p1} \end{cases} \quad (2.10)$$

h^{p1} and h^{p2} is the height at point P_1 and P_2 respectively. The shape of that sample can be estimated widely in follows:

$$S = A \cdot H \quad (2.11)$$

whose S is:

$$\begin{bmatrix} \frac{s_x^{p1} + s_x^{p2}}{2} \\ \frac{s_x^{p3} + s_x^{p2}}{2} \\ \frac{s_x^{p4} + s_x^{p3}}{2} \\ \dots \end{bmatrix} \quad (2.12)$$

, H is the height grid to be reconstructed, $A \cdot H$ is the reconstruction matrix:

$$\begin{bmatrix} h^{p_1} - h^{p_2} \\ h^{p_3} - h^{p_2} \\ h^{p_4} - h^{p_3} \\ \dots \end{bmatrix} \quad (2.13)$$

We can obtain the solution H as:

$$H = (A^T A)^{-1} A^T S \quad (2.14)$$

The slope estimation(Equ.2.10) can be wrote in other form to improve the accuracy of PMR such as HFLI[24].

2.2.2.2 Testing of monoscopic PMR setup

A simple monoscopic PMR setup is built as depicted in Fig.2.8. After the system calibration, we can reconstruct the 3D profile of a 5cm diameter lens via this monoscopic PMR system. The result is shown in Fig.2.8. The FOV of the monoscopic PMR setup is 5cmx5cm, and the accuracy of the measurement result is affected by the relative position of the sample surface to the reference plane. Since the ambiguous normal vector cannot be resolved by the single camera, the reconstruction iteration of the monoscopic PMR setup is unstable compared with the stereo PMR setup that would be introduced in chapter 3, and the discussion of the accuracy of the monoscopic setup is not included in this thesis.

2.3 Depth from focus (DFF) and deconvolution

2.3.1 Introduction

It is challenging to achieve the lateral resolution in sub-micrometer scale with PMR. To obtain the higher lateral resolution, we apply focus fusion methods based on deconvolution to extract the depth information among an image stack. These methods utilize the focusing mechanism, which proves that objects can produce

FIGURE 2.8: Left image is the monoscopic PMR hardware setup, Right image is a lens measurement results

FIGURE 2.9: Formation of focused and defocused point

a sharp image region on the image sensor when the objects locate on the focal plane. reversely, a blurred image is generated by it out of the focal plane. With the Gaussian lens law:

$$\frac{1}{d_1} + \frac{1}{d_2} = \frac{1}{f} \quad (2.15)$$

The d_1 and d_2 is the distance of the object and the image to the principal centre of the lens respectively, the f is the focal distance of lens as illustrated in Fig.2.9. On the best focus plane, the image of those focused points can be formed on the image sensor. If there is an object whose surface topography exceeds the system depth of field (DOF), and the region exceeds the focal range is blurred. This phenomenon can be explained by the 3D point spread function (PSF) of the optical system. The intensity distribution on the image space of a point source in an imaging system is

defined by PSF. The acquired image arises from the convolution of the real object texture with the PSF. The texture can be reconstructed by the deconvolution of the acquired 2D image with the appropriate PSF. Once we map the correct texture or focal point to the corresponding distance, we can obtain the entire surface depth variations.

The quality of depth measurement results depends on the DOF of the optical system, considering the Rayleigh criteria and the geometrical circle of confusion, the DOF can be calculated by the wave-optical DOF (represented by DOF_w) or the geometrical DOF (represented by DOF_G) as:

$$DOF_w = \frac{\lambda \cdot n}{NA^2} \quad (2.16)$$

$$DOF_G = \frac{n \cdot e}{NA \cdot M^2} \quad (2.17)$$

The sign of λ is the wavelength of the light, and n is the refraction index of the medium, the NA is the the abbreviation of the numerical aperture, and the M is lateral magnification, the e is the circle of confusion size on the image sensor. A stack of images was taken along the optical axis by a single camera, they are commonly defined as z-stack images. Since we move objects or image sensor passing through the system's focal plane, and the focus point on certain images located at certain axis position, which correspond to the depth of that focus point. Hence, the focus point on an image needs to be found accurately to obtain the accurate depth value. The problem can be narrowed down to the problem of finding the focus points in a stack of images.

One type of solutions is using optical techniques to obtain focus points and to avoid the defocus information. Marvin Minsky invented an optical apparatus[25] which was thoroughly researched by C.J.R.Sheppard[26]. It only allow focus point form its image on the image plane, however, the optical method only suitable for particular hardware which usually has complex optical components, for the standard bright field microscopy or other conventional optical systems, Nayar proposed a computational method, which determines the focus point by focus measuring[27]. Another type of computational methods is deconvolution, which is used for improving the contrast of the digital image. The focus measuring methods calculate the best-focused point in the image, which is defined as a criterion to describe the

sharpness of that point. A.P.Pentlands introduced the focal gradients for measuring the number of defocus[28]. Laplacian is used for focus measuring firstly, the size of the Laplacian kernel is determined by the richness of the texture, if the surface has weak-texture, and the traditional methods are not robust enough, as presented by Nayar. The sum of modified Laplacian can be used in local windows instead. With the help of high-performance hardware, we can also apply the deep learning methods to obtain the correct focus points[29].

Even we can obtain correct focus points and its corresponding frame index, however, the information is still discrete. Furthermore, the discrete information needs to be refined by approximation techniques. This technique is assumed that the global surface is comprised of multiple partial surfaces which are formed by several points. It indicate that the focused points on an image are formed by points on the free form 3D surface instead of the 2D plane, i.e., the focus measures should be analyzed on the approximation of 3D surface[30] to improve the accuracy of measurement results. Besides, the approximation is also optimized from a single planar plane to the Bezier surface approximation[31].

Instead of measuring the focus, Aguet proposed a model based methods[32] based on the image formation model. In this model, the accuracy highly depends on the PSF.

2.3.2 Mathematical model and current solutions

It is reasonable to make an assumption that the sample surface texture distributes on the three-dimensional space can be represented by function $o(x, y, z)$:

$$o(x, y, z) = f(x, y)\delta(z - p(x, y)) \quad (2.18)$$

the $f(x, y)$ defines the surface texture on 2D space, the $p(x, y)$ defines the topography of the surface, i.e. the height of the surface. The δ is the Dirac delta function. With such an assumption model combined with the PSF model, we can propose the processing procedure as illustrated in Fig.2.10. Ideally, the images stack is generated as the convolution between the surface function $o(x, y, z)$ and the three-dimensional PSF of the optical system. It can be represented by the equation

FIGURE 2.10: Schematic representation of the topography and texture retrieval algorithm.

as:

$$\tilde{S}(x, y, z) = \int_{\mathbf{R}^3} f(u, v) h(x - u, y - v, z - p(x, y)) du dv \quad (2.19)$$

$h(x, y, z)$ is a 3D PSF, which can be obtained directly from a set of 2D PSFs along the optical axis[33] as illustrated in fig2.11. the $\tilde{S}(x, y, z)$ is the mathematical form

FIGURE 2.11: point source image at different location along optical axis

of the image stack, the error function between the calculated image stack and the

captured image stack is defined as:

$$J(f, p) = \int_{\mathbf{R}^3} (s(x, y, z) - \mathbf{s}(x, y, z))^2 dx dy dz \quad (2.20)$$

Where the $s(x, y, z)$ is the captured image stack. When the error function $J(f, p)$ reaches the minimum value, the corresponding function f and p is the maximum likelihood solution. The solution can be represented as:

$$f, p = \arg \min_{f, q} J(f, q) \quad (2.21)$$

The problem is fairly clear based on the model mentioned in Equ.2.19, therefore it requires that the best texture function f and topography function p are found through iteration methods with suitable algorithm, i.e. finding a suitable texture estimation and topography estimation model and updating these functions as:

$$f_{i+1}^{(\tau+1)}(x, y) = f_i^{\tau+1}(x, y) - \alpha \frac{J}{f(x, y)} \quad (2.22)$$

$$p_{i+1}^{(\tau+1)}(x, y) = p_i^{\tau+1}(x, y) - \alpha \frac{J}{p(x, y)} \quad (2.23)$$

Aguet's methods generate the topography by symmetric B-splines of order n dilated by m at resolution m [32].

$$p_m(\mathbf{r}) = \sum_{\mathbf{k} \in \mathbf{Z}^2} c_m[\mathbf{k}] \beta^n \left(\frac{\mathbf{r}}{m} - \mathbf{k} \right) \quad (2.24)$$

β updated according to m , i.e. $\beta_m^n = \beta^n(\frac{x}{m})$. $c[\mathbf{k}]$ updated with cost gradient function:

$$c_m^{(i+1)}[\mathbf{k}] = c_m^i[\mathbf{k}] - \alpha \frac{\partial J}{\partial c_m} \quad (2.25)$$

α is the gradient step size, which can be obtained by line searching[32].

The texture generation method is based on separable PSF model. We assume that the all focus image is formed by assembling in-focus segmentation of each image in the image stack[34], the PSF can be written as:

$$h(x, y, z) = \delta(x)\delta(y)h(z) \quad (2.26)$$

Substitute the h function Eq.2.26 in the Eq.2.19, we can obtain:

$$\tilde{s}(x, y, z) = f(x, y)h(z - p(x, y)) \quad (2.27)$$

and the error function can be represented as:

$$J(f, p) = \int_{\mathbf{R}^3} (s(\mathbf{X}) - f(x, y)h(z - p(x, y)))^2 d\mathbf{X} \quad (2.28)$$

Not surprisingly, by using the simplified PSF model, we can obtain a close form texture update solution for minimum error function $J(f, q)$:

$$f(x, y) = \frac{\int_{\mathbf{R}} s(x, y, z)h(z - p(x, y))dz}{\int_{\mathbf{R}} h(z - p(x, y))^2 dz} \quad (2.29)$$

in the paper [34], the $h(z)$ equals to 1.

In this method, the drawback is that the texture generation is not accurate enough due to the nature of the simplified PSF approximation, and the B-spline of topography function cannot provide accurate, but instead smooth results. As we mentioned, deconvolution methods are a type of high-density computing solution, and the accuracy of recovered shape highly depends on the PSF model. The same principle holds true as the simplified PSF model that the texture function is formed by the focus parts of each image in the stack. Shape from focus (SFF) methods, the focus parts are measured via focus measure operators, the operator can be represented as:

$$SML_i(x_0, y_0) = \sum_{pixel(x, y) \in U(x_0, y_0)} \left[\left(\frac{\partial^2 g_i(x, y)}{\partial x^2} \right)^2 + \left(\frac{\partial^2 g_i(x, y)}{\partial y^2} \right)^2 \right] \quad (2.30)$$

$U(x_0, y_0)$ is the pixels around point (x_0, y_0) . The $g_i(x, y)$ is the i -th image in the image stack. Generally, it is a focus measuring problem in the first stage, which is summarized by Malik and Choi [35] detail, and various focus measure operator has been analyzed in [36].

After focus measuring, we can only obtain the discrete depth information $i(x, y)$ of each sampling window(focus measure window), which is the frame index of pixel (x, y) , the depth information between two frames and pixel-level is lost. Approximation methods have been developed, Nayar applied interpolation methods

into that discrete depth information[27]. Instead of only measuring the focus energy on a single image, Subbarao and Choi[30] proposed Focused Image Surface Method (FISM), which is the surface formed by the set of points at which the object points are focused by the optical system. According to Gaussian lens law in Equ.2.15, $d_1(x, y)$ is the object shape and $d_2(x, y)$ is the focused image shape. In each sampling windows, we redefine the $i(x, y)$ as a planar surface, i.e. $i(x, y) = i_0 + p_0 * j + q_0 * k$. j, k represents the position at the lateral coordination along x, y axis respectively, in the selected frame index range t_{min} to i_{max} , we apply the following algorithm: The accuracy of measurement result can be improved by

Algorithm 1: Shape from focus based FIS algorithm

FIS: $I_0, L, \delta_i, \delta_p, \delta_q \triangleright$ Where I - initial depth position, L is the Laplacian value of each frame, $\delta_i, \delta_p, \delta_q$ is the searching step size for i, p, q , respectively. **begin**

```

    for  $i_m = i_{min}$  to  $i_{max}$  by step  $\delta_i$  do
        for  $p_m = p_{min}$  to  $p_{max}$  by step  $\delta_p$  do
            for  $q_m = q_{min}$  to  $q_{max}$  by step  $\delta_q$  do
                sum = 0 ;
                for  $j, k = -S$  to  $S$  do
                    sum = sum +  $\sum_{m=l+1}^{m=l-1} (L[m][j][k])^2$ ;
                if sum > maxsum then
                     $[i, p, q]_p = [i, p, q]_m$ 
            return  $i_p(x, y) = i_p + p_p * j + q_p * k \triangleright$  the output of this function is estimation
            results based on focus measuring on FIS

```

choose the appropriate surface mathematical model, such as choose higher surface order instead of planar surface[37], or the fuzzy surface approximation[38].

2.4 Transport of Intensity Equation (TIE)

2.4.1 Introduction

Transport of Equation (TIE) can define the connection between the phase and the derivative of intensity along the optical axis. TIE is a non-interferometric method for phase-measuring in the homogeneous medium; it is valid for bright field microscope with partially coherent light source[39][40]. TIE was proposed by Teague[41] [3] with monochromatic coherent beam, and After that it has been

researched by Roddier in the application of adaptive optics for astronomy[42]. Streibl [43] is the first one who points out the validity of the TIE in the spatially partially coherent imaging system. Under Fresnel approximation, TIE needs 2 defocused images orthogonal to the optical axis at least to calculate the derivative of the intensity, the accuracy of the results and the signal-to-noise ratio is affected by the defocus distance. The resolution of the measurement result is out to the diffraction limit.

TIE techniques have been developed a lot for many years. These researches can be roughly categorized as follows:

- The solution of the partial derivative equation: To solve this second order partial differential equation (PDE) is mainly to find a solution to the boundary value problem. Teague first proposed the Green's function methods and treated the region boundary as Dirichlet boundary [3]. Roddier also applied the Zernike polynomial to divide the TIE solution into a series of Zernike polynomials[44]. With the assumption of periodic boundary condition, Fourier transform is a practical solution that has applied broadly to solve the PDE[45] [46]. However, for large objects that exceeds system field of view, the periodic boundary condition is not accurate enough, to avoid this problem, the discrete cosine transform is applied to solve the TIE, the main contribution of this method is the zero boundary intensity assumption[47].
- Study of partially coherent fields: Firstly, Teague's TIE was extended by Streibl for a general case of partial coherence light. Paganin and Nugent defined the partially coherent light using the concept of the time-averaged Poynting Theory[39]. Gureyev interpreted the partially coherent beams and proposed its use in quantitative imaging[48]. Apart from these space-time correlation functions, the partially coherent light can also be expressed regarding the light field, Wigner proposed a light field distribution function in 1986, which provides a description of the space and the radiance at the same time, has been applied in the field description in the TIE[49].
- Approximation of intensity derivative: Usually the intensity derivative approximation is based on the finite difference method, but when the Δz is too large, the linearity of finite difference assumption will break down. Soto

and Acosta[50] applied multiple images to make the derivative approximation, this method can overcome the noise effects, but for large Δz , the higher order derivatives are treated as errors. A method proposed by Waller using higher order intensity derivatives to account for noise and non-linearity effects[40][51][52].

- Acquisition of intensity: TIE requires several intensity distributions along the optical axis, the goal of the acquisition system is to obtain these images as fast as possible. Traditional methods acquire images by shifting the position of the CCD/CMOS or the objects in time-lapse[53]. To achieve fast acquisition, forming the images at the different position on the image sensor can be used with the sacrifice of the field of view[40] [54]. Apart from these methods that based on the wave propagation or the frequency modulation, the wavelength also a good tool, Since the nature of chromatic dispersion in the optical system, we can obtain the different defocused-image using different wavelength with the sacrifice of accuracy.

2.4.2 Mathematical model and current solutions

In an optical system, Under paraxial approximation, and assuming a monochromatic light field $u(x, y, z)$ propagating along the z direction, (x, y) always located in the plane perpendicular to the optical axis.

$$u(x, y, z) = \sqrt{I(x, y, z)} \exp[ikz + i\phi(x, y, z)] \quad (2.31)$$

the phase $\phi(x, y, z)$ is a low gradient function. Compared to $2\pi/\lambda$, it varies slowly, i.e. $|\nabla\phi| \ll 2\pi/\lambda$. the derivative of intensity along the optical axis contains phase information can be retrieved via TIE[3]:

$$-k \frac{\partial I(\vec{r})}{\partial z} = \nabla \cdot [I(\vec{r}) \nabla \phi(\vec{r})] \quad (2.32)$$

the k in the equation is $2\pi/\lambda$, $\vec{r}$ is the position vector (x, y) . ∇ is the two-dimensional gradient operator over $\vec{r}$. z represents the optical axis, normal to the (x, y) plane. But usually the wave is not pure monochromatic wave, it is quasi-monochromatic wave, represented by a free space scale wave $U(\vec{r}, t)$, $\vec{S}(\vec{r}, t)$

is the Poynting vector fields, According to Poynting's theorem, we can obtain S :

$$\vec{S}(\vec{r}, t) = -\frac{1}{8\pi}[\nabla U(\vec{r}, t)\frac{\partial U^*(\vec{r}, t)}{\partial t} + \nabla U^*(\vec{r}, t)\frac{\partial U(\vec{r}, t)}{\partial t}] \quad (2.33)$$

We define the time-averaged irradiance $I_{avg} = \langle U(\vec{r}, t) \rangle$, $\langle \rangle$ is the time-averaged operation in one period. The $S_{avg}(\vec{r})$ is the time-averaged Poynting vector fields, in paper [39], they used the Helmholtz decomposition theorem, which points out the fields can be represented by the sum of the gradient of a scale part and the curl of a vector part, and backed with Poynting's theorem in time-averaged form:

$$\nabla \cdot \vec{S}_{avg}(\vec{r}) = \nabla \cdot I_{avg}(\vec{r})\nabla\phi_S(\vec{r}) + \nabla I_{avg}(\vec{r}) \cdot [\nabla \times \vec{\phi}_V(\vec{r})] = 0 \quad (2.34)$$

$\phi_S(\vec{r})$ and $\vec{\phi}_V(\vec{r})$ is the scale phase function and the vector phase function respectively in the Helmholtz decomposition theorem.

$$\phi_S(\vec{r}, t) = -\frac{1}{\bar{w}} \int \frac{\nabla \cdot \vec{S}_{avg}(\vec{r}')}{|\vec{r} - \vec{r}'|} d^3\vec{r}' \quad (2.35)$$

$$\vec{\phi}_V(\vec{r}, t) = \frac{1}{\bar{w}} \int \frac{\nabla \times \vec{S}_{avg}(\vec{r}')}{|\vec{r} - \vec{r}'|} d^3\vec{r}' \quad (2.36)$$

In par-axial approximation, the Equ.2.34 can be wrote as:

$$-k \frac{\partial I(\vec{r})}{\partial z} = \nabla[I(\vec{r})\nabla\phi(\vec{r})] + \nabla I(\vec{r}) \cdot [\nabla \times \vec{\phi}_V(\vec{r})] \quad (2.37)$$

If the vector phase part is vanished, which is in the case of coherent fields, Equ.2.37 can be wrote same as Eq.2.32 in the case the wave is quasi-monochromatic and the intensity is uniform. The final phase can be obtain using following equation:

$$\phi(\vec{r}) = -k \frac{1}{I_0} \nabla^{-2} \frac{\partial I(\vec{r})}{\partial z} \quad (2.38)$$

To solve the Equ.2.32, finding an auxiliary function $\Phi(\vec{r})$ is a common technique:

$$\nabla\psi(\vec{r}) = I(\vec{r})\nabla\phi(\vec{r}) \quad (2.39)$$

The TIE Equ.2.32 reduces to Poisson function:

$$\nabla^2\psi(\vec{r}) = -k \frac{\partial I(\vec{r})}{\partial z} \quad (2.40)$$

According to standard PDE solution formula, and with respect to x , the Fourier transform for the n -th partial derivative of the function $f(x, y)$ can be wrote as:

$$\mathcal{F}\left(\frac{\partial^n f(x, y)}{\partial x^n}\right) = i^n x_*^n \mathcal{F}(f(x, y)) \quad (2.41)$$

The solution of auxiliary function $\phi(\vec{r})$ can be given by:

$$\psi(\vec{r}) = \mathcal{F}^{-1}\{(r_*)^{-2} \mathcal{F}[k \frac{\partial I(\vec{r})}{\partial z}]\} \quad (2.42)$$

x_* and r_* are the value of the variable conjugate to x and $\vec{r}$ in the Fourier space respectively. The phase solution can be found under the assumption that the phase satisfies periodic boundary conditions, the solution can be wrote as:

$$\phi(\vec{r}) = -\mathcal{F}^{-1} r_*^{-2} \mathcal{F}(\nabla \cdot I^{-1}(\vec{r}) \nabla \psi(\vec{r})) \quad (2.43)$$

These exist models are limited due to several physical assumptions, To obtain the Equ.2.43, two assumption are made, first one is the mathematical model of light source, the bandwidth of the light source should small enough to make sure the Equ.2.43 is valid, usually a narrow bandwidth filter or laser are used in TIE-based measurement system, and the vector field in Equ.2.37 can be ignored. Second one is the finite derivative approximation, Using Taylor expansion theory, the actual intensity derivative without approximation can be given as:

$$\frac{\partial I}{\partial z} = \frac{I(\Delta z) - I(0)}{\Delta z} - \left[\frac{\Delta z}{2!} \frac{\partial^2 I}{\partial z^2} + \frac{(\Delta z)^2}{3!} \frac{\partial^3 I}{\partial z^3} + \frac{(\Delta z)^3}{4!} \frac{\partial^4 I}{\partial z^4} + \dots \right] \quad (2.44)$$

but in usually the higher order terms is ignored, it will involved unwanted error in some situation, we will give a detail analysis in next subsection.

2.4.3 Current experimental setup

To build a TIE optical system that the main function is obtained from two defocus images along the optical axis, as illustrated in Fig.2.12, according to the TIE, with the approximation:

$$\frac{\partial I_z(x, y)}{\partial z} \approx \frac{I_{z+\Delta z}(x, y) - I_z(x, y)}{\Delta z} \quad (2.45)$$

FIGURE 2.12: Optical System used for TIE

transitional method obtains $I_{z+\Delta z}(x, y)$ and $I_z(x, y)$ by changing the position of the CCD/CMOS or the sample along the optical axis directly[43]. However, it slowed down the acquisition speed and involved the lateral displacement. Instead of changing the image plane or the objects mechanically, a spatial light modulator can be used for modifying the optical distance length[55]. Another method is applying the electrically tunable lens to obtain equivalent defocus distance with higher speed as illustrated in Fig.2.12, which can reach 15 frames per second in Chao Zuo's paper[53]. Usually, the electrically tunable lens is controlled by currents, to obtain quantitative defocusing distance, the accurate and stable lens control system needs to be very necessary, and current-(defocus-distance) curve needs to be calibrated regularly.

Apart from acquiring several images at a different time, different defocused-images can be obtained at a different position on the camera sensor. Images are captured on an image sensor by adding an SLM between 2 tube lenses but with different configuration compared with [55]. The first tube lens relays the back focal plane of the objective on a tilt mirror and an SLM device at the same time with the help of beam split. The second tube lens can generate two images at a different lateral position on image sensor since the tilt mirror[54], in this setup, the pattern on SLM does not need to adjust during the acquisition, and can only obtain two images at the different depth.

With the help of volume hologram (VH), a focal stack of intensity images can be obtained in a single-shot[40]. As illustrated in Fig.2.12, the VH that in the center of the 4-f lens system has several multiplexed holographic gratings. Each grating performs as a spatial-spectral filter modulate the images from a slightly different position along the optical axis of the first tube lens L1 on the different position of the image sensor. To split the focal stack images, we need to set the second tube

FIGURE 2.13: Transmission TIE system(Courtesy of d’Optron)

lens off-axis in angle θ . However, without doubts, the field of view of these two methods will decrease. The registration processes of these images also involve new error.

To overcome the limitation of extra optical elements such as beam splitter, tunable lens, VH plate, the equivalent defocus images can be obtained by chromatic dispersion of the lens system in fig 2.12 with the white light source. The image formed by different wavelength is located at different image plane along the optical axis. By utilizing a color image sensor, we can obtain three images at different axis position at different color channel simultaneously[56].

A local company has implemented a transmissive TIE system. We have evaluated their TIE system as shown in Fig.2.13. This validation of this system has been checked in a biological application, the 3D shape of cells can be shown in Fig.2.14.

2.5 Deep learning methods

The PMR can measure Full-field results of samples, but it will take much time if the multi-scale method is used to measure the whole sample directly. To inspect the regions of interest (ROI) efficiently, we use the deep learning methods to locate the defect, die or other ROI, then use the multi-scale measurement in that small region to get fined inspection results. This section provides a short review of the background and typical applications of the deep learning algorithm.

FIGURE 2.14: Three images dz , 0 and $-dz$ are captured by the TIE system from d’Optron, then 3D profile (Right image) of the biological sample can be reconstructed.

2.5.1 Introduction

Deep learning is a class of machine learning; it is in the family of artificial intelligence methods. Because of the computer gathers knowledge from the history of the world or the physic phenomenon, there is no need for a human operator or very specific or known physic principle formally to specify all of the knowledge needed by the computer. The deep learning layer allows the computer to learn complicated concepts by building them out of simpler ones; a graph of these layers would be many layers deep, which is one of many reasons why we call it deep learning methods.

The relationship between machine learning and deep learning can be illustrated in Fig.2.15. Many deep learning tasks can be done by designing the right set of hallmarks to extract for that task, then providing these hallmarks to a machine learning algorithm. For example, a useful characteristic for PSF model identification from a raw image is an estimation of the parameters of certain PSF model. Another example is applying useful characteristic to find best focus points among an image stack. However, to choose a set of useful features is a big problem.

Finding the mapping from representation to output and to the representation itself by machine learning algorithm is a solution to this problem. This approach is known as representation learning. But in the real world, it can be challenging to obtain such high-level, abstract hallmarks from raw data. Deep learning

FIGURE 2.15: A Venn diagram showing the relationship between deep learning and representation learning.

solves this central problem in representation learning by introducing representations which are expressed regarding other, simpler representations. Learning the right representation for the data provides one perspective on deep learning. In general, Deep learning belongs to the broader family of machine learning that achieves great power and flexibility by representing the world as a nested hierarchy of concepts, with each concept defined with simpler concepts, and more abstract representations computed regarding less abstract ones.

Regarding three-dimensional processing, three-dimensional scanning technological developments have been leading to the expansion in the amount of three-dimensional data. Datasets of three-dimensional CAD models are expanding continuously, predominantly through aggregation of three-dimensional content on the web. Fringe projection and other technology for scanning and reconstruction those are mentioned in chapter 2 are providing increasingly higher fidelity geometric representations of objects that can eventually become CAD-quality models. At the same time, there are many new research problems due to fundamental difficulties in using three-dimensional data, in geometric analysis, there are two basic problems, i.e., computing segmentations of three-dimensional shapes, and establishing correspondences between these segmentations. Recognition of shapes from partial scans is also a research goal in computer vision. Scene understanding from two-dimensional images is a grand challenge in vision that has recently benefited tremendously from three-dimensional CAD models[57][58]. Defects detection and in-situ re-manufacturing and robotic manipulation are several important application[59] that benefit from an understanding of three-dimensional shapes.

FIGURE 2.16: Flow chart of a typical deep learning procedure

Meanwhile, the growing market of crowd-sourcing for three-dimensional modeling and increasing three-dimensional sensing device amount make it possible to use the deep learning algorithm to accomplish these three-dimensional vision tasks, such as three-dimensional classification and three-dimensional segmentation as depicted in Fig.2.17.

2.5.2 Deep learning in 2D machine vision

Image classification and segmentation is a task of assigning an input image one label from a fixed set of categories, and the image is represented as one sizeable 3-dimensional array of numbers. This task is one of the core problems in machine vision. For example, the image is shown in Fig.2.17, the neural network takes a

FIGURE 2.17: Various machine vision tasks(image credits: Slide 19 from cs231n 2017 lecture11)

single image as input and do segmentation in pixel-wise, and assign different colors to the cat, dog, and sofa in the image.

Instead of trying to specify what every one of the categories of interest looks like directly in code, deep learning is referred to as a data-driven approach, since it relies on first accumulating a training dataset of labeled images. There are some popular image datasets:

- MNIST/NIST: handwritten digits dataset.
- Perturbed NIST: handwritten digits with tons of deformations.[60]
- CIFAR: 32x32 wild images with 100 categories.[61]
- Caltech 101/256: images of items belonging to 101/256 categories.[62]
- Caltech Silhouettes: 28x28 binary silhouettes of the picture in Caltech 101 dataset.
- The Street View House Numbers (SVHN) Dataset.[63]

- NORB: binocular images of toys with different pose under different illumination[64].
- Imagenet: image database organized according to the WordNet hierarchy.[2]
- Pascal VOC: various object recognition challenges.
- COIL 20: various items imaged at every angle in a 360-degree.[65]
- MS COCO: a large-scale object detection, segmentation. [66]
- Open Images Dataset: a dataset of almost 9 million URLs for images. These images have been annotated with image-level labels bounding boxes spanning thousands of classes. The dataset contains a training set of 9,011,219 images, a validation set of 41,260 images and a test set of 125,436 images.[67]
- Face image dataset: Labeled Faces in the Wild(13,000 images of faces collected from the web, labeled with the name of the person pictured);
- Yale Face Database: includes gray-scale images of 15 individuals. There are 11 images for every subject, one for every different facial expression: center-light, w/glasses, happy, left-light, with or without glasses, normal, right-light, sad, sleepy, surprised, and wink.[68]

The neural network will take an array of pixels and assign the right label to it. The complete pipeline beginning with a set of N images labeled with different classes which is the training set of our neural network. Then we can update the parameters and weights in our neural network with the training set. In the end, we need to evaluate the quality of the neural model.

2.5.3 Deep learning in 3D machine vision and optical problem

Unlike the 2D image, which has different representation with a regular data structure that can be stored as the matrix, the 3D shape representation has many different representations, such as multi-view RGB-depth (RGB-D) images, volumetric, polygonal mesh, point cloud, primitive-based CAD models. Among those three-dimensional shape representations, multi-view RGB-D images and volumetric can be treated as a regular rasterized form of data or data with regular grids.

FIGURE 2.18: Different 3D shape representations

However, the point cloud and others are represented in geometric form, which is irregular.

Due to the regular grids, the regular three-dimensional representation can be processed by existing CNN network as standard images, but irregular three-dimensional data representations have several problems in conventional deep neural network, also the three-dimensional shape representations is generally high dimensional, which means that it is easily prone to over-fit due to the “curse of dimensionality”. Some unique deep learning algorithms will be introduced in the next subsection.

Computer vision researchers have typically developed object recognition algorithms from two-dimensional features of two-dimensional images. Then they used these features to classify new two-dimensional pictures of those objects. In the three-dimensional shape case, one can directly train recognition algorithms on three-dimensional features such as voxel occupancy or surface curvature. Due to the introduction of sizeable three-dimensional shape repositories[69], it is possible to build a three-dimensional classifier of three-dimensional shapes from three-dimensional representation directly[57]. Instead of building a classifier from three-dimensional representation, Hang Su[70] also try to classify two-dimensional images, which is taken from multi-view from that three-dimensional shape. These methods can be fit for different three-dimensional shape representation, since it converts the three-dimensional shape into 2D images, to maintain more three-dimensional information, we can also render the RGB-D images from the three-dimensional shape[71], the dimension of data is decreased compared with volumetric and primitive-based CAD three-dimensional representation.

However, these solutions do not take advantages of the abundant three-dimensional

data, during the multi-view 2-D images rendering, some useful data will be deposited if some parts of that three-dimensional shape is being hidden, or the three-dimensional shape is represented by using the point cloud, in other words, the viewpoints selection and the input SNR(signal noise ratio) are several practical problems.

Instead of rendering three-dimensional shapes to multi-view 2D images, we can choose different three-dimensional neural network architecture for different three-dimensional shape representations. Several early three-dimensional CNN like the 3D ShapeNet[57], VoxNet[72] were proposed based on two-dimensional CNN. Based on these neural networks, three-dimensional volumetric data can be fully used in the classifier. However, the information loss of primitive-based CAD three-dimensional data still remains at a high level. OctNet[73] and O-CNN[74] can increase the data resolution by storing those volumetric data in a efficient way so called “Octree”. For the point cloud representation, which is unordered and unstructured point data, it can be processed by PointNet[75] directly. The PointNet is a unified framework for various tasks by end-to-end learning for unstructured, unordered point data, these tasks include objects classification, part segmentation, data completing, etc. Different from the three-dimensional CNN[57], and due to the nature of point cloud data, the hierarchical feature of point data is not included, and no local context for each point during the training and inference. PointNet++[76] overcomes this problem by using PointNet in local regions, and aggregating local features by PointNet again. Some paper also used primitive-based three-dimensional data as neural network input[77] to accomplish these machine vision tasks.

The deep learning has recently been applied to solve optical problems such as in super-resolution [78], acceleration of the speed of computed tomography(CT)[79] and holography [80] [81].

2.6 Conclusion

According to the review of the 3D shape measurement with PMR and multi-scale measurement techniques in literature, we find that there is a gap in their adoption in the industry due to the lack of measurement schemes for measuring large objects and to zoom into small regions with higher accuracy simultaneously. For instance,

the FOV in current PMR is not sufficient enough, phase retrieval methods based on TIE are mainly used in the transmissive sample, the depth from focus methods is computation-intensive. We can still make numerous improvements in various aspects within or combining related technical procedures. Therefore, in this project, we put the main effort on both processing algorithms and measurement scheme for samples with different surface characteristics and different sizes.

For the PMR, the following aspects should be investigated:

- During the measurement procedure, the phase is related to the slope of the sample surface, which means a sample with step profiles is difficult to be measured. Giving a solution for profiling step features is essential.
- Numerous PMR hardware configurations only use one camera to obtain the phase; the FOV is not enough for colossal objects. It is an enhancement to acquire phase by more than one camera.
- In many industrial applications, PMR usually gives a full-field result, but for the small region of interest, we cannot apply the PMR due to the huge time consumption. Hence, we need to find some flexible measurement schemes to meet the industrial requirement.

For 3D measurement methods based on the TIE and deconvolution, the investigation in several aspects mentioned below is necessary for measuring samples with multi-scale resolution.

- 3D measurement methods based on TIE do not have sufficient depth of field due to the assumption of TIE. Also, TIE based methods cannot give a reasonable profile for the surface still slope. The depth from focus methods can be combined with TIE to handle the reflective sample with a still slope.
- Methods based on deconvolution is computation-intensive, an efficient processing algorithm will increase the performance of the measurement system.
- More and more applications are required to fully measure large objects with micrometer axial resolutions and at the same time to zoom into small regions with nanometer resolutions. In this aspect, it is necessary to develop a measurement scheme for inspection of specular objects at different scales and

focusing the region of interest in the full-field result with the deep learning method.

Chapter 3

PMR for large objects with specular surface

In this chapter, we utilize fringe methods to obtain the topography of specular samples. The aim of this chapter is to illustrate the necessary procedures of PMR. The first main focus is the experimental setup. Secondly, we would explain the mathematical model of the experimental setup and the calibration methods. In further sections, we introduce a FOV enhanced method and its reconstruction results. Finally, We conclude this chapter with the highlighting of several current developments and potential directions for future research.

3.1 System Configuration

We implement a PMR system with two calibrated camera as illustrated in Fig.2.6. To obtain highly accurate results in this system, we use 19" LCD to display fringes, and utilize two FLIR Blackfly S[©](Model: BFS-U3-16S2M) cameras whose resolution is 1440×1080 and two lenses with 12mm EFL to record deformed fringes. The pixel size of the 19" planar screen in this setup (Fig.3.1) is 0.294 mm.

FIGURE 3.1: The experimental PMD system configuration: left image is side view, right image is top view

3.2 System calibration

To find normal vectors of the sample under measuring, we have to obtain positions of the screen and cameras with respect to the reference plane. Hence, in this section we would introduce the system calibration which includes the camera and screen calibration. Firstly, we utilize conventional least square methods to obtain the intrinsic parameters of the camera and the transformation matrix between two cameras in our system. In the next step, we set a predefined reference plane to calibrate the screen. After the screen calibration, the spatial relation of the screen and the camera which represented by T_{sc}, R_{sc} can be obtained by extrinsic parameters computing function as same as the one used in camera calibration. At the same time, we can obtain the position of the predefined reference plane by calculating the extrinsic parameters (T_{ref}, R_{ref}) of the left or right camera. Finally, based on coordinates of the marks points on the mirror which is the reference plane, we can obtain the geometric information of the whole system. The results can be visualized as Fig.3.2.

3.2.1 Camera calibration

Commonly, we use chess board to calibrate the camera under pinhole camera model[82]. In the fringe profilometry, chess board will affect the phase extraction accuracy, hence a standard calibration mirror plate with dot grids as shown in Fig.3.3 was used in our experimental setup to eliminate the effects of pattern in the

FIGURE 3.2: 3D position of the calibrated system

FIGURE 3.3: Those bigger dots in the dot grid indicate the orientation of the calibration board

phase measuring stage. The camera calibration problem can be solved in 2 stages. In the first stage, we use closed-form solution to obtain the rotation matrix R_{wc} and transformation vector T_{wc} , whose subscript wc represents the relation between the world coordinate and camera coordinate is given a point P_w in the world space.

We can find the respective point P_c in the camera space by applying Equ.3.1.

$$\begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} = R_{wc} \begin{bmatrix} X_p \\ Y_p \\ Z_p \end{bmatrix} + T_{wc} \quad (3.1)$$

x_p, y_p, z_p and X_p, Y_p, Z_p represent the position of point P in the camera coordinates and the world coordinate respectively, but those two cameras only captured the two-dimensional image, which means it is necessary to carry out projective transformation on P_c to obtain the corresponding point P_u , whose subscript c, u indicates camera coordinate and image coordinate respectively. The projective transformation is a kind of intrinsic transformation, which can be represented as

$$s \begin{bmatrix} u_p \\ v_p \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x'_p \\ y'_p \\ 1 \end{bmatrix} z_p \quad (3.2)$$

whose $x'_p = \frac{x_p}{z_p}$ and $y'_p = \frac{y_p}{z_p}$. A represents the first term of the right-hand equation. Therefore, with $\mathbf{H} = AR_{cw}T_{cw}$, the objects point P and its image point p is related as $\mathbf{sp} = \mathbf{HP}$. Hence, the intrinsic and extrinsic parameters can be solved by solving a linear equation. The final step is modeling the lens distortions based on a disparity with the real circumstance of projection. Therefore the equation can be rewritten as

$$\begin{aligned} x'' &= x' \frac{1+k_1r^2+k_2r^4+k_3r^6}{1+k_2r^2+k_4r^4+k_6r^6} + 2p_1x'y' + p_2(2x'^2 + r^2) \\ y'' &= y' \frac{1+k_1r^2+k_2r^4+k_3r^6}{1+k_2r^2+k_4r^4+k_6r^6} + 2p_2x'y' + p_1(2y'^2 + r^2) \\ r^2 &= x'^2 + y'^2 \\ u_p &= f_x x'' + c_x \\ v_p &= f_y y'' + c_y \end{aligned} \quad (3.3)$$

whose $k_1, k_2, k_3, k_4, k_5, k_6$ are radial distortion coefficients. p_1 and p_2 are tangential distortion coefficients. To calibration these coefficients, we can apply a physically meaningful approach i.e maximum likelihood estimation, which can be obtained by solving a nonlinear minimization function:

$$\sum_{i=1}^n \sum_{j=1}^m ||p_{i,j} - P_f(A, \mathbf{k}, R_i, T_i, P_j)||^2 \quad (3.4)$$

FIGURE 3.4: Left points map shows the left camera calibration errors, right image shows the right camera calibration errors

$p_{i,j}$ represents the point j pixels locations in the image on i th position of the calibration plate. $\mathbf{k}$ is the lens distortion vector, in this report, we do not consider high order distortion, but taking the first 5 coefficients. Camera calibration results are depicted by Fig.3.4.

Each point in the Fig.3.4 represents the pixel error $p_{i,j} - P_f(A, \mathbf{k}, R_i, T_i, P_j)$ on image coordinate system of the computer in pixels.

3.2.2 Screen calibration

The screen is so called fringe generator. It is used to provide phase reference as depicted in Fig.3.5, i.e., we can find the standard phase reference according to finding the corresponding point on the image, whose relation can be visualized by the following mapping function:

$$f : \begin{pmatrix} \phi_x \\ \phi_y \end{pmatrix} : \begin{pmatrix} u \\ v \end{pmatrix} \quad (3.5)$$

To find the LCD's the position and the phase distribution on the reference plane, we have to record the phase distribution ϕ on the reference plane and its corresponding world coordinates at the same time. We put a 1/4 wave mirror on the sample's holder to obtain the phase distribution. There are several spots with standard size printed on the mirror as shown in Fig.3.3, and the world coordinates of the screen can be reconstructed based on the camera's intrinsic parameters and positions of spots on the standard calibration plate which is shown in Fig.3.3.

FIGURE 3.5: The left image represents the y-direction phase results with points coordinates of circle marks, the right one shows the x-direction phase, both images are captured by left camera

Unfortunately, the phase value on the spots area is not accurate due to the changing of the reflectance in that area. To acquire the accurate phase distribution, we can utilize the mapping function which is approximated by two polynomials of degree two using Polyfit function of MATLAB based on 10×10 pairs of phase-point, i.e. 10×10 pairs of $(u, v) \rightarrow \phi_x$ and 10×10 pairs of $(u, v) \rightarrow \phi_y$ respectively, the polynomials can be represented as

$$\begin{bmatrix} \phi_x & \phi_y \end{bmatrix} = \begin{bmatrix} 1 \\ v \\ u \\ uv \\ uv^2 \\ u^2 \\ u^2v \\ u^2v^2 \end{bmatrix}^T \cdot \begin{bmatrix} f_x & f_y \end{bmatrix} \quad (3.6)$$

To achieve the simplicity of the formula, we present this polynomials as

$$\phi = \mathbf{u} \cdot \mathbf{F} \quad (3.7)$$

In the practical application, we can easily find the $\mathbf{F}$ if the rank of ϕ and the rank of $[\phi|\mathbf{u}]$ equal to the rank of full rank matrix $\mathbf{F}$. For each spot on the calibration plate, we can find its camera coordinates (x, y) with the intrinsic parameters of cameras. The screen position can be eventually represented in camera coordinates

FIGURE 3.6: Camera pixel errors of screen calibration

as:

$$\begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} = R_{sc} \begin{bmatrix} \phi_x \\ \phi_y \\ \mathbf{0} \end{bmatrix} + T_{sc} \quad (3.8)$$

The reprojection error can be measured in pixel as illustrated in Fig.3.6. The screen calibration procedure can be illustrated as Fig.3.7.

3.3 Shape reconstruction

3.3.1 Surface optimization

We utilized phase shifting methods to obtain the phase information of left camera and right camera. To reconstruct the profile of the surface which is the z in the Fig.3.8, we can use the normal field of that surface, and the profile of the surface can be reconstructed by integrating its normal vectors. After the phase measuring, we can determine the location of W_1 and its corresponding pixel location V_1 through the calibration of the system. However the position of the surface Z remains unknown. As shown in Fig.3.8, there are many potential locations of Z . We can find an approximating solution for this problem by using iterative optimization

FIGURE 3.7: The procedure of screen calibration: Firstly, we obtain the accurate phase value, and screen coordinates can be represented by phase value. Eventually, we can obtain the transformation matrix between screen coordinates and camera coordinates

method. We can define the cost function as:

$$J_{normal}(S_Z) = \int_{S_Z} \|\mathbf{n} - \mathbf{n}_{S_Z}\|^2 dS_Z \quad (3.9)$$

Hence, the optimized surface profile can be represented as:

$$S_{optimized} = \underset{S_Z}{argmin}(J(S_Z)) \quad (3.10)$$

The speed of the iterative optimization can be significantly increased by the initial surface approximation by reducing the steps of iteration. In this system, we can find the initial surface by two cameras. Firstly, we selected 100 pixel points in the common FOV of those two cameras, and based on these points, we can determine their location $\mathbf{W}_1$ and $\mathbf{V}_1$ with respect to right camera. Secondly, for each point, we make a approximation on the Z location, which is the location of the point on the surface. We can easily find the pixel location ($\mathbf{V}_2$) of this point on the left camera and the phase value on that point, which represents the screen location

FIGURE 3.8: Initial surface searching

($\mathbf{W}_2$). Based on $\mathbf{W}_1, \mathbf{V}_1, \mathbf{W}_2, \mathbf{V}_2$, we can obtain two normal vector $\mathbf{n}_1, \mathbf{n}_2$. We can search all the Z values from a particular range to find the correct Z to meet the condition that $\|\mathbf{n}_1 - \mathbf{n}_2\| = 0$. Finally, we can fit a nearest plane that based on these 100 points, and then we can start surface optimization from that plane.

To increase the speed of surface reconstruction, we can execute the surface optimization of two cameras together. Eventually, we can obtain two optimized surface profile from two cameras.

3.3.2 Shape fusion

To extend the FOV of PMR, we proposed a fusion scheme to utilize two cameras in our PMR system. As mentioned above, two surface profiles S_L and S_R are reconstructed based on the left and right camera respectively. The two surface profiles can be visualized in Fig.3.9.

The fusion workflow is shown in Fig.3.11. The registration algorithm is based on ICP algorithm[83]. The fusion result is shown in Fig.3.10. The FOV of PMR system can be increased from $160mm \times 150mm$ to $200mm \times 190mm$ as shown in Fig.3.10.

FIGURE 3.9: Surface profiles reconstructed from two cameras

FIGURE 3.10: Surface profiles fusion

3.4 Experiment results

We packaged the system in Fig.3.1 with black poly methyl methacrylate glass plate to prevent the effect of environmental light source, which can be visualized in Fig.3.12. To validate this system, a $1/4$ wave mirror and a 4-inch wafer were measured, the results are shown in figure.3.16. A straight line in the left image of Fig.3.16 is selected on the wafer randomly, and the comparison was shown in the right image of Fig.3.16.

FIGURE 3.11: Complete profile fusion workflow

3.4.1 The measurement of a plane mirror

We put the 1/4 wave mirror with a diameter of 150mm into the the PMR system as shown in Fig.3.1. We displayed a single fringe on the screen to achieve the phase measuring. The intensity of each pixel on the screen can be described as

$$I_C = 255(a + b\cos(2\pi OC/p_0)) \quad (3.11)$$

FIGURE 3.12: The Packaged PMR system

FIGURE 3.13: Part of captured fringe images of plane mirror

The fringe number p_0 was set to 1, and the OC is the width (phase measuring along the x -axis of the PMR system) or the height (phase measuring along the y -axis) of the screen. The a, b was adjusted to prevent over-exposure and under-exposure of cameras. Since we projected the fringe with a single frequency, the wrapped phase needs to be unwrapped by direct phase unwrapped methods, which will not be discussed in this thesis.

The surface profile of the wave mirror is obtained at different areas as shown in Fig.3.14. The first row in the Fig.3.14 is the mirror image obtained from the left camera, and we selected the $2cm \times 2cm$ rectangle at the different part of the mirror to reconstruct the surface profile which can be visualized in the second row of Fig.3.14. It can be seen that the peak-to-peak height is less than 2.2 micrometer which means that the system error is around 2 micrometer.

FIGURE 3.14: Reconstruction results of different area in a 1/4 wave mirror

FIGURE 3.15: The left image shows the image of the wafer; the right images shows one of recorded fringe images.

3.4.2 The measurement of a wafer

The second experiment is the wafer measurement. We measured a 4-inch wafer which has rectangular patterns and some marks on its surface which can be visualized in Fig.3.15. The measurement configuration is as same as the mirror measurement. The measurement result can be visualized in Fig.3.16. To compare results between the PMR system and the Taylorscan[®] system, we utilized the center mark in the wafer as shown in Fig.3.15. A profile line with a length of 80mm was obtained by the Taylorscan[®] system and the PMR system, and the comparison is shown in the right image of Fig.3.16. The difference between these two profiles is illustrated in Fig.3.17.

FIGURE 3.16: The 4-inch wafer measurement result and the comparison of the line-profile between stylus(Taylor Hobsen Stylus[®]) and PMD

FIGURE 3.17: The difference between two results

FIGURE 3.18: The aluminum foil and its surface texture

3.4.3 The measurement of an aluminum foil

The full-field measurement of a thin aluminum foil was demonstrated to provide the manifest 3D information. The thickness of this aluminum foil is 0.2mm , and the size is $2\text{cm} \times 3.5\text{cm}$. This aluminum sample is shown in Fig.3.18. The image of the surface texture was captured with 10x objective. The recorded fringe images can be partially visualized in Fig.3.19. As we can see from the Fig.3.20. The full-field result can be reconstructed, however, the profile of the surface texture were lost

FIGURE 3.19: Part of captured fringe images of the aluminum foil

FIGURE 3.20: The surface topography of the aluminum foil and the comparison of two profiles

due the limited resolution of the PMR system. Due to the steep slop of the defect, the depth of defects has a large error compared with the result of Taylorscan[®].

3.4.4 Conclusion

In experiments mentioned above, we demonstrated the validity of the PMR system. The difference between the contact measurement system and our PMR system is less than $5\mu\text{m}$ in wafer and mirror sample. However, the accuracy of the PMR system was reduced by the steep slope on the sample, which will be further resolved in next chapter.

3.5 Summary

The research of the PMR presented in this chapter was investigated by the study of surface reconstruction and shape fusion. The main contribution of the research in this chapter is that it proposed a novel scheme to measure larger objects with micrometer resolution at a single glance. With this scheme, the PMR system is able to extend its FOV to match industrial demands by utilizing two or more cameras.

In this thesis, the PMR with large FOV provides with the manifest 3D information for the further inspection of devices or defects. Hence the PMR is applied at the first stage to obtain the full field 3D profile with the micrometer resolution, and the region of interest can be classified and segmented based on the full field 3D profile. Overall, The research of PMR provides powerful measurement techniques to achieve full-field, high axial resolution (few micrometers) measurement capabilities for the defects or devices localization and segmentation.

Chapter 4

Multi-scale measurement subsystem based on DFF and TIE

According to the principles mentioned in the chapter 2, this chapter presented a multi-scale depth measurement system, which is used for the reflective sample. DFF is exploited to measure the dies or defects which have a large gradient or a big step size, it is less accurate to measure them by PMR. The depth of field of this system can also be increased via DFF method. However, the DFF methods still cannot obtain the roughness of the surface at the nanoscale. Hence, we proposed the DFF and TIE based method to implement the measurement subsystem to obtain the measurement results with a multi-scale resolution. The system calibration procedure and processing algorithm are introduced in this chapter, and in the end of this chapter, the experiment results and error will be analyzed in the end of this chapter.

4.1 System configuration and calibration

As mentioned in chapter 2, the vertical scanning capability is required in both TIE and DFF based methods. Instead of moving the sample holder or the image sensor mechanically, the focus plane is moved by phase modulation on the Fourier plane of a 4F system[53] in this research. The experimental system can be separated into two parts, which are the telecentric lens subsystem for scanning along the optical

FIGURE 4.1: The schematic of the multi-scale measurement subsystem

FIGURE 4.2: Mechanical drawing of telecentric lens part(Ballon 2,4,9: adjustable lens tube)

axis and the illumination subsystem for illuminating the reflective sample. The schematic of this system can be visualized in Fig.4.1.

4.1.1 Telecentric lens subsystem

Firstly, we setup a 4F lenses system, and the mechanical drawing of this system is shown in Fig.4.2. The complex amplitude distribution U_f in the focal plane of the

first lens is the Fourier transform of the input image, using Equ.4.1:

$$U_f(u, v) = \frac{\exp[j\frac{k}{2f}(v^2 + u^2)]}{j\lambda f} \times \iint_{-\infty}^{+\infty} U'_l(x, y) \exp[j\frac{k}{2f}(x^2 + y^2)] \exp[-j\frac{2\pi}{\lambda f}(xu + yv)] dx dy \quad (4.1)$$

whose U'_l is the disturbance behind the lens, which is

$$U'_l(x, y) = U_l \exp[-j\frac{k}{2f}(x^2 + y^2)] \quad (4.2)$$

on that focal plane, a tunable lens was added, similar to the Equ.4.2, the disturbance after the tunable lens can be written as

$$U'_f(x, y) = U_f T_l(x, y) \quad (4.3)$$

whose T_l is the phase transformation of the tunable lens without the constant phase factor, $T_l(x, y)$ can be written as $\exp[-j\frac{k}{2f'}(x^2 + y^2)]$, f' is the focal length of the tunable lens. For a normal 4F system without the tunable lens, the $U'_f = U_f$, and will be transformed by the second lens in the 4F system. Including the Fresnel transformation in the free space, the U_f can be rewritten as:

$$U_f(u, v) = \frac{\exp[j\frac{k}{2f}(1 - \frac{d}{f})(u^2 + v^2)]}{j\lambda f} F_0(\frac{u}{\lambda f}, \frac{v}{\lambda f}) \quad (4.4)$$

F_0 represents the Fourier spectrum of the input image, and the d represents the distance between the input image and the first Fourier lens. Substituting Equ.4.4 into Equ.4.3, we have

$$U'_f = \frac{\exp[j(\frac{k}{2f}(1 - \frac{d}{f}) - \frac{k}{2f'})(u^2 + v^2)]}{j\lambda f} \mathcal{F}(U'_f) \quad (4.5)$$

the distant part in Equ.4.5 can be rewritten as

$$\frac{k}{2f}(1 - \frac{d}{f}) = \frac{k}{2f}(1 - \frac{d}{f}) - \frac{k}{2f'} \quad (4.6)$$

hence the distance difference $d' - d$ can be expressed as

$$d' - d = \frac{f^2}{f'} = f^2 P \quad (4.7)$$

FIGURE 4.3: LED characteristic parameter

FIGURE 4.4: Mechanical drawing of illumination part

P is the optical power of the tunable lens. We can adjust the P to acquire the intensity distribution at different z -position.

4.1.2 Illumination subsystem

To obtain the profile of the reflective sample by TIE or DFF methods, a collimated LED illumination was built in this system. The LED illumination only utilized a single LED and a collimation lens. The color control unit and the RGB chip was integrated in a package of 5050 components. The mechanical and optical characteristics of the LED in this system is shown in Fig.4.3. The design of this illumination system is visualized in Fig.4.4. The tunable lens tube was used for adjusting the collimation lens to fit different applications. The beam splitter in Fig.4.4 is represented by **BS**, and it built the connection between the collimated LED and the telecentric subsystem.

FIGURE 4.5: the relation of optical power to current at 30°C The error bars show the standard deviation.

4.1.3 Focus calibration

In the experimental system, we used an electrical tunable lens, which is from Optotune[®]. The optical power was controlled by the current. As depicted in the Fig.4.5, the relation of the optical power to current at 30°C is obtain by experiments. Considering the whole optical system, which is illustrated in Fig.4.6, we changed the position of the sample along the optical axis via a x-y-z stage, and adjust the current to find the best focused image based on the energy of the intensity gradient[84]. According to the curve in Fig.4.7, we can extract the distance

FIGURE 4.6: TIE optical system

FIGURE 4.7: Experimental results of the hardware system, the position along the optical axis is adjusted by a optical travel Translation Stages

along the optical axis according to the currents value.

4.2 DFF based on Convolution LSTM

Given a stack of images, trained researchers can easily determine the depth distribution. According to the image formation model mentioned in chapter 2, the topography of the sample can be retrieved by guessing the right texture and appropriate the point spread function. To simplify this problem, we used the RNN to obtain the depth disparity map of the image stack, and reconstruct the relative profile of the surface with the distance between each two neighbor images. Convolution LSTM is an advanced RNN model for sequential learning problem. In this section, we defined DFF as a sequence-to-weight problem, and they can be solved convolution LSTM RNN.

4.2.1 Concepts

Concepts behind the RNN is to extract the information inside a series of data. RNN is widely applied in the language processing or stock trends prediction. The feature of RNN based DFF is that it takes all the information in the stack into its consideration, and the results yielded by convolution LSTM is based on the global focus distribution and spatial correlation.

Assuming there is a image stack $\mathbf{S}_N$, and we define two neighbor images as S_t and S_{t+1} . The disparity map or the height distribution among these images is represented as $\mathcal{H}$. Hence, the function of the DFF RNN can be given by

$$DFF\ RNN : \mathbf{S}_N \rightarrow \mathcal{H} \quad (4.8)$$

The CNN mentioned in [29] only utilized the two neighbor images to obtain the disparity map or the crucial hidden layer for further processing. The LSTM RNN model takes all the images in the stack into consideration, an additional memory unit C_t is included in this model to “memory” the global information dynamically. As a result, the output H_t of each LSTM cell is leveraged by all images in the stack. The procedure can be expressed as

$$LSTM\ DFF : \begin{cases} S_t, H_{t-1}, C_{t-1} \rightarrow H_t, C_t \\ \mathbf{H}_N \rightarrow \mathcal{H} \end{cases} \quad (4.9)$$

The graph model is visualized in Fig.4.8. The rules of the parameter updating in ConvLSTM is the same as the work studied by Xingjian, S. H. I., et al.[85].

FIGURE 4.8: The graph of LSTM DFF model

4.2.2 Implementation

The basic model is bidirectional RNN[86]. Based on the DFF problem, we implemented 8-steps bidirectional ConvLSTM model, and trained it with simulated image stacks.

The ConvLSTM has an input gate i and a forget gate f to control current inputs and history memories into the memory cell C . It can be mathematically represented

as

$$\begin{bmatrix} i_t \\ f_t \end{bmatrix} = \begin{pmatrix} \sigma(\cdot) \\ \sigma(\cdot) \end{pmatrix} \odot \left(\mathbf{W}_i f * \begin{bmatrix} S_t \\ H_{t-1} \\ C_{t-1} \end{bmatrix} + W_c \odot C_{t-1} + \mathbf{b}_{if} \right) \quad (4.10)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(\mathbf{W}_c * \begin{bmatrix} S_t \\ H_{t-1} \end{bmatrix} + \mathbf{b}_c) \quad (4.11)$$

The output H_t of each ConvLSTM cell is controlled by an output gate o_t , which can be expressed as

$$o_t = \sigma \left(\mathbf{W}_o * \begin{bmatrix} S_t \\ H_{t-1} \end{bmatrix} + W_c \odot C_t + \mathbf{b}_o \right) \quad (4.12)$$

And H_t is

$$H_t = o_t \odot \tanh(C_t) \quad (4.13)$$

The depth disparity map is the final output of this model, which can be expressed as

$$\mathcal{H} = \mathbf{W}_h * \mathbf{H} \quad (4.14)$$

We implemented this ConvLSTM model above by Tensorflow-backend Keras framework. There are 8-time steps, and 8 filters with shape (3,3,8) in the convolution operation in the ConvLSTM model.

Without considering the batch size, the input shape is (8,256,256,1) and the output shape is (256,256,1). There are 8 images in the input tensor, and a matrix of depth disparity $\mathcal{H}$ in the output tensor. It is a tedious work to capture the real image for training. To reduce the labour work, we simulated the image stack with the RGB-D dataset provided by [87][88][89][90]. The RGB image I in this dataset was captured by the camera with small numerical aperture, it is safe to assume that I is a all-in-focus image. We chose an (256,256) area randomly in this image and its depth map $p(x, y)$, and simulating blur images with even depth steps between the nearest distance to the furthest distance. The point spread function $p(x, y, z)$ is a 3D Gaussian function, and the variance of Gaussian σ is modulated by the distance as $\sigma(z) = \sigma_0 + \sigma_1 z$. To generalize the model, we used different pairs of

σ_0, σ_1 to represent the PSFs of different optical systems. The PSF is expressed as

$$h(x, y, z) = \frac{1}{\sigma(z)^2} e^{-\frac{(x^2+y^2)}{2\sigma(z)^2}} \quad (4.15)$$

The simulating procedure can be represented as

$$S_t(x, y) = \int_{\mathbf{R}^3} I(u, v) h(x - u, y - v, z_t - p(x, y)) du dv \quad (4.16)$$

For each image stack, the images at different focus plane is generated with a fixed σ_0, σ_1 , and with z_t as

$$\begin{cases} z_t &= p_{min} + z_{step}t \\ z_{step} &= \frac{p_{max} - p_{min}}{N_{timesteps}} \end{cases} \quad (4.17)$$

And the ground truth is the depth disparity map $\mathcal{H}$, it can be obtained from the depth image p by

$$\mathcal{H} = N_{timesteps} \frac{p}{p_{max} - p_{min}} \quad (4.18)$$

The loss function of this model is expressed as

$$J = \frac{1}{\mathbf{R}^2} \int_{\mathbf{R}^2} (\tilde{\mathcal{H}}(x, y) - \mathcal{H}(x, y))^2 dx dy \quad (4.19)$$

Once the weights of this model was converged with the simulating data, we fixed the weights of ConvLSTM parts in this model, and re-train the CNN parts $\mathbf{W}_h$ with the inputs data captured from the experimental system (shown in Fig.4.1) to fit this model for our practical usage.

4.2.3 Results

The graph of this model is generated based on Tensorboard; it can be visualized in Fig.4.9. We simulated 10325 image stacks based on the SUN RGB-D dataset[87], and with 4 different PSFs. The values of σ_0, σ_1 are (0.2, 0.02), (0.2, 0.04), (0.2, 0.08) and (0.2, 0.12) respectively. One of simulating image stack is visualized in Fig.4.10. The loss of each training step is shown in Fig.4.11, and there are 8325 steps per epoch. We trained this model 15 epochs totally, and at the end of each epoch, the model was validated with 400 validation data image stack which was taken randomly from another 2000 image stacks. The training is completed with

FIGURE 4.9: The Tensorboard graph of DFF RNN

FIGURE 4.10: a-d is the image simulated at different focus plane; e is the all-in-focus image; f is the depth map p ; x-y and x-z is the spatial section and the axial section of a PSF

FIGURE 4.11: The training loss and the validation loss of each training step

FIGURE 4.12: 8 images taken along the optical axis

the Nvidia GTX1060 graphic card for 30 hours, which is, 2 hours per epochs. The validation loss is also shown in Fig.4.11 Real experiments were carried out in the multi-scale measurement subsystem. We utilized a Chameleon3[®] gray camera with resolution of 1928×1448 , and mounted the multi-scale measurement subsystem to a BX51m Olympus[®] microscope with 10x objective. The size of the camera sensor is $3.69\mu m$. A series of image was taken along the optical axis as shown in Fig.4.12, and the inference result obtained via the DFF RNN can be visualized by Fig.4.13.

Based on the focus calibration, we can obtain the distance d_t between two neighbor images according to the current. The reconstructed topography, shown in Fig.4.13, is better than the results reconstructed by the conventional DFF methods mentioned in chapter 2. The comparison between the result obtained by DFF RNN and the result from Taylorscan[®] is shown in Fig.4.14. The top transparent layer in the left image is the profile obtained from Taylorscan[®] stylus scanner and the comparison of the line profile between Taylorscan[®] is shown in the right image of

FIGURE 4.13: The left mesh graph shows the surface profile measured by the Taylorscan[©]sytlus CMM; the right mesh shows the results reconstructed by the DFF RNN

FIGURE 4.14: The comparison between the result reconstructed from DFF RNN and the result from Taylorscan[©]. The profile plot is obtained by the red section plane

Fig.4.14.

4.3 Measurement method based on TIE

According to Chapter 2, the phase can be reconstructed by the FFT Poisson equation solver. The processing strategy is depicted in Fig.4.15. In this section, the TIE reconstruction result is based on images taken from the DFF experiment, we selected a smaller region ($1\text{mm} \times 0.6\text{mm}$) as shown in Fig.4.16, and the result can be visualized in Fig.4.17.

We scanned a line profile with the length of $500\mu\text{m}$ on the surface of the defect, the lateral resolution of Taylorscan is $1\mu\text{m}$, and the lateral resolution of the TIE system is $0.369\mu\text{m}$. The roughness R_a obtained from the TIE system is $0.9621\mu\text{m}$,

FIGURE 4.15: TIE phase retrieval algorithm flow chart based FFT Poisson solver

FIGURE 4.16: Three images selected from the DFF image stack

and the R_a obtained from the Taylorscan[®] is $0.8748\mu m$. Considering the result from Taylorscan[®] was smoothed by the tip of the stylus, we can use the TIE system to obtain a more accurate result. The comparison of roughness profile can be visualized in Fig.4.18.

4.4 Summary

We proposed a new method (DFF RNN) to measure the step object with micrometer resolution. This method constitutes a series of improvements over current

FIGURE 4.17: The surface roughness obtained by TIE method

FIGURE 4.18: The comparison of roughness measurement between TIE method and Taylor Hobson stylus CMM

DFF methods in terms of the speed of reconstruction and the quality of results. The nanoscale measurement technique (reflective TIE) was implemented in this research to measure reflective samples based on TIE.

The research of the multi-scale measurement methods presented in this chapter was investigated by the study of DFF and reflective TIE. The main contribution of the research in this chapter is that it proposed the DFF RNN model to measure objects which has large steps with micrometer resolution, and the nanoscale roughness is able to be measured by the reflective TIE.

Chapter 5

Smart metrology with the deep learning

5.1 Introduction

Firstly, the PMR will provide a processed high-resolution 2D/3D data, and then the deep neural network will highlight the defect and interested items on the full-field image. We are using trained Mask R-CNN[91] and UPerNet[92] neural network to classify the defects. In the two-dimensional image segmentation, Mask R-CNN, which is regional convolution neural network, has two processing stages: the first stage is scanning the image and generating proposals, which is the areas most likely to contain a certain object, the second stage is classifying the proposals and generating bounding boxes and masks. The UPerNet is a unified perceptual parsing neural networks, which can perceive multi-target from a given image. Based on the bounding box, we can obtain the detail profile and roughness of those areas of targets by the multi-scale measurement approach. The whole processing procedure is depicted in Fig.5.1.

In this chapter, we propose a new smart measurement architecture to implement the whole processing procedure mentioned above. The PMR and multi-scale measurement approach are coupled by the mask perceptual neural network, and the DFF methods are enhanced by the DFF RNN.

FIGURE 5.1: Coarase to fine smart measurement scheme

5.2 The neural network configuration

Mask R-CNN was introduced in year 2017 to extend its predecessor Fast R-CNN and back-boned with ResNet50 [93], a popular framework for object detection. Mask R-CNN is used for extending it with instance segmentation among other things. ResNet (residual network) uses short connection connects the feature input to the next x layer output. ResNet50 means a ResNet with 50 layers. Combined with FPN[94] which can improve the feature extraction by adding more feature pyramids at different layers, the backbone feature map can be well generated based on the implementations mentioned above. Once the feature map is obtained, the RPN, which is used for target regions extraction, can be directly applied on it to improve the processing speed. in the next stage, a branch of neural network for perceiving an objective mask is added in parallel with the existing branch for ROI(region of interest) box recognition. The whole neural network configuration is depicted in Fig.5.2.

The additional branch is functioned as a mask perceptron. The MLP(multi-layer perceptrons) is applied to the neural network based on the RPN output, which containing the prediction results and their bounding boxes. The following neural networks focus on mask prediction without concern for their class, and the speed of training and perceiving can be acceleration.

FIGURE 5.2: Neural network for mask perceptual (Courtesy of [https:// skr-ish13.github.io/ articles/2018-03/fair-cv-saga](https://skr-ish13.github.io/articles/2018-03/fair-cv-saga))

5.3 Smart measurement architecture

5.3.1 Concepts

The computer vision system for on-line industrial inspection involves various detection and measurement techniques. To pave the way for the 3D inspection, we propose the smart measurement architecture constituted a series of components as follow:

- **Measurement apparatus:** The Measurement apparatus is comprised by the PMR and the multi-scale measurement subsystem, providing a series of hardware interfaces to the local processing unit. The local processing unit is controlled by the measurement procedures via these interfaces.
- **Processing Units:** Processing units is used for controlling the measurement apparatus to measure the sample with different techniques at different areas, and the multi-scale measurement is applied to a particular area by moving the multi-scale measurement subsystem. Processing units is able to determine the region of interest by the local accelerated processing server or remote servers.
- **Accelerated processing server:** The processing server is used for handling the perceptual task. Different models are loaded based on the request from the processing unit. The perceptual neural network determines the area of interest; the DFF RNN can determine the depth map. The inference result is sent back to the processing unit after finishing the perceptual task.

The smart measurement architecture can be visualized in Fig.5.3.

FIGURE 5.3: Smart measurement architecture

5.3.2 Implements

We implemented the measurement apparatus as shown in Fig. 5.4. In the PMR system, the processing unit is used for controlling the screen and two point-gray cameras, and sending the result obtained from PMR to the local accelerated processing server via TCP/IP. The multi-scale measurement subsystem is mounted inside to complete the measurement in the region of interest, and its inspection area can be controlled by a x-y linear stage.

5.4 Results

5.4.1 Training of mask perceptual neural network

Defects detection experiments are applied two-dimensional images firstly, different defect types will be analyzed based on the output of our neural network, then the defects will be measured by the multi-scale measurement system in detail.

The deep neural network will be trained to classify and segment different components such as capacitors, resistors, copper, scratches and digs. The location and the size of each class can be compared with the standard data to tell whether this sample is good or not. The complete procedure can be depicted in Fig. 5.5.

FIGURE 5.4: Measurement apparatus

FIGURE 5.5: Neural Network Data processing Stream

For the system input, we apply the image-centric training on the SUT image as discussed by [95], the model is implemented based on Keras tensorflow backend framework, and the Nvidia GPU 1060 is used for training the model. The data used for training is captured from 500 real sample images which contain different electronic components were captured under random light illumination which combining the coaxial, ring and dome light. To train and test the neural network for components segmentation and classification, we labeled different components with different names. The ratio of training, testing and validation is 7:2:1. To increase the training speed, we trained the initial model weight based on the ImageNet[96].

The training contains 160 epochs, which takes 160 hours to complete. The loss of the last main layer is recorded in Fig.5.6.

FIGURE 5.6: Global loss, bounding box loss and mask loss

5.4.2 Segmentation and classification

According to the output of the neural network, we can find locations of defects on the sample, and it will be fully measured by the multi-scale measurement equipment developed by our team based on nanoimager from d’Optron a local company of Singapore.

Firstly, the rectified image shown in Fig.5.7 was obtained in the PMR measurement, Simultaneously, the surface topography of this sample is obtained as shown in Fig.5.8.

FIGURE 5.7: The original image from the left camera and the rectified image obtained from PMR

FIGURE 5.8: The surface topography obtained from PMR

FIGURE 5.9: The inference result of a aluminum plate with defects

The segmentation and classification of the defect on a aluminum plate is shown as Fig.5.9.

Once the area of defects is located and highlighted by the MASK RCNN, the multi-scale measurement approach mentioned in chapter 4 can be invoked by the smart measurement system to fully analyze the defect.

5.5 Summary

The main topic we discussed in this chapter is the development of a computer vision system for on-line industrial inspection, which focus on deep learning based on defect detection in two and three dimensions. The smart measurement architecture proposed in this chapter pave the way for combining different measurement techniques into a standard equipment. With the help of this standard equipment,

we can measure large objects with micrometer axial resolution and at the same time to zoom into small regions with nanometer resolution.

Chapter 6

Conclusion

6.1 Summary of results

In this report, we mainly reviewed the following optical metrology: the TIE phase retrieval, the deconvolution and SFF, and the fringe structural light. We also introduced deep learning methods applied to high dimensional data processing in the end.

In addition, the following contributions are made through this research and described in this thesis.

- **The FOV enhanced PMR:** A dual camera PMR for measuring big objects was proposed and built. Based on the stereo PMR, the FOV is enlarged by synthesizing the result that is reconstructed from two camera.
- **The multi-scale measurement system:** A reflective multi-scale measurement system is developed for measuring reflective objects at micro-scale and characterizing roughness at nano-scale.
- **The DFF RNN:** A depth from focus system with deep learning was developed. Based on the RNN and LSTM model, the DFF RNN provides a alternative solution to reconstruct the depth from a image stack. The experiment results show the greater performance over conventional DFF with respect to the reconstruction speed and quality.

- **The smart measurement architecture:** A smart measurement system for smart 2D inspection in multi-scale was developed. A initial implementation of this architecture was described. And experimental results carried on the aluminum plate shows the potential value of this architecture.

6.2 Outlook for future research

More future research are expected on the signal processing and three-dimensional data post-processing based on deep learning. Future works can be 2 parts, the first part is improving the accuracy and speed of optical metrology using deep learning, considering the whole signal processing procedures, a unified deep learning framework needs be studied to handle the noise which could be introduced during the image acquisition. The uncertainty of phase or height retrieval involved in TIE needs to be researched to provide a more robust measurement techniques for reflective objects. For DFF methods, the performance of RNN models can be improved in further studies to replace the cost-insensitive algorithm.

The second is to establish a series inspection benchmark to validate a particular smart measurement system that is implemented based on our smart measurement architecture. The system can be modularized to allow the system integrator to implement their inspection system easily.

List of Author's Awards, Patents, and Publications

Journal Articles

- Yan, Ketao, Yingjie Yu, **Chongtian Huang**, Liansheng Sui, Kemao Qian, and Anand Asundi. "Fringe pattern denoising based on deep learning." *Optics Communications* (2018).
- Sui, Liansheng, MinJie Xu, **Chongtian Huang**, Achyut Adhikari, Ailing Tian, and Anand Asundi. "Multiple-image encryption by space multiplexing based on vector quantization and interference." *OSA Continuum* 1, no. 4 (2018): 1370-1384.

Conference Proceedings

- **Huang, Chongtian**, Hongtao Zhong, and Anand Asundi. "Fun with optics: a novel optical engineering kit." In *Sixth International Conference on Optical and Photonic Engineering (icOPEN 2018)*, vol. 10827, p. 1082714. International Society for Optics and Photonics, 2018.

Bibliography

- [1] H. Golnabi and A. Asadpour. Design and application of industrial machine vision systems. *Robotics and Computer-Integrated Manufacturing*, 23(6):630 – 637, 2007. ISSN 0736-5845. doi: <https://doi.org/10.1016/j.rcim.2007.02.005>. URL <http://www.sciencedirect.com/science/article/pii/S0736584507000233>. 16th International Conference on Flexible Automation and Intelligent Manufacturing. [1](#)
- [2] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3): 211–252, 2015. doi: [10.1007/s11263-015-0816-y](https://doi.org/10.1007/s11263-015-0816-y). [1](#), [33](#)
- [3] Michael Reed Teague. Deterministic phase retrieval: a Greens function solution. *Journal of the Optical Society of America*, 73(11):1434, November 1983. ISSN 0030-3941. doi: [10.1364/JOSA.73.001434](https://doi.org/10.1364/JOSA.73.001434). URL <https://www.osapublishing.org/abstract.cfm?URI=josa-73-11-1434>. [2](#), [22](#), [23](#), [24](#)
- [4] Martin Weis, Jack Lin, Dai Taguchi, Takaaki Manaka, and Mitsumasa Iwamoto. Analysis of transient currents in organic field effect transistor: The time-of-flight method. *The Journal of Physical Chemistry C*, 113(43):18459–18461, 2009. doi: [10.1021/jp908381b](https://doi.org/10.1021/jp908381b). [2](#)
- [5] Sai Siva Gorthi and Pramod Rastogi. Fringe projection techniques: whither we are? *Optics and lasers in engineering*, 48(IMAC-REVIEW-2009-001):133–140, 2010. [2](#), [9](#)
- [6] Mitsuo Takeda and Kazuhiro Mutoh. Fourier transform profilometry for the automatic measurement of 3-d object shapes. *Applied optics*, 22(24):3977–3982, 1983. [7](#), [8](#), [9](#)

-
- [7] V Srinivasan, Hsin-Chu Liu, and Maurice Halioua. Automated phase-measuring profilometry: a phase mapping approach. *Applied optics*, 24(2): 185–188, 1985. [8](#)
 - [8] Koichi Iwata, Fuminori Kusunoki, Kousuke Moriwaki, Hiroki Fukuda, and Takaharu Tomii. Three-dimensional profiling using the fourier transform method with a hexagonal grating projection. *Applied optics*, 47(12):2103–2108, 2008. [8](#)
 - [9] Joan Batlle, E Mouaddib, and Joaquim Salvi. Recent progress in coded structured light as a technique to solve the correspondence problem: a survey. *Pattern recognition*, 31(7):963–982, 1998. [8](#)
 - [10] Joaquim Salvi, Jordi Pages, and Joan Batlle. Pattern codification strategies in structured light systems. *Pattern recognition*, 37(4):827–849, 2004. [8](#)
 - [11] TG Stahs and FM Wahl. Fast and robust range data acquisition in a low-cost environment. In *Close-Range Photogrammetry Meets Machine Vision*, volume 1395, page 13951R. International Society for Optics and Photonics, 1990. [8](#)
 - [12] Yajun Wang and Song Zhang. Optimal pulse width modulation for sinusoidal fringe generation with projector defocusing. *Optics letters*, 35(24):4121–4123, 2010. [8](#)
 - [13] Xian-Yu Su, Wen-Sen Zhou, G Von Bally, and D Vukicevic. Automated phase-measuring profilometry using defocused projection of a ronchi grating. *Optics Communications*, 94(6):561–573, 1992. [8](#)
 - [14] Jingang Zhong and Jiawen Weng. Dilating gabor transform for the fringe analysis of 3-d shape measurement. *Optical Engineering*, 43(4):895–900, 2004. [8](#)
 - [15] Qian Kemaο. Two-dimensional windowed fourier transform for fringe pattern analysis: principles, applications and implementations. *Optics and Lasers in Engineering*, 45(2):304–317, 2007. [8](#)
 - [16] Ali Dursun, Serhat Özder, and F Necati Ecevit. Continuous wavelet transform analysis of projected fringe patterns. *Measurement Science and Technology*, 15(9):1768, 2004. [8](#)

- [17] Yan Tangy, Wen-jing Chen, Xian-yu Su, and Li-qun Xiang. Neural network applied to reconstruction of complex objects based on fringe projection. *Optics Communications*, 278(2):274–278, 2007. [8](#)
- [18] Richard M Goldstein, Howard A Zebker, and Charles L Werner. Satellite radar interferometry: Two-dimensional phase unwrapping. *Radio science*, 23(4):713–720, 1988. [8](#)
- [19] Hock Lim, Wei Xu, and Xiaojing Huang. Two new practical methods for phase unwrapping. In *Geoscience and Remote Sensing Symposium, 1995. IGARSS'95. 'Quantitative Remote Sensing for Science and Applications', International*, volume 1, pages 196–198. IEEE, 1995. [8](#)
- [20] C Quan, CJ Tay, XY He, X Kang, and HM Shang. Microscopic surface contouring by fringe projection method. *Optics & Laser Technology*, 34(7):547–552, 2002. [9](#)
- [21] Liang-Chia Chen and Chung-Chih Huang. Miniaturized 3d surface profilometer using digital fringe projection. *Measurement Science and Technology*, 16(5):1061, 2005. [9](#)
- [22] Stefano Brusaporci. *Handbook of research on emerging digital tools for architectural surveying, modeling, and representation*. IGI Global, 2015. [9](#)
- [23] J Herriot Bruning, Donald R Herriott, JE Gallagher, DP Rosenfeld, AD White, and DJ Brangaccio. Digital wavefront measuring interferometer for testing optical surfaces and lenses. *Applied optics*, 13(11):2693–2703, 1974. [10](#)
- [24] Hongyu Ren, Feng Gao, and Xiangqian Jiang. Improvement of high-order least-squares integration method for stereo deflectometry. *Applied optics*, 54(34):10249–10255, 2015. [15](#)
- [25] M Minsky. Microscopy apparatus: Us, 3013467. [17](#)
- [26] CJR Sheppard, DK Hamilton, and IJ Cox. Optical microscopy with extended depth of field. *Proc. R. Soc. Lond. A*, 387(1792):171–186, 1983. [17](#)
- [27] Shree K Nayar and Yasuo Nakagawa. Shape from focus. *IEEE Transactions on Pattern analysis and machine intelligence*, 16(8):824–831, 1994. [17](#), [22](#)

- [28] Alex Paul Pentland. A new sense for depth of field. *IEEE transactions on pattern analysis and machine intelligence*, (4):523–531, 1987. [18](#)
- [29] Yu Liu, Xun Chen, Hu Peng, and Zengfu Wang. Multi-focus image fusion with a deep convolutional neural network. *Information Fusion*, 36:191–207, 2017. [18](#), [61](#)
- [30] Muralidhara Subbarao and Tao Choi. Accurate recovery of three-dimensional shape from image focus. *IEEE Transactions on pattern analysis and machine intelligence*, 17(3):266–274, 1995. [18](#), [22](#)
- [31] Mannan Saeed Muhammad, Muhammad Tariq Mahamood, and Tae-Sun Choi. Approximating 3d shape using bezier surface. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 753–756. IEEE, 2009. [18](#)
- [32] François Aguet, Dimitri Van De Ville, and Michael Unser. Model-based 2.5-d deconvolution for extended depth of field in brightfield microscopy. *IEEE Transactions on Image Processing*, 17(7):1144–1153, 2008. [18](#), [20](#)
- [33] David A Agard, Yasushi Hiraoka, Peter Shaw, and John W Sedat. Fluorescence microscopy in three dimensions. In *Methods in cell biology*, volume 30, pages 353–377. Elsevier, 1989. [19](#)
- [34] RJ Pieper and Adrianus Korpel. Image processing for extended depth of field. *Applied Optics*, 22(10):1449–1453, 1983. [20](#), [21](#)
- [35] Aamir Saeed Malik and Tae-Sun Choi. A novel algorithm for estimation of depth map using image focus for 3d shape recovery in the presence of noise. *Pattern Recognition*, 41(7):2200–2225, 2008. [21](#)
- [36] Said Pertuz, Domenec Puig, and Miguel Angel Garcia. Analysis of focus measure operators for shape-from-focus. *Pattern Recognition*, 46(5):1415–1432, 2013. [21](#)
- [37] Tae-Sun Choi and Joungil Yun. Three-dimensional shape recovery from the focused-image surface. *Optical Engineering*, 39(5):1321–1327, 2000. [22](#)
- [38] Aamir Saeed Malik, Humaira Nisar, and Tae-Sun Choi. A fuzzy-neural approach for estimation of depth map using focus. *Applied Soft Computing*, 11(2):1837–1850, 2011. [22](#)

- [39] David Paganin and Keith A Nugent. Noninterferometric phase imaging with partially coherent light. *Physical review letters*, 80(12):2586, 1998. [22](#), [23](#), [25](#)
- [40] Laura Waller, Lei Tian, and George Barbastathis. Transport of intensity phase-amplitude imaging with higher order intensity derivatives. *Optics express*, 18(12):12552–12561, 2010. [22](#), [24](#), [27](#)
- [41] Michael Reed Teague. Irradiance moments: their propagation and use for unique retrieval of phase. *JOSA*, 72(9):1199–1209, 1982. [22](#)
- [42] Francois Roddier. Wavefront sensing and the irradiance transport equation. *Applied Optics*, 29(10):1402–1403, 1990. [23](#)
- [43] Norbert Streibl. Phase imaging by the transport equation of intensity. *Optics communications*, 49(1):6–10, 1984. [23](#), [27](#)
- [44] Timur E Gureyev and Keith A Nugent. Phase retrieval with the transport-of-intensity equation. ii. orthogonal series solution for nonuniform illumination. *JOSA A*, 13(8):1670–1682, 1996. [23](#)
- [45] Timur E Gureyev and Keith A Nugent. Rapid quantitative phase imaging using the transport of intensity equation. *Optics communications*, 133(1-6):339–346, 1997. [23](#)
- [46] VV Volkov, Y Zhu, and M De Graef. A new symmetrized solution for phase retrieval using the transport of intensity equation. *Micron*, 33(5):411–416, 2002. [23](#)
- [47] Chao Zuo, Qian Chen, and Anand Asundi. Boundary-artifact-free phase retrieval with the transport of intensity equation: fast solution with use of discrete cosine transform. *Optics express*, 22(8):9220–9244, 2014. [23](#)
- [48] TE Gureyev, DM Paganin, AW Stevenson, SC Mayo, and SW Wilkins. Generalized eikonal of partially coherent beams and its use in quantitative imaging. *Physical review letters*, 93(6):068103, 2004. [23](#)
- [49] TE Gureyev, A Roberts, and KA Nugent. Partially coherent fields, the transport-of-intensity equation, and phase uniqueness. *JOSA A*, 12(9):1942–1946, 1995. [23](#)

- [50] Marcos Soto and Eva Acosta. Improved phase imaging from intensity measurements in multiple planes. *Applied optics*, 46(33):7978–7981, 2007. [24](#)
- [51] Shiling Zheng, Bindang Xue, Wenfang Xue, Xiangzhi Bai, and Fugen Zhou. Transport of intensity phase imaging from multiple noisy intensities measured in unequally-spaced planes. *Optics express*, 20(2):972–985, 2012. [24](#)
- [52] Rui Bie, Xiu-Hua Yuan, Ming Zhao, and Li Zhang. Method for estimating the axial intensity derivative in the tie with higher order intensity derivatives and noise suppression. *Optics Express*, 20(7):8186–8191, 2012. [24](#)
- [53] Chao Zuo, Qian Chen, Weijuan Qu, and Anand Asundi. High-speed transport-of-intensity phase microscopy with an electrically tunable lens. *Optics express*, 21(20):24060–24075, 2013. [24](#), [27](#), [55](#)
- [54] Chao Zuo, Qian Chen, Weijuan Qu, and Anand Asundi. Noninterferometric single-shot quantitative phase microscopy. *Optics letters*, 38(18):3538–3541, 2013. [24](#), [27](#)
- [55] Percival F Almero, Laura Waller, Mostafa Agour, Claas Falldorf, Giancarlo Pedrini, Wolfgang Osten, and Steen G Hanson. Enhanced deterministic phase retrieval using a partially developed speckle field. *Optics letters*, 37(11):2088–2090, 2012. [27](#)
- [56] Laura Waller, Shan Shan Kou, Colin JR Sheppard, and George Barbastathis. Phase from chromatic aberrations. *Optics express*, 18(22):22817–22825, 2010. [28](#)
- [57] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shape modeling. In *Proc. CVPR*, volume 1, page 4, 2015. [30](#), [34](#), [35](#)
- [58] Shuran Song and Jianxiong Xiao. Sliding shapes for 3d object detection in depth images. In *European conference on computer vision*, pages 634–651. Springer, 2014. [30](#)
- [59] WLK Nguyen, A Aprilia, A Khairyanto, WC Pang, GGL Seet, and SB Tor. Morphological box classification framework for supporting 3d scanner selection. *Virtual and Physical Prototyping*, pages 1–11, 2018. [30](#)

- [60] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86 (11):2278–2324, 1998. [32](#)
- [61] Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. Technical report, Citeseer, 2009. [32](#)
- [62] Li Fei-Fei, Rob Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence*, 28 (4):594–611, 2006. [32](#)
- [63] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. In *NIPS workshop on deep learning and unsupervised feature learning*, volume 2011, page 5, 2011. [32](#)
- [64] Yann LeCun, Fu Jie Huang, and Leon Bottou. Learning methods for generic object recognition with invariance to pose and lighting. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–104. IEEE, 2004. [33](#)
- [65] Sameer A Nene, Shree K Nayar, Hiroshi Murase, et al. Columbia object image library (coil-20). 1996. [33](#)
- [66] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312, 2014. URL <http://arxiv.org/abs/1405.0312>. [33](#)
- [67] Dim P. Papadopoulos, Jasper R. R. Uijlings, Frank Keller, and Vittorio Ferrari. We don’t need no bounding-boxes: Training object class detectors using only human verification. *CoRR*, abs/1602.08405, 2016. URL <http://arxiv.org/abs/1602.08405>. [33](#)
- [68] A Georghiades, P Belhumeur, and D Kriegman. Yale face database. *Center for computational Vision and Control at Yale University*, <http://cvc.yale.edu/projects/yalefaces/yalefa>, 2:6, 1997. [33](#)
- [69] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su,

- et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. [34](#)
- [70] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953, 2015. [34](#)
- [71] Evangelos Kalogerakis, Melinos Averkiou, Subhransu Maji, and Siddhartha Chaudhuri. 3d shape segmentation with projective convolutional networks. In *Proc. CVPR*, volume 1, page 8, 2017. [34](#)
- [72] Daniel Maturana and Sebastian Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 922–928. IEEE, 2015. [35](#)
- [73] Gernot Riegler, Ali Osman Ulusoy, and Andreas Geiger. Octnet: Learning deep 3d representations at high resolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 3, 2017. [35](#)
- [74] Peng-Shuai Wang, Yang Liu, Yu-Xiao Guo, Chun-Yu Sun, and Xin Tong. Ocnn: Octree-based convolutional neural networks for 3d shape analysis. *ACM Transactions on Graphics (TOG)*, 36(4):72, 2017. [35](#)
- [75] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 1(2):4, 2017. [35](#)
- [76] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems*, pages 5099–5108, 2017. [35](#)
- [77] Shubham Tulsiani, Hao Su, Leonidas J Guibas, Alexei A Efros, and Jitendra Malik. Learning shape abstractions by assembling volumetric primitives. In *Proc. CVPR*, volume 2, 2017. [35](#)
- [78] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016. [35](#)

- [79] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017. [35](#)
- [80] Yair Rivenson, Yibo Zhang, Harun Günaydin, Da Teng, and Aydogan Ozcan. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light: Science & Applications*, 7(2):17141, 2018. [35](#)
- [81] YoungJu Jo, Sangjin Park, JaeHwang Jung, Jonghee Yoon, Hosung Joo, Minhyeok Kim, Suk-Jo Kang, Myung Chul Choi, Sang Yup Lee, and YongKeun Park. Holographic deep learning for rapid optical screening of anthrax spores. *Science advances*, 3(8):e1700606, 2017. [35](#)
- [82] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22, 2000. [40](#)
- [83] Paul J Besl and Neil D McKay. Method for registration of 3-d shapes. In *Sensor Fusion IV: Control Paradigms and Data Structures*, volume 1611, pages 586–607. International Society for Optics and Photonics, 1992. [47](#)
- [84] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986. [59](#)
- [85] SHI Xingjian, Zhouong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, pages 802–810, 2015. [61](#)
- [86] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681, 1997. [61](#)
- [87] Shuran Song, Samuel P Lichtenberg, and Jianxiong Xiao. Sun rgb-d: A rgb-d scene understanding benchmark suite. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 567–576, 2015. [62](#), [63](#)
- [88] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgb-d images. In *European Conference on Computer Vision*, pages 746–760. Springer, 2012. [62](#)

- [89] Allison Janoch, Sergey Karayev, Yangqing Jia, Jonathan T Barron, Mario Fritz, Kate Saenko, and Trevor Darrell. A category-level 3d object dataset: Putting the kinect to work. In *Consumer depth cameras for computer vision*, pages 141–165. Springer, 2013. [62](#)
- [90] Jianxiong Xiao, Andrew Owens, and Antonio Torralba. Sun3d: A database of big spaces reconstructed using sfm and object labels. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1625–1632, 2013. [62](#)
- [91] Xiaolong Wang, Abhinav Shrivastava, and Abhinav Gupta. A-fast-rcnn: Hard positive generation via adversary for object detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. [69](#)
- [92] Tete Xiao, Yingcheng Liu, Bolei Zhou, Yuning Jiang, and Jian Sun. Unified perceptual parsing for scene understanding. *arXiv preprint arXiv:1807.10221*, 2018. [69](#)
- [93] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. [70](#)
- [94] Tsung-Yi Lin, Piotr Dollár, Ross B Girshick, Kaiming He, Bharath Hariharan, and Serge J Belongie. Feature pyramid networks for object detection. In *CVPR*, volume 1, page 3, 2017. [70](#)
- [95] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015. [73](#)
- [96] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012. [73](#)