NORTHWESTERN UNIVERSITY

Mobile Computational Imaging Systems for Appearance Modeling Based Surface Shape Recovery

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Computer Science

By

Chia-Kai Yeh

EVANSTON, ILLINOIS

September 2020

ABSTRACT

Mobile Computational Imaging Systems for Appearance Modeling Based Surface Shape Recovery

Chia-Kai Yeh

Surface appearance represents the sense impression of the surface. In visual art, the artists try to use the appearance of their artworks to express their mental state and philosophy. Researchers in the cultural heritage community has been trying to use different analysis approaches to interpret artworks. In Computer Graphics and Computational Imaging, surface appearance modeling and shape recovery has been the central challenge. Combining the optimized hardware (camera and illuminants) and reconstruction algorithm, it could estimate the surface geometry and material properties with the measured surface appearances. Unfortunately, the cultural heritage community hardly exploits these powerful tools due to the high requirement of sophisticated system.

In this thesis, we introduce novel computational imaging frameworks that apply commodity hardware such as mobile phone/tablet system, digital single-lens reflex (DSLR) camera, and liquid-crystal display (LCD). Appearance modeling-based shape recovery techniques such as Photometric Stereo (PS) and Phase Measuring Deflectometry (PMD) has been widely used in academic research and industrial sophisticated applications. They have

been intensively used for applications such as human digitization in visual effect and industrial inspection for high-quality geometry and material properties. However, conventional appearance modeling-based methods require complicated hardware systems and constrained environments, limiting its usage to different applications.

In the original PS image formation model, it assumes the light source places at infinity far away. In order to obtain high-quality shape recovery, conventional PS techniques typically require a giant light dome or calibration equipment (*e.g.* mirror ball) with open operation space to carefully control the lighting condition. To break the limitation and improve the robustness, we propose a novel near-light PS framework by leveraging photogrammetry and unique portable dual cameras system to improve the lighting calibration. An uncalibrated photometric stereo setup is augmented by a synchronized secondary witness camera co-located with a point light source. By recovering the witness of the camera's position for each exposure with photogrammetry techniques, we estimate the precise $3D$ location of the light source relative to the photometric stereo camera. We have shown a significant improvement in both light source position estimations and normal map recovery compared to previous uncalibrated photometric stereo techniques. Besides, with the new configuration we propose, we benefit from improved surface shape recovery by jointly incorporating the corrected photometric stereo surface normal and a sparse 3D point cloud from photogrammetry.

Although the proposed PS framework helps improve the robustness and quality of the surface shape, to further improve the portability and accessibility for PS to apply to different applications, we introduce mobile shape-from-shifting (SfS): a simple, low-cost and streamlined photometric stereo framework for scanning planar surfaces with a consumer mobile device coupled to a low-cost add-on component. Our free-form mobile SfS framework relaxes the rigorous hardware and other complex requirements inherent to conventional 3D

scanning tools. This is achieved by taking a sequence of photos with the on-board camera and flash of a mobile device. The sequence of captures is used to reconstruct high-quality normal maps using near-light photometric stereo algorithms, which are of comparable quality to conventional photometric stereo. We demonstrate 3D surface reconstructions with SfS on different materials and scales. Moreover, the mobile SfS technique can be used" in the wild" so that 3D scans may be performed in their natural environment, eliminating the need for transport to a laboratory setting.

PS techniques could welly handle most of the common surface material and provide high-quality surface geometry. However, it would fail with highly reflective surface which disobeyed the Lambertian reflectance used in PS. To cover specular material, we introduce a system that exploits the screen and front-facing camera of a mobile device to perform three-dimensional Deflectometry-based surface modeling. In contrast to current mobile deflectometry systems, our method can capture surfaces with large normal variation and wide field of view (FoV). We achieve this by applying automated multiview panoramic stitching algorithms to produce a large FoV normal map from a hand-guided capture process without the need for external tracking systems, like robot arms or fiducials.

Lastly, we propose an inverse rendering based reflective surface shape reconstruction. Conventional multiview deflectometry techniques require multiview normal map stitching and normal field integration to obtain 3D surface geometry. Instead of trying to contain the noise that introduces in these two procedures, we adopt differential renderer to directly optimize the surface geometry from the appearance measurements, which produce high quality $3D$ surface information.

Acknowledgments

I'm really lucky that there are lots of people giving substantial supports throughout my Ph.D. journey. I would like to express my gratitude for their time and supports.

First of all, I would like to present my sincerest gratitude to my advisor Prof. Oliver Cossairt for having me in the Computational Photography lab for the past five years. Since I started as a master student in the lab, Ollie has generously provided calm and supportive guidance. With his patience, motivation, and immense knowledge, I have learned a lot about the imaging, respectful collaboration with colleagues, and communication. Most importantly, Ollie has showed me how to treat the problem by starting from a holistic perspective to the fine grained detail and always encourage me to think outside of the box. Thanks to the flexible environment that Ollie creates for his students, I have been able to explore freely for different aspects while looking to Ollie for hands-on guidance anytime I needed. He is one of the smartest and kindest people I know and I cannot imagine a better advisor and role model for me.

I am very grateful to the rest of my committee members, Prof. Marc Walton, Prof. Aggelos Katsaggelos, Prof. Jack Tumblin, and Prof. Florian Willomitzer, for their insightful comments and encouragement. Marc has guided me in to the cultural heritage community and showed me how to think outside of the box to look for any possibility. Aggelos and Jack have instilled me to be a critical and creative thinker and researcher. Florian and I have lots

of close collaborations and he has provided hands-on support during my research and the thesis preparation.

Within the last five years, I was fortunate to have the chances to work in top-tier industry labs during the summer. In the summer of 2016, I had the chance to work Dr. Andrew Jones and Dr. Paul Debevec on markerless volumetric performance capture and appearance modeling of human body at University of Southern California Institution of Creative Technologies. Also, I am very grateful to Dr. Sunil Hadap, Dr. Duygu Ceylan and T.J. Rhodes for my internship at Adobe Research to work on Project PhotoGeode (https://research.adobe.com/news/photogeode-3d-faces-from-2d-imagery/) in 2018. These fruitful experiences has broaden my horizons in human digitization, digital geometry processing and appearance modeling.

During my time at Northwestern, I was lucky to be surrounded by a group of wonderful labmates and colleagues. I would like to thanks Dr. Fengqiang Li, Dr. Nathan Matsuda, Dr. Xiang Huang, Danielle Duggins, Dr. Annette Miranda and Bingjie Xu have made many contributions to my researches and projects. Dr. Marina Alterman, Dr. Jason Holloway, Dr. Kuan He, Dr. Zihao Wang, Sushoban Ghosh, Florian Schiffers, Lionel Fiske, Weixin Jiang, Fabian Wagner, Dr. Yudong Yao, and Prof. Ming Zhao for the stimulating discussions and for all the fun time we had in Comp Photo lab. Outside the campus, I have a group of friends and basketball buddies, Pei-Lin Tsai, Kevin Deng, Steven Chang, Scottie Tien, Prof. Shih-Hsun Chen and Tommy Wu. Thank you for making my life in Chicago so colorful.

I would like to express my sincerest appreciation to my family and parents-in-law at Taiwan. My parents G.J., Ping and my sister Nicole provide continuous support and encouragement to my study. Especially my dad G.J. who instilled my passion with imaging and engineering since I was a little boy had encouraged me to take on the path to get a

PhD in the first place. I would like to dedicate this thesis to him in that I've fulfilled his unfinished dreams.

Five years ago, I was a young man that trying to start a fresh life in the state. Now, I've a sweet family with two cutest kids, Ariel and Alvin. I'm glad that I could be closely on their side while they grew up along with this graduated study journey. I will always remember the day that Ariel typed her little keyboard and said "I'm coding!" or Alvin sat on my lap in front of my computer and tried to type-in my thesis document. Last but not least, none of this lovely family nor this doctoral study could been possible without my wife, Vivian. She encouraged me to pursue my dream from the very beginning and adapted the sudden change in location and lifestyle. She is the staunch backing of mine and the best companion through various of ups and downs during the journey. Vivian, Ariel and Alvin, I love you all and looking forward to the new chapter in our life.

This dissertation was supported in part by the NSF CAREER grant IIS-1453192, a generous grant from the Andrew W. Mellon Foundation to NU-ACCESS, NEH grant PR-258900-18 and the Materials Research Center, the Office of the Vice President for Research, the McCormick School of Engineering and Applied Science and the Department of Materials Science and Engineering at Northwestern University.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1. Surface Appearance Modeling

In the history of the visual arts, the artist has spent lots of efforts using different art forms such as painting, drawing, sculpture, ceramics, and architecture to share their perspective with the others. The appearance of the artworks meets the spectators' eyes and evokes emotion and a sense of expression so that the spectators could relate to the artist's philosophy and background characteristic of times. Over the years, the conservators in the cultural heritage community tried to document, preserve, and restore the artifacts for our future generation. In the meantime, conservation science uses different kinds of analysis approaches to interpret the materials and techniques used by artists and study the causes of deterioration. In the last decade, the conservation science community has started to adopt imaging methods such as $3D$ imaging, which is non-destructive and fast as their analytical techniques.

To fully understand the relationship between every component of the surface appearance, surface appearance modeling has been the central challenge in computer vision and computer graphics. Given measurements of the surface, it estimates the geometry and material properties of the surface which have opened up a wide range of compelling applications such as visual effects in movies and video games [24], computer-aided-design for rapid prototyping, quality inspection [143], cultural heritage [70, 23, 51, 58, 32, 22, 110], and biological imaging. For human visual perception, this problem seems trivial to use since we perform it easily and often subconsciously. On the other hand, it is an extremely challenging task for

computers since it perceives the images in a completely different way. In the last decade, graphics researchers have pushed a significant progress on photorealistically delineation appearance of complex real-world material [25, 62, 29, 133]. While there has been remarkable progress in appearance modeling research, still there are some significant disconnections between theory and practice to block the usage in different applications such as conservation science. In this dissertation, we focus on the ability to using appearance information to infer the surface properties, in particular, surface shape. Most importantly, we focus on enabling such capabilities and techniques to perform under relatively unconstrained and user-friendly conditions and also decrease the hardware requirement by using commodity hardware such as DSLRs, tablets, and mobile phones.

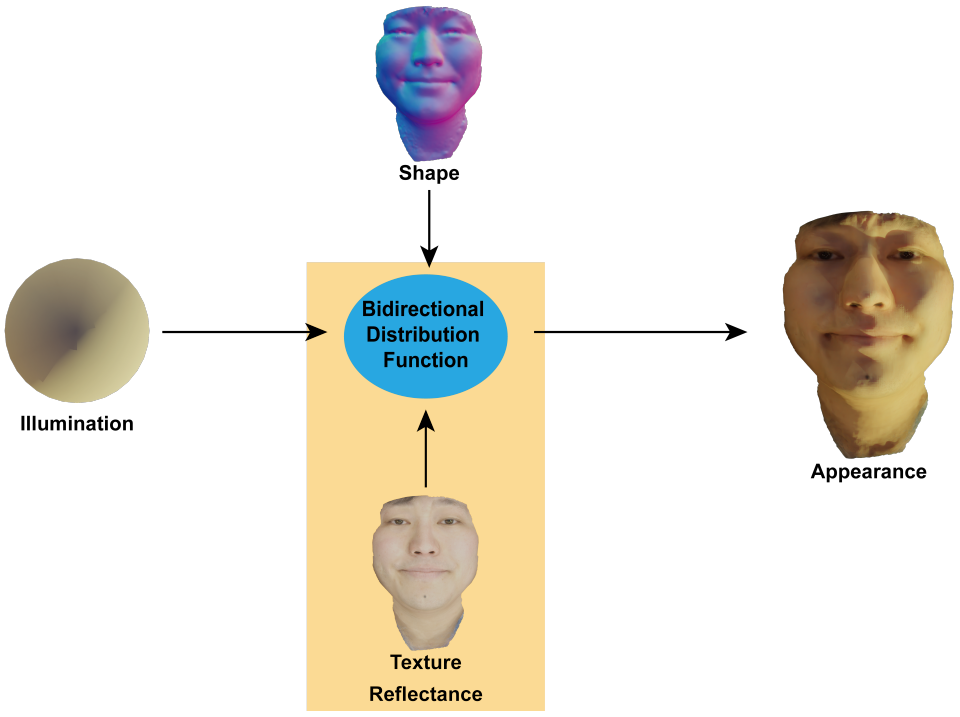


Figure 1.1. **Surface appearance model:** The appearance of a surface combines with its shape, illumination, and reflectance which consist of material bidirectional distribution function and texture.

$$I_{observed} = f_\lambda(n, k, l), \tag{1.1}$$

To achieve these capabilities in surface appearance modeling, we first have to understand how shape, reflectance, and illumination are interacted with each other on/in the surface to reveal its appearance. As Shown in Figure 1.1 and Eq. 1.1, the intensity observed at a surface point I is a complex function f_λ between surface geometry n, materials properties (color texture) k and illumination l. In early computer graphics, there has been plenty of effort devoted to model different material properties and light models. They model the phenomenon by point-based, which describes the portion of the light that is reflected toward the viewpoint when a surface point is lit from a certain direction. This complex function f_λ often formalized by the means of the *bi-directional reflectance distribution function* (BRDF [93]). It describes the ratio between outgoing radiance L_o toward ω_o at surface point x_p and incident irradiance L_i from a direction ω_i:

$$f_\lambda(\theta_o, \phi_o, \theta_i, \phi_i) = \frac{L_o(\theta_o, \phi_o)}{L_i(\theta_i, \phi_i)} \tag{1.2}$$

which (θ_o, ϕ_o) and (θ_i, ϕ_i) is the two spherical directions ω_o and ω_i under spherical coordinates relative to a local surface coordination. The rendering equation [63] which is often used to simulate the light transport in realistic real-world scene could be expressed by integrating the BRDF f_λ, surface geometry n and incident ω_o and outgoing ω_i directions over the positive hemisphere of the surface:

$$L(x_p, \omega_o) = \int_{\Omega_+} f_\lambda(x_o, \omega_o, \omega_i) L(x_p, \omega_i)(\omega_i \cdot n) d\omega_i \tag{1.3}$$

Appearance modeling-based shape recovery methods try to solve this high complexity inverse problem to isolate these factors. There are plenty of different techniques that have been studied in detail in prior researches. Most of them follow the general approach, which estimates the shape and reflectance - in terms of surface orientation and color texture - by assuming the illumination and imposing priors to the reflectance (*e.g.* Lambertian material or Specular material).

$$I_{observed} = k_d N L \tag{1.4}$$

In computer vision, it is a long-time research problem to model the shape and reflectance information with known illumination conditions and prior assumption of the reflectance. For diffuse reflectance material, photometric stereo (PS) [55] is an extensively studied research topic which estimates surface normal from a set of photographs taken with a fixed camera position and multiple known lighting directions. Intensity values in the captured images are modeled as in Eq. 1.4 which a function of lighting angle L, surface normal N, and material reflectance k_d. By inverting this model, PS techniques recover surface normal, which can then be integrated to produce $3D$ Surface shape. The original formulation by Horn assumed lights are infinitely far away, the camera is orthographic, and the object surface is Lambertian and convex (*i.e.*, no shadows or inter-reflections). Since photometric stereo was originally introduced, several researchers have sought to generalize the technique for more practical camera, surface, and lighting models. Belhumeur *et al.* [12] discovered that with an orthographic camera model and uncalibrated lighting, the object's surface could be uniquely determined to within a bas-relief ambiguity. Papadhimitri and Favaro *et al.* [96] recently pointed out that this ambiguity is resolved under the perspective camera model. Several

researchers have also sought to relax the Lambertian reflectance assumption and incorporate effects such as specular highlights and shadows. New techniques have been introduced based on non-Lambertian reflectance models [52, 7, 45], or sophisticated statistical methods to automatically filter non-Lambertian effects [142, 59, 148]. Others consider non-isotropic illuminations [102]. Ackermann *et al.* [3] recently gave a more comprehensive survey on earlier and recent photometric stereo techniques.

Although using the Lambertian prior has achieved great success among most of the objects, unfortunately, real-life objects often involve Specular materials that interact with the light in a totally different way. In order to cover the specular material, researchers had proposed Deflectometry [66, 57], Specular Flow [4, 106], and shape from specularity [127] to model the specular surface. Because of the uniqueness of the specular reflectance that for single incoming light, the specular reflection would only present in one direction, researchers posed this as a correspondence problem. The observed appearance of a specular (mirror-like) surface is a distorted version of the illumination, where the deformation depends on the surface normal distribution of the object surface. From this deformation, the normal vectors of the surface can be calculated with the known correspondence between illumination and the observation point. As compared to diffuse reflection, which would miss detail due to the blurring effects of subsurface scattering, specular reflection could reveal more high-frequency geometry detail of the surface with complex material reflectance.

1.2. Statement of the problem

Computational 3D imaging techniques have had explosive growth in both industry and academic research during the last decade with a variety of applications. Although there are lots of real-world tasks that have been benefited by different kinds of commercial $3D$ imaging

systems, the conventional commercial $3D$ imaging system still suffer from many drawbacks. Among the triangulation-based approaches, Photogrammetry provides high versatility for different setups. Depend on the applications, the system could be either with a single camera that took images from different positions [117] or fixed multiple cameras captured at the same time [39]. The resulting 3D result is within centimeter-scale and comes with a correspondent low-quality color texture of the surface. The result is relatively spare since 3D information is computed only for each feature point, but not each pixel. These sparse 3D points can be interpolated to generate a low-resolution 3D mesh model of the object. It would also suffer from the textureless surface, which could not efficiently generate the correspondences. Structured light system [149, 41] improves the depth resolution to sub-millimeter and solves the textureless issue by projecting designed patterns. However, both of the methods cannot image complex material such as specular and transparent surfaces since the geometry triangulation is based on the photo-consistency property of the Lambertian reflectance.

On the other hand, appearance modeling-based methods such as Photometric Stereo and Deflectometry are modeling surface reflectance to recover dense geometry information and reflectance properties. Both of the surface appearance modeling techniques have been explored extensively in the literature but still have fundamental drawbacks, which make it hard to apply to lots of real-world applications. In order to have high quality and accurate results, surface appearance modeling usually requires a sophisticated and bulked hardware system. In typical setups, the hardware system (*e.g.* a lighting dome or a large curve display screen) could effectively constrain the illumination condition for the modeling. Moreover, they often use high-end, high speed, and well-calibrated camera system to capture the observed appearance. However, this custom hardware solution is expensive and often inaccessible, making

it impractical for different usages. Another major issue with appearance modeling-based methods is the accuracy of the global geometry. With the pixel-based modeling, it is capable of revealing highly detailed surface geometry features. Yet, in the original appearance formation, it only models the surface orientation, which usually represents in surface normal instead of depth information. In order to acquire $3D$ geometry, the further integration step is required. Over the years, many research efforts have spent on this problem [103], but still, it remains a significant gap from the surface normal to the exact depth information.

1.3. Summary and outline of the thesis

In this dissertation, we leverage commodity hardware systems such as DSLR camera, flashlight, display screen, and mobile system with computational imaging algorithms to improve the usability and robustness for different applications. Solving these problems in a mobile setting raises some additional challenges, and those already enumerated above. Compared to the state-of-art appearance modeling system such as the sophisticated and carefully calibrated light stage system [24], the inability to having well-constrain conditions such as the illumination as well as additional challenges in terms of the camera calibration increases the complexity of the problem. Although the significant hardware reduction poses critical challenges, we think the problem that we seek to address in this thesis could enable highly potential applications. For example, in the cultural heritage community, surface appearance modeling could be highly adopted as a collection survey and documenting tool.

Chapter 2 proposes an online calibrated, streamlined framework combining photometric stereo and photogrammetric information for a robust $3D$ acquisition. Chapter 3, we extend the framework into a single mobile device. Chapter 4 studies the surface shape estimation for a highly specular material with single mobile device. In Chapter 5, we extend the

multiview deflectometry with a differential rendering-based depth optimization framework to further improve $3D$ shape information. Lastly, we summarize the discussion contribution and highlight the potential applications that could be benefited by the proposed system in Chapter 6.

1.4. Disclaimer

The mathematical derivations and formulations, results and a large part of the text in the thesis are adapted from:

- C. Yeh, N. Matsuda, X. Huang, F. Li, M. Walton and O. Cossairt, " A Streamlined Photometric Stereo Framework for Cultural Heritage", In European Conference on Computer Vision, pp. 738-752. Springer, Cham, 2016.
- C. 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 International Conference on eScience Workshop on High Throughput Digitization for Natural History Collections (BigDig)
- F. Willomitzer, C. Yeh, V. Gupta, W. Spies, F. Schiffers, A. K. Katsaggelos, M. Walton, and O Cossairt, " Hand-guided qualitative deflectometry with a mobile device", OSA Optics Express 28, 9027-9038 (2020)

CHAPTER 2

Streamlined Near-Light Photometric Stereo Framework with Dual Cameras

2.1. Introduction

Computational Imaging techniques have been widely used for art history analysis and cultural heritage research in the last decade. Digital imaging technologies empower conservation scientists by revealing more information about works of art, helping to better preserve and protect their history for future generations. Accurate, automatic $3D$ surface recovery using only commodity cameras is particularly important for a number of applications in cultural heritage research. Since artifacts of historical significance are often located in public spaces or museums without the possibility of relocation to a laboratory environment, art conservators require $3D$ shape acquisition techniques that are portable, inexpensive, non-destructive, and fast, in order to uncover previously unknown information about artist techniques and materials. Two commonly used techniques that fit these requirements are Reflectance Transformation Imaging (RTI) and Photogrammetry (PG).

RTI is a visualization technique that allows users to probe the appearance of an artwork under arbitrary illumination conditions computationally, in a post-processing step. RTIs are created from multiple photographs of the object captured by a camera with fixed position and varying illumination. Researchers use RTI to virtually relight an object under arbitrary illumination conditions. Computational relighting can reveal fine details of the subject's $3D$ surface, for instance, when strong raking light is used to visualize the surface appearance.

However, because RTI is merely a visualization technique, it provides no direct access to depth information. Similar to RTI, Photometric Stereo (PS) is a well-known technique in computer vision and computer graphics. By observing the surface appearance and modeling the photometric invariant properties from the same viewpoint but under varying lighting conditions, it seeks to estimate the shape of the surface.

Alternatively, Photogrammetry (PG) uses images taken at different camera positions, using triangulation to compute $3D$ surface shape. Following the photo-consistency under the Lambertian reflectance assumption, Structure from Motion (SfM) techniques take feature points that common between/among multiple views to jointly solve for both $3D$ location of the surface points and the corresponding camera geometry. With the sparse point cloud and calibrated camera geometry, Multi-view Stereo (MVS) would interpolate per-view depth map to generate a low-resolution $3D$ mesh of the surface.

While there has been remarkable progress in both photometric stereo and photogrammetry, there are still some significant disconnection between theory and practice. For example, accurate PS normal output usually requires pre-calibrated lighting positions. In typical setups, this is achieved using either lighting with a fixed calibrated $3D$ geometry (*e.g.* a lighting dome) or by placing a reflective sphere in the scene to estimate incident lighting directions. $3D$ light position can be accurately pre-calibrated using a lighting dome, but this custom hardware solution is often inaccessible and sometimes impractical. A reflective sphere can accurately measure distant lighting but produces significant errors when light sources violate the far light condition and are actually located near the object (*e.g.* within 4 times the size of the object), typical of many PS capture setups [58]. PG techniques do not require controllable lighting but do require a high number of identifiable correspondence points in order to produce high resolution surface output, precluding the possibility of capturing low-texture

or single-material objects frequently found in a wide variety of natural scenes. Furthermore, at large standoff distances, depth precision for PG methods is relatively coarse, while PS solutions are capable of capturing highly detailed depth features.

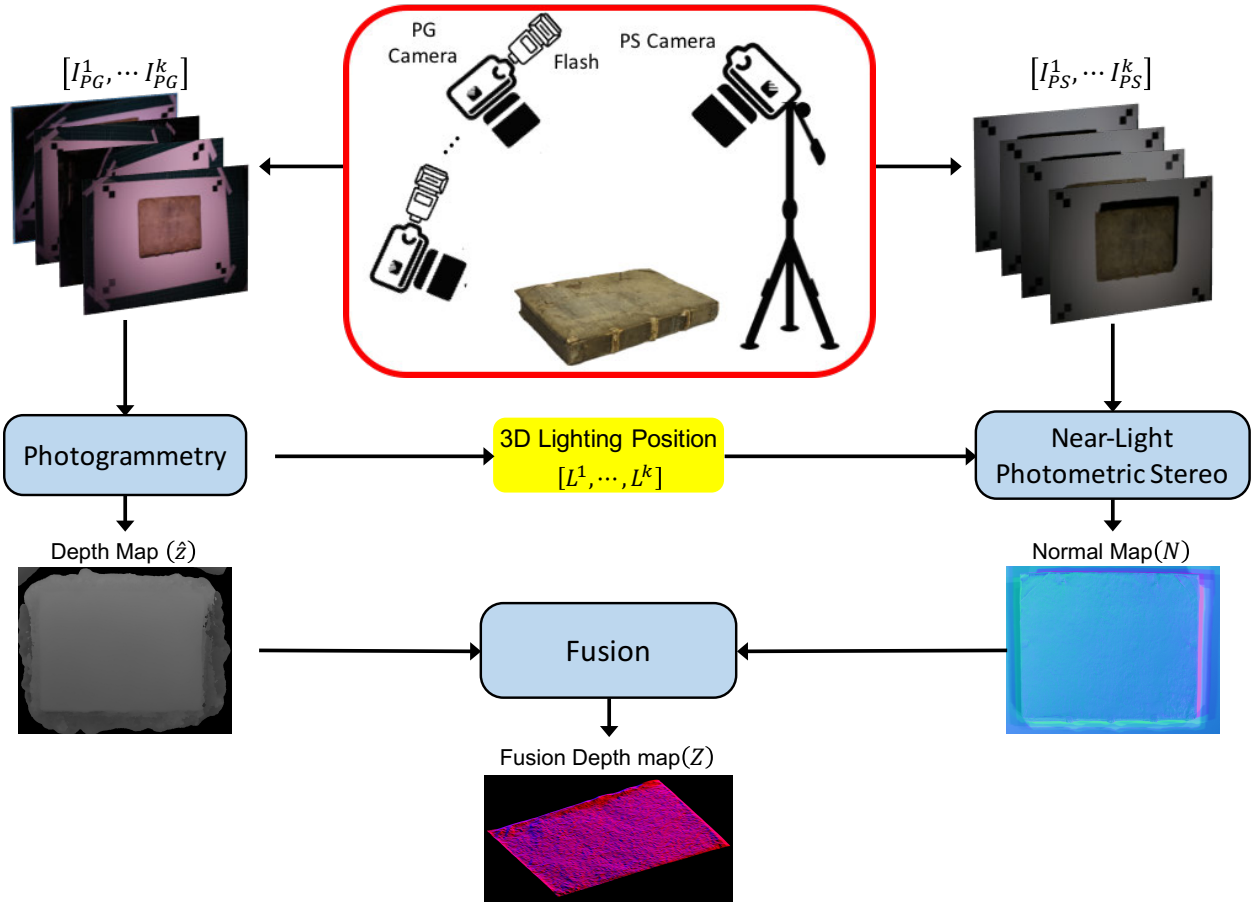


Figure 2.1. **Overview of the Streamlined photometric stereo framework for cultural heritage:** We use photogrammetry to find the $3D$ light positions $[L^1, \cdots, L^k]$ relative to a stationary photometric stereo (PS) camera. The estimated $3D$ light positions then allow us to compute accurate surface normal N from the PS camera. We fuse the computed normal map with a depth map $\hat{z}$, computed using photogrammetry, to generate globally accurate $3D$ shapes Z with high-quality micro surface details.

In this chapter, we present a robust $3D$ shape recovery capture framework for cultural heritage, as shown in Figure 2.1. Throughout the remainder of the chapter, we will refer

to these two cameras as the PS camera, capturing reflectance information from a fixed position, and the PG camera, affixed to the light source and capturing scene structure for photogrammetry from multiple views of the object. The PG camera images are processed using existing SfM algorithms to recover the camera position for each frame, and thus the lighting positions for the PS camera as well. Using these computed $3D$ lighting positions, we then produce an accurate PS normal map. Because we have generated a point cloud from the PG algorithm as well, we can fuse this sparse $3D$ information with the PS normal map to produce a $3D$ surface with both the fine surface detail typical of PS techniques and the absolute depth accuracy typical of PG techniques. The technique introduces minimal complexity beyond a conventional photometric stereo capture setup, yet can be used to significantly improve the accuracy of $3D$ surface reconstructions.

2.1.1. Contributions

- **A simple, robust $3D$ capture system:** We present a simple system for the free-form photometric stereo capture system using just two cameras with wireless synchronize triggers and an on-camera ring light. We show that our system simplifies reflectance capture and results in more accurate $3D$ surface reconstruction.
- **More accurate light position estimation:** Previous techniques estimate $3D$ light position directly from images from radiometric measurements [58], which are easily corrupted by shadows and specularities. In contrast, our light position estimation is based on geometric triangulation using SfM and is, therefore, largely independent of scene reflectance and illumination.
- **Improved near-light PS surface recovery:** Traditional PS techniques assume infinitely distant light sources. Under this assumption, the lighting direction can be

calibrated by placing a mirror ball in the scene. Our approach removes this far light assumption and eliminates the need for a lighting calibration object. Instead, $3D$ light position is estimated using a PG camera attached to the light source. We show that by accurately measuring the $3D$ location of the light sources, we can recover more accurate $3D$ surface shapes when using a PS setup that violates the far light assumption.

- **Large scale, high precision $3D$ reconstructions:** We show experimentally that our setup can be used to generate large field of view $3D$ shape reconstructions with high precision. This is done by fusing the fine details from dense normal estimation using PS, with the sparse $3D$ point clouds from our PG camera.

2.2. Previous Work

2.2.1. Reflectance Transformation Imaging

Reflectance transformation imaging is widely popular among art conservators through the use of the CHI RTI Builder and Viewer software suites [23]. RTI, originally known as Polynomial Texture Mapping (PTM), was first proposed by Malzbender [85] as a way to use a polynomial basis function for computational relighting. Later, the hemispherical harmonics (HSH) version [31] was introduced to reduce the directional bias in computational relighting results. Palma *et al.* [95] estimated normal from PTM RTIs by fitting the pixel intensity to a local bi-quadratic function of light angles and then setting the derivative to zero, which has the effect of finding the direction of the brightest pixel. Conservators use the CHI software to interactively explore image relighting and normal maps in the RTI Viewer, and also export those images offline for further research.

2.2.2. Photometric Stereo

In the original photometric stereo formulation introduced by Horn [53], light sources are assumed infinitely distant, the camera is orthographic, and the object surface is Lambertian reflectance and convex shape (*i.e.*, no shadows or inter-reflections). There are the vast majority of researchs are devoted to generalize the technique with the real-life scenes. Papadhimitri and Favaro *et al.* [96] tried to solve the ambiguity of the shape estimation with the perspective camera model. However, less attention has been paid to relaxing assumptions on the lighting model. Several researchers [132, 97, 58] recently investigated removing the far-light assumption to improve the accuracy of photometric stereo. Queau *et al.* [102] consider non-isotropic illumination to accurately model the real-world lighting condition. Shi *et al.* [115] recently came up with a benchmark dataset to quantitatively evaluate photometric stereo methods for general scene conditions.

2.2.3. Photogrammetry

Developed in the 1990s, this technique has its origins in the computer vision community and the development of automatic feature-matching algorithms from the previous decade. To determine the $3D$ location of points within a scene, traditional photogrammetry methods require the $3D$ location and pose of the cameras, or the $3D$ location of a series of control points to be known. Later, Structure-from-Motion (SfM) relaxed this requirement, simultaneously reconstructing camera pose and scene geometry through the automatic identification of matching features in multiple images [116, 118].

2.2.4. Combining Photometric Stereo and Photogrammetry

Although PS provides relatively accurate surface normal, it is still challenging to reconstruct a globally accurate surface shape. Some work has aimed to combine PG and PS techniques, such as the multi-view photometric stereo method by Hernandez *et al.* [32], which used RANSAC to estimate the light sources position and reconstruct $3D$ surfaces of Lambertian objects. For calibrated light sources, Birkbeck *et al.* [15] employed a variational method to estimate the surface and handle specular reflections using a Phong reflectance model. Ahmed *et al.* [6] used calibrated illumination and multi-view video to capture normal fields and improve the geometry templates. Wu *et al.* [141] performed a spherical harmonic lighting approximation to combine multi-view photometric stereo. Sabzevariuse *et al.* [108] used the $3D$ metric information computed with SfM from a set of $2D$ landmarks to solve for the bas-relief ambiguity for dense PS surface estimation. All of these algorithms require really critical environment constraint, either accurate light-source calibration under far light model or careful illumination design. Nehab *et al.* 's [91] hybrid reconstruction algorithm focused on leveraging Poisson system to combine depths and normal information. Their fusion algorithm produces high-quality reconstruction of $3D$ surfaces with a given parametric surface.

Our method relaxes the hardware setup constraints relative to these prior methods. To our knowledge, ours is the first system to work on the fusion between near-light PS model and PG. Besides having more accurate light position and surface normal estimates, our method also can leverage the surface estimate obtained using photogrammetry. By fusing PS and PG results, we can produce an improved $3D$ surface that retains the advantages of both PS and PG techniques.

2.3. Our Streamlined Photometric Stereo Framework

2.3.1. Hardware Setup

Our system setup consists of two Canon 5D Mark III DSLR cameras with $50mm$ Canon Prime lenses. One of these, the PS camera was affixed to a tripod above the imaging area. A Polaroid 18 Super Bright Macro SMD LED Ring Light was mounted to the PG camera lens. Both cameras were attached to a PocketWizard Flex TT5 wireless trigger system to ensure synchronized exposures. From a systems perspective, the cameras and flash components could be miniaturized (*e.g.* replaced with a point-and-shoot camera or even mobile system that comes with both camera and flashlight) to allow for greater freedom by the operator and quicker overall capture times. Lastly, a printed set of corner fiducial markers was affixed to the imaging area to provide a means to scale the PS and PG image sets to match the physical distances between the markers.

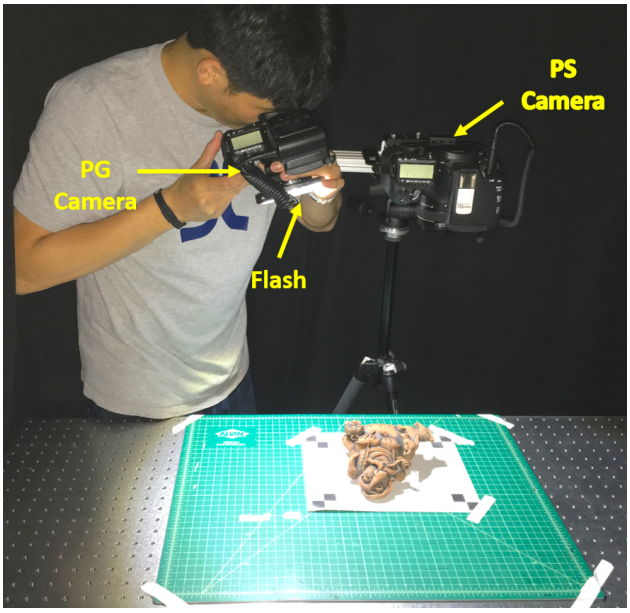


Figure 2.2. **capture setup:** We use two Canon $5D$ Mark III cameras with $50mm$ prime lens. The PS camera is placed about $0.5m$ away from the object.

2.3.2. Framework Work Flow

We begin by capturing an image at each of k different PG camera positions (see Fig. 2.2). A ring light is placed around the lens of the PG camera so that the centroid location of the illumination coincides with the optical center of the lens. The PG camera captures a set of images $[I^1_{PG}, ..., I^k_{PG}]$ of the scene from a unique viewing location. The PS camera also captures k images $[I^1_{PS}, ..., I^k_{PS}]$, but from a fixed position. For the PG camera, illumination is always aligned with the camera axis. For the PG camera, a diversity of illumination directions is captured. The PG images $[I^1_{PG}, ..., I^k_{PG}]$ are input into an off-the-shelf photogrammetry software Agisoft PhotoScan [5], which outputs the camera centers corresponding to the $3D$ light source positions $[L^1, ..., L^k]$. In addition, the software computes a sparse point cloud estimate of the objects $\hat{z}$. An image from the PS camera is input together with the PG camera images so that the extrinsic parameters from all cameras are determined in a unified global coordinate frame. Note that our PG images do not all have the same lighting and contain specularity and shadows under different lighting environments, none of which is ideal for typical passive multi-view stereo matching. However, we have sufficiently dense views under similar-enough lighting for the matching algorithm to find enough matching features between the images to reconstruct a photogrammetry model, which is accurate to within a few millimeters.

Next, the $3D$ light positions $[L^1, ..., L^k]$ are used as input into a PS algorithm to accurately recover normal and albedo based on the spatially-varying incident lighting position at each point p in the scene. To accomplish this, we solve a least squares problem to iteratively solve for the albedo a and normal N, given captured images $[I^1_{PS}, ..., I^k_{PS}]$ and corresponding 3D

light positions $[L^1, ..., L^k]$, similar to the work by Papadhimitri *et al.* [97]:

$$\arg\min_{a'_p, N_p} \sum_{p,k} \left(a'_p N_p^T \frac{\boldsymbol{L^k} - \boldsymbol{p}}{\|\boldsymbol{L^k} - \boldsymbol{p}\|} - {I'_{PS}}^k \right)^2, \tag{2.1}$$

where the near-light compensated pixel intensity

$${I'_{PS}}^k = I_{PS}^k \frac{\|\boldsymbol{L^k} - \boldsymbol{p}\|^2}{e_k}. \tag{2.2}$$

Finally, the PS algorithm generates a normal map $N = (n_x, n_y, n_z)$ for each pixel in the image. I, The relationship between the estimated normal and the depth map z is then $(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}) = (p, q)$, where $(p, q) \triangleq (-\frac{n_x}{n_z}, -\frac{n_y}{n_z})$. The PG algorithm produces a depth map $\hat{z}$ of the scene only for a sparse subset of pixels. We assume $\hat{z}$ is transformed to the PS camera frame using the extrinsic parameters computed from the PG/SfM software. We then recover the PS-PG fused depth z_i for each pixel i by solving the following least squares problem:

$$\begin{aligned} \text{minimize} \sum_{i \in I} \left\| \begin{bmatrix} \partial z_i / \partial x_i \\ \partial z_i / \partial y_i \end{bmatrix} - \begin{bmatrix} p_i \\ q_i \end{bmatrix} \right\|_2^2 + \lambda \sum_{i \in \hat{I}} (z_i - \hat{z}_i)^2 \\ = \|\nabla Z - \Gamma\|_2^2 + \lambda \|MZ - \hat{Z}\|_2^2, \end{aligned} \tag{2.3}$$

where $Z, \hat{Z}, \text{and} \Gamma$ are the lexicographically vectorized versions of $z_i, \hat{z}_i, \text{and} (p_i, q_i)$, ∇ is the gradient matrix, M is a binary selection matrix that only selects the pixels that have valid PG depths, and λ is the parameter depends on the confidence of PG depth.

Note that this formulation does not rely on any linear constraints or statistical priors; it is simply a weighted least-squares approach that attempts to satisfy, on average, the conditions observed by both the PS and PG recovery techniques. A wide variety of variations on this optimization could be employed depending on the type of object and intended usage of the

recovered surface, but a detailed analysis of such possibilities is beyond the scope of this paper. We simply aim to demonstrate that the combination of both sets of simultaneously captured data, even with a rudimentary approach to optimization, characterizes the surface significantly better than either approach alone.

2.4. Experiments and Results

2.4.1. Light Position Estimation

First, we evaluated the accuracy and stability of our PG camera-based method for light position estimation. In order to compare to known physical lighting positions, we affixed the tripod mount of the PG camera onto an optical mounting post, which we then inserted sequentially into optical post holders at known locations on an optical table. Though we do not consider this manual procedure sufficient to provide ground truth data, the sub-millimeter tolerances of the machined optical table and mounting posts can demonstrate the extent to which the recovered lighting positions can be relied upon.

In Table 2.1, we repeated the three fixed lighting positions 5 times, which resulted in an average error relative to our measured positions of less than $10mm$, or well under 1% error. The standard deviation of these values was less than $1mm$, indicating good repeatability of the technique. Compared to Huang *et al.* [58] using image intensity to estimate the lighting position, our approach using PG/SfM has more accurate lighting position estimation for a near-light photometric stereo model.

	P_1	P_2	P_3
Ground truth distance (mm)	869.98	896.28	756.12
ΔL (Our method) (mm)	1.25	9.09	8.27
ΔL (Huang *et al.* [58]) (mm)	242.46	239.76	216.11
SD δ (Our method) (mm)	0.26	0.23	0.09
SD δ (Huang *et al.* [58]) (mm)	2.09	4.18	2.09

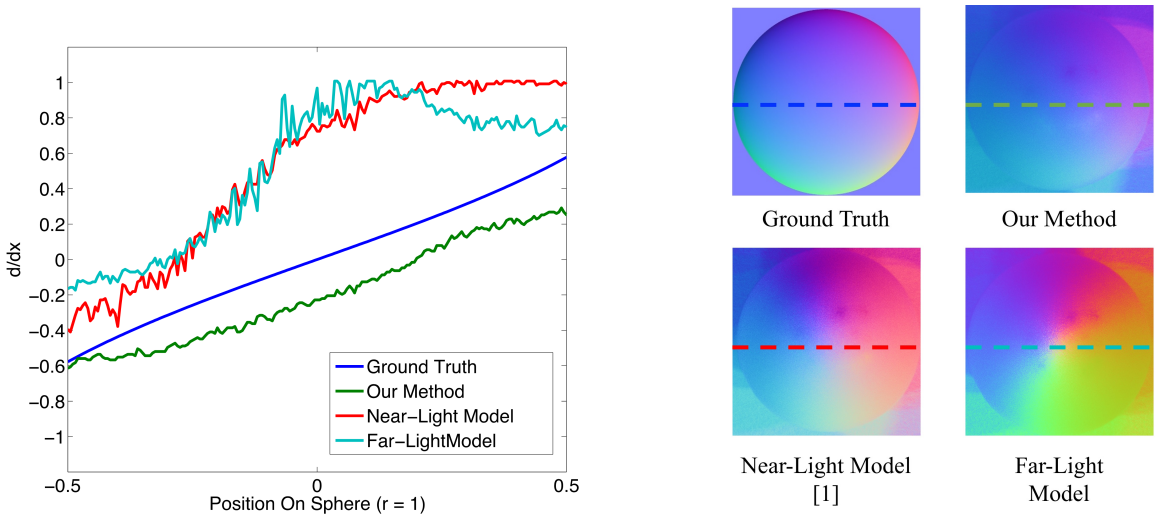
Table 2.1. **Measure value** *v.s.* **ground truth for three light positions P_1, P_2, and P_3:** The first row shows the ground truth distance between the PG camera optical center and the $3D$ location of light sources P_1, P_2 and P_3. The second and third rows show ΔL values for our technique and that of Huang *et al.* [58], respectively. The ΔL values reported are the distance between the estimated $3D$ position of the PG camera's optical center, and the ground truth $3D$ position, averaged over five measurements. The fourth and fifth rows report the standard deviation of the distance between the estimated and ground truth $3D$ location of the PG camera.

2.4.2. Normal Map Accuracy

To confirm that PG lighting position estimation produces a more accurate PS normal map, we compare normal map recovery for a sphere using our method, the near-light model in Huang *et al.* [58], a conventional distant-light PS model, and ground truth.

Figure 2.3 shows the X-component of the normal map sampled through the center of a sphere for the ground truth, conventional distant-light PS model, near-light model from Huang *et al.* [58], and our PG light estimation. Our method clearly demonstrates increased fidelity in normal map estimation.

This method is a unique use case for PG techniques in surface reconstruction because it can be applied to textureless objects that would normally be a failure case for PG. So long as there are sufficient correspondence features to perform bundle adjustment somewhere in



(a) N_X Comparison (b) Normal Map

Figure 2.3. **Normal Map Accuracy for a sphere:** Comparison between the x component of the estimated normal map for a sphere. The ground truth (shown in blue) normal for the sphere closely resembles a line (the gradient of a parabola is exactly a line). The normal estimate computed using the far light assumption (shown in cyan) and the uncalibrated photometric stereo method from Huang *et al.* [58] (shown in red) both produce significant errors. Our method (shown in green) accurately estimates $3D$ light position, and therefore produces the most accurate $3D$ normal.

the PG camera field of view, our technique will produce accurate lighting positions, and thus more accurate normal maps, regardless of the amount of texture in the target object.

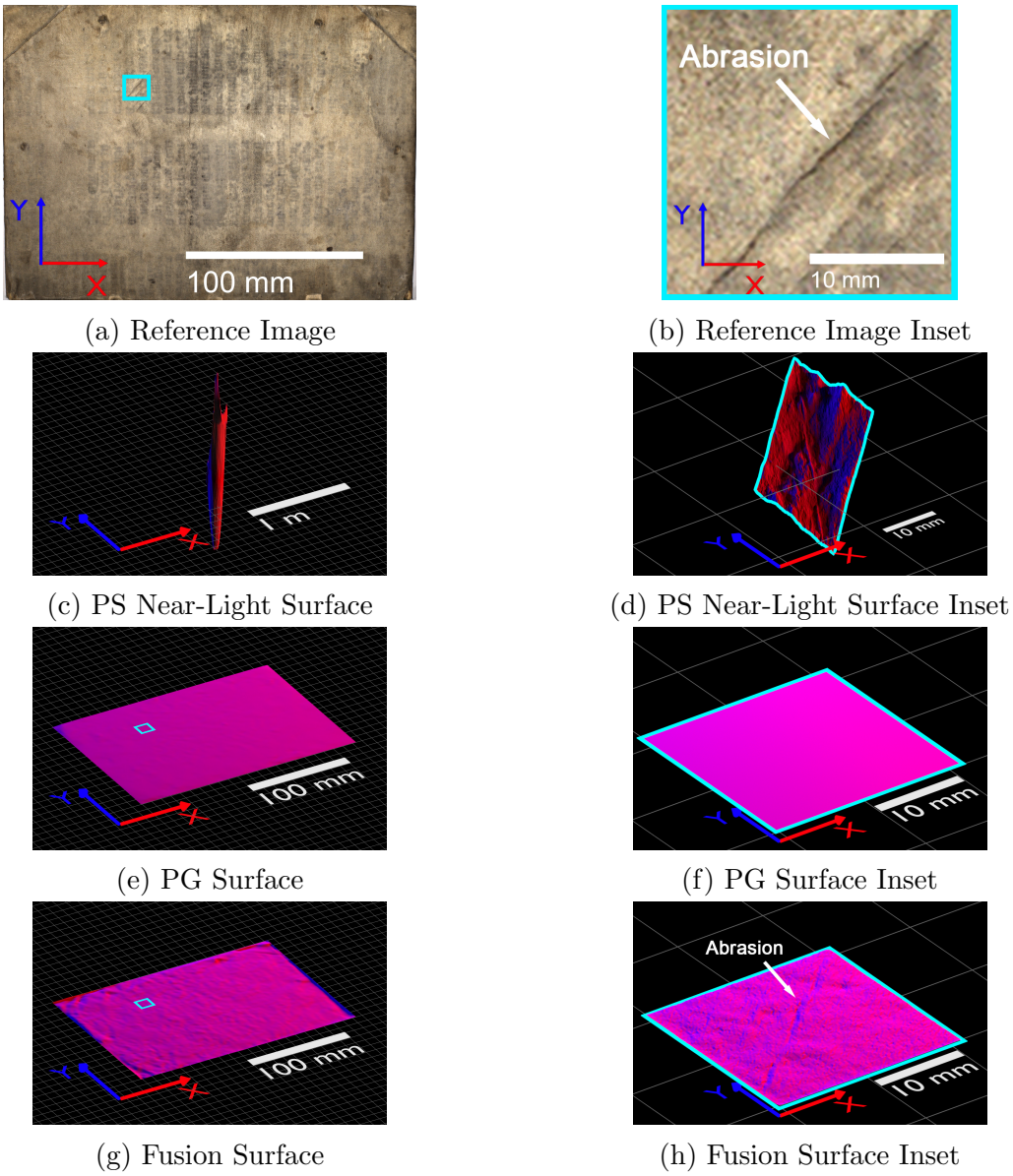
2.4.3. Fusion Surface Reconstruction

When objects have enough surface texture for the PG algorithm alone to produce a sparse point cloud, we can leverage this data for more globally accurate surface reconstruction. Surface shape recovery remains a significant challenge for all PS techniques since small errors in normal recovery will produce incorrect geometry upon integration, and the absolute position of the surface can never be recovered. The formulation in Equation 2.3 retains the fine surface detail recovered by PS and the gross geometric shape recovered by PG.

We chose to test the visual fidelity of surface fusion reconstructions using a cultural heritage object from our University's rare book collection, an object representative of the intended use case for this technique. Shown in Figure 2.4, this 16th-century reprinting of Hesiod's 'Works and Days', was covered with a reused parchment from an early manuscript that was scraped down to remove the letters from the top surface. Small ridges on the surface are aligned with the direction of the scraping motion. We hope to observe these abrasions in the context of the largely flat overall surface geometry. PS techniques alone will not retain the course flatness but will reveal the small ridges, while PG techniques alone will retain the flat surface but will not resolve the ridges at all. This object is thus an example of a surface of our PS and PG fusion technique is well suited to recover.

Figure 2.4. **Test object:** a $16th$ century book covered with reused parchment. Small surface abrasions on the surface are of interest to historians.

The λ parameter in Equation 2.3 was set to 0.15, a value found experimentally that retained surface detail while preventing the large-scale PS errors to propagate into the final output.



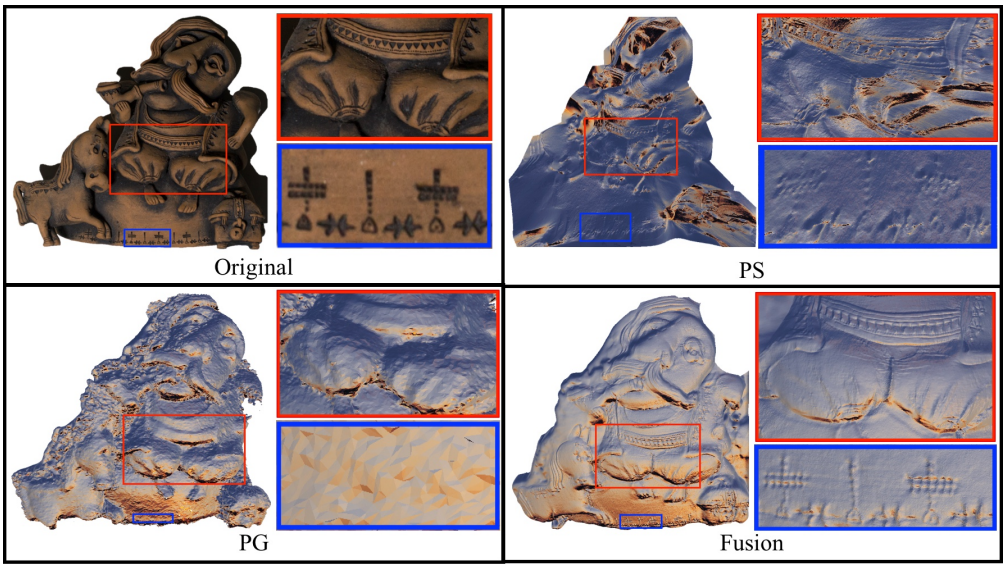
(a) Reference Image (b) Reference Image Inset

(c) PS Near-Light Surface (d) PS Near-Light Surface Inset

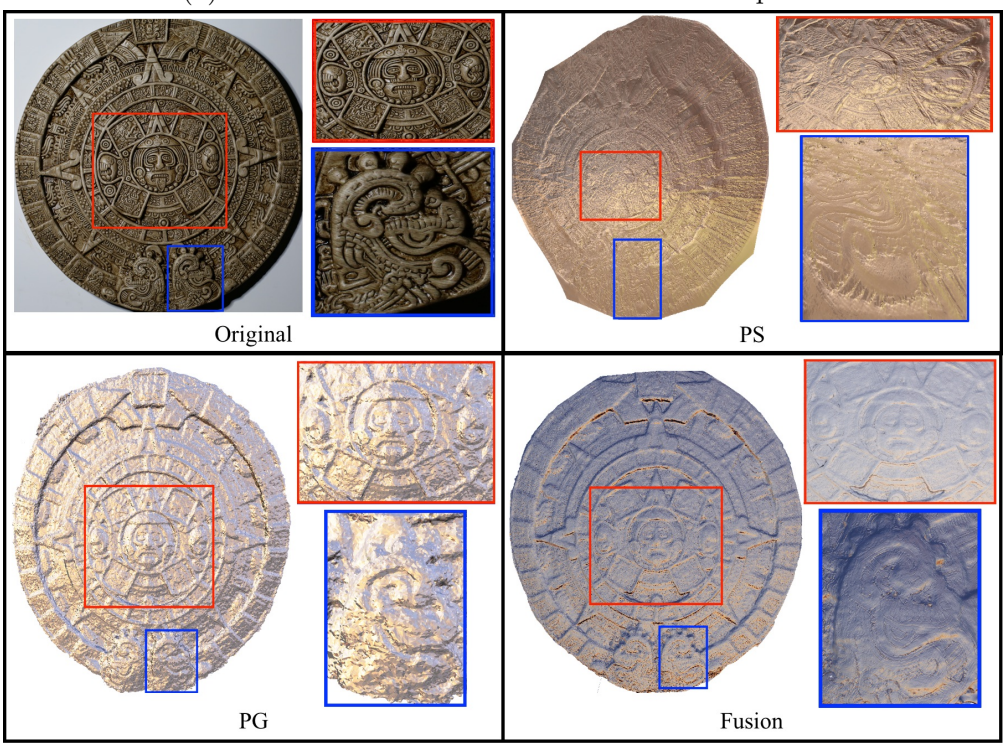
(e) PG Surface (f) PG Surface Inset

(g) Fusion Surface (h) Fusion Surface Inset

Figure 2.5. **Reconstruction Results:** Comparison of reconstruction methods on a $16th$ century book shown in (a), and hi-resolution inset (b), corresponding to the outlined region to the left. After surface recovery, these results are depicted in orthographic perspective and illuminated by a red directional light along the x-axis and a blue directional light along the y-axis to reveal surface details without exaggerating the scale of the z-axis. The PS reconstructions using the method from [58], shown in (c) and (d), exhibit severe global geometry errors due to lack of absolute reference points (the scale in these images were reduced to accommodate the extreme range of z-axis values). PG output from Agisoft Photoscan is shown in (e) and (f). Our fusion results, produced by optimizing the surface for consistency with both PS and PG results are shown in (g), (h). Note that the fusion results exhibit a balance of course geometric accuracy (a flat book surface) while retaining small surface variations.



(a) Hutsul Ceramics Ukraine Terracotta Sculpture



(b) Chinese Jade BI Carvings

Figure 2.6. **Experimental Results using our Framework:** We tested our framework on several objects with complex geometry and fine surface detail. These objects demonstrate that our system produces a good balance between global geometric accuracy and micro surface details. $3D$ reconstruction results using only photometric stereo (PS), and photogrammetry (PG) are shown for comparison. Our fusion results clearly demonstrate superior $3D$ reconstruction quality.

In Figure 2.5, we show side-by-side comparisons between the full surface and an inset revealing small details. The top row contains a reference image from the PS data set - the full book surface on the left, followed by the pink inset region expanded on the right. These regions are used in subsequent rows, where surface reconstructions are depicted in orthographic renders using a white Lambertian material and raking angle lights to highlight surface variation in blue along the y-axis and red along the x-axis. The $2nd$ row shows the surface output from the photometric stereo algorithm, which despite recovering small surface details, exhibits extreme geometric errors, which would significantly limit any objective analysis based on the surface height. The $3rd$ row shows the PG surface mesh output from Agisoft Photoscan. The PG results correctly recover the general flatness of the object, but lose all fine surface detail. Finally, in the bottom row, we show our optimized PS+PG fusion results. We retain both the overall flat shape of the book surface while recovering the small wrinkles and abrasions present on the surface of the book.

In order to test our framework in more general settings, we have tested our method on several additional objects with complex geometry and fine surface detail. As shown in Figure 2.6, our framework produces accurate $3D$ reconstructions that maintaining both global accuracy and high precision. The results are far superior to $3D$ reconstructions using either PS or PG alone.

2.5. Conclusion and Future Work

We have presented a new technique using a PG camera attached to a flashlight source to estimate $3D$ lighting positions for more robust photometric stereo $3D$ surface reconstructions. The resulting light position estimates are more accurate than conventional far-light

directional estimates or near-light position estimates and consequently produce more accurate normal maps. We also demonstrate that the PG surface information can be fused with the PS normal map output for surface reconstruction that retains both the fine details from PS and accurate global geometry from PG. We have demonstrated how to use a simple setup to acquire high quality $3D$ reconstruction results of several cultural heritage objects. Our initial results also give rise to another question: if fusion between poor normal recovery and good PG data produces a reasonable result, is the improved PS performance by accurate light position estimation even necessary? Further analysis is necessary to conclusively compare our results to fusion results that do not attempt to improve PS performance, but we believe that at the very least better input data from PS will not perform worse than other fusion methods and is likely in most cases to perform better. We hope our method will empower conservators and conservation scientists with new tools for simple, inexpensive, $3D$ acquisition of cultural heritage artifacts. It is our belief that doing so will open the doors to new applications in monitoring the deterioration of objects and help inform new methods of damage prevention and preservation.

There are several possible directions for future work. Photometric stereo, technically, is a fix-view $2.5D$ reconstruction method that could not deal with the scene with lots of depth changes. In the future, we are interested in merging multi-view information to account for artifacts that photometric stereo can create and produce a high-quality surface detail model. On the other hand, our light source estimation method could be extended to non-point or non-isotropic light sources, an extension applicable to nearly all real-world use cases. By performing PG camera pose estimation on both the PS camera and the PG camera, the full surface of a convex or more complicated surface shape may be recovered. From a systems perspective, the PG camera and flash component could be miniaturized (*e.g.* replaced with

a point-and-shoot camera) to allow for greater freedom by the operator and quicker overall capture times. Two or more of these camera/flash units could be synchronized and processed to capture bidirectional reflectance information and ultimately used to recover more sophisticated material characterization jointly with surface shape. We are also interested in investigating more sophisticated PS algorithms that can handle difficult cases such as shadows and non-Lambertian reflectance. Last but not least, although it is quite difficult to have a real ground truth to benchmark a $3D$ reconstruction system, we would still like to compare our framework with state of the art $3D$ acquisition method on cultural heritage application in the nearly future.

CHAPTER 3

Uncalibrated Photometric Stereo with Single Mobile Device

3.1. Introduction

3D imaging techniques have had explosive growth in both industry and academic research during the last decade with a variety of applications. In the community of cultural heritage, 3D imaging has gained widespread popularity as a tool for documenting object condition [71, 23, 51, 58, 32, 22, 110]. 3D imaging methods can be loosely divided into two groups: passive and active 3D imaging. Passive based 3D imaging, such as photogrammetry [117, 39] relies on the reflected radiance from an object lit with ambient illumination to reconstruct the object's 3D surface shape. Active 3D imaging, such as photometric stereo (PS) [55], uses a controlled light source, such as a flashlight, to illuminate the object and recover the 3D surface shape. PS is a highly sensitive technique that is capable of recovering 3D surface shape information on the scale of micrometers. For this reason, it has been widely used for the visualization of works of art and artifacts. While PS has been explored extensively, it still faces many fundamental challenges that limit its ease of use and has prevented its widespread adoption as a collection survey tool in the cultural heritage community.

PS estimates the surface normal/shape from photos taken by a fixed position camera but with varying lighting position and direction. By modeling the measured image intensity as a function of the incident lighting angle, one can recover the surface normal and material reflectance of each point on the object. The depth information of the object can then be recovered by integrating the reconstructed surface normals. The material's reflectance can

also be interactively manipulated by the users for the purposes of visualization and may be integrated into virtual reality (VR) head-mounted displays (HMDs) or augmented reality (AR) displays. Conventional PS reconstruction techniques use a far-light assumption, which assumes the position of the illumination light source is infinitely far away from the object. In practice, this assumption is frequently violated due to the space limitations of the acquisition setup. Without proper correction of the forward model, the violation of this assumption introduces errors in the reconstruction of the surface normal.

In 2015, Huang *et al.* [58] introduced a near-light model that automatically estimates the light source positions at the same time as the surface normal estimations increasing the accuracy of both. However, to achieve these results, a light dome and high-end digital single-lens reflex (DSLR) cameras, remote triggers, and flashlights were required. In this chapter, we ask the question: Is it possible to achieve the same results as achieved in [58] using a simplified imaging setup that is the user-friendly operation and portable enough to be used in remote locations?

To this end, in Chapter 1, we proposed a streamlined framework with a dual-camera setup and near-light model for PS reconstruction [145]. In this setup, DSLR cameras and a flash are synchronized together: a fixed camera is used for PS capture, and a second camera is attached to a flash to estimate lighting direction. A sequence of photos is taken as the second camera is moved along an arbitrary path in 3D space. Detected features on the object in each image are fed into a Structure from Motion (SfM) algorithm, which recovers the pose of the second camera, and in turn, provides the 3D position of the flash fixed to this camera. High-quality surface normals are then recover using a near-light PS algorithm, such as the one proposed in [58]. In addition, a course point cloud recovered from SfM is fused with the recovered surface normals, producing a high-quality depth map.

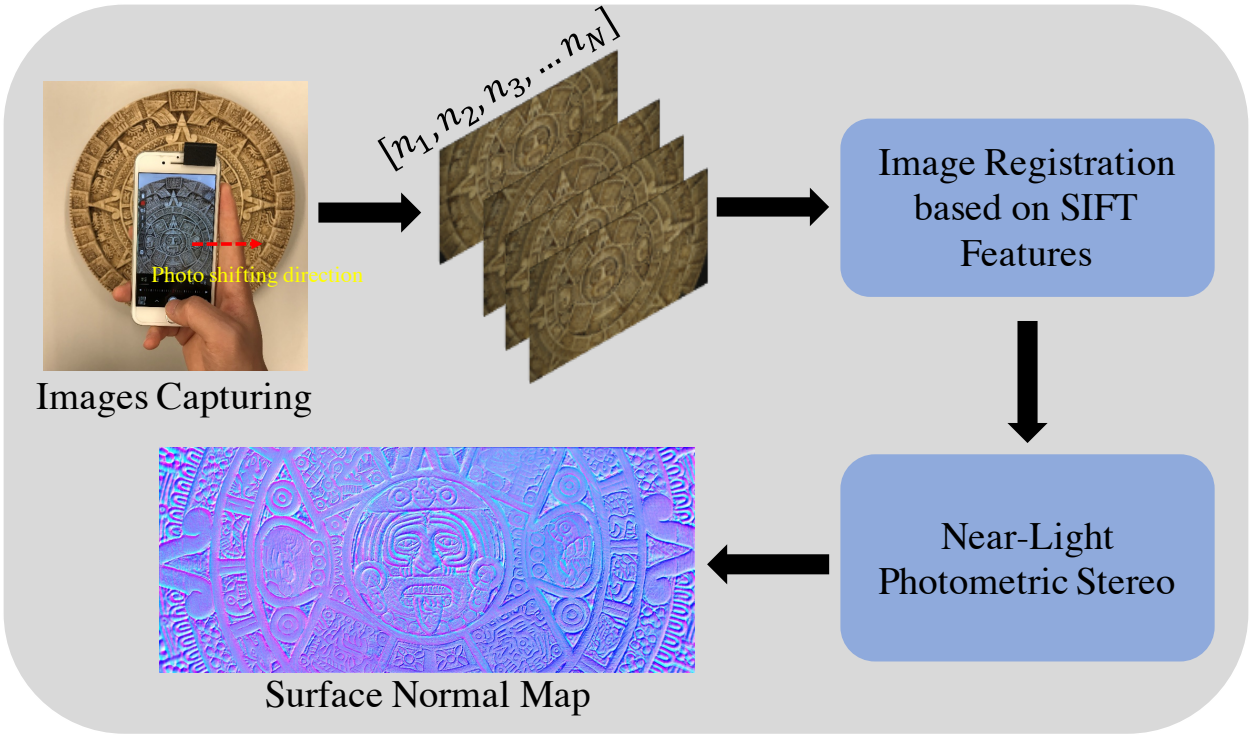


Figure 3.1. **Overview of Shape from Shifting:** Our shape-from-shifting technique uses a mobile device camera to capture images around the object with the built-in flash used as a source of illumination. SIFT-based image registration renders the images to the same viewpoint but each with a unique illumination direction. The synthesized images are further processed by un-calibrated photometric stereo to acquire dense surface normal vector maps.

In this chapter, we try to further relax the system complexity. We draw inspiration from this previous work to simplify photometric stereo acquisitions and innovate a simpler un-calibrated method for surface acquisition of works of art. Here we introduce a near-light PS technique that uses just a single camera for surface normal reconstruction. We propose mobile shape-from-shifting (SfS), a robust 3D surface shape recovery framework that can be used on a mobile device as shown in Figure 3.1. SfS uses just a mobile device and a custom 3D printed widget fitted with crossed polarization filters placed respectively in front of the camera lens and on the flash LED. Hand-held capturing is possible due to the compact

size of the mobile device. The polarization filters are used to separate diffuse and specular reflectance, in order to suppress normal reconstruction errors caused by specular reflections. During capture, users simply turn on the built-in flash and capture a sequence of images of the object surface. These images are pre-processed using scale-invariant feature transform (SIFT) [82] to register object features in each frame. Then, a near-light PS algorithm is used to recover surface normals from these pre-processed images. In Table 3.1, we summarize and compare the proposed Mobile SfS with the 3D imaging techniques frequently used in cultural heritage. Our proposed method significantly reduces the complexity of PS acquisition so that images may be captured in nearly any setting. We show that our method produces similar quality 3D surface normal reconstructions to those achieved using a lighting dome in a laboratory.

Imaging Techniques	**Price**	**System Complexity**	**Acquisition Time**	**In the Wild**	**Portable**
Laser Scanner [71]	$$$	High	Hours	Yes	Yes
PS with Dome [25]	$$$$	Medium	$\sim$ 30 Minutes	No	No
Mobile SfS	$	Low	$<$ 1 minute	Yes	Yes

Table 3.1. Comparisons of specs between our proposed Mobile SfS and two standard techniques.

3.1.1. Contributions

- **A novel PS surface normal reconstruction framework (SfS):** In this chapter, we propose a novel PS surface normal reconstruction framework that uses only a

single camera and flash on a mobile device. We develop an image processing pipeline and near-light PS reconstruction algorithms for the novel framework.

- **A Simple, Cost-effective Solution:** Our technique requires only a mobile device. Optionally, an add-on widget may be used to increase performance for highly specular objects. Previously, polarization has been widely used in light dome setups [25] to improve normal reconstruction, but calibrating polarizing filters is time-consuming and error-prone. Mobile SfS uses a 3D printed widget with just two polarizers, making it cost-effective and easy to use.
- **Portability and Accessibility:** Since Mobile SfS only requires a mobile device, such as an iPhone, and a small widget, it is very portable and user friendly. We believe Mobile SfS will be a powerful tool for conservators because it drastically simplifies the 3D surface acquisition process by allowing objects to be scanned in their natural environment and without the need of calibration hardware such as a mirrored ball.
- **Mobile SfS for different materials and scales:** In this paper, we have demonstrated that Mobile SfS can work for ceramic, stone, and paper objects of various physical proportions.

3.2. Related Works

3.2.1. Image-Based modeling and Photogrammetry

Developed in the 1990s, image-based modeling is a technique that utilizes a collection of images to create a three dimensional model [27]. To determine the 3D location of points within a scene, traditional photogrammetry methods require the 3D location and pose of the cameras, or the 3D location of a series of control points to be known. Structure-from-Motion

(SfM) [48] removed this requirement, simultaneously reconstructing camera pose and scene geometry through the automatic identification of matching features in multiple images.

3.2.2. Photometric Stereo

Photometric stereo [140] has often been used to recover high fidelity surface shape from image intensity. Original PS formulation assumed Lambertian reflectance and single color point light source. However, those constraints make it impractical to real-world conditions. Anderson *et al.* [8] adapted multi-spectral light source into the illumination model. Georghiades [42] used Torrance and Sparrow [128] to model both diffuse and specular reflectance information with uncalibrated PS. Zickler *et al.* [150] recovered specular surface with photometric invariants under known illuminant color. Researchers also tried to use different cameras and reflectance model to attack the problem. The recent work of Mecca *et al.* [88] considered a perspective camera model to deal with non-Lambertian effects. Khanian *et al.* [64] combine Blinn-Phong [16] reflectance model and perspective projection to deal with complex real-world applications with many specular highlights.

3.2.3. Separation of Diffuse and Specular Reflections

Surface reflectance is a well-studied research problem in computer graphics. In 1985, Shafer [114] proposed the Dichromatic Reflectance model and utilized color images analysis to separate surface reflection into "diffuse" and "specular" components. Because of the different spectral distributions for diffuse and specular reflection under dielectrics, the method easily separates them in RGB color space. Klinker *et al.* [65] also developed a method based on color histograms. To get more accurate and concise results, researchers kept pushing the color-based methods to the limit. [69, 76, 75, 84, 122, 123, 146] In addition to color-based techniques,

several hardware-based approaches have been introduced. Lamond *et al.* [68] use controlled illumination to exploit specific frequency behaviors of reflectance functions for separating diffuse and specular components. Polarization can also be used to separate diffuse and specular reflections. Wolff [138] demonstrated the use of cross-polarized filters, using two images captured with vertical and horizontal polarizers in front of the camera to efficiently separate diffuse and specular reflection components. Nayar *et al.* [89] combined polarization and color information to separate diffuse and specular reflections. Significant work from Ma *et al.* [83] described the use of polarization differential images and spherical gradient illumination to perform photometric stereo for acquiring high-quality surface normal with a small number of images.

3.3. Mobile SfS

3.3.1. Hardware Setup and History Collection Samples

Hardware Setup:

Two polarizers with opposite polarization directions (marked with blue boxes in Figure 3.2(a)) were attached to the camera and flashlight on an iPhone 6 through a custom 3D printed add-on component as shown in Figure 3.2(b). Specifically, these two polarizers were cut from the same polarization film. The polarizer on the flashlight was first glued on the add-on component, and the orientation of the polarizer on the camera was carefully tuned and fixed at the position of extinction. The total cost for the polarizers and 3D printed component is less than 5 USD.

Historcal Artworks:

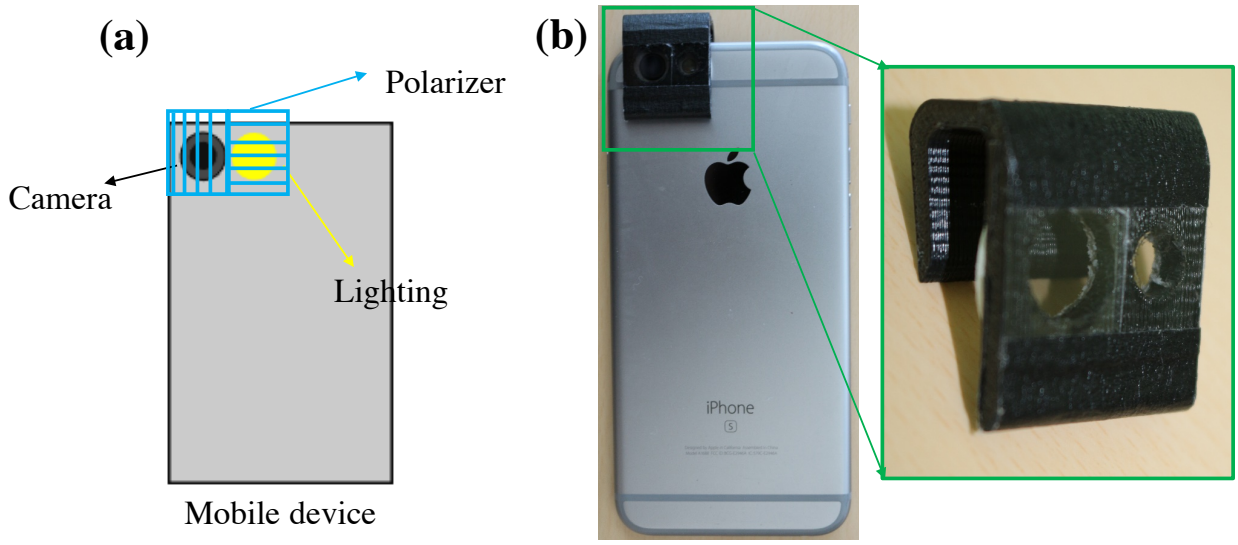
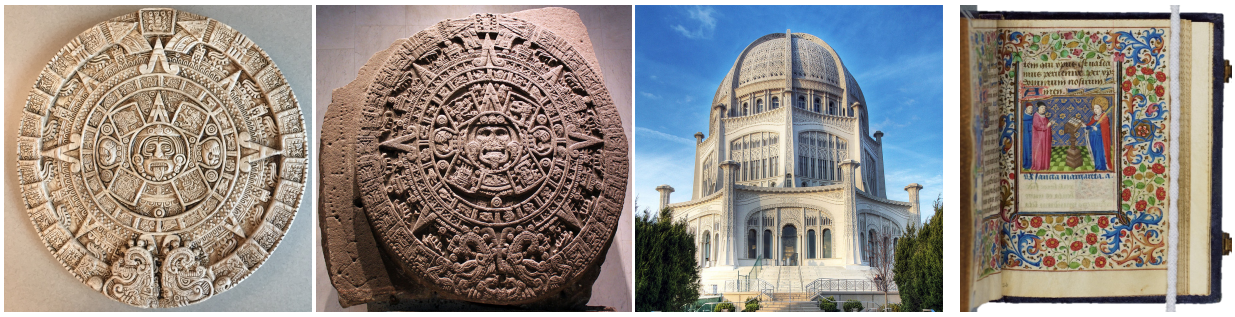


Figure 3.2. **Mobile SfS hardware:** (a). two polarizers with opposite polarization are attached to the camera and the flash light of the mobile device, respectively; (b). photo of our prototype Mobile SfS with an iPhone 6 and a custom 3D printed widget.



(a). A duplicate of Aztec calendar stone (b). Original Aztec calendar stone (c). Bahá'í Temple (d). French illuminated manuscript

Figure 3.3. **Samples for evaluation with Mobile SfS:** A duplicate (a) of Aztec calendar stone (b); a portion of the wall of Bahá'í Temple (c); and a page of an old manuscript (d).

Three artworks with different materials and scales were evaluated with our prototype Mobile SfS. The first sample is a duplicate (Figure 3.3(a)) of Aztec calendar stone [1] which dates back to 15^{th} century in Mexico and is housed in the Mexico National Anthropology

Museum as shown in Figure 3.3(b). The second sample is a portion of wall of Bahá'í Temple [2] (Figure 3.3(c)) built in the 1930s and located in Evanston, IL. The third sample is a parchment page of a French illuminated manuscript, as shown in Figure 3.3(d), Suffrages from a Book of Hours [10], dating from the 1460s-1490s, which belongs to the permanent collection of the Isabella Stewart Gardner Museum in Boston.

3.3.2. Acquisition Procedure

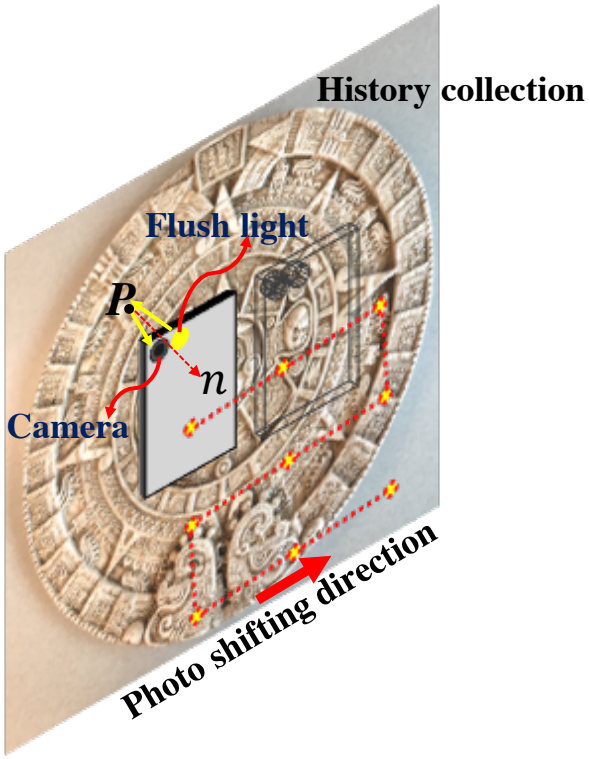


Figure 3.4. **Acquisition procedure with Mobile SfS:** Hand-held Mobile SfS faces to the object. Slightly shift the phone and take one image at each position. Nine images are taken in this paper for the surface normal reconstruction.

As shown in Figure 3.4, the phone is held by hand at about 30 cm from the samples (depending on the size of the imaging area and the scale of the surface profile). A sequence

of photos (9 images in this paper) are captured by sequentially moving the camera marked with a dashed line in the figure. During capture, the focus, white balance, and exposure of the phone camera are set to manual. Due to the limited power of the iPhone flash, the examples shown were captured in a dark room or in the evening to minimize the effect of ambient light.

3.3.3. Reconstruction Algorithms

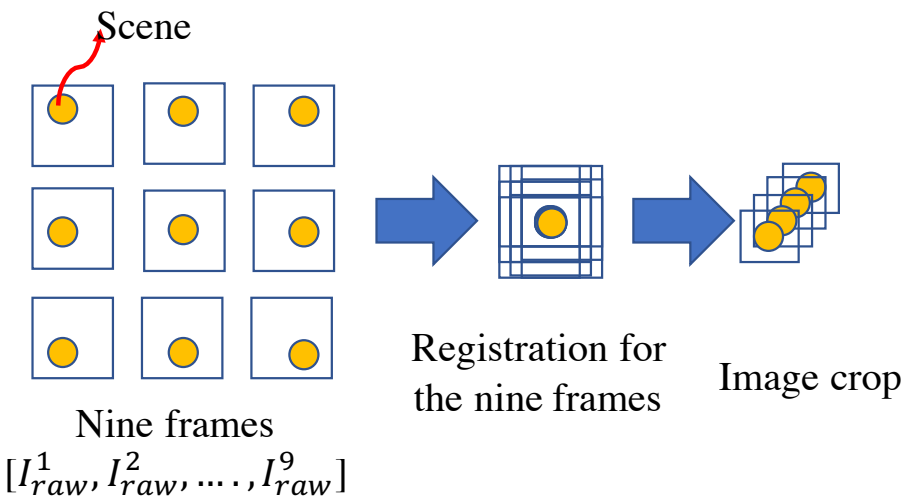


Figure 3.5. **Image pre-processing:** Step 1: nine images are captured; Step 2: images are registered with the SIFT function; Step 3: images are cropped to display the same region.

Using this pipeline, we have acquired a set of diffuse reflection images where each surface point is illuminated by k lights and k different viewpoints $[I^1_{raw}, ..., I^k_{raw}]$. In order to establish the conditions necessary to measure the surface shape by photometric stereo, all the images need to be captured from the same point of view. Likewise, a Lambertian surface will diffusely reflect light with an intensity proportional to the cosine of the illumination angle, regardless of the observer's angle of view. We meet these requirements by performing a

geometric transformation to render each captured image with a single viewpoint, as shown in Figure 3.5.

The transform is calculated by identifying a set of features in each of the captured images $[F_1^1, ..., F_j^k]$ using the SIFT feature detection algorithm [82]. Correspondences between features in all images, as well as identification of outliers, are found using RANSAC [35]. With these data as inputs, the transform between the source and target images is used to estimate a homography matrix $[H_1, ..., H_{k-1}]$. Once each image is transformed through the homography matrix, the collection of images is thus registered $[I_{reg}^1, ..., I_{reg}^k]$ to the target viewpoint.

Next, the registered images $[I_{reg}^1, ..., I_{reg}^k]$ are used as the input into a near-light uncalibrated PS algorithm to accurately recover surface normal vector and albedo images. Similar to Xiang *et al.* [58], we assume the mobile phone camera to be linear and model surface intensity by minimizing the following energy function:

$$E(\boldsymbol{L_k}, e_k, N_p, \boldsymbol{p}, A'_p) = \sum_{p,k} \left(I_{pk} - \frac{N_p^T(\boldsymbol{L_k} - \boldsymbol{p})}{\|\boldsymbol{L_k} - \boldsymbol{p}\|^3} e_k A'_p \right)^2 . \tag{3.1}$$

where p is the each surface point, $\hat{N} = (\hat{N}_x, \hat{N}_y, \hat{N}_z)$, $A'_p = (A'_{px}, A'_{py})$, denote the surface normal and albedo, L_k denotes the 3D position of the k-th light source and e_k is represented the lighting intensity. To simplify and accelerate the reconstruction process, we also take the 3D position from homography estimation as the initial 3D lighting position. The 3D lighting position would interactively update while estimating the albedo and surface normal.

Note that conventional photometric stereo usually requires more than 10 images to get quality surface reconstruction. With Mobile SfS, we are able to use only 9 images to reconstruct the surface geometry.

3.4. Experiments and Results

Three case studies using artworks, as described above, are performed with our prototype Mobile SfS setup to demonstrate the reconstruction of surface normal vector maps and its versatility on a variety of materials.

3.4.1. Plaster Replica of an Aztec Calendar Stone

To qualitatively evaluate our mobile SfS, we first performed the experiment with the replica of an Aztec calendar stone and compared our normal map reconstruction to one captured, from the same object, in a light dome. The Mobile SfS device was placed about 30 cm away from the object surface, facilitating a small region of interest to be captured but at high resolution. Larger fields of view can be captured, albeit with the trade-off of lower resolution. The same region of the stone was imaged with the light dome. We also compared the normal map reconstruction with the same mobile phone, with and without polarizers. Sequences of nine images are used for the normal map reconstruction for our Mobile SfS, while eighty-one images are used for the normal map generation for photometric stereo with a light dome.

In Figure 3.6(a) an RGB photo of the imaged portion of the stone is shown with a close-up view of the nose region. Normal maps produced respectively from a light dome and by Mobile SfS without polarization filters are shown in Figure3.6 (b) and (c). Compared to the normal map from the light dome (Figure 3.6(b)), the normal map generated using the proposed method (Figure 3.6(d)) retains the global shape information of the test sample. There are, however some notable differences in all three captures. Specularly reflected light from the object is removed for the normal map reconstruction in Figure 3.6(d), via polarizing

filters, to avoid introducing errors in the normal map generation. Thus when compared to Figure 3.6(c), where specular reflections are strong, fewer high-frequency fine details such as the lines on the close-up nose region may be observed in Figure 3.6(d). Likewise, the light dome capture in Figure3.6 (b) also lacks these high-frequency details but for the reason that the camera is placed much farther away from the surface than in the SfS setup.

If we reconsider the mathematical model for photometric stereo, most algorithms, including the one described in this chapter, assume that the object has a Lambertian surface. However, no object in the real world is pure 'Lambertian'. Therefore, our proposed method takes advantage of crossed-polarization to suppress specular reflections to better obey the cosine illumination conditions imposed by the photometric stereo model. Moreover, with the specular highlight suppression, it could contribute to better viewpoint registration in the first step. Using the SfS setup, we are thus achieving improved accuracy of the reconstructed normal map, as will be discussed below in more detail.

3.4.2. In-Situ Measurement of Architectural Elements: Bahá'í Temple

To evaluate the Mobile SfS on different materials and its capability for imaging 'in the wild', we performed experiments on the wall of Bahá'í Temple located in Wilmette, Illinois. Nine images were taken at night, placing the camera about 30 cm from the wall as had been done in the last example.

Compared to RGB photo (Figure 3.7(a)), the surface information of the wall can be generated with our proposed method as shown in Figure 3.7(b). The surface information matches with the RGB photo, and the proposed method provides fine surface information about the wall structure. Another important observation is that the generated surface shape reveals more detail than the RGB image, which lacks contrast and texture.

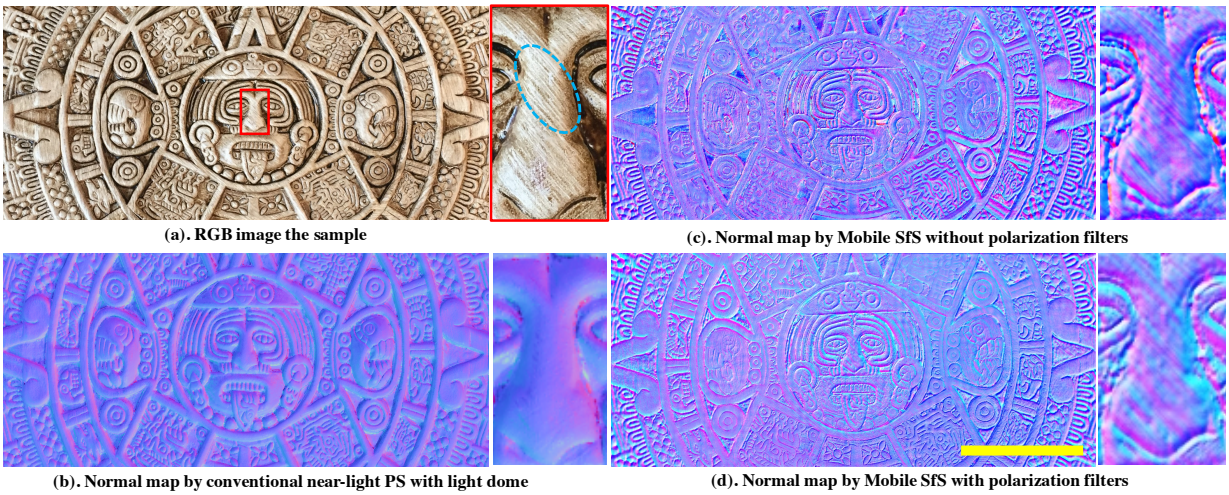


Figure 3.6. **Surface normal for the plaster replica of an Aztec calendar stone:** (a). the photo of the imaged sample; (b). the surface normal generated with near-light PS using light dome; (c). the surface normal generated using mobile SfS but without polarizers; (d). the surface normal reconstructed with the proposed mobile SfS. Close-up images show the "nose" on the plate. Scale bar: 5 *cm*.

Since the normal map of the sample is generated, the object can be rendered as shown in Figure 3.7(c). All structure and fine details can be seen from the 3D rendering. This helps digitally document the priceless artwork and protect the samples.

3.4.3. Surface Shape Measurement of An Illuminated Manuscript Page

To assess the capability of the Mobile SfS to discern features no larger than a few hundred micrometers across, we performed a further experiment on a mid-to-late 15th-century French illuminated manuscript page. Nine images were extracted from a much larger array of images acquired by a macro web-camera rastered across the surface. The overall RGB image is shown in Figure 3.8(a). These shots were acquired at about 2-centimeter away from a page of the manuscript. The technique makes it possible to observe individual brush strokes and obtain in-depth information on how the manuscript was technically constructed, as shown

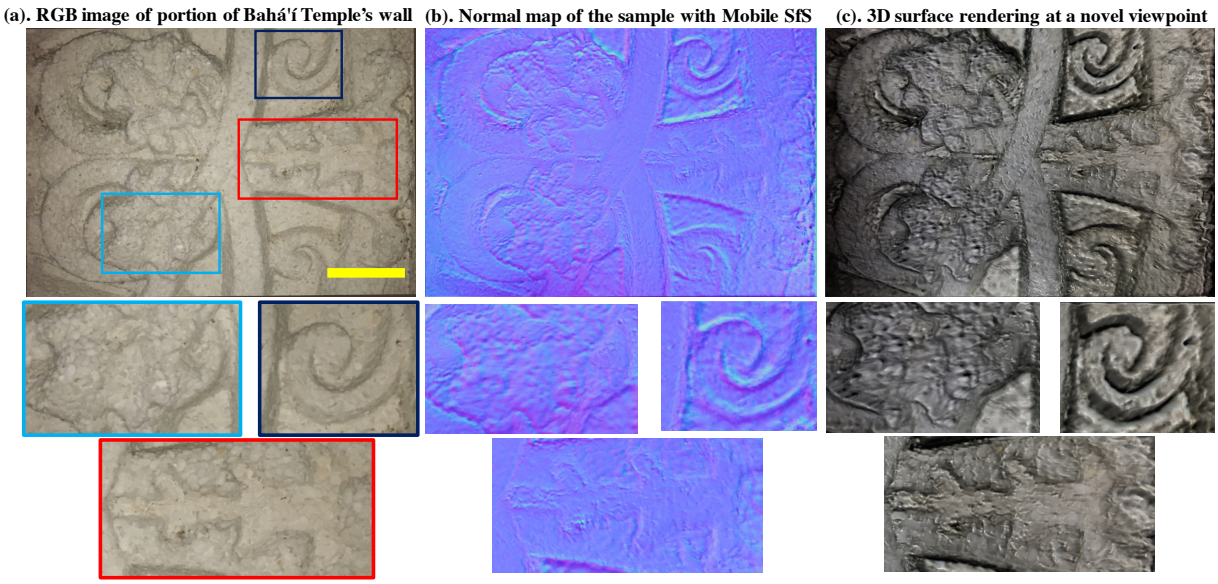
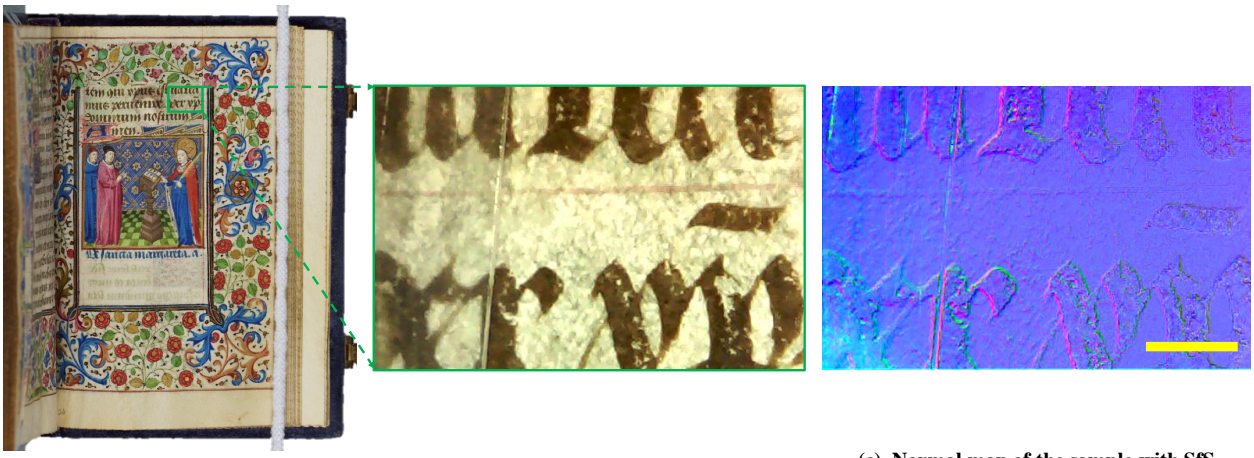


Figure 3.7. **Surface normal for the architectural elements:** (a). photo of the region of Bahá'í Temple which is been scanned; (b). the surface normal generated with mobile SfS; (c). the rendering with surface normal generated with mobile SfS. Close-up shows the detailed feature on the stone. Scale bar: 10 *cm*.

in Figure 3.8(b). We can also apply the 3D rendering to show the object's topography by separating color from the surface shape, which allows us to determine the presence and extent of 19th-century restorations. Hence, using these techniques, it may be possible to differentiate between the original illuminations and later restorations by correlating these topographic differences with the stylistic variations.

3.5. Discussion

Many historical artworks have specular surface such as those made of plaster, paper, and stone materials. The intensity of these specular reflections is often too strong to make accurate surface shape reconstructions of these materials. The reason for this is that in photometric stereo, the surface of the object is assumed to be characterized by a Lambertian



(a). RGB photo of the page (b). RGB photo of the interested region (c). Normal map of the sample with SfS

Figure 3.8. **Surface normal for a page of the manuscript page:** (a). the RGB photo of portion of the page of the manuscript; (b). the surface normal reconstruction with mobile SfS for that region. Scale bar: 5 mm.

model so that photons can be collected from different illumination angles and the intensity photons directed to the camera is only affected by the angle of the incoming illumination. It thus follows that under conditions where specular reflections dominate the photometric stereo will not produce an accurate assessment of the surface shape. In practice, reconstruction made using devices such as a light dome simply ignore specular reflections and assume that Lambertian reflections dominate the total light received by the camera. Clearly, this practice introduces error to the reconstruction. We show that it is possible to eliminate this source of error using the mobile SfS framework described here. In our set-up, the specular reflections are removed by the crossed polarizers covering the light source and detector. From the visualization point of view, specular information is desirable since it contains high high-frequency information on the surface shape of the object. Based on this observation, we are currently designing a method to separate the diffuse and specular reflection, but record both of them to be used for accurate rendering of object surfaces.

We have evaluated the prototype mobile SfS on different materials, as described above. Since artworks are made with different materials as well as stored and/or situated in different areas, it is very useful for the cultural heritage community to develop methods that can work under these varied conditions. We believe that our method affords conservators an accurate and precise surface shape evaluation method with simple and widely available tools– a smartphone device. We believe this method will improve the efficiency of art conservators to document the condition of many works of art rapidly.

We also note that taking photos with a flying drone has gained popularity and attention recently. It allows researchers to take photos of landscapes that were once difficult to reach, such as the top of the Bahá'í temple. However, these regions are critical for the conservation since conservators can not check frequently and ignore the problem due to the limited access. Since only a small piece of add-on component is needed in our proposed method, we can combine our small-size setup with the flying drone to provide a solution for the conservators to check the 3D surface of those regions. We believe this would benefit the community of architectural conservation who often need high-resolution data to assess a building's condition.

3.6. Conclusion

In summary, we have proposed a portable and cost-effective surface-shape imaging technique with an off-the-shelf mobile device. Our proposed method is demonstrated to scan historical artworks with different materials and scales. Moreover, our Mobile SfS can be operated in the wild, which greatly assists in the conservation of artworks in the wild. We believe our Mobile SfS can be a very useful tool for lots of different applications, such as the community of historical artworks.

CHAPTER 4

Highly Reflective Surface Shape Estimation with Mobile Device

4.1. Introduction

Three-dimensional (3D) imaging techniques are now omnipresent in a multitude of scientific and commercial disciplines. Industrial 3D inspection, medical 3D imaging as well as 3D documentation and analysis of art or cultural heritage are only a few examples of the broad range of applications. The work introduced in this chapter is motivated by a specific and challenging application of 3D imaging: The 3D measurement and analysis of highly reflective surfaces *in the wild*, *i.e.*, , for objects that cannot be transported to a laboratory for

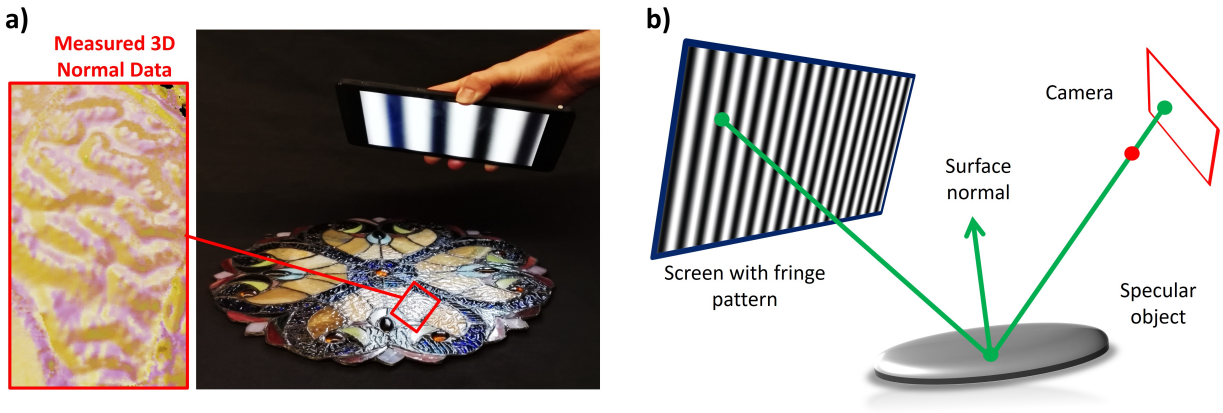


Figure 4.1. a) Handheld measurement of a stained glass painting with a mobile device. The reflections of the screen are visible on parts of the glass surface and reveal its three-dimensional structure. The measurement result (normal map) is displayed in the zoomed inset. b) Basic principle of 'Phase Measuring Deflectometry' (PMD): A screen with a fringe pattern is observed over the reflective surface of an object. The normal map of the object surface can be calculated from the deformation of the fringe pattern in the camera image.

measurement. As a concrete example, we study the *3D measurement and analysis of stained glass paintings*, such as larger glass artworks, church windows, or glass reliefs. The shape of the small glass pieces in a stained glass artwork is not necessarily flat! Over the centuries, several glass manufacturers developed unique techniques to imprint unique three-dimensional structures to the glass surface that reflects and refracts light in a very distinct way. These unique 3D structures in the glass piece are a powerful tool to match the small glass pieces in a stained glass painting to the individual manufacturers and to trace the circulation of stained glass around the globe. The latter is of significant interest for the cultural heritage community. We present a comprehensive 3D measurement tool that can perform this task in a hand-guided fashion with unprecedented ease of use, to be adopted by a broad audience of users with little to no technical expertise. In particular, we wish to provide 3D surface measurement capability to untrained personnel like museum conservators and tourists.

3D image acquisition techniques can be roughly divided into methods for two categories of surfaces: *(diffuse) scattering* and *specular*. Diffusely scattering surfaces are commonly measured by projecting a temporally or spatially structured light beam onto the object and evaluating the back-scattered signal. 'Time-of-Flight' [113] or Active Triangulation ('Structured Light') [135, 112] are prominent examples. Another well-known principle is 'Photometric Stereo' [139], where the object surface is sequentially flood illuminated with 'point' light sources from different angles.

Unfortunately, the application of these principles to *specular surfaces* yields only limited success. The reason for this is simple: specular reflections from a point light source scarcely find their way back into the camera objective. A straightforward solution to this problem is to extend the angular support of the illumination sources. This is the basic principle behind *'Deflectometry'* [49, 57, 66], where a patterned screen replaces the 'point-like' light source

(see Fig 4.1). This screen can be self-illuminated (TV Monitor) or printed. In Deflectometry systems, the screen and camera face the object, which means that the camera observes the *specular reflection* of the screen over the object surface. The observed pattern in the camera image is a deformed version of the image on the screen, where the deformation depends on the surface normal distribution of the object surface (Fig 4.1b). From this deformation, the normal vectors of the surface can be calculated. In order to calculate a normal vector for each camera pixel, correspondence between camera pixels and projector pixels must be determined. A common technique to achieve this is with the phase-shifting of sinusoidal fringes. The resulting *'Phase-Measuring Deflectometry'* (PMD) [49, 66] has established itself as a powerful technique that is used with great success in industrial applications, *e.g.* to test the quality of optical components or to detect defects on metallic parts like car bodies. Given a proper calibration, PMD reaches a precision close to interferometric methods [33, 94, 50, 13].

The task of digitizing specular 3D surfaces 'in the wild' leads to several fundamental and technical challenges of great scientific interest. Our goal is to develop a surface measurement method for objects that are large and, therefore, cannot be transported to a controlled lab environment. Besides a large FoV, the desired method should support a large variation in surface normals and also achieve high spatial resolution. In principle, this can be achieved using large-screen PMD systems, but these setups are bulky and cannot be applied 'in the wild'.

Similar to the method mention in last chapter [144], we tackle this problem with mobile devices (smartphone, tablet) to perform PMD measurements, *i.e.*, using the screen to display the patterns and the front-facing camera to image the object surface. Since the screen size of mobile devices is limited, only a small angular range of surface normals can be measured

in any single view [20, 129, 105, 46]. We overcome this limitation using an automated feature-based registration applied to PMD measurements acquired from different viewing angles. The multi-view measurements can be acquired in a hand-guided fashion. The features are extracted directly from captured images so that external markers or fiducials are not necessary.

For our mobile PMD system, we do not perform photometric and geometric calibration, necessary to recover quantitative surface shape information. This is because accurate calibration severely complicates the acquisition setup and makes it difficult to capture 3D shape for objects 'in the wild', which is the primary goal of this chapter. Without calibration, the accuracy of our method is compromised for low spatial frequencies of 3D surfaces that are reconstructed. This low-frequency bias produces limitations in the quantitative 3D surface information that can be extracted. We sidestep this problem by exploiting *a-priori* knowledge about our objects of interest (*e.g.* the stained glass examples in Fig. 4.2, or paintings in Fig. 4.5). Their overall shape is mostly flat but also contains high-frequency 3D surface shape information. This information is captured with high quality and can be used as features to help recognize an object's identity, *e.g.* by applying feature matching techniques (e.g., 'SIFT' [81]) to register normal maps captured from different viewpoints.

4.1.1. Contributions

- We demonstrate a hand-held Deflectometry system, able to measure specular 3D surfaces 'in the wild' over a large FoV. The system consists only of an off-the-shelf mobile device, like a tablet or a smartphone.
- We introduce the idea of exploiting a prior knowledge about surface shape to avoid the tedious calibration process necessary for multi-view registration and stitching

of arbitrary 3D surfaces. Our method works well for objects that contain a small amount of low-frequency 3D surface information but also posses interesting high-frequency 3D surface features.

- We apply automated feature-based registration to stitch together different 'normal maps' of an extended object surface into a panoramic, wide-FoV normal map. To our knowledge, our method is the first to enable hand-guided deflectometric measurements without the need for a prior 3D pose information, tracking, or external fiducials.
- We demonstrate the first registered and stitched normal map of an extended specular object with large angular normal variation that was captured with a hand-held system 'in the wild' - a stained glass artwork (see Fig. 4.4e). In addition, we show numerous examples of surface normal maps recovered from a variety of objects captured 'in the wild' from a single viewpoint.

4.2. Related work

'Phase Measuring Deflectometry' (PMD) is just one of many techniques that have been introduced to measure the 3D surface of specular objects. As discussed, deflectometric methods are widely used in the optical metrology community for the ultra-precise measurement of optical components, such as lenses, astronomical mirrors, or other kinds of free form surfaces. The power of the related approaches has been impressively demonstrated by many researchers over the last decades [57, 66, 50, 120, 129, 56, 49, 34, 13]. It has been shown that the principle is by far not limited to the procedure of phase-shifting sinusoidal fringes (PMD). Correspondence between the screen and camera can be established in many different ways [134], including the utilization of binary patterns [19], patterns multiplexed in

color space [129], or the application of the single-sideband demodulation trick, known from 'Fourier Transform Profilometry' (FTP) [121, 56, 79, 92].

Considering the vast potential of the deflectometric principle, it is not surprising that the computer vision community makes extensive use of it as well. However, the names of the proposed methods mostly lack the word 'Deflectometry,' and the related applications differ from high precision metrology tasks in many cases. Nevertheless, similar techniques using color fringes [124], lines [28], or even a light field created from two stacked LED screens [126] are known. Passive methods that do not require a self-illuminated screen at all are used as well: In [17], the reflection of color-coded circles observed by multiple cameras is exploited (which also resolves the bas-relief ambiguity). Completely 'screenless' methods, such as [44, 61] analyze the environment or track prominent features (*e.g.* straight lines in buildings) used to obtain information about the slope of specular surfaces. In general, the deflectometric principle allows for *any known pattern or structure* to be used as a reference.

Of course, each of the techniques mentioned above comes with benefits and drawbacks. For example, some of the techniques that use a static pattern instead of temporally phase-shifted sinusoids are capable of 'single-shot' acquisition [56, 79, 92, 129]. However, this comes not without a price: Many related methods deliver restricted lateral resolution or require the object surface to be sufficiently smooth. Shifting the correspondence problem to the color space (by applying a colored pattern) implies certain assumptions about the texture and reflectivity of the object surface. All this might not be a big problem for the measurement of lenses or mirrors, but it presents a significant challenge for cultural heritage applications like the measurement of stained glass surfaces.

It should be noted that even 'Photometric Stereo' techniques can perform the desired tasks under certain limitations. For example, [60] and [130] use known reflectance maps of

object surfaces to measure their 3D structure. Such approaches are especially beneficial for partially specular surfaces, but fail when the surface is too shiny. Other techniques exploit sparse specular reflections produced by photometric stereo measurements for 3D surface reconstruction or refinement [21, 111, 90].

It should be noted as well, that mobile versions of Deflectometry have also been demonstrated. The authors of [107], built a custom Deflectometry device compact enough to be used inside diamond turning machines to measure milled free form surfaces in-situ without rechucking. The authors of [105, 20, 129] even exploit the LCD screen and front camera of a smartphone or tablet to perform deflectometric measurements. However, these 'mobile device' systems only demonstrate results with limited FoV and coverage of surface normals. The 3D surface measurement of objects with high-frequency surface information is not addressed in these researches. The authors of the previously mentioned research [107] circumvent the problem of insufficient coverage of surface normals by rotating the object under the device and fusing normal maps taken at different rotation angles. The respective transformations are received from the rotation stages of the diamond turning machine. A similar approach is used in [46]. However, free-hand guidance over the object with subsequent pose calculation of the device is not possible. In comparison to previous work, we introduce a system capable of free-hand guided 3D surface measurement 'in the wild' for extended specular surfaces with large normal variations.

This section describes the image acquisition and processing steps that enable uncalibrated 3D Deflectometry measurements with mobile devices. We demonstrate a set of qualitative surface measurements that can be used to identify and compare characteristic surface structures for highly specular objects.

4.3. Setup and image acquisition process

Our hand-held PMD system implementation consists of a consumer tablet that serves as a measurement device (for the results shown in this work we used an NVIDIA Shield K1 or an Apple iPad Pro 10.5"). An application runs on the mobile device to perform the image acquisition process and transfer data to a host computer that performs the surface normal calculation and panoramic stitching.

During image acquisition, the tablet displays phase-shifted sinusoidal patterns and observes the object with its front camera (see Fig. 4.1a). The tablet is positioned approximately $200mm$ above the object surface. PMD is a *multi-shot principle*, meaning that *a sequence of temporally acquired images* has to be used to calculate *one 3D image*. During the measurement, the display projects four 90°-phase-shifted versions of a sinusoid in horizontal and vertical direction. Different frequencies of the sinusoid can optionally be used as well. The position of the tablet relative to the object has to remain fixed during the whole acquisition process. Depending on the speed of projection and image acquisition, this can be a hard task for the inexperienced user, if a handheld measurement is desired. For an optimal measurement result, the tablet can be fixed with a respective mount. We discuss possible extensions of our system towards a *single-shot principle* in section 4.6.

The front-facing camera objectives of mobile device cameras commonly have a short focal length, which results in a large FoV. Unfortunately, this large FoV cannot be exploited in its entirety by our system. This is because the device cannot be held closer to the object surface than the minimum possible focus distance, and the LCD screen has limited angular coverage. A valid PMD measurement can only be taken at image pixels that *observe a display pixel*

over the specular surface. As a result, the number of pixels that produce valid measurements can be as small as 25% of the imaging FoV.

4.4. Evaluation, results and discussion

In the following, we evaluate the surface normal map of several stained glass tiles [67], and a large, 300mm diameter stained glass artwork. A photo of the stained glass objects is shown in Fig. 4.2. The tiles have an approximately squared shape with an edge length of about 50 mm and demonstrate a significant variation in the distribution of surface normals. We first demonstrate the measurement and evaluation of the small stained glass tiles from a single viewpoint, then demonstrate a 'multi-view measurement' of the large stained glass painting.

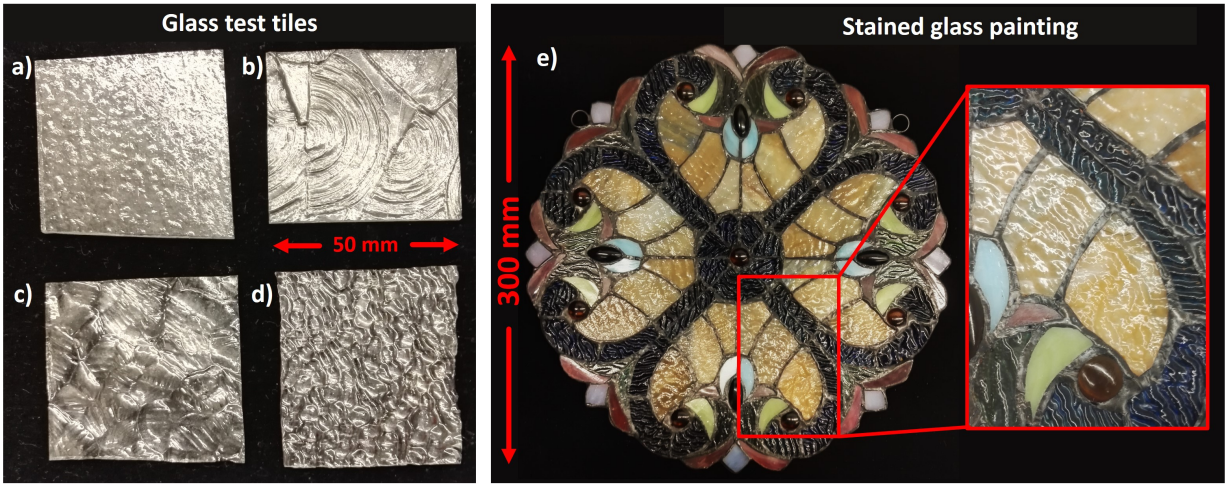


Figure 4.2. Photograph of objects to be measured with our system. a-d) Stained glass test tiles from the Kokomo glass factory [67], each with an edge length of $\sim 50mm$. Surface structure complexity and angular distribution of surface normals increase from a to d: '33KDR' (a), '33RON' (b), '33WAV' (c), '33TIP' (d). e) Large stained glass painting (diameter 300mm), scanned with our multi-view technique by 14 views from different angles and positions.

4.4.1. Single-view measurement

Most of the tiles in our test set display a size and surface normal distribution small enough to be evaluated from one single view. Each tile is placed at a position in the field of view, where the reflected display can be observed. The intensity $I(x', y')$ at each image pixel (x', y') can be expressed as

$$I(x', y') = A(x', y') + B(x', y') \cdot \cos(\phi(x', y')) \ . \tag{4.1}$$

Eq. (4.1) contains *three unknowns per pixel*: The (desired) phase $\phi(x', y')$ of the sinusoidal pattern, that correlates display pixels with image pixels, but also $A(x', y')$ and $B(x', y')$, which contain information about the unknown bias illumination and object reflectivity. This means that at least *three equations* are required per pixel to calculate $\phi(x', y')$. For each pattern direction, these equations are taken from the four acquired phase-shift images (the *four* phase-shift algorithm is very simple and, in addition, insensitive to second order nonlinearities), where the intensity in each image pixel for the m-th phaseshift is

$$I_m(x', y') = A(x', y') + B(x', y') \cdot \cos(\phi(x', y') - \phi_m) \ , \tag{4.2}$$

with

$$\phi_m = (m-1)\frac{\pi}{2} \ . \tag{4.3}$$

Finally $\phi(x', y')$ can be evaluated by

$$\phi(x', y') = \arctan \frac{I_2(x', y') - I_4(x', y')}{I_1(x', y') - I_3(x', y')} \tag{4.4}$$

This has to be done for each pattern direction, leading to phase maps $\phi_x(x', y')$ and $\phi_y(x', y')$ for the horizontal and vertical fringe direction respectively. The acquired phase maps are equivalent to the surface gradient in the horizontal and vertical direction *plus* a low-frequency phase offset that is dependent on the relative position between device and object, and any distortion present in the camera objective [66, 94]. In conventional PMD setups, this offset is removed by employing a calibration process whereby the phase map is first measured for a planar mirror, then subtracted from the measured phase. We avoid this step by exploiting a priori knowledge about our objects, namely that their overall shape is known to be mostly flat so that low spatial frequencies in the surface normal measurements can be ignored. In this case, the unknown phase offset can be removed by simply high pass

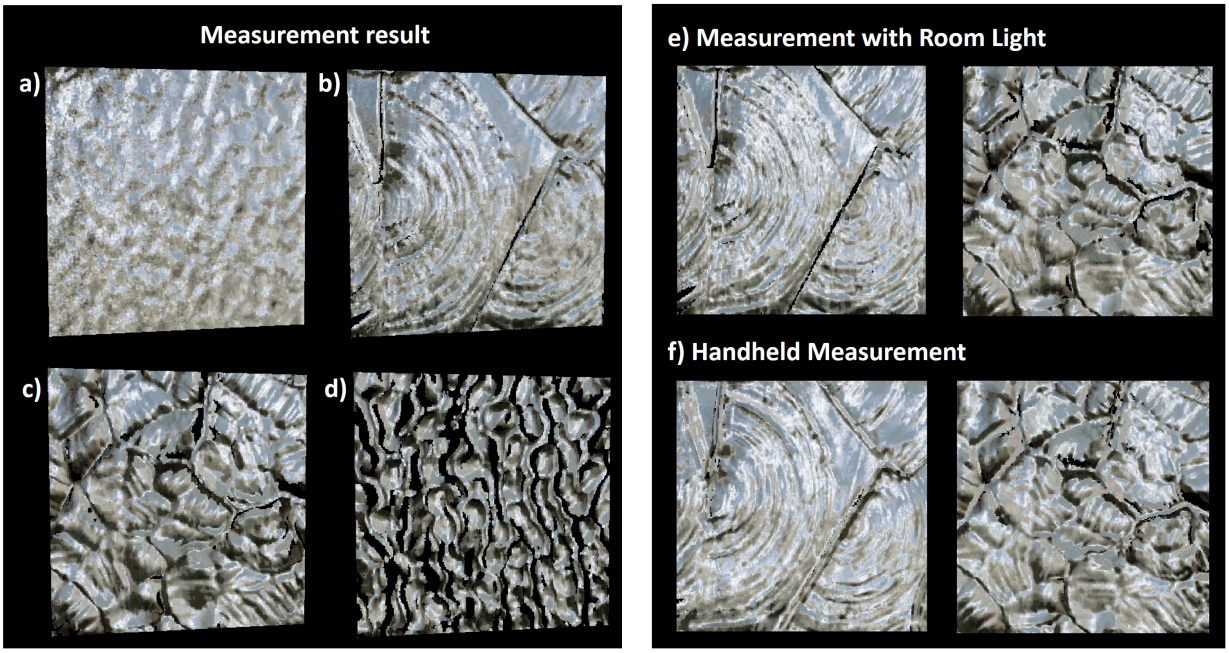


Figure 4.3. Single-view 3D reconstructions (surface normal maps) of Kokomo glass test tiles. '33KDR' (a), '33RON' (b), '33WAV' (c), '33TIP' (d). Measurements are performed with mounted tablet and no room lights. e) Reconstructions of '33RON' and '33WAV' measured under normal office light ($\sim 500lx$). e) Reconstructions for a handheld measurement of'33RON' and '33WAV'.

filtering the unwrapped phase map. The high pass filtered phase maps $\tilde{\phi}_x$ and $\tilde{\phi}_y$ are then equivalent to the surface gradient maps in the x- and y- directions. It should be noted that the filtering operation also compensates for the nonlinear photometric responses of the display and camera, avoiding an additional calibration procedure. Moreover, the assumption of a mostly flat object resolves the depth-normal ambiguity of deflectometry measurements, which typically requires two cameras to resolve [66].

The surface normal can be computed directly from the estimated phase maps via

$$\vec{n} = \frac{1}{\sqrt{\tilde{\phi}_x^{\,2} + \tilde{\phi}_y^{\,2} + 1}} \cdot \begin{pmatrix} \tilde{\phi}_x \\ \tilde{\phi}_y \\ -1 \end{pmatrix} , \tag{4.5}$$

where $\tilde{\phi}_x$ and $\tilde{\phi}_y$ denote the gradient for the horizontal and vertical direction, respectively. Figure 4.3 shows the calculated normal maps of all four tiles. The normal maps are shaded with a specular finish and are slightly tilted for visualization purposes. It can be seen that the characteristic surface structures important for the identification process are well resolved. The black spots in the normal maps are produced by surface points where the surface normal resulted in no measured signal, i.e., the camera was not able to see the display.

To test the robustness of our qualitative measurement results against different environmental conditions, we additionally acquired measurements for two of the four tiles with ambient room lighting and with performing a hand-held measurement without mounting the device. The results are shown in Fig. 4.3 e and f.

The measurement captured with ambient room lighting (Fig. 4.3e) shows no significant degradation in performance. This is understandable because the brightness of the room light

was moderate (regular office lights, illuminance $\sim 500lx$), and the SNR was not reduced significantly. Under these conditions, the four-phaseshift algorithm effectively compensates for bias illumination. For the free-hand guided measurement, motion artifacts in the evaluated phase map are expected. These artifacts can be seen at the slightly blurred edges in Fig.4.3f. The fact that the visible artifacts occur 'only' at edges is a consequence of the low frequency $\nu = 1$ (corresponding to one sinusoidal period displayed over the entire width of the screen) used to acquire these measurements. Higher frequencies would result in more prominent artifacts, for example, commonly observed in triangulation-based fringe projection.

4.4.2. Multi-view measurement

A single view measurement is not enough to capture a sizeable specular object with large normal variation in its entirety. This is not only because of the limited effective FoV of mobile devices but also because the large normal variation of some surfaces cannot be captured from a single viewing angle (see *e.g.* Fig. 4.3d). As discussed, our solution to this problem is to acquire and register multiple phase maps of the object surface, while our system is positioned by hand at different viewpoints. In this section, we show qualitative results that demonstrate our approach. We study a circular shaped glass painting with a diameter of 300 mm. From the magnification window in Fig. 4.2e, it can be seen that the glass pieces in this painting exhibit high-frequency surface features. Moreover, some glass pieces are milky. For the results shown below, we scanned one half of the glass painting by acquiring *14 single views* under different viewing angles and positions.

To assist in registration, we acquired an additional 'white image' (image of glass painting only illuminated by diffuse room light) at each viewing position. The registration transformation for the normal maps acquired at each single view is calculated from these 'white

images'. Performing registration with the 'white images' was found to be more robust than registration with calculated normal maps. For registration, we used the feature-based registration algorithms provided by the Matlab Computer Vision Toolbox. It should be noted that the usage of images which are captured under diffuse illumination is beneficial in this case since the diffuse illumination makes the object look similar from different viewing angles. No strong specular reflections (which look different from different viewing angles) disturb the feature extraction of the registration algorithm. With this trick, we are able to register subsequent views without applying markers or other fiducials onto the object surface, just by using the texture of the object itself. Figure 4.4 shows 'white images' of two subsequent views, their detected and mapped features, as well as the registration result.

It can be seen that the feature extraction and the subsequent registration transformation is applied on the *whole* FoV of the camera (not only on the limited effective FoV in the middle) in order to detect a large number of features with high quality. In this case, it can be beneficial to perform a simple internal calibration of the front camera (*e.g.* with a checkerboard) to compensate for distortion. This can reduce the registration error significantly. It should also be noted that such a distortion correction was avoided for the previous single-view measurements since most of the signal was measured in the middle of the FoV, where the distortions are small. In the future, we plan to develop methods that estimate the distortion parameters of the camera *during registration* without the need for an explicit calibration procedure. Figure 4.4e shows all 14 views after registration and stitching. Most parts of the object's surface are densely reconstructed, and the high-frequency structures of the individual glass pieces are visible. However, some normals are still missing, mostly from the blue glass pieces in the painting. The structure of these pieces displays extraordinary

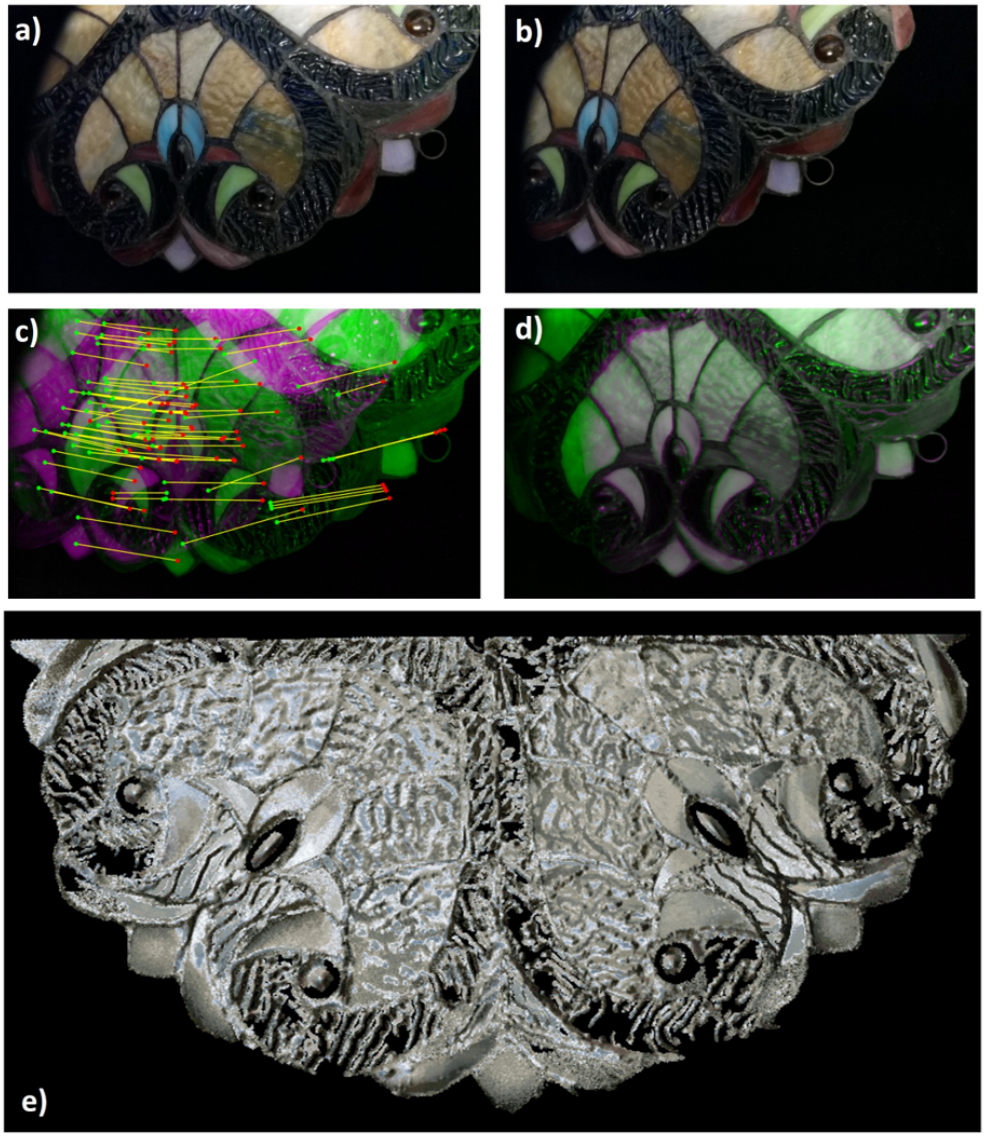


Figure 4.4. Multi-view normal map 3D reconstruction of large stained glass painting using image-based registration. a) and b) 'White images' (images captured with black screen and diffuse room light illumination) before distortion correction. c) Detected and mapped features in the two subsequent 'white images' (color-coded by green and magenta). d) Registered 'white images'. e) Visualization of stitched multi-view normal map result, consisting of 14 registered single-views.

high hills and deep craters, producing a wide distribution of normals that would require more than 14 views to be measured effectively.

4.5. Additional experimental results

Although the presented method was motivated by the 3D measurement of stained glass artworks, the system is in no case restricted to this specific object type. A 3D surface acquisition with our uncalibrated method is possible as long as the overall shape of the object is flat, and the surface under test is relatively shiny.

Figure 4.5a-c displays the surface measurement of an oil painting. The three-dimensional analysis of painting surfaces is also of great interest to the cultural heritage community. The ability to separate surface texture from its shape or slope data is a valuable tool to understand different painting techniques (*e.g.* by looking at the directions of brush strokes). Monitoring of pigment degradation in paintings [100, 119, 109] is another application that does not work reliably by only looking at captured 2D images. Our mobile 3D imaging method is well suited for the analysis of paintings 'in the wild', *i.e.*, directly on the museum wall. Figure 4.5a shows an image of a measured oil painting. The surface normals of the black region in the red box (approximately $70mm \times 80mm$) are acquired with our method. For better visualization of the hills and valleys of the brushstrokes, the acquired normal map is integrated into a depth map, using the Frankot-Chellappa surface integration algorithm [37]. Figures 4.5b and c show the calculated depth map from two different perspectives (z-component exaggerated for display purposes). The brush strokes, and even the underlying canvas can nicely be resolved.

Another potential field of application is the 3D acquisition of technical metallic surfaces. Figure 4.5e displays the acquired normal map of a metallic key ($70mm$ height, Fig. 4.5d),

shaded with a specular finish. The normal maps of a 5 cent and a 10 cent US coin (21mm / 18mm diameter) are shown in Fig. 4.5h. Imprinted letters or symbols can be resolved,

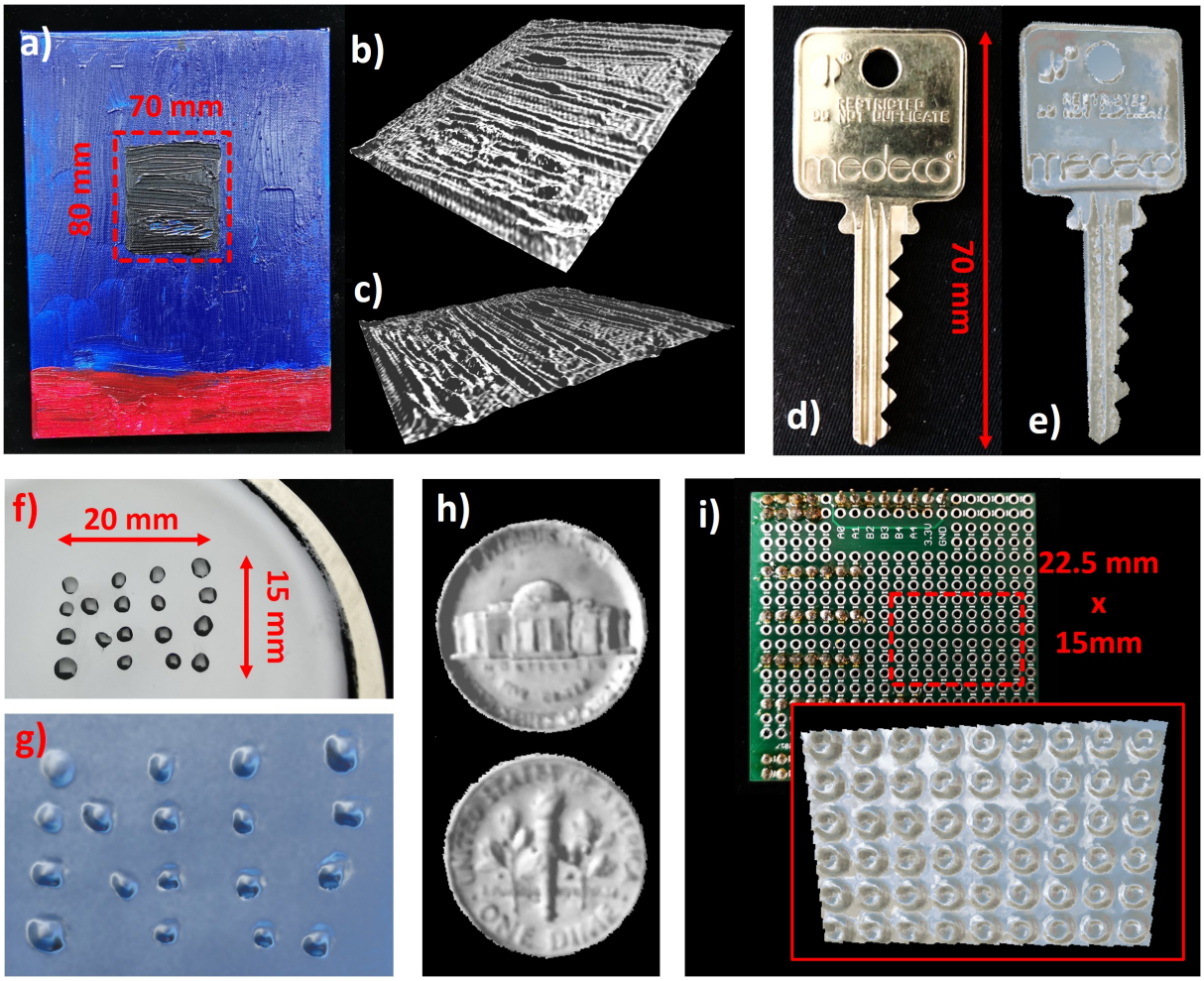


Figure 4.5. Deflectometric measurements of different surfaces: Paintings, technical, metallic, enameled ceramic, and fluid surfaces. a) Image of measured painting with marked 70mm × 80mm measurement region. b) and c) Surface shape of the marked region, calculated by integration of the acquired normal map. Brushstrokes and canvas can nicely be resolved. d) Image of measured key (70mm length). e) Measured normal map of the key. f) Water drops (20mm × 15mm) on an enameled ceramic surface (coffee mug). g) Evaluated normal map. h) Normal maps of a 5 cent and a 10 cent coin. i) Circuit board with marked 22.5mm × 15mm measurement region and measured normal map. Each metallic circle has a diameter of ∼ 2mm

both for the key as well as for the coins. Figure 4.5i displays the normal map of a circuit board. The diameter of one single metallic ring is only about $2mm$. In the last example, we demonstrate the capability of our method to measure fluid surfaces, *e.g.* for the analysis of surface tension. Figure 4.5g shows the normal map that was acquired from water drops on an enameled ceramic surface (coffee mug). The water drops are arranged to form the letters 'N U' (Fig. 4.5f). The shape of each drop is clearly visible from the normal map. In the future, we plan to use our system to measure *dynamic fluid surfaces* with a single-shot PMD technique, such as [56, 79, 92, 129]. In addition, we are developing algorithms capable of recovering surface normals from objects with much more complicated reflectivity.

4.6. Conclusion and outlook

In this chapter, we presented a mobile Deflectometry system that is able to measure specular surfaces with a high normal variation and much larger than the system's initial FoV in a hand-guided fashion. In order to sample the entire object surface densely with high resolution over a large FoV, we applied a feature-based registration to stitch normal maps from different viewing angles and positions. The system can be moved by freehand from one viewpoint to the other. No external guidance or fiducials affixed to the object are necessary.

We demonstrated the 3D surface measurement of stained glass surfaces using both single view and registered multi-view measurements. As a proof of principle, we scanned one half of circular stained glass artwork with 300 mm diameter by stitching together *14 single views*. In a second experiment not shown in the work, we tried registration with 28 views. However, global registration errors were significant so that the first and last views did not fit together after one pass. This is a well-known problem for surface measurements with registration [9]. Reducing the global registration error is one of our main goals for future work.

Our evaluation process exploits a prior knowledge about the object to avoid extensive fringe and display calibration, which also solves the depth-normal ambiguity problem without the use of a second camera [66]. In the future, we seek to develop self-calibrating algorithms for multi-view measurements. Our plan is to apply a non-rigid registration on our data and obtain the information about the distortion from the calculated deformation fields. Moreover, we will work towards obtaining *quantitative measurements without calibration*. This work will build upon previously demonstrated self-calibrating PMD setups, *e.g.* shown in [94].

Although we have shown that hand-held measurements are possible with our system, PMD is commonly a multi-shot principle, and can, therefore, introduce motion blur. Single-shot PMD techniques that rely on single-sideband demodulation, *e.g.* like introduced in [56, 79, 92] will not work on objects like stained glass paintings because of the severe bandwidth restrictions. In the future, we want to explore other single-shot and/or motion-robust Deflectometry techniques that exploit additional modalities to solve the ambiguity problem. Examples of how such problems are solved in the field of line triangulation can be found in [9, 136, 135, 134, 101]. Our future goal is to develop similar methods for Deflectometry. Ideally, the user only needs to continuously wave around his device in front of the object to obtain a dense 3D reconstruction after a few seconds.

Lastly, to foster the adoption of our technique by a broad audience, we plan to make our measurement App publicly available so that anyone with a mobile device can make 3D surface measurements of specular objects. Each user will be able to transform his phone or tablet into a 3D measurement instrument. We envision this framework will serve as a platform for crowd-sourced aggregation of surface shape acquisition/fingerprinting of unattributed artworks around the globe.

CHAPTER 5

Inverse Rendering Based Specular Surface Shape Estimation

5.1. Introduction

In the last decade, there are increasing interests utilizing surface shape measurement for many applications in science and industry, such as biometrics, industrial inspection, and cultural heritage. Common $3D$ scanning techniques such as laser scanning, structured light, and photogrammetry has provided robust and reliable shape recovery for different real-world applications. These commercialize $3D$ scanning methods could handle most of the surface except for complex material such as a highly reflective surface, which is still an open challenge in computer vision and computer graphics due to unique material reflectance properties. Appearance modeling based shape recovery techniques such as Deflectometry adapt specular reflectance properties in the image formation model, which use an extended light source to recover the geometry of the highly reflective surface densely. Due to the law of reflection, to densely recover different surface angle, a bulky and expensive hardware solution is usually required for the deflectometry measurement and which constrain the development of this powerful technique to utilize in the real-world tasks. Recent computational imaging works **??** address the hardware-intensive problem for deflectometry using a compact hardware system or even a mobile system. Nevertheless, these approaches are assumed relatively simple surface geometry or leverage the other reflectance component for registration.

Another open challenge for deflectometry is the surface gradient field integration. In the previous chapters, we propose several novel portable appearance-based shape recovery

systems that cover both Lambertian and highly reflective surfaces. We demonstrate the robustness and portability of the systems using only off-the-shelf community hardware components such as DSLR cameras, flashlights, mobile phones, and tablets. Similar to conventional appearance-based shape recovery techniques, the image formation model of deflectometry use surface orientation to represent the target surface. The surface gradients are normalized directional information in which there are multiple possible solutions of the surface along the same normal direction. It is possible that by instead modeling the surface normal, directly obtain the surface depth information that improves the surface reconstruction. If an accurate and fully differentiable image forward model is available to provide an unbiased simulation, this inverse problem could be solved by optimizing the differences between the simulation and the experimental observation.

This chapter introduces a novel appearance-based shape recovery framework that leverages the differentiable rendering as the forward imaging model and a simple direct depth optimization of the surface instead of integrating the surface from surface gradient field. The direct optimization can correct the biased measurements and avoid the bias introduced by the normal integration to produce a more accurate surface geometry estimation. We also incorporate the multiview observation into the joint optimization that could truly apply onto a highly reflective surface that would fail with image feature-based registration. We compare this approach to ground truth in simulation and show experimental results using a simple portable system which is possible to apply on to mobile device or conventional computer system.

5.1.1. Contributions

- **Multiview portable Deflectometry system:** We present a simple system for the free-form deflectometry capture system. Our system architecture is possible to further extend to only using an off-the-shelf mobile device such as mobile phone and tablet.
- **Joint Multi-view optimization:** Previous techniques estimate per-view surface normal and then manually blend in multiple normal maps, which could potentially increase the error in geometry estimation due to either bad viewpoint estimation or noise per-view surface normal estimation. By contrast, we frame the normal blending as a joint optimization problem that could jointly constrain the noisy measurement.
- **Direct depth optimization:** Conventional appearance based shape recovery methods model the surface geometry as surface gradient. In order to obtain the $3D$ information, it typically requires a normal integration, which could lead to the low-frequency geometry bias due to the depth/scale ambiguities. Our methods leverage rendering pipeline directly, optimization the depth information of the surface.

5.2. Previous Work

5.2.1. Multiview Deflectometry

'Phase Measuring Deflectometry' is one of the specular surface metrology methods which have been widely adopted into industrial inspection applications. By projecting phase-shifting sinusoidal fringes pattern with an extending light source (e.g., a display screen), it

could model the specular reflection to recover the surface shape information. Deflectometry methods have been widely used in the optical metrology community for high precision measurements of highly reflective surfaces such as lenses, automobile bodies, and integrated circuit boards. In principle, given the narrow specular reflection lobe, in order to cover the wide range of surface geometry, either the extending light source or the camera should densely sample across the whole positive hemisphere of the surface. This constraint limits the portability and the robustness of the deflectometry system. Commercial PMD system [131] requires a large LED screen and multiple cameras to cover different angles. In recent years, researchers start to tackle this problem and come up with mobile versions of Delfectometry systems with the minimal system. Röttinger *et al.* [107] customize a compact system for machining inspection. The others [105, 20, 129, 137] adopt the LCD screen and camera on the mobile system such as smartphone and tablet for robust and user-friendly deflectometry capturing. However, with the limited size of the LCD and camera Field-of-View (FoV), these systems could only demonstrate results with related simple surface geometry that the surface normal could be covered by the miniature systems. Researchers start to leverage multiview methods to cover the surface geometry with multiple measurements. Balzer *et al.* [11] introduce a compact system bundled on an industrial robotic arm. They track the transformation between different measurement positions and register the measurement with the transformation. In the last chapter, we introduce the image-based registration pipeline that exploits the diffuse component in the stained glass painting for feature extraction and matching. Nevertheless, for highly reflective surface, it is still an open challenge to adapt multiview measurement for the shape reconstruction.

5.2.2. Normal Field Integration

For shape reconstructions techniques such as photometric stereo (PS) and deflectometry, they model the shape information as the surface gradient field. It is necessary to perform gradient integration process to get $3D$ information of the surface. In theory, the integral of a surface point gradient/normal along any closed-loop (path) should be zero as the inverse problem of the definition of the surface normal. In practice, the estimated surface normal rarely fulfills this condition due to the noise aggregated in the estimation process. Moreover, since surface normal is $2.5D$ information, the depth/scale ambiguities make this problem trickier than in theory. By using the Poisson equation, Horn and Brooks [54] model the surface with second-order finite differences approximations of the Laplacian and the divergence to integrate the surface from gradient. It is robust to the noise input data and performs well without any prior or parameters. However, it is really slow since it uses a Jacobi iteration to solve the larger linear system, especially if the initialization is far from the solution. On the other hand, Frankot and Chellappa [37] try to solve the problem with Fourier analysis, which boosts the processing time and maintains robustness. However, their method still suffers from the depth discontinuities and ambiguities problem. Queau *et al.* [103] came up with a comprehensive review on many popular normal integration techniques. Nevertheless, a robust and reliable normal integration method could be highly beneficial for the application of PS and deflectometry.

5.2.3. Differentiable Rendering

Creating photorealistic images has been a major focus in the computer graphics community. This effort has led to mathematical models and algorithms that can generate predictive and

physically realistic images from known scene appearance factors such as camera position, surface shape, surface reflectance, and the illumination of the scene. In order to physically obtain realistic appearance factors from the real-world scene, inverse rendering problem has also been investigated since the very beginning of the field [14, 99, 147, 87]. By comparing the change between observed images and rendered images, inverse rendering aims to infer the surface shape, reflectance, and scene illumination of real-world photography. In order to seamless optimize these parameters along with the rendering process, *Differential Renderer* (*i.e. Neural Rendering*) has been introduced to solve the problem.

Gkioulekas *et al.* [43] propose an optimization framework with material dictionaries to model translucent materials with stochastic gradient descent. To deal with the more general scene, Mansinghka *et al.* [86] use Bayesian probabilistic graphics model to infer the appearance parameters. With the rising development of the convolutional neural network in vision and graphics community, many learning-based 3D reconstruction and material inference techniques [74, 77, 151, 125, 73] have incorporated a *differrentiable rendering layer* to adapt inverse rendering into the end-to-end network architecture. A general, flexible, and easily integrated with deep learning differentiable renderer would come in handy. Similar to Mansinghka *et al.*, Loper and Black [80] introduce *OpenDR*, an approximate differentiable renderer which could obtain derivatives of the model parameters by incorporating into probabilistic programming framework. More recently, Li *et al.* [72] release *redner* with a novel edge sampling algorithm for Monte Carlo Ray Tracing without approximation. Their method could efficiently handle secondary effects such as shadows or global illumination with gradient-based optimization, which has lots of potential for different applications in inverse rendering and the generation of adversarial for neural networks. Liu *et al.* [78] propose Soft Rasterizer, which is a truly differentiable renderer framework by modeling the rasterization

process with probabilistic model blending, which "soften" the traditional discrete rasterization and the nondifferentiabilities. In 2020, Ravi *et al.* [104] introduce *PyTorch3D*, a open-source $3D$ library with optimized operators, heterogeneous batching capabilities, and a modular truly differentiable renderer that could smoothly integrate with machine learning framework *PyTorch* [98].

Our method exploits the novel differentiable rendering framework as the image forward model. By incorporating the differentiable rendering and simple direct depth optimization, our framework improves the multiview blending and normal integration challenges commonly encountered in deflectometry techniques with joint multiview depth optimization. Also, our hardware setup constraints relative to prior deflectometry methods could potentially extend onto mobile device for a more portable and better user experience.

5.3. Direct Depth Estimation with Inverse Rendering Based Optimization

Appearance-based shape recovery methods have extensive history solving the inverse problem with appearance factors. From photometric stereo [140] that model Lambertian reflectance to deflectometry which specific for specular reflectance, there are intensive of prior researches that work on high complexity inverse problem to isolate the surface shape from appearance. In order to simplify the complexity of this problem, conventional approaches come up with a particular image formation model that model surface geometry as surface gradient and solve it with different mathematical approaches such as matrix inversion or numerical optimization. To acquire $3D$ information of the surface, a normal integration step is required. Moreover, these image formation models only model the appearance from a single viewpoint for idealization.

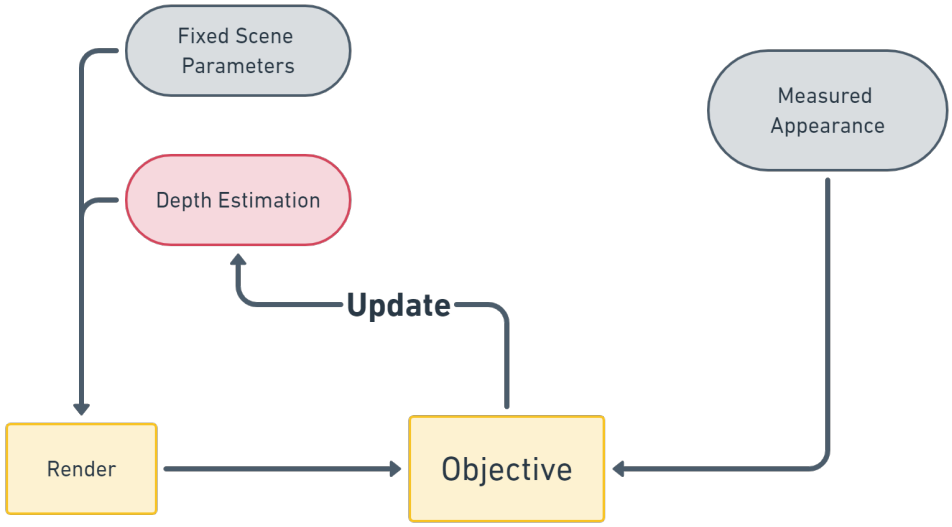


Figure 5.1. **Inverse Rendering Based Shape Optimization** Our framework directly optimizes the surface depth information by exploiting the differential renderer, which could compute the derivative of the target parameter (*i.e.*, depth) during the forward rendering process. The depth estimation can then be iteratively updated to incrementally minimize the objective function between the measured appearance with the back-propagation process.

On the other hand, the conventional rendering systems use either rasterization or ray-tracing to simulate the appearance of the $3D$ surface. If we instead have a render system which could take surface depth and fixed scene parameters such as camera, light source and reflectance information as input to render the appearance of the surface, we could directly model the depth information and reference different viewpoints to form an optimization problem as shown in Fig 5.1 with real measured appearance to reconstruct the $3D$ surface information. However, without any knowledge of the render forward model, the optimization would be exhaustive and time-consuming. Instead, we utilize a differentiable renderer that could trace over the parameter gradient so that the optimization would be first-order instead of zero-order, which could be more efficient and provide a better result.

In our objective function, we try to minimize the error E, between the measured surface image $\hat{I}$ and the output of the forward operator R given the depth estimate D. All other scene parameters λ are assumed to be fixed upon initializing the renderer.

$$E = \underset{x \in D}{\text{minimize}} \|\hat{I} - R_\lambda(G(D))\|_2 \tag{5.1}$$

where G is the smoothing operator that enforcing the coarse-to-fine search with referencing neighborhood surface point and constrain the noisy data input.

We choose the PyTorch3D [104] differentiable renderer, which is a free, open-source software library written in Python as our forward renderer. It provides efficient operations on triangle meshes and smooth integration with PyTorch. Both PyTorch and PyTorch3D could utilize GPUs for acceleration, which could significantly improve the reconstruction runtime performance. We design a custom shader in PyTorch3D, where it models the specular reflectance for our rendering process. Also, in order to efficiently simulate the extended light source (*i.e.*, LCD display) as in our system, we customize the environment mapping process to project the illumination patterns on to the surface instead of integrating the array of the point light which would be very time-consuming and memory intensive.

Data preparation, scene setup, optimization, and profiling are implemented in Python using PyTorch and PyTorch3D with GPU acceleration. We project the target depth map to the triangle mesh and construct the rendered scene with the other fixed scene parameter such as camera locations, display locations, and fixed shader maps. The renderer would storage the derivative of the depth parameter in PyTorch tensors within the rendering. After each rendering, we setup the back-propagation to optimize the depth map with Adam solver and

the loss function in Eq 5.1. The optimization process will iteratively update the surface depth information until the current surface appearance matches the real measured appearance.

5.4. Simulated Results

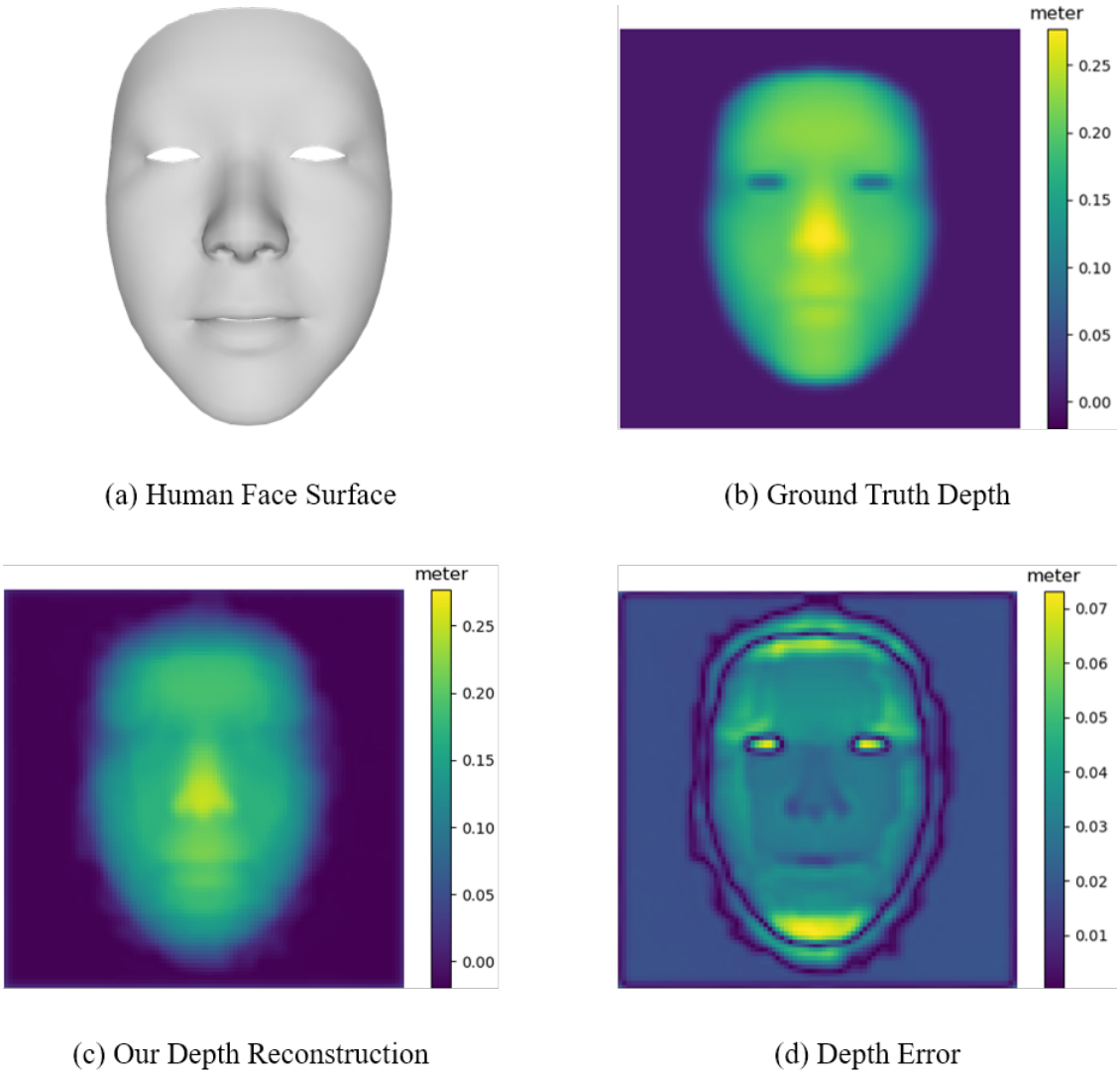
We conducted performance evaluations on simulated data using the rendering system due to the difficulty in acquiring precise ground truth measurements over moderate-sized physical scenes. We chose to test the quantitative fidelity of our framework using a human face geometry, a convex surface, including a wide range of surface orientations.

We construct the rendered scene with pure specular reflectance and a deflectometry system similar to our portable system. For each viewing position, we project 8 phase-shifted sinusoid patterns similar to conventional PMD methods (4 vertical and 4 horizontal). To simulate the captured appearance of the surface, we render a 4 by 4 grid capturing sequences with 16 different viewing position, which we found it is efficient to cover various surface orientations of the face. We start our optimization from a planar surface and then run the optimization with Adam solver and learning rate 0.0003. We reduce our learning by 0.0001 for every 800 iterations, requiring about 2.4k iterations in total for convergence.

Figure shows the simulated result with the human face. Our method's root-mean-square error (RMSE) is 1.45 cm and peak signal-to-noise ratio (PSNR) is 35.75 dB. This shows that our optimization framework achieve good depth reconstruction considering the overall scale and the diversity of surface angle.

5.5. Acquisition System Architecture

To capture experimental data and demonstrate our framework's robustness, we set up this portable system consisting of a FLIR Blackfly S machine vision 5.0MP polarization cameras



(a) Human Face Surface

(b) Ground Truth Depth

(c) Our Depth Reconstruction

(d) Depth Error

Figure 5.2. **Simulated Human Face Depth Reconstruction Results** Comparison of our depth optimization reconstruction framework with human face surface shown in (a), and ground truth depth map (b). (c) Our inverse rendering base depth optimization result with simulated data produce high-fidelity depth estimation. Depth error map compute the difference between optimized result and ground truth which achieved good overall reconstruction quality with $< 1.5cm$ RMSE and $> 30dB$ PSNR for 3D surface reconstruction.

with (BFS-U3-51S5PC-C) with Navitar 16mm C-Mount lens and a 15.6 inch Portable LCD display. Similar to the concept and approaches in Chapter chapter 3, We utilize the linear

polarization filters on both camera and the LCD display to perform cross-polarization imaging to extract specular reflectance component. We leverage this novel sensor with on-sensor polarizing filters. Each sensor pixel has its own linear polarizing filter, and the neighborhood 2 by 2 pixels grid contains 4 ($0\,\text{deg}, 45\,\text{deg}, 90\,\text{deg}, and 135\,\text{deg}$) equal distributed filter orientation for polarimetry analysis. Although it only supports with industrial and scientific-grade machine vision cameras, with its powerful functionalities, we foresee in the future that it would be widely adapted on to consumer devices for different applications. The camera and the display are rigid bundle together with a filming stand and which the system could be either handheld by the user or mount on a tripod for optimal capture results. During the acquisition, the LCD display project user-defined patterns and observes the surface reflection with the camera. Lastly, a printed set of corner fiducial makers (*i.e.* ArUco Markers [40]) were affixed to the imaging area to provide tracking points for the pose estimation. Our system architecture is relatively simple compared to the conventional deflectometry system. The framework is highly possible to extend the capturing on to mobile device similar in Chapter 4.

For inverse rendering measurement, it is critical to have the whole system to be carefully calibrated, including the camera and light source. However, calibrating the geometric and photometric characteristics for both camera and light source usually requires an extensive calibration procedure. In the spirit of using a robust and user-friendly system for 3D reconstruction, our calibration procedure requires minimal equipment, and it is robust enough for users without previous experiences and skillsets. Our calibration consists of two parts, Geometry Calibration and Photometric Calibration for both camera and display.

5.5.1. Geometric Calibration

In order to accurately reconstruct the shape information, first, we need to model the geometric relationship for both camera and our light source.

For the camera, geometric calibration models each camera's imaging transformation from world to image coordination at each viewing locations. For a pinhole camera model, this could be expressed in homogeneous coordinates as:

$$p_i = AMp_w, \quad p_w = \begin{pmatrix} X_w & Y_w & Z_w & 1 \end{pmatrix}^T \in P^3,$$
$$A = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \quad M = \begin{bmatrix} r_{00} & r_{01} & r_{02} & t_1 \\ r_{10} & r_{11} & r_{12} & t_2 \\ r_{20} & r_{21} & r_{22} & t_2 \end{bmatrix} \tag{5.2}$$

where p_w denotes the point in world coordinates and p_i as in image coordinates, A is the intrinsic matrix that project the points from camera coordinate to image coordinate and M is the extrinsic matrix that describe the camera pose.

The camera calibration is usually separated into two parts: intrinsic and extrinsic calibration. In our system, we are using standard implementation of the OpenCV library [18] to determine intrinsic parameter and camera distortion parameters in a one-time off-line procedure. As for the extrinsic parameters, we utilize the ArUco Markers that were attached to the corner of the imaging area and the intrinsic parameters to solve Perspective-n-Point(PnP) problem for camera poses for each viewpoint.

On the other hand, for the light source, we also require knowledge of the posture and position of the display with respect to the camera. If the camera faces the display directly, the calibration is simple which project a patter on the display and solve the PnP problem

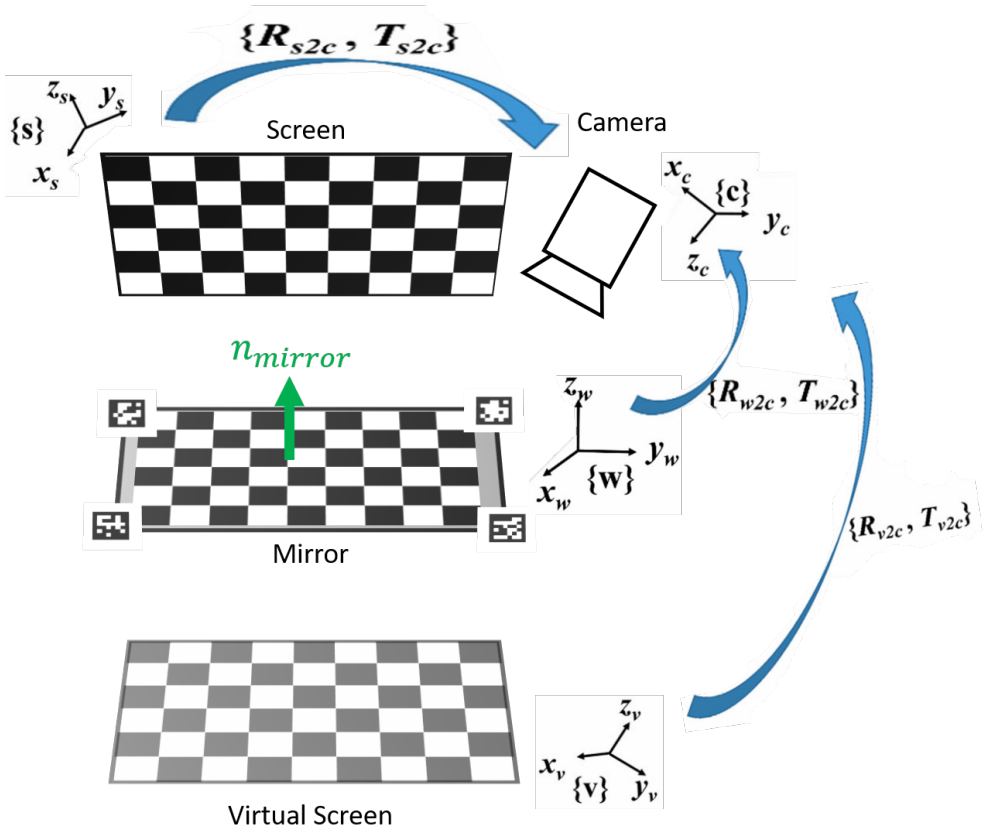


Figure 5.3. **Display geometric calibration** We leverage a planar mirror to calibration the display geometric relationship with respect to the camera.

as the camera intrinsic. Unfortunately, the display is not in the view of the camera in our setup. This camera-display calibration problem [47, 38, 36] has been addressed for different applications such as structure light system and conventional phase measuring deflectometry systems. In order to solve this problem with minimal user effort, we adapt an planar mirror to accomplish the calibration. The planar mirror is placed on the target plane which is in front of the camera and the display, the geometric relationship of the components is shown in Fig 5.3. The original point of the camera coordinate system $\{c\}$ is set at optical center of the camera and the original point of the display coordinate system $\{s\}$ is set at the center of the display. As shown in Fig 5.3, four ArUco markers were placed on the corner of the mirror which would

be used to determine the transformation between the mirror plane (world coordination $\{w\}$) and the camera coordination. The camera captures the virtual image of the display $\{c\}$ which will be mirror-symmetrical to the original display. The transformation relationship between the real display and the virtual display could be model by the n_mirror which is the normal vector of the mirror respect to the camera coordination and d_{w2c} refers to the distance between the mirror and the camera. By modeling the Householder transformation [30], the relation between $\{s\}$ and $\{c\}$ can be calculated as:

$$(I_3 - 2n_{mirror}n_{mirror}^T)R_{s2c} = R_{v2c}(I_3 - 2e_3e_3^T) \tag{5.3}$$

$$(I_3 - 2n_{mirror}n_{mirror}^T)T_{s2c} = T_{v2c} + 2d_{w2c}n_{mirror} \tag{5.4}$$

where I_3 is the 3 by 3 identity matrix, $e_3 = [0, 0, 1]^T$, $n_{mirror} = R_{w2c}[0, 0, 1]^T$ and $d_{w2c} = |n_{mirror}^T \cdot T_{w2c}|$.

5.5.2. Photometric Calibration

In order to accurately model how light propagates from the display to the surface and reflect back to the camera, we need to know how the relationship between the radiance that the display projected out and also the radiance arrived at the camera, which been interpret as the pixel values in the images. First, we perform a photometric calibration with the camera similar to [26], which captures multiple exposure values of the camera. By recovering the camera response function, we could figure out the mapping between the reflected radiance from the surface and the pixel values in the images.

Typically, to achieve optimal visualization quality, the display would apply gamma correction to compensate for the nonlinearity of the human perception response. Moreover, the

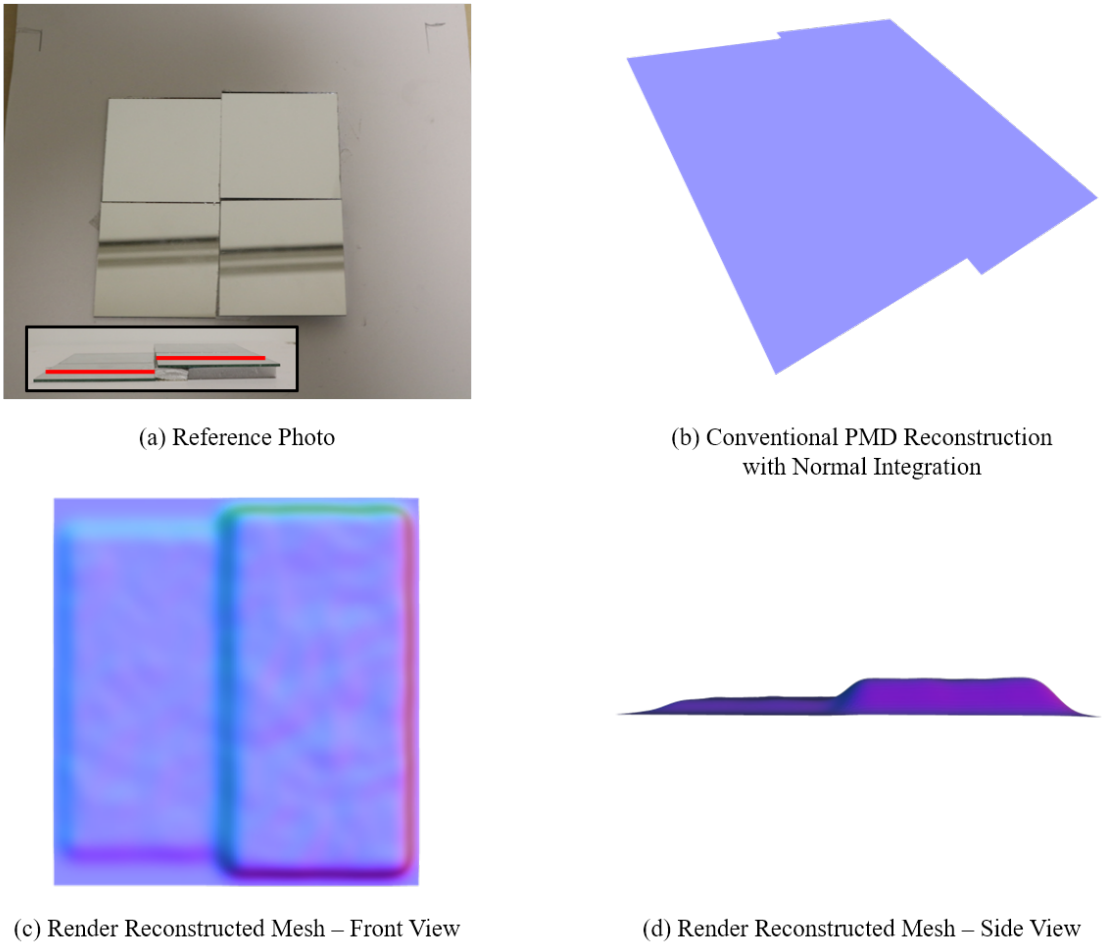
LEDs of the LCD panel would also have nonlinearity response to the intensity. Since we are using the LCD display as the extended light source to project grayscale ramp pattern, it is crucial to model the mapping between the assigned pixel values and the projected intensity. For the display radiometric calibration, there are plenty of off-the-shelf monitor calibrators that could provide reliable photometric calibration for the display. Instead of using the expensive off-the-shelf solution, we solve this problem with the existing component in the system. We use the display to project series of pixel intensity (*i.e.*, 0 to 255 for 8 bits display) and capture the intensity by either direct observation or through a planar mirror. With these images, we could construct an inverse look-up table, which we refer to as Camera-Display Transfer Curve (CDTC). During the capturing process, we apply the projected pattern with CDTC to maintain the linearity of the light source.

5.6. Experimental Results

To test out the capability and demonstrate the advantages of our framework in more realistic situations, we choose two extreme cases a step mirror and a convex mirror that fulfill the problems that we try to solve. These objects are both highly reflective surfaces. Due to the memory limitation of our reconstruction pipeline, The appearance images are measured with limited resolution(512×512 pixel), and the optimized depth maps are set to (41×41 pixel) for the whole scene.

Depth ambiguities with normal integration: Fig 5.4 show our experimental result for a scene with depth ambiguities problem. The test case consists of two planar mirror planes with different depth levels. Conventional PMD technique model surface normal information would only reveal both planes with flat surface orientation, which the normal integration could not integrate the depth change. Our framework directly models the depth information

in the optimization, which could reveal the depth difference between the two mirror planes. Note that our experimental result operates with limited appearance images and depth maps, which would produce noise for the reconstruction.



(a) Reference Photo

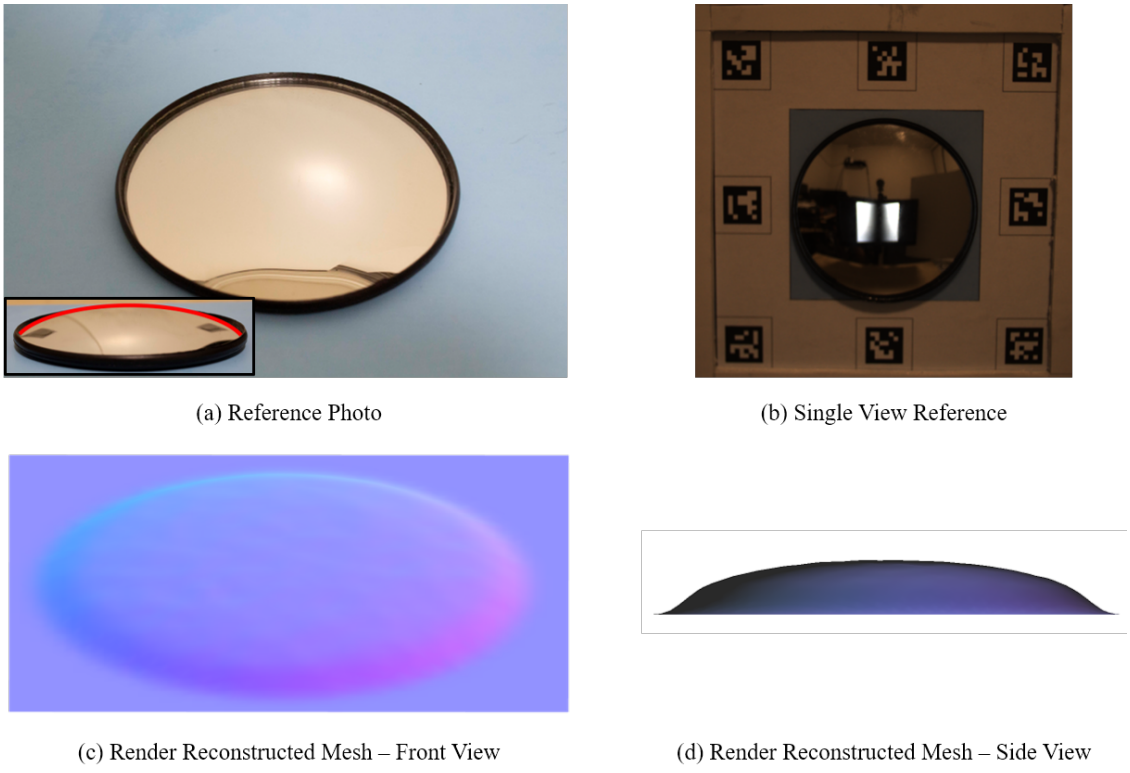
(b) Conventional PMD Reconstruction with Normal Integration

(c) Render Reconstructed Mesh – Front View

(d) Render Reconstructed Mesh – Side View

Figure 5.4. **Experimental Capture, Step Mirror** Comparison between our depth optimization framework with conventional PMD reconstruction with normal integration. The step mirror surface is shown in (a). Conventional PMD reconstruction with normal integration shown in (b). (c) and (d) Our inverse rendering base depth optimization result with render mesh.

Wide variety of surface angle: We capture a convex mirror which contain wide range of surface angle variation to show the capability of our joint multiview optimization. As shown in Fig 5.5, our compact system could only cover about one-sixth of the surface angle of this convex mirror within single view capture. We perform 9 different viewpoints with efficient overlapping to cover the whole curved surface. With our joint multiview optimization, we directly recover the smooth surface curve.



(a) Reference Photo

(b) Single View Reference

(c) Render Reconstructed Mesh – Front View

(d) Render Reconstructed Mesh – Side View

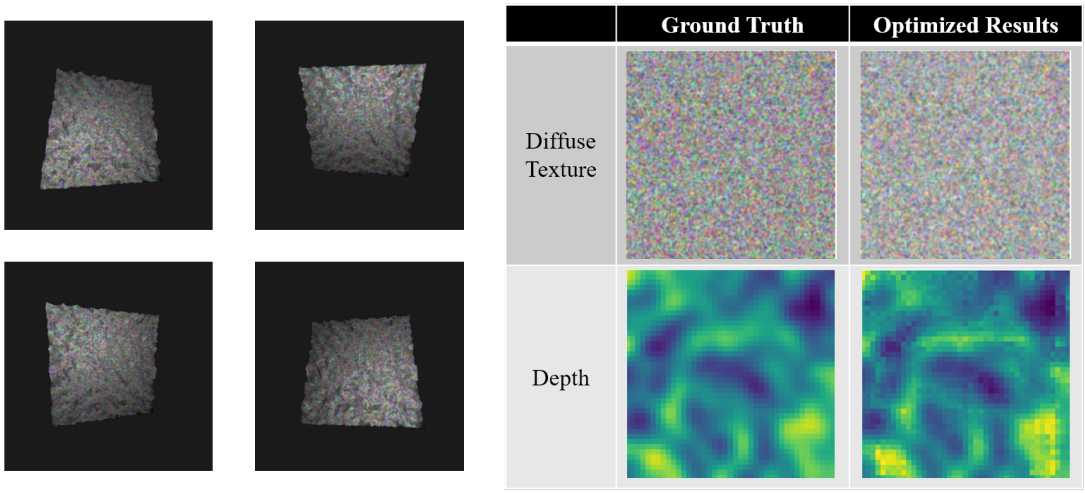
Figure 5.5. **Experimental Capture, Convex Mirror** The Convex mirror shown in (a). (b) Limited single view coverage of the variety of surface angle. (c) and (d) Our inverse rendering base depth optimization result with render mesh.

5.7. Conclusion and Future Work

To conclude this chapter, we introduce a novel inverse rendering based optimization framework with a portable capture system for surface shape recovery with highly reflective surface. We combine conventional multiview deflectometry with render-based optimization to jointly optimize the surface shape without individual measurement blending. Moreover, our framework directly optimizes the surface depth information, which could resolve the common depth/scale ambiguities in conventional appearance based shape recovery methods. We have demonstrated how to use a simple setup to acquire high quality $3D$ reconstruction results of the challenging surface. We hope our method would create a new field for appearance based shape recovery and enable more different new applications with this powerful tool.

There are still several limitations, and we think it could be solved by possible future direction. First, due to the reflectance model that we are currently using, we could only reconstruct mirror-like object, which is perfectly aligned with pure specular reflectance. However, our render base pipeline could be generalized for more complex reflectance properties. We have been testing the capabilities to deal with more complex materials with the Blinn-Phong model. From our preliminary simulation results, as shown in Fig 5.6, it could reconstruct not only the depth but also the diffuse texture. For the real data, we could further leverage the polarization camera in our system to separate diffuse and specular components for better modeling of real world material.

For the result that we showed, the reconstruct surfaces are a little bit noisy in terms of the depth change. To better constrain the optimization to deal with the noise, we are interested in changing into Mesh optimization instead of depth. With mesh optimization,



(a) Selected simulated appearance measurements

(b) Comparison of optimized diffuse texture and depth

Figure 5.6. **Simulation of complex reflectance model and joint parameters optimization** Preliminary results for complex reflectance model and joint parameters optimization with simulation. The selected simulated appearance measurement are shown in (a). (b) The diffuse texture and depth comparison between ground truth and optimized results.

we could impose more shape priors such as template mesh and regularization terms such as laplacian graph and volumetric constraint to achieve smooth surface reconstruction. Also, due to the memory limitation with the single GPU, currently, we could only do limited resolution for capture image and the optimized depth map. This could be improved with a more optimized pipeline and utilize with multiple GPUs. Moreover, currently, we use the sinusoid pattern similar to conventional deflectometry. It is an interesting problem to see the optimal illumination pattern for fewer patterns or more complex surface material. Lastly, we only focus on the geometry reconstruction with fixed scene parameters now. It is also interesting to see the possibility of having more joint scene parameters optimization such as BRDF fitting or relax the camera and display calibration, which could also potentially extend our framework to less constrained systems such as mobile devices.

CHAPTER 6

Conclusion

In the previous chapters, this dissertation introduced four novel mobile computational imaging systems for appearance modeling based surface shape recovery. The first part of the thesis focus on robust and portable photometric stereo for the Lambertian surface. In chapter 2, we solve the near light challenge, which is the most fundamental limitation in conventional photometric stereo methods by efficiently leveraging photogrammetry for lighting position estimation with geometric triangulation, which could provide better lighting estimation for the near light photometric stereo algorithm. In addition, with the coarse geometry from photogrammetry, we could regularize the low-frequency bias that produces from the normal integration step in photometric stereo results to recover the surface shape on both coarse and fine-scale with a compact portable and user-friendly $3D$ imaging solution. After this dual-camera approach, we keep thinking about relaxing the hardware constrain and also isolating the different reflectance signals to achieve better shape information. In chapter 3, we innovate a simpler un-calibrated photometric stereo framework with a single mobile device and a custom $3D$ printed widget. We exploit the linear polarization filter to efficiently suppress the specular reflection in our measurements. With the specular reflectance suppression from our unique hardware setting, it not only reduces the noise from different reflectance signals but also enables our robust hardware setup and streamlined acquisition process for the ability to capture in the wild. We believe this could greatly benefit different applications to apply surface shape recovery.

The second part of the thesis moves to address the same surface shape measurement problem but with highly reflective surface. We first propose a novel mobile deflectometry system that could measure surface shape with specular objects in the wild. By applying automated multi-view registration and blending algorithms, our method could still capture surface with large normal variation and wide field of view with compact tablet system under a hand-guided capture process. Finally, in chapter 5, inspired by the previous experience, we propose a novel rendering Based optimization for specular surface shape Estimation. We combine conventional multi-view deflectometry with render-based optimization to jointly optimize the surface shape without individual measurement blending. Also, our framework directly optimizes the surface depth information, which could provide unbias 3D information.

This thesis uses these distinct examples to highlight the soul of computational imaging, which is leveraging unique hardware settings and novel image processing algorithms to enable new capabilities in conventional cameras and mobile systems. We think the portable and robust shape recovery systems that we address in this thesis could enable lots of useful potential applications. For instance, in the cultural heritage community, it could be a powerful and user-friendly tool for collection surveys and documenting tools. For medical applications, we foresee that it could be utilized by dermatologists for skin disease follow through and remote diagnosis. Also, it could be applied to virtual reality and augmented reality applications for more immersive user experience. Last but not least, during this COVID-19 pandemic and close down situation, we start to think about the possibility of applying our framework to use the same video conferencing hardware that a significant portion of the world already has in front of them right now as they are forced to work remotely. We think it is highly possible to adapt our methods for high-quality facial modeling for different usages.

References

[1] https://en.wikipedia.org/wiki/Aztec_calendar_stone.

[2] https://www.bahai.us/bahai\protect\discretionary{\char\hyphenchar\font}{}{}temple/.

[3] Jens Ackermann and Michael Goesele. "A Survey of Photometric Stereo Techniques". In: *Found. Trends. Comput. Graph. Vis.* 9.3-4 (Nov. 2015), pp. 149–254. ISSN: 1572-2740. DOI: 10.1561/0600000065. URL: http://dx.doi.org/10.1561/0600000065.

[4] Y. Adato et al. "Toward a Theory of Shape from Specular Flow". In: *2007 IEEE 11th International Conference on Computer Vision*. 2007, pp. 1–8.

[5] Agisoft LLC. *Agisoft PhotoScan*. http://www.agisoft.com.

[6] Naveed Ahmed et al. "Robust fusion of dynamic shape and normal capture for high-quality reconstruction of time-varying geometry". In: *IN PROC. IEEE CONF. ON COMPUTER VISION AND PATTERN RECOGNITION*. 2008, pp. 1–8.

[7] Neil Alldrin, Todd Zickler, and David Kriegman. "Photometric stereo with non-parametric and spatially-varying reflectance". In: *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR* (2008). ISSN: 1063-6919. DOI: 10.1109/CVPR.2008.4587656.

[8] R. Anderson, B. Stenger, and R. Cipolla. "Color photometric stereo for multicolored surfaces". In: *2011 International Conference on Computer Vision*. 2011, pp. 2182–2189.

[9] Oliver Arold et al. "Hand-guided 3D surface acquisition by combining simple light sectioning with real-time algorithms". In: *arXiv e-prints*, arXiv:1401.1946 (Jan. 2014), arXiv:1401.1946. arXiv: 1401.1946 [physics.optics].

[10] Janet Backhouse. "A Victorian Connoisseur and His Manuscripts: The Tale of Mr. Jarman and Mr. Wing". In: *The British Museum Quarterly* 32.3/4 (1968), pp. 76–92.

[11] J. Balzer, S. Holer, and J. Beyerer. "Multiview Specular Stereo Reconstruction of Large Mirror Surfaces". In: *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*. CVPR '11. USA: IEEE Computer Society, 2011, 2537–2544. ISBN: 9781457703942. DOI: 10.1109/CVPR.2011.5995346. URL: https://doi.org/10.1109/CVPR.2011.5995346.

[12] Peter N Belhumeur, David J Kriegman, and Alan L Yuille. "The Bas-Relief Ambiguity". In: *IJCV* 35.1 (1999), pp. 33–44.

[13] Ralf B. Bergmann, Jan Burke, and Claas Falldorf. "Precision optical metrology without lasers". In: *International Conference on Optical and Photonic Engineering (icOPEN 2015)*. Ed. by Anand K. Asundi and Yu Fu. Vol. 9524. International Society for Optics and Photonics. SPIE, 2015, pp. 23 –30. DOI: 10.1117/12.2183451. URL: https://doi.org/10.1117/12.2183451.

[14] P. J. Besl. "Geometric modeling and computer vision". In: *Proceedings of the IEEE* 76.8 (1988), pp. 936–958.

[15] Neil Birkbeck et al. "Computer Vision – ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I". In: ed. by Aleš Leonardis, Horst Bischof, and Axel Pinz. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. Chap. Variational Shape and Reflectance Estimation Under Changing Light and Viewpoints, pp. 536–549. ISBN: 978-3-540-33833-8. DOI: 10.1007/11744023_42. URL: http://dx.doi.org/10.1007/11744023_42.

[16] James F. Blinn. "Models of Light Reflection for Computer Synthesized Pictures". In: *Proceedings of the 4th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '77. San Jose, California: Association for Computing Machinery, 1977, 192–198. ISBN: 9781450373555. DOI: 10.1145/563858.563893. URL: https://doi.org/10.1145/563858.563893.

[17] Thomas Bonfort and Peter Sturm. "Voxel carving for specular surfaces". In: *Proceedings Ninth IEEE International Conference on Computer Vision*. 2003, 591–596 vol.1. DOI: 10.1109/ICCV.2003.1238401.

[18] G. Bradski. "The OpenCV Library". In: *Dr. Dobb's Journal of Software Tools* (2000).

[19] Guillaume P. Butel, Greg A. Smith, and James H. Burge. "Binary pattern deflectometry". In: *Appl. Opt.* 53.5 (2014), pp. 923–930. DOI: 10.1364/AO.53.000923. URL: http://ao.osa.org/abstract.cfm?URI=ao-53-5-923.

[20] Guillaume P. Butel, Greg A. Smith, and James H. H. Burge. "Deflectometry using portable devices". In: *Optical Engineering* 54 (2015), pp. 54 –54 –9. DOI: 10.1117/1.OE.54.2.025111. URL: https://doi.org/10.1117/1.OE.54.2.025111.

[21] Tongbo Chen, Michael Goesele, and Hans-Peter Seidel. "Mesostructure from Specularity". In: *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*. CVPR '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 1825–1832. ISBN: 0-7695-2597-0. DOI: 10.1109/CVPR.2006.182. URL: http://dx.doi.org/10.1109/CVPR.2006.182.

[22] Paolo Cignoni and Roberto Scopigno. "Sampled 3D models for CH applications: A viable and enabling new medium or just a technological exercise?" In: *Journal on Computing and Cultural Heritage (JOCCH)* 1.1 (2008), p. 2.

[23] *Cultural Heritage Imaging: Reflectance Transformation Imaging (RTI)*. 2013. URL: http://culturalheritageimaging.org/Technologies/RTI/index.html.

[24] Paul Debevec. "The light stages and their applications to photoreal digital actors". In: *SIGGRAPH Asia* 2.4 (2012).

[25] Paul Debevec et al. "Acquiring the Reflectance Field of a Human Face". In: *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '00. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 2000, pp. 145–156. ISBN: 1-58113-208-5. DOI: 10.1145/344779.344855. URL: http://dx.doi.org/10.1145/344779.344855.

[26] Paul E. Debevec and Jitendra Malik. "Recovering High Dynamic Range Radiance Maps from Photographs". In: *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '97. USA: ACM Press/Addison-Wesley Publishing Co., 1997, 369–378. ISBN: 0897918967. DOI: 10.1145/258734.258884. URL: https://doi.org/10.1145/258734.258884.

[27] Paul E Debevec, Camillo J Taylor, and Jitendra Malik. "Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach". In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM. 1996, pp. 11–20.

[28] Y. Ding, J. Yu, and P. Sturm. "Recovering specular surfaces using curved line images". In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009, pp. 2326–2333. DOI: 10.1109/CVPR.2009.5206624.

[29] Julie Dorsey, Holly Rushmeier, and Franois Sillion. *Digital Modeling of Material Appearance*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2008. ISBN: 9780080556710, 9780122211812.

[30] A. A. Dubrulle. "Householder Transformations Revisited". In: *SIAM Journal on Matrix Analysis and Applications* 22.1 (2000), pp. 33–40. DOI: 10.1137/S0895479898338561. eprint: https://doi.org/10.1137/S0895479898338561. URL: https://doi.org/10.1137/S0895479898338561.

[31] Shireen Y. Elhabian, Ham Rara, and Aly a. Farag. "Towards accurate and efficient representation of image irradiance of convex-Lambertian objects under unknown near lighting". In: *Proceedings of the IEEE International Conference on Computer Vision* (2011), pp. 1732–1737. ISSN: 1550-5499. DOI: 10.1109/ICCV.2011.6126437.

[32] Carlos Hernandez Esteban, George Vogiatzis, and Roberto Cipolla. "Multiview Photometric Stereo". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30.3 (2008), pp. 548–554. ISSN: 0162-8828. DOI: http://doi.ieeecomputersociety.org/10.1109/TPAMI.2007.70820.

[33] Christian Faber et al. "Deflectometry challenges interferometry: the competition gets tougher!" In: *Proc.SPIE*. Vol. 8493. 2012, pp. 8493 –8493 –15. DOI: 10.1117/12.957465. URL: https://doi.org/10.1117/12.957465.

[34] M. Fischer, M. Petz, and R. Tutsch. *Model-Based Deflectometric Measurement of Transparent Objects*. Fringe 2013 – 7th International Workshop on Advanced Optical Imaging and Metrology, Springer (2013).

[35] Martin A Fischler and Robert C Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography". In: *Communications of the ACM* 24.6 (1981), pp. 381–395.

[36] Y. Francken, C. Hermans, and P. Bekaert. "Screen-Camera Calibration Using Gray Codes". In: *2009 Canadian Conference on Computer and Robot Vision*. 2009, pp. 155–161.

[37] Robert T. Frankot and Rama Chellappa. "A Method for Enforcing Integrability in Shape from Shading Algorithms". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 10.4 (July 1988), 439–451. ISSN: 0162-8828. DOI: 10.1109/34.3909. URL: https://doi.org/10.1109/34.3909.

[38] S. Fu et al. "Automatic extrinsic calibration of display-camera system using an annular mirror". In: *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*. 2015, pp. 451–456.

[39] Y. Furukawa and C. Hernández. *Multi-View Stereo: A Tutorial*. 2015.

[40] S. Garrido-Jurado et al. "Automatic Generation and Detection of Highly Reliable Fiducial Markers under Occlusion". In: *Pattern Recogn.* 47.6 (June 2014), 2280–2292. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2014.01.005. URL: https://doi.org/10.1016/j.patcog.2014.01.005.

[41] Jason Geng. "Structured-light 3D surface imaging: a tutorial". In: *Advances in Optics and Photonics* 3.2 (2011), pp. 128–160.

[42] Athinodoros S. Georghiades. "Recovering 3-D Shape and Reflectance from a Small Number of Photographs". In: *Proceedings of the 14th Eurographics Workshop on Rendering*. EGRW '03. Leuven, Belgium: Eurographics Association, 2003, 230–240. ISBN: 3905673037.

[43] Ioannis Gkioulekas et al. "Inverse Volume Rendering with Material Dictionaries". In: *ACM Trans. Graph.* 32.6 (Nov. 2013). ISSN: 0730-0301. DOI: 10.1145/2508363.2508377. URL: https://doi.org/10.1145/2508363.2508377.

[44] C. Godard et al. "Multi-view Reconstruction of Highly Specular Surfaces in Uncontrolled Environments". In: *2015 International Conference on 3D Vision*. 2015, pp. 19–27. DOI: 10.1109/3DV.2015.10.

[45] Dan B. Goldman et al. "Shape and spatially-varying BRDFs from photometric stereo". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.6 (2010), pp. 1060–1071. ISSN: 01628828. DOI: 10.1109/TPAMI.2009.102.

[46] L. R. Graves et al. "Infinite deflectometry enabling 2 pi -steradian measurement range". In: *Opt. Express* 27.5 (2019), pp. 7602–7615. DOI: 10.1364/OE.27.007602. URL: http://www.opticsexpress.org/abstract.cfm?URI=oe-27-5-7602.

[47] Hao Han, Shiqian Wu, and Zhan Song. "An Accurate Calibration Means for the Phase Measuring Deflectometry System". In: *Sensors* 19.24 (2019). ISSN: 1424-8220. DOI: 10.3390/s19245377. URL: https://www.mdpi.com/1424-8220/19/24/5377.

[48] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. 2nd ed. New York, NY, USA: Cambridge University Press, 2003. ISBN: 0521540518.

[49] Gerd Häusler. *Verfahren und Vorrichtung zur Ermittlung der Form oder der Abbildungseigenschaften von spiegelnden oder transparenten Objekten*. Patent DE19944354A1, (1999).

[50] Gerd Häusler et al. "Deflectometry vs. interferometry". In: *Proc.SPIE*. Vol. 8788. 2013, pp. 8788 –8788 –11. DOI: 10.1117/12.2020578. URL: https://doi.org/10.1117/12.2020578.

[51] Tim Hawkins, Jonathan Cohen, and Paul Debevec. "A Photometric Approach to Digitizing Cultural Artifacts". In: *Proceedings of the 2001 Conference on Virtual Reality, Archeology, and Cultural Heritage*. VAST '01. Glyfada, Greece: Association for Computing Machinery, 2001, 333–342. ISBN: 1581134479. DOI: 10.1145/584993.585053. URL: https://doi.org/10.1145/584993.585053.

[52] 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.* 1 (2003), pp. 1–8. ISSN: 1063-6919. DOI: 10.1109/CVPR.2003.1211400.

[53] B. K. P. Horn. "Obtaining shape from shading information". In: *The Psychology of Computer Vision* (1975), pp. 115–155.

[54] Berthold K. P. Horn and Michael J. Brooks. "The Variational Approach to Shape from Shading". In: *Comput. Vision Graph. Image Process.* 33.2 (Feb. 1986), 174–208. ISSN: 0734-189X. DOI: 10.1016/0734-189X(86)90114-3. URL: https://doi.org/10.1016/0734-189X(86)90114-3.

[55] Berthold KP Horn and Michael J Brooks. *Shape from shading*. MIT press, 1989.

[56] Lei Huang, Chi Seng Ng, and Anand Krishna Asundi. "Dynamic three-dimensional sensing for specular surface with monoscopic fringe reflectometry". In: *Opt. Express* 19.13 (2011), pp. 12809–12814. DOI: 10.1364/OE.19.012809. URL: http://www.opticsexpress.org/abstract.cfm?URI=oe-19-13-12809.

[57] Lei Huang et al. "Review of phase measuring deflectometry". In: *Optics and Lasers in Engineering* 107 (2018), pp. 247 –257. ISSN: 0143-8166. DOI: https://doi.org/10.1016/j.optlaseng.2018.03.026. URL: http://www.sciencedirect.com/science/article/pii/S0143816618300599.

[58] Xiang Huang et al. "Near Light Correction for Image Relighting and 3D Shape Recovery". In: *International Congress on Digital Heritage - Theme 2 - Computer Graphics And Interaction*. Ed. by Gabriele Guidi, Roberto Scopigno, and Pere Brunet. IEEE, 2015. ISBN: 978-1-5090-0048-7. DOI: TBA.

[59] Satoshi Ikehata et al. "Robust photometric stereo using sparse regression". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern*

Recognition. Vol. 1. 1. 2012, pp. 318–325. ISBN: 9781467312264. DOI: 10.1109/CVPR.2012.6247691.

[60] Katsushi Ikeuchi. "Determining Surface Orientations of Specular Surfaces by Using the Photometric Stereo Method". In: *Shape Recovery*. Ed. by Lawrence B. Wolff, Steven A. Shafer, and Glenn E. Healey. USA: Jones and Bartlett Publishers, Inc., 1992, pp. 268–276. ISBN: 0-86720-452-4. URL: http://dl.acm.org/citation.cfm?id=136740.136778.

[61] B. Jacquet et al. "Real-World Normal Map Capture for Nearly Flat Reflective Surfaces". In: *2013 IEEE International Conference on Computer Vision*. 2013, pp. 713–720. DOI: 10.1109/ICCV.2013.94.

[62] Henrik Wann Jensen et al. "A Practical Model for Subsurface Light Transport". In: *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '01. New York, NY, USA: ACM, 2001, pp. 511–518. ISBN: 1-58113-374-X. DOI: 10.1145/383259.383319. URL: http://doi.acm.org/10.1145/383259.383319.

[63] James T. Kajiya. "The Rendering Equation". In: *Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '86. New York, NY, USA: Association for Computing Machinery, 1986, 143–150. ISBN: 0897911962. DOI: 10.1145/15922.15902. URL: https://doi.org/10.1145/15922.15902.

[64] Maryam Khanian, Ali Sharifi Boroujerdi, and Michael Breuß. "Photometric stereo for strong specular highlights". In: *Computational Visual Media* 4.1 (2018), pp. 83–102. ISSN: 2096-0662. DOI: 10.1007/s41095-017-0101-9. URL: https://doi.org/10.1007/s41095-017-0101-9.

[65] Gudrun J Klinker, Steven A Shafer, and Takeo Kanade. "A physical approach to color image understanding". In: *International Journal of Computer Vision* 4.1 (1990), pp. 7–38.

[66] Markus C. Knauer, Jürgen Kaminski, and Gerd Häusler. "Phase measuring deflectometry: a new approach to measure specular free-form surfaces". In: *Proc.SPIE*. Vol. 5457. 2004, pp. 5457 –5457 –11. DOI: 10.1117/12.545704. URL: https://doi.org/10.1117/12.545704.

[67] *Kokomo Opalescent Glass Co.*

[68] Bruce Lamond, Pieter Peers, and Paul E Debevec. "Fast image-based separation of diffuse and specular reflections." In: *SIGGRAPH Sketches* 6 (2007).

[69] S. H. Lee et al. “Stochastic Approach to Separate Diffuse and Specular Reflections”. In: *2006 International Conference on Image Processing*. 2006, pp. 3305–3308. DOI: 10.1109/ICIP.2006.312879.

[70] Marc Levoy et al. “The digital Michelangelo project: 3D scanning of large statues”. In: *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. 2000, pp. 131–144.

[71] Marc Levoy et al. “The Digital Michelangelo Project: 3D Scanning of Large Statues”. In: *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH ’00. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 2000, pp. 131–144. ISBN: 1-58113-208-5. DOI: 10.1145/344779.344849. URL: http://dx.doi.org/10.1145/344779.344849.

[72] Tzu-Mao Li et al. “Differentiable Monte Carlo Ray Tracing through Edge Sampling”. In: *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 37.6 (2018), 222:1–222:11.

[73] Zhengqin Li et al. “Inverse Rendering for Complex Indoor Scenes: Shape, Spatially-Varying Lighting and SVBRDF From a Single Image”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020.

[74] Zhengqin Li et al. “Learning to Reconstruct Shape and Spatially-Varying Reflectance from a Single Image”. In: *ACM Trans. Graph.* 37.6 (Dec. 2018). ISSN: 0730-0301. DOI: 10.1145/3272127.3275055. URL: https://doi.org/10.1145/3272127.3275055.

[75] Stephen Lin and Heung-Yeung Shum. “Separation of diffuse and specular reflection in color images”. In: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. Vol. 1. IEEE. 2001, pp. I–I.

[76] Stephen Lin et al. “Diffuse-specular separation and depth recovery from image sequences”. In: *European conference on computer vision*. Springer. 2002, pp. 210–224.

[77] G. Liu et al. “Material Editing Using a Physically Based Rendering Network”. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017, pp. 2280–2288.

[78] Shichen Liu et al. “Soft Rasterizer: A Differentiable Renderer for Image-based 3D Reasoning”. In: *The IEEE International Conference on Computer Vision (ICCV)* (2019).

[79] Y. Liu et al. “Fast and accurate deflectometry with crossed fringes”. In: *Advanced Optical Technologies 3(4), pp. 441-445* (2014).

[80] Matthew M. Loper and Michael J. Black. "OpenDR: An Approximate Differentiable Renderer". In: *Computer Vision – ECCV 2014*. Ed. by David Fleet et al. Cham: Springer International Publishing, 2014, pp. 154–169. ISBN: 978-3-319-10584-0.

[81] David Lowe. *Method and apparatus for identifying scale invariant features in an image and use of same for locating an object in an image*. Patent US6711293B1, (2000).

[82] David G Lowe. "Object recognition from local scale-invariant features". In: *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*. Vol. 2. IEEE. 1999, pp. 1150–1157.

[83] Wan-Chun Ma et al. "Rapid Acquisition of Specular and Diffuse Normal Maps from Polarized Spherical Gradient Illumination". In: *Proceedings of the 18th Eurographics Conference on Rendering Techniques*. EGSR'07. Grenoble, France: Eurographics Association, 2007, pp. 183–194. ISBN: 978-3-905673-52-4. DOI: 10.2312/EGWR/EGSR07/183-194. URL: http://dx.doi.org/10.2312/EGWR/EGSR07/183-194.

[84] Satya P Mallick et al. "Beyond lambert: Reconstructing specular surfaces using color". In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 2. IEEE. 2005, pp. 619–626.

[85] Tom Malzbender, Dan Gelb, and Hans Wolters. "Polynomial texture maps". In: *Proceedings of SIGGRAPH 2001, Annual Conference Series*. New York, New York, USA: ACM Press, 2001, pp. 519–528. ISBN: 158113374X. DOI: 10.1145/383259.383320.

[86] Vikash K. Mansinghka et al. "Approximate Bayesian Image Interpretation Using Generative Probabilistic Graphics Programs". In: *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1*. NIPS'13. Lake Tahoe, Nevada: Curran Associates Inc., 2013, 1520–1528.

[87] Stephen Robert Marschner. "Inverse Rendering for Computer Graphics". AAI9839924. PhD thesis. USA, 1998. ISBN: 0591937395.

[88] R. Mecca, E. Rodolà, and D. Cremers. "Realistic photometric stereo using partial differential irradiance equation ratios". In: *Computers Graphics* 51 (2015). International Conference Shape Modeling International, pp. 8 –16. ISSN: 0097-8493. DOI: https://doi.org/10.1016/j.cag.2015.05.020. URL: http://www.sciencedirect.com/science/article/pii/S0097849315000679.

[89] S. K. Nayar, X. S. Fang, and T. Boult. "Removal of specularities using color and polarization". In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. 1993, pp. 583–590. DOI: 10.1109/CVPR.1993.341071.

[90] S. K. Nayar et al. “Specular surface inspection using structured highlight and Gaussian images”. In: *IEEE Transactions on Robotics and Automation* 6.2 (1990), pp. 208–218. ISSN: 1042-296X. DOI: 10.1109/70.54736.

[91] Diego Nehab et al. “Efficiently Combining Positions and Normals for Precise 3D Geometry”. In: *ACM Trans. Graph.* 24.3 (July 2005), pp. 536–543. ISSN: 0730-0301. DOI: 10.1145/1073204.1073226. URL: http://doi.acm.org/10.1145/1073204.1073226.

[92] M.T. Nguyen, Y. Ghim, and H. Rhee. “Single-shot deflectometry for dynamic 3D surface profile measurement by modified spatial-carrier frequency phase-shifting method”. In: *Sci Rep 9, 3157* (2019).

[93] F. E. Nicodemus et al. “Geometrical Considerations and Nomenclature for Reflectance”. In: *Radiometry*. USA: Jones and Bartlett Publishers, Inc., 1992, 94–145. ISBN: 0867202947.

[94] Evelyn Olesch, Christian Faber, and Gerd Häusler. “Deflectometric Self-Calibration for arbitrary specular surfaces”. In: *Proceedings of DGaO*. 2011.

[95] Gianpaolo Palma et al. “Dynamic shading enhancement for reflectance transformation imaging”. In: *Journal on Computing and Cultural Heritage* 3.2 (2010), pp. 1–20. ISSN: 15564673. DOI: 10.1145/1841317.1841321.

[96] Thoma Papadhimitri and Paolo Favaro. “A new perspective on uncalibrated photometric stereo”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2013), pp. 1474–1481. ISSN: 10636919. DOI: 10.1109/CVPR.2013.194.

[97] Thoma Papadhimitri, Paolo Favaro, and Universität Bern. “Uncalibrated Near-Light Photometric Stereo”. In: *Proceedings of the British Machine Vision Conference*. 2014, pp. 1–12.

[98] Adam Paszke et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[99] Gustavo Patow and Xavier Pueyo. “A Survey of Inverse Rendering Problems”. In: *Computer Graphics Forum* 22.4 (2003), pp. 663–687. DOI: 10.1111/j.1467-8659.2003.00716.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2003.00716.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8659.2003.00716.x.

[100] Sid Perkins. “New app reveals the hidden landscapes within Georgia O’Keeffe’s paintings”. In: *Science Magazine* (2019).

[101] Jiaming Qian et al. “High-resolution real-time 360° 3D model reconstruction of a handheld object with fringe projection profilometry”. In: *Opt. Lett.* 44.23 (2019), pp. 5751–5754. DOI: 10.1364/OL.44.005751. URL: http://ol.osa.org/abstract.cfm?URI=ol-44-23-5751.

[102] Yvain Quéau and Jean-denis Durou. “Some Illumination Models for Industrial Applications of Photometric Stereo”. In: 2015, QCAV.

[103] Yvain Quéau, Jean-Denis Durou, and Jean-François Aujol. “Normal Integration: A Survey”. In: *J. Math. Imaging Vis.* 60.4 (May 2018), 576–593. ISSN: 0924-9907. DOI: 10.1007/s10851-017-0773-x. URL: https://doi.org/10.1007/s10851-017-0773-x.

[104] Nikhila Ravi et al. *PyTorch3D*. https://github.com/facebookresearch/pytorch3d. 2020.

[105] J. Riviere, P. Peers, and A. Ghosh. “Mobile Surface Reflectometry”. In: *Computer Graphics Forum* (2016). DOI: 10.1111/cgf.12719.

[106] S. Roth and M. J. Black. “Specular Flow and the Recovery of Surface Structure”. In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*. Vol. 2. 2006, pp. 1869–1876.

[107] Christian Röttinger et al. “Deflectometry for Ultra-Precision Machining - Measuring without Rechucking”. In: *Proceedings of DGaO*. 2011.

[108] Reza Sabzevari, Alessio Del Bue, and Vittorio Murino. “Structure from Motion and Photometric Stereo for Dense 3D Shape Recovery.” In: *ICIAP (1)*. Ed. by Giuseppe Maino and Gian Luca Foresti. Vol. 6978. Lecture Notes in Computer Science. Springer, 2011, pp. 660–669. ISBN: 978-3-642-24084-3. URL: http://dblp.uni-trier.de/db/conf/iciap/iciap2011-1.html#SabzevariBM11.

[109] Johanna Salvant et al. “Photometric stereo by uv-induced fluorescence to detect protrusions on Georgia O’Keeffe’s paintings”. In: *Metal Soaps in Art* (2019).

[110] Johanna Salvant et al. “Photometric stereo by uv-induced fluorescence to detect protrusions on georgia o’keeffe’s paintings”. In: *Metal Soaps in Art*. Springer, 2019, pp. 375–391.

[111] A. C. Sanderson, L. E. Weiss, and S. K. Nayar. "Structured highlight inspection of specular surfaces". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 10.1 (1988), pp. 44–55. ISSN: 0162-8828. DOI: 10.1109/34.3866.

[112] Martin Schaffer, Marcus Grosse, and Richard Kowarschik. *High-speed pattern projection for three-dimensional shape measurement using laser speckles.* Appl. Opt. 49(18), 3622-3629 (2010).

[113] Rudolf Schwarte et al. "New electro-optical mixing and correlating sensor: facilities and applications of the photonic mixer device (PMD)". In: *Proc.SPIE.* Vol. 3100. 1997, pp. 3100 –3100 –9. DOI: 10.1117/12.287751. URL: https://doi.org/10.1117/12.287751.

[114] Steven A Shafer. "Using color to separate reflection components". In: *Color Research & Application* 10.4 (1985), pp. 210–218.

[115] B. Shi et al. "A Benchmark Dataset and Evaluation for Non-Lambertian and Uncalibrated Photometric Stereo". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.2 (2019), pp. 271–284.

[116] Keith N. Snavely. "Scene Reconstruction and Visualization from Internet Photo Collections". PhD thesis. Seattle, WA, USA, 2009.

[117] Noah Snavely, Steven M. Seitz, and Richard Szeliski. "Photo Tourism: Exploring Photo Collections in 3D". In: *ACM Trans. Graph.* 25.3 (July 2006), 835–846. ISSN: 0730-0301. DOI: 10.1145/1141911.1141964. URL: https://doi.org/10.1145/1141911.1141964.

[118] Noah Snavely, Steven M. Seitz, and Richard Szeliski. "Skeletal graphs for efficient structure from motion". In: *Proc. Computer Vision and Pattern Recognition.* 2008.

[119] Lily Strelich. "Why Are Georgia O'Keeffe's Paintings Breaking Out in Pimples?" In: *Smithsonian Magazine* (2019).

[120] Peng Su et al. "Software configurable optical test system: a computerized reverse Hartmann test". In: *Appl. Opt.* 49.23 (2010), pp. 4404–4412. DOI: 10.1364/AO.49.004404. URL: http://ao.osa.org/abstract.cfm?URI=ao-49-23-4404.

[121] Mitsuo Takeda and Kazuhiro Mutoh. "Fourier transform profilometry for the automatic measurement of 3-D object shapes". In: *Appl. Opt.* 22.24 (1983), pp. 3977–3982. DOI: 10.1364/AO.22.003977. URL: http://ao.osa.org/abstract.cfm?URI=ao-22-24-3977.

[122] R. T. Tan and K. Ikeuchi. "Reflection components decomposition of textured surfaces using linear basis functions". In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 1. 2005, 125–131 vol. 1. DOI: 10.1109/CVPR.2005.298.

[123] R. T. Tan, K. Nishino, and K. Ikeuchi. "Separating reflection components based on chromaticity and noise analysis". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26.10 (2004), pp. 1373–1379. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2004.90.

[124] Marco Tarini et al. "3D acquisition of mirroring objects using striped patterns". In: *Graphical Models* 67.4 (2005), pp. 233 –259. ISSN: 1524-0703. DOI: https://doi.org/10.1016/j.gmod.2004.11.002. URL: http://www.sciencedirect.com/science/article/pii/S1524070304001043.

[125] A. Tewari et al. "Self-Supervised Multi-level Face Model Learning for Monocular Reconstruction at Over 250 Hz". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018, pp. 2549–2559.

[126] S. Tin et al. "3D reconstruction of mirror-type objects using efficient ray coding". In: *2016 IEEE International Conference on Computational Photography (ICCP)*. 2016, pp. 1–11. DOI: 10.1109/ICCPHOT.2016.7492867.

[127] Tongbo Chen, M. Goesele, and H. . Seidel. "Mesostructure from Specularity". In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. Vol. 2. 2006, pp. 1825–1832. DOI: 10.1109/CVPR.2006.182.

[128] K. E. Torrance and E. M. Sparrow. "Theory for Off-Specular Reflection from Roughened Surfaces". In: *Radiometry*. USA: Jones and Bartlett Publishers, Inc., 1992, 32–41. ISBN: 0867202947.

[129] Isaac Trumper, Heejoo Choi, and Dae Wook Kim. "Instantaneous phase shifting deflectometry". In: *Opt. Express* 24.24 (2016), pp. 27993–28007. DOI: 10.1364/OE.24.027993. URL: http://www.opticsexpress.org/abstract.cfm?URI=oe-24-24-27993.

[130] Borom Tunwattanapong et al. "Acquiring Reflectance and Shape from Continuous Spherical Harmonic Illumination". In: *ACM Trans. Graph.* 32.4 (July 2013), 109:1–109:12. ISSN: 0730-0301. DOI: 10.1145/2461912.2461944. URL: http://doi.acm.org/10.1145/2461912.2461944.

[131] Viztronics. *TranZ Multiview Deflectometry 3D Inspection For Glass Metal*. URL: https://viztronics.com/product/transz-3d-for-glass-and-metal/.

[132] Aaron Wetzler et al. "Close-Range Photometric Stereo with Point Light Sources". In: *2014 2nd International Conference on 3D Vision*. 2014, pp. 115–122. ISBN: 978-1-4799-7000-1. DOI: 10.1109/3DV.2014.68.

[133] Tim Weyrich et al. "Principles of appearance acquisition and representation". In: *Foundations and Trends in Computer Graphics and Vision* 4.2 (2008), pp. 75–191. DOI: http:/dx.doi.org/10.1561/0600000022.

[134] Florian Willomitzer. *Single-Shot 3D Sensing Close to Physical Limits and Information Limits*. Dissertation, Springer Theses (2019).

[135] Florian Willomitzer and Gerd Häusler. "Single-shot 3D motion picture camera with a dense point cloud". In: *Opt. Express* 25.19 (2017), pp. 23451–23464. DOI: 10.1364/OE.25.023451. URL: http://www.opticsexpress.org/abstract.cfm?URI=oe-25-19-23451.

[136] Florian Willomitzer et al. "Flying triangulation - A motion-robust optical 3D sensor for the real-time shape acquisition of complex objects". In: *AIP Conference Proceedings* 1537.1 (2013), pp. 19–26. DOI: 10.1063/1.4809687. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.4809687. URL: https://aip.scitation.org/doi/abs/10.1063/1.4809687.

[137] Florian Willomitzer et al. "Hand-guided qualitative deflectometry with a mobile device". In: *Opt. Express* 28.7 (2020), pp. 9027–9038. DOI: 10.1364/OE.383475. URL: http://www.opticsexpress.org/abstract.cfm?URI=oe-28-7-9027.

[138] L. B. Wolff. "Using polarization to separate reflection components". In: *Computer Vision and Pattern Recognition, 1989. Proceedings CVPR '89., IEEE Computer Society Conference on*. 1989, pp. 363–369. DOI: 10.1109/CVPR.1989.37873.

[139] Robert J. Woodham. "Photometric Method For Determining Surface Orientation From Multiple Images". In: *Optical Engineering* 19 (1980), pp. 19 –19 –6. DOI: 10.1117/12.7972479. URL: https://doi.org/10.1117/12.7972479.

[140] Robert J. Woodham. "Photometric Stereo: A Reflectance Map Technique For Determining Surface Orientation From Image Intensity". In: *Image Understanding Systems and Industrial Applications I*. Ed. by Ram Nevatia. Vol. 0155. International Society for Optics and Photonics. SPIE, 1979, pp. 136 –143. DOI: 10.1117/12.956740. URL: https://doi.org/10.1117/12.956740.

[141] Chenglei Wu et al. "Fusing multiview and photometric stereo for 3d reconstruction under uncalibrated illumination". In: *Transactions on Visualization and Computer Graphics* (), p. 2011.

[142] Lun Wu et al. "Robust photometric stereo via low-rank matrix completion and recovery". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6494 LNCS (2011), pp. 703–717. ISSN: 03029743. DOI: 10.1007/978-3-642-19318-7_55.

[143] Zheng Yang, Alexander Kessel, and Gerd Häusler. "Better 3D inspection with structured illumination: signal formation and precision". In: *Applied optics* 54.22 (2015), pp. 6652–6660.

[144] C. Yeh et al. "Shape-from-Shifting: Uncalibrated Photometric Stereo with a Mobile Device". In: *2017 IEEE 13th International Conference on e-Science (e-Science)*. 2017, pp. 551–558. DOI: 10.1109/eScience.2017.89.

[145] Chia-Kai Yeh et al. "A Streamlined Photometric Stereo Framework for Cultural Heritage". In: *European Conference on Computer Vision*. Springer. 2016, pp. 738–752.

[146] K. j. Yoon, Y. Choi, and I. S. Kweon. "Fast Separation of Reflection Components using a Specularity-Invariant Image Representation". In: *2006 International Conference on Image Processing*. 2006, pp. 973–976. DOI: 10.1109/ICIP.2006.312650.

[147] Yizhou Yu et al. "Inverse Global Illumination: Recovering Reflectance Models of Real Scenes from Photographs". In: *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '99. USA: ACM Press/Addison-Wesley Publishing Co., 1999, 215–224. ISBN: 0201485605. DOI: 10.1145/311535.311559. URL: https://doi.org/10.1145/311535.311559.

[148] Mingjing Zhang. "Robust surface normal estimation via greedy sparse regression". PhD thesis. 2014.

[149] Z. Zhang. "Microsoft Kinect Sensor and Its Effect". In: *IEEE MultiMedia* 19.2 (2012), pp. 4–10. ISSN: 1070-986X. DOI: 10.1109/MMUL.2012.24.

[150] T. Zickler et al. "Color Subspaces as Photometric Invariants". In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. Vol. 2. 2006, pp. 2000–2010.

[151] J. Zienkiewicz, A. Davison, and S. Leutenegger. "Real-time height map fusion using differentiable rendering". In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2016, pp. 4280–4287.