NORTHWESTERN UNIVERSITY

Computational Imaging for Accurate 3D Modeling in Diverse Fields

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Computer Science

By

Bingjie (Jenny) Xu

EVANSTON, ILLINOIS

June 2023

ABSTRACT

This dissertation introduces several novel computational imaging techniques that capture and analyze the 3D surface shapes and internal layered materials. The research proposes user-friendly and non-invasive imaging systems, constructed using only commercial off-the-shelf (COTS) components, which provide accurate measurement of 3D information that was previously inaccessible. The dissertation focuses on two areas of 3D modeling: surface and volumetric modeling. For surface modeling, the research introduces a mobile photometric stereo system called 'shape-from-shifting', and two mobile phase measuring deflectometry systems that capture surface features of large-scale mostly-flat objects and mostly-spherical objects. It showcases their implementation with cultural heritage applications, by detecting protrusions in oil paintings, determining the provenance of stained glass pieces, and by tracking eye movements for virtual reality (VR) headsets. In comparison to previous studies that mostly relied on bulky and lab-specific systems, our imaging systems use a mobile device screen or a wall, and commercial camera(s) or a projector, accompanied by open-source Python packages and tutorials, enabling flexible usage and accessibility to non-experts. For volumetric modeling, I present a novel time-domain optical coherence tomography (OCT) system that uses off-the-shelf optical components and a dynamic-focus approach to compensate for low-energy problems in commercial low-cost continuum laser sources. This system can penetrate multi-layer oil paintings with IR wavelengths that reach up to 2 μm, making it the first low-cost time-domain OCT system at 2 μm for cultural heritage applications. To analyze the chemical composition of multi-layer oil paintings, the research proposes a deep-learning framework that uses X-Ray Fluorescence (XRF) spectra, enabling automatic identification of mixed pigments for multi-layer oil paintings, which is the first study to use deep learning for this purpose.

ACKNOWLEDGEMENTS

I am honored and grateful to have the opportunity to express my heartfelt appreciation to those who have contributed to the completion of my Ph.D. dissertation.

First and foremost, I extend my deepest gratitude to my supervisor, Prof. Jack Tumblin, whose unwavering guidance, support, and encouragement have been invaluable throughout my research journey. His insightful feedback and critical insights have been instrumental in shaping my research and improving the quality of this dissertation.

I also want to express my gratitude to the members of my dissertation committee, Prof. Marc Walton, Prof. Florian Willomitzer, Prof. Aggelos Katsaggelos, and Prof. Oliver Cossairt, for their time, effort, and valuable feedback along my entire research journey. Their expertise and critical insights have helped me to refine my research and broaden my perspective. I am also grateful to Prof. Matthias Alfeld for his support and guidance on the pigment identification project.

I would like to acknowledge the contributions of my colleagues from Northwestern, Meta Reality Labs, and Adobe Research. Special thanks to Chia-Kai (Kai) Yeh for the protrusion detection project; Yunhao Li and Xun Xia for the stained glass provenance project; Jiazhang Wang, Tianfu Wang, and Nathan Matsuda for the VR deflectometry project; Alexander Fix, Hossein Daraei, Chris Aholt, and Yatong An for the glint-only eye tracking project; Pengxiao Hao, Kuan He, and Jian Gao for the time-domain OCT project; Yunan Wu, Pengxiao Hao, Marc Vermeulen, and Alicia McGeachy for the pigment identification project; Jiwon Choi, Thomas Bochynek, Alexis Baudron, Florian Schiffers, Merlin Nau, Eunyee Koh, and Shunan Guo for all other side projects I have worked on. I am deeply grateful for their support, intelligence, and feedback, which have been invaluable. I am also grateful to everyone from the Comphoto Lab, the Center for Scientific Studies in the Art at Northwestern, Eye Tracking Team at Meta Reality Labs, AI Experience Team

at Adobe Research, for their collaboration and support.

I would like to acknowledge the camaraderie, encouragement, and moral support of my friends Pengxiao Hao, Haoru Zhang, Hongyu Zheng, Dawei Xue, Yingyi Luo, Jingyang Zhang, Jiamin Zhang, and Luyu Ren. I am grateful for their friendship, which has made the journey more enjoyable.

My family and my boyfriend Hongyuan Yang deserve special recognition for their unwavering love and support. Their encouragement and sacrifices have made this achievement possible. I am forever grateful to them.

Finally, I would like to express my gratitude to the funding agencies National Science Foundation (NSF) for the Partnerships for International Research and Education (PIRE) grant and the Andrew W. Mellon Foundation for providing me with the financial support that enabled me to pursue this research. Without their support, this achievement would not have been possible.

To all those who have supported and encouraged me throughout my research journey, I extend my sincerest thanks and appreciation. Your contributions have been invaluable, and I am deeply grateful.

TABLE OF CONTENTS

Acknowledgments 3

List of Figures 9

List of Tables 14

Chapter 1: Introduction 15

Chapter 2: Background 20

2.1 3D Surface Modeling 21

2.1.1 Photometric Stereo 23

2.1.2 Phase Measuring Deflectometry 24

2.2 Volumetric Modeling 27

2.2.1 Optical Coherence Tomography 30

Chapter 3: 3D Surface Modeling for Works of Art 32

3.1 Protrusion Detection in Georgia O'Keeffe's painting 35

3.1.1 Shape from Shifting 35

3.1.2 Protrusion Detection 36
3.1.3 Discussion 38
3.2 Provenance of Stained Glass Pieces in Tiffany Artworks 39
3.2.1 Large-Scale Deflectometry for Flat Specular Objects 41
3.2.2 Discussion 45
3.3 Conclusion 47

Chapter 4: 3D Surface Modeling for Eye Tracking 48
4.1 Eyeball Surface Modeling using Deflectometry 49
4.1.1 Stereo Deflectometry 50
4.1.2 Camera-Display System Calibration 50
4.1.3 Single-shot Stereo Deflectometry 54
4.2 Eye Tracking with High Fidelity Eye Models 56
4.2.1 Multi-view Glint-only Eye Tracking 58
4.2.2 Experiments and Discussion 64

Chapter 5: Volumetric Modeling for Oil Paintings 68
5.1 Time-domain Optical Coherence Tomography (OCT) for Paintings 69
5.1.1 Dynamic focus time domain OCT 71
5.1.2 Experiments and Results 74
5.1.3 Discussion 76

5.2 Automatic Identification of Layered Pigments from XRF Data 78

5.2.1 Dataset and Preparation 80

5.2.2 Deep Learning Model for Pigment Identification 87

5.2.3 Results and Model Performance 90

5.2.4 Discussion and Conclusion 95

Chapter 6: Conclusion and Future Work 106

References 124

LIST OF FIGURES

2.1 **Problem of specular surface reconstruction for common active methods.** 25

2.2 **Typical Deflectometry System.** 26

2.3 **Ambiguity problem of deflectometry with one camera.** 26

2.4 **Illustration of optical absorption and scattering with particles inside the object.** Figure Courtesy of Regnima et.al [51] with modification. 28

2.5 **Volumetric Imaging Systems.** Figure courtesy of Drexler et al. [52] 29

2.6 **Schematic Diagram of time domain OCT.** 31

3.1 **Hidden information in surface measurement of a color woodblock print** (Katsushika Hokusai (1760 – 1849) "Interior of a Brothel in Yoshiwara" 1811), Private Collection 32

3.2 **Shape from Shifting.** 36

3.3 **Protrusions in Georgia O'Keeffe's painting.** (a) The color image of Georgia O'Keeffe's painting. (b) Measured surface normal map by PS. (c) Enlarged result of the surface normal map. (d) Schematic model of a protrusion: a half sphere of about 20 μm to 500 μm. 37

3.4 **Protrusion Detection.** (a) An example area of the surface normal map. (b) The detection result of this area. The detected protrusions are marked in red. 38

3.5 **Glass features of the Tiffany Stained-Glass.** (a) The color image of the Tiffany Stained-Glass. (b) Measured surface normal map of one glass feature. 40

3.6 **Deflectometry System Setup.** (a) Basic concept of deflectometry. (b) Our system setup contains one camera, one project, the white wall as the screen, and the object. 42

3.7 **Geometric calibration for the display-camera system using a planar mirror.** . . 44

3.8 **Spherical Mirror Evaluation Results.** (a) Measured normal map of the mirror (b) Angular errors between normal maps. 45

3.9 **Normal maps of four example features in the database of Kokomo glass rollers.** 45

3.10 **Cause of normal difference on the same object at different positions.** 46

4.1 **Stereo Deflectometry Setup.** 51

4.2 **Deflectometry-based VR eye-tracking.** (a) Basic concept of deflectometry. (b) Current prototype setup using Stereo Deflectometry. (c) An example of captured images 52

4.3 **Geometric camera(s)-display-system calibration.** (a) The position and tilt of the phone screen are found by imaging a displayed checkerboard pattern, which acts as a "marker". (b) Eventually, the position and tilt of the display are found by observing a displayed checkerboard pattern reflected over the phone screen. The phone screen is switched off and acts as a "mirror". (c) Reprojection error: Three dimensional checker positions of the calculated (virtual) display are traced back to the camera chip (blue crosses). Comparison with detected checker positions in the real image (red circles) reveals a reprojection RMSE of 0.42 camera pixels 54

4.4 **Eye tracking using single-shot stereo Deflectometry.** (a) Prototype setup. (b,c) Captured camera images of realistic eye model and real human eye, overlaid with the calculated surface normals and estimated gaze direction. 55

4.5 **Multi-view Glint-only Eye Tracking Framework.** 59

4.6 **Raw images from the capture.** Multiview camera images from 9 views. 59

4.7 **The ground truth specular mask obtained from the differential lighting sequence.** 62

4.8 **Labeled specularities superimposed on the input image.** Each light is assigned a different color for visualization. 62

4.9 **Pose Refinement with Glint Centroids.** 63

4.10 **Comparison of rendered specularity against ground truth original image and specular mask.** 66

5.1 **Dynamic focus OCT.** (a) Schematic Diagram of light paths in conventional and dynamic focus OCT. The red spot and dashed line represent coherence gate. (b) Conventional and dynamic focus OCT apparatus. The blue region depicts conventional OCT, with reference mirror moved by the translation stage; the red region depicts dynamic focus OCT with fixed reference mirror and sample arm moved by the translation stage with galvanometer mirror and scanning lens attached. 72

5.2 **Zoom-in interference pattern at the air/glass interface.** The FWHM is 4.85 μm. The axial resolution of our system achieves a high resolution of $\sim$4.85 μm at the air/glass boundary. 75

5.3 **Scanning results of a layer of indigo-linseed-oil on a glass slide** by (a) lab-built 2 μm TD-OCT and (b) commercial FD-OCT with a central wavelength of 900 nm. (c) The physical thickness of the paint film measured by a profilometer. 76

5.4 **The structure of mock-up samples.** The first and second layers both contain multiple strips of pigment layers, each strip with its pigment combination selected from Table 5.1; the third layer is a ground layer consisting of calcium carbonate. . . 82

5.5 **The Architecture of the Deep Learning Model.** It contained 1D convolutional layers (Conv), normalization layers (Norm), max pooling layers (Max Pool), a flatten layer, a dropout layer and a fully-connected layer. The input size of each XRF spectrum was 3815$\times$1. The output predictions were in 11 classes (one per pigment). # kernel and kernel size stood for the number and the size of the kernel at the corresponding convolutional layer, respectively. 87

5.6 ***Poèmes Barbares* (1896)**, oil on canvas, 64.8 x 48.3 cm (unframed), painted by the French artist Paul Gauguin (1848–1903), Harvard Art Museums/Fogg Museum, Bequest from the Collection of Maurice Wertheim, Class of 1906. Object Number: 1951.49 ©President and Fellows of Harvard College. 92

5.7 ***The Bathers* (1899-1904)**, oil on canvas, 51.3 × 61.7 cm, painted by the French artist Paul Cezanne (1839–1906), The Art Institute of Chicago, Amy McCormick Memorial Collection. Object Number: 1942.457 ©The Art Institute of Chicago. . . 94

5.8 **Pigment identification results of *Poèmes Barbares*** from the model without finetuning (left half), and the results from the finetuned model (right half), with results from one pigment on each row. The first, second and third columns in each row display the pigment map, element map(s) and the comparison figure and scatter plots, respectively. 96

5.9 **Pigment identification results of *The Bathers*** from the finetuned model, including cobalt blue, vermilion, and emerald green from top to bottom. The first, second and third columns display the pigment map, element map(s) and the comparison figure and scatter plots, respectively. 97

5.10 **Photo image of one mock-up** that contains Top 2, Top 1, Top 3, Top 4 (from top to bottom) as top layers and Bottom 2 with three different mass fractions (A, B, C), Bottom 1 with three different mass fractions (A, C, B) (from left to right) as bottom layers. 98

5.11 **Comparison between pigment maps and elemental maps for the mock-up.** (a) The ground truth of the location of vermilion (VM). (b) The Hg-L element map directly generated from PyMCA. (c) The VM pigment map generated from the model without finetuning identifies Hg in the 1, 3, and 4 rows and the 3 and 4 columns. It can barely detect Hg in the 5 and 6 columns. (d) The VM pigment map generated from the finetuned model indicates VM presence in the 1, 3, and 4 rows and 1, 3, 4, 5, and 6 columns with high probabilities. The only missing column of the VM pigment map is the second column, which contains 0.6% of vermilion. . . 99

5.12 **Our model distinguishes Prussian blue (PB) and iron oxide (IO) in one mock-up.** (a) The ground truth of PB and IO locations. (b) The Fe element map directly generated from PyMCA contains both PB and IO. (c) The PB pigment map generated from our model. (d) The IO pigment map generated from our model. 102

5.13 **XRF spectra of the mock-up painting comparing the effect of highly absorbing pigments.** (a) One single layer M that contains lead white (LW), red lead (RL), vermilion (VM) and zinc white (ZW). The lead peaks marked in red have high intensities. (b) The single layer M covered by one layer of vermilion (VM). The existence of VM in the top layer significantly blocks the XRF signal of the lead-containing pigments at the bottom, challenging our model's ability to detect the hidden lead element. 103

LIST OF TABLES

4.1 Evaluation of estimated gaze direction for single-shot Stereo-Deflectometry method and optimization-based method. 56

5.1 Pigment library for selected pigments. 84

5.2 Summary of pigment mixtures in the mock-ups. 85

5.3 The classification results for each pigment class among the models trained from three different datasets: the simulation dataset, the experimental dataset without finetuning, and the experimental dataset acquired from the mock-ups with finetuning[a]. 90

CHAPTER 1
INTRODUCTION

All knowledge of nature originates from observation. In fact, phenomena such as the reflection and refraction of light, and the ability of certain shaped crystals to produce magnified images, have been recorded and observed in different parts of the world since ancient times [1]. In 300 BC, the ancient Greek mathematician Euclid correctly described in his book "Optics" the law that the angle of reflection of light on a surface is equal to the angle of incidence. He also observed the phenomenon of light refraction, but at that time a reasonable explanation was not provided. The Greek philosopher Plato attributed this to people's imperfect perception of reality.

In ancient China, the earliest recorded optical research can be found in the Mo Jing(墨经) [2], which was compiled around 388 BC and is a part of the book of Mozi(墨子), a well-known name for every Chinese for his Mohist school of thought. Mo Jing contains an experiment on pinhole imaging: "The scene, being illuminated by a person, is vivid as if it were being shot. The person below is high, while the person above is low. The foot blocks the light below, so the scene is formed above, and the head blocks the light above, so the scene is formed below..." It points out that the fundamental reason for the inverted image produced by a pinhole is the straight propagation of light. Mozi and his students record eight laws of optical experience and their annotations, respectively. Among them, the eighth law states: "When a mirror reflects a circle, the image appears to be the same size as the circle if the mirror is nearby, and the image appears to be smaller than the circle if the mirror is far away. However, the image will always be upright, and the reflection will be distorted if the mirror is not positioned correctly." It describes the laws of imaging by a convex mirror, explaining that the image will always be upright, and that its size is

related to the position of the object.

Over the past two thousand years, the field of optical imaging has undergone constant development and evolution since the time of Mozi. As a member of Generation Z, I've had the pleasure of experiencing the rise of digital cameras and the incredible evolution of smartphones, both of which owe their success in part to the advancements made in imaging technology. In recent years, the design of imaging systems has shifted towards a new concept known as computational imaging. This approach balances the processing capabilities of optics and electronics by optimizing the imaging system through the comprehensive consideration of all technologies involved, including the formation of optical images, image detection, and image data processing. With the help of computational imaging, the potential for imaging technology has expanded to previously unimaginable heights, where electronics and signal processing techniques have pushed the boundaries of what was previously thought possible.

My interest in the field of computational imaging was sparked during a lab visit in my second year of undergraduate studies in 2016. During the visit, we witnessed the instant 3D modeling of our faces and bodies by standing in front of a camera while blue stripes were projected onto us. This technique is known as fringe projection for 3D imaging, and the experience left a lasting impression on me. The ability to obtain a 3D model of oneself from 2D camera images felt like something out of a sci-fi movie. It was then that I joined the computational imaging lab during undergrad and began working on Fourier-based single-pixel imaging [3], [4], which was inspired by ghost imaging and represents the very beginnings of computational imaging. I was so captivated by computational imaging that I decided to pursue a Ph.D. degree in this area at Northwestern University after completing my undergraduate studies. Throughout my research journey, I've been particularly fascinated by the measurement of 3D information and the potential for 3D imaging to reveal both surface and subsurface information about objects, which has continued to captivate and

pique my curiosity.

My investigations have focused on both 3D surface modeling and volumetric modeling techniques. As I pursued my Ph.D. studies at Northwestern, I had the opportunity to explore various applications of these techniques in both cultural heritage and virtual reality contexts. Through this exploration, I gained a deeper understanding of the underlying principles and specific use cases for these applications, which have helped to inform the work presented in this thesis.

Accurate 3D modeling of objects is an essential aspect of many fields, including medicine, engineering, archaeology, and art. It can provide important information about the geometry, shape, material, and size of objects, and it can be useful for visualization, simulation, and analysis purposes. 3D modeling can be achieved through surface modeling, where the outer surface of an object is captured, or volumetric modeling, where the interior of an object is reconstructed based on multiple imaging modalities. Computational imaging has emerged as a solution that allows for more precise, efficient, and cost-effective 3D modeling. By utilizing advanced algorithms and mathematical techniques, novel optical components, and optical arrangements, computational imaging can create 3D models that are not possible using traditional imaging techniques.

In recent years, 3D modeling has become increasingly important and widespread in various fields of daily life, such as medicine and video creation. In the medical field, 3D modeling is widely used to acquire surface and subsurface information of the human body, enabling better diagnosis and treatment. Techniques such as X-ray and computed tomography scans utilize 3D modeling with computational imaging [5], [6]. Additionally, 3D modeling techniques have also been widely adopted in the entertainment industry, with 3D models of human actors' faces and bodies being created using computational imaging and computer graphics techniques for visual effects in movies, such as the light stages created by Debevac et al. [7], [8]. Moreover, the popularity of mobile applications, that can generate 3D models of individuals and daily objects using mobile

devices in real-time, further demonstrates the impact and versatility of 3D modeling [9], [10].

The process of generating 3D models of objects is highly dependent on the properties of the object itself, such as surface roughness, reflectivity, and light penetration into the material. Therefore, there is a need for novel techniques that can be applied to a wide range of applications for 3D modeling. This is particularly relevant in areas such as cultural heritage and virtual reality, where accurate 3D modeling can enhance preservation and immersive experiences. The aim of this thesis is to develop new computational imaging techniques that can achieve highly accurate 3D modeling across a range of fields and further analysis based on the measurements, with a specific focus on cultural heritage and virtual reality. By exploring and developing these techniques, this research will contribute to the broader field of computational imaging and advance our ability to accurately capture and represent 3D information.

The upcoming chapters will delve into the underlying techniques for 3D surface and volumetric imaging. I will first introduce two user-friendly and non-invasive surface modeling systems that work for different artworks. These systems can help monitor the protrusion and growth of metallic soap blobs in oil paintings and identify the origin of each piece of colorful stained glass on church windows. Compared with previous techniques, these systems are low-cost, easy to use and set up. I will also propose two novel high-accuracy eye tracking techniques that utilize high-fidelity surface models for the first time to further improve the state-of-the-art eye tracking performance. Finally, I will employ two different technologies to investigate the physical structure and material compositions of layered oil paintings, including developing the first low-cost scanning system for paint layers and using machine learning to automatically predict mixed pigments.

The structure of the following chapters in this thesis is as follows:

Chapter 2 will provide an overview of 3D modeling, including 3D surface modeling and volumetric modeling. I will introduce several common computational imaging methods for both surface

modeling and volumetric modeling, with their underlying principles, key techniques, and examples of suitable applications. This chapter will also include a detailed summary of previous works on photometric stereo, deflectometry, and optical coherence tomography (OCT) techniques, which form the basis of the computational imaging techniques used in this thesis.

Chapter 3 will explore how cultural heritage benefits from 3D surface modeling. To illustrate the application of computational imaging techniques in this field, two case studies will be presented. The first case study investigates the use of photometric stereo for detecting protrusions in oil paintings. The second case study describes the use of deflectometry to create "fingerprint" and distinguish the glass textures of stained glass windows.

Chapter 4 will focus on the use of 3D surface modeling in eye tracking for virtual reality headsets. Two different techniques will demonstrate the potential applications of computational imaging in this field for 3D modeling. The first technique uses deflectometry to reconstruct the surface of the eyeball for eye tracking. The second technique employs the recovered 3D shape of the eyeball surface to re-pose and re-render with changing eye pose, creating a reference eye model for eye tracking.

Chapter 4 will focus on volumetric modeling for oil painting, including its internal physical structure and material composition. The first project proposes a novel non-invasive optical coherence tomography system to see through paint layers. The second project introduces an automatic pigment identification framework to analyze the X-ray Fluorescence spectra of layered oil paintings and directly predict the existence of pigments for the entire painting.

Finally, the conclusion will discuss the significance and contributions of this thesis to 3D modeling in the field of computational imaging. This discussion will include potential future directions and applications of the proposed techniques.

CHAPTER 2
BACKGROUND

Computational imaging is a relatively new approach to imaging in recent years and has garnered significant attention since the 1990s because it generalizes imaging – it replaces fixed optics, fixed sensors and fixed lighting with dynamic, programmable adaptive systems that cooperate to extract useful scene information. Traditional imaging tended to optimize each portion independently: lens systems for the sharpest, brightest, best-quality image on a 2-D plane; sensors for the highest sensitivity, speed and resolution at the lowest noise; lighting systems for the best color, visibility and contrast of scene contents. Computational imaging methods offer more choices: imaging systems may actively control the light illuminating the scene, may choose any useful optics to transfer light from the scene to the sensor, and may devise any useful configuration of sensor systems.

The sensor may collect any information-carrying light that can be computationally decoded, such as the early "coded aperture" methods devised for non-refractive x-ray and gamma-ray telescopes that replaced lenses with an invertible binary mask (well introduced in this instructional site by Paul Carlisle [11]). Computational imaging systems drew perhaps its first widespread attention in the pioneering work of Cathey and Dowski [12] whose purposeful lens aberrations enabled later deconvolutions to computationally increase depth-of-field in the captured images. Joseph Mait et al. in their publication titled "Evolutionary paths in imaging and recent trends" [13] later made the first use of the term "computational imaging" that I could find in its current usage; as an integrated computational imaging system that incorporates optics, optoelectronics, and signal processing.

Computational imaging techniques generally combine conventional optical imaging with opti-

cal coding and/or digital decoding, adding extra computing steps between the scene and the sensor that assesses it. By taking advantage of more sophisticated optics components and/or controlled lighting that assists in scene measurement tasks, computational imaging enables the extraction of more and better information from the scene than traditional methods, where a sensor simply copies the sharp 2D image formed on its surface by a lens. Moreover, signal processing techniques applied to the sensed signal, possibly extended by machine learning methods, can extract or infer more scene information indirectly, such as materials and 3D shapes [14].

The field of computational imaging has opened up new possibilities for imaging. One of the areas where computational imaging has found extensive use, and where this thesis aims its focus, is 3D capture and modeling. In this thesis, 3D modeling refers to the sensing and reconstruction of the three-dimensional surface shape and internal volume of an object. In the following sections, we will discuss 3D surface modeling and volumetric modeling separately to provide a more detailed understanding of these techniques.

2.1 3D Surface Modeling

The research area of 3D surface modeling seeks better ways to sense and reconstruct the shape and appearance of physical objects from available sensor data. Recently, 3D surface modeling has emerged as an indispensable tool for a wide range of applications, including industrial manufacturing [15] and medical imaging [16], [17].

The surface of an object is its outermost layer that we see, touch, or photograph. The classification of surfaces based on their light reflection type is a crucial aspect when discussing 3D surface modeling. Surface reflectance can be categorized into three broad types: diffuse, specular, and mixed. Diffuse surfaces appear matte and scatter reflected light uniformly in all directions, while mirror-like specular surfaces are polished and reflect strong light in a narrow range of directions.

Mixed surfaces exhibit a combination of both diffuse and specular reflections. Accurately understanding how light interacts with an object's surface is critical for designing imaging systems that can capture the surface's properties with precision.

Häusler et al. [18], [19] classify 3D surface modeling methods into three primary categories based on the physical origin of the fundamental distance uncertainty: triangulation, interferometry, and slope-measuring methods. Triangulation is the process of determining the location of a point by measuring the angles to it from two known points rather than measuring distances directly. A diverse range of 3D sensors can be classified as triangulation-based methods, including laser triangulation [20], stereo-photogrammetry [21], [22], fringe projection triangulation [23], and structured illumination microscopy [24].

Interferometry involves the interferometric measurement of the phase [25] or time-of-flight [26], [27] of a coherent or partially coherent signal. Interferometry systems utilize the interference of light waves to measure the distance between a camera and a subject. This method typically splits a laser beam into two paths, and two beams of light combine at the sensor, where the interference pattern created by their superposition encodes the lengths of the unknown paths.

Slope-measuring methods refer to techniques that intrinsically measure the surface slope, lateral derivative or perpendicular "surface normal" vector at all surface points. This category includes both incoherent and coherent methods. Coherent methods utilize shearing interferometry [28] and its modifications, such as differential interference contrast. On the other hand, incoherent methods, such as photometric stereo [29], [30] and phase measuring deflectometry, utilize non-coherent light sources.

2.1.1 Photometric Stereo

Photometric stereo is a widely used method for estimating diffuse shape of diffuse surfaces from image intensity. It enables the computation of surface normals at each pixel in images captured under known and controlled lighting conditions. A surface normal is a unit vector that is perpendicular to the surface at a given point. In the latter part of this thesis, we will refer to the surface normal map, which is a map of the surface normal vector for each pixel in a camera image. This surface normal map is typically presented as an RGB image, with the red, green, and blue color channel values representing the x, y, and z directions of the surface normal, respectively.

The original formulation of photometric stereo, presented by Horn [29], assumed that the lights are infinitely far away, the camera is orthographic, and the object surface is Lambertian (entirely diffuse) and convex, which implies no inter-reflections and no shadows. Photometric stereo spans both passive [31] and active methods [30]. Despite its ease of setup and less stringent illumination and calibration requirements, passive photometric stereo has certain limitations, including higher noise levels and the inability to extract features from textureless regions [32]. In contrast, active photometric stereo has the advantage of precise information on light source positions and ability to control the lighting conditions, thus improving the accuracy of the reconstruction for any diffuse surface, even for dark-colored or strongly self-shadowing shapes. As a result, it is preferred for applications that demand high accuracy, robustness, and reliability. Therefore, this thesis will focus on active photometric stereo since they can overcome these limitations and provide better results for cultural heritage applications discussed in Chapter 3.

To reconstruct the object's surface using active photometric stereo, the camera's image of the surface captures the radiance equal to the diffuse reflectance (the albedo) times the projected pattern's irradiance times the dot-product of the surface normal and the unit vector from the surface point to the projector's aperture. As all are known except the surface normal, we solve for normals

as a function of camera pixel values. The main equation used in active Photometric Stereo is given by:

$$I(x,y) = \rho(x,y)\mathbf{l}(x,y) \cdot \mathbf{n}(x,y) + b(x,y) \tag{2.1}$$

where $I(x,y)$ is the intensity of the image at pixel location (x,y), $\rho(x,y)$ is the surface reflectance, $\mathbf{l}(x,y)$ is the known lighting direction, $\mathbf{n}(x,y)$ is the surface normal, and $b(x,y)$ is an additive constant representing ambient lighting or other sources of illumination. By using a known sequence of lighting directions, it is possible to solve for the surface normals at each pixel location, which can be used to create a 3D reconstruction of the object's surface.

Since its introduction, several researchers have extended the PS technique to more practical camera, surface, and lighting models. Belhumeur et al. [33] found that, with an orthographic camera model and uncalibrated lighting, the object's surface could be uniquely determined up to a bas-relief ambiguity. Recently, Papadhimitri and Favaro et al. [34] showed that even this ambiguity can be resolved under the perspective camera model. Researchers have also attempted to remove the assumption of Lambertian reflectance and include effects such as specular highlights and shadows. New techniques based on non-Lambertian reflectance models [35]–[37] or sophisticated statistical methods for automatically filtering non-Lambertian materials and effects [38], [39] have been explored with some success.

2.1.2 Phase Measuring Deflectometry

Phase measuring deflectometry is a well-established technique in surface metrology that reconstructs the 3D surface of specular objects, such as freeform lenses, car windshields, or other shiny mechanical parts [40]–[42]. (Note that, in the rest of this thesis, "deflectometry" only refers to phase measuring deflectometry.) The limited efficacy of common 3D surface modeling techniques

on specular surfaces occurs because specular surfaces reflect incoming light beams in a single direction following the law of specular reflection and may have little or no diffuse reflectance to redirect light in other directions. This limitation causes incomplete data acquisition when the light source is limited to a few points, as shown in Figure 2.1. To address this issue effectively, increasing the area and angular range of the light source is necessary, which is the principle of deflectometry.

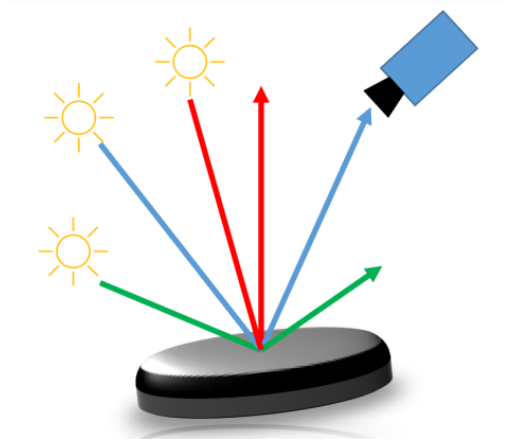

Figure 2.1: **Problem of specular surface reconstruction for common active methods.**

A typical deflectometry setup consists of three main components: a screen, a camera, and the shiny or "specular" object under test as shown in Figure 2.2. The technique relies on observing the reflection of a screen displaying a known pattern (such as a sinusoid) after it reflects from the surface of the specular object. By analyzing the deformation of the pattern in the camera image, the normal vectors of the surface can be calculated, and converted to 3D surface shape.

To calculate the slope of the test surface accurately, it is crucial to determine which screen pixel reflected from the test surface corresponds to which camera pixel. Phase measuring deflectometry uses phase information calculated from deformed patterns to retrieve this mapping, and therefore gets the correspondence points (the screen and camera pixel). However, as depicted in Figure 2.3,

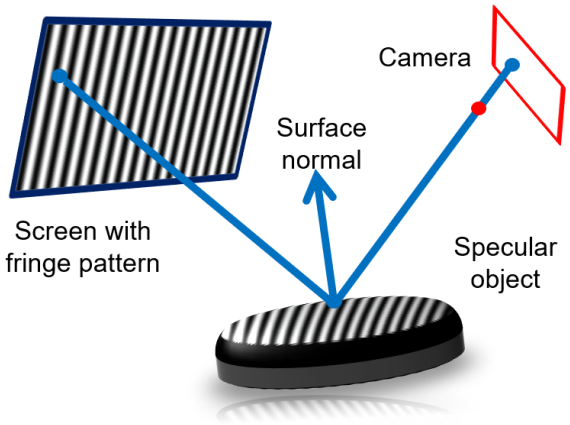

Figure 2.2: **Typical Deflectometry System.**

the correspondence points between the camera and screen, which are calculated using the deformed patterns, do not provide enough information to determine a specific surface location, and therefore, can lead to different surface normals. This phenomenon is known as the "depth-normal-ambiguity" in deflectometry. To address this problem, we will introduce two different solutions in Section 3.2.1 and Section 4.1.1.

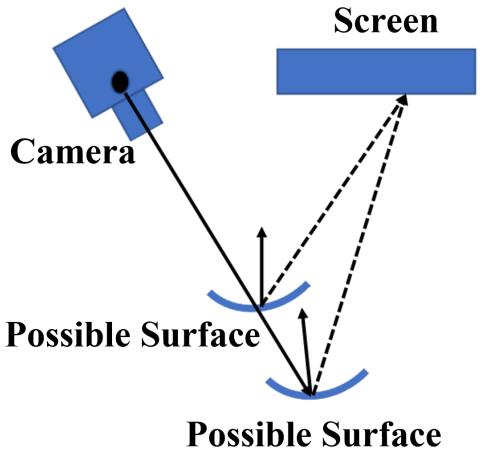

Figure 2.3: **Ambiguity problem of deflectometry with one camera.**

As slope measuring methods, photometric stero and deflectometry both directly determine a

normal map to describe the object surface. Although in this thesis, we mainly use surface map only for further analysis, there are several approaches [43], [44] that enforce the integrability constraint over the whole normal map to estimate the depths to reconstruct the actual 3D surface shape from the normals.

In this thesis, I will focus on these two non-coherent slope-measuring methods, photometric stereo and deflectometry. These methods were chosen due to their low cost and ease of setup, which will be further explained in Section 2.1.1 and 2.1.2, respectively. Chapter 3 will examine the relationship between surface types and surface modeling techniques. Additionally, I will develop different 3D surface imaging systems for matte oil paintings and specular stained glass. A comprehensive analysis of their surface shape will also be provided. In Chapter 4, I will discuss 3D surface modeling for eye tracking in virtual reality. Specifically, I will focus on surface shape modeling of the transparent cornea and sclera of the human eyeball using deflectometry. I will also explore the further application of surface shape for eye tracking purposes.

2.2 Volumetric Modeling

Volumetric modeling allows for the visualization and analysis of internal structures of objects in three dimensions, in contrast to 3D surface modeling. This capability is crucial for various applications, including medical imaging [45], conservation science [46], and virtual reality [47], as it enables the creation of detailed, high-resolution models of complex internal structures. This thesis will focus solely on non-invasive imaging modalities.

The quality of volumetric models depends heavily on the optical properties of the material being investigated. Scattering and absorption are two fundamental properties that influence the propagation of light through materials, as illustrated in Fig. 2.4, and consequently affect the penetration depth in volumetric modeling. Scattering occurs when light interacts with particles or

microstructures in a material, causing the light to be redirected randomly and diffused in all directions. Conversely, absorption is defined as the wavelength-selective removal of light energy as light propagates through the material. These properties play a crucial role in determining the optical behavior of materials and can significantly affect the performance of volumetric modeling techniques [48]–[50].

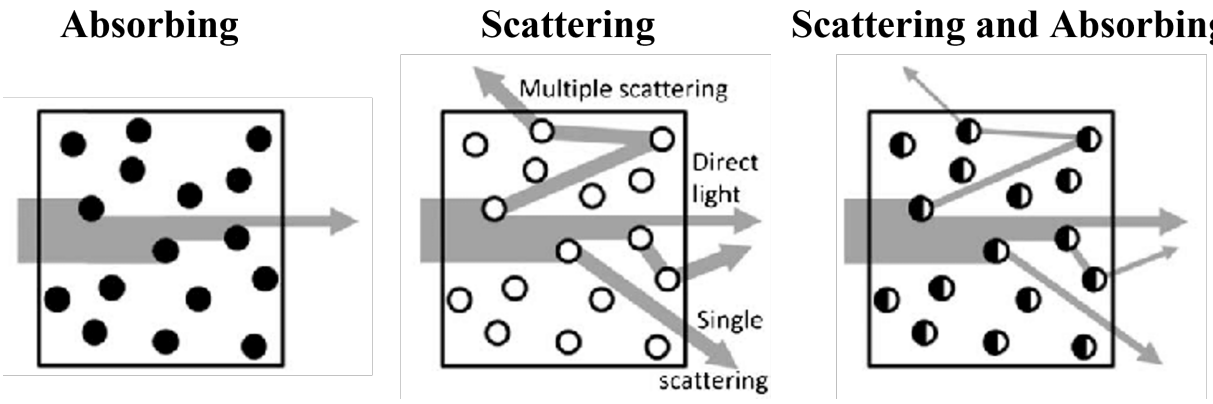

Figure 2.4: **Illustration of optical absorption and scattering with particles inside the object.** Figure Courtesy of Regnima et.al [51] with modification.

Currently, three primary types of volumetric imaging systems predominate: ultrasound, confocal microscopy, and optical coherence tomography(OCT). Figure 2.5 provides a summary of the typical penetration depth and resolution of each volumetric imaging system technology used for biomedical applications [52], highlighting the distinctions among these techniques.

Ultrasound [53] uses sound waves to penetrate materials and generate images. The sound waves diffract and scatter internal structures, and the echoes are detected and used to reconstruct a 3D model. Ultrasound is particularly useful for imaging high optical scattering objects that are difficult to image using traditional optical techniques due to its high penetration. However, the long wavelengths of ultrasound limit its maximum achievable resolution to feature sizes far lower than the two other imaging techniques [52].

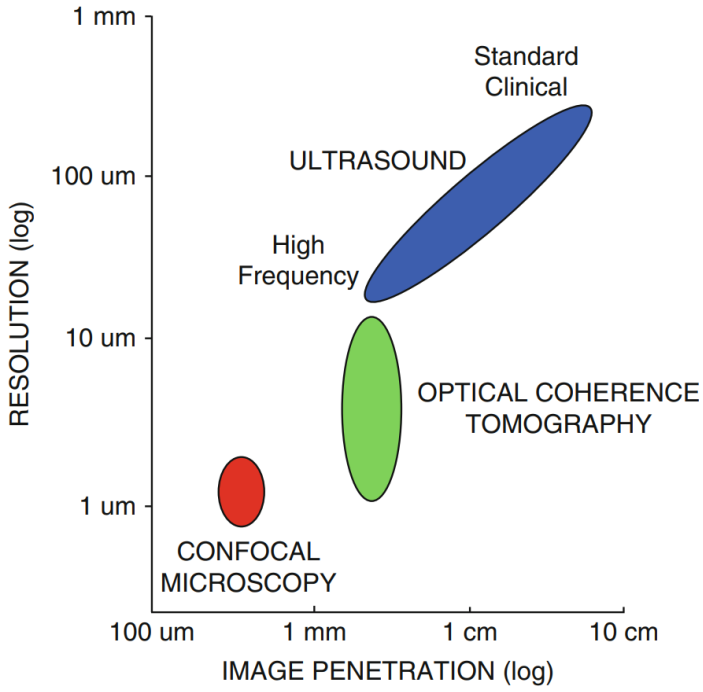

Figure 2.5: **Volumetric Imaging Systems.** Figure courtesy of Drexler et al. [52]

Confocal microscopy [54] uses point illumination via a spatial pinhole to eliminate out-of-focus signals. The excitation light in confocal microscopy is usually provided by a laser, which generates high intensities of fluorescence or reflectance from the focal spot. Confocal microscopy can provide high optical resolution and contrast of an image of a specimen, but is limited by the penetration depth of imaging within thick samples [55].

Optical coherence tomography (OCT) [52] uses self-interference of laser-generated light waves to create cross-sectional images. The light waves are reflected back from different layers of the object, and the interference patterns are detected and used to generate a 3D model. Compared with ultrasound and confocal microscopy, OCT can provide high-resolution images and typically has a penetration depth up to 1 cm, as shown in Figure 2.5.

2.2.1 Optical Coherence Tomography

OCT systems rely on low-coherence or "broadband" or "white-light" interferometry to precisely compare the lengths of two separate optical paths, as shown in Fig. 2.6. Similar to a Michelson interferometer, the "beam-splitter" sends light to both the "sample arm" and the "reference arm", and combines the light that returns from them. Light sent into the "reference arm" reflects from a mirror and returns to the beam-splitter after traversing a path exactly twice as long as the distance to the mirror. Light sent into the "sample arm" illuminates the painting, and each layer of paint may return some light to the beam-splitter by reflection and back-scattering. This small amount of light, which left the beam-splitter and returned by back-scattering from one painting layer, has traversed a path exactly twice as long as the optical distance to the layer (higher index of refraction increases optical distances). If the "reference" and "sample" path lengths match closely (within the coherence length of the light source), resonant wave-fronts combine constructively and destructively at the beam-splitter.

A time-domain OCT system moves the reference-arm mirror to find the distances where interference varies the incident power at its broad-band detector, finding back-scattering depths one-at-a-time. This moving-mirror method may be slow, but applies all available light power to measuring back-scatter of individual, sequential depths. A frequency-domain OCT system keeps its reference mirror fixed during measurements, and replaces the single detector with a spectroscope that measures light power vs. wavelength. The Fourier transform of this measured spectrum then yields a map of all back-scatter strength vs all depths, enabling faster measurements but requiring spectrally-selective detection that must compromise between spectral selectivity, sensitivity, and SNR.

In Chapter 5, we will further compare these techniques for oil paintings and delve into the principles of Optical Coherence Tomography (OCT) and the enhanced OCT system we have developed

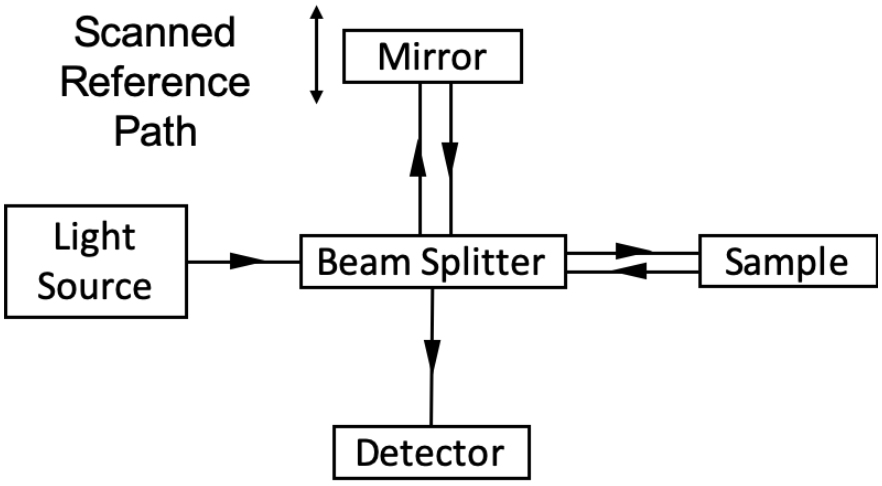

Figure 2.6: **Schematic Diagram of time domain OCT.**

for modeling the internal structure of oil paintings. To investigate the multilayered composition of oil paintings in further detail, we will employ X-Ray Fluorescence(XRF), a chemical analysis technique which will be introduced comprehensively in Section 5.2. Finally, we will employ machine learning methods to predict pigments in multilayered paintings based on the data obtained from the aforementioned imaging and analysis techniques.

CHAPTER 3

3D SURFACE MODELING FOR WORKS OF ART

High-resolution 3D surface modeling is a crucial component in the study of cultural heritage objects, enabling the discovery of intricate details and features that are not visible through traditional 2D photography. This is exemplified in Figure 3.1, which shows how the color texture of a woodblock painting can mask subtle indentations, valuable information that can reveal the painting's provenance and conservation history. (Please note this chapter restates text and figures from our previously published work [56], [57], and also shared work previously appearing in Yunhao Li's master thesis [58].)

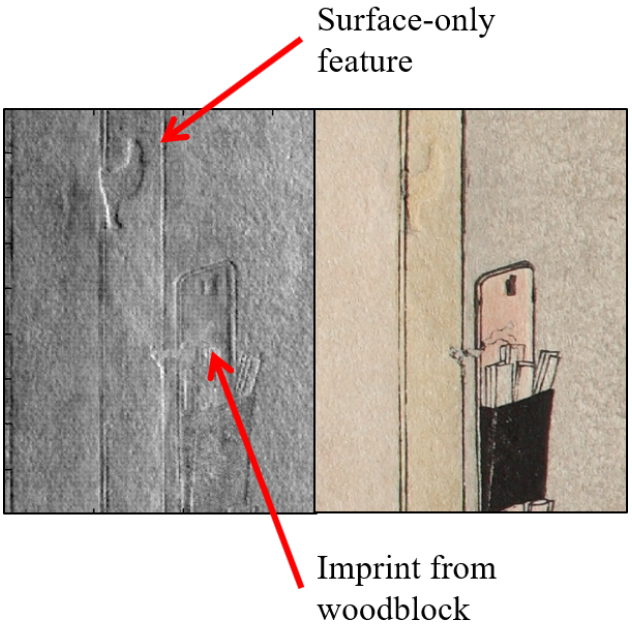

Figure 3.1: **Hidden information in surface measurement of a color woodblock print** (Katsushika Hokusai (1760 – 1849) "Interior of a Brothel in Yoshiwara" 1811), Private Collection

To ensure the accuracy of 3D surface evaluation, it is important to select an appropriate imaging technique for the object being measured. As discussed in Section 2.1, the choice of imaging technique is strongly influenced by the object's microscopic surface structure, which can be divided into two general categories: diffuse (opaque, "matte-finish" surfaces) and specular ("shiny" or mirror-like surfaces). Diffuse surfaces are commonly evaluated using projection-based techniques such as 'Time-of-Flight' (ToF) imaging [26] or "Active Triangulation" methods [23] (also known as "Structured Light" systems). Another well-known technique is active "Photometric Stereo" (PS) [29](Section 2.1.1), which use two or more point-light sources to sequentially flood the surface with illumination from well defined directions. Compared to ToF and Active Triangulation, PS does not require a projector and can collect data for accurate measurement of the surface's "Bidirectional Reflectance Distribution Function" (BRDF) to assist in a more accurate representation of an object's appearance than diffuse and specular terms alone. However, as explained in Section 2.1.2, the use of these techniques on specular surfaces is challenging, and results can often be improved by increasing the spatial extent of the light source. This is the basic principle behind "deflectometry" [59], which addresses the limitations of PS and other techniques.

Although most active PS or deflectometry systems are bulky and confined to laboratory settings [8], [60], some artworks, such as church windows or wall murals, require *in-situ* measurements. To address this need, our lab has developed two methods for measuring diffuse and specular surfaces using handheld mobile devices. Kai Yeh et al. introduced mobile Shape-from-Shifting (SfS), a simple, low-cost, and streamlined active PS framework for scanning planar surfaces with a consumer mobile device and a low-cost add-on lens component [61]. Florian Willomitzer et al. introduced mobile deflectometry system that uses only a mobile handheld device and exploits the screen and front camera for deflectometry-based surface measurements [62]. Both methods are fast, portable, and easy to use for non-experts.

On the basis of their work, we conduct two case studies that highlight potential applications of these new 3D surface measurement tools. The first case study focuses on paintings by Georgia O'Keeffe, where small white protrusions gradually grew, erupted, and fell away, leaving craters on the paint surface. High-resolution 3D surface assessments using mobile SfS [61] enable the monitoring of these protrusions and may aid in early detection while helping to identify the environmental conditions that contribute to their formation. In the second case study, we focus on the stained-glass artworks produced by the Tiffany and La Farge studios in the late 19^{th} and early 20^{th} centuries. The provenance of the stained glass often remained unknown, but using the deflectometry system we built inspired by [62], we gathered characteristic 3D surface texture patterns from works by Tiffany and compared them with glass samples from the well-known supplier "Kokomo Opalescent Glass" producing similar materials at that time. A database of 3D surface information gathered from historic Kokomo glass may eventually help to solve the provenance problem of the Tiffany windows and other contemporary works. Meanwhile, we developed a complete and open-source tutorial and Python code package for our deflectometry system [63] to guide people to build their own systems from scratch.

To provide a clear and detailed understanding of our contributions, we will present the background, underlying principles, experiments, and results of our two case studies individually. In Section 3.1, we begin by providing a review of the methods in the mobile Shape-from-Shifting (SfS) system [61]. We then explain how we used the measured surface normal textures from the SfS system to detect protrusions. In Section 3.2, we will first introduce the large-scale deflectometry system that we developed and then present the experiments we conducted and the results obtained from our stained glass surface modeling approach.

3.1 Protrusion Detection in Georgia O'Keeffe's painting

One of the most comprehensive areas of focus in the realm of cultural heritage science pertains to the long-term behavior of materials in museum environments. Throughout her lifetime, Georgia O'Keeffe observed the emergence of small, unobtrusive protrusions on the surfaces of her paintings [64]. Initially, she attributed these protrusions to sand grains embedded in the paint. However, over the course of several decades, the protrusions grew slowly and some even erupted, leaving small craters in the paint. These protrusions were subsequently identified as metal soap nodules which form and develop gradually within the paint [65]–[68]. The detection of new nodules can be quite challenging as they are difficult to discern based on color, particularly within dark pigments. Therefore, high-resolution 3D surface modeling techniques may rapidly and reliably reveal their locations and dimensions.

3.1.1 Shape from Shifting

Shape from Shifting(SfS) [61] is a portable and robust shape measurement method best suited for finding the 3D shape of diffuse or near-diffuse surfaces. As shown in Fig. 3.2, SfS relies only on a mobile phone with a small 3D printed attachment. The method uses the phone's built-in flash as the controlled light source and the built-in camera as the image detector. By free-hand guiding the cell phone above the object, users capture an overlapping sequence of flash-lit images of the object surface. Image registration based on the scale-invariant feature transform (SIFT) [69] features aligns these images and determines multiple illumination directions for all surface locations. The near-light compensated active PS algorithm then computes the surface normal map for the surface from this array of registered images.

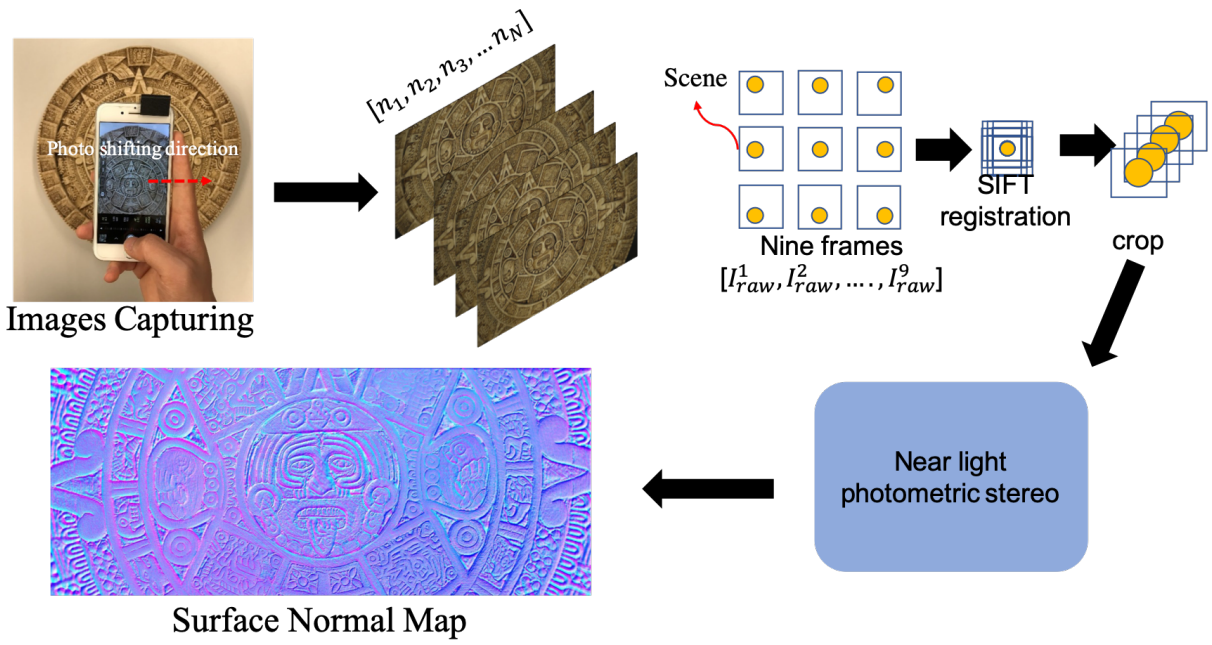

Figure 3.2: **Shape from Shifting.**

3.1.2 Protrusion Detection

The Georgia O'Keeffe oil painting *Pedernal* (Fig. 3.3(a)) has been chosen to illustrate the processing and analysis of active PS data. Imaging was undertaken on a region of the painting, demarcated in Figure . 3.3(a)), that exhibits a high number of protrusions and is representative of the overall palette [64]. From the surface normal map measured by active PS (Fig. 3.3(b)(c)), we can clearly notice the existence of protrusions as blob-like features, whose reconstructed 3D shape is similar to a half sphere of about 20 mum to 500 mum (Fig. 3.3(d)).

Salvant et al. [64] manually identified the distribution of these protrusions in various pigmented area and created 3D surface renderings to help find incipient eruptions for preventive treatment, but to further improve the efficiency of protrusion detection, I developed an automatic protrusion detection method based on the surface normal map. As the shape of the protrusions is well-

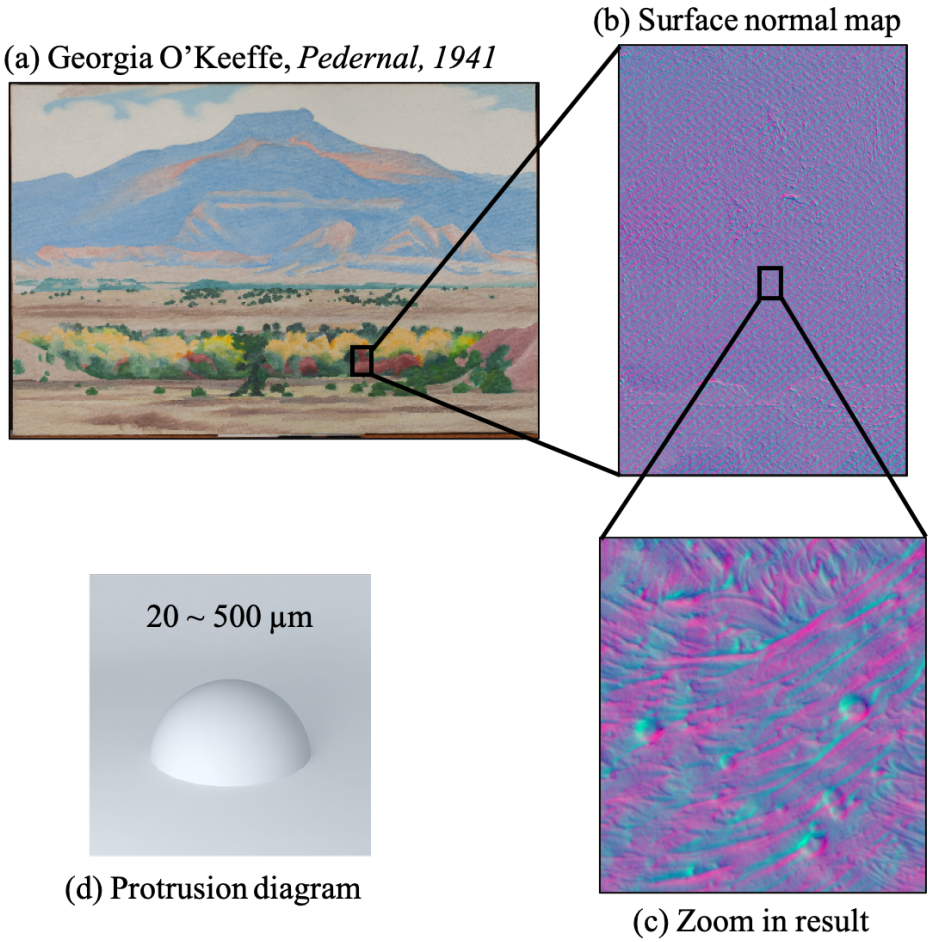

Figure 3.3: **Protrusions in Georgia O'Keeffe's painting.** (a) The color image of Georgia O'Keeffe's painting. (b) Measured surface normal map by PS. (c) Enlarged result of the surface normal map. (d) Schematic model of a protrusion: a half sphere of about 20 μm to 500 μm.

described by half-spheres or ellipsoids, and the Laplacian of the shape is consistently positive in the center with negative surrounds, I can calculate Laplacians of the painting surface shape and applied thresholds to create a binary image. Classifying the binary image by connected-components with well-chosen circularity and eccentricity coefficients provided a robust and accurate map of every detectable metal soap protrusion shown as Figure 3.4. Particularly, for the normal map of this painting, I utilize the `imfindcircles` function from MATLAB [70] with option of two-stage circular Hough transform, and sensitivity as 0.78.

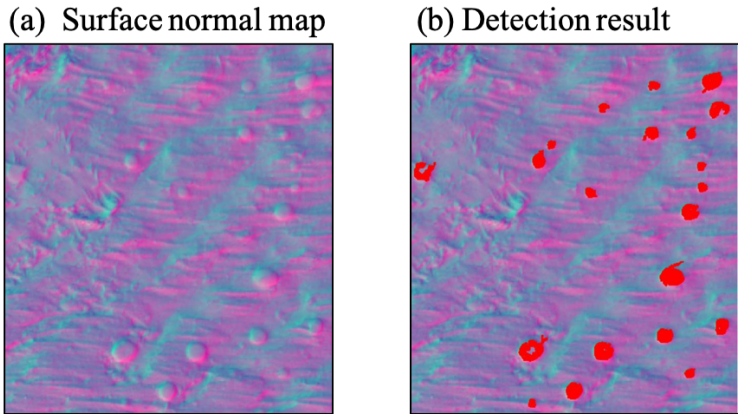

Figure 3.4: **Protrusion Detection.** (a) An example area of the surface normal map. (b) The detection result of this area. The detected protrusions are marked in red.

3.1.3 Discussion

The protrusion detection method proposed in this study facilitates the documentation of protrusions and helps to identify the environmental conditions that contribute to their formation. One hypothesis for the formation of these protrusions is that they arise due to a chemical reaction between the canvas and pigments, resulting in the protrusions beginning below the paint layer and gradually tunneling to the surface [68]. To further investigate this hypothesis, future research could integrate the surface shape information obtained from the protrusion detection method with inter-

nal 3D shape information collected from volumetric modeling systems, such as the 2μm optical coherence tomography system described in Section 2.2.1. By combining these approaches, it may be possible to gain a more comprehensive understanding of the formation mechanisms underlying protrusions in paintings.

3.2 Provenance of Stained Glass Pieces in Tiffany Artworks

As previously mentioned, 3D surface imaging can aid in the determination of materials' provenance. The Tiffany stained glass window presented in Figure 3.5(a) features colored glass pieces with surface textures formed by various techniques, and first introduced between 1890 and 1910 by prominent stained glass artists such as Tiffany, Kimberly, Duffner, and LaFarge. They revolutionized the field by utilizing colored and textured glass to create three-dimensional effects. These techniques were particularly influential in the Art Nouveau movement throughout Europe and the burgeoning US architectural designs. Despite the popularity of textured glass, few manufacturers produced it, with the Kokomo opalescent glass factory in Indiana being one of the primary sources. Although L. C. Tiffany used Kokomo glass in other works, but object records for most of these windows are scarce, and the provenance of textured glass in this window and many others in Europe remains unknown. Even now, the Kokomo glass factory still keeps and applies many of the same large, century-old handmade metal rollers, each with a uniquely carved pattern, to create various surface textures on the glass. After imaging these old rollers at the factory, and after combining them with images of other Kokomo glass samples with known provenance, we built a database of texture-image "fingerprints" to help reliably identify glass textures made from the Kokomo factory rollers.

(a) Tiffany Stained-Glass, *Driehaus Museum, 1900*

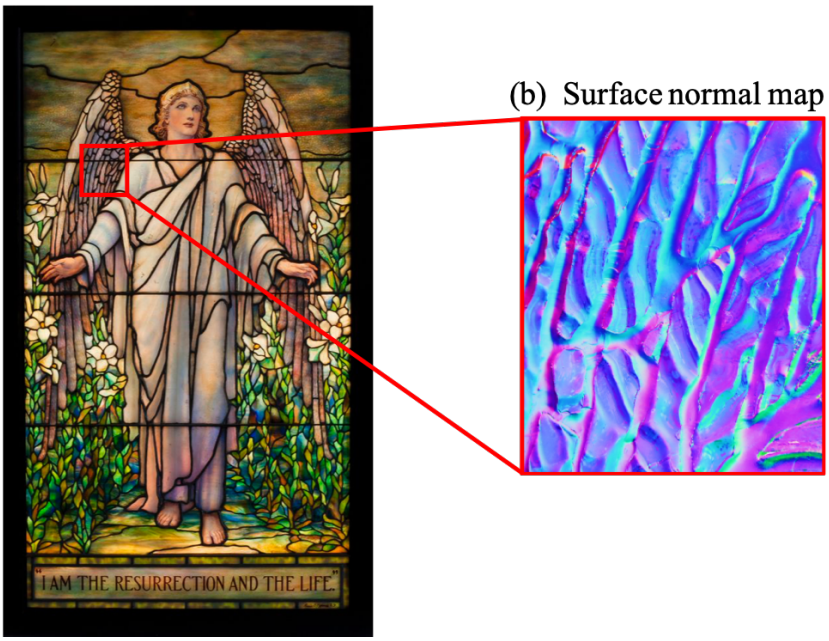

Figure 3.5: **Glass features of the Tiffany Stained-Glass.** (a) The color image of the Tiffany Stained-Glass. (b) Measured surface normal map of one glass feature.

3.2.1 Large-Scale Deflectometry for Flat Specular Objects

The reconstruction of large-scale specular objects, particularly stained glass windows, has been a significant challenge for years due to their extensive specular surface areas and finely-textured or "high spatial frequency" 3D surface information. Although these windows display distinct details and 3D features, their overall shape is commonly flat with little or no self-occlusion. Therefore, we present a low-cost, portable, and high-accuracy 3D reconstruction method for large-scale specular objects based on deflectometry.

Our setup (Fig. 3.6) sets a camera to view the specular surface from an oblique angle, and carefully positions a large flat display screen symmetrically, so that the camera sees the display reflected in the specular surface we wish to measure. The display, which may be an active LCD or OLED display, or a viewing screen illuminated by a digital projector, shows a short sequence of sinusoidal "fringe" patterns (e.g. sine waves at 0, 90, 180 and 270 degree phase shifts) that uniquely identify every point on the display screen by its sequence of light intensities. The camera, aimed at the specular surface, captures the sequence of light values reflected from every specular surface point, and that sequence uniquely specifies the display location. From this location, the system determines the vector s $\mathbf{n}$s $= [n_x, n_y, n_z]^{\mathrm{T}}$). By replacing the LCD screen commonly used in conventional deflectometry setups with a projector and an entire diffuse-white wall, our system can create an effective screen large enough to "cover" the angular range and surface area of large-scale specular objects with high spatial frequencies on its surface. This system can be assembled/disassembled within 5 minutes and can be deployed at any place with a surface large enough to serve as the "screen".

In our approach, we address the normal-depth-ambiguity problem (described in Section 2.1.2) that is inherent to each deflectometry method by utilizing prior knowledge of the object under test, specifically that its overall shape is flat. To achieve this, we place markers on the object

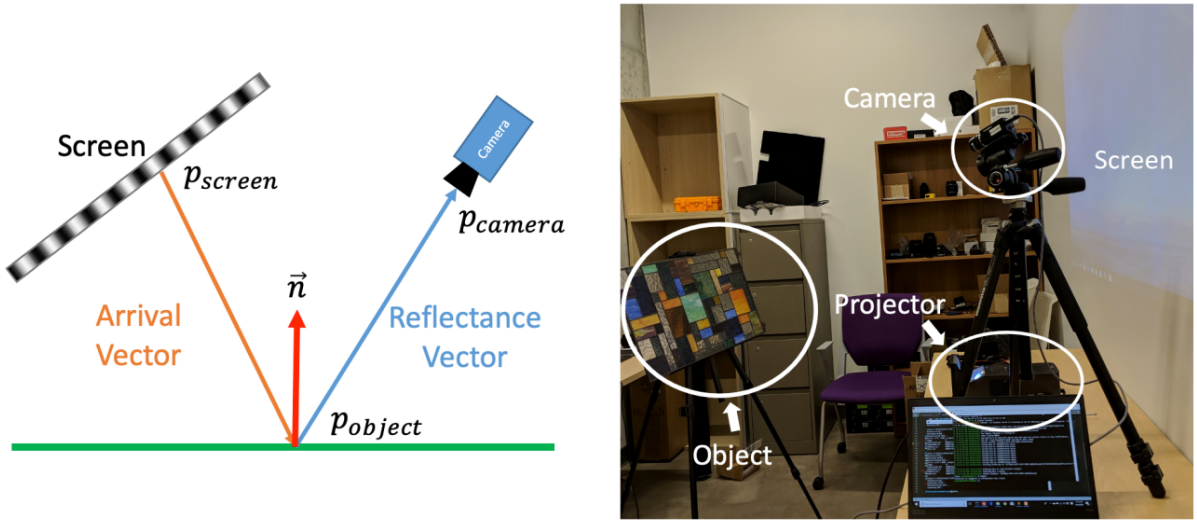

Figure 3.6: **Deflectometry System Setup.** (a) Basic concept of deflectometry. (b) Our system setup contains one camera, one project, the white wall as the screen, and the object.

and use our calibrated camera to detect their positions in space. We then fit a plane between the marker positions, which is used to approximate the depth of each point on the object's surface. This approximation allows for an unambiguous calculation of the surface normal. To compute the surface normal, we require calibration of the camera intrinsic matrix and the position and pose of the screen and object. To achieve this, we integrate a system calibration procedure into our system.

System calibration comprises three subtasks: Intrinsic Calibration, Geometric Calibration, and Radiometric Calibration. Intrinsic calibration returns the camera intrinsic matrix and distortion coefficient, geometric (or 'extrinsic') calibration returns the camera, object, and screen's pose and position, while radiometric calibration removes the camera and projector's radiometric error, due to the nonlinear mapping from pixel values to display irradiance, and from scene radiance values captured by the camera's lens to pixel values from its image sensor.

Intrinsic calibration measures the camera and image coordinate systems, and finds the 4x4

matrix (the “camera intrinsic matrix”) and a set of distortion coefficients that map 3D positions in front of the camera to 2D pixel locations. We use the OpenCV library [71] augmented with a ChArUco board [72] to achieve higher accuracy than the chessboard alone. ArUco markers are placed at the white squares on the board, and both the chessboard corners and ArUco markers are detected during calibration.

Geometric calibration links the calibrated camera coordinate system to the unified ‘world’ coordinate system that specifies the specular object and the display screen locations, and defines the specular object’s surface normal vectors. As illustrated in Fig. 3.7, it supplies both the object-camera and screen-camera transformation matrices. The system computes the object-camera transformation matrix from detected image locations of the ArUco markers attached to the object’s corners. As the camera cannot directly observe the screen, we incorporate a planar mirror, usually located at the desired object position, to enable the camera to capture the entire screen through reflection. To obtain the screen-camera transformation, we first calculate the mirror-camera transformation from the camera images of the planar mirror with ArUco markers on its corners. We then display a checkerboard pattern on the screen and capture its images with the camera, which provides the virtual image location of the screen through the mirror. Combining this information with the mirror-camera transformation, we can obtain the screen-camera transformation matrix.

Radiometric calibration removes the nonlinearity that connects pixel values in the camera to pixel values in the display or projector. I used the Debevec et al. [73] approach to find response curves mapping radiance to pixel value and reflectivity calculation using display-camera radiance. To obtain the camera response curve, photos of an arbitrary scene with different camera exposure values are taken. To calculate the display response curve, plain images with varying pixel values are projected onto the wall, and pictures of different pixel values with fixed exposure time are captured.

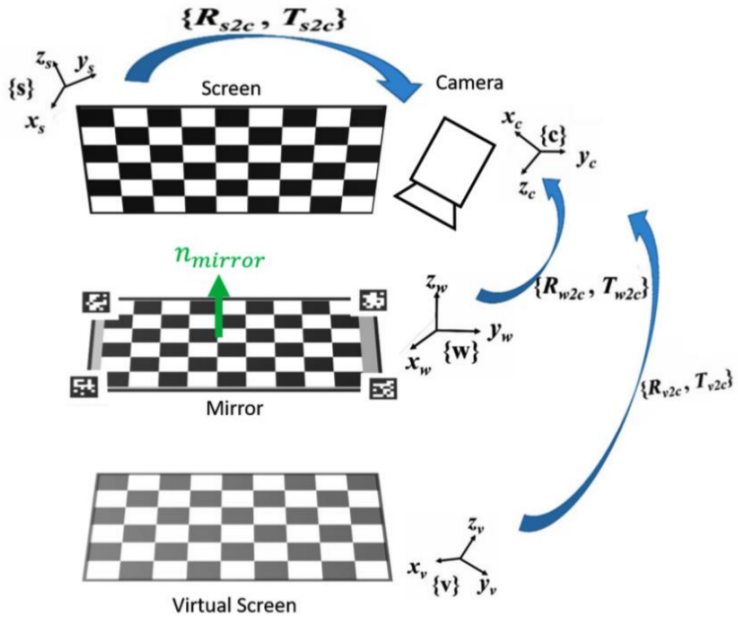

Figure 3.7: **Geometric calibration for the display-camera system using a planar mirror.**

You can find the open-source Python package and complete tutorial for building your own deflectometry system here [63].

In order to assess the accuracy of our proposed system, we conducted an experiment using a lab-grade spherical mirror [74]. The measured normal map obtained from our system was compared with the ground truth normal map, generated using the manufacturer's specifications. The results of this comparison are shown in Figure 3.8(a), where the normal maps of the mirror are displayed in terms of their nx, ny, and nz components. Additionally, Figure 3.8(b) presents the absolute angular errors between the measured and ground truth normal maps. Our evaluation reveals that the average absolute angular error is 0.67 degrees, while the root mean square angular error is 0.75 degrees. These results demonstrate that our system is capable of accurately recovering the surface normals of specular objects.

With our large-scale deflectometry system, we can develop a database of "fingerprints" of

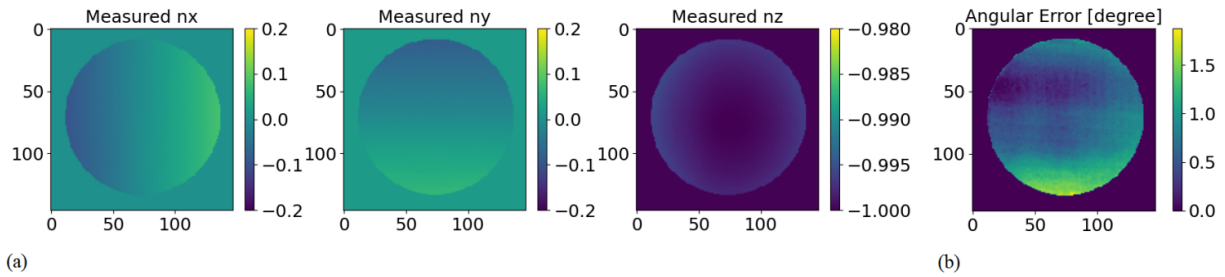

Figure 3.8: **Spherical Mirror Evaluation Results.** (a) Measured normal map of the mirror (b) Angular errors between normal maps.

Kokomo glass rollers as shown in Figure 3.9.

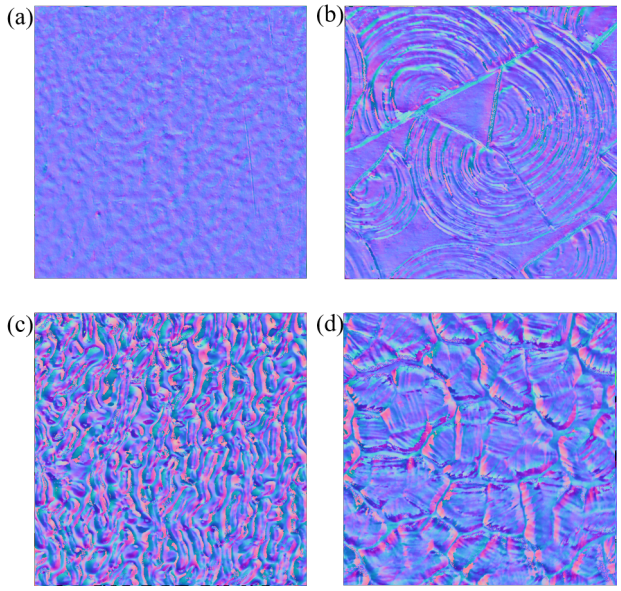

Figure 3.9: **Normal maps of four example features in the database of Kokomo glass rollers.**

3.2.2 Discussion

This large-scale deflectometry system has a limitation regarding the angular range of the reconstructed normal vector. The normal map obtained from this system is unable to recover the surface normal of all specular regions on the object, resulting in meaningless noise data. This effect can

occur due to the limited screen area that prevents the light from the light source from reaching some regions of the surface, or because the object surface within these regions does not reflect the light into the camera's field of view, or self-occlusion due to sharp peaks and valleys in the specular surface prevent the camera from seeing small portions of the specular surface. Additionally, the computed normal map varies when the object is shifted to different positions within the camera's field of view, due to different parts of the object's specular surface reflecting light from the screen into the camera's field of view, causing differences in the computed normal maps, as shown in Figure 3.10.

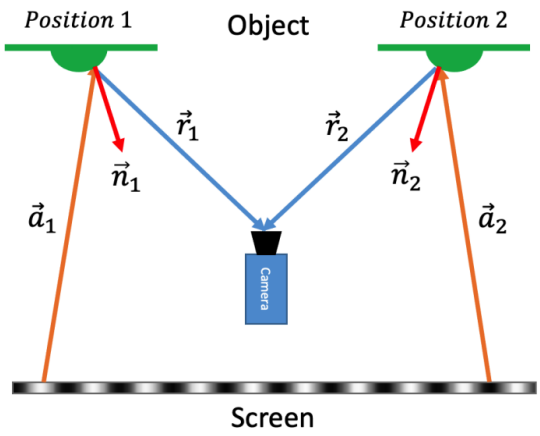

Figure 3.10: **Cause of normal difference on the same object at different positions.**

I propose two solutions to remove these effects. The first solution is to expand the display area, which is limited by the size of the wall that restricts the display area of the projected screen. The second solution is to introduce multiple cameras to build a multi-view deflectometry system [42]. However, this solution presents a challenge in terms of the additional error introduced by calibration of multiple cameras and the fusion techniques of normal maps retrieved from different perspectives.

3.3 Conclusion

In this chapter, our objective of simplifying the surface shape reconstruction process while maintaining accuracy and accessibility has been achieved. We have replaced the traditional complex and expensive setup with two novel methods, SfS and large-scale deflectometry, which are simple, cost-effective, portable, and have been proven effective with different materials and scales. Furthermore, we have presented two case studies that exemplify the practical applications of these methods: the protrusion detection problem using PS and the stained glass identification problem using deflectometry.

Surface modeling plays a critical role in various fields, including cultural heritage applications. The results of this chapter provide a basis for future research in combining these two methods to measure the surface shapes of more complex materials. The potential of these techniques for fieldwork and remote measurements opens up new opportunities for non-invasive and non-destructive examination of objects and artifacts. In the next chapter, we will further explore surface modeling techniques in virtual reality applications.

CHAPTER 4

3D SURFACE MODELING FOR EYE TRACKING

Eye tracking is a key component in human-computer interaction, especially for AR/VR applications such as visual search [75], foveated rendering [76], and human interaction [77], [78]. In the field of eye-tracking, head-mounted, non-contact systems (systems that determine gaze direction by aiming head-mounted cameras and lights at eyes) can be divided into two main categories: appearance-based and feature-based approaches [79]. Appearance-based methods [80]–[82] directly use image contents as input without explicitly extracting features like pupils and glints, aiming to estimate the point of attention, relevant features and personal variation implicitly without the need for scene geometry or camera calibration. However, these methods are known to require extensive user-specific training datasets to achieve better performance [83]. (Please note this chapter restates text and figures from our previously published work [84]–[86]. The work in section 4.2 is completed during my internship at Meta Reality Labs.)

On the other hand, feature-based approaches utilize extracted local features, including contours, eye corners, and reflections, that encompass aspects related to the geometry of the system and eye physiology. Two types of feature-based approaches currently predominate: 2D regression-based and 3D model-based methods. The former assumes a particular parametric or nonparametric form, such as a polynomial [87] or neural networks [88], [89] for the mapping from image features to gaze coordinates (2D or 3D). These methods avoid explicit calculation of the intersection between the gaze direction and gazed object and do not require camera or geometric calibration. However, they may struggle to handle head pose changes or still require initial training with large datasets.

In contrast, 3D model-based methods [90]–[93] directly calculate gaze direction from eye fea-

tures by sensing the pose of a geometric model of the eye. The main sources of eye features are corneal reflections [94]–[96], pupil centers and iris edges [97], [98]. These methods commonly assume spherical or ellipsoidal surfaces for the eyeball and cornea and fit mathematical parameters to represent the shape using detected features. However, this may neglect higher spatial-frequency shape variations, leading to errors and reduced accuracy. Therefore, in this chapter, I will combine 3D surface modeling with model-based eye-tracking methods to further improve performance. I will first present an eyeball surface modeling technique with deflectometry in Section 4.1. Then I will introduce a mesh-based eye-tracking method that uses high-frequency cornea and sclera shape for each person to achieve high-accuracy eye tracking in Section 4.2.

4.1 Eyeball Surface Modeling using Deflectometry

Again, current model-based eye-tracking methods use features detected from 2D eye images or sparse reflections of point light sources on the eye surface (corneal/scleral reflections) to fit spherical or ellipsoidal models to estimate the eyeball surface and gazing direction. However, the limited precision of such methods is due to the lack of higher spatial-frequency shape variations. The field of ophthalmology has made significant advancements in medical imaging techniques, resulting in precise models of the eyeball that incorporate corneal topography and tomography [99], [100], generated using various imaging methods such as Placido disk-based keratoscopyy [101], [102], slit-scanning elevation topography [103], [104], Scheimpflug imaging [105], [106], and optical coherence tomography (OCT) [107], [108].

Despite the advantages of these imaging methods, they rely on complicated stand-alone optical systems and components, which can be challenging to implement in VR headsets. To overcome these limitations, we propose an approach that significantly increases the information content obtained from corneal or scleral reflections by using deflectometry to obtain a dense and precise 3D

model of the eye surface. Our method, which uses off-the-shelf hardware, can acquire approximately one million surface points per measurement step. Moreover, we can utilize the retrieved surface normals and dense 3D features estimated via deflectometry to accurately estimate the gazing direction.

4.1.1 Stereo Deflectometry

In Section 2.1.2, I introduce deflectometry as a well-established method in surface metrology that is used to measure the surface normals of specular objects and, when needed, reconstruct the 3D shapes they describe. However, there is the depth-normal-ambiguity problem as shown in Figure 2.3. To address this problem, in Section 3.2.1, for mostly flat specular objects, e.g., stained glasses, we assume that the surface is a plane and use the markers to calibrate the plane to get the depth information for each surface point. Consequently, we are able to obtain one and only one position and its corresponding surface normal vector for each camera-to-display correspondence point pair as shown in Figure 2.3. However, this method cannot be used for non-flat objects such as the eyeball. To overcome this inherent depth-normal-ambiguity problem, we incorporate a second camera into our system resulting in a "stereo deflectometry" system [40], as shown in Figure 4.1. In this system, we calibrate the two cameras and the screen in the same world coordinate using our novel calibration method, which we will introduce in the next section. After calibration, two pairs of correspondence points are obtained for each object point, and there exists one depth and normal vector direction that can bisect both sets of correspondence rays.

4.1.2 Camera-Display System Calibration

The accuracy of system calibration significantly influences the accuracy of deflectometry measurements. Calibration involves both intrinsic and geometric calibration of all cameras involved,

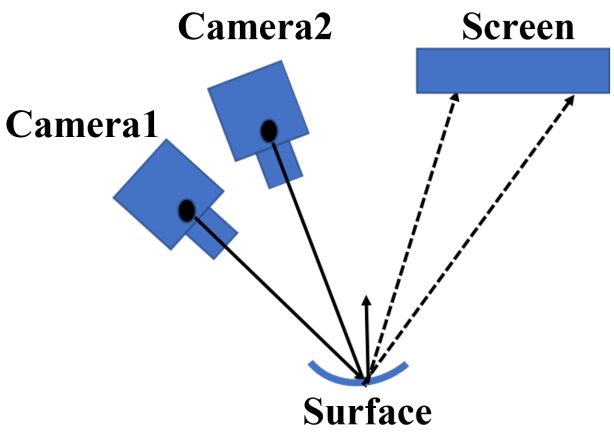

Figure 4.1: **Stereo Deflectometry Setup.**

as well as geometric calibration of the entire camera(s)-display system. Typically, camera calibration is performed using Zhang's method [109], which employs a black and white checkerboard with known dimensions as a calibration target. However, for small fields of view (FoV), such as those required to calibrate a deflectometry system for measuring the eye gazing vector in VR headsets, this method can be challenging to execute. Self-printed calibration targets often fail to meet the accuracy requirements, and customized high-accuracy targets, such as chrome-on-glass targets, can be prohibitively expensive.

Geometric camera(s)-display-system calibration for deflectometry systems poses additional challenges. Usually, the position and tilt of the display are calibrated by observing it with the camera(s) over an auxiliary mirror. However, accurately applying markers onto the mirror surface to determine the position and tilt of that mirror can introduce more errors. Additionally, the markers often occupy a significant fraction of the camera(s) FoV, which can limit the useful angular extent for the screen and reduce the span of measurable surface normal directions.

To address these challenges, we propose a flexible and straightforward calibration approach that replaces all necessary calibration targets and mirrors with a small, sufficiently flat screen, such

as a mobile phone. The "calibration screen" can be remotely controlled to emulate any required target or marker. Furthermore, the screen can serve as a mirror when it is not displaying any patterns.

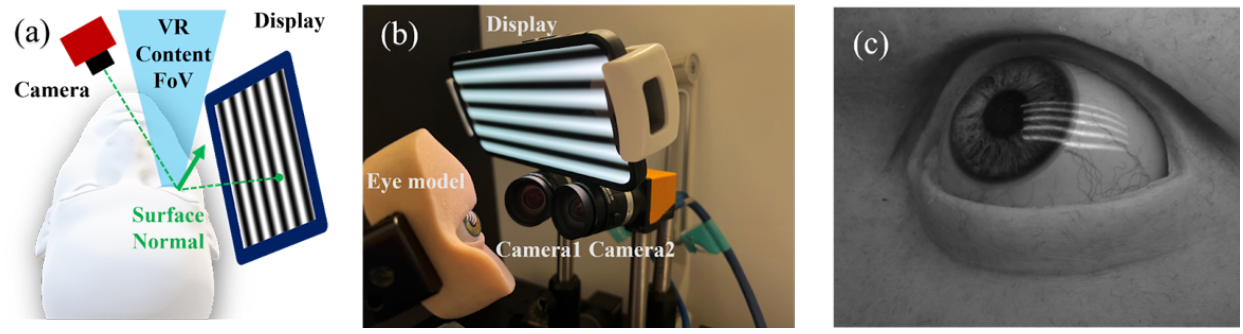

Figure 4.2: **Deflectometry-based VR eye-tracking.** (a) Basic concept of deflectometry. (b) Current prototype setup using Stereo Deflectometry. (c) An example of captured images

To implement this approach, we use an iPhone 12 Pro screen with 460 pixels per inch, as depicted in Figure 4.2(b). The calibration process consists of three steps: intrinsic camera calibration, stereo calibration, and geometric camera(s)-display-system calibration.

The first step involves the intrinsic calibration of both cameras. To use the phone screen as a target for the camera calibration, we display a checkerboard pattern with 8 rows and 11 columns. Each checker has a width of 28 pixels, which corresponds to 1.546mm. By precisely moving and tilting this small "emulated" calibration target within the measurement volume, we perform an intrinsic camera calibration with a mean reprojection error of 0.08 pixels. Our experiments show that previous attempts using handmade printed calibration targets yielded a mean reprojection error of 0.35 pixels, indicating the superior accuracy of our proposed approach.

In the second step, we calibrate the relative position between both cameras used in the prototype system. To achieve this, we move the "emulated" calibration target within the field of view of both cameras and capture stereo image pairs to perform a stereo calibration.

Finally, we perform the geometric camera(s)-display-system calibration to estimate the relative position between the camera and the display used to project the sinusoidal patterns during the actual deflectometry measurement process (see Fig. 4.2). This calibration step is crucial, as even minor errors in this process can result in significant errors in the reconstructed 3D surface. In conventional deflectometry systems, an auxiliary flat mirror is utilized to locate the display's position, as it is commonly not directly visible to the camera(s). For this calibration, the mirror's position in space needs to be accurately determined, which is commonly done by applying several markers on the mirror. In our proposed approach, we use the flat phone screen to replace the flat mirror. To determine the position of the phone screen in space, we first display a checkerboard pattern on the screen (see Fig. 4.3). This eliminates the need for applying markers to the mirror. We then remotely switch off the phone screen (without changing its position), causing it to act as a mirror that reflects the pattern displayed on the actual screen (see Fig. 4.3b). By displaying a known pattern with known dimensions on the actual display (e.g., another checkerboard pattern - see Fig. 4.3b), we can calculate the display's position and tilt in space. To evaluate the quality of our geometric camera(s)-display-system calibration, we trace the three-dimensional checker positions of the calculated (virtual) display back to the camera chip and compare it with the real images captured by the camera (see Fig. 4.3c). Our experiments show that the root mean square error (RMSE) of this reprojection error is 0.42 camera pixels.

Although the proposed calibration method offers higher accuracy compared to conventionally printed targets, there are potential drawbacks that need to be considered. One such drawback is the possibility of manufacturing tolerances in the pixel size and position of the phone screen, which can affect the calibration outcome. Additionally, the cover glass of the screen can cause refraction of the light rays emitted from each pixel. While we found these effects to be negligible in our experiments, they can introduce significant errors when the screen is viewed at oblique

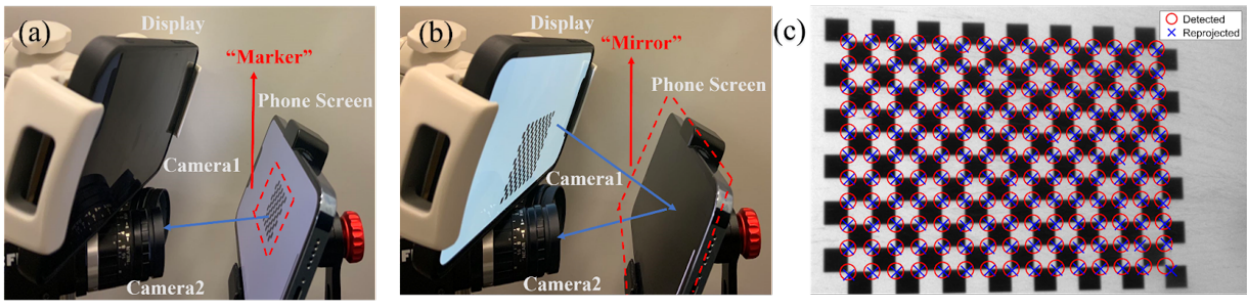

Figure 4.3: **Geometric camera(s)-display-system calibration.** (a) The position and tilt of the phone screen are found by imaging a displayed checkerboard pattern, which acts as a "marker". (b) Eventually, the position and tilt of the display are found by observing a displayed checkerboard pattern reflected over the phone screen. The phone screen is switched off and acts as a "mirror". (c) Reprojection error: Three dimensional checker positions of the calculated (virtual) display are traced back to the camera chip (blue crosses). Comparison with detected checker positions in the real image (red circles) reveals a reprojection RMSE of 0.42 camera pixels

angles. Future research can focus on developing models to rigorously address these effects to further enhance the calibration accuracy.

4.1.3 Single-shot Stereo Deflectometry

We introduce a novel technique for measuring the surface normal map of the eye using a single-shot Stereo-Deflectometry approach. This involves projecting a crossed fringe pattern onto the eye's surface and evaluating the phase information for both fringe directions in single-shot through a 2D continuous wavelet transform approach, as detailed in [110], in order to establish correspondence between the screen and the camera. By utilizing a second camera, we are able to resolve the normal-depth ambiguity problem inherent in Deflectometry, as outlined in Section 4.1.1, which results in a unique reconstruction of the shape and normal map of the eye surface in a single shot.

To determine the gaze direction, we trace the calculated surface normals back to the center of the eye. Due to the cornea and sclera's spherical shape, but with vastly different radii, the back-traced normals intersect at two distinct points inside the eye - the centers of the cornea and of the

sclera. Connecting these two points provides the optical axis of the eye and the gaze direction. It is worth noting that this approach works even if the cornea and sclera are not perfectly spherical but still exhibit rotational symmetry. In such cases, all back-traced normals intersect along a line that coincides with the optical axis of the eye.

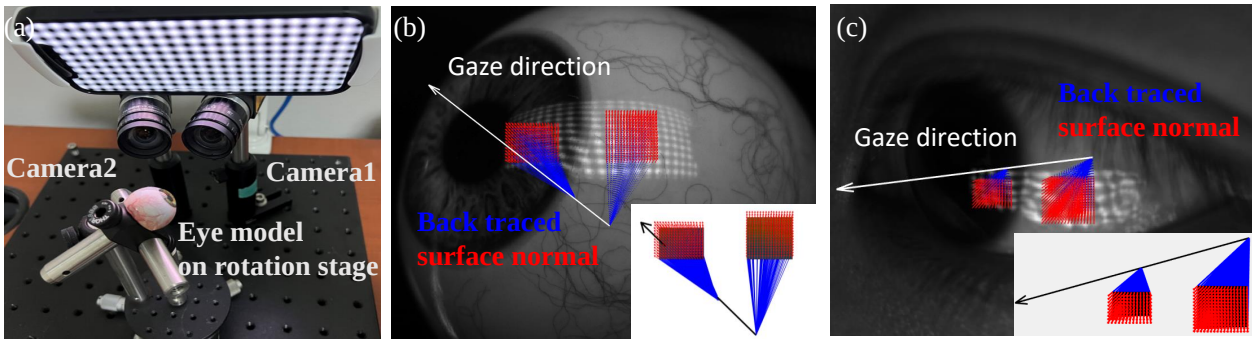

Figure 4.4: **Eye tracking using single-shot stereo Deflectometry.** (a) Prototype setup. (b,c) Captured camera images of realistic eye model and real human eye, overlaid with the calculated surface normals and estimated gaze direction.

We conduct a real-world experiment to evaluate the effectiveness of our method. To do so, we utilize a realistic eye model mounted on a rotation stage, as depicted in Fig. 4.4a. The eye model is rotated to various positions ($-3°$, $0°$, $3°$, and $6°$), and we determine the gaze angle θ_a at each rotation position using the methodology described previously. Since the absolute gaze angle is unknown, we calculate the relative gaze angle between two rotation positions and compare the result with the angle we rotate the eye model (i.e. the "ground truth"). We collect 20 measurements for each rotation position, for a total of 80 measurements. The eye model is always moved or rotated before each measurement is taken.

We calculate the mean relative error $\epsilon_{0°}$ at the rotation position $0°$ with respect to all other rotation positions a, which is given by $\epsilon_{0°} = ||\bar{\theta_a} - \bar{\theta_{0°}}| - |a - 0°||$. The results are presented in Tab. 4.1, where it can be observed that $\epsilon_{0°}$ is below $0.35°$ for all measurements, demonstrating the effectiveness of our method.

Rotation position a	-3°	3°	6°
Mean relative error $\epsilon_{0°}$	0.34°	0.24°	0.17°

Table 4.1: Evaluation of estimated gaze direction for single-shot Stereo-Deflectometry method and optimization-based method.

To demonstrate the feasibility of our approach under real-world conditions, we also perform a qualitative single-shot measurement on a living human eye. The result is illustrated in Fig. 4.4c. The measurement on the human eye exhibits similar properties to those observed in the model eye, with the evaluated surface normals intersecting at two points, and the evaluated gaze vector being qualitatively correct. Note that a procedure for the quantitative evaluation of the absolute gaze direction and further development is still ongoing work at Prof. Florian Willimizter's Computational 3D Imaging and Measurement (3DIM) Lab.

4.2 Eye Tracking with High Fidelity Eye Models

In eye-tracking research, a re-renderable model of the eye-region that can extrapolate to novel viewpoints and lighting conditions is highly valuable for rendering avatars [111] and generating synthetic data for machine learning-based eye tracking [112]. However, capturing the surface of the cornea and sclera with computer vision is particularly challenging. The surface is highly specular, and its mirror-like reflection from its curved and bumpy surface makes accurate capture difficult to reliably achieve. Furthermore, the cornea is transparent, so it cannot be directly imaged by cameras; its position and shape can only be inferred from reflections and refractions of other parts of the scene. In Section 4.1, we introduce a method for modeling the eyeball surface using deflectometry. Another widely used method in commercial eye-tracking systems is to place lights at known positions in the scene, infer the surface location by matching the reflected images of the lights (called "glints") to the lights that created them, and then use 3D geometry and the law of

reflection to optimize the surface location. These glints, often in combination with a pupil tracker, have been used in many geometric eye trackers to date [113]. Such methods can even be used to track the eye in situations where the pupil is not visible, such as the Event Camera glint tracking method of [114].

However, a limitation of prior methods is that they mostly rely on the "sphere-on-sphere" model for the eye, which models the cornea as a spherical surface located as a bump on a larger spherical surface for the sclera (white of the eye). This model has been successful for real-time tracking methods like [113], but it works best when the glints are reflected off the central portion of the cornea, which is the most spherical part of the human eye. For glints that reflect off the transition area between the cornea and the sclera, these models break down because the surface position and normal can be quite different from the sphere-on-sphere model being used for 3D geometric optimization. Such non-central glints are especially common for eye trackers in AR/VR contexts, where the cameras and lights typically view the eye from an oblique angle on the frame of the glasses or headset.

Although model-based eye-tracking methods mostly rely on simplified parametric models of the eyeball, recent research in computer graphics has explored using image-based capture systems to reconstruct more realistic eye models for photo-realistic animation. In 2014, Bérard et al. [115] reconstructed the spatiotemporal shape of all visible parts of the eye, including the sclera, cornea, and iris. They used facial scanning [116] to create partial sclera reconstructions for each gaze direction, then integrated these scans into a complete eyeball model. To model the transparent cornea surface, they used highlights from different eye poses to fit an open uniform B-spline surface with 100 control points. Most recently, Kerbiriou et al. [117] proposed a 3D reconstruction methodology to generate 3D meshes of the eye region. They fit a generic eyeball mesh on each scan for visual purposes, with an estimated sphere representing the cornea.

In this section, I will propose an alternative approach to the widely used sphere-on-sphere model for tracking the human eye. Our approach leverages the use of more complex cornea surfaces that better capture the real shape of the eye. We obtain these surfaces from existing eye scanning systems that perform 3D scans of the eye prior to tracking. To demonstrate the effectiveness of our method, I will show how to track these complex surfaces in multiview images of the eye using a 3D geometric pose fitting approach that uses the glints from known lights.

We evaluate our approach on a dataset of 10 individuals and show qualitative results of improved glint rendering from novel views and lighting conditions. Our results demonstrate the effectiveness of our approach and the potential for more accurate eye tracking using complex cornea surfaces.

4.2.1 Multi-view Glint-only Eye Tracking

I present a multi-view glint-only method for tracking the pose of the eye using a previously scanned mesh of the cornea and sclera. Using a lighting scheme called differential lighting, I am able to isolate the specular appearance of the eye from our eye-tracking images, and demonstrate fitting the eye pose using both a geometric loss term and an appearance-based loss. Figure 4.5 presents an overview of our eye-tracking framework based on mesh-based eye models. Figure 4.6 shows one example set of raw camera images from 9 views.

Eye Region Segmentation

I first utilize a pre-trained SegNet model [118] to perform image segmentation on the captured images from multiple camera views. The SegNet model enables me to accurately segment the eye region and cornea region in each camera image. Specifically, the segmentation results allow me to tell which specular pixels and glints belong to the cornea (in the pupil and iris segmentation

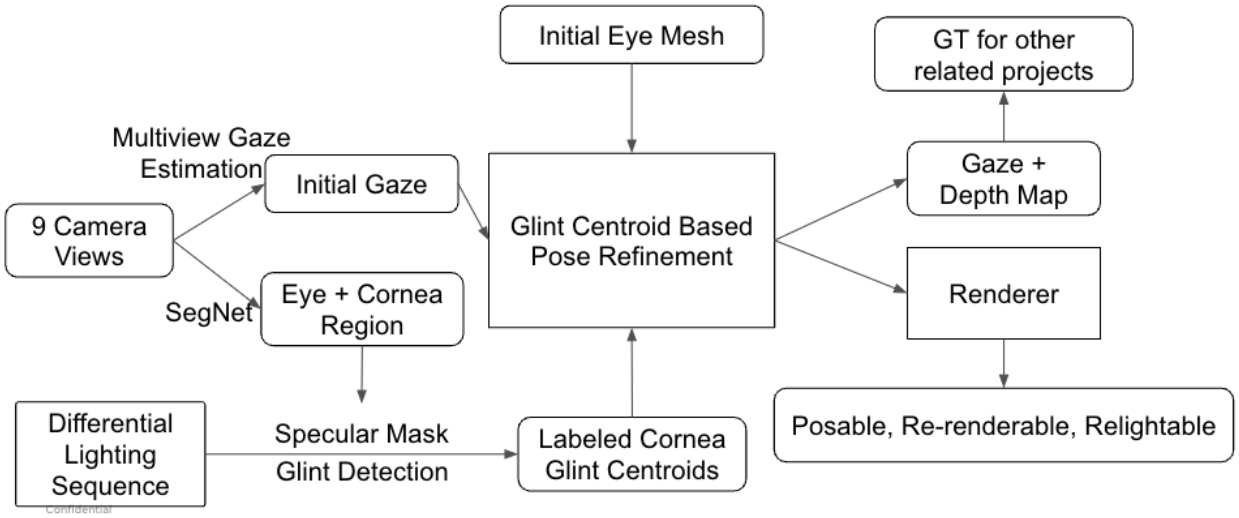

Figure 4.5: **Multi-view Glint-only Eye Tracking Framework.**

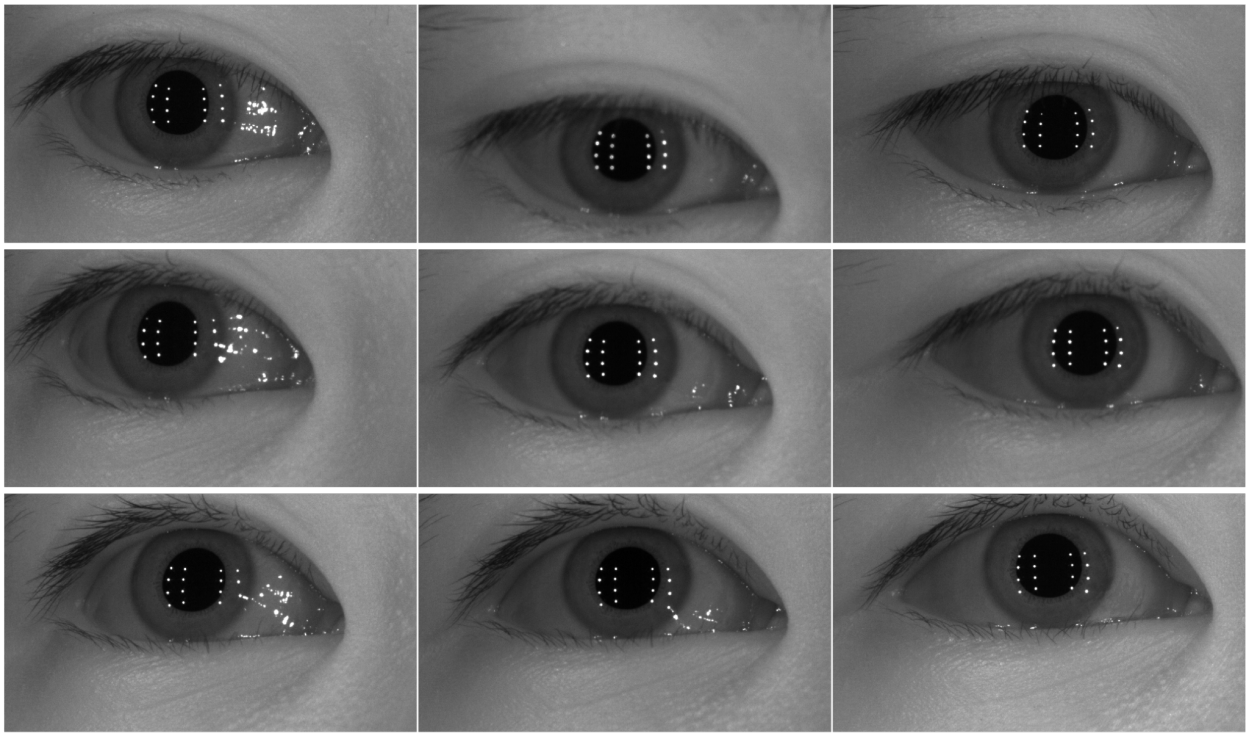

Figure 4.6: **Raw images from the capture.** Multiview camera images from 9 views.

regions) vs. specular reflections from the sclera. Furthermore, I can increase the robustness and performance in the further steps of the algorithm by discarding pixels belonging to the face category of the segmentation, since these are not informative for the specularity of the eye. The use of the pre-trained SegNet model has been shown to be effective in previous research [118], allowing for accurate segmentation of the eye and cornea regions in a range of different image settings.

Multi-view Gaze Estimation

To get an initial pose for the eyeball, I apply the multi-view gaze estimation method from Guestrin and Eizenman [113]. The gaze estimation process involves several steps. First, the eye regions are detected in each view using a combination of skin color segmentation and feature-based approaches. The pupil center is then detected in each view using a circular Hough transform, and the 3D position of the pupil center is estimated using the calibrated camera parameters. The gaze direction is estimated in each view by computing the direction vector from the 3D position of the pupil center to the center of the eye. The gaze directions from all nine views are then combined using a Bayesian framework to obtain a more accurate estimate of the gaze direction.

Glint Labeling with Differential Lighting

Glint labeling is a long-standing challenge in glint-based eye tracking. In order to achieve high model performance, it is crucial to determine the correct correspondence of glints to lights in different camera images. Additionally, the presence of blockage by the nose and eyelashes, as well as variant glint shapes due to irregular surfaces, pose further challenges for glint labeling. To address these issues, I propose adding a differential lighting sequence capture for each eye pose in our model. This involves capturing a series of frames with different light conditions and then separating the diffuse component of the scene from the specular component for each LED by using

the difference between these frames.

The differential lighting pattern sequence contains 16 one-off frames for each LED (where all but one light is on in each frame), 2 half-off frames, and 1 all-on frame. I calculate two types of specular masks for further use: a specular mask for each LED based on one-off images; a specular mask for all LEDs based on half-off images to minimize the influence of eye movement during the capture sequence. The specular mask for each LED is given by subtracting two one-off images of this LED from the sum of the one-off images of the previous and the next LED and then dividing the result by two. The whole specular mask is determined by taking the maximum value of the difference between the all-on image and the half-off image and the difference between the all-on image and the other-half-off image.

With calculated specular masks, I can further detect glints on the cornea for each frame and camera (Fig. 4.7). The cornea region segmented in each camera image helps to separate specular reflection on the cornea and sclera. Only cornea glints are used for our further pose refinement, as the centroids of irregular glint shapes on the sclera are undetermined. I then use blob detection to acquire cornea glint centroids in specular masks for each LED. The cornea glint for each LED is labeled using the cornea glint centroid detected in the all-on frame to eliminate error due to eye movement during the capture sequence. One example result of labeled glints is shown in Figure 4.8

Pose Refinement with Glint Centroids

We utilize a geometric cost that adheres to the physical law of reflection. According to this law, the surface normal must be equal to the vector bisector between the point-to-camera and point-to-light vectors. Initially, we compute the point of intersection between the ray corresponding to the observed glint centroid and the cornea surface at a non-optimal surface location. Subsequently, we determine the vector pointing towards the light and calculate the vector bisector. Finally, we calcu-

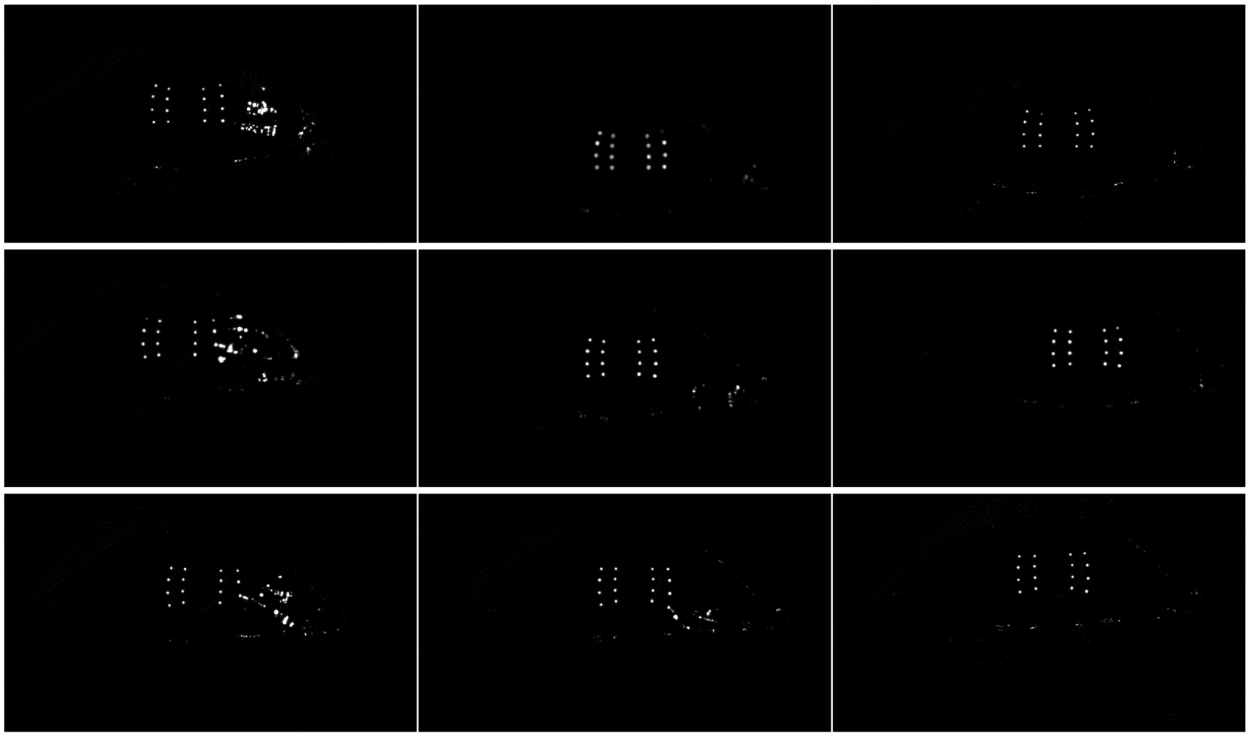

Figure 4.7: **The ground truth specular mask obtained from the differential lighting sequence.**

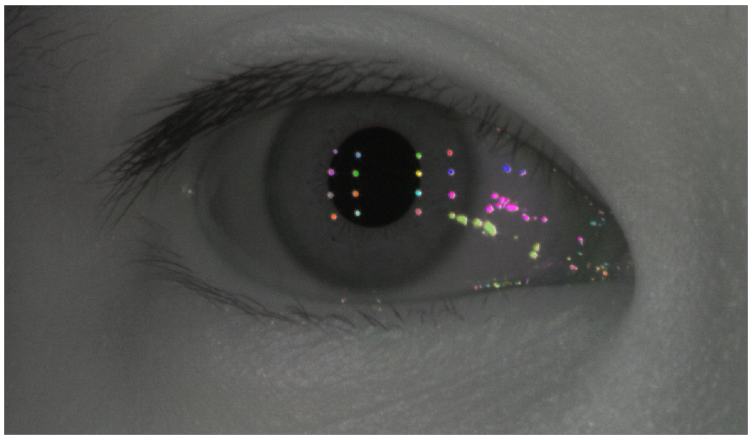

Figure 4.8: **Labeled specularities superimposed on the input image.** Each light is assigned a different color for visualization.

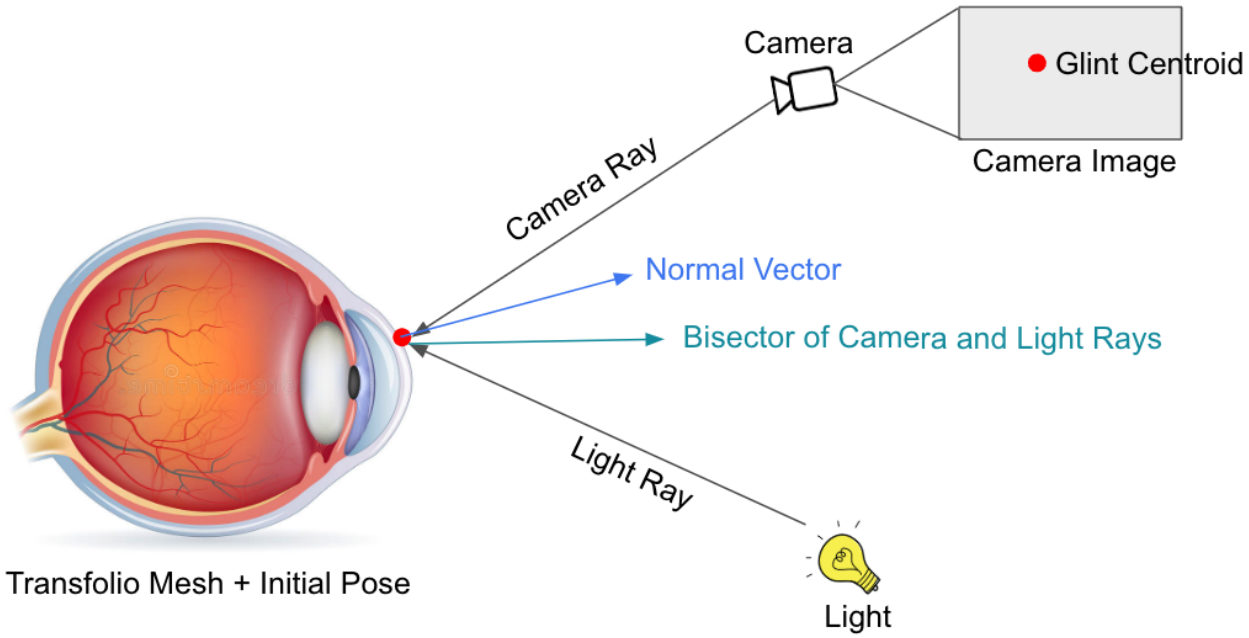

Figure 4.9: **Pose Refinement with Glint Centroids.**

late the difference between this bisector and the surface normal to obtain a loss, which represents the deviation from the optimal surface location.

As shown in Figure 4.9, to determine the glint centroids on the surface of the eyeball, I utilize a geometric cost that adheres to the physical law of reflection. According to this law, the surface normal must be equal to the vector bisector between the point-to-camera and point-to-light vectors. Initially, I compute the point of intersection between the ray corresponding to the observed glint centroid and the cornea surface at a non-optimal surface location. With an initial pose obtained from multiview gaze estimation, I use our ray tracing function based on NVIDIA OptiX package [119] to map each detected glint centroid in the camera image to a corresponding 3D surface point on the eyeball, which is defined by a specific camera ray and surface normal vector at that point. Then, connecting the 3D point with its corresponding LED position creates a light ray. However, the initial eye pose may not guarantee the matching between the normal vector and

the bisector of the camera and light rays. Therefore, I calculate the difference between the surface normal and the bisector to obtain a loss, which represents the deviation from the optimal surface location, and further optimize to minimize the difference between them for each cornea glint in all camera views. For optimization, I use the autograd function in PyTorch to back-propagate the mean squared error (MSE) loss between the normal vector and bisector, where the translation and rotation of the eyeball serve as learning parameters. During training, I employ the 6D rotation matrix [120] to enhance the rotation continuity.

4.2.2 Experiments and Discussion

To evaluate our multi-view glint-only eye tracking framework, we conduct experiments with 10 individuals. All experimental data is captured using a multi-camera multi-light capture rig, consisting of 9 cameras and 16 lights. The cameras are placed with 8 cameras in a circle surrounding a central 9th camera. All cameras are aimed at a common point 250mm in front of the central camera, and the angle between the center and outer cameras is 15 deg. The lighting consists of 16 850nm IR LEDs approximating point sources, which are placed in a 4x4 grid on the surface of a sphere of radius 275mm centered at the user's eye covering $\pm$40 deg horizontally and $\pm$30 deg vertically. The camera intrinsics and extrinsics are calibrated using traditional methods [121]. Light positions are calibrated using a reflective sphere moved through the working volume using a motorized XYZ translation stage.

Camera timing and light control are chosen to enable a burst capture of a static pose of the eye under varied but known lighting conditions. All cameras have a resolution of 2048 $\times$ 2048 and were set to their maximum framerate of 60 Hz. Cameras are synchronized using a single hardware trigger signal, and lights are synchronized with the camera exposure using the same signal. Lights are driven using a microprocessor, which could individually control LED on/off

states in a programmable repeating pattern. We capture a burst of 19 images, taking 0.32s at 60 Hz. The high-fidelity eye model we choose is from Transfolio: Eye3D™ Blue output [122], which contains an outer layer mesh for the cornea and the sclera.

Qualitative Results

We evaluated the results of our eye surface fitting qualitatively: our primary aim is to improve the ability to simulate light reflections off the surface of the cornea and sclera via rendering. Example input images from the multiview capture are given in Figure 4.6, showing the 9 camera views, cropped to just the region of the eye containing the specular reflections off cornea and sclera. In these images, the specular pixels are generally the saturated pixels, however, to more precisely separate the image into specular and diffuse components, we use the method of differential lighting which gives a specularity map for both all specular pixels in the image, as well as the per-led specular map of the specular contribution of each light. These specular masks are visualized in Figure 4.7.

The overall results of our specular renders are presented in Figure 4.10. This figure shows side-by-side comparison of the ground truth images, the extracted specular masks, and the rendered specularities for the glint-only pose refinement method.

Discussion

Based on the previous results, we can notice that the specular reflection rendered on the sclera still exhibits noticeable differences when compared to the ground truth. This discrepancy is primarily due to the imprecise sclera shape of the high-fidelity model employed in our study. Therefore, to enhance the performance of our method, an additional step that jointly refines the pose and shape of the eyeball based on appearance seems promising. To this end, I conduct initial trials utilizing

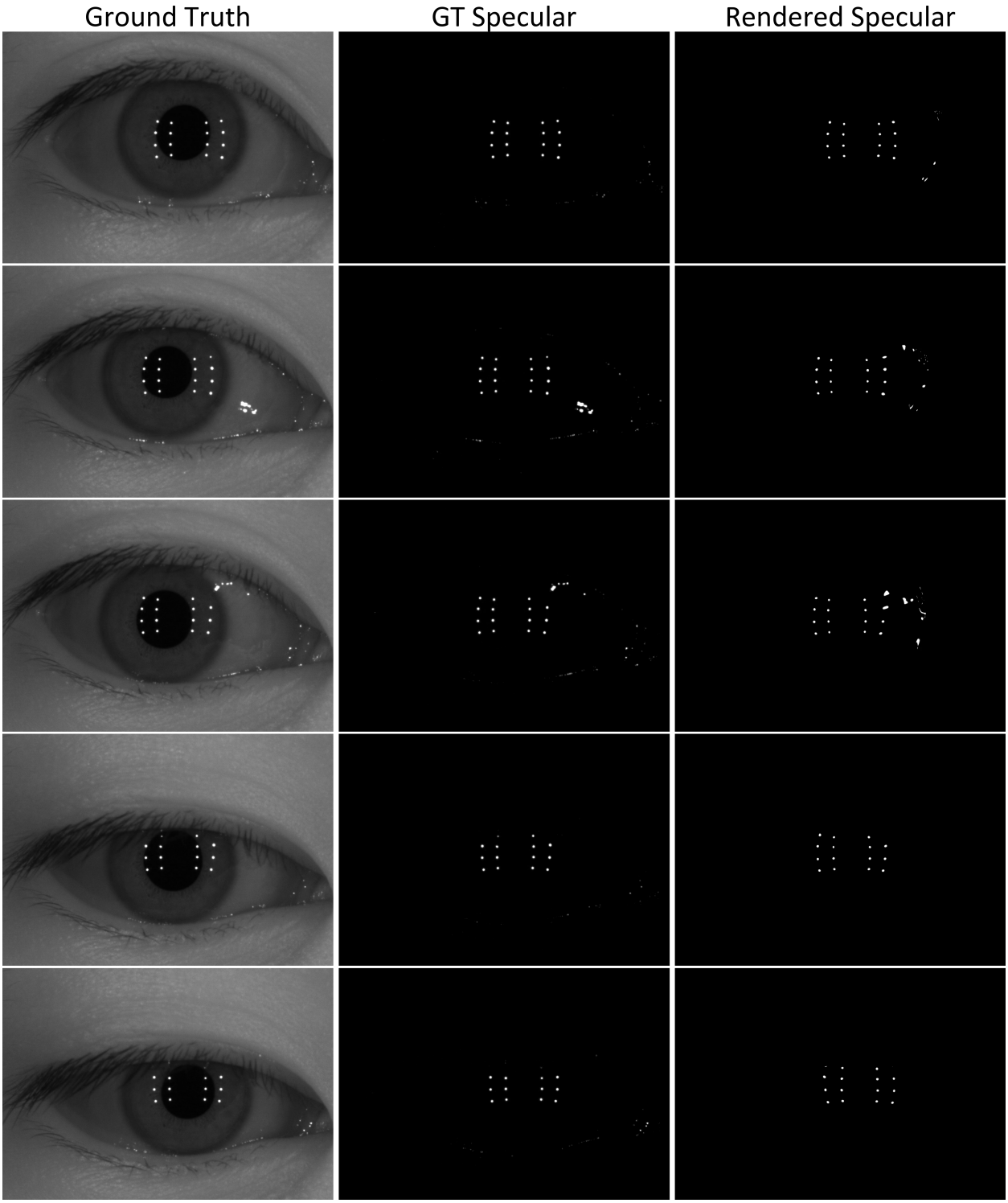

Figure 4.10: **Comparison of rendered specularity against ground truth original image and specular mask.**

the renderer function based on ray tracing [119] to generate the current appearance of the eyeball in each camera view. The texture of the entire eyeball was initially set to zero, and using the autograd function in Pytorch, the shape of the eyeball and its pose were optimized simultaneously to minimize the cost function. This cost function measures the dissimilarity between the rendered appearance images and captured camera images. However, I was not able to achieve satisfactory outcomes during the internship period, but I believe that this direction is worthwhile for further investigation.

CHAPTER 5
VOLUMETRIC MODELING FOR OIL PAINTINGS

The analysis and understanding of artworks such as paintings depend greatly on accurate measurements of their geometric shape, their internal structure and their chemical composition. In Chapter 3, I propose surface modeling techniques for artworks with diffuse and specular surfaces. This chapter, however, focuses on volumetric modeling for oil paintings, which often have complex layers and delicate materials, but their high value and uniqueness preclude anything but the sparsest sample-based measurements by traditional microtomy (thin cross-sections from paint chips or hollow-needle punctures). In the last decade, optical coherence tomography (OCT) has enabled new non-destructive measurements that gather dense, point-wise depths of layered surfaces to create 3D images with axial resolutions at micron scales. While commercial OCT systems at biologically-useful wavelengths (900 nm to 1.3 μm) can reveal some painting layers, strong scattering and absorption at these wavelengths limit their useful penetration depth. In Section 5.1, I will present an improved lower-cost time-domain OCT (TD-OCT) system for deeper, high-resolution 3D imaging of painting layers. This chapter will restate some text and figures from my previously published work [123], [124].

Conservators typically use X-ray fluorescence spectroscopy (XRF) methods to identify the pigments in oil paintings, and analysis of these data to diagnose and treat objects and investigate artists' materials [125]. XRF imaging also provides the opportunity for spatial analysis of pigment distributions based on their elemental composition, using a raster scan to acquire spectra pixel-wise across artworks. However, conventional XRF-based pigment identification relies on time-consuming elemental mapping that requires interpretation of measured spectra by experts.

Recent studies have applied machine learning techniques to cluster similar XRF spectra for data analysis to identify the most likely pigments, but implementing automatic pigment identification remains challenging, as they must include strategies to directly tackle the complex structure of real paintings, e.g. pigment mixtures and layered pigments. In addition, high noise levels in pixel-by-pixel XRF can further impede pigment identification. To solve this problem, Section 5.2 introduces a deep-learning based pigment identification framework to fully automate the process.

5.1 Time-domain Optical Coherence Tomography (OCT) for Paintings

As introduced in Section 2.2.1 OCT has enjoyed widespread commercialization for producing high-resolution 3-D images of the eye and other biological tissues [126], [127], and is, in its essence, a high-resolution 3D scanning Michelson interferometer [128]. Unlike confocal microscopy, which relies on the numerical aperture of the microscope objective lens for depth resolution, OCT employs the principle of low coherence interferometry to achieve depth resolution determined by the bandwidth of the source spectrum. Recent advances now allow OCT systems to achieve sub-micron resolutions through the use of specialized ultra-wideband coherent light sources [129], such as super-luminescent diodes (SLDs), Kerr lens mode-locked lasers, and supercontinuum laser sources. These sources have wide bandwidth and short coherence lengths, which contributes to the high depth resolution achieved by OCT.

Compared to confocal microscopy, OCT produces images with a higher range of intensities by measuring only interference signals instead of overall reflected light amounts. Additionally, as illustrated in Section 2.2.1, OCT amplifies the weak signal from the sample arm by combining it with the strong signal from the reference arm. For measurements in highly scattering media such as paint layers, OCT provides approximately double the penetration depth of confocal microscopy [130]. Earlier research on OCT for 3D retinal imaging directed great attention to

non-contact measurements and to avoiding tissue damage, and hence these OCT systems perform high-resolution scanning at comfortable working distances, typically just a few centimeters. This feature is also enables safe scanning of valuable artworks.

Therefore, OCT has become attractive in the field of cultural heritage due to its ability to probe the hidden layers of artworks in a non-invasive and non-destructive way [130], [131]. While OCT systems with the central wavelengths from 900 nm to 1.3 μm are commercially available for biomedical uses, a central wavelength at 2.2 μm is more effective in detecting pigment layers due to its larger penetration depth [132]. Liang et al. demonstrated the effectiveness of a custom-built 2.2 μm Fourier domain OCT (FD-OCT) system in improving the imaging depth through oil-paint pigment layers such as titanium white, cobalt blue, and yellow ochre [133]. However, this FD-OCT system included a customized high-power laser source and a costly broadband spectroscopic camera. To make the system more accessible to a broader user community in the field of cultural heritage, I propose an inexpensive alternative assembled from commercially-available parts; a 2 μm time domain OCT (TD-OCT) system with comparable resolution (axial resolution of 4.85 μm), signal quality, and penetration depth. The system reduces cost and equipment complexity substantially, but does require a modest increases in measurement time.

Our 2 μm TD-OCT system incorporates a commercial Q-switch supercontinuum source [134] with power density of 0.04 mW/nm over the wavelength of 1.8 to 2.2 μm, a 2x2 fiber coupler, and an InGaAs photodiode detector to reduce the cost. To compensate for the relatively weak laser intensity, I implement a dynamic focus method using a voice-coil translation stage to move the sample arm. This approach enhances the interference amplitude and improves the signal-to-noise ratio (SNR). I fully automate the measurement, control, synchronization, evaluation, and display of data using Python. Users can easily initiate a scanning process without requiring any further guidance.

5.1.1 Dynamic focus time domain OCT

In conventional TD-OCT, the generation of a depth scan (axial or "A-scan") relies on the mechanical movement of the reference mirror. However, this method presents a challenge when scanning at deeper layers, as the coherence gate (the single depth at which returned light causes interference) becomes out of focus, leading to low returned light intensity. This issue becomes more pronounced when using a low-intensity laser, as the interference amplitude may be insufficient for effective detection. To address this problem, I devised a dynamic focus TD-OCT setup to enhance the reflective intensity, which I later discover is similar to the method proposed by Hughes et al.[135]. The implementation is simple; instead of moving the reference mirror with the voice-coil translation stage, I keep the mirror fixed and move the sample arm assembly, requiring no additional optical parts. By matching the reference path length with the focus point of the scanning lens, the coherence gate remains fixed at or near the beam focus depth throughout the imaging process (Fig. 5.1(a)). Conventional OCT yields low reflective intensity in deeper layers due to keeping focus fixed on the painting's top surface, with interior layers being out of focus. In dynamic focus OCT, layers that match the path-length are in focus and possess high reflective intensity. Figure 5.1(b) displays the differences in setup between conventional and dynamic focus TD-OCT.

The design of the 2 μm TD-OCT system

We design a fiber-based OCT system to replace the mechanically-vulnerable free-space beam-splitter shown in Figure 2.6. Our system integrates easily with a near-infrared light source, and is more reliable due to the replacement of the free-space beam-splitter. A low-cost supercontinuum white light laser (NKT SuperK COMPACT) generates light over a wide range of wavelengths, 450-2400 nm, filtered by a bandpass filter (Thorlabs FB2000-500) centered at 2000 nm with 500

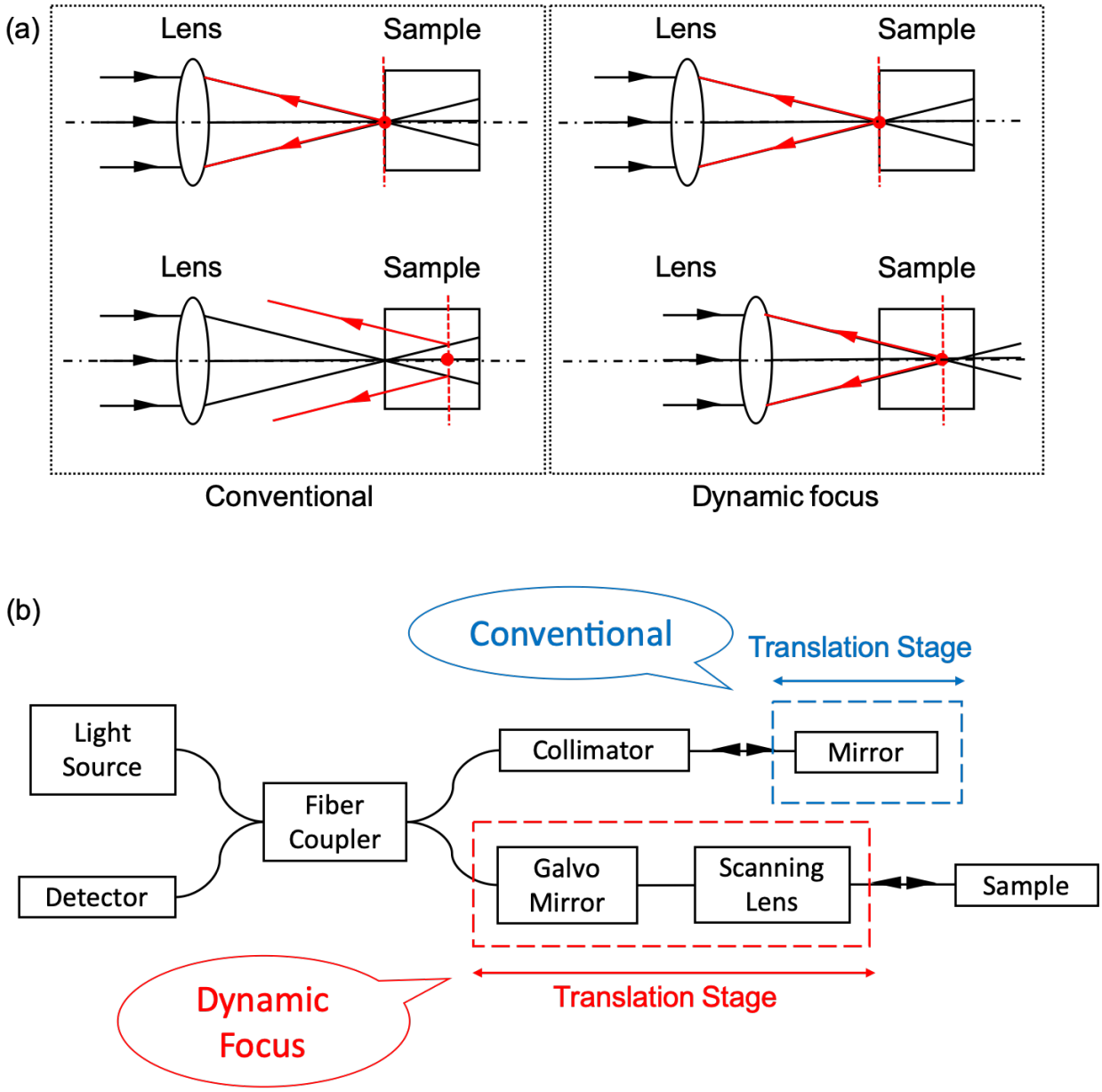

Figure 5.1: **Dynamic focus OCT.** (a) Schematic Diagram of light paths in conventional and dynamic focus OCT. The red spot and dashed line represent coherence gate. (b) Conventional and dynamic focus OCT apparatus. The blue region depicts conventional OCT, with reference mirror moved by the translation stage; the red region depicts dynamic focus OCT with fixed reference mirror and sample arm moved by the translation stage with galvanometer mirror and scanning lens attached.

nm bandwidth. A collimator couples the free-space light beam source into a 50:50 fiber coupler (2000$\pm$200 nm, Thorlabs TW2000R5A2A); the fiber coupler further splits the beam into the reference and sample arms. The scanner on the sample arm consists of a 2D galvanometer system (Thorlabs GVS002) and a scanning lens (Thorlabs SL50-3P) for lateral scan. By putting the scanner on a translation stage (Zaber X-DMQ12P-DE52), the scanning lens can focus at different depths of the sample for the axial scan. The reference arm contains an angular mirror and a neutral density filter to match the beam intensity of the two arms. The fiber coupler further combines the beam from the two arms into the detector (Thorlabs PDA10DT) that records the interference pattern. A multifunction Data Acquisition (DAQ) module (NI USB6003) controls the galvanometer system and captures detector data for use in the MATLAB host computer.

The supercontinuum laser is a pulsed source. Since the time width of each pulse is less than 200 ns, a sampling frequency larger than 10 MHz is necessary to capture each pulse. To decrease the cost of the data acquisition equipment, I use the detector to integrate the signal: the output signal from the detector becomes continuously constant with a bandwidth setting of 500 Hz and a gain setting of 70 dB. This setting enables coherence signal to be obtained without capturing individual pulse peaks by high sample rate data acquisition equipment. Given these parameters, the signal-to-noise ratio (SNR) is $\sim$75 dB with no sample.

In-line with the reference arm, an adjustable neutral density filter helps produce high SNR values, since the amplitude of the interference pattern achieves the maximum when the reflective intensities from the two arms are identical. I also find improvements by using a 3-sided, right-angle corner-reflector or ''retroreflector" as the fixed reference mirror, as it requires little or no alignment with the fiber optic coupler to return maximum light power, and is largely insensitive to environmental vibrations.

On the sample arm, the galvanometer scanning mirrors and the scanning lens are attached to

the translation stage. For the following experiment, the lateral resolution of the cross-sectional B-scan is 100 pixels/mm. The velocity of the stage is 1 mm/s; the total scanning depth is 2.5 mm.

I wrote a Python application to control both the DAQ and the voice coil stage simultaneously. The DAQ reads the signal from the detector and generates the analog voltage that controls the galvanometer system. In our TD-OCT system, capturing well-aligned B-scan data requires precise synchronization between the detector data acquisition and the translation stage positioning commands. By setting the Windows 10 process priority to its maximum ("real-time priority"), I limit unpredictable delays to less than 100 nS, which ensures good results.

5.1.2 Experiments and Results

Preparation and characterization of the paint layer

To test our system, we mixed Indigo dye ($C16H10N2O2$, KREMER, genuine) and linseed oil (GAMBLIN, cold pressed) to form a thin layer of paint on a microscope slide. Subsequently, we heated the slide at 60 °C for one week to dry the paint layer. We then characterized the freshly dried paint layer using our lab-built 2 μm TD-OCT. To compare our results, we also characterize the same slide using a commercial SD-OCT with a central wavelength at 900 nm (Thorlabs GAN220, bandwidth 150 nm). Furthermore, we measure the physical thickness of the paint layer using a Stylus Profilometer (Dektak 150) to establish a ground truth for comparison.

Resolution of the 2 μm TD-OCT

Our laser light has a fixed but non-Gaussian spectrum, which produces a overly- broad and non-Gaussian interference pattern at our detector. To restore the Gaussian pattern shape and reduce this broadening, I applied the deconvolution method of Bash [136]. To measure the axial resolution of the 2 μm TD-OCT system, we use a microscope glass slide with known thickness as the reference

sample. The results (shown in Fig. 5.2) indicate that our OCT system achieves a resolution of $\sim$4.85 μm, which is comparable to the theoretical resolution of 4.41 μm.

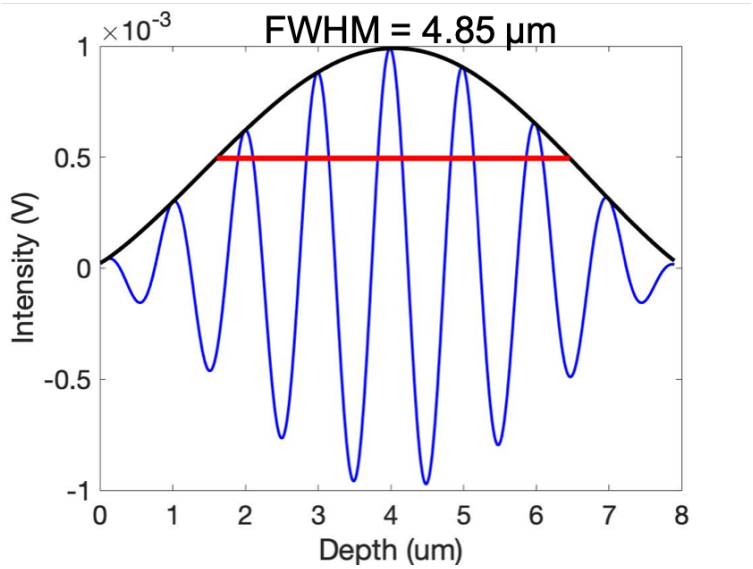

Figure 5.2: **Zoom-in interference pattern at the air/glass interface.** The FWHM is 4.85 μm. The axial resolution of our system achieves a high resolution of $\sim$4.85 μm at the air/glass boundary.

Comparison with the commercial 900 nm FD-OCT

Figure 5.3 presents a comparison of the imaging results obtained with our 2 μm TD-OCT and a commercial 900 nm FD-OCT for a layer of indigo-linseed-oil on a glass slide, demonstrating the improved penetration depth of long wavelength OCT. At 900 nm, the strong scattering of the indigo layer prevents the visualization of the indigo/glass interface, which is clearly visible in our 2 μm measurements(top of Fig. 5.3a). Moreover, the 2 μm result allows us to observe the bottom of the glass slide, confirming the penetration of the indigo layer and the ability to measure multiple underlying layers. The OCT image accurately matched the physical thickness of the indigo layer (Fig. 5.3c), and thus validated the results of our 2 μm OCT. Furthermore, the 2 μm laser demonstrated better penetration ability than the 900 nm laser for several other pigments, such

as yellow ochre (iron oxyhydroxide) and alizarin crimson (1,2-dihydroxyantraquinone).

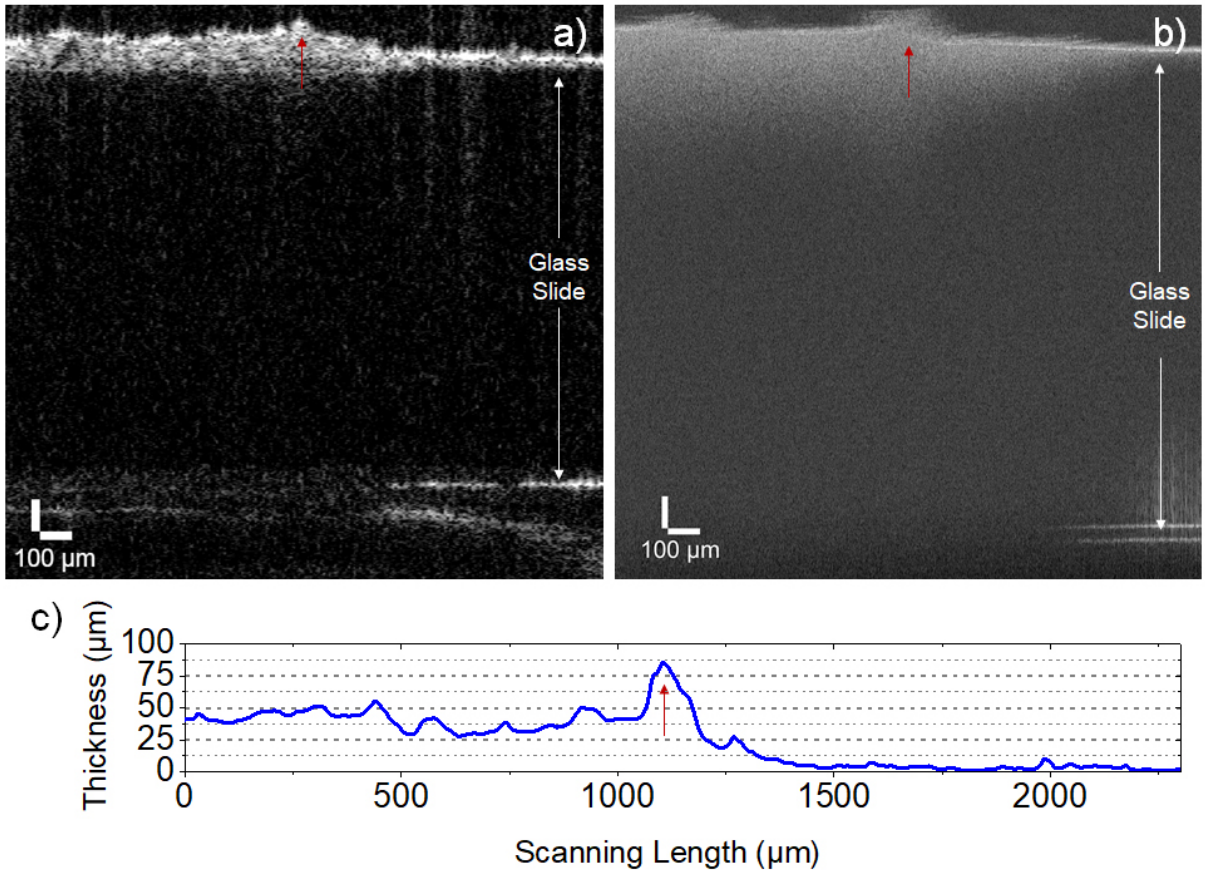

Figure 5.3: **Scanning results of a layer of indigo-linseed-oil on a glass slide** by (a) lab-built 2 μm TD-OCT and (b) commercial FD-OCT with a central wavelength of 900 nm. (c) The physical thickness of the paint film measured by a profilometer.

5.1.3 Discussion

This dynamic-focus time domain OCT system at 2 μm provides resolution and penetration depth suitable for cultural heritage applications. Compared with the commercial OCT systems, a central wavelength at 2 μm improves the depth of penetration for pigments that are opaque in shorter wavelengths. In addition, a broad bandwidth of 400 nm enables an axial resolution of 4.85 μm.

Our system applies the dynamic focus setup by placing the translation stage under the galvanometer mirror and scanning lens in the sample arm instead of reference arm. It has greatly improved the system SNR by maintaining the coherence gate near the beam focus throughout the imaging process. All parts in the system are inexpensive commercial off-the-shelf (COTS) equipment.

However, there is still room for further improvement. I use fibers that do not preserve or maintain the polarization supplied by our laser, better polarization control and matching may further improve system SNR. In addition, the attenuator for matching the intensities of the two arms has been adjusted only at the start of each experiment and kept fixed during measurement, despite weaker signal returns for deeper layers. Dynamically adjusting the attenuator may improve the SNR in the scanning process.

Moreover, physical movement of the reference/sample arm limits the speed of time domain OCT systems. To overcome this limitation, I conducted initial trials to develop a low-cost Fourier domain OCT by replacing the expensive swept-source laser with a low-cost super-continuum laser in our time domain system, along with a digital micro-mirror device (DMD). However, for over a year I faced technical difficulties with microcontroller software in the DMD device and with spectrum wavelength calibration, preventing us from further developing this approach. Nonetheless, I believe that this is a promising avenue for developing low-cost Fourier domain OCT.

Despite these limitations, we demonstrated that our OCT system is experimentally capable of producing high-resolution high-penetration scanning results of paintings. This simple and inexpensive OCT system can be widely used in cultural heritage area. Moreover, the scanning results of our system can be fused with pigment identification methods in the next section for further studies.

5.2 Automatic Identification of Layered Pigments from XRF Data

X-ray fluorescence spectroscopy (XRF) is a well-established workhorse technique for elemental analysis in a wide range of scientific fields [137], such as geochemistry [138]–[140], forensic science [141], [142] and archaeology [143]. Few areas of research can obtain more beneficial use than the investigation of cultural heritage, as it often necessitates *in-situ* investigations that take place under ambient conditions as is offered by open-architecture and hand-held versions of the XRF instrument. To fully characterize the heterogeneous nature and complex history of many artworks, XRF often requires a high number of measurements that are best implemented in the form of raster-based imaging. To form an XRF image, the instrument moves across the surface of an object, such as a painting, while collecting spectra point-by-point that are spatially redressed to their 2-D locations [144]. These forms of XRF imaging allow for spatial analysis of pigment distributions based on their elemental composition.

Conventional XRF-based pigment identification uses spectrum evaluation methods [145] to generate elemental maps, which existing XRF analysis software, for example, PyMCA [146], can fully support. However, identifying the pigments that cause these elemental maps and their spectra requires input from experts that have prior knowledge of the painting technique. Moreover, many artworks consist of varying pigment mixtures layered in complicated stratigraphies [147]–[149]. To assist the manual work of evaluating spectra, machine learning techniques have recently been applied to pigment identification by clustering pigment-related spectral features [150]. For example, XRFast, an open-source unsupervised sparse dictionary learning algorithm developed recently by our group, finds maps of correlated elements to help in pigment identification, which is an improvement over the traditional approach that calculates image correlations [151]. Today, deep learning (DL) has been widely applied to assist with XRF analysis and has the potential to perform

fully automatic identification of elements and their sources. For example, Shugar et al. identified 48 different wood species with the XRF dataset using convolutional neural networks, reaching an accuracy of 99% [152]. They found the range between 0.7-1.7 keV the most important portion of the spectra for wood classification, which covers the elements calcium, aluminum and magnesium. Moreover, Kim et al. applied a neural network on micro XRF data to generate mineral maps on natural rocks. They showed that DL was a good way to improve the description of mineral reactivity to rock samples of different origins, sizes, and thickness [153]. Most recently, Cerys et al. proposed a deep-learning-based method to directly identify pigments from XRF spectra [154]. By training a convolutional neural network, they classified XRF spectra into one of the 15 pigment classes with high accuracy, but claimed that it was still challenging to apply the model to more complex scenarios, including layered pigments and pigment mixtures. Therefore, focusing on the complicated stratigraphies of real paintings, this work builds on these previous studies by proposing a convolutional neural network to automatically identify pigment mixtures in layered structures and to display 2D pigment maps based on the probability of their occurrence.

Here I propose an end-to-end pigment identification framework, including pigment library creation, XRF spectra simulation, mock-up preparation, a pigment identification DL model, and 2D pigment map generation. As a case study, I apply our framework to a 19th-century painting, Paul Gauguin's *Poèmes Barbares* (1896), focusing on a set of 19-century pigments previously identified in this painting [155]. In addition, previous analysis revealed a hidden painting beneath the surface, which tremendously increased the difficulty of pigment interpretation from XRF spectra. Therefore, our DL model targets pigment identification in the multilayered matrices of the painting. By training the DL model using 16224 simulated XRF spectra of three-layered pigments, followed by finetuning of the model using 20% of the experimental XRF spectra (i.e., 1320 XRF spectra) from mock-ups, the DL model demonstrates satisfying performance of pigment identification on

the mock-ups as well as the painting *Poèmes Barbares*. In particular, a high sensitivity toward identifying pigments present in low concentrations is shown.

To further demonstrate the applicability of this approach, I apply the finetuned model to Paul Cezanne's *The Bathers* (1899-1904). This is a single-layered painting created from a comparable time period to *Poèmes Barbares*, composed of similar but fewer pigments. The DL model achieved high probabilities in identifying the pigments in *The Bathers*, suggesting our model's generalizability and stability.

In all, our framework provides an automatic and quick pigment identification strategy based on non-invasive XRF imaging, in particular targeting the paintings' complex layered structure to the XRF response. The trained model does not require expertise or extensive familiarity working with XRF and pigments, but directly answers where the pigment might exist. Although the type of pigments are limited to those trained in the current work, our framework shows great potential for extension to other types of pigments and paintings, as well as XRF-based identification problems in the fields beyond cultural heritage.

5.2.1 Dataset and Preparation

We gather the XRF data from three sources to build, train, and test our framework: from existing paintings, from oil-paint "mock-ups" of crossed paint stripes of different pigment mixtures, and the spectra of multilayered pigments simulated by PyMCA.

We selected Paul Gaugin's *Poèmes Barbares* (1896) to test our pigment identification approach. This painting is representative of the challenging pigment identification tasks present in many 19th-century paintings, which often involve a plethora of pigments made available after the industrial revolution. In addition, Marc Vermeulen et al. [155] previously studied the types of pigments and the structure of the paint layers of *Poèmes Barbares* by combining XRF, reflectance imaging

spectroscopy, and cross-section analysis, providing us with extensive and reliable ground-truth measurements of this historical work.

These previous studies show that *Poèmes Barbares* consists of a visible (top) painting and a hidden (bottom) painting, each constructed with multiple paint layers and various mixtures of pigments [155]. To accommodate this complex layer structure for the DL approach, I simplify it into a three-layer structure: one top pigment layer, one bottom pigment layer, and one ground layer. Referring to the chemical analysis of multiple cross-sections of the painting, we chose 11 distinct pigments of interest (Table 5.1) and calcium carbonate as the ground layer.

Next, we made our own set of layered oil paint mock-ups to better measure how the layers affect the pigments' XRF spectra. For these layered paintings (Figure 5.4), we choose a range of pigment fractions, binder ratios, and layer thicknesses as related to *Poèmes Barbares*. We then create sets of 3-layer mock-ups consisting of crossed strips of paint with these values and prepare them for XRF measurement, generating 6605 XRF spectra including 64 pigment layer structures. Unfortunately, the experimental data causes over-fitting of the deep learning model due to the lack of variation in the pigment layer structure. To tackle this problem, I further generate a simulation dataset for these same pigments and thicknesses to train and validate the deep learning model.

Experimental Dataset

The experimental dataset contains XRF spectra of mock-ups with known pigment layer structures to train and test the deep learning model. We prepare the three-layered mock-ups with various combinations of pigments: six pigment mixtures as the bottom paint layer (mainly varying in the pigments' mass fractions, providing 16 bottom layers in total) and four mixtures as the top paint layer, making up 64 different layer structures (Table 5.2). Figure 5.4 illustrates the structure of our mock-up samples: the first and second layers both contained multiple strips of pigment layers,

where the combination of pigments in each strip was selected from Table 5.2; the third layer is a fixed ground layer of calcium carbonate.

We used a tape casting coater (model MSK-AFA-HC100, MTI Corporation (Richmond, CA)) to deposit the paint layers sequentially with precisely adjusted thicknesses, and paint a new layer after the previous layer was completely dry. In all mock-ups, I apply a calcium carbonate ground layer of 150 – 200 μm thickness to match that of *Poèmes Barbares*. As for the pigment layers, the type of mixtures simulates the palette of both the bottom and top painting in *Poèmes Barbares*. To include the XRF effects commonly found in layered systems and pigment mixtures, such as shielding and matrix effects, I varied the pigment fractions in the bottom paint layers (e.g. Bottom 1A, Bottom 1B, Bottom 1C in Table 5.2), while the top paint layer differs only in their layer thicknesses (30 – 200 μm).

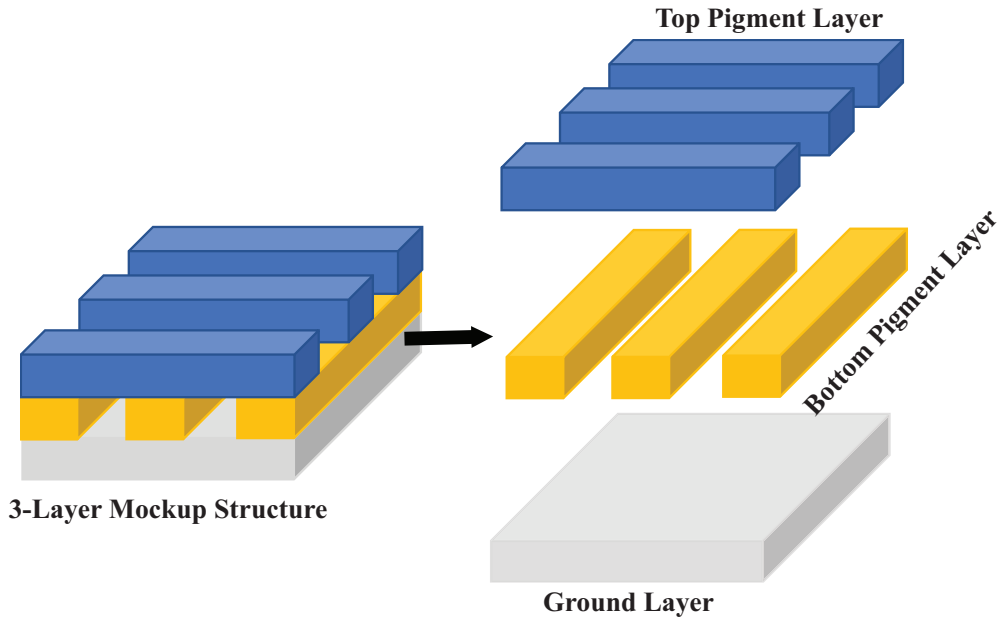

Figure 5.4: **The structure of mock-up samples.** The first and second layers both contain multiple strips of pigment layers, each strip with its pigment combination selected from Table 5.1; the third layer is a ground layer consisting of calcium carbonate.

In preparing the mock-ups, we hand-ground commercial pigments and the binder for 10 min to obtain a uniform mixture [156]. Since the organic binder would not significantly affect the XRF signal, we use Galkyd Lite (Gamblin Artists Colors (Portland, OR)) for its fast-drying property. A pigment-to-binder ratio (w%:w%) of 3:1 was applied to achieve the mobility required by the tape-casting coater. For lead-containing paint mixtures which appeared dryer, we added drops of Gamsol odorless mineral spirits (Gamblin Artists Colors) to further dilute the mixture. All pigment mixtures were deposited on pH-neutral art boards (Crescent Cardboard Company LLC, Wheeling, IL).

To collect the experimental XRF dataset, we scan the mock-ups with the XGLab ELIO XRF imaging spectrometer system and then acquire the XRF spectra of the mock-ups at 40 kV and 40 μA. We set the acquisition time at 10.0 seconds per point to increase the signal-to-noise ratio required for deep learning, and raster scan the mock-ups using a 100 x 100 mm motorized X-Y linear stage mount (Zaber T-LSM100A) with a step size of 1 x 1 mm.

Simulation Dataset

To ensure a sufficient dataset size for training our DL model, I generated a simulation dataset of 16224 XRF spectra in total by calculating the spectra using the `matrixSpectrum` function in `PyMca5.PyMcaGui.physics.xrf.McaAdvancedFit` of the PyMCA Python package [146]. I extract these functions from the PyMCA graphical user interface (GUI) and run the code in a Jupyter Notebook running Python. You can find basic functionality and use on the NU-ACCESS Github (https://github.com/NU-ACCESS). Based on the fundamental parameter approach [157], the `matrixSpectrum` function simulates XRF spectra for multilayer samples. In generating the simulation dataset, I apply a three-layer structure (top pigment layer - bottom pigment layer - ground layer) similar with the mock-up paintings. Each of the top and the bottom layers consists

Table 5.1: Pigment library for selected pigments.

Index	Pigment[a]	Chemical Formula
1	Calcium carbonate	$CaCO_3$
2	Chrome oxide green (CrG)	Cr_2O_3
3	Chrome yellow (CrY)	$PbCrO_4$
4	Cobalt blue (CB)	$CoO{\cdot}Al_2O_3$
5	Emerald green (EG)[b]	$Cu(CH_3COO)_2{\cdot}3Cu(AsO_2)_2$
6	Iron oxide (IO)	Fe_2O_3
7	Lead white (LW)	$2PbCO_3{\cdot}Pb(OH)_2$
8	Prussian blue (PB)	$Fe[Fe_2(CN)_6]_3$
9	Red lead (RL)	Pb_3O_4
10	Carmine (CM)[c]	SnO_2
11	Vermilion (VM)	HgS
12	Zinc white (ZW)	ZnO

a Calcium carbonate, chrome oxide green, cobalt blue, verdigris, iron oxide red (120 M), lead white, Prussian blue, and vermilion are purchased from Kremer Pigmente (New York, NY). Chrome yellow and red lead are purchased from Rublev Colours (Willits, CA). Zinc oxide is obtained from Gamblin Artists Colors (Portland, OR). Sodium arsenite ($\geq 90\%$) and tin oxide (99.99%) were obtained from Sigma Aldrich. *b* Due to the current unavailability of commercial emerald green pigment, we mix the copper carbonate pigment verdigris (VG, $\mathrm{Cu(CH_3COO)_2{\cdot}2Cu(OH)_2}$) and sodium arsenite (SA, $\mathrm{NaAsO_2}$) to approximate the XRF signal of emerald green, in which the Cu-As mass ratio was set accordingly. *c* Since the chemical analysis suggests that tin oxide ($\mathrm{SnO_2}$) is the support of carmine, we solely use the $\mathrm{SnO_2}$ powder in preparing the mock-ups to represent carmine in the XRF dataset.

Table 5.2: Summary of pigment mixtures in the mock-ups.

Pigment Mixture	Compound[a]	Mass fraction (%)
Top 1	LW, CM, PB, CB	55, 10, 25, 10
Top 2	VM, CM, CB	30, 30, 40
Top 3	VM, CrY, IO, VG, SA	39, 6, 39, 6, 10
Top 4	VM	100
Bottom 1A	ZW, LW, RL, VM	10, 10, 10, 70
Bottom 1B	ZW, LW, RL, VM	10, 20, 50, 20
Bottom 1C	ZW, LW, RL, VM	10, 50, 20, 20
Bottom 2A	ZW, LW, VM	10, 85, 5
Bottom 2B	ZW, LW, VM	10.7, 88.7, 0.6
Bottom 2C	ZW, LW, VM	10, 50, 40
Bottom 3A	ZW, LW, CB, CrG	10, 15, 15, 60
Bottom 3B	ZW, LW, CB, CrG	10, 15, 60, 15
Bottom 3C	ZW, LW, CB, CrG	10, 60, 15, 15
Bottom 4A	ZW, LW, CrY, VG, SA	10, 45, 15, 12, 18
Bottom 4B	ZW, LW, CrY, VG, SA	10, 30, 30, 12, 18
Bottom 4C	ZW, LW, CrY, VG, SA	10, 15, 15, 24, 36
Bottom 5A	ZW, VM, CrY, VG, SA	10, 5, 5, 32, 48
Bottom 5B	ZW, VM, CrY, VG, SA	78, 6, 6, 4, 6
Bottom 5C	ZW, VM, CrY, VG, SA	10, 15, 15, 24, 36
Bottom 6	VM	100

a The pigments and their corresponding abbreviations used in this manuscript are listed as below: CB: cobalt blue, CrG: chrome oxide green, CM: carmine, CrY: chrome yellow, EG: emerald green, IO: iron oxide, LW: lead white, PB: Prussian blue, RL: red lead, SA: sodium arsenite, VG:verdigris, VM: vermilion, ZW: zinc white.

of a single pigment from the pigment library (Table 5.1), with a layer thickness of 50 – 200 μm (10 μm interval) and 100 – 150 μm (10 μm interval), respectively.

Data Preprocessing

Before feeding the experimental and simulation datasets into the DL model, I conduct several preprocessing steps. First, based on preliminary ablation studies, the overlaps between the sulfur-K lines (2.31 keV) and the lead-M (2.34 keV) or mercury-M lines (2.20 keV) confuses the DL model in distinguishing these elements. As a result, I optimized both the experimental and simulation datasets to start at 2.80 keV to improve the performance of the DL model.

Also, the simulated XRF spectra lack the underlying spectral background signal caused by X-ray scattering and the equipment properties. Therefore, I estimate this spectral background from the mean of the measured spectra using Statistics-sensitive Non-linear Iterative Peak-clipping (SNIP) algorithm [158], which contains a low statistics digital filter and a multipass peak clipping loop. Then I add the estimated background to the simulation dataset to better mimic their experimental counterpart, without significantly changing the key elemental peaks.

Since the XRF sensitivity to different elements varies, both the experimental and simulated spectra exhibit skewness of more than one order of magnitude with generally comparable elemental concentrations. To reduce this skewness, we apply a log-log square root transformation (Eq.5.1) of the original spectrum X followed by normalization to X'. This normalization step further enhance the recognition of the elements with low peak intensities in the spectrum. X' is the final input features to the DL model.

$$X' = \log_e(1 + \log_e(1 + \sqrt{X})), \tag{5.1}$$

5.2.2 Deep Learning Model for Pigment Identification

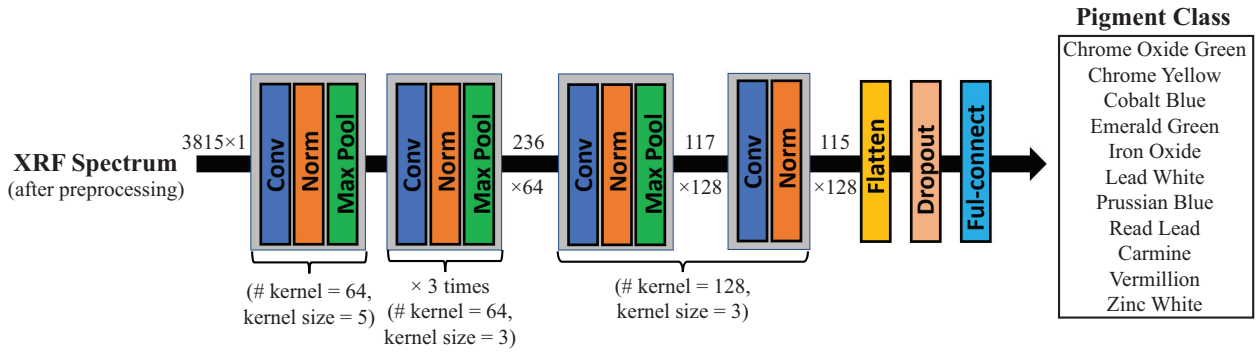

Figure 5.5: **The Architecture of the Deep Learning Model.** It contained 1D convolutional layers (Conv), normalization layers (Norm), max pooling layers (Max Pool), a flatten layer, a dropout layer and a fully-connected layer. The input size of each XRF spectrum was 3815×1. The output predictions were in 11 classes (one per pigment). # kernel and kernel size stood for the number and the size of the kernel at the corresponding convolutional layer, respectively.

The model consists of 5 convolutional blocks, where each block contains a 1D convolutional layer, an activation function LeakyReLU [159], a batch normalization layer, and a max-pooling layer, as shown in Fig. 5.5. According to preliminary ablation studies, we set the number and size of the kernels of each 1D convolutional layer at 64,64,64,64,128 and 5,3,3,3,3, respectively. The activation function LeakyReLU, i.e., $f(x) = max(0.01x, x)$, keeps the positive part of its input while preventing "the dead ReLU issue" by using a small value when the input was negative [159]. The batch normalization layer sped up the training process by distributing the input for every layer around the same mean and standard deviation. The max-pooling layer downsamples the dimension of the input to half. Finally, the model ends with a post-convolutional layer with 128 kernels with a size of 3, a normalization layer, a dropout layer with a rate of 0.25, and one fully-connected layer of 11 classes, outputting the probabilities of each class (each pigment) between 0 to 1. The output layer uses a sigmoid activation function $\sigma(z_j) = \frac{1}{1+e^{-z_j}}$, where z_j is the predicted score from the

model of each class. The probabilities as predicted further build the pigment maps of the paintings. We utilize two datasets, the simulation and experimental datasets, to train the model and compare the results. The two training processes apply the same architecture as described but differ in the initial weights. The model trained with the simulation dataset uses randomly initialized weights. We then apply it as a pre-trained model, using its weights as the initial value, and further finetune it with the experimental dataset using Transfer Learning (TL) [160].

TL is a popular technique that uses the pre-trained weights from an initial model as the starting point on another model, which reduces or eliminates the risk of overfitting and allows for better training speed and model performance [161]. Accordingly we pre-train the model on the simulation dataset and then refine it using the experimental dataset. Specifically, when training with the experimental dataset, we first fix (i.e., untrain) the pre-trained weights from the simulation dataset in all convolutional layers, whereas only finetune the fully connected layers with the targeted dataset. Next, we train all layers together and further finetune the model with the experimental dataset. Section 5.2.3 will compare the model performances with and without finetuning to show the effectiveness of TL.

Moreover, we calculate the loss to optimize the performance of the model by averaging the binary cross entropy of each predicted class, as defined in Eq.5.2:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K}[\hat{z}_j \cdot log(z_j) + (1-\hat{z}_j)\cdot log(1-z_j)], \tag{5.2}$$

where $\hat{z}$ was the ground truth label, z was the score predicted from the model for each class, K was the number of the class, and N was the number of batch size.

Training Strategy

We complete the training process in two steps. We first pre-train the model proposed in Fig. 5.5 on the simulation dataset by splitting off a total of 16224 simulated XRF spectra into a testing dataset and a training dataset with a ratio of 1:4. In training the model, we apply the five-fold cross-validation [162] by further dividing the training dataset into '5' groups of equal size and iteratively selecting one group as the validation set, while the rest remaining as the training set. Therefore, with the full iteration, we evaluate the model's performance by the testing dataset five times. After preprocessing all cohorts of the datasets following Section 5.2.1, we train the model with the Adam optimizer [163], due to its robustness, less convergence time and fewer parameters for tuning, with an initial learning rate of 0.001. The batch size is 64, and on average, it takes about 0.95 hours for each fold in the 5-fold cross-validation for 150 epochs with early stopping settings. Next, the experimental dataset has 6604 XRF spectra from the two-layer pigment areas, which I manually pick from all mock-ups with known ground truth. To confirm the effect of finetuning on pigment identification, we test the model before and after finetuning by the experimental dataset. Specifically, we use all experimental data as the testing dataset to test the model before finetuning, which we train with only the simulation dataset. Then we initialize the model after finetuning with the weights that performed the best among the five-fold cross-validation and further finetune it with 20% of the randomly selected data from the experimental dataset, i.e. 1320 XRF spectra. The testing dataset is the remaining 80% (5284 XRF spectra). Similar to the training process of the model trained with the simulation dataset, we choose Adam optimizer with a lower learning rate of 0.0005 to train the finetuned model. The batch size is 64, and it takes an average of 0.2 hours for each of the 5 groups or "folds". All training and testing processes run on an NVIDIA GeForce RTX 2070 GPU using Tensorflow 2.0 in Python 3.7.

5.2.3 Results and Model Performance

Table 5.3: The classification results for each pigment class among the models trained from three different datasets: the simulation dataset, the experimental dataset without finetuning, and the experimental dataset acquired from the mock-ups with finetuning[a].

Pigment Class	**Simulation**			**Experimental (no finetune)**			**Experimental (finetune)**		
	Accuracy	Sensitivity	F1	Accuracy	Sensitivity	F1	Accuracy	Sensitivity	F1
Cobalt Blue	0.950	0.997	0.782	0.870	0.973	0.890	0.899	0.985	0.916
Emerald Green	0.964	1.0	0.861	0.773	0.821	0.820	0.859	0.870	0.871
Iron Oxide	0.882	0.732	0.687	0.659	0.506	0.550	0.998	0.995	0.997
Prussian Blue	0.878	0.572	0.592	0.869	0.771	0.746	0.996	0.994	0.993
Carmine	1.0	1.0	1.0	0.653	0.664	0.685	0.995	0.992	0.995
Vermilion	0.987	0.997	0.955	0.918	0.994	0.952	0.947	0.978	0.970
Zinc White	0.952	1.0	0.813	0.916	0.918	0.956	0.916	0.953	0.954
Chrome Yellow	0.933	0.981	0.708	0.631	0.642	0.687	0.831	0.862	0.842
Chrome Oxide Green	0.942	0.982	0.781	0.563	0.569	0.587	0.798	0.887	0.620
Red Lead	0.871	0.653	0.626	0.508	0.458	0.556	0.868	0.687	0.634
Lead White	0.858	0.634	0.629	0.755	0.767	0.860	0.836	0.957	0.886

a The results sum up the number of pigment class predictions of both the top- and bottom-layered pigments. The results are averaged from five-fold cross-validation.

Table 5.3 shows the effectiveness of different training approaches for pigment classification; first trained solely with simulation data, then with experimental data without finetuning, then with the experimental data with finetuning. The overall accuracy, sensitivity, and F1 score are calculated for each class (each pigment) averaged from the five fold validation groups of experimental data on the testing datasets. Accuracy and Sensitivity are defined in Eq.5.3 and 5.4.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}, \tag{5.3}$$

$$Sensitivity = \frac{TP}{TP + FN}. \tag{5.4}$$

F1 score (the harmonic mean of precision and sensitivity) evaluates the imbalanced classes, as

defined in Eq.5.5.

$$F1 = \frac{2 * Precision * Sensitivity}{Precision + Sensitivity}, \tag{5.5}$$

where

$$Precision = \frac{TP}{TP + FP}. \tag{5.6}$$

The output predictions are classified into TP, TN, FP, and FN, which are short for True Positive, True Negative, False Positive, and False Negative, respectively. True or False denotes whether the class exists or not according to the ground truth. Positive or negative suggests if the (pigment) class is predicted as existing or non-existing. All four evaluation metrics range from 0 to 1, and the closer to 1, the better the performance of the model. The analyses are performed on the same GPU using the scikit-learn package in Python 3.7.

As shown in Table 5.3, the model trained from the simulation dataset generally provided satisfactory accuracy, which ranges from 0.858 to 1.0. Sensitivity varies from 0.572 to 1.0, and F1 score varies from 0.592 to 1.0. According to the sensitivity and F1 score, two groups of pigments - iron-containing-only Prussian blue and iron oxide, and lead-containing-only red lead and lead white - perform worse than other pigments due to the similar elemental profiles within each group. This model can be generalized to the experimental data without finetuning but with relatively worse performance on carmine (tin-based, according the the cross-section analysis), chrome oxide green, and chrome yellow. Compared with the simulation data, tin has a much lower concentration in carmine [164], causing a bigger error and therefore a lower accuracy for the experimental data. In the simulation dataset, there is no pigment mixture. Chrome oxide green only contains Cr and chrome yellow contains Cr and Pb, which helps the model to separate them, but Pb also exists in other pigments in the mock-ups, which confuses the model. However, finetuning the model on the experimental data significantly improves the classification results, reaching an overall accuracy

ranging from 0.798 to 0.998, sensitivity from 0.687 to 0.995, and F1 from 0.634 to 0.997. The next section will further explain our finetuning strategy.

Tests on Paul Gauguin's Poèmes Barbares

Building on the ability to identify the pigments in the mock-up samples, we applied the models to the XRF dataset obtained from *Poèmes Barbares* (Fig. 5.6), which was collected also by the XGLab ELIO XRF imaging spectrometer system [155]. The red rectangle Fig. 5.6 marks the area of investigation.

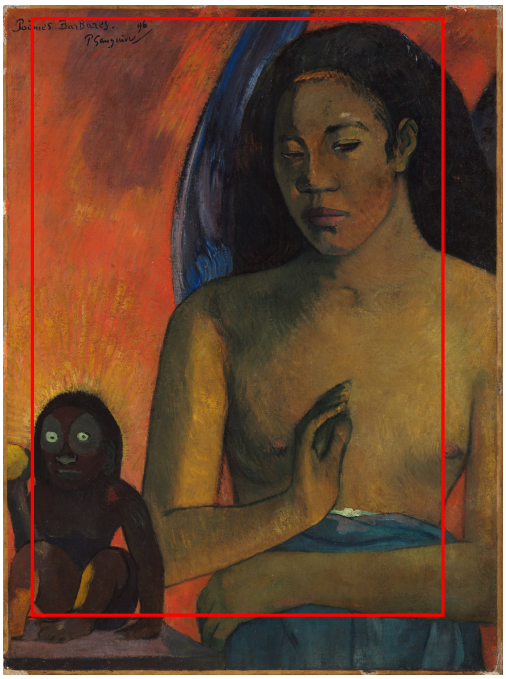

Figure 5.6: ***Poèmes Barbares* (1896)**, oil on canvas, 64.8 x 48.3 cm (unframed), painted by the French artist Paul Gauguin (1848–1903), Harvard Art Museums/Fogg Museum, Bequest from the Collection of Maurice Wertheim, Class of 1906. Object Number: 1951.49 ©President and Fellows of Harvard College.

As shown in Fig. 5.8, the pigment maps suggest the probability of pigments' existence as predicted by the pigment identification model, where an increased pixel brightness suggests a higher

probability. The probability ranges from 0 - 100%, calculated directly from the XRF spectrum image. Their corresponding 2D elemental maps generated from PyMCA are also shown in Fig. 5.8 for comparison, where a brighter pixel suggests a higher elemental concentration. In addition, each pigment map and its corresponding elemental map were merged to better visualize their pixel differences. Note here that as the elemental maps have a higher dynamic range than the pigment maps, we adjusted the max limit of the color bar to 99th percentile of the concentration data to balance the brightness between the two maps. To further evaluate the performance of the pigment identification model, Figure 5.8 also includes a series of scatter plots, which summarize the relationship between the predicted pigment probability and its actual elemental concentration at each pixel location. With a higher accuracy in identifying the pigments, the points remain higher in the plot. If the model identifies a pigment accurately, the probability (y value) will approach to zero at zero concentration, but will rapidly increase with any non-zero elemental concentration (x value).

As shown in Fig. 5.8, the DL model trained and finetuned solely on the mock-ups is applicable to the paintings and simulations as well. Figure 5.8 displays the results of six pigments: cobalt blue, carmine, vermilion, zinc white, and emerald green, each with a unique elemental spectral profile. The first column of images in each row shows the pigment map, with highest probability in red. The second column of images show the elemental map(s) for the pigment calculated by PyMCA, with highest concentrations in green or blue. The third column images overlay the first two for comparison: it combines the red pigment map with the elements maps in green and (sometimes) blue. Yellow or white areas depict strong agreement between pigment maps and their corresponding elemental maps. The fourth column scatter plot compares pigment probability and element concentration data for all image points, where element concentration sets the x-axis value (or for the bottom row, the minimum concentration of two elements), and the pigment probability sets the y-axis value. The right half of the figure shows how finetuning our DL model improves its

results, and depicts images in the same arrangement used in the figure's left half. By comparing the models with and without finetuning, the scatter plots suggest significant effect of finetuning on improving the sensitivity and accuracy in identifying all pigments, particularly for the situation of low elemental concentrations such as carmine. As for the remaining six pigments (chrome oxide green, chrome yellow, iron oxide, Prussian blue, red lead, lead white), the current model has not yet been able to distinguish the pigments that share the same elements, as discussed below.

Tests on Paul Cezanne's The Bathers

Figure 5.7: ***The Bathers* (1899-1904)**, oil on canvas, 51.3 $\times$ 61.7 cm, painted by the French artist Paul Cezanne (1839–1906), The Art Institute of Chicago, Amy McCormick Memorial Collection. Object Number: 1942.457 ©The Art Institute of Chicago.

To further demonstrate that our model is also applicable to other paintings, we applied our same finetuned model to the XRF dataset of Paul Cezanne's *The Bathers* (Fig. 5.7). The XRF mapping was executed with a MA-XRF system at 40 kV and 1 A, with an acquisition time of 100

ms per point and a step size of 1 mm. As shown in Fig. 5.9, pigment maps of cobalt blue, carmine and emerald green perfectly match with their elemental maps and achieve high probabilities with concentrations increasing. The pigment identification of this painting shows results comparable to the Gauguin painting, highlighting the broader applicability and stability of our model.

5.2.4 Discussion and Conclusion

Finetune the model

Finetuning is a general technique popularized in deep learning models, especially on 2D images, to take advantage of weights trained on a huge dataset for another similar but smaller dataset. This technique has shown success in many fields, such as image recognition [165], medical diagnosis [166] and unsupervised learning [167]. In applying this method, the model initially learned the spectral features from the large simulation dataset followed by finetuning on the limited experimental data. Table 5.3 suggests that finetuning, even using a small subset of the dataset, can significantly improve the performance of pigment identification. In particular for the cases of low elemental concentrations, such as carmine, the accuracy, sensitivity, and F1 score increased from 0.653 to 0.995, 0.664 to 0.992, and 0.685 to 0.995, respectively.

To better visualize the effect of finetuning, I generate 2D pigment maps for one mock-up painting (Fig. 5.10) as an example. As shown in Fig. 5.11(a), the vermilion pigment (VM) was present in three horizontal paint strips (top pigment layer) and all six vertical strips (bottom pigment layer), but at different concentrations. However, neither the Hg elemental map nor the VM pigment map detected it reliably without finetuning. While Figure 5.11(b) and (c)) may reveal the existence of VM at low concentrations, finetuning significantly improved the identification result of VM at low concentrations in Fig. 5.11(d), reaching its limit near 0.6% concentration as shown in the second vertical strip.

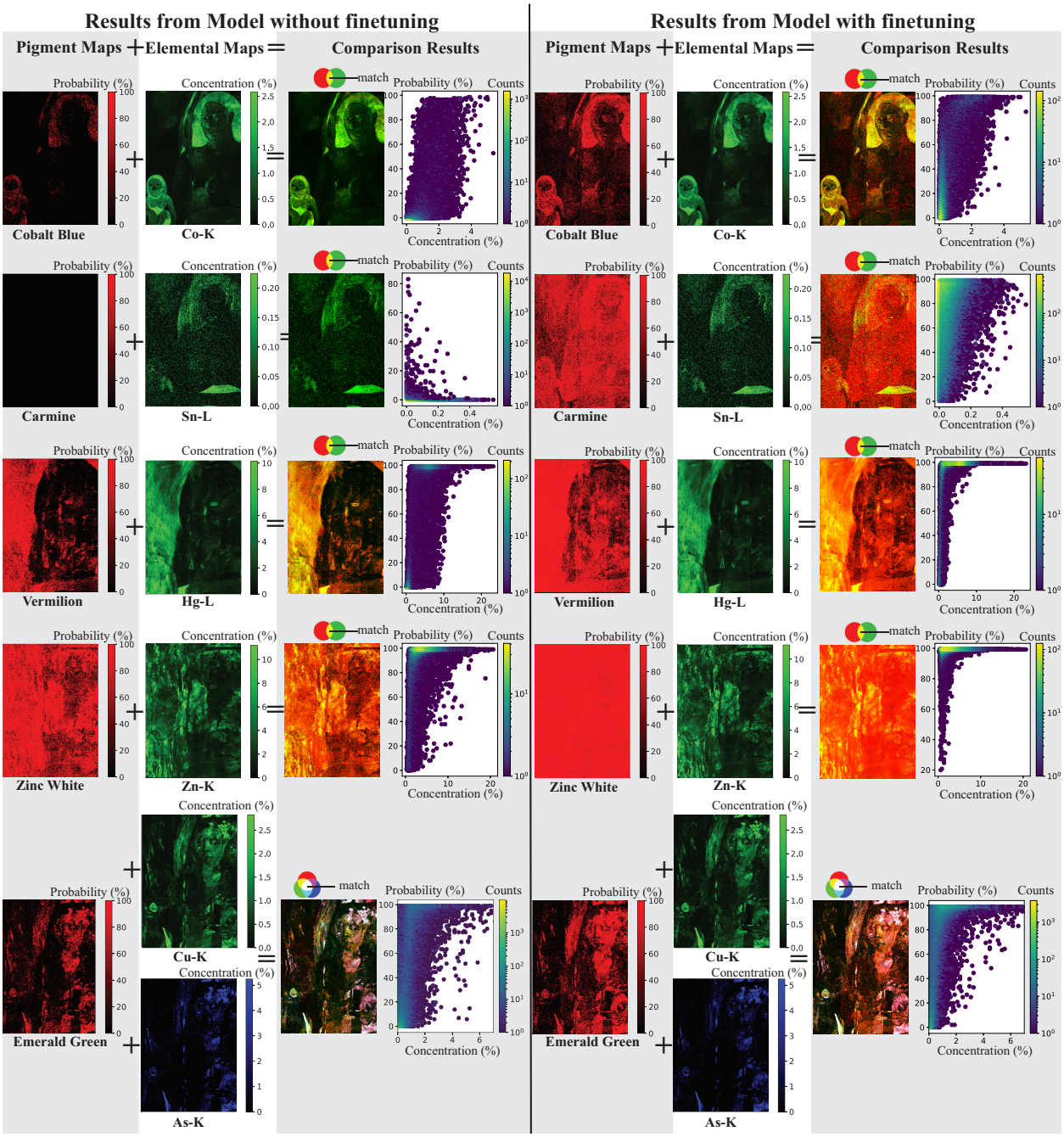

Figure 5.8: **Pigment identification results of *Poèmes Barbares*** from the model without finetuning (left half), and the results from the finetuned model (right half), with results from one pigment on each row. The first, second and third columns in each row display the pigment map, element map(s) and the comparison figure and scatter plots, respectively.

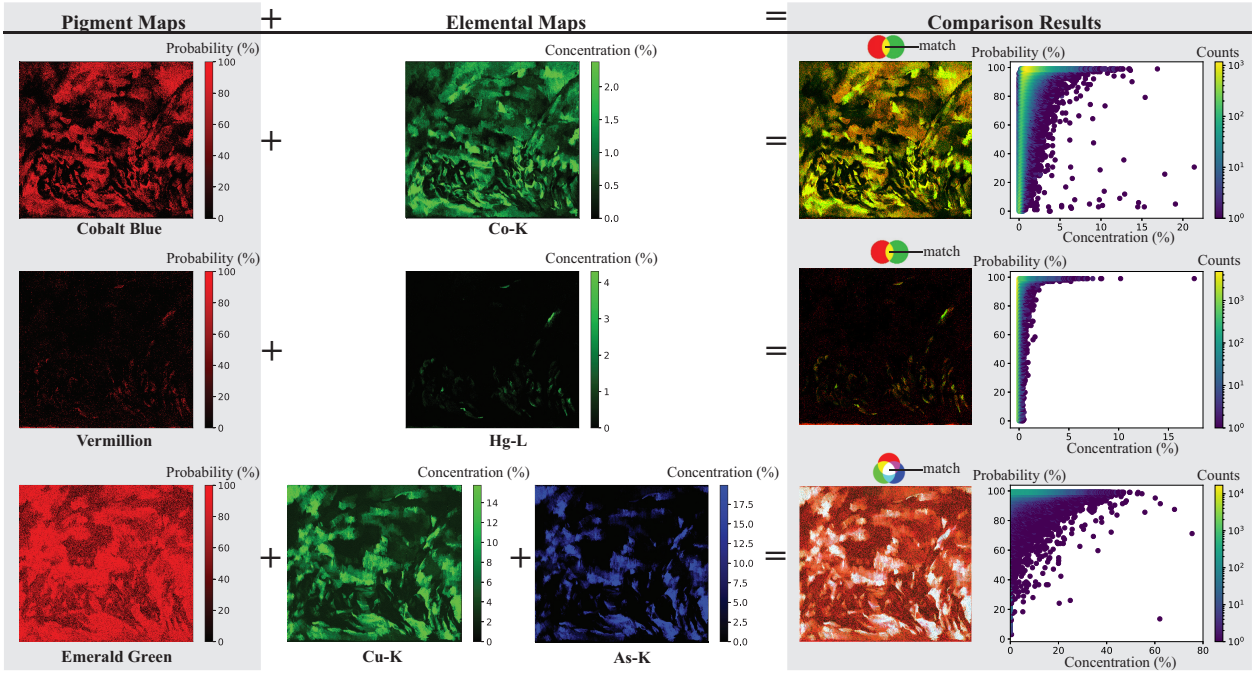

Figure 5.9: **Pigment identification results of *The Bathers*** from the finetuned model, including cobalt blue, vermilion, and emerald green from top to bottom. The first, second and third columns display the pigment map, element map(s) and the comparison figure and scatter plots, respectively.

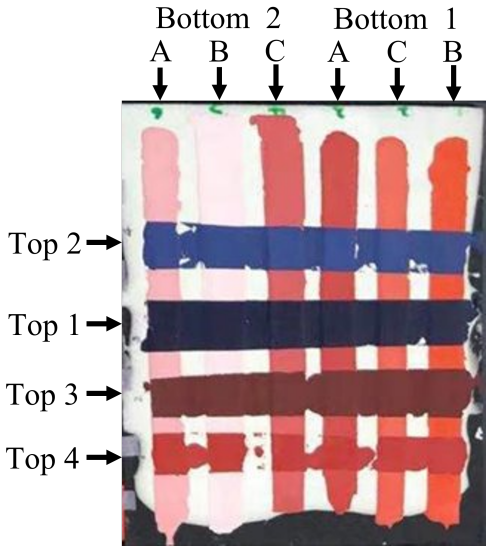

Figure 5.10: **Photo image of one mock-up** that contains Top 2, Top 1, Top 3, Top 4 (from top to bottom) as top layers and Bottom 2 with three different mass fractions (A, B, C), Bottom 1 with three different mass fractions (A, C, B) (from left to right) as bottom layers.

To our best knowledge, this work is the first to apply transfer learning to pigment classification using XRF spectra. By finetuning on only 20% randomly selected of the experimental data, we observe significant improvements in pigment identification. On one hand, it improved the model performance even with a limited training dataset. On the other hand, since mock-ups are hard to make, it releases the pressure of preparing a huge experimental dataset. However, as mentioned previously, since the pigment combination in the experimental dataset is limited, the finetuned model may overfit. Nevertheless, the finetuning technique contributes to extracting the features related to some specific pigment mixtures in the painter's palette, which helps the identification of pigments, especially in a specific painting, with similar painting styles. Therefore, this finetuning technique can be applied to many different fields using XRF spectra.

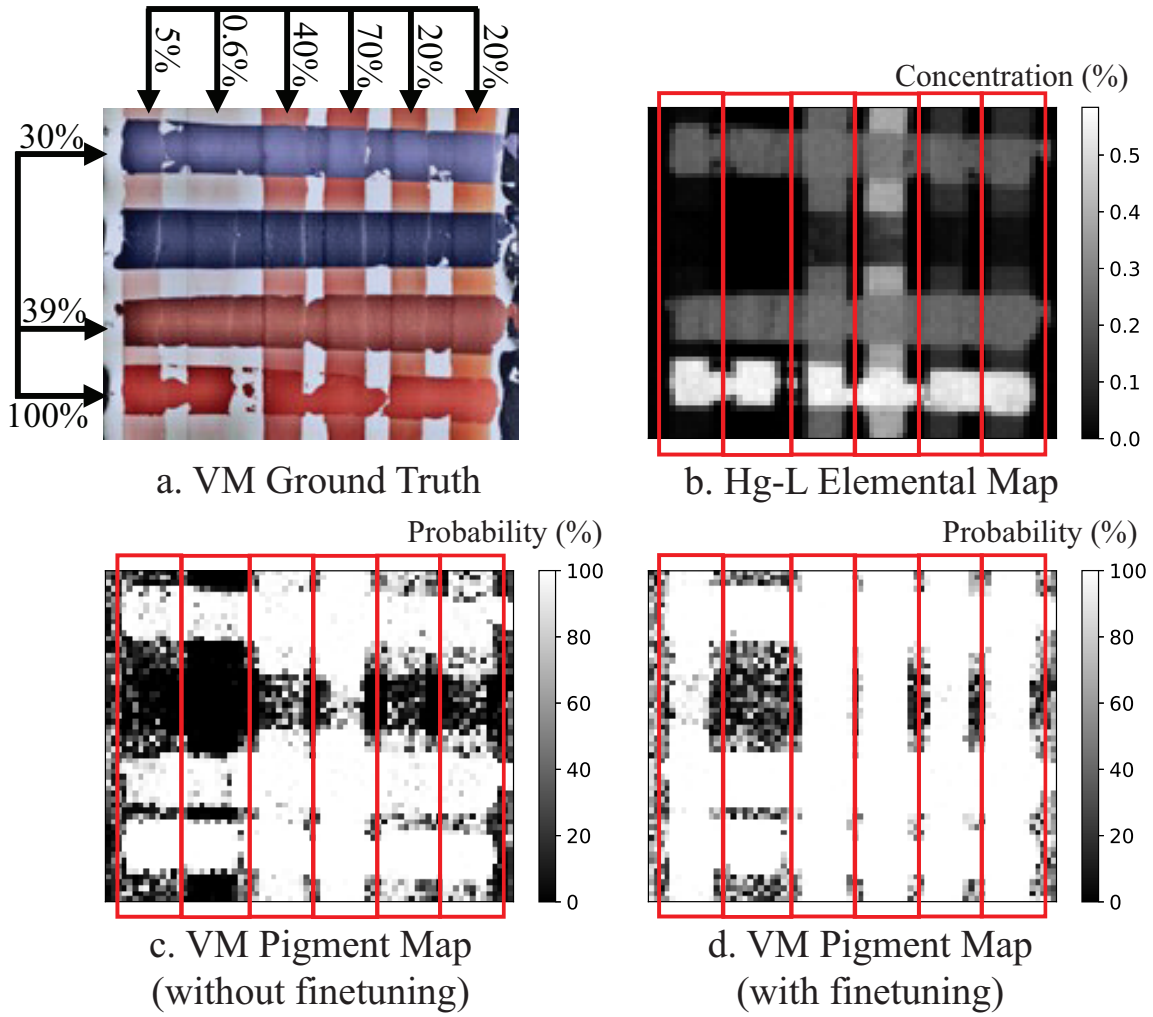

Figure 5.11: **Comparison between pigment maps and elemental maps for the mock-up.** (a) The ground truth of the location of vermilion (VM). (b) The Hg-L element map directly generated from PyMCA. (c) The VM pigment map generated from the model without finetuning identifies Hg in the 1, 3, and 4 rows and the 3 and 4 columns. It can barely detect Hg in the 5 and 6 columns. (d) The VM pigment map generated from the finetuned model indicates VM presence in the 1, 3, and 4 rows and 1, 3, 4, 5, and 6 columns with high probabilities. The only missing column of the VM pigment map is the second column, which contains 0.6% of vermilion.

Pigments with similar elemental profiles

Three groups of pigments in our pigment library posed challenges due to similar elemental profiles: the chromium-containing group (chrome oxide green and chrome yellow), the iron-containing-only group (Prussian blue and iron oxide), and the lead-containing-only group (lead white and red lead). Although the two pigments in the chromium-containing group slightly vary in their elemental map (chrome oxide green only contains Cr and chrome yellow contains Cr and Pb), the model failed to distinguish between these two pigments. This is possibly caused by Pb, which exists almost everywhere in the painting, often mixed with other pigments (e.g. lead white).

For the other two pigment groups that share similar elemental profiles, the model before finetuning cannot distinguish them and shows low to medium probability for all areas that contain the element(s). However, the finetuned model can distinguish between those pigments when present in different pigment mixtures, but this only applies when the testing dataset contains the same pigment mixtures as the finetuning dataset. One special case in our result is that the finetune model can distinguish Prussian blue (PB) and iron oxide (IO) in our mock-ups. For example, in the mock-up in Fig. 5.10, PB exists in the second row, and IO appears in the third row. (Fig. 5.12(a)). The only element detectable by XRF in Prussian blue and iron oxide is Fe. Therefore, the Fe element map (Fig. 5.12(b)) shows the location of both pigments and cannot be used to separate one from the other. However, Prussian blue pigment map (Fig. 5.12(c)) and iron oxide pigment map (Fig. 5.12(e)), generated from our fintuned model, show promising results of distinguishing these two pigments. The two pigment maps both have high probabilities at the ground truth location of the pigment and fairly low probabilities at the ground truth location of the other pigment. At the same time, we notice the pigment maps incorrectly identify where the Fe element has low concentration or does not exist, especially for the Prussian blue pigment map, which may be related to the high tinting strength, and therefore low concentration where used, of Prussian blue [155],

[168], [169]. This situation also appears in the Gauguin and Cezanne painting results: when the pigment concentration is low, the probability might vary from 0 to 100%. We infer that this is due to the normalization process during the spectrum preprocessing. Specifically, the small peaks (low intensities) are enlarged after the normalization, enhancing both useful information and noise level. In the Fe element map (Fig. 5.12(b)), Fe concentration is much higher in iron oxide than in Prussian blue. Therefore, the iron oxide pigment map only has a few noisy points at the blank area (only the ground layer), while the Prussian blue pigment map shows high probabilities at the 1, 2, 5, 6 columns and the blank area.

X-Ray absorbent materials

Some pigments can strongly absorb photons at energies needed for X-ray fluorescence measurements, and absorption by top layers of a painting can severely distort the spectra detected from fluorescence of elements in the bottom layers. For example, the presence of vermilion in the top layer diminishes the lead peaks of lead white and red lead in the bottom layer (Fig. 5.13). In particular, as the energy of the Pb-Lβ radiation is above the Hg-L3 edge, it is absorbed by vermilion in the top layer, decreasing the Pb-Lβ to Pb-Lα ratio. At the same time, the absorption gives rise to florescent emission from Hg-Lα, increasing the Hg-Lα to Hg-Lβ ratio. The shifts in line ratios for both elements create an extraordinary case the simulated data did not adequately describe and our model cannot distinguish from the underlying lead-containing pigments. This effect was previously observed in manual data evaluation as well [170].

Conclusions

These XRF-based pigment identification problems have long required expert analysis and previous knowledge. In this paper, we pursue an automatic XRF data evaluation framework using deep

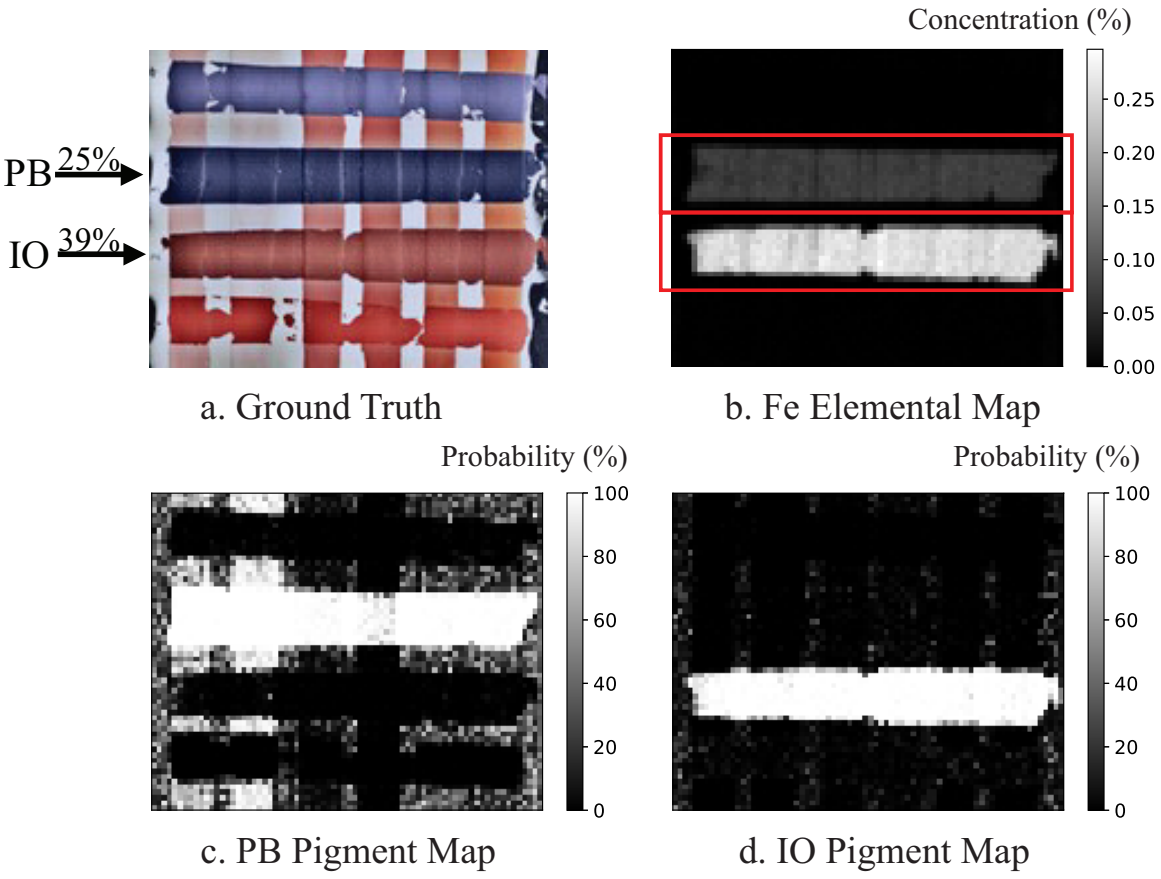

Figure 5.12: **Our model distinguishes Prussian blue (PB) and iron oxide (IO) in one mock-up.** (a) The ground truth of PB and IO locations. (b) The Fe element map directly generated from PyMCA contains both PB and IO. (c) The PB pigment map generated from our model. (d) The IO pigment map generated from our model.

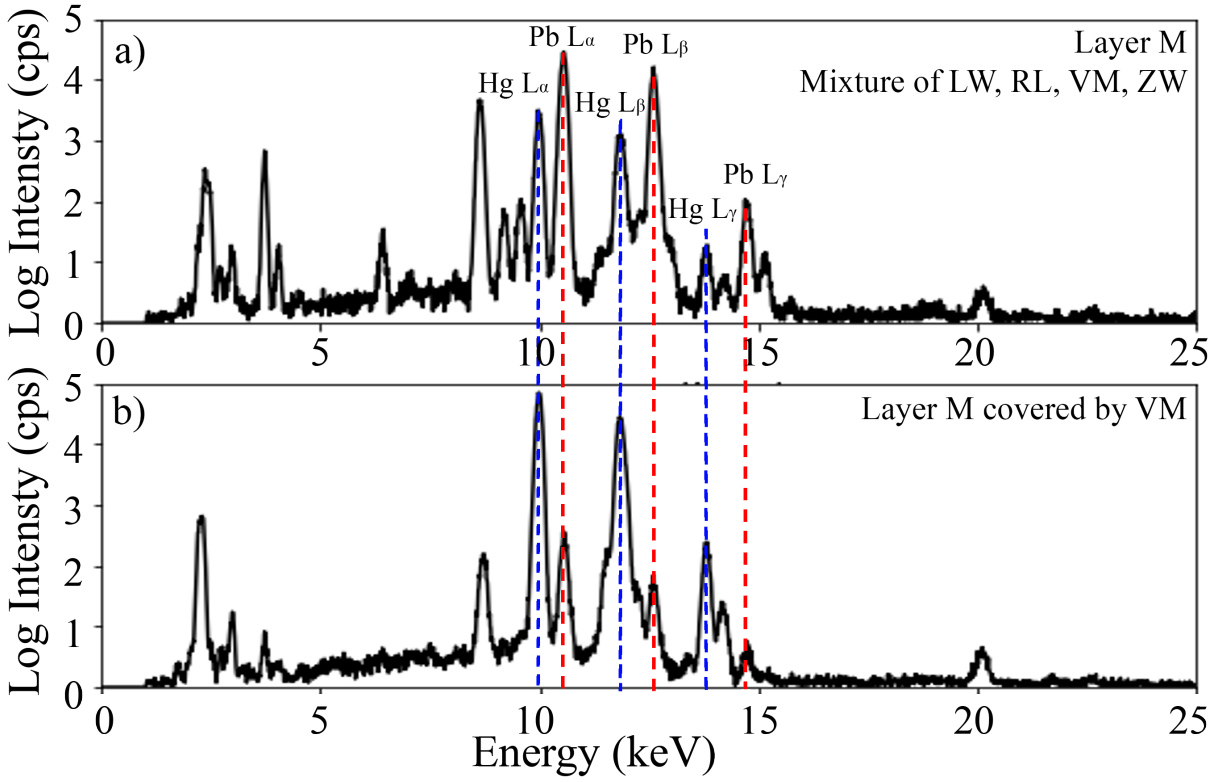

Figure 5.13: **XRF spectra of the mock-up painting comparing the effect of highly absorbing pigments.** (a) One single layer M that contains lead white (LW), red lead (RL), vermilion (VM) and zinc white (ZW). The lead peaks marked in red have high intensities. (b) The single layer M covered by one layer of vermilion (VM). The existence of VM in the top layer significantly blocks the XRF signal of the lead-containing pigments at the bottom, challenging our model's ability to detect the hidden lead element.

learning. Our initial attempts at automatically identifying individual and overlapped pigments directly from XRF spectra show promise. While our model only tests a small number of pigments (11) and a small number of layers (2 + base) in this current stage, it automatically identified pigments in two different 19th-century paintings and in the training mock-ups and simulations they inspired. We intend this work to stimulate further work in deep-learning assisted XRF studies for layered-pigment identification, and prompt more discussion of their feasibility and practicality for broader uses.

We focus on a set of representative pigments identified or considered present in Paul Gauguin's *Poèmes Barbares* (1896) [155] as a starting point and then created mock-ups to generate experimental datasets to capture nonlinear effects of layer structures. We add a simulation dataset to reduce the need for more extensive, prohibitively tedious and difficult mock-up preparation. After data preprocessing, we train the convolutional neural network with the simulation dataset and then finetuned it with the experimental dataset, to converget to the pigment identification model.

Pigment maps are the visualization of the probability output of our model. The comparison of pigment maps and their corresponding element maps shows that our model can successfully identify pigments, especially in low concentration or in overpainted layers. However, the model still has some shortcomings: (a) The model cannot always distinguish pigments with similar elemental profiles; (b) It does not work when high absorbing pigments block the radiation emitted from the hidden layer; (c) The finetuned model needs experimental data with at least a small set of ground truth measurements to prevent limited pigment mixtures in the mock-ups from causing model overfitting and wrong predictions.

This research is still in its early stages, and there are multiple directions to extend the current work. First of all, including data from other techniques, such as spectral imaging to get molecular structures, might better differentiate pigments and compensate for shortcomings (a) and (b) which

are common problems with using XRF to identify pigments. Second, we only try to identify presence or absence of pigments, and not their depth or the layers' sequence from front-to-back, an important but much more challenging problem. Third, the current pigment library is limited, and training the model with a larger range of pigments can make it more accessible to different paintings. With a sufficient pigment dataset, more advanced deep learning algorithms can be applied to further boost the pigment identification performance of the model. For example, Recurrent Neural Networks (RNNs) [171] typically solve problems with sequential input signals because their internal cells store the information retrieved from the previous point in time and use them to generate the next point in time. It will be interesting to try how RNNs work on XRF data with similar time series structures. In addition, some unsupervised learning methods, such as autoencoders [172], have been proposed to extract latent features of signals without any ground truth labels. This is another interesting approach to try in future work that may solve the problem of limited unlabeled datasets in this research area.

CHAPTER 6
CONCLUSION AND FUTURE WORK

The current research employs computational imaging techniques to capture and analyze 3D surface shapes and internal layered materials.

The research seeks to address the following central questions:

1. How can we capture surface shapes for diffuse and specular artworks?

2. How can we measure and utilize the surface shape of the eyeball for eye tracking?

3. How can we identify the mixed pigments inside layered oil paintings?

To address these questions, I presented four imaging systems for surface and volumetric modeling: shape-from-shifting, large-scale deflectometry, single-shot stereo deflectometry, and dynamic focus time domain OCT, which are all user-friendly and non-invasive, utilizing only commercial off-the-shelf (COTS) components, allowing for easy set-up without extensive prior expertise. For further analysis of 3D information, I introduce protrusion detection, multi-view glint-only eye tracking, and automatic pigment identification algorithm separately. Chapters 3, 4 and 5 include limitations and short-term future work for each work in detail. Additionally, I would like to discuss some long-term ideas to further extend our research.

First, I have introduced two slope-measuring principles, photometric stereo, and deflectometry for surface modeling, both of which use illumination and camera images to calculate surface normals. We have developed various imaging systems based on these two principles to measure surface normals for diffuse and specular surfaces. However, the idea of having one system for any type of surface (diffuse, specular, or mixed surface) has been a topic of discussion for some time.

My thought is to create a new system that incorporates a point light source, a screen, and camera(s) or a movable camera. We will calibrate all components and conduct image capture for photometric stereo and deflectometry in a sequence. The primary challenge would be determining how diffuse/specular light affects both methods and how to combine the calculated normal maps from both techniques to reconstruct the mixed surface.

Second, we have utilized the single-shot stereo deflectometry system to measure the surface shape of the eyeball for eye tracking purposes. Additionally, I have discussed two methods for calculating the gaze direction from the surface shape. The first solution involves a direct approach, while the second solution, which I developed during my internship, employs the surface shape and conducts a new multi-view experiment with glints to refine the eyeball pose.

It is worth noting that the second solution could be seen as redundant, as deflectometry already captures specular reflections, making it unnecessary to conduct another experiment to capture the glints, which are also specular reflections. My colleague Tianfu Wang et al. [173] have conducted preliminary research that explores a similar idea. Specifically, they use optimization methods for eye tracking by leveraging information from deflectometry. I find their results to be very promising and believe that further exploration in this direction is warranted.

Last but not least, to identify pigments in layered oil paintings, I have proposed an OCT system to reconstruct the internal layer structure and a deep learning model that utilizes XRF spectra to identify pigments in simplified three-layer paintings. However, OCT only provides information about the internal layer structure, while XRF only provides information about the composition of the pigments. Therefore, our original idea was to combine these two techniques to build a machine-learning model that could simultaneously obtain information about the internal layer structure and composition. However, as this was one of the first studies on automatic pigment identification in layered paintings, we encountered more obstacles than expected, and our progress has been

limited. Despite this, our current results provide a promising starting point for further research along this avenue.

REFERENCES

[1] J. N. Mait, "A history of imaging: Revisiting the past to chart the future," *Optics and Photonics News*, vol. 17, no. 2, pp. 22–27, 2006.

[2] *Optics history*, https://zh.wikipedia.org/wiki/%E5%85%89%E5%AD%B8%E5%8F%B2.

[3] H. Jiang, S. Zhu, H. Zhao, B. Xu, and X. Li, "Adaptive regional single-pixel imaging based on the fourier slice theorem," *Optics express*, vol. 25, no. 13, pp. 15 118–15 130, 2017.

[4] B. Xu, H. Jiang, H. Zhao, X. Li, and S. Zhu, "Projector-defocusing rectification for fourier single-pixel imaging," *Optics Express*, vol. 26, no. 4, pp. 5005–5017, 2018.

[5] S. Hosseinian and H. Arefi, "3d reconstruction from multi-view medical x-ray images–review and evaluation of existing methods.," *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 40, 2015.

[6] H. P. Hiriyannaiah, "X-ray computed tomography for medical imaging," *IEEE signal Processing magazine*, vol. 14, no. 2, pp. 42–59, 1997.

[7] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar, "Acquiring the reflectance field of a human face," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 2000, pp. 145–156.

[8] P. Debevec, "The light stages and their applications to photoreal digital actors," UNIVERSITY OF SOUTHERN CALIFORNIA LOS ANGELES, Tech. Rep., 2012.

[9] T. H. D. Nguyen, T. C. T. Qui, K. Xu, *et al.*, "Real-time 3d human capture system for mixed-reality art and entertainment," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 6, pp. 706–721, 2005.

[10] K. Kolev, P. Tanskanen, P. Speciale, and M. Pollefeys, "Turning mobile phones into 3d scanners," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3946–3953.

[11] P. Carlisle, *Coded aperture imaging*, https://www.paulcarlisle.net/codedaperture/.

[12] E. R. Dowski and W. T. Cathey, "Extended depth of field through wave-front coding," *Applied optics*, vol. 34, no. 11, pp. 1859–1866, 1995.

[13] J. N. Mait, R. Athale, and J. van der Gracht, "Evolutionary paths in imaging and recent trends," *Optics Express*, vol. 11, no. 18, pp. 2093–2101, 2003.

[14] G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," *Optica*, vol. 6, no. 8, pp. 921–943, 2019.

[15] G. S. Cheok, M. Juberts, M. Franaszek, and A. M. Lytle, *3D Imaging systems for manufacturing, construction, and mobility*. US Department of Commerce, National Institute of Standards and Technology, 2010.

[16] A. J. Ireland, C. McNamara, M. Clover, *et al.*, "3d surface imaging in dentistry–what we are looking at," *British dental journal*, vol. 205, no. 7, pp. 387–392, 2008.

[17] C.-H. J. Tzou and M. Frey, "Evolution of 3d surface imaging systems in facial plastic surgery," *Facial Plastic Surgery Clinics*, vol. 19, no. 4, pp. 591–602, 2011.

[18] G. Häusler and S. Ettl, "Limitations of optical 3d sensors," *Optical measurement of surface topography*, pp. 23–48, 2011.

[19] G. Häusler and F. Willomitzer, "Reflections about the holographic and non-holographic acquisition of surface topography: Where are the limits?" *Light: Advanced Manufacturing*, vol. 3, no. 2, pp. 226–235, 2022.

[20] R. G. Dorsch, G. Häusler, and J. M. Herrmann, "Laser triangulation: Fundamental uncertainty in distance measurement," *Applied optics*, vol. 33, no. 7, pp. 1306–1314, 1994.

[21] A. Meydenbauer, "Die photometrographie," *Wochenblatt des Architektenvereins zu Berlin*, vol. 1, no. 14, pp. 125–126, 1867.

[22] E. M. Mikhail, J. S. Bethel, and J. C. McGlone, *Introduction to modern photogrammetry*. John Wiley & Sons, 2001.

[23] F. Willomitzer and G. Häusler, "Single-shot 3d motion picture camera with a dense point cloud," *Optics express*, vol. 25, no. 19, pp. 23 451–23 464, 2017.

[24] M. Saxena, G. Eluru, and S. S. Gorthi, "Structured illumination microscopy," *Advances in Optics and Photonics*, vol. 7, no. 2, pp. 241–275, 2015.

[25] P. Hariharan, *Basics of interferometry*. Elsevier, 2010.

[26] R. Schwarte, Z. Xu, H.-G. Heinol, *et al.*, "New electro-optical mixing and correlating sensor: Facilities and applications of the photonic mixer device (pmd)," in *Sensors, Sensor Systems, and Sensor Data Processing*, SPIE, vol. 3100, 1997, pp. 245–253.

[27] L. Li *et al.*, "Time-of-flight camera—an introduction," *Technical white paper*, no. SLOA190B, 2014.

[28] M. Riley and M. Gusinow, "Laser beam divergence utilizing a lateral shearing interferometer," *Applied Optics*, vol. 16, no. 10, pp. 2753–2756, 1977.

[29] B. K. Horn and M. J. Brooks, *Shape from shading*. MIT press, 1989.

[30] J. J. Clark, "Active photometric stereo.," in *CVPR*, vol. 92, 1992, pp. 29–34.

[31] V. Argyriou and M. Petrou, "Photometric stereo: An overview," *Advances in Imaging and Electron Physics*, vol. 156, pp. 1–54, 2009.

[32] Y. Liu, N. Pears, P. L. Rosin, and P. Huber, *3D imaging, analysis and applications*. Springer, 2020.

[33] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille, "The bas-relief ambiguity," *International journal of computer vision*, vol. 35, no. 1, pp. 33–44, 1999.

[34] T. Papadhimitri and P. Favaro, "A new perspective on uncalibrated photometric stereo," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1474–1481.

[35] A. Hertzmann and S. M. Seitz, "Shape and materials by example: A photometric stereo approach," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, IEEE, vol. 1, 2003, pp. I–I.

[36] N. Alldrin, T. Zickler, and D. Kriegman, "Photometric stereo with non-parametric and spatially-varying reflectance," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2008, pp. 1–8.

[37] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz, "Shape and spatially-varying brdfs from photometric stereo," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 6, pp. 1060–1071, 2009.

[38] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma, "Robust photometric stereo via low-rank matrix completion and recovery," in *Computer Vision–ACCV 2010: 10th Asian Conference on Computer Vision, Queenstown, New Zealand, November 8-12, 2010, Revised Selected Papers, Part III 10*, Springer, 2011, pp. 703–717.

[39] S. Ikehata, D. Wipf, Y. Matsushita, and K. Aizawa, "Robust photometric stereo using sparse regression," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2012, pp. 318–325.

[40] M. C. Knauer, J. Kaminski, and G. Hausler, "Phase measuring deflectometry: A new approach to measure specular free-form surfaces," in *Optical Metrology in Production Engineering*, SPIE, vol. 5457, 2004, pp. 366–376.

[41] L. Huang, M. Idir, C. Zuo, and A. Asundi, "Review of phase measuring deflectometry," *Optics and Lasers in Engineering*, vol. 107, pp. 247–257, 2018.

[42] F. Willomitzer, C.-K. Yeh, V. Gupta, *et al.*, "Hand-guided qualitative deflectometry with a mobile device," *Optics express*, vol. 28, no. 7, pp. 9027–9038, 2020.

[43] N. Petrovic, I. Cohen, B. J. Frey, R. Koetter, and T. S. Huang, "Enforcing integrability for surface reconstruction algorithms using belief propagation in graphical models," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, IEEE, vol. 1, 2001, pp. I–I.

[44] Y. Quéau, J.-D. Durou, and J.-F. Aujol, "Variational methods for normal integration," *Journal of Mathematical Imaging and Vision*, vol. 60, pp. 609–632, 2018.

[45] D. Verellen, M. De Ridder, K. Tournel, *et al.*, "An overview of volumetric imaging technologies and their quality assurance for igrt," *Acta Oncologica*, vol. 47, no. 7, pp. 1271–1278, 2008.

[46] D. C. Adler, J. Stenger, I. Gorczynska, *et al.*, "Comparison of three-dimensional optical coherence tomography and high resolution photography for art conservation studies," *Optics Express*, vol. 15, no. 24, pp. 15 972–15 986, 2007.

[47] A. C. Venjakob and C. R. Mello-Thoms, "Review of prospects and challenges of eye tracking in volumetric imaging," *Journal of Medical Imaging*, vol. 3, no. 1, pp. 011 002–011 002, 2016.

[48] G. Satat, "Imaging through scattering," Ph.D. dissertation, Massachusetts Institute of Technology, 2015.

[49] K. Khaksari and S. J. Kirkpatrick, "Combined effects of scattering and absorption on laser speckle contrast imaging," *Journal of biomedical optics*, vol. 21, no. 7, pp. 076 002–076 002, 2016.

[50] J. F. P.-J. Abascal, J. Aguirre, J. Chamorro-Servent, *et al.*, "Influence of absorption and scattering on the quantification of fluorescence diffuse optical tomography using normalized data," *Journal of biomedical optics*, vol. 17, no. 3, pp. 036 013–036 013, 2012.

[51] G.-O. Regnima, T. Koffi, O. Bagui, *et al.*, "Quantitative measurements of turbid liquids via structured laser illumination planar imaging where absorption spectrophotometry fails," *Applied optics*, vol. 56, no. 13, pp. 3929–3938, 2017.

[52] W. Drexler, J. G. Fujimoto, *et al.*, *Optical coherence tomography: technology and applications*. Springer, 2015, vol. 2.

[53] T. G. Leighton, "What is ultrasound?" *Progress in biophysics and molecular biology*, vol. 93, no. 1-3, pp. 3–83, 2007.

[54] A. Nwaneshiudu, C. Kuschal, F. H. Sakamoto, R. R. Anderson, K. Schwarzenberger, and R. C. Young, "Introduction to confocal microscopy," *Journal of Investigative Dermatology*, vol. 132, no. 12, pp. 1–5, 2012.

[55] C. L. Smithpeter, A. K. Dunn, A. Welch, and R. Richards-Kortum, "Penetration depth limits of in vivo confocal reflectance imaging," *Applied optics*, vol. 37, no. 13, pp. 2749–2754, 1998.

[56] B. Xu, F. Willomitzer, C.-K. Yeh, *et al.*, "3d surface measurement and analysis of works of art," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, IEEE, 2019, pp. 1779–1782.

[57] Y. Li, C.-K. Yeh, B. Xu, *et al.*, "A low-cost solution for 3d reconstruction of large-scale specular objects," in *Computational Optical Sensing and Imaging*, Optica Publishing Group, 2021, CW4H–3.

[58] Y. Li, "Large scale specular object 3d reconstruction and pattern recognition," Master's thesis, Northwestern University, Evanston, IL, 2021.

[59] G. Häusler, "Verfahren und vorrichtung zur ermittlung der form oder der abbildungseigenschaften von spiegelnden oder transparenten objekten," *Patent DE*, vol. 19944354, A1, 1999.

[60] C. Faber, E. Olesch, R. Krobot, and G. Häusler, "Deflectometry challenges interferometry: The competition gets tougher!" In *Interferometry XVI: techniques and analysis*, SPIE, vol. 8493, 2012, pp. 232–246.

[61] C.-K. Yeh, F. Li, G. Pastorelli, M. Walton, A. K. Katsaggelos, and O. Cossairt, "Shape-from-shifting: Uncalibrated photometric stereo with a mobile device," in *2017 IEEE 13th international conference on e-Science (e-Science)*, IEEE, 2017, pp. 551–558.

[62] F. Willomitzer, C.-K. Yeh, V. Gupta, *et al.*, "Uncalibrated deflectometry with a mobile device on extended specular surfaces," *arXiv preprint arXiv:1907.10700*, 2019.

[63] N. U. C. for Scientific Studies in the Arts, *Ghostscan: A non-invasive tool for imaging beneath painted surfaces*, `https://github.com/NU-ACCESS/GhostScan`, 2021.

[64] J. Salvant, M. Walton, D. Kronkright, *et al.*, "Photometric stereo by uv-induced fluorescence to detect protrusions on georgia o'keeffe's paintings," *Metal soaps in art: conservation and research*, pp. 375–391, 2019.

[65] P. Noble, A. van Loon, and J. J. Boon, "Chemical changes in old master paintings ii: Darkening due to increased transparency as a result of metal soap formation," in *Proceedings of 14th Triennial ICOM-CC Meeting in The Hague (I. Verger ed), James & James, London. p. 496*, vol. 503, 2005.

[66] P. Noble, J. J. Boon, and H. M. Parkin, "Metal soap degradation of oil paintings: Aggregates, increased transparency and efflorescence," in *AIC Paintings Specialty Group Postprints, Providence, Rhode Island, June 16-19, 2006*, vol. 19, 2007, pp. 1–15.

[67] Y. Shimadzu and K. Van Den Berg, "On metal soap related colour and transparency changes in a 19th c painting by millais," *Reporting Highlights of the De Mayerne Programme*, vol. 2006, pp. 43–52, 2006.

[68] S. A. Centeno and D. Mahon, "The chemistry of aging in oil paintings: Metal soaps and visual changes," *The Metropolitan Museum of Art Bulletin*, vol. 67, no. 1, pp. 12–19, 2009.

[69] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, Ieee, vol. 2, 1999, pp. 1150–1157.

[70] MathWorks, *MATLAB software*, MathWorks website, 2021.

[71] G. Bradski, "The opencv library.," *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, vol. 25, no. 11, pp. 120–123, 2000.

[72] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marın-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

[73] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM SIGGRAPH 2008 classes*, 2008, pp. 1–10.

[74] *Convex spherical mirrors*, `https://www.edmundoptics.com/f/convex-spherical-mirrors/13761/`.

[75] H. Sattar, M. Fritz, and A. Bulling, "Visual decoding of targets during visual search from human eye fixations," *arXiv preprint arXiv:1706.05993*, 2017.

[76] A. Patney, J. Kim, M. Salvi, *et al.*, "Perceptually-based foveated virtual reality," in *ACM SIGGRAPH 2016 emerging technologies*, 2016, pp. 1–2.

[77] R. J. Jacob and K. S. Karn, "Eye tracking in human-computer interaction and usability research: Ready to deliver the promises," in *The mind's eye*, Elsevier, 2003, pp. 573–605.

[78] H. Chu, S. Ma, F. D. la Torre, S. Fidler, and Y. Sheikh, "Expressive telepresence via modular codec avatars," in *European Conference on Computer Vision*, Springer, 2020, pp. 330–345.

[79] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 3, pp. 478–500, 2009.

[80] W. Sewell and O. Komogortsev, "Real-time eye gaze tracking with an unmodified commodity webcam employing a neural network," in *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, 2010, pp. 3739–3744.

[81] F. Lu, Y. Sugano, T. Okabe, and Y. Sato, "Adaptive linear regression for appearance-based gaze estimation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 10, pp. 2033–2046, 2014.

[82] K. Krafka, A. Khosla, P. Kellnhofer, *et al.*, "Eye tracking for everyone," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2176–2184.

[83] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "Appearance-based gaze estimation in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 4511–4520.

[84] J. Wang, B. Xu, T. Wang, *et al.*, "Vr eye-tracking using deflectometry," in *Computational Optical Sensing and Imaging*, Optica Publishing Group, 2021, CF2E–3.

[85] J. Wang, B. Xu, O. Cossairt, and F. Willomitzer, "Easy and flexible calibration approach for deflectometry-based vr eye-tracking systems," in *Computational Optical Sensing and Imaging*, Optica Publishing Group, 2022, CTh5C–1.

[86] J. Wang, T. Wang, B. Xu, O. Cossairt, and F. Willomitzer, *Accurate and fast vr eye-tracking using deflectometric information*, 2023. arXiv: 2304.09862 [cs.HC].

[87] C. H. Morimoto and M. R. Mimica, "Eye gaze tracking techniques for interactive applications," *Computer vision and image understanding*, vol. 98, no. 1, pp. 4–24, 2005.

[88] Q. Ji and X. Yang, "Real-time eye, gaze, and face pose tracking for monitoring driver vigilance," *Real-time imaging*, vol. 8, no. 5, pp. 357–377, 2002.

[89] W. Fuhl, H. Gao, and E. Kasneci, "Neural networks for optical vector and eye ball parameter estimation," in *ACM Symposium on Eye Tracking Research and Applications*, 2020, pp. 1–5.

[90] A. Villanueva, R. Cabeza, and S. Porta, "Eye tracking: Pupil orientation geometrical modeling," *Image and Vision Computing*, vol. 24, no. 7, pp. 663–679, 2006.

[91] C.-C. Lai, S.-W. Shih, and Y.-P. Hung, "Hybrid method for 3-d gaze tracking using glint and contour features," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 1, pp. 24–37, 2014.

[92] K. Dierkes, M. Kassner, and A. Bulling, "A novel approach to single camera, glint-free 3d eye model fitting including corneal refraction," in *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, 2018, pp. 1–9.

[93] E. Wood, T. Baltrušaitis, L.-P. Morency, P. Robinson, and A. Bulling, "A 3d morphable eye region model for gaze estimation," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, Springer, 2016, pp. 297–313.

[94] D. H. Yoo and M. J. Chung, "A novel non-intrusive eye gaze estimation using cross-ratio under large head motion," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 25–51, 2005.

[95] Z. Zhu and Q. Ji, "Eye gaze tracking under natural head movements," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, IEEE, vol. 1, 2005, pp. 918–923.

[96] Z. Zhu, Q. Ji, and K. P. Bennett, "Nonlinear eye gaze mapping function estimation via support vector regression," in *18th International Conference on Pattern Recognition (ICPR'06)*, IEEE, vol. 1, 2006, pp. 1132–1135.

[97] J. Chen and Q. Ji, "3d gaze estimation with a single camera without ir illumination," in *2008 19th International Conference on Pattern Recognition*, IEEE, 2008, pp. 1–4.

[98] D. W. Hansen and A. E. Pece, "Eye tracking in the wild," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 155–181, 2005.

[99] J. M. Jalife-Chavira, G. Trujillo-Schiaffino, P. G. Mendoza-Villegas, D. P. Salas-Peimbert, M. Anguiano-Morales, and L. F. Corral-Martınez, "Optical methods for measuring corneal topography: A review," *Opt. Pura Apl.*, vol. 52, no. 2, p. 051 016, 2019.

[100] R. Fan, T. C. Chan, G. Prakash, and V. Jhanji, "Applications of corneal topography and tomography: A review," *Clinical & experimental ophthalmology*, vol. 46, no. 2, pp. 133–146, 2018.

[101] S. D. Klyce, "Computer-assisted corneal topography. high-resolution graphic presentation and analysis of keratoscopy.," *Investigative ophthalmology & visual science*, vol. 25, no. 12, pp. 1426–1435, 1984.

[102] L. Sorbara and K. Dalton, "The use of video-keratoscopy in predicting contact lens parameters for keratoconic fitting," *Contact Lens and Anterior Eye*, vol. 33, no. 3, pp. 112–118, 2010.

[103] G. Cairns, A. Collins, and C. McGhee, "A corneal model for slit-scanning elevation topography," *Ophthalmic and Physiological Optics*, vol. 23, no. 3, pp. 193–204, 2003.

[104] G. Cairns, C. N. McGhee, M. J. Collins, H. Owens, and G. D. Gamble, "Accuracy of orbscan ii slit-scanning elevation topography," *Journal of Cataract & Refractive Surgery*, vol. 28, no. 12, pp. 2181–2187, 2002.

[105] T. Koller, H. P. Iseli, F. Hafezi, P. Vinciguerra, and T. Seiler, "Scheimpflug imaging of corneas after collagen cross-linking," *Cornea*, vol. 28, no. 5, pp. 510–515, 2009.

[106] R. Ambrósio Jr, I. Ramos, A. Luz, *et al.*, "Dynamic ultra high speed scheimpflug imaging for assessing corneal biomechanical properties," *Revista Brasileira de Oftalmologia*, vol. 72, pp. 99–102, 2013.

[107] D. Huang, E. A. Swanson, C. P. Lin, *et al.*, "Optical coherence tomography," *science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[108] I. Grulkowski, J. J. Liu, B. Potsaid, *et al.*, "Retinal, anterior segment and full eye imaging using ultrahigh speed swept source oct with vertical-cavity surface emitting lasers," *Biomedical optics express*, vol. 3, no. 11, pp. 2733–2751, 2012.

[109] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[110] H. Liang, T. Sauer, and C. Faber, "Using wavelet transform to evaluate single-shot phase measuring deflectometry data," in *Applications of Digital Image Processing XLIII*, SPIE, vol. 11510, 2020, pp. 404–410.

[111] S. Lombardi, J. Saragih, T. Simon, and Y. Sheikh, "Deep appearance models for face rendering," *ACM Transactions on Graphics (ToG)*, vol. 37, no. 4, pp. 1–13, 2018.

[112] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2107–2116.

[113] E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *IEEE Transactions on biomedical engineering*, vol. 53, no. 6, pp. 1124–1133, 2006.

[114] T. Stoffregen, H. Daraei, C. Robinson, and A. Fix, "Event-based kilohertz eye tracking using coded differential lighting," in *"Winter Conference on Applications of Computer Vision (WACV)"*, 2022.

[115] P. Bérard, D. Bradley, M. Nitti, T. Beeler, and M. H. Gross, "High-quality capture of eyes.," *ACM Trans. Graph.*, vol. 33, no. 6, pp. 223–1, 2014.

[116] T. Beeler, B. Bickel, P. Beardsley, B. Sumner, and M. Gross, "High-quality single-shot capture of facial geometry," in *ACM SIGGRAPH 2010 papers*, 2010, pp. 1–9.

[117] G. Kerbiriou13, Q. Avril, F. Danieau, and M. Marchal12, "Detailed eye region capture and animation,"

[118] S. J. Garbin, Y. Shen, I. Schuetz, R. Cavin, G. Hughes, and S. S. Talathi, "Openeds: Open eye dataset," *arXiv preprint arXiv:1905.03702*, 2019.

[119] NVIDIA, *NVIDIA OptiX™ Ray Tracing Engine*, NVIDIA Developer - OptiX Ray Tracing.

[120] Y. Zhou, C. Barnes, J. Lu, J. Yang, and H. Li, "On the continuity of rotation representations in neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5745–5753.

[121] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[122] Eye3D, *Eye3D™ capture research*, eye3D – Giving eyes a voice.

[123] B. Xu, K. He, P. Hao, *et al.*, "Time-domain optical coherence tomography can measure artworks with high penetration and high resolution," in *Optics for Arts, Architecture, and Archaeology VII*, International Society for Optics and Photonics, vol. 11058, 2019, p. 110580M.

[124] B. J. Xu, Y. Wu, P. Hao, *et al.*, "Can deep learning assist automatic identification of layered pigments from xrf data?" *Journal of Analytical Atomic Spectrometry*, vol. 37, no. 12, pp. 2672–2682, 2022.

[125] L. D. Glinsman, "The practical application of air-path x-ray fluorescence spectrometry in the analysis of museum objects," *Studies in conservation*, vol. 50, no. sup1, pp. 3–17, 2005.

[126] R. Leitgeb, W. Drexler, A. Unterhuber, *et al.*, "Ultrahigh resolution fourier domain optical coherence tomography," *Optics express*, vol. 12, no. 10, pp. 2156–2165, 2004.

[127] S. Zhao, "Advanced monte carlo simulation and machine learning for frequency domain optical coherence tomography," Ph.D. dissertation, California Institute of Technology, 2016.

[128] A. Fercher, K. Mengedoht, and W. Werner, "Eye-length measurement by interferometry with partially coherent light," *Optics letters*, vol. 13, no. 3, pp. 186–188, 1988.

[129] A. Brunetti, B. Golosio, T. Schoonjans, and P. Oliva, "Use of monte carlo simulations for cultural heritage x-ray fluorescence analysis," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 108, pp. 15–20, 2015.

[130] J. Fujimoto and W. Drexler, "Introduction to optical coherence tomography," *Optical coherence tomography: Technology and applications*, pp. 1–45, 2008.

[131] D. Stifter, "Beyond biomedicine: A review of alternative applications and developments for optical coherence tomography," *Applied Physics B*, vol. 88, pp. 337–357, 2007.

[132] H. Liang, R. Lange, B. Peric, and M. Spring, "Optimum spectral window for imaging of art with optical coherence tomography," *Applied Physics B*, vol. 111, pp. 589–602, 2013.

[133] H. Liang, C. Cheung, J. Daniel, M. Tokurakawa, W. Clarkson, and M. Spring, "High resolution fourier domain optical coherence tomography at 2 microns for painted objects," in *Optics for Arts, Architecture, and Archaeology V*, SPIE, vol. 9527, 2015, pp. 7–12.

[134] M. Maria, I. B. Gonzalo, T. Feuchter, *et al.*, "Q-switch-pumped supercontinuum for ultrahigh resolution optical coherence tomography," *Optics letters*, vol. 42, no. 22, pp. 4744–4747, 2017.

[135] M. Hughes and A. G. Podoleanu, "Simplified dynamic focus method for time domain oct," *Electronics letters*, vol. 45, no. 12, pp. 623–624, 2009.

[136] M. Bashkansky, M. Duncan, J. Reintjes, and P. Battle, "Signal processing for improving field cross-correlation function in optical coherence tomography," *Applied optics*, vol. 37, no. 34, pp. 8137–8138, 1998.

[137] C. van Hoof, J. R. Bacon, U. E. A. Fittschen, and L. Vincze, "Atomic spectrometry update – a review of advances in X-Ray Fluorescence pectrometry and its special applications," *Journal of Analytical Atomic Spectrometry*, vol. 36, no. 9, pp. 1797–1812, 2021.

[138] H. Rowe, N. Hughes, and K. Robinson, "The quantification and application of handheld energy-dispersive x-ray fluorescence (ed-xrf) in mudrock chemostratigraphy and geochemistry," *Chemical Geology*, vol. 324, pp. 122–131, 2012.

[139] T. D. T. Oyedotun, "X-ray fluorescence (xrf) in the investigation of the composition of earth materials: A review and an overview," *Geology, Ecology, and Landscapes*, vol. 2, no. 2, pp. 148–154, 2018.

[140] P. Sarala, "Comparison of different portable xrf methods for determining till geochemistry," *Geochemistry: Exploration, Environment, Analysis*, vol. 16, no. 3-4, pp. 181–192, 2016.

[141] K. Langstraat, A. Knijnenberg, G. Edelman, *et al.*, "Large area imaging of forensic evidence with ma-xrf," *Scientific reports*, vol. 7, no. 1, pp. 1–11, 2017.

[142] K. Nakano, C. Nishi, K. Otsuki, Y. Nishiwaki, and K. Tsuji, "Depth elemental imaging of forensic samples by confocal micro-xrf method," *Analytical chemistry*, vol. 83, no. 9, pp. 3477–3483, 2011.

[143] M. S. Shackley, "An introduction to x-ray fluorescence (xrf) analysis in archaeology," in *X-Ray Fluorescence spectrometry (XRF) in geoarchaeology*, Springer, 2011, pp. 7–44.

[144] M. Alfeld and L. de Viguerie, "Recent developments in spectroscopic imaging techniques for historical paintings-a review," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 136, pp. 81–105, 2017.

[145] P. J. van Espen and K. H. Janssens, "Spectrum evaluation," in *Handbook of X-Ray spectrometry: Methods and techniques*, Marcel Dekker, Inc., New York, NY, 1993, ch. 5, pp. 181–293.

[146] V. Solé, E. Papillon, M. Cotte, P. Walter, and J. Susini, "A multiplatform code for the analysis of energy-dispersive x-ray fluorescence spectra," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 62, no. 1, pp. 63–68, 2007.

[147] M. Alfeld and K. Janssens, "Strategies for processing mega-pixel x-ray fluorescence hyperspectral data: A case study on a version of caravaggio's painting supper at emmaus," *Journal of analytical atomic spectrometry*, vol. 30, no. 3, pp. 777–789, 2015.

[148] F. P. Romano, C. Caliri, P. Nicotra, *et al.*, "Real-time elemental imaging of large dimension paintings with a novel mobile macro x-ray fluorescence (ma-xrf) scanning technique," *Journal of Analytical Atomic Spectrometry*, vol. 32, no. 4, pp. 773–781, 2017.

[149] M. Alfeld, W. De Nolf, S. Cagno, *et al.*, "Revealing hidden paint layers in oil paintings by means of scanning macro-xrf: A mock-up study based on rembrandt's "an old man in military costume"," *Journal of Analytical Atomic Spectrometry*, vol. 28, no. 1, pp. 40–51, 2013.

[150] S. Kogou, L. Lee, G. Shahtahmassebi, and H. Liang, "A new approach to the interpretation of xrf spectral imaging data using neural networks," *X-Ray Spectrometry*, vol. 50, no. 4, pp. 310–319, 2021.

[151] M. Vermeulen, A. McGeachy, B. Xu, *et al.*, "Xrfast a new software package for processing of ma-xrf datasets using machine leaning," *Journal of Analytical Atomic Spectrometry*, 2022.

[152] A. N. Shugar, B. L. Drake, and G. Kelley, "Rapid identification of wood species using xrf and neural network machine learning," *Scientific reports*, vol. 11, no. 1, pp. 1–10, 2021.

[153] J. J. Kim, F. T. Ling, D. A. Plattenberger, A. F. Clarens, and C. A. Peters, "Quantification of mineral reactivity using machine learning interpretation of micro-xrf data," *Applied Geochemistry*, vol. 136, p. 105 162, 2022.

[154] C. Jones, N. S. Daly, C. Higgitt, and M. R. Rodrigues, "Neural network-based classification of x-ray fluorescence spectra of artists' pigments: An approach leveraging a synthetic dataset created using the fundamental parameters method," *Heritage Science*, vol. 10, no. 1, pp. 1–14, 2022.

[155] M. Vermeulen, K. Smith, K. Eremin, G. Rayner, and M. Walton, "Application of uniform manifold approximation and projection (umap) in spectral imaging of artworks," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 252, p. 119 547, 2021.

[156] L. F. Sturdy, "Coating characterization with the quartz crystal microbalance," Ph.D. dissertation, Northwestern University, 2016.

[157] L. De Viguerie, V. A. Sole, and P. Walter, "Multilayers quantitative x-ray fluorescence analysis applied to easel paintings," *Analytical and bioanalytical chemistry*, vol. 395, no. 7, pp. 2015–2020, 2009.

[158] C. Ryan, E. Clayton, W. Griffin, S. Sie, and D. Cousens, "Snip, a statistics-sensitive background treatment for the quantitative analysis of pixe spectra in geoscience applications," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 34, no. 3, pp. 396–402, 1988.

[159] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *arXiv preprint arXiv:1505.00853*, 2015.

[160] F. Zhuang, Z. Qi, K. Duan, *et al.*, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.

[161] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding transfer learning for medical imaging," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32, Curran Associates, Inc., 2019.

[162] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the royal statistical society: Series B (Methodological)*, vol. 36, no. 2, pp. 111–133, 1974.

[163] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[164] R. Dapson, "The history, chemistry and modes of action of carmine and related dyes," *Biotechnic & Histochemistry*, vol. 82, no. 4-5, pp. 173–187, 2007.

[165] H.-W. Ng, V. D. Nguyen, V. Vonikakis, and S. Winkler, "Deep learning for emotion recognition on small datasets using transfer learning," in *Proceedings of the 2015 ACM on international conference on multimodal interaction*, 2015, pp. 443–449.

[166] S. Khan, N. Islam, Z. Jan, I. U. Din, and J. J. C. Rodrigues, "A novel deep learning based framework for the detection and classification of breast cancer using transfer learning," *Pattern Recognition Letters*, vol. 125, pp. 1–6, 2019.

[167] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," in *Proceedings of ICML workshop on unsupervised and transfer learning*, JMLR Workshop and Conference Proceedings, 2012, pp. 17–36.

[168] M. Vermeulen, A. S. O. Miranda, D. Tamburini, S. E. R. Delgado, and M. Walton, "A multi-analytical study of the palette of impressionist and post-impressionist puerto rican artists," *Heritage Science*, vol. 10, no. 1, pp. 1–22, 2022.

[169] L. D. Glinsman *et al.*, "The application of x-ray fluorescence spectrometry to the study of museum objects," Ph.D. dissertation, Universiteit van Amsterdam [Host], 2004.

[170] C. Neelmeijer, I. Brissaud, T. Calligaro, *et al.*, "Paintings—a challenge for xrf and pixe analysis," *X-Ray Spectrometry: An International Journal*, vol. 29, no. 1, pp. 101–110, 2000.

[171] A. Sherstinsky, "Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132 306, 2020.

[172] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proceedings of ICML workshop on unsupervised and transfer learning*, JMLR Workshop and Conference Proceedings, 2012, pp. 37–49.

[173] T. Wang, J. Wang, O. Cossairt, and F. Willomitzer, "Optimization-based eye tracking using deflectometric information," *arXiv preprint arXiv:2303.04997*, 2023.

COMPUTATIONAL IMAGING FOR ACCURATE 3D MODELING IN DIVERSE FIELDS

Approved by:

Jack Tumblin
Computer Science
Northwestern University

Marc Walton
Conservation and Research
M+ Museum

Florian Willomitzer
Optical Sciences
University of Arizona

Aggelos Katsaggelos
Electrical & Computer Engineering
Northwestern University

Oliver Cossairt
Computer Science
Northwestern University

Date Approved: May 4, 2023