

HIGH-RESOLUTION IMAGING FOR E-HERITAGE

LU, ZHENG

NATIONAL UNIVERSITY OF SINGAPORE

2011

HIGH-RESOLUTION IMAGING FOR E-HERITAGE

LU, ZHENG

B.Comp.(Hons.), NUS

A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE
NATIONAL UNIVERSITY OF SINGAPORE

2011

To my parents and wife

Acknowledgements

I owe my deepest gratitude to my advisor Michael S. Brown for his enthusiastic and patient guidance, for his brilliant insights and extremely encouraging advice, for his generous support both technically and financially, and much more. For all the wonderful things he has done for me, I am and will always be thankful.

I am heartily thankful to my mentor Moshe Ben-Ezra in Microsoft Research Asia (MSRA), who is always enlightening and supportive in both software and hardware, during and even after my internship in MSRA. I also would like to thank Yasuyuki Matsushita, Bennett Wilburn, Yi Ma (马毅) and all other members of visual computing group in MSRA, for their valuable comments and suggestions on my research work. Thanks to Moshe for allowing the use of images in this thesis.

I would like to thank my co-author Wu Zheng (吴铮), Deng Fanbo (邓凡博) and Tai Yu-Wing (戴宇荣) for their great contribution to my research work. Extra thanks to Yu-Wing. He has helped me tremendously in technical and other aspects. Besides Wu Zheng, Fanbo and Yu-Wing, I would like to thank all my other lab-mates at the National University of Singapore and friends in MSRA. Our friendship has made my life as a graduate student very colorful and enjoyable.

Thanks to Dunhuang Academy staff, Sun Zhijun (孙志军), Wang Xudong (王旭东), Yu Tianxiu (俞天秀), Jin Liang (金良), Qiao Zhaofu (乔兆福), Sun Hongcai (孙洪才) and the unsung heroes of Dunhuang Academy. Their unbelievable enthusiasms and devotion to Dunhuang are inspiring and admirable. Special thanks to Sun Zhijun for his great help and hospitality when I was in Dunhuang. Thanks to the Dunhuang Academy for allowing the use of images in this thesis.

Lastly, I would like to express my great gratitude to my parents for their un-failing love and unselfishly support, even through I rarely have the patience to explain and share my feelings. I would like to thank my wife, Tong Yu (童昱), who has been at my side since we first met. This thesis would not have been possible if without her love, understanding and support. I will keep my promise to her in St. Peter's Basilica, till the end of time.

Abstract

In the context of imaging for e-heritage, several challenges manifest, such as increased time and effort of capturing and processing data, accumulation of errors, and shallow depth-of-field. These challenges are mainly originated from the high-resolution imaging requirement and restricted working environments commonly found at cultural heritage sites. This thesis addresses problems in high-resolution 2D and 3D imaging for e-heritage under restricted working environments. In particular, we first discuss our feasibility study on imaging Buddhist art in a UNESCO cultural heritage site using a large-format digital camera. We describe lessons learned from this field study as well as remaining challenges inherent to such projects. We then devise a framework that can capture high-resolution 3D data that combines high-resolution imaging with low-resolution 3D data. Our high-resolution 3D results show much finer surface details even compared with the result produced by a state-of-the-art laser scanner. To our best knowledge, the proposed framework can produce the highest surface sampling rate demonstrated to date. Last, we introduce a method that can produce more accurate surface normals in the situation of shallow depth-of-field and show how we can improve the reconstructed 3D surface without additional setup. Our synthetic and real world experiment results show improvement to both surface normals and 3D reconstruction.

Contents

| | |
|--|-----------|
| List of Figures | x |
| List of Tables | xi |
| List of Algorithms | xii |
| 1 Introduction | 1 |
| 1.1 Challenges of 2D and 3D Imaging for E-Heritage | 4 |
| 1.1.1 High-Resolution Imaging | 4 |
| 1.1.2 Restricted Environment | 6 |
| 1.2 Overview of 2D and 3D Imaging | 7 |
| 1.3 Objective | 10 |
| 1.4 Contributions | 11 |
| 1.5 My Other Work Not in the Thesis | 13 |
| 1.6 Road Map | 13 |
| 2 Large-Format Digital Camera | 14 |
| 2.1 Overview | 15 |
| 2.2 Hardware | 16 |
| 2.2.1 Central Components | 17 |
| 2.2.2 Peripheral Components | 21 |

| | |
|---|-----------|
| 2.3 Software | 24 |
| 2.3.1 Main Functions | 25 |
| 2.3.2 User Interface | 27 |
| 2.4 My Effort | 28 |
| 3 Field Study of 2D Imaging with Large-Format Digital Camera | 30 |
| 3.1 Introduction and Motivation | 31 |
| 3.2 First Field Deployment | 36 |
| 3.2.1 Results | 38 |
| 3.3 Discussion and Summary | 41 |
| 3.3.1 Lessons Learned | 41 |
| 3.3.2 Summary | 42 |
| 4 High-Resolution 3D Imaging | 44 |
| 4.1 Introduction | 45 |
| 4.2 Related Work | 48 |
| 4.3 System Setup | 49 |
| 4.4 Surface Reconstruction Algorithm | 51 |
| 4.4.1 Surface from Normals | 51 |
| 4.4.2 Low-Resolution Geometry Constraint | 53 |
| 4.4.3 Boundary Connectivity Constraint | 55 |
| 4.4.4 Multi-Resolution Pyramid Approach | 57 |
| 4.5 Results | 60 |
| 4.6 Summary | 64 |

| | |
|---|-----------|
| 5 Photometric Stereo using Focal Stacking | 65 |
| 5.1 Introduction | 66 |
| 5.2 Related Work | 68 |
| 5.3 Focal Stack Photometric Stereo | 70 |
| 5.3.1 Focal Stack and Normals | 70 |
| 5.3.2 Normals Refinement Using Deconvolution | 71 |
| 5.3.3 Depth-from-Focus Exploiting Photometric Lighting | 73 |
| 5.3.4 Surface Reconstruction | 76 |
| 5.4 Experimental Results | 77 |
| 5.4.1 Synthetic Examples with Ground Truth | 77 |
| 5.4.2 Real Objects | 79 |
| 5.5 Summary and Discussion | 81 |
| 6 Conclusion | 87 |
| 6.1 Summary | 88 |
| 6.2 Review of Objective | 89 |
| 6.3 Future Directions | 90 |
| A The dgCam Project: A Digital Large-Format Gigapixel Camera User Manual | 91 |
| A.1 Scope | 91 |
| A.2 User Interface | 91 |
| A.2.1 Main Window | 92 |
| A.2.2 Capture Control Window | 93 |
| A.2.3 Calibration Window | 96 |

| | |
|---|------------|
| A.2.4 Manual Focus Window | 97 |
| A.3 Working with the Software | 98 |
| A.3.1 Start or Stop Camera | 98 |
| A.3.2 Snapshot | 101 |
| A.3.3 Manual Focus | 102 |
| A.3.4 Calibrate Dark Current, White Images, Vignetting or White Balance | 103 |
| A.3.5 Cameras Alignment | 103 |
| A.3.6 Capture Image | 107 |
| A.3.7 Fast Stitch | 108 |
| A.3.8 Focal Stacking | 108 |
| A.4 Summary | 108 |
| Bibliography | 109 |

List of Figures

| | | |
|-----|--|----|
| 1.1 | A cultural heritage site, the Mogao Caves, Dunhuang, China. | 2 |
| 1.2 | Comparison of high-resolution and low-resolution 2D and 3D of the same scene | 6 |
| 2.1 | Camera overview schematics - top | 17 |
| 2.2 | Camera overview schematics - side | 18 |
| 2.3 | Camera overview schematics - back | 19 |
| 2.4 | The skeleton of the camera | 20 |
| 2.5 | Image plane stage scanning | 21 |
| 2.6 | Camera pier | 23 |
| 2.7 | Illumination | 24 |
| 2.8 | Main window | 27 |
| 3.1 | Dunhuang cave structure | 32 |
| 3.2 | Dunhuang current and proposed imaging method | 33 |
| 3.3 | Our high-resolution large-format digital camera in a cultural heritage project in Mogao Cave #46 | 37 |
| 3.4 | An image from part of the north wall of Mogao Cave #46 | 39 |
| 3.5 | An image from part of the east wall of Mogao Cave #418 | 40 |

| | | |
|------|---|----|
| 4.1 | Comparison of our results against a state-of-the-art industrial 3D scanner | 46 |
| 4.2 | Our 3D imaging setup | 50 |
| 4.3 | The osculating arc constraint [101] for surface reconstruction | 52 |
| 4.4 | Example of surface reconstruction with/without including the geometric constraints | 54 |
| 4.5 | Effect of the boundary connectivity constraints | 56 |
| 4.6 | Example of the benefits of the multi-resolution scheme | 58 |
| 4.7 | Evolution of our 3D surface up the multi-resolution pyramid | 59 |
| 4.8 | 3D reconstruction of the <i>elephant</i> figurine | 61 |
| 4.9 | 3D reconstruction of the <i>man</i> figurine | 62 |
| 4.10 | Full-size comparison with an industrial laser scanner | 63 |
| | | |
| 5.1 | Work flow of our system | 67 |
| 5.2 | The estimated normals, with and without, deconvolution refinement | 74 |
| 5.3 | The estimated depth map using depth-from-focus, with and without photometric lighting | 76 |
| 5.4 | Synthetic examples in Maya | 80 |
| 5.5 | An example with heavy texture and pitted surface | 82 |
| 5.6 | Normal map and 3D reconstruction result of <i>statue</i> figurine | 83 |
| 5.7 | Normal map and 3D reconstruction result of <i>angel</i> figurine | 84 |
| 5.8 | Normal map and 3D reconstruction result of <i>duck</i> figurine | 85 |
| | | |
| A.1 | Main window | 92 |
| A.2 | Capture control window | 94 |

LIST OF FIGURES

| | | |
|------|--|-----|
| A.3 | Calibration window | 95 |
| A.4 | Manual focus window | 98 |
| A.5 | Properties window - image properties tab | 99 |
| A.6 | Properties window - image properties 2 tab | 99 |
| A.7 | Properties window - camera control tab | 100 |
| A.8 | Properties window - video proc amp tab | 100 |
| A.9 | Properties window - video format | 101 |
| A.10 | Cameras alignment window | 104 |
| A.11 | View window of the auxiliary video camera | 105 |
| A.12 | View window of the main camera | 106 |
| A.13 | Selected region to be captured | 107 |

List of Tables

- 5.1 Comparison on average angular error (in degrees) of normals among our method and the all-in-focus methods with and without bilateral filtering. Comparisons are on textured objects. The last row shows the results are virtually identical when the object is textureless. 79

List of Algorithms

| | | |
|---|--|----|
| 1 | Depth-from-focus using photometric stereo lighting | 75 |
|---|--|----|

Chapter 1

Introduction

Photographers deal in things which are continually vanishing and when they have vanished there is no contrivance on earth which can make them come back again.

Henri Cartier-Bresson

Through the long history of human civilization, our ancestors have left large amounts of precious cultural heritage. Culture heritage can be generally categorized into two groups: tangible artifacts, such as architectural structures, paintings, statues, and frescos; and intangible attributes that describe a particular society or culture, such as folklore, traditions, language, and knowledge. Figure 1.1 shows a picture of a cultural heritage site [92], the Mogao Caves, in Dunhuang, China. The Mogao Caves contain a large number of tangible artifacts over 1000 years old including paintings, frescos, and architectural structures.

Cultural heritage is considered worthy of preservation for the future due to various reasons, such as its significance to archeology, architecture, and science



Figure 1.1: A cultural heritage site, the Mogao Caves, Dunhuang, China.

and technology of a culture. However, due to climate, environmental, human, and other factors, tangible cultural heritage that survives today is often in the danger of degradation or disappearance. Hence, the preservation of tangible cultural heritage becomes an important task to the contemporary generation. Motivated by the prominence and the challenges faced, this thesis concerns itself with the digital presentation of tangible culture heritage. In the rest of the thesis, culture heritage preservation refers to this tangible aspect.

Cultural heritage preservation includes archaeological excavation, tangible archaeological research, active management, conservation, exhibition, and so on. With the prevalence of computers, computing technologies are increasingly playing a significant role in culture heritage preservation. This is commonly referred

to as *e-heritage*. Existing efforts for e-heritage can be classified into three categories: 1) *data acquisition* which involves tasks such as cultural heritage imaging in both 2D and 3D, and meta information collection; 2) *data processing and analysis* which are responsible for processing and analyzing the large amount of data captured in data acquisition to benefit cultural heritage preservation; and 3) *data utilization* which mainly serves the purpose of education and exhibition of cultural heritage using digital means.

While applications in above three categories may vary in terms of objectives, their success hinges on the high quality presentation of cultural heritage in a digital format. Hence, 2D and 3D imaging of cultural heritage is a cornerstone for all the e-heritage initiatives.

There are two main characteristics of 2D and 3D imaging for e-heritage: high-resolution imaging requirement and restricted working environment. Note that resolution refers to the sampling rate on the object, instead of pixel count. In other words, we hope to sample and resolve more points per unit area on the target object. Higher resolution not only increases the time and effort of capture, but also brings problems such as high demand of processing time and computer power, accumulation of errors due to various reasons and so on. However, in the current literature pertaining to e-heritage, few works specifically aim to address problems caused by high resolution.

Different from the high-resolution requirement, 2D and 3D imaging for e-heritage can also be constrained by the physical environment. In the context of e-heritage, imaging is often carried out in a restricted working environment such as a cultural heritage site. Such restricted environments often pose additional difficulties to the imaging process. One such difficulty is shallow depth-of-field. This

may due to large aperture used under weak illumination or close object-camera distance. Under such circumstances, focal stacking, i.e. capturing multiple images at various focus distance, is a well-known way to extend the depth-of-field. Though many previous works focus on extending depth-of-field using focal stack, little attention is given to the utilization of focal stack data in the context of 3D imaging. In fact, very few prior works in computer vision and graphics discuss real field work undertaken in a cultural heritage site and summarize lessons learned.

In the rest of this chapter, some prominent challenges of imaging for e-heritage are highlighted, followed by a short review of research work in 2D and 3D imaging to motivate this thesis. Afterwards, the research objectives and contributions of this thesis are presented followed by a brief description of my other work not in this thesis but still in the context of e-heritage. The chapter is concluded with the road map of this thesis.

1.1 Challenges of 2D and 3D Imaging for E-Heritage

This section highlights three prominent challenges manifested from 2D and 3D imaging for e-heritage: increased time and effort of capturing and processing data, accumulation of errors and shallow depth-of-field. These challenges are discussed under the two main characteristics of imaging for e-heritage: high-resolution requirement and restricted working environments.

1.1.1 High-Resolution Imaging

One of the purposes of imaging for e-heritage is to preserve physical artifacts in a digital format. This allows people to access a digital copy in the future if the

artifact has degraded or disappeared permanently. Hence, in the context of e-heritage, higher resolution imaging is very important and always desired. For example, museums require a minimum density of $500ppi$ on the object (20 samples per mm^2)¹ for digital archiving of paintings [59]. Figure 1.2 shows examples of high and low resolution 2D and 3D of the same scene.

With high-resolution imaging, many challenges manifest. First, the time and efforts for imaging high-resolution data are significantly increased. Taking 2D imaging as an example, a full image of a wall of fresco sized $4m \times 5m$ with $75ppi$ has size of 14764×11811 pixels. To capture the full image by translating a conventional digital single-lens reflex (DSLR) camera (assuming the sensor size is 5616×3744 pixels) with only 25% overlapping, it needs to take at least 12 images and the total data size will be about 1.4*giga* bytes (16 bits data without compression). By increasing the resolution to $300ppi$, the size of the full image will be 59055×47244 pixels and there requires at least 238 images captured with the same camera. The total data size will be 28.6*giga* bytes (16 bits without compression). Second, high-resolution usually requires the camera to be closer to the object, hence reducing the depth-of-field of the images captured. As sharpness is also an important requirement for e-heritage, extending the depth-of-field is a must to ensure the whole object will remain in focus. Third, the large amount of data will not only increase the precessing time but also accumulate errors due to various reasons such as computational error, sensor noise, or camera distortion. For example, as the number of images been stitched increases, slight inaccuracy in the pair-wise stitching will be accumulated to larger errors in the final output even with bundle

¹As a convention, this thesis uses *ppi* (pixel per inch) to describe the resolution of 2D images, and samples per mm^2 for 3D data.

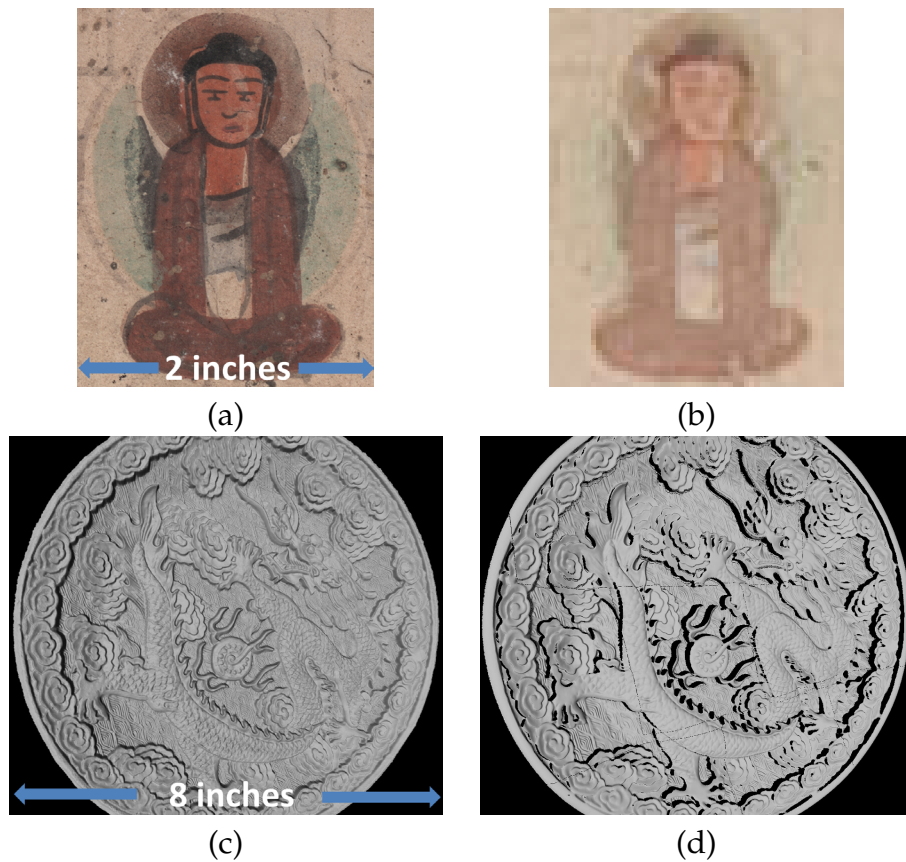


Figure 1.2: Comparison of high-resolution and low-resolution: (a) 2D image from a large-format digital camera, 400ppi , (b) 2D image from a conventional DSLR, 30ppi (interpolated to the same size as (a)), (c) 3D object, $600\text{ samples per } \text{mm}^2$, (d) 3D object from a state-of-the-art laser scanner, $168\text{ samples per } \text{mm}^2$.

adjustment.

1.1.2 Restricted Environment

In many cultural heritage sites, the working environments of 2D and 3D imaging setups are often restricted. Restriction refers to physical inflexibility such as limited use of hardware or space. The following highlights some typical restrictions and shows that such restrictions may create potential problems for imaging.

Weak Illumination Certain types of cultural heritage such as historical documents or paintings, are strictly inhibited from long exposure of strong lights. Illumination exposure can significantly change an object's attributes such as color, or reduce the life of the object. This weak illumination requirement may potentially force the use of a large aperture that significantly decreases the depth-of-field.

Spatial Restriction Besides restrictions on illumination, most of artifacts are not allowed to be touched to avoid damage. In some situations, an additional safe distance is kept to further reduce the risk of breaking the target object. On some cultural heritage sites, such as caves or tomes, the working space may be very small due to reasons such as indoor structure or outdoor geography. All these spatial restrictions may limit types of devices or setups that can be used. For example, some of the Dunhuang caves only allow one person inside. In such case, the distance of the camera to the object may be short. As a result, the depth-of-field will be significantly decreased.

1.2 Overview of 2D and 3D Imaging

In the area of computer vision and graphics, several efforts have been made to investigate 2D and 3D imaging for e-heritage, e.g. the Great Buddha Project [63], the Digital Bayon Archival Project [43], and the Digital Michelangelo Project [51]. This section briefly describes the recent efforts in high-resolution 2D and 3D imaging.

High-Resolution 2D Imaging Despite the significant resolution improvement of DSLR and medium-format digital cameras in recent years, there is still considerable demand for building large-format digital camera in areas that desire higher reso-

lution, such as e-heritage. In 2001, Flint introduced a semi-digital high-resolution large-format camera [25]. The camera uses analog photographic film for initial capture, which is scanned to digital images. While Flint's camera can produce gigapixel images, each capture costs approximately \$50 due to the use of film. Focal-plane-array technology using array of sensors, is commonly used to obtain gigapixel images in the area of astronomy, e.g. Pan-Starrs [91]. The expensive cost limits the use of this type of cameras only in telescopes. In attempt to lower the cost, Wang *et al.* [96] introduced a low-cost high-resolution scan camera capable of capturing *490megapixel* images in 2004. However, this camera suffers from several problems such as the need of multiple scans to produce a color image, limited gain and exposure control, and scanline artifacts. Among commercial solutions for large-format imaging, Anagramm David [4] provides the highest resolution, up to *340megapixel* images. However, due to the tri-linear sensor used, the camera can only capture one column of pixels per capture, and hence requires a long exposure time.

Instead of building large-format digital camera, high-resolution 2D imaging can be approached through image mosaicing. However, current image mosaicing techniques [61, 62, 15, 16, 17, 85, 86] use homography-based image alignment which only works in the ideal case of a purely planar surface or very distance scenes. Close objects or scenes that are not perfectly planar will cause visible seams or alignment artifacts in the final mosaic. While image blending and/or seam-cutting usually hide seams well, perceptually masking errors may not be an acceptable solution for e-heritage.

In 2010, Ben-Ezra [10] introduced a low cost large-format tile-scan digital camera specifically designed for high-resolution imaging in museums and cultural heritage

sites. The camera is capable of capturing (and truly resolving) more than a gigapixel image. And the camera is designed to be able to operate at close ranges. In addition, the camera can be programmed to automatically capture focal stack images to significantly enhance the depth-of-field. Given these benefits, most of the data in this thesis are captured with this camera. The details of the camera are described in Chapter 2.

General 3D Imaging 3D imaging has been an active topic in computer vision and graphics for many years. Existing approaches can be categorized into three types: passive, active, and hybrid methods that combine two or more methods.

Passive methods include multi-view stereo [24, 70, 74, 82, 33, 29, 84, 14, 30], and shape from shading [44, 26, 39, 67, 106, 23, 100, 71, 77, 75, 98]. Active methods include structured-light [95, 9, 79, 103, 46] and photometric approaches such as photometric stereo [19, 83, 7, 40, 37, 87, 64, 28, 8, 31]. On the one hand, while approaches such as multi-view stereo and structured-light can provide very accurate estimation of global shape of object surface, the main drawback lies in their inability to provide high-resolution surface details. For example, most multi-view stereo approaches need to re-sample the scene points. This resampling decreases the spatial resolution. Similarly, the resolution produced by a structured-light system depends on the resolution of the projector, which usually has much lower resolution compared with camera sensor. On the other hand, photometric stereo approaches are good at capturing very fine surface details of the target object. The current state-of-the-art approaches in photometric stereo can capture more subtle surface details than those in multi-view stereo and structured-light systems. However, one well-known drawback of photometric stereo approaches is that they

cannot adequately capture the global shape of the target object.

Observing the pros and cons of the above methods, researchers opt for hybrid methods that integrate two or more methods. Hybrid methods include approaches that combine shape from motion and photometric stereo [105, 52, 45, 38], positional (3D points) data and normals [89, 6, 27, 49, 42, 12, 18, 66], silhouette or visual hull and normals [13, 94, 35], and recently normals and volume carving [93]. As expected, the obvious advantage of hybrid methods is their ability to obtain both good global shape and fine surface details. Drawbacks of such hybrid methods include complicated system setup and increase of capturing time due to multiple methods used.

While hybrid methods are able to produce satisfactory results, they do not specifically address problems caused by the high-resolution requirement for e-heritage. For example, Nehab *et al.* [66] proposed a method fusing positional data and surface normals using a linear formulation. However, in their work, the surface geometry and the photometric data were of nearly the same resolution. In addition, as mentioned in Section 1.1, in real field work of e-heritage, restricted working environment may cause shallow depth-of-field. No previous work combines 3D imaging with methods that extend depth-of-field such as focal stacking.

1.3 Objective

The main objective of the research presented in this thesis is to improve high-resolution 2D and 3D imaging for e-heritage. More specifically, we aim to address the following problems:

- In cultural heritage sites, how feasible is high-resolution 2D image capture

using a large-format digital camera? What are the problems and challenges in such real world projects?

- How can we obtain high-resolution 3D data with a setup consists of high-resolution photometric stereo and a structured-light system with much lower resolution?
- How can we improve 3D imaging in a restricted environment where one can only rely on a camera with a shallow depth-of-field and without the benefit of auxiliary depth information such as that obtained from a structured-light system?

1.4 Contributions

In terms of original contributions to the research community, this thesis makes the following:

1. Discusses recent field work that captured high-resolution images of Buddhist art at the UNESCO world heritage site, the Mogao Caves. As a feasibility study, the thesis reports on the current challenges faced by the Dunhuang Academy in their imaging efforts and how the use of a large-format digital camera can improve the quality of the imaging process while reducing time and efforts. The thesis also describes lessons learned from this field study as well as remaining challenges inherent to such projects. This work is currently conditionally accepted (i.e. minor revision) by the ACM Journal of Computer and Cultural Heritage (JOCCH).

2. Develops an imaging framework to acquire 3D surface scans at high-resolution (exceeding 600 samples per mm^2). The approach couples a standard structured-light setup and photometric stereo using a large-format digital camera. To deal with the significant asymmetry in the resolution between the low-resolution positional data and the high-resolution surface normals, a multi-resolution patch-wise surface reconstruction scheme is proposed. In addition, boundary constraints are used to ensure patch coherence. Our imaging framework can produce 3D scans that show exceptionally detailed 3D surfaces far exceeding existing technologies. This work is published in CVPR'10 [53].
3. Develops a unique setup that combines photometric stereo with focal stacking. In some restricted situation, a photometric stereo framework can only rely on a camera with a shallow depth-of-field and without the benefit of auxiliary depth information from structured-light or range-scanning. Such narrow depth-of-field requires focal stacking to properly image the object. The proposed method utilizes this additional information in the photometric stereo process. In particular, the proposed approach regularizes the normals against the varied focused images to improve normal estimation. It is also discussed that how the photometric lighting can be used to improve estimations for depth-from-focus which can be incorporated into the overall framework. Our results show that the proposed framework produces better 3D data than naive approach.

1.5 My Other Work Not in the Thesis

During my PhD candidature, my primary research interest is using vision techniques to solve problems in e-heritage. Besides works on 2D and 3D imaging detailed in this thesis, I have worked also on problems of historical document image restoration. In particular, I worked on ink-bleed problem that aims to reduce ink-bleed that penetrates from the opposite side of the document, published in CVPR'09 [54]. I have also worked on binarization problem that aims to separate document foreground (the text ink) from its background (the paper), published in WACV'09 [55]. In addition, I co-developed a software framework, Binarazation-Shop, that combines a series of binarization approaches that have been tailored to exploit user assistance, published in JCDL'10 [21]. Interested readers can refer to these publications for details.

1.6 Road Map

The rest of this thesis is organized as follows: Chapter 2 describes the gigapixel large-format digital camera used in this thesis as well as my involvement in the development of this camera. Chapter 3 describes a real field work examining the use of a large-format digital camera in a cultural heritage setting. Chapter 4 presents a high-resolution 3D imaging framework and Chapter 5 shows how we use focal stack data to refine normal estimation and use photometric stereo to improve depth-from-focus. Finally, Chapter 6 concludes the thesis along with discussions of future research directions.

Chapter 2

Large-Format Digital Camera

If your pictures aren't good enough,
you aren't close enough.

Robert Capa

In this chapter, we describe the large-format digital camera [11] used in this thesis. The camera is specifically designed for high-resolution imaging in museums and cultural heritage sites. In this thesis, most of the data are captured with this camera. While the camera is designed and built by Dr. Moshe Ben-Ezra from Microsoft Research Asia, I developed the entire operation software including the user manual when I was an intern in Microsoft Research Asia from August 2009 to June 2010. Section 2.1 gives an overview of the camera. Section 2.2 describes the hardware design, followed by the software in Section 2.3. Section 2.4 concludes the chapter by summarizing my involvement in the development of the camera. The details of the algorithms related to the camera, such as demosaicing, image stitching, and focal stacking, are not included in this thesis. Please see [11] for such

information.

2.1 Overview

In recent years, DSLR cameras have increased their resolution significantly. While the resolution is satisfiable for general usage, higher resolution is desirable for applications such as e-heritage. To satisfy such a need, Ben-Ezra from Microsoft Research Asia developed a large-format tile-scan digital camera capable of acquiring high-quality and high-resolution images of static scene [11]. The main advantages of the camera can be summarized in the following:

1. The camera can capture (and truly resolve) images that are larger than one gigapixel in size. Please see [11] for the detailed resolution evaluation of the camera.
2. The camera is very flexible and can operate at close range, and can capture objects with wide selection of resolutions ranging from approximately 300ppi to over 3000ppi, depending on lens used.
3. The camera can be programmed to automatically capture focal stack images that can be used to produce results with a significantly enhanced depth-of-field using the camera operation software.
4. The camera can capture photometric stereo data with the same high-resolution by attaching controllable lights.
5. The camera is made of mostly from off-the-shelf components so that the overall cost is low compared to cameras with similar capability.

As a tile-scan camera, the capture and processing time highly depend on the exposure time and the number of tile images to be captured. To illustrate the acquisition speed of the camera, we show a typical e-heritage example that requires maximum number of tiles (240 tiles), four seconds to expose a single tile, and five focal stack images to keep the target object in focus. Under such setting, it takes approximately two hours to capture all the tile images (a total of 1200 tiles). On a Xeon E5540 2.53-GHz machine (using a single core), it takes approximately one hour to automatically process all the tiles, including various calibration, de-mosaicing, computing all-in-focus image from focal stack, and the final stitching (see Section 2.3).

2.2 Hardware

In a nutshell, the large-format tile-scan digital camera consists of a large-format lens attached on a motorized translation stage in front, the *focusing stage*, and a sensor attached to the two motorized translation stages, the *horizontal and vertical image plane stages*, at the back. Images are captured by moving the sensor in a grid fashion along the horizontal and vertical image plane stages. Figure 2.1, Figure 2.2 and Figure 2.3 show the camera overview schematics from different views. In the following, we categorize the main components of the camera based on whether the component is central or peripheral to the camera. For each main component, we illustrate its purpose and the related specification.

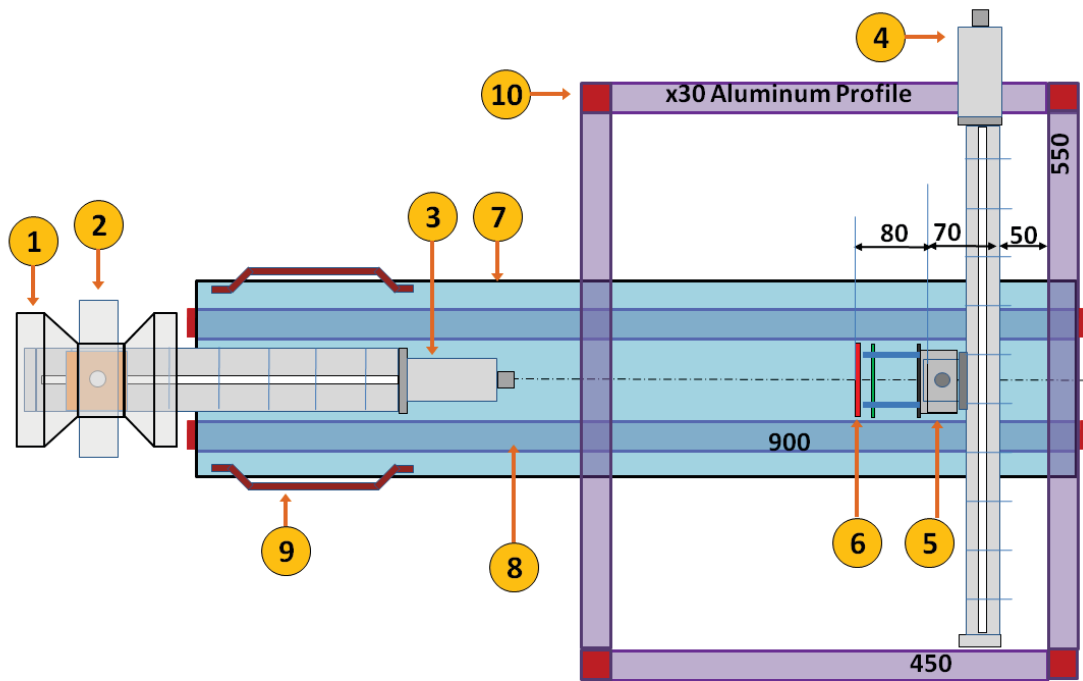


Figure 2.1: Camera overview, top view. The front of the camera (left) consists of the main lens (1) that is attached to the lens holder (2) that is attached to the focusing stage (3). The back of the camera holds the main sensor (6), which is attached to the vertical image plane stage (5) that is attached to the horizontal image plane stage (4). All stages are attached to the main breadboard (7), which is supported by two aluminum rails (8). Covers (10) are placed at the open ends of the rails. Two handles (9) are attached to the front of the main board. (Image courtesy of Moshe Ben-Ezra)

2.2.1 Central Components

The central part of the camera consists of three components: lens, sensor, and mechanical skeleton.

Lens In this camera, the main lens used is a Schneider Optics Apo Symmar L 480/8.4 [80]. This lens is selected primarily based on the following two reasons.

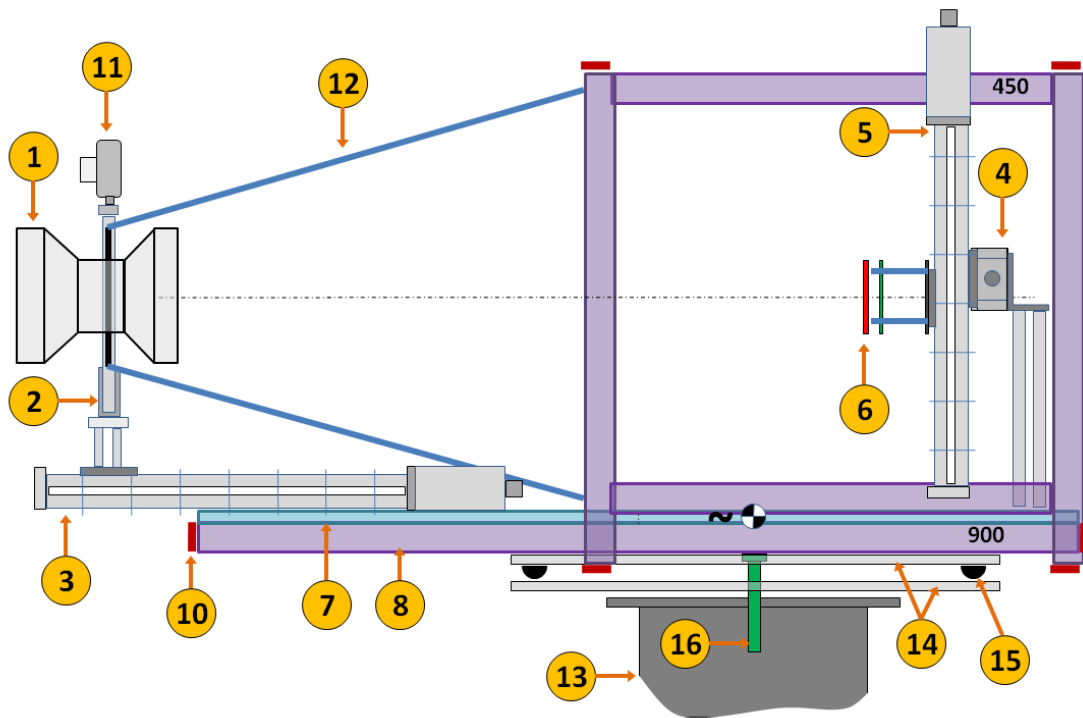


Figure 2.2: Camera overview, side view. Item (1-10) are the same as in the top view, additionally we can see the auxiliary video camera (11) that is attached to the main lens holder and moves with it, the bellow (12), and the base of the camera. The camera is carried by a telescope pier (13) (only top is shown (13)). Two aluminum plates (14) and four Sorbothane bumpers provide vibration isolation/damping. A safety screw (16) holds the two plates together while allowing relative motion (see back view for more details). (Image courtesy of Moshe Ben-Ezra)

First, this large-format lens can produce a large image circle of 500mm at $f/22$ and a standard field-of-view of 56° . Such a large image circle has the capability of covering all locations of the image plane stages. Second, the lens also has the advantage of low distortion and uniform resolution throughout its field-of-view.

Sensor A Kodak KAI11002 CCD [47] is used as the main sensor of the camera. This sensor has 11megapixel with $9\mu\text{m}$ per pixel. The CCD is taken from a

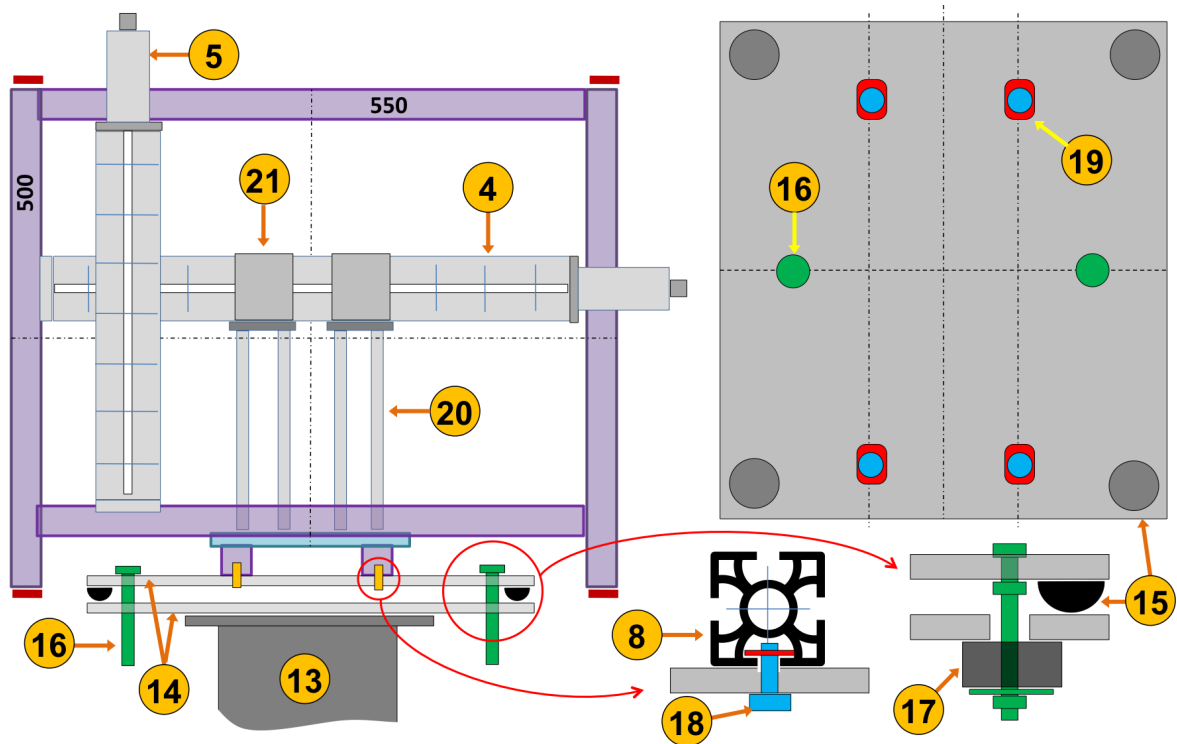


Figure 2.3: Camera overview, back view. This view shows the steel posts (20) and the angle brackets (21) used to attach the horizontal image plane stage (4) to the breadboard. The vibrations suspension plates are also shown with magnified inserts on the right. The safety screw (16) is firmly attached to the top plate (14), and goes through a wider hole in the bottom plate. (Image courtesy of Moshe Ben-Ezra)

Lumenera [56] camera because it provides a flexible SDK and USB 2.0 interface.

Mechanical Skeleton To provide firm support for the lens and sensor, the skeleton of the camera is made of several off-the-shelf optical components. Figure 2.4 shows an overview of the skeleton of the camera. These components, such as the lens holder, the breadboard, and motorized translation stages, are rigid and accurate to prevent the camera from unnecessary image blurring and distortion.

As shown in Figure 2.4, the lens holder is firmly attached to the lens board that

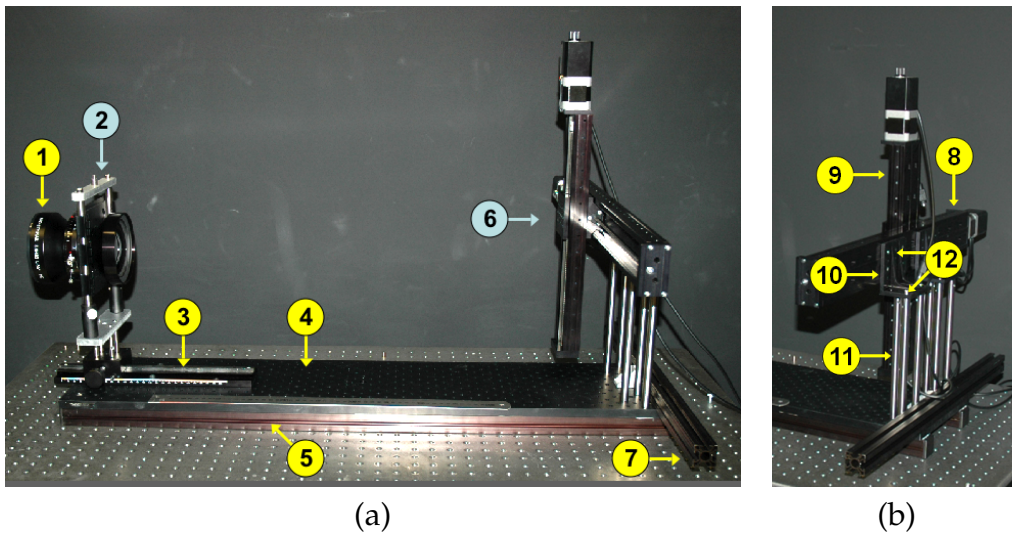


Figure 2.4: The skeleton of the camera. (1) lens, (2) lens holder, (3) Focusing Stage (manual), (4) breadboard, (5) supporting rail, (6) image plane stages, (7) first enclosure rail, (8) horizontal image plane stage, (9) vertical image plane stage, (10) angle bracket, (11) steel post (12mm), (12) M6 screws. (Images courtesy of Moshe Ben-Ezra)

has grooved edges align against the posts. The lens holder is made of optical table posts and custom made aluminum bars. Besides firm support, the holder allows for easy changes of various lens. It is also important to note that the lens holder has a mounting point where a low-resolution auxiliary video camera is firmly attached. The auxiliary video camera mainly serves as the viewfinder of the main camera.

As the backbone to connect various central components of the camera, the baseplate used is a $900mm \times 200mm \times 12.7mm$ double density breadboard from Thorlabs [90]. Two construction rails are attached at the bottom to support the baseplate.

Three motorized translation stages from Zaber [104] are used in the camera. The first motorized translation stage is located under the lens holder and also attached to the baseplate. This translation stage acts as the focusing stage. Focusing of the

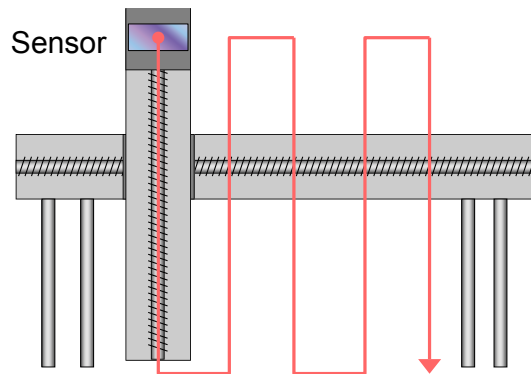


Figure 2.5: Image plane stage scanning. The ZigZag motion minimizes the moving mass. (Image courtesy of Moshe Ben-Ezra)

camera can be achieved by moving the lens holder along the focusing stage. The other two motorized translation stages are located at the other end of the baseplate. These two translation stages are responsible for moving the sensor. The stages are arranged to be perpendicular to each other. As shown in Figure 2.5, by moving the sensor in a zig-zag fashion, the moving mass is minimized to prevent motion blur. In addition, through such arrangement, the camera could scan the whole image plane step by step, and more importantly, capture multiple images at the same location. This is important for capturing photometric stereo data because there is no need to align the images under different illuminations.

2.2.2 Peripheral Components

Peripheral components are also salient parts of the camera that support and complement the central components. Four peripheral components are briefly discussed below: auxiliary video camera, enclosure box, pier, and illumination.

Auxiliary Video Camera The auxiliary video camera used in our camera is a

Dragonfly camera from Point Gray [73]. This camera has a 1/3" CCD with a varifocal lens. The video camera provides a low-resolution continuous view of the scene and serves as the viewfinder of the main camera. Under certain circumstances, this video camera can also be used for other purposes, e.g., capturing positional data using structured-light system as discussed in this thesis. By allowing the user to specify the corresponding points in the scene, our camera operation software is able to align the field of view of the video camera and the main camera (see Appendix A). The alignment needs only to be performed once unless the focal length of either cameras is changed.

Enclosure Box The enclosure box of the camera consists of aluminum frames with rails and walls of Styrofoam or extruded-polystyrene. Several parts are connected to the enclosure box. First, at the back of the box, a thermoelectric active cooling setup is used to regulate the internal temperature. This is important as the camera is intended to be used in low light conditions such as museums or cultural heritage sites. Under such circumstances, the cooling setup can take effect to lower the internal temperature of the camera and hence reduce sensor noise. Second, a custom made Neoprene coated nylon bellow is used to connect the lens to the main frame. Third, Sorbothane bumpers are placed under the enclosure box to dampen vibrations from floors or from other camera mechanical parts (e.g., movement of vertical image plane stage).

Pier A telescope pier is used to mount the camera because the weight of the camera exceeds the maximum load of a conventional camera tripod (see Figure 2.6). Using the pier, one can easily move the camera with wheels, lock the camera in a fix



Figure 2.6: Camera pier. The camera is mounted on a telescope pier that allows adjusting the height of the camera. (Images courtesy of Moshe Ben-Ezra)

location and elevate the camera up and down.

Illumination Two types of illumination are used in this camera: a static ring illumination and computer controlled directional illumination (see Figure 2.7 for an illustration). The former consists of 12 halogen lights (Osram Cool Blue [69]) arranged around the lens. This can be used to shorten the exposure time while still obtaining the correct color information. The latter consists of four fluorescent light bulbs that are placed at the four corners of the camera frame and are controlled by a USB relay from Phidgets [72]. Such directional illumination is mainly used for special purposes such as photometric stereo and multi-spectral imaging.

It is important to note that in order to reduce specular reflections for tasks that require diffused surface, such as capturing photometric stereo data, as discussed in this thesis, cross polarization needs to be used. This can be achieved by putting polarizers on the lens and in front of the directional illumination.

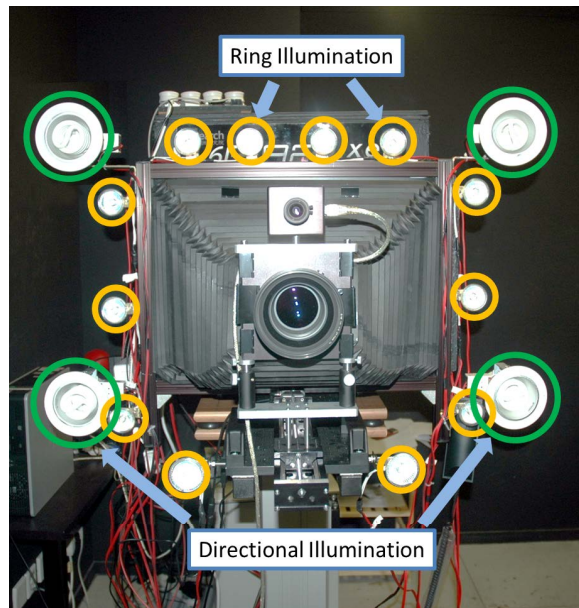


Figure 2.7: Illumination. The camera has a ring of dim-able halogen lights that provides spatial and spectral full balanced and four computer controlled directional lights used primarily for photometric stereo. (Image courtesy of Moshe Ben-Ezra)

2.3 Software

The camera is connected to a PC with the camera operation software that allows the user to perform tasks such as viewing the current scene, selecting parts of the scene for capture, moving the lens along the translation stage for focusing, or adjusting exposure parameters. In this section, the software is briefly described, including main functions and most frequently used user interfaces. For a complete user manual of the software, please see Appendix A.

2.3.1 Main Functions

Image Capture As a tile-scan camera, the scene is captured by moving the sensor along a predefined grid and capturing one tile at a time. These individual tile images are then automatically stitched to form the final output image. To facilitate internal stitching, the software allows the user to specify the percentage of the overlapping region between neighboring tiles. While our camera could capture images over more than one gigapixels that covers almost the whole image plane, there are situations that the user is only interested in a small region of the current scene. The software allows the user to specify the region of interest from the viewfinder. The camera will only capture the image covering that region.

Conventional Capture Parameters The software allows users to specify conventional parameters for image capturing, such as exposure time including bracketing, gain (ISO), white balance types, and so on. Note that due to hardware limitation, the aperture of the main lens can only be set manually. The user can also choose the image file format including both 8 bits and 16 bits formats, and the output folder. The output is organized as a folder which includes all capture parameters, the tile images, calibration images, and the final image.

Manual Focusing and Parameter Adjusting While the software has a viewfinder area to show the entire view seen by the camera (resized to fit the window), the user can view the real time scene seen by the main sensor, at the current location on the image plane. By looking at the real time view from the main sensor, the user could manually focus on the designated object by firstly moving the sensor along the vertical and horizontal image plane stages then moving the lens along

the focusing stage. The software also allows the user to test the focus and exposure settings by taking a snapshot image at the current location.

Calibration and Post-Processing The software supports capturing calibration images and post-processing including: 1) *dark current*¹ *processing* which removes the fixed pattern noise due to bad pixels on the sensor; 2) *demosaicing* which converts the raw format images to RGB format with gamma correction; 3) *white image*² *processing* which corrects dark regions in the input image due to sensor/lens dusts using white images; and 4) *color calibration* which corrects the color of the images using images captured with Macbeth chart.

Illumination While the ring illumination can only be turned on and off using physical switches, the directional illumination is controllable using our software. The software gives options such as *fix mode* and *photometric mode*. Fix mode allows the user to turn on and off certain lights from the four directional lights. Photometric mode turns on one light at a time to allow the camera to take a shot under single directional illumination.

Focal Stacking The basic idea of focal stacking is to capture the scene multiple times while varying the focus of the camera. The multiple images are then composited to produce an all-in-focus image. Our camera operation software can automatically capture focal stack images and produce the final image with large depth-of-field. The user needs to specify the starting focus distance, the number of focus steps, and the step size. During image capture, at each sensor location (on the grid), the lens will move step by step along the focusing stage. At each focus position, the

¹Raw images that are captured when the aperture is completely closed.

²Images that are captured when white card is in front of the camera and defocused.



Figure 2.8: The main window contains a viewfinder on the left and a control panel on the right. Note the dimmed area is the scene that will not be captured by the camera.

camera will take one shot.

2.3.2 User Interface

Main Window The main window of the software, shown in Figure 2.8, contains a viewfinder on the left and a control panel on the right. The viewfinder displays the video from the auxiliary video camera. Through the viewfinder, the user can send the sensor to a certain location, or select a region of interest to be captured. The viewfinder can also show the grid along which the sensor moves. The control panel contains the buttons that can be clicked to activate sub-windows of the software.

Sub-windows include capture control window, calibration window, manual focus window, and etc. These sub-windows provide the main functionalities of the camera such as starting/stopping the camera, adjusting capture parameters, manual focusing, camera calibration, snapshot, capturing image, fast stitching, and focal stacking. In the following paragraphs, two sub-windows that are slightly more complicated than the rest are further described.

Capture Control Window The capture control window allows the user to adjust various parameters used for capturing an image, including: exposure time with bracketing, gains (ISO), manual or auto focus with bracketing, overlapping percentage between each image plane stage movement (default is 25%), illumination methods such as photometric stereo mode and fix mode, white balance, image saving parameters including format and project directory, etc.

Manual Focus Window A separate window displaying real time video from the main sensor is opened when the manual focus window is started. By looking at the real time video, the user can move the lens along the focusing stage by clicking forward and backward buttons on the manual focus window. Similarly the user can move the sensor of the main camera along the horizontal and vertical image plane stages. The user can move the lens and main sensor repeatedly till the target object is in focus.

2.4 My Effort

My primary effort to the camera is to develop its entire operation software that has a user-friendly graphical user interface (GUI). The GUI contains not only the

functions similar to that of conventional DSLR, but also the processing tools that are specifically developed for the camera such as calibration and post-processing, e.g., focal stacking. Significant effort has been expended on developing this operation software. My effort can be generally summarized into three areas: 1) dealing with low level SDK and drivers; 2) deep understanding the camera/sensor mechanisms and the GUI design; and 3) implementing and improving well-known camera calibration and post-processing algorithms. The development of the operation software took approximately 100 man-days to complete and was solely completed by myself when I was a full-time intern in Microsoft Research Asia involved in the dgCam project ³ [22].

³When I joined the project, Dr. Moshe Ben-Ezra has finished designing and building the camera. He also has written a simple software that can move the sensor and capture snapshots.

Chapter 3

Field Study of 2D Imaging with Large-Format Digital Camera

In my view you cannot claim to have
seen something until you have
photographed it.

Emile Zola

In the previous chapter, we described a large-format digital camera that can capture images that are larger than one gigapixel in size. In this chapter, we present recent field work that uses this camera to capture high-resolution images of Buddhist art at the UNESCO world heritage site, the Mogao Caves. This field work is undertaken by Microsoft Research Asia and the Dunhuang Academy. I was the main student who worked on this field work project when I was an intern in Microsoft Research Asia. The project is intended as a feasibility study examining the use of a large-format digital camera to capture high-resolution images in a cultural heritage setting. In particular, we describe on the current challenges faced by the

Dunhuang Academy in their imaging efforts and how the use of a large-format digital camera can improve the quality of the imaging process while reducing time and effort. We also discuss lessons learned from this field study as well as remaining challenges inherent to such projects.

3.1 Introduction and Motivation

The *Mogao Caves*, often referred to as the Dunhuang caves, consist of a network of over 400 various-sized caves housing some of China's oldest Buddhist artwork. The artwork, which spans a 1000 year period, includes statues, religious texts, and wall paintings that cover over $42000m^2$. The Mogao Caves have been a UNESCO world heritage site since 1987 [92] and are managed by the Dunhuang Academy which was established in 1943 by the Chinese government. As part of the many duties of the Dunhuang Academy, imaging the caves for purposes of preservation, studying, or virtual tourism [57] has been on going since 1999. In the last two years (2010 and 2011), this imaging task has turned its focus to high-resolution digital imaging, with the long term goal of capturing images of each cave's artwork at full $300ppi$ resolution, the recommended standard for archival purposes. To date, frescos in only 47 caves have been imaged at only $75ppi$. This imaging task has proved so labor intensive that the Dunhuang Academy is now considering alternative options over its current procedure involving small-format cameras. The following discusses the physical environment, the current imaging procedure, and potential benefits of large-format imaging.

Environment Photographing the caves at a minimum resolution of $300ppi$ presents

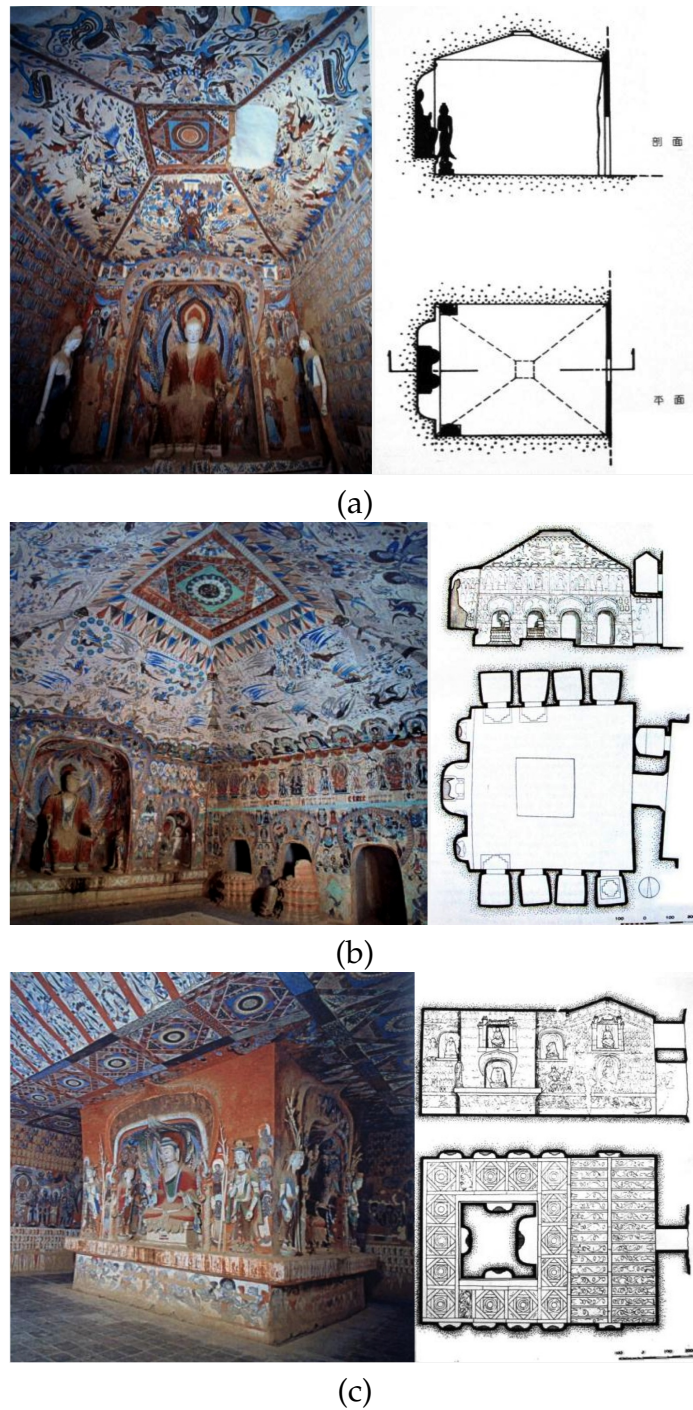


Figure 3.1: Dunhuang cave structure: (a) small cave example; (b) medium cave example with rooms; and (c) large cave example with central pillar.

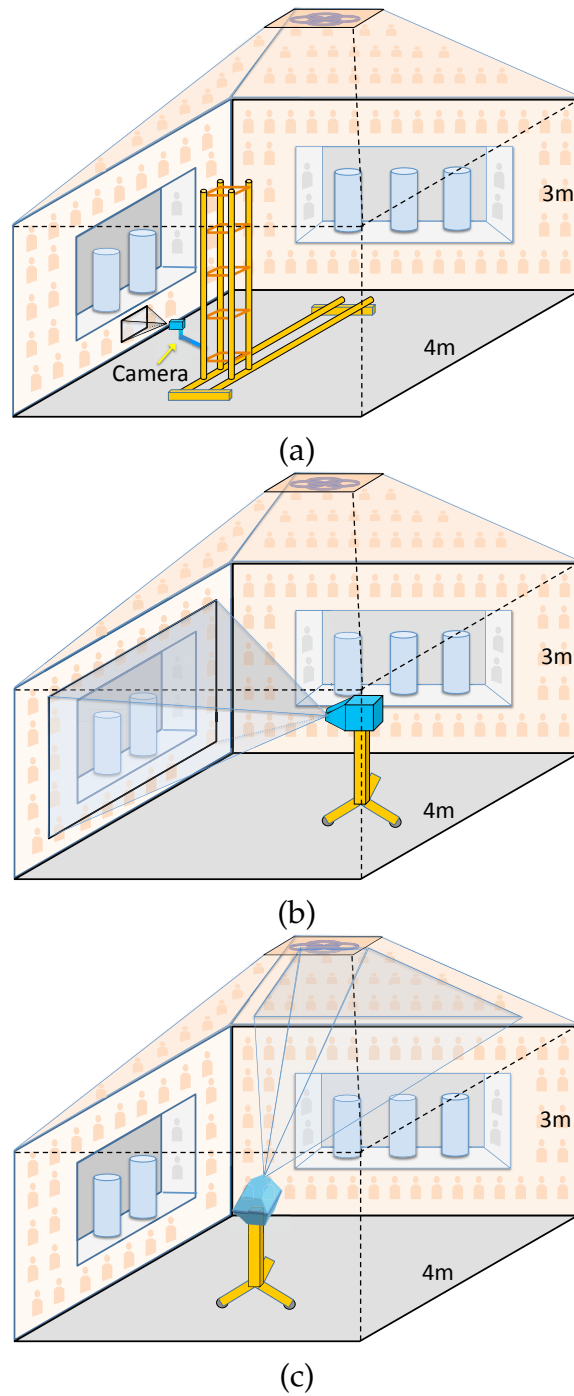


Figure 3.2: Dunhuang current and proposed imaging method: (a) current imaging method, (b-c) proposed method for wall, niche and ceiling imaging using our large-format digital camera.

a unique challenge due to large variations in cave size, structure, and sometimes difficult access. Figure 3.1(a-c) shows three typical layouts of the Dunhuang caves. A relatively small cave, shown in (a), includes frescos on the walls as well as on the pyramidal ceiling. In addition, a few statues are located in niches in the wall. Special interest is given to the painted center of the ceiling known as the *caisson*. A medium size cave, shown in (b), can have attached small rooms that often house statues. Larger caves like the example shown in (c) often have one or more central pillars, with niches and statues. In most of the caves, all areas of the walls, ceilings, and niches are covered by frescos. Given the complexity of imaging 3D objects at 300ppi, the Dunhuang Academy has focused its efforts instead on imaging the frescos and other paintings in the grottos. However, even for these near-planar surfaces, acquiring high-resolution images is challenging.

Current Approach and Limitations The current approach is a result of Mellon International Dunhuang Archive project [60], a recently completed collaboration between the Dunhuang Academy and Northwestern University. Figure 3.2(a) shows the current procedure used, where a rail and a tower structure are erected in the cave and a small-format camera, typically a DSLR camera, is attached to the tower and placed close to the wall (closer than in the illustration). A sequence of images is then taken while manually moving the camera. To obtain the final mosaiced image over an area of several square meters at 75ppi resolution requires taking an enormous number of overlapping images using the small-format camera. To put this in perspective, the Dunhuang Academy reported that it requires 812 images using a 10megapixel DSLR camera to obtain a 75ppi resolution image of a 5148mm × 3329mm fresco located on the north wall of Mogao Cave #220. In addition, the camera needs

to be carefully translated to facilitate the imaging process. The total capture time for the wall is about four to five days excluding time for hardware setup.

After image capture, the images are stitched together to form a single mosaic image. While image mosaicing is often considered a solved problem in computer vision and computer graphics, it is not necessarily the case when imaging artwork. Current image mosaicing techniques [61, 62, 15, 16, 17, 85, 86] use homography-based image alignment which only works in the ideal case of a purely planar surface or very distance scenes. Cave walls, however, are not perfectly planar and this causes errors in the final mosaic. The strategy taken by almost all image mosaicing techniques is to hide these errors through image blending and/or seam-cutting. Perceptually masking errors is not an acceptable solution for the Dunhuang Academy. Instead, they use image-editing software, i.e., Photoshop [1], to get an initial alignment. This is typically done using the mosaicing feature in Photoshop. Photoshop, however, uses seam-cutting and color adjustments in its processing (applied as masks to the image layers). The Dunhuang Academy turns off these features and uses only the estimated planar warps as a starting point. The images are then manually adjusted using local warping functionality in Photoshop to compensate for the non-planar misalignment. In some case, Photoshop is unable to provide even the initial alignment. In these situations, the Dunhuang Academy is forced to align all images by hand using Photoshop local warping tools. As per the previous example for Mogao Cave #220, it took approximately 54 man-days to stitch all 812 images. In fact, this mosaicing procedure is currently the major bottle neck in the Dunhuang academy's imaging pipeline, with 50 caves worth of images backlogged for processing.

Large-Format Camera Imaging In contrast to the previously described procedure, a well designed large-format camera can capture a 10 square meter area at 300ppi resolution without moving the camera or lens. This allows capturing of statues and reliefs in a single image as shown in Figure 3.2(b). The camera can also help capture the ceiling by only tilting the camera as shown in Figure 3.2(c). The Dunhuang Academy was interested in exploring the use of a large-format digital camera as it could significantly reduce overall time and labor as well increasing imaging resolution. A large-format digital camera cannot completely solve the imaging problems in a place as diverse as Dunhuang caves, as some caves and statues that are simply too large, and in some the space is too restricted. This would still require manual editing to stitch together the large-format images. However, its benefits were enticing enough to perform a initial study in the Dunhuang caves.

The remainder of this chapter describes the results of our initial field deployment in Section 3.2. A discussion and summary conclude the chapter in Section 3.3.

3.2 First Field Deployment

We shipped the large-format digital camera described in Chapter 2 to the Mogao Caves and carried out the first field deployment in March 2010 (see Figure 3.3). This section describes the image capture and post-processing of this field work. We also show several results obtained from our field deployment.

Among the hundreds of caves, Mogao Cave #46, #418, and #420, were selected for our field work with the following considerations: the size of our camera, the size of and complexity of the cave, and the artistic value of artifacts inside the cave. Cave #46 is a small size cave whose walls contain both statues and frescos. Cave



Figure 3.3: Our high-resolution large-format digital camera in a cultural heritage project in Mogao Cave #46. The frescos and statues in the cave are from the Tang Dynasty (618 - 907 AD). (Image courtesy of the Dunhuang Academy)

#420 contains frescos that have significant 3D texture, while Cave #418 has frescos that consist of several layers of paint from different dynasties.

In the three caves, high-resolution 2D images of the frescos together with statues were captured using our camera. All the images were captured in 16bits raw format under proper exposure. When the required depth-of-field of the scene exceeded the lens depth-of-field, focal stacking was used to extend the depth-of-field.

Before each scene was captured, we first imaged a dark current¹, a white image², and a Macbeth color chart for post-processing. When images were captured with

¹Raw images that are captured when the aperture is completely closed.

²Images that are captured when white card is in front of the camera and defocused.

focal stacking, the method described in [11] was used to extend the depth-of-field. In total, two complete walls containing both fresco and statues, and portions of frescos from the three caves were captured.

3.2.1 Results

In Figure 3.4 and Figure 3.5, we show two examples from images captured in our field study. Figure 3.4 captures part of the north wall in Cave #46. The walls and statues captured are from the Tang Dynasty (618 - 907 AD). This example required focal stacking with five images of varying focal distance. The size of the image is 45320×32645 (1.4 *gigapixel*) with 415 *ppi*. The total capture time was 124 minutes including the time to capture focal stacking. Figure 3.4(a-b) show the overall image and zoomed region of the image and Figure 3.4(c-f) show further zoomed regions. Note that Figure 3.4(e) shows a close look of the slanted wall that requires extended depth-of-field.

Figure 3.5 shows part of the frescos on the east wall in Cave #418. The size of the image is 37392×31718 (1.1 *gigapixel*). Although the fresco appears to be flat, focal stacking was still required. The total capture time was 77 minutes including the time for focal stacking. Figure 3.5(a) shows the full image. Figure 3.5(b-h) show regions at different zoom levels. Note that this image shows frescos with two layers. While Figure 3.5(b) shows a Buddha from the Western Xia Dynasty (1038 - 1227 AD), Figure 3.5(c-h) not only show fresco fragments from the Western Xia Dynasty, but also reveal a Buddha from the Sui Dynasty (518 - 618 AD).

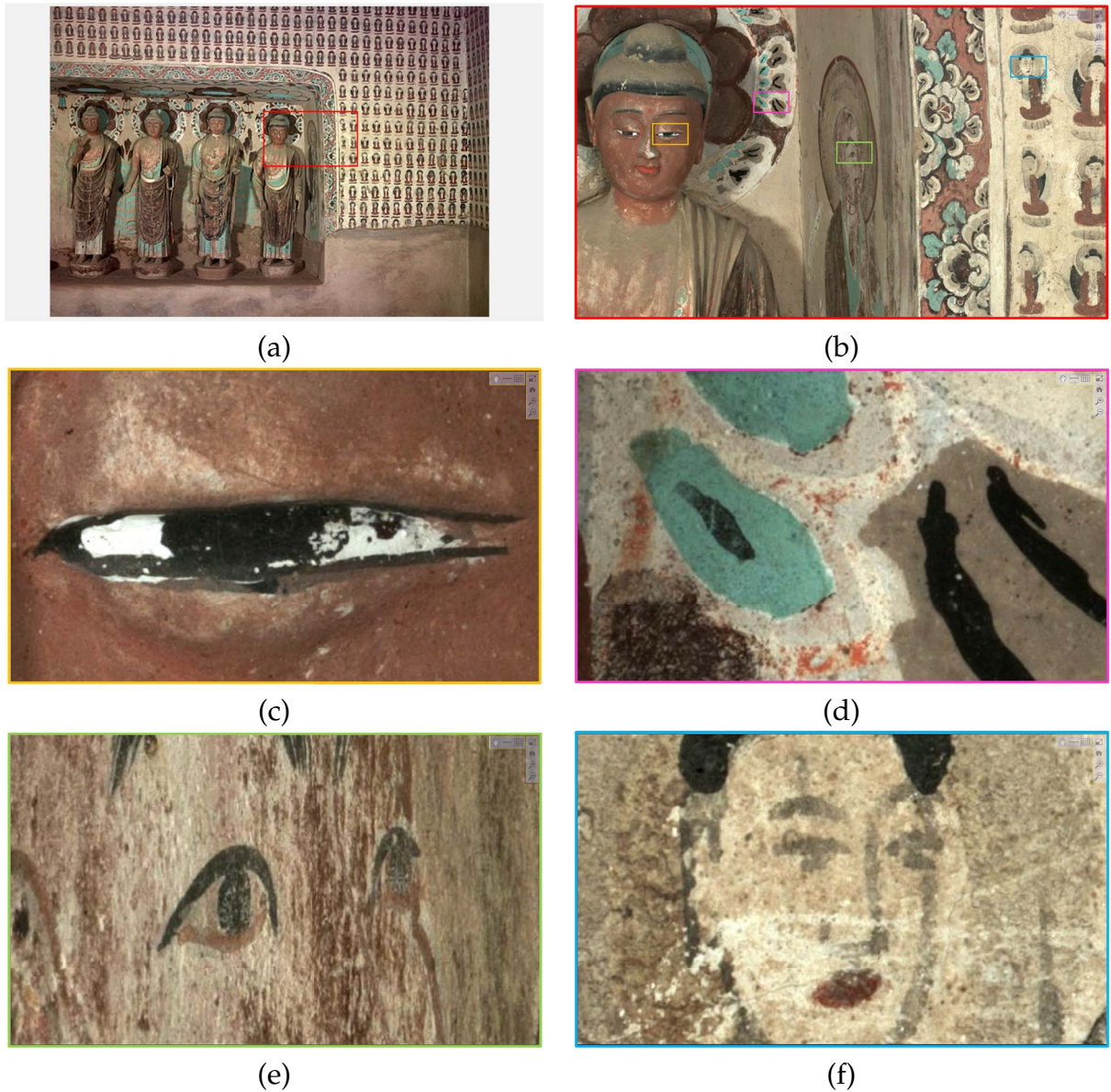


Figure 3.4: An image from part of the north wall of Mogao Cave #46: (a) the final image, (b) zoomed region of the image, (c-f) further zoomed regions. Note that (e) shows a close look of a wall in the niche with large depth-of-field. The walls and statues captured are from the Tang Dynasty (618 - 907 AD).

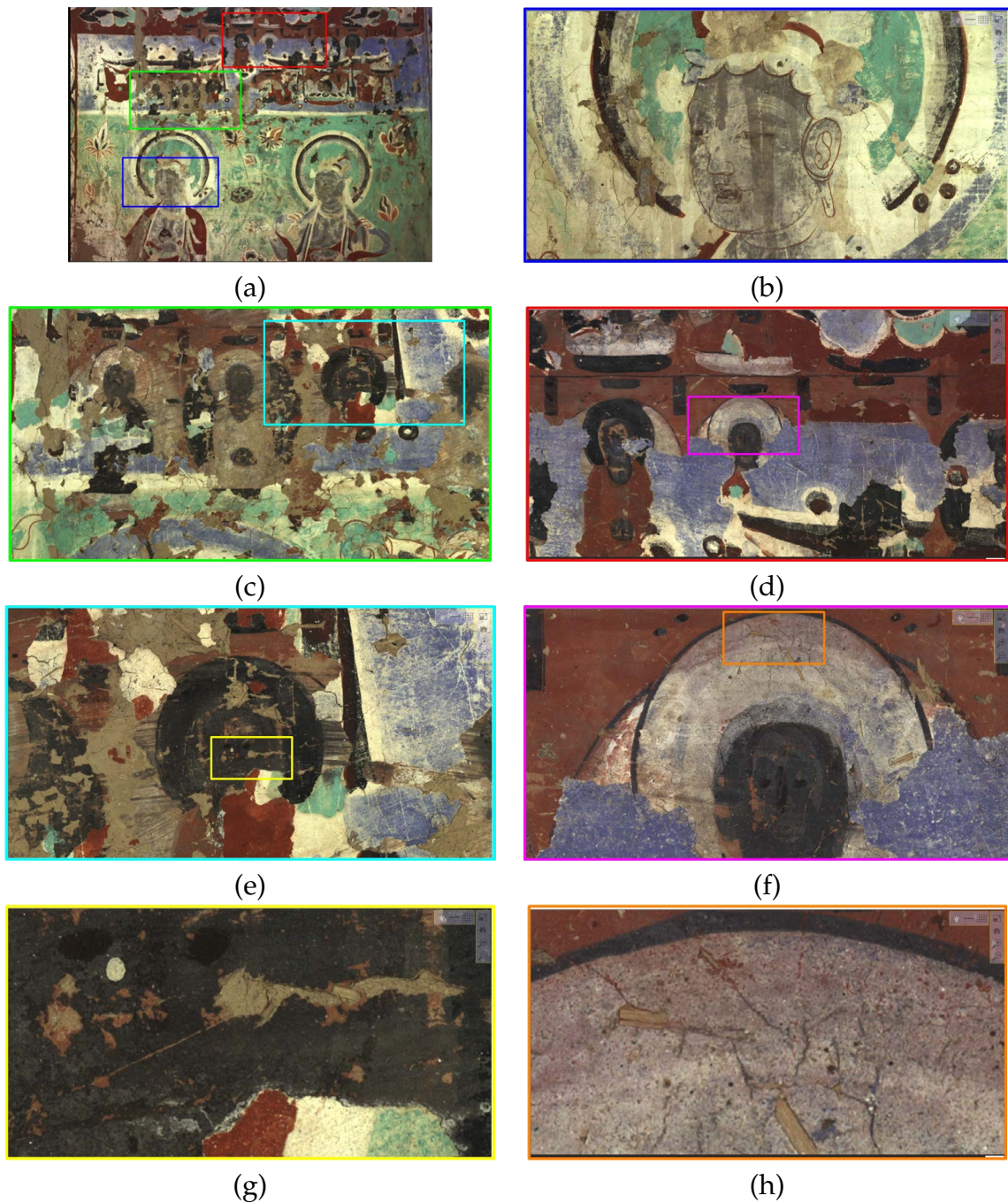


Figure 3.5: An image from part of the east wall of Mogao Cave #418: (a) the final image, (b-d) zoomed regions of the image, (e-f) further zoomed regions of (c-d), (g-h) further zoomed regions of (e-f). This image captures frescos with two layers. While (b) shows a Buddha from the Western Xia Dynasty (1038 - 1227 AD), (c-h) not only show fresco fragments from the Western Xia Dynasty, but also reveal a Buddha from the Sui Dynasty (518 - 618 AD).

3.3 Discussion and Summary

3.3.1 Lessons Learned

Several valuable lessons learned from this field work, from both a technical and cultural heritage perspective, are summarized in the following.

Dust Unlike a lab setting, imaging inside a cave involves in a great deal of dust. Together with bad sensor pixels, dust on the lens/sensor become artifacts in the final image. Since our large-format digital camera design uses a translating sensor, dust on the sensor creates artifacts that form repeated patterns in the final stitched images. To deal with this problem we needed to capture a dark current and white image, before each imaging session, to help detect both bad sensor pixels and dust.

Focal Stacking While we have already mentioned the use of focal stacking to extend the depth-of-field of our camera, we were surprised how often this was necessary. Focal stacking was even necessary for capturing frescos that appear flat but had depth variation. Apparently this is also an issue for the small-format cameras. From the feedback of the Dunhuang Academy, our camera's ability of automatically capture and process the focal stack is very valuable for their work.

Color Management Proper handling of color is absolutely essential for imaging heritage site such as the Mogao Caves. Color needs to be carefully managed before and after capturing. For a large-format camera, one capture covers a large area of the scene. Hence we need to place several high-quality lights in the cave to ensure proper illumination. Devices such as hand-held light meters are used to ensure the scene is as uniformly illuminated as possible. For each capture, we

imaged a Macbeth chart under the same illumination for color calibration in the post-processing step. Furthermore, all the output devices need to be calibrated too (e.g., monitors, printers, etc.). Both technical staff and professional achievers/artists need to work together to fine-tune the color if required.

Remaining Challenges for Mosaicing Our camera does not completely remove the need for image mosaicing. For large scene, multiple images from our camera will be required. The problem with non-planar scene remains, however, our large-format digital camera does require less amount of manual stitching, since large portions of the image are correct. This points to the direction for more effective mosaicing algorithms that are capable of overcoming 3D surface variations in a proper manner, versus hiding artifacts using seam cutting and blending.

3.3.2 Summary

Following the success of our field study, we initiated design and assembly of a second generation large-format digital camera. The new camera, while still over a gigapixel, is slightly smaller and better protected from dust. The new design will use a 300mm lens instead of 480mm to cover wider area with slightly lower resolution. The smaller design will also allow operation in a tighter space.

In conclusion, this chapter described recent field work for imaging Buddhist art at the UNESCO world heritage site, the Mogao Caves, in Dunhuang, China. We outlined current challenges faced by the Dunhuang Academy using conventional mosaicing techniques with a small-format camera. The need for human intervention in the mosaicing step was the major bottle neck in the Dunhuang Academy imaging. Our preliminary results showed that using a large-format digital camera

CHAPTER 3. Field Study of 2D Imaging with Large-Format Digital Camera

not only improves the imaging resolution, but also significantly reduces time and effort. We concluded the chapter with several lessons learned from our field work.

Chapter 4

High-Resolution 3D Imaging

Wherever there is light, one can
photograph.

Alfred Stieglitz

In the previous chapter, we presented recent field work that uses a large-format digital camera to capture high-resolution 2D images in a cultural heritage setting. We have shown the current challenges faced by practitioners, the preliminary result using our approach as well as the lessons learned from the field work. In this chapter, instead of 2D, we present a 3D imaging framework to acquire surface scans at high-resolutions (exceeding 600 samples per mm^2). Our approach couples a standard structured-light setup and photometric stereo using a large-format digital camera. While previous applications have employed similar hybrid imaging systems to fuse positional data with surface normals, what is unique to our approach is the significant asymmetry in the resolution between the low-resolution geometry and the high-resolution surface normals. To deal with these resolu-

tion differences, we propose a multi-resolution surface reconstruction scheme that propagates the low-resolution geometric constraints through different frequency bands while gradually fusing in the high-resolution photometric stereo data. In addition, to deal with the high-resolution images, our surface reconstruction is performed in a patch-wise fashion and additional boundary constraints are used to ensure patch coherence. Based on this multi-resolution reconstruction scheme, our imaging framework can produce 3D scans that show exceptionally detailed 3D surfaces far exceeding existing technologies.

4.1 Introduction

As one of the purposes of imaging for e-heritage is to preserve cultural heritages in digital format, it is important and always desirable to acquire high-resolution data. While high-resolution 2D images reveal very fine surface details of the target object as shown in Chapter 3, high-resolution 3D imaging allows such details to be viewed from different angles and under different lighting direction. In this chapter, we propose a high-resolution 3D imaging framework that targets surface sampling at more than 600 samples per mm^2 . To our best knowledge, this is the highest sampling rate demonstrated to date.

The fundamental problem for 3D imaging at high-resolution is that while it is possible to obtain very dense samples of the surface using a high-resolution digital sensor, it is difficult to perform structured-light at these resolutions. Virtually all projector lenses are designed to magnify the projected screen and therefore cannot produce very dense samples on an object's surface. While lasers can be focused on an object at much finer resolution, lasers used in 3D scanners rarely can produce an

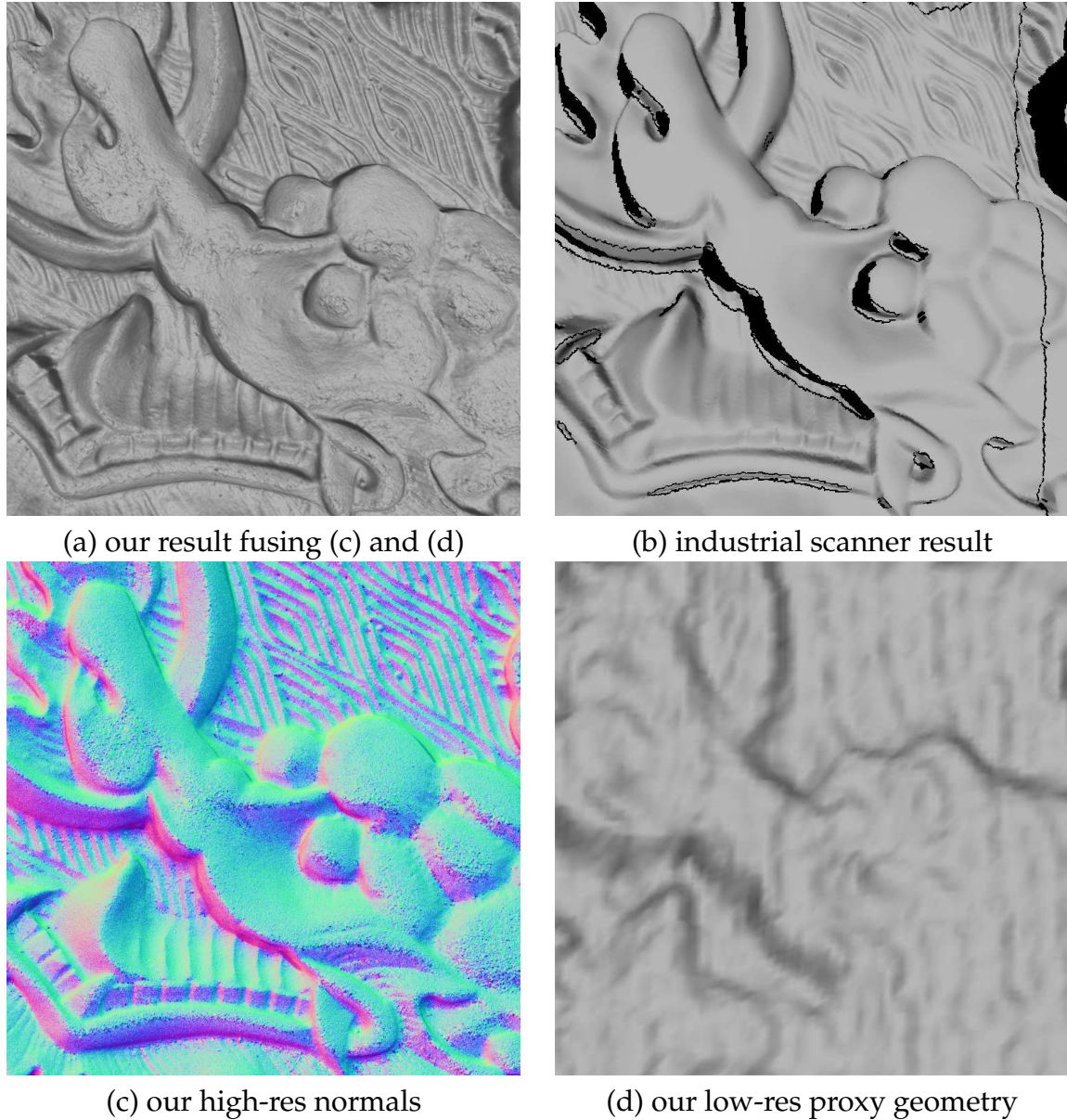


Figure 4.1: Comparison of our results against a state-of-the-art industrial 3D scanner: (a) our reconstructed surface obtained by combining (c) and (d); (b) surface reconstructed by a Konica Minolta Range 7 (168 samples per mm^2); (c) our input from photometric stereo of (over 600 samples per mm^2); (d) our low-resolution geometry (6.25 samples per mm^2) obtained via structured-light.

illuminated point less than 10 microns at very close ranges (i.e. a few centimeters away from the object) to only several hundred microns at ranges $50\text{cm} - 100\text{cm}$ away. Thus, high-resolution laser scanning at these resolutions is limited to small areas that must be stitched together.

To appreciate this difference, Figure 4.1 shows a 3D surface scanned using our high-resolution imaging and the same object scanned by a high-end commercial 3D scanner designed for industrial inspection (Konica Minolta Range 7 [48]). Our 3D surface is scanned at more $3\times$ the reported surface resolution of the commercial scanner. The commercial scanner also had to acquire multiple scans that are stitched together (using Konica’s software) because the object is bigger than its effective area.

To overcome the limitations of structured-light, hybrid imaging approaches that combine positional data from structured-light together with fine surface normals captured from photometric stereo have been proposed (e.g., [12, 66, 93]). We adopt this hybrid imaging approach, however unique to our framework is the significant asymmetry in the resolutions between the two inputs. Existing work dealt with sample differences of $4\times$ resolution between the surface normals and low-resolution geometry ([12]), while our photometric surface detail is scanned at around $100\times$ the resolution of the low-resolution geometry (see Figure 4.1(a) and (b)). This high-resolution and significant resolution difference in the hybrid framework create difficulties not yet encountered by previous approaches.

To address this significant difference in the resolution between the surface normals and low-resolution geometry, we propose a multi-resolution surface reconstruction scheme that fuses the low-resolution geometric with the photometric stereo data at increasing levels of details. To deal with the large data from the high-resolution input, we adopt a patch-based scheme that uses additional bound-

ary constraints to maintain patch coherence at the boundaries. The results of our approach are 3D images of surface captured at an exceptionally high level of detail.

The remainder of this chapter is organized as follows: Section 4.2 discusses related work; Section 4.3 describes our system setup to capture high-resolution images for normal estimation, and our structured-light system to capture surface geometry; Section 4.4 presents our main algorithm for surface reconstruction; Section 4.5 presents our results. A summary of this work is presented in Section 4.6.

4.2 Related Work

There is a vast amount of literature on 3D imaging. Readers are directed to [81] and [97] for broad overviews; here only representative examples are mentioned.

3D imaging has been approached using passive triangulation methods such as conventional stereo (e.g., [78]), passive photometric methods such as shape from shading (e.g., [39]), active triangulation methods such as structured-light (e.g., [79]) and active photometric methods such as photometric stereo (e.g., [99]). Hybrid methods that integrate two or more methods include approaches that combine shape from motion and photometric stereo (e.g., [38]), positional (3D points) data and normals (e.g., [89, 6, 27, 49, 42, 12, 18, 66]), visual hull and normals [35]¹, and recently normals and volume carving [93].

All of the previously mentioned hybrid methods have the potential to be adapted to handle very high-resolution imagery as in our application. We decided on a solution that is closely related to that presented by Nehab et al. [66]. The system presented in [66] used two cameras in a structured-light setup with one of

¹Also see <http://carlos-hernandez.org/gallery/>.

the cameras also used for the photometric stereo. Positional and surface normals were fused using a linear formulation that resulted in a sparse-linear system. In their work, the surface geometry and the photometric data were of the nearly the same resolution. In our work, we have $100\times$ more estimated surfaces normals than we do 3D points. At our resolutions (e.g., $5K\times 5K$ surface normals), the sparse matrix proposed [66] would have approximately 150 million entries. Solving such a large matrix is not straight forward, even for out-of-core linear solvers such as [76]. As such, we adopt a patch-wise strategy to the fusion process. In addition, to deal with the significant difference in resolutions, we use a multi-resolution pyramid approach to adaptively incorporate the geometric constraint from the low-resolution geometry during the surface integration.

4.3 System Setup

Our hybrid system uses a large-format digital camera together with four controllable lights. This is combined with a separate structured-light rig composed of a low-res camera and projector. Figure 4.2(a) shows our setup. The four light sources are calibrated using a mirrored sphere. We placed polarized filters in front of the lights, together with a polarized filter on the high-resolution camera to reduce light scattering. The two cameras and projectors are calibrated by a physical calibration pattern. The following gives more details to the two main components.

High-Resolution Photometric Stereo High-density photometric stereo images can be captured using a conventional DSLR camera equipped with a macro lens. Due to the zoom factor of the lens, this approach will only be able to scan very small

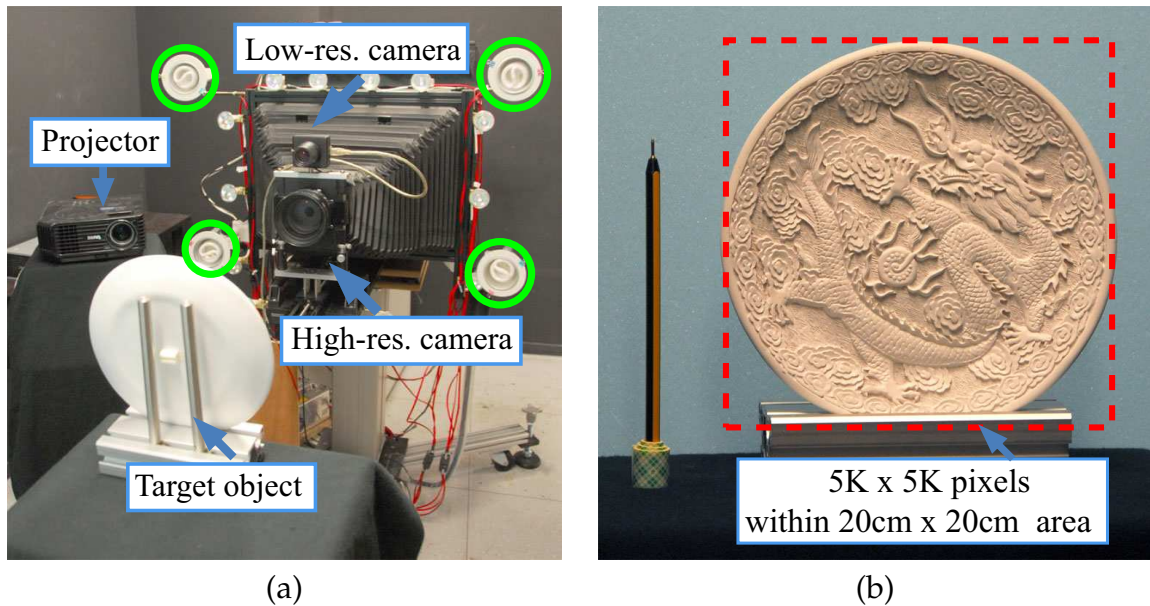


Figure 4.2: Our 3D imaging setup: (a) Our experimental setup consists of a high-resolution camera with four lights used photometric stereo. A low-resolution video camera and digital light projector form the structured-light system. (b) shows the effective resolution about one of our objects. Note the scale of the physical object, versus the pixel resolution. This results in a pixel resolution of over 600 samples per mm^2 .

surface patches. While it is possible to integrate several surface patches together, we have opted to use a large-format digital camera. Commercial large-format digital cameras up to 340megapixel exist on the market [4], however, in this work we use the gigapixel large-format digital camera described in Chapter 2.

To estimate normals, we use the high-resolution camera to capture four images of the surface with varying illumination from the four light sources. The photometric stereo technique in [99] is used to obtain the surface normals. Figure 4.2(b) shows the scanning setup for the *dragon plate* used as a running example in this chapter. The plate has a diameter of 20cm. These images of the object are at a resolution of $5K \times 5K$ pixels.

Structured-Light Geometry Our structured-light system consists of a Benq MP624 projector and a 1024×768 video camera. Standard binary gray-code patterns [79] is used to estimate the low-resolution geometry. In our setup (see Figure 4.2(b)), we get approximately 6 samples per mm^2 from the structured-light scanner, this resolution varies based on the projector’s distance to the object but is representative of a typical structured-light system. Figure 4.1(d) shows a small example of the 3D surface geometry estimated using our method. There are slight pixelization-like artifacts due to inaccuracies in estimating the projected patterns’ boundaries, however, since the low-resolution geometry serves only as a soft constraint in the surface reconstruction process our approach is insensitive to these errors. Because our high-resolution tile-scan camera requires roughly a minute to capture an image, we opted to use an auxiliary video camera to perform the structured-light procedure instead of the large-format camera itself.

4.4 Surface Reconstruction Algorithm

This section describes the surface reconstruction algorithm. The basic algorithm to reconstruct a surface from normals is described first. This is followed by a description on how to include the low-resolution geometry constraint and boundary connectivity constraint into the algorithm. Finally, the steps of the multi-resolution strategy is detailed.

4.4.1 Surface from Normals

Given a dense set of normals the goal is to reconstruct a surface that satisfies the normals’ orientation constraints. We use the recent approach presented by Wu *et*

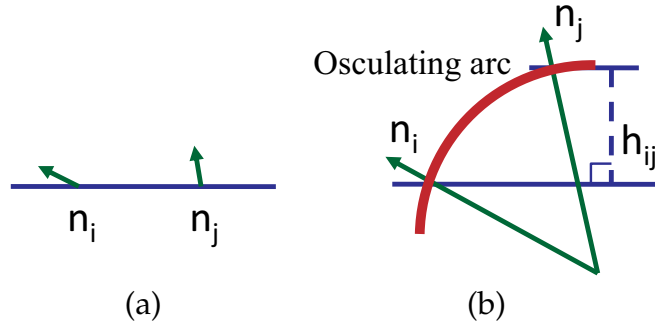


Figure 4.3: The osculating arc constraint [101] for surface reconstruction. Given the normal configuration $\{n_i, n_j\}$ between neighborhood pixel i and j in (a), we can uniquely define the relative height h_{ij} in (b) by using an osculating arc to connect n_i and n_j with minimum curvature.

al. [101] for obtaining a surface from normals that constrains the estimated surface using an osculating arc between neighboring normals (see Figure 4.3). This problem can be cast as a least-square problem that minimizes the following energy function:

$$E(S|\vec{n}) = \sum_i^N \sum_{j \in \mathcal{N}(i)} ((S_i - S_j) - h_{ij})^2 \quad (4.1)$$

where S is the surface we want to reconstruct, $(S_i - S_j)$ is the first order derivative of S in discrete form, h_{ij} is the relative height defined by the osculating arc constraint between neighborhood pixels, $\mathcal{N}(i)$ is the first order neighborhood of a pixel, and N is number of pixels. The main advantage of the method proposed by Wu *et al.* [101] is its ability to globally distribute the normal reconstruction error to the low frequency component. This is especially well suited to our framework which uses low-resolution geometry constraint to improve the global shape of the surface. A qualitative comparison of the osculating arc constraint with other surface from normals algorithms can be found in [101].

Equation (4.1) can be solved using Gauss-Seidel iteration. At each iteration, the

surface height is updated according to the following equations:

$$\begin{aligned} S_i^{t+1} &= S_i^t + \lambda_1 \xi_1, \\ \xi_1 &= \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} (h_{ij} - (S_i^t - S_j^t)) \end{aligned} \quad (4.2)$$

where $|\mathcal{N}(i)|$ is the number of neighborhood pixels, $\lambda_1 = 0.9$ is the step size and t is iteration index. Note that h_{ij} is the same for all iterations and can be pre-computed.

4.4.2 Low-Resolution Geometry Constraint

Because photometric stereo inherently captures only local reflection information rather than global structure, many surface from normal reconstruction approaches do not accurately reflect the real surface geometry. As discussed in Section 4.2, one strategy to overcome this is to incorporate positional information in the reconstruction process.

Our low-resolution geometry constraint is modeled using the following equation:

$$E(S|L) = \sum_i^M (|d(h(S_i)) - L_i| - \Delta)^2 \quad (4.3)$$

where L is the low-resolution geometry captured by the structured-light setup, M is the number of pixels in low-resolution geometry proxy, $h(\cdot)$ is a Gaussian convolution process with radius equals to two times the downsample rate, $d(\cdot)$ is a downsample operation to match the high-resolution normals to the low-resolution geometry, and $|\cdot|$ is the L1 norm (absolute value) of the errors. The term Δ is a parameter controlling the amount of depth tolerance for surface details to be reconstructed and refined by the normals.

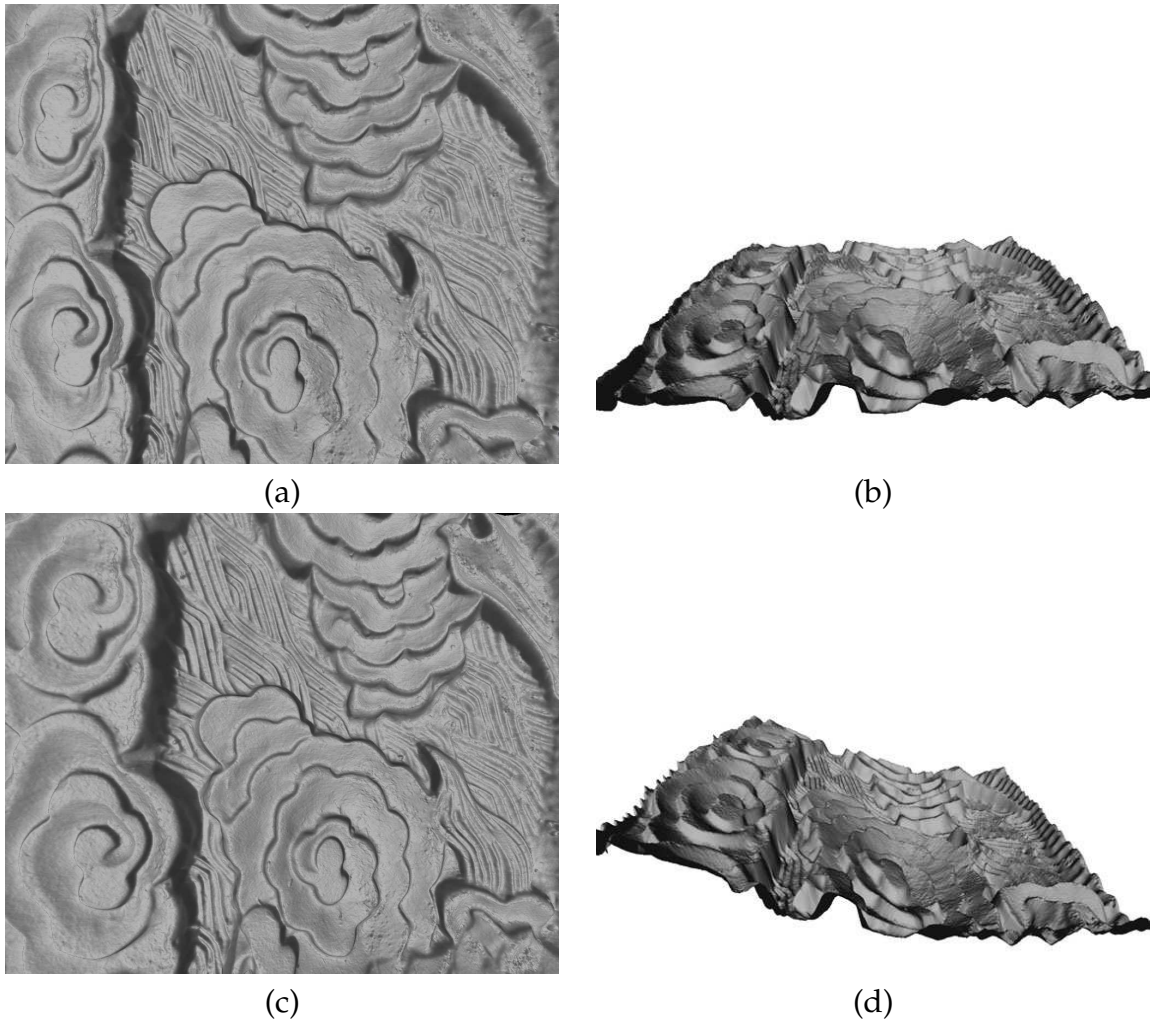


Figure 4.4: Example of surface reconstruction with/without including the geometric constraints: (a) reconstructed surface from normals only; (b) a side view of (a); (c) reconstructed surface with low-resolution geometry constraints; (d) a side view of (c).

With the additional low-resolution geometry constraint, the iterative update equation in Equation (4.2) can be updated as follows:

$$\begin{aligned} S_i^{t+1} &= S_i^t + \lambda_1 \xi_1 + \lambda_2 \xi_2, \\ \xi_2 &= \begin{cases} h(u(L_i - d(h(S_i)))) , & \text{if } |d(h(S_i)) - L_i| > \Delta \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (4.4)$$

where $u(\cdot)$ is an upsample operator. The effect of our low-resolution geometry constraint is shown in Figure 4.4. The value of Δ can be estimated according to the variance of surface details reconstructed from normals and can be spatially varying.

4.4.3 Boundary Connectivity Constraint

As discussed in Section 4.2, the high-resolution of the photometric stereo component makes it challenging to perform integration on the entire surface in one pass. To overcome this the surface can be subdivided into more manageable sized patches which can be reconstructed individually. This leads to a problem that the boundaries of adjacent patches may not be properly aligned after reconstruction. To overcome this, we add a boundary connectivity constraint described by the following equation:

$$E(S|B) = \sum_{i \in \Omega} (S_i - B_i)^2 \quad (4.5)$$

where Ω is the overlapping area of neighborhood surface patch, B is a surface computed by blending the intermediate reconstructed surface in Ω between neighborhood patches using linear feathering. Adding the boundary constraint into Equ-

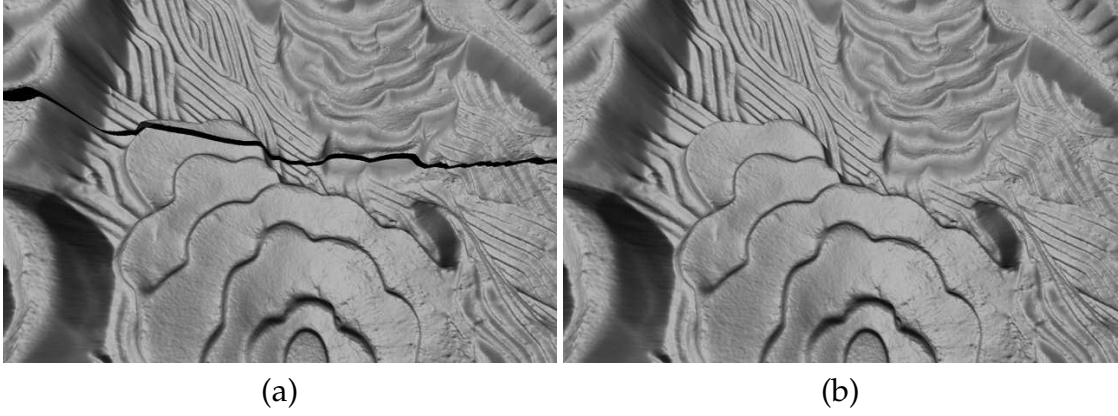


Figure 4.5: Effect of the boundary connectivity constraints: (a) without the boundary constraint, and (b) with boundary connectivity constraint.

tion (4.2), we get:

$$\begin{aligned}
 S_i^{t+1} &= S_i^t + \lambda_1 \xi_1 + \lambda_2 \xi_2 + \lambda_3 \xi_3, \\
 \xi_3 &= \begin{cases} B_i - S_i, & \text{if } i \in \Omega \\ 0, & \text{otherwise} \end{cases} \quad (4.6)
 \end{aligned}$$

For this boundary connectivity constraint, the weight of λ_3 during the update iterations need to be adjusted as the system is iterated. In the initial estimation, λ_3 equals to zero, and its weight is gradually increased as the number of iterations increases. This allows the surface patch to be reconstructed freely at initial iterations and later refined to meet the boundary of neighborhood patches. In our implementation, λ_3 and B are updated at every 100 iterations. With this boundary connectivity constraint, surface reconstructed can be done in parallel and the problem of resolution is no longer an issue. The effect of this boundary constraint is shown in Figure 4.5. For the results in this chapter, surface patches are divided to be of size 1024×1024 with overlaps of 100 pixel boundary overlap (i.e., 10%).

4.4.4 Multi-Resolution Pyramid Approach

Due to the very large differences in resolutions between the surface normals and the low-resolution geometry, directly adding the surface normals to the low-resolution geometry will result in noisy reconstruction as shown in Figure 4.6. To avoid this, our surface reconstruction is done in a multi-resolution pyramid fashion. The main purpose of using the pyramid approach is to correct the low-resolution geometry using normals at the equivalent level before we use it as a soft constraint at a higher resolution. The multi-resolution pyramid approach also allows us to resolve small misalignments between the high-resolution normals and low-resolution geometry due to device calibration errors.

We divide the pyramid uniformly into 5 different levels starting at the resolution used to capture the low-resolution geometry (i.e., 1024×768). For each level, instead of downsampling the estimated high-resolution normals, we downsampled the high-resolution input images and estimate the normals from the downsampled images. This downsampling helps reduce some of the camera's sensor noise when estimating the normals. We run our surface reconstruction algorithm described in Equation (4.6) with the results from previous level as the low-resolution constraint. For the lowest resolution, the low-resolution geometry estimated by structured-light is used. Figure 4.7 shows our intermediate surface reconstruction results (i.e., the evolution) at different levels in the pyramid.

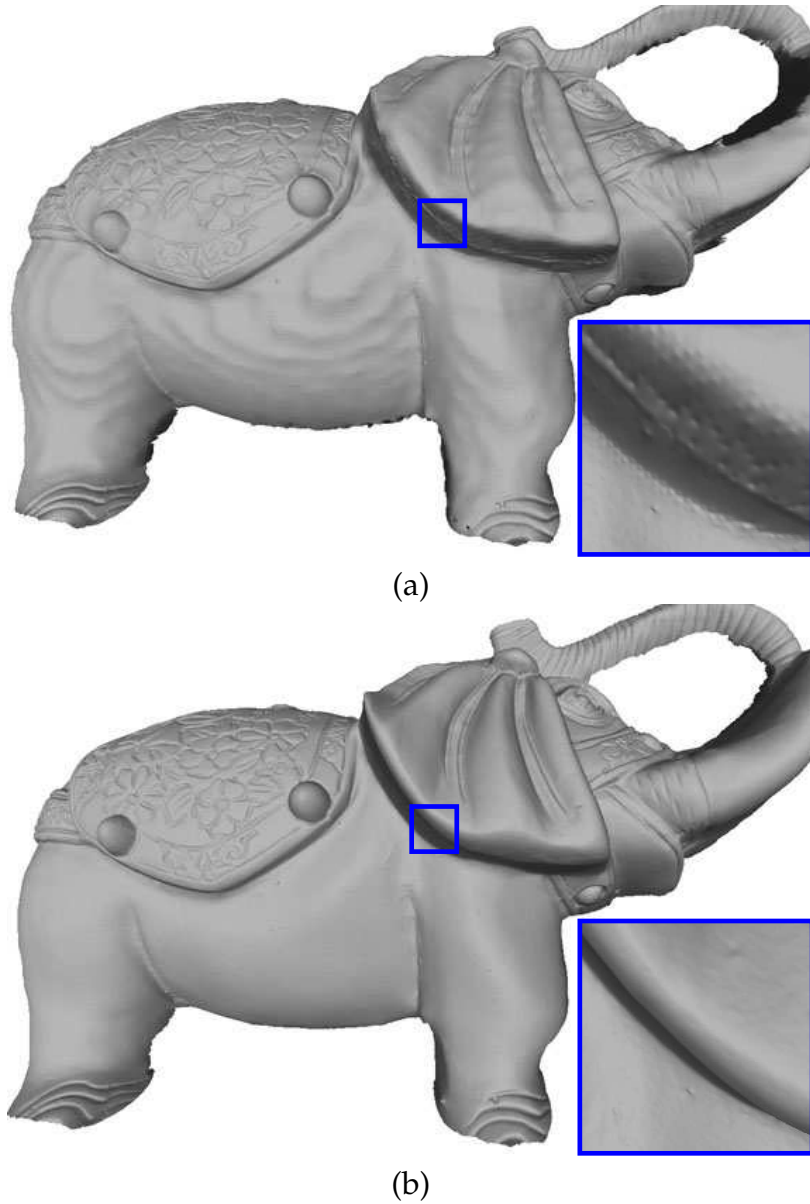


Figure 4.6: Example of the benefits of the multi-resolution scheme: (a) our surface reconstruction directly using the low-resolution geometry; (b) reconstructed surface using the multi-resolution scheme. The 3D surface in (a) shows noticeable quantization errors due to the low-resolution geometry.

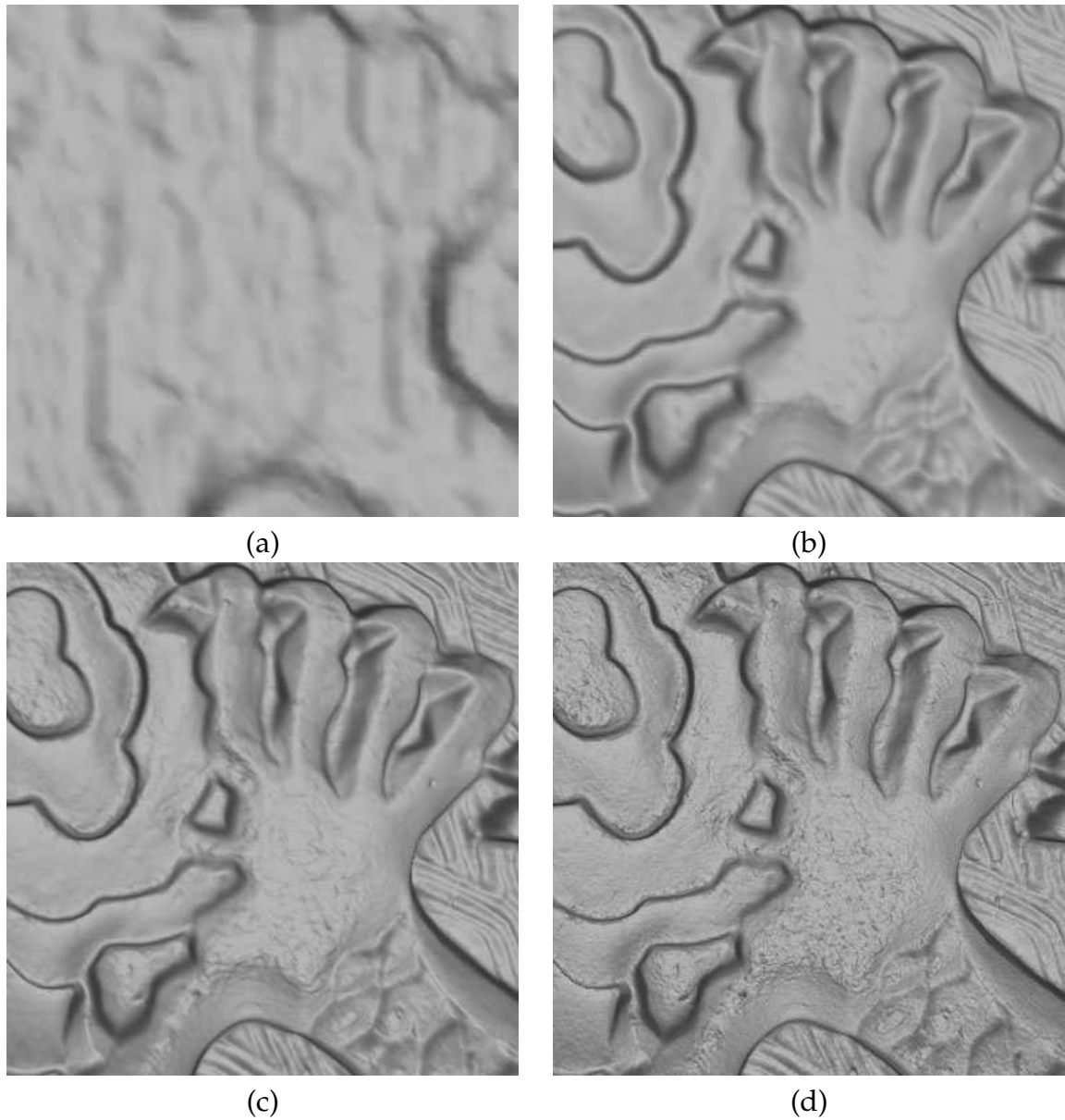


Figure 4.7: Evolution of our 3D surface up the multi-resolution pyramid: (a) low-resolution geometry; (b) intermediate result at the lowest level of the pyramid; (c) the third level; (d) the last level and final 3D reconstruction.

4.5 Results

This section shows several results captured by our system. Each surface was generated using 800 iterations (per patch) of our surface reconstruction algorithm with the boundary constraint applied once after every 100 steps. Because we are working with resolutions beyond the capabilities of existing devices, providing quantitative comparisons proved very challenging. As a result, we can only show visual results for some of our results. For the *dragon plate* we took the object to an industrial scanning facility using a Konica Minolta Range 7. This serves as our baseline comparison against a state-of-the-art industrial laser scanner.

Figure 4.8 shows a 3D reconstruction of an *elephant* figurine which is approximately 15cm wide. Figure 4.9 shows example of a *man* figurine roughly 12cm high. The elephant required 9 patches and resulted in about 6.5 million reconstruct 3D points, while the man required 12 patches and resulted in about 4.5 million reconstructed 3D points. Both of these results show exceptional surface detail. The *man* figurine is further zoomed to reveal detail that would require a magnifying glass to be seen (properly) with the unaided eyes.

Finally, we compare our result a scanned *dragon plate* with that obtained from an industrial standard high end laser scanner (Konica Minolta Range 7) in Figure 4.10. The finest scanning resolution that can be obtained by the laser scanner is 168 samples per mm^2 , while our sampling rate is 600 samples per mm^2 . The plate required 25 patches and resulted in about 21.5 million reconstructed 3D points. The state-of-the-art scanner reports to have a scanning accuracy of 40 microns. We can see that on the double-zoom of these two surfaces, we reveal detail while the result from the laser scanner is almost completely flat. Note that in order to capture

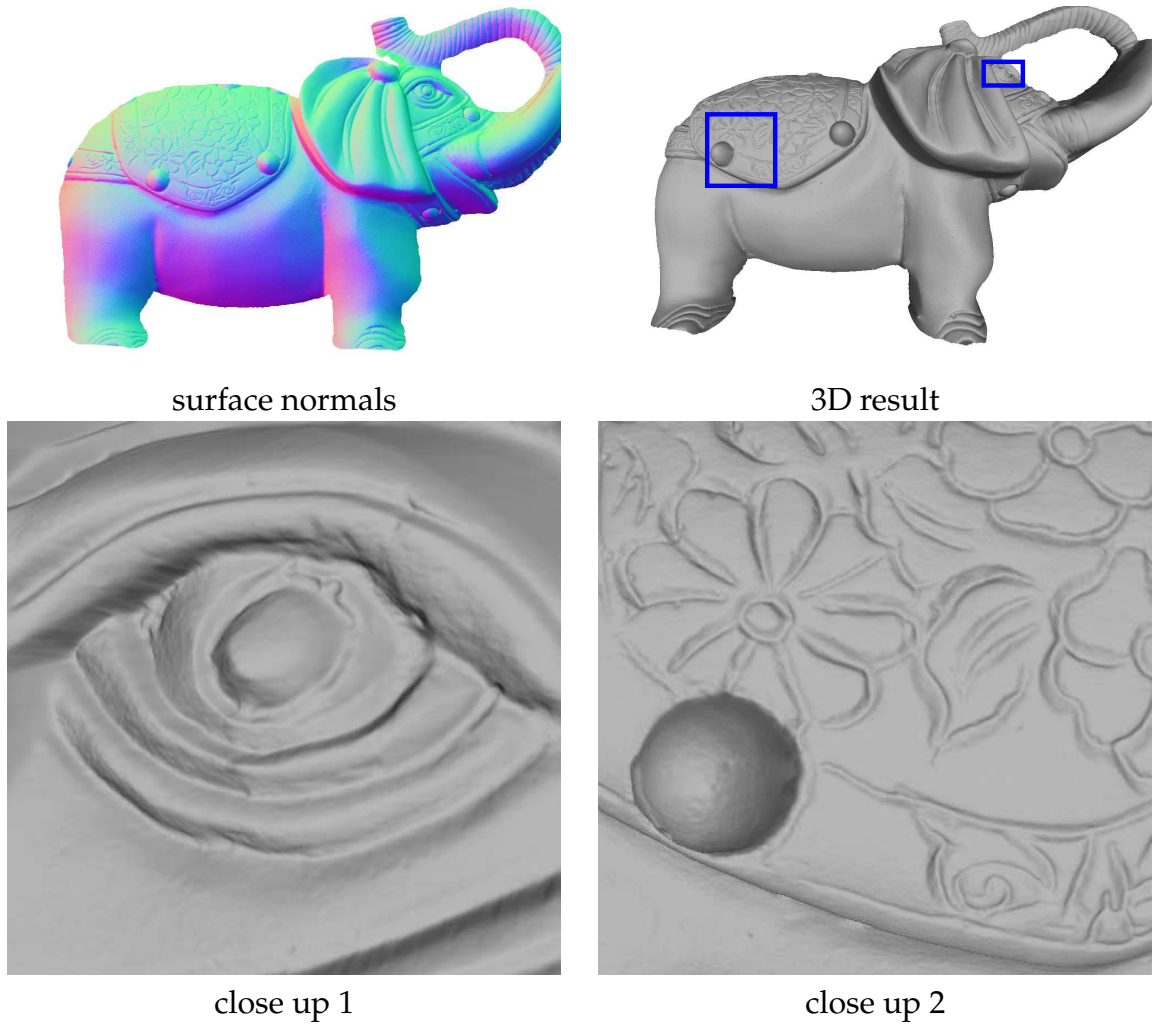


Figure 4.8: 3D reconstruction of the *elephant* figurine. The zooms show exceptional detail on the surface of the object.

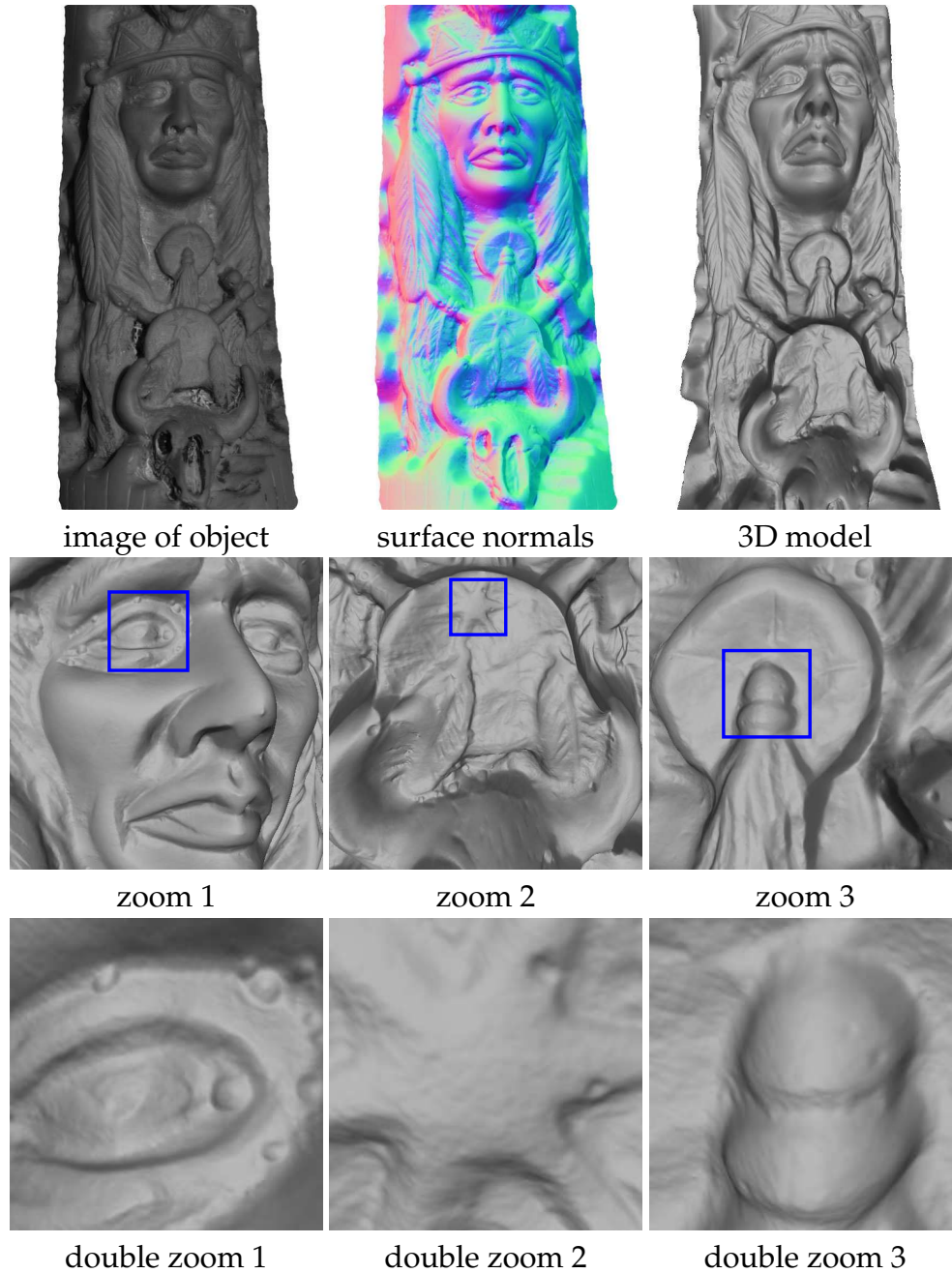


Figure 4.9: 3D reconstruction of the *man* figurine. Due to the high-resolution of the 3D scan, we can show a zoom and “double zoom” of the 3D surface. This double zoom reveals detail that would require a magnifying glass to see properly.

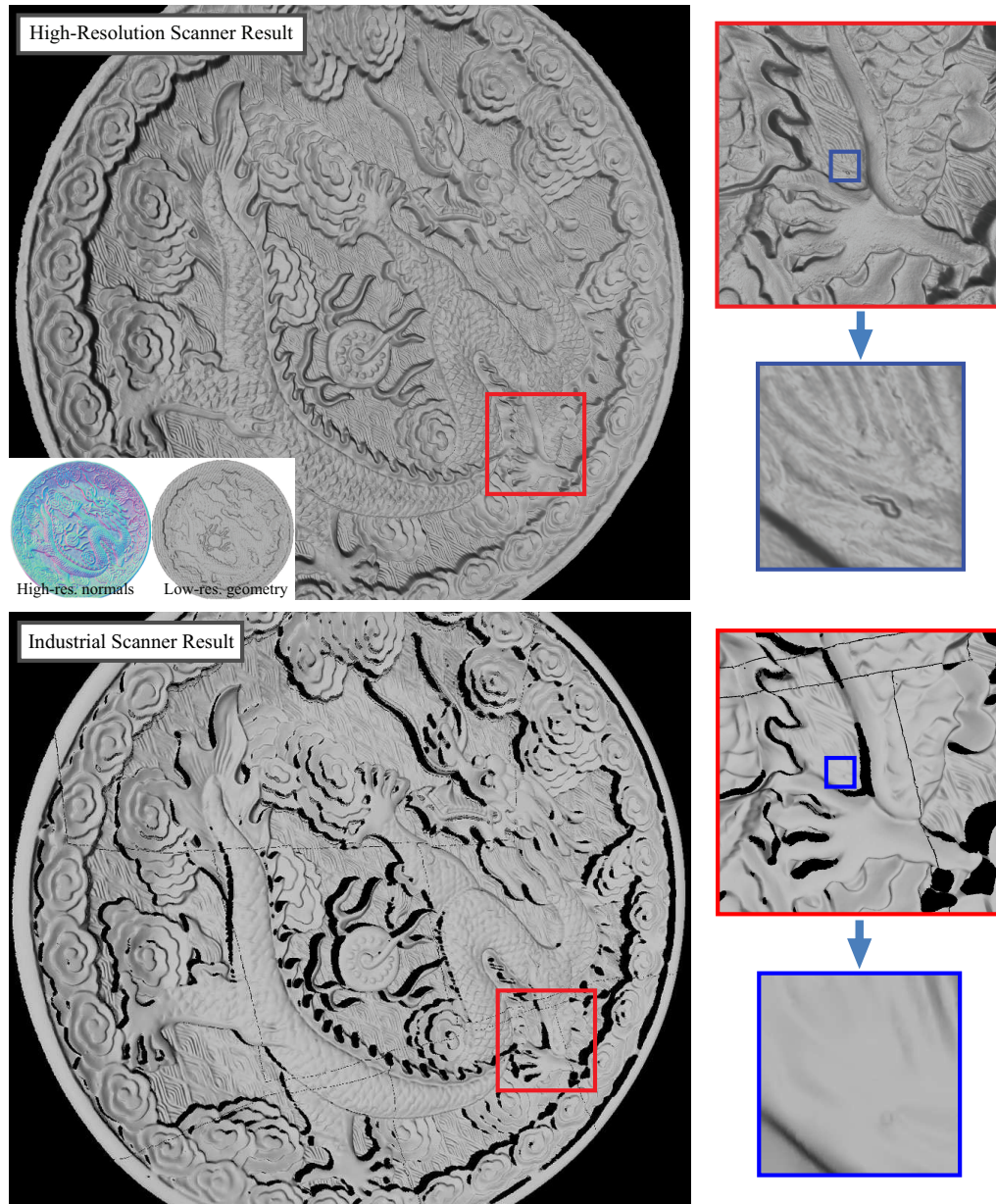


Figure 4.10: Full-size comparison with an industrial laser scanner. Shown are the full 3D reconstruction from our approach and that from a Konica Minolta Range 7 industrial scanner. Insets for our approach show the surface normals and low-resolution geometry. Zoomed and double zoomed regions show that while the two scans reveal that our result contains considerable more surface detail, both appear to reflect the correct geometry. Note that the Konica Minolta Range 7 specification states a scanning accuracy of up to ± 40 microns.

the whole plate by the laser scanner, several scans were performed and stitched together. Our approach, on the other hand, was able to image the entire 3D object in one pass. In addition, even though we use a patch-wise approach in surface integration, our surface does not contain any blocking or pixelation artifacts. This helps demonstrate the effectiveness of our boundary connectivity constraint and multi-resolution pyramid approach. Comparing the surface depth, our estimated surface depths are consistent with surface depths captured by the laser scanner.

4.6 Summary

This chapter presented a framework to capture high-resolution surfaces using a hybrid system that consists of a high-resolution photometric stereo system and a structured-light system. Currently, this strategy appears to be one of the only ways to surpass the resolution of structured-light systems. There are real concerns that had to be addressed when dealing with such massive amounts of data and with significant resolution asymmetry in the respective subcomponents. To address these issues, a multi-resolution pyramid approach was introduced to reconstruct the high-resolution surface progressively and adaptively. We have also discussed how to reconstruct surface in a patch-wise fashion and seamlessly stitched the reconstructed surface patches together.

We can envision that other strategies could be used to produce a similar 3D imaging system. Moreover, as this work is not focused on improving photometric stereo, we inherit all the issues known to affect normal estimations (e.g Lambertian surface assumption, albedo estimation, use of light sources and their calibration, etc.).

Chapter 5

Photometric Stereo using Focal Stacking

There is nothing worse than a sharp image of a fuzzy concept.

Ansel Adams

In the previous chapter, we presented a high-resolution 3D imaging framework that consists of a high-resolution photometric stereo system and a structured-light system. However, in many e-heritage settings, one can only rely on a camera with a shallow depth-of-field and without the benefit of auxiliary depth information from structure-light or range scanning. In this chapter, we introduce a unique setup that aims to capture 3D surfaces under such restricted working environment. We observe that using a camera with a narrow depth-of-field often requires focal stacking to properly image the target object. The idea of our approach is to utilize this additional information in the photometric stereo process. In particular, we introduce an

approach to regularize the normals against the varied focused images to improve normal estimation. Experimental results show that this added regularization using the focal stack data outperforms methods that build an all-in-focus normal map by compositing normals computed at each focus distance. We also discuss how the photometric lighting can be used to improve estimations for depth-from-focus which can be incorporated into the overall framework.

5.1 Introduction

This chapter addresses an unconventional photometric stereo framework that relies on using a camera with a very shallow depth-of-field. The shallow depth-of-field means that the entire object cannot be taken in focus within a single image. As discussed in Chapter 1, the working environments of the 2D and 3D imaging for e-heritage are often restricted. In many scenarios, hybrid setups, such as the one introduced in Chapter 4, cannot be used due to spatial restriction. Moreover, both weak illumination and spatial restriction may lead to a shallow depth-of-field. In such case, capturing multiple images at varying focal lengths, i.e., *focal stacking*, is required to extend the depth-of-field to capture the object. This scenario has interesting implications when used in a photometric stereo framework as we now have significantly more input images than these from a conventional setup.

Figure 5.1 shows an illustration of our overall framework, where calibrated lights perform photometric stereo imaging at varying focal settings. From this input our aim is to acquire the object's surface normals. The most obvious method would be to build an all-in-focus normal map by compositing the estimated normals computed for each focus image. This approach however does not exploit the

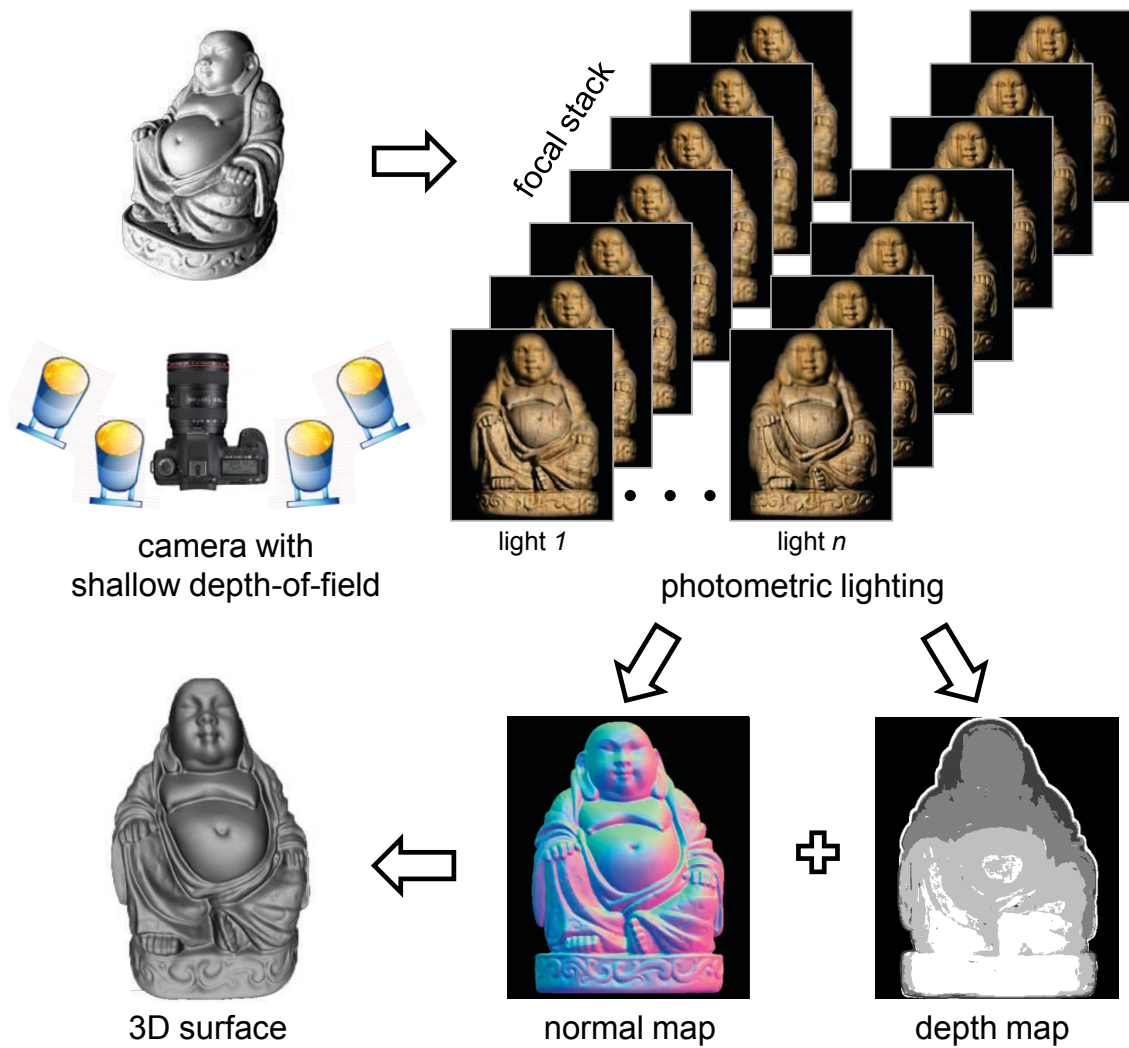


Figure 5.1: Work flow of our system. Both focal stack and photometric stereo data are captured at the same time and exploited to compute the surface normals and coarse depth from which the 3D surface is reconstructed.

redundant observations of each normal present in the focal stack. Our main contribution of this work is to demonstrate how to improve the normal estimation by globally optimizing the estimated normals against each focal image. Experimental results demonstrate that this focal stack regularization produces better results than methods that composite an all-in-focus normal map, even with post-processing by bilateral filtering. Our results are evaluated quantitatively using synthetic data generated by *Maya* [5] with an accurate lens model to simulate the shallow depth-of-field effect, as well as qualitatively using real world examples.

A secondary contribution in this work is with regards to depth-from-focus. A common procedure for photometric stereo is to incorporate positional information to help in the surface reconstruction process. A natural choice for our imaging setup is to use depth-from-focus. To this end, we also show how we can use the photometric lighting to improve depth-from-focus results, again improving the overall results.

The remainder of this chapter is organized as follows: Section 5.2 discusses related work; Section 5.3 describes our approach including normal regularization, depth-from-focus estimation, and surface reconstruction; Section 5.4 demonstrates examples on both real and synthetic inputs; a discussion and summary conclude the chapter in Section 5.5.

5.2 Related Work

While there are a variety of well-studied techniques for 3D imaging, this work limits its focus to photometric stereo and depth-from-focus. For general surveys on 3D imaging, see [81] and [97].

Photometric stereo is a technique to estimate surface normals of an object by observing the object under different lighting conditions. There is a vast amount of literature on photometric stereo, with representative works including [39, 6, 42, 99, 83, 50, 87, 36, 31, 34, 88]. Among those previous works, many of them are based on the Lambert's model [68] which describes the observed intensity by a simple dot product equation between surface normals and lighting directions. Since there are three unknowns in this simplest reflectance model, at least three lighting directions are required to solve the normal direction estimation problem [99]. When there are more than three input images, the optimum normal direction for each pixel can be obtained by using least square fitting. In this work, we will also assume our surface reflectance satisfies this Lambert's model.

After normals estimation, it is common to reconstruct the surface from normals. Since surface normals capture only the relative 3D location between neighborhood pixels, it does not contain any absolute depth information. The reconstructed surface can be distorted when there are errors in the estimated surface normals. More recently, researchers have focused methods to fuse photometric stereo results with other information, such as auxiliary positional data (e.g. [89, 6, 27, 49, 42, 12, 18, 41, 66, 53]), visual hulls and volumes (e.g., [35, 93]), and motion (e.g., [52, 38]).

Our work also intends to utilize auxiliary depth information to improve the results of the final integrated surface using depth-from-focus. These methods estimate an object's surface structure from two or more images of the object with varying focus parameters. Notable examples include [20, 65, 102, 58]. The basic idea involves determining when a point becomes in focus and relating that to the (calibrated) focus distance to the camera. While our approach draws on elements from prior techniques, as far as we are aware, our combination of photometric

stereo and depth-from-focus with focal stacking is unique. In addition, our use of the focus images and photometric stereo lighting to improve the normal estimation and depth-from-focus results presents a new method for 3D imaging in restricted working environments.

5.3 Focal Stack Photometric Stereo

In this section we describe our framework starting with a short discussion on the focal stack imaging and its relationship to the surface normals. This is followed by a description of our normal estimation algorithm which globally optimizes constraints of defocus normals over the entire focal stack. Finally, a simple method to improve the depth-from-focus result using the available photometric stereo lighting is described. Details to our physical and synthetic setup are discussed in Section 5.4.

5.3.1 Focal Stack and Normals

To understand the additional information pertaining to normals in the focal stack, let us first study the relationship of the surface normals when the captured images are out-of-focus. We assume the captured surface follows the Lambertian lighting model. For each point of the input image, we can represent the effect of defocus blur by the following convolution equation:

$$\begin{aligned}
 I(x, y) &= \sum_{(m,n)} I^*(x', y') K(m, n) \\
 &= \sum_{(m,n)} \rho(x', y') |N(x', y') \cdot L(x', y')| K(m, n),
 \end{aligned} \tag{5.1}$$

where I is the captured defocus image, I^* is the ideal all-in-focus image, K is the spatially varying defocus blur kernel with its size proportionates to the depth of the scene, N is the surface normals, L is the lighting direction, ρ is the surface albedo, and $(x' = x - m, y' = y - n)$ is the local neighborhood of (x, y) .

In photometric stereo, we assume the L is a directional light source and therefore constant within the same input image. Now, suppose for a local region, if ρ is constant, we can simplify Equation (5.1) as follows:

$$I(x, y) = \rho L \cdot \left(\sum_{(m,n)} N(x', y') K(m, n) \right). \quad (5.2)$$

Solving Equation (5.2) using standard least square methods [99], we obtain the normals which have undergone defocus blur:

$$\tilde{N}(x, y) = \sum_{(m,n)} N^*(x', y') K(m, n) \quad (5.3)$$

Note that although the assumption of local constant albedo appears to be restrictive, in practice, we find that with our high-resolution setting this assumption is valid in most places. This will be demonstrated in our results in Section 5.4.

5.3.2 Normals Refinement Using Deconvolution

Although within the focal stack, there should be one estimation per pixel that is in-focus (or the best in focus), we found that simply compositing the in-focus normals resulted in noisy result. This is due to the quantization effect in the focal stack and noise in estimating where a surface patch is in focus. Our aim is instead to estimate the normals globally using the entire focal stack.

For a focal stack with M number of levels, we can obtain M observations of normals as follows:

$$\begin{aligned}
 \tilde{N}_1(x, y) &= \sum_{(m,n)} N^*(x', y') K_1(m, n) \\
 \tilde{N}_2(x, y) &= \sum_{(m,n)} N^*(x', y') K_2(m, n) \\
 &\vdots \\
 \tilde{N}_M(x, y) &= \sum_{(m,n)} N^*(x', y') K_M(m, n)
 \end{aligned} \tag{5.4}$$

Solving each individual equation alone in Equation (5.4) is well-known to be an ill-posed problem. However, when we combine all the equations together, this problem becomes well-posed as shown in [3]. This can be formulated into a set of linear equations by rewriting Equation (5.4) into $AN^* = b$ with:

$$\begin{aligned}
 A &= \begin{bmatrix} K_1 \\ \vdots \\ K_M \end{bmatrix}^T \begin{bmatrix} K_1 \\ \vdots \\ K_M \end{bmatrix} + w(G_x^T G_x + G_y^T G_y), \\
 b &= \begin{bmatrix} K_1 \\ \vdots \\ K_M \end{bmatrix}^T \begin{bmatrix} \tilde{N}_1 \\ \vdots \\ \tilde{N}_M \end{bmatrix}
 \end{aligned} \tag{5.5}$$

where G_x and G_y are the x - and y - derivative filters used as regularization to help suppress noise and ringing artifacts in the deconvolution. The term w is the regularization weight. Since the defocus kernel is different at each pixel, the deconvolution process is performed for each pixel individually. In our implementation,

we calibrated the defocus kernel K using textured pattern for each level of defocus in the focal stack. Note that if the lens optics are known, the K can be computed directly. The defocus kernel is selected according to the depth map estimated from depth-from-focus which will be detailed in next subsection. While the estimated defocus kernel might be inaccurate, the multiple observations of the blurry normals and the neighborhood regularization in Equation (5.5) help protect us from these estimation errors. Figure 5.2 shows the comparisons between our estimated normals and all-in-focus normals. All-in-focus normals are composited from a stack of normals that are computed from photometric images in a focal stack. While these look similar, on careful inspection it is clear the simple all-in-focus normal map (Figure 5.2(b)) is more noisy than that obtained with regularized normal map (Figure 5.2(a)). Quantitative evaluations and comparisons of our normal estimation method will be given in Section 5.4.

5.3.3 Depth-from-Focus Exploiting Photometric Lighting

Techniques for depth-from-focus return a depth map from the focal stack via edges/textures sharpness analysis. These measurements depend greatly on the rich texture information on the object surface. In situations where the object's surface does not contain any texture, the above measurements might fail since there is no obvious difference between the in-focus image and the out-of-focus images in the focal stack.

In photometric stereo, scenes are captured with varying illuminations under different lighting directions. The controlled light sources produce shadings or shadows according to the geometry and curvature of object surface, regardless of

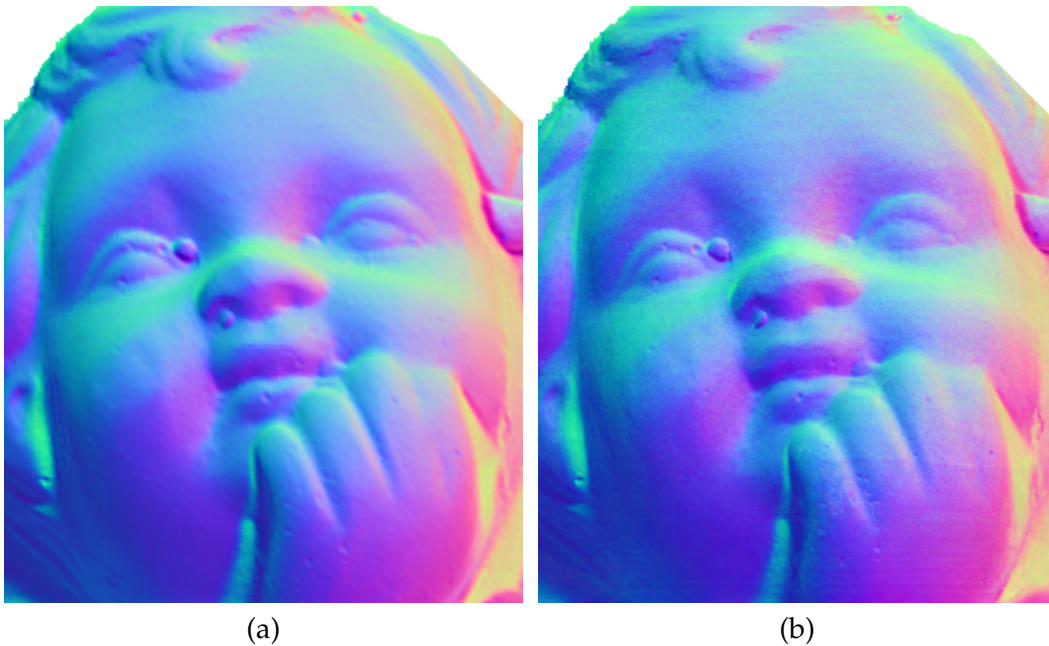


Figure 5.2: The estimated normals, with (a) and without (b), deconvolution refinement. While similar at first glance, on careful inspection it is apparent that the normal map in (b) exhibits more noise than the normals in (a).

local albedo. In other words, the shadings/shadows can still exist when the local albedo is the same. With the induced discontinuities from shadings or shadows, we can analyze the amount of defocus even the object is homogeneous in color. Note that when more lighting directions are used in photometric stereo, we can get more accurate depth since we have more observations of shading/shadows discontinuities. Without the directional light sources, the image of the object surface may look flat under ambient lighting condition.

We take the photometric stereo images at each focus distance as a trade off of additional capturing time. The lighting directions are fixed and calibrated so that at each focus distance, we can find the corresponding photometric stereo images in the focal stack for focus analysis. The pseudo code of our depth-from-focus algorithm is detailed in Algorithm 1. F_k^i is the focus measured under lighting direction i at

Algorithm 1 Depth-from-focus using photometric stereo lighting

```

1:  $F_{max} = 0$ 
2:  $\bar{d} = 0$ 
3: for each illumination  $i$  do
4:    $F_{max}^i = 0$ 
5:    $d^i = 0$ 
6:   for each focus distance  $k$  do
7:     Compute focus measure  $F_k^i$ 
8:     if  $F_{max}^i < F_k^i$  then
9:        $F_{max}^i = F_k^i$ 
10:       $d^i = d_k$ 
11:     end if
12:   end for
13:   if  $F_{max} < F_{max}^i$  then
14:      $F_{max} = F_{max}^i$ 
15:      $\bar{d} = d^i$ 
16:   end if
17: end for

```

the focus distance k . We use the Sum-Modified-Laplacian introduced by Nayar and Nakagawa [65] for focus analysis. The term d_k is the depth corresponding to the focus distance k . The term \bar{d} is our resulting depth estimation. In addition to the depth map, we will also compute a confidence map C which measures the reliability of the depth value computed at each pixel location. Our confidence map is computed using the following equation:

$$C(x, y) = \frac{\sum_{(j \in \mathcal{J})} F_{max}^j(x, y)}{\sum_{(i \in \mathcal{L})} F_{max}^i(x, y)}, \quad (5.6)$$

where $\mathcal{J} = \{i | d^i = \bar{d}\}$, \mathcal{L} is the set of all illuminations, \bar{d} is the overall depth estimate and d^i is depth estimate for i th illumination, as computed in Algorithm 1. Figure 5.3 shows the comparison between our estimated depth using photometric lighting and usual lighting. The zoomed-in region shows that the estimated depth is less

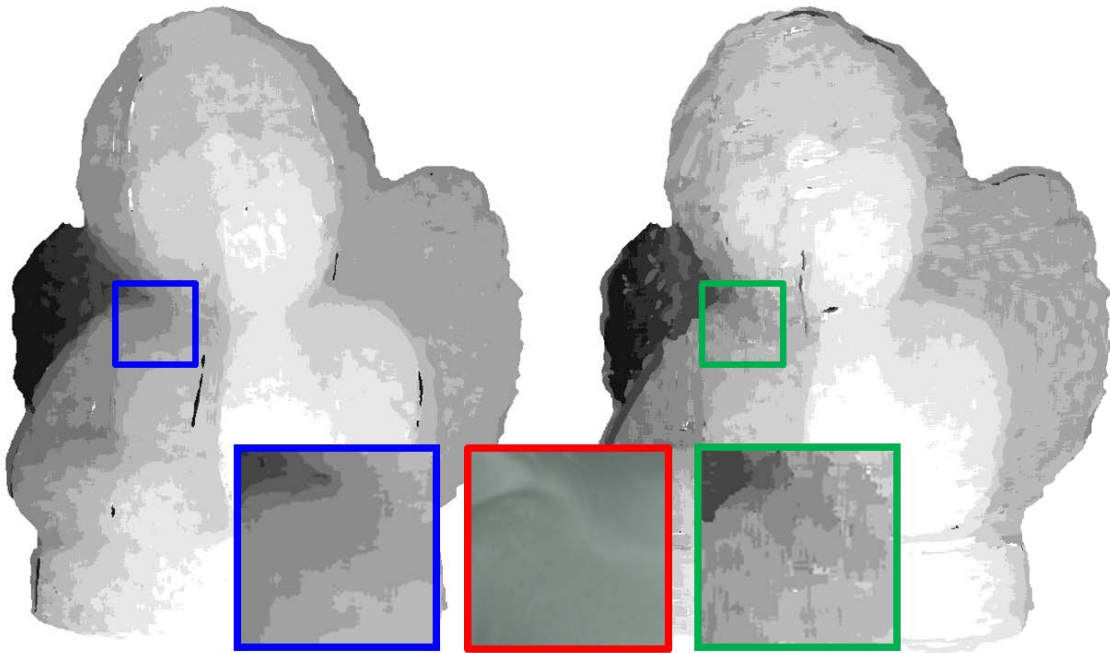


Figure 5.3: The estimated depth map using depth-from-focus, with (left) and without (right) photometric lighting. The zoomed-region shows that the estimated depth is less noisy with photometric lighting. Note that the red box in the middle shows the zoomed-region of an input image.

noisy when the photometric lighting is used.

5.3.4 Surface Reconstruction

Using our acquired normal map and depth-from-focus results we reconstruct the 3D surface using the method presented in Chapter 4. Note that although the technique outlined in Chapter 4 focused on fusing normals and positional data obtained at significantly different resolutions, the technique still works when the resolutions are similar. The work in Chapter 4, however, assumed that the positional information came from a structured-light scanner and each 3D coordinate had the same confidence. In our case, however, the depth-from-focus data are more noisy. We

therefore use the computed confidence map C to weight the positional information, which can be easily done using the formulation in Chapter 4.

5.4 Experimental Results

Our approach is tested on synthetic and real objects. Since ground truth normals are difficult to obtain for this type of setup, we opted to perform a series of synthetic experiments using the *Maya* rendering software to simulate shallow depth-of-field imaging with 3D models from which we could compute normals for ground truth. Experiments are also performed on real objects which show the difference in quality using both our estimated normals and depth-from-focus algorithm.

5.4.1 Synthetic Examples with Ground Truth

Generating Synthetic Data We use four objects with known geometry and surface normals. We select the object material to be purely Lambertian. To unify the camera setup, all the objects are scaled to the similar size. To simulate the focal stack, we set the camera in *Maya* at a fix distance looking at the object. We turn on *Maya's* depth-of-field setting and set the camera aperture to $f1.0$ to simulate a shallow depth-of-field at each focus distance. We then render the objects using this synthetic camera model. In order to test the effect of texture on our method, we render each object both with and without texture.

The camera is focused at four different distances from the closest point of the object to the furthest point observable from the camera. To simulate photometric stereo, we set up four different directional lights. The intensity of all the lights are

set to the same value. At each focus distance, four images that each with only one light on are rendered. In total, 16 images are rendered and serve as input of our methods.

Calibrating Defocus Kernels To perform our normal refinement, we need to calibrate the defocus kernels for each focus distance. While we could compute this from the lens model used by *Maya*, to provide a more realistic experiment, the defocus kernels are estimated in the same manner as with real objects. This is done by placing a textured plane at the nearest focus distance. Then we render multiple images with the camera focused at each focus distance. As a result, the textured plane is blurred with different defocus kernels at different focus distance. These images are then used for defocus kernel estimation.

Comparison and Results With the focal stack photometric stereo images rendered from *Maya*, we compute the normals using our method described in Section 5.3. As a comparison, we also compute all-in-focus normals using two focal stack methods: a classic all-in-focus method [32] and a recent method based on graph-cut [2]. For the latter, we use the code provided by the authors' webpage. Figure 5.4 shows our synthetic examples with and without texture, as well as normals computed using our method. In Table 5.1, we show the mean angular errors (in degrees) between normals computed by ours and the two all-in-focus methods and the ground truth. To show our method is not just a matter of filtering, we applied bilateral filtering on normals computed by the two all-in-focus methods. While the overall error is relatively small for all methods, our approach is consistently better. We also applied our method on the four examples without texture (see the last two rows

| Textured | Ex1 | Ex2 | Ex3 | Ex4 |
|---------------------------------|------|------|------|------|
| Hausler[32] | 2.51 | 1.07 | 0.97 | 2.12 |
| Agarwala <i>et al.</i> [2] | 2.44 | 0.88 | 1.07 | 1.63 |
| Hausler[32] + BL | 2.71 | 1.12 | 1.10 | 2.22 |
| Agarwala <i>et al.</i> [2] + BL | 2.12 | 0.88 | 0.97 | 1.50 |
| Ours | 2.06 | 0.77 | 0.69 | 1.41 |
| Ours Textureless | 2.05 | 0.78 | 0.70 | 1.41 |

Table 5.1: Comparison on average angular error (in degrees) of normals among our method and the all-in-focus methods with and without bilateral filtering. Comparisons are on textured objects. The last row shows the results are virtually identical when the object is textureless.

of Table 5.1) with virtually the same results. This shows our approach’s ability to handle textured and textureless inputs.

5.4.2 Real Objects

This section shows several real objects captured by our system. For each object, normals and depths are estimated using the methods described in Section 5.3. Then the surfaces are reconstructed using the technique described in Chapter 4.

We use a Canon EOS 1Ds Mark III with a Canon EF 50mm f/2.5 Compact Macro Lens for our experiments. The camera is mounted on a translation stage. To capture the focal stack photometric stereo data, we first focus on the nearest point on the target object. Then we change the focus level by moving the camera $3mm$ closer to the object along the stage while keeping the focus distance fixed. This has the same effect as changing the focal distance while keeping the camera position fixed. In the setup, we make sure that the depth-of-field for each image is about $3mm$ so that each point on the object is in focus in at least one image. For each focus level, the same four lights are used for photometric stereo. The light directions are calibrated

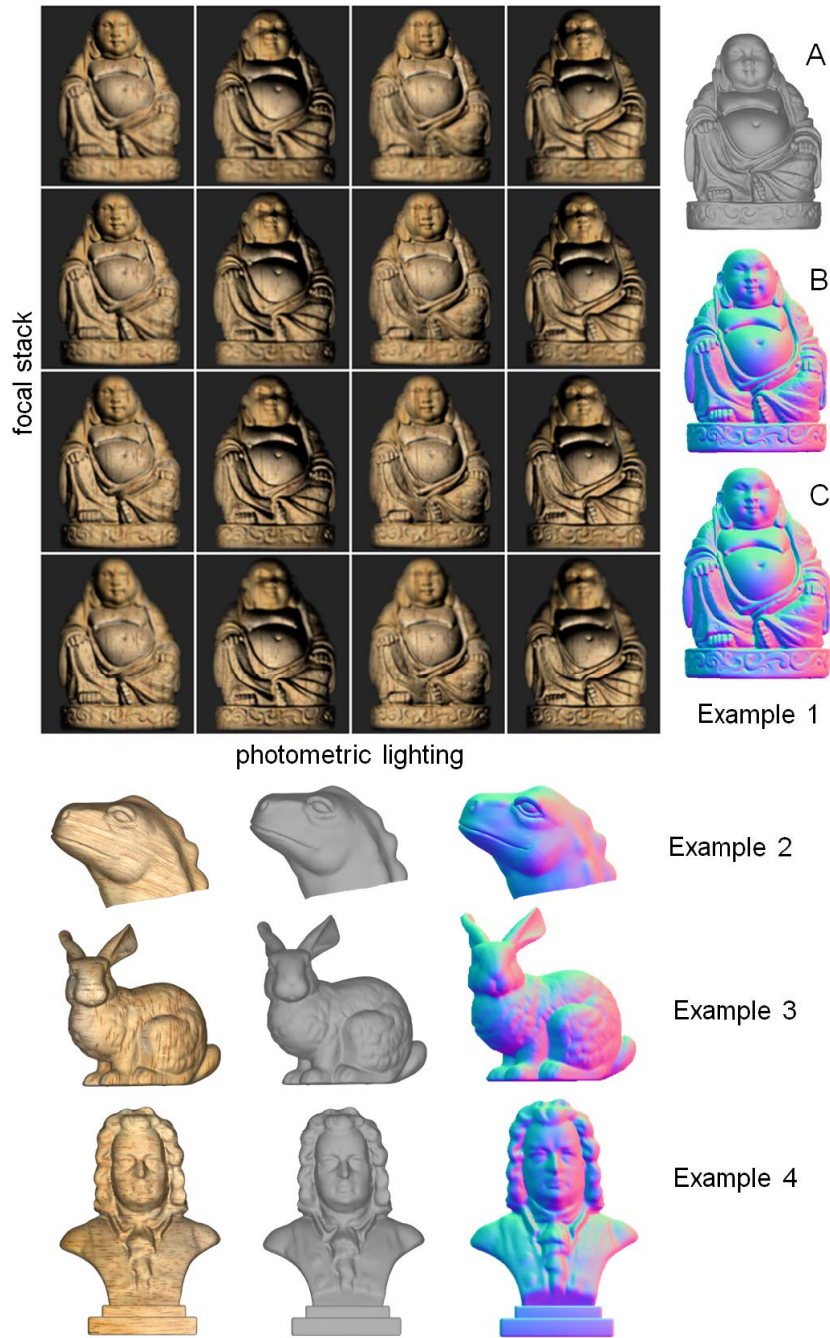


Figure 5.4: On the top, we show the full input images of example 1. The object without texture (A), normal map of the ground truth (B) and our method (C) are shown too. On the bottom, example 2, 3, and 4 are shown with and without texture, as well as normal maps computed from our method.

by using a mirror ball placed in the scene. Similar to the synthetic experiments, we pre-calibrated the defocus kernels by using a patterned board.

In Figure 5.5, we show an example with heavy texture and pitted surface. The zoomed-region shows the difference of normals and relighted images from our method and all-in-focus method ([32]). Figure 5.6 and Figure 5.7 show estimated normals and 3D reconstructions of a *statue* and an *angel* figurine. The zoomed-regions on the right show the comparisons of our method and all-in-focus methods ([32] and [2]). Figure 5.8 shows the estimated normals and the 3D reconstruction of a *duck* figurine. The zoomed-regions on the bottom show the comparisons of our methods with and without normal refinement, and depth-from-focus with and without exploiting photometric lighting. The improvement in the results using the depth-from-focus algorithm are clear, producing less noticeable artifacts in the coarse depth estimation. The normal map improvements are more subtle, but on close inspect reveal that our approach has less noisy normals, resulting in smoother results that still contain small details present on the objects surface. For the examples produced using [2], seams are often noticeable in the final results due to the graph-cut algorithm (see the corresponding zoomed-regions in the figures).

5.5 Summary and Discussion

This chapter has addressed an imaging framework to acquire surface normals and 3D scans of objects via focal stacking and photometric stereo. While admittedly an unusual imaging setup, it is well suited for 3D imaging in a very restricted environment. When it becomes necessary to work within this imaging scenario, we have offered two technical contributions that can improve the results by exploiting

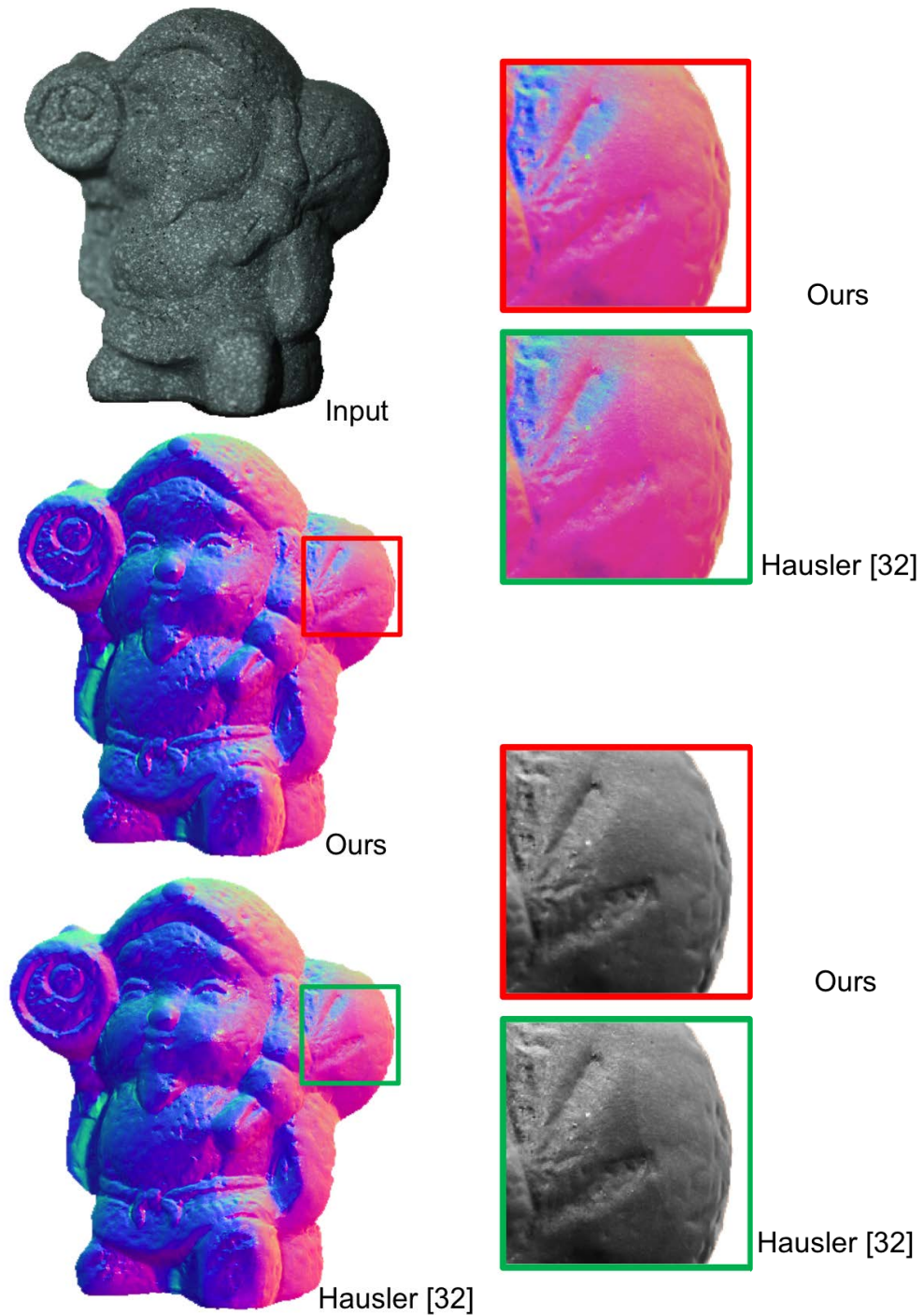


Figure 5.5: An example with heavy texture and pitted surface. (Left) An example of input image, normal map computed by our method and Hausler [32]. (Right) the comparison of a zoomed-region of our and Hausler [32]. Note that bottom right shows re-lighted images from normals.

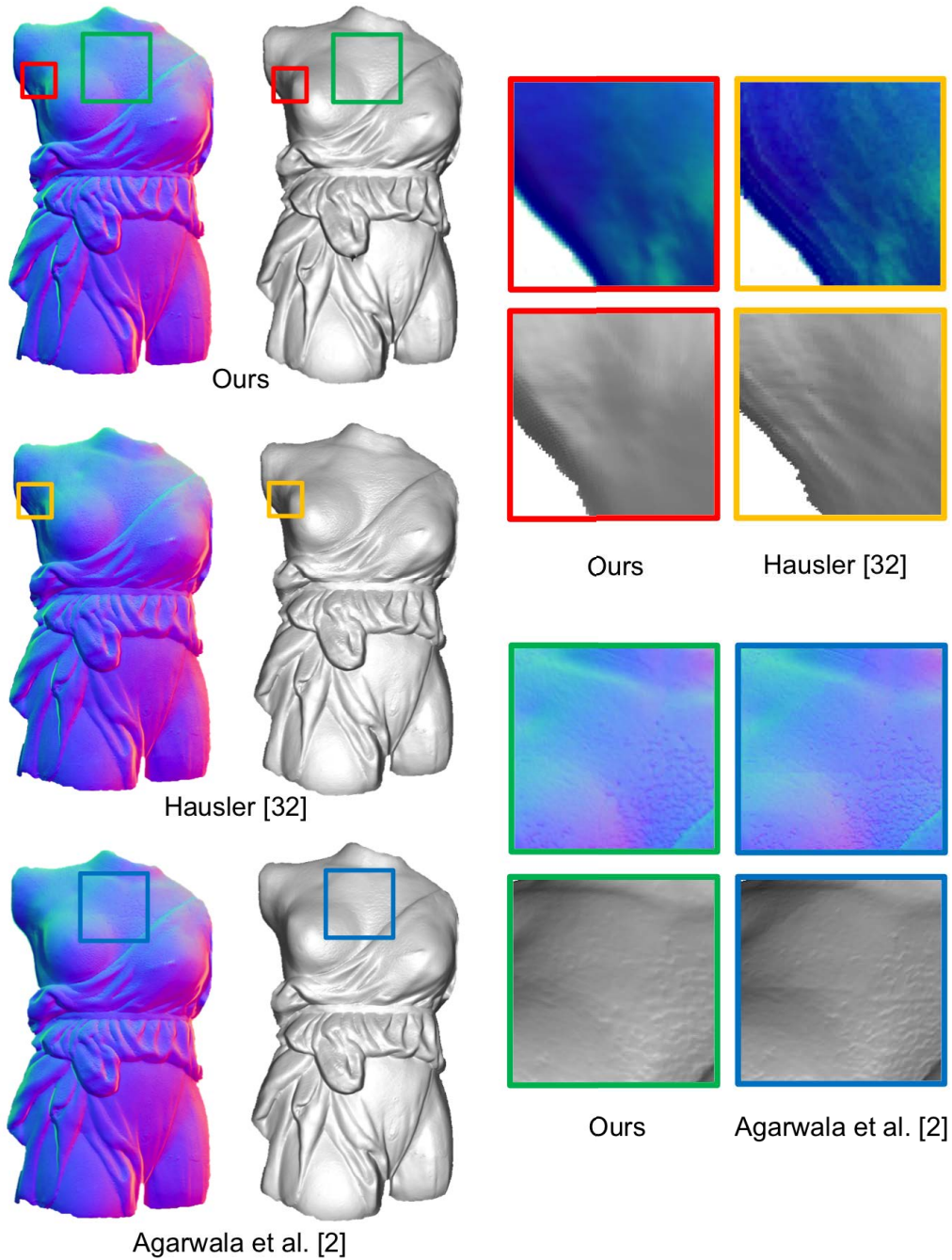


Figure 5.6: Normal map and 3D reconstruction result of *statue* figurine. (Left) Normal map and 3D reconstruction of our approach, Hausler [32] and Agarwala *et al.* [2]. (Top right) the comparison of a zoomed-region of our and Hausler [32]. (Bottom right) the comparison of a zoomed-region of our and Agarwala *et al.* [2].

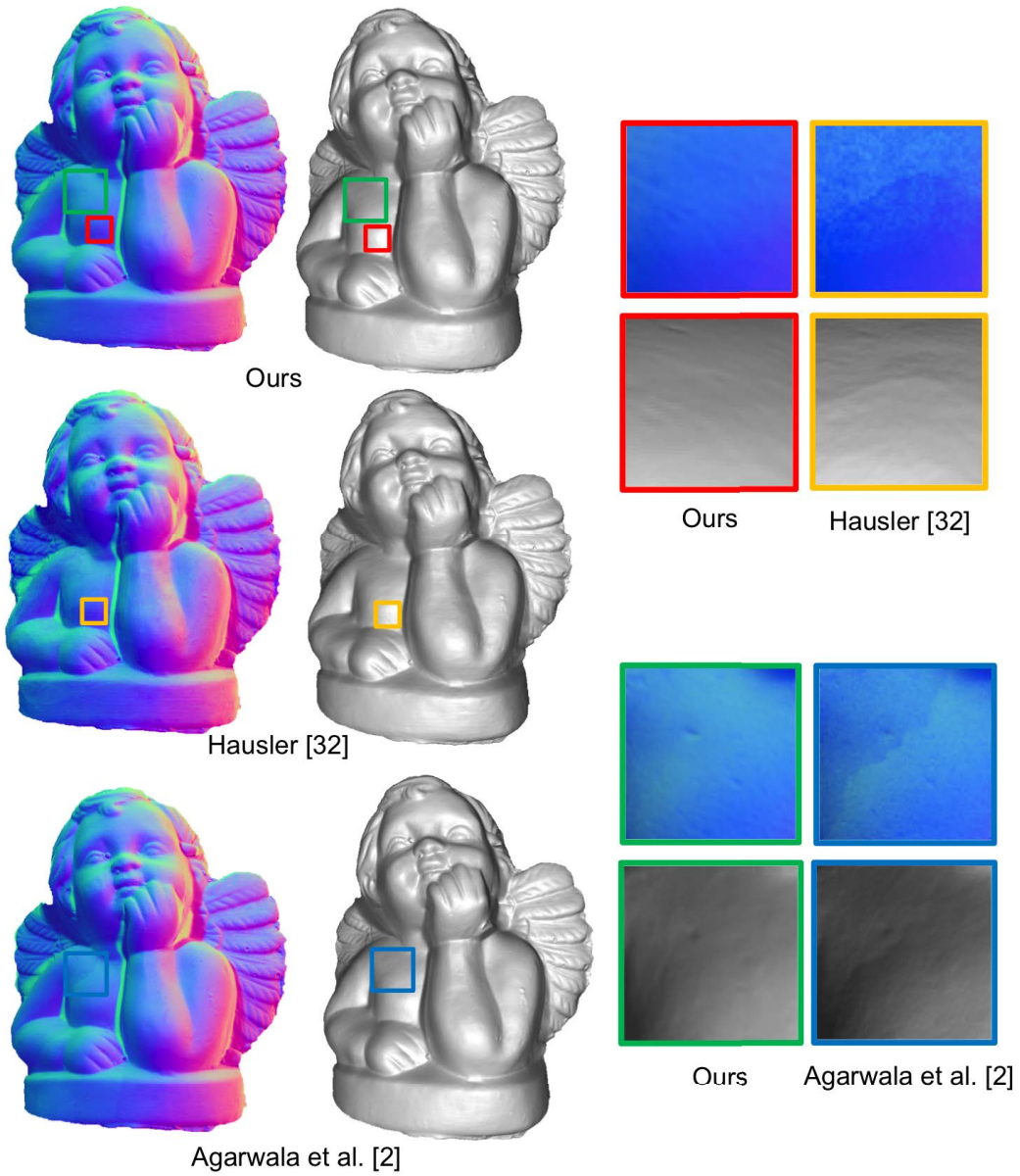


Figure 5.7: Normal map and 3D reconstruction result of *angel* figurine. (Left) Normal map and 3D reconstruction of our approach, Hausler [32] and Agarwala *et al.* [2]. (Top right) the comparison of a zoomed-region of our and Hausler [32]. (Bottom right) the comparison of a zoomed-region of our and Agarwala *et al.* [2].

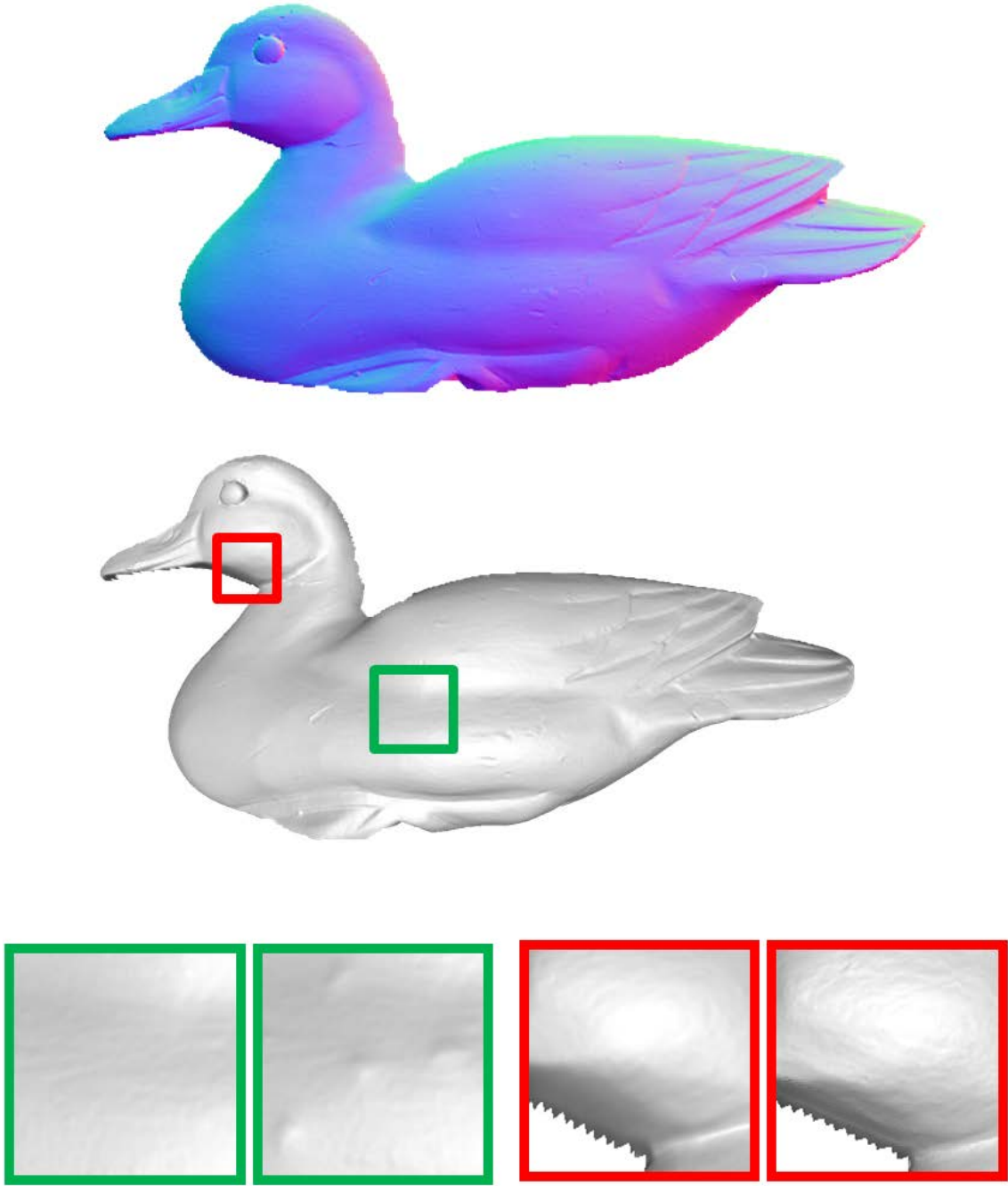


Figure 5.8: Normal map and 3D reconstruction result of *duck* figurine. (Bottom left) the comparison of a zoomed-region with (left) and without (right) photometric depth-from-focus. (Bottom right) the comparison of a zoomed-region with (left) and without (right) normal refinement.

the information available within the rich focal stack data.

Our main contribution is to show how to regularize the normal estimation against the focal stack images. Both real and synthetic experimental results demonstrate that the normal estimation using this approach is better. Our secondary contribution is to exploit the varied lighting offered by the photometric stereo setup to improve the confidence measure for depth-from-focus. While the depth-from-focus is a simple use of the photometric lighting, it does show an overall improvement in the results.

The frameworks proposed in this chapter and Chapter 4 do not compete with each other. On the one hand, our normal refinement technique described in Section 5.3 can be used as a pre-processing step of the framework described in Chapter 4 when focal stack is captured. On the other hand, the surface reconstruction algorithm described in Section 4.4 is not restricted to the system setup introduced in Section 4.3. The low-resolution/sparse data can be obtained from any method, including structured-light system as well as depth-from-focus (by removing less confidence points).

From a practitioner's point of view, we recommend obtaining low-resolution geometry using structured-light system/range scanning whenever the working environment allows, since such system provides better 3D data than depth-from-focus in general. When the imaging scenario does not allow the use of additional hardware setup, the framework in this chapter should be used. In addition, our formulation does not limit the use of lights. One can use one light being moved around or multiple lights at fix positions, as long as the illumination direction is known and there are at least three photometric images for each focus distance.

Chapter 6

Conclusion

In photography, the smallest thing can be a great subject. The little, human detail can become a Leitmotiv.

Henri Cartier-Bresson

This chapter concludes this thesis by giving a short summary for the work described in previous chapters, including field work to capture high-resolution images using a large-format digital camera, an imaging framework coupling a standard structured-light setup and photometric stereo, and a unique setup that combines photometric stereo with focal stacking. This is followed by the review of objective of the thesis. Finally, a short description on possible future research directions is discussed.

6.1 Summary

The goal of this thesis is to improve high-resolution 2D and 3D imaging for e-heritage. This includes addressing challenges faced by the high-resolution requirement and restricted working environments. Chapter 1 introduced and motivated this thesis by briefly describing e-heritage and current research in 2D and 3D imaging, followed by the main contributions of this thesis. To facilitate the understanding of our work, Chapter 2 describes the large-format digital camera used in most of the work in this thesis as well as my involvement in the development of the camera.

Chapter 3 presented recent field work for imaging Buddhist art at the UNESCO world heritage site, the Mogao Caves, in Dunhuang, China. Current challenges faced by the Dunhuang Academy using conventional mosaicing techniques with a small-format camera were outlined. The preliminary results showed that using a large-format digital camera not only improves the resolution of the imaging, but also significantly reduces time and effort. The chapter was concluded with several lessons learned from our field work.

Chapter 4 proposed a framework to capture high-resolution 3D surfaces using a hybrid system that consists of a high-resolution photometric stereo system and a structured-light system. Given such massive amounts of data and with significant resolution asymmetry in the respective subcomponents, a multi-resolution pyramid approach was introduced to reconstruct the high-resolution surface progressively and adaptively. It was also discussed that how to reconstruct surface in a patch-wise fashion and seamlessly stitch the reconstructed surface patches together. Our high-resolution 3D results show much finer surface details even

compared with the result produced by a state-of-the-art laser scanner [48].

Chapter 5 proposed an imaging framework to acquire surface normals and 3D surfaces of objects via focal stacking and photometric stereo. The proposed approach regularizes the normals against the varied focused images to improve normal estimation. It was also discussed that how the photometric lighting can be used to improve estimations for depth-from-focus which can be incorporated into the overall framework. To show our method can produce more accurate normals, both real and synthetic experimental results were compared with all-in-focus normals using two focal stack methods [32, 2] with/without simple bilateral filtering. The advantage of our algorithm was further demonstrated by showing improved reconstructed 3D surfaces.

6.2 Review of Objective

The objective of this thesis is achieved as follows:

- We showed that the use of large-format digital camera for high-resolution 2D imaging is feasible in a real world project. We also reported the lessons learned and remaining challenges from our field work study.
- We developed a high-resolution 3D framework combining high-resolution photometric data and low-resolution positional data. To our best knowledge, the proposed framework can produce the highest sampling rate demonstrated to date.
- We developed a high-resolution 3D framework that relies only on a camera with a shallow depth-of-field without the benefit of auxiliary depth informa-

tion such as that obtained from a structured-light system.

6.3 Future Directions

There are several future research directions for the work presented in this thesis. These are summarized in the following:

Non-Lambertian Surface The approaches proposed in this thesis assume object surfaces satisfy Lambertian model. Many historical artifacts satisfy this assumption due to aging or simply the dust on the surface. However there are still objects that are more specular such as metal and jade, which violate the Lambertian assumption. Relaxing this assumption will make the proposed methods more applicable in e-heritage settings.

Refinement of Depth-from-Focus with Photometric Stereo Chapter 5 shows photometric stereo can improve depth-from-focus. Further work can be done in this direction. For example, using the estimated defocus kernels to improve the depth-from-focus by helping better interpolate the depth values via kernel interpolation.

High-Resolution Visualization As our methods could produce 3D scans with 600 samples per mm^2 , one problem from such high-resolution is how to visualize such large amounts of data. Even more challenging is how to visualize such 3D data over the Internet, similar to that of the Google Art Project (2D only). To our best knowledge, there is no good existing solution to this problem.

Appendix A

The dgCam Project: A Digital Large-Format Gigapixel Camera User Manual

A.1 Scope

This document describes how to operate the dgCam. This document does neither describe how the camera *works* nor *how it should be built*. These parts are described in the “Computational tutorial” document and the “Construction tutorial” document that are available at the camera web site: <http://www.dgcam.org>.

A.2 User Interface

This section describes the most frequently used user interface windows of the software including the main window, camera control window, calibration window, and



Figure A.1: The main window contains a viewfinder on the left and a control panel on the right.

manual focus window. The functionalities of these windows are briefly described too. To complete a specific task such as capturing image, camera calibration, please refer to Section A.3.

A.2.1 Main Window

The main window of the software shown in Figure A.1 contains a viewfinder on the left and a control panel on the right.

Viewfinder displays the video from the auxiliary video camera. Note that the maximum region can be captured by the main camera is highlighted while the

rest field-of-view is dimmed. You could perform the following tasks through the viewfinder:

1. Send the sensor to certain location by right clicking on the highlighted region.
2. Select a region of interest (highlighted by blue dashed rectangle) to be captured by left clicking and dragging.

Control Panel provides the main functionalities such as starting/stopping the camera, adjusting capture parameters, manual focusing, camera calibration, snapshot, capturing image, fast stitching, and focal stacking. Section A.3 will elaborate the details of these functions.

A.2.2 Capture Control Window

The capture control window shown in Figure A.2 is activated by clicking on “Capture Control” button in the main window. The capture control window allows you to adjust various parameters used for capturing an image, including the following:

- **Exposure Time** You can select auto exposure or input a fixed exposure time. Exposure bracketing is supported too.
- **Gain** You can select auto gain or input fixed gains for R, G1, G2 and B.
- **Focus** You can select auto focus or manual focus. Focus bracketing is supported too.
- **Overlap** You can select the overlapping percentage between each stage movement. The default selection is 25%.

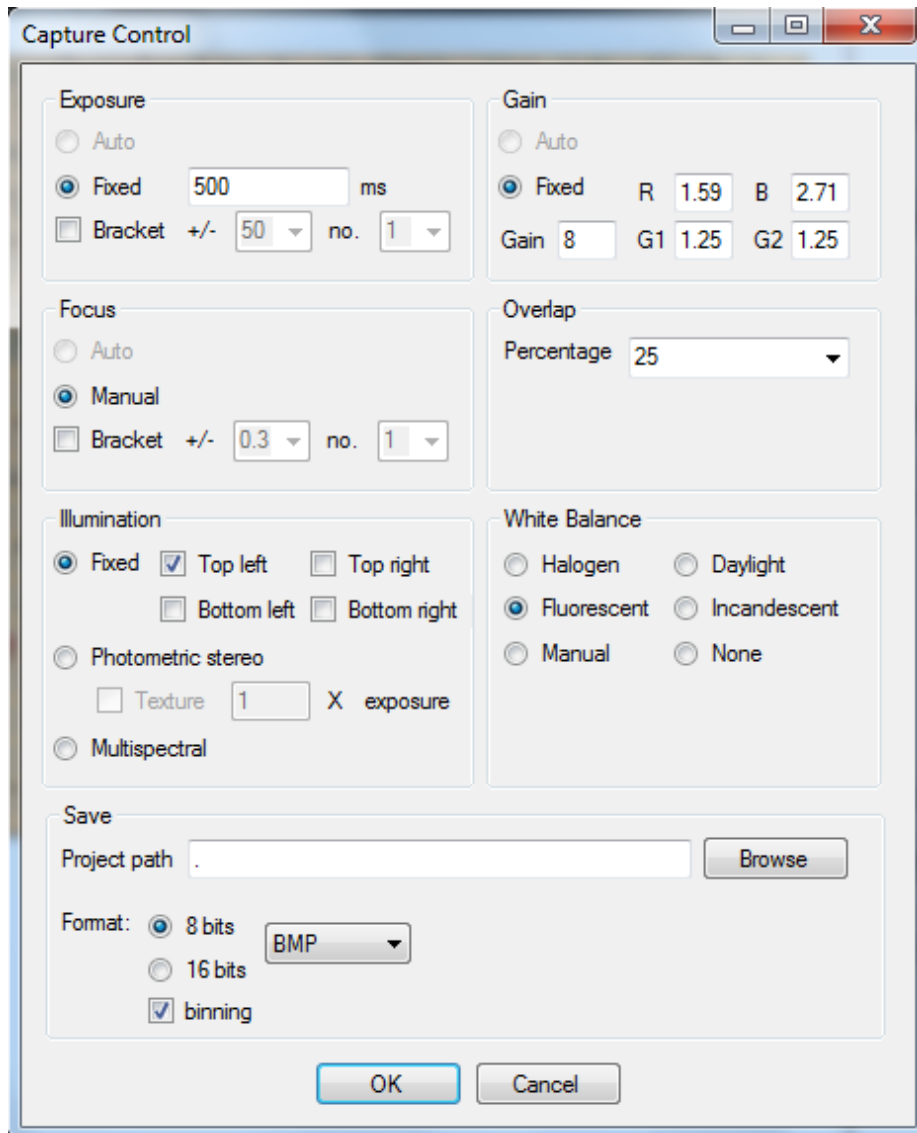


Figure A.2: The capture control window allows you to adjust various parameters used for capturing an image, including exposure time, gain, focus, overlap, illumination, white balance, and save options.

- **Illumination** The illumination methods, *fixed*, *photometric stereo* and *multispectral* are supported. *Fixed* allows you to turn on any combination of the four main lights attached on the camera. *Photometric stereo* is used to

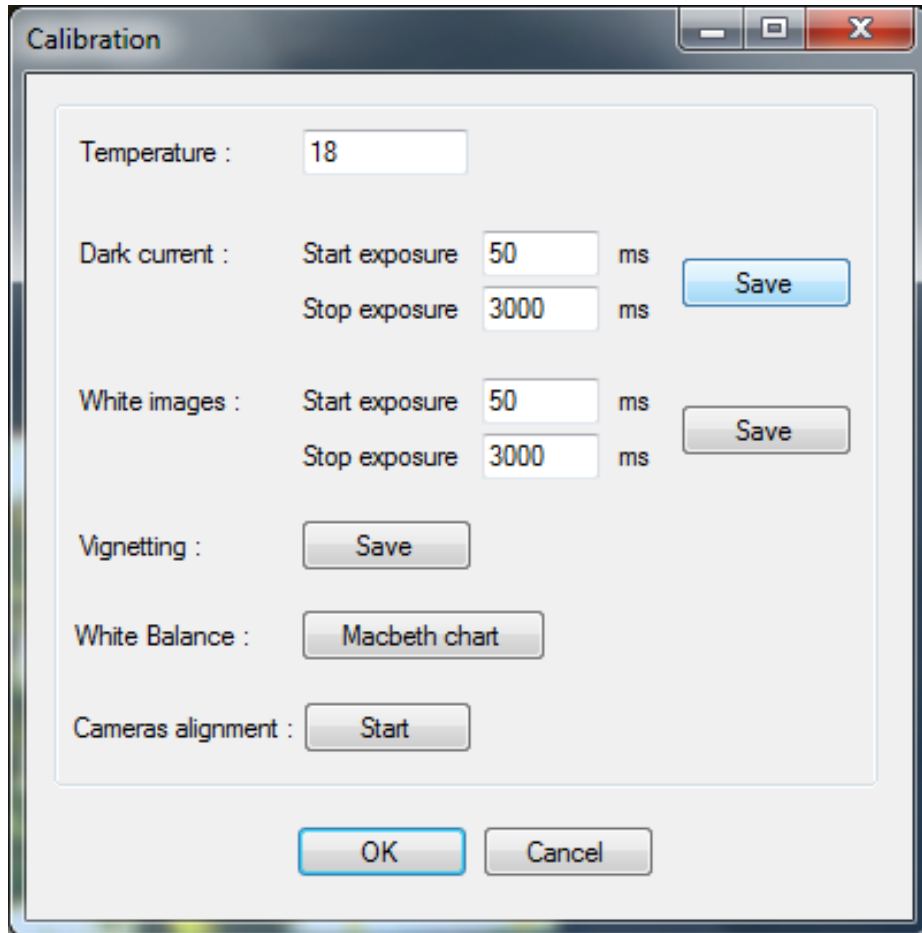


Figure A.3: The calibration window allows you to calibrate dark current, white images, white balance, and camera alignments.

capture photometric data that requires to turn on the four main lights one by one during the capturing procedure. When *photometric stereo* is selected, the binning function is used. To capture color image in addition to the four grayscale images, you need to check “Texture” box and input the multiplier of the exposure time.

- **White Balance** Various white balance options are available including *halogen*, *daylight*, *fluorescent*, *incandescent*, *manual* and *none*.

- **Save** You can select project path (where images and all the calibration settings are saved) and image format. Selecting “Binning” will use the binning function of the main camera and save images in grayscale.

A.2.3 Calibration Window

The calibration window shown in Figure A.3 is activated by clicking on “Calibration” button in the main window. The calibration window allows you to calibrate the following:

- **Dark Current** Input the start and stop exposure time. When clicking on “Save” button, a series of images with different exposure time will be saved under the current input temperature.
- **White Images** Similar to dark current calibration, input the start and stop exposure time, and click on “Save” button.
- **Vignetting** Click on “Save” button to save images for calibrating vignetting. You need to ensure that the image is properly exposed.
- **White Balance** Click on “Macbeth Chart” button to save images for calibrating white balance. You need to put a Macbeth chart in front of the camera.
- **Cameras Alignment** Click on “Start” button to align the field of views of the main camera and auxiliary video camera. The detailed procedure will be described in Section A.3.

A.2.4 Manual Focus Window

The manual focus window shown in Figure A.4 is activated by clicking on “Manual Focus” button in the main window. A separate window displaying video from the main camera is started too. The manual focus window allows you to manually move the lens and the sensor of the main camera, as described in the following:

- **Moving Lens** You can move the lens to focus on far/near objects by using move buttons (▶ ◀), step move buttons (▶◀◀▶) and stop button (■), or send the lens home (focus on infinity) using **H** button. Use the track bar to adjust the moving speed.
- **Moving Sensor** You can move the sensor freely along the x or y stage by using move buttons (▶ ◀ ▼ ▲), step move buttons (▶◀◀▶ ▼ ▲) and stop button (■), send the lens to the center of field of view by using **C** button or send the sensor home by using **H** button. Use the track bar to adjust the moving speed. Clicking on “Align with grid” button will align the sensor with the nearest grid¹ to the current location. The function is especially useful when calibrating cameras alignment (see Section A.3).
- **Adjust the Setting of the Sensor** Click on “Video” button to further adjust the setting of the sensor such as brightness or gamma (see Figure A.5, A.6, A.7, A.8). Click on “Video format” button to adjust video format setting of the sensor such as output format, output size or frame rate (see Figure A.9). Note that the most of the settings adjusted using “Video” and “Video format” buttons

¹The sensor moves along the x and y stages in an grid fashion. The step size of each sensor movement is determined by the overlap percentage specified in capture control window (see Section A.2.2).

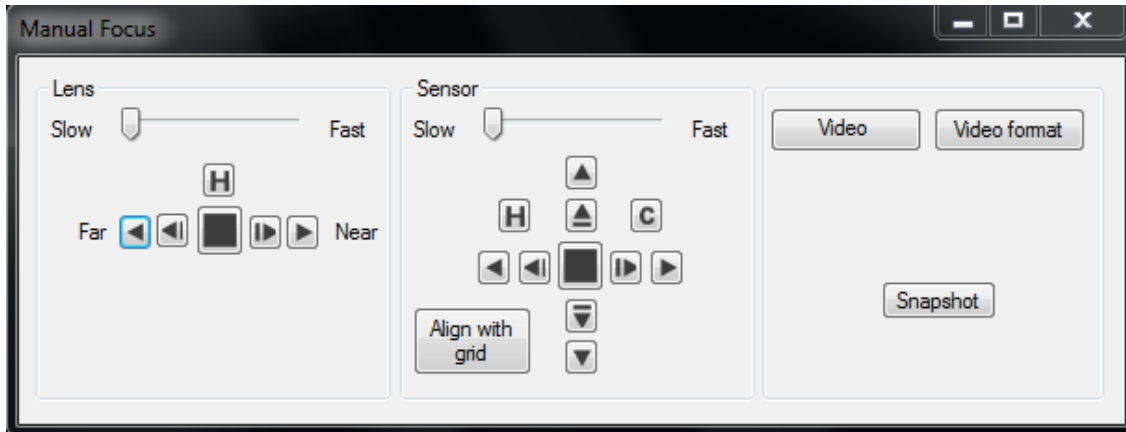


Figure A.4: The manual focus window allows you to move the lens and the sensor as well as adjusting the video display of the sensor.

only affect the video of the main camera not still images captured. As an exception, the settings under “Image properties 2” tab of “Properties” window activated by clicking on “Video” button will affect both both video and still images.

A.3 Working with the Software

This section describes the tasks that are frequently performed when working with the camera using the software.

A.3.1 Start or Stop Camera

When the software starts, all the hardware devices including the main camera, the auxiliary video camera, the relay to the four main lights, and the translating stages are not started yet. Click on “Start” button in the main window to start all the

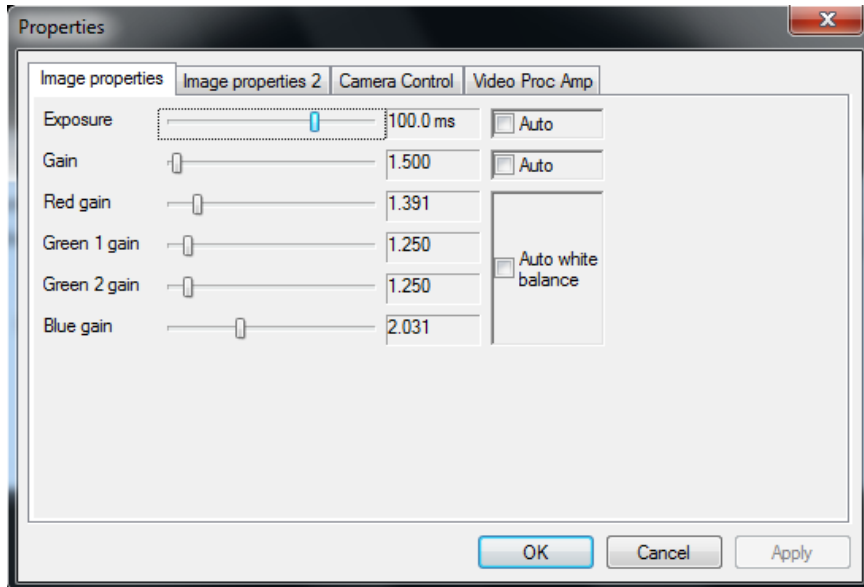


Figure A.5: "Image properties" tab of the properties window activated by clicking on "Video" button in the manual focus window.

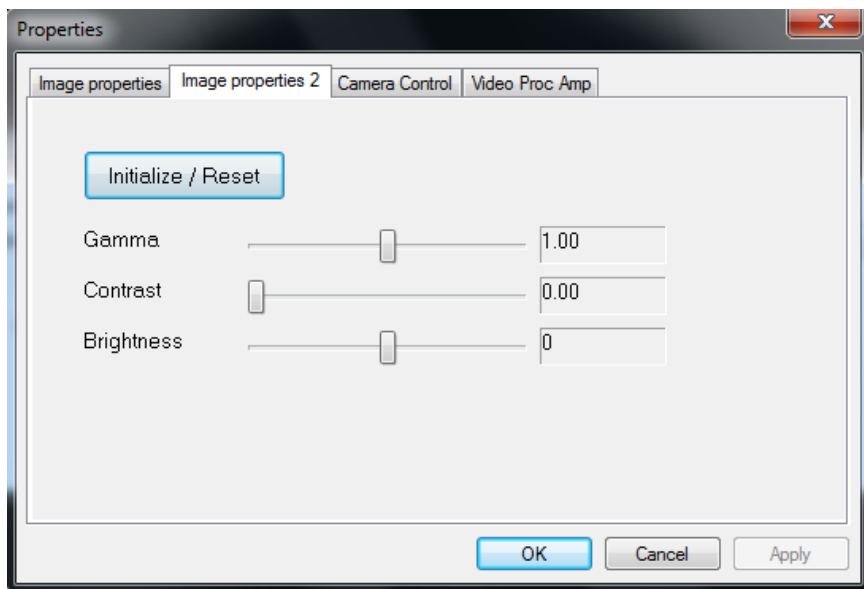


Figure A.6: "Image properties 2" tab of the properties window activated by clicking on "Video" button in the manual focus window.

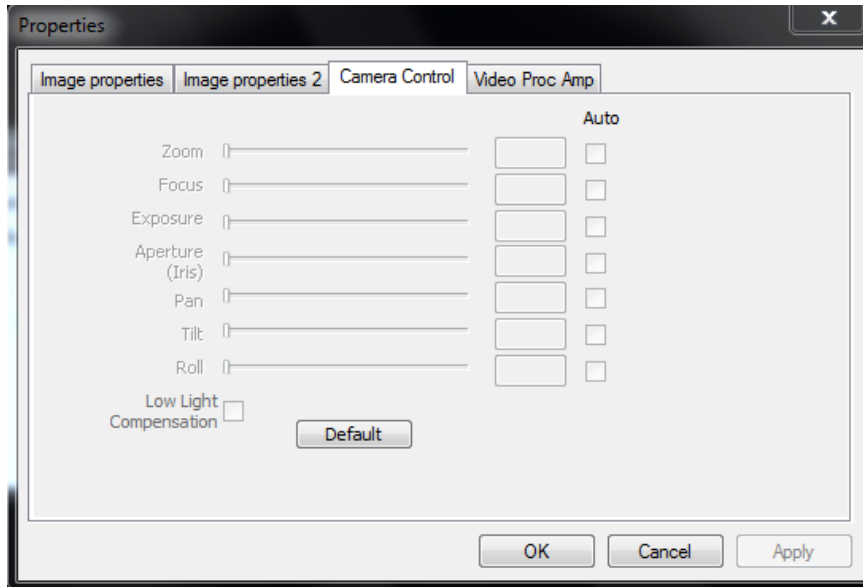


Figure A.7: "Camera control" tab of the properties window activated by clicking on "Video" button in the manual focus window.

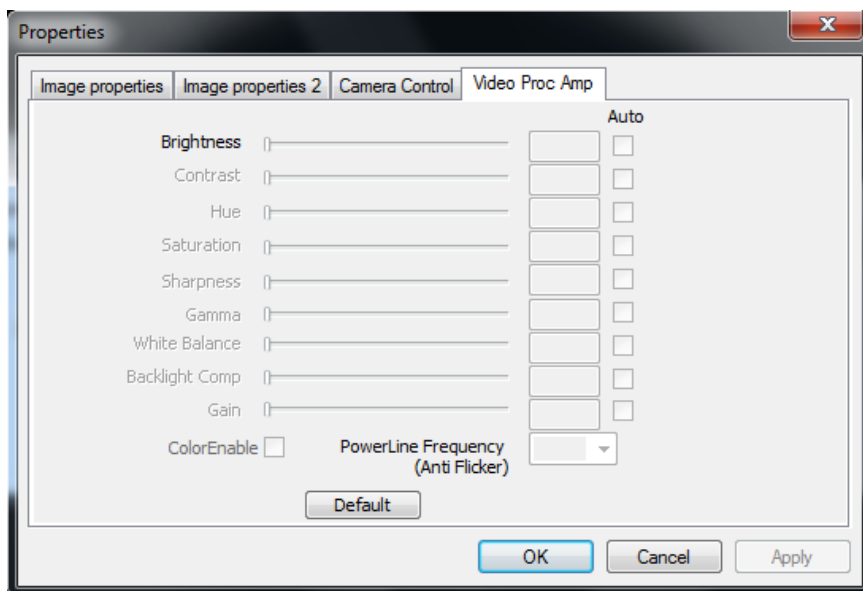


Figure A.8: "Video proc amp" tab of the properties window activated by clicking on "Video" button in the manual focus window.

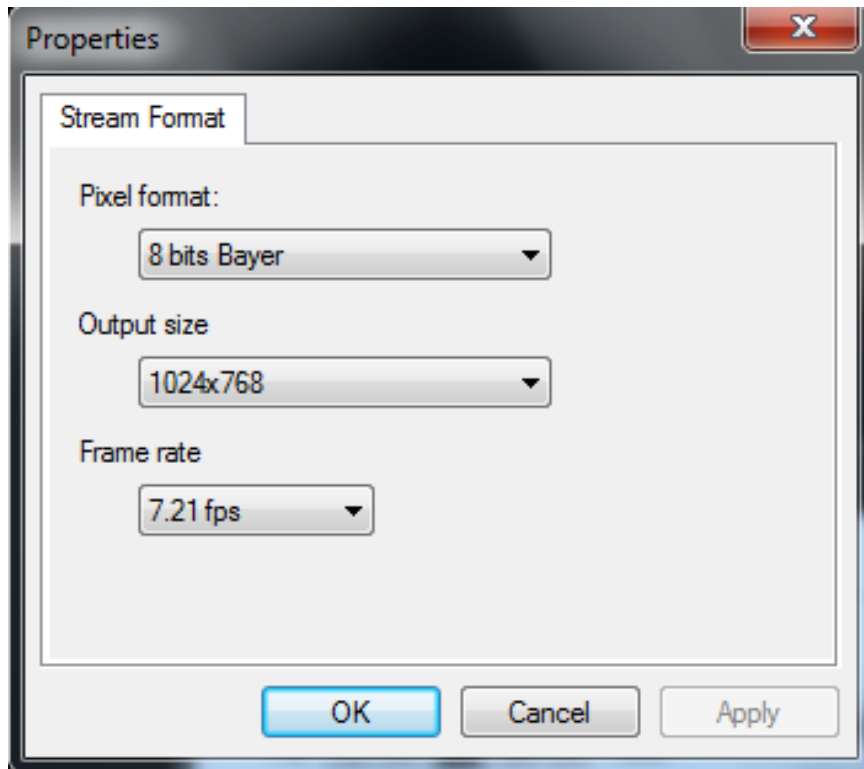


Figure A.9: The properties window activated by clicking on “Video format” button in the manual focus window.

necessary hardware devices. Note that before clicking on the button you need to ensure all the devices are powered and connected to the PC properly. Click on the same button again to stop the camera.















A.3.2 Snapshot

The snapshot feature is very useful for setting the camera parameters used for capturing data. Click on “Snapshot” button in the main window or manual focus window to take a snapshot of the current view of sensor of the main camera. A window containing the snapshot image will pop up. The snapshot uses the

parameters set in the capture control window (see Section A.2.2).

A.3.3 Manual Focus

To manually focus on an object, you need to do the following:

1. Click on “Manual focus” button in the main window to activate the manual focus window and video window of the main camera.
2. Click on “Video” button in the manual focus window and adjust settings including gain, brightness, contrast, and gamma to obtain proper exposure. You may need to turn on additional lights if the video is too dark. It is recommended to maximize the aperture during the focusing and turn it back afterwards.
3. Click on **H** button on both “Lens” and “Sensor” panel to send all the stages home. Ignore this step if the stages have been sent home previously.
4. Use move buttons (   ), step move buttons (   ) and stop button () on “Sensor” panel to move the sensor to the location to be focused. Click on the track bar to control the moving speed.
5. Use move buttons ( ), step move buttons ( ) and stop button () on “Lens” panel to focus. Click on the track bar to control the moving speed.
6. Perform previous two steps interchangeably till the object is in focus.

A.3.4 Calibrate Dark Current, White Images, Vignetting or White Balance

You can calibrate dark current, white images , vignetting and white balance using the calibration window. You need to do the following:

1. Prepare for the calibration, such as defocusing for white images calibration, shut off aperture for dark current calibration, or placing Macbeth chart in front of the camera for white balance calibration.
2. Open the calibration window by clicking on “Calibration” button in the main window.
3. Input the current temperature.
4. Input the start exposure time and stop exposure time.
5. Click “Save” button to start the calibration process. It may take few minutes. The calibration images will be saved in the project path specified in the capture control window.

A.3.5 Cameras Alignment

The software needs to align the main camera and the auxiliary camera by establishing the correspondence between both views. Without the proper alignment, the highlighted region in the viewfinder in the main window may be wrong and the final image may not capture the user selected region of interest. To align the cameras, you need to do the following:

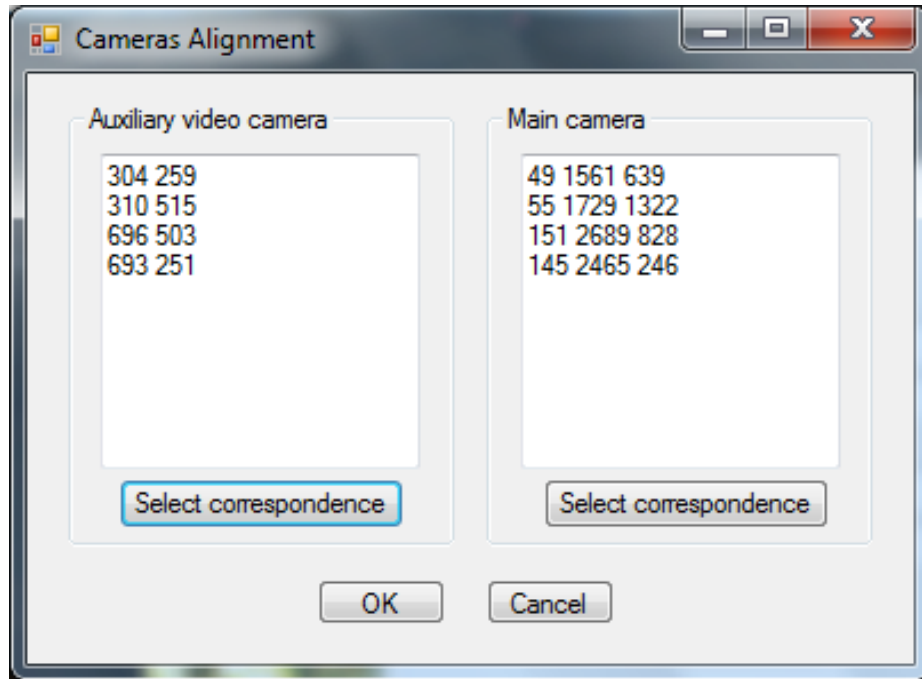


Figure A.10: The cameras alignment window allows you to align the main camera and the auxiliary camera by selecting four pairs of corresponding points from both views.

1. Click on "Calibration" button in the main window to activate the calibration window.
2. Click on "Start" button in the calibration window. The dialog box shown in Figure A.10 will pop up.
3. Select four points from the auxiliary video camera as following:
 - (a) Click on "Select correspondence" button in "Auxiliary video camera" panel. A window containing the current view of the auxiliary video camera will pop up (see Figure A.11).
 - (b) Select four points by clicking on four locations on the view window.



Figure A.11: Four points (marked as red) are selected in the view window of the auxiliary video camera.

- (c) Close the view window.
4. Select the corresponding four points from the main camera as following:
- (a) Open the manual focus window and move the sensor to the location contains the point corresponding to one of the points selected in previous step (see Section A.3.3).
 - (b) Click on “Align with grid” button in the manual focus window to ensure the sensor is aligned with the grid. This step is important for correct calibration. Note that after clicking on the button, the desired location may be out of the current view. Please move the sensor around by repeating previous step and click “Align with grid” again till the desired location is in the current view.

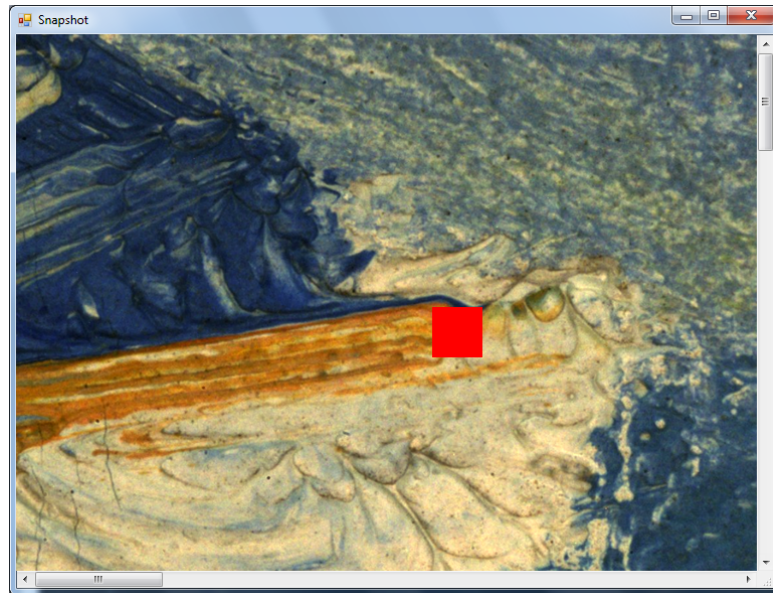


Figure A.12: One point (marked as red) is selected in the view window of the main camera. Note that the point corresponds to one of the points selected in the auxiliary video camera.

- (c) Click on “Select correspondence” button in “Main camera” panel. A window contains the current view of the main camera will pop up (see Figure A.12).
 - (d) Select the corresponding point by clicking on the location on the view window.
 - (e) Repeat previous three steps three times to select three more points. Note that the order of the points selected should be the same as the order of the points selected in step 3.
5. Click on “OK” button to calibrate cameras alignment.



Figure A.13: You could select the region to be captured by clicking and dragging on the viewfinder in the main window. The region to be captured will be highlighted by a blue dashed rectangle.

A.3.6 Capture Image

To capture a full image, you should do the following:

1. Select the region to capture by clicking and dragging on the viewfinder in the main window. The region to be captured will be highlighted by a blue dashed rectangle, as shown in Figure A.13. It may be necessary to align the cameras (see Section A.3.5) to ensure the region selected corresponds to the actual region to be captured. Note that if this step is omitted, the maximum region will be captured.
2. Set the parameters in capture control window if necessary. Use the snapshot

feature to help adjust the parameters (see Section A.3.2).

3. Click on “Capture Image” button to start the capturing process. It may take several minutes to complete. You can stop the process at any time by clicking on the same button again. Note that if the capturing process is stopped in the middle. Some images may be already saved.
4. When the capturing process completes, the set of images and a text file (“setting.txt”) containing the parameters used for capturing are stored in the location specified in the capture control window.

A.3.7 Fast Stitch

Click on “Stitch Image” button in the main window to select the folder containing the captured images and fast-stitch them. The stitching process uses the estimated grid location and may take several minutes to complete.

A.3.8 Focal Stacking

To create an all-in-focus image, you need to capture an image with focus stacking (see Section A.2.2). Then click on “Focal Stack” button in the main window to select the folder containing the captured images with focus stacking and create an all-in-focus image. The process may take several minutes to complete.

A.4 Summary

Congratulations! You have just completed reading the users manual for our digital large-format tile-scan camera. You should be ready to operate the camera now.

Bibliography

- [1] Adobe Photoshop. <http://www.photoshop.com> [visited on August 1, 2011].
- [2] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. *ACM Transactions on Graphics (SIGGRAPH'04)*, 23(3):294–302, 2004.
- [3] A. Agrawal, Y. Xu, and R. Raskar. Invertible motion blur in video. *ACM Transactions on Graphics (SIGGRAPH'09)*, 28(3):1–8, 2009.
- [4] Anagramm And Digital Reproduction. <http://www.linhofstudio.com> [visited on August 1, 2011].
- [5] Autodesk. Maya. <http://www.autodesk.com/maya> [visited on August 1, 2011].
- [6] S. Banerjee, P. Sastry, and Y. Venkatesh. Surface reconstruction from disparate shading: An integration of shape-from-shading and stereopsis. In *Proceedings of IAPR International Conference on Pattern Recognition (IAPR'92)*, 1992.
- [7] S. Barsky and M. Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1239–1252, 2003.

- [8] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, 2007.
- [9] J. Batlle, E. Mouaddib, and J. Salvi. Recent progress in coded structured light as a technique to solve the correspondence problem a survey. *Pattern Recognition*, 31(7):963–982, 1998.
- [10] M. Ben-Ezra. High resolution large format tile-scan camera design, calibration, and extended depth of field. In *International Conference on Computational Photography (ICCP'10)*, 2010.
- [11] M. Ben-Ezra. A digital gigapixel large-format tile-scan camera. *IEEE Computer Graphics and Applications*, 31(1):49–61, 2011.
- [12] F. Bernardini, H. Rushmeier, I. M. Martin, J. Mittleman, and G. Taubin. Building a digital model of Michelangelo’s Florentine Pieta. *IEEE Computer Graphics and Applications*, 22(1):59–67, 2002.
- [13] N. Birkbeck, D. Cobzas, P. Sturm, and M. Jagersand. Variational shape and reflectance estimation under changing light and viewpoints. In *Proceedings of European Conference on Computer Vision (ECCV'06)*, 2006.
- [14] D. Bradley, T. Boubekeur, and W. Heidrich. Accurate multi-view reconstruction using robust binocular stereo and surface meshing. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08)*, 2008.
- [15] M. Brown and D. Lowe. Recognising panoramas. In *Proceedings of IEEE International Conference on Computer Vision (ICCV'03)*, 2003.
- [16] M. Brown and D. Lowe. Automatic panoramic image stitching using invari-

- ant features. *International Journal of Computer Vision*, 74(1):59–73, 2007.
- [17] D. Capel and A. Zisserman. Automated mosaicing with super-resolution zoom. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'98)*, 1998.
- [18] C. Y. Chen, R. Klette, and C. F. Chen. Shape from photometric stereo and contours. In *Proceedings of Computer Analysis of Images and Patterns (CAIP'03)*, 2003.
- [19] E. Coleman and R. Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Computer Graphics and Image Processing*, 18(4):309–328, 1982.
- [20] T. Darrell and K. Wohn. Pyramid based depth from focus. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'88)*, 1988.
- [21] F. Deng, Z. Wu, Z. Lu, and M. S. Brown. Binarizationshop: a user-assisted software suite for converting old documents to black-and-white. In *Proceedings of Joint Conference on Digital Libraries (JCDL'10)*, 2010.
- [22] dgCam: a digital large-format gigapixel camera. <http://www.dgcam.org> [visited on August 1, 2011].
- [23] P. Dupuis and J. Oliensis. Direct method for reconstructing shape from shading. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'92)*, 1992.
- [24] O. Faugeras and R. Keriven. Variational principles, surface evolution, PDE's, level set methods, and the stereo problem. *IEEE Transactions on Image Processing*, 7(3):336–344, 1998.

- [25] G. Flint. Gigapxl project. <http://www.gigapxl.org> [visited on August 1, 2011].
- [26] R. T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):439–451, 1988.
- [27] P. Fua and Y. G. Leclerc. Using 3-dimensional meshes to combine image-based and geometry-based constraints. In *Proceedings of European Conference on Computer Vision (ECCV'94)*, 1994.
- [28] A. S. Georghiades. Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In *Proceedings of IEEE International Conference on Computer Vision (ICCV'03)*, 2003.
- [29] M. Goesele, B. Curless, and S. M. Seitz. Multi-view stereo revisited. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2006.
- [30] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. M. Seitz. Multi-view stereo for community photo collections. In *Proceedings of IEEE International Conference on Computer Vision (ICCV'07)*, 2007.
- [31] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and spatially-varying BRDFs from photometric stereo. In *Proceedings of IEEE International Conference on Computer Vision (ICCV'05)*, 2005.
- [32] G. Hausler. A method to increase the depth of focus by two step image processing. *Optics Communications*, 6(1):38–42, 1972.
- [33] C. Hernández and F. Schmitt. Silhouette and stereo fusion for 3d object

- modeling. *Computer Vision and Image Understanding*, 96(3):367–392, 2004.
- [34] C. Hernández, G. Vogiatzis, G. J. Brostow, B. Stenger, and R. Cipolla. Non-rigid photometric stereo with colored lights. In *Proceedings of IEEE International Conference on Computer Vision (ICCV'07)*, 2007.
- [35] C. Hernández, G. Vogiatzis, and R. Cipolla. Multi-view photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):548–554, 2008.
- [36] A. Hertzmann and S. M. Seitz. Shape and materials by example: a photometric stereo approach. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03)*, 2003.
- [37] A. Hertzmann and S. M. Seitz. Example-based photometric stereo: shape reconstruction with general, varying BRDFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1254–1264, 2005.
- [38] T. Higo, Y. Matsushita, N. Joshi, and K. Ikeuchi. A hand-held photometric stereo camera for 3D modeling. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'09)*, 2009.
- [39] B. Horn and M. Brooks. *Shape from shading*. MIT Press, 1989.
- [40] B. Horn, R. Woodham, and W. Silver. Determining shape and reflectance using multiple images. In *MIT AI Memo*, 1978.
- [41] I. Horovitz and N. Kiryati. Depth from gradient fields and control points: bias correction in photometric stereo. *Image and Vision Computing*, 22(9):681–694, 2004.
- [42] K. Ikeuchi. Determining a depth map using a dual photometric stereo. *Inter-*

national Journal of Robotics Research, 6(1):15–31, 1987.

- [43] K. Ikeuchi, K. Hasegawa, A. Nakazawa, J. Takamatsu, T. Oishi, and T. Masuda. Bayon digital archival project. In *Proceedings of Virtual Systems and Multimedia*, 2004.
- [44] K. Ikeuchi and B. K. P. Horn. Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, 17(1-3):141–184, 1981.
- [45] N. Joshi and D. J. Kriegman. Shape from varying illumination and viewpoint. In *Proceedings of IEEE International Conference on Computer Vision (ICCV'07)*, 2007.
- [46] H. Kawasaki, R. Furukawa, R. Sagawa, and Y. Yagi. Dynamic scene shape reconstruction using a single structured light pattern. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08)*, 2008.
- [47] Kodak. <http://www.kodak.com> [visited on August 1, 2011].
- [48] Konica Minolta Range 7 laser scanner. <http://www.konicaminolta.com> [visited on August 1, 2011].
- [49] H. Lange. Advances in the cooperation of shape from shading and stereo vision. In *Proceedings of IEEE International Conference on 3-D Digital Imaging and Modeling (3DIM'99)*, 1999.
- [50] K. Lee and C. Kuo. Shape reconstruction from photometric stereo. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'92)*, 1992.
- [51] M. Levoy. The digital Michelangelo project. In *Proceedings of IEEE International Conference on 3-D Digital Imaging and Modeling (3DIM'00)*, 2000.

- [52] J. Lim, J. Ho, M. H. Yang, and D. J. Kriegman. Passive photometric stereo from motion. In *Proceedings of IEEE International Conference on Computer Vision (ICCV'05)*, 2005.
- [53] Z. Lu, Y. W. Tai, M. Ben-Ezra, and M. S. Brown. A framework for ultra high resolution 3D imaging. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'10)*, 2010.
- [54] Z. Lu, Z. Wu, and M. S. Brown. Directed assistance for ink-bleed reduction in old documents. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'09)*, 2009.
- [55] Z. Lu, Z. Wu, and M. S. Brown. Interactive degraded document binarization: an example (and case) for interactive computer vision. In *IEEE Workshop on Applications of Computer Vision (WACV'09)*, 2009.
- [56] Luminera Corperation. <http://www.lumenera.com> [fisited on August 1, 2011].
- [57] B. Lutz and M. Weintke. Virtual Dunhuang art cave: a cave within a CAVE. *Computer Graphics Forum*, 18(3):257–264, 1999.
- [58] A. S. Malik and T. S. Choi. A novel algorithm for estimation of depth map using image focus for 3d shape recovery in the presence of noise. *Pattern Recognition*, 41(7):2200–2225, 2008.
- [59] K. Martinez, J. Cupitt, D. Saunders, and R. Pillay. Ten years of art imaging research. *Proceedings of the IEEE*, 90(1):28–41, 2002.
- [60] MIDA: Mellon Interational Dunhuang Archive. <http://www.artstor.org/what-is-artstor/w-html/col-mellon-dunhuang.shtml>

- [visited on August 1, 2011].
- [61] D. Milgram. Computer methods for creating photomosaics. *IEEE Transactions on Computers*, C-24(11):1113–1119, 1975.
- [62] D. Milgram. Adaptive techniques for photomosaicking. *IEEE Transactions on Computers*, 26(11):1175–1180, 1977.
- [63] D. Miyazaki, T. Oishi, T. Nishikawa, R. Sagawa, K. Nishino, T. Tomomatsu, Y. Takase, and K. Ikeuchi. The great Buddha project: Modeling cultural heritage through observation. In *Modeling from Reality*, volume 640 of *The Kluwer International Series in Engineering and Computer Science*, pages 181–193. Springer US, 2002.
- [64] S. K. Nayar, K. Ikeuchi, and T. Kanade. Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE Transactions on Robotics and Automation*, 6(4):418–431, 1990.
- [65] S. K. Nayar and Y. Nakagawa. Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):824–831, 1994.
- [66] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. *ACM Transactions on Graphics (SIGGRAPH'05)*, 24(3):536–543, 2005.
- [67] R. Onn and A. Bruckstein. Integrability disambiguates surface recovery in two-image photometric stereo. *International Journal of Computer Vision*, 5(1):105–113, 1990.
- [68] M. Oren and S. K. Nayar. Generalization of Lambert’s reflectance model. In *Proceedings of ACM SIGGRAPH'94*, 1994.
- [69] Osram. <http://www.osram.com> [visited on August 1, 2011].
- [70] S. Paris, F. X. Sillion, and L. Quan. A surface reconstruction method using global graph cut optimization. *International Journal of Computer Vision*, 66(2):141–161, 2006.

- [71] N. Petrovic, I. Cohen, B. J. Frey, R. Koetter, and T. S. Huang. Enforcing integrability for surface reconstruction algorithms using belief propagation in graphical models. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'01)*, 2001.
- [72] Phidgets Inc. <http://www.phidgets.com> [visited on August 1, 2011].
- [73] Point Gray Research Inc. <http://www.ptgray.com> [visited on August 1, 2011].
- [74] J. P. Pons, R. Keriven, and O. Faugeras. Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. *International Journal of Computer Vision*, 72(2):179–193, 2006.
- [75] B. Potetz. Efficient belief propagation for vision using linear constraint nodes. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)*, 2007.
- [76] J. K. Reid and J. A. Scott. An out-of-core sparse cholesky solver. *ACM Transactions on Mathematical Software*, 36(2):1–33, 2009.
- [77] F. Sartori and E. R. Hancock. An evidence combining approach to shape-from-shading. In *Proceedings of International Conference on Pattern Recognition (ICPR'02)*, 2002.
- [78] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(1-3):7–42, 2002.
- [79] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03)*, 2003.
- [80] Schneider Optics. <http://www.schneideroptics.com> [visited on August 1, 2011].

- [81] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2006.
- [82] S. N. Sinha, P. Mordohai, and M. Pollefeys. Multi-view stereo via graph cuts on the dual of an adaptive tetrahedral mesh. In *Proceedings of IEEE International Conference on Computer Vision (ICCV'07)*, 2007.
- [83] F. Solomon and K. Ikeuchi. Extracting the shape and roughness of specular lobe objects using four light photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(4):449–454, 1996.
- [84] C. Strecha, R. Fransens, and L. Van Gool. Combined depth and outlier estimation in multi-view stereo. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2006.
- [85] R. Szeliski and H. Y. Shum. Creating full view panoramic image mosaics and environment maps. In *Proceedings of ACM SIGGRAPH'97*, 1997.
- [86] R. Szeliski. Image alignment and stitching: a tutorial. *Foundations and Trends in Computer Graphics and Vision*, 2(1):1–104, 2006.
- [87] H. D. Tagare and R. J. P. deFigueiredo. A theory of photometric stereo for a class of diffuse non-Lambertian surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):133–152, 1991.
- [88] P. Tan, S. Lin, and L. Quan. Subpixel photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(8):1460–1471, 2008.
- [89] D. Terzopoulos. The computation of visible-surface representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):417–438, 1988.
- [90] Thorlabs. <http://www.thorlabs.hk> [visited on August 1, 2011].

- [91] J. Tonry, P. Onaka, B. Burke, and G. Luppino. Pan-STARRS and gigapixel cameras. *Astrophysics and Space Science Library*, 336(1):53–62, 2006.
- [92] UNESCO. Mogao Caves, 1987. <http://whc.unesco.org/en/list/440> [visited on August 1, 2011].
- [93] D. Vlastic, P. Peers, I. Baran, P. Debevec, J. Popović, S. Rusinkiewicz, and W. Matusik. Dynamic shape capture using multi-view photometric stereo. *ACM Transactions on Graphics (SIGGRAPH-ASIA'09)*, 28(5):1–11, 2009.
- [94] G. Vogiatzis, C. Hernández, and R. Cipolla. Reconstruction in the round using photometric normals and silhouettes. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2006.
- [95] P. Vuylsteke and A. Oosterlinck. Range image acquisition with a single binary-encoded light pattern. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(2):148–164, 1990.
- [96] S. Wang and W. Heidrich. The design of an inexpensive very high resolution scan camera system. *Computer Graphics Forum*, 23(3):441–450, 2004.
- [97] C. Whöler. *3D computer vision: efficient methods and applications*. Springer, 2009.
- [98] J. Wilhelmy and J. Krüger. Shape from shading using probability functions and belief propagation. *International Journal of Computer Vision*, 84(3):269–287, 2009.
- [99] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.
- [100] P. L. Worthington and E. R. Hancock. New constraints on data-closeness and needle map consistency for shape-from-shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(12):1250–1267, 1999.
- [101] T. P. Wu, J. Sun, C. K. Tang, and H. Y. Shum. Interactive normal reconstruction from

- a single image. *ACM Transactions on Graphics (SIGGRAPH-ASIA'08)*, 27(5):1–9, 2008.
- [102] Y. Xiong and S. Shafer. Depth from focusing and defocusing. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'93)*, 1993.
- [103] M. Young, E. Beeson, J. Davis, S. Rusinkiewicz, and R. Ramamoorthi. Viewpoint-coded structured light. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)*, 2007.
- [104] Zaber Technologies. <http://www.zaber.com> [visited on August 1, 2011].
- [105] L. Zhang, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and motion under varying illumination: unifying structure from motion, photometric stereo, and multi-view stereo. In *Proceedings of IEEE International Conference on Computer Vision (ICCV'03)*, 2003.
- [106] Q. Zheng and R. Chellappa. Estimation of illuminant direction, albedo, and shape from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):680–702, 1991.