Generalization of Photometric Stereo: Fusing Photometric and Geometric Approaches for 3-D Modeling

照度差ステレオの一般化：
3次元モデル化のための光学的手法と幾何学的手法の融合

BY

Tomoaki Higo
肥後 智昭

A Doctoral Dissertation

Submitted to the Graduate School of
the University of Tokyo

in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Information Science and Technology

December 2010

Committee:

Kiyoharu AIZAWA (Chair)
Keikichi HIROSE
Jun ADACHI
Shunsuke KAMIJO
Takeshi NAEMURA

Supervisor:

Katsushi IKEUCHI

ABSTRACT

Three-dimensional (3-D) models created by computer vision and computer graphics are widely used in many important applications and for a variety of purposes. However, creating 3-D models manually involves significant costs. Therefore, automatic generation of such models is attracting considerable interest, and this process in turn requires efficient 3-D scanning.

Numerous systems exist for 3-D scanning using a laser range sensor and image-based methods such as multi-view stereo, structured light, and photometric stereo. These are classified into two approaches: a photometric approach and a geometric approach. Generally, the photometric approach, as typified by photometric stereo, estimates surface normals to be suitable for representing fine details of the surfaces. On the other hand, the geometric approach recovers a depth of the target object to acquire a rough shape.

This dissertation proposes efficient 3-D modeling methods that combine photometric and geometric approaches. Combining photometric stereo with a laser range sensor or multi-view stereo allows us to introduce practical constraints for 3-D modeling. Also, we propose practical methods of photometric stereo for handling object with various reflection properties.

The first method fuses a laser range sensor and a camera with an attached camera flash for 3-D modeling. The laser range sensor captures the basic shape of the target object. Meanwhile, photometric stereo estimates surface normals as a bump map for detailed surfaces. Then accurate reflection parameters are estimated by using the surface normals.

The second method combines photometric stereo and multi-view stereo to simultaneously estimate shape and surface normals. The method uses a simple configuration with a camera and a camera flash. Furthermore, by using color information, the method is extended to robustly handle specularities and occlusions.

Generally, photometric stereo assumes that the target object has only diffuse reflection. However, many materials have specular reflection components, which cause an estimation error in photometric stereo. To overcome this problem, a real-time specular removal method is proposed. With a known light source color, the method enables the removal of specular reflection components faster than conventional methods by using a defined color space.

In addition, a new photometric stereo method is proposed to handle a wide range of surface reflectances. The method avoids imposing restricting assumptions on surface

reflectances and expands the applicability of photometric stereo using three reflection properties. Moreover, our method eliminates the necessity of radiometric calibration and any dependency on ambient illumination.

In this dissertation, the theory of these methods is presented, and qualitative and quantitative experiments are performed to demonstrate the effectiveness of each method.

論文要旨

近年コンピュータビジョンやコンピュータグラフィックスの技術を駆使した3次元モデルは,「3D元年」という言葉にも代表されるように，多くのアプリケーションや様々な目的のために幅広く利用されてきている．しかし3次元モデルを人手で作るコストや手間は膨大であるため，3次元モデルを自動で生成しようとする試みに関心が高まってきているとともに，効果的な3次元計測手法が必要とされてきている．

3次元モデルを計測する方法には，レーザーレンジセンサを用いたり，画像ベースで推定をおこなう多視点ステレオ，光切断法，照度差ステレオなどの様々な手法が存在する．これらの手法は大きく分けて光学的手法と幾何学的手法に分けることができる．一般的に光学的手法は，照度差ステレオに代表されるように，物体表面の法線を推定する手法であり，細かな凹凸や滑らかな面を推定するのに適している．一方で幾何学的手法は対象物体の奥行きを計測する手法であり，物体の外形を推定するのに適している．

本論文ではこれらの相反する光学的手法と幾何学的手法を組合せることで，より効果的な3次元モデル化手法の提案をおこなう．具体的には，光学的手法である照度差ステレオと、レーザーレンジセンサや、多視点ステレオとを組み合わせることで，新たな拘束を導いて3次元モデル化をおこなう．さらに，様々な反射特性の物体に対して照度差ステレオを扱えるようにするための手法を提案する．

1つ目の手法はレンジセンサとフラッシュ付きのカメラを融合した3次元モデル化手法である。大まかな形状をレーザレンジセンサで取得し，細かな凹凸などは照度差ステレオによって法線を求め，バンプマップとして表現する．さらに得られた法線を用いて正確な反射パラメータを推定する．

2つ目の手法は照度差ステレオと多視点ステレオを組合せることで，法線と形状を同時に推定する手法である．本手法はレーザーレンジセンサを使わず，カメラとフラッシュだけの簡易なセットアップで3次元モデルを推定することができる．さらに色情報を利用することによって本手法を拡張し，鏡面反射や遮蔽にロバストな手法を提案する．

一般的に照度差ステレオを用いる場合は，対象物体が拡散反射であることを仮定する．しかし物体の中には鏡面反射成分が含まれる物体も数多く存在し，これが推定に悪影響を与えてしまう，そこでリアルタイムに鏡面反射成分を除去する手法を提案する．本手法は光源色を既知として独自の色空間を用いることで，従来手法よりも高速に鏡面反射成分を除去することができる．

さらに，鏡面反射成分だけでなく，幅広い反射特性の物体の法線を推定するための手法を提案する．特定の反射モデルを仮定せず，基本的な反射特性から導いた拘束を用いて法線を推定する．本手法はカメラの特性や環境光の影響に対してもロバストである．

本論文では，これらの手法の原理を示し，定性的及び定量的な実験を通してそれぞれの手法の実用性を示す．

Acknowledgements

I would first like to express my gratitude to my advisor, Prof. Katsushi Ikeuchi, for sharing his vast knowledge through discussions and for the support and encouragement. He gives me the freedom to follow my interests, and the guidance to keep me on track. I have learned from him how to enjoy doing research in computer vision.

I would also like to thank my committee members, Prof. Kiyoharu Aizawa, Prof. Keikichi Hirose, Prof. Jun Adachi, Prof. Shunsuke Kamijo, and Prof. Takeshi Naemura for giving valuable advice on this dissertation. I would like to thank Prof. Toshihiko Yamasaki for his valuable feedback at the adviser meetings.

I would also like to express my deepest gratitude to Dr. Yasuyuki Matsushita, my mentor at Microsoft Research Asia. I was very fortunate to have him as my mentor; his advice, support, and insightful comments were very valuable and significant for the improvement of my capabilities. I also wish to thank other researchers, Dr. Neel Joshi, Dr. Moshe Ben-Ezra, and Dr. Bennett Wilburn for the interesting discussions we had about research in the field of computer vision. My good friends at Microsoft Research Asia, Hyeongwoo Kim, Yuki Arase, Nan Huang, Yuji Oyamada, and Boxin Shi, also deserve special thanks for inspiring me with many good ideas and made me relax.

Many thanks go to my colleagues at the Computer Vision Laboratory at the University of Tokyo. My special thanks go to Daisuke Miyazaki and Rei Kawakami for helping me achieve a good start, giving me many interesting ideas, and holding important discussions with me, and my thanks go to Koichi Ogawara, Jun Takamatsu, Takeshi Oishi, Shunsuke Kudoh, Atsuhiko Banno, Shintaro Ono, Akifumi Ikari for instructing me in the earlier years of my study. I am also very proud of, and feel fortunate to have worked with, the talented people in the Photometry group. I would also like to thank Masataka Kagesawa, Kiminori Hasegawa, Keiko Motoki, Yoshiko Matsuura, Mikiko Yamaba, Kaoru Kikuchi, and Yoshihiro Sato for their constant and warm support.

Although, due to limited space, I cannot name everyone who has helped me, I am very grateful to all the people I have met and interacted with in this lab.

I am also very grateful to Joan Knapp and Robert Knapp for spending considerable time proofreading my English manuscripts. Their kind support enabled me to complete my thesis.

Finally, I would like to thank my friends and my family for their patience throughout my long-term student life. I could not have finished this work without their strong support.

December 2010

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Three-dimensional (3-D) models created by computer vision and computer graphics are widely used in many important applications in archeology, medicine, and in the film and video game industries. These models allow visualization in the fields of research, education, and entertainment and thus present many possibilities for progress. Currently, most 3-D models are reconstructed by manual operation [Aut, Goo]. However, manual creation of 3-D models involves significant cost, and therefore more sophisticated techniques for modeling a target object are needed for supplying 3-D data at lower cost. Thus, automation for creating 3-D models has attracted considerable interest as the need for such models has increased, and the creation of these models requires efficient systems for 3-D scanning.

Numerous systems exist for 3-D scanning using a laser range sensor and image-based methods such as multi-view stereo, structured light, and photometric stereo. These are classified into two approaches: a photometric approach and a geometric approach. Generally, the photometric approach estimates surface normals, while the geometric approach recovers a depth of the target object.

As a photometric approach for 3-D modeling, photometric stereo is well known as a means of estimating surface normal from an image sequence taken from a fixed viewpoint under varied directional lighting. Estimated surface normals are useful in representing fine details of the target object, such as a bumpy surface vs. a smooth surface. Photometric stereo has a long history. After the early work of Woodham

[Woo80] and Silver [Sil80], many researchers studied the approach to make it work under more generalized conditions, such as with specularity [Ike81, CJ82, SW93, SI96, BP03, WTTW06, VVG08, TLQ08], and with shadows [CAK07, HVC08b, OSS09]. Other research included example-based approaches for general reflectance [HI84, Ike87, HS05, AZK08], and approaches under such conditions as uncalibrated lighting [Hay94, Geo03, SOYS07, SMW*10, SZP10], a near light source [IWTI94, CP99], and color lighting [BP01, HVB*07, KWBE10].

The geometric approach also has a long history. Many methods for depth acquisition have been proposed. The use of a laser range sensor [BM92, CL96, LPC*00, HH03, INHO03, IHN*04, MKH*06] allows us to directly acquire a depth map, but this approach has some drawbacks. The laser range sensor is a time-of-flight scanning sensor; therefore, it requires more measurement time than image-based methods. Its accuracy is limited since range sensors depend on step intervals of mechanical scanning. Moreover, its cost is high; in the $100K range. One of the most popular image-based methods is multi-view stereo [OK93, SD99, KS00, PVGV*04]. Multi-view stereo estimates positions of feature points by using a triangulation method. Most early works in multi-view stereo tended to reconstruct all scene points independently. In recent years, various approaches typically cast this as a variation problem: depth map merging based methods [GCS06, BBBH08, LCDX09, LLC*10], featured-region growing and expansion based methods [LQ05, FP07, GSC*07, LPK07, BBBH08], 3-D volumetric based methods [HK06, SMP07, VHTC07], graph based approaches [HK06, VHTC07], methods for large-scale reconstruction [SSS06, SSS07, WCL*08, MK09, FCSS10], real-time methods [PNF*08, ND10], anisotropic metric [KPC10], and closed-form solution [WYJT10]. An excellent survey of most of these approaches can be found in Seitz [SCD*06, Mid]. Another image-based method is structured light [BK87, SBM98, HHR02, DNRR05, KVG06, YBD*07, AX08, KFSY08, YX10]. Structured light methods actively generate geometric correspondences between projectors and cameras. These geometric approaches are good for position estimation but not good at characterizing fine details of the surface (such as smooth vs. bumpy) because neighboring points pose no constraints regarding surface continuity.

This dissertation attempts to generalize photometric stereo especially for estimating accurately and efficiently a 3-D model with fusion of photometric and geometric approaches. The photometric approach estimates surface normals, while the geometric approach estimates the position or depth of the target object. Moreover, we introduce physics-based reflection models to estimate reflection properties of the target object as

well as the shape. Obtaining both shape and reflection properties could be very useful for the realistic rendering in movies and video games and for use in a digital museum, because the proposed process can reconstruct an appearance of the target object from any viewpoint under any illumination conditions.

To obtain these properties, our approach fuses photometric stereo with a geometric approach. This is useful because in standard photometric stereo, input images taken from a fixed viewpoint by definition contain feature correspondence, but multi-view images are a challenge because of the difficulty in finding these correspondences. On the other hand, geometric approaches usually begin with multi-view images, so this information can be used to determine the needed image correspondence so that varying illumination conditions can be handled; *i.e.*, so that photometric stereo can be used to satisfactorily process images from different viewpoints under different illuminations. Moreover, since photometric stereo often assumes a very simple model and an ideal case, we have to provide a more robust solution that will apply in real-world cases.

1.2 Research Objective

This dissertation describes two research objectives: one is to estimate accurate 3-D models with fine details of the target object and the other is to overcome various limitations of photometric stereo. Estimating accurate 3-D models is important for visualization in the fields of research, education, and entertainment. Automation for creating 3-D models has attracted much interest since manual operation for 3-D models involves significant cost. Generally, to make a 3-D model, a laser range sensor is used to acquire shape information, and a textured image taken with a camera is mapped onto the surface. However, in such a process, the problems are that using a laser range sensor makes it difficult to measure fine details such as a bumpy surface, and that texture mapping with a particular image only visualizes an appearance under the same illumination condition in which the image was taken. To solve these problems, this dissertation proposes the careful combination of photometric stereo and the geometric approach.

First we propose a method that fuses a laser range sensor and a camera with an attached camera flash to estimate the 3-D shape and reflection parameters. The fusion of these sensors gives us new constraints for efficient estimation. In this method, we use the laser range sensor to acquire the basic shape of the target object, and use multi-view photometric stereo to estimate surface normals as a ”bump map” for real-world

surfaces. Accurate reflection properties are robustly estimated with the normal map and clustering.

The second method we propose only uses a camera and an attached light source without any range sensors. It is a very simple configuration for 3-D modeling. This method simultaneously estimates depth and surface normals by combining multi-view stereo and photometric stereo. Furthermore, by using color information and a view constraint, we extend this method for robust estimation despite specularities and occlusions.

For handling specular reflection in the above methods, we propose a real-time method for specular removal. Based on the dichromatic reflection model and the neutral interface reflection assumption, we define our color space and quickly remove specular components from one image with known illumination color. Moreover, we can quickly generate a specular-free image that is appropriate for input to photometric stereo.

Thus, by fusing photometric stereo and geometric information, this dissertation proposes methods to extend photometric stereo not only for estimation of surface normals but also for estimation of shape and for handling specularities.

The assumption of diffuse Lambertian reflection and specularity is, however, not appropriate for all materials. Therefore, we propose a photometric stereo method that works with a wide range of surface reflectances using an image sequence taken from a fixed viewpoint under different directional lighting. Instead of assuming a specific parametric reflectance model, such as Lambertian, we assume only three reflectance properties that are often observed in real-world scenes: monotonicity, visibility, and isotropy. Each of these three properties independently gives a possible solution space of the surface orientation. By taking the intersection of the solution spaces, our method determines the surface normal in the form of a consensus. In addition, our method eliminates the necessity of radiometric calibration and has no dependency on the ambient illumination.

1.3 Thesis overview

Chapter 2 describes a method that estimates shape and reflection parameters of the target object by fusing a laser range sensor and a camera. The chapter shows that the sensor fusion provides efficient constraints for 3-D modeling. Estimation of surface normal on the surface of an object can be used to represent fine details of the object's

appearance and to obtain accurate reflectance properties. After briefly reviewing the previous works and the setup of our proposed method, the chapter explains estimation for geometric data, such as a basic shape and normal map. Surface normals are estimated with multi-view photometric stereo using near light source. Then, robust estimation for reflection property is presented, and experimental results on both synthetic and real data are provided. Finally, we summarize our proposed method.

Chapter 3 proposes a simple configuration method that estimates shape and reflectance of the target object by combining photometric stereo and multi-view stereo. After reviewing previous work, we first describe near-light photometric stereo, specular removal, both an intensity-based method and a color-based method, and a more efficient formulation of the algorithms. We show that using color information extends the method to be more efficient and robust. We explain simultaneous estimation of depth and surface normals as well as refinement of the surface shape. We then describe an implementation of our whole system and full 3-D reconstruction. Experimental results using simulation and real-world scenes are provided. Finally, the chapter provides discussion and a summary.

Chapter 4 describes a real-time method to remove specular components from an input image taken under uniform illumination. We first describe color properties and our color space, and then explain white balance correction. The color space is based on the dichromatic reflection model and the neutral interface reflection assumption. We present and prove the theory of our proposed method for specular removal from the color space. After that, we describe the algorithm of our real-time specular removal system and provide experimental results. We discuss the limitations of our method, and summarize.

Chapter 5 proposes a photometric stereo method that works with a wide range of surface reflectances by assuming three reflectance properties – monotonicity, visibility, and isotropy – instead of assuming a specific parametric reflectance model. After reviewing previous work, we explain our proposed consensus approach to limit the solution space for a surface normal. We then describe its implementation using a voting method and describe how the consensus approach can be turned into an energy minimization scheme for an especially efficient implementation. After the implementation details, we describe the theoretical relationship between the number of lighting directions and the accuracy of surface normals. We also describe experimental validations using simulation and real-world images. Finally, we provide discussion and a summary.

Chapter 6 concludes this dissertation by summarizing the research and discussing possible future research directions.

Chapter 2

Efficient Estimation and Representation of 3-D model with Sensor Fusion

Estimation of surface normal on the surface of an object can be used to represent fine details of the object's appearance and to obtain accurate reflectance properties. This chapter provides an efficient 3-D modeling method with *sensor fusion* of a laser range sensor and a camera. The novelty of the method is efficient estimation and representation of the 3-D model. Detailed surfaces can be estimated as a normal map using multi-view photometric stereo on the basic shape measured with the laser range sensor. Accurate reflection properties are robustly estimated with the normal map and clustering.

Experimental results show that realistic 3-D models are obtained. From thorough verification of a normal map and robust estimation, our method can represent the fine appearance and estimate accurate reflection parameters with a small number of input images.

2.1 Introduction

Three-dimensional models created by computer vision and graphics techniques are used in a wide variety of areas, such as mechanical, medical, and architectural industries, and for a variety of purposes, such as visualization in the fields of research, entertainment, and education. Automation for creating 3-D models has also attracted considerable interest as the need for such models has increased. Currently, most 3-D

models are reconstructed by manual operation [Aut, Goo], causing a significant increase in cost, and therefore more sophisticated techniques for modeling a target object are needed for supplying 3-D data at lower cost. For this purpose, a number of methods that reconstruct 3-D shapes with sensors have been developed, such as a laser range sensor. In general, however, it is difficult to measure surface details with a laser range sensor. A small interval between scan acquisitions is costly. Furthermore, it is still difficult to measure surface details with a small scan interval. Meanwhile, for the accurate appearance of the target object, we have to know not only the object's shape but also its surface reflectance properties. Once we get the reflectance properties, we can simulate the appearance of the object under any illumination. For estimating surface reflectance properties, we require object appearances with known shape and known illumination conditions. As shape information, surface orientation is especially important for the estimation.

In this chapter, we propose a new method for 3-D modeling with sensor fusion. Combination of a laser range sensor and a camera provides efficient constraints for 3-D modeling. Basic shape is measured by a laser range sensor and detailed surface is represented as a normal map estimated with multi-view photometric stereo. The normal map also achieves an accurate estimation of surface reflectance. Moreover, clustering and robust estimation are used for estimation of specular reflection parameters.

2.2 Related Work

Many methods that estimate 3-D models have been proposed. Here, we address three related works that use both a laser range sensor and a camera.

Sato *et al*. [SWI97] used surface normal to estimate reflection parameters. They calculated an eigenvector of nearby sampling 3-D points from a range image taken with the laser range sensor, and used it as surface normal to estimate reflection parameters. This method is very effective to obtain a smooth surface normal; however, difficulties occur when representing the details of a bumpy surface. Many input images are required to separate diffuse and specular components for estimation of reflection parameters.

Nehab *et al*. [NRDR05] proposed a method that refines shape data from a laser range sensor with estimated surface normal by using photometric stereo. Their acquisition was from a fixed viewpoint for photometric stereo, so only one aspect could be refined. Moreover, they did not estimate any reflection properties.

Lensch *et al*. [LKG*03] obtained surface normal by estimating reflection properties of the target object using a non-linear optimization algorithm and refined shape measured with a laser range sensor. They clustered regions that have similar reflection properties using a tree structure, and then estimated reflection parameters on the clustered regions. However, this method had various constraints for measurement; for example, they required very accurate initial shape and some mirror spheres for light direction estimation. Since they simultaneously estimated surface normal and reflection parameters, non-linear optimization might converge to the local minimum.

In our method, we fuse the laser range sensor and the camera with the camera flash. These are relatively fixed to each other, and this helps to get light source positions and to achieve near light photometric stereo for accurate estimation of the surface normal. Unlike previous approaches, we can handle a near light source as well as reflection with specularities. Since calibration of the laser range sensor and the camera provides correspondence among multi-view images through shape, the number of required input images is only one for each viewpoint for multi-view photometric stereo. Using bump mapping with estimated surface normals represents details of the target geometric model. Therefore, acquired low-resolution shape data *e.g.*, basic shape, with the laser range sensor is acceptable. We can then effectively acquire input data. Furthermore, using our estimated geometric model, we can estimate reflection parameters accurately and robustly.

The rest of this chapter is structured as follows. We first describe the setup of our proposed method in Section 2.3. We present robust estimation for the shape and the reflection parameters in Section 2.4 and Section 2.5. Then we present results in Section 2.6 followed by discussion and conclusions.

2.3 Setup of Our Proposed Method

For data acquisition, we use a laser range sensor, a camera, and a camera flash. The laser range sensor and the camera are relatively fixed. Using a reference object whose shape is known, we first calculate camera intrinsic and extrinsic parameters that represent projection from 3-D world coordinates to 2-D image coordinates. We also measure relative position between the camera and the camera flash. Moreover, we acquire light intensity E_C (C = R,G,B) of the camera flash by capturing white reference under only the flash light. These calibrations are done before data acquisition.

We measure a target object by moving the target object or the entire setup for multi-

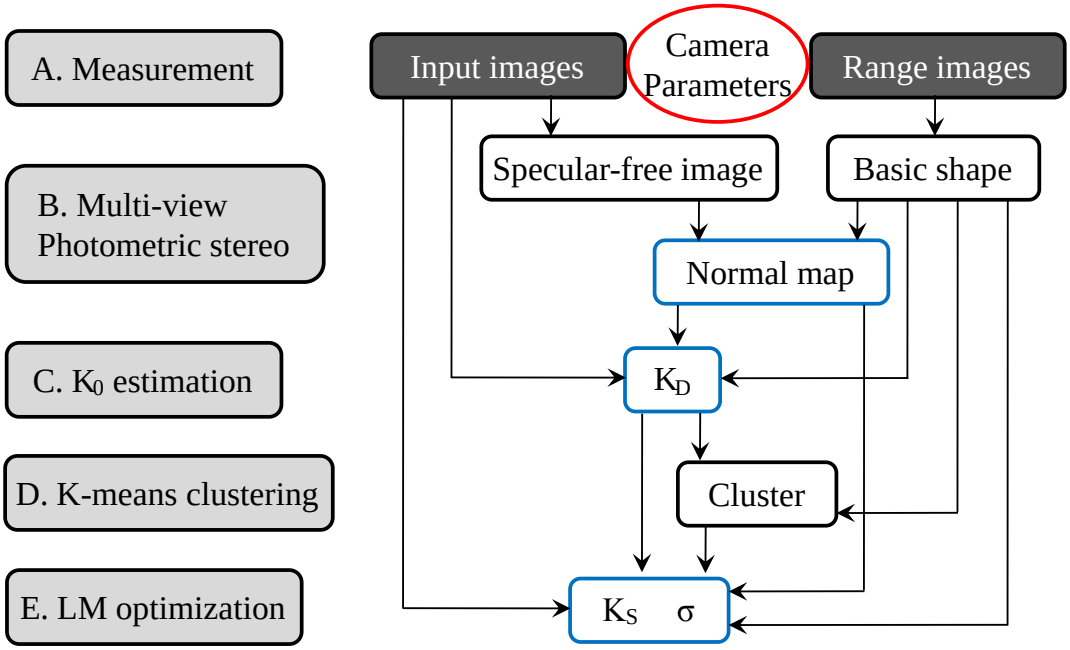


Figure 2.1: Flowchart of our proposed method.

view acquisition. We capture an image sequence with the camera and measure range data with the laser range sensor. In case ambient lighting exists, we capture the target object under both flash light on and flash light off, and then we extract the difference for acquiring an image sequence affected from only flash lighting.

In our method, we remove shadow pixels from the estimation by simple thresholding. Since the camera flash illuminates a target object near the camera viewpoint, shadows are observed in few areas.

Fig. 2.1 shows a flowchart of our estimation for shape and reflection properties. First, (A) our setup captures a 2-D image sequence and corresponding range images. Then a basic shape is made from the range images as shown in subsection 2.4.1. Second, (B) multi-view photometric stereo is applied to estimate the normal map that is mapped onto the surface of the basic shape from subsection 2.4.2 to 2.4.4. Third, (C) using estimated geometric information, diffuse reflection parameters are estimated in subsection 2.5.1. Then, (D) clustering is applied in subsection 2.5.2. Finally, (E) specular reflection parameters are estimated in each clustered segment in subsection 2.5.3.

2.4 Estimation for Geometric Data

2.4.1 Acquision for Basic Shape with Laser Range Sensor

Data acquired with a laser range sensor is a range image that discretely represents surface depth. We first align several range images measured from different viewpoints, and then merge them to get surface shape as a polygon mesh. Then we reduce the number of polygon faces as appropriate to make it a basic shape. A smaller size of data for the basic shape improves computational costs for 3-D rendering and other 3-D contents.

As an advantage, the smaller the number of polygon faces for the basic shape is, the smaller the computational cost becomes. However, since too small number of polygon faces loses details of the shape, especially surface edges, the number of polygon faces should be adjusted corresponding to the target object. In experimental results, we use QSlim made by Garland *et al*. [GH97] to reduce the number of polygon faces for a basic shape.

2.4.2 Normal Map

We use a normal map to represent high frequency components of the target shape. Here, each triangle mesh of the basic shape is divided into micro regions as shown Fig. 2.2. Then, we apply surface orientation in each micro region and call it a normal map. The number of divisions is defined depending on the size of the triangle mesh and the resolution of input images. It would be most effective to have one micro region correspond to one pixel of the input image. The normal map is mapped onto the polygon mesh of the basic shape as bump mapping. In our method, we estimate surface normals in each micro region using multi-view photometric stereo assumed near light sources as shown in Section 2.4.3.

2.4.3 Surface Normal Estimation with Multi-view Photometric Stereo using Near Light Source

Photometric stereo is a surface normal estimation method with an image sequence taken under different illuminations [Woo80]. The image sequences are taken from a fixed viewpoint, which makes it possible to have correspondences among input images. It is difficult to handle near light condition and specular reflections with conventional

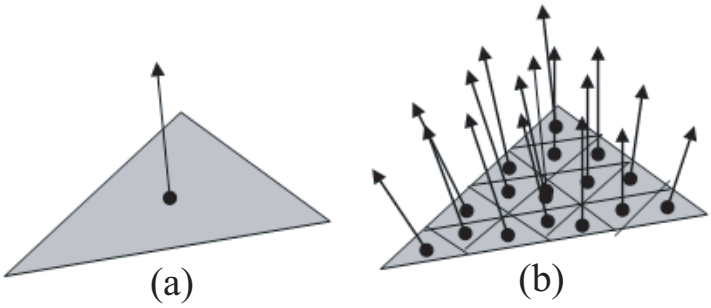


Figure 2.2: Normal map: (a) triangle mesh and its normal, (b) applying a normal map.

photometric stereo, but our proposed multi-view photometric stereo can effectively solve near light conditions and reflections with specularities in the same framework as conventional photometric stereo, The calibrated laser range sensor and the camera make it possible to have correspondences among multi-view images by alignment of the range images. Therefore, even if a target object has no texture and no characteristic shape, it is still easy to have correspondences among them.

Handling Near Light Source Most photometric stereo methods assume a distant light source. This assumption works out in cases where a target is relatively small enough and there is distance between the target object and the light source. However, if the assumption does not hold, a planar surface is wrongly estimated to be a curved surface using photometric stereo.

To solve this problem, a photometric stereo with near light source has been proposed in the past [IWTI94, OD97, CP99]. Without knowing the shape of the target object, these methods need to simultaneously estimate surface normal and shape. Therefore, [IWTI94] assumes unit reflectance of the target object, [OD97] assumes that a light is located on the optical center, and [CP99] uses a special light source. In this way, these methods are not practical because of various limitations.

By using a basic shape acquired by the laser range sensor, we achieve simple photometric stereo with a near light source. In our method, since the position of the light source is known, it is easy to estimate the light source vector from a micro region on the surface of the basic shape to the light source. Moreover, we can consider attenuation of light intensity against the distance between the light source and the object surface.

Handling specular reflection Generally, most photometric stereo methods estimate surface normal by removing specular pixels. However it is difficult to detect areas that specular components affect, especially in case specularity with low intensity is observed across a wide area. Unfortunately, using the camera flash mounted on the camera, specularity is often observed in many areas because of the similar direction of both the camera view and the light. In our method, we use specular-free images as shown in Fig. 2.3, which preserve shading information and are based on Lambertian low, as input images for photometric stereo. Converting to specular-free images, we effectively use specular pixels as well as diffuse pixels as input for photometric stereo.

With known light color intensities (E_r, E_g, E_b), a specular-free image can be independently calculated for each pixel from only one image as follows. Suppose a pixel observation is (i_r, i_g, i_b). First, observation is normalized with light color intensities as $\boldsymbol{i}' = (i'_r, i'_g, i'_b) = (i_r/E_r, i_g/E_g, i_b/E_b)$. Here, assuming that the color of the specular components is the same as the color of the light source, a specular component is $(1, 1, 1)$ direction in the normalized color space. Then, specular-free image $\hat{\boldsymbol{i}} = (\hat{i}_r, \hat{i}_g, \hat{i}_b)$ represents the following equation:

$$\hat{\boldsymbol{i}} = \boldsymbol{i}' - (i'_r + i'_g + i'_b)/3 + a\sqrt{(i'_r - i'_g)^2 + (i'_g - i'_b)^2 + (i'_b - i'_r)^2}, \tag{2.1}$$

where a is a constant value. In our experiments, we use $a = 1$. Mallick *et al*. [MZKB05] also remove specular components to handle them with photometric stereo. But our method is faster than their process because the specular-free image can be simply calculated at the expense of diffuse reflectances that are different from ground truth. However, the difference is no problem for surface normal estimation.

Surface Normal Estimation Suppose $\boldsymbol{n}$ denotes surface normal in micro region **A**, where $|\boldsymbol{n}| = 1$, ρ is the diffuse reflection parameter, $\boldsymbol{s} = \rho\boldsymbol{n}$ is scaled normal, and $\boldsymbol{l}$ is a light vector from micro region **A** to a point light source, where $|\boldsymbol{l}| = 1$, then pixel intensity i of an input image is described as

$$i = \frac{E}{d^2}\boldsymbol{s} \cdot \boldsymbol{l}, \tag{2.2}$$

where d is distance between micro region **A** and the light source, and E is intensity of the camera flash. Irradiance at micro region **A** from the flash light is based on the inverse square low attenuation as $\frac{E}{d^2}$. For surface normal estimation, we use mean intensity among R, G, and B components. Using three input images that can observe

(a) (b)

Figure 2.3: Example of specular-free image: (a) input image, (b) specular-free image.

micro region **A**, we can get the following equations from Eq. (2.2):

$$i_j = \frac{E_j}{d_j^2} \boldsymbol{s} \cdot \boldsymbol{l}_j \quad (j = 1, 2, 3). \tag{2.3}$$

Here, Eq. (2.3) means

$$\mathbf{L} = \begin{pmatrix} \frac{E_1}{d_1^2} l_{1x} & \frac{E_2}{d_2^2} l_{2x} & \frac{E_3}{d_3^2} l_{3x} \\ \frac{E_1}{d_1^2} l_{1y} & \frac{E_2}{d_2^2} l_{2y} & \frac{E_3}{d_3^2} l_{3y} \\ \frac{E_1}{d_1^2} l_{1z} & \frac{E_2}{d_2^2} l_{2z} & \frac{E_3}{d_3^2} l_{3z} \end{pmatrix}, \tag{2.4}$$

where l_{jx}, l_{jy}, l_{jz} are respectively x, y, z components of light vector in image j. Then, Eq. (2.4) is simplified as

$$\boldsymbol{i} = \boldsymbol{s}\mathbf{L}. \tag{2.5}$$

Now $\boldsymbol{i}$ and $\mathbf{L}$ are known, we calculate $\mathbf{L}^{-1}$ to get $\boldsymbol{s}$. Since $\boldsymbol{n}$ is unit vector, we can finally get surface normal as $\boldsymbol{n} = \frac{\boldsymbol{s}}{|\boldsymbol{s}|}$.

2.4.4 Determination of Surface Normal

As we mentioned before, since three input images give one estimated surface normal on each micro region, k input images can provide ${}_kC_3$ candidates for surface normal. To

determine one plausible surface normal from multiple candidates, we remove outliers and estimate it robustly.

We use a voting method based on similarity evaluation to determine surface normal from $\tilde{k}\,(={}_kC_3)$ candidates on micro region $\mathbf{A}$. The similarity between two vectors is calculated as an angle between the two; the smaller the angle, the more similar the two vectors are. Given that $\tilde{k}$ normal vectors $\boldsymbol{n}_j$ $(j = 1, 2, \cdots, \tilde{k})$ are candidates and $V(\boldsymbol{b})$ is a vector group that consists of similar vectors to a vector $\boldsymbol{b}$, where $\boldsymbol{b}$ is one of the candidates. $V(\boldsymbol{b})$ is provided as

$$V(\boldsymbol{b}) = \{\boldsymbol{n}_j \mid \frac{\boldsymbol{b} \cdot \boldsymbol{n}_j}{|\boldsymbol{b}||\boldsymbol{n}_j|} > T_1\}, \tag{2.6}$$

where T_1 is a threshold that defines how similar they are. Let $V(\tilde{\boldsymbol{b}})$ denote a vector group that maximizes the number of vectors in $V(\boldsymbol{b})$ with regard to $\boldsymbol{b}$. Then we determine surface normal $\tilde{\boldsymbol{n}}$ as follows;

$$\tilde{\boldsymbol{n}} = \frac{1}{|V(\tilde{\boldsymbol{b}})|} \sum_{\boldsymbol{n}_j \in V(\tilde{\boldsymbol{b}})} \boldsymbol{n}_j, \tag{2.7}$$

where $|V(\tilde{\boldsymbol{b}})|$ is the number of vectors in $V(\tilde{\boldsymbol{b}})$. In our experiments in Section 2.6, we use $T_1 = 0.99$.

In case the number of input images is very large, it is inefficient to choose $\boldsymbol{b}$ from all ${}_kC_3$ candidates and it would appear that the RANSAC algorithm should be used in order to evaluate whether the sampled surface normal satisfies input images. In our method, however, since the number of input images is at most 20, we calculated with all $\tilde{k}\,(={}_kC_3)$ candidates of surface normal.

Theoretically, all $\tilde{k}\,(={}_kC_3)$ candidates of surface normal should be the same vector, because a specular-free image is also based on the Lambertian model. Practically, however, various errors cause candidate normal vectors to be widely distributed, for example, image noises, shadow pixels, occlusion, and discrepancy as shown Fig. 2.4. Since this kind of candidate vectors affected by a large error would be outliers and be removed, our method can robustly estimate the accurate surface normal.

2.5 Robust Estimation for Reflection Property

To estimate reflection parameters robustly and accurately, first we estimate diffuse reflection parameters in the same manner as the surface normal estimation. Second we cluster micro regions based on the estimated diffuse reflection parameters. Finally we

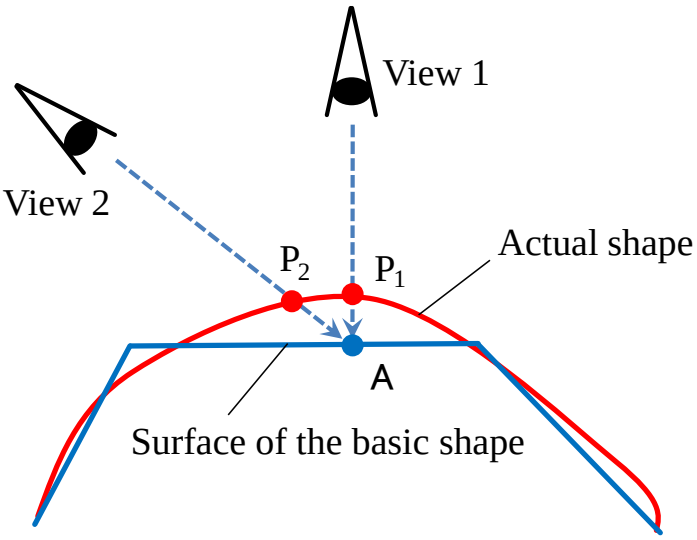


Figure 2.4: The discrepancy between a real shape and a basic shape: micro region **A** corresponds to pixel P_1 with view 1, while it corresponds to P_2 with view 2.

estimate specular reflection parameters on each clustered segment. In our method, the clusters help to estimate specular reflection parameters effectively and accurately.

2.5.1 Estimation of Diffuse Reflection Parameters

Diffuse reflection parameters $K_{D,C}$ (C = R,G,B) are calculated based on the Lambertian model as follows:

$$K_{D,C} = \frac{I_{D,C}}{\boldsymbol{n} \cdot \boldsymbol{l}} \tag{2.8}$$

where $I_{D,C}$ is an input intensity normalized with a light intensity, $\boldsymbol{n}$ is the estimated surface normal, and $\boldsymbol{l}$ is a light vector that is known.

$K_{D,C}$ is calculated on each micro region in the same manner as the normal map. When we have k input images that observe micro region **A**, k candidates of $\boldsymbol{K_D}$ are provided from Eq. (2.8). Therefore we use robust estimation for $\boldsymbol{K_D}$ color vectors with similarity evaluation that is the same as the surface normal estimation.

Here, similarity is defined with an angle between two vectors and scales of these vectors; the smaller the angle and the more similar the scale of the vectors, the more similar the two vectors are. Given that k color vectors $\boldsymbol{K}_{Dj}$ ($j = 1, 2, \cdots, k$) are candidates of diffuse reflection parameters and $V(\boldsymbol{b})$ is a vector group that consists of similar vectors

to a vector $\boldsymbol{b}$, where $\boldsymbol{b}$ is one of the candidates. $V(\boldsymbol{b})$ is provided as

$$V(\boldsymbol{b}) = \left\{ \boldsymbol{K}_{Dj} | \frac{\boldsymbol{b} \cdot \boldsymbol{K}_{Dj}}{|\boldsymbol{b}||\boldsymbol{K}_{Dj}|} > T_2, T_3 < \frac{|\boldsymbol{K}_{Dj}|}{|\boldsymbol{b}|} < T_4 \right\}, \tag{2.9}$$

where T_2, T_3, T_4 are thresholds. Let $V(\tilde{\boldsymbol{b}})$ denote a vector group that maximizes the number of vectors in $V(\boldsymbol{b})$ with regard to $\boldsymbol{b}$. Then we determine a diffuse reflection parameter $\tilde{\boldsymbol{K}_D}$ as follows:

$$\tilde{\boldsymbol{K}_D} = \frac{1}{|V(\tilde{\boldsymbol{b}})|} \sum_{\boldsymbol{K}_{Dj} \in V(\tilde{\boldsymbol{b}})} \boldsymbol{K}_{Dj}, \tag{2.10}$$

where we use $T_2 = 0.99$, $T_3 = 0.9$, $T_4 = 1.1$ in our experiments.

The main reasons for estimation errors are the discrepancy as shown in Fig. 2.4, measurement error, shadow pixels, and specularities. However since these errors are in the minority, our method can remove these outliers with the voting algorithm.

2.5.2 Clustering

Since specular reflection is observed in few areas, many input images are required in order to estimate specular reflection parameters for each micro region. Instead of using many input images, we cluster micro regions and assume that each clustered segment has unit parameters of specular reflection to estimate them effectively and robustly.

As definitions of the clustered segment, in case each micro region satisfies the two following rules, they belong to the same clustered segment if

- 3-D positions are near each other
- Diffuse reflection parameters are similar to each other

In other words, neighboring micro regions whose object colors, *i.e.*, diffuse reflection parameters, are similar to each other, could be made of the same material, so these two rules are appropriate for various objects to assume unit parameters of specular reflection.

Based on the two rules, we use K-means clustering in six-dimensional feature space; three dimensions are for 3-D position of micro regions and the other three are 3-D color vector $\boldsymbol{K}_D$. Distance in the six-dimensional feature space is defined as standard squared Euclidean distance. Let $\boldsymbol{v} = (v_1, v_2, v_3, v_4, v_5, v_6)$ be a six-dimensional vector of

the feature space, then distance D_{pq} between $\boldsymbol{v_p}$ and $\boldsymbol{v_q}$ is represented as

$$D_{pq} = \sum_{j=1}^{6} \frac{(v_{p,j} - v_{q,j})^2}{s_j^2}, \tag{2.11}$$

where s_j^2 is a variance of v_j.

Size of the clustered segment is defined as at least more than one specular observations are given in each clustered segment. Suppose N is the number of micro regions, M is the number of pixels of the target object in the input image sequence, and k is the number of pixels of specular observation, we iteratively merge micro regions until the number of micro regions in each clustered segment is over $\frac{2k}{M}N$.

2.5.3 Estimation of Specular Reflection Parameters

We use the Torrance-Sparrow model [TS67] as follows for representation of specular reflection and estimate K_S and σ in each clustered segment.

$$I_S = \frac{K_S}{\cos\theta_r} \exp\Big(-\frac{\alpha^2}{2\sigma^2}\Big), \tag{2.12}$$

where θ_r is an angle between a viewing vector and a normal vector and α is an angle between the normal vector and a half vector that equally divides an angle between a viewing angle and a light angle.

For estimation of specular reflection parameters, we use some parts of specular pixels $I_{S,C}$ that are calculated by subtracting diffuse pixels $I_{D,C}$ from normalized input pixels I_C. The specular pixels should satisfy the following constraints for the estimation.

- $I_{S,C} > T_{S,C}$, where $T_{S,C}$ is a threshold.
- $\cos^{-1}(\boldsymbol{E} \cdot \boldsymbol{I_S}/|\boldsymbol{E}||I_S|) < \theta_T$, where **E** is light color vector and θ_T is threshold angle.
- $\alpha < \alpha_T$, where α_T is a threshold angle.

Suppose k is the number of specular observations in the clustered segment, A_j ($j = 1, 2, \cdots, k$) is a micro region where specularity is observed, and I_{S,A_j} is intensity of specularity observed in micro region A_j. Then error function $Er(K_S, \sigma)$ is calculated as the following equation based on the Torrance-Sparrow model;

$$Er(K_S, \sigma) = \sum_{j=1}^{k} \frac{1}{2}\Big(I_{S,A_j} - \frac{K_S \exp(-\alpha_{A_j}^2/2\sigma^2)}{\cos\theta_{r,A_j}}\Big)^2. \tag{2.13}$$

Figure 2.5: 3 of 13 input images in verification experiments.

By minimizing this error function, we can estimate K_S and σ.

We solve this non-linear optimization problem with the Levenberg-Marquardt algorithm. Since this algorithm quickly converges but possibly to local minimum depending on initial parameters, we iteratively optimize to change initial parameters with simulated annealing.

In order to robustly estimate the parameters, we also use M-estimation with the Levenberg-Marquardt algorithm. Since some of the input data include estimate error of $\boldsymbol{K_D}$ and the discrepancy as shown Fig. 2.4, a simple fitting approach would come to harm from these outliers. Therefore, these outliers are given low weight with M-estimation to be fitted again. In our method, iterative fitting and M-estimation provide robust estimation against noise.

In our experiments, we use a weighted function that gives zero weight to ten percent of the data most distant from the fitting line. Then we process the remaining 90 percent of the data iteratively using the Levenberg-Marquardt algorithm fitting until convergence occurs.

2.6 Experiment

2.6.1 Simulation Results

To evaluate the effectiveness of the proposed method, we first show quantitative evaluation using the simulation data Fig. 2.5. We use six different resolutions of basic shape data for verification evaluation of estimate accuracy in the difference among them. The numbers of polygonal faces for a basic shape are 300, 500, 1000, 2000, 12000, and 20000 as shown in Fig. 2.6. The number of input images is 13 as shown in Fig. 2.5.

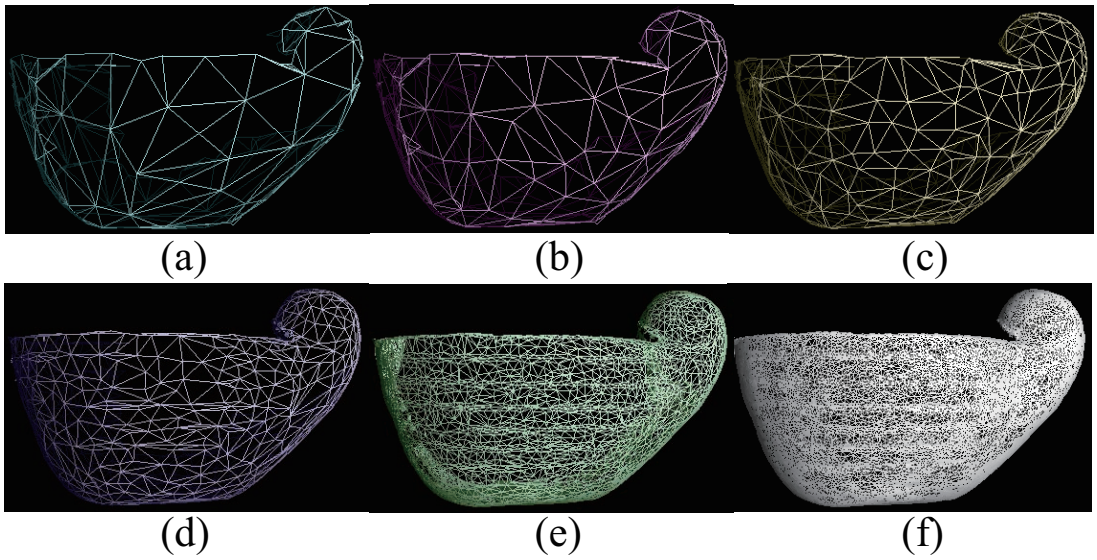


Figure 2.6: Basic shapes in verification: (a) 300, (b) 500, (c) 1000, (d) 2000, (e) 12000, (f) 20000 [faces].

Fig. 2.7 shows the error ratio of the estimated diffuse reflection parameter in the different resolutions of basic shapes with or without normal map. Here, error ratio = $\frac{1}{N}\sum \frac{|\text{difference between ground truth and estimate value}|}{\text{ground truth}}$, where N is the number of sampling points. With the estimated normal map, the errors of the estimated diffuse reflection parameter are around $1 \sim 2\%$ even if the resolution of the basic shape is low. Similarly, Fig. 2.8 shows the error ratio of the estimated specular reflection parameter with or without normal map and also M-estimation. With the estimated normal map, the errors of the estimated diffuse reflection parameter are quite low. It is difficult to accurately estimate the reflection parameter only with normal map because of the adverse effect of outliers. By iteratively using M-estimation and curved line fitting with the Levenberg-Marquardt method, the errors of the specular reflection parameter are extremely low, less than 0.5%.

Fig. 2.9 shows synthesized images using estimated normal map and reflection parameters. Fig. 2.9 (a) is the ground truth image. Fig. 2.9 (b) is the synthesized image with 2000 faces. Fig. 2.9 (c) is the synthesized image with 500 faces. Even if the resolution of basic shape is low, the synthesized image is as good as the high resolution one. However, when the resolution of the edge part is low, the outline of the shape becomes different from the ground truth, and the appearance of the result become degraded. From these results, we see that the resolution of the basic shape should be

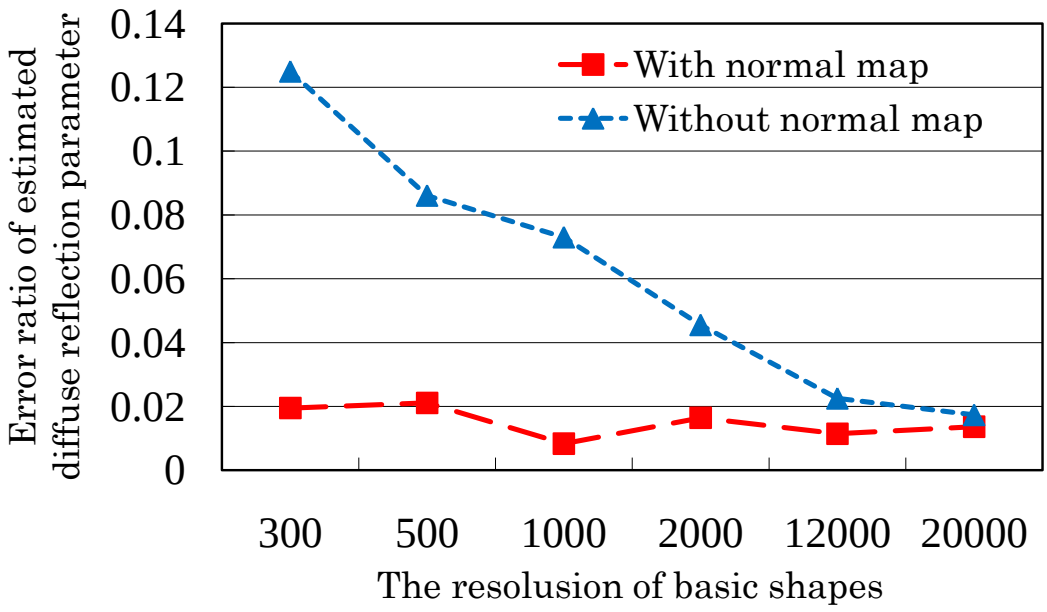


Figure 2.7: Relationship between the resolution of basic shapes and error ratio of estimated diffuse reflection parameters.

determined with accuracy of the edge part.

2.6.2 Real-world Results

We applied our method to various objects for 3-D modeling. We captured basic shape using two different laser range sensors: VIVID made by Konica Minolta that is for small objects, and HDS-3000 made by Leica that is for large objects. We recorded the scenes using a D1x camera made by Nikon. The scenes were illuminated only by a Nikon SPEEDLIGHT SB-80DX camera flash that is mounted on the camera.

Table. 2.1 shows details of target objects. Fig. 2.10 and Fig. 2.11 show the results of the diffuse statue and the diffuse textured tube scene, respectively. (a) is one of the input images, (b) is basic shape, (c) and (d) are synthesized images. Fig. 2.12 shows a magnified part of Fig. 2.11. Our method can handle not only simple diffuse objects but also dense textured objects because of estimated dense normal map with reflection parameters.

Fig. 2.13 shows the result of the dinosaur with specularity. Fig. 2.13 (a) is basic shape, (b) is the clustering result, (c) (e) (g) are three of the input images, and (d) (f) (h) are synthesized images rendered from the same viewpoints. Fig. 2.14 shows magnified

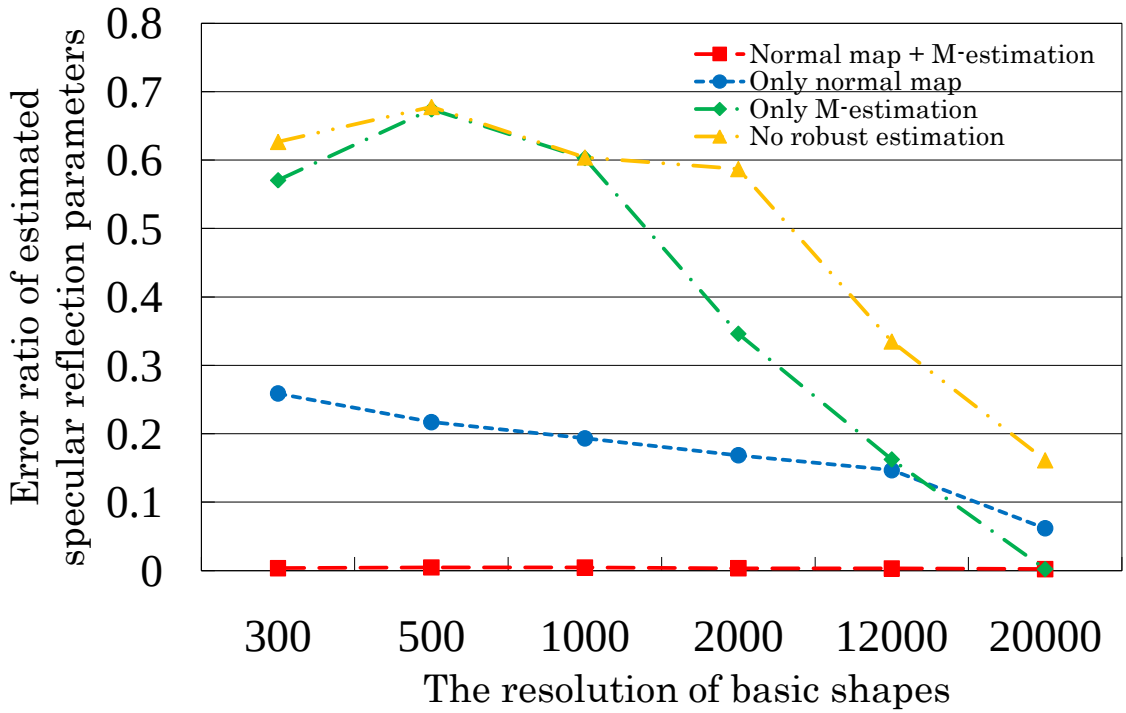


Figure 2.8: Relationship between the resolution of basic shapes and error ratio of estimated specular reflection parameters.

parts of the geometric appearance. Even if the resolution of range image captured by the laser range sensor is quite high as shown (d), its appearance looks noisy. On the other hand, appearance with estimated normal map shown (b) shows fine details from a low-resolution basic shape. Moreover, our method can reconstruct low-intensity and widely distributed specular reflection such as Fig. 2.13 (g).

Fig. 2.15 shows the result of a replica of the Takamatsu tomb as a large object whose scale is 1.6m×2.0m×4.5m. Takamatsu tomb is an ancient tomb built in the 8th century. Fig. 2.15 (a) is one of input images, (b) is basic shape, (c) (d) are synthesized images. Generally, when a target object is large, parallel light can not be assumed. However, since our method obtains the position of the light source, light directions are accurately defined each surface point, then our method estimates surface normal and reflection parameters.

We applied our method for digital 3-D contents. The target is an ancient tomb named Segonko (千金甲) as shown in Fig. 2.16 (a). It was built at Oshimashimomatchi, Kumamoto-shi, Kumamoto-ken (熊本県熊本市小島下町) in the late 5th century. In the stone chamber, we recorded geometric data with a laser range sensor as shown Fig. 2.16

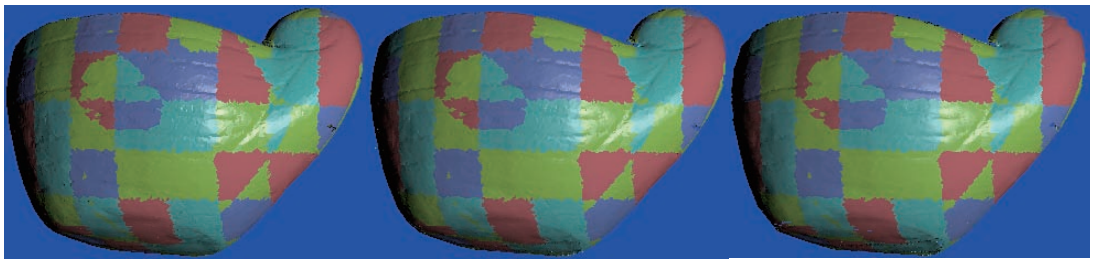


Figure 2.9: Synthesized images in the simulation results: (a) ground truth, (b) 2000 faces, (c) 500 faces.

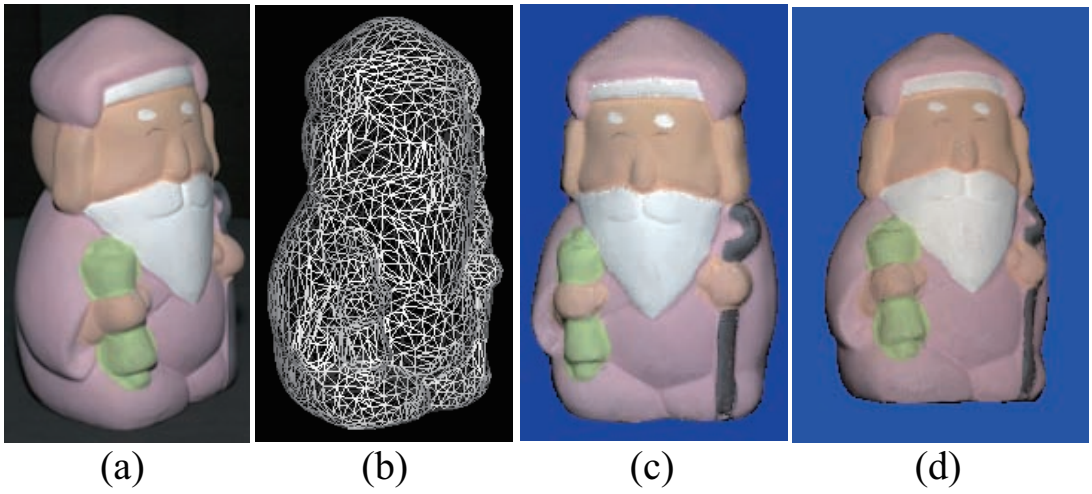


Figure 2.10: Diffuse object: (a) one of the input images, (b) basic shape, (c)(d) synthesized images.

(b). There are sculptured concentric circles and diagonal lines on the inside stone wall. We represented their appearance with an estimated normal map and estimated reflection parameters. We made a 3-D model from these estimated results, and made a 3-D content that allows us to see inside of the stone chamber from a free viewpoint. Fig. 2.17 (a) shows an example of our content. Fig. 2.17 (b) (c) (d) (e) show magnified parts of the sculptured concentric circle pattern. (b) is estimated normal map, (c) is estimated diffuse reflectance, (d) is the synthesized image with normal map, and (e) is the synthesized image without normal map. Using estimation with normal map, we can effectively reconstruct the appearance of the concentric circle pattern.

Table 2.1: Details of the target objects in experiments.

target object	# of input images	size	distance from camera	# of faces	division number of micro regions
diffuse statue	18	20cm	80cm	5000	121
textured tube	15	10cm	1.0m	1000	900
specular dinosaur	18	20cm	80cm	5000	121
big tomb	18	1.6m	4.5m	3500	1600
Segonko	24	1.2×2.6m	1.0m	30000	529

Figure 2.11: Textured diffuse tube: (a) one of the input images, (b) basic shape, (c)(d) synthesized images.

2.7 Conclusion

We proposed a novel method for 3-D modeling using a fusion of a laser range sensor, a camera, and a camera flash. This combination provides dense normals and surface colors that can be mapped on a 3-D model, whereas conventional sensors only output point clouds of the 3-D geometry. Furthermore, the fusion enables formulations to be made simply and practically. Multi-view photometric stereo was used for estimating the fine normal distribution with a basic shape measured by the laser range sensor. Our photometric stereo can easily handle near-light formulation and specularity. Detailed surfaces can be shown by applying the normal map as bump mapping to the basic shape. Robust estimation and clustering were used for estimating reflection parame-

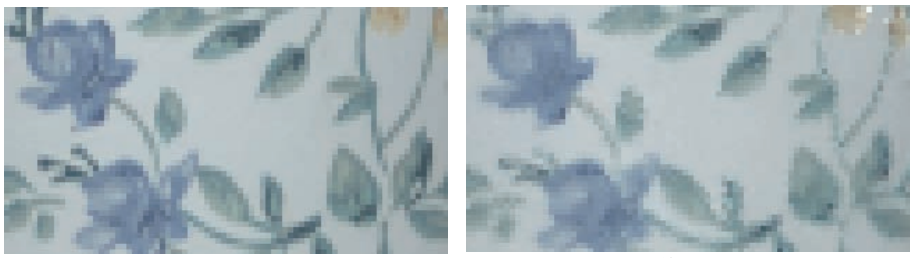


Figure 2.12: A magnified part of the synthesized image in Fig. 2.11: (a) ground truth, (b) synthesized image.

ters. Results demonstrated that our method could estimate highly accurate reflection parameters and provide fine surface appearances using only a small amount of data. The effectiveness and the practicality of our method was shown by an application that displayed 3-D contents.

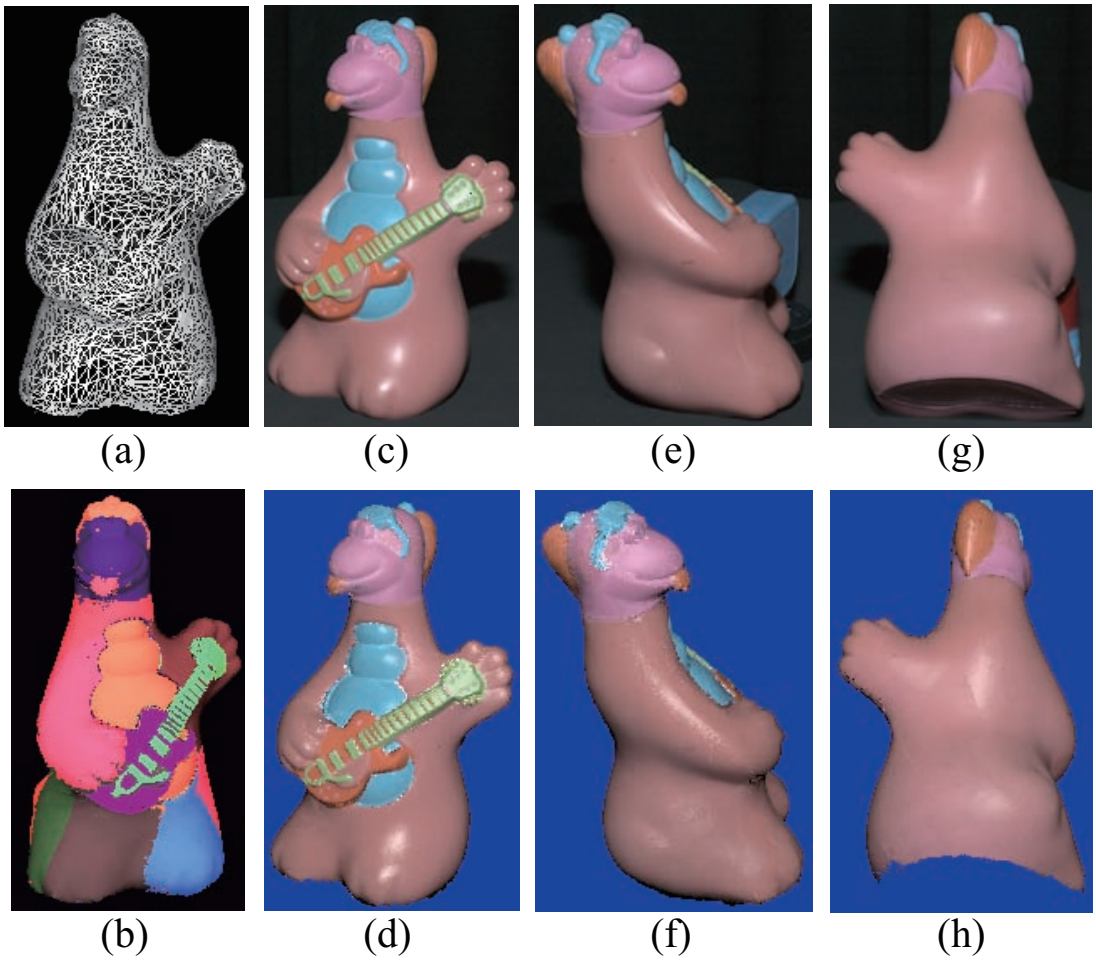


Figure 2.13: Dinosaur scene with specularity: (a) basic shape, (b) clustering result, (c)(e)(g) three of the input images, (d)(f)(h) synthesized images.

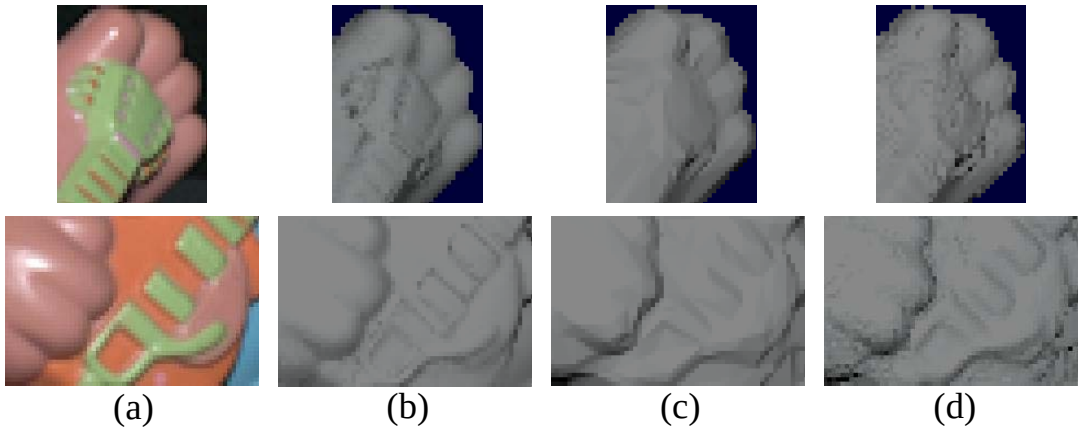


Figure 2.14: Magnified parts of the dinosaur's geometric appearance: (a) input images, (b) our geometric appearance using a normal map, (c) basic shape, (d) 10 times denser range data than basic shape.

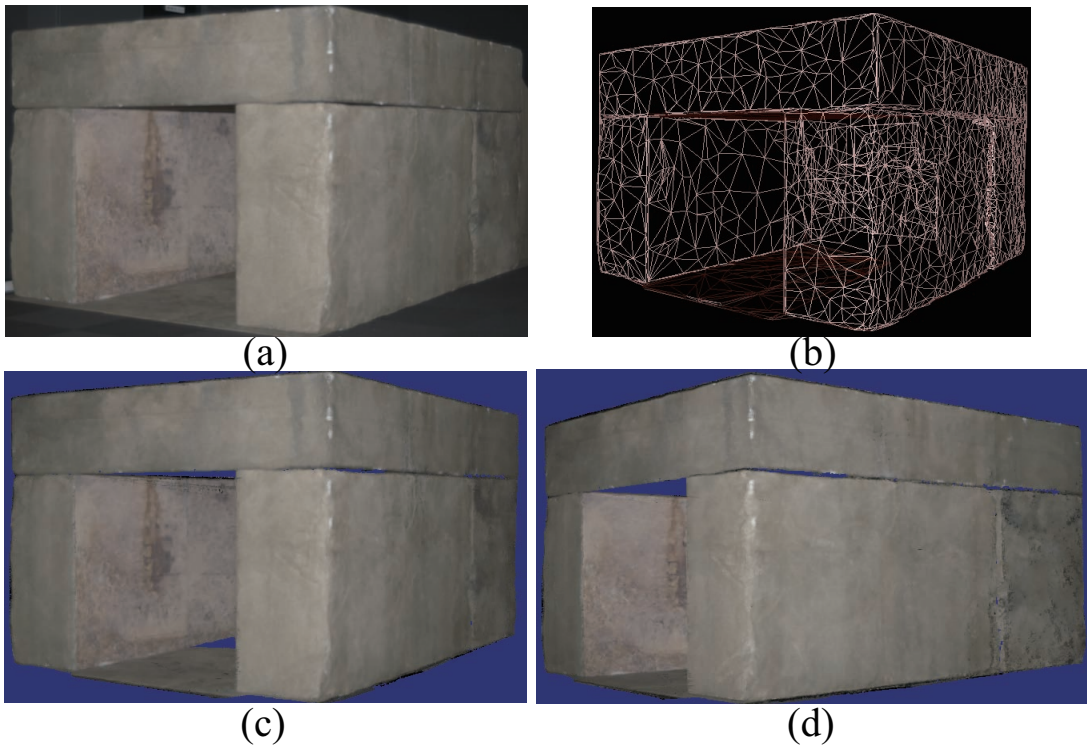


Figure 2.15: Large object (Takamatsu tomb): (a) one of the input images, (b) basic shape, (c)(d) synthesized images.

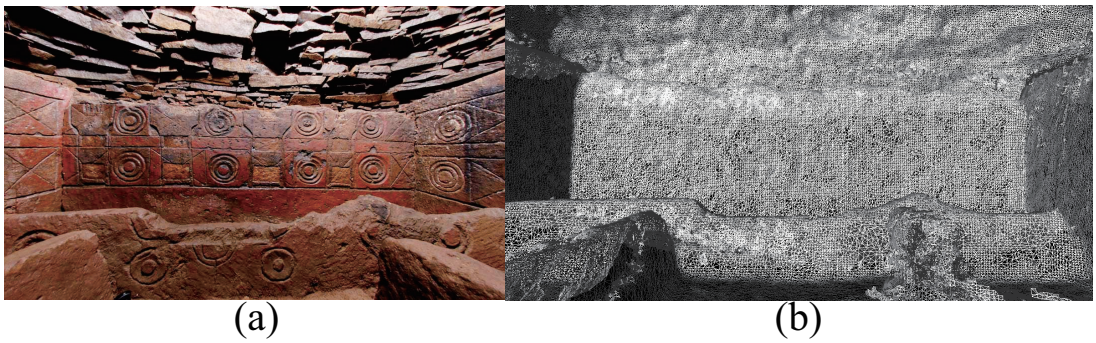


Figure 2.16: Segonko: (a) photograph, (b) basic shape.

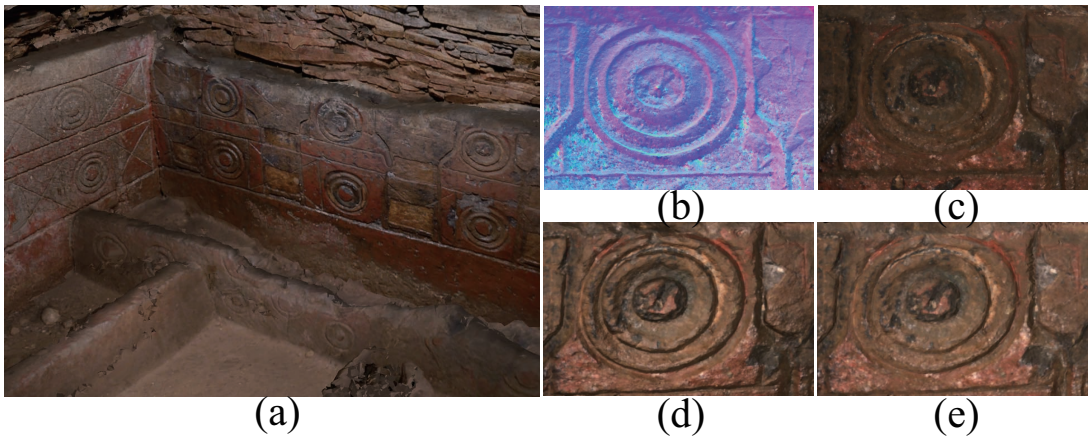


Figure 2.17: An example of 3-D contents: (a) synthesized image from a viewpoint, (b) estimated normal map, (c) estimated diffuse albedo, (d) with normal map, (e) without normal map.

Chapter 3

A Hand-held Photometric Stereo Approach for Full 3-D Modeling

A photometric constraint achieves simultaneous estimation of shape, surface normal, and reflectance from a set of images taken from different viewpoints under various directional lighting. This chapter presents a simple yet practical 3-D modeling method using a hand-held camera with an attached point light source. Unlike prior approaches, we formulate the problem using realistic assumptions of a near light source, non-Lambertian surfaces, perspective camera model, and the presence of ambient lighting.

Our simultaneous estimation works well but requires high computational cost, so we propose another extended method that uses color information and efficient formulation to remove outliers and to reduce the computational cost. Removing outliers help to robustly estimate a full 3-D model of the target object. The effectiveness of the proposed method and the comparison between the proposed method and the extended method are verified using simulated and real-world scenes.

3.1 Introduction

Three-dimensional (3-D) shape acquisition and reconstruction is a challenging problem with many important applications in archeology, medical, and film and video game industries. Numerous systems exist for 3-D scanning using methods such as multi-view stereo, structured light, and photometric stereo; however, the use of 3-D modeling is limited by the need of large, expensive, and costly hardware setups that require

extensive calibration procedures. As a result, 3-D modeling is often neither a practical nor accessible option for many application scenarios. In this chapter we present a simple, low-cost method for object shape and reflectance acquisition using a hand-held camera with an attached point light source.

When an object is filmed with our camera setup, its appearance changes both geometrically and photometrically. These changes provide clues to the shape of an object; however, their simultaneous variation prohibits the use of traditional methods for 3-D reconstruction. Standard multi-view stereo and photometric stereo assumptions fail when considered independently; however, when considered jointly their complementary information enables high-quality shape reconstruction.

The particular concept of jointly using multi-view and photometric clues for shape acquisition is not new to this work and has become somewhat popular in recent years [ZCHS03, LHYK05, JK07]; however, these previous works have several limitations that keep them from being used in practice: the need for fixed or known camera and light positions, a dark room, an orthographic camera model, and a Lambertian reflectance model. It is often difficult to fit all these constraints in real world situations, *e.g.*, to adhere to an orthographic camera and distant point light source model, one has to film the object at a distance from the camera and light, which makes hand-held acquisition impossible. Furthermore, most real-world objects are not Lambertian. Our work improves upon previous work by removing all of these constraints. The proposed method extends our previous work [HMJI09] by efficiently removing outliers using color information and view constraint and by reducing computational cost using efficient computation. This leads to robust estimation and achieves full 3-D reconstruction that was not feasible with previous approach.

The primary contributions of the present work include:

1. Development of an auto-calibrated, hand-held multi-view photometric stereo camera,

2. A new shape estimation algorithm that considers perspective effect of the camera, near light configuration, ambient illumination, and specular surfaces.

3. A method for simultaneous estimation of depth and surface normal.

4. An efficient outlier rejection method that uses color information and view constraint.

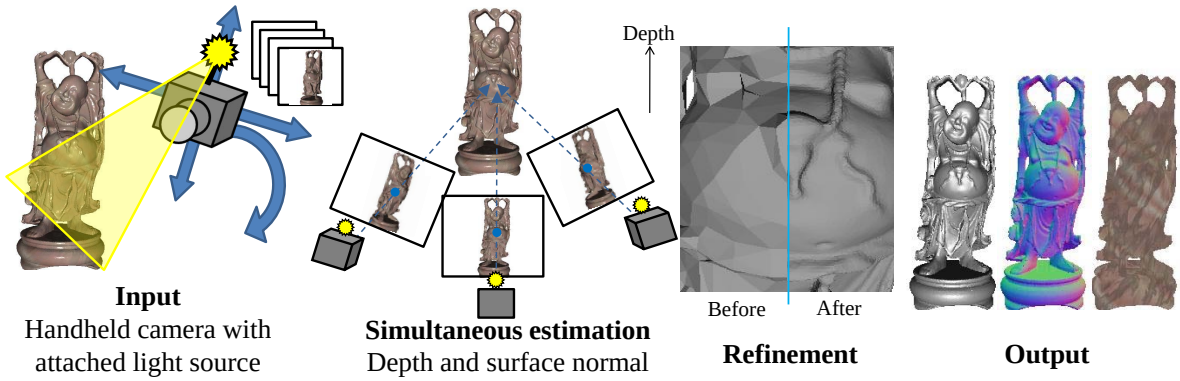


Figure 3.1: Overview of the proposed approach.

In the next section, we discuss the previous work of shape estimation methods. In Section 3.2, we describe the framework of our method, and Section 3.5 describes implementation details. We present results in Section 3.6 followed by a discussion and our conclusions.

3.1.1 Previous work

Shape reconstruction has a long, storied history in computer vision, and, unfortunately, cannot be fully addressed within the scope of this dissertation. At a high-level, typical approaches use either multi-view information or photometric information separately. Multi-view stereo methods often require elaborate setups [ZKU*04, SCD*06] and, while they can excel at recovering large-scale structures, they often fail to capture high-frequency details [NRDR05]. Photometric stereo setups can be more modest, but they still require known or calibrated light positions [MWGA06] and often have inaccuracies in the low-frequencies components of the shape reconstruction [NRDR05].

Recent work has merged the benefit of these to methods using either two separate datasets [NRDR05, WMP*06] or jointly using one dataset. Maki *et al*. [MWW02] use a linear subspace constraint with several known correspondences to estimate light source directions up to an arbitrary invertible linear transform, but they do not recover surface normals. Simakov *et al*. [SFB03] merge multi-view stereo and photometric constraints by assuming that the relative motion between the object and the illumination source is known. While this motion is recoverable in certain situations, there can be ambiguities. Additionally, their process can only recover normals up to an ambiguity along a plane.

In contrast, our method automatically finds correspondences to recover camera parameters, with a known relative light position, and solves depth and normals without any remaining ambiguity. More recently, Birkbeck *et al*. [BCSJ06] and Hernández *et al*. [HVC08a] show impressive surface reconstruction results by exploiting silhouette and shading cues using a turntable setup.

Our work is similar in spirit to that of Pollefeys *et al*. [PVGV*04] who perform 3-D modeling with a perspective camera model, but use standard multi-view clues and no photometric clues, thus they do not recover normals as we do. Our work also is closely related to the work by Zhang *et al*. [ZCHS03], Lim *et al*. [LHYK05], and Joshi and Kriegman [JK07]. Zhang *et al*. present an optical flow technique that handles illuminations changes, which requires numerous images from a dense video sequence. Lim *et al*. start with very sparse initial estimate of the shape computed from the 3-D locations for a sparse set of features and refine this shape using iterative procedure. Joshi and Kriegman extend a sparse multi-view stereo algorithm with a cost-function that uses a rank-constraint to fit the photometric variations. Our work shares some similarity with Joshi and Kriegman's approach for simultaneous estimation of depth and normals. In contrast with these three previous works, we use a known, near light position and can handle using a perspective camera and non-Lambertian objects.

3.2 Proposed method

Our method uses a simple configuration, *i.e.*, one LED point light source attached to a camera. Fig. 3.2 shows a prototype of the hand-held photometric stereo camera. This configuration has two major advantages. First, it gives a photometric constraint that allows us to efficiently determine surface normals. Second, it enables a completely hand-held system that is free from heavy rigs.

Fig. 3.3 illustrates the flow of the proposed method. After calibrating camera intrinsics and vignetting (step 1), we take images of a scene from different view points using the camera with the LED light always turned on. Given such input images, our method first determines camera extrinsics and light source position in steps 2 and 3. In step 4, our method performs simultaneous estimation of shape, normals, and albedos. We use an efficient discrete optimization to make the problem tractable. Step 5 refines the estimated surface shape by a simple optimization method. We first describe the photometric stereo formulation for our configuration in Section 3.2.1, and then describe the algorithmic details of our two major stages (steps 4 and 5) in Sections 3.3 and 3.4.

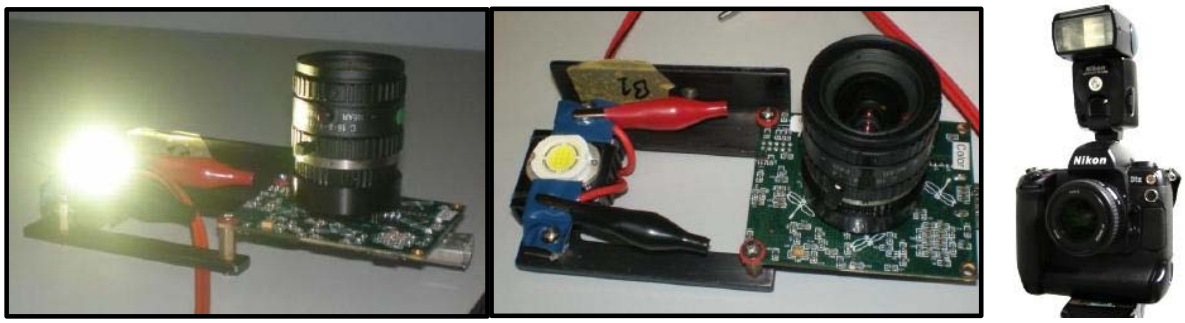


Figure 3.2: Left: Our prototype implementation of the hand-held photometric stereo camera. Right: Commercially available camera of Nikon D1 with a camera flash. Our method can be handled with each camera.

3.2.1 Photometric stereo under a near-light source

This section formulates the photometric stereo for Lambertian objects under a near-light source with ambient illumination. Our method handles specular reflection and shadows as outliers that deviates from this formulation.

Suppose $\boldsymbol{s}$ is a light position vector that is known and fixed in the camera coordinate. Let us consider a point $\boldsymbol{x}$ on the scene surface with a surface normal $\boldsymbol{n}$ in the world coordinate. In the i-th image, the light vector $\boldsymbol{l}_i$ from the surface point $\boldsymbol{x}$ to the light source is written as

$$\boldsymbol{l}_i = \boldsymbol{s} - (\mathbf{R}_i\boldsymbol{x} + \boldsymbol{t}_i), \tag{3.1}$$

where $\mathbf{R}_i$ and $\boldsymbol{t}_i$ are, respectively, the rotation matrix and translation vector from the world coordinate to the camera coordinate. With the near light source assumption, intensity observation o_i is computed with accounting the inverse-square law as

$$o_i = E\rho\frac{\boldsymbol{l}_i \cdot (\mathbf{R}_i\boldsymbol{n})}{|\boldsymbol{l}_i|^3} + a, \tag{3.2}$$

where E is the light source intensity at a unit distance, ρ is surface albedo, and a is the magnitude of ambient illumination. Defining a scaled normal vector $\boldsymbol{b} = \rho\boldsymbol{n}$, normalized pixel intensity $o_i' = o_i/E$, and normalized ambient effect $a' = a/E$, Eq. (3.2) becomes

$$o_i' = \frac{\boldsymbol{l}_i \cdot (\mathbf{R}_i\boldsymbol{b})}{|\boldsymbol{l}_i|^3} + a' = \frac{(\mathbf{R}_i^T\boldsymbol{l}_i) \cdot \boldsymbol{b}}{|\boldsymbol{l}_i|^3} + a'. \tag{3.3}$$

According to Eq. (3.3), we can compute $\boldsymbol{n}$, ρ, and a' from at least 4 observations. Given the rotation matrix $\mathbf{R}_i$, translation vector $\boldsymbol{t}_i$, and position vector $\boldsymbol{x}$, we can easily compute the light vector $\boldsymbol{l}_i$ from Eq. (3.1). Once we know the light vector $\boldsymbol{l}_i$, we can

1. **Calibrate the Camera** (Section 3.5.1)

   Calibrate camera intrinsics and estimate vignetting.

2. **Estimate Camera Projection Matrices** (Section 3.5.2)

   Using Structure from Motion/Bundle adjustment, recover the camera projection matrices for each frame.

3. **Estimate light source position** (Section 3.5.2)

   Resolve the scale ambiguity by using our photo consistency on feature points from the structure from motion process.

4. **Compute Dense Depth and Normal Map** (Section 3.3)

   Find the dense depth map and normals by minimizing our near light-source, multi-view photometric constraint using a graph cut.

5. **Compute Final Surface** (Section 3.4)

   Recover the final surface by fusing the recovered dense depth map and normal field.

Figure 3.3: Our shape reconstruction algorithm.

estimate the scaled normal vector $\boldsymbol{b}$ on each surface point with photometric stereo. According to Eq. (3.3), we can compute $\boldsymbol{n}$, ρ, and a' in straightforward way from at least 4 observations as

$$\begin{bmatrix} o'_1 \\ o'_2 \\ o'_3 \\ o'_4 \end{bmatrix} = \begin{bmatrix} \boldsymbol{l}'^T_1 & 1 \\ \boldsymbol{l}'^T_2 & 1 \\ \boldsymbol{l}'^T_3 & 1 \\ \boldsymbol{l}'^T_4 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ a' \end{bmatrix}, \tag{3.4}$$

where we define the near light vector $\boldsymbol{l}'_i = \mathbf{R}_i^T \boldsymbol{l}_i / |\boldsymbol{l}_i|^3$. By solving the linear system, we can estimate $\boldsymbol{b}$ and a'.

The above derivation shows how to recover normals using near-light source photometric stereo once image correspondence is known; however, for our setup where we want to leverage multi-view clues, correspondence is unknown and must be es-

timated. Estimating the unknown correspondence is one of the key concerns of this work. To efficiently achieve this goal, we propose a method to utilize color information to effectively remove outliers.

3.2.2 Color based approach

This section describes a color based approach for removing erroneous matches. With the color information, our method effectively removes outliers that are due to incorrect matches or specular reflections. Moreover, we remove the ambient lighting effect by subtraction only using inliers. We show that this changes the formulation above and leads to an efficient computation.

Specular removal

We use color information to remove specularity based on the dichromatic reflection model in the RGB color space. For color representation, we use 3-D color vector, *e.g.*, $\boldsymbol{o} = (o^R, o^G, o^B)$ instead of intensity $o = o^R + o^G + o^B$. We assume that our inlier observation consists diffuse lambertian reflection component and ambient effect. However, some observations $\boldsymbol{o}$ contain specular reflection component as

$$\boldsymbol{o} = \boldsymbol{I}_{dif} + \boldsymbol{I}_{sp} + \boldsymbol{a}, \tag{3.5}$$

where $\boldsymbol{I}_{dif}$ is diffuse reflection component, $\boldsymbol{I}_{sp}$ is specular reflection component, and $\boldsymbol{a}$ is ambient effect. In case an observation $\boldsymbol{o}$ contains specular reflection component, it becomes outlier from our assumption. To reduce ratio of the outliers, we remove specular reflection component and use the observation as an inlier.

At first, we show descriptions of specular reflection and diffuse reflection. The color of specular reflection component is known to be the same with the light source color $\boldsymbol{E} = (E^R, E^G, E^B)$. This is known as neutral interface reflection assumption. Then the color vector of specular reflection $\boldsymbol{I}_{sp}$ is as follows:

$$\boldsymbol{I}_{sp} = m_{sp}\boldsymbol{E}, \tag{3.6}$$

where m_{sp} is scaling factor of the specularity and $\boldsymbol{E}$ is a unit vector ($|\boldsymbol{E}| = 1$). On the other hand, the color of diffuse reflection component is defined with both the light source

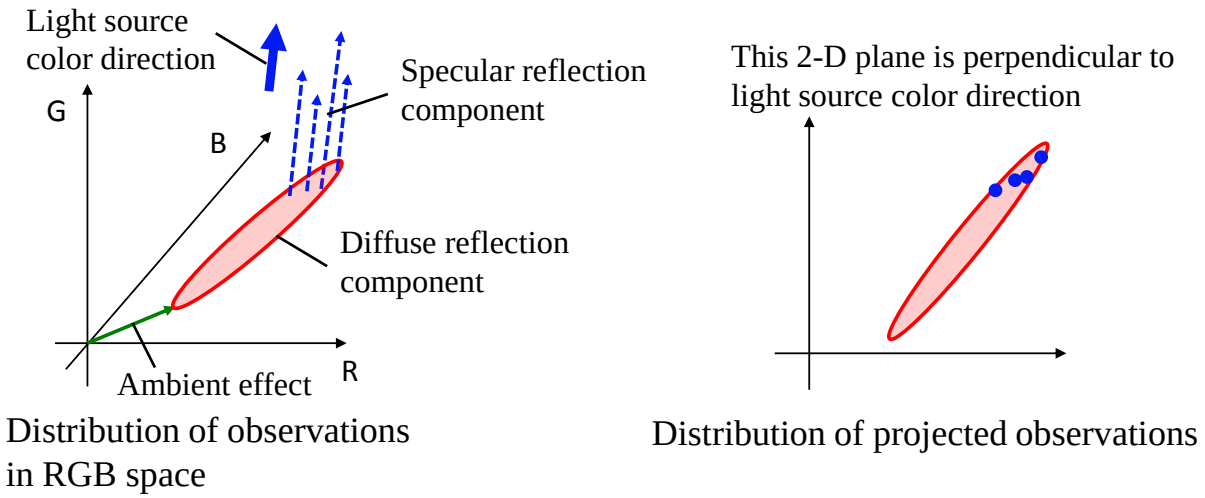


Figure 3.4: Left: In RGB space diffuse reflection observations are distributed on a straight line. Right: When observations are projected to a plane that is perpendicular to light source color direction, specular pixels only depend on diffuse reflection component.

color and the surface color. Let $\boldsymbol{D}$ denote a color vector of diffuse reflection. Then the diffuse reflection $\boldsymbol{I}_{dif}$ is as follows:

$$\boldsymbol{I}_{dif} = m_{dif}\boldsymbol{D}, \tag{3.7}$$

where m_{dif} is scaling factor of diffuse reflection and $\boldsymbol{D}$ is a unit vector.

Before the specular removal, we first pick up observations that has the same diffuse color independent of specular reflection. In 3-D RGB space, distributions of diffuse reflection and specular reflection are based on respective manners. The left figure of Fig. 3.4 shows observation distribution in RGB space. Diffuse reflection is distributed on a straight line whose directional vector is $\boldsymbol{D}$. The diffuse line deviates from the origin because of the ambient effect. Specular reflection is distributed on the light source color direction $\boldsymbol{E}$ from the diffuse line. Suppose we know the light source color $\boldsymbol{E}$, we project observations to the plane that is perpendicular to light source color direction $\boldsymbol{E}$, then the projected observations are specular-invariant and distributed on the one straight line. The projected observation $\hat{\boldsymbol{o}}$ is calculated as

$$\begin{aligned} \hat{\boldsymbol{o}} &= \boldsymbol{o} - (\boldsymbol{o}\cdot\boldsymbol{S})\boldsymbol{S} \\ &= m_{dif}\Big(\boldsymbol{D} - (\boldsymbol{D}\cdot\boldsymbol{S})\boldsymbol{S}\Big) - (\boldsymbol{a}\cdot\boldsymbol{S})\boldsymbol{S}. \end{aligned} \tag{3.8}$$

This equation is independent of m_{sp}, *i.e.*, independent of specular reflection. According to Eq. (3.8), projected observations $\hat{\boldsymbol{o}}_i$ form a straight line distribution on the 2-D plane that is perpendicular to $\boldsymbol{S}$ as shown the right figure of Fig. 3.4 if the observations have correct correspondences and the same diffuse color. We use RANSAC algorithm to fit the straight line and to remove outliers deviated from the straight line.

After removing the outliers, we detect the straight line of the diffuse reflection in 3-D RGB color space. Again, we use RANSAC algorithm to fit the straight line and estimate the color vector of diffuse reflection $\tilde{\boldsymbol{D}}$ and ambient effect $\tilde{\boldsymbol{a}}$. Here, specularities are outliers. In order to remove specular reflection, we should estimate a diffuse scale factor m_{dif} as follows:

$$m_{dif} = \left[\frac{1}{1 - (\boldsymbol{S} \cdot \tilde{\boldsymbol{D}})^2} \tilde{\boldsymbol{D}} - \frac{\boldsymbol{S} \cdot \tilde{\boldsymbol{D}}}{1 - (\boldsymbol{S} \cdot \tilde{\boldsymbol{D}})^2} \boldsymbol{S} \right] \cdot (\boldsymbol{o} - \tilde{\boldsymbol{a}}). \tag{3.9}$$

Once we get m_{dif}, we can calculate specular removal observation $\boldsymbol{o}_{\boldsymbol{dif}}$ as follows:

$$\boldsymbol{o}_{\boldsymbol{dif}} = m_{dif} \tilde{\boldsymbol{D}} + \tilde{\boldsymbol{a}}. \tag{3.10}$$

In this way, specularities are removed and the observation gets to be consist of only diffuse reflection and ambient effect, that is an inlier in our observation assumption. In case $\boldsymbol{S} \cdot \tilde{\boldsymbol{D}} \approx 1$, *i.e.*, surface color of the target object is white, our method of specular removal does not work and we handle specularities as outliers to be removed.

3.2.3 Efficient formulation

According to Eq. (3.4), we need to calculate a 4×4 inverse matrix to solve for the scaled normal $\boldsymbol{b}$ and ambient effect a'. Here, we cancel out ambient effect a'; then we can estimate scaled normal $\boldsymbol{b}$ by calculating a 3×3 inverse matrix. To do this we first pick up one inlier observation o'_0 that is on the diffuse reflection line in RGB space and then subtract from the other inliers o'_i as

$$O_{i0} = o'_i - o'_0 = (\boldsymbol{l}'_i - \boldsymbol{l}'_0) \cdot \boldsymbol{b}. \tag{3.11}$$

Then we can solve the following linear system with 3×3 matrix instead of 4×4 matrix as

$$\begin{bmatrix} O_1 \\ O_2 \\ O_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{l}'^T_1 - \boldsymbol{l}'^T_0 \\ \boldsymbol{l}'^T_2 - \boldsymbol{l}'^T_0 \\ \boldsymbol{l}'^T_3 - \boldsymbol{l}'^T_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \end{bmatrix}. \tag{3.12}$$

The computational cost of calculating a 3×3 inverse matrix is much less than the cost of calculating 4×4 inverse matrix.

3.3 Simultaneous estimation of depth and normal

Our method simultaneously estimates depth, normal, and surface albedo. To do this we estimate correspondence to get position information and use photometric clues to get normals – these two are fused to get the final depth. To compute correspondence, we run a stereo algorithm, where we replace the traditional match function that uses brightness constancy with one that uses the photometric clues, normal consistency, and surface smoothness. We formulate the problem in a discrete optimization framework.

Let us first assume the camera positions and light position are known – the estimation of these parameters is discussed in detail in Section 3.5.2. Suppose that we have m images taken from different view points with our camera. We recover correspondence by performing plane-sweep stereo. For each depth in the plane-sweep, we warp the set of images from different view points to align to one reference view. In this reference camera coordinate frame, the depth planes are assumed in the z direction parallel to the xy plane at a regular interval Δ_z as shown in Fig. 3.6.

Specifically, we warp each image to the reference camera coordinate for depth $z_j = z_0 + j\Delta_z$ using a 2-D projective transform $\boldsymbol{H}_{ij}$ that converts pixel location from a view i to the reference view. $\boldsymbol{H}_{ij}$ is calculated as

$$\boldsymbol{H}_{ij} = \boldsymbol{A}_i \left(\boldsymbol{R} + \frac{\boldsymbol{t}\boldsymbol{\nu}^T}{-z_j} \right) \boldsymbol{A}_0^{-1}, \tag{3.13}$$

where $\boldsymbol{A}_i$ and $\boldsymbol{A}_0$ are intrinsic matrices in the view i and the reference view respectively, $\boldsymbol{R}$, $\boldsymbol{t}$, and $\boldsymbol{\nu}$ are described as follows:

$$\boldsymbol{R} = \boldsymbol{R}_i \boldsymbol{R}_0^T, \tag{3.14}$$

$$\boldsymbol{t} = \boldsymbol{t}_i - \boldsymbol{R}\boldsymbol{t}_0, \tag{3.15}$$

$$\boldsymbol{\nu} = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}, \tag{3.16}$$

where $\boldsymbol{R}_i$ and $\boldsymbol{R}_0$ are rotation matrices in the view i and the reference view respectively, $\boldsymbol{t}_i$ and $\boldsymbol{t}_0$ are translation vectors in the view i and the reference view, and $\boldsymbol{\nu}$ is a unit vector from a depth plane to the origin of the reference camera coordinate. Using the

Inputs
z_0, z_n, $\triangle_z$: depth range , $\boldsymbol{C}$: camera parameters , $\boldsymbol{I}$: an image sequence , $\boldsymbol{L}$: light direction
Outputs
$\boldsymbol{D}$: depth map, $\boldsymbol{N}$: normal map, $\boldsymbol{R}$: reflectance map
Parameters
z: depth, m: the number of input images, $\boldsymbol{H}$: homography matrix, p: the number of pixels, o: intensity of the observation, $\boldsymbol{S}$: observation group

---

for $z = z_0$ to z_n with a step $\triangle_z$ **do**
  for $i = 0$ to m with a step 1 **do**
    $\boldsymbol{H}_i$ = CalcHomography($\boldsymbol{C}$)
  end for
  for $k = 0$ to p with a step 1 **do**
    for $i = 0$ to m with a step 1 **do**
      $\boldsymbol{S} = \{o_i \,|\, o_i = \text{Warping}(\boldsymbol{H}_i, \boldsymbol{I}_i, z)\}$
      $(\boldsymbol{D}, \boldsymbol{N}, \boldsymbol{R})$ = SimulEstimation($\boldsymbol{S}$, $\boldsymbol{L}$)
    end for
  end for
end for

Figure 3.5: Algorithm of simultaneous estimation for depth, surface normal, and reflectance.

2-D projective transform $\boldsymbol{H}_{ij}$, we warp each image to the reference camera coordinate as

$$\boldsymbol{p}_w = \boldsymbol{H}_{ij}\boldsymbol{p}_o, \tag{3.17}$$

where $\boldsymbol{p}_w$ and $\boldsymbol{p}_o$ represent the warped pixel location and the original pixel location, respectively, described by $\boldsymbol{p} = [u\, v\, 1]^T$ in the image coordinate system. Then we perform an optimization over this set of warped images to find the optimal per-pixel depth z_j that gives the best agreement among the registered pixels (given pixel p in the reference view and corresponding pixels in the warped images $I_{ij}(p) \quad (i = 1, 2, \ldots, m)$). This is done according to three criteria: photo consistency, a surface normal constraint, and a

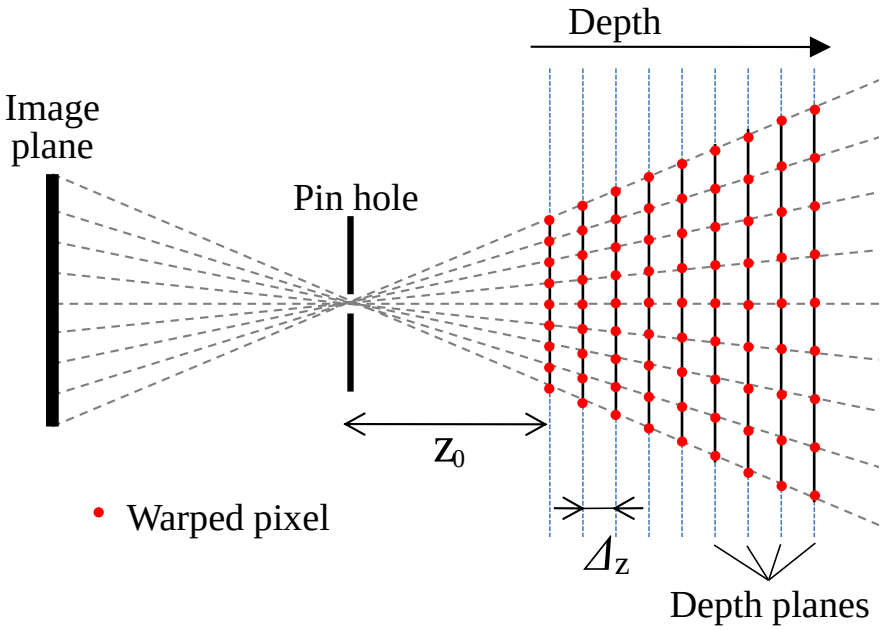


Figure 3.6: Performing plane-sweep stereo in the reference camera coordinate.

smoothness measure.

Photo consistency

Intensity based approach

Our photo consistency measure is defined to account for varying lighting, since the light source is attached to the moving camera. To explicitly handle shadows, specular reflections, and occlusions, we use a RANSAC [FB81] approach to obtain the initial guess of surface normal $\boldsymbol{n}_p$, surface albedo $\boldsymbol{\rho}_p$, and ambient $\boldsymbol{a}_p$ using the near-light photometric stereo assumption described in Section 3.2.1. The vector form of surface albedo $\boldsymbol{\rho}_p$ and ambient $\boldsymbol{a}_p$ contain elements of three color channels. Using the initial guess, the photo consistency g is checked with each of other $m-4$ images at a given pixel p as

$$g_i(\boldsymbol{n}_p, \boldsymbol{\rho}_p, \boldsymbol{a}_p) = \sum_{c=\{R,G,B\}} |O_i^c(p) - E^c \rho_p^c \boldsymbol{l'} \cdot \boldsymbol{n}_p - a_p^c|. \tag{3.18}$$

We also compute the number of images that satisfy the photo consistency N as

$$N = |\{i \,|\, g_i(\boldsymbol{n}_p, \boldsymbol{\rho}_p, \boldsymbol{a}_p) < \tau\}|, \tag{3.19}$$

Input
z_0, z_n, $\triangle_z$: depth range , $\boldsymbol{C}$: camera parameters , $\boldsymbol{I}$: an image sequence , $\boldsymbol{L}$: light direction
Output
$\boldsymbol{D}$: depth map , $\boldsymbol{N}$: normal map , $\boldsymbol{R}$: reflectance map
Parameter
z: depth , m: the number of input images , $\boldsymbol{H}$: homography matrix , p: the number of pixels , o: intensity of the observation , $\boldsymbol{V}$: viewing vector for subtraction , $\boldsymbol{S}$: observation group , $\boldsymbol{S}_D$: diffuse observation group , $\boldsymbol{S}'$: subtracted observation group , $\boldsymbol{L}'$: subtracted light direction

---

for $z = z_0$ to z_n step $\triangle_z$ **do**
  for $i = 0$ to m step 1 **do**
    $\boldsymbol{H}_i$ = CalcHomography($\boldsymbol{C}$)
  end for
  for $k = 0$ to p step 1 **do**
    for $i = 0$ to m step 1 **do**
      $\boldsymbol{S} = \{o_i \,|\, o_i = \text{Warping}(\boldsymbol{H}_i, \boldsymbol{I}_i, z)\}$
      $\boldsymbol{S}$ = ViewConstraint($\boldsymbol{S}$)
      $\boldsymbol{S}$ = EstimateDiffuseLine($\boldsymbol{S}$)
      ($\boldsymbol{S}_D$, $\boldsymbol{V}$) = SpecularRemoval($\boldsymbol{S}$)
      ($\boldsymbol{S}'$, $\boldsymbol{L}'$) = Subtraction($\boldsymbol{S}_D$, $\boldsymbol{V}$, $\boldsymbol{L}$)
      ($\boldsymbol{D}$, $\boldsymbol{N}$, $\boldsymbol{R}$) = SimulEstimation($\boldsymbol{S}'$, $\boldsymbol{L}'$)
    end for
  end for
end for

Figure 3.7: Algorithm of simultaneous estimation with the color based approach.

where τ is a threshold for photo consistency. The RANSAC process above computation is repeated to find the best estimates of $\boldsymbol{n}_p$, $\boldsymbol{\rho}_p$, and $\boldsymbol{a}_p$ that maximizes N at each p and depth label j. Finally, the photo consistency cost E_p is evaluated as

$$E_p(p, j) = \eta \frac{1}{N} \sum_{i \in N} g_i(\boldsymbol{n}_p, \boldsymbol{\rho}_p, \boldsymbol{a}_p) - N, \tag{3.20}$$

where η is a scaling constant. The first term in the cost function assesses the overall photo consistency, and the second term evaluates the reliability of the photo consistency, *i.e.*, when it is supported by many views (number of N), it is more reliable. These two criteria are combined together using a scaling constant term η. In our implementation, we fixed η as $\eta = 1/\tau$.

Color based approach

Our photo consistency measure is modified slightly for the color based approach. To explicitly handle outliers, we select three subtracted observations $O_i(p)$ and use a RANSAC approach to obtain an initial guess of scaled normal $\boldsymbol{b}_p$ using the near-light photometric stereo assumption described in Section 3.2.1. Using this initial guess, instead of Eq. (3.18),the photo consistency g is checked against each of the other subtracted observations at a given pixel p as

$$g_i(\boldsymbol{b}_p) = |O_i(p) - (\boldsymbol{l}_i'^T - \boldsymbol{l}_0'^T)\boldsymbol{b}_p|. \tag{3.21}$$

Similar to Eq. (3.19), we also compute the number of images that satisfy the photo consistency N as

$$N = |\{i \,|\, g_i(\boldsymbol{b}_p) < \tau\}|, \tag{3.22}$$

where τ is a threshold for photo consistency. The RANSAC process above computation is repeated to find the best estimates of $\boldsymbol{b}_p$ that maximizes N at each p and depth label j. Finally, similar to Eq. (3.20), the photo consistency cost E_p is evaluated as

$$E_p(p, j) = \eta \frac{1}{N} \sum_{i \in N} g_i(\boldsymbol{b}_p) - N. \tag{3.23}$$

Surface normal constraint

Preferred depth estimates are those which are consistent with the surface normal estimates. We use a surface normal cost function $E_n(p, j)$ to enforce this criterion. Let j' be the depth label of the neighboring pixel p' that is located nearest in 3-D coordinates to the plane specified by the site (p, j) and its surface normal as explained in details in Appendix 3.A. Sometimes, the site (p', j') does not have a valid surface normal due to unsuccessful fitting of a surface normal by RANSAC. In that case, we take the next nearest site as (p', j'). Once the appropriate j' is found within $|j - j'| < T_j$, a vector $\boldsymbol{d}_{(p,j)}^{(p',j')}$ that connects (p, j) and (p', j') in the 3-D coordinate is defined on the assumed plane. We then compute the agreement of the surface normal at (p', j') with the depth estimate by evaluating if these two vectors are perpendicular to each other. The surface normal cost function is defined as

$$E_n(p, j) = \begin{cases} \sum_{p'} (|j - j'| + 1) \boldsymbol{n}_{p'j'} \cdot \boldsymbol{d}_{(p,j)}^{(p',j')} & \text{if } |j - j'| < T_j \\ C_0 \, (= \text{const.}) & \text{otherwise.} \end{cases}, \tag{3.24}$$

Smoothness constraint

We use a smoothness constraint on depth to penalize large discontinuities. Suppose p and p' are neighboring pixels whose depth labels are j and j' respectively. The smoothness cost function E_s is defined as

$$E_s(j, j') = |z_j - z_{j'}| = \Delta_z |j - j'|. \tag{3.25}$$

Energy function

Finally, the energy function E is defined by combining above three constraints as

$$E(p, j, j') = E_p(p, j) + \lambda_n E_n(p, j) + \lambda_s E_s(j, j'). \tag{3.26}$$

We use a 2-D grid graph cut framework to optimize the energy function. The 2-D grid corresponds to the pixel grid, *i.e.*, we define each pixel p as a site and the depth label j is associated. We use Boykov *et al*. [BVZ01, KZ04, BK04]'s graph cut implementation to solve the problem. By solving Eq. (3.26), we obtain the estimates of depth, surface normal, surface albedo, and ambient lighting.

3.4 Shape Refinement

The depth estimate obtained by the solution method described in the previous section is discretized, and therefore it is not completely accurate due to the quantization error. To refine the depth estimate, we perform a regularized minimization of a position error, normal constraint, and smoothness penalty, to derive the optimal surface Z. The optimization method is based on Nehab *et.al.* [NRDR05], and we define the error function following the work of Joshi and Kriegman [JK07]:

$$J(Z) = E^P + E^N + E^S. \tag{3.27}$$

The position error E^P is the sum of squared distances between the optimized positions S_p and original positions S'_p in the 3-D coordinate:

$$E^P = \lambda_1 \sum_p \|S_p - S'_p\|^2, \tag{3.28}$$

where λ_1 is the relative weighting of the position constraint versus the normal constraint. To evaluate the position error, depth values are transformed to distances from the center of the perspective projection:

$$\begin{aligned} \|S_p - S'_p\|^2 &= \mu_p^2 (z_p - z'_p)^2, \\ \mu_p^2 &= \left(\frac{x}{f_x}\right)^2 + \left(\frac{y}{f_y}\right)^2 + 1, \end{aligned} \tag{3.29}$$

where f_x and f_y are the camera focal lengths in pixels, and z'_p is the depth value of the original position p'.

The normal error constrains the tangents of the final surface to be perpendicular to the input normals:

$$E^N = (1 - \lambda_1) \sum_p \left(\left(\boldsymbol{n}_p \cdot \boldsymbol{T}_p^x \right)^2 + \left(\boldsymbol{n}_p \cdot \boldsymbol{T}_p^y \right)^2 \right), \tag{3.30}$$

where $\boldsymbol{T}_p^x$ and $\boldsymbol{T}_p^y$ represent the tangent vectors:

$$\boldsymbol{T}_p^x = \left[-\frac{1}{f_x}\left(x \frac{\partial Z_p}{\partial x} + Z_p \right), -\frac{1}{f_y} y \frac{\partial Z_p}{\partial x}, \frac{\partial Z_p}{\partial x} \right]^T, \tag{3.31}$$

$$\boldsymbol{T}_p^y = \left[-\frac{1}{f_x} x \frac{\partial Z_p}{\partial y}, -\frac{1}{f_y}\left(y \frac{\partial Z_p}{\partial y} + Z_p \right), \frac{\partial Z_p}{\partial y} \right]^T. \tag{3.32}$$

The smoothness constraint penalizes high second-derivatives by penalizing the Laplacian of the surface:

$$E^S = \lambda_2 \sum_p \nabla^2 Z_p. \tag{3.33}$$

λ_2 is a regularization parameter to control the amount of smoothing.

Each pixel generates at most 4 equations: one for the position error, one for the normal error in each of x and y directions, and one for the smoothness. Therefore, the minimization can be formulated as a large, sparse over-constrained system to be solved by least squares:

$$\begin{bmatrix} \lambda_1 \mathcal{I} \\ (1-\lambda_1)\mathcal{N}\cdot\mathcal{T}^x \\ (1-\lambda_1)\mathcal{N}\cdot\mathcal{T}^y \\ \lambda_2 \nabla^2 \end{bmatrix} [Z] = \begin{bmatrix} \lambda_1 z \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{3.34}$$

where $\mathcal{I}$ is an identity matrix and $\mathcal{N}\cdot\mathcal{T}^x$ and $\mathcal{N}\cdot\mathcal{T}^y$ are matrices that, when multiplied by the unknown vector Z, evaluate the normal constraints $(1-\lambda_1)\boldsymbol{n}\cdot T^x$ and $(1-\lambda_1)\boldsymbol{n}\cdot T^y$. We solve this system using a conjugate gradient method for sparse linear least squares problems [PS82].

3.5 Implementation

3.5.1 Calibration

Before data acquisition, we calibrate the intrinsic parameters of the camera and vignetting. We use Camera Calibration Toolbox for Matlab [Bou07] to estimate the camera intrinsics. For vignetting correction, we take images under a uniform illumination environment with a diffuser to create a vignetting mask. During the data acquisition, we move the camera system with the LED light on, without changing the intrinsic parameters of the camera.

3.5.2 Structure from motion

From the image sequence, we use the state-of-the-art structure from motion implementation *Bundler* [SSS06] to estimate camera extrinsics and 3-D positions of feature points. Here, all we need is camera extrinsics with absolute scale: the scale of camera position should be the same as the scale of the measured distance between the camera and the attached LED light source. You can use any structure from motion or any SLAM method instead of *Bundler*.

Unfortunately, the estimated 3-D positions of feature points have a scaling ambiguity because of the fundamental ambiguity of structure from motion. We solve the

ambiguity with the following two ways: one uses photo consistency and the other uses measured distance. The scale k can affect the light vector estimation in Eq. (3.1) as

$$\boldsymbol{l}_i = \boldsymbol{s} - k(\mathbf{R}_i\boldsymbol{x} + \boldsymbol{t}_i). \tag{3.35}$$

We resolve this ambiguity using our photo consistency measure on feature points $\mathcal{F}$. The photo consistency cost E_p of Eq. (3.20) varies with the scaling parameter k. We find the optimal k that minimizes the score of $E_p(k)$ using the feature points $\mathcal{F}$ as

$$E_p(k) = \sum_{p\in\mathcal{F}}\Big[\eta\frac{1}{N}\sum_i g_i(\boldsymbol{n}_p, \boldsymbol{\rho}_p, \boldsymbol{a}_p) - N\Big]. \tag{3.36}$$

We minimize $E_p(k)$ by simply sweeping the parameter space of k to obtain the solution.

On the other hand, if distance between two feature points is known as absolute value, we can solve the scaling parameter with normalization of the known distance.

3.5.3 Coarse-to-fine implementation

The simultaneous estimation method described in Section 3.3 gives good estimates; however, the computational cost becomes high when the image resolution is large and also when many depth labels are considered. We adopt a coarse-to-fine approach to avoid this issue.

First, image pyramids are created for the registered images after image warping by Eq. (3.17). At the coarsest level, the simultaneous estimation method is applied using full depth labels. In the finer level of the pyramid, we expand the depth labels from the earlier level and use them as the initial guess. From this level, we prepare only a small range of depth labels around the initial guess for each site p. Using the minimum and maximum depth labels, $j_{\min}$ and $j_{\max}$, of the site and its neighboring sites, the new range is defined as $[j_{\min} - 1, j_{\max} + 1]$. We also use a finer Δ_z in the finer level of the pyramid. We set $\Delta_z \leftarrow \Delta_z/2$ when moving to the finer level of the pyramid.

3.5.4 Full 3-D reconstruction

Our method estimates one depth image and normal map from one reference view, as discussed above. By making depth and normal maps from several different reference views, we can get a full set of surface points with surface orientations. Then we use the Poisson surface reconstruction method [KBH06] to reconstruct a full 3-D shape.

However, simultaneous estimation using input images for full 3-D reconstruction has another problem in that a lot of occlusions occur and many outliers are therefore created, so it is difficult to apply our method as it is. For reducing the number of outliers resulting from occlusions, a simple view constraint is applied. For full 3-D reconstruction, we capture images on rings around the target object from different elevation angles. Since we do not know whether a point on the surface is visible or not from a particular view point, we simply assume that a similar view from the reference view is less likely to have occlusion. Suppose θ_i is an angle between a particular view direction i and the reference view direction, the view constraint is defined that we only use image i satisfying $\theta_i < T_\theta$ for simultaneous estimation. In our experiments, we use $T_\theta = 45$ [deg.].

3.6 Experiments

We use a Point Grey DragonFly camera (640×480) with an attached point light source as our prototype system. The camera can sequentially capture images, and we use this capability for ease of data acquisition. During the capturing process, the point light source is always turned on. We also use a commercially available Nikon D1 camera (1024×672 resized) with a camera flash. The camera records images from different view points with the camera flash on.

In this section, we show the effectiveness of our both the intensity based method and the color based method with the efficient algorithm formulation as shown in Section 3.2.3. We first show quantitative evaluation using synthetic data in Section 3.6.1. We use three real-world scenes that have different properties to verify the applicability of the proposed method in Section 3.6.2. We further show comparisons with other state-of-the-art 3-D modeling methods using the real-world scenes. Finally we show a full 3-D reconstruction result. Throughout the experiments, we use $\tau = [6.0, 8.0]$, $\lambda_n = 7.5$ and $\lambda_s = [1.5, 3.0]$, $\lambda_1 = [0.01, 0.1]$, $\lambda_2 = [0.5, 1.5]$, $C_0 = 5$, and initial $\Delta_z = 8.0$[mm].

3.6.1 Simulation results

In the simulation experiments, we render synthetic scenes by simulating the configuration of our photometric stereo camera. We created a baseline scene which is textured, Lambertian, and has no ambient lighting. By changing the settings so that the objects

Approach	Condition	Depth [%]		Normal [deg.]		Albedo [%]	
		mean	med	mean	med	mean	med
Intensity based	Baseline	1.73	**0.42**	10.5	**4.27**	9.14	4.95
	Textureless	3.05	0.46	11.2	4.74	9.23	4.99
	Specular	1.77	**0.42**	10.0	4.63	9.43	5.38
	Ambient	2.68	0.47	10.0	4.44	9.44	5.09
Color based	Baseline	1.20	0.46	**8.71**	4.30	**8.51**	**4.44**
	Textureless	**1.05**	0.45	8.96	4.39	8.79	4.77
	Specular	1.07	0.45	9.02	4.51	9.38	4.91
	Ambient	1.28	0.66	11.0	5.49	11.7	6.92

Table 3.1: Quantitative evaluation using synthetic scenes. "mean" and "med" indicate mean and median errors, respectively. The upper group is estimated with the intensity based method, the lower group is estimated with the color based method.

were (1) textured, (2) have specular reflectance, and (3) the scene has ambient lighting, we are able to assess the performance variation in comparison with the baseline case.

Table. 3.1 shows the summary of the evaluation. The upper side is the intensity based approach, while the lower side is the color based approach. In each side, from top to bottom, the results of the baseline, textureless, specular, and ambient cases are shown. The errors are evaluated using the ground truth depth map, normal map, and albedo map by looking at the mean and median errors. The depth error is represented by percentage, using [maximum depth - minimum depth] as 100%. The surface normal error is evaluated by the angular error in degrees, and albedo error is computed by taking the average of the percentage difference in R, G, and B channels, using the ground truth of the reflectance as 100%. The mean error is sensitive to outliers, while the median error is not. Looking at the median error, the estimation accuracy is quite stable across the table. The ambient case in the color based approach produces slightly larger errors, and this indicates that the process of canceling ambient effect described in Section 3.2.3 tends to affect errors. Fig. 3.8 shows the result on the simulated scene with specularity. The upper side is the intensity based result, while the lower side is the color based result.

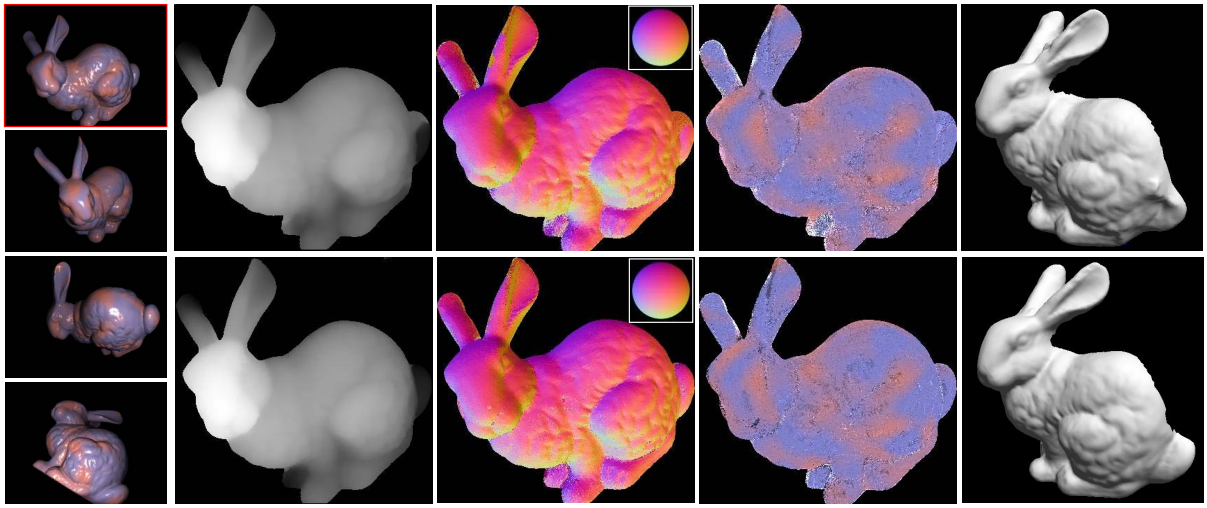


Figure 3.8: Simulation result using the bunny scene. The left images are input images (reference view in the top). The top images are results of the intensity based method, the bottom images are results of the color based method. From left to right, the estimated depth map, normal map, albedo, and a final rendering of the surface are shown. In the depth map, brighter is nearer and darker is further from the camera. In the normal map, a reference sphere is placed for better visualization. 62 images are used as input.

3.6.2 Real-world results

We applied our method to various different real-world scenes. We show three scenes: (1) a statue scene (textureless, roughly Lambertian), (2) a bag scene (textured, glossy surfaces), and (3) a toy scene (various reflectance properties, complex geometry).

Fig. 3.9 shows the result of the statue scene. The left images are input images, the upper side is the intensity based method, and the lower side is the color based method. To produce the result, we manually masked out the background portion of the statue in the reference image. Our method can recover the surface and normal map as well as surface albedo from a textureless scene. Fig. 3.10 and Fig. 3.11 show the results of the bag scene and toy scene, respectively. These scenes contain textured surfaces as well as specularities. Our method can handle these cases as well because of our robust estimation scheme to handle specularities. Our handheld camera is particularly useful for measuring scenes like the toy scene that are difficult to move to a controlled setup.

To demonstrate the effectiveness of our photometric constraint, we have performed

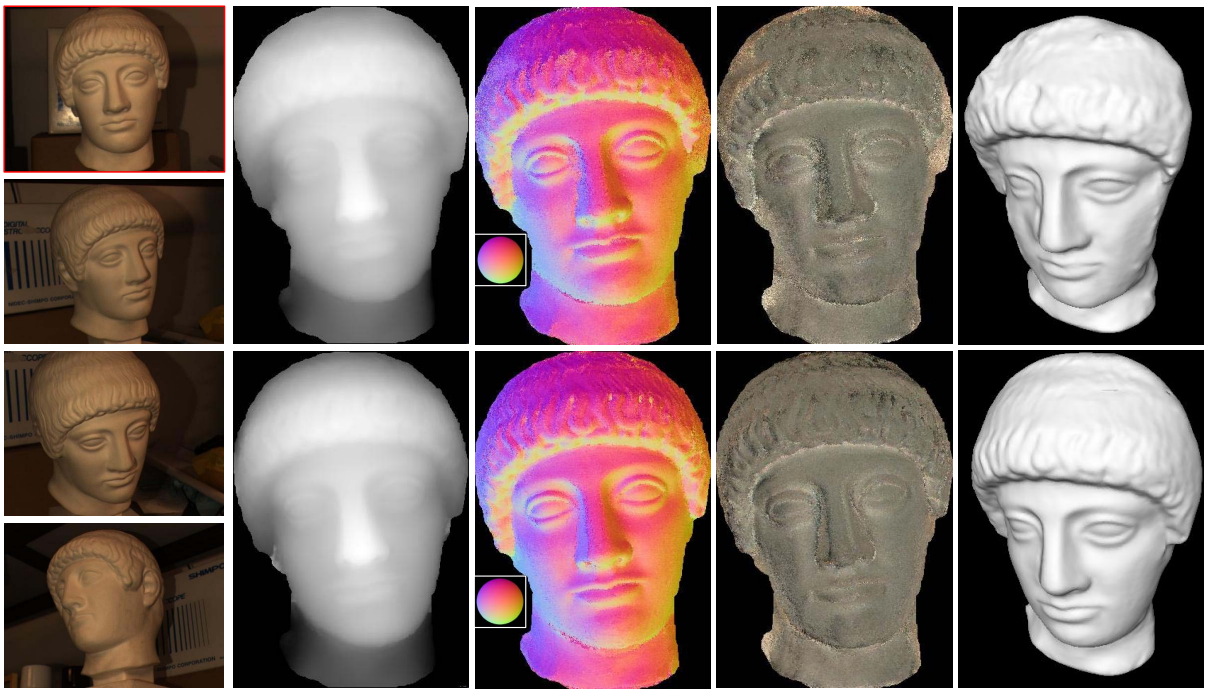


Figure 3.9: Result of the statue scene. The left images are input images (reference view in the top). The top images are results of the intensity based method, the bottom images are results of the color based method. From left to right, the estimated depth map, normal map, albedo, and a final rendering of the surface are shown. 93 images are used as input.

a comparison with a state-of-the-art multi-view stereo method [GCS06] that does not use a photometric constraint. The input data is obtained by fixing a camera at each view point and capturing two images with the attached point light source on and off. The images without the point light source but under environment lighting are used as input for the multi-view stereo method. Fig. 3.12 shows the rendering of three surfaces recovered by both our intensity based and color based method and the multi-view stereo method. Typical multi-view stereo algorithms can only establish a match in areas with some features (texture, geometric structure, or shadows), and this example is particularly difficult for them as it lacks such features in the most of the areas. On the other hand, our method works well because of the photometric constraint.

We also compare our normal method to a result from Joshi and Kriegman's method [JK07] using a gray-scale image sequence. In their method, far-distant lighting and ortho-

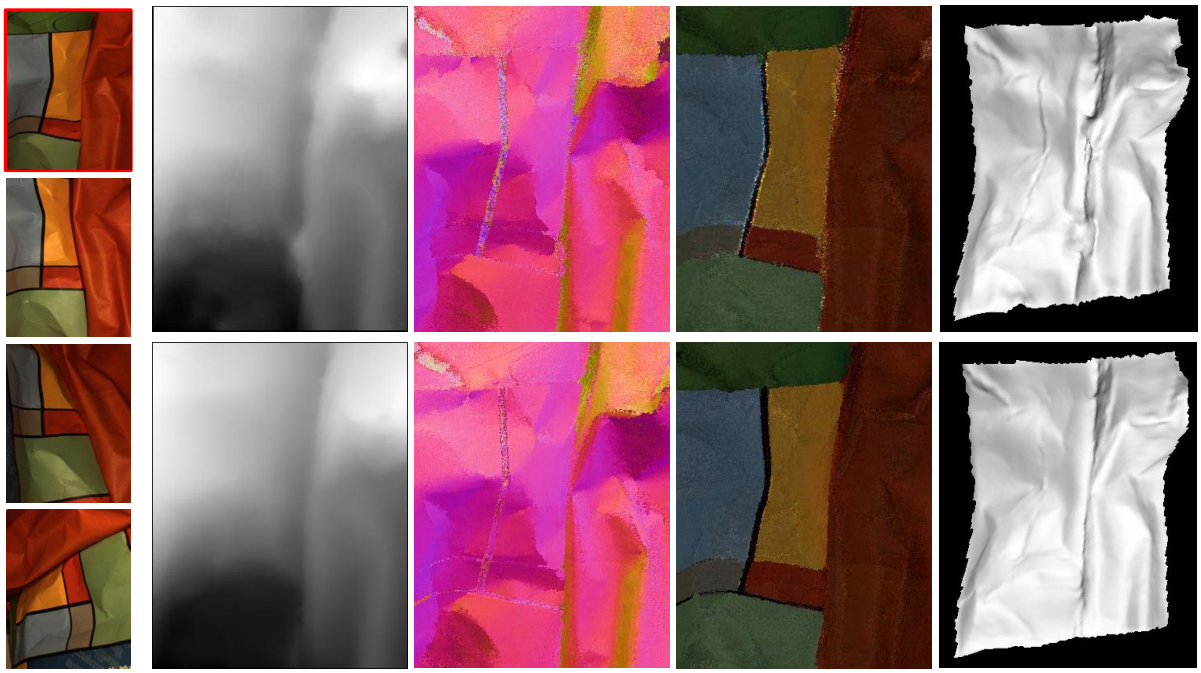


Figure 3.10: Result of the bag scene. The left images are input images (reference view in the top). The top images are results of the intensity based method, the bottom images are results of the color based method. From left to right, the estimated depth map, normal map, albedo, and a final rendering of the surface are shown. 65 images are used as input.

graphic projection are assumed. We use the same dataset from their experiment and approximate their assumptions by diminishing light-fall off term $(1/|\mathbf{l}_i|^2)$ in Eq. (3.2) and using large focal lengths f_x and f_y. The side-by-side comparison is shown in Fig. 3.13. Our intensity based method can produce a result with equal quality to their method.

Fig. 3.14 shows the result of the dinosaur scene for full 3-D reconstruction with the commercial camera attached the camera flash. To get this result, we merge eight reference views. Although this scene contains specularities, our method can handle this well, achieving specular removal. In the synthetic images, we mapped estimated albedo on the surface and use manually adjusted reflection parameters of specularity.

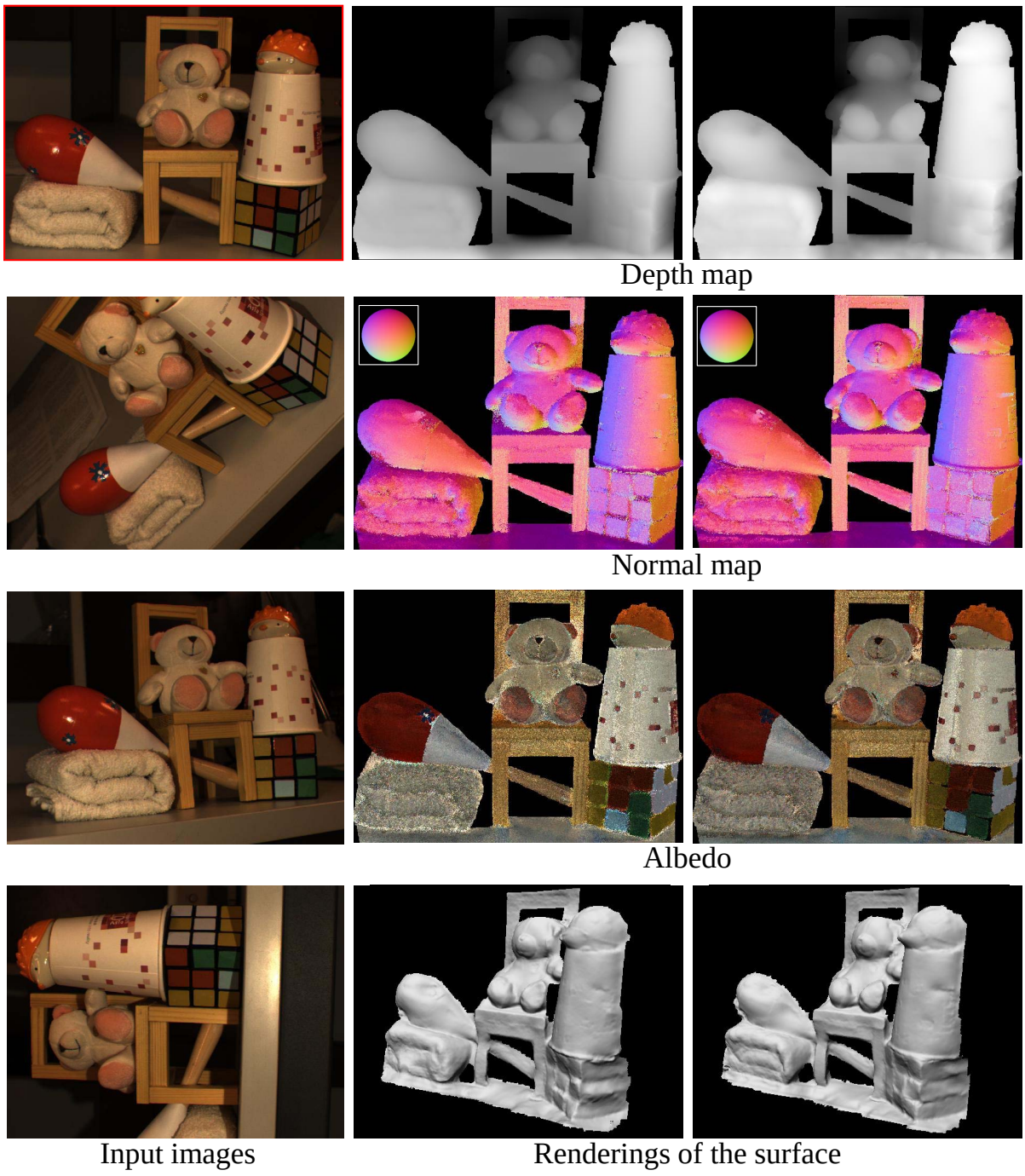
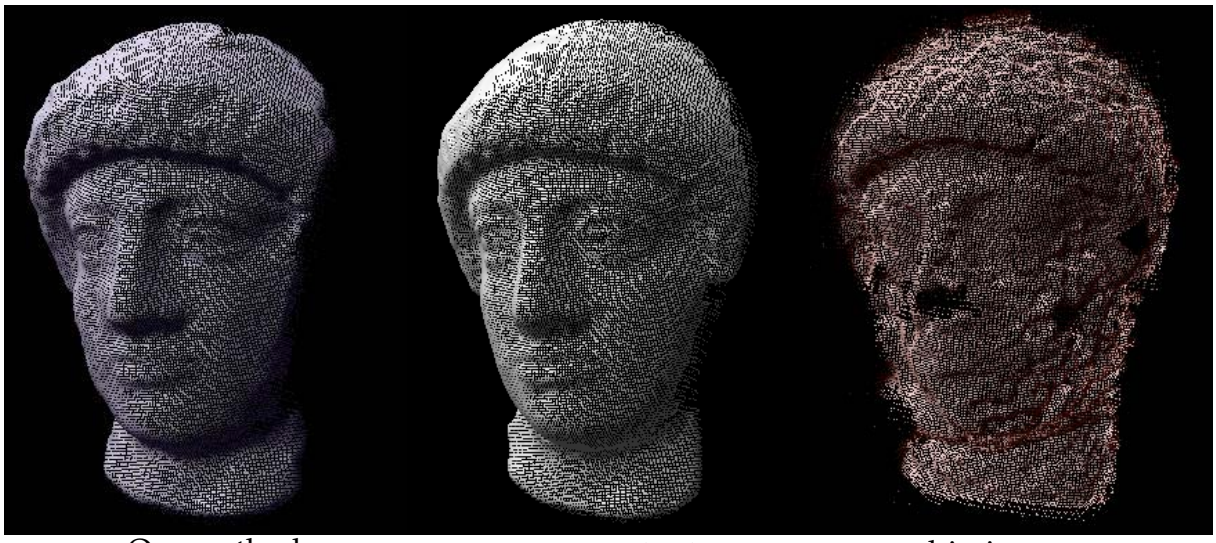


Figure 3.11: Result of the toy scene. The scene contains various color and reflectance properties. The left images are input images (reference view in the top). The middle column is the intensity based method, the right column is the color based method. From top to bottom, estimated depth map, and normal map, the estimated albedo map, and renderings of the final surface. 84 images are used as input.



Our method intensity based | Our method color based | multi-view stereo method [GCS06]

Figure 3.12: Comparison with a multi-view stereo method without a photometric constraint [GCS06] using the statue scene. 93 images are used as input for both methods.

3.7 Discussion and Future Work

We presented a simple, low-cost method for high-quality object shape and reflectance acquisition using a hand-held camera with an attached point light source. Our system is more practical than those in previous work and can handle hand-held filming scenarios with a broad range of objects under realistic filming conditions. Moreover, we extend our method to make it robust and accurate using color information and to reduce computational cost using the efficient computation. Nevertheless, there are some limitations and several avenues for future work.

One current limitation is that we only implicitly account for self-occlusions, shadowing, and inter-reflections. Our robust fitting method addresses these properties by treating them all as outliers from a Lambertian shading model. The view constraint also handles occlusions by simply removing inappropriate views against the reference view. While this works well in practice, it is very likely that explicitly accounting for these factors would improve our results. We are investigating methods that could be used to explicitly model outlier pixels as self-occlusions, shadows, and inter-reflections [BP03, CKK05, CAK07]. Not only would this help refine the 3-D shape and reflectance model, but it should also enable higher quality rendering of scanned objects.

Input image | JK [JK07] | Our method intensity based

Figure 3.13: Comparison with Joshi and Kriegman's method (JK) using the cat scene. Eight images are used as input for both methods. Note that rendering parameters are different as the original parameters are not available.

3.A How to find the depth label j' for the surface normal constraint

In this appendix, we present how to find the depth label j' for the surface normal constraint described in Section 3.3. As we mentioned in the section, we use a 2-D grid graph cut framework for energy minimization problem. The 2-D grid corresponds to the pixel grid: we define each pixel p as a site and the depth label j is associated. Here, j' is the depth label of the neighboring pixel p' that is located nearest in 3-D coordinates to the plane specified by the site (p, j) and its surface normal $\boldsymbol{n}$ as shown in Fig. 3.15.

Consider the horizontal neighboring pixel p'. The origin of the image plane is the center of the image for p and p'. Suppose z_j is depth corresponding to the depth label j, f is a focal length of the reference camera, a line l_p and $l_{p'}$ represent rays, and a dashed line l_n is an orthogonal line to the surface normal $\boldsymbol{n}$. We can calculate depth $\tilde{z}$ of an intersection between $l_{p'}$ and l_n as

$$\tilde{z} = \frac{(n_z/n_x)f + p}{(n_z/n_x)f + p'} z_j, \tag{3.37}$$

where n_x and n_z are respectively x and z component of the surface normal $\boldsymbol{n}$. Here, $p' = p_\pm = p \pm 1$ because p' is neighboring pixel of p. Using $z_j = z_0 + j\Delta_z$, then, $\tilde{z}$ is

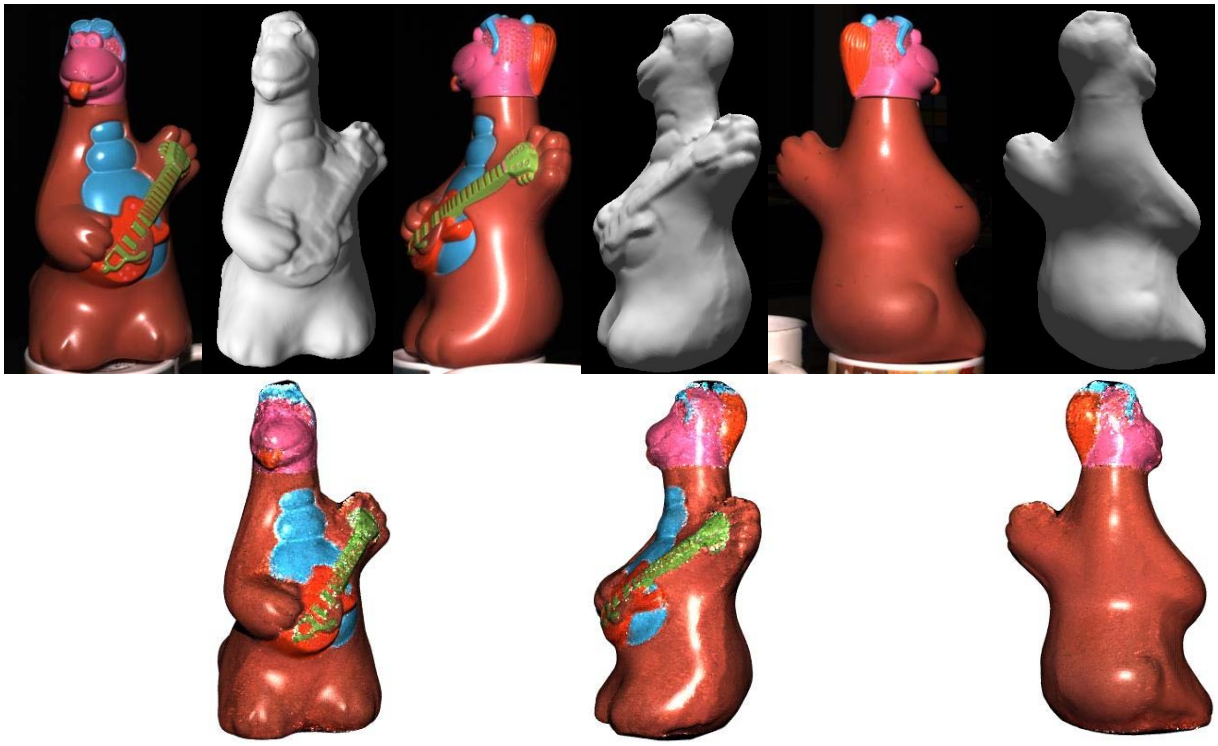


Figure 3.14: Results of the full 3-D reconstruction. The scene is captured with a commercially available Nikon D1 camera with a camera flash. The scene contains specularities. Top figures show input images and the 3-D model with no textured rendering. Bottom figures show the 3-D model mapped with estimated diffuse albedo. In the 3-D model, specular reflection parameters are manually adjusted. 86 images are used as input.

described as

$$\begin{aligned} \tilde{z} &= \left(1 \mp \frac{1}{(n_z/n_x)f + p_\pm}\right)(z_0 + j\Delta_z) \\ &= z_0 + \left\{ j \mp \left(\frac{z_0}{\Delta_z} + j\right)\frac{1}{(n_z/n_x)f + p_\pm}\right\}\Delta_z. \end{aligned} \tag{3.38}$$

Suppose $\lfloor \alpha \rfloor$ represents maximum integer not greater than α, we can calculate the integer depth label j' as follows:

$$j' = j + \left\lfloor \mp\left(\frac{z_0}{\Delta_z} + j\right)\frac{1}{(n_z/n_x)f + p_\pm} + 0.5 \right\rfloor. \tag{3.39}$$

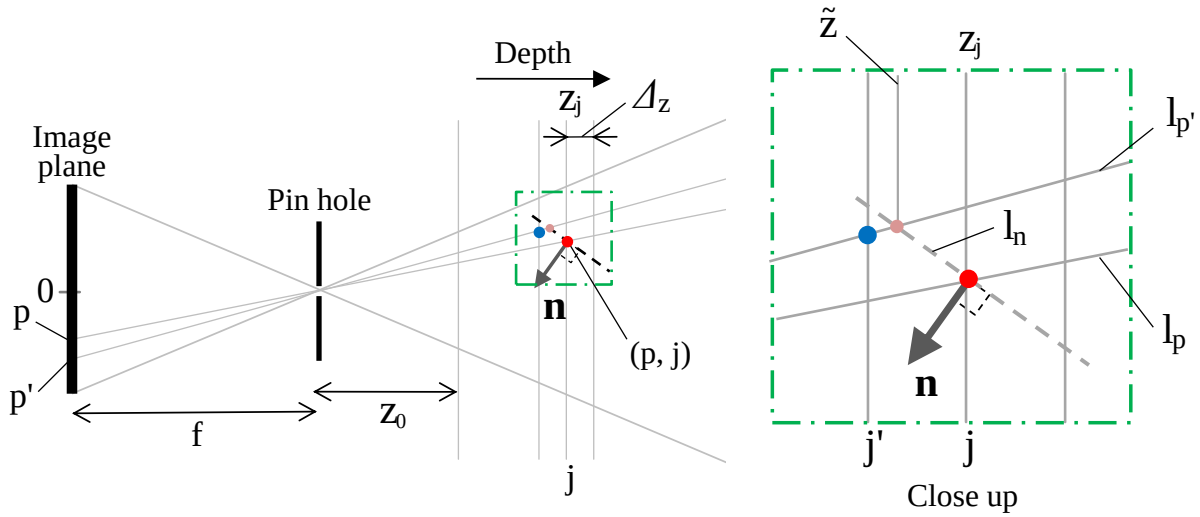


Figure 3.15: Find the depth label j' for the surface normal constraint. Top figure shows an overview and the bottom figure shows the close up around the site (p, j).

Chapter 4

Real-time Specular Removal

From one input image taken under uniform illumination, the specular reflection component can be removed while preserving shading information. This chapter proposes a method for specular removal using a color space based on a dichromatic reflection model and a neutral interface reflection assumption. The novelty of the method is that it uses the color space to remove the specular component very quickly to generate a specular-free image. A specular-free image has no specularity and has a different diffuse albedo from the original. Since it also preserves shading information based on the Lambertian low, it is appropriate for an input image of photometric stereo.

The handling of a specular-free image and specular reflection component in other chapters is based on this chapter.

4.1 Introduction

In inhomogeneous objects, reflections are linear combinations of diffuse and specular reflection components. We call this the dichromatic reflection model. Diffuse reflection represents object color and is scattered in all directions with the same intensity. On the other hand, specular reflection is a mirror-like reflection, and the color of it is the same as the color of the light source. This is known as a neutral interface reflection assumption. Since the appearance of specular reflection is different in both view direction and light source direction, specular reflections often cause error or outliers in various methods based on diffuse reflections, *e.g.*, object recognition, photometric stereo, and stereo matching. Therefore, a number of methods have been proposed to separate or remove

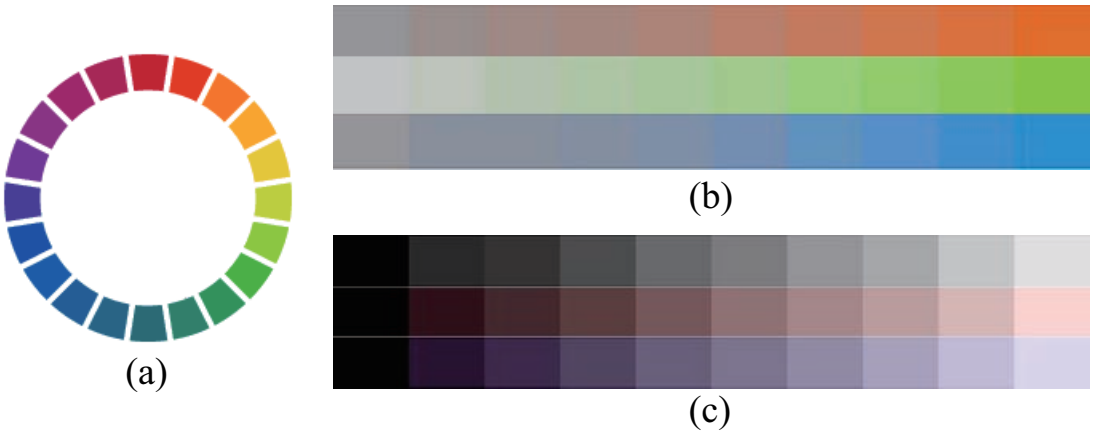


Figure 4.1: Color components. (a): Hue, (b): Saturation, (c): Brightness

specular reflection: using a polarizer [Wol89, NFB97, MTHI03], using more than one image [SI94, LS01, SKS*02], and using only one image [Sha85, KSK88, BLL96, TLQS03, TI05, MZKB05].

In this chapter, we propose a fast specular removal method using our color space. Moreover, we show how to generate a specular-free image which is appropriate for input of photometric stereo. The specular-free image is a specular removal image but has a different diffuse color from the original image while preserving shading information.

The rest of this chapter is as follows. We first describe our color space in Section 4.2. We formulate the specular removal method in Section 4.3 and present results in Section 4.4 followed by discussions and conclusions.

4.2 Color Space

4.2.1 Hue, Saturation, and Brightness

When we try to understand a *color* intuitively, three components of color are generally used as shown Fig. 4.1; hue, saturation, and brightness. Hue is one of the main properties of color that is described as red, green, blue, and yellow. Saturation is a colorfulness of color relative to its own intensity. Brightness is an intensity of color. These three properties are easily and intuitively understood for color representation. Based on these three color components of hue, saturation, and brightness, we define our own color space to easily remove specular components in real time.

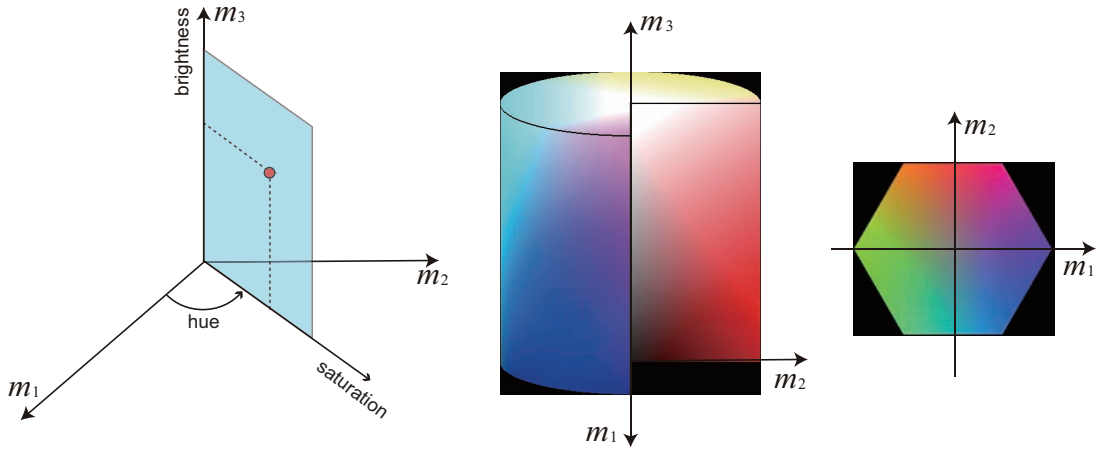


Figure 4.2: Proposed color space

4.2.2 Proposed Color Space

Fig. 4.2 shows our color space. In this color space, hue represents the azimuthal angle from m_1 axis on the $m_1 - m_2$ plane. Saturation is defined as the distance from the origin on the $m_1 - m_2$ plane. Brightness represents the m_3 component.

Similar to the S space proposed by Bajcsy [BLL96], this color space is good for handling hue, saturation, and brightness. Unlike the S space, our color space is easy to convert to from RGB color space. Moreover, since the color space has a symmetric property, it is easily and intuitively understood and analyzed. Though our color space is also similar to HSI color space, the definition of saturation is different. Because of the difference, our method can remove specular components with first-order approximation, as described in Section 4.3.1.

Our color space whose axes are (m_1, m_2, m_3) is described with RGB color space as

$$\begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} = \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} r \\ g \\ b \end{pmatrix}. \tag{4.1}$$

Then hue, saturation, and brightness are calculated as follows:

$$\text{hue} = \arctan \frac{m_2}{m_1} \tag{4.2}$$

$$\text{saturation} = \sqrt{m_1^2 + m_2^2} \tag{4.3}$$

$$\text{brightness} = m_3 \tag{4.4}$$

4.2.3 Correcting White Balance

Our method of specular removal assumes a white light source. Therefore, in case the color of the light source is not white, we need to correct the white balance as the light source should theoretically be dealt as white. As discussed below, since correcting the white balance is a linear and invertible operation, it does not interfere at all with converting observations from the RGB color space to our color space and vice versa.

First we capture the image of the white reference under the color light source. Then the mean values of each RGB component on the white reference region are normalized as (l_r, l_g, l_b). Given an observation of the target input image (L_r, L_g, L_b) taken under the color light source, correcting RGB value $(\hat{L}_r, \hat{L}_g, \hat{L}_b)$ is calculated as follows:

$$\begin{pmatrix} \hat{L}_r \\ \hat{L}_g \\ \hat{L}_b \end{pmatrix} = \begin{pmatrix} L_r/l_r \\ L_g/l_g \\ L_b/l_b \end{pmatrix}. \tag{4.5}$$

4.3 Real-time Specular Removal

In this section, we describe and prove the theory of our proposed method for specular removal from the color space. After that we describe the algorithm of our real-time specular removal system.

4.3.1 Theory

For specular removal, we assume the following three assumptions:

Assumption 1. The target scene is illuminated with uniform color light source.

Assumption 2. The color of the specular components is the same as the color of the light source, *i.e.*, the neutral interface reflection assumption.

Assumption 3. Each hue has only one surface color.

Based on these assumptions, we explain how to remove specular components and prove that they have been removed.

First we project the RGB value of the target image to our color space. Let us focus on a plane defined with a particular hue value as shown in the left figure of Fig. 4.2. According to assumption 3, each plane represents surface color; we call the plane a

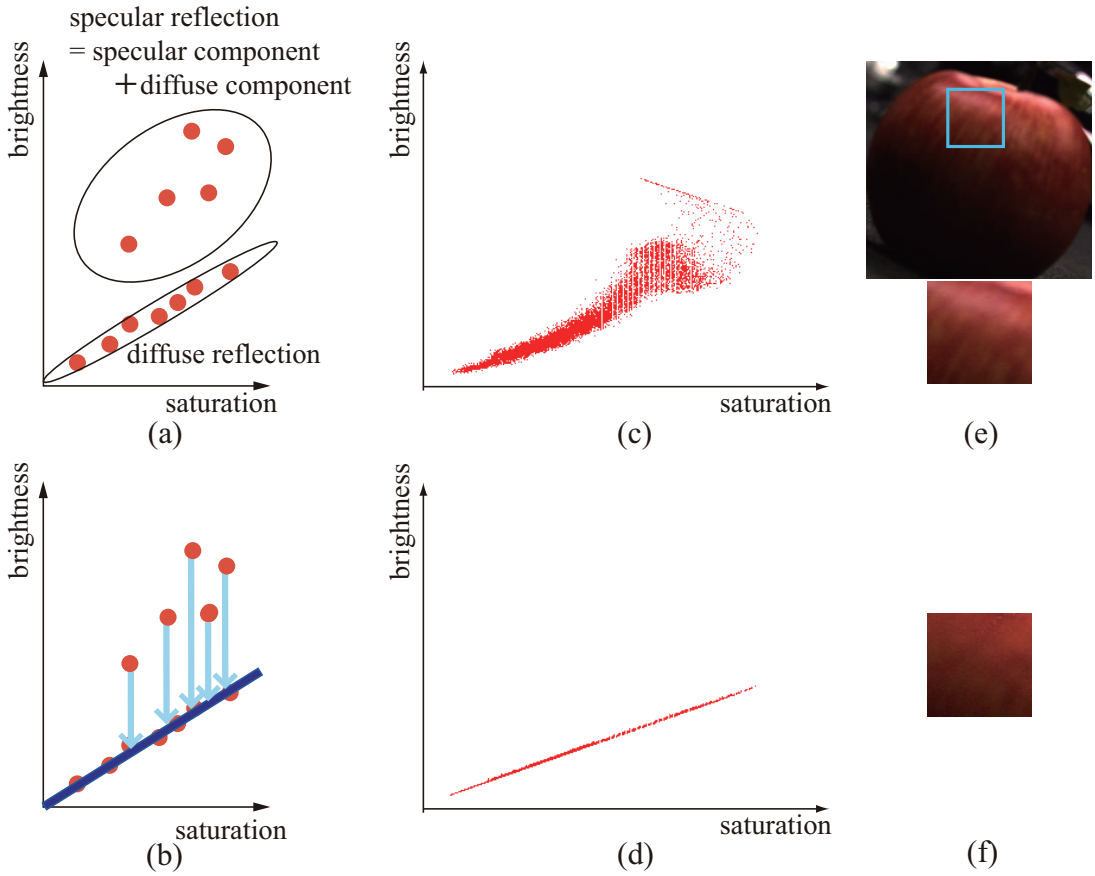


Figure 4.3: (a): plots on the surface color plane, (b): specular removal, (c): plots for the apple scene, (d): plots after the process of specular removal, (e): input image of (c), (f): result of the specular removal

surface color plane. Since the hue is a continuous value, an infinite number of the surface color planes may exist. Practically, we separate hue angle into finite regions and each region provides one surface color plane. So in each region, we define one surface color plane. On a surface color plane, the horizontal axis denotes saturation and the vertical axis denotes brightness. According to the property of our color space, plots of diffuse reflection components exist in a directly proportional line between saturation and brightness as shown Fig. 4.3 (a). Using this property, we remove the specular reflection components as follows. When the color of the light source is white, the color of specularity is also white and the color vector of specular reflection components is parallel to the brightness axis based on assumption 1 and assumption 2. Therefore, if two observations have the same saturation, their diffuse reflection components are also the same independent of the specular reflection components. In other words, we can

calculate the brightness of diffuse reflection components by fitting the line as shown in Fig. 4.3 (b), which means specular removal. If the color of the light source is not white, we should first correct the white balance using the procedure described in Section 4.2.3

For removing specular reflection based on the above theory, we have to prove the following three propositions.

Proposition I When two pixels have the same surface color but one has only diffuse reflection and the other has specular reflection, they have the same specified hue value independent of their shading.

Proposition II When two pixels have the same diffuse components but have different specular components, their saturation values are the same independent of the specular components.

Proposition III Plots of diffuse observations form a direct proportional line on the surface color plane.

According to assumption 3, each hue has only one surface color, so we have only to prove proposition II and III with regard to one particular hue value.

Let us begin with our image formation based on the dichromatic reflection model [Sha85]. Suppose $\boldsymbol{I_S}$ is a specular reflection three-dimensional vector, each component denotes red, green, and blue intensity respectively. Since the color of the light source becomes white by correcting the white balance and the color of the specular component is the same as the color as the light source based on the neutral interface reflection assumption, normalized $\boldsymbol{I_S}$ is described as

$$\boldsymbol{I_S} = s \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

where s is a scalar value that represents specular reflectance. Based on the Lambertian model, normalized diffuse reflection vector $\boldsymbol{I_D}$ is provided as

$$\boldsymbol{I_D} = \begin{pmatrix} \alpha_r \\ \alpha_g \\ \alpha_b \end{pmatrix} \cos\theta,$$

where θ is an angle between surface normal and the light direction vector, and $\cos\theta$ represents shading of the Lambertian model. $(\alpha_r, \alpha_g, \alpha_b)^T$ is respectively red, green,

and blue diffuse reflectance that defines surface color. Then normalized observation $(r, g, b)^T$ is calculated with dichromatic reflection model as

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} = s \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \alpha_r \\ \alpha_g \\ \alpha_b \end{pmatrix} \cos\theta. \tag{4.6}$$

in case $s = 0$, Eq. (4.6) denotes diffuse reflection, on the other hand, in case $s > 0$, Eq. (4.6) contains specularity.

Proof of proposition I

From Eq. (4.6) and Eq. (4.1), m_1 and m_2 are calculated as

$$\left.\begin{aligned} m_1 &= \left(\alpha_r - \tfrac{1}{2}\alpha_g - \tfrac{1}{2}\alpha_b\right)\cos\theta \\ m_2 &= \left(\tfrac{\sqrt{3}}{2}\alpha_g - \tfrac{\sqrt{3}}{2}\alpha_b\right)\cos\theta \end{aligned}\right\}. \tag{4.7}$$

Then we can calculate its hue value from Eq. (4.2) as

$$\text{hue} = \arctan\frac{m_2}{m_1} = \arctan\left(\frac{\frac{\sqrt{3}}{2}\alpha_g - \frac{\sqrt{3}}{2}\alpha_b}{\alpha_r - \frac{1}{2}\alpha_g - \frac{1}{2}\alpha_b}\right). \tag{4.8}$$

According to this equation, we can find that the hue value is specified with $(\alpha_r, \alpha_g, \alpha_b)$ independent of shading, *i.e.*, $\cos\theta$. Moreover, we can also find that Eq. (4.8) is independent of s, in other words, specular reflection components do not affect the hue value at all. Q.E.D.

Proof of proposition II

We can calculate saturation from Eq. (4.3) and Eq. (4.7) as

$$\begin{aligned} \text{saturation} &= \sqrt{m_1^2 + m_2^2} \\ &= \frac{\cos\theta}{\sqrt{2}}\sqrt{(\alpha_r - \alpha_g)^2 + (\alpha_g - \alpha_b)^2 + (\alpha_b - \alpha_r)^2}. \end{aligned} \tag{4.9}$$

According to this equation, saturation is defined independent of s. In other words, the saturation of diffuse reflection is the same as the saturation of the specular reflection. Q.E.D.

Proof of proposition III

For a particular hue, the relationship between saturation and brightness of the diffuse reflection is provided as follows. First we can calculate the brightness of the diffuse reflection as

$$\text{brightness} = m_3 = \frac{\alpha_r + \alpha_g + \alpha_b}{3} \cos\theta. \tag{4.10}$$

According to Eq. (4.9) and Eq. (4.10), the relationship between saturation and brightness is directly proportion as shown

$$\text{brightness} = A \times \text{saturation}, \tag{4.11}$$

where A is as follows:

$$A = \frac{\sqrt{2}}{3} \frac{\alpha_r + \alpha_g + \alpha_b}{\sqrt{(\alpha_r - \alpha_g)^2 + (\alpha_g - \alpha_b)^2 + (\alpha_b - \alpha_r)^2}}. \tag{4.12}$$

Therefore we can find that the relationship between saturation and brightness for a particular hue is directly proportion, and that its gradient A is specified with only $(\alpha_r, \alpha_g, \alpha_b)$. Q.E.D.

The proofs of proposition I, II , and III show that we can remove specular reflection components by the method shown above. Moreover, we can find that specular removal preserves shading information $\cos\theta$.

After calculating gradient A, we can get diffuse brightness m_3' as

$$m_3' = A \times \text{saturation}. \tag{4.13}$$

Then we convert estimated values from our color space to the RGB color space as

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & 0 & 1 \\ -\frac{1}{3} & \frac{1}{\sqrt{3}} & 1 \\ -\frac{1}{3} & -\frac{1}{\sqrt{3}} & 1 \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ m_3' \end{pmatrix}. \tag{4.14}$$

Meanwhile, a specular-free image is calculated in using any positive value as gradient A. Since it does not require any calculation about computing A, the specular-free image is generated faster than the specular removal image.

One limitation of our method is that it cannot handle grayscale pixels such as white, gray, and black. The reason is that grayscale pixels are not on the straight line shown in Fig. 4.3 (a), since grayscale pixels are $r \approx g \approx b$ and then saturation ≈ 0 from Eq. (4.3).

Of course our method cannot handle saturated pixels because saturated pixels do not follow any optical rules about the observations. Therefore we require careful attention not to exceed camera sensitivity against irradiance, so we have to adjust exposure time and lens diaphragm.

4.3.2 Algorithm

The algorithm of the specular removal system is as follows:

Step 1: convert pixel observations from RGB color space to our color space and calculate hue, saturation, and brightness.

Step 2: plot the observations on the surface color plane in each divided hue and then fit the direct straight line with diffuse pixels to get the gradient A.

Step 3: recalculate brightness of all pixels by using the gradient A.

Step 4: convert recalculated pixels from our color space to the RGB color space.

Concretely, first (**Step 1**) we convert pixel observations from RGB color space to our color space. Then we calculate hue, saturation, and brightness based on Eq. (4.2), Eq. (4.3), and Eq. (4.4). Second, (**Step 2**) we separate hue to create the surface color plane in each clustered hue. Then we plot the observations on the plane as shown in Fig. 4.3 (b). Here we only plot minimum brightness pixels in each saturation value in order to easily detect diffuse reflection. These plots form a direct proportional line as mentioned in proof of proposition III. Therefore by fitting the straight line, we calculate a gradient A in each hue. Since specular reflection and diffuse reflection have same saturation but have different brightness (**Step 3**), we can recover diffuse brightness for all pixels by using the gradient A. Finally, (**Step 4**) converting calculated observations from our color space to the RGB color space provides the specular removal image.

The computational cost of fitting the straight line involves linear order, so our specular removal achieves real-time performance.

4.3.3 Speed-up technique

The computational cost of our method is proportional to the size of the input image because our process is done with all pixels. However, one of the four steps, **Step 2**, which is a process for estimation of gradient A, can be speeded up. Not all plotted

Figure 4.4: Specular removal under white light source.

observations are required for the estimation of gradient A. For example, when we choose every two pixels to plot for straight line fitting, compared to using all pixels, we can reduce the cost of the fitting process by half. However, although using fewer sampled pixels for the line fitting achieve faster estimation, it could possibly have less accuracy. We should consider the trade-off between speed and accuracy.

Furthermore, if the user can confine regions for specular removal instead of all pixels, we can reduce the cost of all four steps.

4.4 Experiments

We captured the scenes using a camera (512×384): a Sony color digital camera XCD-X710CR that has a linear response function. The experiments were run on a laptop with 2.0GHz Intel Core 2 CPU. Our process of specular removal is as follows: first, **Step 1** to **Step 4** are done in numbered order. Then the output image is displayed, and after that a frame image taken at the same time is captured as an input image and returned back to **Step 1**. In case grayscale pixels are observed, it is left as it is to return the same color as the input. Since we tried to capture the image without a saturation of the camera, the image looked totally dark. Therefore, we used Photoshop to adjust the brightness with level correction for better visualization. We found that our method of specular removal can achieve real-time rendering at 8.8 fps.

Fig. 4.4 shows the result of the scene under white light source. The left figure

shows the input image and the right figure shows the result of specular removal with our proposed method. The bottom right figures are close-up results of the side surface of the cup. Our method can remove specular components while preserving shading information.

Fig. 4.5, Fig. 4.6, and Fig. 4.7 show the results under different color light sources. We correct white balance using a white reference. Output images are already corrected for white balance. Results under red light source and under green light source are somewhat noisy. The reason why is that input images are recorded with a short exposure time and are very dark and noisy because the scene under a red or green light source tends to be too bright for camera sensitivity. Fig. 4.5, Fig. 4.6, and Fig. 4.7 have also been adjusted with level correction for better visualization.

4.5 Discussion

We presented a real-time method for specular removal using our color space while fitting a straight line. Our method is practical because it is very fast and uses only one image with known color of illumination.

Our current limitation is that some target objects do not satisfy one of the three assumptions described in Subsection 4.3.1; each hue has only one surface color. In this case, we can still remove specular reflection components, but the color of diffuse reflections is different from the original input image.

For example, Fig. 4.8 shows an input image of the Macbeth color checker (left) and an output image processed by our method (right). Looking at the flesh color and brown patches at the bottom right in each image, these two colors have a different surface color but have the same hue. Therefore, we estimate a wrong gradient A and then the output image has a different surface color from the original one as shown in Fig. 4.9.

Figure 4.5: Specular removal under red light source.

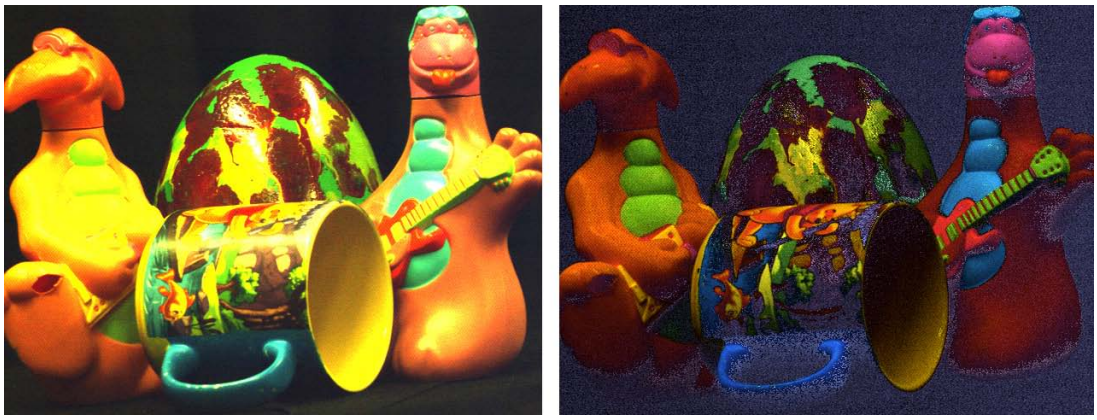


Figure 4.6: Specular removal under green light source.

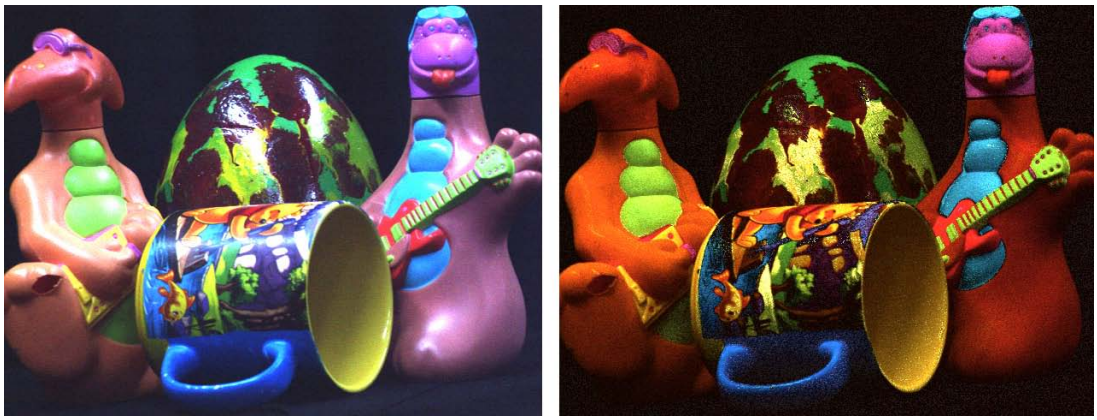


Figure 4.7: Specular removal under blue light source.

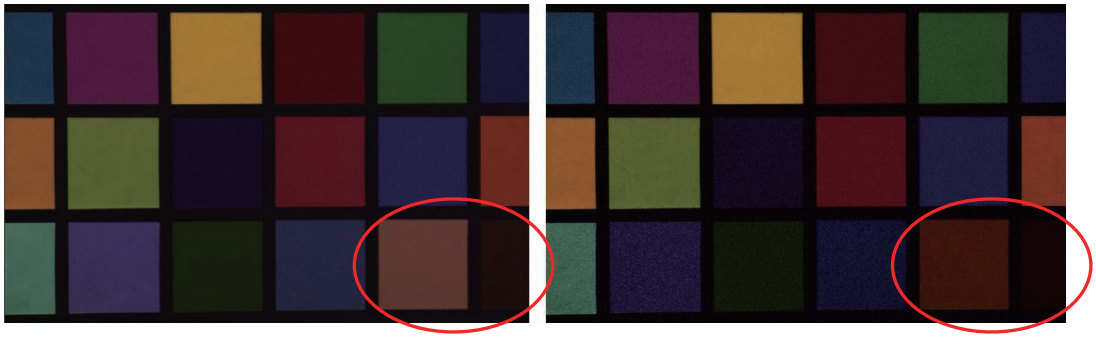


Figure 4.8: Example of two surface colors in one hue.

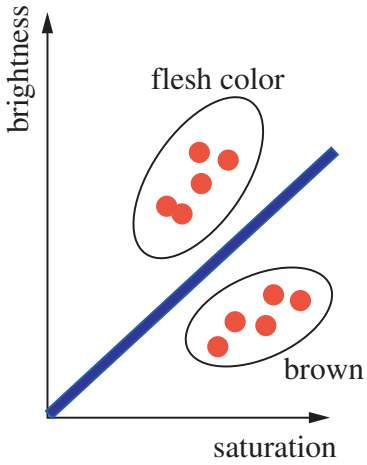


Figure 4.9: Estimation of gradient A for the case of two surface colors in one hue.

Chapter 5

Consensus Photometric Stereo for Non-Lambertian Surfaces

Reflection properties of monotonicity, visibility, and isotropy can be useful clues to estimate surface normal of a variety of surface reflectance conditions. This chapter describes a photometric stereo method that works with a wide range of surface reflectances. The novelty of the method is not only in handling various kinds of surface reflectance, but also in avoiding radiometric calibration and alteration of the ambient lighting. We derive a theoretical relationship between the number of input images and the expected accuracy of surface normal estimates. The effectiveness of the proposed method is demonstrated using simulated and real-world scenes that contain a variety of diffuse and specular surfaces.

5.1 Introduction

Photometric stereo estimates surface orientation from a set of images taken from a fixed viewpoint under different lighting directions. The original work of photometric stereo by Woodham [Woo80] and Silver [Sil80] assume a Lambertian surface illuminated by a distant point light source. Given three or more images, the Lambertian photometric stereo method recovers surface orientations of the scene. After this early work, many researchers have studied the approach to make it work under more general conditions. Still, many photometric stereo methods are built upon specific parametric reflectance models, so are naturally restricted to limited classes of reflectances. One of the most im-

portant milestones in making photometric stereo practical would be to handle various surface reflectances.

In this work, we present a new computational approach for solving photometric stereo's problem in handling a wide range of surface reflectances. Instead of assuming a specific parametric reflectance model, such as Lambertian, we assume only three reflectance properties that are often observed in real-world scenes, *i.e.*, *monotonicity*, *visibility*, and *isotropy* of reflectance with respect to the cosine of the surface normal and light direction. In fact, these reflectance properties are observed in many common materials such as plastics, ceramics, rubber, opaque glasses, and smooth glossy paints.

Our method uses input images of a static scene, possibly composed of spatially varying reflectances, taken from a fixed viewpoint under varying and known directional lightings. From the intensity observations per pixel, we establish a set of inequalities derived from the monotonicity, visibility, and isotropy properties. These inequalities specify convex cones in the solution space of surface orientations. By taking the intersection of the convex cones, our method obtains a smaller solution space for the surface orientation. As more input images are given, the solution space becomes more restricted. We show, in this chapter, the relationship between the number of input images (lighting directions) and the estimation accuracy. To that end, we show that given about 50 images our method can achieve an accuracy of less than one degree.

Our consensus approach avoids imposing restricting assumptions on surface reflectances and expands the applicability of photometric stereo. We show that the method can also deal with surfaces with only specular reflections using the same scheme by assuming the monotonicity and isotropy properties with respect to the cosine of the surface normal and the bisector between the lighting direction and viewing direction. In addition, our method is naturally free from radiometric calibration. Because radiometric response functions are monotonic, the monotonicity, visibility, and isotropy properties are maintained in the observation even with any non-linear radiometric response functions. This allows our method to work without knowing the camera response function.

The rest of the chapter is organized as follows. After briefly discussing previous approaches in Section 5.1.1, Section 5.2 describes the theory of the proposed method. We then describe an implementation using a voting method for illustrating the consensus approach in Section 5.3. In Section 5.4, we describe how the consensus approach can be turned into an energy minimization scheme for an efficient implementation. After the implementation details, we describe the theoretical relationship between

the number of lighting directions and the accuracy of surface normal estimates in Section 5.6. Section 5.7 shows the experimental validations using simulation and real-world images. Finally, Section 5.8 concludes the chapter with discussions on future research directions.

5.1.1 Previous work

Photometric stereo has a long history since the pioneering works by Woodham [Woo80] and Silver [Sil80]. Early methods made strong assumptions on the surface reflectance, often the Lambertian model. There have been many studies to weaken the constraints on the reflectance model.

For handling specularity, Coleman and Jain [CJ82] use four images and discard one observation that is most likely a highlight for each pixel. Barsky and Petrou [BP03] extend the method to handle highlights as well as shadows by using four color images. These methods treat non-Lambertian effects as outliers. Solomon and Ikeuchi [SI96] recover surface roughness using the similar four-light setup. Provided there are enough images, non-Lambertian reflectance parameters can be estimated with their method. Nayer *et al*. [NIK90] apply photometric stereo using a hybrid reflectance model that is a linear combination of Lambertian and specular components. Tagare and de Figueiredo [TdF91] consider diffuse non-Lambertian surfaces and solve the problem using an m-lobed reflectance map. Georghiades [Geo03] considers both diffuse and specular reflections and estimates surface normals as well as reflectance parameters based on the Torrance-Sparrow model with unknown light directions.

Some early works [HI84][Ike87] use a reference object for photometric stereo. Recently, Hertzmann and Seitz [HS05] proposed an example-based surface reconstruction method with arbitrary bidirectional reflectance distribution functions (BRDFs). Goldman *et al*. [GCHS05] consider object surfaces modeled by a linear combination of two fundamental materials and remove the need for a reference object by iteratively estimating the basis BRDFs and surface normals.

There are other approaches for the generalization of reflectance properties that are based on isotropy [LL99], Helmholtz stereopsis [ZBK02], bilateral symmetry [AK07], isotropic reflectance [AZK08], reflective symmetry of the halfway vector [HLHZ08], and monotonicity [SH10]. Lu and Little [LL99] proposed a hybrid method with controlled lightings and object poses to estimate both the surface and a non-parametric reflectance map. The method requires that the BRDF is both isotropic and uniform across the

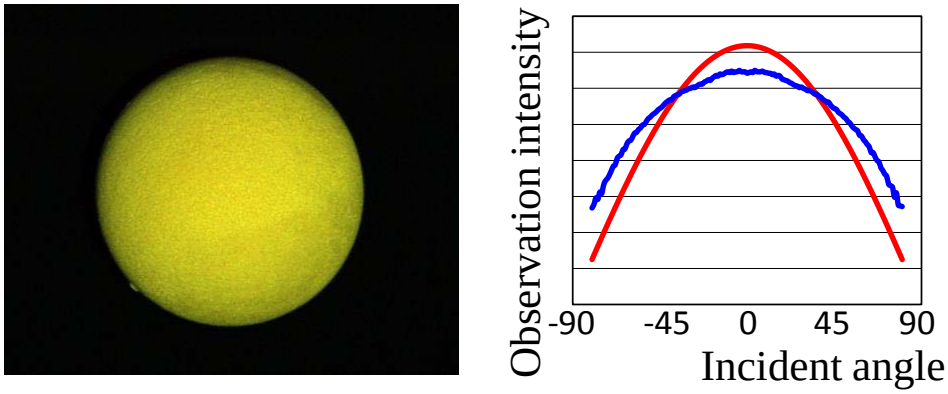


Figure 5.1: Measured reflectance of a diffuse yellow sphere painted with a poster color containing gum Arabic (blue line) and Lambertian fitting (red line).

surface. Zickler *et al*. [ZBK02] use Helmholtz reciprocity to recover both depth and surface normals independent of reflectance. Alldrin and Kriegman [AK07] estimate surface normals using symmetries along intensity profiles from view-centered circles of light directions. Alldrin *et al*. [AZK08] represent the class of isotropic reflectances using a linear basis of general non-parametric bivariate functions to simultaneously estimate shape and reflectance. Holroyd *et al*. [HLHZ08] use a dense sampling to resolve both the normal direction as well as tangent vectors using the symmetry property of reflectance. These symmetries are very general and apply to both isotropic and anisotropic materials. Smith *et al*. [SH10] estimate facial surface reflectance properties by fitting a curve with a monotonicity constraint.

There have been some diffuse reflection models for dielectric materials. Reichman [Rei73] derives diffuse reflection and transmission from the media of arbitrary optical thickness. Wolff *et al*. [Wol94] provide an azimuth-independent diffuse reflection model with accounting isotropic subsurface scattering and Fresnel boundary effects. Oren and Nayar [ON95] propose a generalized diffuse reflectance model by taking surface roughness into account. Once the surface roughness is known, it is reported that it works well for estimating surface orientations. However, in practice, it is difficult to know the surface roughness beforehand. While our model theoretically does not entirely cover the Oren-Nayer model because of forward and backward scattering effects, as shown later in our experimental results, our method is still able to handle rough surfaces such as the one shown in Fig. 5.1.

Our approach is close to Chen *et al*.'s work [CGS06] in that both methods do not use

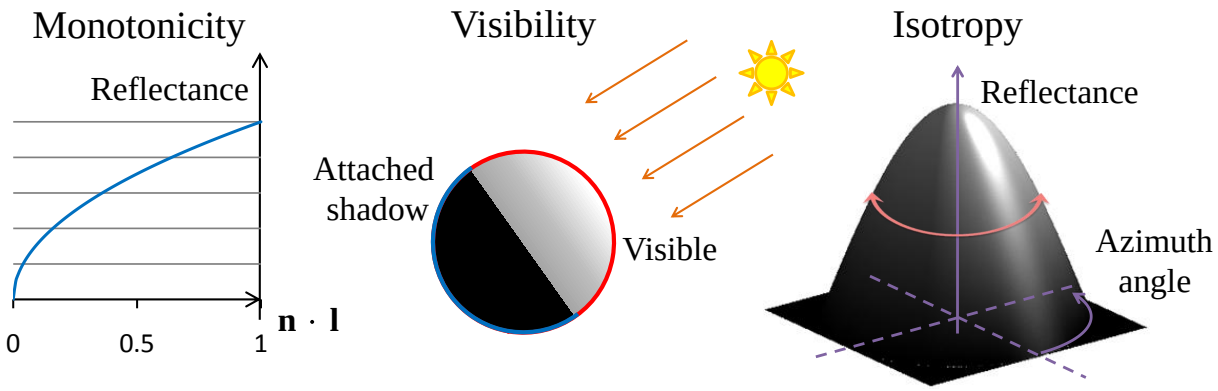


Figure 5.2: Monotonicity, visibility, and isotropy properties of reflectances. Left: The reflectance r monotonically increases with $\boldsymbol{n} \cdot \boldsymbol{l}$. Middle: The reflectance becomes zero when $\boldsymbol{n} \cdot \boldsymbol{l} \leq 0$. Right: The reflectance r gives the same value when $\boldsymbol{n} \cdot \boldsymbol{l}_i = \boldsymbol{n} \cdot \boldsymbol{l}_j$.

a specific parametric reflectance model. Their method determines surface orientations by taking the bisector of view and lighting directions using specular highlight. Unlike Chen *et al.*'s approach, our method can handle diffuse surfaces as well as specular surfaces using monotonicity, visibility, and isotropy properties.

5.2 Consensus approach

Let us begin with our image formation model. An intensity observation o_i is described using the light source intensity E, scalar ambient lighting a, surface normal $\boldsymbol{n}$, incident light direction $\boldsymbol{l}_i$, reflectance function r and radiometric response function f as

$$o_i = f\Big(E\, r(\boldsymbol{n} \cdot \boldsymbol{l}_i) + a\Big), \tag{5.1}$$

where $\boldsymbol{n} \cdot \boldsymbol{l}_i$ is the dot product of $\boldsymbol{n}$ and $\boldsymbol{l}_i$.

In this work, we assume three properties about the surface reflectance r: monotonicity, visibility, and isotropy (Fig. 5.2). These reflectance properties are observed in a wide range of diffuse reflectances. In fact, it is pointed out that many existing diffuse materials deviate from the Lambertian model in prior studies [Wol94][ON95]. Fig. 5.1 shows an actual measurement that deviates from the Lambertian model. Our reflectance model covers such diffuse reflections as well as the Lambertian model as a special case where $r(\boldsymbol{n} \cdot \boldsymbol{l}) = \rho \boldsymbol{n} \cdot \boldsymbol{l}$, where ρ is surface albedo.

Using these properties of monotonicity, visibility, and isotropy, we derive three constraints in the form of inequalities that specify possible solution spaces of the surface

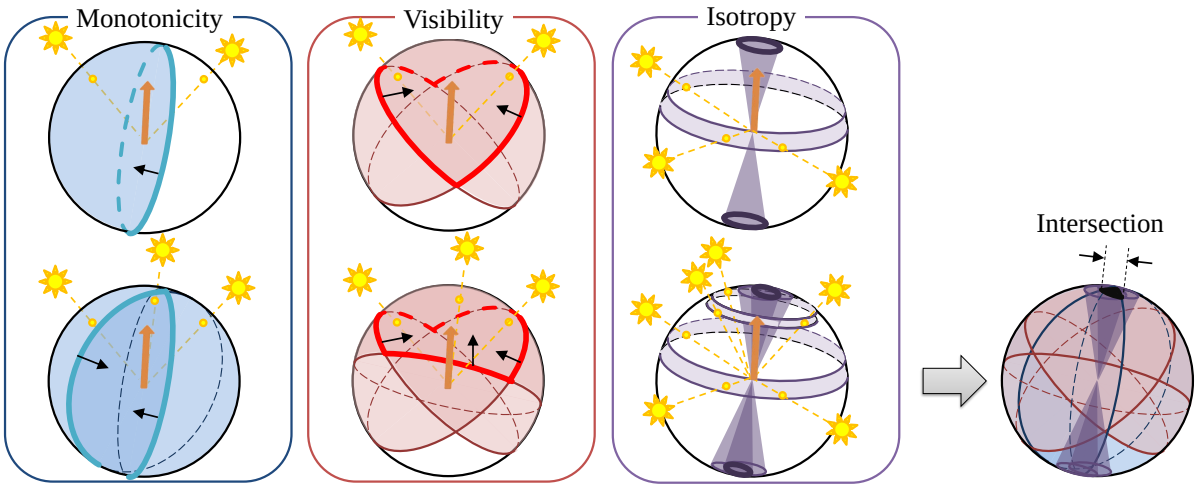


Figure 5.3: Monotonicity, visibility, and isotropy constraints. Each of these three constraints gives a solution space of the surface orientation. By taking the intersection of the solution spaces, our method obtains a smaller solution space of the surface orientation. The narrow arrows represent the solution space, and the bold ones correspond to the true surface orientation. The two rows show how the solution space becomes smaller as the number of observations increases.

orientation. Each of these three constraints independently gives a solution space. Our method estimates surface orientations by taking the intersection of these solution spaces. We only use illuminated pixels for these constraints. The solution space of surface normal $\boldsymbol{n}$ is initialized on a Gaussian sphere because the surface normal $\boldsymbol{n}$ is a unit vector.

5.2.1 Monotonicity constraint

We assume the following monotonicity of the reflectance function r:

$$\boldsymbol{n} \cdot \boldsymbol{l}_i > \boldsymbol{n} \cdot \boldsymbol{l}_j \Leftrightarrow r(\boldsymbol{n} \cdot \boldsymbol{l}_i) > r(\boldsymbol{n} \cdot \boldsymbol{l}_j). \tag{5.2}$$

This monotonicity says that as the dot product of surface normal $\boldsymbol{n}$ and lighting direction $\boldsymbol{l}$ increases, the reflectance r increases.

These constraints hold even for unknown radiometric response functions. In Eq. (5.1), the radiometric response function f is also monotonically increasing, and E and ρ are non-negative. Therefore, $f(r(x))$ is also monotonically increasing, so the

intensity observation o monotonically increases as $\boldsymbol{n} \cdot \boldsymbol{l}$ increases. This property eliminates the necessity of radiometric calibration for our method and allows us to directly use the following relationship regardless of the shape of the response function f:

$$\boldsymbol{n} \cdot \boldsymbol{l}_i > \boldsymbol{n} \cdot \boldsymbol{l}_j \Leftrightarrow o_i > o_j. \tag{5.3}$$

Using inequalities (5.3) obtained from multiple observation pairs (o_i, o_j), the solution space $\mathcal{N}_1$ of the surface orientation $\boldsymbol{n}$ can be determined by taking the intersection of multiple observations as

$$\mathcal{N}_1 = \left\{ \boldsymbol{n} \in \mathbb{R}^3 \mid \bigcap_{i,j} \Big((\boldsymbol{l}_i - \boldsymbol{l}_j) \cdot \boldsymbol{n} > 0 \Big) \right\}, \tag{5.4}$$

for pairs of $\boldsymbol{l}_i$ and $\boldsymbol{l}_j$ that satisfy $r(\boldsymbol{n} \cdot \boldsymbol{l}_i) > r(\boldsymbol{n} \cdot \boldsymbol{l}_j)$. The pair of $(\boldsymbol{l}_i, \boldsymbol{l}_j)$ makes the solution space specified on the north hemisphere whose pole is $(\boldsymbol{l}_i - \boldsymbol{l}_j)$ as illustrated in Fig. 5.3 (Left).

5.2.2 Visibility constraint

When a scene point is illuminated by a light source $\boldsymbol{l}$, the surface normal $\boldsymbol{n}$ should lie in the hemisphere $\boldsymbol{n} \cdot \boldsymbol{l} > 0$. When $\boldsymbol{n} \cdot \boldsymbol{l} \leq 0$, the scene point is in the attached shadow, *i.e.*, the scene point is not visible from the light source. The visibility is defined as

$$\mathcal{N}_2 = \left\{ \boldsymbol{n} \mid \bigcap_{i} (\boldsymbol{n} \cdot \boldsymbol{l}_i > 0) \right\}, \tag{5.5}$$

for all lighting directions $\boldsymbol{l}_i$ that illuminate the scene point. A similar constraint is used by Belhumeur and Kriegman [BK98] for describing possible light source directions. Because our method only uses illuminated pixels, it is not necessary to identify whether the pixel is in an attached or cast shadow.

5.2.3 Isotropy constraint

In addition to the above two constraints, we use the isotropy constraint when multiple similar intensity observations are obtained. Suppose an ideal case where more than two observations under different lighting directions show the same intensity value. In this case, given the different lighting directions $\boldsymbol{l}_i$, $\boldsymbol{l}_j$, and $\boldsymbol{l}_k$, the surface normal $\boldsymbol{n}$ should fall on the direction that is perpendicular to the plane spanned by the lighting vectors. It can be determined up to a sign ambiguity by taking the cross-product:

$$\pm(\boldsymbol{l}_i - \boldsymbol{l}_j) \times (\boldsymbol{l}_i - \boldsymbol{l}_k). \tag{5.6}$$

Such ideal situations are rare in practice, so we use a relaxed near-equality constraint:

$$\boldsymbol{n} \cdot \boldsymbol{l}_i \simeq \boldsymbol{n} \cdot \boldsymbol{l}_j \Leftrightarrow o_i \simeq o_j. \tag{5.7}$$

The constraint says that when similar intensity observations o are obtained, the cosines of incident lighting direction and surface normal are also similar.

When we have similar observations o_i, $(i = 1, 2, \ldots, k)$ under different lighting directions $\boldsymbol{l}_i$, we can expect that the surface normal $\boldsymbol{n}$ lies near to the direction where the variance of $\boldsymbol{n} \cdot \boldsymbol{l}$ is minimized:

$$\boldsymbol{n} = \pm \min_{\boldsymbol{n}} \sum_{i=1}^{k} \left(\boldsymbol{n} \cdot \boldsymbol{l}_i - \overline{\boldsymbol{n} \cdot \boldsymbol{l}} \right)^2, \tag{5.8}$$

where $\overline{\boldsymbol{n} \cdot \boldsymbol{l}}$ is the mean of the dot products. Using m such normal directions $\boldsymbol{n}_m$ obtained from m-sets of lighting directions, our method determines the solution space $\mathcal{N}_3$ that is represented by a convex cone spanned by $\boldsymbol{n}_m$ as

$$\mathcal{N}_3 = \left\{ \boldsymbol{n} \mid \boldsymbol{n} = \sum_m a_m \boldsymbol{n}_m,\ a_m \geq 0 \right\}. \tag{5.9}$$

5.2.4 Consensus solution

Each of the monotonicity, visibility, and isotropy constraints independently gives a solution space (Fig. 5.3). Our method takes the intersection of these to form a smaller solution space $\mathcal{N}$as

$$\mathcal{N} = \mathcal{N}_1 \cap \mathcal{N}_2 \cap \mathcal{N}_3. \tag{5.10}$$

As the number of images increases, it is expected that the solution space $\mathcal{N}$ becomes smaller.

5.2.5 Extension to specular surfaces

Our method can be extended naturally to handle specular reflections by assuming monotonicity and isotropy for specular lobes. Here, we assume only specular reflection and no diffuse reflection, like for metallic surfaces. In this case, as shown in Fig. 5.4, the monotonicity and isotropy are assumed with respect to the cosine of surface orientation $\boldsymbol{n}$ and the bisector $\boldsymbol{h}$ $(= (\boldsymbol{l}+\boldsymbol{v})/\|\boldsymbol{l}+\boldsymbol{v}\|)$ between the light direction $\boldsymbol{l}$ and the view direction $\boldsymbol{v}$. The right-hand side of the figure depicts the diffuse case for reference. In this way,

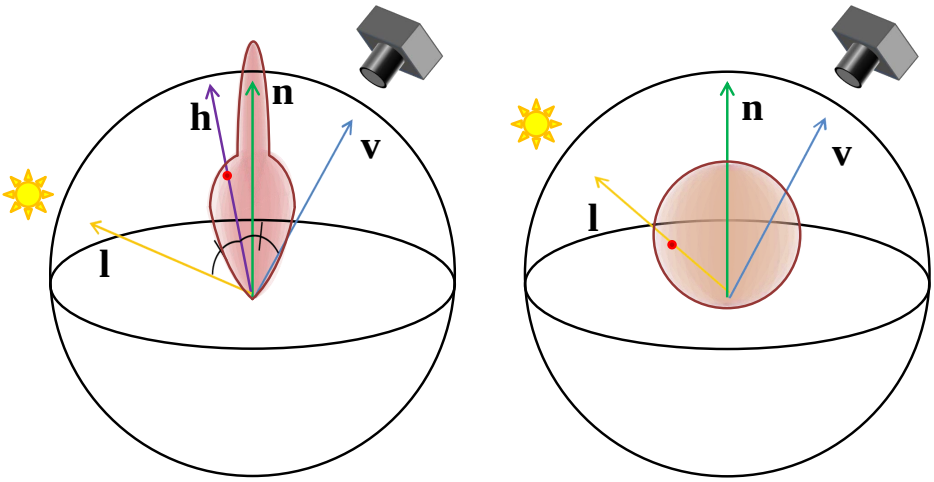


Figure 5.4: Monotonicity and isotropy constraints for the case of specular reflection. Left: The case of specular reflection. The reflectance r monotonically increases with $\boldsymbol{n} \cdot \boldsymbol{h}$ and gives the same value when $\boldsymbol{n} \cdot \boldsymbol{h}_i = \boldsymbol{n} \cdot \boldsymbol{h}_j$. We use the bisector $\boldsymbol{h}$ replacing the light vector $\boldsymbol{l}$ in Eq. (5.1) for the specular lobes. Right: The case of diffuse reflection.

by replacing the light vector $\boldsymbol{l}_i$ with the bisector $\boldsymbol{h}_i$ in Eq. (5.1), the previous discussion holds for the specular lobes. For the visibility constraint, since we do not know the width of specular lobes, we still use $\boldsymbol{n} \cdot \boldsymbol{l} > 0$ as the constraint.

5.3 Consensus Photometric Stereo by Voting

To illustrate an intuitive implementation, we show a voting method to find the solution space as an intersection segment. We use a geodesic sphere for defining the entire solution space of the surface orientation per pixel. In this representation, the vector from the geodesic sphere center to a vertex represents a surface orientation. As shown in Fig. 5.3, the voting approach gives a score to vertices of the geodesic sphere when the vertices are inside the region of the solution space. This process is regarded as a histogram-based approach. As more images under different lighting conditions are used, the smaller number of vertices would have the highest score. Finally, the surface normal is estimated by taking the direction from the geodesic sphere center to the vertex with the highest score. When more than one vertices have the highest score, we take the mean vector to produce the surface normal estimate. As shown in Fig. 5.3, monotonicity and visibility constraints give a vote to a hemispherically distributed region. On the other hand, the isotropy constraint defined by a group of similar intensity observations

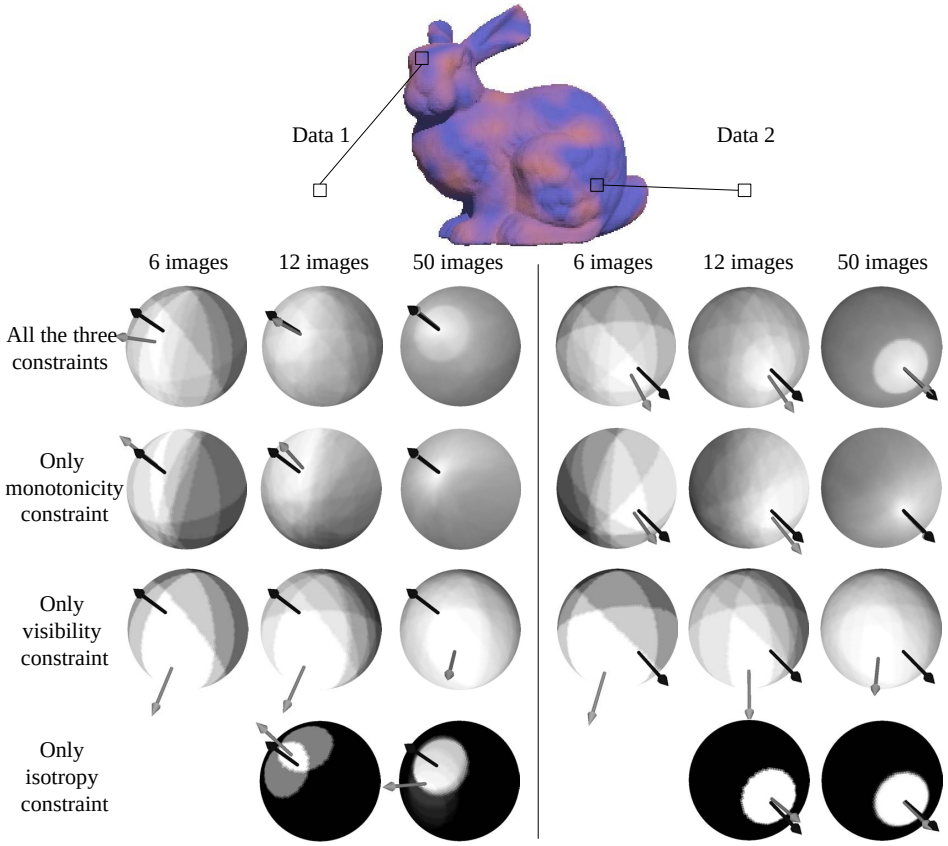


Figure 5.5: Voting results with regard to the number of input images. A black arrow represents the ground truth of the surface normal. A gray arrow represents the estimated surface normal. The brighter area has the higher score, which indicates the solution space.

gives a convex cone region when voting. The size of the cone is defined by the variance of surface normals given by Eq. (5.6) with all the combinations from the group. The smaller the variance is, the smaller the size of the cone region becomes for limiting the solution space.

Fig. 5.5 shows voting results with respect to the number of input images, *i.e.*, lighting directions. A black arrow represents the ground truth of the surface normal. A gray arrow represents the estimated surface normal. Brighter areas have high scores, while darker ones have low scores. We can see that more input images give a smaller solution space for surface normal in each constraint. Table. 5.1 shows the accuracy of the surface normal. In the only isotropy constraint test, with 6 images, isotropy constraint with 6 images, there is no group of similar intensity observations, therefore we cannot get

Table 5.1: The table shows the accuracy of the estimated surface normals [deg.] with regard to the number of input images with an intuitive voting method. Data 1 shows the result of the images on the left in Fig. 5.5, and Data 2 shows the result of the images on the right in Fig. 5.5. Mean error and median error of all estimated pixels are shown below.

		6 images	12 images	50 images
Data 1	All the three constraints	11.772	2.558	1.019
	Only monotonicity	11.024	4.429	1.019
	Only visibility	50.187	48.633	36.323
	Only isotropy	-	11.686	18.506
Data 2	All the three constraints	8.265	6.304	1.024
	Only monotonicity	4.376	7.008	1.342
	Only visibility	36.293	26.538	22.160
	Only isotropy	-	3.499	1.498
Mean error	All the three constraints	21.208	6.831	1.158
	Only monotonicity	36.487	9.363	1.172
	Only visibility	42.329	39.779	38.160
	Only isotropy	-	10.035	6.708
Median error	All the three constraints	14.953	3.510	1.140
	Only monotonicity	33.858	3.951	1.159
	Only visibility	42.806	39.786	37.084
	Only isotropy	-	4.588	2.707

results.

From these results, we can see that monotonicity and isotropy are very powerful constraints for limiting the solution space. In addition, these two constraints are complementary to each other. When many groups of similar intensities are observed, the isotropy constraint effectively works. On the other hand, when various intensities are observed, the monotonicity constraint becomes more effective to estimate the surface normal. Compared with these two constraints, the visibility constraint does not produce a small solution space. Nevertheless, it steadily limits the solution space. Actually, when we only use monotonicity and isotropy constraints in data 2 with 50 input images, the accuracy is 1.342 [deg.]; however, a combination of all three constraints

give a more accurate surface normal estimate (1.024 [deg.]).

The voting method described in this section can produce good results with simple implementation. However, the accuracy of surface normal estimates is limited by the resolution of the geodesic sphere. The number of vertices of the geodesic sphere that we used in Fig. 5.5 and Table. 5.1 is 10242, where the angle between two neighboring vertices is about 2 [deg.]. It indicates that even if we use much more input images, the accuracy of surface normal estimates is limited to an average error of about 1 [deg.] in the best case. Therefore, with this implementation, there is a trade-off between the resolution of the geodesic sphere and the estimation accuracy.

5.4 Efficient Implementation with Energy Minimization

The previous section describes a straightforward implementation of consensus photometric stereo. To efficiently estimate surface orientation $\boldsymbol{n}$, we cast the consensus approach to an energy minimization problem. For this purpose, we develop energy terms for monotonicity, visibility, and isotropy constraints, respectively. These energy terms are computed at each pixel.

Monotonicity term From Eq. (5.4), we develop an energy term that favors $\boldsymbol{n}\cdot(\boldsymbol{l}_i-\boldsymbol{l}_j) > 0$ being satisfied for observations $o_i > o_j$. Using a sigmoid-like function, we formulate this constraint as

$$E_1(\boldsymbol{n}) = \frac{1}{N_1}\sum_{i,j}\frac{1-k\boldsymbol{n}\cdot(\boldsymbol{l}_i-\boldsymbol{l}_j)}{1+\exp\Big(t\boldsymbol{n}\cdot(\boldsymbol{l}_i-\boldsymbol{l}_j)\Big)}, \tag{5.11}$$

for all pairs of (i, j) where $o_i > o_j$. We use the sigmoid-like function to equally give a small cost when $\boldsymbol{n}\cdot(\boldsymbol{l}_i-\boldsymbol{l}_j) > 0$. In the energy term, t is a gain, and N_1 is the number of pairs (i, j) that are used for the computations. The numerator is designed to form a slope that is determined by the factor k so that more deviations from the constraint are penalized. With such a slope, the optimization becomes more efficient and quickly converges. Fig. 5.6 shows the form of the function $s(x) = (1-kx)/(1+e^{tx})$ that is used for E_1.

For efficient computation, we re-sample all possible combinations (i, j) to reduce the number of pairs. We select N_M observations o_j that are close to o_i while satisfying $o_i > o_j$, because the similar intensity observation pairs (o_i, o_j) tend to give smaller solution spaces. However, the combinations that are used for the isotropy constraint are excluded because of the condition $o_i > o_j$.

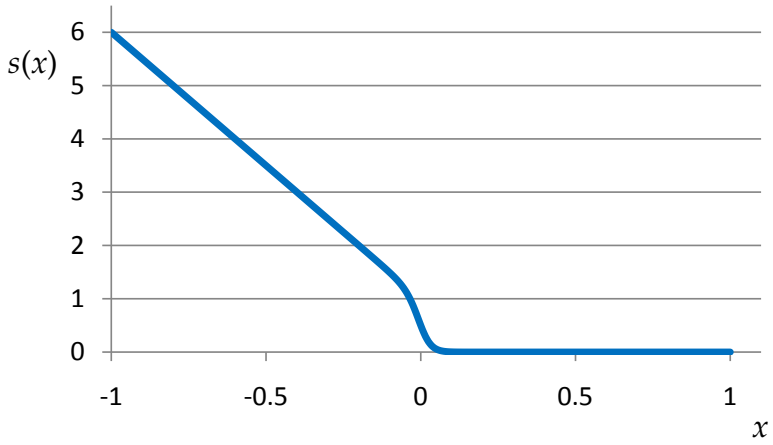


Figure 5.6: Plot of a function $s(x) = (1 - kx)/(1 + e^{tx})$ used to design energy terms. $(k, t) = (5, 50)$ is used for the plot.

Visibility term We formulate the visibility constraint of Eq. (5.5) in a similar manner with the monotonicity constraint Eq. (5.11). Using a sigmoid-like function, the visibility term E_2 is formulated as

$$E_2(\boldsymbol{n}) = \frac{1}{N_2} \sum_i \frac{1 - k\boldsymbol{n} \cdot \boldsymbol{l}_i}{1 + \exp(t\boldsymbol{n} \cdot \boldsymbol{l}_i)}, \tag{5.12}$$

where N_2 is the number of observations that are illuminated, *i.e.*, the number of observations used for the estimation.

Isotropy term The isotropy constraint Eq. (5.9) gives a solution space from a set of lighting directions that produce similar intensity observations. The more similar the intensity observations are, the smaller the solution space becomes. We formulate this as an energy function E_3 that favors smaller variances of the dot product $\boldsymbol{n} \cdot \boldsymbol{l}$ in each set of similar observations. Given m sets of similar observations S, the energy term is defined as

$$E_3(\boldsymbol{n}) = \frac{1}{\sum_i^m |S_i|} \sum_i^m \sum_{j \in S_i} \left(\boldsymbol{n} \cdot \boldsymbol{l}_j - (\overline{\boldsymbol{n} \cdot \boldsymbol{l}})_i \right)^2, \tag{5.13}$$

where S_i is the i-th set of observation indices, $|S_i|$ represents the number of elements in the set, and $(\overline{\boldsymbol{n} \cdot \boldsymbol{l}})_i$ is the mean of the dot product $\boldsymbol{n} \cdot \boldsymbol{l}$ in S_i.

Energy function The energy function E is defined by combining the above three constraints and an additional constraint of a unit normal length as

$$E(\boldsymbol{n}) = \lambda_1 E_1(\boldsymbol{n}) + \lambda_2 E_2(\boldsymbol{n}) + \lambda_3 E_3(\boldsymbol{n}) + (1 - \|\boldsymbol{n}\|^2)^2, \tag{5.14}$$

where λ_i represents a weighting factor. We use the Levenberg-Marquardt method [BW88] to minimize the multivariate function to estimate a surface normal vector per pixel. For initialization, we use the lighting direction vector that shows the highest intensity (without saturation) as the initial guess of the normal vector. Before the optimization, we exclude low intensity observations as shadow pixels and use only illuminated observations o_i.

5.5 Comparison between voting and energy minimization approaches

We compared estimation error and convergence with respect to the number of light directions between the voting and energy minimization approaches. The comparison is performed using a simulation environment, where a planar patch is illuminated by random lighting directions. To obtain the statistical result, the ground truth surface normal is also randomly generated, and the procedure is repeated 5000 times. In this simulation, we avoid considering cast shadowing effects. For the voting approach, the number of vertices of the geodesic sphere is set to 10242.

Fig. 5.7 shows a graph of the comparison. When the number of light directions is small, the solution space remains to be large due to the small number of constraints. In such cases, the energy minimization approach converges at the edge of the solution space; therefore, it becomes less accurate than the voting approach, which takes the center of the solution space. On the other hand, when a sufficiently large number of observations are given, the energy minimization approach can produce a more accurate result than the voting approach. As mentioned in Section 5.3, the accuracy of the voting approach is limited to at most 1 [deg.] in this setting, even when more observations are provided.

In terms of computational cost, the energy minimization approach works significantly faster than the voting approach. For this experiment, the voting approach takes 86.27 [sec.] while the energy minimization approach takes 8.67 [sec.] in estimating 5000 different normals. The energy minimization approach runs about 10 times faster than the voting approach.

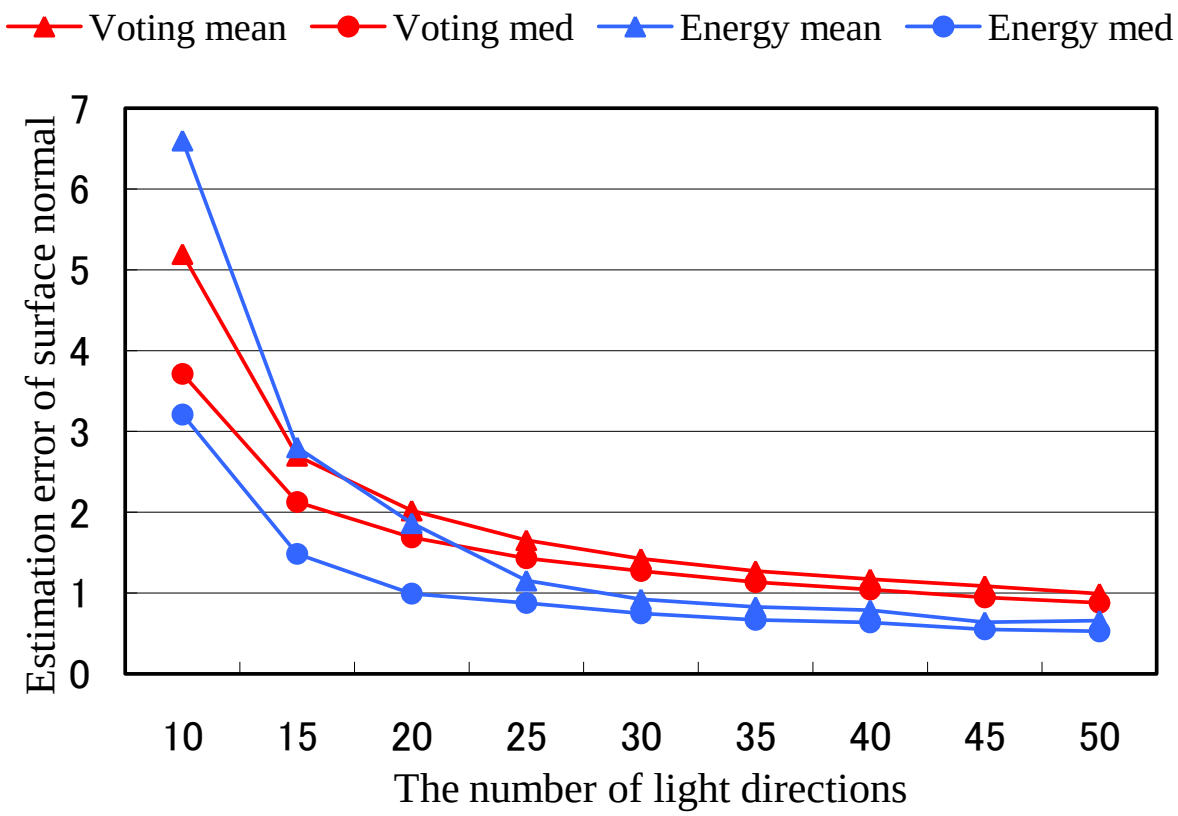


Figure 5.7: Estimation error of the surface normal by the voting and energy minimization approach. In the voting approach, the number of vertices of the geodesic sphere is set to 10242. In the plots, mean and median errors of each method are shown.

5.6 How many lighting directions are required?

As mentioned in Section 5.2, as the number of light directions increases, the solution space becomes smaller. In this section, we analyze the statistical relationship between the error of the estimated surface normal $\triangle\theta$ and the number of light directions N_l, by taking the monotonicity property as an example.

First, we develop an expression for the number of regions $S(k)$ on the hemisphere divided by k great circles derived from light directions. Here the hemisphere represents the entire solution space of the surface normal, because the camera can be placed at a fixed viewpoint. The problem of obtaining the number of distinct aspects $S(k)$ is equivalent to obtaining the number of regions into which k lines divide a 2-D plane. Therefore, $S(k)$ also denotes the number of regions that k lines divide a 2-D plane into. We can derive the formula of the number of regions $S(k)$ in an inductive manner.

Suppose we add the $k+1$-th line after k lines have already been drawn. This new line intersects the existing k lines at k points in the maximal case, which divide the new line into $k+1$ segments. Each segment on the new line divides one old region into two new regions. Therefore, this operation adds $k+1$ new regions:

$$S(k+1) = S(k) + k + 1. \tag{5.15}$$

By solving this, we obtain

$$S(k) = \frac{1}{2}(k^2 + k + 2). \tag{5.16}$$

Let us now count the number of great circles $K(t)$ given by t light directions by taking the monotonicity property as an example. Every pair of light directions provides one great circle as a perpendicular bisector plane between the two light directions. In other words, every pair of two light direction provides a perpendicular line in the 2-D plane between the two points, which represent light directions. We call these light-direction points. Therefore,

Lemma 1. *t light directions provide ${}_tC_2$ lines.*

Since all lines are perpendicular bisectors, three lines of them have one intersection as a circumcenter of the three light-direction points. One intersection among three lines reduce one segment. In other words, every three light-direction points reduce the number of segments by one. It leads to the following lemma:

Lemma 2. *t light directions reduce the number of segments by ${}_tC_3$.*

From Eq. (5.16) and the above lemmas, we can calculate the number of regions $R(t)$ divided by great circles computed from the monotonicity property with t light directions as the following:

$$\begin{aligned} R(t) &= \frac{1}{2}\left[({}_tC_2)^2 + {}_tC_2 + 2\right] - {}_tC_3 \\ &= \frac{1}{24}(3t^4 - 10t^3 + 21t^2 - 14t + 24). \end{aligned} \tag{5.17}$$

Theoretically, the monotonicity property gives the number of division segments described by Eq. (5.17), however, not all the pairs of light directions are useful for limiting the solution space. For example, a selected pair may not be informative when it does not shrink the solution space. Therefore, we use the following selection strategy in our implementation for computational efficiency.

For each lighting direction, our method chooses not all combinations but N_M pairs of light directions. When a pair of observations has a similar intensity to each other, they tend to efficiently restrict the solution space. However, when the pair has almost the same intensity, there may be an observation error. Therefore, in that case, our method does not use such pairs. Taking these into account, our method chooses pairs using the following simple strategy. First, we sort light directions by the observed intensity magnitudes of the target pixel. Then, light directions with the minimum or close to minimum intensity are removed as a shadow pixel. The light directions associated with saturated or very high intensity observations are also removed as specular pixels. Second, we remove the light directions with almost the same intensity observations. These are instead used for the isotropy constraint. After removing these light directions, our method finally chooses the top N_M light directions that have similar intensity observations. These pairs finally give the perpendicular bisectors as the monotonicity constraint.

Now we describe how many segments are needed to achieve the expected accuracy of the estimation based on the above discussion. The number of segments N_{sgm} on a hemisphere divided by N_{gc} great circles is the same as that of a 2-D plane divided up by N_{gc} lines. It becomes

$$N_{sgm} = \frac{1}{2}(N_{gc}^2 + N_{gc} + 2). \tag{5.18}$$

In an ideal case, assuming that N_{sgm} equal-size disks with the area $\pi(\triangle\theta)^2$ cover the entire solid angle of a hemisphere (2π), we obtain an optimistic estimate of the accuracy as follows:

$$\triangle\theta = \sqrt{2/N_{sgm}}. \tag{5.19}$$

Let N_v be the number of illuminated observations. The monotonicity property provides N_M constraints for each observation o_i ($i = 1, 2, \ldots, N_v$) without duplication, therefore, the total number of the great circles obtained from the monotonicity property becomes $N_{gc} = (N_v - N_M)N_M$. Hence, from Equations (5.18)(5.19), the estimation accuracy becomes:

$$\triangle\theta = \sqrt{\frac{4}{N_v^2 N_M^2 - 2N_v(N_M^3 + N_M) + N_M^4 - N_M^2 + 2}}. \tag{5.20}$$

Assuming that half the number of light directions illuminate the patch of interest among the entire N_l light directions, *i.e.*, $N_v = \frac{1}{2}N_l$, and by setting $N_M = 8$, this analysis

from Eq. (5.20) indicates that given about 45 input images, our method can achieve an accuracy of less than one degree. Therefore, in our experiments Section 5.7, we use about 45 input images for various target objects.

Simulation. To verify the theory described above, we conduct a simulation of the estimation error and compare the result of the energy minimization approach with the theoretical error. In this simulation, a surface is rendered synthetically under various lighting directions that are randomly generated. From the input data, we estimate the surface normal from the observations with the known lighting directions using the energy minimization approach. The procedure is repeated using various surface orientations and lighting directions. Finally, the errors are computed from the ground truth normals, and the median error is computed. In this simulation, we assume there are no cast shadow while the attached shadows exist.

Fig. 5.8 shows the plot of the result. The horizontal axis represents the number of light directions, *i.e.*, the number of input images, and the vertical axis is the estimation error [deg.]. Since this simulation assumes no occlusion, the energy minimization approach achieves a better result than the theoretical error ("Theory1") described above. On the other hand, "Theory2" represents a theoretical error that assumes no occlusion, *i.e.*, $N_v = N_l$. Because our energy minimization approach does not take the self-occluded observations, *i.e.*, $\boldsymbol{n} \cdot \boldsymbol{l} < 0$, the estimation error becomes greater than that of "Theory2". Therefore, "Theory1" and "Theory2" can be regarded as the upper and lower bounds of the actual error, respectively.

5.7 Experiments

To evaluate the effectiveness of the proposed method, we performed experiments using both simulation and real-world scenes. We first show a quantitative evaluation using the simulation data in Section 5.7.1. Second, we evaluate our method using five real-world scenes in Section 5.7.2. Throughout the experiments, we used parameters $k = 5$, $t = 50$, $N_M = 8$, $\lambda_1 = 8$, $\lambda_2 = 1$, and $\lambda_3 = 300$ for diffuse objects. For specular objects, we only changed the weighting factor of the isotropy term to $\lambda_3 = 30$.

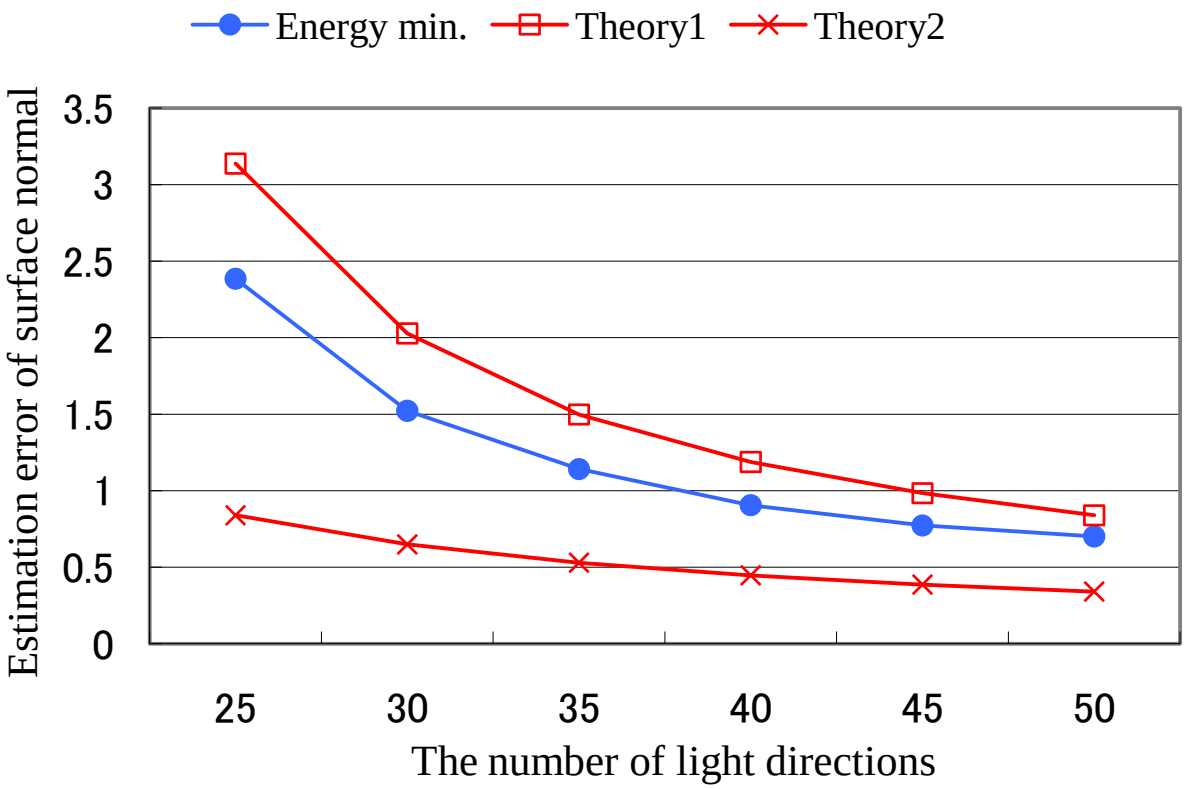


Figure 5.8: Estimation error of the surface normal based on the monotonicity constraint using the energy minimization approach and the theoretical errors. "Theory1" represents the theoretical error with occlusions, and "Theory2" assumes no occlusion, *i.e.*, $N_v = N_l$.

5.7.1 Simulation results

The simulation experiment is designed to quantitatively examine the performance of the proposed method. We use combinations of different settings;

1. Linear/non-linear camera response functions.
2. Lambertian/non-Lambertian reflectances.
3. With/without ambient lighting.

We represent these settings by **Y**es or **N**o for {linear response function, Lambertian surface, ambient lighting}. For example {Y, N, Y} indicates the combination of linear response function, non-Lambertian surface, and with ambient illumination. A synthetic scene is rendered using these settings, and our method is applied to each of these

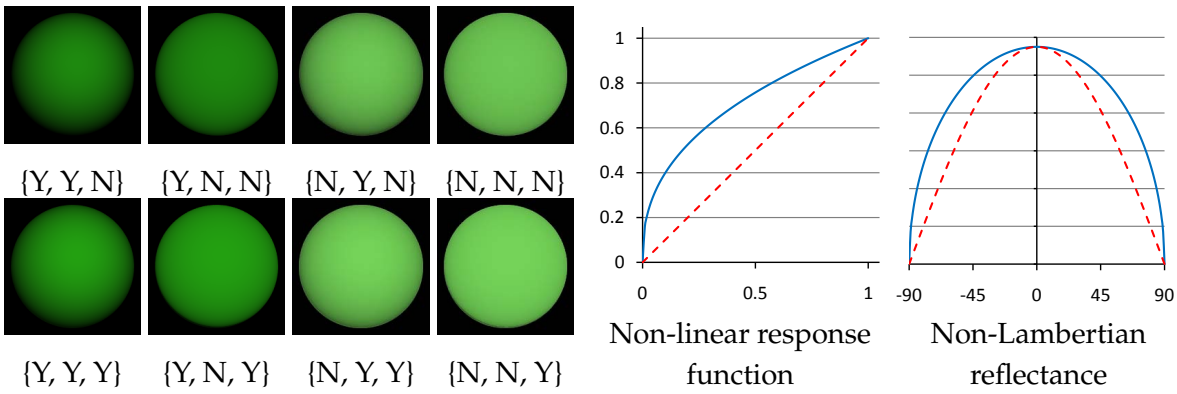


Figure 5.9: Simulation setup and results. Left shows the reference spheres rendered with the combinations of {(1) linear response function **Y**es/**N**o, (2) Lambertian: **Y**es/**N**o, (3) ambient illumination: **Y**es/**N**o }. In the right, the shapes of the non-linear response function and non-Lambertian reflectance that are used in this simulation are shown.

datasets. Fig. 5.9 (left) shows the reference spheres rendered under these settings. On the right hand-side of the figure, the shapes of the non-linear response function and non-Lambertian diffuse reflectance used for this experiment are shown. The numerical results are summarized in Table. 5.2. As shown in the table, our method is not susceptible to ambient lighting, non-Lambertian diffuse reflection, or the non-linear response function, while the standard photometric stereo method suffers from these non-ideal conditions. The estimation error becomes consistently small with our method except for the completely ideal situation.

5.7.2 Real-world results

We applied our method to various diffuse objects and specular objects and compared them with the standard photometric stereo method. We used five different scenes under various conditions: (1) Yellow sphere scene (Fig. 5.1, non-Lambertian), (2) Terracotta scene (non-linear response function), (3) Statue scene (with ambient illumination), (4) Relief scene (non-linear response function with ambient illumination), and (5) Clip scene (specular lobes).

We recorded the scenes using two different cameras; a Sony color digital camera XCD-X710CR that has a linear response function, and a Nikon D1x camera with a non-linear response function. The scenes were illuminated by a moving LED point light source and recorded from a fixed viewpoint. To obtain the light source directions, we used a mirror sphere placed in the scene. We compare our method with the standard

Table 5.2: Mean and median RMSE [deg.] evaluation of the estimated surface normals under corresponding rendering settings described in Fig. 5.9.

	Our method		Standard PS	
	mean	med	mean	med
{Y, Y, N}	0.708	0.617	**0.000**	**0.000**
{Y, N, N}	**0.740**	**0.651**	8.479	7.999
{N, Y, N}	**0.719**	**0.634**	3.866	3.640
{N, N, N}	**0.737**	**0.647**	8.541	7.855
{Y, Y, Y}	**0.705**	**0.622**	5.320	4.920
{Y, N, Y}	**0.741**	**0.658**	8.412	7.793
{N, Y, Y}	**0.721**	**0.633**	7.105	6.576
{N, N, Y}	**0.723**	**0.627**	8.709	8.020

photometric stereo method based on the Lambertian model (referred to as 'standard PS' in the following). To use the same input to both methods, shadowed pixels were excluded when the standard PS was applied.

Fig. 5.10 shows the result of the yellow sphere scene recorded by a Sony XCD-X710CR. Our method recovers surface normals from a non-Lambertian diffuse reflectance scene more accurately than the standard PS. The error maps in the fourth and fifth figures in Fig. 5.10 clearly show the difference. With the standard PS, the error tends to become greater especially around the boundary of the sphere where surface normals are off from the viewing direction.

In Fig. 5.11, we show the result of the Terracotta scene taken with the Nikon D1x camera with a non-linear response function. The results of our method and the standard PS appear to be similar, but our method results in more vivid surface orientations, *e.g.*, on the top of the left hand of the terracotta soldier. Similar to the case of the yellow sphere scene, the standard PS results in a rather flat surface normal field.

Fig. 5.12 shows the result of the statue scene under ambient illumination taken by a Sony XCD-X710CR. Our method produces faithful surface orientations, while the standard photometric stereo produces overly smooth surface normals.

The relief scene of Fig. 5.13 was taken with a Nikon D1x camera with a non-linear response function, under ambient lightings. Our method faithfully recovers surface orientations even when the imaging condition significantly deviates from classical

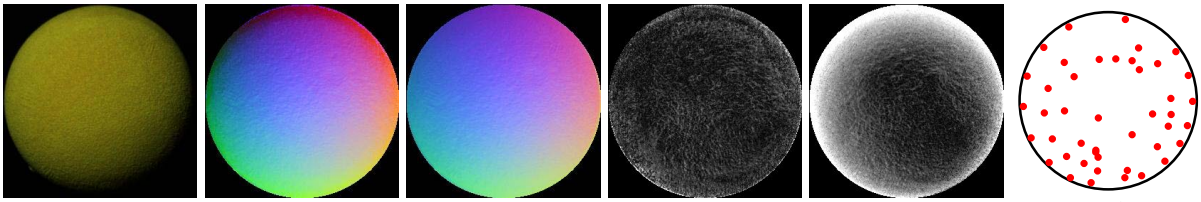


Figure 5.10: Result of our method applied to the yellow sphere with a non-Lambertian surface. From left to right, one of the input images, the estimated normal map with our method, that with standard photometric stereo method, the corresponding errors from the ground truth, and the sampled light directions are shown. The higher intensity in the error maps indicates the greater errors. 43 images are used as input.

assumptions, *i.e.*, Lambertian reflectance, no ambient illumination, and a linear camera response.

To assess the significance of the three properties individually, we show the effect of these in Fig. 5.14. The monotonicity and isotropy properties are the strongest constraints, however large errors are found in places. For the isotropy constraint, inaccurate surface normal estimates are observed at locations where the zenith angle of surface normals is large, *i.e.*, outward-looking surface normals. The visibility constraint prevents large errors, especially for the outward-looking surface normals. These three constraints work in a complementary manner, and the combination of all the three constraints gives the best result.

Fig. 5.15 shows the result of the clip scene that is made of specular surfaces. The result shows that our method can estimate surface normals from the specular lobes as well.

Finally, Fig. 5.16 shows the rendering of 3D surfaces and relighting. The relief scene (Top) is the reconstructed 3D, and the others are relighting results. The reference spheres depict the lighting directions.

5.8 Discussion

We present a consensus approach for photometric stereo for a generalized reflectance model that holds three properties: Monotonicity, visibility, and isotropy. These properties are naturally observed in a wide variety of diffuse reflections as well as in specular

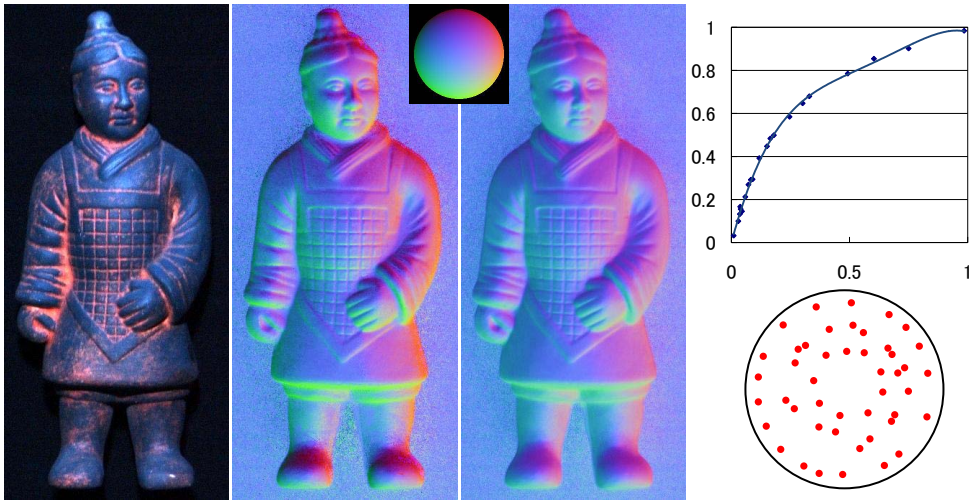
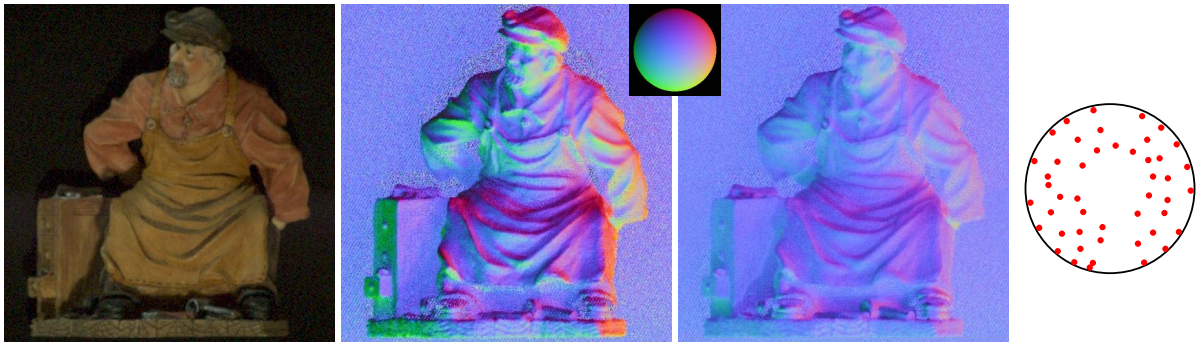


Figure 5.11: Result of the terracotta scene taken with a Nikon D1x camera with a non-linear response function without ambient illumination. From left to right, one of the input images, the estimated normal map with our method, that with the standard photometric stereo method, the measured response function, and lighting directions are shown. 46 images are used as input.

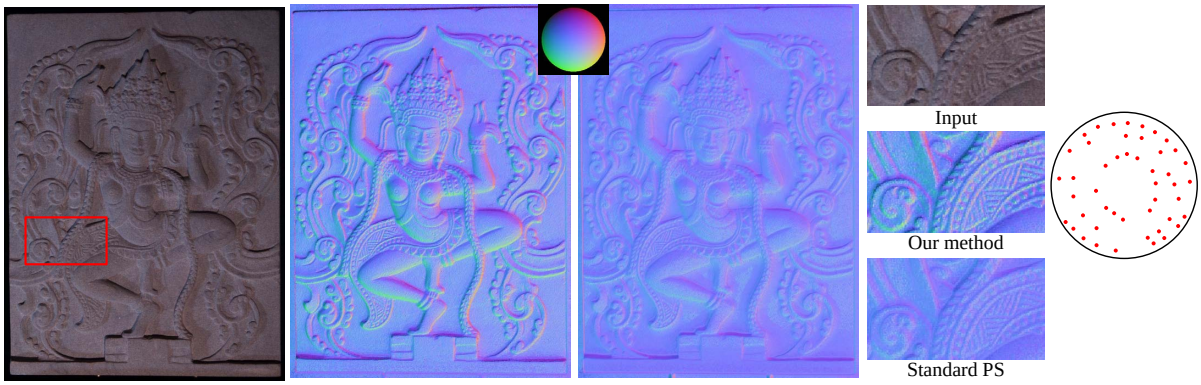
lobes. In addition, our method eliminates the necessity of radiometric calibration and any dependency on the ambient illumination.

Currently, our method is limited to work with surfaces that show either diffuse or specular reflection. To handle surfaces that have both diffuse and specular reflections, we are interested in applying a color subspace method [ZMKB08] for separating these reflections. We are also interested in using shadowed pixels. In our current method, we only use illuminated pixels for estimation; however, it has been shown in the previous work [OSS09] that shadow can be used as a cue for estimation.



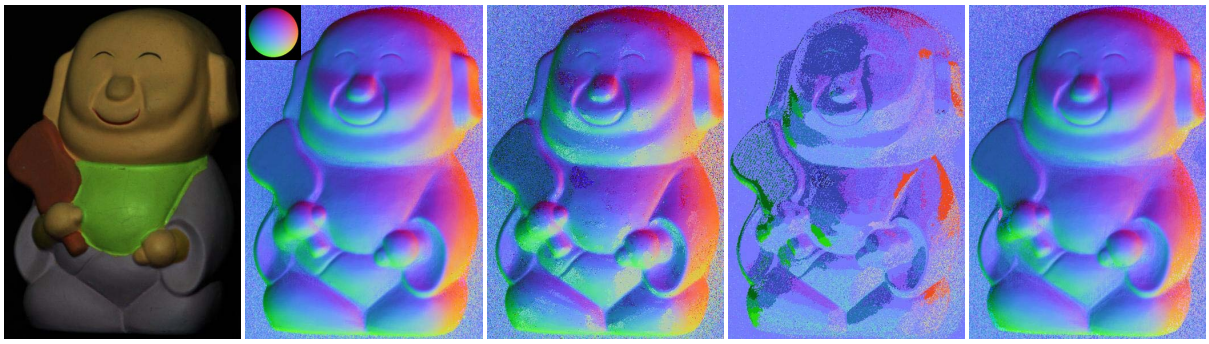
Input Our method Standard PS Light directions

Figure 5.12: Result of the statue scene recorded by a Sony XCD-X710CR camera with a linear response function under ambient illumination. From left to right, one of the input images, the estimated normal map with our method, and that of the standard photometric stereo method are shown. The reference sphere is overlaid in the middle of the second and third figures. The light source directions are shown on the right. 47 images are used as input.



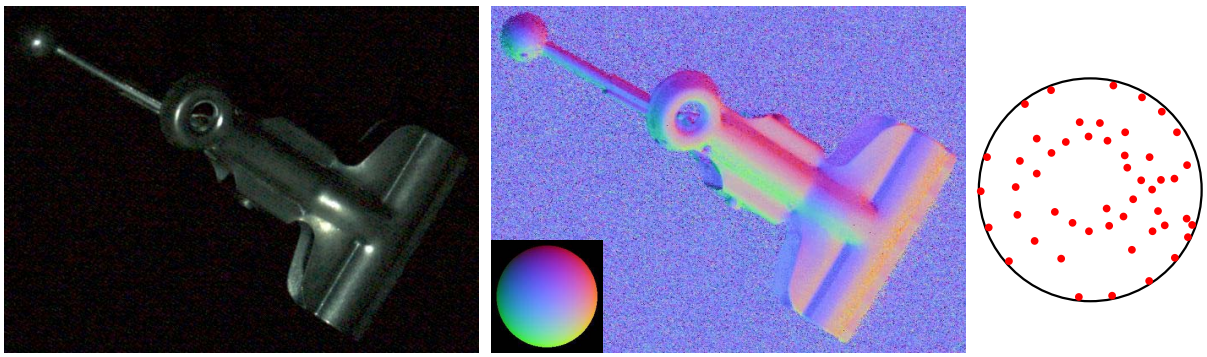
Input Our method Standard PS Close up Light directions

Figure 5.13: Result of the relief scene taken with a Nikon D1x camera with a non-linear response function under ambient illumination. From left to right, one of the input images, the estimated normal map with our method, that with standard photometric stereo method, and the light directions. 47 images are used as input.



Input | All the three constraints | Only monotonicity | Only visibility | Only isotropy

Figure 5.14: Results with all the three and individual constraints. Captions below the figures indicate the constraints that are used.



Input | Normal map | Light directions

Figure 5.15: Result of the clip scene captured with a Sony XCD-X710CR camera. From left to right, one of the input images, the estimated normal map, and the light directions are shown. 50 images are used as input.

Figure 5.16: 3D reconstruction and relighting results. Top figures show 3D reconstruction of the relief scene from top view and close-up side view. Bottom figures show relighting results of terracotta, statue, and clip scenes. The reference spheres show rendering parameters.

Chapter 6

Conclusion

6.1 Summary

The ultimate purpose of this dissertation is to provide efficient 3-D modeling approaches based on photometric stereo. To accomplish this, we have proposed two new methods for 3-D modeling by combining photometric and geometric approaches: a method that fuses a laser range sensor and a camera with attached camera flash, and a method based on a simple lighting configuration, *i.e.*, one LED point light source attached to a camera. Both methods estimate surface normals to efficiently recover 3-D models, especially fine details of the object surfaces. Moreover, for the purpose of handling specularities in these two methods, a real-time method that removes specular reflection components has been proposed. Furthermore, we have proposed a photometric stereo method that works with a wide range of surface reflectances.

6.1.1 Efficient Estimation and Representation of 3D model with Sensor Fusion

In Chapter 2, we proposed a method for 3-D modeling using a fusion of a laser range sensor, a camera, and a camera flash. The fusion provides dense normals and surface colors that can be mapped on a 3-D model and enables formulations to be made simply and practically. Multi-view photometric stereo was used for estimating the fine normal distribution with a basic shape measured by the laser range sensor. Our photometric stereo can easily handle near-light formulation and specularity. Detailed surfaces can

be shown by applying the normal map as "bump mapping" to the basic shape. Robust estimation and clustering were used for estimating reflection parameters. Our results show that our method could estimate highly accurate reflection parameters and provide fine surface appearances using only a small amount of data. The effectiveness and the practicality of our method were shown by an application that displayed 3-D contents.

6.1.2 A Hand-held Photometric Stereo Approach for Full 3-D Modeling

In Chapter 3, we presented a simple and practical 3-D modeling method that simultaneously estimates depth, surface normals, and reflectance from a set of images. We used a hand-held camera with an attached point light source to combine photometric stereo and multi-view stereo. We used photometric clues to get surface normals, which make it possible to find correspondences among multi-view images.

Moreover, using color information extended the method to be more efficient and robust for handling input images for full 3-D modeling. Efficient formulation is also applied to reduce the computational cost. Both simulation and real-world experiments showed the effectiveness and robustness of our method.

6.1.3 Real-time Specular Removal

In Chapter 4, we proposed a real-time method for specular removal using our color space while fitting a straight line. Our method was fast and practical with only one image taken under uniform illumination whose color was known. We also presented a process for making a specular-free image appropriate for an input image of photometric stereo. Theoretical propositions were proved, and experimental results showed the effectiveness of our method amidst changes of illumination color.

The current limitation of our method is that when more than one surface color is present in one hue, our method provides unexpected behavior. In such a case, we can still remove specular reflection components, but the color of diffuse reflections is different from the original input image.

6.1.4 Consensus Photometric Stereo for Non-Lambertian Surfaces

In Chapter 5, we presented a consensus approach for photometric stereo for a generalized reflectance model that satisfies three properties: monotonicity, visibility, and isotropy. These properties are naturally observed in a wide variety of diffuse reflections as well as in specular lobes. In addition, our method eliminates the necessity of radiometric calibration and any dependency on ambient illumination.

We implemented both voting and energy minimization approaches, and from the synthetic evaluation we showed that the energy minimization approach was faster and more accurate than the voting approach. Experimental evaluation was done, and it showed the effectiveness of our method.

6.2 Contributions

The main contributions of this dissertation are as follows:

- *Development of an efficient method for 3-D modeling using effective constraints from sensor fusion.*

  The advantage of the method is that the calibration among a laser range sensor, a camera, and a camera flash provides effective constraints for finding correspondences among multi-view images and for solving near light conditions in photometric stereo.

- *Accurate and robust estimation of reflection parameters and good appearance by using normal map.*

  Detailed surfaces are estimated as a normal map even if the resolution of the basic shape is quite low. Using the estimated normal map, reflection properties are accurately and robustly estimated with specular removal, clustering, and M-estimation.

- *Development of a simple and practical 3-D modeling method.*

  Our simple configuration is a camera and an attached point light source for 3-D modeling. Experimental results also showed that a commercial camera and an attached camera flash could work well in our method.

- *Full 3-D modeling with a hand-held camera.*

Input images for full 3-D reconstruction present the problem of many occlusions. We use color information and view constraints to overcome the problem.

- *Development of a real-time method for specular removal with a single image.*

  Using our color space lets us remove the specular reflection component very quickly and also generate a specular-free image from a single image taken under uniform illumination when the illumination color is known.

- *A new photometric stereo framework to work with a wide range of surface reflectances.*

  We introduced three basic reflection properties to derive constraints that specify solution spaces of the surface normal. Moreover, the method naturally avoids radiometric calibration and does not disturb ambient lighting.

- *Theoretical relationship between the number of input images and the expected accuracy of surface normal estimates.*

  We theoretically showed how many lighting directions are required with the monotonicity constraint, and conducted a simulation comparison between the estimation result and the theoretical error to verify the theory.

6.3 Future Directions

We conclude our discussion by mentioning several open problems and future improvements that we believe are important to pursue.

Interactive 3-D modeling with photometric stereo camera

A method for extending the hand-held photometric stereo camera so that it could reconstruct a target object in real-time. Owing to recent advances in Structure from Motion, such as *PTAM* [KM07], we are able to estimate a live camera pose and a sparse point cloud for 3-D reconstruction [ND10]. Adding a photometric constraint to the live reconstruction would achieve more dense and accurate results – ideally in real-time.

Surface normal estimation of both diffuse and specular surface using consensus approach

The consensus photometric stereo method described in Chapter 5 is limited to surfaces that show either diffuse or specular reflection. To handle surfaces that have both diffuse and specular reflections, we are interested in applying a color subspace method [ZMKB08] or our specular removal method described in Chapter 4.

Multi-view photometric stereo with a wide range of surface reflectance

Combining a hand-held photometric stereo camera described in Chapter 3 with the consensus photometric stereo algorithm presented in Chapter 5 should achieve robust and dense 3-D reconstructions for an object which displays varied surface reflectances. At this point, simultaneous estimation in order to find correspondences among multi-view images and for defining surface normals would be a challenging task to solve, but a task we believe to be important for further machine-vision technology innovation.

References

[AK07] Alldrin N., Kriegman D.: Toward reconstructing surfaces with arbitrary isotropic reflectance: a stratified photometric stereo approach. *Proc. of Int'l Conf. on Computer Vision* (2007).

[Aut] Autodesk, 3ds Max:. `http://usa.autodesk.com`.

[AX08] Aliaga D., Xu Y.: Photogeometric structured light: A self-calibrating and multi-viewpoint framework for accurate 3D modeling. *Proc. of Computer Vision and Pattern Recognition* (2008).

[AZK08] Alldrin N., Zickler T., Kriegman D.: Photometric stereo with non-parametric and spatially-varying reflectance. *Proc. of Computer Vision and Pattern Recognition* (2008).

[BBBH08] Bradley D., Boubekeur T., Berlin T., Heidrich W.: Accurate Multi-View Reconstruction Using Robust Binocular Stereo and Surface Meshing. *Proc. of Computer Vision and Pattern Recognition* (2008).

[BCSJ06] Birkbeck N., Cobzas D., Sturm P., Jagersand M.: Variational Shape and Reflectance Estimation under Changing Light and Viewpoints. *Proc. of European Conf. on Computer Vision* (2006).

[BK87] Boyer K., Kak A.: Color-encoded structured light for rapid active ranging. *IEEE Trans. on Pattern Analysis and Machine Intelligence 9*, 1 (1987), 14–28.

[BK98] Belhumeur P., Kriegman D.: What is the set of images of an object under all possible illumination conditions? *Int'l Journal of Computer Vision 28*, 3 (1998), 245–260.

[BK04] Boykov Y., Kolmogorov V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence 26*, 9 (2004), 1124–1137.

[BLL96] Bajcsy R., Lee S., Leonardis A.: Detection of diffuse and specular interface reflections and inter-reflections by color image segmentation. *Int'l Journal of Computer Vision 17*, 3 (1996), 241–272.

[BM92] Besl P., McKay N.: A method for registration of 3-D shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence* (1992), 239–256.

[Bou07] Bouguet J. Y.: *Camera calibration toolbox for matlab*. Tech. rep., 2007. `http://www.vision.caltech.edu/bouguetj/calib_doc/`.

[BP01] Barsky S., Petrou M.: Colour photometric stereo: simultaneous reconstruction of localgradient and colour of rough textured surfaces. *Proc. of Int'l Conf. on Computer Vision 2* (2001).

[BP03] Barsky S., Petrou M.: The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Trans. on Pattern Analysis and Machine Intelligence 25*, 10 (2003), 1239–1252.

[BVZ01] Boykov Y., Veksler O., Zabih R.: Fast approximate energy minimization via graph cuts. *IEEE Trans. on Pattern Analysis and Machine Intelligence 23*, 11 (2001), 1222–1239.

[BW88] Bates D. M., Watts D. G.: *Nonlinear Regression and Its Applications*. Wiley, New York, 1988.

[CAK07] Chandraker M., Agarwal S., Kriegman D.: ShadowCuts: Photometric Stereo with Shadows. *Proc. of Computer Vision and Pattern Recognition* (2007).

[CGS06] Chen T., Goesele M., Seidel H.: Mesostructure from Specularity. *Proc. of Computer Vision and Pattern Recognition 2* (2006), 1825–1832.

[CJ82] Coleman E. N., Jain R.: Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Computer Graphics and Image Processing 18* (1982), 309–328.

[CKK05] Chandraker M. K., Kahl F., Kriegman D. J.: Reflections on the generalized bas-relief ambiguity. In *Proc. of Computer Vision and Pattern Recognition* (2005), pp. 788–795.

[CL96] Curless B., Levoy M.: A volumetric method for building complex models from range images. *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (1996), 303–312.

[CP99] Clark J., Pekau H.: Integral formulation for differential photometric stereo. *Proc. of Computer Vision and Pattern Recognition 1* (1999), 119–124.

[DNRR05] Davis J., Nehab D., Ramamoorthi R., Rusinkiewicz S.: Spacetime stereo: a unifying framework for depth from triangulation. *IEEE Trans. on Pattern Analysis and Machine Intelligence 27*, 2 (2005), 296–302.

[FB81] Fischler M. A., Bolles R. C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM 24* (1981), 381–395.

[FCSS10] Furukawa Y., Curless B., Seitz S., Szeliski R.: Towards internet-scale multi-view stereo. *Proc. of Computer Vision and Pattern Recognition* (2010).

[FP07] Furukawa Y., Ponce J.: Accurate, dense, and robust multiview stereopsis. *Proc. of Computer Vision and Pattern Recognition* (2007).

[GCHS05] Goldman D., Curless B., Hertzmann A., Seitz S.: Shape and spatially-varying BRDFs from photometric stereo. *Proc. of Int'l Conf. on Computer Vision 1* (2005), 341–348.

[GCS06] Goesele M., Curless B., Seitz S.: Multi-view stereo revisited. *Proc. of Computer Vision and Pattern Recognition 2* (2006), 2402–2409.

[Geo03] Georghiades A.: Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo. *Proc. of Int'l Conf. on Computer Vision* (2003), 816–823.

[GH97] Garland M., Heckbert P.: Surface simplification using quadric error metrics. *Proc. of ACM SIGGRAPH 97*, 31 (1997), 209–216.

[Goo] Google SketchUp:. `http://sketchup.google.com`.

[GSC*07] Goesele M., Snavely N., Curless B., Hoppe H., Seitz S.: Multi-View Stereo for Community Photo Collections. *Proc. of Int'l Conf. on Computer Vision* (2007).

[Hay94] Hayakawa H.: Photometric stereo under a light source with arbitrary motion. *Journal of the Optical Society of America A 11*, 11 (1994), 3079–3089.

[HH03] Huber D., Hebert M.: 3d modeling using a statistical sensor model and stochastic search.

[HHR02] Hall-Holt O., Rusinkiewicz S.: Stripe boundary codes for real-time structured-light range scanning of moving objects. *Proc. of Int'l Conf. on Computer Vision 2* (2002), 359–366.

[HI84] Horn B., Ikeuchi K.: The mechanical manipulation of randomly oriented parts. *Scientific American 251*, 2 (1984), 100–111.

[HK06] Hornung A., Kobbelt L.: Hierarchical volumetric multi-view stereo reconstruction of manifold surfaces based on dual graph embedding. *Proc. of Computer Vision and Pattern Recognition 1* (2006), 503–510.

[HLHZ08] Holroyd M., Lawrence J., Humphreys G., Zickler T.: A photometric approach for estimating normals and tangents. *Int'l Conference on Computer Graphics and Interactive Techniques* (2008).

[HMJI09] Higo T., Matsushita Y., Joshi N., Ikeuchi K.: A hand-held photometric stereo camera for 3-d modeling. *Proc. of Int'l Conf. on Computer Vision* (2009).

[HS05] Hertzmann A., Seitz S.: Example-based photometric stereo: shape reconstruction with general, varying BRDFs. *IEEE Trans. on Pattern Analysis and Machine Intelligence 27*, 8 (2005), 1254–1264.

[HVB*07] Hernandez C., Vogiatzis G., Brostow G., Stenger B., Cipolla R.: Non-rigid Photometric Stereo with Colored Lights. *Proc. of Int'l Conf. on Computer Vision* (2007).

[HVC08a] Hernández C., Vogiatzis G., Cipolla R.: Multiview Photometric Stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence 30*, 3 (2008), 548–554.

[HVC08b] Hernández C., Vogiatzis G., Cipolla R.: Shadows in three-source photometric stereo. *Proc. of European Conf. on Computer Vision* (2008).

[IHN*04] Ikeuchi K., Hasegawa K., Nakazawa A., Takamatsu J., Oishi T., Masuda T.: Bayon Digital Archival Project. *Proc. of the Tenth Int'l Conference on Virtual System and Multimedia* (2004), 334–343.

[Ike81] Ikeuchi K.: Determining Surface Orientations of Specular Surfaces by Using the Photometric Stereo Method. *IEEE Trans. on Pattern Analysis and Machine Intelligence 3*, 6 (1981), 661–669.

[Ike87] Ikeuchi K.: Determining a depth map using a dual photometric stereo. *The Int'l Journal of Robotics Research 6*, 1 (1987), 15–31.

[INHO03] Ikeuchi K., Nakazawa A., Hasegawa K., Ohishi T.: The Great Buddha Project: Modeling Cultural Heritage for VR Systems through Observation. *Proc. of the The 2nd IEEE and ACM Int'l Symposium on Mixed and Augmented Reality* (2003).

[IWTI94] Iwahori Y., Woodham R., Tanaka H., Ishii N.: Moving Point Light Source Photometric Stereo. *IEICE Trans. on Information and Systems 77*, 8 (1994), 925–929.

[JK07] Joshi N., Kriegman D.: Shape from Varying Illumination and Viewpoint. *Proc. of Int'l Conf. on Computer Vision* (2007).

[KBH06] Kazhdan M., Bolitho M., Hoppe H.: Poisson surface reconstruction. *ACM Int'l Conference Proceeding Series; Proc. of the fourth Eurographics symposium on Geometry processing 256* (2006).

[KFSY08] Kawasaki H., Furukawa R., Sagawa R., Yagi Y.: Dynamic scene shape reconstruction using a single structured light pattern. *Proc. of Computer Vision and Pattern Recognition* (2008).

[KM07] Klein G., Murray D.: Parallel tracking and mapping for small AR workspaces. *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality* (2007).

[KPC10] Kolev K., Pock T., Cremers D.: Anisotropic minimal surfaces integrating photoconsistency and normal information for multiview stereo. *Proc. of European Conf. on Computer Vision* (2010).

[KS00] Kutulakos K., Seitz S.: A Theory of Shape by Space Carving. *Int'l Journal of Computer Vision 38*, 3 (2000), 199–218.

[KSK88] Klinker G., Shafer S., Kanade T.: The measurement of highlights in color images. *Int'l Journal of Computer Vision 2*, 1 (1988), 7–32.

[KVG06] Koninckx T., Van Gool L.: Real-Time Range Acquisition by Adaptive Structured Light. *IEEE Trans. on Pattern Analysis and Machine Intelligence 28*, 3 (2006).

[KWBE10] Kim H., Wilburn B., Ben-Ezra M.: Photometric stereo for dynamic surface orientations. *Proc. of European Conf. on Computer Vision* (2010).

[KZ04] Kolmogorov V., Zabih R.: What Energy Functions Can Be Minimized via Graph Cuts? *IEEE Trans. on Pattern Analysis and Machine Intelligence* (2004), 147–159.

[LCDX09] Liu Y., Cao X., Dai Q., Xu W.: Continuous depth estimation for multi-view stereo. *Proc. of Computer Vision and Pattern Recognition* (2009), 2121–2128.

[LHYK05] Lim J., Ho J., Yang M., Kriegman D.: Passive Photometric Stereo from Motion. *Proc. of Int'l Conf. on Computer Vision 2* (2005), 1635–1642.

[LKG*03] Lensch H., Kautz J., Goesele M., Heidrich W., Seidel H.: Image-based reconstruction of spatial appearance and geometric detail. *ACM Trans. on Graphics (TOG) 22*, 2 (2003), 234–257.

[LL99] Lu J., Little J.: Reflectance and shape from images using a collinear light source. *Int'l Journal of Computer Vision 32*, 3 (1999), 213–240.

[LLC*10] Li J., Li E., Chen Y., Xu L., Zhang Y.: Bundled depth-map merging for multi-view stereo. *Proc. of Computer Vision and Pattern Recognition* (2010).

[LPC*00] Levoy M., Pulli K., Curless B., Rusinkiewicz S., Koller D., Pereira L., Ginzton M., Anderson S., Davis J., Ginsberg J., et al.: The Digital Michelangelo Project: 3D Scanning of Large Statues. *Proc. of the 27th annual conference on Computer graphics and interactive techniques* (2000), 131–144.

[LPK07] Labatut P., Pons J., Keriven R.: Efficient Multi-View Reconstruction of Large-Scale Scenes using Interest Points, Delaunay Triangulation and Graph Cuts. *Proc. of Int'l Conf. on Computer Vision* (2007).

[LQ05] Lhuillier M., Quan L.: A quasi-dense approach to surface reconstruction from uncalibrated images. *IEEE Trans. on Pattern Analysis and Machine Intelligence 27*, 3 (2005), 418–433.

[LS01] Lin S., Shum H.: Separation of diffuse and specular reflection in color images. *Proc. of Computer Vision and Pattern Recognition 1* (2001), 341–346.

[Mid] Middlebury:. `http://vision.middlebury.edu/mview/`.

[MK09] Micusik B., Kosecka J.: Piecewise planar city 3D modeling from street view panoramic sequences. *Proc. of Computer Vision and Pattern Recognition* (2009), 2906–2912.

[MKH*06] Miyazaki D., Kamakura M., Higo T., Okamoto Y., Kawakami R., Shiratori T., Ikari A., Ono S., Sato Y., Oya M.: 3D Digital Archive of the Burghers of Calais. *12th Int'l Conference on Virtual Systems and Multimedia(VSMM2006)* (10 2006).

[MTHI03] Miyazaki D., Tan R., Hara K., Ikeuchi K.: Polarization-based inverse rendering from a single view. *Proc. of Int'l Conf. on Computer Vision* (2003), 982–987.

[MWGA06] Malzbender T., Wilburn B., Gelb D., Ambrisco B.: Surface enhancement using real-time photometric stereo and reflectance transformation. *Proc. of EGSR* (2006).

[MWW02] Maki A., Watanabe M., Wiles C.: Geotensity: Combining Motion and Lighting for 3D Surface Reconstruction. *Int'l Journal of Computer Vision 48*, 2 (2002), 75–90.

[MZKB05] Mallick S., Zickler T., Kriegman D., Belhumeur P.: Beyond Lambert: Reconstructing Specular Surfaces Using Color. *Proc. of Computer Vision and Pattern Recognition 2* (2005), 619–626.

[ND10] Newcombe R., Davison A.: Live dense reconstruction with a single moving camera. *Proc. of Computer Vision and Pattern Recognition* (2010).

[NFB97] Nayar S., Fang X., Boult T.: Separation of Reflection Components Using Color and Polarization. *Int'l Journal of Computer Vision 21*, 3 (1997), 163–186.

[NIK90] Nayar S., Ikeuchi K., Kanade T.: Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE Trans. on Robotics and Automation 6*, 4 (1990), 418–431.

[NRDR05] Nehab D., Rusinkiewicz S., Davis J., Ramamoorthi R.: Efficiently combining positions and normals for precise 3D geometry. *Proc. of ACM SIGGRAPH 24*, 3 (2005), 536–543.

[OD97] Okatani T., Deguchi K.: Shape Reconstruction from an Endoscope Image by Shape from Shading Technique for a Point Light Source at the Projection Center. *Computer Vision and Image Understanding 66*, 2 (1997), 119–131.

[OK93] Okutomi M., Kanade T.: A multiple-baseline stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence 15*, 4 (1993), 353–363.

[ON95] Oren M., Nayar S.: Generalization of the Lambertian model and implications for machine vision. *Int'l Journal of Computer Vision 14*, 3 (1995), 227–251.

[OSS09] Okabe T., Sato I., Sato Y.: Attached shadow coding: estimating surface normals from shadows under unknown reflectance and lighting conditions . *Proc. of Int'l Conf. on Computer Vision* (2009).

[PNF*08] Pollefeys M., Nistér D., Frahm J., Akbarzadeh A., Mordohai P., Clipp B., Engels C., Gallup D., Kim S., Merrell P., et al.: Detailed real-time urban 3d reconstruction from video. *Int'l Journal of Computer Vision 78*, 2 (2008), 143–167.

[PS82] Paige C., Saunders M.: LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares. *ACM Trans. on Mathematical Software (TOMS) 8*, 1 (1982), 43–71.

[PVGV*04] Pollefeys M., Van Gool L., Vergauwen M., Verbiest F., Cornelis K., Tops J., Koch R.: Visual Modeling with a Hand-Held Camera. *Int'l Journal of Computer Vision 59*, 3 (2004), 207–232.

[Rei73] Reichman J.: Determination of absorption and scattering coefficients for nonhomogeneous media. 1: Theory. *Applied Optics 12*, 8 (1973), 1811–1815.

[SBM98] Salvi J., Batlle J., Mouaddib E.: A robust-coded pattern projection for dynamic 3d scene measurement. *Pattern Recognition Letters 19*, 11 (1998), 1055–1065.

[SCD*06] Seitz S., Curless B., Diebel J., Scharstein D., Szeliski R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. *Proc. of Computer Vision and Pattern Recognition 1* (June 2006), 519–526.

[SD99] Seitz S., Dyer C.: Photorealistic Scene Reconstruction by Voxel Coloring. *Int'l Journal of Computer Vision 35*, 2 (1999), 151–173.

[SFB03] Simakov D., Frolova D., Basri R.: Dense shape reconstruction of a moving object under arbitrary, unknown lighting. In *Proc. of Int'l Conf. on Computer Vision* (2003), pp. 1202–1209.

[SH10] Smith W., Hancock E.: Estimating Facial Reflectance Properties Using Shape-from-Shading. *Int'l Journal of Computer Vision 86* (2010), 152–170.

[Sha85] Shafer S.: Using Color to Separate Reflection Components. *COLOR Research and Application 10*, 4 (1985), 210–218.

[SI94] Sato Y., Ikeuchi K.: Temporal-color space analysis of reflection. *Journal of the Optical Society of America A 11*, 11 (1994), 2990–3002.

[SI96] Solomon F., Ikeuchi K.: Extracting the shape and roughness of specular lobe objects using four light photometric stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence 18*, 4 (1996), 449–454.

[Sil80] Silver W.: Determining shape and reflectance using multiple images. *Master's thesis, MIT* (1980).

[SKS*02] Swaminathan R., Kang S., Szeliski R., Criminisi A., Nayar S.: On the motion and appearance of specularities in image sequences. *Proc. of European Conf. on Computer Vision 1* (2002), 508–523.

[SMP07] Sinha S., Mordohai P., Pollefeys M.: Multi-View Stereo via Graph Cuts on the Dual of an Adaptive Tetrahedral Mesh. *Proc. of Int'l Conf. on Computer Vision* (2007).

[SMW*10] Shi B., Matsushita Y., Wei Y., Xu C., Tan P.: Self-calibrating photometric stereo. *Proc. of Computer Vision and Pattern Recognition* (2010).

[SOYS07] Sato I., Okabe T., Yu Q., Sato Y.: Shape Reconstruction Based on Similarity in Radiance Changes under Varying Illumination. *Proc. of Int'l Conf. on Computer Vision* (2007).

[SSS06] Snavely N., Seitz S. M., Szeliski R.: Photo tourism: Exploring photo collections in 3d. *Proc. of ACM SIGGRAPH* (2006), 835–846. `http://phototour.cs.washington.edu/bundler/`.

[SSS07] Snavely N., Seitz S., Szeliski R.: Modeling the World from Internet Photo Collections. *Int'l Journal of Computer Vision* (2007).

[SW93] Schlüns K., Wittig O.: Photometric stereo for non-lambertian surfaces using color information. *Proc. 7th Int. Conf. on Image Analysis and Processing* (1993), 505–512.

[SWI97] Sato Y., Wheeler M., Ikeuchi K.: Object shape and reflectance modeling from observation. *Proc. of the 24th annual conference on Computer graphics and interactive techniques* (1997), 379–387.

[SZP10] Sunkavalli K., Zickler T., Pfister H.: Visibility Subspaces: Uncalibrated Photometric Stereo with Shadows. *Proc. of European Conf. on Computer Vision* (2010), 251–264.

[TdF91] Tagare H., de Figueiredo R.: A theory of photometric stereo for a class of diffuse Non-Lambertian surfaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence 13*, 2 (1991), 133–152.

[TI05] Tan R., Ikeuchi K.: Separating reflection components of textured surfaces using a single image. *IEEE Trans. on Pattern Analysis and Machine Intelligence 27*, 2 (2005), 178–193.

[TLQ08] Tan P., Lin S., Quan L.: Subpixel photometric stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence 30*, 8 (2008), 1460–1471.

[TLQS03] Tan P., Lin S., Quan L., Shum H.: Highlight Removal by Illumination-Constrained Inpainting. *Proc. of Int'l Conf. on Computer Vision 1* (2003), 164–169.

[TS67] Torrance K., Sparrow E.: Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America 57*, 9 (1967), 1105–1114.

[VHTC07] Vogiatzis G., Hernández C., Torr P., Cipolla R.: Multiview Stereo via Volumetric Graph-Cuts and Occlusion Robust Photo-Consistency. *IEEE Trans. on Pattern Analysis and Machine Intelligence* (2007), 2241–2246.

[VVG08] Verbiest F., Van Gool L.: Photometric stereo with coherent outlier handling and confidence estimation. *Proc. of Computer Vision and Pattern Recognition* (2008).

[WCL*08] Wu C., Clipp B., Li X., Frahm J., Pollefeys M.: 3D Model Matching with Viewpoint-Invariant Patches (VIP). *Proc. of Computer Vision and Pattern Recognition* (2008).

[WMP*06] Weyrich T., Matusik W., Pfister H., Bickel B., Donner C., Tu C., McAndless J., Lee J., Ngan A., Jensen H. W., Gross M.: Analysis of human faces using a measurement-based skin reflectance model. *ACM Trans. Graph. 25*, 3 (2006), 1013–1024.

[Wol89] Wolff L.: Using polarization to separate reflection components. *Proc. of Computer Vision and Pattern Recognition* (1989), 363–369.

[Wol94] Wolff L.: Diffuse-reflectance model for smooth dielectric surfaces. *Journal of the Optical Society of America A 11*, 11 (1994), 2956–2968.

[Woo80] Woodham R.: Photometric method for determining surface orientation from multiple images. *Optical Engineering 19*, 1 (1980), 139–144.

[WTTW06] Wu T., Tang K., Tang C., Wong T.: Dense Photometric Stereo: A Markov Random Field Approach. *IEEE Trans. on Pattern Analysis and Machine Intelligence 28*, 11 (2006), 1830–1846.

[WYJT10] Wu T., Yeung S., Jia J., Tang C.: Quasi-dense 3D reconstruction using tensor-based multiview stereo. *Proc. of Computer Vision and Pattern Recognition* (2010).

[YBD*07] Young M., Beeson E., Davis J., Rusinkiewicz S., Ramamoorthi R.: Viewpoint-coded structured light. *Proc. of Computer Vision and Pattern Recognition* (2007).

[YX10] Yamazaki M., Xu G.: 3D reconstruction of glossy surfaces using stereo cameras and projector-display. *Proc. of Computer Vision and Pattern Recognition* (2010).

[ZBK02] Zickler T., Belhumeur P., Kriegman D.: Helmholtz stereopsis: exploiting reciprocity for surface reconstruction. *Int'l Journal of Computer Vision 49*, 2 (2002), 215–227.

[ZCHS03] Zhang L., Curless B., Hertzmann A., Seitz S.: Shape and motion under varying illumination: unifying structure from motion, photometric stereo, and multiview stereo. *Proc. of Int'l Conf. on Computer Vision* (2003), 618–625.

[ZKU*04] Zitnick C. L., Kang S. B., Uyttendaele M., Winder S., Szeliski R.: High-quality video view interpolation using a layered representation. *ACM Trans. Graph. 23*, 3 (2004), 600–608.

[ZMKB08] Zickler T., Mallick S., Kriegman D., Belhumeur P.: Color subspaces as photometric invariants. *Int'l Journal of Computer Vision 79*, 1 (2008), 13–30.

List of Publications

Journal Papers

1. 肥後智昭, 宮崎大輔，池内克史，“センサフュージョンによる効率的な 3 次元モデルの推定と表現”，映像情報メディア学会誌，Vol.64, No.1 , January. 2010.

International Conferences

1. Tomoaki Higo, Yasuyuki Matsushita, Katsushi Ikeuchi, “Consensus Photometric Stereo,” *Proceedings of Computer Vision and Pattern Recognition (CVPR2010)*, San Francisco, June, 2010.

2. Tomoaki Higo, Yasuyuki Matsushita, Neel Joshi, Katsushi Ikeuchi, “A Hand-held Photometric Stereo Camera for 3-D Modeling,” *Proceedings of International Conference on Computer Vision(ICCV2009)*, Kyoto, September, 2009.

3. Daisuke Miyazaki, Mawo Kamakura, Tomoaki Higo, Yasuhide Okamoto, Rei Kawakami, Takaaki Shiratori, Akifumi Ikari, Shintaro Ono, Yoshihiro Sato, Mina Oya, Masayuki Tanaka, Katsushi Ikeuchi, Masanori Aoyagi, “3D Digital Archive of the Burghers of Calais,” *In the Proceedings of the International Conference on Virtual Systems and Multimedia (VSMM2006),* Lecture Notes in Computer Science (LNCS), October, 2006.

Domestic Magazines

1. 肥後智昭，宮崎大輔，池内克史，“物体の全体形状と反射パラメータの同時推定”, 画像ラボ，Vol.18No.11，日本工業出版，pp. 31-36，2007 年 11 月.

Domestic Conferences

1. 肥後智昭，松下康之，池内克史，“非ランバート拡散反射に対する照度差ステレオ”，画像の認識・理解シンポジウム (MIRU2010)，釧路，2010 年 7 月. **優秀論文賞受賞**

2. 肥後智昭，宮崎大輔，池内克史，“疎な形状を利用した多視点照度差ステレオ法”，画像の認識・理解シンポジウム (MIRU2008)，軽井沢，2008 年 7 月.

3. 肥後智昭，宮崎大輔，池内克史，“陰影からの形状と反射パラメータの同時推定”，画像の認識・理解シンポジウム (MIRU2007)，広島，2007 年 7 月．

Workshops

1. Tomoaki Higo, Yasuyuki Matsushita, Neel Joshi, Katsushi Ikeuchi, “A Hand-held Photometric Stereo Camera for 3-D Modeling,” *5th international joint workshop of KAIST/UT on robust vision technology*, Tokyo, December, 2009.

2. 池内克史，小野晋太郎，高松淳，影沢政隆，森本哲郎，川上玲，鎌倉真音，肥後智昭，大蔵苑子，“火山被災遺跡における形状と色彩再現：ソンマヴェスビアーナとポンペイを例として”，火山噴火罹災地の文化・自然環境復元，ソンマヴェスビアーナ，指宿，浅間，戦略的学融合研究 2007，東京，2008 年 2 月．

3. Tomoaki Higo, Daisuke Miyazaki, Katsushi Ikeuchi, “3D Modeling: Laser range sensor + Photometric stereo,” *3rd international joint workshop of KAIST/UT on robust vision technology*, Tokyo, November, 2007.

4. 肥後智昭，宮崎大輔，池内克史，“多視点フォトメトリックステレオを用いた全体形状と反射パラメータの同時推定”，三次元映像のフォーラム 第 80 回研究会，東京，2007 年 6 月．

5. 肥後智昭，宮崎大輔，池内克史，“二色性反射モデルに基づくリアルタイム鏡面反射成分除去”，情報処理学会コンピュータビジョンとイメージメディア研究会 (CVIM)，福岡，2006 年 9 月．

Awards

1. 画像の認識・理解シンポジウム (MIRU2010) 優秀論文賞 2010 年 7 月．

2. Microsoft Research Asia Fellowship Award, September, 2008.